Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

3-10-2016 12:00 AM

Validation Strategies Supporting Clinical Integration of Prostate Segmentation Algorithms for Magnetic Resonance Imaging

Maysam Shahedi Bagh Khandan
The University of Western Ontario

Supervisor
Aaron Fenster
The University of Western Ontario Joint Supervisor
Aaron D. Ward
The University of Western Ontario

Graduate Program in Biomedical Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Maysam Shahedi Bagh Khandan 2016

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biomedical Engineering and Bioengineering Commons

Recommended Citation

Shahedi Bagh Khandan, Maysam, "Validation Strategies Supporting Clinical Integration of Prostate Segmentation Algorithms for Magnetic Resonance Imaging" (2016). *Electronic Thesis and Dissertation Repository*. 3669.
https://ir.lib.uwo.ca/etd/3669

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

المنارة للاستشارات

www.manaraa.com

Abstract

Three-dimensional (3D) segmentation of the prostate in medical images is useful for prostate cancer diagnosis and therapy guidance. However, manual segmentation of the prostate is laborious and time-consuming with inter-observer variability at the prostatic apex, mid-gland, and base. The focus of this thesis was on: (1) accuracy, reproducibility and procedure time measurement for prostate segmentation on T2-weighted endorectal (ER) magnetic resonance (MR) imaging (MRI), and (2) assessment of the potential of a computer-assisted segmentation technique to be translated to clinical practice for prostate cancer management.

We collected 42 ER MR images from patients with biopsy-confirmed prostate cancer. Prostate border delineation was manually performed by one observer on all the images and by two other observers on a subset of 10 images.

We developed a novel semi-automatic and automatic prostate segmentation algorithms that identify candidate prostate boundary points using learned local prostate border appearance characteristics, which were regularized according to learned prostate shape information to produce the final segmentation. The main novelties of the algorithms are that the segmentations was based on local appearance similarity of the prostate border across patients, rather than on the appearance of the entire image. This makes the appearance-based segmentation more robust to inter-patient variation of prostate gland internal appearance that could be caused by variable spatial distribution of cancer in the patients. We evaluated our method against the manual reference segmentations using mean absolute distance (MAD), Dice similarity coefficient (DSC), recall rate, precision rate, and volume difference as a complementary boundary-, region-

المنارة للاستشارات

www.manaraa.com

and volume-based error metric set to measure the different types of observed segmentation errors. We applied this evaluation for expert manual segmentation as well as semi-automatic and automatic segmentation approaches before and after manual editing by expert physicians. Physicians were instructed to edit the segmentations to their satisfaction for use in clinical procedures, as would be done with any prostate segmentation technique integrated into the clinical workflow. We recorded the time needed for user interaction to initialize the semi-automatic algorithm, algorithm execution, and manual editing where applicable.

On 42 images, comparing to a single-observer manual segmentation reference, the measured errors for semi-automatic and automatic algorithm on whole prostate gland were, respectively, MAD of 2.0 mm and 3.2 mm; DSC of 82% and 71%; recall of 77% and 69%; precision of 88% and 76%; and ΔV of -4.6 cm^3 and -3.6 cm^3. These results compared favourably with observed difference between manual segmentation and a simultaneous truth and performance level estimation (STAPLE) reference for a subset of 10 images (whole gland differences as high as MAD = 3.1 mm, DSC = 78%, recall = 66%, precision = 77%, and ΔV = 15.5 cm^3). For each 3D image, the semi-automatic algorithm required about 30 seconds, on average, to be initialized. Using an unoptimized Matlab research platform on a single CPU core, the average execution times for semi-automatic and automatic algorithm were 85 seconds and 54 seconds, respectively, to segment a prostate MRI in 3D. We also measured average editing times of 330 and 390 seconds for the semi-automatic and automatic segmentation results, respectively, whereas an expert spent 210 seconds on average editing manual segmentations performed by another expert. Inter-operator variability resulting from using our computer-assisted

www.manaraa.com

algorithms to generate starting segmentations for manual editing was not substantially higher than that resulting from using expert manual segmentations as starting segmentations, suggesting a role for our (semi-)automated segmentation algorithm in this context.

The presented algorithms used learned local appearance characteristics and prostate shape separately to segment the prostate and regularize the segmentation, respectively. The semi-automatic algorithm needed minimal user interaction of approximately 30 seconds to be initialized; this was replaced by about 3 seconds of computational time using the automatic segmentation. Both algorithms are highly parallelizable.

The main conclusions of this thesis were that: (1) computer-assisted segmentation approaches reduced the inter-observer segmentation variability compared to manual segmentation, (2) the accuracy of the computer-assisted approaches was near to or within the range of observed variability in manual prostate segmentation performed by experts, (3) manual editing of semi-automated and automated segmentation approaches improved the accuracy and inter-operator variability, (4) the recorded procedure time for prostate segmentation was reduced using computer-assisted segmentation approaches followed by manual editing compared to fully manual segmentation, and (5) starting the manual segmentation from an initial computer-assisted segmentation label could yield lower variability in the final segmentations and the choice of automatic vs. semi-automatic segmentation comes down to operator preference.

المنارة للاستشارات
www.manaraa.com

Keywords

Magnetic resonance imaging, MRI, prostate cancer, segmentation, validation, T2-weighted, endorectal coil.

www.manaraa.com

Co-Authorship Statement

This thesis is presented in an integrated article format, the chapters of which are based on the following publications that are either published or in preparation for submission:

Chapter 2: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, E. Gibson, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. D. Ward, "Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods," Medical Physics 41 (11), 113503 (2014).

My contributions to this work included defining the research questions; designing and implementing the segmentation method; data preparation; designing, implementing and analyzing the experiments; operating the segmentation algorithm; and drafting the manuscript. D. W. Cool, C. Romagnoli, G. S. Bauman, and M. Bastian-Jordan contributed to collectiion and preparation of the data set, manual segmentation of the MR images, and operating the semi-automatic segmentation algorithm. E. Gibson, G. Rodrigues, B. Ahmad, and M. Lock contributed to operating the semi-automatic segmentation algorithm. A. D. Ward contributed to defining the research questions, designing and analyizing the experiments, operating the semi-automatic segmentation algorithm, interpreting the results, and drafting the manuscript. A. Fenster motivated the initial research direction. All authors helped in reviewing and editing the manuscript. The work was performed under supervision of A. D. Ward and A. Fenster.

Chapter 3: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, A. Fenster, and A. D. Ward, "Accuracy and acceptability of an automated method for prostate segmentation in magnetic resonance imaging.".

www.manaraa.com

My contributions to this work included defining the research questions; designing and implementing the segmentation method; data preparation; designing, implementing and analyzing the experiments; and drafting the manuscript. D. W. Cool, C. Romagnoli, G. S. Bauman, and M. Bastian-Jordan contributed to collection and preparation of the data set and manual segmentation of the MR images. A. D. Ward contributed to defining the research questions, designing and analyizing the experiments, interpreting the results, and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under supervision of A. D. Ward and A. Fenster.

Chapter 4: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. D. Ward, "Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging.".

My contributions to this work included defining the research questions; data preparation; designing, implementing and analyzing the experiments; and drafting the manuscript. D. W. Cool and G. S. Bauman contributed to collecting and preparation of the data set, manual segmentation of the MRI images, and editing the segmentations. C. Romagnoli and M. Bastian-Jordan contributed to collection and preparation of the data set and manual segmentation of the MR images. G. Rodrigues, B. Ahmad, and M. Lock contributed to editing the segmentations. A. D. Ward contributed to defining the research questions, designing and analyizing the experiments, interpreting the results, and drafting the manuscript. All authors helped in reviewing and editing the manuscript. The work was performed under supervision of A. D. Ward and A. Fenster.

المنارة للاستشارات
www.manaraa.com

Dedication

To my mother and father,

and to all those who have taught me throughout my life's journey

www.manaraa.com

Acknowledgements

The support, participation and assistance of so many people were contributory to the completion of the work presented in this thesis.

First, I would like to thank Prof. Aaron Fenster for giving me this opportunity to work in his laboratory. I acknowledge his supervision, guidance and support throughout my graduate studies. He taught me a lot by asking critical questions that showed me hidden aspects of problems and led me to better solutions. I also learned a lot from him by watching him managing the projects and the research group.

I would also like to thank Prof. Aaron Ward who was not only a supervisor to me but also a wise mentor, a great teacher, and a nice friend. It was really a good opportunity for me to join his lab and work with him closely. He deeply cares about his students and always loves to teach them something new. I can definitely say that in all of the meetings with him, even the informal and non-scientific ones, I learned something valuable from him. It was a great chance and an absolute pleasure for me to work in his group and I cannot forget what he has done for me throughout the years of my PhD program.

I would like to thank my advisors Prof. Eugene Wong and the late Prof. Cesare Romagnoli for their advice, guidance and support. They provided me with invaluable insights and suggestions that improved my knowledge and elevated my understanding of the problems in the technical and clinical domains. Prof. Romagnoli's passing in the last year of my PhD program was a big loss for me. Even during his illness, he was always willing to help and spent a lot of time assisting with data preparation and answering my questions. May he rest in peace.

المنارة للاستشارات
www.manaraa.com

I am hugely indebted to Dr. Derek Cool for his unconditional help and support at each step of the project. He spent hours reviewing my segmentation labels and editing them. His experience and knowledge in both technical and clinical aspects made him an incredible collaborator on this project.

This project gave me the opportunity to collaborate with an excellent group of clinicians: Drs. Glenn Bauman, George Rodrigues, Michael Lock, Belal Ahmad and Matthew Bastian-Jordan. I would like to thank all of them for their assistance and invaluable advice. I would also like to thank Dr. David Palma who always seemed more than happy to help and answer my questions related to the clinical side of the project.

I would also thank all my colleagues at Baines Imaging Laboratory and Robarts Research Institute for their help and support, especially Dr. Eli Gibson, Dr. Tharindu De Silva, Hamid Sadeghi Neshat and Dr. Ali Tavallaei, Sarah Mattonen, Carol Johnson, Yiwen Xu, Peter Martin, Andrew Warner, Mehrnoush Salarian, Wenchao Han, Dr. Timothy Yeung, Derek Soetemans, Sachi Elkerton, Dr. David Tessier, Dr. Chandima Edirisinghe, Dr. Mohammad Kayvanrad, Dr. Lena Gorelick, Dr. Eranga Ukwatta and so many other colleagues in different labs at Western University that I cannot enumerate them all.

My sincere gratitude also goes to Ashley Kewayosh Samuel, the Internationalization Programming Coordinator of The International and Exchange Student Centre, for her unconditional support and encouragement from the very beginning of my trip to Canada and also to Maria Jardine, housekeeping staff at the Baines Imaging Laboratory, who always reminded and encouraged me to look at life positively.

المنارة للاستشارات
www.manaraa.com

I am grateful to my friends in London who helped and supported me during recent years. Special thanks to Omid Farhangi, Arman Abtahi, Sahar Samimi, Naeeme Modir, Elmira Rezanavaz, Maryam Mojalal, Vahid Dehnavi, Nasrin Farhangi, Malihe Mehdizadeh, Maysam Mirahmadi, Ehsan Omidi, Pooriya Khorshidi, Lida Radan, Amir Owrangi and Samaneh Kazemi. I would also thank my friends all over the world for their supports, especially Alireza Aghasi, Omeed Momeni, Hasan Mashayekh, Afshin Safari, Amir Mir Mohammad Sadeghi, Ehsan Mahmoodi, Mohammad Ansari, Hoda Moazzen, Behrouz Golzarpour, Samaneh Shadmehr, Mahdi Saniei and Bahare Aminian.

Finally, I would like to wholeheartedly thank my parents, Akram Panahandeh and Mohammad Shahedi, my siblings, Ehsan, Yasser and Sama, and other extended family for their love and incredible support in every step of my life’s journey.

للاستشارات المنارة

www.manaraa.com

Table of Contents

Abstract i

Co-Authorship Statement v

Dedication vii

Acknowledgements viii

Table of Contents xi

List of Tables xv

List of Figures xvi

List of Abbreviations xvii

Chapter 1. Introduction 1
- 1.1 Anatomy of the prostate gland 3
- 1.2 Prostate cancer and its prevalence 5
- 1.3 Prostate cancer screening and diagnosis 5
  - 1.3.1 Digital rectal examination 7
  - 1.3.2 Prostate specific antigen test 7
  - 1.3.3 Prostate biopsy 8
  - 1.3.4 Grading and staging of prostate cancer 9
  - 1.3.5 Other diagnosis methods 11
  - 1.3.6 Risk groups 12
  - 1.3.7 Early diagnosis clinical workflow 13
- 1.4 Prostate cancer treatment 14
  - 1.4.1 Radical treatments 14
  - 1.4.2 Lesion-directed treatments 17
- 1.5 Prostate cancer imaging 17
  - 1.5.1 Ultrasound imaging 18
  - 1.5.2 Computed tomography imaging 19
  - 1.5.3 Magnetic resonance imaging 19
  - 1.5.4 Nuclear imaging 26
- 1.6 The role of MRI in diagnosis and treatment of prostate cancer 26
  - 1.6.1 MRI-targeted TRUS-guided biopsy 26
  - 1.6.2 MRI-CT fusion radiotherapy planning 27
  - 1.6.3 MRI-guided biopsy and focal therapy 28
- 1.7 Prostate contouring on MRI 28
  - 1.7.1 Challenges in manual prostate contouring in MRI 29
  - 1.7.2 Computer-assisted prostate segmentation on MRI 30
  - 1.7.3 Validation challenges 34
- 1.8 Hypothesis 36

www.manaraa.com

1.9 Objectives ... 37
1.10 Thesis outline ... 37
    1.10.1 Chapter 2 - Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods ... 37
    1.10.2 Chapter 3 - Accuracy and acceptability of an automated method for prostate segmentation in magnetic resonance imaging ... 39
    1.10.3 Chapter 4 - Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging ... 40
1.11 References ... 42

Chapter 2. Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods **53**
2.1 Introduction ... 53
2.2 Materials and Methods ... 57
    2.2.1 Materials ... 57
    2.2.2 Semi-automated segmentation ... 58
    2.2.3 Validation metrics ... 65
2.3 Experiments ... 69
    2.3.1 Inter-operator variability: Manual segmentation ... 69
    2.3.2 Accuracy and Inter-operator variability: Semi-automatic segmentation ... 70
    2.3.3 Sensitivity to initialization: Semi-automatic segmentation ... 71
    2.3.4 Source of Variability: Semi-automatic segmentation ... 72
2.4 Results ... 73
    2.4.1 Inter-operator variability: Manual segmentation ... 73
    2.4.2 Accuracy and inter-operator variability: Semi-automatic segmentation ... 74
    2.4.3 Sensitivity to initialization: Semi-automatic segmentation ... 80
    2.4.4 Source of variability: Semi-automatic segmentation ... 81
2.5 Discussion ... 81
    2.5.1 Inter-operator variability: manual segmentation ... 82
    2.5.2 Accuracy and inter-operator variability: Semi-automatic segmentation ... 83
    2.5.3 Sensitivity to initialization: Semi-automatic segmentation ... 86
    2.5.4 Source of variability: Semi-automatic segmentation ... 86
    2.5.5 Limitations ... 87
    2.5.6 Conclusions ... 88
2.6 References ... 89

Chapter 3. Accuracy and acceptability of an automated method for prostate segmentation in magnetic resonance imaging **92**
3.1 Introduction ... 92
3.2 Materials and Methods ... 97
    3.2.1 Materials ... 97
    3.2.2 Automated segmentation ... 98
    3.2.3 Validation ... 102

المنارة للاستشارات
www.manaraa.com

3.3 Experiments 104
  3.3.1 Comparison of automatic and semi-automatic segmentation: accuracy and time 104
  3.3.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation 105
3.4 Results 107
  3.4.1 Comparison of automatic and semi-automatic segmentation: accuracy and time 107
  3.4.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation 108
3.5 Discussion 112
  3.5.1 Comparison of automatic and semi-automatic segmentation: accuracy and time 112
  3.5.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation 114
  3.5.3 Limitations 116
  3.5.4 Conclusions 117
3.6 References 119

Chapter 4. Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging 121
4.1 Introduction 121
4.2 Materials and Methods 125
  4.2.1 Materials 125
  4.2.2 Study design 126
  4.2.3 Spatial editing magnitude and inter-operator variability 127
  4.2.4 Required editing time and correlation with spatial error metrics 130
4.3 Results 132
  4.3.1 Spatial editing magnitude and inter-operator variability 132
  4.3.2 Required editing time and correlation with spatial error metrics 134
  4.3.3 Correlation of editing time with the metric values 135
4.4 Discussion 139
  4.4.1 Spatial editing magnitude and inter-operator variability 139
  4.4.2 Required editing time and correlation with spatial error metrics 141
  4.4.3 Limitations 144
  4.4.4 Conclusions 145
4.5 References 147

Chapter 5. Conclusions and directions for future work 150
5.1 Conclusions 150
5.2 Applications and future directions 153
  5.2.1 Applications in ongoing clinical research studies 153
  5.2.2 Suggestions for future work 154
5.3 References 156

Appendix A 158

المنارة للاستشارات
www.manaraa.com

Curriculum Vitae **159**

www.manaraa.com

List of Tables

Table 1.1: Prostate cancer staging using the TNM system. 11

Table 1.2: Prostate cancer risk groups. 13

Table 1.3: A survey of prostate MRI segmentation algorithms. 35

Table 2.1: Inter-operator variability in manual segmentation: Range of mean MAD, DSC, recall, precision, and ΔV (bilateral MAD and |ΔV| was used for “Operator vs Operator” section). 74

Table 2.2: Accuracy and variability for semi-automatic segmentation: mean±standard deviation of MAD, DSC, recall, precision, and ΔV. Corresponding letters show statisticaly significant differences between each error value of our method and the corresponding error value of another method where applicable. ($p < 0.05$). 75

Table 2.3: The average results across the nine operators and 10 images compared to the single operator results across 42 images: mean±standard deviation of MAD, DSC, recall, precision, and ΔV. Corresponding letters indicate statistically significant differences between two modes ($p < 0.05$). 77

Table 2.4: Consistency of the manual and the semi-automatic segmentations: average of means (average of standard deviations) of MAD, DSC, recall, precision, and ΔV across 3 manual and 3 semi-automatic segmentations of the prostate by 3 expert operators. 78

Table 2.5: Sensitivity of the semi-automatic algorithm to initialization (center points and anteroposterior symmetry axes): mean±standard deviation of MAD, DSC, recall, precision, and ΔV offsets from reference measurements across 1000 repetitions × 42 patients. 81

Table 3.1: Accuracy of automatic segmentation: mean ± standard deviation of MAD, DSC, recall, precision, and ΔV. 108

Table 4.1: Reported segmentation errors for prostate segmentation algorithms intended for use on T2w ER MRI. 125

Table 4.2: User manual interaction time for ready to use prostate segmentation in T2w MRI. 134

Table 4.3: Correlation between editing time and spatial editing magnitude measured using five metrics. 136

للاستشارات المنارة

www.manaraa.com

List of Figures

Figure 1.1: Prostate location in the reproductive system of the male human body. 4

Figure 1.2: Prostate gland zonal anatomy. 4

Figure 1.3: Digital rectal examination (DRE)... 7

Figure 1.4: Diagram depicting transrectal prostate biopsy. 9

Figure 1.5: Three standard imaging planes in radiology. 20

Figure 1.6: Axial view of T2w prostate MRI acquired (a) without, and (b) with ER coil... 21

Figure 1.7: ER coil distortion on MRI 25

Figure 2.1: Algorithm block diagram. 59

Figure 2.2: Training. 61

Figure 2.3: Segmentation. 65

Figure 2.4: Inter-observer variability. 74

Figure 2.5: Qualitative and quantitative results for three sample prostates. 76

Figure 2.6: Inter-observer variability. 79

Figure 3.1: Automatic coarse localization of the prostate. 102

Figure 3.2: Elements used to calculated the DSC, recall, and precision validation metrics.. 104

Figure 3.3: Accuracy of the computer-based segmentations vs. inter-operator variability of manual segmentation.. 110

Figure 3.4: Accuracy of the computer-based segmentations vs. inter-operator variability of manual segmentation.. 111

Figure 4.1: Study design showing the workflow for a particular operator #i. 126

Figure 4.2: User manual interaction time on manual, semi-automatic (S.A.) and automatic segmentations. 134

Figure 4.3: Inter-operator variability. 137

Figure 4.4: Editing magnitude, showing the differences between the segmentations pre- and post-editing for each of the three conditions. 138

المنارة للاستشارات

www.manaraa.com

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ADC	Apparent diffusion coefficient (mm^2/sec.)
ADT	Androgen deprivation therapy
ANOVA	Analysis of variance
AP	Anteroposterior
APSA	Anteroposterior symmetry axis
CT	Computed tomography
CZ	Central zone
DCE	Dynamic contrast-enhanced
DRE	Digital rectal examination
DSC	Dice similarity coefficient (%)
DWI	Diffusion-weighted imaging
EBRT	External beam radiotherapy
ER	Endorectal receiver
FDG	Fluorodeoxyglucose
FN	False negative (cm^3)
FP	False positive (cm^3)
HD	Hausdorff distance (mm)
HDR	High dose-rate
HIFU	High intensity focused ultrasound
HRQOL	Health-related quality of life
IGRT	Image-guided radiotherapy
IMRT	Intensity modulated radiotherapy
IS	Inferior-superior
LINAC	Linear accelerator
LDR	Low dose-rate
MAD	Mean absolute distance (mm)
MG	Mid-gland

المنارة للاستشارات

www.manaraa.com

MICCAI	Medical Image Computing and Computer Assisted Intervention
MR	Magnetic resonance
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
NCC	Normalized cross correlation (unitless)
NCCN	National Comprehensive Cancer Network
NVB	Neurovascular bundle
PCa	Prostate cancer
PDM	Point distribution model
PET	Positron emission tomography
PROMISE	Prostate MR image segmentation
PSA	Prostate-specific antigen
PZ	Peripheral zone
RF	Radio frequency
RMS	Root-mean-square
ROI	Regions of interest
RT	Radiation therapy
SA	Semi-automatic
SD	Standard deviation
SNR	Signal-to-noise ratio (unitless)
STAPLE	Simultaneous truth and performance level estimation
T2w	T2-weighted
TNM	Tumor Node Metastasis
TP	True positive (cm^3)
TRUS	Transrectal ultrasound
TZ	Transition zone
WG	Whole gland
ΔV	Volume difference (cm^3)

www.manaraa.com

Chapter 1.

Introduction

Three-dimensional (3D) segmentation of the prostate in medical images is useful for diagnosis and treatment planning of prostate cancer (PCa) [1, 2]. Magnetic resonance imaging (MRI) is increasingly being utilitized for PCa diagnosis and staging [3]. More specifically, T2-weighted (T2w) MRI is superior to other MRI sequences in anatomy visualization and is most commonly used for contouring the prostate boundary [3]. Endorectal receiver (ER) coil MRI provides improved image quality by increasing the contrast and signal-to-noise ratio [4-6]. However, the ER coil causes substantial deformation of the prostate tissue [7, 8] and also renders fine details and edges more salient in the magnetic resonance (MR) images, presenting an additional challenge to segmentation algorithms developed for use on MR images acquired without an ER coil. Manual contouring of the prostate on MRI is a time-consuming task with substantial inter-observer variability [9]. This is an important issue in clinical trials involving multiple investigators performing radiation therapy planning, in which inter-operator variation in contouring could materially impact the trial results. The impact of inter-operator contouring variability on clinical outcomes is unknown and has not yet been extensively studied [1].

Manual contouring of the prostate is a labourious and time-consuming task [10, 11]. Employing a computer-assisted algorithm for prostate segmentation could facilitate the contouring task by making it faster with minimal required user interaction. It also could help in the establishment of clinical methods that require the prostate to be

المنارة للاستشارات
www.manaraa.com

segmented more than once during the course of the treatment. For example, recontouring the prostate over the course of radiation therapy (RT) could help radiotherapists to adjust the plan based on the evolving condition of the patient [12].

Several algorithms have been presented in the literature for 3D segmentation of the prostate on T2W MRI acquired with an ER coil [13-16]. Predominantly, in the published studies the accuracy of the segmentation results was evaluated by comparison to a single-observer reference standard manual segmentation of each image in the data set. However, due to high inter-observer variation in manual contouring of the prostate by experts, there is no gold standard for prostate contouring on MRI [1] and this challenges the comparison of error metric values obtained from different segmentation algorithms on different data sets. Furthermore, in most studies, the choice of segmentation error metrics is somewhat arbitrary and not determined based on clinical demands [1]. The use of a set of complementary error metrics to capture most types of medically relevant segmentation errors is beneficial to assessing an algorithm's suitability for clinical translation. Comparing two commonly-used metrics, surface-based error metrics such as the mean absolute distance (MAD) between segmentation boundaries are usually more sensitive to local misalignments of the shape surfaces. Region-based metrics such as the Dice similarity coefficient (DSC), are measured based on the overlap area or volume and they are less sensitive to error types like sharp local surface misalignments. The appropriateness of a particular method depends on the intended use of the segementation. For example, a local surface misalignment in radiation therapy could be neglected using a region-based error metric but this error might cause an overdose of healthy tissues or an underdose of the tumour volume.

المنارة للاستشارات
www.manaraa.com

The focus of this thesis is to comprehensively evaluate the prostate segmentation accuracy and inter-operator variability and to compare novel semi-automatic and fully automatic computer-assisted segmentation algorithms with manual segmentation in terms of accuracy, reproducibility and required operator interaction time. The remainder of this chapter describes PCa prevalence and its clinical diagnosis and treatment methods; the role of medical imaging, specifically MRI, in diagnosis and treatment planning of the PCa; and the role of prostate contouring in MRI-guided or MRI-targeted procedures.

1.1 Anatomy of the prostate gland

The prostate is a part of the reproductive and urinary system of the male human body. A normal and healthy prostate is usually a walnut-sized gland that is located just below the bladder, anterior to the rectum and posterior to the pubic symphysis. It has an inverted pyramid shape with the base abutting the bladder and the apex abutting the urogenital diaphragm (Figure 1.1). It surrounds the bladder neck and the urethra. Additional structures such as seminal vesicles, neurovascular bundles (NVBs) and muscles also surround the prostate gland. Figure 1.1 shows the prostate location in the male reproductive system.

The prostate gland is divided into three different zones with different embryologic origins: the peripheral zone, transition zone, and central zone. In a healthy prostate, the peripheral zone forms about 70% of the prostate gland volume. It surrounds the urethra at the prostate apex and extends posterolaterally to the prostate base. The central zone contains the ejaculatory ducts and accounts for about 25% of the prostate tissue. The transition zone comprises only 5% to 10% of the gland volume and is located between

www.manaraa.com

the peripheral and the central zones. Figure 1.2 shows the anatomy of the prostate gland and its zones.

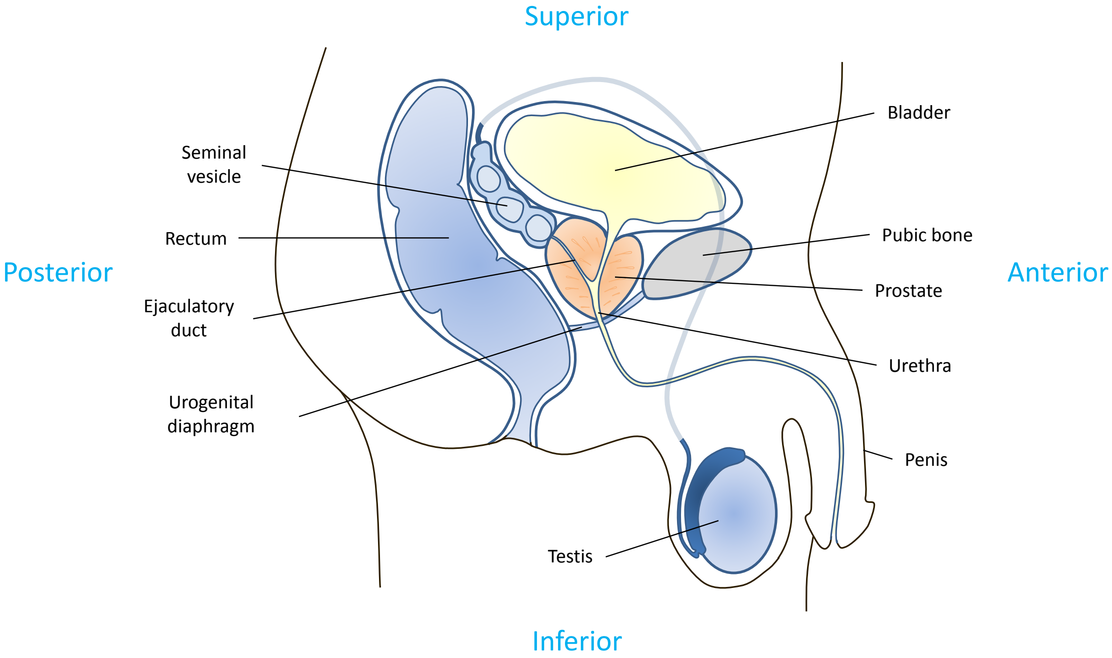


Figure 1.1: Prostate location in the reproductive system of the male human body. Adapted from [17].

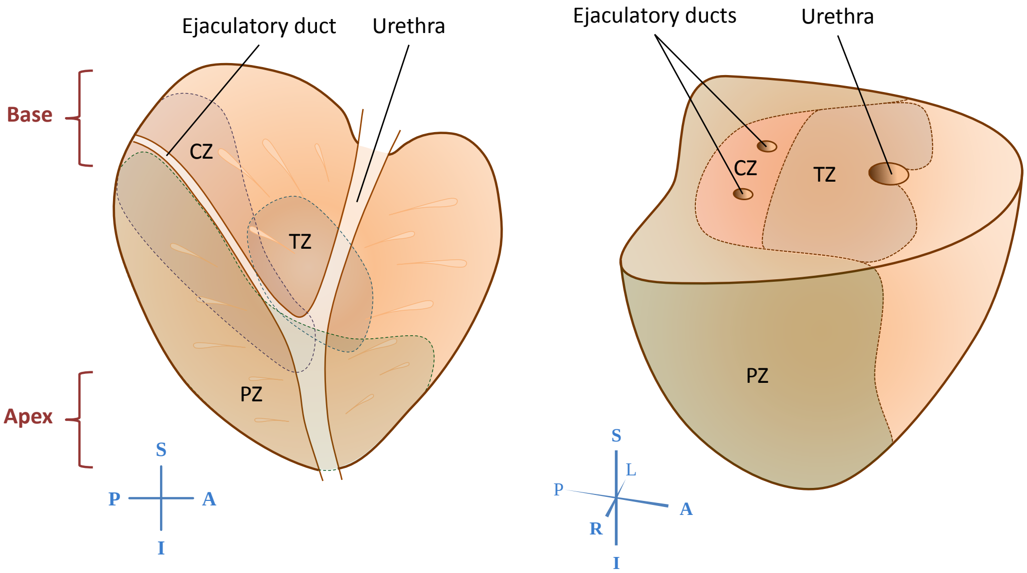


Figure 1.2: Prostate gland zonal anatomy. CZ: central zone, TZ: transition zone, and PZ: peripheral zone

www.manaraa.com

1.2 Prostate cancer and its prevalence

PCa is the most commonly diagnosed cancer in men in North America, excluding skin cancer [18, 19]. The American Cancer Society predicts that in 2015 over 220,000 new cases of PCa will be diagnosed; this represents approximately one quarter of all cancers among men. The American Cancer Society also predicts over 27,000 deaths from PCa in the United States in 2015 [18]. In Canada, one out of eight men will be diagnosed with PCa within his lifetime. Approximately 24,000 new cases of PCa are predicted to be diagnosed in Canada in 2015. PCa is also the second cause of death by cancer among Canadian men [19]. About 98% of PCa cases occur in men aged 50 years and older [20].

Prostate tumour cells develop in widely different patterns, at different growth rates, and with different aggressiveness and metastatic ability. There also exist benign prostate tumours that are not cancerous. Benign tumours can cause problems, such as pain and urination difficulties, but usually they are not life threatening and they do not spread to other organs of the body. Due to the wide variation in prostate tumour types, accurate and reliable diagnosis of PCa is vital for planning effective treatment that is appropriate to the aggressiveness level of the disease. Tumour size, location and extent, and the type of carcinoma are considered by clinicians during treatment selection.

1.3 Prostate cancer screening and diagnosis

In general, screening means testing for a disease in healthy and asymptomatic populations to identify and treat the disease at earlier stages, whereas diagnosis refers to detecting disease among individuals having symptoms and signs. Since PCa is usually asymptomatic in the early stages, there are some screening tests such as digital rectal examination (DRE), prostate-specific antigen (PSA) blood testing and transrectal

المنارة للاستشارات
www.manaraa.com

ultrasound (TRUS) imaging [21] to help in identifying the PCa in its early stages. It has been shown that about 75% of PCa would not be diagnosed early without PSA screening in the population [22] and there is no doubt that early diagnosis of aggressive PCa is helpful to preventing cancer spread to other organs. An early and accurate diagnosis helps to treat the patient more efficiently and effectively. However, screening in PCa is still a controversial subject [23-26]. On one hand, there is evidence that PCa screening could reduce the rate of advanced and metastatic PCa [25, 26]. On the other hand, there is a higher probability of missing a fast-growing tumour in the interval between PCa screening tests [24], implying that screening is a less effective approach for diagnosis of fast-growing PCa tumours compared to slow-growing tumours that are usually less life-threatening. Moreover, early diagnosis of PCa through screening tests could be also harmful due to potential overdiagnosis and overtreatment of the disease [23]. Overtreatment of PCa is especially concerning in patients for whom treatment is associated with minimal benefit compared to active surveillance (e.g. patients older than 65 years [27]).

The success of screening processes depends on two principal conditions: (1) there are available tests to detect the disease at the early stages, and (2) there are effective treatments for the disease at the early stages [24]. Therefore, to improve the benefits of screening it is necessary to improve the accuracy, sensitivity and specificity of the cancer detection methods as well as the efficiency of the treatment procedures.

In the following subsections, we briefly introduce the standard tests that are typically used for PCa screening and diagnosis.

المنارة للاستشارات
www.manaraa.com

1.3.1 Digital rectal examination

DRE is a PCa screening test. During the test, the physician uses a gloved lubricated finger to palpate the prostate gland through rectum to examine the prostate for any irregularities in shape, size and texture (Figure 1.3). The detection rate of DRE by itself is low [28]. Sensitivities of 40% to 55% has been reported in the literature for DRE [28-30]. However, there are cancers that are detected by DRE alone or could be diagnosed through DRE earlier than with PSA blood testing and ultrasound imaging [31]. Therefore, considering also the simplicity and availability of the test, DRE is routinely used for screening.

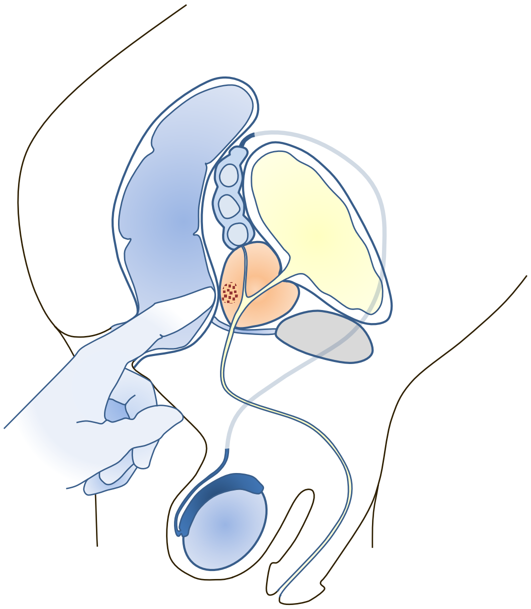


Figure 1.3: Digital rectal examination (DRE) Adapted from [17].

1.3.2 Prostate specific antigen test

PSA is a protein that is produced by the prostate gland and released into the bloodstream. Most of the time, when an abnormality such as PCa occurs in the prostate,

www.manaraa.com

more PSA is released into the blood stream. During a PSA test, a small amount of blood is taken and the level of PSA in the blood is measured. A high level of PSA in the blood or a rapid elavation of the PSA level over time are considered as signs of suspicion for PCa.

There is not any level of PSA in the blood defined as normal. However, traditionally, PSA levels of 4 ng/mL are considered as a cutoff point to distinguished normal from abnormal. PSA levels above 10 ng/mL are usually considered as high PSA level that is suspicious for advanced or metastatic PCa. According to American Cancer Society guidlines for early detection of PCa, if the PSA level is lower than 2.5 ng/mL, screening could be conducted every two years, and for PSA levels of 2.5 ng/mL and above the screening interval should be one year. It also suggests biopsy for men at average risk for PCa whose PSA level is 4.0 ng/mL or greater. For individuals at high risk for cancer when the PSA level is within the range of 2.5 ng/mL to 4.0 ng/mL individualized diagnosis planning is suggested [32].

1.3.3 Prostate biopsy

Patients with abnormal DRE or high or elevated PSA levels are referred for prostate biopsy [32]. Currently, prostate needle biopsy is the clinical standard for diagnosis of PCa. This is an outpatient procedure which is done under local anesthesia. During the biopsy process a thin needle is inserted through either the rectal wall (transrectal biopsy) or the perineum and usually 6 to 24 (typically about 12) small samples of the prostate are taken from different parts of the gland [33]. Transrectal prostate biopsy is most common. This process is usually done under two-dimensional (2D) TRUS imaging guidance. Figure 1.4 shows a schematic depicting transrectal

المنارة للاستشارات
www.manaraa.com

prostate biopsy. The biopsy samples are sent to a pathology laboratory, where a pathologist looks at the specimens under a microscope and reports on the presence and grade of PCa. The pathologist categorizes cancerous foci using a standard grading system called the Gleason grading system [34]. The number of biopsy samples that are cancerous and the percentage of cancer in each biopsy core are also reported. Since about 30% of cancers are missed during the first TRUS-guided prostate biopsy [35, 36], for individuals with persistently elevated PSA or positive DRE whose initial biopsy did not detect cancer, repeat biopsy is required [36].

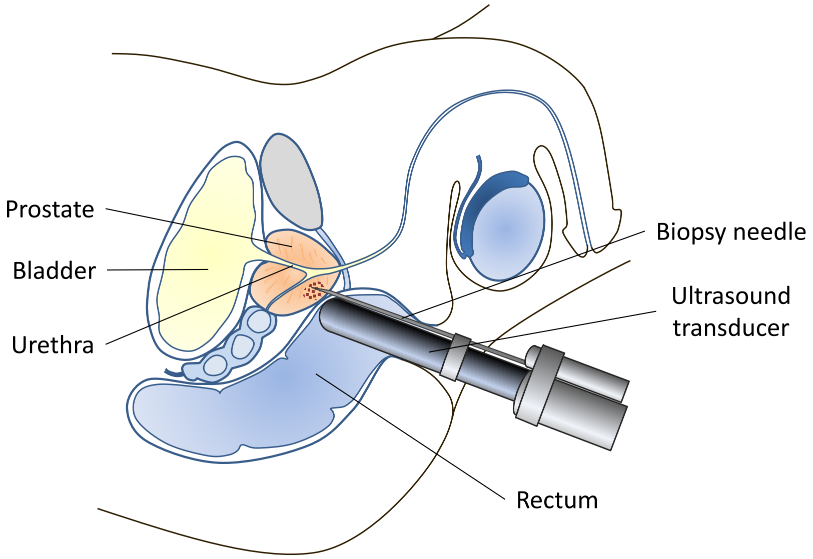


Figure 1.4: Diagram depicting transrectal prostate biopsy. Adapted from [17].

1.3.4 Grading and staging of prostate cancer

PCa grading: The Gleason system is one of the most commonly used systems for gradig PCa in pathological samples. The Gleason system was first presented in 1966 by Donald F. Gleason [37] and it became more popular in North America in the late 1980s and early 1990s. The grading system is based on tissue architecture at the cellular level

www.manaraa.com

and refers to identification of aggressiveness of the cancer cells based on their histologic pattern of arrangement [34]. In the Gleason grading system, nine fundamental tumour cell patterns are defined under 5 grades; grade 1 to grade 5, with lower grades being closer to normal tissue and higher grades being more aggressive. Within the prostate, the first and second most predominant Gleason grades are added together and reported as *Gleason score* or *Gleason sum*. The higher the Gleason score, the higher chance of harbouring PCa tumours with potential to grow and spread quickly.

PCa staging: PCa is also characterized based on how much cancer has spread within or beyond the prostate border; this is referred to as PCa staging. There is a strong relationship between PCa stage and the probability of curative treatment. The tumour-node-metastasis (TNM) staging system presented by the American Joint Committee on Cancer is one of the most commonly used cancer staging methods worldwide [38]. It categorizes PCa into four main stages; taking the size of the tumour, the extent to which lymph nodes are involved, and the presence of metastases into account. Gleason grading of the tumour is also considered in the staging process. In stage I, cancer foci are usually microscopic and cannot be detected during DRE, the PSA level is lower than 10 ng/mL, and the Gleason score is less than or equal to 6. In this stage, cancer is usually detected through biopsy and few of the obtained samples are cancerous. In stage II, cancer is confined to the prostate gland and may be detected by DRE. In this stage, PSA could rise up to 20 ng/mL or higher. In stage III, the tumour has extended beyond the prostate capsule, but no regional lymph node metastasis or distant metastasis is detected. In stage IV, cancer has invaded adjacent tissues and organs. In this stage, metastasis in regional

المنارة للاستشارات
www.manaraa.com

lymph node(s) and/or other organs might be found [38]. Table 1.1 presents a brief and general overview of the TNM staging system for PCa.

Table 1.1: Prostate cancer staging using the TNM system. This table is adapted and summarized from [38]. N0 means no regional lymph node metastasis and N1 means regional lymph node metastasis. M0 means no distant metastasis and M1 means metastasis in other organs beyond the prostate gland.

Stage	***Tumour***	***Regional lymph node metastasis***	***Distant metastasis***	***PSA level*** * ***(ng/mL)***	***Gleason score*** *
I	Microscopic nonpalpable tumour confined to prostate capsule	N0	M0	< 10	≤ 6
IIA	Tumour confined to prostate capsule and involving 50% or less of one lobe	N0	M0	<20	≤ 7
IIB	Tumour confined to prostate capsule and involving either more than 50% of one lobe, or both lobes	N0	M0	any	any
III	Tumour expansion beyond the prostate	N0	M0	any	any
IV	Tumour invasion of adjacent structures beyond the prostate	N0/N1	M0/M1	any	any

* Where available

1.3.5 Other diagnosis methods

Medical imaging: Ultrasound imaging is one of the imaging methods currently performed as a clinical follow up method to the DRE and PSA blood tests to measure the prostate gland size, as well as the PCa tumour size, location and extent. Ultrasound imaging is also used for guidance of clinical procedures such as prostate biopsy or brachytherapy. 2D TRUS imaging is the most common method used for PCa diagnosis or guidance of some of the clinical procedures. Between 25% and 40% of PCa tumours have

www.manaraa.com

been reported as isoechoic [39-41], meaning that they are not detected through ultrasound imaging.

Bone scan: Bones are usually the first target of metastasis in prostate cancer. Therefore a bone scan or bone scintigraphy is usually used as a follow up test for high-grade and/or high-stage PCa. The bone scan is a nuclear imaging method in which a low-level radioactive material (called a radiotracer) is injected into a vein and this material is absorbed by bones. A gamma camera, which is a radiation-sensitive device, scans the body and detects the radiation emitted by the radiotracer. The more active the bone, the more radiotracer will be absorbed and detected by the camera. Some tumours, infections, bone abnormalities and bone damage show up as sites of increased radiotracer uptake and are demonstrated as hot spot areas on imaging. The hot spots might suggest cancer metastasis to the bone or they might be detected because of some other bone abnormalities.

1.3.6 Risk groups

PCa patients are categorized into six different risk groups based on the initial clinical assessment: very low-, low-, intermediate-, high-, very high-risk, or metastatic cancer. These risk groups are used to choose the appropriate treatment option for the patients. If a patient’s risk group changes over a period of time, this is strongly suggestive of cancer progression and indicates radical treatments such as surgery or radiotherapy [42]. Table 1.2 shows the six risk factors and their characteristics. In the next subsection (1.3.7) we describe the clinical treatment plan that is suggested for each risk group.

المنارة للاستشارات

www.manaraa.com

1.3.7 Early diagnosis clinical workflow

The National Comprehensive Cancer Network (NCCN) has published a guideline that suggests a clinical workflow for PCa diagnosis and treatment planning [42]. In this guideline for early detection of the PCa, initial risk assessment based on DRE and PSA is suggested. Biopsy is usually offered based on DRE and PSA results. For individuals of 45 to 75 years of age with normal DRE and PSA level lower than 1 ng/mL, repeating DRE and PSA at 2-4 years intervals is suggested. If the PSA level is equal to or higher than 1ng/mL, or the individual’s age is above 75 years with a PSA level lower than 3 ng/mL, DRE and PSA testing are offered every 1–2 years. For individuals with PSA level above 3.0 ng/mL, the NCCN guidline suggests workup for benign disease; i.e., either TRUS-guided biopsy or PSA and DRE testing every 6–12 months [43].

Table 1.2: Prostate cancer risk groups. This table is adapted from [44].

Risk group	*PCa extent*	*Staging*		*Gleason score*		*PSA level (ng/mL)*
Very low	Clinically localized	Stage I ≤ 5% tissue involvement < 3 positive biopsy cores (< 50% cancer in each)	and	≤ 6	and	< 10
Low	Clinically localized	Stage I or II ≤ 50% of one lobe is involved	and	2 to 6	and	< 10
Intermediate	Clinically localized	Stage II > 50% of one lobe or both lobes are involved	or	7	or	10-20
High	Clinically localized	Stage III Extracapsular extension	or	8-10	or	> 20
Very high	Locally advanced	Stage III or IV Seminal vesicle(s) invasion		any		any
Metastatic	Metastatic	Regional lymph and/or distant metastasis		any		any

For treatment planning, NCCN guidelines suggest a specific strategy for each risk group. The recommended strategies are usually based on the estimated life expectancy of the patient and PCa growth and progression over time. The suggested treatment options

المنارة للاستشارات
www.manaraa.com

could be selected from active surveillance (i.e. active monitoring of disease progression), radiotherapy (brachytherapy or external beam radiotherapy; EBRT) and surgery (radical prostatectomy with or without pelvic lymph node dissection). In some cases, androgen deprivation therapy or ADT is also recommended, usually combined with the radical treatments such as EBRT or radical prostatectomy [44].

1.4 Prostate cancer treatment

1.4.1 Radical treatments

Surgery: Currently, one of the clinical standard PCa treatments is to surgically remove the whole prostate gland and the attached seminal vesicles, also known as radical prostatectomy. This surgery is usually done for patients with clinically localized PCa that is progressive and aggressive. Some times the local lymph nodes are also removed during the same surgery [44]. Radical prostatectomy is sometimes followed by other treatment or monitoring options such as radiotherapy, chemotherapy, ADT or active surveillance to avoid the risk of PCa recurrence [45].

Since the prostate is surrounded by the sphincter urethrae muscle, as well as nerves and blood vessels that are critical for erections, and is attached to many organs such as rectum and bladder, radical prostatectomy can have severe side effects such as urinary incontinence and erectile dysfunction [46-48]. Prostatectomy has minimal post-surgery bowel function-related symptoms [49]. The NCCN guidelines recommend radical prostatectomy for patients with 10 or more years of estimated life expectancy who do not have any serious health conditions that would contraindicate the surgery [44].

External beam radiation therapy: EBRT is another common radical treatment option for PCa, where ionizing radiation (e.g. X-ray) is generated and delivered to the

المنارة للاستشارات
www.manaraa.com

target by a computer-controlled linear accelerator (LINAC). The LINAC targets the prostate and directs the radiation from outside of the body at the prostate gland to kill the cancerous cells. Intensity modulated radiotherapy (IMRT) is one of the state-of-art radiathion therapy methods in radiation oncology used for PCa treatment. With IMRT, compared to traditional radiation therapy, oncologists can plan the radiation therapy with the aim of delivering a higher dose to the tumour and minimizing radiation exposure to the healthy surrounding tissues. For accurate radiation delivery to the target, prostate localization is performed by image-guided radiotherapy (IGRT) [44]. In IGRT, for each radiation delivery secssion the target is tracked by an intra-operative imaging system such as ultrasound imaging, X-ray imaging or cone-beam computed tomography (CT) to increase the accuracy of the targeting and compensate tissue movment.

Prostate EBRT dose planning is usually done under CT image guidance because CT provides 3D anatomical localization of the pelvis and also provides the electron density information of the tissues that is required for radiation dose calculation. Radiation oncologists usually use inverse planning for radiation dose planning in IMRT. In inverse planning the oncologists first delineate the prostate border as well as the surfaces of all organs at risk in 3D. They then use advanced software to prioritize the dose delivery and limitations for the organs at risk and run the software to design the dose plan. The dose plan is used in a computer-controlled LINAC for radiation therapy delivery.

The limitation with CT based planning is the low soft tissue contrast in CT images. Therefore, CT cannot provide accurate and repeatable contour delineation for the prostate and some of the surrounding organs at risk such as the rectum, bladder and NVBs [50, 51].

www.manaraa.com

In terms of health-related quality of life (HRQOL), in general, patients who undergo EBRT have less urinary incontinence but worse bowel function compared to prostatectomy patients [52, 53]. HRQOL improves over time post-treatment for PCa patients treated with EBRT [54]. EBRT also avoids surgery-associated risks and complications such as bleeding and transfusion-related risks, and anesthesia-associated side effects [44].

Brachytherapy: Brachytherapy, as an internal radiotherapy, is another radical treatment method usually used for lower-risk PCa cases [44]. In this method, radioactive sources are placed within the prostate tissue to kill the cancerous cells. Prostate brachytherapy is an outpatient procedure that is performed under either general or spinal anesthesia. The treatment is usually planned using ultrasound and/or MR imaging. The radioactive sources are usually placed in the prostate through transperineal insertion under the guidance of an imaging technique like TRUS [55].

Low dose-rate (LDR) and high dose-rate (HDR) brachytherapy are the two main types of brachytherapy treatment approaches for PCa. In HDR brachytherapy a catheter is inserted into the prostate and a high-dose radiation is delivered to the cancerous tissue. In LDR brachytherapy a number of small radioactive seeds are permanently implanted in the prostate gland to deliver low dose radiation to the tumour cells within a longer period of time compared to HDR brachytherapy. Brachytherapy as monotherapy is recommended to patients with low-risk PCa. For intermediate-risk PCa, brachytherapy is combined with EBRT with or without ADT. Brachytherapy rarely is a useful option for high-risk PCa treatment [44].

المنارة للاستشارات
www.manaraa.com

Brachytherapy is usually performed within a day and the patient can return to normal activities in a short time [44]. Less erectile dysfunction is reported after brachytherapy compared to EBRT and prostatectomy [49]. The incidence of urinary continence is lower after brachytherapy compared to prostatectomy, and bowel dysfunction is comparable to EBRT [49, 52].

1.4.2 Lesion-directed treatments

In a subset of prostate cancer patients with organ-confined cancer, PCa consists of a dominant high-grade tumour surrounded by primarily non-cancerous tissue. Therefore, a number of emerging therapy methods such as cryotherapy and high intensity focused ultrasound (HIFU) suggest preserving as much healthy parenchyma as possible and delivering the treatment to the tumour site [44]. In these local therapy methods (also known as focal therapies) the treatment is focused on the tumour cells to spare healthy tissues from destruction. This leads to minimally invasive treatments with fewer and less-severe risks and side effects compared to radical treatments like prostatectomy and radiotherapy.

1.5 Prostate cancer imaging

There are many different imaging modalities that are being used for PCa diagnostic and therapeutic procedures. For each clinical procedure, the imaging modality to be utilitzed is chosen according to the features required for that type of procedure. Sometimes it is required or more effective to use combination of two or more imaging methods. Ultrasound, CT, MRI and positron emission tomography (PET) are the most

المنارة للاستشارات

www.manaraa.com

popular imaging modalities that are currently being used for PCa diagnosis or treatment in clinical procedures.

1.5.1 Ultrasound imaging

As is also mentioned in section 1.3.5, ultrasound imaging and more specifically TRUS imaging is the most common imaging modality used for PCa diagnosis and treatment. It is an inexpensive and safe imaging modality that is available in most clinical centres. In TRUS imaging, a transrectal ultrasound transducer is inserted into the rectum and acquires images from the prostate gland through the rectal wall. TRUS is capable of displaying the anatomy of the prostate and provides real time imaging with rates of up to 30 frames per second. There are two types of TRUS probe available that provide different views; end-firing and side-firing probes. Both probe types are currently used in clinical procedures such as TRUS-guided biopsy but the preference of one over the other is still a matter of debate [56-59]. However, for prostate biopsy, end-firing probes are recommended because they provide greater freedom of biopsy plane manipulation and they enable better access to the peripheral zone, where PCa tumours are most likely to be found [57, 59]. Side-firing probes are mostly used in transperineal biopsy or brachytherapy.

TRUS is one of the imaging modalities used for accurate estimation of the prostate gland volume [60]. It is also used for PCa detection and tumour volume estimation, however ultrasound imaging is not able to detect all prostate tumours; about 25 to 40 percent of PCa tumours have been reported as isoechoic [39-41].

المنارة للاستشارات
www.manaraa.com

1.5.2 Computed tomography imaging

CT is an imaging modality based on X-ray irradiation of the body from different angles and processing the acquired data by a computer to generate the images. CT provides each 3D image in the form of a set of cross-sectional images. The CT image intensities are directly correlated with the electron density of the tissues. This is an exclusive feature of CT imaging that is required for radiation dose calculation during the radiotherapy planning process. However, X-rays forms the CT images, image contrast is lower for soft tissues compared to the image contrast for hard tissues (e.g. bones). For prostate imaging, although CT imaging provides a useful 3D anatomical image of the pelvis, prostate contouring on CT images is challenging and subject to high inter-observer variability compared to other imaging modalities such as ultrasound and MRI [9, 61, 62].

1.5.3 Magnetic resonance imaging

MRI is known as a noninvasive medical imaging method. MRI uses a strong magnetic field (usually 0.5 to 3.0 Tesla) and radio frequency (RF) pulses (with frequency of ~42.5 MHz/Tesla) to generate the cross-sectional images of the body. MRI yields high-contrast, detailed images of soft tissues. However, for air and bone imaging the quality of MR images is poor, and additional techniques such as using contrast agents are required.

MRI is capable of producing 3D images in the form of a set of cross-sectional 2D images. There are three orthogonal standard imaging planes defined in radiology to present cross-sectional views: the axial, sagittal, and coronal imaging planes. The axial plane (also known as the transverse plane) divides the body into superior and inferior

المنارة للاستشارات

www.manaraa.com

parts (Figure 1.5 (a)). The sagittal imaging plane (also known as the lateral plane) is perpendicular to the axial plane and divides body into left and right parts (Figure 1.5 (b)). The coronal imaging plane (also known as the frontal plane) is perpendicular to the axial and sagittal planes and divides the body into anterior (ventral) and posterior (dorsal) parts (Figure 1.5 (c)).

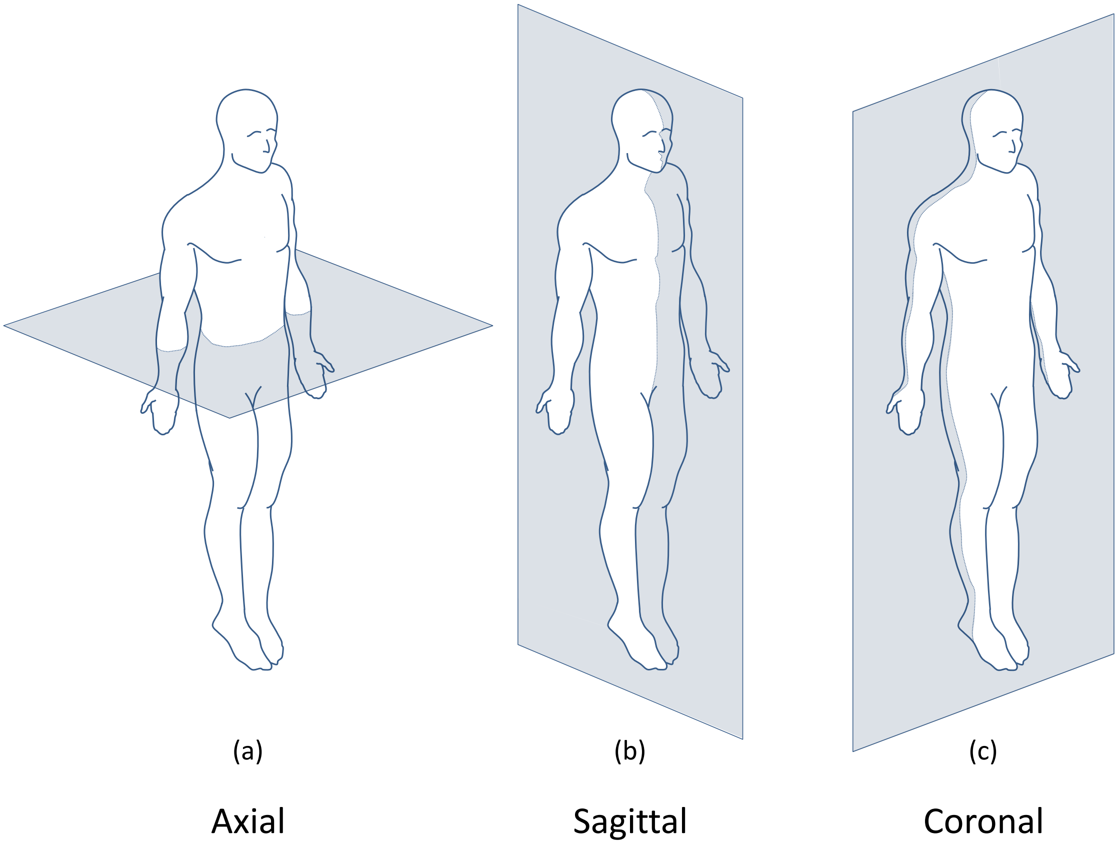


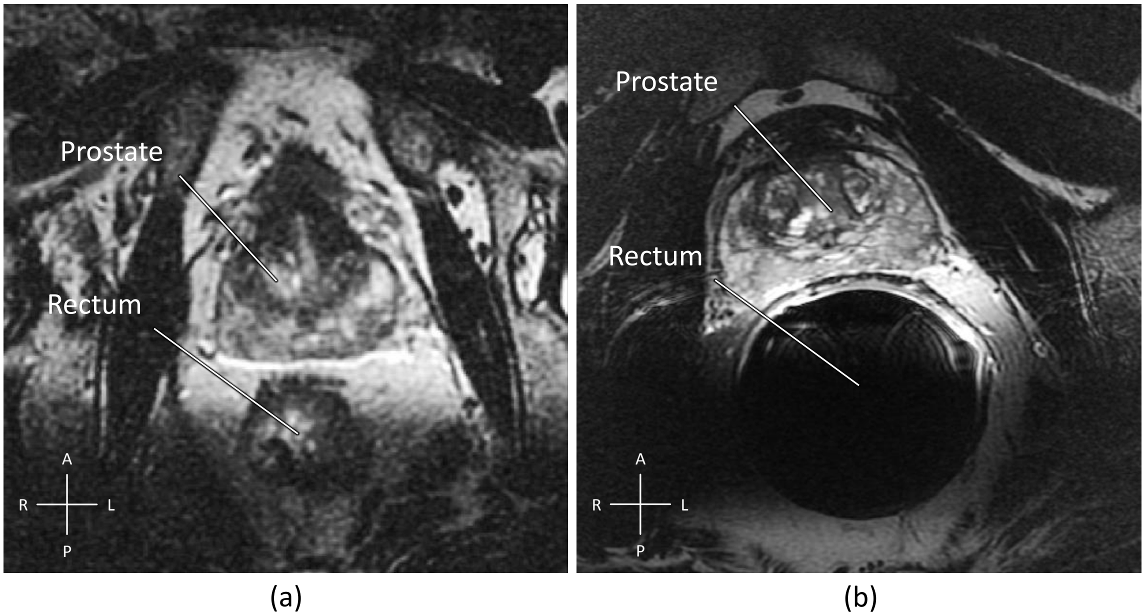
Figure 1.5: Three standard imaging planes in radiology. (a) Axial plane, (b) sagittal plane, and (c) coronal plane.

1.5.3.1 Prostate MRI

Although MRI is not used as a clinical standard test for PCa [63-65], MRI has demonstrated its potential and important role as an imaging modality for PCa management [63-68]. Over the past two decades, in many centres MR imaging has been

www.manaraa.com

used for PCa diagnosis, staging, treatment planning and therapy guidance [65-67, 69]. Most commonly, prostate MRI is performed at 1.5 or 3.0 Tesla magnetic field strength. In some centres, an ER coil and/or pelvic phased array coil (also known as a body coil) are used for prostate MRI to increase the signal-to-noise ratio (SNR) and improve the spatial resolution of the images [7, 70]. Although the optimal use of the ER coil for prostate MRI is still under study, there is some evidence of improvement in diagnosis and staging of PCa using ER MRI [7, 70, 71]. Figure 1.6 shows the same axial cross-section of the prostate on T2w MRI acquired with and without ER coil from the same patient.



(a)

(b)

Figure 1.6: Axial view of T2w prostate MRI acquired (a) without, and (b) with ER coil. Both images are midgland slices of the same patient

There are several different MR imaging pulse sequences available for prostate that form multiparametric MRI; e.g. T1-weighted MRI, T2w MRI, dynamic contrast enhanced (DCE) MRI, diffusion weighted imaging (DWI), and MR spectroscopy (MRS) [65].

المنارة للاستشارات
www.manaraa.com

1.5.3.2 Multiparametric MRI

T1- and T2-weighted MRI: T1- and T2-weighted MRI are both used for PCa detection [72]. T1-weighted MRI is also used for detection of hemorrhage after prostate biopsy [72]. The zonal anatomy of the prostate is appreciated better on T2w MRI compared to T1-weighted MRI, and therefore usually T2w MRI is used as a main imaging approach for anatomy description of the prostate and adjacent tissues [73-75]. In T2w MRI of the healthy prostate, the peripheral zone appears brighter than the central and transitional zones, which are mixed dark to semi-bright regions on the image [76]. In T2w MRI, a hypointense area within the peripheral zone is considered to be PCa unless a hyperintense area (i.e. usually associated with the post-biopsy hemorrhage) is observed at the same location on T1-weighted MRI [77]. However, in some cases, PCa is challenging to detect on T2w MRI, because PCa can occur within an isointense region or even a hyperintense area compared to the background [70]. Sensitivity and specificity for T2w MRI in PCa detection have been reported as 52% to 83%, and 46% to 83%, respectively [78].

Contouring of the prostate on MRI is used for localizing the prostate border with surrounding tissues to help clinicians deliver the treatment to the prostate gland, and more specifically, to the PCa tumour sites while preserving healthy surrounding tissues from harm. Due to its better anatomical definition, the contouring task is often performed on T2w MR images. Furthermore, T2w MRI is useful in assessing the PCa extent and its spread beyond the prostate border. Hence, prostate border localisation on this MRI sequence could be helpful to staging.

المنارة للاستشارات
www.manaraa.com

Dynamic contrast enhanced MRI: For acquisition of DCE MRI, an MRI contrast agent is injected into the body and the changes in contrast agent uptake and washout by the prostate tissue are measured through acquisition of a time series of T1-weighted MR images. DCE MRI is useful for detecting, localising, and staging of PCa [79, 80]. It has shown high sensitivity and specificity for early detection of PCa [79].

Diffusion weighted imaging: DWI is another type of MRI in which the mobility of water molecules at the microscopic level is measured. DWI measures the apparent diffusion coefficient (ADC) value that reflects the water diffusion pattern in the tissue. The idea behind clinical DWI is that, in general, the water motion in healthy human body tissues with intact cell microstructures is oriented and anisotropic. In a pathological change in tissue these microstructures are destroyed, therefore, the pattern of water diffusion in the tissue is more isotropic [81]. DWI has a short acquisition time and usually provides high-contrast between PCa and normal tissue and is useful in PCa diagnosis. However, because of low SNR, the spatial resolution of DWI is low [82].

MR spectroscopy: MRS is used in combination with MRI to provide more information about tissue characteristics [83]. Similar to MRI, MRS is also based on the nuclear magnetic resonance phenomenon. It provides information about the metabolic activity of the prostate by measuring the quantities of some metabolites (e.g. choline, citrates, creatine and polyamines) within the prostate gland. The metabolite quantities or the ratio between them indicate different abnormalities of the prostate [82]. One of the most important metabolite change in PCa is related to the level of citrate [84].

المنارة للاستشارات
www.manaraa.com

1.5.3.3 Endorectal receiver coil

In MRI and MRS, the smaller the receiver RF coil and the closer the coil is located to the target, the lower the noise level. Body and ER coils are two common RF receiver coils that are used in prostate MRI to enhance the quality of MR images in terms of spatial resolution and SNR [70]. The most common ER coil used in prostate MRI is an inflatable ER coil that consists of a probe with a inflatable latex cover, also called a balloon. The balloon is filled by either air, perfluorocarbon, or barium after insertion into rectum for better positioning and coverage, and less coil motion [85]. Typically, the inflated ER coil has a cylindrical shape with about 8.5 cm length and about 4.5 cm diameter after inflation [7, 70].

Despite improvement in image quality via the ER coil, the ER coil complicates some aspects of imaging. For example, the ER coil substantially displaces and deforms the prostate [7]. On average, it compresses the prostate gland about 15% anteroposteriorly, and expands it about 8% in the left-right direction [7]. In MRI-targeted image-guided procedures, MRI information is often combined with another imaging modality (such as intra-procedural TRUS). Therefore, the deformation of the prostate shape challenges image alignment between MRI and the other imaging modality. In EBRT, CT imaging (the standard imaging modality for dose calculation) is acquired with no prostate gland deformation and in TRUS-guided procedures, although the endorectal transducer is used, the shape of the transducer and the way it is located inside the rectum is different and therefore the prostate shape is deformed in a different way [7]. Another limitation related to the use of the ER coil is the presence of some image distortion and artifacts such as magnetic susceptibility, coil flare, and rectum movement artifacts [86].

www.manaraa.com

Figure 1.7 shows some types of imaging artifacts that occur on ER MRI. Some distortions, e.g. magnetic susceptibility, occur because of the air-inflated ER coil and can be reduced by replacing air with other ER coil balloon filling materials [85]. Furthermore, since the ER coil is placed posterior to the prostate, it generates an inhomogeneity in the received signal and, accordingly, in the image intensities [87]. In MRI, the voxel intensities are higher close to the coil.

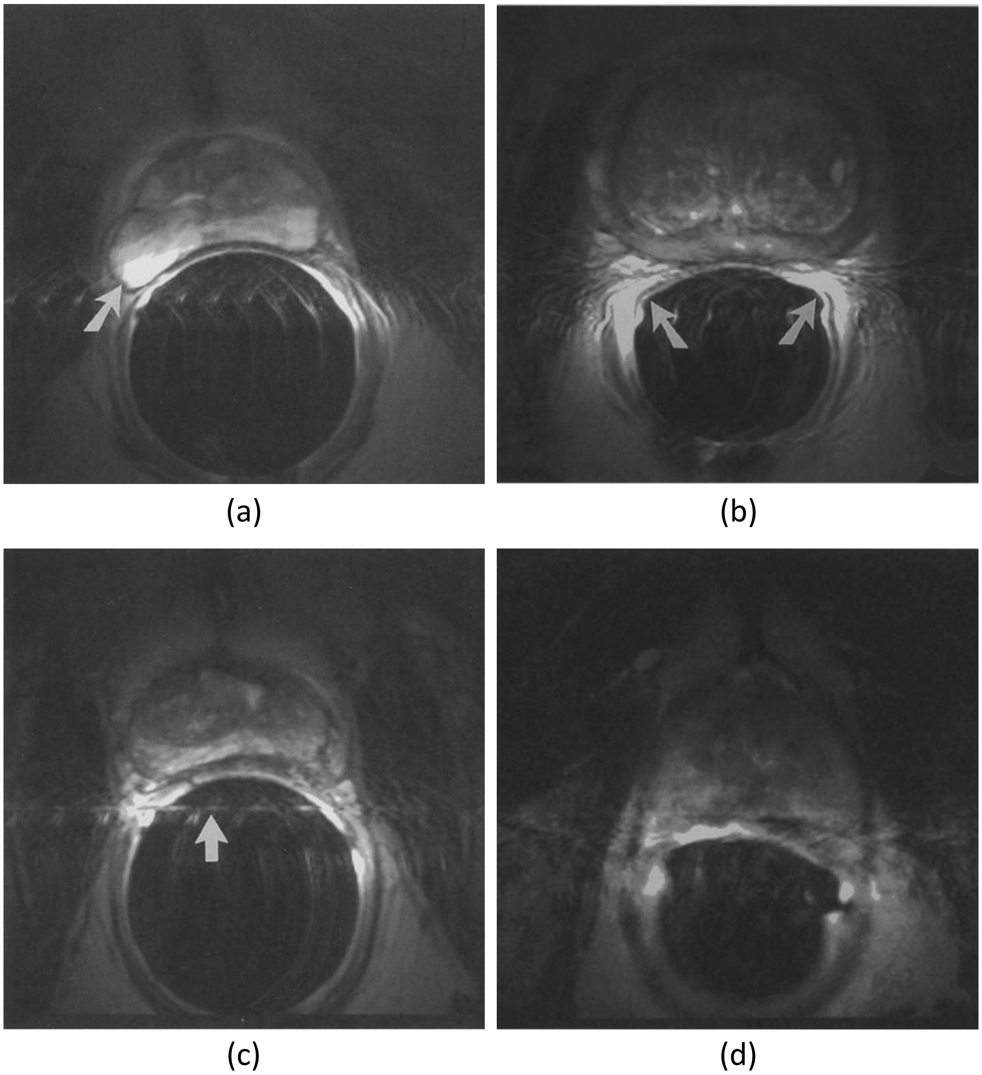


Figure 1.7: ER coil distortion on MRI [86]: (a) gland distortion, (b) near-field coil flare artifact, (c) coil-related artifact because of air-inflated balloon, and (d) rectal movement distortion.

www.manaraa.com

Although the ER coil improves image quality overall and some studies have shown a positive impact of using ER coil on MRI-based PCa diagnosis [6, 69, 71, 88-90], the use of the ER coil for prostate MRI is still debated because the coil is not comfortable for the patients and generates image distortions. One study suggested that the ER coil does not significantly improve MRI power in diagnosis of PCa [86]. Another study [91] has shown that in terms of staging accuracy, non-ER 3.0 Tesla MRI is equivalent to ER 1.5 Tesla MRI.

1.5.4 Nuclear imaging

Some types of nuclear imaging methods such as PET are also used for prostate imaging. PET scanning cannot provide accurate anatomical information; however, it can detect tumours based on the metabolic functionality of the tissues. Sometimes a combination of PET and another imaging modality such as CT (called PET/CT) is used to generate high-resolution anatomic images fused with functional images [92]. Since the metabolic glucose activity of PCa is low, fluorodeoxyglucose (FDG)-PET scanning is less useful for PCa diagnosis particularly in the early stages, but is usually used in metastasis detection [92]. There are studies that show the role of other radiotracers in PET imaging for early detection of PCa [93].

1.6 The role of MRI in diagnosis and treatment of prostate cancer

1.6.1 MRI-targeted TRUS-guided biopsy

Due to lack of visibility of many PCa tumours on TRUS, the standard TRUS-guided prostate biopsy is usually performed based on a systematic sampling approach from different regions of the prostate gland [33]. About 35% of PCa is not detected

المنارة للاستشارات
www.manaraa.com

during the first attempt at TRUS-guided biopsy [36]. Since MRI yields improved ability for detecting and localizing of PCa [63-69], there are some recently developed biopsy systems that utilize MR images to define biopsy targets, and map those targets to the real-time intra-operative TRUS images to help clinicians direct the biopsy needles to the pre-defined suspicious regions. It has been shown that this has increased the detection rate of biopsy and decreased the rate of repeat biopsy [94-96].

1.6.2 MRI-CT fusion radiotherapy planning

The CT scan is very important for dose calculation in radiation therapy planning because it provides the electron density distribution of body tissues as well as useful anatomical information. For dose planning, it is also important to identify the boundaries of the prostate and surrounding sensitive tissues and organs. Since the soft tissue contrast on CT images is lower than on MRI, contouring the prostate on CT could result in lower accuracy and higher intra- and inter-observer variability [9, 50]. It is also nearly impossible to detect or localize prostate tumours in CT images. One way to account for inter-observer variability in radiotherapy planning is to use an expanded safety margin around the boundary, but this can cause undesirable irradiation of surrounding healthy tissues. Another way is to improve the accuracy and consistency of the border delineation using MR imaging. However, in MRI, there is not a unique correspondence between pixel intensity and electron density. Poor imaging of bones and image distortions are the other disadvantages of using MRI for dose planning. To address these challenges, one approach is to use MRI-to-CT image fusion that enables using MRI-based delineated borders on CT images for radiotherapy planning. It has been shown in the literature that using MRI guidance for prostate EBRT planning could increase the accuracy and

المنارة للاستشارات
www.manaraa.com

repeatability of the planning [62, 97, 98]. There are also studies on MRI-only radiotherapy planning methods in which CT imaging has been omitted [99]. To estimate the electron density information of the tissues, the MR image is segmented and different values are assigned to different regions based on the characteristics of the tissues. The MRI-only methods overcome the image registration and fusion errors. However, the accuracy of dosimetry is affected due to lack of accurate electron density information within the tissues.

1.6.3 MRI-guided biopsy and focal therapy

The increasing potential of MRI for diagnosis, localisation, and staging of PCa has driven the development of diagnostic and therapeutic devices that are compatible with the MRI magnetic field and imaging approach and can be used inside the MRI bore. MRI-guided biopsy [75] and MRI-guided focal therapy [100] are two examples of MRI-guided procedures in which MRI-compatible devices are used. In these procedures, the traditional intra-procedure imaging modality is replaced by MRI to increase the accuracy of the procedures by avoiding image fusion and registration errors. This comes with the compromise of increased cost of the procedure and awkward patient positioning issues due to the confines of the MRI bore.

1.7 Prostate contouring on MRI

Delineation of the prostate gland on MRI plays an important role in some diagnostic and therapeutic procedures. It helps to define the anatomy of the organ and to measure its volume. Measurement of the volume is useful for diagnosis and treatment planning. For example, the PSA level is usually interpreted in the context of prostate

المنارة للاستشارات
www.manaraa.com

volume [101]. Contouring of the prostate on MRI could be also helpful either in planning and delivering an MRI-guided therapy, or in the fusion of MR images to other imaging modalities (e.g. CT or ultrasound) for running an MRI-targeted image-guided process.

However, there are uncertainties and challenges around manual contouring of the prostate on MRI, described below.

1.7.1 Challenges in manual prostate contouring in MRI

Accuracy and reproducibility: The prostate is a soft tissue organ that is surrounded by other soft tissue structures such as the bladder, seminal vesicles, muscles, NVBs, and penile bulb. It has been shown using histology that the prostate is not a fully encapsulated gland, and the adjacent tissues in some parts are blended with the periprostatic tissues [102]. Thus, for some portions of the prostate, there does not exist a discrete, “true” boundary, even when viewed under the microscope. This poses challenges to prostate boundary delineation on medical imaging, rendering manual contouring a challenging task that is subject to relatively high intra- and inter-observer variability [9, 103]. This variability is even higher within some parts of the prostate such as base and apex regions [103]. This contouring variability could potentially influence the outcomes of clinical procedures, and also could cause a lack of performance consistency of a similar procedure between different clinical centres in multi-centre trials [1]. Therefore, any approach that helps to reduce this variability and improve the reproducibility of the task could be helpful from clinical point of view.

Timing: Contouring time is another issue with manual contouring of the prostate on MRI. There are several reports that report manual prostate contouring times, with the

المنارة للاستشارات
www.manaraa.com

average contouring time for the whole prostate in 3D varying between five minutes [10] up to approximately 20 minutes [11].

1.7.2 Computer-assisted prostate segmentation on MRI

In some clinical applications, computer-assisted contouring of the images (also called image segmentation) can provide more accurate and reproducible results in a shorter time. Segmentation is an image processing method in which the image usually is divided into two non-overlapped homogeneous regions with respect to some image characteristics such as intensity or texture [104]. One region is the region of interest (ROI) or object and the other is the background.

There are different types of approaches available for image segmentation in medical imaging. Segmentation algorithms work based on the features that are extracted from the image; e.g. image intensities, textures, intensity gradients or edges [105]. Some methods like thresholding and pixel clustering are based on pixel classificaltion and some others could be based on edge, boundary or shape detection. Sometimes a combination of multiple image-derived features is used to segment an image. There is also a group of segmentation methods that segment an image based on prior knowledge about image structure and characteristics obtained from a training image set.

Segmentation algorithms are usually designed or modified to optimize the result for specific applications. There are several presented image segmentation algorithms available in the literature for prostate segmentation in MRI, as described in a recent survey [106]. These algorithms have been developed to make the image contouring either faster, more accurate and/or more repeatable.

المنارة للاستشارات
www.manaraa.com

1.7.2.1 User interaction

There are two types of segmentation algorithm; semi-automatic and automatic. In semi-automatic segmentation, some operator interaction is required. Interaction allows for incorporation of the operator’s domain knowledge into the process of the image segmentation. Usually, operator interaction improves the accuracyof the algorithm and makes the algorithm more robust, but it could make the algorithm laborious and time-consuming to use. In automatic segmentation, the computer segments the image with no operator interaction required. However, automatic segmentation algorithms usually require parameter tuning by a user for initialization [104].

1.7.2.2 Prostate MRI segmentation challenges

As explained earlier, using the ER coil improves MR image quality from a clinical point of view, but can render computer-assisted segmentation more challenging due to the higher contrast within the prostate that reveals many details and edges that are not pertinent to the prostate boundary itself. Segmentation on ER MRI is also challenged by intensity inhomogeneity artifacts [85] and other artifacts as described in subsection 1.5.3.3. Thus, prostate segmentation on ER MRI is a substantially different problem, compared to prostate segmentation on MRI acquired with a body coil.

1.7.2.3 Prostate ER MRI segmentation techniques

There are several techniques have been presented in the literature for segmentation of the prostate on T2w MRI acquired with an ER coil. Martin *et al.* [13] presented a semi-automatic atlas-based method using intensity information combined with few landmarks to register an atlas to a test image. They evaluated their algorithm within different ROIs, including the midgland, base and apex, using a distance-based

المنارة للاستشارات

www.manaraa.com

error metric, and for the whole gland using region-based metrics. They reported some difficulties using atlas registration for small prostates with volume less than 25 cm^3 that resulted in higher segmentation errors. Vikal *et al.* [14] utilized shape modeling for a slice-by-slice 3D segmentation of the prostate on T2w MRI. Their semi-automatic method needed one centre point for initialisation and each slice segmentation was used as the initialisation for the segmentation of the next slice. They evaluated their method on three T2w ER MR images acquired at 3.0 Tesla using the MAD and DSC metrics to measure performance. Toth and Madabhushi [15] presented a semi-automatic segmentation method using a landmark-free active appearance model. They used a level set-based shape representation for their method. The method has been evaluated using the MAD and DSC error metrics selectively for different ROIs. Liao *et al.* [16] presented a hierarchical automatic segmentation using a multi-atlas-based method for coarse segmentation of the target image followed by a semisupervised regularization for the final fine segmentation. They evaluated their method on 66 T2w MR images using MAD, DSC, and Hausdorff distance (HD) metrics for the whole gland. Cheng *et al.* [107] presented an automatic atlas-based approach for T2w prostate MRI segmentation. Their algorithm is a slice-by-slice segmentation in which first an adaptive active appearance model is used to provide an initial coarse segmentation and then a support vector machine-based approach is used to refine the segmentation. Their evaluated their method using region based metrics on the whole gland.

In 2012, the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference held a prostate MR image segmentation (PROMISE12) challenge in which 11 teams were involved. The challenge evaluated the prostate T2w algorithms

المنارة للاستشارات
www.manaraa.com

presented by the teams and compared their performance in two parts; an online challenge and a live challenge. The data set contained both ER and non-ER MR images. DSC, MAD, 95% HD, and the percentage of the relative volume difference metrics were used to evaluate the algorithms. The metrics were applied to the whole gland, as well as the base and apex regions separately. PROMISE12 is a valuable study that measured and compared the segmentation errors of different state-of-the-art methods using the same data set to test and a single reference to evaluate [108].

Alvarez *et al.* [109] presented an automatic segmentation method for T2w prostate and tested their algorithm on 50 images from the PROMISE12 data set, including 24 ER MR images. In their method, for each test image a subset of similar training images are selected using a multi-scale analysis, and then the segmentation labels from the training images are registered to the test image and locally combined using a patch-based approach. Their results were sensitive to the number of atlases used and the size of the patches. They used the DSC measured on the whole gland to evaluate their method against a manual reference segmentation. Table 1.3 provides a high-level comparison of all of these approaches.

Table 1.3 gives a brief overview on all the mentioned segmentation methods. Although there are several segmentation algorithms available in the literature for which the segmentation accuracy is asymptotically approaching the observed range of differences between experts in manual segmentation, there remain some important limitations. For example, for some of the techniques the complexity of the algorithms is high. This complexity resulted in longer computational time ([15, 16]) compared to the methods with less complexity, but did not make a meaningful difference in segmentation

المنارة للاستشارات
www.manaraa.com

accuracy. Furthermore, some methods are not readily amenable to speed-up through parallel computing implementation.

1.7.3 Validation challenges

1.7.3.1 Lack of gold standard

The high inter-observer variability in manual contouring of the prostate MRI challenges the preparation a single gold-standard reference segmentation for each image. The absence of a reference to define the true extent of an object makes it difficult to validate the absolute accuracy of the contouring results [1]. The use of a single observer's manual contours as a reference standard thus complicates the interpretation of the results. Where different observers' reference standard segmentations were used in different studies, this poses a challenge to comparing different algorithms since differing results could be attributed to inter-observer variability in manual contouring rather than in true differences between algorithm performance. Combining a set of different contours from a group of experts as a consensus of opinion to make one reference standard is an approach to mitigate this issue and simplify validation. Simultaneous truth and performance level estimation (STAPLE) [110] is one of the most common approaches for combining a set of segmentation using a weighted voting scheme.

المنارة للاستشارات
www.manaraa.com

Table 1.3: A survey of prostate MRI segmentation algorithms.

Authors	***Techniques***	***Field strength*** *(Tesla)*	***ER coil***	***Validation regions***	***Validation metrics***	***Data set size***	***Number of references***
Martin *et al.* [13]	Atlas-based	X	yes	WG, A, MG, B	MAD	18	one
				WG	Recall, Precision		
Vikal *et al.* [14]	Shape modeling (slice-by-slice)	3.0	yes	A, MG, B	MAD, DSC	3	one based on two experts's agreement
Toth and Madabhushi [15]	Landmark-free active appearance model	3.0	yes	WG, A, MG, B	DSC	108	one for 108 images and two for a subset of 17 images
				WG	MAD, DSC		
Liao *et al.* [16]	Multi-atlas-based	X	X	WG	MAD, DSC, HD	66 (test) 9 (atlas)	one
Cheng *et al.* [107]	Atlas-based	3.0	X	WG	TP, FN, FP, DSC, ΔV%	100 (training) 40 (test)	one
Alvarez *et al.* [109]	Atlas-based	X	24 out of 50	WG	DSC	50	one

WG: whole gland, A: apex, MG: mid-gland, B: base

1.7.3.2 Lack of a standard validation methodology

Despite the lack of a straightforward gold standard, computer-assisted segmentation algorithms require validation to support clinical translation. This evaluation needs (1) a set of error metrics that are sensitive to different, clinically relevant types of contouring errors and (2) a method for evaluation of the contouring in different anatomic regions of interest within the prostate. The validation approach must take inter-observer variability in manual reference contours into account. To the best of our knowledge, there is no accepted standard set of error metrics use for evaluation of prostate contouring on medical imaging. Currently, most research groups have used one or two error metrics, and these choices have not generally been connected to any specific clinical procedures [1]. There are several classes of error metrics that have been used. In one class of metrics, the distances between corresponding points on the automatic and reference segmentations

www.manaraa.com

are calculated and aggregated (e.g. the MAD). In another class of metrics, the overlap region of two shapes or volumes is measured in various ways (e.g. the DSC). However, each metric is able to detect one or few types of errors but not all different types of errors; *e.g.* local surface misalignment, partial regional overlap, and volume difference. Therefore, comprehensive segmentation algorithm evaluation requires a set of complementary error metrics that covers the range of errors types that are relevant to clinical procedures of interest.

Furthermore, since the contouring is generally more challenging to perform (for both manual and automatic methods) at the inferior (apex) and superior (base) ends of the prostate, as compared to the midgland region, reporting the overall segmentation error for the whole prostate gland does not provide enough information about the local accuracy of the segmentation method under evaluation. Large errors in the base and apex can be compensated by small errors in the midgland, with an apparently favourable overall error reported that is discordant with large errors in the apex and base. Measuring segmentation errors separately within these different anatomic regions mitigates this issue. This helps the clinician to evaluate the readiness of an algorithm for clinical translation.

1.8 Hypothesis

The central hypothesis of this thesis is as follows: computer-assisted 3D prostate segmentation on T2w ER MRI will (1) decrease the time required for an expert physician to achieve a clinically acceptable segmentation, and (2) reduce inter-observer variability in segmentation, as compared to manual segmentation.

المنارة للاستشارات
www.manaraa.com

1.9 Objectives

To test the central hypothesis, the three major objectives of this thesis are:

I. **(Chapter 2)** To develop a semi-automatic prostate segmentation algorithm for T2w prostate ER MRI, and evaluate it against multi-observer manual reference standard segmentations.

II. **(Chapter 3)** To develop and evaluate a fully automated prostate segmentation algorithm for T2w prostate ER MRI, and evaluate it against multi-observer manual reference standard segmentation.

III. **(Chapter 4)** To measure the inter-observer variability and total segmentation time resulting from the use of the semi-automatic (Objective I) and automatic (Objective II) segmentation methods, followed by expert manual editing to yield clinically acceptable segmentations.

1.10 Thesis outline

1.10.1 Chapter 2 - Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods

The purpose of this work was to develop an approach for evaluation of a semi-automatic prostate segmentation algorithm for T2w MRI acquired with an ER coil and compare it to manual segmentation in terms of accuracy and repeatability within the whole gland, and separately within the apex, mid-gland, and base regions. We collected MR images from 42 prostate cancer patients. The prostate border was delineated manually by one observer on all images and by two other observers on a subset of 10

المنارة للاستشارات

www.manaraa.com

images. We used complementary boundary-, region-, and volume-based metrics to elucidate the different types of segmentation errors that we observed. Compared to manual segmentation, our semi-automatic approach reduced the necessary user interaction by only requiring an indication of the anteroposterior orientation of the prostate and the selection of prostate center points on the apex, base, and midgland slices. Based on these inputs, the algorithm identified the prostate boundary using learned boundary appearance characteristics and performed regularization based on learned prostate shape information.

In contrast with the active appearance model, our segmentation algorithm was based on local appearance characteristics. Furthermore, our algorithm optimized the segmentation first based on the appearance features and then further optimized based on shape features, rendering it more amenable to parallel computing implementation.

The algorithm required an average of 30 seconds of user interaction time for each 3D segmentation. Comparing the semi-automatic segmentations against a single-operator manual segmentation, the results of this chapter showed a MAD of 2.0 mm, DSC of 82%, recall of 77%, precision of 88%, and ΔV of = -4.6 cm^3 for the whole gland on average. We found that overall, midgland segmentation was more accurate and repeatable than the segmentation of the apex and base, with the base posing the greatest challenge. The semi-automatic approach reduced interobserver segmentation variability. Its accuracy, as well as the accuracies of recently published methods from other groups, were within the range of observed expert variability in manual segmentation. Further efforts in the development of computer-assisted segmentation would be most productive if focused on improvement of segmentation accuracy and reduction of variability within the prostatic apex and base.

www.manaraa.com

1.10.2 Chapter 3 - Accuracy and acceptability of an automated method for prostate segmentation in magnetic resonance imaging

In this chapter, we developed a fully automatic segmentation algorithm and evaluated its accuracy within different regions of interest (i.e. whole gland, apex, midgland, and base regions) using a complementary set of error metrics. We compared it to the semi-automatic approach (Chapter 2) and the inter-observer variability in manual segmentation. We used the same data set used in Chapter 2. In our automatic approach, we coarsely localized the prostate in the image using the prior measured dimensions of the gland that are readily available from the clinical TRUS examination before MRI acquisition. This localization is used to define the search space and to initialize the segmentation algorithm. Consequently, no user interaction is required for running the algorithm.

We evaluated the algorithm using a set of region- bouandary- and volume-based metrics; i.e., MAD, DSC, recall, precision and ΔV. We compared the accuracy of the automatic segmentation approach to the semi-automatic approach. We also compared the accuracy of both computer-assisted approaches to the range of inter-observer variation in manual segmentation.

The automatic algorithm needed less than a minute to segment the prostate in 3D. Comparing the segmentation results to single-observer manual segmentation, for the whole gland we measured a MAD of 3.2 mm, DSC of 71%, recall of 69%, precision of 76%, and ΔV of -3.6 cm^3. In a multi-observer study, we measured a MAD of 3 mm, DSC of 72%, recall of 74%, precision of 74%, and ΔV of -0.3 cm^3, whereas the difference between two observers' manual segmentations were as high as MAD of 2.8 mm, DSC of

المنارة للاستشارات

www.manaraa.com

74%, recall of 87%, precision of 60%, and |ΔV| of 18.3 cm^3. The results of the comparison of semi-automatic and automatic segmentation algorithm performance were mixed. Overall, the previously presented semi-automatic approach outperformed the automatic approach in terms of most of the metrics within some of prostatic regions. However, there were some metrics such as recall and ΔV that revealed superior performance from the automatic approach on some prostatic regions, compared to semi-automatic segmentation.

The results of this chapter show that (1) concordant with results from other published algorithms, accuracy was highest in the mid-gland and lower in the apex and base regions of the prostate, (2) the fully automatic approach requires no user interaction and needs 3 seconds of computation time, (3) the differences between the automatic and semi-automatic segmentation error metrics were consistently smaller than the differences observed between manual contours performed by expert observers, (4) The segmentation error metric values were near to or within the range of expert manual segmentation variability for most of the metrics at most of the prostatic regions.

1.10.3 Chapter 4 - Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging

Segmentation of the prostate gland on T2w MRI is an important part of several diagnostic and therapeutic procedures for PCa. Since manual segmentation is time-consuming and subject to high inter-expert operator variability, it has been widely recognized that these clinical procedures could benefit from a rapid and repeatable computer-assisted prostate segmentation technique. Many such algorithms have been

المنارة للاستشارات
www.manaraa.com

proposed in the literature [13-16, 107, 109], usually evaluated against manual reference segmentations performed by a single operator, with reported error metric values for recently published methods asymptotically approaching inter-operator variability in manual segmentation. Despite the tremendous volume of work performed in this area, the translation of computer-assisted segmentation algorithms to clinical care is rare, and manual segmentation is still routinely performed in clinic. As a step toward addressing this issue, in this chapter we focused on measuring the suitability of computer-assisted segmentation algorithms for clinical translation, based on measurements of inter-operator segmentation variability (which contributes to consistency of patient care) and measurements of the segmentation editing time required to yield clinically acceptable segmentations (which contributes to physician affinity to uptake of new segmentation tools, and patient throughput). We performed a pilot study with five expert operators under three pre- and post-editing conditions: manual segmentation, semi-automatic segmentation, and fully automatic segmentation. As expected, the results of this chapter showed that the amount of editing performed by the operators was directly related to the amount of automation involved in producing the starting segmentations. The provision of a starting segmentation using computer-assisted techniques reduced editing time and post-editing inter-operator variability, compared to manual segmentation. The amount of editing time was not correlated with the values of typically used segmentation error metrics such as the MAD between boundaries or the DSC, implying that the necessary post-segmentation editing time needs to be measured directly for multiple operators in order to evaluate an algorithm's suitability for clinical translation.

المنارة للاستشارات
www.manaraa.com

Chapter 5 will conclude with a summary of the advances in knowledge stemming from this thesis work. This chapter also discusses the practical applications of this work and potential directions for future research.

1.11 References

1. M. G. Jameson, L. C. Holloway, P. J. Vial, S. K. Vinod and P. E. Metcalfe, "A review of methods of analysis in contouring studies for radiation oncology," J Med Imaging Radiat Oncol **54**, 401-410 (2010).

2. F. A. Jolesz, A. Nabavi and R. Kikinis, "Integration of interventional MRI with computer-assisted surgery," J Magn Reson Imaging **13**, 69-77 (2001).

3. B. N. Bloch, R. E. Lenkinski and N. M. Rofsky, "The role of magnetic resonance imaging (MRI) in prostate cancer imaging and staging at 1.5 and 3 Tesla: the Beth Israel Deaconess Medical Center (BIDMC) approach," Cancer Biomark **4**, 251-262 (2008).

4. H. Hricak, P. L. Choyke, S. C. Eberhardt, S. A. Leibel and P. T. Scardino, "Imaging prostate cancer: a multidisciplinary perspective," Radiology **243**, 28-53 (2007).

5. M. Fuchsjager, A. Shukla-Dave, O. Akin, J. Barentsz and H. Hricak, "Prostate cancer imaging," Acta Radiol **49**, 107-120 (2008).

6. S. W. Heijmink, J. J. Futterer, T. Hambrock, S. Takahashi, T. W. Scheenen, H. J. Huisman, C. A. Hulsbergen-Van de Kaa, B. C. Knipscheer, L. A. Kiemeney, J. A. Witjes and J. O. Barentsz, "Prostate cancer: body-array versus endorectal coil MR imaging at 3 T--comparison of image quality, localization, and staging performance," Radiology **244**, 184-195 (2007).

7. Y. Kim, I. C. Hsu, J. Pouliot, S. M. Noworolski, D. B. Vigneron and J. Kurhanewicz, "Expandable and rigid endorectal coils for prostate MRI: impact on prostate distortion and rigid image registration," Med Phys **32**, 3569-3578 (2005).

8. M. Hirose, A. Bharatha, N. Hata, K. H. Zou, S. K. Warfield, R. A. Cormack, A. D'Amico, R. Kikinis, F. A. Jolesz and C. M. Tempany, "Quantitative MR imaging assessment of prostate gland deformation before and during MR imaging-guided brachytherapy," Acad Radiol **9**, 906-912 (2002).

9. W. L. Smith, C. Lewis, G. Bauman, G. Rodrigues, D. D'Souza, R. Ash, D. Ho, V. Venkatesan, D. Downey and A. Fenster, "Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR," Int J Radiat Oncol Biol Phys **67**, 1238-1247 (2007).

المنارة للاستشارات
www.manaraa.com

10. S. Martin, G. Rodrigues, N. Patil, G. Bauman, D. D'Souza, T. Sexton, D. Palma, A. V. Louie, F. Khalvati, H. R. Tizhoosh and S. Gaede, "A multiphase validation of atlas-based automatic and semiautomatic segmentation strategies for prostate MRI," Int J Radiat Oncol Biol Phys **85**, 95-100 (2013).

11. N. Makni, P. Puech, R. Lopes, A. S. Dewalle, O. Colot and N. Betrouni, "Combining a deformable model and a probabilistic framework for an automatic 3D segmentation of prostate on MRI," Int J Comput Assist Radiol Surg **4**, 181-188 (2009).

12. J. J. Battista, C. Johnson, D. Turnbull, J. Kempe, K. Bzdusek, J. Van Dyk and G. Bauman, "Dosimetric and radiobiological consequences of computed tomography-guided adaptive strategies for intensity modulated radiation therapy of the prostate," Int J Radiat Oncol Biol Phys **87**, 874-880 (2013).

13. S. Martin, V. Daanen and J. Troccaz, "Atlas-based prostate segmentation using an hybrid registration," Int J CARS **3**, 8 (2008).

14. S. Vikal, S. Haker, C. Tempany and G. Fichtinger, "Prostate contouring in MRI guided biopsy," Proc SPIE **7259**, 72594A (2009).

15. R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: application to prostate MRI segmentation," IEEE Trans Med Imaging **31**, 1638-1650 (2012).

16. S. Liao, Yaozong Gao, Yinghuan Shi, Ambereen Yousuf, Ibrahim Karademir, Aytekin Oto, and Dinggang Shen, "Automatic prostate MR image segmentation with sparse label propagation and domain-specific manifold regularization," Information Processing in Medical Imaging, 511-523 (2013).

17. P. Carroll and G. Grossfeld, "American Cancer Society Atlas of Clinical Oncology: Prostate Cancer," (Hamilton, ON: BC Decker Inc, 2002).

18. R. L. Siegel, K. D. Miller and A. Jemal, "Cancer statistics, 2015," CA Cancer J Clin **65**, 5-29 (2015).

19. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2015*. 2015

20. T. Navaneelan and T. Janz, *Cancer in Canada: focus on lung, colorectal, breast and prostate*. (Statistics Canada, 2011).

21. P. Tenke, J. Horti, P. Balint and B. Kovacs, "Prostate cancer screening," Recent Results Cancer Res **175**, 65-81 (2007).

المنارة للاستشارات

www.manaraa.com

22. R. Etzioni, R. Cha, E. J. Feuer and O. Davidov, "Asymptomatic incidence and duration of prostate cancer," Am J Epidemiol **148**, 775-785 (1998).

23. M. Djulbegovic, R. J. Beyth, M. M. Neuberger, T. L. Stoffs, J. Vieweg, B. Djulbegovic and P. Dahm, "Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials," BMJ **341**, c4543 (2010).

24. H. G. Welch, *Should I be tested for cancer?: maybe not and here's why*. (Univ of California Press, 2004).

25. I. W. van der Cruijsen-Koeter, M. J. Roobol, M. F. Wildhagen, T. H. van der Kwast, W. J. Kirkels and F. H. Schroder, "Tumor characteristics and prognostic factors in two subsequent screening rounds with four-year interval within prostate cancer screening trial, ERSPC Rotterdam," Urology **68**, 615-620 (2006).

26. G. Aus, S. Bergdahl, P. Lodding, H. Lilja and J. Hugosson, "Prostate cancer screening decreases the absolute risk of being diagnosed with advanced prostate cancer--results from a prospective, population-based randomized controlled trial," Eur Urol **51**, 659-664 (2007).

27. A. Bill-Axelson, L. Holmberg, F. Filen, M. Ruutu, H. Garmo, C. Busch, S. Nordling, M. Haggman, S. O. Andersson, S. Bratell, A. Spangberg, J. Palmgren, H. O. Adami and J. E. Johansson, "Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial," J Natl Cancer Inst **100**, 1144-1154 (2008).

28. F. H. Schroder, P. van der Maas, P. Beemsterboer, A. B. Kruger, R. Hoedemaeker, J. Rietbergen and R. Kranse, "Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer," J Natl Cancer Inst **90**, 1817-1823 (1998).

29. K. Mistry and G. Cable, "Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma," J Am Board Fam Pract **16**, 95-101 (2003).

30. W. J. Catalona, J. P. Richie, F. R. Ahmann, M. A. Hudson, P. T. Scardino, R. C. Flanigan, J. B. deKernion, T. L. Ratliff, L. R. Kavoussi, B. L. Dalkin and et al., "Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men," J Urol **151**, 1283-1290 (1994).

31. O. T. Okotie, K. A. Roehl, M. Han, S. Loeb, S. N. Gashti and W. J. Catalona, "Characteristics of prostate cancer detected by digital rectal examination only," Urology **70**, 1117-1120 (2007).

المنارة للاستشارات

www.manaraa.com

32. A. M. Wolf, R. C. Wender, R. B. Etzioni, I. M. Thompson, A. V. D'Amico, R. J. Volk, D. D. Brooks, C. Dash, I. Guessous, K. Andrews, C. DeSantis and R. A. Smith, "American Cancer Society guideline for the early detection of prostate cancer: update 2010," CA Cancer J Clin **60**, 70-98 (2010).

33. J. C. Presti, "Prostate biopsy: current status and limitations," Rev Urol **9**, 93-98 (2007).

34. P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," Mod Pathol **17**, 292-306 (2004).

35. K. A. Roehl, J. A. Antenor and W. J. Catalona, "Serial biopsy results in prostate cancer screening study," J Urol **167**, 2435-2439 (2002).

36. B. Djavan, V. Ravery, A. Zlotta, P. Dobronski, M. Dobrovits, M. Fakhari, C. Seitz, M. Susani, A. Borkowski, L. Boccon-Gibod, C. C. Schulman and M. Marberger, "Prospective evaluation of prostate cancer detected on biopsies 1, 2, 3 and 4: when should we stop?," J Urol **166**, 1679-1683 (2001).

37. D. F. Gleason, "Classification of prostatic carcinomas," Cancer chemotherapy reports. Part 1 **50**, 125-128 (1966).

38. S. B. Edge, D. R. Byrd, C. C. Compton, A. G. Fritz, F. L. Greene and A. Trotti, *AJCC cancer staging manual*. (Springer New York, 2010).

39. M. Norberg, M. Häggman, T. Andersson, C. Busch and A. Magnusson, "Evaluation of localised prostate cancer: a comparative study of transrectal ultrasonography versus histopathology," European Radiology **3**, 166-173 (1993).

40. P. Narayan, V. Gajendran, S. P. Taylor, A. Tewari, J. C. Presti, Jr., R. Leidich, R. Lo, K. Palmer, K. Shinohara and J. T. Spaulding, "The role of transrectal ultrasound-guided biopsy-based staging, preoperative serum prostate-specific antigen, and biopsy Gleason score in prediction of final pathologic diagnosis in prostate cancer," Urology **46**, 205-212 (1995).

41. K. Shinohara, T. M. Wheeler and P. T. Scardino, "The appearance of prostate cancer on transrectal ultrasonography: correlation of imaging and pathological examinations," J Urol **142**, 76-82 (1989).

42. J. Mohler, R. R. Bahnson, B. Boston, J. E. Busby, A. D'Amico, J. A. Eastham, C. A. Enke, D. George, E. M. Horwitz, R. P. Huben, P. Kantoff, M. Kawachi, M. Kuettel, P. H. Lange, G. Macvicar, E. R. Plimack, J. M. Pow-Sang, M. Roach, 3rd, E. Rohren, B. J. Roth, D. C. Shrieve, M. R. Smith, S. Srinivas, P. Twardowski and P. C. Walsh, "NCCN clinical practice guidelines in oncology: prostate cancer," J Natl Compr Canc Netw **8**, 162-200 (2010).

المنارة للاستشارات
www.manaraa.com

43. J. Mohler, A. Armstrong and R. Bahnson, "NCCN clinical practice guidelines for prostate cancer [Internet]," (2015).

44. J. L. Mohler, P. W. Kantoff, A. J. Armstrong, R. R. Bahnson, M. Cohen, A. V. D'Amico, J. A. Eastham, C. A. Enke, T. A. Farrington, C. S. Higano, E. M. Horwitz, C. J. Kane, M. H. Kawachi, M. Kuettel, T. M. Kuzel, R. J. Lee, A. W. Malcolm, D. Miller, E. R. Plimack, J. M. Pow-Sang, D. Raben, S. Richey, M. Roach, 3rd, E. Rohren, S. Rosenfeld, E. Schaeffer, E. J. Small, G. Sonpavde, S. Srinivas, C. Stein, S. A. Strope, J. Tward, D. A. Shead and M. Ho, "Prostate cancer, version 2.2014," J Natl Compr Canc Netw **12**, 686-718 (2014).

45. M. Bolla, H. van Poppel, L. Collette, P. van Cangh, K. Vekemans, L. Da Pozzo, T. M. de Reijke, A. Verbaeys, J. F. Bosset, R. van Velthoven, J. M. Marechal, P. Scalliet, K. Haustermans and M. Pierart, "Postoperative radiotherapy after radical prostatectomy: a randomised controlled trial (EORTC trial 22911)," Lancet **366**, 572-578 (2005).

46. G. Steineck, F. Helgesen, J. Adolfsson, P. W. Dickman, J.-E. Johansson, B. J. Norlén and L. Holmberg, "Quality of life after radical prostatectomy or watchful waiting," New England Journal of Medicine **347**, 790-796 (2002).

47. P. C. Walsh, P. Marschke, D. Ricker and A. L. Burnett, "Patient-reported urinary continence and sexual function after anatomic radical prostatectomy," Urology **55**, 58-61 (2000).

48. J. B. Madalinska, M. L. Essink-Bot, H. J. de Koning, W. J. Kirkels, P. J. van der Maas and F. H. Schroder, "Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer," J Clin Oncol **19**, 1619-1628 (2001).

49. M. G. Sanda, R. L. Dunn, J. Michalski, H. M. Sandler, L. Northouse, L. Hembroff, X. Lin, T. K. Greenfield, M. S. Litwin, C. S. Saigal, A. Mahadevan, E. Klein, A. Kibel, L. L. Pisters, D. Kuban, I. Kaplan, D. Wood, J. Ciezki, N. Shah and J. T. Wei, "Quality of life and satisfaction with outcome among prostate-cancer survivors," N Engl J Med **358**, 1250-1261 (2008).

50. E. Berthelet, M. C. Liu, A. Agranovich, K. Patterson and T. Currie, "Computed tomography determination of prostate volume and maximum dimensions: a study of interobserver variability," Radiother Oncol **63**, 37-40 (2002).

51. X. Gual-Arnau, M. V. Ibanez-Gual, F. Lliso and S. Roldan, "Organ contouring for prostate cancer: interobserver and internal organ motion variability," Comput Med Imaging Graph **29**, 639-647 (2005).

52. S. J. Frank, L. L. Pisters, J. Davis, A. K. Lee, R. Bassett and D. A. Kuban, "An assessment of quality of life following radical prostatectomy, high dose external beam

المنارة للاستشارات

www.manaraa.com

radiation therapy and brachytherapy iodine implantation as monotherapies for localized prostate cancer," J Urol **177**, 2151-2156; discussion 2156 (2007).

53. M. S. Litwin, J. L. Gore, L. Kwan, J. M. Brandeis, S. P. Lee, H. R. Withers and R. E. Reiter, "Quality of life after surgery, external beam irradiation, or brachytherapy for early-stage prostate cancer," Cancer **109**, 2239-2247 (2007).

54. D. C. Miller, M. G. Sanda, R. L. Dunn, J. E. Montie, H. Pimentel, H. M. Sandler, W. P. McLaughlin and J. T. Wei, "Long-term outcomes among localized prostate cancer survivors: health-related quality-of-life changes after radical prostatectomy, external radiation, and brachytherapy," Journal of Clinical Oncology **23**, 2772-2780 (2005).

55. B. J. Davis, E. M. Horwitz, W. R. Lee, J. M. Crook, R. G. Stock, G. S. Merrick, W. M. Butler, P. D. Grimm, N. N. Stone, L. Potters, A. L. Zietman and M. J. Zelefsky, "American Brachytherapy Society consensus guidelines for transrectal ultrasound-guided permanent prostate brachytherapy," Brachytherapy **11**, 6-19 (2012).

56. M. Rom, A. Pycha, C. Wiunig, A. Reissigl, M. Waldert, T. Klatte, M. Remzi and C. Seitz, "Prospective randomized multicenter study comparing prostate cancer detection rates of end-fire and side-fire transrectal ultrasound probe configuration," Urology **80**, 15-18 (2012).

57. C. B. Ching, A. S. Moussa, J. Li, B. R. Lane, C. Zippe and J. S. Jones, "Does transrectal ultrasound probe configuration really matter? End fire versus side fire probe prostate cancer detection rates," J Urol **181**, 2077-2082; discussion 2082-2073 (2009).

58. R. Hara, Y. Jo, T. Fujii, N. Kondo, T. Yokoyoma, Y. Miyaji and A. Nagai, "Optimal approach for prostate cancer detection as initial biopsy: prospective randomized study comparing transperineal versus transrectal systematic 12-core biopsy," Urology **71**, 191-195 (2008).

59. R. Paul, C. Korzinek, U. Necknig, T. Niesel, M. Alschibaja, H. Leyh and R. Hartung, "Influence of transrectal ultrasound probe on prostate cancer detection in transrectal ultrasound-guided sextant biopsy of prostate," Urology **64**, 532-536 (2004).

60. B. E. Weiss, A. J. Wein, S. B. Malkowicz and T. J. Guzzo, "Comparison of prostate volume measured by transrectal ultrasound and magnetic resonance imaging: is transrectal ultrasound suitable to determine which patients should undergo active surveillance?," Urol Oncol **31**, 1436-1440 (2013).

61. Z. Gao, D. Wilkins, L. Eapen, C. Morash, Y. Wassef and L. Gerig, "A study of prostate delineation referenced against a gold standard created from the visible human data," Radiother Oncol **85**, 239-246 (2007).

www.manaraa.com

62. C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk and J. V. Lebesque, "Definition of the prostate in CT and MRI: a multi-observer study," Int J Radiat Oncol Biol Phys **43**, 57-66 (1999).

63. J. Kurhanewicz, D. Vigneron, P. Carroll and F. Coakley, "Multiparametric magnetic resonance imaging in prostate cancer: present and future," Curr Opin Urol **18**, 71-77 (2008).

64. H. U. Ahmed, A. Kirkham, M. Arya, R. Illing, A. Freeman, C. Allen and M. Emberton, "Is it time to consider a role for MRI before prostate biopsy?," Nat Rev Clin Oncol **6**, 197-206 (2009).

65. A. Shukla-Dave and H. Hricak, "Role of MRI in prostate cancer detection," NMR Biomed **27**, 16-24 (2014).

66. M. L. Schiebler, M. D. Schnall, H. M. Pollack, R. E. Lenkinski, J. E. Tomaszewski, A. J. Wein, R. Whittington, W. Rauschning and H. Y. Kressel, "Current role of MR imaging in the staging of adenocarcinoma of the prostate," Radiology **189**, 339-352 (1993).

67. G. M. Villeirs and G. O. De Meerleer, "Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning," Eur J Radiol **63**, 361-368 (2007).

68. A. E. Wefer, H. Hricak, D. B. Vigneron, F. V. Coakley, Y. Lu, J. Wefer, U. Mueller-Lisse, P. R. Carroll and J. Kurhanewicz, "Sextant localization of prostate cancer: comparison of sextant biopsy, magnetic resonance imaging and magnetic resonance spectroscopic imaging with step section histology," J Urol **164**, 400-404 (2000).

69. M. D. Schnall, Y. Imai, J. Tomaszewski, H. M. Pollack, R. E. Lenkinski and H. Y. Kressel, "Prostate cancer: local staging with endorectal surface coil MR imaging," Radiology **178**, 797-802 (1991).

70. D. J. Gilderdale, N. M. deSouza, G. A. Coutts, M. K. Chui, D. J. Larkman, A. D. Williams and I. R. Young, "Design and use of internal receiver coils for magnetic resonance imaging," Br J Radiol **72**, 1141-1151 (1999).

71. J. Nakashima, A. Tanimoto, Y. Imai, M. Mukai, Y. Horiguchi, K. Nakagawa, M. Oya, T. Ohigashi, K. Marumo and M. Murai, "Endorectal MRI for prediction of tumor site, tumor size, and local extension of prostate cancer," Urology **64**, 101-105 (2004).

72. A. C. Westphalen, D. A. McKenna, J. Kurhanewicz and F. V. Coakley, "Role of magnetic resonance imaging and magnetic resonance spectroscopic imaging before and after radiotherapy for prostate cancer," J Endourol **22**, 789-794 (2008).

المنارة للاستشارات

www.manaraa.com

73. P. R. Carroll, F. V. Coakley and J. Kurhanewicz, "Magnetic resonance imaging and spectroscopy of prostate cancer," Rev Urol **8 Suppl 1**, S4-S10 (2006).

74. O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino and H. Hricak, "Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging," Radiology **239**, 784-792 (2006).

75. R. J. Stafford, S. E. McRae and K. Ahrar, "MRI-Guided Prostate Biopsy," in *Percutaneous Image-Guided Biopsy,* (Springer, 2014), pp. 297-311.

76. A. Graser, A. Heuck, B. Sommer, J. Massmann, J. Scheidler, M. Reiser and U. Mueller-Lisse, "Per-sextant localization and staging of prostate cancer: correlation of imaging findings with whole-mount step section histopathology," AJR Am J Roentgenol **188**, 84-90 (2007).

77. C. K. Kim, B. K. Park and B. Kim, "Localization of prostate cancer using 3T MRI: comparison of T2-weighted and dynamic contrast-enhanced imaging," J Comput Assist Tomogr **30**, 7-11 (2006).

78. S. W. Heijmink, J. J. Futterer, S. S. Strum, W. J. Oyen, F. Frauscher, J. A. Witjes and J. O. Barentsz, "State-of-the-art uroradiologic imaging in the diagnosis of prostate cancer," Acta Oncol **50 Suppl 1**, 25-38 (2011).

79. N. Hara, M. Okuizumi, H. Koike, M. Kawaguchi and V. Bilim, "Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer," Prostate **62**, 140-147 (2005).

80. P. Kozlowski, S. D. Chang, E. C. Jones, K. W. Berean, H. Chen and S. L. Goldenberg, "Combined diffusion-weighted and dynamic contrast-enhanced MRI for prostate cancer diagnosis--correlation with biopsy and histopathology," J Magn Reson Imaging **24**, 108-113 (2006).

81. R. Bammer, B. Acar and M. E. Moseley, "In vivo MR tractography using diffusion imaging," Eur J Radiol **45**, 223-234 (2003).

82. A. Abdellaoui, S. Iyengar and S. Freeman, "Imaging in prostate cancer," Future Oncol **7**, 679-691 (2011).

83. S. K. Gujar, S. Maheshwari, I. Bjorkman-Burtscher and P. C. Sundgren, "Magnetic resonance spectroscopy," J Neuroophthalmol **25**, 217-226 (2005).

84. J. Kurhanewicz, D. B. Vigneron, S. J. Nelson, H. Hricak, J. M. MacDonald, B. Konety and P. Narayan, "Citrate as an in vivo marker to discriminate prostate cancer from benign prostatic hyperplasia and normal prostate peripheral zone: detection via localized proton spectroscopy," Urology **45**, 459-466 (1995).

المنارة للاستشارات

www.manaraa.com

85. S. M. Noworolski, J. C. Crane, D. B. Vigneron and J. Kurhanewicz, "A clinical comparison of rigid and inflatable endorectal-coil probes for MRI and 3D MR spectroscopic imaging (MRSI) of the prostate," J Magn Reson Imaging **27**, 1077-1082 (2008).

86. J. E. Husband, A. R. Padhani, A. D. MacVicar and P. Revell, "Magnetic resonance imaging of prostate cancer: comparison of image quality using endorectal and pelvic phased array coils," Clin Radiol **53**, 673-681 (1998).

87. S. M. Noworolski, G. D. Reed, J. Kurhanewicz and D. B. Vigneron, "Post-processing correction of the endorectal coil reception effects in MR spectroscopic imaging of the prostate," J Magn Reson Imaging **32**, 654-662 (2010).

88. J. J. Futterer, M. R. Engelbrecht, G. J. Jager, R. P. Hartman, B. F. King, C. A. Hulsbergen-Van de Kaa, J. A. Witjes and J. O. Barentsz, "Prostate cancer: comparison of local staging accuracy of pelvic phased-array coil alone versus integrated endorectal-pelvic phased-array coils. Local staging accuracy of prostate cancer using endorectal coil MR imaging," Eur Radiol **17**, 1055-1065 (2007).

89. H. Hricak, S. White, D. Vigneron, J. Kurhanewicz, A. Kosco, D. Levin, J. Weiss, P. Narayan and P. R. Carroll, "Carcinoma of the prostate gland: MR imaging with pelvic phased-array coils versus integrated endorectal--pelvic phased-array coils," Radiology **193**, 703-709 (1994).

90. M. D. Schnall, R. E. Lenkinski, H. M. Pollack, Y. Imai and H. Y. Kressel, "Prostate: MR imaging with an endorectal surface coil," Radiology **172**, 570-574 (1989).

91. B. K. Park, B. Kim, C. K. Kim, H. M. Lee and G. Y. Kwon, "Comparison of phased-array 3.0-T and endorectal 1.5-T magnetic resonance imaging in the evaluation of local staging accuracy for prostate cancer," J Comput Assist Tomogr **31**, 534-538 (2007).

92. P. Oehr and K. Bouchelouche, "Imaging of prostate cancer," Curr Opin Oncol **19**, 259-264 (2007).

93. H. Jadvar, "Molecular imaging of prostate cancer: PET radiotracers," AJR Am J Roentgenol **199**, 278-291 (2012).

94. S. Xu, J. Kruecker, B. Turkbey, N. Glossop, A. K. Singh, P. Choyke, P. Pinto and B. J. Wood, "Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies," Comput Aided Surg **13**, 255-264 (2008).

95. D. W. Cool, J. Bax, C. Romagnoli, A. D. Ward, L. Gardi, V. Karnik, J. Izawa, J. Chin and A. Fenster, "Fusion of MRI to 3D TRUS for mechanically-assisted targeted prostate biopsy: system design and initial clinical experience," in *Prostate Cancer*

المنارة للاستشارات

www.manaraa.com

Imaging. Image Analysis and Image-Guided Interventions, (Springer, 2011), pp. 121-133.

96. E. Baco, E. Rud, O. Ukimura, L. Vlatkovic, A. Svindland, T. Matsugasumi, J. C. Bernhard, J. C. Rewcastle and H. B. Eggesbo, "Effect of targeted biopsy guided by elastic image fusion of MRI with 3D-TRUS on diagnosis of anterior prostate cancer," Urol Oncol **32**, 1300-1307 (2014).

97. K. Kagawa, W. R. Lee, T. E. Schultheiss, M. A. Hunt, A. H. Shaer and G. E. Hanks, "Initial clinical assessment of CT-MRI image fusion software in localization of the prostate for 3D conformal radiation therapy," Int J Radiat Oncol Biol Phys **38**, 319-325 (1997).

98. G. M. Villeirs, K. Van Vaerenbergh, L. Vakaet, S. Bral, F. Claus, W. J. De Neve, K. L. Verstraete and G. O. De Meerleer, "Interobserver delineation variation using CT versus combined CT + MRI in intensity-modulated radiotherapy for prostate cancer," Strahlenther Onkol **181**, 424-430 (2005).

99. M. Kapanen, J. Collan, A. Beule, T. Seppala, K. Saarilahti and M. Tenhunen, "Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate," Magn Reson Med **70**, 127-135 (2013).

100. O. Raz, M. A. Haider, S. R. Davidson, U. Lindner, E. Hlasny, R. Weersink, M. R. Gertner, W. Kucharczyk, S. A. McCluskey and J. Trachtenberg, "Real-time magnetic resonance imaging-guided focal laser therapy in patients with low-risk prostate cancer," Eur Urol **58**, 173-177 (2010).

101. A. W. Partin, W. J. Catalona, P. C. Southwick, E. N. Subong, G. H. Gasior and D. W. Chan, "Analysis of percent free prostate-specific antigen (PSA) for prostate cancer detection: influence of total PSA, prostate volume, and age," Urology **48**, 55-61 (1996).

102. A. G. Ayala, J. Y. Ro, R. Babaian, P. Troncoso and D. J. Grignon, "The prostatic capsule: does it exist? Its importance in the staging and treatment of prostatic carcinoma," Am J Surg Pathol **13**, 21-27 (1989).

103. S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring and J. P. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information," Med Phys **35**, 1407-1417 (2008).

104. D. L. Pham, C. Xu and J. L. Prince, "Current methods in medical image segmentation," Annu Rev Biomed Eng **2**, 315-337 (2000).

105. L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M. Vaidyanathan, L. O. Hall, R. W. Thatcher and M. L. Silbiger, "MRI segmentation: methods and applications," Magn Reson Imaging **13**, 343-368 (1995).

www.manaraa.com

106. S. Ghose, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibe and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," Comput Methods Programs Biomed **108**, 262-287 (2012).

107. R. Cheng, B. Turkbey, W. Gandler, H. K. Agarwal, V. P. Shah, A. Bokinsky, E. McCreedy, S. Wang, S. Sankineni, M. Bernardo, T. Pohida, P. Choyke and M. J. McAuliffe, "Atlas based AAM and SVM model for fully automatic MRI prostate segmentation," Conf Proc IEEE Eng Med Biol Soc **2014**, 2881-2885 (2014).

108. G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," Med Image Anal **18**, 359-373 (2014).

109. C. Álvarez, F. Martínez and E. Romero, "An automatic multi-atlas prostate segmentation in MRI using a multiscale representation and a label fusion strategy," in *Tenth International Symposium on Medical Information Processing and Analysis*, (International Society for Optics and Photonics, 2015), pp. 92870D-92870D-92875.

110. S. K. Warfield, K. H. Zou and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans Med Imaging **23**, 903-921 (2004).

المنارة للاستشارات
www.manaraa.com

Chapter 2.

Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods[†]

2.1 Introduction

PCa is the most common non-cutaneous cancer and was the second leading cause of cancer death among North American men in 2012 [1]. Three-dimensional (3D) prostate segmentation in medical images is useful to the planning of diagnosis and therapy procedures [2, 3]. Recent developments in magnetic resonance imaging (MRI) have demonstrated its usefulness for PCa detection and staging[4-6] with T2 weighted (T2w) MRI most commonly used for prostate boundary delineation due to its superior anatomic visualization [6]. Endorectal (ER) coil imaging provides improved image quality[4, 5, 7], but this coil induces substantial tissue deformation [8, 9] and the resulting higher contrast images contain more details and edges, presenting an increased challenge to segmentation algorithms designed for use on non-ER coil imaging. Manual segmentation of the prostate on MRI is a time-consuming task and is subject to

---

[†] A version of this chapter has been published: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, E. Gibson, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward, “Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semiautomated methods.” Medical Physics 41:11 (2014).

www.manaraa.com

substantial inter-observer variation [10], motivating the need for a fast and reproducible segmentation algorithm for 3D segmentation of the prostate on T2W ER MRI.

Several methods have been published in the literature for 3D segmentation of the prostate on T2W ER MRI. Martin et al. [11] presented a semi-automatic method based on the registration of an atlas to a test image using a combination of intensity-based and landmark-based methods, and evaluated it within different regions of interest including mid-gland, base and apex using a distance-based metric. They also used region-based evaluation for the whole gland. Vikal et al. [12] presented a semi-automatic slice-by-slice 3D method using a shape model, evaluated on 3 images using the mean absolute distance (MAD) and Dice similarity coefficient [13] (DSC). Toth et al. [14] used a semi-automatic multi-feature landmark-free active appearance model, and selectively used the MAD and DSC for evaluation of different anatomic regions. Liao et al. [15] presented an automatic multi-atlas-based segmentation method followed by a semi-supervised regularization. They evaluated their method using DSC, MAD and Hausdorff distance metrics on the whole gland. In 2012, a prostate MR image segmentation (PROMISE12) challenge was held as part of the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference and involved 11 teams. This challenge compared the performance of the teams' submitted prostate T2W MRI segmentation algorithms. It consisted of two main parts: an online challenge and a live challenge. The image data set in the challenge contained both ER and non-ER MR images, some acquired at 1.5 Tesla and some at 3.0 Tesla magnetic field strengths. Four experts each manually segmented 25 images out of 100 (i.e. each image was segmented by one expert and not more than one expert segmented each image). Next, an additional expert reviewed all the manual

المنارة للاستشارات
www.manaraa.com

segmentations and edited them for consistency as deemed necessary, yielding a single manual reference segmentation for accuracy measurement in each of the 100 cases (a second manual segmentation by an inexperienced non-clinical observer with two years' experience in prostate MRI research was used for a ranking score calculation based on the error metrics, but was not used in any inter-operator variability measurements). The algorithms' segmentations were compared with the manual reference using DSC, MAD, 95% Hausdorff distance, and the percentage of the relative volume difference as the metrics. These metrics were reported on the whole gland, as well as the base and apex regions separately. The results of the live challenge on 20 images showed a range of 2.0 mm to 4.2 mm for MAD, 65% to 89% for DSC, and 1.5% to 43.1% for absolute relative volume difference on whole gland [16].

The PROMISE study measured the performance of different segmentation methods on the same set of images with the same manual reference. By holding the images and reference segmentations constant and measuring the performance of different algorithms, this study provided highly valuable measurements of variability in segmentation errors arising from the use of different state-of-the-art algorithms. By contrast, our study holds the algorithm constant (i.e. we tested a single algorithm) and used a reference standard based on multiple operators, addressing the question of the accuracy and variability of a segmentation algorithm's results compared to inter-operator variability in the manual segmentation, such as one could observe in routine clinical practice. Thus, our study and the PROMISE study achieved complementary aims; in the future, a grand challenge-style study comparing different segmentation algorithms against a multiple-operator reference standard would be highly valuable.

المنارة للاستشارات
www.manaraa.com

In all of the reviewed published studies, the segmentation results were evaluated by comparison to a single-observer reference, with the inter-operator variability in manual segmentation and its effect on accuracy measurement not measured or taken into account in interpreting the segmentation results. In these studies, the accuracy was usually measured using one or at most two types of error metrics and none used a complementary set of error metrics capturing different types of errors such as surface/boundary misalignments, regional overlap errors, and volume differences. The use of a complementary set of error metrics is supportive of a comprehensive segmentation accuracy measurement, permitting the end user to focus on the metrics capturing performance aspects of importance to the user's intended application of the technique. In addition, several previous studies report on segmentation accuracy only for the prostate gland as a whole, without reporting on spatial variations in the error through gland sub regions such as the apex, mid-gland and base. It is well-known that for both human experts and contemporary algorithms, mid-gland segmentation is usually performed with lower error and variability compared to segmentation of the apex and base, which are considered to be more challenging tasks. Thus, reporting of segmentation error on a whole-gland basis alone challenges the interpretation of the segmentation results in the interventional context, where accurate apex and base segmentations are critical to sparing harm to surrounding critical structures.

To address the need for a 3D method, fully evaluated using a comprehensive set of metrics, we present here an interactive algorithm for 3D prostate segmentation on T2W ER MRI, based on learned local appearance of the prostate border and learned variability of prostate shape. We used a set of complementary boundary-based, overlap-

www.manaraa.com

based, and volume-based metrics to evaluate the segmentation over the whole gland and separately within each anatomic region of interest (prostate apex, mid-gland, and base). The method has two main steps, training and segmentation. During training, we captured the local image appearance of the prostate at the boundary on cross-sections from superior to inferior by computing a set of circular mean intensity image patches on each slice. The prostate shape variability on each axial slice was measured using a point distribution model (PDM) [17]. The segmentation algorithm requires minimal user input to initialize a radial-based search for candidate boundary points that are regularized using the PDM to produce the final result. The algorithm segmentations were validated against manual segmentations using complementary boundary-based (MAD), regional overlap (DSC, recall rate, and precision rate) and volume difference (ΔV) metrics, and inter-operator variability was measured.

2.2 Materials and Methods

2.2.1 Materials

We used 42 axial T2W fast spin echo ER MR images acquired as follows: 23 images with TR = 4000–13000 msec, TE = 156–164 msec, NEX = 2 and 19 images with TR = 3500–7320 msec, TE = 102–116 msec, NEX = 1–2. Some images were acquired at 1.5 Tesla (9 images) and some at 3.0 Tesla (33 images), with voxel sizes from 0.27×0.27×2.2 mm to 0.44×0.44×6 mm (covering a range of voxel sizes typically seen in clinical prostate MRI). The images were acquired using four different scanners: Signa Excite, Discovery MR 750 (General Electric Healthcare, Waukesha, WI, USA), MAGNETOM Avanto, and MAGNETOM Verio (Siemens Medical Solutions, Malvern, PA, USA). All of the images were acquired from patients diagnosed with PCa based on

www.manaraa.com

needle biopsy. The study was approved by the research ethics board of our institution, and written informed consent was obtained from all patients prior to enrolment. Each of the 42 images was segmented manually by one observer, with the segmentation subsequently reviewed and adjusted as deemed necessary by an expert radiology resident with experience in reading >100 prostate MRI cases. The initial manual segmentations were performed either by a radiologist or by a graduate student under the advisement of a radiologist. For inter-operator comparison, two additional observers (a radiologist and a radiation oncologist) each performed manual segmentations on a subset of 10 images to provide a total of three independent manual segmentations per patient. To select this subset of 10 images, we qualitatively assigned easy-, moderate- and difficult-to-segment labels to a set of images acquired at our institution, and randomly selected 10 images from all three categories. The prostate volumes were calculated based on the available manual segmentations for the whole image set and ranged from 15 cm^3 to 89 cm^3 with mean ± standard deviation (SD) of 35±14 cm^3.

2.2.2 Semi-automated segmentation

Our algorithm consists of two main parts: training and segmentation. Figure 2.1 shows the algorithm's block diagram, illustrating the training and segmentation components, described in detail in sections 2.2.2.1 and 2.2.2.2 below.

2.2.2.1 Training

2.2.2.1.1 Spatial normalization. As a spatial normalization step, we parameterized the slice locations in the training images according to slices identified by the operator at specific anatomic locations. Our inferior-superior parameterization was from 0 (apex) to

www.manaraa.com

1 (base) and was used to define inter-subject axial slice correspondence. Therefore, for each (x, y, z) point in the MR Cartesian space, we have an $(x, y, \hat{z})$ point in the normalized coordinate system, where $\hat{z}$ is a real unitless value in the range of [0,1]. We map $\hat{z}$ values to corresponding slice numbers in the MR space by using a nearest neighbor inter-slice interpolation. We also chose the smallest physical pixel size along x- and y-axes (0.273 mm × 0.273 mm) in the data set as the reference pixel size and resampled all the training images with different pixel sizes to that reference pixel size, using bicubic interpolation.

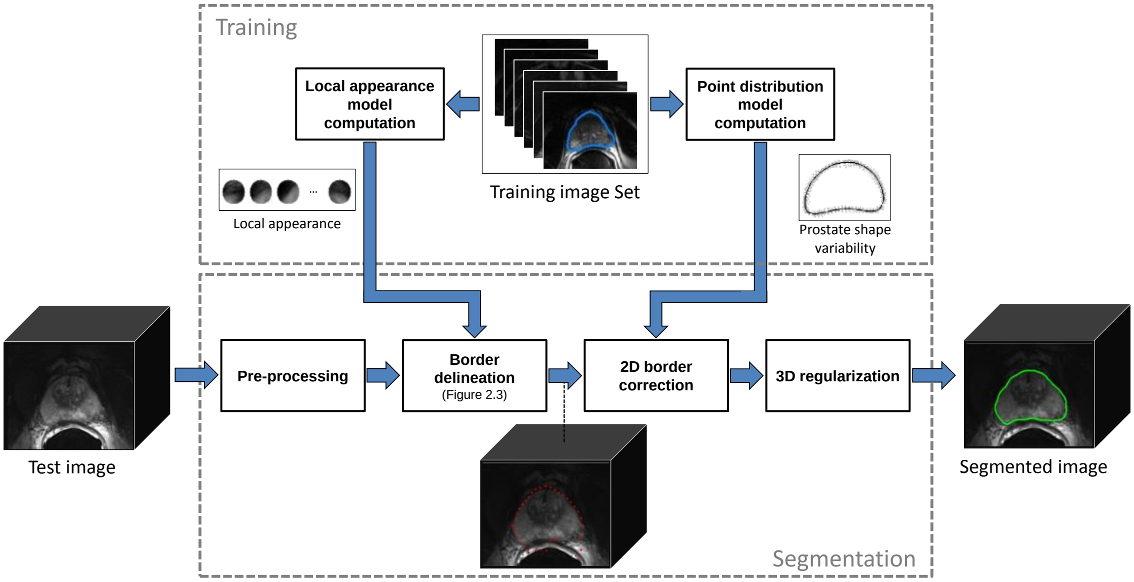


Figure 2.1: Algorithm block diagram. The training images are manually delineated. The candidate boundary points are shown on the test image after “border delineation” step. The final segmentation result is shown on the test image after the “3D regularization” step.

2.2.2.1.2 Prostate border landmark selection. For each training image slice, we manually defined 4 corresponding landmark points on the prostate border: the anterior-most point, the opposite posterior point on the rectal wall, and two points approximately the midpoints of the portions of the prostate boundary touching the neurovascular bundles

www.manaraa.com

(NVBs); see Figure 2.2. We used equal angle interpolation between each neighboring landmark pair, using the mid-point of the line segment defined by anterior and posterior landmarks as the central point, to define 32 additional landmarks, for a total of 36. For a slice with a parameterized axial position of $\hat{z}$, the i^{th} landmark is $\boldsymbol{l}_{i,\hat{z}} = \left(x_{i,\hat{z}}, y_{i,\hat{z}}, \hat{z}\right)$ and $i \in \{1,2, \dots ,36\}$. $x_{i,\hat{z}}$ and $y_{i,\hat{z}}$ are the x- and y-coordinates of the i^{th} landmark on the slice. We observed that in general, the anterior and NVB landmarks are separated by ~120 degrees, and the NVB and posterior landmarks are separated by ~60 degrees. We therefore interpolated ~2/3 of the 32 additional landmarks between the NVB and anterior landmarks (11 interpolated landmarks on the left and right), and ~1/3 of the 32 additional landmarks between the NVB and posterior landmarks (5 interpolated landmarks on the left and right), as shown in Figure 2.2.

2.2.2.1.3 Image patch. A circular image patch $\boldsymbol{p}(m)$, centered at $\left(x_\phi, y_\phi, \hat{z}\right)$ on the slice at axial position $\hat{z}$, is defined as a vector of M consistently ordered image intensities

$$\boldsymbol{p} = \Phi\left(x_\phi, y_\phi, \hat{z}\right) = \left\{I(x, y, \hat{z}) \middle| D\left((x, y, \hat{z}), (x_\phi, y_\phi, \hat{z})\right) \le r_\phi\right\}, \tag{2.1}$$

where r_ϕ is the patch radius, and D is the Euclidean distance function.

2.2.2.1.4 Training image patches. We defined a circular image patch $\boldsymbol{p}_{i,\hat{z}} = \Phi\left(\boldsymbol{l}_{i,\hat{z}}\right) = \Phi\left(x_{i,\hat{z}}, y_{i,\hat{z}}, \hat{z}\right)$, centered on i^{th} landmark of the 36 landmarks ($\boldsymbol{l}_{i,\hat{z}}$) in the training images. The intensity-normalized patch corresponding to the i^{th} landmark on the k^{th} training image (I^k) is defined as

$$\widehat{\boldsymbol{p}}_{i,\hat{z}}^{k} = \frac{\boldsymbol{p}_{i,\hat{z}} - \boldsymbol{\mu}_\phi}{\sigma_\phi}, \tag{2.2}$$

www.manaraa.com

where μ_ϕ and σ_ϕ are the mean and standard deviation of $\boldsymbol{p}_{i,\hat{z}}$, respectively. For each set of corresponding slices in the training set, we calculated the mean intensity of each corresponding set of patch pixels, yielding a set of 36 mean intensity patches. The mean intensity patch corresponding to the i^{th} landmark at slice position $\hat{z}$ across the N training images is defined as

$$\bar{\boldsymbol{p}}_{i,\hat{z}} = \frac{1}{N}\sum_{k=1}^{N} \hat{\boldsymbol{p}}_{i,\hat{z}}^{k} . \tag{2.3}$$

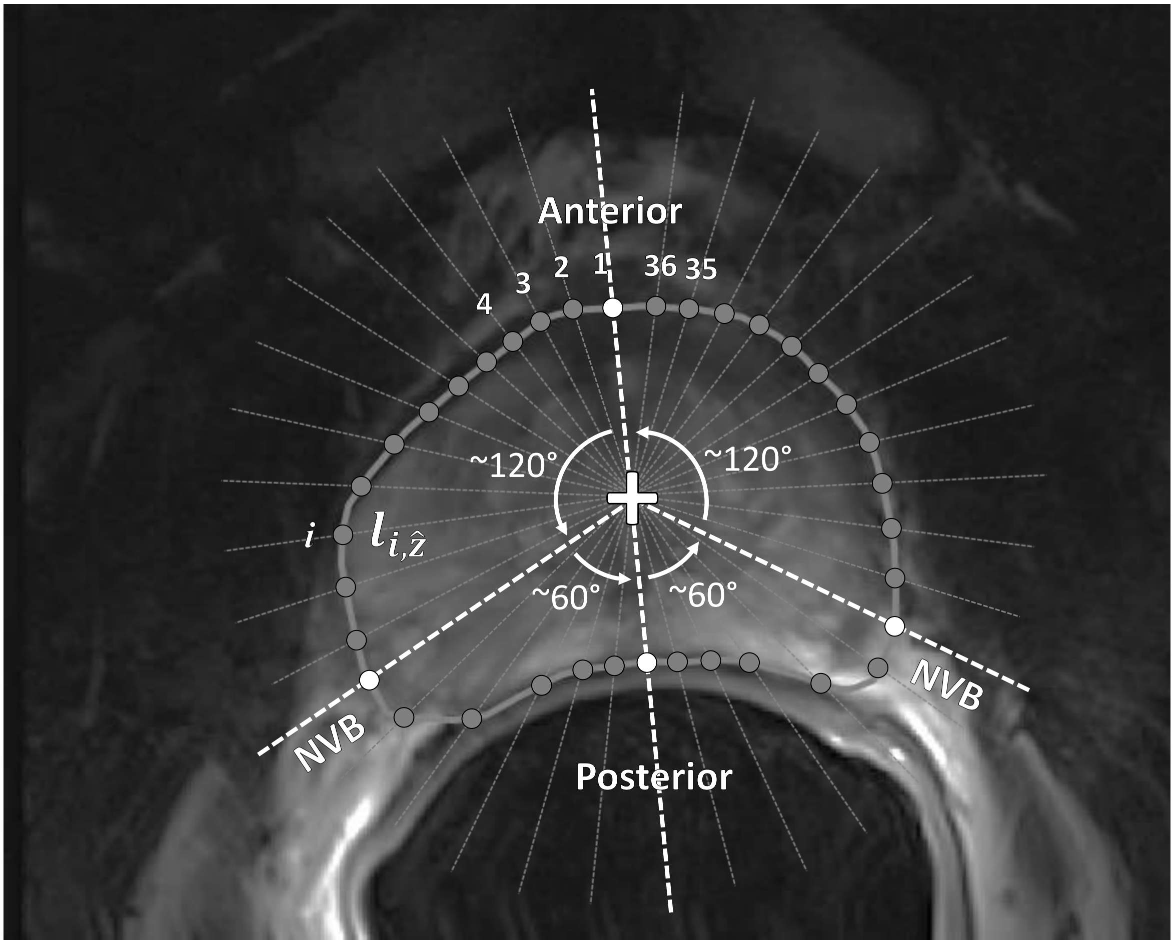


Figure 2.2: Training. determination of 36 prostate border landmarks. The four white dots are user-selected landmarks, the gray dots are interpolated landmarks, and the white cross is the origin.

المنارة للاستشارات
www.manaraa.com

2.2.2.1.5 Point distribution model. For each set of corresponding slices at slice position $\hat{z}$, we also used the 36 landmarks to compute a PDM capturing prostate shape variability at each inferior-superior anatomic position. We used generalized Procrustes analysis[18] to align (translating, rotating and scaling) all the segmentations by minimizing the least squares error between the points. Principal component analysis was then used to compute the eigenvectors and eigenvalues of the covariance matrix for all of the training landmark coordinates[17].

2.2.2.2 Segmentation

The segmentation algorithm incorporates a small set of inputs from the operator to define the inferior-superior extents of the prostate, as well as its center and orientation. These inputs are: (1) the apex-most and base-most slice numbers (z); (2) the points at the center of the prostate on the apex- and base-most slices, and on the slice within the mid-gland equidistant to these two slices; and (3) the anteroposterior (AP) orientation of the prostate as seen on this mid-gland slice. We developed a customized graphical user interface to efficiently collect these operator inputs.

Using these operator inputs, we parameterized the axial slice positions of the test image as in training, permitting the extraction of the corresponding mean intensity patches and PDM corresponding to each axial slice from the training stage. The center points for all prostate slices were estimated by interpolating the three operator-provided center points on the base, mid-gland and apex slices. Therefore, a center point $(x_C(\hat{z}), y_C(\hat{z}), \hat{z})$ was available for each slice at position $\hat{z}$. We approximated the orientation of the prostate in all axial slices from base to apex using the mid-gland AP symmetry axis (APSA). The segmentation was performed on each prostate axial slice

www.manaraa.com

within the image volume, resulting in a 3D segmentation of the prostate from the base, through the mid-gland, to the apex.

2.2.2.2.1 Preprocessing. Before delineating the prostate border, first we applied a median filter (using a 5 × 5 pixel sliding window) as an edge-preserving low-pass filter to each axial slice, in order to reduce image noise.

2.2.2.2.2 Appearance-based boundary point selection. For each axial slice $\hat{z}$ in the 3D volume, we used the prostatic center point $(x_C(\hat{z}), y_C(\hat{z}), \hat{z})$ and the APSA to define 36 rays emanating from the center point, intended to be homologous to the orientations of the training landmarks. We used a radial search strategy to choose a set of 36 candidate points for the prostate border on each slice. As the search space was small, we used an exhaustive search to maximize the normalized cross correlation (NCC) of each mean intensity patch with the image region under the patch along the corresponding ray (Figure 2.3), i.e.:

$$\left(\dot{x}_{i,\hat{z}}, \dot{y}_{i,\hat{z}}\right) = \arg\max_{(x,y)} NCC\left[\bar{\boldsymbol{p}}_{i,\hat{z}}, \Phi(x,y,\hat{z})\right], \tag{2.4}$$

$$(x,y) \in \{(x,y)|(x,y,\hat{z}) \in R_i, r_{min} < D[(x,y,\hat{z}),(x_C(\hat{z}),y_C(\hat{z}),\hat{z})] < r_{max}\}, \tag{2.5}$$

where $\left(\dot{x}_{i,\hat{z}}, \dot{y}_{i,\hat{z}}\right)$ is the optimal point with the highest NCC along i^{th} ray (R_i), r_{min} and r_{max} indicate the search start point and stop point on each ray, respectively, and

$$NCC(\boldsymbol{p}_1, \boldsymbol{p}_2) = \frac{1}{M}\sum_{m=1}^{M} \frac{\boldsymbol{p}_1(m) - \mu_{\phi 1}}{\sigma_{\phi 1}} \times \frac{\boldsymbol{p}_2(m) - \mu_{\phi 2}}{\sigma_{\phi 2}}, \tag{2.6}$$

where M is the number of pixels in patches $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$, and $\mu_{\phi 1}$ and $\mu_{\phi 2}$ are mean voxel intensities, and $\sigma_{\phi 1}$ and $\sigma_{\phi 1}$ are the standard deviations of pixel intensities of patches $\boldsymbol{p}_1$

www.manaraa.com

and $\boldsymbol{p}_2$, respectively. This process yielded 36 candidate border points on each slice, with one on each ray.

2.2.2.2.3 Shape-based boundary regularization [two-dimensional (2D)]. We aligned the mean shape in the PDM to each set of 36 candidate points using Procrustes analysis and extracted the parameters from the model that represented the shape of the candidate points. Then we calculated the parameters of the nearest shape in the PDM to the 36 candidate points by restriction of each extracted parameter in the model to the range of $[-1.5\lambda, 1.5\lambda]$ where λ is the corresponding eigenvalue. We defined those points with absolute distances greater than 1.5 standard deviation to the nearest shape in the model as outlier boundary points, and corrected them by replacing them with the corresponding points of the model shape. This procedure was iterated until all outliers were eliminated or a specified maximum number of iterations was reached. This resulted in a set of shape-regularized boundary points, yielding a plausible prostate shape.

2.2.2.2.4 3D regularization. After applying two-dimensional (2D) shape regularization to all of the prostate slices, for ray *i*, a second order curve was fitted to all the boundary points from base to apex in order to regularize the prostate shape in 3D. By interpolating the points with a spline from apex to base, we obtained a smooth, continuous 3D segmentation of the prostate.

www.manaraa.com

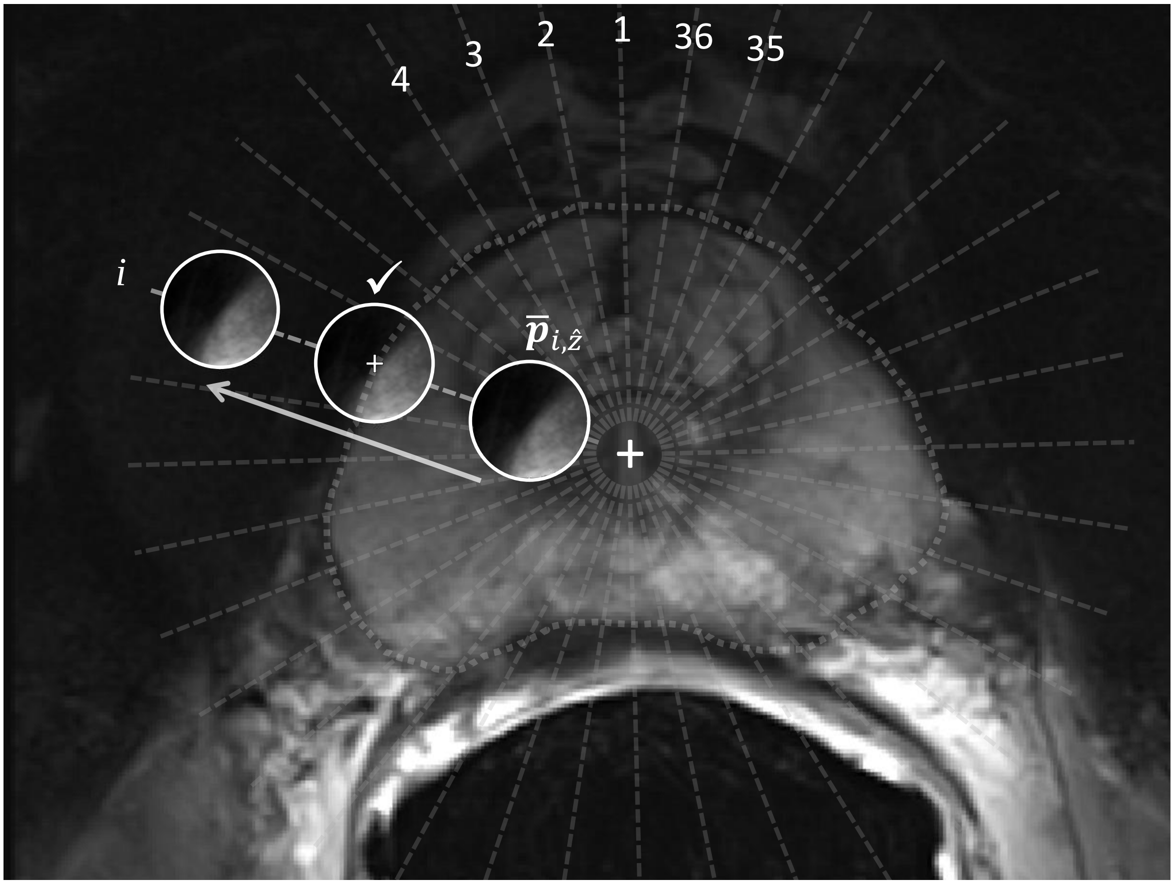


Figure 2.3: Segmentation. 36 rays and patch translation along one ray (the dotted manual segmentation is overlaid for reference but is not provided to the segmentation algorithm).

2.2.3 Validation metrics

We evaluated our method against manual segmentations with complementary boundary-based (MAD), regional overlap (DSC, recall and precision) and volume difference (ΔV) metrics. The metrics are explained later in this section. To develop a reference against which to compare the metrics resulting from our segmentation algorithm, we measured the inter-operator variability in expert manual prostate border delineation on a subset of 10 of our 42 3D images. Each of these images was manually segmented in 3D by three observers: one radiologist, one radiation oncologist, and one

المنارة للاستشارات
www.manaraa.com

radiology resident, all specializing in prostate MRI. We calculated the metrics (1) in 2D for each slice, (2) in 3D for the whole gland, (3) in 3D for the superior-most third of the prostate (corresponding approximately to the base), (4) in 3D for the middle third (mid-gland), and (5) in 3D for the inferior-most third (apex). In cases where the operator's selected apex and base slices differed from those of the reference segmentation, we calculated the 2D metrics on the slices that were common to both segmentations. To apply our 3D metrics to the defined base, mid-gland and apex regions, the middle third of the slices common to both the operator's and the reference segmentations were defined as the mid-gland region, and the remaining inferior and superior parts were considered to be the apex-most and base-most components, respectively. The operator interaction time was measured as well as inter-operator time and accuracy differences.

2.2.3.1 Mean absolute distance

The mean absolute distance (MAD) is a metric that measures the disagreement between two curves (in 2D) or surfaces (in 3D) as an aggregate of Euclidean distances between corresponding sets of points on these surfaces. We defined two modes for computing the MAD: unilateral and bilateral. In unilateral MAD, one surface is the reference surface and points are corresponded by finding the closest point on the reference surface to each point on the other surface. The MAD is then the average of the distances between corresponding points, defined as

$$MAD(X,Y) = \frac{1}{K}\sum_{p\in X} \min_{q\in Y} D(p,q)\,, \tag{2.7}$$

where X and Y are the point sets (Y is the reference set), K is the number of points of X, $p = (x_p, y_p, z_p)$ is a point in X and $q = (x_q, y_q, z_q)$ is a point in Y, and $D(p,q)$ is the 3D

المنارة للاستشارات

www.manaraa.com

Euclidean distance between p and q. The bilateral MAD is defined as the mean of two unilateral MADs, calculated with each of the two surfaces as the reference. We reported MAD in mm, with $MAD = 0$ mm indicative of perfect alignment between shapes, and larger MAD values indicating increasing levels of shape disagreement.

2.2.3.2 Dice similarity coefficient

The Dice similarity coefficient (DSC) [13] measures the misalignment between two shapes in terms of their overlap region. The DSC of two 3D shapes is

$$DSC(X,Y) = \frac{2(X \cap Y)}{X + Y} = \frac{2TP}{FP + 2TP + FN}, \tag{2.8}$$

where TP is the true positive (correctly identified) region, and FN is the false negative (incorrectly ignored) region. We reported DSC as a percentage. A DSC value of 100% indicates perfect alignment, and a DSC value of 0% indicates no overlap of the two shapes.

2.2.3.3 Recall rate

The recall rate, or sensitivity, is the proportion of the reference, which is identified correctly, defined as

$$Recall(X,Y) = \frac{TP}{TP + FN}. \tag{2.9}$$

In this chapter, $Recall$ is the proportion of the reference prostate segmentation that is within the segmentation provided by the algorithm and is reported as a percentage. An ideal $Recall$ value of 100% indicates that the segmentation provided by the algorithm covers the entire reference segmentation, plus potentially some additional regions outside of the reference segmentation.

المنارة للاستشارات
www.manaraa.com

2.2.3.4 Precision rate

The precision rate is the proportion of the segmentation which is true positive, and is defined as

$$Precision(X,Y) = \frac{TP}{TP+FP}. \tag{2.10}$$

where FP is the false positive (incorrectly identified) region. In this chapter, $Precision$ is the proportion of the segmentation provided by the algorithm that is within the reference prostate segmentation. An ideal $Precision$ value of 100% indicates that the segmentation provided by the algorithm lies entirely within the reference segmentation, but may or may not completely overlap the reference segmentation. An ideal segmentation algorithm would yield $Precision$ of 100% and $Recall$ of 100%. Computing and interpreting these metrics both separately and together provides a means for understanding both the magnitude and the meaning of the types of regional overlap errors made by a segmentation algorithm, complementing the information provided by DSC.

2.2.3.5 Volume difference (ΔV)

The signed volume difference (ΔV) is the subtraction of the volume of the reference segmentation from the volume given by the segmentation algorithm:

$$\Delta V = V_{algorithm} - V_{reference}, \tag{2.11}$$

where $V_{algorithm}$ is the volume of segmentation result, and $V_{reference}$ is the volume of the prostate in the manual segmentation. ΔV is a signed metric and was reported in cm^3 in this chapter. A negative value of ΔV indicates under-segmentation and a positive value indicates over-segmentation in terms of the volume of the prostate.

www.manaraa.com

2.3 Experiments

For all of the experiments in this chapter, we used a single value for the patch radius r_ϕ, chosen by systematic search. We selected a representative subset of 6 images from the data set. We applied our algorithm using patch radii in the range of $3\ mm \leq r_\phi \leq 17\ mm$, using leave-one-out cross-validation to define training and test images. For each image, we measured the MAD, DSC, ΔV, as well as the average of recall and precision values (as one metric) for each patch radius. Then, we calculated the average of the four metrics across the 6 images, yielding four mean values, one for each metric. We ranked the radii based on each metric, resulting in 4 rankings; thus, each radius had four rank values. We calculated the average of the 4 rank values for each radius and chose the radius ($r_\phi = 5$) having the lowest average rank.

The radial search started on each ray from a distance of 2 mm from the center point ($r_{min} = 2\ mm$) and ended at 35 mm ($r_{max} = 35\ mm$). This range was chosen based on our observed prostate size and imaging field of view in the data set. The maximum number of iterations for shape-based 2D regularization was set to be 25.

2.3.1 Inter-operator variability: Manual segmentation

We compared the observers' segmentations in a pairwise fashion using our 3D validation metrics. We also compared each observer's segmentation to the simultaneous truth and performance level estimation (STAPLE) [19] segmentation derived from all three observers' segmentations. STAPLE is a method that is intended to estimate a single reference segmentation from a set of reference segmentations using a weighted voting scheme.

المنارة للاستشارات
www.manaraa.com

2.3.2 Accuracy and Inter-operator variability: Semi-automatic segmentation

We performed a single-operator evaluation of the semi-automated segmentation algorithm on all 42 3D images in our data set, and compared the results to a single manual reference segmentation. We used a leave-one-out cross-validation methodology to split the 42 images into training (41 images) and test (one image) sets in each of 42 rounds of testing, with metrics averaged over all rounds.

We performed a multiple-operator evaluation of the semi-automated segmentation algorithm in terms of accuracy, inter-operator variability, and operator interaction time. We partitioned our data set into non-overlapping training and test sets of 32 and 10 images, respectively. The 10-image test set was the same as the data set used in Section 2.3.1. Nine operators, including 4 radiation oncologists, one radiologist, one radiology resident, one imaging scientist and two graduate students, all with research and/or clinical experience with prostate imaging, used our semi-automated segmentation algorithm to segment each of the 10 images. We computed aggregate 3D segmentation metrics by averaging across all operators, and we also compared the metrics for each operator with all other operator results to measure the inter-operator variability. Since operators' judgments regarding anteroposterior prostate orientation and the locations of the apex-most and base-most slices differed, we measured the inter-operator variability in base and apex slice selection and prostate orientation definition. We calculated the mean standard deviation of the operators' selected apex and base slice numbers, as well as the APSA angle with respect to anterior-posterior axis of the MRI coordinate system. To measure the inter-operator variability in apex, base and mid-gland center point selection, we first determined the superior-most (base) and inferior-most (apex) slices that were common to

المنارة للاستشارات

www.manaraa.com

the segmentations of all observers, as well as the mid-gland slice equidistant to both. We then calculated the means of the 9 actual or interpolated center points at each of these slice locations. On each slice, we then measured the Euclidean distance of each of the 9 center points to the mean point.

We had three manual segmentations available on the same 10-image test set as was used in Section 2.3.1. To perform a direct comparison of our segmentation error metrics for manual and semi-automatic segmentation, we used those three manual segmentations to compute a STAPLE reference standard segmentation from each of the 10 images. Our error metrics were calculated with respect to the STAPLE reference for the manual segmentations, as well as for semi-automatic segmentations performed by the same experts on the same 10 images. The mean and standard deviation of these metrics for the manual and semi-automatic scenarios were calculated to measure differences in accuracy and observer variability arising from using manual vs. semi-automatic segmentation.

2.3.3 Sensitivity to initialization: Semi-automatic segmentation

To examine the sensitivity of the semi-automated segmentations to the operator’s center point selection, we performed a simulation study wherein our 42 images were repeatedly segmented 1000 times using perturbed (in accordance with the previously observed inter-operator variability) prostate center points at each iteration. We calculated perturbed center point positions within the prostate by randomly sampling from 2D Gaussian distributions (three in total: one for each of the apex, mid-gland, and base slices) with means defined at “ideal” center points defined on the midpoint of the line segment between the most-anterior and the most-posterior prostate border landmarks

الـمنارة للاستشارات
www.manaraa.com

used in the training. The standard deviations of these Gaussian distributions were estimated based as the root mean square (RMS) distances to the means of the center points collected from the nine operators in Section 2.3.2. In this test, the sensitivity for each image was measured as the difference of the metrics based on the perturbed center points and the metrics based on the “ideal” center points. Therefore, for N images and 1000 repetitions, we have 1000N measured differences. We reported the mean and the standard deviation of these 1000N values for each metric.

To measure the sensitivity of the results to the selection of the anteroposterior symmetry axes, we performed another simulation study wherein our 42 images were repeatedly segmented 1000 times using randomly modified axis angles. For that purpose, at each iteration, we randomly selected a set of 42 angles from a Gaussian distribution with zero mean and the same standard deviation as the standard deviation of the observed angle across the nine operators in Section 2.3.2, and added them to the symmetry axis angles used in the single-operator experiment. We measured the sensitivity as the differences between metrics based on the randomly generated angles and the metrics based on the reference angles used in the single-operator experiment. We reported the mean and standard deviation of these differences across all patients and 1000 repetitions.

2.3.4 Source of Variability: Semi-automatic segmentation

To measure the relative contributions of different sources of variability for our semi-automatic segmentation algorithm, we designed a three-way analysis of variance (ANOVA) test, with *reference*, *trainer*, and *operator* factors. We used the same subset of 10 images as used in the three-operator experiment, and two observers (denoted Observer #1 and Observer #2) who were selected due to the discordance of their manual

المنارة للاستشارات

www.manaraa.com

segmentations of these 10 images observed in the results of experiment described in Section 2.3.1. These two observers also executed the semi-automated segmentation on each of these 10 images. For these two observers, all possible configurations of manual segmentations used for *reference* (used to calculate the validation metrics), manual segmentations used as the *trainers* for the semi-automated tool, and semi-automated segmentation *operators* were tested. This yielded a set of segmentation error metrics for each configuration. We performed an ANOVA test for each of our five metrics and each region of interest including whole gland, mid-gland, base and apex to test the following null hypotheses:

$H0_1$: The trainer has no significant impact on the error.

$H0_2$: The operator has no significant impact on the error.

$H0_3$: The reference has no significant impact on the error.

2.4 Results

2.4.1 Inter-operator variability: Manual segmentation

The key result of this experiment was a substantially high inter-operator variability in manual segmentation. Table 2.1 shows the range of 3D metrics in pair-wise comparison between operators and also between each operator and STAPLE reference. Since in this experiment, the segmentations in each pair-wise comparison were both performed manually, the MAD values were calculated in bilateral mode and the absolute volume difference ($|\Delta V|$) was calculated. MAD values were calculated in unilateral mode with STAPLE as the reference, and the signed volume difference (ΔV) was reported. Figure 2.4 qualitatively shows the inter-observer variability in prostate segmentation.

المنارة للاستشارات
www.manaraa.com

Table 2.1: Inter-operator variability in manual segmentation: Range of mean MAD, DSC, recall, precision, and ΔV (bilateral MAD and |ΔV| was used for “Operator vs Operator” section).

		Range of mean metric values				
	Region of interest	**MAD (mm)**	**DSC (%)**	**Recall (%)**	**Precision (%)**	**ΔV (cm^3)**
Operator vs Operator	Whole gland	[1.0,2.8]	[74,90]	[87,99]	[60,94]	[1.9,18.3]
	Mid-gland	[0.7,1.8]	[88,96]	[96,99]	[81,95]	[0.1,3.3]
	Apex	[1.1,3.0]	[65,88]	[83,98]	[51,94]	[0.5,6.1]
	Base	[1.3,3.5]	[66,86]	[79,99]	[52,93]	[1.5,7.7]
Operator vs STAPLE	Whole gland	[0.2,3.1]	[78,98]	[66,100]	[87,98]	[-2.8,15.5]
	Mid-gland	[0.2,1.9]	[89,98]	[82,100]	[96,99]	[-0.5,3.2]
	Apex	[0.2,3.4]	[70,99]	[58,100]	[84,98]	[-0.8,5.3]
	Base	[0.2,3.7]	[72,98]	[60,100]	[80,98]	[-1.8,7.0]

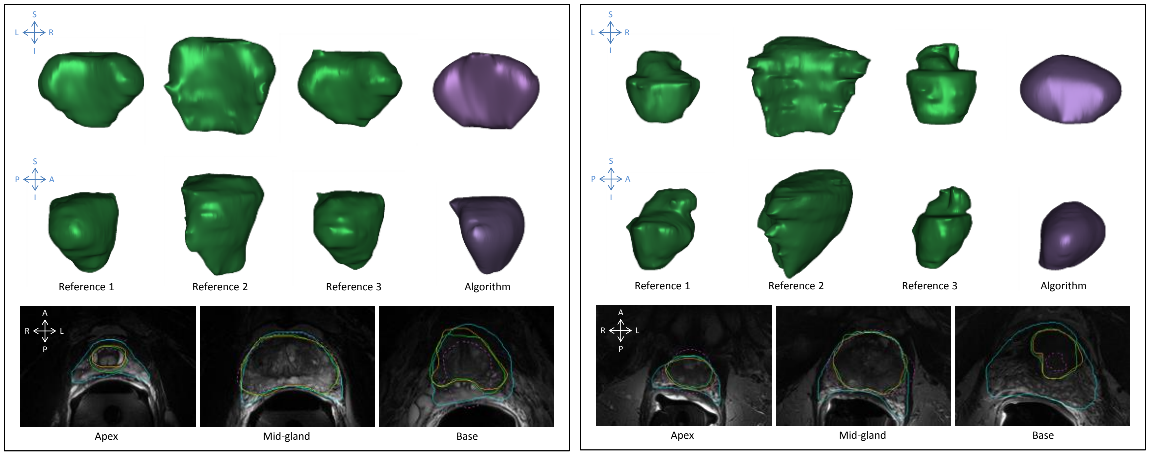


Figure 2.4: Inter-observer variability. The 3D surfaces show the three manual segmentations and the algorithm results. The three solid contours show the three observers’ manually drawn contours. The dashed contours show the algorithm’s results.

2.4.2 Accuracy and inter-operator variability: Semi-automatic segmentation

The key results of this experiment were that (1) the accuracy measured for our algorithm based on one reference, similar to the accuracies of most of the other segmentation algorithms presented in the literature, are within the inter-operator variability range for manual segmentation on our data set; (2) the variability observed between different operators in the measured errors using a multi-operator study was not significant based on most of the metrics and for most regions of interest. The results of the single-operator evaluation of the semi-automated segmentation algorithm on all 42

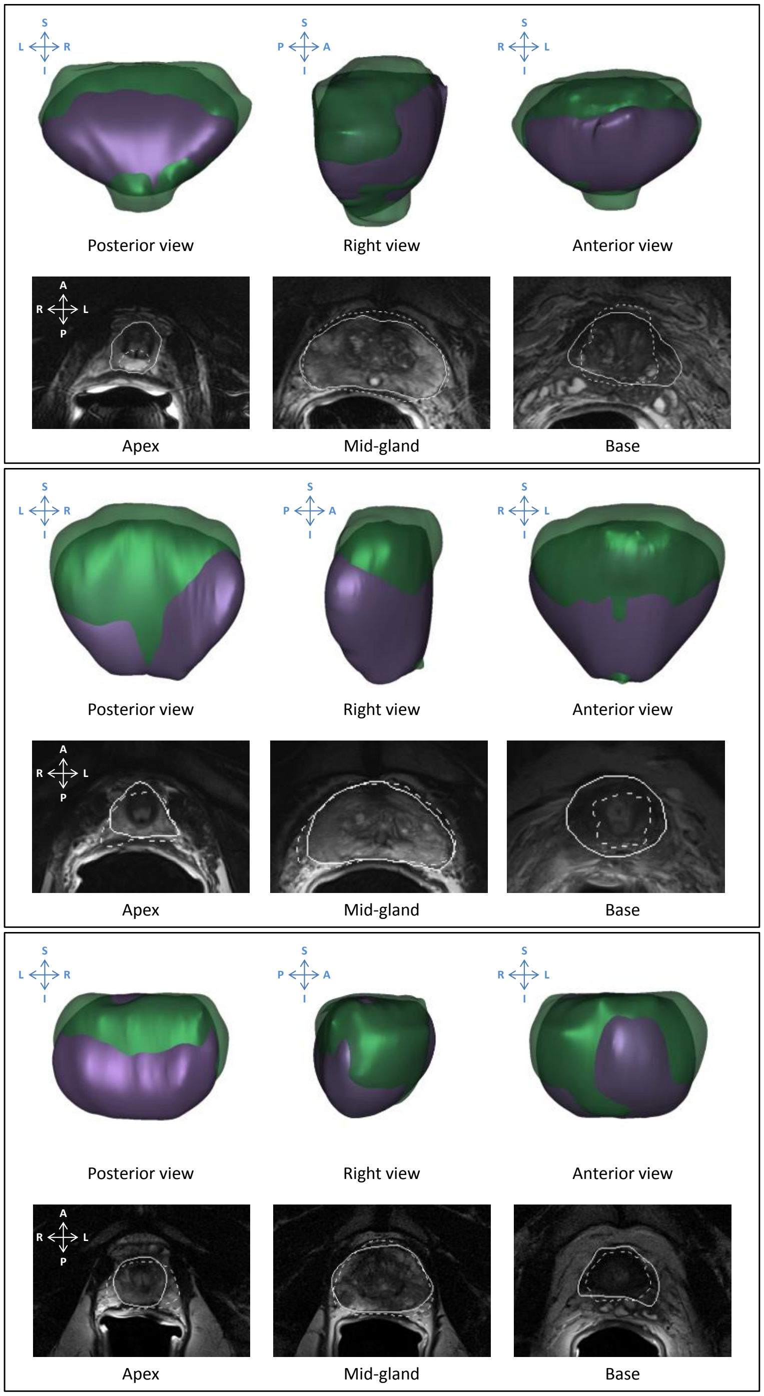
المنارة للاستشارات
www.manaraa.com

3D images in our data set are shown in Table 2.2 and compared to previous work. For the whole gland, we measured a mean±standard deviation MAD (unilateral) of 2.0±0.5 mm, DSC of 82±4%, recall of 77±9%, precision of 88±6% and ΔV of -4.6±7.2 cm^3. The measured mean±standard deviation execution time using an unoptimized Matlab research platform on a single CPU core was 85±20 sec. Under the assumption of normal distribution of the error metric values, we conducted one-tailed heteroscedastic t-tests [20] to compare our results to previous work, in each case testing the null hypothesis regarding the relative performance of the methods. With $\alpha=0.05$, corresponding letters show where the null hypothesis was rejected in Table 2.2. Figure 2.5 shows qualitative and quantitative results for three sample prostates.

Table 2.2: Accuracy and variability for semi-automatic segmentation: mean±standard deviation of MAD, DSC, recall, precision, and ΔV. Corresponding letters show statisticaly significant differences between each error value of our method and the corresponding error value of another method where applicable. ($p < 0.05$).

Methods	**N**	**Region of interest**	**MAD (mm)**	**DSC (%)**	**Recall (%)**	**Precision (%)**	**ΔV (cm^3)**
Our method	42	Whole gland	2.0±0.5 **bi**	82±4 **ac**	77±9 **m**	88±6 **n**	-4.6±7.2
		Mid-gland (1/3)	1.6±0.5 **j**	90±3**d**	90±7	91±6	-0.1±2.0
		Apex (1/3)	2.0±0.7 **gk**	79±6 **e**	82±14	80±13	0.1±3.3
		Base (1/3)	2.6±0.8 **l**	73±10 **fh**	61±14	93±6	-4.5±3.7
Liao *et al* [15]	66	Whole gland	1.8±0.9	88±3 **a**	-	-	-
Toth *et al* [14]	108	Whole gland	1.5±0.8 **b**	88±5 **c**	-	-	-
		Mid-gland (1/3)	-	91±4 **d**	-	-	-
		Apex (1/3)	-	84±9 **e**	-	-	-
		Base (1/3)	-	88±6 **f**	-	-	-
Vikal *et al* [12]	3	Mid-gland (9/13)	2.0±0.6	93±3	-	-	-
		Apex (2/13)	3.8±0.9 **g**	80±5	-	-	-
		Base (2/13)	3.9±1.8	86±8 **h**	-	-	-
Martin *et al* [11]	17	Whole gland	3.4±2.0 **i**	-	89±6 **m**	78±12 **n**	-
		Mid-gland	2.4±1.3 **j**	-	-	-	-
		Apex	2.9±1.3 **k**	-	-	-	-
		Base	4.3±2.0 **l**	-	-	-	-

المنارة للاستشارات
www.manaraa.com



	W.G.	M.G	Apex	Base
MAD (mm)	1.7	1.3	1.3	2.1
DSC (%)	85	92	85	80
Recall (%)	84	97	83	73
Precision (%)	87	88	86	87
ΔV (cm3)	-1.3	1.4	-0.2	-2.5

	W.G.	M.G	Apex	Base
MAD (mm)	1.8	1.2	1.8	2.4
DSC (%)	82	92	80	75
Recall (%)	79	96	98	61
Precision (%)	86	88	68	99
ΔV (cm3)	-2.4	0.9	2.4	-5.6

	W.G.	M.G	Apex	Base
MAD (mm)	2.2	1.7	2.3	2.4
DSC (%)	85	91	83	82
Recall (%)	84	89	97	71
Precision (%)	86	93	73	97
ΔV (cm3)	-0.9	-0.5	3.2	-3.6

Figure 2.5: Qualitative and quantitative results for three sample prostates. In the left column, the semi-transparent surfaces show the manual segmentation as reference, and the solid surfaces show the algorithm results. On the 2D cross sections, the manual segmentation is shown with a solid line, and the algorithm's segmentation is shown with a dashed line. The most inferior and the most superior slices that contain both reference and algorithm contours were, respectively, shown as the apex and base. In the right column, the tables show the measured error metrics for that corresponding cases in whole gland (W.G.), as well as apex, mid-gland (M.G.), and apex.

www.manaraa.com

For our multiple-operator evaluation of the semi-automated segmentation, Figure 2.6 shows the results for each of the 9 operators in comparison with STAPLE. The average interaction time across 9 operators and 10 images was measured as 28±14 sec. To determine whether there are significant differences between the means of the error metrics for each operator we conducted one-way ANOVA followed by Bonferroni's pairwise tests with the null hypothesis that the means of the metrics for all the 9 operators were the same. We showed the post ANOVA test results in Figure 2.6 for each region of interest, where ANOVA detected significant inter-operator differences in terms of any of the metrics (α=0.05). Table 2.3 also shows the average results across all 9 operators and 10 images and compares it to the results based on one operator on 42 images reported in Table 2.2. For each metric, we applied a t-test with the null hypothesis that the means of the metric resulting from the 9 operators' segmentations of 10 images are the same as the mean of the metric for one operator's segmentation of 42 images (i.e. comparing the top row of Table 2.3 to the bottom row, for each metric and within each anatomic region) .

Table 2.3: The average results across the nine operators and 10 images compared to the single operator results across 42 images: mean±standard deviation of MAD, DSC, recall, precision, and ΔV. Corresponding letters indicate statistically significant differences between two modes ($p < 0.05$).

Methods	**N**	**Region of interest**	**MAD (mm)**	**DSC (%)**	**Recall (%)**	**Precision (%)**	**|ΔV| (cm^3)**
One operator (Table 2.2)	42	Whole gland	2.0±0.5	82±4 **b**	77±9 **d**	88±6	-4.6±7.2
		Mid-gland	1.6±0.5	90±3	90±7	91±6	-0.1±2.0
		Apex	2.0±0.7	79±6	82±14	80±13	0.1±3.3
		Base	2.6±0.8 **a**	73±10 **c**	61±14 **e**	93±6	-4.5±3.7
Nine operators	10	Whole gland	2.2±0.7	77±8 **b**	72±12 **d**	86±10	-4.0±5.5
		Mid-gland	1.7±0.7	89±4	88±8	91±7	-0.1±1.7
		Apex	2.0±1.0	78±12	84±15	78±17	0.6±3.3
		Base	2.9±0.8 **a**	65±12 **c**	54±17 **e**	92±11	-4.5±3.4

For each image, the apex and base slices were manually selected by each of the 9 operators. We calculated the resulting standard deviation of the slice positions for each

المنارة للاستشارات
www.manaraa.com

image and obtained their average across all 10 images. The mean standard deviation of the operators' selected apex and base slices were 1.8 slices (4 mm) for the apex and 1.3 slices (2.9 mm) for the base. The range of the maximum inter-operator difference at the apex was 3 to 9 slices (6.6 mm to 19.8 mm) with a mean of 5.7 slices (12.5 mm), and this range for base was 2 to 10 slices (4.4 mm to 22 mm) with a mean of 3.8 slices (8.4 mm). The mean standard deviation of the APSA angle with respect to anteroposterior axis of the MRI coordinate system was 3.2 degrees. The differences between operators ranged from 4.0 to 17.8 degrees with a mean of 9.8 degrees.

For the center points, the measured distances ranged from 0 mm to 3.9 mm with an average of 1.1 mm at the apex, 0.2 mm to 4.3 mm with an average of 1.3 mm at the mid-gland, and 0.1 mm to 4.9 mm with an average of 1.7 mm at the base. The RMS of the point distances were 1.3 mm, 1.5 mm, and 2.0 mm at the apex, mid-gland and base, respectively. The actual range of distances by which the center points were perturbed were [-5 mm, +5mm], [-7 mm, +7 mm], and [-8 mm, +8 mm] for the apex, midgland, and base, respectively.

Table 2.4 shows the of manual and semi-automatic segmentations performed by the same three operators on 10 images.

Table 2.4: Consistency of the manual and the semi-automatic segmentations: average of means (average of standard deviations) of MAD, DSC, recall, precision, and ΔV across 3 manual and 3 semi-automatic segmentations of the prostate by 3 expert operators.

	N	Region of interest	MAD (mm)	DSC (%)	Recall (%)	Precision (%)	ΔV (cm^3)
Manual segmentation	10	Whole gland	1.3 (1.6)	90 (11)	88 (19)	94 (6)	3.9 (10.1)
		Mid-gland	0.8 (0.9)	95 (5)	93 (10)	97 (2)	0.8 (2.1)
		Apex	1.4 (1.8)	86 (15)	85 (24)	93 (8)	1.4 (3.3)
		Base	1.6 (1.9)	86 (14)	86 (22)	91 (11)	1.6 (4.8)
Semi-automatic segmentation	10	Whole gland	1.9 (0.3)	80 (4)	75 (7)	88 (3)	-3.2 (2.9)
		Mid-gland	1.5 (0.3)	90 (2)	90 (3)	91 (2)	0.2 (0.7)
		Apex	1.8 (0.4)	82 (4)	87 (8)	80 (8)	0.8 (1.2)
		Base	2.7 (0.5)	68 (7)	57 (12)	93 (6)	-4.2 (2.5)

المنارة للاستشارات
www.manaraa.com

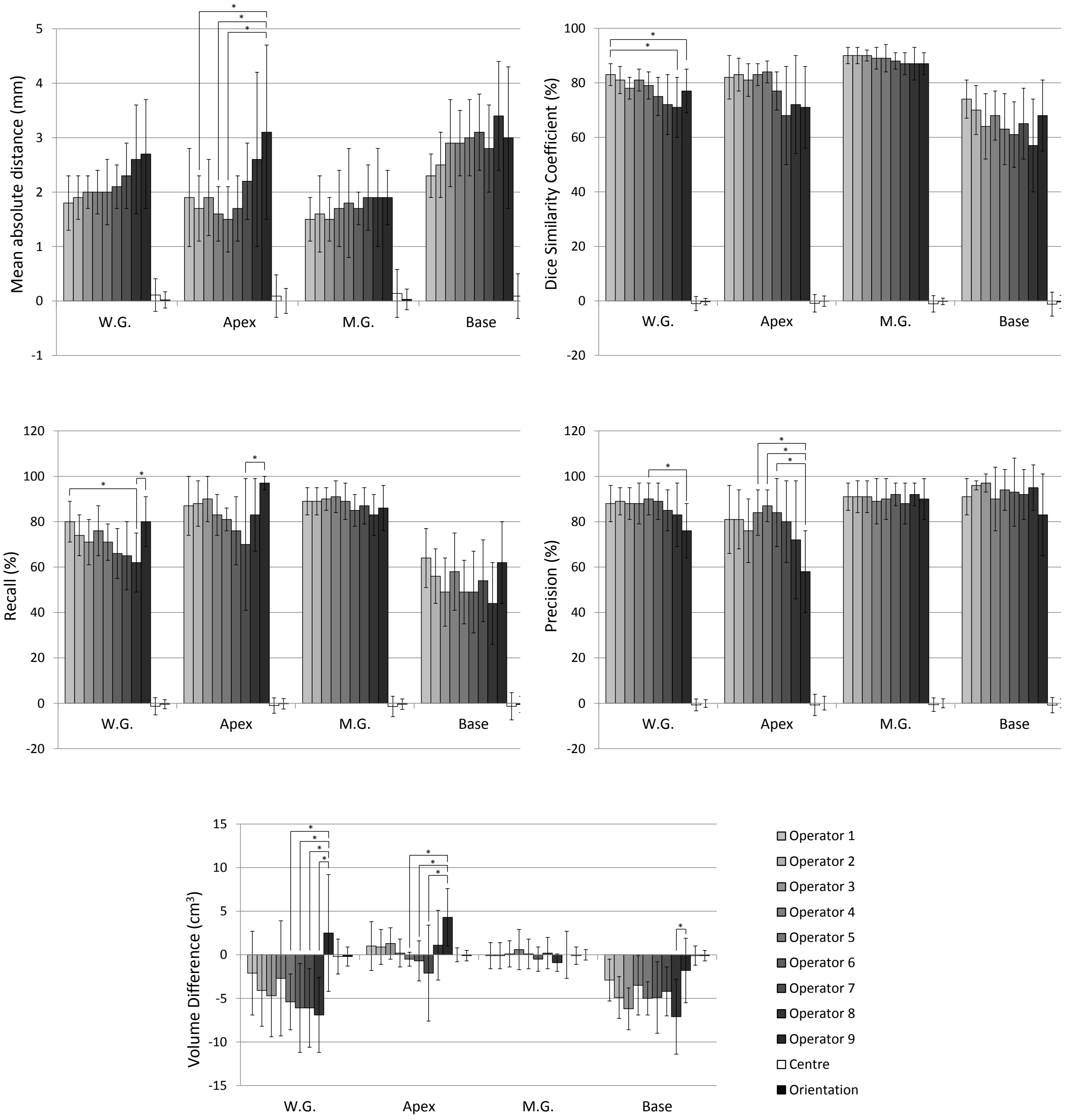


Figure 2.6: Inter-observer variability. Mean±standard deviation (a) MAD, (b) DSC, (c) recall, (d) precision, and (e) ΔV for each of the 9 operators for whole gland (W.G.), apex, mid-gland (M.G.), and base ($P < 0.05$). The last two columns in each section show the average variations of the metric using perturbed prostate center point, and anteroposterior symmetry axes, respectively.

المنارة للاستشارات
www.manaraa.com

2.4.3 Sensitivity to initialization: Semi-automatic segmentation

The key result of this experiment was that the sensitivity of the algorithm accuracy to center point and anteroposterior symmetry axis selection was substantially lower than the measured error metric values. The means and standard deviations of the variation of metrics with regards to center point and anteroposterior symmetry axes variations are shown in Figure 2.6 for the whole gland as well as apex, mid-gland, and base regions.

2.4.3.1 Sensitivity to centre point selection

For the whole gland, the mean and the range of variation ([minimum, maximum]) for MAD, DSC, recall, precision and ΔV, respectively, were 0.1 mm ([-1.3 mm, 2.0 mm]), -1% ([-24%, 11%]), -1% ([-33%, 16%]), -1% ([-19%, 16%]), and -0.2 cm^3 ([-18.4 cm^3, 15.4 cm^3]). The mean and standard deviation of the differences between results based on randomly generated center points and reference center points across 42 patients and 1000 repetitions are shown in Table 2.5.

2.4.3.2 Sensitivity to anteroposterior symmetry axes selection

For the whole gland, the mean and the range of variation ([minimum, maximum]) for MAD, DSC, recall, precision and ΔV, respectively, were 0.0 mm ([-0.7 mm, 0.9 mm]), 0% ([-8%, 5%]), 0% ([-15%, 8%]), 0% ([-11%, 8%]), and -0.2 cm^3 ([-9.2 cm^3, 4.5 cm^3]). Table 2.5 shows the mean and one standard deviation of the differences between results based on randomly generated angles and the reference measurements across 42 patients and 1000 repetitions.

www.manaraa.com

Table 2.5: Sensitivity of the semi-automatic algorithm to initialization (center points and anteroposterior symmetry axes): mean±standard deviation of MAD, DSC, recall, precision, and ΔV offsets from reference measurements across 1000 repetitions × 42 patients.

	N	# of Iterations	Region of interest	MAD (mm)	DSC (%)	Recall (%)	Precision (%)	ΔV (cm^3)
Sensitivity to center point selection	42	1000	Whole gland	0.11±0.30	-1.0±2.6	-1.3±3.8	-0.7±2.6	-0.2±2.0
			Mid-gland	0.14±0.44	-1.1±3.0	-1.4±4.5	-0.6±3.0	-0.1±1.0
			Apex	0.09±0.39	-0.9±3.2	-1.0±3.4	-0.7±4.7	0.0±0.8
			Base	0.09±0.41	-1.2±4.4	-1.3±6.0	-0.8±3.4	-0.1±1.1
Sensitivity to anteroposterior symmetry axes selection	42	1000	Whole gland	0.02±0.15	-0.3±1.2	-0.4±2.0	-0.1±1.7	-0.2±1.1
			Mid-gland	0.03±0.19	-0.2±1.2	-0.4±2.3	0.0±2.0	0.0±0.6
			Apex	0.00±0.23	-0.1±1.9	-0.2±2.3	0.1±3.0	-0.1±0.6
			Base	0.04±0.22	-0.4±2.4	-0.5±3.6	0.0±1.9	-0.1±0.6

2.4.4 Source of variability: Semi-automatic segmentation

The key result of this experiment was that the operator has less impact on the algorithm accuracy, as compared to the reference and trainer. For all of the null hypotheses tested by ANOVA, rejection was reported at the $p < 0.05$ level. For the mid-gland, there was a significant effect of the trainer on all the metrics and reference had a significant impact on three of the metrics (ΔV, recall, and precision). There was no significant impact of the operator on the metrics. For the apex, the reference had a significant impact on 4 of the metrics (DSC, ΔV, recall, and precision) and the trainer and operator had no significant impact on the metrics. For the base, the reference had a significant impact on all of the metrics, the operator had significant impact on four metrics (DSC, ΔV, recall and precision) and the trainer had significant impact on two metrics (recall and precision) and a marginally significant ($p < 0.1$) impact on DSC.

2.5 Discussion

In the presented semi-automatic segmentation algorithm, we first trained our algorithm to capture the inter-patient local appearance of the prostate border as well as the prostate shape characteristics on different axial cross-sections. Then for an unseen

www.manaraa.com

MR image, the prostate border was locally defined based on the learned appearance characteristics of the prostate border at the corresponding location. The defined border was regularized on each 2D axial slice using the corresponding 2D shape model obtained from training. Finally, a 3D shape regularization was applied to the result. In the statistical modeling method referred to as an active appearance model [21], the global appearance of the image is used in combination with the shape model of the prostate to segment the prostate. Therefore, an inter-patient internal appearance variation of the prostate gland that could be caused *e.g.* by differently-located prostate tumours or benign prostatic hyperplasia nodules challenges the segmentation. Furthermore, using the combination of the shape and appearance modeling challenges the simultanous shape modeling when there is a local appearance difference between the test image and the appearance model. We addressed this issue by separating the shape model from the appearance based segmentation.

2.5.1 Inter-operator variability: manual segmentation

We observed substantial inter-operator variability in manual segmentation of the prostate on T2W ER MRI (Table 2.1), with differences between operators ranging between 0.7 mm and 3.5 mm in terms of MAD, and between 65% and 96% in terms of DSC, depending on the observer pair and the anatomic location. There was more inter-operator consistency in delineation of the mid-gland, with greater discordance at the apex and base. These results suggest that measured errors for prostate segmentation algorithms on T2W ER MRI may vary substantially as a function of the manual segmentation used as the reference. Therefore, it is challenging to define a “gold standard” for this task. Consequently, segmentation results reported for an algorithm using a single-operator

المنارة للاستشارات
www.manaraa.com

reference may change substantially if a different operator were to delineate a set of reference segmentations on the same data set. This inter-observer variability also renders comparison of algorithm performance challenging when different data sets and reference segmentations are used in different published results. One approach to mitigate this effect is to evaluate algorithm performance against multiple expert reference segmentations, and assess the algorithm's segmentation error in the context of inter-operator variability in manual segmentation on the same data set.

2.5.2 Accuracy and inter-operator variability: Semi-automatic segmentation

For comparison with previous work, we conducted a single-operator, single-reference experiment measuring the accuracy of our presented semi-automatic segmentation algorithm. Some statistically significant differences were found with respect to other published methods (Table 2.2) but were within the observed ranges of human expert variability in manual delineation on our data set (Table 2.1). Concordant with previously published results, our data show that segmentation of the apex and (especially) the base is considerably more challenging than the segmentation of the mid-gland, with errors contributed not only by the unusual shape and appearance of these structures in some patients (e.g. the shape of the base on the manual segmentation shown in the right-hand panel of Figure 2.4), but also by the substantial variability we measured in experts' selections of the apex-most and base-most slices. Our results in Table 2.2 compare favorably with many of the segmentation error metric values reported from the PROMISE12 challenge [16]; however, the different nature of the data sets in terms of ER coil usage challenges the interpretation of this comparison. For the prostate as a whole, our algorithm and the top performing algorithms in PROMISE12 appear to be

www.manaraa.com

asymptotically approaching human performance as reflected by inter-observer variability in manual contouring. These observations suggest that further improvement of algorithms for computer-assisted segmentation of the mid-gland are unlikely to provide measurable impact, and that efforts toward improved accuracy and consistency in prostate apex and base segmentation are a higher priority. Informal observations of our results suggest negligible impact of magnetic field strength on segmentation error; however, this would be an interesting area of future research on a larger data set of 1.5 Tesla and 3.0 Tesla images.

We observed maximum inter-operator differences in apex-most and base-most slice selection of 12.5 mm at the apex and 8.4 mm at the base on average; these represent aggregates of the largest distances one might observe between these observers' apex and base-most slices, respectively. On the other hand, when the mean surface-to-surface distances (i.e. the MAD values) were calculated for the same nine observers using the semi-automatic algorithm (Table 2.3) and for a subset of three of the observers doing manual contouring (Table 2.1), smaller values (~2–4 mm) were observed. Although the mean surface-to-surface distances would be expected to be smaller than the measured maxima (as observed), the magnitudes of the observed differences in our data suggest that the bulk of the surfaces at the apex and base are in better spatial agreement than are the extrema of the prostate, which cover a relatively smaller surface area and thus have less influence on the calculated MAD metric values. Thus, there appear to be spatial differences in terms of where most of the inter-observer variability lies; there is greater variability in localizing the superior-most end of the apex and inferior-most end of the base, compared to the variability in contouring the apex and base as a whole.

المنارة للاستشارات
www.manaraa.com

Although our use of signed prostate volume differences as well as recall and precision rates as complementary evaluation metrics is unusual with respect to previously published work in this area, these metrics can be helpful in distinguishing different types of segmentation errors in a way that could facilitate the understanding the clinical applicability of the algorithm and facilitate adoption. For instance, in the lower-most segmentation shown in Figure 2.5, the algorithm over-contours the apex and under-contours the base overall. These two errors are quite different in terms of potential clinical impact; for example, in a radiation oncology context, under-contouring could result in untreated cancer whereas over-contouring could result in radiation damage to surrounding healthy tissue. The table adjacent to this example in Figure 2.5 indicates that the MAD and DSC metrics, frequently reported in previous literature, are nearly identical for the apex and base. However, for the apex, recall is substantially larger than precision, and vice-versa for the base, capturing the nature of this segmentation error. The ΔV metric also directly captures this error in a complementary fashion.

Our nine-operator experiment resulted in small degradations of accuracy (Table 2.3), although the results were still within the range of expert variability in manual segmentation (Table 2.1). Our three-operator experiment directly comparing segmentation error and variability for manual and semi-automated segmentation of the same 10 images compared to the same STAPLE reference standard indicated an increase in error with a concomitant decrease in variability when the semi-automated tool was used. This suggests the presence of a tradeoff between segmentation accuracy and variability that is related to the use of automation; computer-assisted delineation may increase the consistency of segmentations at the expense of some accuracy. This lost

المنارة للاستشارات
www.manaraa.com

accuracy could in principle be recovered through minor segmentation editing, but this remains to be tested. The observed reduction in segmentation variability also suggests that the semi-automatic segmentation tool could be valuable in the hands of the novice radiologist or radiation oncologist, providing useful guidance in the form of a segmentation that is consistent with a training set constructed based on segmentations provided by experienced experts.

2.5.3 Sensitivity to initialization: Semi-automatic segmentation

Our data indicate that variability in the semi-automatic segmentation results arising from varying the inputs to the algorithm (Table 2.5) is substantially smaller than the segmentation accuracy and variability observed for both manual and semi-automatic segmentation. This suggests that the algorithm is robust to the placement of center points and orientation of the gland by the user, and helps to explain the accuracies we obtained despite minimal user interaction; users do not need to exercise a high degree of time-consuming accuracy and precision in interacting with this tool. Moreover, since the prostate APSA is very close to the image AP axis, it might be possible to replace it with the image AP axis and minimize the user interaction without loss of accuracy.

2.5.4 Source of variability: Semi-automatic segmentation

Our ANOVA test results indicated that in all regions of the prostate, the reference segmentation used for evaluation had the most significant impact on segmentation error. This reinforces our earlier observation (Section 2.5.1) that measurements of a segmentation algorithm’s performance based on single reference segmentation could vary

www.manaraa.com

substantially according to the particular expert reference segmentation used for evaluation.

With the exception of the base region, our test did not detect a significant impact of the operator on any of the error metrics. This suggests that the use of the proposed semi-automated segmentation tool could result in improved inter-operator consistency of mid-gland and apex delineations, but that further work would be useful to improve the consistency of delineation of the challenging area of base, where the prostate meets the bladder neck.

The significant impact of the trainer on all the metrics (within the mid-gland), in conjunction with the above observation regarding the impact of the operator, suggests that in this region the semi-automatic segmentation algorithm might provide outputs that mimic the trainer more than the operator. Hence, this tool could be useful in the hands of an expert trainer and a relatively more novice operator. Further work involving a larger sample size will be required to elucidate the impact of the trainer on the apex and base regions.

2.5.5 Limitations

The results of this work must be considered in the context of its strengths and limitations. To the best of our knowledge, this work represents the first use of the complementary MAD, DSC, recall, precision and ΔV error metrics in the evaluation of a prostate segmentation algorithm for T2W ER MRI, using multiple operators and multiple reference standard segmentations. However, our study was limited in several ways. First, our sample size (42 images for the single-operator experiment and 10 images for the multiple-operator experiment) is small and therefore the results of this study should be

المنارة للاستشارات
www.manaraa.com

considered to be hypothesis-generating, and our conclusions should be interpreted accordingly. Second, the only MR appearance information used by our segmentation algorithm to delineate the border is MR image intensity; no derived quantities such as image texture measures were utilized. Although using such features may add complexity and computation time to the method, such an approach could yield improved results, especially in the context of high variability of shape of the base and apex where our shape regularization step is less applicable. Third, we did not provide the operators with the opportunity to edit the semi-automatic segmentations to their satisfaction; an interesting avenue of further work would be to measure the time required for the user to obtain a satisfactory segmentation using the output of the semi-automatic tool as a starting point. Finally, since all of the images in our data set are from patients with confirmed PCa, the appearance of the prostate could have been locally modified in the presence of lesions near the capsule, increasing the challenge of accurate prostate segmentation using a local model of appearance; thus, our patient selection may have pessimistically affected our reported semi-automatic segmentation results.

2.5.6 Conclusions

We presented a comprehensive evaluation of a 3D segmentation algorithm for prostate T2W ER MRI, comprising boundary-, region-, and volume-based metrics computed separately for the whole gland, mid-gland, apex and base. We tested the algorithm using multiple reference segmentations and multiple operators, and observed reduced inter-operator variability via the use of this semi-automated tool. Minimal operator interaction of less than 30 sec, on average, was required. Based on our results, further work in this area should be focused on improving segmentation accuracy and

المنارة للاستشارات
www.manaraa.com

variability at the prostatic base and apex, including reducing inter-observer variability in selecting the apex-most and base-most slices of the prostate. Due to high inter-operator variability in the manual prostate segmentation, particularly at the apex and base, it appears to be challenging to interpret reported improvements in segmentation algorithm accuracy based on a single-operator manual reference standard. We anticipate that our comprehensive approach to segmentation evaluation will facilitate the assessment and adoption of our algorithm by clinical end users, who can interpret the segmentation metrics as appropriate to their clinical use cases of interest.

2.6 References

1. R. Siegel, D. Naishadham and A. Jemal, "Cancer statistics, 2012," CA Cancer J Clin **62**, 10-29 (2012).

2. M. G. Jameson, L. C. Holloway, P. J. Vial, S. K. Vinod and P. E. Metcalfe, "A review of methods of analysis in contouring studies for radiation oncology," J Med Imaging Radiat Oncol **54**, 401-410 (2010).

3. F. A. Jolesz, A. Nabavi and R. Kikinis, "Integration of interventional MRI with computer-assisted surgery," J Magn Reson Imaging **13**, 69-77 (2001).

4. H. Hricak, P. L. Choyke, S. C. Eberhardt, S. A. Leibel and P. T. Scardino, "Imaging prostate cancer: a multidisciplinary perspective," Radiology **243**, 28-53 (2007).

5. M. Fuchsjager, A. Shukla-Dave, O. Akin, J. Barentsz and H. Hricak, "Prostate cancer imaging," Acta Radiol **49**, 107-120 (2008).

6. B. N. Bloch, R. E. Lenkinski and N. M. Rofsky, "The role of magnetic resonance imaging (MRI) in prostate cancer imaging and staging at 1.5 and 3 Tesla: the Beth Israel Deaconess Medical Center (BIDMC) approach," Cancer Biomark **4**, 251-262 (2008).

7. S. W. Heijmink, J. J. Futterer, T. Hambrock, S. Takahashi, T. W. Scheenen, H. J. Huisman, C. A. Hulsbergen-Van de Kaa, B. C. Knipscheer, L. A. Kiemeney, J. A. Witjes and J. O. Barentsz, "Prostate cancer: body-array versus endorectal coil MR imaging at 3 T--comparison of image quality, localization, and staging performance," Radiology **244**, 184-195 (2007).

المنارة للاستشارات
www.manaraa.com

8. Y. Kim, I. C. Hsu, J. Pouliot, S. M. Noworolski, D. B. Vigneron and J. Kurhanewicz, "Expandable and rigid endorectal coils for prostate MRI: impact on prostate distortion and rigid image registration," Med Phys **32**, 3569-3578 (2005).

9. M. Hirose, A. Bharatha, N. Hata, K. H. Zou, S. K. Warfield, R. A. Cormack, A. D'Amico, R. Kikinis, F. A. Jolesz and C. M. Tempany, "Quantitative MR imaging assessment of prostate gland deformation before and during MR imaging-guided brachytherapy," Acad Radiol **9**, 906-912 (2002).

10. W. L. Smith, C. Lewis, G. Bauman, G. Rodrigues, D. D'Souza, R. Ash, D. Ho, V. Venkatesan, D. Downey and A. Fenster, "Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR," Int J Radiat Oncol Biol Phys **67**, 1238-1247 (2007).

11. S. Martin, V. Daanen and J. Troccaz, "Atlas-based prostate segmentation using an hybrid registration," Int J CARS **3**, 8 (2008).

12. S. Vikal, S. Haker, C. Tempany and G. Fichtinger, "Prostate contouring in MRI guided biopsy," Proc SPIE **7259**, 72594A (2009).

13. L. R. Dice, "Measures of the amount of ecologic association between species," Ecology **26**, 297-302 (1945).

14. R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: application to prostate MRI segmentation," IEEE Trans Med Imaging **31**, 1638-1650 (2012).

15. S. Liao, Yaozong Gao, Yinghuan Shi, Ambereen Yousuf, Ibrahim Karademir, Aytekin Oto, and Dinggang Shen, "Automatic prostate MR image segmentation with sparse label propagation and domain-specific manifold regularization," Information Processing in Medical Imaging, 511-523 (2013).

16. G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," Med Image Anal **18**, 359-373 (2014).

17. T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active shape models-their training and application," Computer vision and image understanding **61**, 38-59 (1995).

18. J. C. Gower, "Generalized procrustes analysis," Psychometrika **40**, 33-51 (1975).

المنارة للاستشارات

www.manaraa.com

19. S. K. Warfield, K. H. Zou and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans Med Imaging **23**, 903-921 (2004).

20. R. F. Woolson and W. R. Clarke, *Statistical methods for the analysis of biomedical data*. (John Wiley & Sons, 2011).

21. T. F. Cootes, G. J. Edwards and C. J. Taylor, "Active appearance models," in *Computer Vision—ECCV'98,* (Springer, 1998), pp. 484-498.

www.manaraa.com

Chapter 3.

Accuracy and acceptability of an automated method for prostate segmentation in magnetic resonance imaging †

3.1 Introduction

Prostate cancer (PCa) is the most commonly diagnosed cancer in men in North America, excluding skin carcinoma. More than 30,000 deaths from PCa are predicted in the United States and Canada for 2015 [1, 2]. Magnetic resonance (MR) imaging (MRI), due to its promising potential in diagnosis and staging of PCa [3, 4], is one of the imaging modalities utilized in multiple emerging diagnosis and therapeutic procedures. Contouring of the prostate on MRI could assist with PCa diagnosis and therapy planning. More specifically, T2-weighted (T2w) MRI is superior to other MRI sequences for anatomic depiction of the prostate gland and the surrounding tissues [5]. The use of an endorectal (ER) receive coil helps MRI acquisition performance in terms of image quality and spatial resolution [6]. However, it deforms and displaces the prostate gland [7], produces some ER coil-based imaging artifacts [8], and detects more edges and details that challenge the adaptation of computer-assisted prostate contouring algorithms designed for non-ER MRI to this context.

---

† A version of this chapter has been submited: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, A. Fenster, and A. D. Ward, " Accuracy and acceptability validation of an automated method for prostate segmentation in magnetic resonance imaging," Medical Physics.

www.manaraa.com

Manual segmentation of prostate MRI is a laborious and time-consuming task that is subject to inter-observer variability [9]. This motivates the need for fast and reproducible segmentation algorithms for T2w ER prostate MRI. There have been several algorithms published in the literature for segmentation of the prostate on T2w ER MRI. Martin *et al.* [10] presented a semi-automatic algorithm for segmentation of the prostate on MRI based on registration of an atlas to the test image. They evaluated their method on 17 MR images using manual segmentations performed by a single operator as the reference standard. To measure the accuracy of their method, they used a surface-based metric for different regions of interest (ROIs) including the whole prostate gland, base, midgland and apex regions. They also used region based metrics, but for the whole gland only. They reported higher atlas registration error, yielding to higher segmentation error, for their methods on small prostates (less than 25 cm^3) compared to the atlas registration error on the larger prostates. Vikal *et al.* [11] developed a two-dimensional (2D) slice-by-slice segmentation algorithm based on shape modeling for three-dimensional (3D) segmentation of the prostate on T2w MRI. Their semi-automatic method was initialized by user selection of prostate centre point on one of the central slices of the prostate. In their method, segmentation starts from the selected central slice. The segmentation on each 2D slice is used as an initialization for segmenting its adjacent slice. They evaluated their method on three images using the mean absolute distance (MAD) and Dice similarity coefficient [12] (DSC), compared to a single reference standard developed by consensus of two expert observers. Toth and Madabhushi [13] developed a semi-automatic segmentation algorithm based on a landmark-free active appearance model and level set shape representation method. To evaluate their method they applied the

www.manaraa.com

algorithm to 108 T2w ER MRI and compared the results to manual segmentations performed by one observer using the MAD for the whole gland only and the DSC for whole-gland, apex, midgland and base. Although results were reported for a second observer on a subset of 17 images, inter-observer variability of their method was not reported. Liao *et al.* [14] presented a coarse-to-fine hierarchical automatic segmentation algorithm for prostate segmentation on T2w MRI. They used the MAD, DSC and Hausdorff distance error metrics for evaluation of their method on the whole gland using a manual reference segmentation performed by one observer on 66 T2w MR images. Cheng *et al.* [15] developed an automatic approach consisting of two main steps: first, a coarse segmentation based on an adaptive appearance model and then a segmentation refinement using a support vector machine. They used region-based metrics computed only within the whole gland to evaluate their method, using manual reference segmentations verified by one radiologist. In 2012, 11 teams were involved in a challenge for prostate MRI segmentation, called PROMISE12, held as part of the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference. The challenge tested the performance of the segmentation algorithm presented by each team in two steps; online and live challenges. The image data set used by the challenge contained both ER and non-ER MR images and the results were evaluated against one set of manual segmentations provided by one expert and reviewed and edited, if necessary, by another expert using surface-, region- and volume-based metrics for the whole gland, apex and base regions [16].

In most previously published work, the segmentation result has been evaluated by comparison against a single manual reference segmentation. However, there is high inter-

للاستشارات المنارة
www.manaraa.com

observer variability in contouring the prostate in MRI [9] and changing the manual reference segmentation used for segmentation evaluation likely has a substantial impact on the reported segmentation performance. Therefore, it is necessary to consider this variation when validating segmentation algorithms. Furthermore, in most published studies, the algorithm results have been evaluated using only one or two error metrics. Since each metric is sensitive to certain types of errors (e.g. the MAD is sensitive to large, spatially localized errors, whereas the DSC is sensitive to smaller, global errors), there is not a single globally-accepted metric for comprehensive evaluation of segmentation algorithms. Thus, using a set of metrics that are sensitive to different types of error such as surface disagreement, regional misalignment, and volume differences, yields a more comprehensive algorithm evaluation. Moreover, the accuracy and repeatability of the prostate segmentation varies for different parts of the gland in manual [9] and computer-based [10, 11, 13, 17] segmentations. Some groups reported segmentation error only for the whole prostate gland without reporting the error for the gland subregions such as the apex, mid-gland and base. Segmentation error metrics computed for the whole gland are challenging to interpret, since large errors in the apex and base regions can be offset by smaller errors in the mid-gland. When the segmentations are used to guide radiation or ablative interventions, this is especially important since the apex and base are near to sensitive structures such as the bladder, urethra, and penile bulb.

We previously described a semi-automatic segmentation approach for ER prostate MRI based on local appearance and shape characteristics and evaluated its performance in comparison with manual segmentation in terms of accuracy and inter-operator

المنارة للاستشارات
www.manaraa.com

variability [17]. We applied our evaluation using different types of error metrics (i.e. surface-, region- and volume-based metrics) and assessed the performance of the algorithm over the whole prostate gland as well as within the apex, midgland and base subregions. Our semi-automatic segmentation method required that the user select four initial points to run the contouring algorithm. Thus, the algorithm's segmentation results depended on the user's judgment of the correct loci for these points. This included a requirement that the user indicate the apex-most and base-most slices of the prostate, which is a challenging task with substantial inter-observer variability.

Although many segmentation algorithms have been proposed, an operator-independent algorithm that has been comprehensively validated using multiple complementary error metrics against a multi-observer reference standard remains elusive. In this chapter, we build on our previous semi-automated segmentation algorithm to develop a fully automated approach that has no dependence on user input. We compare the fully automatic segmentation performance to the semi-automatic and manual approaches. We address the following four research questions in this chapter. (1) What is the accuracy of the automated segmentation algorithm when compared to a single-observer manual reference standard? (2) What is the difference in the time required to use our automated segmentation algorithm and our semi-automated segmentation algorithm? (3) What is the difference in accuracy between our automated segmentation algorithm and our semi-automated segmentation algorithm? (4) Is the measured misalignment between the computer-assisted segmentations and manual segmentations within the range of inter-expert variability in manual segmentation?

المنارة للاستشارات
www.manaraa.com

3.2 Materials and Methods

3.2.1 Materials

The data set contained 42 axial T2w fast spin echo ER MR images acquired from patients with biopsy-confirmed PCa. 23 of the images were acquired with TR = 4000–13000 ms, TE 156–164 ms, NEX = 2, and for the other 19 images TR = 3500–7320 ms, TE = 102–116 ms, NEX = 1–2. Nine and 33 images were obtained with 1.5 and 3.0 Tesla field strengths, respectively. The voxel sizes varied from 0.27 × 0.27 × 2.2 mm to 0.44 × 0.44 × 6 mm, covering the range typically seen in clinical prostate MRI. Four different MRI scanners were used for image acquisition: MAGNETOM Avanto, MAGNETOM Verio (Siemens Medical Solutions, Malvern, PA), Discovery MR 750 and Signa Excite (General Electric Healthcare, Waukesha, WI). The study was approved by the research ethics board of our institution, and written informed consent was obtained from all patients prior to enrolment. All 42 MR images were initially segmented manually by one observer (either a radiologist or a graduate student under advisement of a radiologist) followed by review and adjustment of the contours by an expert senior radiology resident with experience reading >100 prostate MRI scans. Two additional manual segmentations were performed on a subset of 10 images performed by two expert observers (one radiologist and one radiation oncologist). To select this subset of 10 images, we qualitatively assigned easy-, moderate- and difficult-to-segment labels to a set of images acquired at our institution and randomly select 10 images from all the three categories. The prostate volumes in the data set calculated based on the available manual segmentations ranged from 15 to 89 cm^3 with mean ± standard deviation of 35 ± 14 cm^3.

المنارة للاستشارات
www.manaraa.com

3.2.2 Automated segmentation

Our automatic segmentation approach consists of two main parts: training and segmentation, described in sections 3.2.2.1 and 3.2.2.2, respectively. In this chapter, we focus on automation of the manual steps of our previously-published semi-automated method. Thus, we describe elements common to our automatic and semi-automatic approaches at a high level; full details on these elements are available in [17].

3.2.2.1 Training

We use the approach to training described in reference [17] reporting on our semi-automated segmentation method. The training method is described at a high-level here. During training, the algorithm learns the local appearance of the prostate border by extracting 36 locally defined circular mean intensity image patches, and generates a 2D statistical shape model for the prostate on each axial cross-section of the prostate. To extract the mean intensity image patches, we first spatially normalized all the prostates in the training set to define a spatial correspondence between axial slices of all the training images. For each slice in a set of corresponding axial slices, a set of 36 anatomically corresponding points was defined on the prostate border and for each point, a circular patch centered at that point was selected. By computing the average of the intensities of the corresponding pixels across all the patches obtained from the corresponding points, a set of 36 mean intensity patches were generated, each corresponding to one anatomical point on the prostate border. The 36 defined border points were also used for building a statistical point distribution model (PDM) of prostate shape on each selected axial cross-section.

المنارة للاستشارات

www.manaraa.com

3.2.2.2 Segmentation

To segment the prostate in a new MR image, the algorithm first coarsely localizes the region containing the prostate by automatically positioning a template shaped similarly to a prototypical prostate on the mid-sagittal plane (blue polygon in Figure 3.1). The algorithm then searches within a region defined according to this template to define the 3D prostate boundary. This high-level process resolves to a four-step procedure: (1) anterior rectal wall boundary determination, (2) inferior bladder boundary determination, (3) coarse prostate localization by template fitting, and (4) 3D prostate boundary localization. Each of these four steps is described in detail below.

The first step was to fit a line to the anterior rectal wall boundary on the mid-sagittal slice of the MRI. Candidate points lying on the anterior rectal wall boundary were selected by finding loci of minimum first derivative along line intensity profiles oriented parallel to the axial planes and running from anterior to posterior on the mid-sagittal plane. This approach was chosen due to the observation that the intensity generally transitions sharply from bright to dark at the rectal wall boundary. To reduce the search space, we restricted our search to a domain covering 50% of the width of the mid-sagittal plane in the anteroposterior direction, offset 20% from the posterior-most extent of the mid-sagittal plane. Within this domain, 10 equally-spaced lines (every second line) nearest to the mid-axial plane were searched. For robustness to outlier candidate points, we computed a least-trimmed squares fit [18] line to the candidate points, with the optimizer tuned to treat 40% of the candidate points as outliers. We took the resulting best-fit line to represent the anterior rectal boundary (posterior-most yellow dashed line in Figure 3.1).

www.manaraa.com

The second step was to fit a curve to the inferior bladder boundary on the mid-sagittal slice of the MRI. Candidate points lying on the inferior bladder boundary were selected by finding loci of minimum first derivative along line intensity profiles oriented parallel to the anterior rectal boundary determined in the previous step and running from superior to inferior on the mid-sagittal plane. This approach was chosen due to the observation that the intensity generally transitions sharply from bright to dark at the inferior bladder boundary. To reduce the search space we restricted our search to line segments lying within the superior half of the image, starting 5 mm anterior to the rectal wall with 2 mm spacing between them. We eliminated implausible candidate points in two stages. In the first stage, points forming a locally concave shape near the posterior side, inconsistent with anatomy of the inferior aspect of the bladder, were eliminated. In the second stage, we computed a least-trimmed squares fit [18] polynomial curve (second-order curve in the case of a point configuration yielding a convex shape; first-order curve otherwise) to the remaining candidate points, with the optimizer tuned to treat 20% of the candidate points as outliers. We took the resulting curve to represent the inferior bladder boundary (superior-most yellow dashed curve in Figure 3.1).

The third step was to fit the prostate template (described by the dimensions shown in Figure 3.1) to the image using the anatomic boundaries found in the first and second steps. This was done by defining the dimensions of the template to match the anteroposterior (AP) and inferior-superior (IS) dimensions of the prostate on the test image; this information is readily available in every clinical case from the prostate ultrasound examination conducted prior to MRI. The template was then positioned parallel to and 3 mm anterior to the rectal wall line (along a line perpendicular to the

المنارة للاستشارات

www.manaraa.com

rectal wall line), inferior to the bladder boundary curve with a single point of contact between the bladder boundary and the template (Figure 3.1).

The fourth and final step was to define the 3D surface of the prostate detected and localized by the template. After fitting the template to the image, we extract a set of three points (three blue crosses in Figure 3.1) from the template: the prostate centre points on (1) the apex-most slice, (2) the base-most slice, and (3) the midgland slice equidistant to the apex- and base-most slices. We then interpolate these three centre points using piecewise cubic interpolation to estimate the centre points for all of the axial slices between the apex and base. We then use the approach to prostate boundary localization described in reference [17] reporting on our semi-automated segmentation method. The approach is described at a high level here. For each slice, we oriented a set of 36 equally spaced rays emanating from the centre point, one corresponding to each of the learned mean intensity patches. For each ray we translated the corresponding mean intensity patch to find the point whose circular image patch has the highest normalized cross-correlation with the corresponding mean intensity path. Shape regularization was performed within each slice using the corresponding PDM, followed by 3D shape regularization. Full details are available in [17].

المنارة للاستشارات

www.manaraa.com

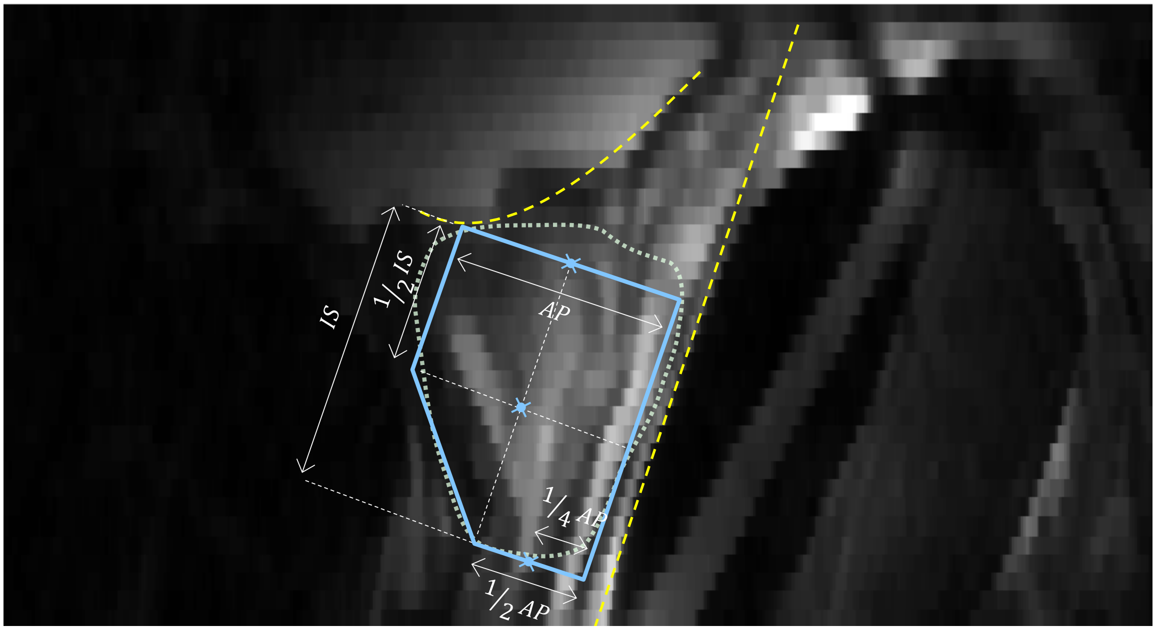


Figure 3.1: Automatic coarse localization of the prostate. The dashed line shows the estimated tangent line to the rectal wall. The dashed curve shows the estimated bladder border. The solid line polygon is the template used to select the centre points for apex, midgland and base. The prostate border based on manual segmentation has been overlaid in dotted line as a reference. AP and IS are ,respectively, anterioposterior and inferior-superior dimensions of the prostate measured during routine clinical ultrasound imaging. The three indicated points on the template define the three estimated centre points for the prostate.

3.2.3 Validation

To evaluate the accuracy of the segmentation algorithm, we used complementary boundary-based, regional overlap-based, and volume-based metrics. This allows the user of the method to understand its applicability to a specific intended workflow. For instance, the use of this algorithm for planning whole-prostate radiation would increase the importance of low error in a boundary-based metric, whereas the use of the algorithm in a retrospective study correlating prostate size with clinical outcome would focus on accuracy of a volume-based metric. We used the MAD as the boundary-based error metric; the DSC, recall rate and precision rate as regional overlap-based error metrics;

المنارة للاستشارات
www.manaraa.com

and the volume difference (ΔV) metric to evaluate the automatic segmentation against manual segmentation. We measured the metrics in 3D for the whole prostate gland and also for the inferior-most third of the gland (corresponding to the apex region), the middle third of the gland (corresponding to the midgland region) and the superior-most third of the gland (corresponding to the base region).

MAD measures the misalignment of two surfaces in 3D in terms of absolute Euclidean distance. To calculate the MAD in a unilateral fashion, the surface of each shape is defined as a set of points, with one of the two shapes designated as the reference. The MAD is the average of the absolute Euclidean distances between each point on the non-reference set to the closest point on the reference set. Specifically,

$$MAD(X,Y) = \frac{1}{N}\sum_{p\in X} \min_{q\in Y} D(p,q)\,, \tag{3.1}$$

where X and Y are the point sets (Y is the reference set), N is the number of points in X, p is a point in X, q is a point in Y, and $D(p,q)$ is the Euclidean distance between p and q.

The MAD is an oriented metric and is therefore not invariant to the choice of reference shape. This can be addressed by calculating the bilateral MAD, which is the average of the two unilateral MAD values calculated taking each shape as the reference.To calculate the DSC [12], recall rate and precision rate [17], we measured the volume overlap between the two 3D shapes. Figure 3.2 and equations (3.2), (3.3) and (3.4) define DSC, recall and precision, respectively.

$$DSC(X,Y) = \frac{2(X\cap Y)}{X+Y} = \frac{2TP}{FP+2TP+FN} \tag{3.2}$$

$$Recall(X,Y) = \frac{TP}{TP+FN} \tag{3.3}$$

www.manaraa.com

$$Precision(X,Y) = \frac{TP}{TP + FP} \tag{3.4}$$

We subtract the volume of the reference shape from the volume of the test shape to calculate the signed volume difference (ΔV) metric

$$\Delta V(X,Y) = V_{algorithm} - V_{reference}\,, \tag{3.5}$$

where $V_{algorithm}$ and $V_{reference}$ are the prostate volumes given by the segmentation algorithm and manual reference segmentation, respectively. Negative and positive values of ΔV indicate under-segmentation and over-segmentation, respectively.

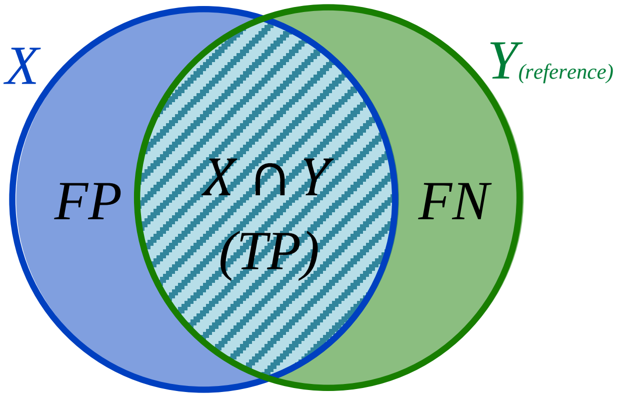


Figure 3.2: Elements used to calculated the DSC, recall, and precision validation metrics. X and Y are the two shapes, with Y taken as the reference shape. FP: false positive, TP: true positive, FN: false negative.

3.3 Experiments

For all of the experiments in this chapter, all algorithm parameters were tuned identically to those used in reference [17] to allow for direct comparison of the results.

3.3.1 Comparison of automatic and semi-automatic segmentation: accuracy and time

We ran the automatic segmentation algorithm on our data set of 42 3D images and compared the results to a single manual reference segmentation using leave-one-patient-

المنارة للاستشارات
www.manaraa.com

out cross validation. We compared each segmentation result against the reference using our five error metrics on the four ROIs; the whole gland, apex, midgland and base regions. We applied one-tailed heteroscedastic *t*-tests [19] to compare the performance of the automatic segmentation to the semi-automatic segmentation. We measured the average execution time for the automatic segmentation approach across the 42 images and compared it to the average of semi-automatic execution time across the same data set and identical running conditions, using a one-tailed *t*-test.

3.3.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation

We ran the automatic algorithm on the subset of 10 images for which we had three manual reference segmentations. For comparison, we also applied our semi-automatic algorithm [17] to the same data set using nine different operators (four radiation oncologists, one radiologist, one senior radiology resident, one imaging scientist, and two graduate students, all with clinical and/or research experience with prostate imaging). We used the remaining 32 images for training both algorithms. We compared each segmentation result against the manual reference segmentations using our five error metrics on the four ROIs; the whole gland, apex, midgland and base regions.

For the automatic segmentation method, we calculated the mean and standard deviation of each metric for each ROI across all 10 images and three references, defined as

$$\bar{M}^{a}_{Metric} = \frac{1}{M \times K} \sum_{i=1}^{M} \sum_{j=1}^{K} Metric\left(L_i^a, L_i^k\right) \text{ and } \tag{3.6}$$

المنارة للاستشارات
www.manaraa.com

$$\sigma^{a}_{Metric} = \sqrt{\frac{1}{(M \times K - 1)} \sum_{i=1}^{M} \sum_{j=1}^{K} \left[Metric\left(L_i^a, L_i^k\right) - \bar{\mathcal{M}}_1^a \right]^2}, \tag{3.7}$$

where $Metric$ is a function computing any one of the five metrics (e.g. MAD); $\bar{\mathcal{M}}^{a}_{Metric}$ is the mean value of the metric for automatic segmentation across all the images and all the references; $\sigma^{a}_{\mathcal{M}1}$ is the standard deviation of the metric; *M=10* and *K=3* are the number of images and references, respectively; L_i^k is the manual segmentation by the *k*th operator on the *i*th image; and L_i^a is the automatic segmentation on the *i*th image. For the semi-automatic segmentation, we calculated the mean and standard deviation of each metric for each ROI across all 10 images, three references and nine operators, defined as

$$\bar{\mathcal{M}}^{s}_{Metric} = \frac{1}{M \times N \times K} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} Metric\left(L_i^{sj}, L_i^k\right) \text{ and} \tag{3.8}$$

$$\sigma^{s}_{Metric} = \sqrt{\frac{1}{(M \times N \times K - 1)} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{K} \left[Metric\left(L_i^{sj}, L_i^k\right) - \bar{\mathcal{M}}_1^s \right]^2}, \tag{3.9}$$

where $\bar{\mathcal{M}}^{s}_{Metric}$ is the mean value of the metric across all the semi-automatic labels, all the images and all the references; $\sigma^{s}_{\mathcal{M}1}$ is the standard deviation of the metric; *N*=9 is the number of operators; and L_i^{sj} is the semi-automatic segmentation by the *j*th operator on the *i*th image.

We used Simultaneous Truth and Performance Level Estimation (STAPLE) [20] to generate one reference segmentation from each triplet of manual segmentations performed on each image. We then computed $\bar{\mathcal{M}}^{a}_{Metric}$ and $\bar{\mathcal{M}}^{s}_{Metric}$ using the STAPLE reference exactly as in Equations (3.6)–(3.9), with $K = 1$ (reflecting the use of a single STAPLE reference rather than 3 manual references).

المنارة للاستشارات
www.manaraa.com

We compared the semi-automatic and automatic approaches separately for both explained scenarios (three manual references and single STAPLE reference) using one-tailed heteroscedastic *t*-tests. We defined the range of mean values of each metric ($B^{L}_{Metric}\ to\ B^{H}_{Metric}$) when we compared three manual segmentations pairwise reported in [17] as follows:

$$B^{L}_{Metric} = \min_{m,n} \bar{\mathcal{M}}^{m,n}_{Metric} \text{ and} \tag{3.10}$$

$$B^{H}_{Metric} = \max_{m,n} \bar{\mathcal{M}}^{m,n}_{Metric}, \tag{3.11}$$

Where

$$\bar{\mathcal{M}}^{m,n}_{Metric} = \frac{1}{M}\sum_{i=1}^{M} Metric(L^{m}_{i}, L^{n}_{i}) \text{ and} \tag{3.12}$$

L^{m}_{i} and L^{n}_{i} are the manual segmentations for *i*th image by observers *m* and *n*, respectively. Wecompared the mean metric values for semi-automatic and automatic segmentation ($\bar{\mathcal{M}}^{a}_{Metric}$ and $\bar{\mathcal{M}}^{s}_{Metric}$) to this range. If the average of a metric at one ROI is within this manual segmentation variability range or even the observed average error is below the range, we took it into account as an improvement in accuracy and variability of the algorithms compared to manual segmentation.

3.4 Results

3.4.1 Comparison of automatic and semi-automatic segmentation: accuracy and time

The results in this section address research questions (1), (2), and (3) as described in the introduction. Table 3.1 shows our automatic segmentation accuracy on 42 T2w MR images against one manual reference segmentation. The results of the *t*-tests (with

www.manaraa.com

α=0.05) showed that using the automatic algorithm significantly increased the error in terms of MAD and DSC in all the ROIs. Recall rates significantly decreased for the whole gland, apex and midgland and significantly increased for the base when we used the automatic segmentation algorithm. The precision rate also showed more error within the whole gland, midgland and base. No significant changes were detected within the apex in terms of the precision rate. We did not detect a significant increase in error for the whole gland and midgland in terms of ΔV. The absolute value of ΔV was significantly increased within the apex and significantly decreased within the base.

The mean ± standard deviation execution time using an unoptimized MATLAB platform on a single CPU core for coarse prostate localization was 3.2 ± 2.1 sec. and for 3D segmentation was 54 ± 13 sec.

Table 3.1: Accuracy of automatic segmentation: mean ± standard deviation of MAD, DSC, recall, precision, and ΔV. § and * show statistically significant accuracy gain and loss, respectively, when compared to the results of semi-automatic segmentation in [17] ($p < 0.05$).

Region of Interest	**MAD (mm)**	**DSC (%)**	**Recall (%)**	**Precision (%)**	**ΔV (cm^3)**
Whole Gland	3.2 ± 1.2*	71 ± 11*	69 ± 15*	76 ± 12*	-3.6 ± 10.4
Apex	2.8 ± 1.3*	66 ± 15*	62 ± 23*	81 ± 17	-3.3 ± 5.1*
Midgland	2.8 ± 1.1*	82 ± 9*	82 ± 15*	84 ± 10*	-0.5 ± 3.6
Base	3.8 ± 1.7*	64 ± 15*	71 ± 21§	69 ± 22*	0.2 ± 7.5§

3.4.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation

The results in this section address research question (4) as described in the introduction. In this experiment, the key result was that the accuracy of semi-automatic and automatic segmentation algorithms approaches the observed inter-operator variability range in manual segmentation. Figure 3.3 shows the mean ± standard deviation of the

المنارة للاستشارات

www.manaraa.com

five metric values for each ROI for semi-automatic and automatic segmentation algorithms, compared with the range of the mean of each metric within each ROI in pair-wise comparison of the three manual reference segmentations. Figure 3.4 shows the mean ± standard deviation values for the five metrics for each region of interest for semi-automatic and automatic segmentation algorithms in comparison with STAPLE reference segmentations. We overlaid the results of each metric lower and upper bounds at each ROI in comparison of the three manual reference segmentations against STAPLE reference using dashed lines. Note that in both Figure 3.3 and Figure 3.4 if the metric value for each algorithm located within the range or at the lower error side it means that the algorithm accuracy reached the observed inter-expert observer variation in manual segmentation, and if the metric value located beyond the higher error bound that means there could be still room for improvement of the algorithm accuracy. As these figures show depends on the metric and ROI each of the algorithms might have outperformed the other. In terms of some of the metrics at some of the ROIs no statistically significant difference were detected between semi-automatic and automatic algorithms.

المنارة للاستشارات
www.manaraa.com

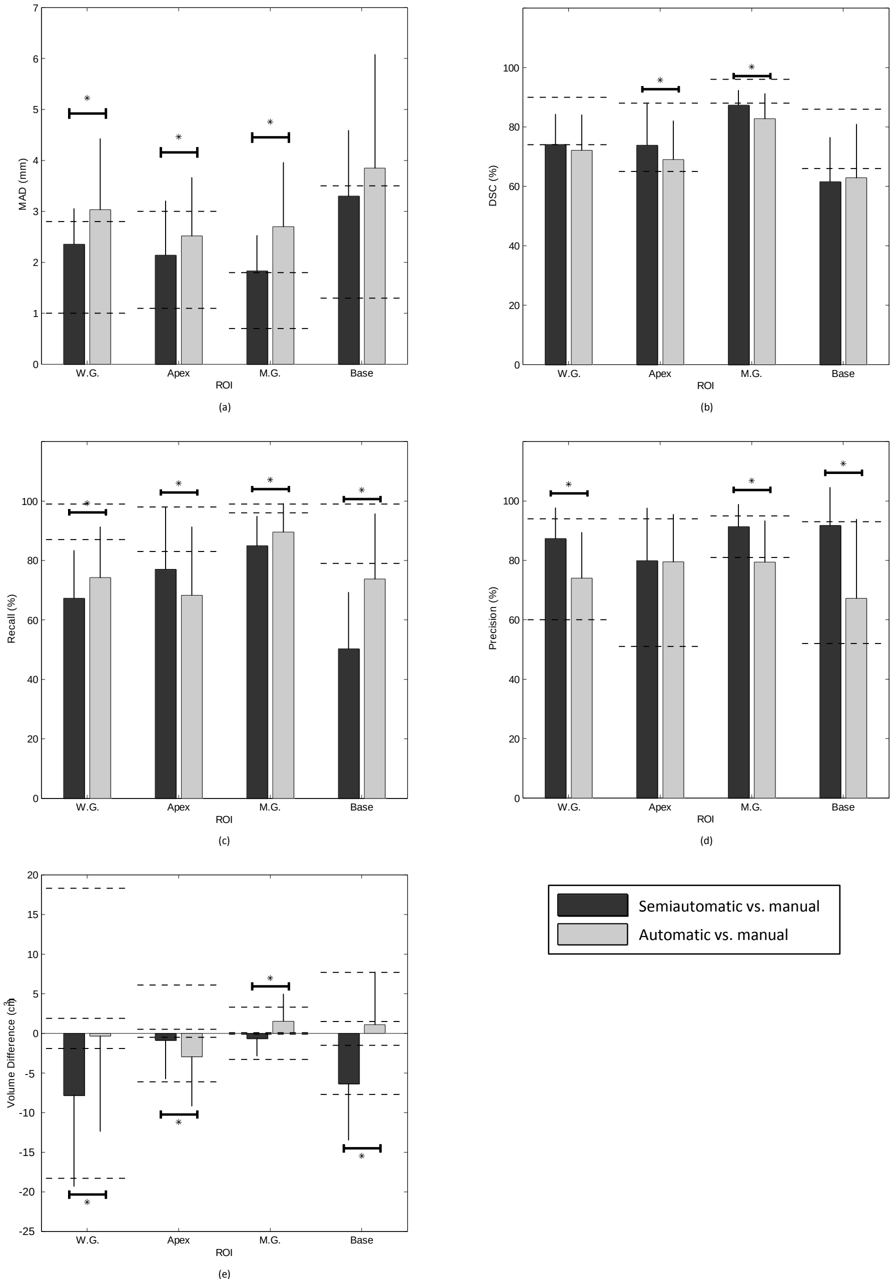


Figure 3.3: Accuracy of the computer-based segmentations vs. inter-operator variability of manual segmentation. The average accuracy of one set of 10 automatic and nine sets of 10 semi-automatic segmentations in comparison with three manual reference segmentations in terms of (a) MAD, (b) DSC, (c) recall, (d) precision and (e) ΔV. The dashed line segments show the observed range of each metric at each ROI in pair-wise comparison between three manual segmentations. For ΔV, the ranges are based on the absolute value of ΔV due to lack of reference in comparison of two manual segmentations. The significant differences detected between semi-automatic and automatic segmentation at different ROIs have been indicated on the graphs (*p*-value < 0.05).

www.manaraa.com

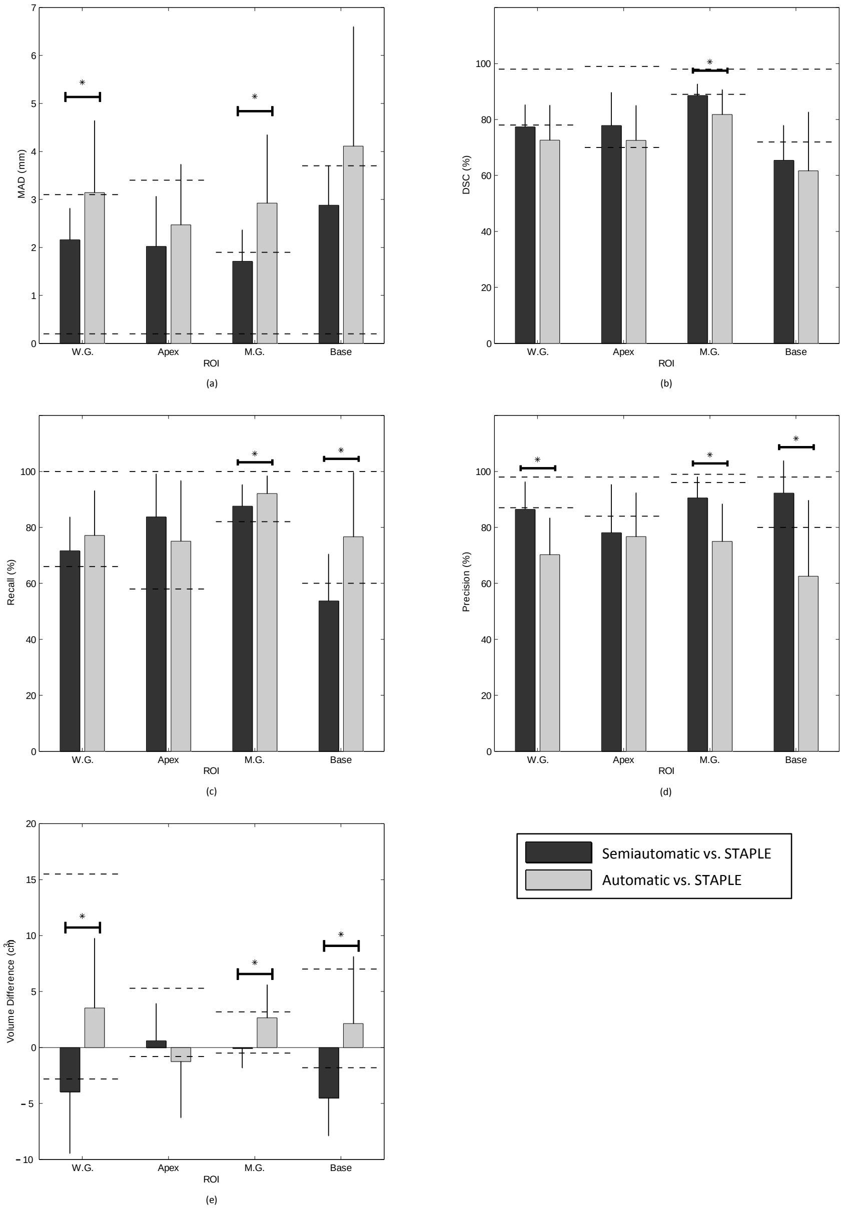


Figure 3.4: Accuracy of the computer-based segmentations vs. inter-operator variability of manual segmentation. The average accuracy of one set of 10 automatic and nine sets of 10 semi-automatic segmentations in comparison with STAPLE reference segmentation in terms of (a) MAD, (b) DSC, (c) recall, (d) precision and (e) ΔV. The dashed line segments show the observed range of each metric at each ROI in comparison between three manual segmentations and STAPLE reference. The significant differences detected between semi-automatic and automatic segmentation at different ROIs have been indicated on the graphs (p-value < 0.05).

المنارة للاستشارات
www.manaraa.com

3.5 Discussion

In this work, we measured the segmentation accuracy gained or lost when using a fully-automatic version of a previously-published semi-automatic segmentation algorithm. Such comparisons are routinely performed in the literature, often using a small number of validation metrics and a single-observer reference standard. In this work, we extended our analysis beyond this traditional approach to include a comparison of the algorithm performance differences to inter-observer variability in segmentation error metrics resulting from different expert manual segmentations. Measuring performance differences between algorithms – those presented in this chapter or in other literature – in the context of expert manual segmentation variability is important to understanding the practical importance of algorithm performance differences.

3.5.1 Comparison of automatic and semi-automatic segmentation: accuracy and time

For comparison to our previous results and other published work, we conducted an experiment using a single manual reference segmentation to measure the accuracy of our automatic algorithm. In terms of most of the metrics, there was a statistically significant difference between automatic and semi-automatic segmentation errors. On average, by switching from semi-automatic segmentation to automatic segmentation, MAD increases by 1.2 mm, DSC decreases by 11%, recall decreases by 8%, precision decreases by 12%, and the error in prostate volume decreases by 1 cm^3 for the whole gland. According to the results based on our multi-reference and/or multi-operator experiments (Figure 3.3), the absolute value of the average ΔV based on automatic

www.manaraa.com

segmentation on whole gland significantly decreased from approximately 7 cm^3 to less than 1 cm^3. This illustrates the complementary nature of the validation metrics and the varying utility of different segmentations for different purposes. Whereas the automatic segmentations may be less preferable to the semi-automatic segmentations for therapy planning, the automatic segmentations may be preferable for correlative studies involving prostate volume and clinical outcomes.

The nature of the data set used in PROMISE12 challenge is different from our data set in terms of the consistent use of the an ER coil for MRI acquisition; our data set contained only images acquired using the ER coil, whereas the PROMISE12 data set contained a some with and some without the ER coil. However if we compare our results in Table 3.1 to the published results in [16] where applicable, our results are within the range of the metric values reported for the PROMISE12 challenge.

In the semi-automatic approach, the operator provided coarse prostate localization, whereas in the automatic approach, this was done entirely by the algorithm. To compare the time required for this step in both contexts, the mean measured operator interaction time for semi-automatic segmentation was approximately 30 seconds [17], whereas the mean measured time required for automatic coarse prostate localization was measured in this study to be approximately 3 seconds using unoptimized MATLAB code on a single CPU core.

www.manaraa.com

3.5.2 Comparison of automatic and semi-automatic segmentation versus inter-operator variability in manual segmentation

The measured accuracy differences between the automatic and semi-automatic approaches are nearly always smaller than the measured differences between manual observer contours (differences between gray and black bars versus differences between dashed lines on Figure 3.3), and also smaller than the measured differences between manual observer contours and a STAPLE consensus contour (Figure 3.4). This suggests that the performance differences measured between these two algorithms may be less than the differences we would expect when comparing different observers' manual contours.

We observe that the top of the dark gray bar corresponding to the MAD metric in Figure 3.3 for the whole gland lies within the range of variability between expert observers' manual contours. This indicates that on average, the semi-automatic segmentation algorithm's whole-gland segmentation accuracy, as measured by MAD, is within the range of human expert variability in manual contouring. This means that further investment of engineering efforts to improve this metric for this algorithm may not be beneficial to the ultimate clinical workflow, since the algorithm's error is already smaller than the difference that might be observed between expert observers' manual contours. The fact that the top of the light gray bar in the same part of the figure lies higher than the range given by the dashed lines indicates that this is not the case for the fully automatic algorithm; further accuracy improvement in terms of MAD on the whole gland may be warranted, with the caveat that such improvement must be measured using a multi-observer reference standard. Inter-observer variability in manual segmentation

المنارة للاستشارات
www.manaraa.com

would likely mask small improvements in the MAD; this is evidenced by the size of the gap between the dashed lines (1.8 mm), compared to the 0.6 mm improvement in the MAD that would be necessary to yield equal performance to the semi-automatic algorithm. We observe in Figure 3.3 that for the MAD, DSC, precision, and ΔV metrics, algorithm performance is near or within the range of human expert variability; this is the case more often for the semi-automatic algorithm. The performance of the algorithms in terms of the recall metric suggest that overall, both algorithms tend to undersegment the prostate to an extent where there is practically important room for improvement. This is especially true for the base region of the prostate. Interestingly, in terms of the recall metric, the automatic algorithm had statistically significantly better performance than the semi-automatic algorithm for every anatomic region except for the apex, with substantially better performance in the base region. This is concordant with our observations [17] of large inter-observer variability in determining the slice location of the base during initialization of the semi-automated algorithm; determining where the prostate base ends and the bladder neck begins is a challenging task even for expert physicians. The observations made in Figure 3.4, where the range of observer variability relative to a STAPLE reference are shown, are generally concordant with observations made on Figure 3.3.

Taken as a whole, these observations highlight the value of measuring inter-observer variability in manual segmentation, using complementary segmentation error metrics, and measuring segmentation error in different anatomic regions known to pose varying levels of challenge to expert operators and automated algorithms. Analysis of these quantities as performed above allows us to determine the best ways to focus further

www.manaraa.com

engineering efforts to improve automated segmentation algorithms. A clinical end user can identify the segmentation error metrics of greatest relevance to the user's intended application of the algorithm and use the plots in Figure 3.3 to determine whether a particular algorithm's accuracy in terms of those metrics is within the range of human expert variability in manual segmentation. If so, the algorithm is ready to be moved forward for full retrospective validation and then prospective testing within the intended clinical workflow. If the analysis shows there is room for improvement to bring the algorithm within the range of human performance for one or more anatomic regions, further engineering efforts can be specifically focused accordingly. We anticipate that this form of segmentation performance analysis will enrich future studies of automated segmentation algorithms intended for use on the prostate and other anatomic structures, enabling a means for determining the point at which an algorithm is ready to move forward from bench testing toward clinical translation.

3.5.3 Limitations

The results of our work should be considered in the context of its strengths and limitations. First, although the automatic segmentation algorithm does not require any user interaction with the images, it does depend on the IS and AP dimensions of the prostate as determined on the routine clinical ultrasound imaging that is performed as part of guided biopsy before any MRI study would be conducted. In this study, the IS and AP dimensions taken from manual MRI prostate segmentation were used as surrogates for the measurements that would be taken during clinical ultrasound, and the performance sensitivity of the automatic segmentation method to these measurements was not determined. Second, our 3D segmentation algorithm requires the AP symmetry axis of

المنارة للاستشارات
www.manaraa.com

the prostate for orientation information. Since during MRI acquisition the scanner operator aligns the midsagittal plane of the scan to the midsagittal plane of the prostate using localizer scans, we assumed that the AP symmetry axis of the prostate gland is oriented parallel to AP axis of the image and assumed that all three prostate centre points (at the apex, mid-gland and base) are located on the mid-sagittal plane of the image. These assumptions are supported by our observations that segmentation algorithm is robust to perturbations of the AP symmetry axis and centre point selection [17] but nevertheless we felt it important to acknowledge these assumptions. Third, the small size of our data set (42 single-reference images and 10 multi-reference images) limits the strength of the conclusions of our work. Finally, we used MR image intensity as the only image feature for prostate border detection and we did not use other image-derived features such as image texture. Using other features might add complexity to the method and may make the algorithm slower, however it could improve the accuracy of the segmentation. Moreover, to have a more reliable assessment on the segmentation algorithm, we still need to study the effects of post-segmentation manual editing on prostate segmentation time, accuracy and reproducibility; this is the subject of our ongoing work.

3.5.4 Conclusions

In this work, we described an automatic 3D prostate segmentation method intended for use on T2w prostate MRI acquired using an endorectal receive coil. We compared it to a semi-automated algorithm using complementary error metrics separately in the apex, mid-gland, and base. We evaluated the algorithms’ accuracies in the context of expert variability in manual segmentation. We addressed four key research questions

المنارة للاستشارات

www.manaraa.com

described in the introduction section of this chapter, the answers to which are enumerated accordingly here. (1) When compared to a single-observer reference standard, the automatic algorithm has an average MAD of 2.8 mm, DSC of 82%, recall of 82%, precision of 84%, and volume difference of 0.5 cm^3 in the mid-gland. Concordant with results from other published algorithms, accuracy was highest in the mid-gland and lower in the apex and base regions of the prostate. (2) The use of the automated algorithm eliminated the need for 30 seconds of user interaction to perform coarse localization of each prostate, replacing this step with a fully automatic approach requiring no user interaction and needing 3 seconds of computation time. (3) The automatic algorithm's accuracy did not differ from the semi-automatic algorithm's accuracy by more than 1 mm in terms of MAD; 5% in terms of DSC, precision, and recall; and 8 cm^3 in terms of volume. The differences between the automatic and semi-automatic segmentation error metrics were consistently smaller than the differences observed between manual contours performed by experts. (4) The segmentation error metric values were near to or within the range of expert manual segmentation variability for all but the recall metric, especially in the prostatic base. This suggests that for our algorithms, engineering efforts should be focused on further improvement of the segmentation of the base, which is challenging even for human experts. The analysis approach taken in this chapter provides a means for determining the readiness of a segmentation algorithm for translation toward clinical trial for a specific purpose, and for focusing further engineering efforts on the most practically relevant performance issues, supporting eventual achievement of clinical translation.

المنارة للاستشارات
www.manaraa.com

3.6 References

1. R. L. Siegel, K. D. Miller and A. Jemal, "Cancer statistics, 2015," CA Cancer J Clin **65**, 5-29 (2015).

2. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian Cancer Statistics 2015*. 2015

3. J. Kurhanewicz, D. Vigneron, P. Carroll and F. Coakley, "Multiparametric magnetic resonance imaging in prostate cancer: present and future," Curr Opin Urol **18**, 71-77 (2008).

4. A. Shukla-Dave and H. Hricak, "Role of MRI in prostate cancer detection," NMR Biomed **27**, 16-24 (2014).

5. O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino and H. Hricak, "Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging," Radiology **239**, 784-792 (2006).

6. D. J. Gilderdale, N. M. deSouza, G. A. Coutts, M. K. Chui, D. J. Larkman, A. D. Williams and I. R. Young, "Design and use of internal receiver coils for magnetic resonance imaging," Br J Radiol **72**, 1141-1151 (1999).

7. Y. Kim, I. C. Hsu, J. Pouliot, S. M. Noworolski, D. B. Vigneron and J. Kurhanewicz, "Expandable and rigid endorectal coils for prostate MRI: impact on prostate distortion and rigid image registration," Med Phys **32**, 3569-3578 (2005).

8. J. E. Husband, A. R. Padhani, A. D. MacVicar and P. Revell, "Magnetic resonance imaging of prostate cancer: comparison of image quality using endorectal and pelvic phased array coils," Clin Radiol **53**, 673-681 (1998).

9. W. L. Smith, C. Lewis, G. Bauman, G. Rodrigues, D. D'Souza, R. Ash, D. Ho, V. Venkatesan, D. Downey and A. Fenster, "Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR," Int J Radiat Oncol Biol Phys **67**, 1238-1247 (2007).

10. S. Martin, V. Daanen and J. Troccaz, "Atlas-based prostate segmentation using an hybrid registration," Int J CARS **3**, 8 (2008).

11. S. Vikal, S. Haker, C. Tempany and G. Fichtinger, "Prostate contouring in MRI guided biopsy," Proc SPIE **7259**, 72594A (2009).

12. L. R. Dice, "Measures of the amount of ecologic association between species," Ecology **26**, 297-302 (1945).

المنارة للاستشارات

www.manaraa.com

13. R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: application to prostate MRI segmentation," IEEE Trans Med Imaging **31**, 1638-1650 (2012).

14. S. Liao, Y. Gao, Y. Shi, A. Yousuf, I. Karademir, A. Oto and D. Shen, "Automatic Prostate MR Image Segmentation with Sparse Label Propagation and Domain-Specific Manifold Regularization," (Springer, 2013), pp. 511-523.

15. R. Cheng, B. Turkbey, W. Gandler, H. K. Agarwal, V. P. Shah, A. Bokinsky, E. McCreedy, S. Wang, S. Sankineni, M. Bernardo, T. Pohida, P. Choyke and M. J. McAuliffe, "Atlas based AAM and SVM model for fully automatic MRI prostate segmentation," Conf Proc IEEE Eng Med Biol Soc **2014**, 2881-2885 (2014).

16. G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," Med Image Anal **18**, 359-373 (2014).

17. M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, E. Gibson, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster and A. D. Ward, "Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semiautomated methods," Med Phys **41**, 113503 (2014).

18. A. C. Atkinson and T.-C. Cheng, "Computing least trimmed squares regression with the forward search," Statistics and Computing **9**, 251-263 (1999).

19. R. F. Woolson and W. R. Clarke, *Statistical methods for the analysis of biomedical data*. (John Wiley & Sons, 2011).

20. S. K. Warfield, K. H. Zou and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans Med Imaging **23**, 903-921 (2004).

المنارة للاستشارات
www.manaraa.com

Chapter 4.

Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging [†]

4.1 Introduction

In 2014, prostate cancer (PCa) was one of the most commonly diagnosed types of cancer and the second leading cause of death from cancer among men in North America [1, 2]. Due to its high soft tissue contrast, magnetic resonance (MR) imaging (MRI) has demonstrated potential for detection, localization and staging of PCa [3-6] and therefore in several centers MRI is being used for PCa diagnosis, treatment planning and therapy guidance [3, 6-8]. Using an endorectal receiver (ER) coil during MRI acquisition yields images with higher resolution and improved signal-to-noise ratio, with reported positive impact on PCa diagnosis [7, 9, 10].

Delineation of the prostate capsule on MRI is required for several clinical procedures in which MR images are employed. T2-weighted (T2w) prostate MRI plays an important role in anatomy description [11, 12], PCa detection and localization [13], and therefore, prostate contouring is usually performed on T2w MRI. However, three-

---

[†] A version of this chapter is in preparation for submission: M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. D. Ward, " Impact of physician editing on repeatability and time for manual and computer-assisted prostate segmentation on magnetic resonance imaging," (in preparation).

www.manaraa.com

dimensional (3D) manual prostate contour delineation is laborious and time-consuming, and subject to substantial inter-operator variability [14].

Several algorithms have been presented in the literature for 3D segmentation of the prostate on T2w MRI, as described in a recent survey [15]. However, a minority of these methods have been validated for use on T2w MRI acquired using an ER coil (henceforth referred to as ER MRI). Although ER MRI can improve PCa detection, its improved contrast results in the presence of additional high-frequency details in the images. This makes automatic segmentation more challenging, especially for algorithms designed for use on non-ER MRI where the intraprostatic signal is more homogeneous. Furthermore, the ER coil deforms and displaces the prostate gland and produces MRI artifacts [16] that further challenge automatic segmentation. We have previously reported on semi-automatic [17] and automatic (Chapter 3) segmentation algorithms developed in our laboratory. Full details are available in the original publications; we describe details relevant to this study here. These methods are based on prostate shape and appearance models learned from a training set. Segmentation is performed in two steps: coarse localization of the prostate, followed by 3D segmentation boundary detection and refinement. In the semi-automated approach, coarse localization is performed by the operator with four mouse clicks requiring approximately 30 seconds of user interaction time. In the automated approach, coarse localization is performed automatically within 3 seconds of computation time, with no requirement for user interaction.

A range of segmentation accuracy values has been reported in the literature for automated and semi-automated algorithms (Table 4.1). Typically, reported error metrics include the mean absolute distance (MAD) between the boundaries of the automatic and

المنارة للاستشارات
www.manaraa.com

manual segmentations, and/or the Dice similarity coefficient (DSC). Reported MAD values range from 1.5–3.4 mm [17-20], and reported DSC values range from 82%–91% [17-21]. Reasons for the range of different error values reported include algorithm design, the use of single-operator manual reference segmentations for validation in most studies, and the use of different imaging data sets. These differences notwithstanding, the errors yielded by state-of-the-art segmentation methods are asymptotically approaching the differences observed between human expert operators [14]. It is thus timely to shift the focus of research in this area to studies aimed at enabling clinical translation of these techniques so that they can be of benefit to those suffering from cancer.

For reasons of diagnostic accuracy and patient safety, the integration of any computer-assisted segmentation algorithm, fully automatic or otherwise, into clinical use will require that an expert reviews and edits each segmentation as necessary before proceeding. This will always be necessary since regardless of the reported accuracy of a given segmentation algorithm, unusual cases will occur in the clinic that result in poor-quality computer-assisted segmentations, with potentially disastrous consequences to the patient if such segmentations were used to guide treatment. Therefore, the clinical utility of a method will depend not only on its accuracy metric values, such as the MAD and DSC, determined on a testing data set, but also on the amount of editing deemed necessary by expert physicians in order to render the segmentation suitable for clinical use. This editing can be measured spatially using standard metrics such as MAD and DSC, to compare the segmentation as output by the algorithm to the segmentation after editing, and these metrics can be computed on anatomically distinct regions to learn about the portions of the prostate requiring the most editing. Potentially of even greater

المنارة للاستشارات
www.manaraa.com

importance, the amount of required editing time can be measured. For a segmentation algorithm to have clinical utility, it must allow the expert physician to obtain a segmentation deemed clinically acceptable by him/her in less time than would be required to perform a manual segmentation. This statement holds true regardless of the reported segmentation accuracy metrics (e.g. MAD, DSC) for an algorithm in the literature. To the best of our knowledge, questions of editing magnitude and time have not been extensively studied for ER MRI prostate segmentation algorithms reported in the literature.

In this chapter, we conducted a user study to answer four research questions. (1) How much spatial segmentation editing do expert operators perform to obtain clinically useful segmentations? (2) What is the inter-operator variability in segmentation? (3) How much segmentation editing time do expert operators require to obtain clinically useful segmentations? (4) Can the necessary time requirement for segmentation editing be predicted from spatial segmentation error metrics? Questions (1), (2), and (3) were answered and compared under three conditions, where the segmentations provided to the operators for editing came from (a) our automatic segmentation algorithm, (b) our semi-automatic segmentation algorithm, and (c) manual segmentation performed by another expert operator. As the scope of question (4) is limited to evaluation of computer-assisted segmentation algorithms, it was answered under conditions (a) and (b) only.

www.manaraa.com

Table 4.1: Reported segmentation errors for prostate segmentation algorithms intended for use on T2w ER MRI.

Algorithm	Technique	Data set size	Accuracy	Segmentation time
Our semi-automatic algorithm [17] (described in Chapter 2)	Local appearance and shape model (semi-automatic)	**42** (test and training)	Whole gland: MAD: **2.0** ± **0.5 mm** DSC: **82%** ± **4%** Recall: **77%** ± **9%** Precision: **88%** ± **6%** ΔV: **-4.6** ± **7.2 cm^3**	Operator interaction: **28** ± **14 sec.** (across 10 images and 9 operators) Execution: **85** ± **20 sec.** (across 42 images, one operator)
Our automatic algorithm (described in Chapter 3)	Local appearance and shape model (automatic)	**42** (test and training)	Whole gland: MAD: **3.2** ± **1.2 mm** DSC: **71%** ± **11%** Recall: **69%** ± **15%** Precision: **76%** ± **12%** ΔV: **-3.6** ± **10.4 cm^3**	Execution: **54** ± **13 sec.** (across 42 images)
Cheng *et al.* [21]	Atlas-based (automatic)	**100** (training) and **40** (test)	Whole gland: TP: **91.2%** DSC: **87.6%** ΔV: **8.4%**	NA
Liao *et al.* [18]	Multi-atlas-based (automatic)	**66** (test) **9** (atlas)	Whole gland: MAD: **1.8** ± **0.9 mm** DSC: **88%** ± **3%**	Execution: **2.9 min.**
Toth and Madabhushi [19]	Active appearance model (semi-automatic)	**108**	Whole gland: MAD: **1.5** ± **0.8 mm** DSC: **88%** ± **5%**	Execution: **150 sec.**
Vikal *et al.* [22]	Shape model (semi-automatic)	**3**	Has not reported for whole gland	Execution: **23 sec.**
Martin *et al.* [20]	Atlas-based (semi-automatic)	**1** (reference) **17** (test)	Whole gland: MAD: **3.4** ± **2.0 mm** Recall: **89%** ± **6%** Precision: **78%** ± **12%**	NA

MAD: mean absolute distance, DSC: Dice similarity coefficient, ΔV: Volume difference, TP: true positive

4.2 Materials and Methods

4.2.1 Materials

Our sample consisted of 10 axial T2w fast spin echo ER MRI acquired at 3.0 Tesla field strength, all from patients with biopsy-confirmed PCa. Images were acquired with TR = 4000–13000 ms, TE 156–164 ms, NEX = 2. The voxel sizes were 0.27 × 0.27 × 2.2 mm as is typically seen in clinical prostate MRI. The images were acquired using a Discovery MR750 (General Electric Healthcare, Waukesha, WI). The study was approved by the research ethics board of our institution, and written informed consent was obtained from all patients prior to enrolment. All 10 MR images were segmented manually by three operators: one radiologist, one radiation oncologist and an expert

www.manaraa.com

radiology resident with >3 years' experience reading >100 prostate MRI studies in tandem with a board-certified radiologist as part of a trial conducted at our centre. Editing was conducted by four radiation oncologists with genitourinary specialization and the same expert radiology resident. The ITK-SNAP software tool [23] was used for manual segmentation.

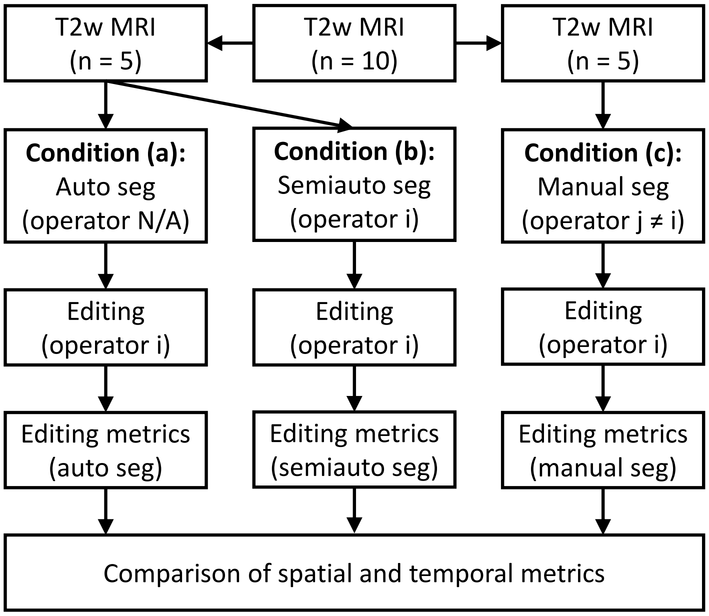

T2w MRI (n = 5)
T2w MRI (n = 10)
T2w MRI (n = 5)
Condition (a): Auto seg (operator N/A)
Condition (b): Semiauto seg (operator i)
Condition (c): Manual seg (operator j ≠ i)
Editing (operator i)
Editing (operator i)
Editing (operator i)
Editing metrics (auto seg)
Editing metrics (semiauto seg)
Editing metrics (manual seg)
Comparison of spatial and temporal metrics

Figure 4.1: Study design showing the workflow for a particular operator #i. The operator edited three sets of segmentations: five automatic segmentations, five semi-automatic segmentations performed by the operator, and five semi-automatic segmentations performed by a different operator #j. Spatial and temporal segmentation metrics were collected to measure the editing task and compared across the three conditions.

4.2.2 Study design

Our study design is shown in Figure 4.1. Each operator #i edited a total of 15 segmentations under three conditions: (a) five automatic segmentations, (b) five semi-automatic segmentations performed based on the operator's own inputs as the semi-automatic segmentation algorithm operator, and (c) five manual segmentations performed by a different expert operator #j. Operator #j was the same individual throughout the

المنارة للاستشارات
www.manaraa.com

entire experiment; operator #j only provided manual reference segmentations and did not take part in this editing study in any other way. Editing was performed in slice-by slice mode using the ITK-SNAP interface on axially-oriented slices. Changes were applied only on the axial slices but sagittal and coronal views were also provided to the operator during editing, so the operator could check for spatial coherence of the segmentations in these views. The operators used the adjustable-size paint brush tool in ITK-SNAP to add/remove area to/from the segmentation labels. They were able to adjust window and level and zoom in and out during editing. Spatial and temporal metrics were collected for each of the three conditions to compare the editing that was performed within each operator and between operators. To enable direct comparison of the editing of the automatic and semi-automatic segmentations, we used the same subset of 5 MRI scans for each operator for these two conditions. To mitigate possible effects of the order of MRI scan presentation on the experiment, the 15 segmentations were presented in a different randomized order for each operator, with a constraint that between any two presentations of the same MRI scan to the operator (i.e. once for automatic segmentation, and again with the same scan for semi-automatic segmentation), there were at least six MRI scans from other patients presented.

4.2.3 Spatial editing magnitude and inter-operator variability

We compared the pre-editing segmentations to the post-editing segmentations in each of the three conditions shown in Figure 4.1, *answering research question (1)*. We used five different metrics, including MAD, DSC, recall, precision and volume difference (ΔV), to perform comparisons in terms of surface disagreement, regional misalignment

المنارة للاستشارات
www.manaraa.com

and volume difference. Where applicable, the post-editing segmentation was defined as the reference segmentation. These metrics are defined in detail below.

4.2.3.1 Mean absolute distance

The MAD metric measures the disagreement between two 3D surfaces as the average of a set of Euclidean distances between corresponding surface points of two shapes. For each point on one surface, the closest point on the other surface is defined as the corresponding point. Equation (4.1) shows the MAD of *X* and *Y* as two surface point sets, where *D(p,q)* is the Euclidean distance between points *p* and *q*. A MAD of zero indicates ideal agreement between two shapes.

$$MAD(X,Y) = \frac{1}{N}\sum_{p\in X}\min_{q\in Y} D(p,q) \tag{4.1}$$

The MAD calculation needs to consider one of the shapes as the reference (*e.g.* point set *Y* is the reference in equation (4.1). Therefore, when two segmentations are to be compared and there is no reference segmentation, we use the bilateral MAD which is the average of the two MAD values obtained using each segmentation as the reference.

4.2.3.2 Dice similarity coefficient

The DSC is a region-based metric that measures the proportion of the volume of the overlap region between two shapes and the average of their volumes in 3D (equation (4.2)). The DSC is a unitless metric and will be 100% in the case of ideal segmentation and 0 when there is no overlap.

www.manaraa.com

4.2.3.3 Recall and precision rates

Recall (or sensitivity) and precision are also unitless error metrics that measure the regional misalignment in terms of the overlap region with 100% and 0 as the ideal and worst-case measurement values, respectively. To calculate recall and precision, we need to consider one shape as the reference. Recall measures the proportion of the reference that is within the segmentation (equation (4.3)) and precision measures the proportion of the segmentation that is within the reference (equation (4.4)).

$$DSC(X,Y) = \frac{2(X \cap Y)}{X + Y} = \frac{2TP}{FP + 2TP + FN} \times 100\,, \tag{4.2}$$

$$Recall(X,Y) = \frac{TP}{TP + FN} \times 100\,, \tag{4.3}$$

$$Precision(X,Y) = \frac{TP}{TP + FP} \times 100\,, \tag{4.4}$$

where TP is the true positive or correctly identified region, FP is the false positive or incorrectly identified region, and FN is the false negative or incorrectly ignored region (see Figure 3.2).

4.2.3.4 Volume difference

To calculate ΔV we subtract the reference shape volume from the segmentation shape volume. Therefore ΔV is a signed error metric; *i.e.* negative values of ΔV show that the segmentation is smaller than the reference and positive values of ΔV show that the segmentation is larger than the reference.

To quantify inter-operator variability in segmentation and editing (*answering research question (2)*), we calculated simultaneous truth and performance level estimation (STAPLE) [24] consensus segmentations from the five operator segmentations before and after editing under all three conditions, with two exceptions. In the case of the

المنارة للاستشارات

www.manaraa.com

pre-editing automatic segmentations, no operators were involved, so no STAPLE segmentation was calculated. In the case of the pre-editing manual segmentations, only the segmentations of a single operator #j were edited in this study. To obtain a measure of inter-operator variability in pre-editing manual segmentations, we computed a STAPLE segmentation from manual segmentations performed by three of our operators on the same five images that were used for manual segmentation editing in our study. There were thus five sets, each containing five segmentations performed by different operators, with accompanying STAPLE consensus segmentation: (1) pre-editing semi-automatic, (2) post-editing semi-automatic, (3) post-editing automatic, (4) pre-editing manual, and (5) post-editing manual. Within each of these five sets, our five segmentation error metrics were computed to compare each operator's segmentation to the corresponding STAPLE segmentation, with the means of the metric values indicating the amount of inter-operator variability. We used one-tailed pairwise heteroscedastic *t*-tests to test for statistical significance of differences in these inter-operator variability measurements between paired elements of the five sets. This allows us, for instance, to measure whether there is a statistically significant reduction of inter-operator variability in edited semi-automatic segmentations, versus edited automatic segmentations.

4.2.4 Required editing time and correlation with spatial error metrics

For each label, we recorded the interaction time that was required to have a clinically acceptable segmentation using manual, semi-automatic and automatic segmentation methods, *answering research question (3)*. We recorded the time from the moment when the operator began reviewing and editing the segmentation until the moment the operator verbally confirmed that the segmentation was ready to be used in

المنارة للاستشارات
www.manaraa.com

clinic. The editing time included browsing through the slices in the 3D volume, reviewing the segmentation, adding to and removing from the segmentation, window and level adjustment, editing tool selection and adjustment, and zooming in and out. For each of the three conditions, we calculated the mean and standard deviation of the interaction time across the five presented MRI scans separately for each operator, and also in aggregate across all five operators. For the semi-automatic algorithm we measured the interaction time required for algorithm operation and included this interaction time as part of the time required for the condition involving semi-automatic segmentation.

We measured the degree to which measured spatial error metric values can be used as surrogates for the amount of editing time needed to achieve a segmentation that is satisfactory to the operator*, answering research question (4)*. To do this, we calculated all five of our error metrics for the whole gland, apex, mid-gland, and base, comparing the pre-editing segmentation to the post-editing segmentation for the automatic and semi-automatic segmentations (conditions 1 and 2 in Figure 4.1), using the post-editing segmentation as the reference where applicable. We measured the monotonicity of the relationship between each metric value and editing time using Spearman's rank-order correlation (ρ). We tested the statistical significance of the correlation coefficients using the null hypothesis that there was no association between the error metric values and editing time values. For all tests, the sample size was 50 (10 images each contoured by 5 operators).

المنارة للاستشارات
www.manaraa.com

4.3 Results

4.3.1 Spatial editing magnitude and inter-operator variability

Figure 4.4 shows the spatial magnitude of editing required for automatic, semi-automatic, and manual segmentations for operators to achieve final edited segmentations suitable for clinical use. As might be expected, the general trend is that the automatic segmentations required the most editing, followed by the semi-automatic and manual segmentations. However, this trend was not reflected in all of the error metrics. For instance, looking at the DSC and recall metrics, we detected no significant difference in the amount of editing applied to the automatic vs. semi-automatic segmentations. Operator editing of manual segmentation consistently decreased segmentation volume without substantially affecting precision. This suggests that the manual pre-editing segmentations were deemed by the operators to be oversegmentations, and editing drew the boundaries inward by an amount reflected by the MAD metric values in Figure 4.4 (MAD < 1 mm in general). Figure 4.3 shows the inter-operator variability in segmentation before and after editing, reported using the mean of each segmentation error metric across all operators for each image, with respect to a STAPLE reference standard. This analysis revealed significant differences in inter-operator variability for most of the conditions, for all metrics expect for the volume difference. Note the substantial inter-operator variability in manual segmentation (reflected by large mean metric values and large variability indicated by the whiskers) for many metrics, relative to the inter-operator variability in semi-automatic and automatic segmentations, even when manual editing is applied (e.g. compare the “manual-pre” measurements to the other measurements for the MAD metric in Figure 4.3). Overall, post-editing variability

المنارة للاستشارات
www.manaraa.com

is lower than pre-editing variability, with post-editing automatic and semi-automatic segmentations having similar variability. The MAD, DSC, and precision metrics revealed that editing reduced the amount of inter-operator variability for the semi-automatic segmentation condition (compare SA (pre) to SA (post) in Figure 4.3 for these three metrics). Interestingly, a similar pattern was observed for the manual segmentations. No significant differences were found between pre-editing manual segmentations and computer-assisted segmentations for any of the conditions or metrics. Post-editing automatic segmentation consistently demonstrated lower variability than pre-editing semi-automatic segmentation. No significant differences were found between post-editing automatic segmentation and post-editing semi-automatic segmentation.

Table 4.2: User manual interaction time for ready to use prostate segmentation in T2w MRI.

Segmentation labels	**No. of images**	**No. of Operators**	**User interaction time**
Manual	5	5	213 ± 90 sec (3:33 ± 1:30 min)
Semi-automatic	5	5	328 ± 126 sec (5:28 ± 2:06 min)
Semi-automatic (user interaction time included)	5	5	351 ± 128 sec (5:51 ± 2:08 min)
Automatic	5	5	393 ± 146 sec (6:33 ± 2:26 min)

www.manaraa.com

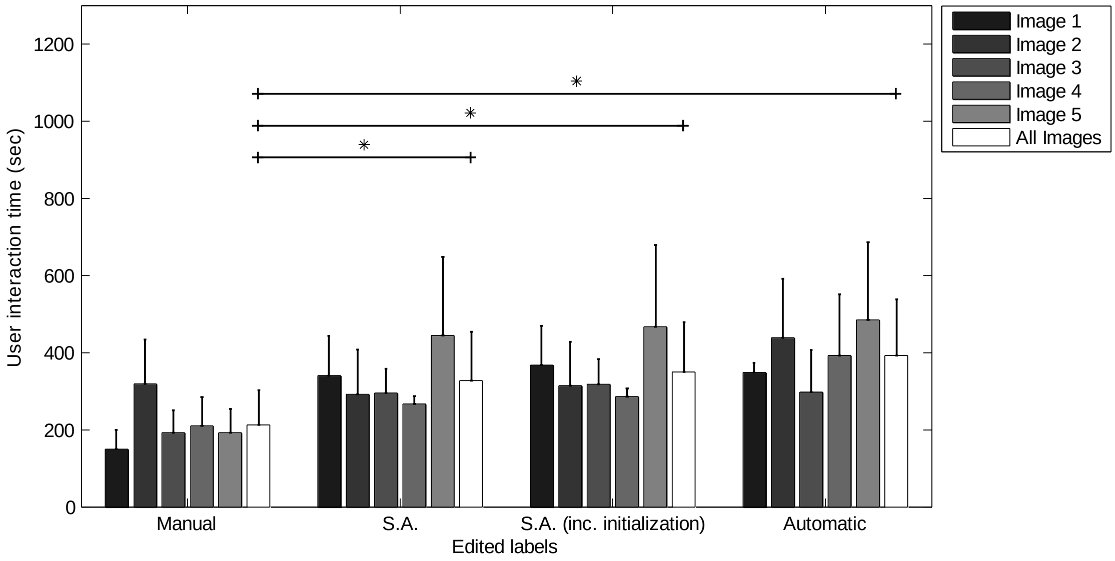


Figure 4.2: User manual interaction time on manual, semi-automatic (S.A.) and automatic segmentations.The statistically significant differences indicated with * on the averages of the groups across all the five images ($p < 0.05$).

4.3.2 Required editing time and correlation with spatial error metrics

Table 4.2 shows the mean ± standard deviation of the recorded time required for each of the three conditions. For the semi-automatic condition, the time required only for editing, as well as the time required for editing plus the time required to interact with the semi-automated algorithm, are reported separately. Figure 4.2 shows the breakdown of these editing times for each image. Significant differences were found between editing times for all conditions, except when comparing automatic segmentation to semi-automatic segmentation. To provide context for these editing times, according to the literature, the time required for manual prostate delineation on MRI can range from approximately 5 minutes [25] to approximately 20 minutes per patient [26], or about 1.6 minutes for each 2D slice [27]. Our experience is concordant with this reported time range; timing of manual segmentation on the five images used in conditions (a) and (b)

المنارة للاستشارات
www.manaraa.com

for one expert operator yielded a mean ± standard deviation segmentation time of 564 ± 162 sec (9:20 ± 2:42 min). Based on Table 4.2, we observe that operators spent approximately 2–3 additional minutes editing computer-assisted segmentations, compared to the amount of time spent editing manual segmentations performed by a different expert operator.

4.3.3 Correlation of editing time with the metric values

Table 4.3 shows the correlations between editing time and spatial editing magnitudes as measured using our segmentation error metrics. There were few significant correlations and none had magnitude > 0.5. Significant correlations were predominantly in the base of the gland. In the base, recall was positively correlated with editing time, and precision and volume difference were negatively correlated with editing time. This pattern was observed in the whole gland as well but only weakly in the mid-gland and not in the apex.

Table 4.3: Correlation between editing time and spatial editing magnitude measured using five metrics. Each value is the Spearman's correlation coefficient between the value of each error metric and editing time. The bold numbers indicate statistically significant correlations ($p < 0.05$)

Anatomic region	MAD	DSC	Recall	Precision	ΔV
Whole gland	0.204	0.18	**0.361**	**-0.341**	**0.417**
Apex	0.206	-0.081	-0.194	-0.138	0.092
Mid-gland	**0.263**	-0.149	0.149	**-0.282**	**0.312**
Base	-0.14	**0.367**	**0.428**	**-0.305**	**0.406**

www.manaraa.com

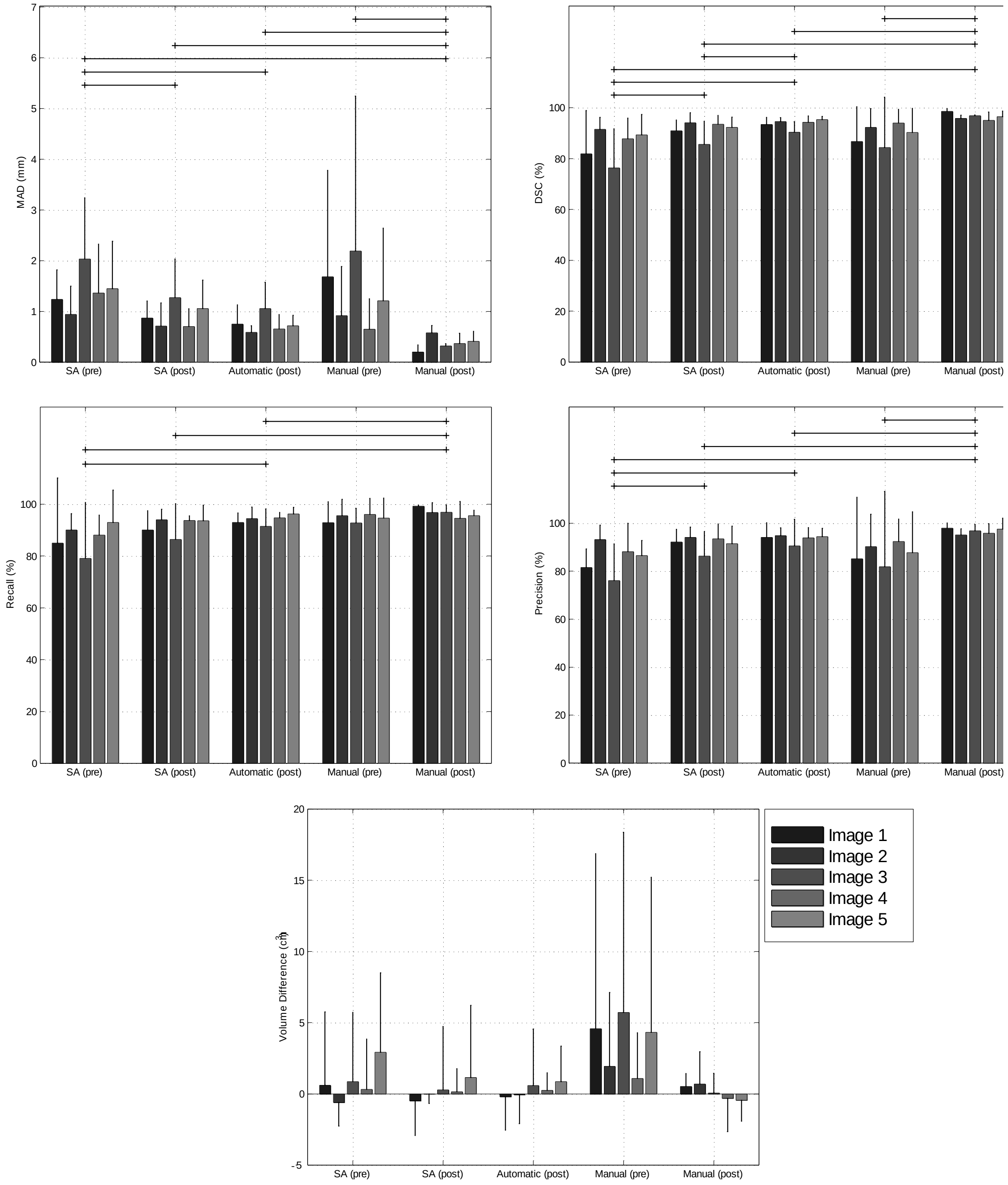


Figure 4.3: Inter-operator variability. Each bar shows the average metric value for one image across five operators. The error bars indicate one standard deviation. The horizontal lines indicated statistically significant differences on the averages of the groups across all the five operators and five images ($p < 0.05$). SA: semi-automatic.

www.manaraa.com

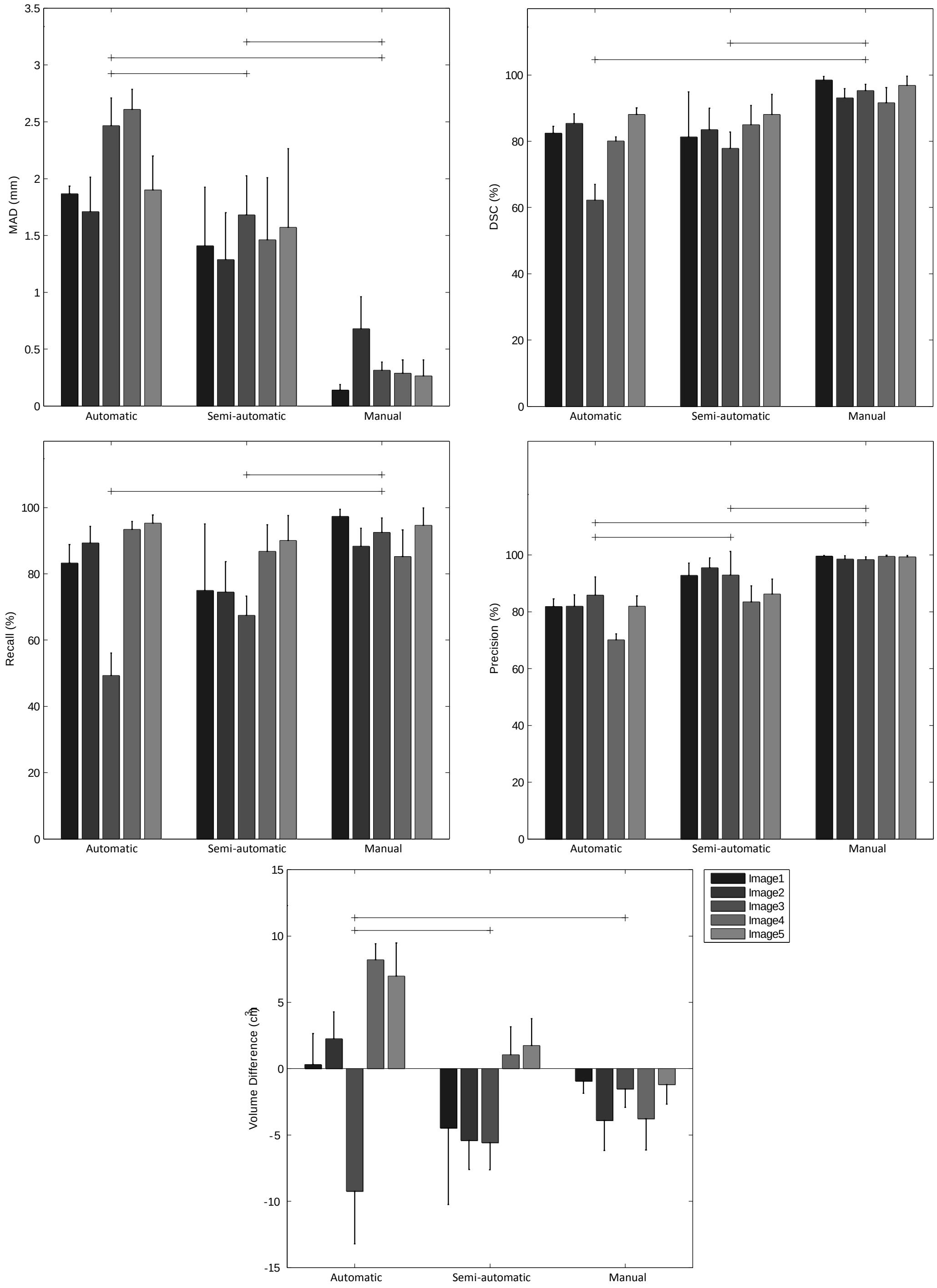


Figure 4.4: Editing magnitude, showing the differences between the segmentations pre- and post-editing for each of the three conditions.Each bar shows the average metric value for one image across five operators. The error bars indicate one standard deviation. The horizontal lines indicated statistically significant differences on the averages of the groups across all the five operators and five images ($p < 0.05$).

www.manaraa.com

4.4 Discussion

4.4.1 Spatial editing magnitude and inter-operator variability

As shown in Figure 4.4, there was a nonzero difference between pre-editing and post-editing expert manual segmentations for all metrics. The amount of editing performed on the manual segmentations provides valuable perspective on the amount of editing performed on the automatic and semi-automatic segmentations. One might expect that improvements to computer-assisted segmentation algorithms would require amounts of editing asymptotically approaching the amounts of editing that operators deem necessary for expert manual segmentations provided by other experts (i.e. expert operators would elect to edit outputs from even an ideal computer-assisted segmentation algorithm). For studies of computer-assisted segmentation algorithms using single-operator manual reference standard segmentations for validation, this observation is especially important; this suggests that algorithms yielding segmentation error metric values within the range observed in expert editing of manual expert segmentations could be considered to have essentially the same performance. For instance, Figure 4.4 would suggest that two algorithms reporting DSC values of 94% and 96% would be considered to perform equally, as these values are well within the range of manual editing of manual segmentations. This observation could have ramifications for the ranking schemes used for segmentation grand challenges (such as PROMISE12 [28]), suggesting a practical equivalence of some top-ranked algorithms and a potential means for deciding when top-ranked algorithms are ready to be moved to the next stage of translation to clinical use. Although some metrics revealed a significant difference in the amount of editing required for automatic vs semi-automatic segmentations, this significance (and the magnitude of

www.manaraa.com

the difference) varied across metrics. This observation emphasizes the need for multiple, complementary spatial metrics to comprehensively assess the performance of a segmentation algorithm.

Our analysis in Figure 4.3 indicates that in general, allowing operators to edit provided segmentations reduces inter-operator variability in segmentation, compared to the inter-operator variability resulting from manual segmentations performed from scratch. The trend held even when comparing manual segmentations performed from scratch to manual segmentations that have been edited to satisfaction by another operator. This result underscores the value of providing operators with a starting segmentation for editing as this could improve the reproducibility of prostate segmentation, which is important for multi-operator clinical trials and consistency of patient care in clinical practice. Whereas the lowest inter-operator variability resulted from giving operators a starting segmentation performed manually by another expert, in clinical practice this is clearly impractical. From this perspective, the automatic segmentation could be seen as a practical alternative approach to obtain the starting segmentation. Although the difference in inter-operator variability between post-editing manual segmentations and post-editing automatic segmentations was statistically significant, inspection of Figure 4.3 reveals that this difference is very small from a practical perspective. This leads to the hypothesis that providing operators with an automatic segmentation with accuracy metric values similar to ours (Table 4.1) as a starting point will yield superior inter-operator reproducibility even after editing, compared to manual segmentations performed from scratch. This hypothesis needs to be tested in a larger study covering a broader range of segmentation

المنارة للاستشارات
www.manaraa.com

algorithms, a larger data set, and a larger pool of operators having different experience levels.

4.4.2 Required editing time and correlation with spatial error metrics

Our results suggest that the use of automatic or semi-automatic segmentation algorithms to provide a starting segmentation for editing should reduce the total amount of time required to achieve a clinically acceptable segmentation, relative to typical reported times required for manual segmentations performed from scratch. Our results also suggest that the difference in total time required to use our automatic vs semi-automatic segmentation algorithms for this purpose is small, when the time required to interact with the semi-automatic segmentation algorithm is taken into account. Thus, the choice in this regard may come down to operator preference; the semi-automatic segmentation algorithm allows the operator to specify the apex-to-base extent of the prostate, reducing the need for editing involving adding or removing entire slices in these regions. This comes at the cost of needing to wait for < 60 seconds for the segmentation to be computed online, whereas the automatic segmentations can be computed offline immediately after MRI scanning and thus would appear instantaneously to the operator at time of editing. Our results also showed that operators spent more time editing the computer-assisted segmentations, compared to the time spent editing manual segmentations by another expert operator. We posit that this difference in editing time is an important metric for determining the suitability of a computer-assisted segmentation algorithm for translation to clinical use in scenarios where for safety or other reasons, expert operator verification for necessary editing will be performed on every segmentation. From this perspective, there is room for improvement in our semi-

المنارة للاستشارات
www.manaraa.com

automatic and automatic algorithms of approximately 2–3 minutes of editing time per prostate in order to achieve concordance with the amount of editing performed on manual segmentations.

Table 4.3 indicates a consistent negative correlation of the precision metric value with editing time, with statistically significant correlations in all anatomic regions except for the apex. This implies that the greater the false positive area in a computer-assisted segmentation, the greater the time that will be required to edit the segmentation to a clinically acceptable level. This is corroborated by the consistent positive correlation with the volume difference metric (again, significant everywhere except the apex), implying that the greater the amount of oversegmentation performed by computer-assisted segmentation algorithm, the more editing time that will be required. Comparing the correlation coefficients for precision and volume difference within the apex, mid-gland, and base, the strongest correlations were found in the base region. This implies that the above relationships are especially applicable for false positive regions and oversegmentation of the base. However, based on these observations one could make only a weak recommendation that the amount of necessary editing time could be estimated based on the precision and volume difference spatial error metric values; although the correlation coefficients are statistically significant in many cases, they do not have high magnitude.

The lack of strong correlations in Table 4.3 implies weak relationships between editing time and spatial editing magnitudes as measured by our segmentation error metrics. The observations in the previous paragraph notwithstanding, this implies that in general, one cannot use spatial metrics such as the MAD, DSC, precision, recall, and

المنارة للاستشارات
www.manaraa.com

volume difference to estimate the amount of time that an operator will require to produce a clinically acceptable segmentations using the output of a segmentation algorithm as a starting point. This is an important observation since in most clinical workflows, time is a scarce and valuable resource; if it takes (nearly as) long to edit a segmentation from an algorithm as it does to perform a manual segmentation from scratch, the clinician may be inclined toward the simpler approach of performing manual segmentation. We surmise that this issue is a major contributor to the present state of affairs, where the academic literature has produced many hundreds of computer-assisted segmentation algorithms and yet very few of them have moved forward to clinical use. This leads to the conjecture that the most important metrics to compute when evaluating the suitability of an algorithm for clinical translation are operator variability, measured using spatial metrics such as MAD, DSC, etc., and editing time, measured directly using a sample of multiple operators. Viewed through this lens, the ideal segmentation algorithm would yield low operator variability and low editing time. This suggests that a potential reevaluation of the use spatial metrics for measuring segmentation *accuracy* may be in order, since in most practical clinical workflows, the final segmentation as edited and approved by the clinician will be used for its clinical purpose and could be considered 100% “accurate” for practical purposes. This observation supports engineers and computer scientists aiming for the concrete goal of *producing a clinically useful segmentation in a minimum amount of* time, in lieu of setting our aims according to the nebulous notion of accuracy, with all of its attendant issues (e.g. differing expert opinions on what constitutes a correct segmentation, issues regarding whether “gold standard” expert segmentations truly delineate the histologic boundary of the target of interest).

المنارة للاستشارات
www.manaraa.com

Comparing our algorithms to a hypothetical segmentation algorithm that demonstrated better performance based all five introduced error metrics in a multi-operator and multi-reference study, and noting that no such algorithm has been reported in the literature due to a lack of comprehensive validation:

(1) We expect that less editing of the segmentation results would be required, but not less than the amount of editing applied by an expert on to a manual segmentation provided by another expert.

(2) Since there was not a big difference in terms of editing time measured for automatic and semi-automatic segmentation algorithms, and no strong correlation between segmentation error metric values and the required editing time, we cannot speculate as to whether this hypothetical algorithm would result in reduced editing time.

(4) We acknowledge that our observed lack of correlation between spatial error metric values and editing time only applies to the range of spatial error metric values that we observed for our algorithms. Such a correlation is possible for different error metric value ranges; e.g. containing the better metric values given by this hypothetical algorithm.

4.4.3 Limitations

This work must be considered in the context of its strengths and limitations. We acknowledge that given our image sample size and number of operators participating in the study, in some aspects, this is a descriptive, hypothesis-generating study that points the way to potentially fruitful studies on larger sample sizes with sufficient statistical power to draw firmer conclusions. We also acknowledge that although the editing interface we used, involving a mouse-driven variable-sized paintbrush tool, is concordant

للاستشارات المنارة
www.manaraa.com

it its mode of operation with the interfaces used in many clinical workflows, it does constitute only a single mode of performing segmentation editing. Thus, our study generates no knowledge about the impact of the choice of editing tool on editing times, and this would be a subject of valuable further study. Finally, in this user study we tested only two computer-assisted segmentation algorithms; a more comprehensive future study involving a broader cross-section of current algorithms is warranted.

4.4.4 Conclusions

In this chapter, we conducted a user study measuring the amount of spatial editing performed by expert users on segmentations generated manually, semi-automatically, and automatically. We measured the inter-operator variability in segmentation before and after editing, and measured the relationship between editing magnitude and time spent editing. With reference to the enumerated research questions in the introduction section of this chapter, we have reached four main conclusions, with the acknowledgment that our sample size implies that these conclusions should be considered as hypotheses to test in future, larger studies. (1) As would be expected, the operators performed the most spatial segmentation editing on the automatic segmentations, followed by the semi-automatic segmentations, and the least amount of editing on the manual segmentations. The measured editing magnitudes varied according to the error metric used, reinforcing the value of using multiple, complementary error metrics in segmentation studies, rather than focusing on one or two typically used metrics (e.g. the MAD and DSC). (2) Providing operators with a starting segmentation for editing, either performed manually by another operator or (semi-)automatically via an algorithm, yielded lower inter-operator variability in the final segmentation, compared to inter-operator variability in

المنارة للاستشارات

www.manaraa.com

manual segmentations performed from scratch (as is frequently performed in clinical workflows currently). Inter-operator variability resulting from using our automatic algorithm to generate starting segmentations was not substantially higher than that resulting from using expert manual segmentations as starting segmentations, suggesting a role for our automated segmentation algorithm in this context. (3) The use of our automatic or semi-automatic segmentation algorithms to generate starting segmentations for editing is expected to decrease the total required segmentation time, compared to the time required to perform manual segmentations from scratch, and the choice of automatic vs. semi-automatic segmentation for this purpose comes down to operator preference. (4) The necessary time requirement for segmentation editing cannot be reliably predicted from spatial segmentation error metrics in all anatomic regions of the prostate. Thus, for the many clinical workflows where manual segmentation review and editing will be performed for safety and other reasons, and minimization of editing time is a primary goal, the fact that one algorithm outperforms another in terms of spatial metrics such as the MAD and DSC does not imply that the algorithm is more suitable for clinical translation. In such contexts, where the medical expert's final edited segmentation is taken as correct for practical purposes, the ideal segmentation algorithm supports the expert's obtaining of a clinically acceptable segmentation in a minimum amount of time while minimizing inter-operator segmentation variability. This increases the volume of patients that can be treated and simultaneously supports consistent quality of the intervention patients receive.

المنارة للاستشارات

www.manaraa.com

4.5 References

1. R. Siegel, J. Ma, Z. Zou and A. Jemal, "Cancer statistics, 2014," CA: a cancer journal for clinicians **64**, 9-29 (2014).

2. Canadian Cancer Society's Advisory Committee on Cancer Statistics. *Canadian cancer statistics 2014*. Canadian Cancer Society; 2014

3. M. L. Schiebler, M. D. Schnall, H. M. Pollack, R. E. Lenkinski, J. E. Tomaszewski, A. J. Wein, R. Whittington, W. Rauschning and H. Y. Kressel, "Current role of MR imaging in the staging of adenocarcinoma of the prostate," Radiology **189**, 339-352 (1993).

4. J. Kurhanewicz, D. Vigneron, P. Carroll and F. Coakley, "Multiparametric magnetic resonance imaging in prostate cancer: present and future," Curr Opin Urol **18**, 71-77 (2008).

5. H. U. Ahmed, A. Kirkham, M. Arya, R. Illing, A. Freeman, C. Allen and M. Emberton, "Is it time to consider a role for MRI before prostate biopsy?," Nat Rev Clin Oncol **6**, 197-206 (2009).

6. A. Shukla-Dave and H. Hricak, "Role of MRI in prostate cancer detection," NMR Biomed **27**, 16-24 (2014).

7. M. D. Schnall, Y. Imai, J. Tomaszewski, H. M. Pollack, R. E. Lenkinski and H. Y. Kressel, "Prostate cancer: local staging with endorectal surface coil MR imaging," Radiology **178**, 797-802 (1991).

8. G. M. Villeirs and G. O. De Meerleer, "Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning," Eur J Radiol **63**, 361-368 (2007).

9. J. Nakashima, A. Tanimoto, Y. Imai, M. Mukai, Y. Horiguchi, K. Nakagawa, M. Oya, T. Ohigashi, K. Marumo and M. Murai, "Endorectal MRI for prediction of tumor site, tumor size, and local extension of prostate cancer," Urology **64**, 101-105 (2004).

10. J. J. Futterer, M. R. Engelbrecht, G. J. Jager, R. P. Hartman, B. F. King, C. A. Hulsbergen-Van de Kaa, J. A. Witjes and J. O. Barentsz, "Prostate cancer: comparison of local staging accuracy of pelvic phased-array coil alone versus integrated endorectal-pelvic phased-array coils. Local staging accuracy of prostate cancer using endorectal coil MR imaging," Eur Radiol **17**, 1055-1065 (2007).

11. O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino and H. Hricak, "Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging," Radiology **239**, 784-792 (2006).

المنارة للاستشارات
www.manaraa.com

12. P. R. Carroll, F. V. Coakley and J. Kurhanewicz, "Magnetic resonance imaging and spectroscopy of prostate cancer," Rev Urol **8 Suppl 1**, S4-S10 (2006).

13. S. W. Heijmink, J. J. Futterer, S. S. Strum, W. J. Oyen, F. Frauscher, J. A. Witjes and J. O. Barentsz, "State-of-the-art uroradiologic imaging in the diagnosis of prostate cancer," Acta Oncol **50 Suppl 1**, 25-38 (2011).

14. W. L. Smith, C. Lewis, G. Bauman, G. Rodrigues, D. D'Souza, R. Ash, D. Ho, V. Venkatesan, D. Downey and A. Fenster, "Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR," Int J Radiat Oncol Biol Phys **67**, 1238-1247 (2007).

15. S. Ghose, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibe and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," Comput Methods Programs Biomed **108**, 262-287 (2012).

16. J. E. Husband, A. R. Padhani, A. D. MacVicar and P. Revell, "Magnetic resonance imaging of prostate cancer: comparison of image quality using endorectal and pelvic phased array coils," Clin Radiol **53**, 673-681 (1998).

17. M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, E. Gibson, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster and A. D. Ward, "Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semiautomated methods," Med Phys **41**, 113503 (2014).

18. S. Liao, Y. Gao, Y. Shi, A. Yousuf, I. Karademir, A. Oto and D. Shen, "Automatic Prostate MR Image Segmentation with Sparse Label Propagation and Domain-Specific Manifold Regularization," (Springer, 2013), pp. 511-523.

19. R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: application to prostate MRI segmentation," IEEE Trans Med Imaging **31**, 1638-1650 (2012).

20. S. Martin, V. Daanen and J. Troccaz, "Atlas-based prostate segmentation using an hybrid registration," Int J CARS **3**, 8 (2008).

21. R. Cheng, B. Turkbey, W. Gandler, H. K. Agarwal, V. P. Shah, A. Bokinsky, E. McCreedy, S. Wang, S. Sankineni, M. Bernardo, T. Pohida, P. Choyke and M. J. McAuliffe, "Atlas based AAM and SVM model for fully automatic MRI prostate segmentation," Conf Proc IEEE Eng Med Biol Soc **2014**, 2881-2885 (2014).

22. S. Vikal, S. Haker, C. Tempany and G. Fichtinger, "Prostate contouring in MRI guided biopsy," Proc SPIE **7259**, 72594A (2009).

المنارة للاستشارات

www.manaraa.com

23. P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," Neuroimage **31**, 1116-1128 (2006).

24. S. K. Warfield, K. H. Zou and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," IEEE Trans Med Imaging **23**, 903-921 (2004).

25. S. Martin, G. Rodrigues, N. Patil, G. Bauman, D. D'Souza, T. Sexton, D. Palma, A. V. Louie, F. Khalvati, H. R. Tizhoosh and S. Gaede, "A multiphase validation of atlas-based automatic and semiautomatic segmentation strategies for prostate MRI," Int J Radiat Oncol Biol Phys **85**, 95-100 (2013).

26. N. Makni, P. Puech, R. Lopes, A. S. Dewalle, O. Colot and N. Betrouni, "Combining a deformable model and a probabilistic framework for an automatic 3D segmentation of prostate on MRI," Int J Comput Assist Radiol Surg **4**, 181-188 (2009).

27. D. Flores-Tapia, G. Thomas, N. Venugopal, B. McCurdy and S. Pistorius, "Semi automatic MRI prostate segmentation based on wavelet multiscale products," Conf Proc IEEE Eng Med Biol Soc **2008**, 3020-3023 (2008).

28. G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," Med Image Anal **18**, 359-373 (2014).

المنارة للاستشارات
www.manaraa.com

Chapter 5.

Conclusions and directions for future work

5.1 Conclusions

The work in this thesis has resulted in the following advances in technology and knowledge, enumerated according to the objectives given in Section 1.9:

1. Chapter 2 described a novel system for semi-automatic prostate segmentation on T2w ER MRI. An important aspect of this system is that the main computational step of the segmentation is in computing loci of candidate boundary points, and our algorithm is such that the estimation of each locus is completely independent of all other loci, with regularization as a post-processing step. Therefore, the algorithm has high potential for further speedup via a parallel computing implementation. The accuracy and inter-observer variability of this system was measured using a set of complementary error metrics on multiple anatomic regions of interest within the prostate. We used a validation methodology in which three different types of error (surface disagreement, region misalignment and volume difference) in segmentation results were quantified using five error metrics. The error measurements were applied to the prostate gland as a whole, and to the apex, midgland and base regions separately. The system improved the reproducibility of the prostate segmentation, compared to manual segmentation, *supporting the central hypothesis of this thesis*. The system was shown to require minimal user interaction (30 seconds). Our results showed that the choice of manual reference segmentation had the biggest impact on segmentation variability, reinforcing

للاستشارات المنارة
www.manaraa.com

the need for multi-operator reference standard segmentations for algorithm validation. This study also showed that not only is prostate apex and base challenging (both semi-automatically and manually), but in fact there is high inter-observer variability in even defining the apex-most and base-most extents of the prostate. This sheds light on the most critical areas of focus for future development of prostate segmentation algorithms on MRI.

2. Chapter 3 described a novel system for automatic prostate segmentation on T2w ER MRI, with accuracy and inter-observer variability measured as in Chapter 2. This method takes advantage of the availability of prostate size measurements that are obtained during clinical standard TRUS imaging prior to MRI to facilitation an automatic coarse localisation of the prostate prior to segmentation refinement using the same algorithm as presented in Chapter 2. This method replaced the 30 seconds of manual operator interaction time required to use the semi-automatic method in Chapter 3 with 3 seconds of computation time, at the expense of a statistically significant but small decrease in accuracy. The use of the automated algorithm substantially mitigated the inter-observer variability observed in Chapter 2 of the segmentation of the base region of the prostate by eliminating the need for an operator to decide on the apex-most and base-most extents of the gland, *supporting the central hypothesis of this thesis*. The automatic algorithm provided improved accuracy, compared to the semi-automatic algorithm, in measuring overall prostate volume.

3. Chapter 4 described an expert user study measuring the impact of using semi-automatic and automatic segmentations on physicians' ability to obtain a clinically acceptable segmentation via editing a provided starting segmentation, compared to

المنارة للاستشارات
www.manaraa.com

achieving the same end via fully manual segmentation. Although the amount of editing required was directly proportional to the amount of automation used to produce the starting segmentation (i.e. via fully automated, vs semi-automated, vs manual segmentation by another operator), the use of an automatically generated starter segmentation yielded lower inter-operator segmentation variability in the final segmentation, compared to from-scratch manual segmentation, *supporting the central hypothesis of this thesis*. Using such a starter segmentation was also found to reduce the total time required to achieve a clinically acceptable segmentation, also *supporting the central hypothesis of this thesis*. Finally, we found that spatial error metrics such as the MAD and DSC are not strongly correlated with the amount of editing time required to render a segmentation clinically acceptable. This observation challenges the comparison between two algorithms' performance based on the values of the spatial metrics. Since for clinical purposes, a segmentation judged by a physician to be clinically acceptable can be taken to be accurate (because the physician will use this segmentation for the intervention at hand), the practical purposes of computer-assisted segmentation are (1) to assist the physician in obtaining a segmentation that is clinically acceptable to him/her in less time than would be required for manual segmentation, and (2) to increase consistency of patient care for procedures depending on prostate segmentation on MRI. Thus, to evaluate an algorithm's suitability for clinical translation, algorithm developers need to directly measure the required editing time for multiple operators to achieve clinically acceptable segmentations; spatial error metrics cannot be used as a surrogate for editing time. Rather, the value of spatial error metrics is in measuring inter-operator segmentation variability; decreasing this variability is a step toward increasing

المنارة للاستشارات
www.manaraa.com

consistency of patient care and increasing consistency of execution of multi-operator and multi-centre clinical trials involving prostate segmentation on MRI.

5.2 Applications and future directions

The segmentation techniques and evaluation methods developed in this thesis support research applications in which prostate segmentation in T2w ER MRI is either being studied or being employed, or clinical applications that require prostate segmentation on T2w ER MRI. In the following section, several applications that could potentially benefit the segmentation algorithm will be discussed. Some remaining gaps in knowledge that could be covered as part of future work will also be discussed.

5.2.1 Applications in ongoing clinical research studies

In an ongoing clinical research studies in our group [1], for a mechanically-assisted targeted prostate biopsy system, surface-based MRI-TRUS registration was required. The TRUS images were segmented using a semi-automatic algorithm. For the MR images, manual segmentation of the prostate was used. Our segmentation algorithms could be used to facilitate the surface-based image registration and MRI-TRUS fusion, to decrease processing time and mitigate inter-observer variabiility. The impact of MRI segmentation error and variability on the MRI-TRUS registration error could be studied before and after applying manual editing to the segmentation labels, with the ultimate endpoint being the impact on positive yield at biopsy as measured in a prospective study.

In another clinical research study [2, 3], manual segmentation of the prostate on T2w MRI has been employed in an MRI-compatible mechatronic system that was developed for MRI-guided needle insertion to the prostate. In this system, a preoperative

المنارة للاستشارات
www.manaraa.com

and an intraoperative ER prostate MR image are manually segmented and registered together for mapping defined targets from the preoperative image to the intraoperative image. Our segmentation algorithms could be applied to both pre- and intra-operative images. Our automated algorithm could be particularly helpful to speeding up the intraoperative MR image segmentation to reduce the total amount of in-bore procedure time required.

There are also other research studies or clinical applications in which ER prostate MRI segmentation is required [4, 5]. In these studies manual prostate segmentation on pre-operative T2w MRI was used for surface-based image registration between pre-operative MRI and either intra-operative MRI or pre-operative ultrasound to localize PCa tumours in image-guided biopsy or focal therapy. Therefore, the impact of using our algorithm could be investigated in terms of processing time and/or procedure accuracy and reproducibility, compared to using manual segmentation.

5.2.2 Suggestions for future work

We have studied the accuracy and reproducibility of a segmentation algorithm as well as the segmentation time that is required to have clinically acceptable contours. To the best of our knowledge, the impact of segmentation error on the final results of a clinical procedure has not been extensively studied; e.g. the impact of MRI segmentation error on PCa targeting in an MRI-targeted TRUS-guided prostate biopsy, or on the results of radiation therapy dosimetry. Hence, it is also important to study the impact of the segmentation error and variability, before and after manual editing, on the performance of some of the clinical applications in which ER prostate MRI segmentation is used.

المنارة للاستشارات
www.manaraa.com

The presented segmentation approaches, although highly amenable to parallel implementation, were implemented sequentially. Hence, an important future step for this work would be implementation of the algorithm for parallel computing on a graphics processing unit (GPU), and we would expect at least an order of magnitude lower computation times from such an implementation. Using an unoptimized implementation of the algorithm on a Matlab research platform and running on a single central processing unit core, the segmentation takes less than 90 seconds, on average. We would expect a GPU implementation to therefore run in well under 10 seconds.

Our observation of no meaningful correlation between the values of the five error metrics used in this work and the required editing time to achieve a clinically acceptable segmentation leads to the recommendation that editing time must be measured directly for multiple observers in order to assess an algorithm's suitability for clinical translation. This renders algorithm evaluation much more expensive in terms of time and effort, and requires the engagement of clinical colleagues which is challenging in many computer science and engineering contexts which may be located distant to clinical centres. There is therefore potential value in future work designing and validating novel spatial error metrics that are more accurately predictive of required editing time. The data set generated as part of our study in Chapter 4 could provide initial validation of novel metrics for this purpose.

Since PCa tumours are most likely to be found within the peripheral zone of the prostate gland, in some the clinical applications the segmentation of the prostate gland into its zones could be helpful. Therefore, zonal segmentation of the prostate gland could be taken into account as another step forward. According to the appearance of the

www.manaraa.com

peripheral zone in T2w MRI (i.e. usually a brighter region compared to the surrounding tissues) the same local appearance-based segmentation method we used for whole prostate gland segmentation might be also applicable for zonal segmentation of the gland.

This segmentation method could also be utilized in segmentation of any other organs or objects that have convex shapes and inter-patient local appearance consistency withing different parts of the object boundary, despite of inter-patient appearance differences inside or outside of the object.

Finally, The conclusions of this work are valid only in the context of our own segmentation techniques, and the results might differ if the operator editing study were conducted with different algorithms. Moreover, due to the small size of the data sets and the number of operators involved, this should be considered to primarily be a set of hypothesis-generating studies that point the way to potentially fruitful studies on larger sample sizes with sufficient statistical power to draw firmer conclusions. We also did not study the effect of MR pulse sequence parameters on the results.

5.3 References

1. D. W. Cool, J. Bax, C. Romagnoli, A. D. Ward, L. Gardi, V. Karnik, J. Izawa, J. Chin and A. Fenster, "Fusion of MRI to 3D TRUS for mechanically-assisted targeted prostate biopsy: system design and initial clinical experience," in *Prostate Cancer Imaging. Image Analysis and Image-Guided Interventions*, (Springer, 2011), pp. 121-133.

2. J. Cepek, U. Lindner, S. Ghai, A. S. Louis, S. R. Davidson, M. Gertner, E. Hlasny, M. S. Sussman, A. Fenster and J. Trachtenberg, "Mechatronic system for in-bore MRI-guided insertion of needles to the prostate: An in vivo needle guidance accuracy study," J Magn Reson Imaging **42**, 48-55 (2015).

3. J. Cepek, B. A. Chronik, U. Lindner, J. Trachtenberg, S. R. Davidson, J. Bax and A. Fenster, "A system for MRI-guided transperineal delivery of needles to the prostate for focal therapy," Med Phys **40**, 012304 (2013).

المنارة للاستشارات
www.manaraa.com

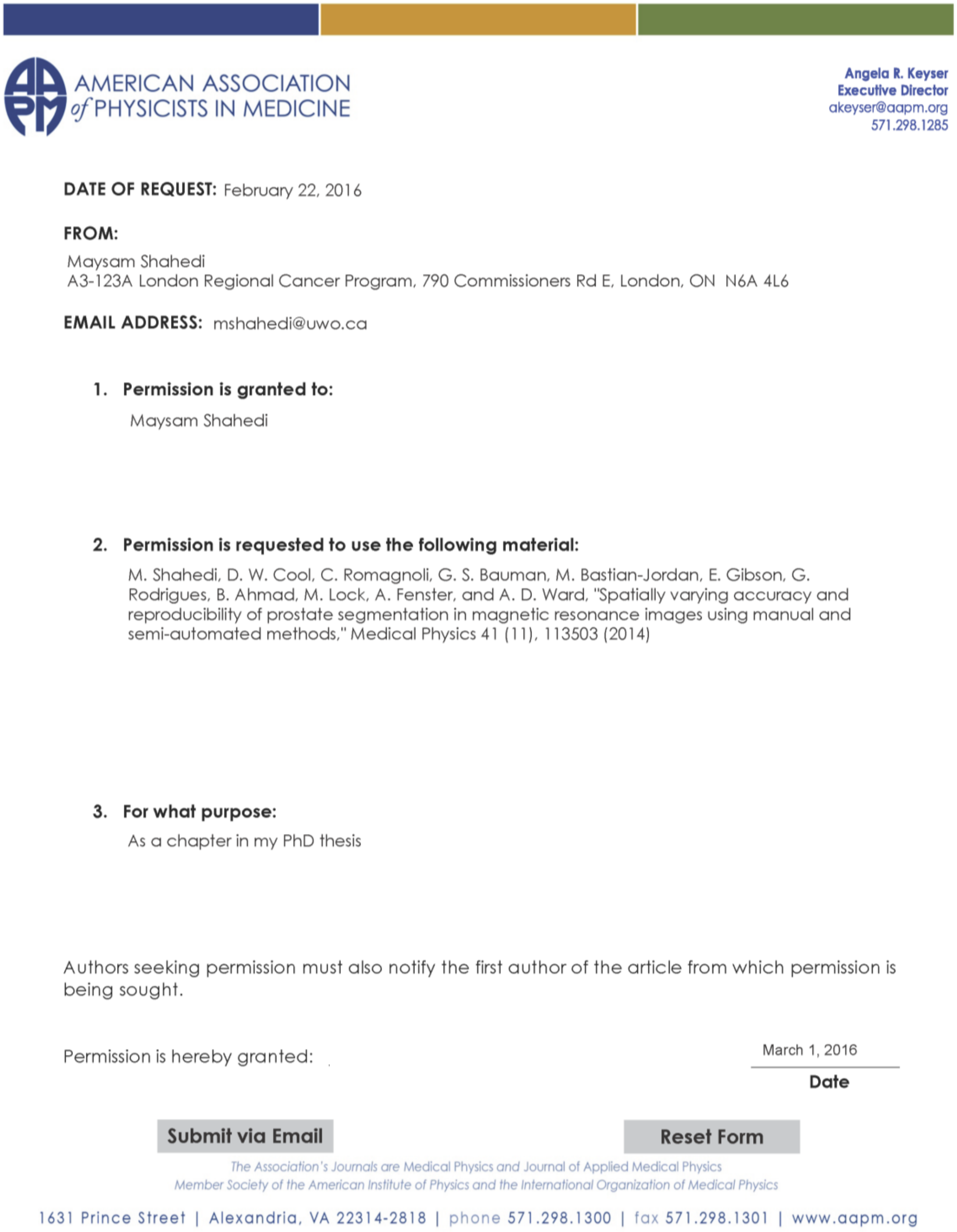
4. B. Marami, S. Sirouspour, S. Ghoul, J. Cepek, S. R. Davidson, D. W. Capson, J. Trachtenberg and A. Fenster, "Elastic registration of prostate MR images based on estimation of deformation states," Med Image Anal **21**, 87-103 (2015).

5. S. Natarajan, L. S. Marks, D. J. Margolis, J. Huang, M. L. Macairan, P. Lieu and A. Fenster, "Clinical application of a 3D ultrasound-guided prostate biopsy system," Urol Oncol **29**, 334-342 (2011).

www.manaraa.com

Appendix A

A.1 Permission to reproduce previously published material in Chapter 2



AMERICAN ASSOCIATION
of PHYSICISTS IN MEDICINE

Angela R. Keyser
Executive Director
akeyser@aapm.org
571.298.1285

DATE OF REQUEST: February 22, 2016

FROM:

Maysam Shahedi
A3-123A London Regional Cancer Program, 790 Commissioners Rd E, London, ON N6A 4L6

EMAIL ADDRESS: mshahedi@uwo.ca

1. **Permission is granted to:**

   Maysam Shahedi

2. **Permission is requested to use the following material:**

   M. Shahedi, D. W. Cool, C. Romagnoli, G. S. Bauman, M. Bastian-Jordan, E. Gibson, G. Rodrigues, B. Ahmad, M. Lock, A. Fenster, and A. D. Ward, "Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semi-automated methods," Medical Physics 41 (11), 113503 (2014)

3. **For what purpose:**

   As a chapter in my PhD thesis

Authors seeking permission must also notify the first author of the article from which permission is being sought.

Permission is hereby granted:

March 1, 2016
Date

The Association's Journals are Medical Physics and Journal of Applied Medical Physics
Member Society of the American Institute of Physics and the International Organization of Medical Physics

1631 Prince Street | Alexandria, VA 22314-2818 | phone 571.298.1300 | fax 571.298.1301 | www.aapm.org

www.manaraa.com

Curriculum Vitae

University Education

2010–	Ph.D. Biomedical Engineering (in progress) *Validation strategies supporting clinical integration of prostate segmentation algorithms for magnetic resonance imaging* Supervisors: Aaron D. Ward, Ph.D.; Aaron Fenster, Ph.D., FCCPM The University of Western Ontario, London, Canada
2007–2009	M.Sc. Electrical Engineering *Automatic Detection of Clustered Microcalcifications in Digital Mammograms* Supervisors: Rasoul Amirfattahi, Ph.D.; Farah Torkamani Azar Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
2001–2006	B.Sc. Electrical Engineering Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran

Undergraduate Honours, Scholarships and Awards

2012	Imaging Network Ontario Symposium — Third place poster award
2010–2015	NSERC CREATE CAMI Training Program ($30,000 stipend over 2 years)
2010–2014	Western Graduate Research Scholarship (WGRS) $6,705/Year
2010–2014	Western Graduate Research Scholarship International (WGRSI) $7,600/Year
2005	Iran Telecommunication Research Center Master Thesis Award

Publications, Presentations and Abstracts

Articles in refereed journals

1. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Spatially Varying Accuracy and Repeatability of Prostate Segmentation in Magnetic Resonance Images using Manual and Semiautomated Methods, Medical Physics, Vol. 41(11), 113503, 2014.

Refereed conference proceedings

1. M. Salarian, E. Gibson, **M. Shahedi**, M. Gaed, J. A. Gómez, M. Moussa, C. Romagnoli, D. W. Cool, M. Bastian-Jordan, J. L. Chin, S. Pautler, G. S. Bauman, A. D. Ward. Accuracy and variability of tumour burden measurement on multi-parametric MRI. In *SPIE Medical Imaging*, San Diego, USA, Feb. 2014. (**podium** presentation by M. Salarian)
2. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, A. D. Ward. Interactive 3D segmentation of the prostate in magnetic resonance images using shape and local appearance similarity analysis. In *SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling*, pp. 86710N, Orlando, USA, Mar. 2013. (**podium** presentation by **M. Shahedi**)
3. M. Salarian, **M. Shahedi**, E. Gibson, M. Gaed, J. A. Gómez-Lemus, M. Moussa, G. S. Bauman, A. D. Ward. Toward quantitative digital histopathology for prostate cancer: comparison of inter-slide interpolation methods for tumour measurement. In *SPIE Medical Imaging: Digital Pathology*, 8676, pp. 86760F, Orlando, USA, Mar. 2013. (**podium** presentation by M. Salarian)
4. **M. Shahedi**, R. Amirfattahi, F. Torkamani Azar, S. Sadri. Accurate Breast Region Detection in Digital mammograms, Using a Local Adaptive Thresholding Method. In *8th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 26, Greece, Jun. 2007.

www.manaraa.com

Refereed conference abstracts

1. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, World Congress on Medical Physics and Biomedical Engineering, Toronto, Ontario, Canada, June 2015. (**podium** presentation by **M. Shahedi**)
2. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, London Health Research Day, London, Ontario, Canada, April 2015.
3. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, Imaging Network Ontario (ImNO) Symposium, London, Ontario, Canada, March 2015.
4. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, Ontario Institute for Cancer Research Scientific Meeting, Toronto, Ontario, Canada, March 2015.
5. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic Methods, Imaging Network Ontario (ImNO) Symposium, Toronto, Ontario, Canada, March 2014.
6. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Semi-Automatic 3D Prostate Segmentation in Magnetic Resonance Images: Accuracy and Inter-operator Variability Measurement, London Health Research Day, London, Ontario, Canada, March 2014.
7. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, G. Bauman, A. D. Ward. 3D Segmentation of the Prostate in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, Canadian Cancer Research Conference, Toronto, Ontario, Canada, November 2013.
8. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, G. Bauman, A. D. Ward. 3D Segmentation of the Prostate in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, Annual Meeting Canadian Organization of Medical Physicists-Canadian Association of Radiation Oncology, Montreal, Canada, September 2013.
9. M. Salarian, **M. Shahedi**, E. Gibson, M. Gaed, J. A. Gómez, M. Moussa, D. W. Cool, C. Romagnoli, G. S. Bauman, A. D. Ward. Imaging validation and quantitative pathology in prostate cancer: shape interpolation methods for tumour measurement, Annual Meeting Canadian Organization of Medical Physicists-Canadian Association of Radiation Oncology, Montreal, Canada, Sep. 2013. (**podium** presentation by M. Salarian)
10. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, G. Bauman, A. D. Ward. 3D Segmentation of the Prostate in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, London Health Research Day, London, Ontario, Canada, June 2013
11. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, A. D. Ward. Interactive 3D segmentation of the prostate in magnetic resonance images using shape and local appearance similarity analysis, Imaging Network Ontario (ImNO) Symposium, Toronto, Ontario, Canada, February 2013. (**podium** presentation by **M. Shahedi**)
12. **M. Shahedi**, A. Fenster, C. Romagnoli, A. D. Ward. Semi-Automatic Segmentation of the Prostate Midgland in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, American Association of Physicist in Medicine (AAPM) 54th Annual Meeting, Charlott, North Carolina, USA, July 2012. (**podium** presentation by **M. Shahedi**)
13. **M. Shahedi**, A. Fenster, C. Romagnoli, A. D. Ward. A Local Appearance Similarity Analysis for Prostate Segmentation in Magnetic Resonance Images, Imaging Network Ontario (ImNO) Symposium, Toronto, Ontario, Canada, February 2012.

المنارة للاستشارات
www.manaraa.com

14. **M. Shahedi**, A. Fenster, C. Romagnoli, A. D. Ward. A Local Appearance Similarity Analysis for Prostate Segmentation in Magnetic Resonance Images, The Canadian Cancer Research Conference, Toronto, Ontario, Canada, November 2011.

Non-refereed conference abstracts

1. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, London Imaging Discovery, London, Ontario, Canada, June 2014.
2. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. Inter-Operator Variability of 3D Prostate MRI Segmentation Using Manual and Semi-Automatic approaches, Oncology Research and Education Day, London, Ontario, Canada, June 2014.
3. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, G. Bauman, A. D. Ward. 3D Segmentation of the Prostate in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, Oncology Research and Education Day, London, Ontario, Canada, June 2013.
4. **M. Shahedi**, D. W. Cool, C. Romagnoli, G. Bauman, M. Bastian-Jordan, E. Gibson, G. B. Rodrigues, B. Ahmad, M. Lock, A. Fenster, A. D. Ward. 3D Segmentation of the Prostate in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, London Imaging Discovery, London, Ontario, Canada, June 2013.
5. **M. Shahedi**, A. Fenster, D. W. Cool, C. Romagnoli, A. D. Ward. Semi-Automatic Segmentation of the Prostate Midgland in Magnetic Resonance Images Using Shape and Local Appearance Similarity Analysis, Oncology Research and Education Day, London, Ontario, Canada, June 2012.

Invited talks

1. **M. Shahedi,** Interactive 3D segmentation of the prostate in magnetic resonance images using shape and local appearance similarity analysis. Citywide Cancer Imaging Seminar. London, Canada. March 2013. Invited by David Palma and Aaron Ward.

Teaching Experience

2010-2014	The University of Western Ontario ENGSCI 1036A – Programming Fundamentals for Engineers – Teaching Assistant
2008	Iranian Academic Center of Education, Culture and Research, Isfahan, Iran Linear Control Systems – Course Lecturer
2007	Iranian Academic Center of Education, Culture and Research, Isfahan, Iran Linear Differential Equations – Course Lecturer
2007-2008	Iranian Academic Center of Education, Culture and Research, Isfahan, Iran Electrical Circuit Laboratory – Laboratory Instruction
2005	Isfahan University of Technology, Isfahan, Iran Digital Image Processing – Teaching Assistant
2005	Isfahan University of Technology, Isfahan, Iran Electronic Circuits Design – Teaching Assistant

المنارة للاستشارات

www.manaraa.com