LINKING AND CHARACTERIZING BIOLOGIC SCALES OF IMAGING DATA: APPLICATIONS TO PROSTATE CANCER DIAGNOSIS

BY RACHEL SPARKS

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

Graduate Program in Biomedical Engineering

Written under the direction of

Anant Madabhushi

and approved by

_____

_____

_____

_____

_____

_____

New Brunswick, New Jersey

Jan, 2014

ABSTRACT OF THE DISSERTATION

Linking and Characterizing Biologic Scales of Imaging Data: Applications to Prostate Cancer Diagnosis

by Rachel Sparks

Dissertation Director: Anant Madabhushi

Prostate cancer is the second most commonly diagnosed cancer of men, an estimated 192,000 men are diagnosed each year in the United States (source: American Cancer Society). The current gold standard for prostate cancer diagnosis is pathologist inspection of prostate needle biopsy samples obtained using transrectal ultrasound (TRUS). TRUS-guided biopsy is routine because TRUS is widely available and acquires real-time imagery. However, TRUS-guided biopsy has a low sensitivity, and initial biopsy misses approximately half of all men with prostate cancer. Multi-parametric Magnetic Resonance Imaging (MRI) has shown promise in detecting, localizing, and grading prostate cancer. MRI-TRUS fusion, whereby MRI is acquired pre-operatively then aligned to TRUS during biopsy, allows for both modalities to be leveraged. MRI-TRUS fusion will enable the construction of joint classifiers, which leverage imaging characteristics on both MRI and TRUS, to detect, localize, and grade prostate cancer. In order to train and validate these classifiers, ground truth spatial extent and aggressiveness of prostate cancer must be obtained. Manual pathologist inspection provides the ultimate definitive diagnosis of prostate cancer, with the Gleason grading system providing a measure of prostate cancer aggressiveness. Therefore whole mount histopathology (WMH) is aligned to fused MRI-TRUS imagery to provide ground truth of cancer location and

aggressiveness. A drawback to this approach is that Gleason grade is subject to inter- and intra-observer variability. Hence there is a need for reproducible, computer assisted grading of pathology which can serve as a surrogate for ground truth prostate cancer aggressiveness. In Aim 1 we develop a novel registration algorithm, multi-attribute probabilistic elastic registration (MAPPER), to align MRI and TRUS prostate imagery. In Aim 2 we align WMH with fused MRI-TRUS imagery (Aim 1). In Aim 3 we develop novel morphologic features to distinguish between aggressive and non-aggressive prostate cancer on histopathology. This will enable WMH to serve as ground truth for prostate cancer aggressiveness in order to train a MRI-TRUS classifier. Future work will leverage the tools developed to combine signatures of prostate cancer appearance across MRI, TRUS, and WMH and enable the development of tools to target biopsy to aggressive prostate cancer.

Preface

This dissertation represents a collection of published and unpublished works of the author [1–7]. It is primarily composed from the content of peer-reviewed conference and journal publications written during the course the author's graduate studies.

Acknowledgements

First and foremost, I want to thank my thesis advisor, Dr. Anant Madabhushi, for all of the support he has provided over the course of my research. Without Dr. Madabhushi's patience and insight this dissertation would not exist in its current form.

I would like to thank Drs. Micheal Feldman, John E. Tomaszewski, and Natalie Shih for their help in understanding the clinical evaluation of prostate histopathology and associated problems, in addition to providing and curating the prostate histopathology datasets used in this work. I would like to thank Drs. Nicholas Bloch, Dean Barratt, Daniel Moses, and Lee Ponsky for their help in providing and curating the prostate Magnetic Resonance Imaging and transrectal ultrasound. Finally, I would like to thank Dr. Ernest Feleppa for providing his expertise on transrectal ultrasound.

I would like to thank the research team, past and present, at the Center for Computational Imaging & Personalized Diagnostics (CCIPD), formerly known as the Laboratory for Computational Imaging and Bioinformatics (LCIB). In particular I would like to thank Drs. James Monaco, Jonathan Chappelow, Jun Xu, Shannon Agner, Gayou Xiao, Mirabela Rusu, Satish Viswanath, and Pallavi Tiwari for their mentorship and advice during the course of my research. I also would like to acknowledge George Lee, Ajay Basavanhally, Robert Toth, and Sahir Ali for their friendship and support during the course of my research.

Finally, I would like to thank all of my friends and family for their unwavering kindness and support. I want to thank Dr. Benjamin Griffel and Nathaniel Bade for always being there when I need them most. I especially want to thank my parents Roy and Amy Sparks and my sisters Sonia Puttonen and Christianna Mallon for their understanding and love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The current gold standard for prostate cancer diagnosis and grading is manual pathologist evaluation of histopathology acquired using transrectal ultrasound (TRUS)-guided biopsy [8]. Due to poor visualization of abnormal prostate tissue on TRUS, biopsies often sample benign tissue unnecessarily and miss prostate cancer nodules [9, 10]. Approximately half of all prostate cancer nodules are not detected using the current clinical protocol [11]. Additionally, $30 - 50\%$ of patients found to have low grade prostate cancer on biopsy will ultimately be determined to have higher grade prostate cancer after radical prostatectomy [12]. Hence, there is a clear clinical need to improve prostate biopsy to enable targeting of (a) regions suspicious for prostate cancer and (b) regions that contain aggressive prostate cancer. The overarching goal of this dissertation is to develop tools that will enable identifying and targeting prostate cancer during biopsy.

1.1 Image-Guided Prostate Biopsy

1.1.1 Transrectal Ultrasound for Guiding Biopsy

B-mode TRUS is useful in assessing gross anatomical details of the prostate, such as locating the borders of the prostate, identifying the seminal vesicles, and guiding insertion of the biopsy needle into selected regions of the prostate [13]. Prostate cancer may appear hypoechoic on TRUS but approximately 40% of prostate cancer nodules are isoechoic and cannot be distinguished from benign prostate tissue [9, 10]. Despite these shortcomings in visualizing prostate cancer, TRUS is routinely used because it is widely available and easy to use.

Due to the poor sensitivity in visualizing prostate cancer on TRUS, the clinical

standard for TRUS-guided needle biopsy is a blinded biopsy procedure. Blinded biopsy is performed by dividing the prostate into six regions, and sampling two biopsy cores from each of the six regions [8]. In addition, biopsy cores may be acquired from regions deemed suspicious for prostate cancer based on manual visual assessment of TRUS. Using the blinded biopsy procedure, TRUS-guided biopsy has a high false negative rate of $30 - 40\%$ [11, 14]. In patients with an initial negative biopsy $10 - 20\%$ will have a cancer detected on a 2nd biopsy and $4 - 5\%$ will have a cancer detected on a 3rd biopsy [11, 15]. Hence, there is a need to develop *in vivo* imaging methods that are more sensitive to prostate cancer and thereby improve biopsy targeting.

1.1.2 Multi-parametric Magnetic Resonance Imaging for Guiding Biopsy

Recent studies have demonstrated that magnetic resonance imaging (MRI) is able to detect prostate cancer with high sensitivity [16–19]. T2-weighted MRI, with a finer spatial resolution than TRUS, provides structural information about prostate cancer [16]. Additional MRI protocols can provide complementary information to T2-weighted MRI, for instance functional information can be obtained from dynamic contrast enhanced (DCE) [17] or diffusion weighted imaging (DWI) [18] and metabolic information can be obtained from magnetic resonance spectroscopy (MRS) [19]. Multi-parametric MRI schemes combine several MRI protocols, enabling prostate cancer to be detected with a high positive predictive value of 80% [20–24]. Additionally, multi-parametric MRI may be able to distinguish between more and less aggressive forms of cancer [25].

MRI-guided biopsy has a low false negative rate of $10 - 20\%$ for prostate cancer detection [26,27]. Despite the lower false negative rate for MRI-guided biopsy compared to TRUS-guided biopsy, there are several limitations. MRI-guided biopsy has long procedure times, is expensive due to the need for specialized equipment and technicians, and is stressful for many patients.

1.1.3 Registration of Magnetic Resonance Imaging with Transrectal Ultrasound for Guiding Biopsy

MRI-TRUS fusion – whereby MRI is spatially aligned to TRUS – enables combining anatomical, structural, functional, and metabolic information obtained from multi-parametric MRI with acoustic and anatomical information obtained from TRUS. Information from both modalities can then be leveraged to guide biopsy. The clinical protocol for MRI-TRUS fusion is as follows. A pre-operative prostate MRI is typically acquired 1 to 2 weeks prior to the biopsy procedure. This obviates the need for specialized biopsy equipment. The prostate MRI may then be analyzed, including pre-processing steps such as bias field correction, delineation of the prostate, and determining regions suspicious for cancer. During the subsequent biopsy procedure, TRUS is acquired to provide real-time guidance of the biopsy procedure. By performing the registration of the pre-operative MRI onto the TRUS both modalities can be used to guide needle placement. Utilizing fused MRI-TRUS to guide biopsy substantially increases the positive yield of prostate biopsies [28–31].

MRI-TRUS-guided biopsy schemes rely on manual visual assessment of multi-parametric MRI to determine needle placement [28–31]. However, identifying prostate cancer on multi-parametric MRI has high inter- and intra-observer variability [32]. The prostate imaging reporting and data system (PI-RADS) has been introduced in an attempt to reduce inter- and intra-observer variability. However, PI-RADS still has significant variability with $\kappa = 0.5 - 0.8$ (moderate to good inter-observer agreement) [33]. Additionally, PI-RADS does not currently contain guidelines for distinguishing between aggressive and non-aggressive prostate cancer. Finally, current MRI-TRUS-guided biopsy schemes do not take into account information from TRUS that may further improve prostate cancer detection. Hence, there is a need to develop computerized decision support tools that can aid clinicians by providing quantitative measures of prostate cancer presence and, in particular, aggressive prostate cancer presence.

1.2 Computerized Decision Support for Prostate Cancer Detection

1.2.1 Transrectal Ultrasound for Automated Prostate Cancer Detection

Despite the limitations of TRUS in visualizing [9,10] several computerized decision support systems have been developed to detect prostate cancer on TRUS [34–37]. Texture features extracted from pixel intensities on B-Mode TRUS have been used to distinguish between healthy tissue and prostate cancer with reported sensitivity $75 - 83\%$ and specificity of $70 - 85\%$ [34,35]. Feature selection to determine the best performing set of texture features on B-Mode TRUS resulted in a computerized decision support system with sensitivity of 80.0% and a specificity of 88.2% [36]. However, conventional B-Mode TRUS does not fully exploit the information contained in the radio-frequency (RF) signals received by ultrasound transducers. Feleppa *et. al.* [37] have developed methods for analyzing the RF signals obtained from TRUS resulting in a reported area under the receiver operator characteristic (ROC) curve of 0.844 for prostate cancer detection. A limitation of these studies is that ground truth was obtained by analysis of biopsy samples, therefore, it is impossible to assess whether prostate cancer nodules were missed in regions of the prostate where no biopsies were acquired.

1.2.2 Magnetic Resonance Imaging for Automated Prostate Cancer Detection

Several groups have developed computerized decision support systems for detecting prostate cancer on MRI [25,38–40]. A two stage classification scheme on multi-parametric MRI was shown to accurately detect prostate cancer with a sensitivity of 0.74, where ground truth was obtained from analysis of biopsy samples [38]. As previously mentioned, ground truth obtained from biopsy samples cannot assess missed prostate cancer nodules. Regions of interest containing prostate cancer can be distinguished from benign regions with an area under the ROC curve of 0.95 [40]. However, this method relies on the ability of a expert to accurately delineate the region of interest to classify. Inter-observer variability in region delineation has demonstrated negative effects

on computerized decision support performance. [41].

An alternative approach to systems that require region selection are decision support systems that operate at the voxel level. Utilizing whole mount histopathology specimens aligned to MRI for ground truth, computerized decision support systems for multi-parameter MRI can detect prostate cancer at the voxel level with an area under the ROC curve of 0.71 [39]. Combining features from multi-parametric MRI can also distinguish between low grade and high grade prostate cancer at the voxel level with an area under the ROC curve of 0.81 [25]. However, assessing prostate cancer aggressiveness, even on histopathology, to obtain ground truth is a difficult task.

1.2.3 Quantitative Histomorphometry for Assessing Prostate Cancer Aggressiveness

Prostate cancer on histopathology is typically assessed according to Gleason grade (from patterns 1 to 5) [42]. Gleason grade is an important predictor of prostate cancer aggressiveness and is often used to guide the treatment a patient receives [43]. Low Gleason grade patterns (≤ 3) are indicative of less aggressive prostate cancer; higher Gleason grade patterns (> 3) are indicative of more aggressive prostate cancer [44]. Manually distinguishing intermediate Gleason grade patterns 3 from 4 on histopathology is a difficult task. Inter-observer agreement between expert pathologists is $\kappa = 0.47 - 0.64$ (reflecting low to moderate agreement) [45]. Hence there is a clear clinical need to develop reproducible, quantitative histomorphometric features to complement pathologists in distinguishing the subtle differences between intermediate Gleason grades.

Quantitative histomorphometric features have been presented to assess Gleason grade on prostate histopathology [46–50]. Texture on histopathology has been shown to correctly identify Gleason grade patterns with $90-95\%$ accuracy [46,47]. Architectural arrangement of nuclei and glands has also been quantified to classify Gleason grade patterns with an accuracy of 76.0% [48]. Morphology of nuclei and glands can distinguish between Gleason grade patterns with an accuracy of $73 - 90\%$ [49, 50]. Recent work suggests that quantitative histomorphometric features that assess nuclei arrangement and gland orientation may be better able to predict prostate cancer aggressiveness

compared to Gleason grade alone [51–53].

1.2.4 Opportunity in Fusion of Magnetic Resonance Imaging - Transrectal Ultrasound for Creating Computerized Decision Support

Currently, no computerized decision support systems exist for detecting prostate cancer on fused MRI-TRUS imagery. As previously discussed in Sections 1.2.1 and 1.2.2, decision support systems for either TRUS or MRI are able to accurately detect prostate cancer. Viswanath *et. al.* [54] demonstrated that combining different imaging modalities can increase the performance of computerized decision support systems compared to any single constituent modality alone. Combining acoustic and anatomic information from TRUS with the structural, function, and metabolic information from MRI should further improve the accuracy of computerized decision support systems to detect prostate cancer.

Additionally, Singanmalli *et. al.* [55] demonstrated that a subset of MRI characteristics are strongly correlated with quantitative histomorphometric features indicative of prostate cancer aggressiveness. Tiwari *et. al.* [25] demonstrated that multi-parametric MRI can distinguish between low grade and high grade prostate cancer. These results suggest that there is a potential to train computerized decision support systems to distinguish between aggressive and non-aggressive prostate cancer on MRI. Hence it is reasonable to assume a MRI-TRUS computerized decision support system may also be able to distinguish between aggressive and non-aggressive prostate cancer.

1.3 Novel Contributions To Enable Fused Magnetic Resonance Imaging - Transrectal Ultrasound Decision Support

To enable the construction of a computerized decision support system for fused MRI-TRUS sophisticated image analysis tools must be developed. Specifically tools must be developed for (a) spatial alignment of the MRI and TRUS images of the prostate to create fused imaging signatures, (b) definitive ground truth acquired from pathologists' annotations on whole mount histopathlogy (WMH) must be aligned to the fused

MRI-TRUS, and (c) surrogate ground truth for prostate cancer aggressiveness must be obtained by developing quantitative histomorphometric features on histopathology. Although, there has been previous work for individual components of such a system to the best of our knowledge these components have never before been combined into a single framework.

In this dissertation a framework that will enable construction of a fused MRI-TRUS computerized decision support system is presented. The framework contains three novel contributions:

1. Multi-attribute probablistic prostate elastic registration (MAPPER) to fuse MRI and TRUS imagery [5].

2. Prostalign, a semi-automated thin-plate spline registration to align WMH onto fused MRI-TRUS imagery, thereby linking histopathology to fused MRI-TRUS imagery.

3. Explicit Shape Descriptors (ESDs), a novel quantitative histomorphometric feature to assess prostate cancer aggressiveness [3, 6, 7].

1.4 Organization of Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 presents fusion of MRI and TRUS prostate imagery via MAPPER with associated previous work, novel contributions, methodology, and experimental evaluation. Chapter 3 describes linking of prostate cancer across WMH, MRI, and TRUS imagery utilizing Prostalign with associated previous work, novel contribution, methodology, and experimental evaluation. In Chapter 4 a method to characterize prostate cancer aggressiveness on histopathology using ESDs is discussed with associated previous work, novel contributions, methodology, and experimental evaluation. Finally, Chapter 5 provides future work and concluding remarks.

Chapter 2

Fusion of Magnetic Resonance Imaging and Transrectal Ultrasound

In this chapter, a novel Magnetic Resonance Imaging (MRI)-Transrectal Ultrasound (TRUS) fusion algorithm, Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER), is presented. Some of the material presented in this chapter is taken from Sparks *et. al.* [5] of which the author of the dissertation is the first author.

Needle biopsy guided by TRUS is the current gold standard for prostate cancer diagnosis [56]. TRUS-guided biopsy is typically performed using a blinded procedure where the prostate is divided into six regions and two biopsy cores are taken from each region [8]. Additional cores may be acquired for regions that appear suspicious for prostate cancer on TRUS. However, approximately 40% of prostate cancer lesions appear isoechoic on TRUS; hence these lesions are difficult to accurately target using TRUS-guided biopsy [9, 10]. TRUS-guided biopsy is associated with a low cancer detection rate of $20 - 25\%$ [11]. Due to the low cancer detection rate for TRUS-guided biopsy, more than one in three men who have a prostate needle biopsy will undergo a repeat biopsy procedure [57].

Multi-parametric MRI is better able to visualize prostate cancer lesions [23]. T2-weighted MRI is able to provide anatomical information about the prostate in addition to structural information about prostate cancer [16]. Other MRI protocols provide complementary functional information, such as dynamic contrast enhanced (DCE) [17] and diffusion weighted imaging (DWI) [18], or metabolic information, such as magnetic resonance spectroscopy (MRS) [19]. The recent introduction of the Prostate Imaging Reporting Data System (PI-RADS) has help standardized the definition of suspicious prostate cancer on multi-parametric MRI, demonstrating moderate to high agreement

between observers [32]. In a study of 67 patients, MRI-guided biopsy in conjunction with PI-RADS was able to detect prostate cancer in 42% of patients [32]. Additional studies confirm these results, with MRI-guided biopsy yielding detection rates of $40 - 55\%$ [27, 58]. However, MRI-guided biopsy requires specialized equipment and technicians. Consequently, these procedures tend to be expensive and time-consuming, and are being done only at very few centers [27, 58].

MRI-TRUS fusion – whereby MRI is spatially aligned to TRUS – enables information from multi-parametric MRI and TRUS to be combined and leveraged to guide prostate needle biopsy procedures. Studies [28–31] have shown that utilizing TRUS in conjunction with MRI substantially increases the positive yield of prostate biopsies. The clinical protocol for MRI-TRUS fusion is typically as follows. A pre-operative prostate MRI is typically acquired 1 to 2 weeks prior to the biopsy procedure. The prostate MRI is then usually subjected to pre-processing steps such as bias field correction, delineation of the prostate, and determining regions suspicious for cancer. During the subsequent biopsy procedure, TRUS is acquired to provide real-time guidance.

There are several challenges to be overcome in MRI-TRUS registration. First, intensity-based metrics traditionally used for image registration, for instance Mutual Information [59], fail because of poor correlation in intensity between MRI and TRUS [60]. Second, differences exist in prostate shape on MRI and TRUS, caused by the different deformations induced by the TRUS probe and, when present, the MRI endorectal coil [61]. Figure 2.1 reveals the differences in prostate deformation between MRI and TRUS with (Figure 2.1(c)) and without (Figure 2.1(a)) an endorectal coil. Finally, the registration must be done in near real-time (< 5 minutes) to minimize the biopsy procedure time and maximize patient comfort. Any manual intervention to determine prostate location on TRUS or guide the registration of the MRI onto the TRUS will result in increased biopsy procedure time and, hence, it is important that all steps after TRUS acquisition involve minimal user interaction.

Current state-of-the-art MRI-TRUS fusion methods require varying degrees of manual intervention to establish spatial correspondence between MRI and TRUS imagery.

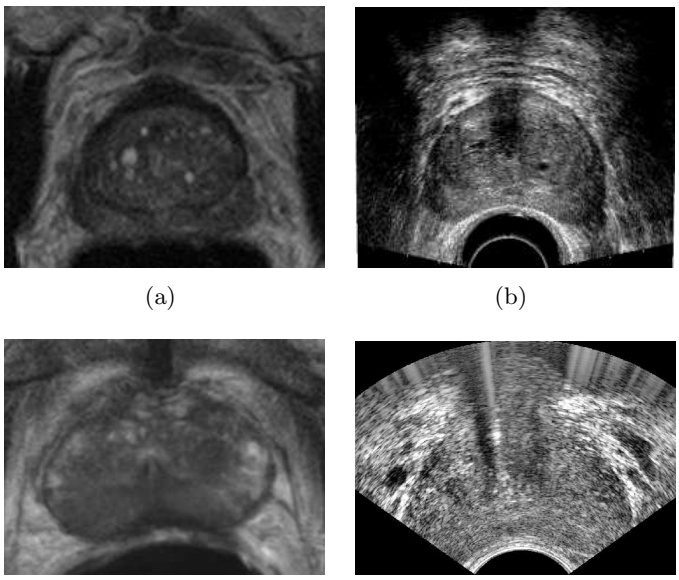


Figure 2.1: Two patient studies with corresponding MRI and TRUS prostate imagery obtained from two different datasets. Dataset 1: (a) 1.5 T MRI acquired with a pelvic phased-array coil and (b) side-firing TRUS probe. Dataset 2: (c) 3.0 T MRI acquired with an endorectal coil and (d) volumetric end-firing TRUS probe. Most current MRI-TRUS fusion schemes are not generalizable to images from different scanners, platforms, and field strengths.

However, manual intervention typically increases procedure time, along with potentially introducing error into the registration if the manually determined correspondence is incorrect. Labanaris *et. al.* [28] performed a study that divided 260 patients into two groups: (1) an 18-core TRUS-guided biopsy with no MRI information added, and (2) a similar biopsy procedure with additional cores sampled from regions suspicious for prostate cancer as determined by T2-weighted MRI. For the group undergoing only TRUS-guided biopsy, the cancer detection rate was 19.4%; the group with additional cores sampled from suspicious regions had a detection rate of 74.9%. Hadaschik *et. al.* [29] obtained a 59.4% cancer detection rate when using a semi-automated MRI-TRUS fusion system in 106 patients.

To overcome the need for time-intensive manual intervention during the biopsy procedure a new multi-modal registration methodology called Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER) to align MRI and TRUS images of the prostate is presented. The MAPPER algorithm involves: (1) Prior to the biopsy procedure, segmenting the prostate on the MRI; (2) During the biopsy procedure, calculating a multi-attribute probabilistic map of the prostate location on TRUS; (3) Maximizing overlap between the prostate segmentation on MRI and the multi-attribute probabilistic map of the prostate on TRUS, thereby driving elastic registration of MRI

and TRUS. MAPPER is well suited to MRI-TRUS fusion as it allows for automatically and accurately determining the location of the prostate on TRUS (Module 2). This step therefore obviates the need for manual intervention to register the images during the biopsy procedure. By utilizing an elastic registration (Module 3), MAPPER can account for differences in prostate deformation on the MRI and TRUS imagery.

The remainder of this chapter is as follows. In Section 2.1 previous work in MRI-TRUS fusion is discussed. Section 2.2 details the novel contributions of MAPPER. Section 2.3 describes the methodology of MAPPER. Section 2.4 describes our experimental design for evaluating MAPPER and Section 2.5 showcases our experimental results. In Section 2.6 concluding remarks on the MAPPER algorithm are provided.

2.1 Previous Work in Magnetic Resonance Imaging-Transrectal Ultrasound Fusion

Attributes	**Algorithm**						
	MAPPER	UroNav [62]	Kaplan *et. al.* [63]	Profuse [64]	Hu *et. al.* [65]	Urostation [66]	Mitra *et. al.* [60]
Requires manual prostate segmentation						✓	✓
Requires manual fiducial selection			✓	✓	✓		
Requires manual correction of prostate segmentation		✓					
Locates prostate automatically	✓						
Recovers prostate deformation	✓			✓	✓	✓	✓

Table 2.1: Comparison between MAPPER and state-of-the-art MRI-TRUS fusion algorithms. The attributes of each algorithm are indicated with a ✓.

Table 2.1 highlights the major differences between MAPPER and current state-of-the-art MRI-TRUS fusion algorithms [30,60,62–67]. MAPPER is able to perform MRI-TRUS fusion with no manual interaction during the biopsy procedure. By comparison, all of the state-of-the-art MRI-TRUS fusion methods require manual intervention to locate the prostate on TRUS [30, 60, 62–67]. The intervention may involve either the delineation of the prostate or the manual selection of fiducials visible on both modalities.

MRI-TRUS fusion methods can be divided into (a) fiducial [60, 62, 63, 66], (b) surface [30, 64, 67], and (c) model-based methods [65].

Fiducial-based methods attempt to find a transformation that minimizes the distance between corresponding landmarks on MRI and TRUS [60, 62, 63, 66]. Early work by Kaplan *et. al.* [63] used manually selected fiducials to determine a rigid transformation between MRI and TRUS imagery. Alignment was assessed qualitatively between the MRI-TRUS images. Bubley *et. al.* [68] evaluated the method of Kaplan *et. al.* [63] in a cohort of 30 prostate cancer patients. 16 out of 30 patients (53%) had a positive biopsy core obtained from a region suspicious for prostate cancer on MRI. Mitra *et. al.* [60] extracted the prostate surface and internal fiducials from a manual segmentation of the prostate; fiducials being used to determine a diffeomorphic transformation between MRI and TRUS imagery.

Xu *et. al.* [62] used fiducials extracted from an automated segmentation of the prostate on MRI and TRUS, with manual refinement of the segmentation, to determine an affine transformation. Using this method Pinto *et. al.* [69] were able to successfully diagnose prostate cancer in 55 out of 101 patients. Furthermore 17 out of 19 regions determined to be highly suspicious for prostate cancer on MRI corresponded to a cancer positive biopsy. The method of Xu *et. al.* [62] was also evaluated on a cohort of 125 patients where multi-parametric MRI showed a low suspicion of prostate cancer [70]. Only 10 patients had a cancer positive biopsy with Gleason score of 7 $(3 + 4)$, an additional 38 patients had a cancer positive biopsy with a Gleason score of 6 $(3 + 3)$, the Gleason score being a risk predictor of prostate cancer aggressiveness. Reynier *et. al.* [66] used fiducials extracted from a manual segmentation of the prostate to calculate an elastic transformation. This methodology has been used to guide brachytherapy [71] and biopsy [31]. Using the methodology of Reynier *et. al.* [66], prostate cancer was diagnosed in 54 out of 80 patients who had regions suspicious for prostate cancer on MRI [31].

Surface-based methods mitigate the need to select fiducials by finding a transformation that minimizes the distance between prostate surfaces on MRI and TRUS

[30, 64, 67]. Narayanan *et. al.* [64] aligned prostate surfaces obtained using a semi-automated segmentation scheme that required manual selection of four or more fiducials on the prostate surface. Karnik *et. al.* [67] used a thin-plate spline registration to align prostate surfaces, surfaces being obtained via a semi-automated segmentation method involving manual selection of 10 or more fiducials. Natarajan *et. al.* [30] extended this approach to (a) require picking of only $4-6$ fiducials on the prostate surface as well as (b) incorporating elastic interpolation when aligning MRI and TRUS. The system was used to guide biopsies in 56 patients and achieved a cancer detection rate of 23%, compared to 7% for systematic, nontargeted biopsies [30]. Further studies using the method of Natarajan *et. al.* [30] have achieved a detection rate of 53% in 171 men [72]; in patients with highly suspicious MRI findings 15 out of 16 had a positive biopsy. The difference in cancer detection rates between Natarajan *et. al.* [30] and Sonn *et. al.* [72] could be a reflection of the different patient populations considered in each study; Sonn *et. al.* [72] considered patients with persistently increased prostate specific antigen (PSA), a patient population at an elevated risk for prostate cancer.

A model-based method utilizes prostate segmentation on MRI to construct a Finite Element Model (FEM); the FEM is then deformed to align the prostate on MRI to TRUS [65]. FEM initialization on TRUS required specifying two fiducials on the base and apex of the prostate, respectively. Dickinson *et. al.* [73] performed a follow up study utilizing a variation of the method of Hu *et. al.* [65] to guide high intensity focused ultrasound (HIFU) ablation of prostate cancer lesions. In Dickinson *et. al.* [73], $10-20$ prostate surface points were selected to guide the fusion of MRI to TRUS in 26 patient studies. The fusion procedure was reported to take between $3-16$ minutes per case.

A limitation to most of the aforementioned MRI-TRUS fusion algorithms is the need for manual intervention during the biopsy. All of the previously described methods rely on user interaction to identify the location of the prostate on TRUS, either by selecting fiducials or delineating the prostate [30, 60, 62–67]. Apart from the increase in procedure time and patient discomfort caused by manual intervention during the biopsy procedure, manual intervention adds a source of variability into the registration algorithm. Inter-observer variability for manual prostate delineation on MRI is reported

to be 2.5 ± 1.2 mm [61]. Inaccuracies in selecting landmarks or delineating prostate boundaries manually may introduce error into the registration, although we are not aware of any studies that have explicitly reported this variation.

2.2 Novel Contributions and Brief Overview of Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER)

Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER) differs from state-of-the-art MRI-TRUS fusion algorithms in that it does not require manual intervention during the biopsy procedure to perform registration. MAPPER presents two novel contributions: a new method to estimate the location of the prostate on TRUS and a novel image registration metric to align a binary mask of the prostate to a probabilistic map of its location.

The first novel contribution of MAPPER allows for estimation of the location of the prostate on TRUS. This estimation is done by creating a probabilistic map of the prostate location that combines texture and spatial priors pertaining to prostate appearance. This approach was motivated by the utility of texture features and spatial location in identifying the prostate location on medical imagery [74–77]. In much the same way MAPPER calculates a probabilistic map of the prostate location on TRUS in order to facilitate registration.

The spatial prior, calculated from a set of training images, describes the probability of a pixel corresponding to the prostate according to spatial location relative to the TRUS probe. The texture prior, calculated as a Gaussian model from a set of texture features, describes the probability of a pixel corresponding to the prostate according to its local texture properties. While, Cosío [77] considered texture and spatial priors to segment the prostate on TRUS, MAPPER differs in several notable ways. (1) Cosío [77] considered only pixel intensity while MAPPER considers local intensity and texture properties, (2) Cosío [77] gave a hard decision (0 or 1) to each pixel while in MAPPER each pixel has a continuous probability value contained in the range of 0 to 1, (3) Cosío [77] considered the texture and spatial priors jointly while in MAPPER we assume the

texture and spatial priors are independent, and (4) Cosío [77] applied their methodology to segment the prostate on TRUS while MAPPER is applied to the registration of prostate MRI and TRUS.

The choice of texture features used to calculate the texture prior has important implications for the accuracy of MAPPER. Texture features which distinguish between prostate and background pixels will result in a more accurate registration compared to texture feature which are unable to distinguish between prostate and background pixels. In this work, several different types of texture features are considered including first-order texture features (mean, median, range, variance), edge detecting texture features (Gabor wavelet), and ultrasound specific features (Rayleigh, Nakagami m-parameter). The Rayleigh and Nakagami m-parameter texture features were considered because of their demonstrated utility in describing the statistics of different tissue types in ultrasound imagery [78, 79].

The inclusion of texture-based probability makes our registration algorithm sensitive to TRUS image appearance. Hence, it is important that TRUS imagery has a consistent appearance, in terms of pixel intensity and texture characteristics. However, ultrasound imagery may have attenuation artifacts, where pixels closer to the ultrasound probe appear brighter than pixels far away. Attenuation is caused by signal loss as the ultrasound waves propagate through tissue [80, 81]. As the TRUS probe is circular, variations in image intensity will be along radial lines from the probe. To account for changes in attenuation, correction methods have been developed [81]. Attenuation correction of TRUS imagery has been demonstrated to be important for segmentation of the heart on echo-cardiograms [81, 82]. In this work, attenuation correction is employed to facilitate and improve image registration.

The second novel contribution of MAPPER is a registration metric to align a binary mask onto a probabilistic model for registration of the T2-weighted MRI segmentation to the probabilistic map of the prostate location on TRUS. The similarity metric is calculated by combining the probability of individual pixels belonging inside and outside the prostate, the goal being to maximize the likelihood of accurate alignment of the prostate segmentation on MRI to the probabilistic map of prostate location on TRUS.

The rationale behind the metric is that it should return a high value for transformations where regions inside the MRI prostate segmentation align with pixels that have a high probability of being prostate. Conversely, the similarity metric should yield a low value for transformations where regions inside the MRI prostate segmentation align with pixels that have a low probability of being prostate.

2.3 Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER)

2.3.1 Notation

A $3D$ MRI volume $\mathcal{C}^m = (C^m, f^m)$ is defined by a set of $3D$ Cartesian coordinates C^m and the image intensity function $f^m(c) : c \in C^m$. The $3D$ prostate segmentation result is represented by $\mathcal{G}^m = (C^m, g^m)$ such that $g^m(c) = i$ for a pixel c belonging to class i, where $i = 1$ represents the prostate and $i = 0$ represents the background. A $3D$ TRUS volume $\mathcal{C}^u = (C^u, f^u)$ is defined in a similar way as $\mathcal{C}^m$. From $\mathcal{C}^u$ a probabilistic map $\mathcal{C}_i^{\mathbf{P}} = (C^u, \mathbf{P}_i(c))$ is calculated, where $\mathbf{P}_i(c) : c \in C^u$ is the probability of the pixel c belonging to class i. Table 2.2 lists the notation used to describe MAPPER. Figure 2.2 displays a flowchart of our methodology which consists of the following three modules:

- **Module 1:** Segment the prostate on MRI prior to TRUS acquisition via a semi-automated algorithm.
- **Module 2:** Create a multi-attribute probabilistic map of prostate location on TRUS. As an initial step attenuation correction is performed on the TRUS imagery. The probabilistic map is created by, (a) determining a spatial-based probability of the prostate on TRUS, (b) calculating a texture-based probability of the prostate on TRUS, and finally (c) estimating the probability of each pixel belonging to the prostate by combining the spatial and texture-based probabilities.
- **Module 3:** Register MRI prostate segmentation and TRUS probabilistic map. Registration is performed via an (a) affine transform to account for translation,

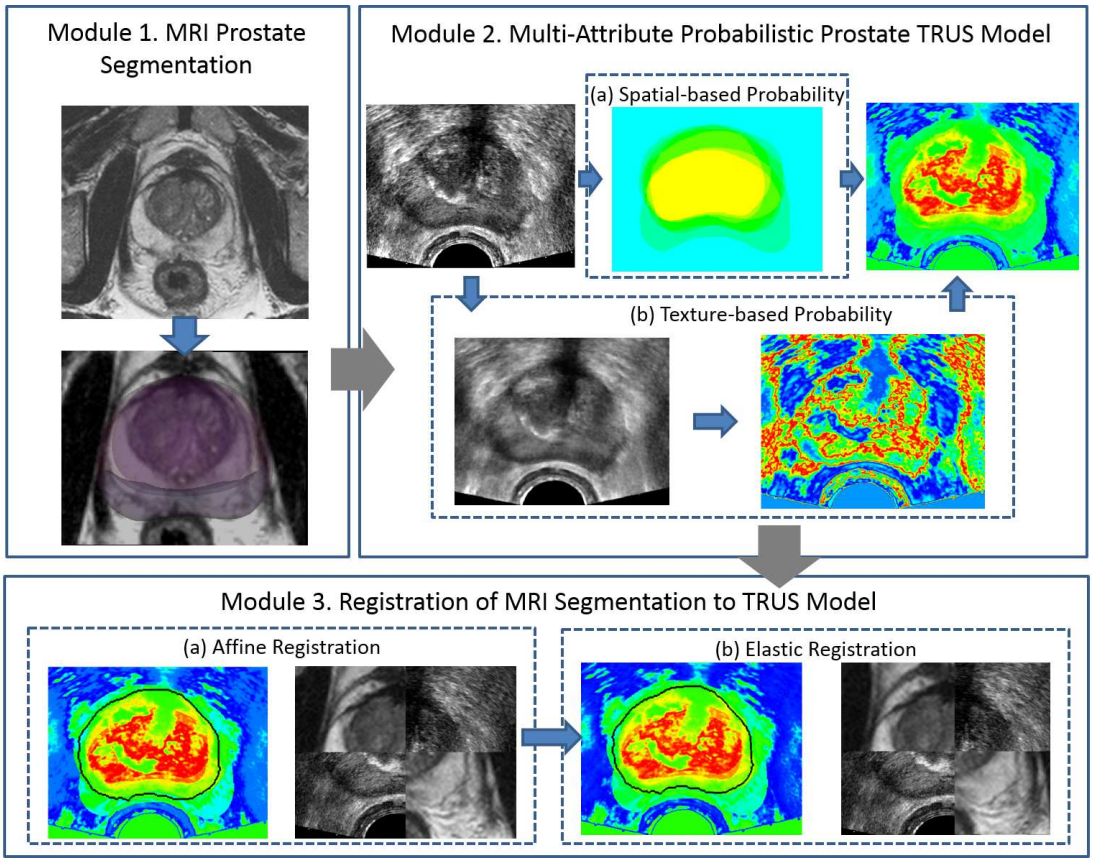


Figure 2.2: Flowchart for MAPPER comprises the following modules: (1) Prostate segmentation on MRI using a semi-automated algorithm (prostate segmentation shown in pink); (2) Construction of multi-attribute probabilistic model of prostate location on TRUS (blue corresponds to pixels least likely to belong to the prostate, red corresponds to pixels most likely to belong to the prostate). The probabilistic prostate estimation model consists of estimating (a) likely location of the prostate on TRUS (spatial-based probability) and (b) likely appearance of the prostate on TRUS (texture-based probability); These features are then combined to obtain a multi-attribute probability map of the prostate location on TRUS. (3) Registration of MRI segmentation to the probabilistic map of prostate location on TRUS by (a) affine (translation, rotation, scale) registration, followed by (b) elastic registration.

Notation	Description	Notation	Description
$\mathcal{C}^m$	$3D$ MRI image scene.	$P_i[F^u(c)]$	Probability of $F^u(c)$ belonging to class i.
C^m	$3D$ grid of pixels of $\mathcal{C}^m$.	Ω_i^u	Collection of pixels in C^u that belong to class i.
$\mathcal{G}^m$	$3D$ MRI prostate segmentation.	μ_i^F	Mean vector of $F^u(c)$ for Ω_i^u.
$\mathcal{C}^u$	$3D$ TRUS image scene.	Σ_i^F	Covariance matrix of $F^u(c)$ for Ω_i^u.
C^u	$3D$ grid of pixels of $\mathcal{C}^u$.	$\tilde{\mathcal{G}}^u$	Estimate of $3D$ TRUS prostate segmentation
$f^u(c)$	TRUS image intensity function for $c \in C^u$.	$T^{m\to u}$	Transformation function.
$\mathcal{G}^u$	$3D$ TRUS prostate segmentation.	$S[T^{m\to u}(\mathcal{G}^m), \mathcal{C}^u]$	Similarity metric for $T^{m\to u}(\mathcal{G}^m)$ and $\mathcal{C}^u$.
$\mathbf{P}_i(c)$	Probability of belonging to class i for $c \in C^u$.	$R(\cdot)$	Regularization metric for a transformation.
Ω_i^m	Collection of pixels in C^m that belong to class i.	p^e	Control point location defined on C^u.
$\tilde{f}^u(c)$	Attenuation corrected TRUS image intensity function for $c \in C^u$.	$E[p^e]$	Expected location of control point p^e.
$F^u(c)$	Set of texture features for $c \in C^u$.	$\mathcal{N}(p^e)$	Set of neighborhood control points for p^e.

Table 2.2: List of notation used to describe MAPPER.

rotation, and scale differences between images followed by (b) elastic transform to account for differences in prostate deformation.

2.3.2 Module 1: Prostate Segmentation on MRI

In this work, the prostate is segmented using a semi-automated algorithm based on the Multi-Feature Appearance (MFA) prostate segmentation scheme described in Toth and Madabhushi 2012 [74]. Figure 2.3 displays the algorithm used to semi-automatically segment the prostate on MRI. Since this work has been previously published, this method is only briefly described here and we refer the reader to Toth and Madabhushi [74] for additional details. The workflow comprises the following main steps.

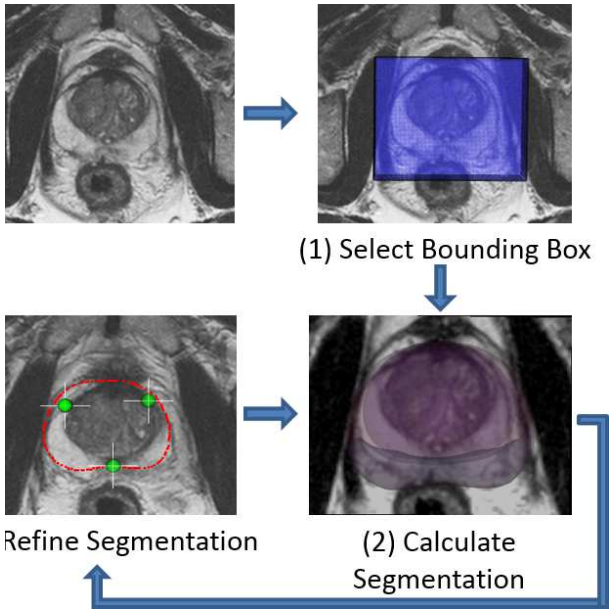


Figure 2.3: Flowchart for the semi-automated prostate segmentation scheme on MRI. Workflow comprises the following steps. (1) Selection of a bounding box (blue) on the MRI containing the prostate. (2) Automated segmentation of the prostate via a Multi-feature Appearance (MFA) model. The MFA estimated prostate surface (pink) is displayed to the user. (3) If manual correction is necessary, landmark points (green) can be placed on the true surface of the prostate, the prostate segmentation is then recalculated. Steps 2 and 3 are then iterated until the user determines the segmentation is accurate.

1. **Select Bounding Box:** A bounding box of the region containing the prostate is manually selected.

2. **Calculate Segmentation:** The MFA algorithm calculates the best segmentation of the prostate within the bounding box region, using shape and appearance features as described in Toth and Madabhushi [74].

3. **Refine Segmentation:** The segmentation may then be refined by selecting landmark points on the surface of the prostate. The landmark points constrain the MFA to always include the points on the surface of the prostate.

4. **Iterative Refinement:** Steps 2 and 3 are repeated until an accurate segmentation is achieved.

The accuracy of the MFA segmentation scheme is dependent on the selection of the bounding box (detailed in Step 1) and the landmark points (detailed in Step 3). Hence a sensitivity analysis of the MFA prostate segmentation scheme is performed (see Section 2.5.4).

2.3.3 Module 2: Probabilistic Model of Prostate Location on TRUS

As an initial step attenuation correction [81] is performed on $\mathcal{C}^u$ to account for spatial variations in image intensity. A probabilistic map of prostate location on TRUS defined as $\mathbf{P}_i(c)$ is then calculated by (1) extraction of texture features from $\mathcal{C}^u$ defined as $F^u(c)$, and (2) estimation of the likely prostate location (spatial prior) and appearance (texture prior).

Attenuation Correction

Attenuation correction is performed as follows. For each pixel $c \in C^u$ with a set of $3D$ Cartesian coordinates expressed as $X_i : i \in 1, 2, 3$ defined such that the probe center is $X_i = 0 : i \in 1, 2, 3$. To perform attenuation correction each pixel is first transformed into $3D$ polar coordinates as follows,

$$\begin{aligned} \theta_1 &= (X_1)^2 + (X_2)^2, \\ \theta_2 &= tan^{-1}(\frac{X_1}{X_2}), \\ \theta_3 &= X_3. \end{aligned} \tag{2.1}$$

Image attenuation is modeled within polar coordinates as,

$$f^u(\theta_i) = \beta(\theta_i)\tilde{f}^u(\theta_i) + \eta(\theta_i), \tag{2.2}$$

where $\tilde{f}^u(\theta_i)$ is the true, unknown TRUS signal associated with the location $\theta_i : i \in 1, 2, 3$. $\eta(\theta_i)$ is additive white Gaussian noise assumed to be independent of $\tilde{f}^u(\theta_i)$ as suggest in Xiao *et. al.* [81]. Similar to Cohen *et. al.* [83], $\beta(\theta_i)$ may be estimated via convolution of a smoothing Gaussian kernel with the image, i.e. a low-pass filtering of the signal. The true underlying signal may then be recovered using the equation,

$$\tilde{f}^u(\theta_i)) = \exp\{\log[f^u(\theta_i)] - \text{lpf}(\log[f^u(\theta_i)])\}, \tag{2.3}$$

where lpf is a low-pass filter. $\tilde{f}^u(\theta_i)$ is then converted back into $3D$ Cartesian coordinates, $\tilde{f}^u(c) : c \in C^u$. Figure 2.4 illustrates an example study where attenuation correction improved the results by over 1 mm.

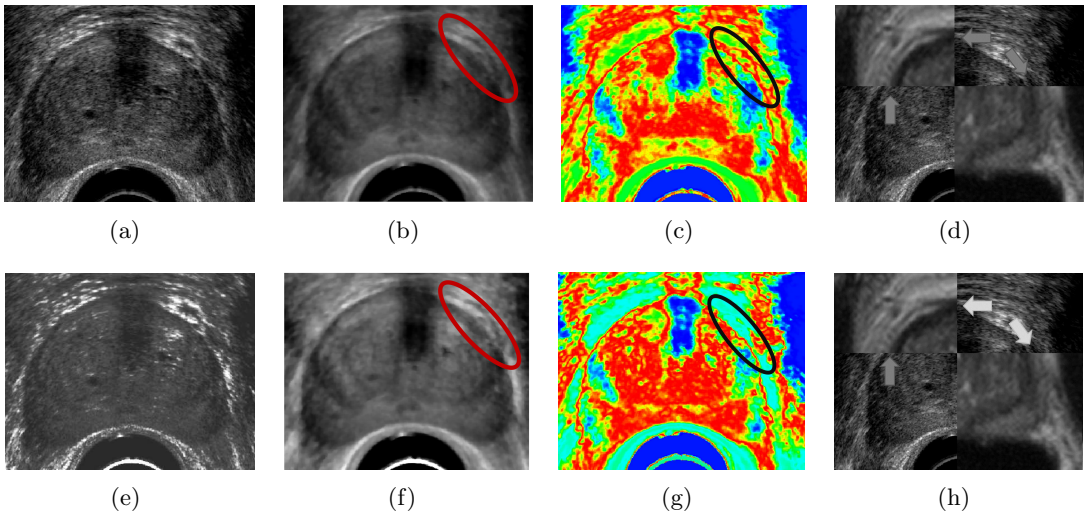
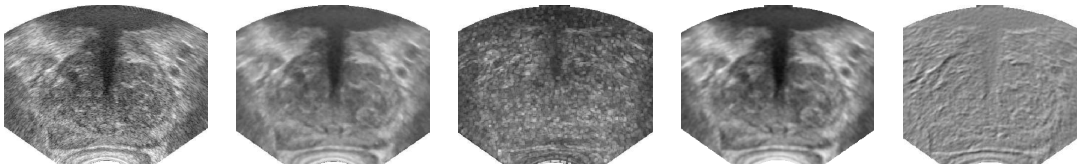


Figure 2.4: The panels reveal the importance of attenuation correction and the corresponding texture feature on MRI-TRUS registration. Panels (a) and (e) display the TRUS image without and with attenuation correction, respectively; (b), (f) the corresponding median feature; (c), (g) the corresponding probability models, where blue corresponds to those pixels least likely to belong to the prostate and red corresponds to those pixels most likely to belong to the prostate; and (d), (h) the final registration results. Light grey arrows in (d) and (h) show boundary regions which are well aligned on MRI and TRUS, while dark grey arrows show boundary regions which are misaligned. The region highlighted by the red circle in (b) and (f) show a region where attenuation correction improved the texture feature contrast between the prostate and background pixels. The corresponding region on the probability models is highlighted by the black circle in (c), (g). Note that the image with attenuation correction (g) is better able to distinguish between pixels belonging to the prostate from the background, resulting in a more accurate registration.

Feature Extraction

For each pixel $\tilde{f}^u(c) : c \in C^u$ a set of texture features $F^u(c)$ are calculated. The texture features chosen describe (a) intensity for a pixel or a region (intensity, mean, median), (b) intensity spread in a region (range), (c) intensity variation (variance, Rayleigh, or the Nakagami m-parameter), (d) edge information (Gabor wavelet). Figure 2.5 illustrates 4 representative texture features: (b) median, (c) range, (d) Rayleigh, and (e) Gabor wavelet.

Features that describe the intensity characteristics of a region are determined by defining a neighborhood of pixels $\mathcal{N}(c)$ for $c \in \mathcal{C}^u$ and then calculating a texture feature



(a) (b) (c) (d) (e)

Figure 2.5: The panels illustrate (a) the original TRUS image and 4 texture features, (b) median, (c) range, (d) Rayleigh, and (e) Gabor wavelet. For all texture features $\mathcal{N}(c)$ is defined A a spherical neighborhood of size 1 mm^3. The Gabor wavelet was calculated with a frequency of .01 Hz and an angle of $\pi/4$.

value. For instance the mean intensity value is calculated as $f_m(c) = \frac{1}{|\mathcal{N}(c)|} \sum_{d \in \mathcal{N}(c)} [\tilde{f}^u(d)]$. The median intensity value defined as $f_d(c)$ is similarly calculated for the median filter operator. The range texture feature defined as $f_r(c)$ describes the range of intensity values within $\mathcal{N}(c)$ for $c \in \mathcal{C}^u$. The range texture feature value is calculated as $f_r(c) = \max_{d \in \mathcal{N}(c)} [\tilde{f}^u(d)] - \min_{d \in \mathcal{N}(c)} [\tilde{f}^u(d)]$.

Intensity variation texture features are calculated to describe the spread of pixel intensity values assuming a specific underlying distribution. For instance the variance texture feature assumes that the underlying pixel distribution is Gaussian and is calculated as,

$$f_v(c) = \sqrt{\frac{1}{|\mathcal{N}(c)|} \sum_{d \in \mathcal{N}(c)} [\tilde{f}^u(d) - f_m(c)]^2}. \tag{2.4}$$

In a similar manner, the Rayleigh texture feature assumes an underlying distribution that describes well formed ultrasound scatter and is defined as,

$$f_y(c) = \sqrt{\frac{1}{2|\mathcal{N}(c)|} \sum_{d \in \mathcal{N}(c)} [\tilde{f}^u(d)]^2}. \tag{2.5}$$

The Nakagami m-parameter defined as f_n describes the shape of a distribution that is generalizable across different scatter conditions on ultrasound. To calculate the Nakagami m-parameter the iterative method described in Greenwood and Durand is used [84].

Finally, edge information is calculated using a set of texture features extracted from Gabor wavelets. Gabor wavelets are calculated by modulating a complex sinusoid with

a Gaussian function as described in [85]. The Gabor wavelets when convolved with the TRUS imagery return high values for regions with strong edges and low values for regions with weak edges. The feature set $F^u(c)$ is then defined as a subset of $[f_m, f_d, f_r, f_v, f_y, f_n, f_g]$.

Calculating Probability Map of Prostate Location on TRUS

The probability of pixel c belong to class i, defined as $\mathbf{P}_i(c)$, is dependent on the location of c and the feature set $F^u(c)$. We define the probability of a location c belonging to tissue class i as $P_i(c)$. Similarly, the probability of a set of features $F^u(c)$ belonging to tissue class i is $P_i[F^u(c)]$.

In this work, we assume that $P_i(c)$ and $P_i[F^u(c)]$ are independent, and hence the final probability $\mathbf{P}_i(c)$ may be expressed as,

$$\mathbf{P}_i(c) = P_i[F^u(c)] \times P_i(c). \tag{2.6}$$

Below the calculation of $P_i(c)$ and $P_i[F^u(c)]$ is described in further detail.

Spatial Probability: $P_i(c)$ is the likelihood of pixel c belonging to class i based on its spatial location. $P_i(c)$ is calculated from a set of J training studies $\mathcal{C}_j^u : j \in \{1, \ldots, J\}$. For each study the prostate has been delineated by an expert yielding the $3D$ prostate segmentation $\mathcal{G}_j^u$. The origin for each study is set as the center of the TRUS probe, so that the location of pixel c has a consistent position relative to the probe across all studies. $P_i(c)$ is the frequency of pixel c being located in the prostate across J training studies and is defined as,

$$P_i(c) = \frac{1}{J} \sum_{j=1}^{J} g_j^u(c). \tag{2.7}$$

Feature Probability: The probability $P_i[F^u(c)]$ is the likelihood of a set of features $F^u(c)$ associated with pixel c belonging to class i. In this work, we assume $F^u(c)$ may be accurately modeled as a multivariate Gaussian distribution with a mean vector μ_i^F and a covariance matrix Σ_i^F for the ith class. Given the Gaussian distribution parameters

μ_i^F and Σ_i^F the probability $P_i[F^u(c)]$ is calculated as,

$$P_i[F^u(c)] = \frac{1}{2\pi^{k/2}(\Sigma_i^F)^{1/2}} e^{[(F^u(c)-\mu_i^F)'(\Sigma_i^F)^{-1}(F^u(c)-\mu_i^F)]}, \tag{2.8}$$

where k is the number of features in $F^u(c)$. However μ_i^F and Σ_i^F are unknown, therefore, these parameters must be estimated.

To estimate μ_i^F and Σ_i^F, first the location of the prostate on TRUS is estimated by assuming an initial rigid transformation T^r (Section 2.3.4). The estimated prostate segmentation is then defined as $\hat{\mathcal{G}}^u = T^r(\mathcal{G}^m)$ where $\hat{\mathcal{G}}^u = (C^u, \hat{g}^u)$ and $\hat{g}^u(c) = i$ for a pixel c estimated to belong in class i. μ_i^F and Σ_i^F are then calculated as, $\mu_i^F = \frac{1}{|\Omega_i^u|}\sum_{\hat{c}\in\Omega_i^u} F^u(\hat{c})$, where Ω_i^u is the collection of pixels in C^u belonging to class i according to $\hat{g}^u(c)$. Σ_i^F is the covariance matrix from class i similarly defined for $F^u(\hat{c})$ and Ω_i^u.

2.3.4 Module 3: Registration of MRI Segmentation and TRUS Probabilistic Model

To goal of image registration is to find a transformation $T^{m\to u}$ to spatially map $\mathcal{C}^m$ onto $\mathcal{C}^u$. In this work, $T^{m\to u}$ is calculated to align $\mathcal{G}^m$ and $\mathcal{C}^u$. $T^{m\to u}$ is calculated via the equation,

$$T^{m\to u} = \underset{T^{m\to u}}{\operatorname{argmax}}\big[S\big[T^{m\to u}(\mathcal{G}^m\big], \mathcal{C}^u) - \alpha R(T^{m\to u})\big], \tag{2.9}$$

where $S\big[T^{m\to u}(\mathcal{G}^m), \mathcal{C}^u\big]$ is a similarity metric between $T(\mathcal{G}^m)$ and $\mathcal{C}^u$. $R(T^{m\to u})$ is a regularization function which penalizes $T^{m\to u}$ for not being smoothly varying and α reflects the weight of $R(\cdot)$ relative to $S(\cdot,\cdot)$.

The similarity metric $S(\cdot,\cdot)$ is calculated as,

$$S(T^{m\to u}(\mathcal{G}^m), \mathcal{C}^u) = \Pi_{i=0}^1 \Pi_{c\in C^u}\big[P_i[F^u(c), c] | T^{m\to u}(\mathcal{G}^m) = \Omega_i^m\big], \tag{2.10}$$

where Ω_i^m is the collection of pixels in C^m belonging to class i. $T^{m\to u}$ is initialized with a rigid transformation T^r such that overlap between $\mathcal{G}^m$ and $P_1(c)$ is maximized. The rigid transform is calculated as,

$$T^r = \underset{T^r}{\operatorname{argmax}}\big[\Pi_{c\in C^u} P_1(c) \times T^r[g^m(c)]\big]. \tag{2.11}$$

Given the initial alignment T^r, an affine registration T^a followed by an elastic registration T^e is used to align the MRI and TRUS images. The final transformation is then calculated as $T^{m\to u} = T^e\big(T^a(T^r)\big)$.

Affine Registration

For the affine transformation T^a no regularization $R(T^a)$ is used since T^a is by definition smoothly varying. Not defining $R(T^a)$ is equivalent to setting $\alpha = 0$.

Elastic Registration

An elastic B-spline-based transformation T^e is used to recover differences in prostate deformation between MRI and TRUS [86]. T^e is defined by a set of knots which determine the transformation T^e for all $c \in C^m$. Each knot, defined by its location $p^e \in C^m$, is allowed to move independently (see Figure 2.6).

The term $R(T^e)$ is added to constrain T^e to only those transformations which are likely to occur. $R(T^e)$ is calculated as,

$$R(T^e) = \sum_{p^e \in T^e} (1 - e^{-\|p^e - E[p^e]\|}), \tag{2.12}$$

where p^e is the location of a B-Spline knot and $E[p^e]$ is the maximum likelihood estimate of the location for knot p^e. In this work $E[p^e]$ is estimated as,

$$E[p^e] = \frac{1}{|\mathcal{N}(p^e)|} \sum_{\hat{p}^e \in \mathcal{N}(p^e)} \hat{p}^e \tag{2.13}$$

where $\mathcal{N}(p^e)$ is the set of knots which neighbor p^e. Thus $E[p^e]$ is the average over the set of knots which neighbor the knot p^e. Figure 2.6 gives a $2D$ illustration of the function $R(T^e)$. In our experiments $R(T^e)$ is calculated in $3D$.

$R(T^e)$ is defined such that if $p^e = E[p^e]$ then the knot p^e will not contribute to the value of $R(T^e)$. As p^e moves farther from $E[p^e]$, the value $1 - e^{-\|p^e - E[p^e]\|}$ increases and contributes more to the value of $R(T^e)$. Hence $R(T^e)$ is lower for evenly spaced, smoothly varying knots compared to randomly spaced, erratically varying knots. Deformations that are not evenly spaced and smoothly varying will only occur if they improve the similarity metric $S(\cdot,\cdot)$.

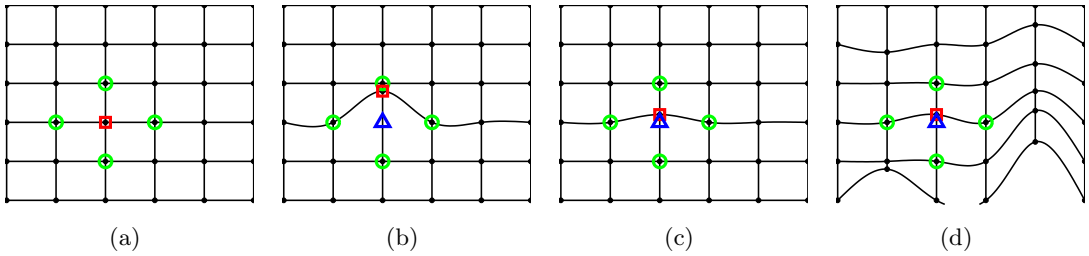


Figure 2.6: A graphical illustration of the regularization constraint $R(T^e)$. Black points correspond to the B-Spline knot locations. A knot of interest p^e is shown enclosed by a red square. (a) The initial knot locations are used to determine neighborhood knots for p^e and denoted as $\mathcal{N}(p^e)$. Knots corresponding to $\mathcal{N}(p^e)$ are shown enclosed by green circles. The expected location of p^e defined as $E[p^e]$ is shown enclosed by a blue triangle. (b) Example where $R(T^e)$ would have a high value because p^e is far from $E[p^e]$. (c) and (d) would give a low $R(T^e)$ value because p^e is near $E[p^e]$. For (d) the deformation not local to p^e is not taken into account when considering $E[p^e]$, other knots may contribute to a higher $R(T^e)$ compared to (c).

2.4 Experimental Design and Results

2.4.1 Dataset Description

MAPPER is evaluated on two different cohorts of MRI and TRUS. The first cohort comprised 6 patients with pelvic phased-array coil MRI and $2D$ ultrasound. The second cohort comprised 7 patients with endorectal coil MRI and $3D$ ultrasound. For all studies an expert radiologist manually selected corresponding fiducials on the MRI and TRUS images. Corresponding fiducials include, the urethra, the center for those locations deemed suspicious for prostate cancer, and the center of small calcifications. In addition, an expert radiologist manually delineated the prostate boundary on MRI and TRUS.

Dataset 1 ($\mathcal{D}_1$): Side-firing Transrectal Probe

T2-weighted MRI was acquired using a Siemens 1.5 T scanner and a pelvic phased-array coil for 6 patients under IRB approval. TRUS imagery was acquired using a B-K Profocus probe that acquires $2D$ transverse B-mode images of the prostate. The TRUS probe was attached to a mechanical stepping device used to translate the probe perpendicular to the axial plane at 2 mm intervals. For each patient one TRUS volume

was acquired, where each volume consists of a set of parallel B-mode planes. A single expert radiologist selected corresponding fiducials between all 6 MRI-TRUS pairs.

Dataset 2 ($\mathcal{D}_2$): Volumetric End-firing Transrectal Probe

T2-weighted MRI was acquired using a General Electric (GE) 3.0 T scanner and an endorectal coil for 7 patients under IRB approval. TRUS imagery was acquired using a GE 4DE7C probe, that acquires $3D$ data in a single, multi-plane sweep of the prostate. For each patient $1-3$ volumes were acquired, where each volume is acquired directly from the ultrasound device. A total of 13 MRI-TRUS pairs were acquired for the 7 patients. Two expert radiologist selected corresponding fiducials between the MRI-TRUS pairs. Expert 1 selected corresponding fiducials for 10 studies and Expert 2 selected corresponding fiducials for 5 studies.

2.4.2 Performance Evaluation

Root Mean Squared Error (RMSE)

RMSE is a measure of how well two corresponding point sets align; a RMSE of 0 represents perfect alignment. A manually selected set of fiducials on MRI is defined as $p_i^m : i \in \{1, \ldots, N\}$. Similarly, a set of fiducials on TRUS is defined as $p_i^u : i \in \{1, \ldots, N\}$, such that p_i^m corresponds to p_i^u. RMSE is then calculated as $\frac{1}{N}\sum_{i=1}^{N}(p_i^m - p_i^u)^2$.

Mean Absolute Deviation (MAD)

MAD is a measure of the average extent of variation between a ground truth manual delineation and an automatically determined delineation. Given a ground truth prostate segmentation $\mathcal{G}^m$, the collection of pixels in C^m belonging to class i is defined as Ω_i^m. Similarly, for an automatically generated prostate segmentation $\hat{\mathcal{G}}^m$, obtained as described in Section 2.3.2, $\hat{\Omega}_i^m$ is defined. MAD is calculated as,

$$\frac{1}{|p|}\sum_{p \in \Omega_i^m} \min_{o \in \hat{\Omega}_i^m} ||p - o||. \tag{2.14}$$

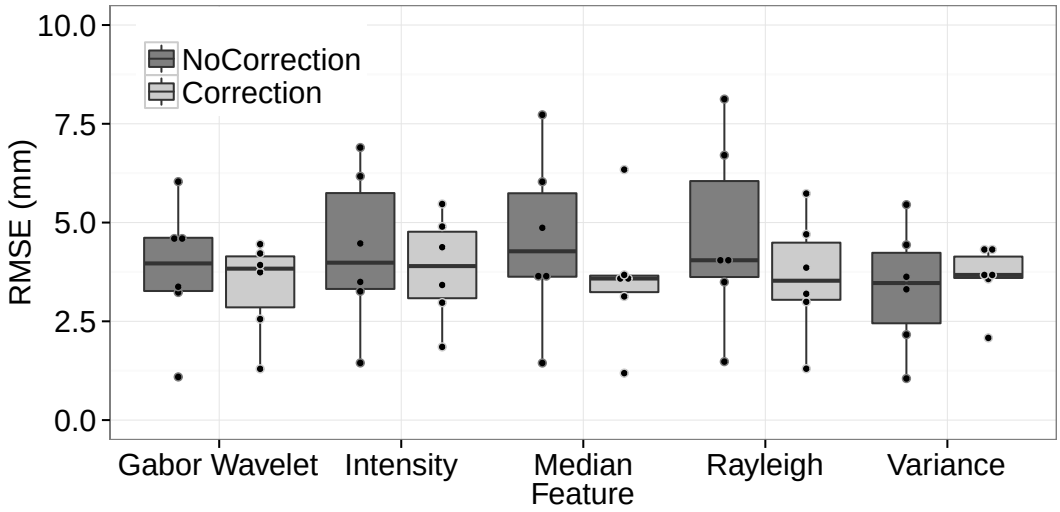


Figure 2.7: RMSE for 5 texture features with and without attenuation correction. Attenuation correction has the positive effect of improve registration accuracy independent of the choice of texture feature.

2.4.3 Implementation Details

All methods described in this Chapter were implemented using the Insight Segmentation and Registration Toolkit (ITK) version 4.5 [87]. All texture features were calculated using a $\mathcal{N}(c)$ with a spherical neighborhood of size 1 mm^3, determined empirically to be large enough to accurately represent local image statistics while small enough to capture only local image statistics. Both T^a and T^e were found via a Powell optimization scheme using a single resolution [88].

2.5 Experimental Results and Discussion

2.5.1 Experiment 1: Effect of Attenuation Correction on Registration Accuracy

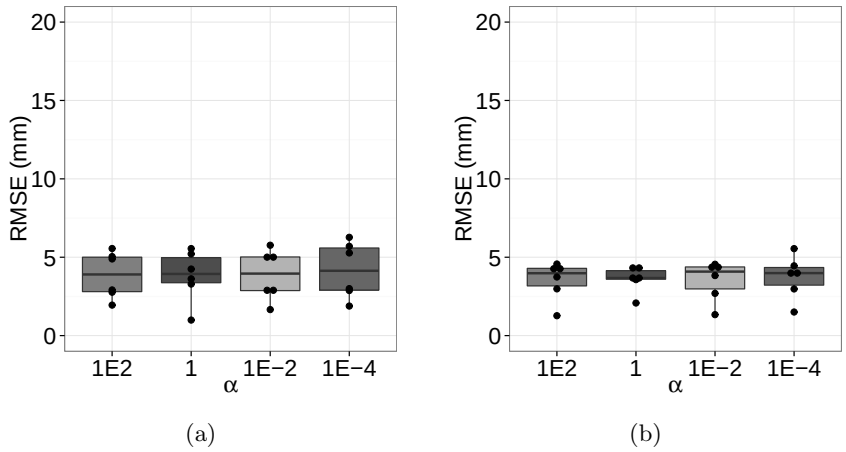
Subtle differences in intensity characteristics across the TRUS image may lead to a probabilistic model $\mathbf{P}_i(c)$ that does not accurately model the prostate location. Incorrect estimation of $\mathbf{P}_i(c)$ can result in sub-optimal image registration. In this experiment, the effects of attenuation correction, as described in Section 2.3.3, are evaluated on registration accuracy in terms of RMSE. MAPPER is evaluated with and without attenuation

correction for $\mathcal{D}_1$.

Figure 2.7 illustrates the quantitative results for T^e with and without attenuation correction for 5 texture features. Attenuation correction has two effects on the registration results (1) it reduces RMSE variation across studies and, therefore, gives a more robust image registration and (2) it lowers RMSE and, hence, provides a more accurate registration accuracy. The positive effects of attenuation correction on registration occur independent of texture feature used.

2.5.2 Experiment 2: Selection of Regularization Weight

The regularization weight α controls the relative importance of a smooth T_e and accurately registering the prostate mask on MRI to the TRUS probabilistic model (i.e. maximizing Equation 2.10). To assess the sensitivity of the performance of MAPPER on the choice of α, we varied α for $\{100, 1, 1\text{E-}2, 1\text{E-}4\}$ and assessed RMSE for $\mathcal{D}_1$.



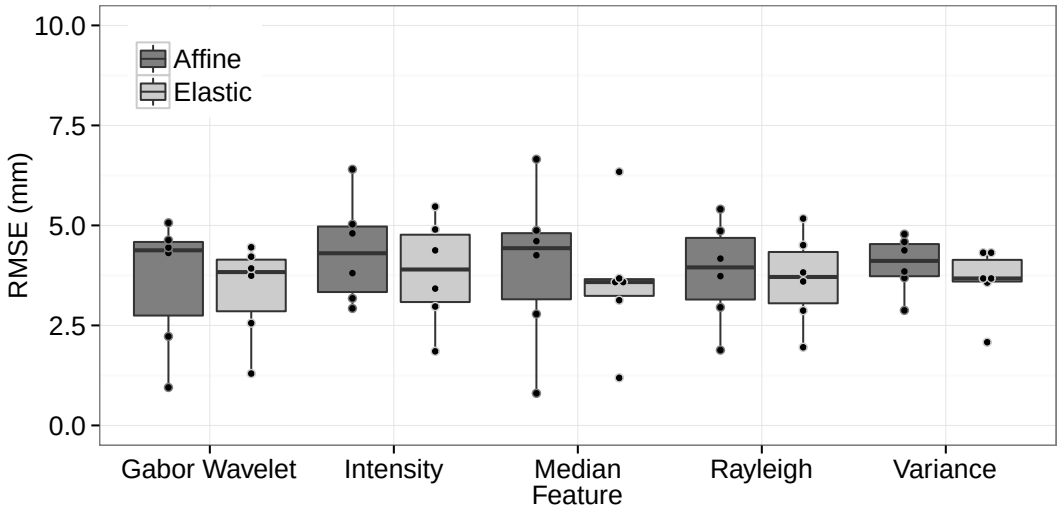
(a) (b)

Figure 2.8: RMSE for $\mathcal{D}_1$ as a function of α for two features: (a) Rayleigh and (b) variance.

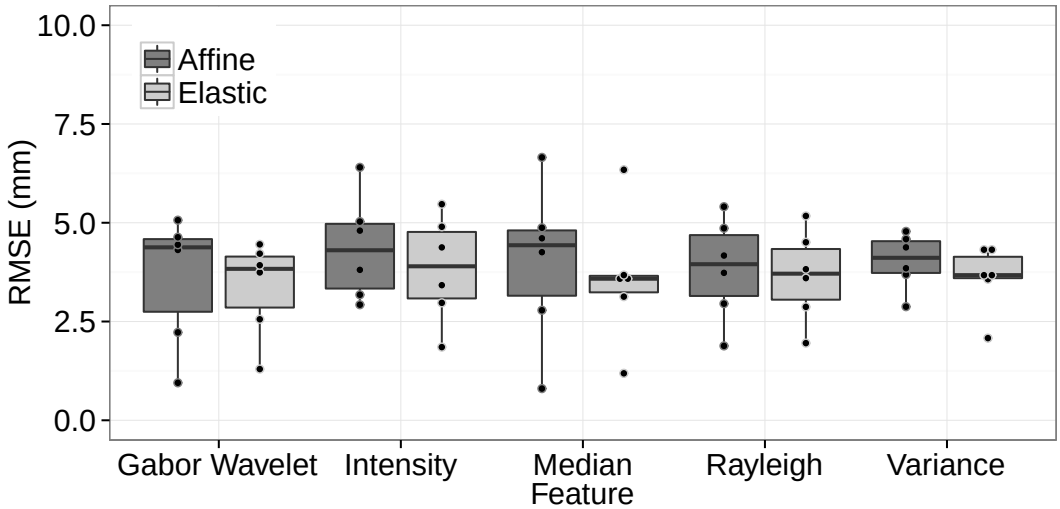
Figure 2.8 illustrates the RMSE for each set of regularization parameters α evaluated for two texture features, (a) Rayleigh and (b) variance. RMSE changes little with respect to α, even across the wide range of values considered ($\alpha \in \{100, 1, 1\text{E-}2, 1\text{E-}4\}$).

2.5.3 Experiment 3: Selection of Features For Creating Probabilistic Map of Prostate on TRUS

The accuracy of $\mathbf{P}_i(c)$ depends on the choice of texture features in $F^u(c)$; texture features which are best able to distinguish prostate from non-prostate tissue will lead to

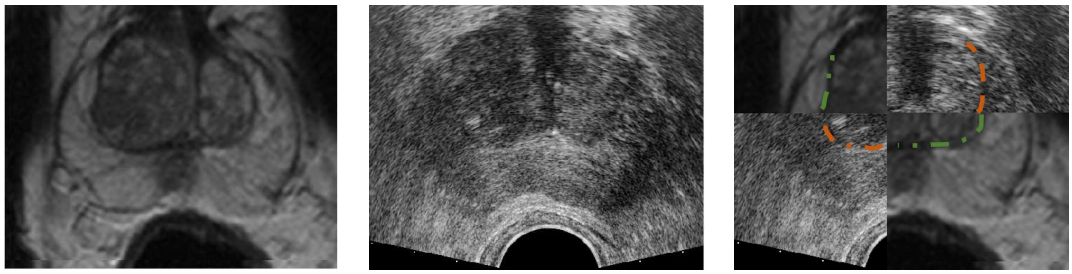


(a)



(b)

Figure 2.9: RMSE for T^e and T^a evaluated over 5 texture features on (a) $\mathcal{D}_1$ and (b) $\mathcal{D}_2$. For $\mathcal{D}_1$ variance and Gabor wavelet texture features were the best performing. For $\mathcal{D}_2$ intensity and Rayleigh were the best performing texture features. The difference in the best performing texture features for $\mathcal{D}_1$ and $\mathcal{D}_2$ demonstrates that prostate appearance may be specific to the TRUS probe.



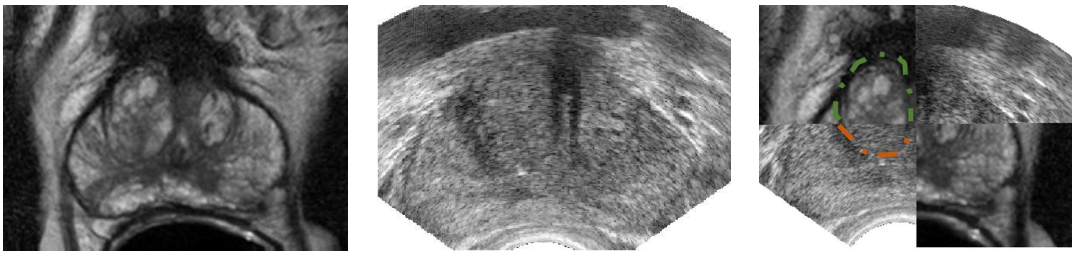
(a) (b) (c)

Figure 2.10: An example MRI-TRUS registration on $\mathcal{D}_1$ for T^e. (c) Checkerboard overlay of the MRI and TRUS images. Dotted lines on checkerboard image reveal the delineation of the central gland surface for MRI (green) and TRUS (orange).

a more accurate $\mathbf{P}_i(c)$ and therefore to a more accurate image registration. MAPPER is evaluated for 7 texture features described in Section 2.3.3 in terms of RMSE for both datasets. Additionally, for $\mathcal{D}_2$ RMSE is compared between expert radiologists to evaluate inter-observer variability.

Figure 2.9 illustrates the RMSE for 5 of the 7 texture features evaluated for (a) $\mathcal{D}_1$ and (b) $\mathcal{D}_2$. Each dataset has a different set of best performing texture features. For $\mathcal{D}_1$, the side-firing TRUS probe, variance and Gabor wavelet texture features were best able to align the MRI and TRUS imagery. For $\mathcal{D}_2$, the end-firing TRUS probe, intensity and Rayleigh were identified as the best performing texture features. The selection of different texture features for $\mathcal{D}_1$ and $\mathcal{D}_2$ most likely reflects differences in imaging characteristics between $\mathcal{D}_1$ and $\mathcal{D}_2$. Although MAPPER was able to align images with an average RMSE of approximately 3 mm the results here clearly reflect the importance of feature selection for accurate registration.

$\mathcal{D}_1$, where MRI was acquired with a pelvic phased-array coil, demonstrates an improvement in RMSE between T^a and T^e. In comparison $\mathcal{D}_2$, where MRI was acquired with an endorectal coil, had relatively little improvement in RMSE between T^a and T^e. These differences in RMSE improvement between T^a and T^e are indicative of $\mathcal{D}_1$ having larger differences in prostate deformation between MRI and TRUS compared to $\mathcal{D}_2$. Figure 2.10 shows the registration result for a representative study from $\mathcal{D}_1$ while Figure 2.11 shows a representative registration result for a study from $\mathcal{D}_2$. For



(a) (b) (c)

Figure 2.11: An example MRI-TRUS registration on $\mathcal{D}_2$ for T^e. (c) Checkerboard overlay of the MRI and TRUS images. Dotted lines on checkerboard image reveal the delineation of the surface of a lateral lobe of the prostate for MRI (green) and TRUS (orange).

both cases MAPPER aligns the prostate surface as well as internal structures which are highlighted by dotted lines.

MAPPER is also evaluated on $\mathcal{D}_2$ with respect to fiducials selected on MRI and TRUS by two expert radiologists. Figure 2.12 presents RMSE for each of the two experts for T^e. For the best performing texture features (Rayleigh, intensity) the difference in RMSE between the two experts is roughly 0.3 mm.

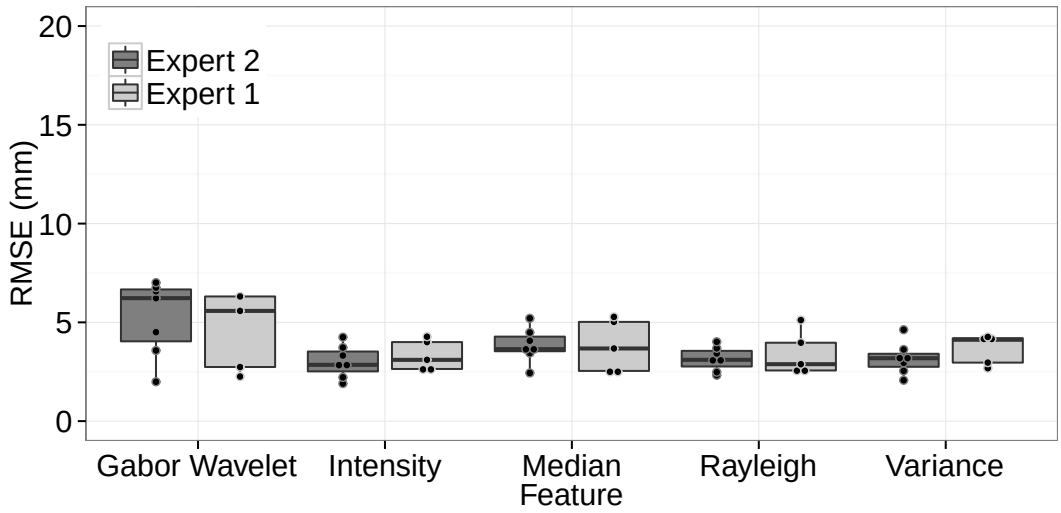


Figure 2.12: RMSE for T^e evaluated on $\mathcal{D}_2$ over two different expert observers and 5 texture features: Gabor wavelet, intensity, median, Rayleigh, and variance.

Method	Reference	RMSE
Urostation	Reynier *et. al.* [66]	2.07 ± 1.57 mm (urethra) 1.11 ± 0.54 mm (prostate surface)
ProFuse	Narayanan *et. al.* [64]	3.06 ± 1.41 mm (phantom)
	Karnik *et. al.* [67]	2.13 ± 0.80 mm
UroNav	Xu *et. al.* [62]	2.3 ± 0.9 mm (phantom)
-	Hu *et. al.* [65]	2.40 mm (median)
MAPPER	-	3.36 ± 1.10 mm ($\mathcal{D}_1$) 3.14 ± 0.75 mm ($\mathcal{D}_2$)

Table 2.3: A comparison of state-of-the-art MRI-TRUS fusion algorithms and MAPPER in terms of RMSE. MAPPER was rigorously evaluated across two datasets from two different institutions, with different TRUS probes, MRI coils and field strengths. Additionally, MAPPER was evaluated with respect to fiducials identified by two different experts. The state-of-art algorithms typically report a single RMSE value evaluated at a single institutional, hence, drawing conclusions about the relative performance of MAPPER compared to state-of-the-art algorithms is difficult.

MAPPER is able to register MRI-TRUS images with a RMSE of approximately 3 mm for $\mathcal{D}_1$ and $\mathcal{D}_2$. MAPPER compares favorably with state-of-the-art MRI-TRUS fusion methods (see Table 2.3). Although MAPPER has a slightly higher RMSE compared to some of the methods listed in Table 2.3, it is automated for all steps after the TRUS acquisition. All other methods listed in Table 2.3 require manual user intervention during the biopsy procedure to align MRI and TRUS images. MAPPER was evaluated for two independent datasets acquired at two different institutions with different TRUS probes, MRI coils and field strength. Hence it is notable that MAPPER is robust in terms of RMSE across $\mathcal{D}_1$ and $\mathcal{D}_2$. It should be noted that comparing performance against previously published methods is challenging at best, due to the use of different (a) datasets, (b) strategies for determining ground truth, and (c) evaluation performed in either $2D$ or $3D$. Given that MAPPER was evaluated on two completely different datasets and against multiple ground truth annotations from different experts, the accuracy of methods in Table 2.3 may reflect a less robust analysis rather than better registration performance.

To further evaluate the accuracy of MAPPER, surface renderings of the prostate are created as shown in Figure 2.13, blue and red represent regions where the MRI was

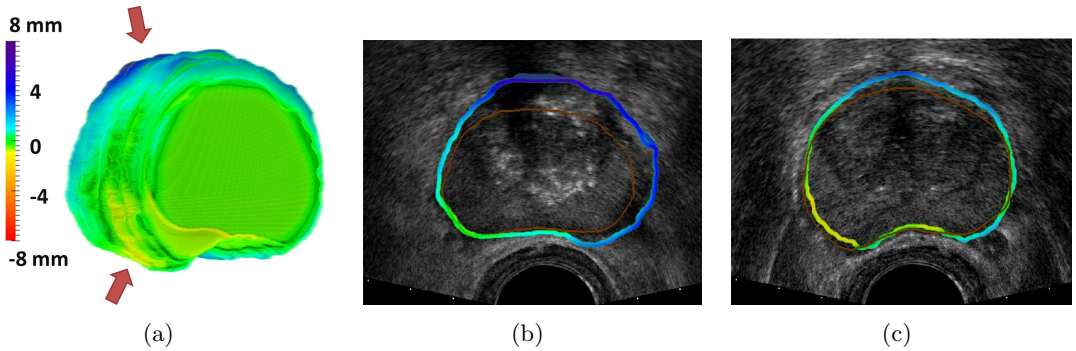
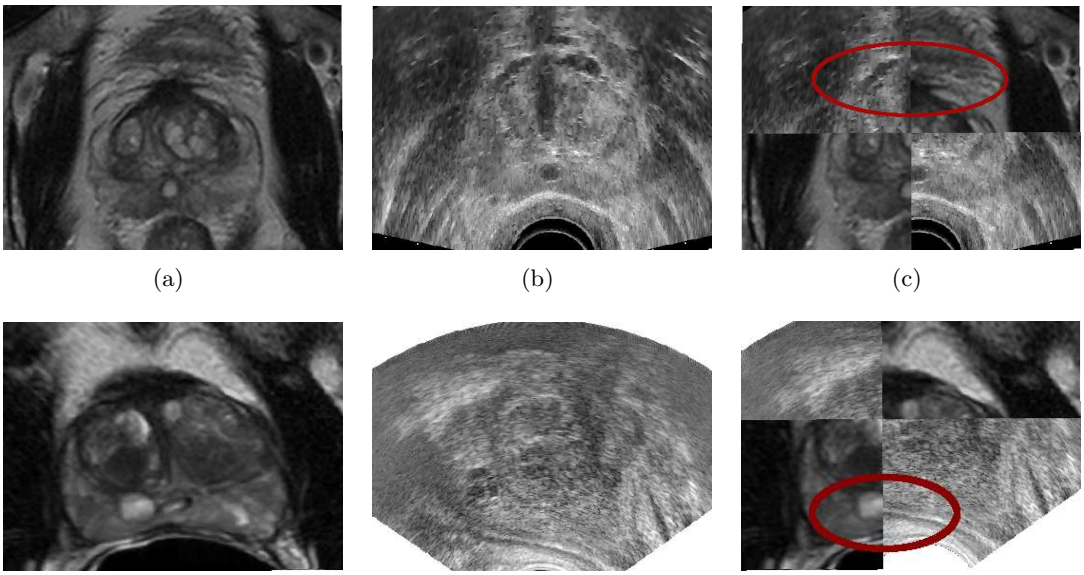


Figure 2.13: (a) Prostate surface rendering with the prostate base facing toward the right, blue and red represent regions where the MRI was misaligned external and internal to the prostate surface on TRUS, respectively. $2D$ axial TRUS image displaying a region of large misalignment (b) distal to the TRUS probe and (c) near the TRUS probe, in both images the brown region represents the expert delineation of the prostate on TRUS.

misaligned external and internal to the prostate surface on TRUS, respectively. In the example shown in Figure 2.13 there are two regions of misalignment, near the rectal wall (yellow) and near the bladder (blue). Figure 2.13(b) illustrates an axial plane of the TRUS displayed with two boundaries overlaid. These represent (1) the axial cross section of the surface rendering shown in Figure 2.13(a) and the true prostate boundary (brown line). The hyperechoic region distal to the TRUS probe caused $\mathbf{P}_i(c)$ to inappropriately model the location of the prostate, resulting in a registration error of $\approx$ 4 mm. Similarly Figure 2.13(c) illustrates a different axial plane of the TRUS displayed with the cross section of the surface rendering shown in Figure 2.13(a) and the true prostate boundary (brown line). Note that this misalignment is much less pronounced, representing a registration error of $\approx$ 1 mm. Near the rectal wall the error is primarily due to T^e being unable to fully account for the subtle differences in prostate deformation.

One shortcoming of MAPPER is that poor TRUS image quality negatively impacts the registration results due to the reliance of MAPPER on TRUS image appearance for the texture-based probability. For other MRI-TRUS fusion methods, manual intervention is used to determine the prostate location. Poor TRUS image quality may also



(a) (b) (c)

(d) (e) (f)

Figure 2.14: Two cases where poor TRUS image quality negatively impacted registration performance. (a)-(c) An example of a poor registration from $\mathcal{D}_1$ where (b) the TRUS has severe intensity artifacts. Note that for this study, intensity on the TRUS appears blurry with poor definition of the prostate boundary. Due to these artifacts, MAPPER fails to determine the location of the prostate on TRUS, leading to a poor registration highlighted by the red circle. (d)-(f) A study from $\mathcal{D}_2$ where (e) the TRUS has a strong shadowing artifacts and abnormal prostate deformation on the right hand side of the image. MAPPER is unable to account for these differences as is visible in (f) the checkerboard image and highlighted by the red circle.

effect these methods, if the expert observer is unable to provide accurate manual intervention. Figure 2.14 shows two examples, one from each dataset evaluated, where poor TRUS image quality resulted in inaccurate alignment of the TRUS and T2-weighted MRI. Both of these studies were clear outliers, in terms of poor image quality and/or large deformation in the prostate.

2.5.4 Experiment 4: Effects of Prostate MRI Segmentation Accuracy

The accuracy of MAPPER is evaluated in the context of variation in segmentation performance on account of different levels of manual intervention to segment the prostate. For both $\mathcal{D}_1$ and $\mathcal{D}_2$ the top performing texture feature identified in Experiment 3 (Section 2.5.3) were used to perform this evaluation. Different levels of user interaction

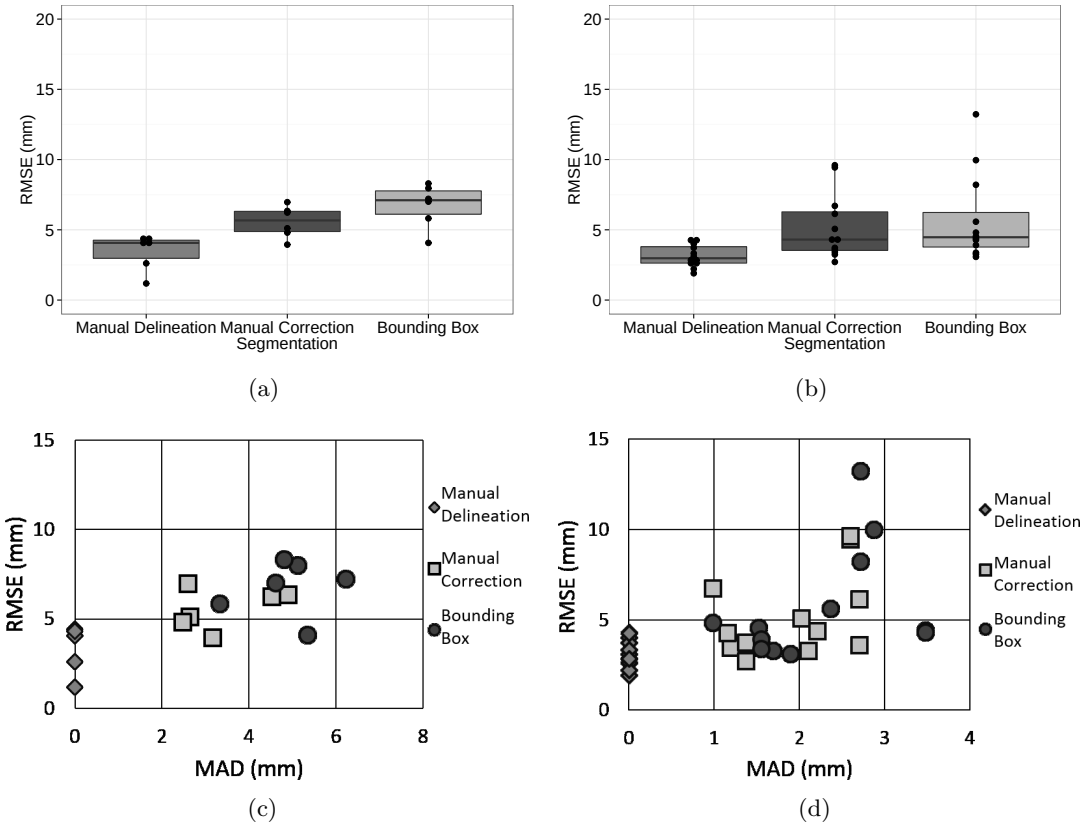


Figure 2.15: RMSE as a function of prostate segmentation algorithm used for (a) variance for $\mathcal{D}_1$ and (b) intensity for $\mathcal{D}_2$. Additionally, registration accuracy versus segmentation accuracy is illustrated for (c) variance for $\mathcal{D}_1$ and (d) intensity for $\mathcal{D}_2$. Accurate segmentation schemes, which require more manual intervention, result in more accurate image registration.

were evaluated via the following strategies.

- **Bounding box:** Manual selection of bounding box of the region containing the prostate prior to MFA segmentation.
- **Manual correction:** Manual selection of bounding box of the region containing the prostate and selection of landmark points to correct the automated segmentation if necessary.
- **Manual delineation:** Manual delineation of the prostate on MRI by an expert radiologist.

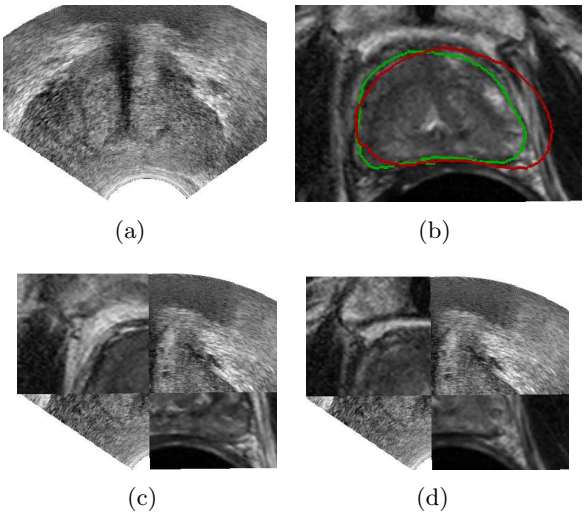


Figure 2.16: 2.16(a) TRUS and (b) corresponding MRI with prostate segmentation for manual (green) and bounding box-based segmentation (red) with corresponding MRI-TRUS registration from $\mathcal{D}_2$ for (c) manual segmentation and (d) bounding-box segmentation. The differences between the bounding-box segmentation and the true location of the prostate (manual segmentation) result in a large misalignment in the prostate surface for (d).

Figure 2.15 illustrates registration accuracy, in terms of RMSE, for each segmentation scheme. Manual prostate delineation, the most accurate segmentation scheme, also has the best registration accuracy. The manual correction of the semi-automated segmentation scheme resulted in an improve registration compared to the semi-automated scheme without manual intervention.

For $\mathcal{D}_2$ there were outliers in terms of RMSE when utilizing the bounding box segmentation method. Figure 2.16 shows one such outlier case, where registration error occurs due to a mis-segmentation of the prostate boundary. The large discrepancies in the prostate segmentation shown in Figure 2.16(b), caused inaccurate registration as shown in Figure 2.16(d).

2.6 Concluding Remarks

This chapter presented a novel registration methodology, Multi-Attribute Probabilistic Prostate Elastic Registration (MAPPER), to spatially align MRI and TRUS images of the prostate. MAPPER was evaluated on 13 patient studies from two datasets– Dataset 1 had 6 studies with a side-firing TRUS probe and 1.5 T surface coil MRI, Dataset 2 had 7 studies with a volumetric end-firing TRUS probe and 3.0 T endorectal coil MRI. RMSE for MAPPER was found to be 3.36 ± 1.10 mm for Dataset 1 and 3.14 ± 0.75 mm for Dataset 2. Unlike previously described MRI-TRUS fusion algorithms [30, 60, 62–67], MAPPER requires no manual intervention. Specifically, MAPPER uses

a semi-automated segmentation scheme on MRI in conjunction with a probabilistic map of the prostate location on TRUS to register MRI onto TRUS. Hence, MAPPER automatically detects and aligns the prostate on MRI and TRUS, whereas state-of-the-art methods rely on manual intervention to either delineate the prostate or select corresponding fiducials on MRI and TRUS.

A limitation of this work is the use of the B-Spline transformations in Module 3 (Section 2.3.4), which recover non-linear deformations with few additional constraints, to account for the difference in deformation of the prostate between MRI and TRUS imagery. In this work, an additional regularization constraint was imposed to ensure the underlying deformation in the prostate was smoothly varying. However, other transformations such as Finite Element Models (FEM), which allow for explicit modeling of tissue physics, could also potentially be used to drive the MRI-TRUS fusion [65]. In future work, other transformations and regularization constraints will be considered to model the differences in deformation of the prostate between MRI and TRUS imagery.

As evidenced by the results in Experiment 4, the accuracy of MAPPER is reliant on an accurate segmentation of the prostate on MRI. The prostate segmentation algorithm is performed offline prior to the biopsy procedure using a Multi-Feature Appearance (MFA) model of prostate appearance on MRI previously described in Toth and Madabhushi [74]. Future work will be directed towards evaluating in detail the performance of MAPPER for (a) independent manual delineations of the prostate and (b) different semi-automated and automated prostate segmentation algorithms.

Chapter 3

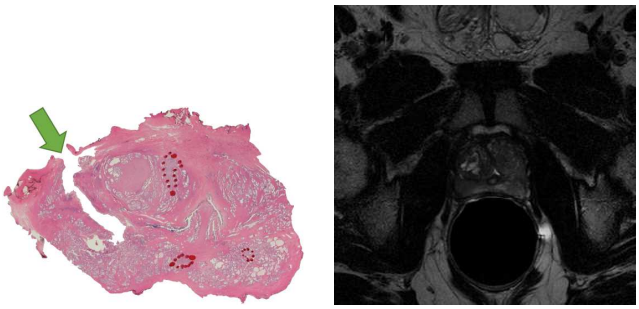
Registration of Whole Mount Histopathology to Fused Magnetic Resonance Imaging-Transrectal Ultrasound

MRI-TRUS-guided biopsy relies on manual visual assessment of the multi-parametric MRI to determine needle placement [28–31]. Identifying and localizing prostate cancer on *in vivo* imagery, such as MRI or TRUS, is a difficult task and associated with a high inter- and intra-observer variability [32]. Hence, there is a need to develop computerized decision support tools that can aid clinicians by computing quantitative measures of prostate cancer presence and, in particular, aggressive prostate cancer presence on fused MRI-TRUS imagery [25, 34–40].

To distinguish cancer from benign prostate tissue on *in vivo* imagery it is necessary to quantitatively model prostate cancer appearance using a set of training images with delineated prostate cancer spatial extent. However, due to the previously described problems identifying prostate cancer location on *in vivo* imagery, utilizing expert annotations obtained on MRI as prostate cancer ground truth results in suboptimal classifier performance [41]. Hence, relying on expert annotations obtained from MRI is highly subjective and not a reliable ground truth of prostate cancer spatial extent.

Prostate cancer spatial extent can only be definitively obtained from manual pathologist annotation on whole mount histopathology (WMH). Therefore there is a need to map the pathologist annotation onto MRI-TRUS to obtain ground truth prostate cancer spatial extent on the fused MRI-TRUS. However there are several unique challenges to registering WMH to *in vivo* radiological imagery, such as MRI or TRUS. The primary challenges to register WMH and *in vivo* radiological imagery are the following [89]:

1. Prostate deformation occurs during *in vivo* imaging due to the use of the TRUS probe, MRI endorectal coil, or changes in bladder and rectum filling.



(a) (b)

Figure 3.1: Example of corresponding (a) $2D$ WMH section and (b) and $2D$ MRI. (a) has a large tear (green arrow) as well as severe deformation to the prostate gland compared to (b) the MRI.

2. Tissue fixation and sectioning results in tissue shrinkage and distortion. Additionally, there is often tissue loss due to surgical intervention and/or tissue processing.

3. Misalignment between the planes and orientation of the WMH tissue section and the *in vivo* imagery.

Figure 3 shows a WMH section and corresponding T2-weighted MRI. This example shows the difference in prostate shape due to the 3.1(b) endorectal coil, 3.1(a) prostate shrinkage and tissue loss (green arrow) due processing of the histopathology sample.

Due to the difficulty to aligning WMH to *in vivo* imaging modalities, there has been limited work in aligning WMH to MRI [90–94] and WMH to TRUS [95, 96], several of these methods require additional *ex vivo* imagery to be acquired to guide the registration scheme. However, acquiring these additional modalities is not part of routine clinical care and, hence, is time consuming, expensive, and may require additional imaging equipment. To overcome these challenges and account for the large differences in prostate deforamtion between the two modalities we present Prostalign, an interactive registration algorithm that aligns WMH and *in vivo* MRI images directly by minimizing the distance between manually selected corresponding fiducials.

The remainder of this chapter is organized as follows. Section 3.1 provides an overview of previous work in aligning *in vivo* prostate imagery with *ex vivo* WMH. Section 3.2 describes the Prostalign algorithm used to align WMH to fused MRI-TRUS imagery. Section 3.3 provides preliminary qualitative results for Prostalign. Finally, concluding remarks on Prostalign are provided in Section 3.4.

3.1 Previous Work Aligning Whole Mount Histopathology and *In Vivo* Imagery

Due to the difficulty to aligning WMH to *in vivo* imaging modalities, there has been limited work in aligning WMH to MRI [90–94] and WMH to TRUS [95, 96]. Methods to align WMH onto MRI often leverage image similarity [93], corresponding fiducials [89, 90, 92, 94], or surface correspondence [95, 96]. To the best of our knowledge WMH has not been aligned to fused MRI-TRUS imagery.

3.1.1 Previous Work Aligning Whole Mount Histopathology and Transrectal Ultrasound

Due to the poor resolution of internal anatomical detail of the prostate on TRUS, most registration algorithms to align WMH to TRUS rely on matching the surface of the prostate [95–97]. These methods first perform a volumetric reconstruction of the $2D$ WMH images, in which individual $2D$ sections are aligned and deformation differences between $2D$ sections are accounted for. The reconstructed $3D$ WMH is then aligning to the prostate gland on $3D$ TRUS using the gland morphology to drive the registration [97]. Moskalik *et. al.* acquired *ex vivo* ultrasound and aligned it to WMH by identifying the prostate surface and urethra location and using these two anatomical structures as fiducials to register the *ex vivo* ultrasound and WMH [95]. The *ex vivo* ultrasound was then aligned to *in vivo* TRUS. Similar methods have been applied to ultrasound elastography [96]. Irregular and/or large deformations to the prostate tissue may occur when sectioning and placing onto glass slides, and these methods cannot account for such deformations if they occur internally to the surface of the prostate gland.

3.1.2 Previous Work Aligning Whole Mount Histopathology and Magnetic Resonance Imaging

MRI has a finer spatial resolution than TRUS, and is able to visualize more internal landmarks. Hence, several methods to align WMH and MRI leverage internal landmarks

Method	$2D/3D$	Description
Park *et. al.* [90]	$3D$	Use of internal and external landmarks to guide WMH registration to block face photography. Block face photography is then aligned to *ex vivo* MRI which is subsequently aligned to *in vivo* MRI.
Orczyk *et. al.* [91]	$3D$	Use of internal landmarks to guide WMH registration to block face photography. Block face photography is then aligned to *in vivo* MRI.
Gibson *et. al.* [92] Ward *et. al.* [89]	$2D$	Strand-like fiduicals are embedding in radical prostatectomy specimens. These fiducals are visible on *ex vivo* MRI and WMH and guide sectioning of the prostate tissue and registration of the *ex vivo* MRI and WMH. *Ex vivo* MRI is subsequently aligned to *in vivo* MRI.
Chappelow *et. al.* [93]	$2D$	Aligning WMH onto *in vivo* MRI leveraging image intensity and texture features.
Zhan *et. al.* [94]	$2D$	Aligning WMH onto *in vivo* MRI leveraging automatically detected landmarks.
Turkbey *et. al.* [98]	$2D$	The use of a patient specific $3D$ mold to determine slice correspondence between MRI and WMH. Corresponding $2D$ images are then registered.

Table 3.1: State-of-the-art MRI-WMH registration algorithms.

that are shared between modalities. Table 3.1 briefly describes the previous methods presented to align MRI and WMH. These methods can be divided into (1) those that first perform a volumetric reconstruction of the $2D$ WMH images and then register the $3D$ WMH to the $3D$ MRI and (2) those that first determine correspondence between $2D$ WMH and $2D$ MRI planar sections and then register $2D$ WMH onto $2D$ MRI.

Methods that first perform $3D$ WMH reconstruction often acquire intermediate *ex vivo* imaging to help aid in the volumetric reconstruction of WMH or the registration to *in vivo* imagery. In Park *et. al.* [90] WMH was aligned to block face photography of the *ex vivo* prostate tissue specimens utilizing internal and external landmarks; the block face photography was then used to guide volumetric reconstruction. The $3D$ block face photography specimen is then aligned to *ex vivo* MRI utilizing internal and external landmarks, and finally, *ex vivo* and *in vivo* MRI are registered using utilizing internal and external landmarks. Another MRI-WMH registration method aligned *ex vivo* MRI to block face photography prior to the prostate being embedding in paraffin; WMH was then aligned to the block face photography and, hence, implicitly aligned

to the MRI [91]. As noted in Park *et. al.* [90] performing several sub-registration steps may be more accurate as the prostate is more similar between each sub-step than directly registering from WMH to *in vivo* MRI. However, *ex vivo* MRI and block face photography are not routinely acquired in a clinical setting. *Ex vivo* MRI can be especially difficult to acquire as it requires access to an MRI scanner.

In Gibson *et. al.* [92] strand-like fiducials were embedded in the prostate after radical prostatectomy. The prostatectomy specimen was then images with *ex vivo* MRI and then sectioned such that the each WMH image corresponded to a single *ex vivo* MRI $2D$ planar image. The fiducials were visible on the *ex vivo* MRI and WMH and helped to guide the registration between the two modalities. Ward *et. al.* [89] extended the methodology of Gibson *et. al.* [92] to align the *ex vivo* MRI onto *in vivo* MRI, thereby allowing for mapping of WMH to *in vivo* MRI. However, manually altering the prostatectomy specimens as done in this approach is often not feasible in a clinical setting.

An automated method to align $2D$ WMH directly onto $2D$ *in vivo* MRI leveraging image intensity and texture features has been presented by Chappelow *et. al.* [93]. The method of Chappelow *et. al.* [93] is highly sensitive to the quality of the WMH imagery, large deformations or tears in the WMH have adverse effects on registration performance. An automated landmark detection scheme is presented in Zhan *et. al.* [94], where corresponding landmarks are determined primarily according to morphological similarity between the contour of the prostate on the $2D$ WMH and $2D$ MRI. Large differences in prostate deformation, that will lead to corresponding landmarks to have large differences in the local morphology, may result in inaccurate landmark correspondence. In both of these works, a necessary prerequisite was to determine which $2D$ WMH corresponded to $2D$ planar images of the MRI.

To overcome the need to manually determine slice correspondence between $2D$ MRI and $2D$ WMH, Turkbey *et. al.* [98] presented a prostate mold approach. In this methodology, a patient specific $3D$ mold was constructed by segmenting the prostate on *in vivo* MRI. The mold enabled sectioning of the WMH such that every $2D$ WMH was oriented and corresponded to a single $2D$ MRI planar image. The $3D$ mold enables a

rigorous method to determine slice correspondence, however, it requires a large amount of preparation to segment each prostate on MRI and print a $3D$ mold prior to tissue processing.

In this work we use Prostalign, an approach to determine corresponding fiducials on $2D$ WMH and $2D$ MRI planar images manually. Fiducials are then aligned using thin-plate spline [99]. Prostalign is presented to (a) obviate the need to acquire additional *ex vivo* image modalites that are not routinely done, require additional time and equipment, and (b) overcome large deformations, tissue loss, and tearing that may occur when creating WMH slides. While this method requires extensive manual intervention, Prostalign will align WMH onto MRI even when the WMH has missing tissue, large deformations, and tissue tearing. Because the fiducials are selected manually, registration accuracy is directly related to the choice and accuracy of the corresponding fiducials selected by the user. In this work, we demonstrate WMH can be aligned to fused MRI-TRUS imagery using Prostalign.

3.2 Prostalign for Mapping Whole Mount Histopathology to Fused Magnetic Resonance Imaging - Transrectal ultrasound

A $3D$ WMH volume $\mathcal{C}^h = (C^h, f^h)$ is defined by a set of $3D$ Cartesian coordinates C^h and the image intensity function $f^h(c) : c \in C^h$. The image intensity function is described by RGB values such that $f^h(c) = [f^h_R(c), f^h_G(c) f^h_B(c)]$ The WMH volume consists of N^h $2D$ WMH slices as defined by $\mathcal{C}^h = [\mathcal{C}^h_1, \ldots, \mathcal{C}^h_N]$.

A $3D$ MRI volume $\mathcal{C}^m = (C^m, f^m)$ and a $3D$ TRUS volume $\mathcal{C}^u = (C^u, f^u)$ are similarly defined. For the MRI and TRUS the corresponding intensity functions, $f^m(c)$ and $f^u(c)$ are a single gray value. The MRI volume consists of N^m $2D$ MRI slices, defined similarly to the WMH slices.

The goal of this section is to align all three image modalities. This is performed in two steps, in step 1 a transform is found that aligns the WMH volume to the MRI volume and in step 2 a transform $T^{m \to u}$ is found via MAPPER as described in Section 2.3 to align the WMH and MRI volumes to TRUS.

Step 1 WMH is pre-processed by: (a) initially aligning MRI and WMH using a visual assessment, (b) removing background information from non-histopathology regions on WMH via thresholding.

Step 2: Correspondences between MRI and WMH are identified by an expert radiologist and pathologist using distances between WMH images and major anatomical landmarks.

Step 3: Corresponding fiducials on WMH and MRI are manually selected and aligned using a thin-plate spline transform [99].

Step 4: Once all $2D$ WMH slices are aligned to the MRI, volumetric reconstruction is performed. Volumetric reconstruction takes into account the spatial location of the corresponding MRI, resulting in a $3D$ WMH volume registered onto the $3D$ MRI.

Step 5: MAPPER as described in Section 2.3 is used to register the joint $3D$ MRI/$3D$ WMH imagery to the TRUS volume. The end result is a fused volume in which each location is associated with a pixel from TRUS, MRI, and WMH.

3.3 Experimental Design and Results for Evaluation of Prostalign

3.3.1 Dataset Description

T2-weighted MRI was acquired using a General Electric (GE) 3.0 T scanner and an endorectal coil for 1 patient under IRB approval. TRUS imagery was acquired using a GE 4DE7C probe, that acquires $3D$ data in a single, multi-plane sweep of the prostate. *Ex vivo* WMH was sliced and stained with hematoxylin and eosin (H&E).

3.3.2 Experimental Results

The registration of WMH to fused MRI-TRUS was evaluated using qualitative assessment. Figure 3.3.2 provides qualitative results for two $2D$ WMH slices. As shown in Figure 3.2(c) and3.2(i) the MRI and WMH have good alignment at the boundary of the prostate gland. There is also good agreement between internal anatomical landmarks,

such as the urethra and lobes. Cancer annotations can also be consistently mapped between MRI, TRUS, and the fused MRI-TRUS.

3.4 Concluding Remarks on Prostalign

In this chapter we have demonstrated qualitatively that we can align whole mount histopathology (WMH) to fused Magnetic Resonance Imaging (MRI)- transrectal ultrasound (TRUS) imagery. This alignment enables mapping ground truth prostate cancer spatial extent onto fused MRI-TRUS imagery, thereby, allowing for the construction of computerized decision support systems for detecting prostate cancer on fused MRI-TRUS. These preliminary results demonstrate that such a fusion is feasible, however, a more robust analysis must be performed to assess the accuracy of Prostalign in greater detail. Future work will involve quantitatively evaluating our methodology to align WMH to fused MRI-TRUS on (a) a larger dataset, and (b) quantifying mis-alignment using a set of corresponding fiducials between MRI, TRUS, and WMH.

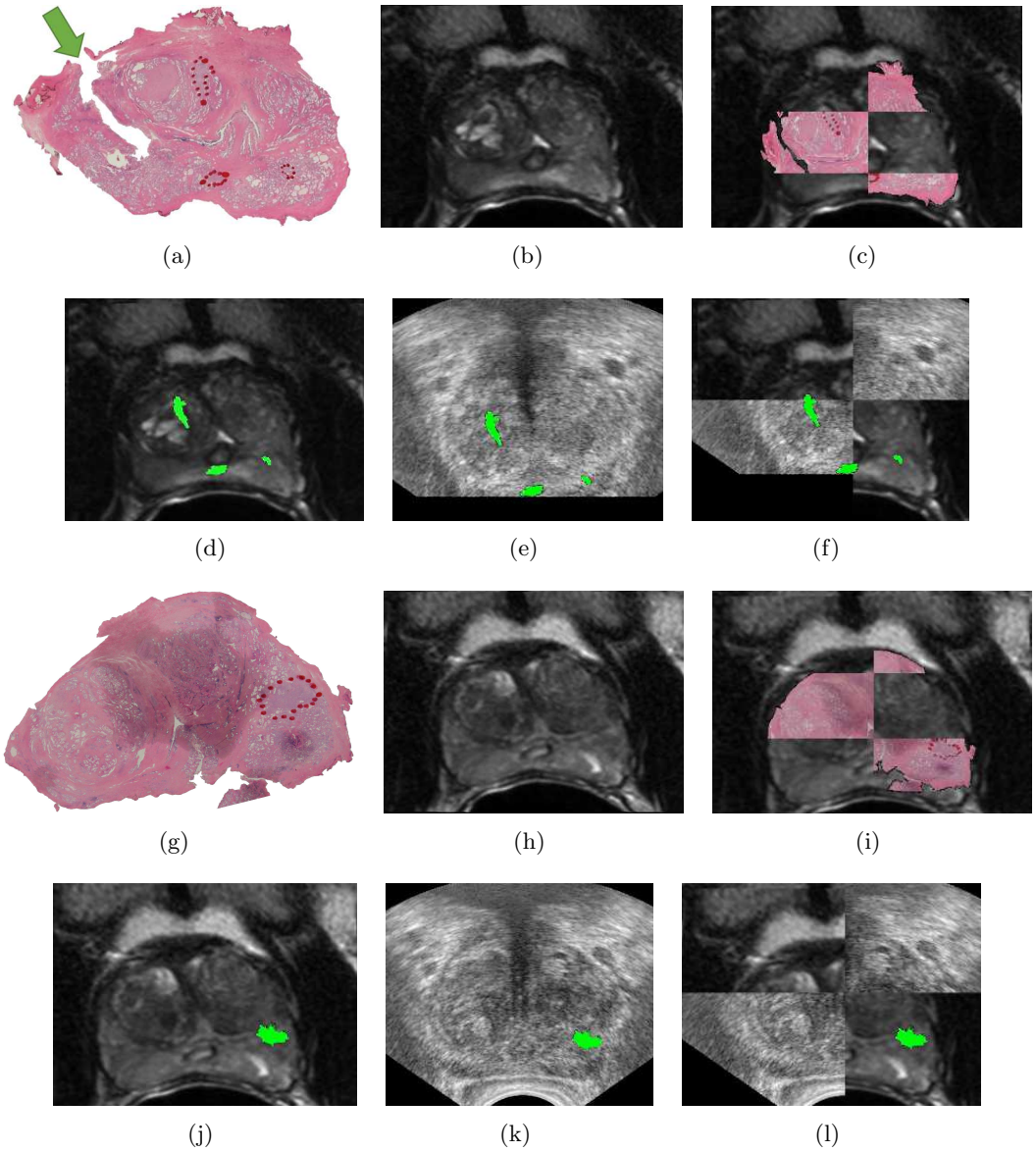


Figure 3.2: Two $2D$ planar images of (a), (g) WMH and (b), (h) corresponding MRI. (c),(i) WMH and MRI checkerboard overlays showing alignment between the two modalities. (d),(j) MRI with cancer annotation obtained from WMH (green). (e),(k) TRUS with cancer annotation obtained from WMH (green). (f),(l) Fused MRI-TRUS images shown as checkerboards with cancer annotation obtained from WMH (green).

Chapter 4

Characterizing Prostate Cancer Aggressiveness on Histopathology

In this chapter we present a novel methodology to assess prostate cancer aggressiveness on histopathology. Some of the material presented in this chapter is taken from published material of which the author of this dissertation is first author [2–4, 6, 7]. Specifically material in Section 4.2 is taken from Sparks and Madabhushi [2] ad Sparks and Madabhushi [6]; material in Section 4.3 is taken from Sparks and Madabhushi [3]; material in Section 4.4 is taken from Sparks and Madabhushi [4] and Sparks and Madabhushi [7].

Gleason grade [42] has been suggested as one of the most important predictors of prostate cancer disease aggressiveness with higher Gleason grade patterns being typically associated with more aggressive disease [44]. Pathologists typically analyze architectural features, the arrangement and morphology of glands and nuclei within the tissue, in order to determine Gleason grade (from grade 1 to grade 5) [44]. In low grade cancers, prostate tissue has a coherent spatial architecture with distinct gland lumen surrounded by cell nuclei. For higher Gleason grade patterns, gland structure begins to breakdown with gland lumen becoming indistinct and crowded with a large density and concentration of cell nuclei [44]. Correctly identifying Gleason grade patterns is critical for determining the appropriate treatment strategy for a patient with prostate cancer [100].

Distinguishing intermediate Gleason grade patterns on histopathology is a difficult task, previous studies have reported an inter-observer agreement between pathologists as low as 0.47-0.64 (reflecting low to moderate agreement) [45]. Low inter-observer agreement reflects the difficulty in distinguishing between objects with very

subtle shape differences (e.g. gland appearance between Gleason grade 3 and grade 4 patterns). The low inter-observer agreement for identifying intermediate Gleason grades [45] on prostate histopathology reflects the need for developing quantitative, reproducible computer-extracted descriptors to complement human observers in distinguishing subtle differences in intermediate Gleason grades. Recent work from our lab has suggest that quantitative histomorphometric features may be better able to predict outcome than Gleason grade alone [51–53].

In this chapter we develop novel quantitative histomorphometric features to assess prostate cancer aggressiveness, as measured by Gleason grade. In this chapter we present three novel but inter-dependent methods to characterize gland morphology on histopathoology. The methods are: (1) Explicit Shape Descriptors (ESDs) to utilize gland morphology to distinguish between intermediate Gleason grades, (2) Out-of-sample Extrapolation using Semi-supervised Learning (OSE-SSL) to improve the computational efficiency and generality of ESDs and (3) Statistical Shape Model of Manifolds (SSMM) to calculate ESDs that are robust to noise and outlier glands.

The remainder of this chapter is organized as follows. Previous work on developing quantitative histomorphometric features for prostate cancer is described in Section 4.1. ESDs to utilize gland morphology to distinguish between intermediate Gleason grades are described in Section 4.2. OSE-SSL to efficiently learn the ESDs described in Section 4.3. Finally, SSMM for ESDs robust to outlier glands are described in Section 4.4.

4.1 Previous Work in Automated Gleason Grading

Pathologists perform Gleason grading of prostate cancer tissue specimens via qualitative, visual evaluation of a tissue section previously stained with Hemotoxilyin and Eosin (H& E) [42]. The primary discriminating traits of Gleason patterns on histopathology are the difference in the arrangement and morphology of the nuclei and glands within a tissue sample [42, 44]. In devising automated pattern recognition methods for distinguishing different Gleason patterns on histopathology, the key questions to consider are (1) what is the best feature set to distinguish between Gleason patterns? and (2) what

is the best method to reduce the dimensionality of the feature set prior to classification?

Jafari *et. al.* [101] characterized tissue patch texture via wavelet features and classified Gleason patterns with an accuracy of 97% for the best performing feature. Huang *et. al.* [46] characterized tissue patch texture via Fractal Dimension and achieved an accuracy of 95%. However, a limitation of these approaches were that the image patches were manually selected to obtain regions which contained only one tissue class on the digitized slide. DiFranco *et. al.* [47] characterized tissue patch texture for each color channel independently showing 90% accuracy classifying images on a per tile. Although tiles were automatically determined, tiles which contained more than one tissue class were removed from the dataset.

Structural features (as opposed to texture features) have also been explored by some researchers for automated categorization of Gleason patterns. Veltri *et. al.* [49] and Ali *et. al.* [50] showed that the quantitative characterization of the shape of individual nuclei on tissue microarrays can distinguish between Gleason patterns with high accuracy. In a preliminary study by Veltri *et. al.* [49] characterization of manually segmented nuclei were able to distinguish between Gleason pattern 3, 4, and 5 with 73-80% accuracy. Ali *et. al.* [50] automated the nuclear segmentation and classification steps in [49], yeilded an 84% accuracy on 80 tissue microarrays. Doyle *et. al.* [48] characterized manually selected image patches according to nuclear arrangement, reporting a predictive positive value of 76.0% in distinguishing between Gleason patterns 3, 4, and 5 within a multi-classification scheme.

Tabesh *et. al.* [102] combined gland morphology, texture features, color channel variance, and nuclear arrangement to classify different Gleason patterns with 81.0% accuracy. Golugula *et. al.* [103] used proteomic data in conjunction with histology derived image features to distinguish between prostate cancer patients who following radical prostatectomy had biochemical recurrence within 5 years from patients who did not.

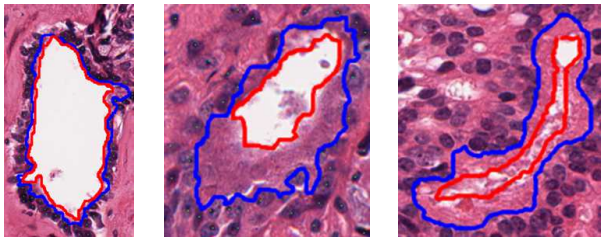
Most automated Gleason grading systems are described by a high dimensional feature space [6, 47, 48, 102, 103]. To perform accurate classification, the high dimensional feature space must be reduced to a lower dimensional space [104]. One approach to

reduce the high dimensional feature space is to perform feature selection, thereby determining a small subset of the original feature space in which accurate classification can be performed [47, 48, 102]. Difranco *et. al.* [47] utilized a random forest feature selection algorithm. Doyle *et. al.* [48] utilized a cascaded classification approach to perform feature selection for a series of pairwise classification tasks. Feature selection schemes have the advantage of selecting those features that give the most accurate classification while discarding features, which may contain noise, that have relatively poorer classification accuracy [48, 102]. However, a limitation of these approaches is that the excluded features may contain important classification information, and there removal may diminish classification accuracy in some tasks [105].

Dimensionality reduction methods learn a low dimensional embedding space which best preserves the original high dimensional feature space [6, 103, 106]. For instance Golugula *et. al.* [103] performed dimensionality reduction via supervised canonical correlation analysis to learn a low dimensional space in which patient classification was performed. Naik *et. al.* [106] demonstrated that GE is well suited for the preservation of a high dimensional feature space which characterized histological differences in texture, nuclear architecture, and gland morphology. All of these schemes have utilized the full dataset to perform ML and then trained a classifier within the low dimensional embedding space. These methods are sensitive to noise in the high dimensional feature space as well as the samples considered when learning the low dimensional space. If newly acquired samples or samples which contain noise are included in these systems they will alter the low dimensional embedding space and may detrimentally affect classification performance. Manifold regularization can alleviate this problem by constraining the manifold shape to only shapes which are most likely to occur.

4.2 Explicit Shape Descriptors (ESDs) for Quantifying Gland Morphology

The morphology of anatomical objects, defined as shape and size characteristics, on medical imagery is often an important cue to determine disease presence and/or disease



(a) (b) (c)

Figure 4.1: Three representative prostate glands on digitized needle biopsy histology specimens with lumen boundary (red) and nuclear boundary (blue) segmentations displayed. (a) A gland from benign prostate tissue, the gland has a regular oval structure. (b) A gland from a prostate cancer region identified as Gleason grade 3, the gland is smaller with greater margin irregularity compared to the benign gland. (c) A gland from a prostate cancer region identified as Gleason grade 4, the gland is highly irregular in shape with a shrunken lumen.

aggressiveness [44,102,107–113]. One application where object morphology is important is in the Gleason grading of prostate cancer which utilizes the appearance of nuclei and glands on histopathology [44, 102, 111, 114]. Figure 4.1 displays examples of prostate glands identified as (a) benign, (b) Gleason grade 3, and (c) grade 4 with lumen (red) and nuclear (blue) boundaries of the glands segmented. Morphological cues are a critical component of Gleason grading, a scheme employed to assess the invasiveness of prostate cancer [44]. Gland morphology is a particularly important feature for distinguishing between intermediate Gleason grades; higher Gleason grades of prostate cancer are characterized by irregularly shaped glands while lower Gleason grades of prostate cancer have smooth margins with a distinct lumen.

4.2.1 Previous Work in Quantifying Object Morphology

Several boundary-based shape descriptors have been previously presented citeRangayyan2007,Tabesh2007,Yang2009,Georgiou2007 to extract specific characteristics from an object margin, determined to be important for a specific task. However, these descriptors typically quantify a single, specific shape characteristic. Boundary-based shape descriptors include fractal dimension citeRangayyan2007, a measure of the self similarity between the object and its parts;

measures of contour variation including symmetry
citeYang2009; and wavelet parameters of the object boundary [115]. Additionally, other measures have included how close to circular an object is [116] or how quickly the contour varies [117]. Such descriptors provide a single global measure of object morphology, and hence may be unable to distinguish between objects with subtle, local shape differences. However, determining the boundary-based shape descriptors that can best distinguish between the prostate glands requires *a priori* information about the domain and classification task.

Alternatively, descriptors that explicitly model object shape, and are hence able to recapitulate the original shape of an object, provide an alternative approach to assessing differences in object shape [118–130]. These descriptors may be used in conjunction with an appropriate similarity metric to quantify differences between shape model representations. Such descriptors typically do not require *a priori* information about the domain and classification task to accurately determine subtle morphologic differences between objects.

Development of model-based descriptors is an active area of research and many approaches to modeling the shape of an object have been presented [118–130]. Point Distribution Models (PDM) describe shape as a collection of points on the surface of an object such as in Active Shape Models (ASMs) [122] or Shape Context [119]. Moment descriptors such as geometric moments [118] or Zernike Moments [130] describe the distribution of pixels contained within the object and hence provide a global measure of object morphology that is robust to subtle, local changes in the contour of an object. Fourier Descriptors (FDs) [123, 125] or Curvature Scale Space [124, 127] describe the shape of the object according to the frequency components contained in the contour and are hence sensitive to subtle changes in the contour of an object; Spherical Harmonics (SPHARM) [121] describe the spherical basis functions contained in the surface of an object. The Laplace-Beltrami shape descriptors
citeReuter2006 and variants, such as Heat Kernels [129], describe the components of some generalized function contained within the object, for instance the Laplace-Beltrami descriptor is defined as the eigenvalues of the Laplacian operator for an object.

Additionally, shape models may extract a graph representation of the object, such as the medial axis shape model (MASM) [120] or Reeb graphs [128], these methods represent the shape of an object using local symmetry to determine a skeletal representation of the object. The MASM describes object morphology as a medial axis, points within an object that are equidistant from two or more locations on the surface of an object, and hence capture the local symmetry of an object [120].

Despite the wide variety of shape modeling approaches available, relatively few methods have been applied to medical imaging applications. PDMs have been proposed in the context of segmentation for cardiac, prostate, and other organs on medical imagery [122]. PDMs have been applied to anatomical structures such as brain hippocampii in conjunction with dimensionality reduction (DR) methods to identify meaningful, local changes to the shape [131, 132]. However, obtaining point correspondence is a difficult task for glands on prostate histopathology, as meaningful substructures are difficult to identify. Heat Kernels have been utilized in conjunction with hippocampal morphology to accurately distinguish between normal patients and those with Alzheimer's disease utilizing [113]. SPHARMs have been also employed to identify Alzheimer's disease utilizing hippocampal morphology [109]. Similar methods have utilized SPHARMs to analyze the atrophy of other brain structures such as the putamen in patients diagnosed with Parkinson's disease [112]. However no method to directly apply either SPHARMs or Heat Kernels to $2D$ shapes as been presented, although a method has been presented to map a $2D$ object into a $3D$ space and then compute SPHARMs [133]. FDs have been applied to the analysis of breast lesions on mammography [115]. Additionally, recent work has shown applications of FDs to distinguishing between normal and abnormal rotator cuffs on MRI [134].

The MASM, also referred to as m-rep, has been applied to medical imaging tasks including segmentation [135, 136], registration [135], and calculation of shape statistics over a population of objects [137]. Recently, the m-rep framework has been used to guide disease classification by exploiting morphologic differences between brain substructures in order to distinguish autistic from normal patients [110]. MASMs have also been applied to the detection of aneurysms [138] as well as the detection of coronary artery

stenosis [139]. The MASM is able to represent a wide range of object morphologies over multiple dimensions and is able to detect and represent subtle differences between objects. The MASM is the shape descriptor we chose to utilize in this work.

Due to the wide application of the MASM to many object recognition and classification tasks, several approaches to efficiently calculate the MASM have been developed [140–145]. Blum presented the idea of a "grassfire" approach to calculating the MASM, the idea being if a fire was set to the boundary of the object the medial axis would be where the propagating flames meet. [141] applied a similar technique to extract the MASM for $3D$ objects. Similarly, [140] presented the idea of shock graphs, calculated by considering the gradient of the level set function for an object. Alternative approaches to calculate the MASM involve iteratively thinning the surface of the object [142]. Some approaches have incorporated pruning, removal of extraneous regions of the axis from the MASM. For instance Dynamic Contour Evolution (DCE), involves first calculating the MASM and then pruning regions on the MASM that lead to small partitions of the contour [143]. Pruning of the MASM can be performed by removing branches which have similar locations and directions as nearby branches [144]. A groupwise medial axis transform was presented in [145] to retain those branches that are consistent across a set of objects while removing those branches that are not preserved across the group.

Calculating similarity between MASMs can be difficult as there is not always a clear correspondence between regions on medial axes. Several approaches have been presented to overcome this problem [140, 144–150]. For instance comparing MASMs may be formalized as a graph matching problem, where the MASM is broken into a set of medial axis branches; shape similarity is calculated as the summation of similarity between matched branches [140]. This method can be extended to utilize additional attributes about each branch, such as branch curvature, when calculating MASM similarity [144,150]. The graph matching problem between MASMs has also been previously formulated for $3D$ structures, where the matching is between hypergraphs which may include $2D$ medial surfaces [146]. Approaches have also been presented to hierarchically match and merge regions to quantify MASM similarity [149]. [147] introduced

the concept of edit-distance, where MASM similarity is calculated as a combination of three possible edits: branch matching, branch removal, and branch addition; where each edit is assigned a similarity value. [148] considered the similarity of paths between end nodes (locations on the MASM that represent where a branch end) to determine similarity. Alternatively, a method to fit a MASM to an object using only likely, predetermined MASM templates have been presented [136]; MASM similarity is calculated as a difference of distance transforms [145].

Researchers have recently begun using nonlinear DR (NLDR) schemes in conjunction with shape descriptors [151–153]. Such methods extract a small set of features which describe the variation in morphology between different objects. For instance, [151] presented a semi-supervised framework, Graph Transduction, to learn a set of discriminating shape descriptors for content-based image retrieval (CBIR) applications. However, the set of shape descriptors learned is dependent on the query object, suggesting that it may not be naturally extensible to classification problems. A k-nearest neighbor approach to finding object similarity in the high dimensional shape space was presented in [152]. [153] presented a NLDR scheme to determine relevant morphologic differences between vertebrae, exploiting the definition of the Procrustes shape space. However, their methodology is only applicable to objects represented by PDMs. [154] utilized a geodesic distance function to distinguish between contours for pairs of objects. The use of NLDR schemes typically improves the ability to measure similarity in object morphology compared to the Euclidean distance in CBIR applications [151–153].

In this section, we present a method to quantify differences in object morphology, where only the contour of the object needs to be defined. We extract a set of Explicit Shape Descriptors (ESDs) to quantitatively represent object morphology. Our framework to extract ESDs involves: (a) utilizing the MASM to quantify object morphology, and (b) obtaining a small set of ESDs via the unsupervised NLDR scheme Graph Embedding [155].

4.2.2 Methodology for Explicit Shape Descriptors (ESDs)

Figure 4.2 presents an overview of our ESD framework and the constituent modules. ESDs are calculated by: (a) representing the shape of each gland using a medial axis shape model (MASM), (b) registering MASMs using a novel diffeomorphic based similarity (DBS) measure, (c) determining parameter correspondence between registered MASMs, (d) extracting a low dimensional representation of morphologic features utilizing the non-linear dimensionality reduction scheme Graph Embedding, (e) classifying the morphologic features using a Support Vector Machine (SVM).

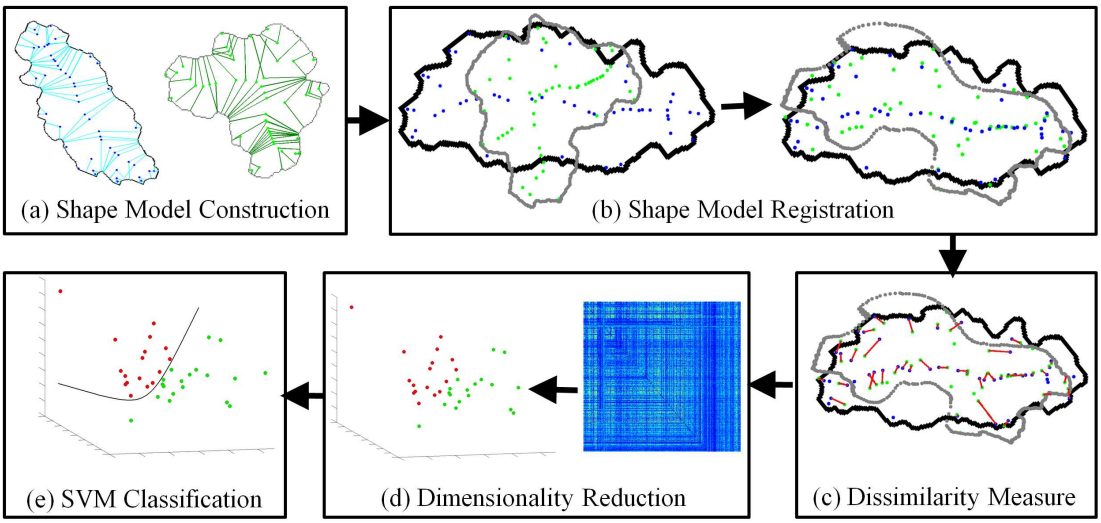


Figure 4.2: An illustration of the main modules for extracting explicit shape descriptors (ESDs). (a) A medial axis shape model (MASM) (blue, green) is fit to each object contour (black, gray). (b) Pairwise registration between MASMs is performed to align medial axes which then aids in (c) determining parameter correspondence between registered MASMs. Subsequently, pairwise differences between object shapes are computed which yields a $N \times N$ affinity matrix. (d) A nonlinear dimensionality reduction scheme, Graph Embedding, is then applied yielding a set of ESDs which quantify shape differences. Finally, (e) a Support Vector Machine (SVM) is trained to learn the optimal hyperplane which separates the ESD feature space into different object classes.

Notation

A histpathology image $\mathcal{C}^h = (C^h, \xi)$ is defined by the d-dimensional grid of voxel locations C^h and an object contour ξ. $c \in C^h$ represents a voxel defined by a d-dimensional

Symbol	Description	Symbol	Description
$\mathcal{C}^h$	d-dimensional histology image scene.	T^a	Affine transformation function.
C^h	d-dimensional grid of voxels.	ϖ_j^k	k^{th} cluster centroid at iteration j.
ξ	Contour for an object of interest.	$P(q\|\varpi_j^k)$	Probability of $q \in Q$ belonging to ϖ_j^k.
$\mathcal{Q}$	Medial axis shape model (MASM).	T^q	Diffeomorphic transformation function.
Q	Set of voxels on a medial axis.	G	Green's function.
$v_1(q), v_2(q)$	Surface vectors for $m \in M$.	$(\breve{q}_a, \breve{q}_b)$	Correspondence between two medial axes.
$f^{sdf}(c)$	signed distance function for $c \in C^h$.	A	$N \times N$ dissimilarity matrix.
X_i	i^{th} direction.	$\mathbf{y}$	Set of n Explicit Shape Descriptors (ESDs).
N	Number of objects.	$\mathcal{E}$	Super quadratic ellipsoid.

Table 4.1: Description of commonly employed notation and symbols relating to Explicit Shape Descriptor(ESD) calculation.

vector that describes its location in C^h. ξ partitions C^h into two regions $\Omega_i^h : i \in 1, 2$, where $i = 0$ denotes background and $i = 1$ denotes foreground such that $\Omega_0^h \cup \Omega_1^h = C^h$. Table 4.1 describes the notation and symbols that appear frequently to describe ESD calculation.

Medial axis shape model construction

The MASM was employed to concisely and explicitly describe the local morphology of an object described by the contour ξ [120]. We define the MASM as $\mathcal{Q} = (q, v_1, v_2)$ where $Q \in C^h$ is a set of voxels on the medial axis and $v_1(q)$, $v_2(q) : q \in Q$ are two functions defining vectors to the first and second closest points on ξ. To find Q, we calculate the gradient magnitude squared of the signed distance function defined as,

$$\hat{f}^{sdf}(c) = \sum_{i=1}^{d} \left(\frac{\partial f^{sdf}(c)}{\partial X_i} \right)^2, \tag{4.1}$$

where $f^{sdf}(c)$ is the signed distance function evaluated over $c \in C^h$ and $\dfrac{\partial(\cdot)}{\partial X_i}$ is the partial gradient along X_i corresponding to the ith direction. For a d-dimensional image

X_i is found for all $i \in \{1, \ldots, d\}$. We use $\hat{f}^{sdf}(c)$ to define the medial atoms as $Q = \{q : q \in C^h, \hat{f}^{sdf}(q) < \tau\}$. Empirically, we determined that $\tau = 0.05 \underset{c \in C^h}{\operatorname{argmax}} \left(\hat{f}^{sdf}(c)\right)$ yields a well defined medial axis. Section 4.2.3 describes the approach we employed to assess whether a MASM was able to accurately and quantitatively describe the morphology of a given object. The use of τ helps avoid spurious branches on the MASM.

The surface vector functions $v_1(q)$ and $v_2(q) : q \in Q$, are calculated as $v_1(q) = \hat{p}_1(q) - q$ and $v_2(q) = \hat{p}_2(q) - q$, where $\hat{p}_1(q)$ and $\hat{p}_2(q)$ represent the two closest points on ξ to q. For q with more than two closest points on the ξ, we chose $\hat{p}_1$ and $\hat{p}_2$ that maximize the angle between $v_1(q)$ and $v_2(q)$.

Framework for quantifying shape differences

Dissimilarity between a set of N MASMs $\mathbf{Q} = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_N\}$ is quantified by comparing differences in corresponding medial atoms for each pair of MASMs. The parameters of a MASM, $\mathcal{Q}_a : a \in \{1, \ldots, N\}$, include Q_a, $v_{1,a}$, and $v_{2,a}$ (defined in Section 4.2.2). We determine correspondence between all $q_a \in Q_a$ and all $q_b \in Q_b : b \neq a, b \in \{1, \ldots, N\}$. By this procedure we attempted to implicitly determine correspondence between the surface vector functions $v_{1,a}(q_a)$ and $v_{1,b}(q_b)$ as well as $v_{2,a}(q_a)$ and $v_{2,b}(q_b)$. To determine correspondence a two step registration was performed by: (1) affine registration of Q_a onto Q_b, followed by (2) a diffeomorphic registration of Q_a onto Q_b. These steps are described in more detail in Sections 4.2.2 and 4.2.2, respectively. The affine registration provides a rough alignment of corresponding regions on the MASMs and is necessary for the initialization of the diffeomorphic registration process. Once Q_a and Q_b have been accurately registered, medial atom correspondence between Q_a and Q_b is determined and used to calculate the dissimilarity between $\mathcal{Q}_a$ and $\mathcal{Q}_b$. Section 4.2.2 describes the calculation of MASM dissimilarity.

Medial atom affine registration

Affine registration between Q_a and Q_b is determined by applying the Iterative Closest Point (ICP) algorithm [130]. Point correspondence between Q_a and Q_b is determined by,

$$(\hat{q}_a, \hat{q}_b) = \underset{\hat{q}_a \in Q_a, \hat{q}_b \in Q_b}{\operatorname{argmin}} ||q_a - q_b||. \tag{4.2}$$

where $\hat{q}_a \in Q_a$ and $\hat{q}_b \in Q_b$ are the set of corresponding points on Q_a and Q_b. An affine transformation T^a is found by minimizing the following function,

$$T^a = \underset{T^a}{\operatorname{argmin}} \left(\sum_{\hat{q}_a, \hat{q}_b} ||\hat{q}_b - T^a(\hat{q}_a)|| \right). \tag{4.3}$$

Estimation of point correspondences (Eq. 4.2) and the affine transformation (Eq. 4.3) are iteratively applied to Q_a until point correspondences, $(\hat{q}_a, \hat{q}_b)$, remain unchanged between iterations. The resulting medial axis $\hat{Q}_a = T^a(Q_a)$ is affinely registered to Q_b.

Medial atom-based non-rigid registration

For two sets of medial atoms $\hat{Q}_a$ and Q_b, which are registered as described in Section 4.2.2, a diffeomorphic registration is then applied to further align $\hat{Q}_a$ and Q_b. We utilized a variation of the diffeomorphic registration method proposed by [156]. A brief overview of our non-rigid registration is as follows: (1) corresponding locations between $\hat{Q}_a$ and Q_b are estimated using a deterministic annealing K-means clustering algorithm [157] (described in greater detail in Section 4.2.2); (2) A diffeomorphic transformation over the image space C^h is calculated to minimize the distance between the estimated corresponding locations on $\hat{Q}_a$ and Q_b. A diffeomorphic transformation was used to ensure a continuous and differentiable transformation field so that the underlying relationship between all $\hat{q}_a \in \hat{Q}_a$ and all $q_b \in Q_b$ are preserved. In Section 4.2.2 we discuss the calculation of the diffeomorphic transformation in greater detail.

Correspondence estimation Individual medial atom correspondence between $\hat{Q}_a$ and Q_b may be difficult to determine accurately. We therefore determine corresponding locations on $\hat{Q}_a$ and Q_b via a deterministic annealing K-means clustering algorithm [157]. We define a set of K cluster centroids at the jth iteration of our registration method as $\varpi^k_{j,a} : k \in \{1, \ldots, K\}$ for all $q_a \in \hat{Q}_a$. Similarly, for Q_b we define a set of cluster centroids $\varpi^k_{j,b} : k \in \{1, \ldots, K\}$. The cluster centroids are initialized such that $\varpi^k_{0,a}$ and $\varpi^k_{0,b}$ represent corresponding points on Q_a and Q_b. As the clusters are

used to determine correspondence between MASMs, K is constrained to be the same for both $\hat{Q}_a$ and Q_b.

We estimate the probability $P(q_a|\varpi^k_{j,a}, \sigma_j)$ of a medial atom $\hat{q}_a \in \hat{Q}_a$ belonging to the $\varpi^k_{j,a}$ cluster as:

$$P(q_a|\varpi^k_{j,a}, \sigma_j) = \frac{e^{-\sigma_j||q_a - \varpi^k_{j,a}||^2}}{\sum_{k=1}^{K} e^{-\sigma_j||q_a - \varpi^k_{j,a}||^2}}. \tag{4.4}$$

Similarly, the probability of $q_b \in Q_b$ belonging to the cluster $\varpi^k_{j,b}$ is given as $P(q_b|\varpi^k_{j,b}, \sigma_j)$. The term $e^{-\sigma_j||q_a - \varpi^k_{j,a}||^2}$ assigns higher values to medial atoms near the centroid $\varpi^k_{j,a}$ and lower values to medial atoms farther away; σ_j determines which medial atoms are considered near and far from the centroid. Convergence of the membership function is enforced by setting $\sigma_j = (\zeta)^j \sigma_0$ where $\zeta > 1$, hence at each iteration the clustering algorithm considers a smaller region to be near the cluster centroid. Therefore at each step in the algorithm fewer medial atoms have a non-zero probability of belonging to the cluster defined by the centroid $\varpi^k_{j,a}$. Ultimately when j is very large, each medial atom is assigned membership (a non-zero probability) to one cluster centroid. The initial weighting term is set as $\frac{1}{\sigma_0} = \max_{q_a \in \hat{Q}_a} ||q_a - \mu_a|| + \max_{q_b \in Q_b} ||q_b - \mu_b||$.

Cluster centroids are updated according to the probability of all $q_a \in \hat{Q}_a$ belonging to the cluster $\varpi^k_{j,a}$. The cluster cluster $\varpi^k_{j,a}$ is updated by the equation,

$$\varpi^k_{j+1,a} = \frac{\sum_{q_a \in \hat{Q}_a} q_a P(q_a|\varpi^k_{j,a}, \sigma_j) + \varpi^k_{j,b}}{1 + \sum_{q_a \in Q_a} P(q_a|\varpi^k_{j,a}, \sigma_j)}. \tag{4.5}$$

The term $\varpi^k_{j,b}$ defines a centroid on Q_b that corresponds to the centroid $\varpi^k_{j,a}$. By taking the average of the two locations $\varpi^k_{j,a}$ and $\varpi^k_{j,b}$ in Equation 4.5, we ensure that the $\varpi^k_{j+1,a}$ remains in a location on Q_a which is proximal to $\varpi^k_{j,b}$ located on Q_b. The centroid $\varpi^k_{j+1,b}$ is determined by a similar equation.

Correspondence registration The goal of correspondence registration is to find a diffeomorphic transformation T^q which best maps the cluster centroids $\varpi^k_{j,a}$ onto $\varpi^k_{j,b}$ for $k \in \{1, \ldots, K\}$. T^q is defined as $\{T^q(t) : t \in \{0, \ldots, t_{\max}\}\}$ where $T^{q,k}(0) = g_a^{k,j}$ and $T^{q,k}(t_{\max}) = \varpi^k_{j,b}$. Hence T^q will enable alignment of $\varpi^k_{j,a}$ to $\varpi^k_{j,b}$. Similar to Twining *et. al.* [158], we use a linear piecewise approximation to solve the energy minimization function,

$$\hat{T^q} = \underset{\hat{T^q}}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{t=0}^{t_{\max}} \omega(t) \cdot \left(\sum_{\eta=1}^{K} \omega^{\eta}(t) G(T^{q,\eta}(t), T^{q,k}(t)) \right), \tag{4.6}$$

where the kernel function G is defined as Green's function: $G(\alpha, \beta) = -(\alpha - \beta)^2 \log(\alpha - \beta)^2$ [99]. Green's function ensures that T^q will be smoothly varying over C^h.

To solve Equation 4.6, an optimization of T^q and the variables $\omega(t)$ and $\omega^{\eta}(t)$ can be found in an iterative fashion [158]. This optimization is performed by first holding $\omega(t)$ and $\omega^{\eta}(t)$ constant and using gradient descent to find the optimal T^q and then repeating the procedure with T^q held constant.

Both the correspondence estimation (Equations 4.4 and 4.5) and correspondence registration (Equation 4.6) are iterated until a user defined threshold, Υ, is reached by the annealing parameter σ_j. The overall algorithm is detailed in Algorithm 1.

Algorithm 1 RegisterMedialAxisShapeModels

Input: MASM $\hat{Q}_a$ and Q_b **Output:** Regsitered MASM $\tilde{Q}_a$

1: Initialize σ_j, $\varpi^k_{j,a}$, $\varpi^k_{j,b}$
2: **while** $\sigma_j < \Upsilon$ **do**
3: Update $P(q_a|\varpi^k_{j,a})$, $P(q_b|\varpi^k_{j,b})$ by Equation 4.4
4: Update $\varpi^k_{j,a}$, $\varpi^k_{j,b}$ by Equation 4.5
5: Update T^q by Equation 4.6
6: $\tilde{Q}_a = T^q(\hat{Q}_a)$
7: **end while**

The two sets of cluster centroids are initialized to be equal $\varpi^k_{0,a} = \varpi^k_{0,b}$ and located at $\varpi^k_{0,a} = \dfrac{\mu_a + \mu_b}{2} + \epsilon$, where ϵ is a random variable with a very small value ($\epsilon \approx 10^{-1}$). The term ϵ is added to ensure that each pair of corresponding cluster centroid $\varpi^k_{0,a}, \varpi^k_{0,b}$ has a unique starting location, and hence will diverge from the other cluster pairs as σ_j increases.

Medial atom correspondence and shape dissimilarity

Given two medial axes $\tilde{Q}_a$ and Q_b registered into a common coordinate frame, we determine point correspondence between $\tilde{Q}_a$ and Q_b as,

$$(\breve{q}_a, \breve{q}_b) = \underset{\breve{q}_a \in \tilde{Q}_a, \breve{q}_a \in Q_b}{\operatorname{argmin}} ||\breve{q}_a - \breve{q}_b||. \tag{4.7}$$

The set of corresponding medial atoms, $(\breve{q}_a, \breve{q}_b)$ determined via Equation 4.7 are then used to calculate dissimilarity between $\tilde{Q}_a$ and Q_b as,

$$A_{ab} = \sum_{\breve{q}_a, \breve{q}_b} \left(||\breve{q}_a - \breve{q}_b|| + ||v_{1,a}(\breve{q}_a) - v_{1,b}(\breve{q}_b)|| + ||v_{2,a}(\breve{q}_a) - v_{2,b}(\breve{q}_b)|| \right). \tag{4.8}$$

For all $\mathcal{Q} \in \mathbf{Q}$, a dissimilarity matrix $A \in \mathbb{R}^{N \times N}$ is constructed such that it represents a N-dimensional space corresponding to morphologic dissimilarity between all $\mathcal{Q} \in \mathbf{Q}$.

Feature extraction via nonlinear dimensionality reduction

GE [155] is applied to the dissimilarity matrix A, which measures the dissimilarity between all all $\mathcal{Q} \in \mathbf{Q}$, to yield a set of ESD features in a low dimensional space. Specifically, a n-dimensional embedding is learned from the N-dimensional matrix A, where $n << N$. The ESD features for all $\mathcal{Q} \in \mathbf{Q}$ are defined as $\mathbf{y} = [y_1, \ldots, y_N]$ where $\mathbf{y}$ represents the top n eigenvectors for the shape space defined by A. $\mathbf{y}$ can be found by minimizing the pairwise reconstruction error measured as,

$$\mathbf{y} = \underset{\mathbf{y}}{\operatorname{argmin}} \sum_{a=1}^{N} \sum_{b=1}^{N} ||y_a - y_b||^2 W_{ab}, \tag{4.9}$$

where $W_{ab} = e^{-A_{ab}/\gamma}$. The term γ is used to normalize A and is dataset specific. The value of γ was determined empirically as described in Section 4.2.3. Equation 4.9 can be rewritten as,

$$\underset{\mathbf{y}}{\operatorname{argmin}} \sum_{a=1}^{N} \sum_{b=1}^{N} (y_a^2 + y_b^2 - 2y_a y_b) W_{ab}. \tag{4.10}$$

Assuming $W_{ab} = W_{ba}$, which will be true since $A_{ab} = A_{ba}$, then Equation 4.10 reduces to,

$$\underset{\mathbf{y}}{\operatorname{argmin}}\ 2 \sum_{a=1}^{N} \sum_{b=1}^{N} (y_a^2 - y_a y_b) W_{ab}. \tag{4.11}$$

The minimization problem can be further simplified by introducing a diagonal matrix defined as $D_{aa} = \sum_b W_{ab}$, making the minimization problem,

$$\underset{\mathbf{y}}{\operatorname{argmin}}\ 2\mathbf{y}(D - W)\mathbf{y}. \tag{4.12}$$

Equation 4.12 is equivalent to the minimum eigenvalue decomposition equation,

$$(D - W)\mathbf{y} = \lambda D \mathbf{y}, \tag{4.13}$$

where the top n eigenvalues in λ correspond to the n eigenvectors $\mathbf{y}$ and the top n ESD features. The top n eigenvalues correspond to the projection of the matrix A into the space $\mathbb{R}^{N\times n}$ such that the pairwise distances between the elements in A, and hence the pairwise distances between objects, are preserved. Furthermore as the eigenvectors $\mathbf{y}$ are orthonormal to each other, each additional feature provides independent information on the shape space represented by A.

Support Vector Machine classification

A SVM classifier [159] can be trained using $\mathbf{y}$, to learn the optimal hyperplane which separates $\mathbf{Q}$ into the classes referenced by the label set $\mathbf{L} = [l_1, \ldots, l_N]$.

The SVM classifier utilizes y_a to determine the distance to the hyperplane $\Phi(\mathcal{Q}_a)$. SVM classifiers are typically used to generate a hard class decision where $\Phi(\mathcal{Q}_a) < 0$ corresponds to assigning a class label of -1 to $\mathcal{Q}_a$. However, a pseudo-threshold can be generated by varying the decision boundary.

Given a specific decision boundary, ϱ, if $\Phi(\mathcal{Q}_a) < \varrho$ and $l_a = -1$ then $\mathcal{Q}_a$ is identified as a true negative (TN); if $\Phi(\mathcal{Q}_a) < \varrho$ and $l_a = 1$ then $\mathcal{Q}_a$ is identified as a false negative (FN); if $\Phi(\mathcal{Q}_a) > \varrho$ and $l_a = -1$ then $\mathcal{Q}_a$ is identified as a false positive (FP); and if $\Phi(\mathcal{Q}_a) > \varrho$ and $l_a = 1$ then $\mathcal{Q}_a$ is identified as a true positive (TP). The number of TN ($N_{TN,\varrho}$), FN ($N_{FP,\varrho}$), FP ($N_{FP,\varrho}$),and TP ($N_{TP,\varrho}$) are calculated over a range of ϱ.

For each ϱ, sensitivity (SN_ϱ), specificity (SP_ϱ), and classifier accuracy (CA_ϱ) can be calculated as,

$$SN_\varrho = \frac{N_{TP,\varrho}}{N_{TP,\varrho} + N_{FN,\varrho}}, SP_\varrho = \frac{N_{TN,\varrho}}{N_{TN,\varrho} + N_{FP,\varrho}}, \text{ and } CA_\varrho = \frac{N_{TP,\varrho} + N_{TN,\varrho}}{N}$$

where N is the total number of objects in the database. By plotting SN_ϱ versus $1-SP_\varrho$ over a range of ϱ a Receiver Operating Characteristic (ROC) curve representing the trade-off between SN and SP for a give feature set is obtained. Area under the ROC curve (AUC) is calculated for each ROC curve.

For our experiments, the SVM classifier was employed with a radial basis function. For the dataset evaluated for our experiments there were three prostate histopathology

classes. To evaluate the prostate histopathology dataset using a SVM the classification task was divided into 4 pairwise classification tasks. The training and evaluation involved a randomized 3−fold cross validation scheme where at each iteration, 2/3 of the dataset was used for training of the classifier, while always maintaining class balance. The remaining 1/3 of the dataset was used for independent testing of the SVM classifier. Training and testing sets were selected such that the training and testing sets never concurrently contained images from the same patient. The cross validation procedure was repeated 5 times where, at each iteration, the training and testing sets were selected randomly always ensuring that there was no overlap between the training and testing sets in terms of patients.

4.2.3 Experimental Design and Results for Evaluation of Explicit Shape Descriptors

Dataset description

Synthetic super quadratic ellipsoids

Super quadratic ellipsoids represent a class of $3D$ objects with a closed topology such that the shape of a super quadratic ellipsoid is fully determined by the 5 parameters: α_1, α_2, α_3, ϵ_1, and ϵ_2 [160]. Super quadratic ellipsoids with similar shapes were generated by carefully modulating the model parameters so that the discriminability of the ESD features in a synthetic setting could be evaluated. The boundary of a super quadratic ellipsoid is defined as,

$$\mathcal{E}(x_1, x_2, x_3) = \begin{cases} x_1 = \alpha_1 \cos(\theta) \cos(\psi)^{\epsilon_1} \\ x_2 = \alpha_2 \sin(\theta) \cos(\psi)^{\epsilon_2} \\ x_3 = \alpha_3 \cos(\theta) \sin(\psi), \end{cases} \tag{4.14}$$

where $\theta \in \{-\pi, \ldots, \pi\}$ and $\psi \in \{-\pi/2, \ldots, \pi/2\}$ [160]. For the purpose of this study the following parameters were combined: $\epsilon_1, \epsilon_2 \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$ and $\alpha_1, \alpha_2, \alpha_3 \in \{0.7, 0.8, 0.9, 1\}$ resulting in 4096 possible shape combinations. Note however, that several shapes will be scaled versions of each other (e.g. $\epsilon_1 = \epsilon_2 = 0.5$,

$\alpha_1 = \alpha_2 = \alpha_3 = 0.7$ and $\epsilon_1 = \epsilon_2 = 0.5$, $\alpha_1 = \alpha_2 = \alpha_3 = 0.8$ will be the same shape with only a scale difference). Two parameters, ϵ_1 and ϵ_2, are referred to as shape parameters and were used to control the concavity/convexity of the super quadratic ellipsoids. The other three parameters, α_1, α_2, and α_3, are scaling parameters and determine the length, width, and depth of the object respectively. By selecting parameters that are close together, we obtain a set of objects with subtle shape differences.

Prostate histopathology

Prostate tissue biopsy cores obtained from 58 patient studies were stained with Hemotoxylin and Eosin (H & E) and digitized using a ScanScope CS^{TM} whole-slide scanning system at 40× optical magnification. An expert pathologist selected regions of interests on the digitized biopsy image, to obtain a total of 102 regions. The expert pathologist then classified each region as benign (24 regions), Gleason grade 3 (67 regions), or Gleason grade 4 (11 regions). Every gland contained within each region was segmented by a human expert to obtain lumen and nuclear boundaries. Glands which did not contain either a nuclear or lumen boundary, or where the contour was not fully contained within the region were removed from the study, resulting in a total of 888 glands containing both lumen and nuclear boundary segmentations. These glands were distributed across the three classes: benign ($N = 93$), Gleason grade 3 ($N = 748$), and Gleason grade 4 ($N = 47$).

Features for comparison against Explicit Shape Descriptors

Our novel ESD features are compared against three morphologic feature sets: Boundary-based features (referred to as Boundary) [161], Fourier Descriptors [125], and a MASM path similarity measure (referred to as Path) [148]. Below, we briefly describe the calculation of each of these feature sets.

Boundary-based features

The Boundary feature set consists of 6 morphologic features that have been previously used with computerized decisions support systems for determining Gleason grade using

prostate gland morphology [102]. The formulation for each of the shape features is presented in Table 4.2 and reflects the (a) circularity of an object (area overlap ratio, compactness), (b) how much does object contour vary with respect to the shape of a circle (normalized average radial distance ratio, standard deviation of distance ratio, variance of distance ratio), and (c) how quickly does the object contour change (smoothness).

Boundary Feature	Description
Normalized Average Radial Distance Ratio	$\frac{\frac{1}{\lvert\xi\rvert}\sum_{p\in\xi}\lvert\lvert p-\bar{p}\rvert\rvert}{\max_{p\in\xi}\lvert\lvert p-\bar{p}\rvert\rvert}$ where $\bar{p}=\frac{1}{\lvert\xi\rvert}\sum_{p\in\xi}p$
Area Overlap Ratio	$\frac{\lvert\xi\rvert}{\pi r^2}$ where $r=\max_{p\in\xi}\lvert\lvert p-\bar{p}\rvert\rvert$
Standard Deviation of Distance Ratio	$\sigma_\Gamma=\sqrt{\Gamma(p)-\mu_\Gamma^2}$ where $\Gamma(p)=\frac{\lvert\lvert p-\bar{p}\rvert\rvert}{\max_{p\in\xi}\lvert\lvert p-\bar{p}\rvert\rvert}$ and $\mu_\Gamma=\frac{1}{\xi}\sum_{p\in\xi}\Gamma(p)$
Variance of Distance Ratio	σ_Γ^2
Compactness	$\frac{F(\xi)^2}{\lvert\xi\rvert}$ where $F(\xi)=\sum_{p\in\xi,j\in\{1,\ldots,J\}}\lvert\lvert p^{j+1}-p^j\rvert\rvert$
Smoothness	$\sum_{p\in\xi,j\in\{1,\ldots,J\}}B(p^{(j)})$ where $B(p^{(j)})=\lvert\lvert p^{(j)}-\bar{p}\rvert\rvert-\frac{\lvert\lvert p^{(j-1)}-\bar{p}\rvert\rvert+\lvert\lvert p^{(j+1)}-\bar{p}\rvert\rvert}{2}$

Table 4.2: A listing of the 6 boundary-based features utilized to evaluate object morphology and compared against our ESD feature set. Note that $|\xi|$ represents the cardinality of set ξ.

Fourier Descriptors

The Fourier Descriptor feature set comprised the first 50 frequency components calculated from the contour of an object ξ [125]. The frequency of the contour was calculated as follows: a set of ordered points around the contour $p^{(j)} \in \xi : j \in \{1, \ldots, J\}$ were found. The magnitude of the points was calculated as $\rho^j = ||p^j - \bar{p}||$ where $\bar{p} = \frac{1}{|\xi|}\sum_{p\in\xi} p$. The Fourier transform of ρ is calculated and is used to derive the first 50 frequency components of each contour.

Path features

We calculated an alternative shape dissimilarity matrix, Z, using a path-based measure of MASM dissimilarity previously presented by [148]. This MASM similarity measure has been demonstrated to perform similarly or better than edit-distances on shock graphs [147]. Given two MASMs $\mathcal{Q}_a$ and $\mathcal{Q}_b$, a set of medial atoms $q'_a \in Q_a$ and $q'_b \in Q_b$ comprising the end nodes of the medial axes are identified. End nodes are defined as those medial atoms with only one neighbor on the medial axis. A path between pairs of end nodes $q_a^{'w} \in Q_a$ and $q_a^{'z} \in Q_a$ is defined as $\nu(q_a^{'w}, q_a^{'z})$. Similarly, $\nu(q_b^{'\hat{w}}, q_b^{'\hat{z}})$ is defined for end node pairs in Q_b. The dissimilarity between Q_a and Q_b is then found by,

$$Z(a,b) = \min_{w,z,\hat{w},\hat{z}} \left[\Delta(\nu(q_a^{'w}, q_a^{'z}), \nu(q_b^{'\hat{w}}, q_b^{'\hat{z}}))\right]. \tag{4.15}$$

The function Δ gives a measure of path similarity which is defined as the summation over the radius and path length [148]. The dissimilarity matrix Z is a high dimensional representation of the shape space, and as with our dissimilarity matrix A, Graph Embedding was employed to return the top n eigenvectors.

Experiment 1: medial axis shape model ability to capture morphology

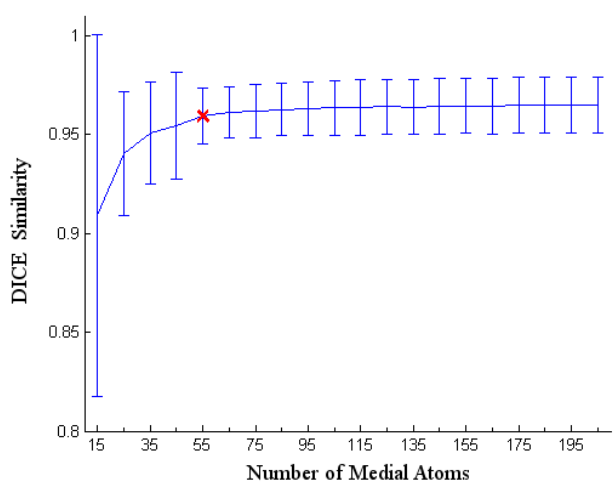


Figure 4.3: Prostate gland reconstruction accuracy as measured by DICE over number of medial atoms (blue). The number of medial atoms determined to give the highest reconstruction accuracy for the least computational cost is displayed (red cross). At first there is a large increase in DICE as more medial atoms are added to the MASM. After a certain point, adding more medial atoms does not significantly increase DICE.

We tested the hypothesis that the MASM can accurately represent the morphology of a wide variety of shapes. We assumed that MASM reconstruction accuracy reflects

the accuracy of the MASM to describe object shape. To evaluate this quantitatively, we reconstructed all of the objects in each dataset. For each dataset we varied the number of medial atoms contained in the MASM and for each set of medial atoms we reconstructed the object and denoted it as Ω_r. Ω_r is determined as a set of pixels belonging to an object given $\mathcal{Q}$ such that,

$$\Omega_r = \{c : ||q - c|| < r(q), q \in Q, c \in C\} \tag{4.16}$$

where the function $r(q)$ is defined by the equation,

$$r(q) = \frac{||(q - v_1(q))|| + ||(q - v_2(q))||}{2}. \tag{4.17}$$

For each Ω_r, we measured how close it is to Ω_1 using the edge based measures - (a) mean absolute distance (MAD) and (b) Hausdorff distance [162]; and the area based measures - (a) Dice's coefficient (DICE) [163] and (b) Positive Predictive Value (PPV). MAD describes on average the extent of variation between the ground truth shape contour and the reconstructed shape and is formally defined as,

$$\frac{1}{|p|} \sum_{p \in \Omega_1} \min_{o \in \Omega_r} ||p - o||. \tag{4.18}$$

Hausdorff distance [162] measures the performance of the worst case disparities between two shapes and is defined as,

$$\max_{p \in \Omega_1} \left(\min_{o \in \Omega_r} ||p - o|| \right). \tag{4.19}$$

DICE [163] is a measure of overlap between two shapes, in this case it reflects the extent of overlap between the reconstructed shape and the ground truth shape and is defined as,

$$\frac{|\Omega_1 \cap \Omega_r|}{|\Omega_1| + |\Omega_r|}. \tag{4.20}$$

PPV in this case is used to evaluate the proportion of pixels in the reconstructed shape accurately identified as belonging to the foreground of the object and is defined as,

$$\frac{\Omega_r \cap \Omega_1}{|\Omega_r|}. \tag{4.21}$$

For each dataset the fewest number of medial atoms that achieved high DICE and PPV were selected to represent all objects in the database for calculation of ESDs.

Figure 4.2.3 illustrates an example of the medial atom evaluation performed for the prostate histopathology dataset, with a red cross displayed at the optimal number of medial atoms for representing prostate morphology. The object reconstruction accuracy for the optimal number of medial atoms is reported in Table 4.3. As seen in Table 4.3 the synthetic dataset which contains $3D$ objects needs more medial atoms to more accurately represent the morphology compared to the prostate histopathology dataset which contains $2D$ objects. This is to be expected since $3D$ objects can have more complex shapes compared to their $2D$ counterparts.

Performance Measure	Dataset (Dimensionality)	
	Ellipsoid ($3D$)	Prostate Gland ($2D$)
Medial Atoms	1000	55
MAD	3.40 ± 3.54	0.01 ± 0.01
Hausdorff	24.46 ± 12.95	1.51 ± 1.16
DICE	0.87 ± 0.12	0.95 ± 0.03
PPV	0.94 ± 0.06	0.98 ± 0.02

Table 4.3: Object reconstruction accuracy for 2 datasets using the previously determined optimal number of medial atoms for each dataset. MAD and Hausdorff distances are shown in units of pixels. DICE and PPV are unitless ratios. Note that $3D$ objects require more medial atoms to accurately represent object morphology.

Experiment 2: registration evaluation

In this experiment the diffeomorphic registration algorithm presented in Section 4.2.2 was evaluated in terms of its ability to (a) recover a large range of non-linear deformations applied to MASMs and (b) determine accurate correspondences between medial atoms on Q_a and Q_b. We conducted a total of 200 experiments in which 20 randomly chosen MASMs and 10 deformation fields were considered. A deformation field denoted as T' was generated by varying the type and magnitude of the deformation applied. We applied T' to the image space C^h containing $\mathcal{Q}_a$ and then used our diffeomorphic registration algorithm to approximate the inverse transformation T^{-1}. We then computed the mean residual error as $\frac{1}{|Q_a|}\sum_{q \in Q_a} = ||q - T^{-1}(T'(q))||$, where $|Q_a|$ is the number of medial atoms contained in $\mathcal{Q}_a$.

Mean residual error was on average 1.09 ± 0.24 pixels, where accurate cluster centroid correspondence between MASMs allowed for close alignment between MASMs. The

worst case registration for a given MASM was 4.94 ± 2.22 pixels. For this specific case the MASM had several branches so that the cluster centroids on the original MASM and the deformed MASM failed to correspond to the equivalent locations on the original MASM. Incorrect correspondence determination between the cluster centroids may cause the diffeomorphic registration to be unable to approximate T^{-1}. In these 10 cases, on average 2.8 ± 0.3 cluster centroids did not have correct correspondence. This effect was only seen in 10 of 200 MASMs and only during the application of large deformations; deformations that had a magnitude greater than 10% of the area of Ω_1, the foreground region of the object of interest.

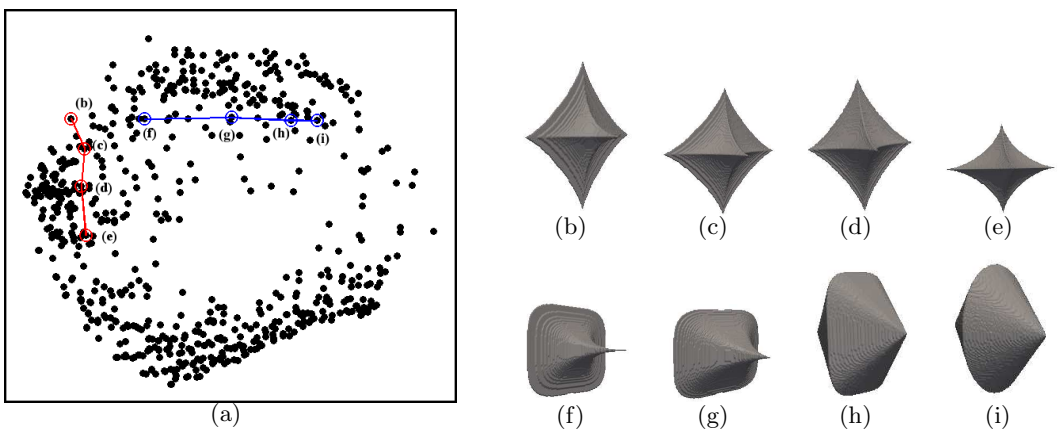


Figure 4.4: (a) The first and second ESDs are plotted on the X and Y axes respectively. Note that the manifold is curvilinear with the two axes corresponding roughly to the variation in ϵ_1 (red) and ϵ_2 (blue) respectively. (b)-(e) Ellipsoids with all parameters held equal except ϵ_2, resulting in subtle differences between object morphology. (f)-(i) Ellipsoids with all parameters held equal except ϵ_1, resulting in subtle differences between object shape. Finally note that the two ellipsoids farthest on the manifold, (e) and (i), are the most dissimilar.

Experiment 3: distinguishing between super quadratic ellipsoids with differing shape parameters

We constructed a set of 4096 super quadratic ellipsoids, denoted by S, to evaluate the ability of our ESD features to represent subtle shape variations between objects (see Section 4.2.3). Dissimilarity between the pair $\mathcal{E}_a$ and $\mathcal{E}_b$ for known sets of shape parameters was measured as $\Pi(\mathcal{E}_a, \mathcal{E}_b) = \sum_{\varpi} ||\varpi_a - \varpi_b|| : \varpi \in \{\alpha_1, \alpha_2, \alpha_3, \epsilon_1, \epsilon_2\}, a, b \in$

$\{1, \ldots, N\}$. $\Pi(\mathcal{E}_a, \mathcal{E}_b)$ represents the total shape dissimilarity.

The correlation between a set of ESD features $\mathbf{y}' = [y_1, \ldots, y_N]$ for a set of N objects and a set of known shape differences Π was calculated as follows. We first define the distance between two objects in the ESD feature space as, $\Xi(\mathcal{E}_a, \mathcal{E}_b) = ||y_a - y_b||$. We then calculate Pearson's correlation coefficient between Π and Ξ [164]. This allows us to quantitatively evaluate the ability for the ESD feature space, represented by $\Xi(\mathcal{E}_a, \mathcal{E}_b)$, to reflect known shape differences (Π).

Pearson's correlation coefficient was determined to be R=0.82, demonstrating a strong correlation between the know shape parameters and the underlying ESD feature space. Figure 4.4 displays all objects in the first 2 dimensions of the ESD feature space with representative S displayed in Figures 4.4(b) - 4.4(i). Note that a curvilinear manifold that contains the subspace of these shapes is clearly visible. The first dimension of the ESD feature space correlates to changes in ϵ_1, corresponding to the red line. The second dimension of the ESD feature space correlates to ϵ_2, corresponding to the blue line. Note that similar super quadratic ellipsoids are embedded adjacent to each other in the feature space while dissimilar super quadratic ellipsoids are embedded far apart. These results on synthetic data suggest that ESDs are able to differentiate between subtle changes in shape.

Experiment 4: Gleason grading of prostate histopathology

We evaluated the ability of four feature sets (Boundary, FD, Path, ESD) to accurately distinguish between Gleason grade 3 (G3), grade 4 (G4), and benign (BE) prostate glands as seen on histopathology using a SVM classifier. As this is a multiclass problem we evaluated SVM classifiers for the following 4 pairwise classification tasks: BE versus other (G3 and G4), G3 versus other (BE and G4), G4 versus other (BE and G3), and G3 versus G4. SVM training and evaluation was performed as described in Section 4.2.2.

The ESD and Path feature sets were evaluated over $1 \leq n \leq 30$ and $1 \leq \gamma \leq 2000$ for each classification problem (results not shown). SVM CAs and corresponding n and γ for each features set is shown in Table 4.3(a). For all classification problems, $n = 4$

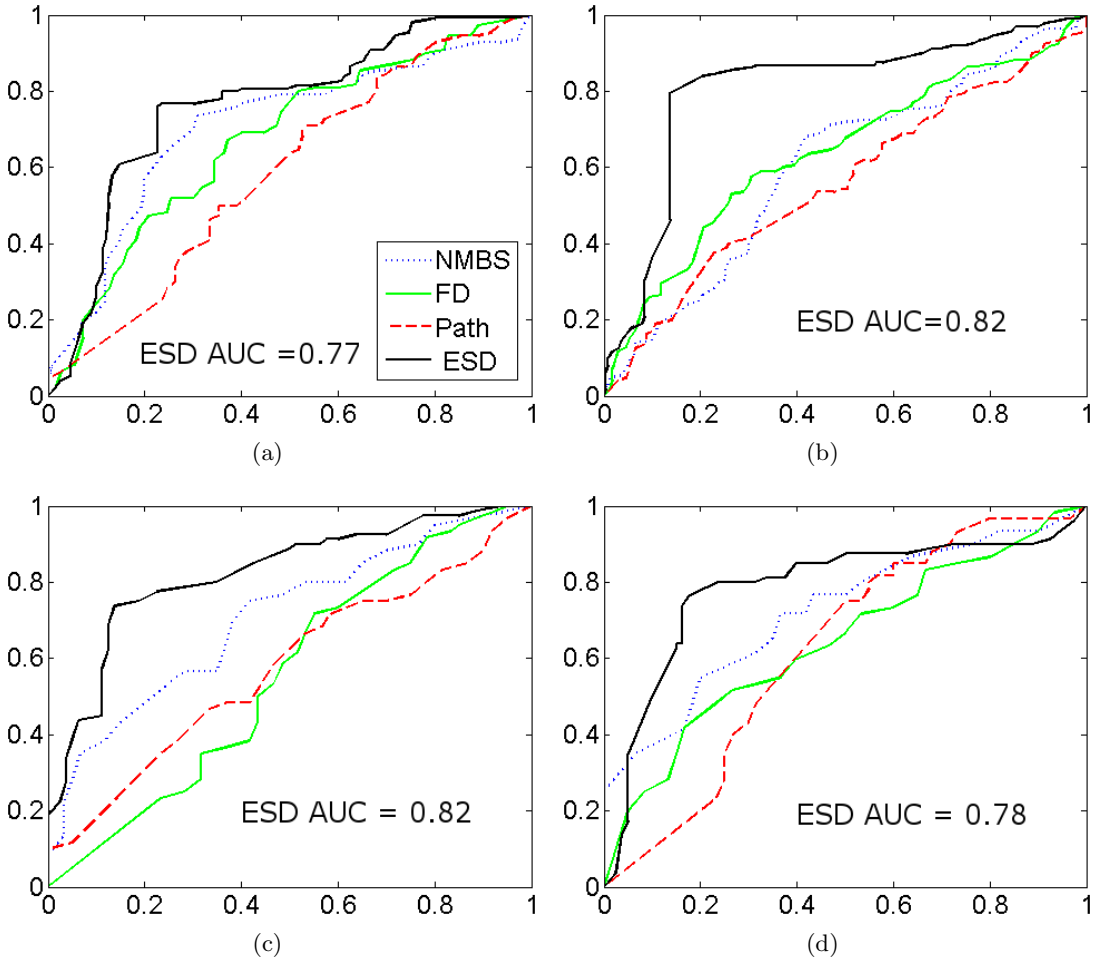


Figure 4.5: ROC curves for a SVM classifier trained using 3-fold cross validation on 888 prostate glands for Gleason grade classification of prostate glands as seen on histopathology in four tasks: (a) BE versus Other (G3 and G4), (b) G3 versus other (BE and G4), (c) G4 versus other (BE and G3), and (d) G3 versus G4. Four feature sets were evaluated, Boundary (blue), FD (green), Path (red), and ESD (black).

Classification Task		Feature			
		Boundary	FD	Path	ESD
BE v. Other	CA	0.80 ± 0.05	0.70 ± 0.06	0.60 ± 0.05 $\gamma = 100$, $n = 9$	$\mathbf{0.83 \pm 0.04}$ $\gamma = 1.2$, $n = 4$
	p-value	**0.005**	**0.001**	**0.005**	-
G3 v. Other	CA	0.70 ± 0.08	0.65 ± 0.04	0.61 ± 0.08 $\gamma = 1025$, $n = 14$	$\mathbf{0.87 \pm 0.07}$ $\gamma = 2.1$, $n = 4$
	p-value	0.013	**0.006**	**0.003**	-
G4 v. Other	CA	0.75 ± 0.06	0.58 ± 0.14	0.65 ± 0.10 $\gamma = 1$, $n = 4$	$\mathbf{0.85 \pm 0.03}$ $\gamma = 1.5$, $n = 4$
	p-value	**0.005**	0.024	0.016	-
G3 v. G4	CA	0.68 ± 0.08	0.60 ± 0.10	0.64 ± 0.06 $\gamma = 151$, $n = 5$	$\mathbf{0.89 \pm 0.06}$ $\gamma = 1.2$, $n = 4$
	p-value	**0.004**	**0.001**	**0.002**	-

(b)

Classification Task		Feature			
		NMBS	FD	Path	ESD
BE v. Other	AUC	0.71 ± 0.07	0.67 ± 0.08	0.61 ± 0.09 $\gamma = 100$, $n = 9$	$\mathbf{0.77 \pm 0.05}$ $\gamma = 1.2$, $n = 4$
	p-value	0.15	0.059	**0.007**	-
G3 v. Other	AUC	0.60 ± 0.09	0.64 ± 0.06	0.56 ± 0.08 $\gamma = 1025$, $n = 14$	$\mathbf{0.81 \pm 0.05}$ $\gamma = 2.1$, $n = 4$
	p-value	**0.002**	**0.001**	**0.0004**	-
G4 v. Other	AUC	0.71 ± 0.08	0.57 ± 0.06	0.56 ± 0.20 $\gamma = 1$, $n = 4$	$\mathbf{0.82 \pm 0.07}$ $\gamma = 1.5$, $n = 4$
	p-value	0.042	**0.0003**	0.023	-
G3 v. G4	AUC	0.72 ± 0.04	0.58 ± 0.17	0.64 ± 0.17 $\gamma = 151$, $n = 5$	$\mathbf{0.78 \pm 0.11}$ $\gamma = 1.2$, $n = 4$
	p-value	0.25	0.065	0.162	-

Table 4.4: (a) CA and (b) AUC for a SVM classifier trained using 3-fold cross validation on 888 prostate glands for distinguishing between Gleason grades on prostate histopathology. SVM classifiers were trained with 4 feature sets (Boundary, FD, Path, ESD). In total 16 classification studies (4 feature sets, 4 pairwise classification tasks) were performed. p-values comparing ESD to the comparison feature sets are reported, statistically significant p-values ($p < 0.01$) are bolded. The best CA and AUC across the feature sets is bolded.

was empirically determined to yield the consistently best results in the ESD feature space. A narrow range of $1 \leq \gamma \leq 2$ was identified as yielding the best performance in the ESD feature space, however γ can be adjusted to obtain better performance for a specific classification task. In contrast, Path had a wide range of n and γ which yielded high CA and AUC values. For all classification problems considered, the ESD features outperformed the Boundary, FD, and Path feature sets. ROC curves for each classification task are displayed in Figure 4.5 and corresponding AUC values are reported in Table 4.3(b).

Figure 4.6 displays the first 2 ESD features of the 888 glands; representative glands from three different classes are shown along with their corresponding locations in the ESD feature space. Misclassified glands, shown in the far right row often display characteristics very similar to glands of other Gleason grades. Consequently, some misclassifications may be attributed to glands displaying atypical attributes.

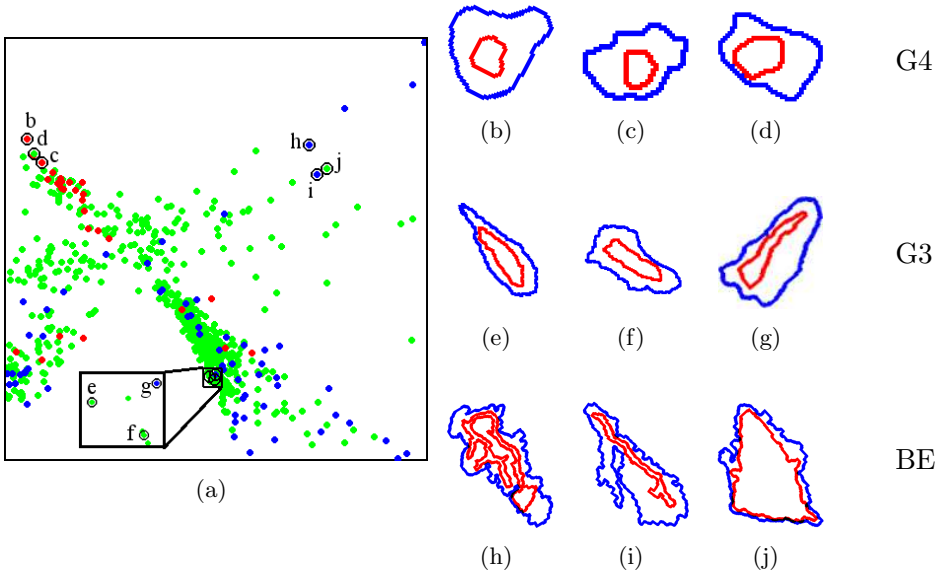


Figure 4.6: (a) ESD feature space for prostate digital histopathology with BE (blue), G3 (green), and G4 (red) glands. The first and second ESDs are plotted on the X and Y axes respectively. Lumen (red) and nuclear (blue) layers are shown, for glands labeled (b)-(d) G4, (e)-(g) G3, and (h)-(j) BE. Ground truth for mislabeled glands, displayed in the far right row, are (d) G3, (g) BE, (j) G3. Glands with similar shapes are embedded adjacent to each other on the manifold while glands with dissimilar shapes are embedded far apart.

Experiment 5: evaluation of gland misclassification

We performed additional evaluation of our classification results to better understand the reasons behind gland misclassification. As stated previously, glands may be misclassified if they exhibit traits atypical to the Gleason grade they are assigned, however, other reasons for misclassification exists. Each step in the ESD framework may introduce errors that result in misclassification. We have identified that misclassifications may occur due to the following reasons:

1. Glands may be inaccurately represented by the MASM.
2. MASMs may be misaligned during the registration step. As the subsequent steps in the ESD methodology involves computing shape distances between aligned MASMs, glands dissimilarity will be inaccurate if the registration is inaccurate.
3. Misclassification may occur if the pairwise distances between MASMs are not preserved when projecting from the original high dimensional feature space A into the lower dimensional space y.

4 Glands that exhibits atypical traits the Gleason grade they belong to may be misclassified.

As each source of misclassification is due to a unique underlying cause we have assessed sources of misclassification within our framework. However, it is impossible to assess whether glands demonstrate atypical shape characteristics except by manual inspection, and have not tried to quantify glands that may be misclassified due to this cause.

Experiment 5a: gland misclassification due to medial axis shape model

Glands for which the shape variance may not be fully characterized by the MASM, we anticipate will have a lower than expected reconstruction accuracy. We have attempted to identify these outlier glands or shapes via use of the DICE [163]. DICE [163] is a measure of overlap between two shapes, in this case it reflects the extent of overlap

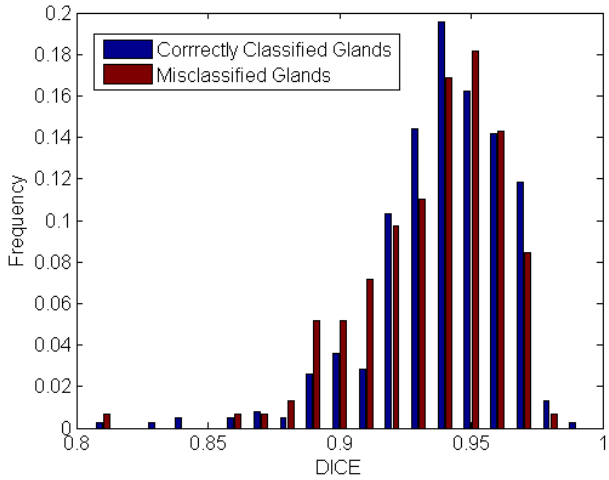
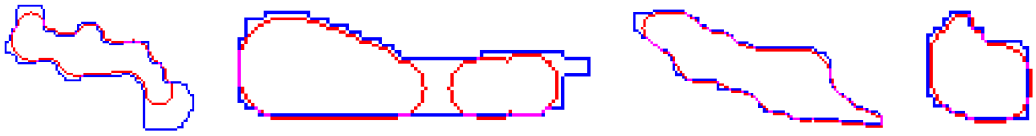


Figure 4.7: Histogram of DICE values for correctly classified (blue) and misclassified (red) glands. A DICE value of 1 represents a MASM that can accurately reconstruct the original gland shape. A DICE value of 0 represents a complete inability to reconstruct the original gland shape.



(a) DICE=0.88 (b) DICE=0.81 (c) DICE=0.91 (d) DICE=0.94

Figure 4.8: (a), (b) Representative misclassified glands due to poor MASM reconstruction and (c), (d) representative correctly classified glands. The blue contour represents the original shape of the gland and red represents the gland shape reconstructed from the MASM. Note that the glands which are correctly classified ((c), (d)) have fewer discrepancies between the original shape and the reconstructed shape compared to misclassified glands ((a), (b)).

between the reconstructed gland shape and the ground truth gland shape. For glands where the MASM preserves the shape characteristics we expect a DICE value near 1, for glands where the MASM cannot preserve the original shape we expect a low DICE value. Figure 4.2.3 shows a histogram plot of the Dice coefficient for correctly classified (blue) and misclassified (red) glands. We determined approximately 20% of misclassified glands had DICE coefficient values below 0.90, compared to correctly classified glands where only 7% have Dice coefficient values below 0.9. Figure 4.8 shows representative gland shapes (blue), for both correctly classified and misclassified glands, and their reconstructed shapes (red).

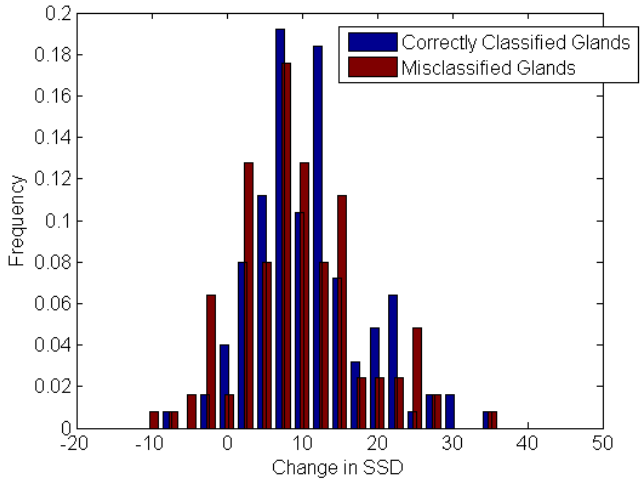


Figure 4.9: Histogram of the change in SSD, represented in real world coordinates, during the diffeomorphic registration.

Experiment 5b: gland misclassification due to registration

There is no ground truth for correspondence between MASMs, we evaluated the alignment (or misalignment) between MASMs for different glands by quantifying the sum of squared differences (SSD) between corresponding medial atoms for a pair of MASMs. To evaluate the registration algorithm, we compute SSD before and after the diffeomorphic registration, and then calculate the change in SSD between these values. In this experiment we make the implicit assumption that glands with small changes in SSD are more likely to have incorrect correspondence. However, we recognize that this is only a surrogate measure to determine how well MASMs are aligned via our registration algorithm. Figure 4.2.3 shows a histogram plot of change in SSD for correctly classified (blue) and misclassified (red) glands. We identified that approximately 40% of misclassified glands had a change in SSD of less than 1 μm in real world coordinates, corresponding to approximately 5 pixels, a low change in SSD compared to correctly classified glands. Over 85% of correctly classified glands have a change in SSD greater than 1 μm.

Experiment 5c: gland misclassification due to dimensionality reduction

We evaluated if pairwise distances between gland representations in A and $\mathbf{y}'$ are preserved by the following procedure. We determined a set of glands N_a in the local

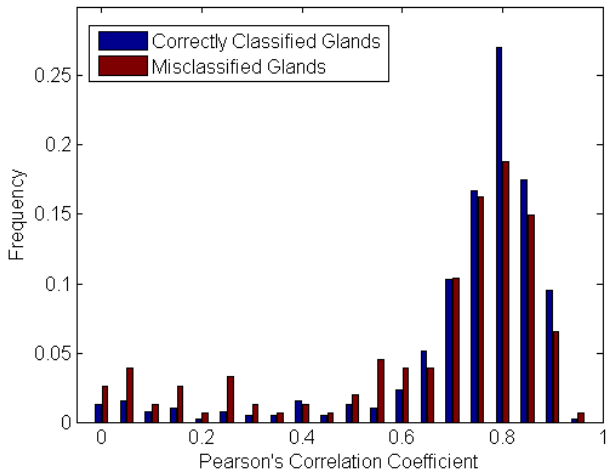


Figure 4.10: Histogram of Pearsons correlation coefficient values between Δ_{HD} and Δ_{LD} for correctly classified (blue) and misclassified (red) glands.

neighborhood of a single gland such that $W(a,b) < r : b \in N_a$, where r is the threshold distance that determines if two glands are considered neighbors. $\Delta_{HD} = W(a,b)$ represents the distance between two glands in the original high dimensional space.

We calculated the distance between glands in the low dimensional space as $\Delta_{LD} = ||y_a - y_b|| : b \in N_a$.

We then evaluated Pearson's correlation coefficient between Δ_{HD} and Δ_{LD} [164]. Low correlation values reflect glands for which GE does not preserve neighborhood relationships between glands while high correlation values reflect glands where neighborhood relationships are preserved.

Figure 4.2.3 shows a histogram plot of the Pearsons correlation coefficient for correctly classified (blue) and misclassified (red) glands. We determined that approximately 25% of misclassified glands had Pearsons correlation coefficient values below 0.60, while less than 5% of correctly classified glands had values below 0.6. This suggests that GE does not accurately preserve pairwise distances between A and $\mathbf{y}'$ in the 25% of misclassified glands with a low Pearsons correlation coefficient.

4.2.4 Concluding Remarks on Explicit Shape Descriptors

Explicit Shape Descriptors (ESDs) are calculated by (a) fitting a medial axis shape model (MASM), (b) calculating diffeomorphic based similarity (DBS), and (c) applying Graph Embedding (GE) to the shape dissimilarity matrix to find a set of ESDs. The

individual modules contained in our ESD calculations (e.g. MASM [120], GE [155]) have been previously described, however, our methodology represents a novel integration of each of these methods in order to describe object morphology with a concise set of features. ESDs were able to distinguish between subtle differences in super quadratic ellipsoids and were also able to distinguish between prostate glands on histopathology with subtle morphologic differences with a maximum accuracy of 89% for 888 prostate glands acquired from 58 patient needle core biopsies.

ESDs offer distinct advantages compared to previously reported methods which combine shape models with NLDR. ESDs, unlike that of previously presented work [151, 152], is unique in that it can be applied to data where class information is not known. ESDs are derived from the eigenvector decomposition of the high dimensional shape dissimilarity matrix and hence are guaranteed to be of low dimensionality while simultaneously preserving pairwise class relationships between objects. For the datasets considered in this work, ESDs were better able to accurately capture morphologic differences between objects compared to other classes of feature descriptors including boundary-based features [102], Fourier Descriptors (FDs) [125], and MASM path similarity [148].

The results from Experiment 1 (Section 4.2.3) show that the MASM is able to accurately model the shapes of glands as seen on prostate histopathology. However, a relatively minor portion of glands (approximately 20% of misclassified glands) may be misclassified due to an inability of the MASM to capture subtle differences in shape.

The results from Experiment 2 (Section 4.2.3) demonstrate that DBS is able to accurately determine correspondence between a wide variety of MASMs over a range of deformations. However, determining MASM correspondence is a difficult task. Furthermore, evaluation of the registration step demonstrated that in approximately 40% of misclassified glands the inability to correctly determine medial atom correspondence, and hence accurately determine shape dissimilarity, may be responsible for misclassification (results not shown).

Finally, gland misclassification may be caused by GE being unable to accurately preserve the relationships between samples when projecting the high dimensional space

into a lower dimensional space. We found roughly 25% of misclassified glands may be on account of classes relationships not being preserved during this step. Future work will evaluate other NLDR algorithms (e.g. Locally Linear Embedding (LLE) [165], Isomaps [166]) within our framework.

For approximately 15% of glands that were misclassified we were unable to determine which step in our algorithm may be responsible for the misclassification. Visual inspection of a subset of these glands showed that some misclassifications may be attributed to glands displaying atypical shape attributes. In future work we will integrate the ESD features with other types of histologic image attributes, such as the shape and arrangement of nuclei [111] or texture [48, 167], to build classification tools for Gleason grading of prostate histopathology that may be better able to classify glands with atypical shape attributes.

4.3 Out-Of-Sample Extrapolation Utilizing Semi-Supervised Learning for Efficient Explicit Shape Descriptors

In this section we present out-of-sample extrapolation utilizing semi-supervised learning (OSE-SSL) for computationally efficient calculation of ESDs for never before seen images. OSE-SSL is evaluated in the context of Content-based Image Retrieval (CBIR) where images are retrieved from a database according to their similarity to a query image. In the context of medical imagery, images which are visually similar often have similar pathologies. A CBIR system for histopathology images could serve as a useful training tool for pathology residents, fellows, and medical students and could potentially serve as a decision-support tool in diagnosis and grading of pathologies [168–177]. CBIR systems are particularly relevant in the context of histopathology imagery where (a) the images can be extremely large and described by a very large set of image descriptors, and (b) differences between pathologies may be very subtle and not immediately appreciable visually. Additionally, with the recent advent of whole-slide digital scanners, pathology labs will soon be routinely generating very large amounts of digitized histopathology imagery, necessitating intelligent and efficient image retrieval systems [178].

CBIR systems attempt to retrieve images from a database identified as being the most similar to the query image in terms of quantitative image descriptors obtained from the query and database images. In the context of medical imagery, images which are visually similar often have similar pathologies. A CBIR system for histopathology images could serve as a useful training tool for pathology residents, fellows, and medical students and could potentially serve as a decision-support tool in diagnosis and grading of pathologies [168–177]. CBIR systems are particularly relevant in the context of histopathology imagery where (a) the images can be extremely large and described by a very large set of image descriptors, and (b) differences between pathologies may be very subtle and not immediately appreciable visually. Additionally, with the recent advent of whole-slide digital scanners, pathology labs will soon be routinely generating very large amounts of digitized histopathology imagery, necessitating intelligent and efficient image retrieval systems [178].

The remainder of the section is organized as follows. In Section 4.3.1 we review previous work in CBIR systems for histopathology and describe the novel contributions of OSE-SSL. Section 4.3.2 describes the theoretical foundations of OSE-SSL and the OSE-SSL algorithm. Section 4.3.3 discusses our experimental design and results for OSE-SSL. Section 4.3.4 presents our concluding remarks.

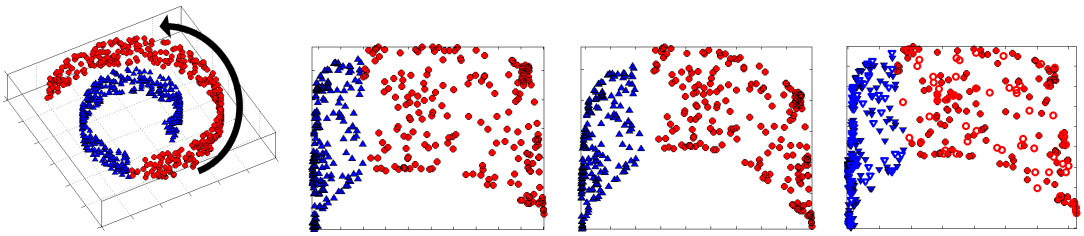
4.3.1 Previous Work in Content-based Image Retrieval for Histopathology

CBIR systems typically comprise two components: (1) a module for extraction of domain specific image descriptors to quantitatively characterize the images, and (2) a module for computation of the similarity between the query and database images in terms of the quantitative image descriptors. Several CBIR methods for radiological medical imagery have been presented [179, 180]. Such CBIR systems extract relatively few image descriptors and hence are able to accurately perform image retrieval in the original high dimensional image descriptor space. In comparison, CBIR systems for histopathology imagery extract a very large number of features to describe the complex imagery [168, 170–177]. Such medical imagery can be represented by a high dimensional

space, where each dimension corresponds to a single image descriptor.

A high dimensional image descriptor space makes the calculation of similarity between image descriptors difficult as (a) the number of database images may be small compared to the number of image descriptors giving rise to the curse of dimensionality problem [181], and (b) images often cluster densely in small regions of the high dimensional space [182]. Hence relationships between image descriptors may be important when calculating image similarity. Consequently, a few researchers have proposed dimensionality reduction methods [168,171,172,175,177,183] to map the high dimensional image descriptors into a low dimensional representation so that image similarity calculation and retrieval can be performed directly in the low dimensional space. Retrieval performed in a low dimensional space is often more accurate than retrieval performed in the original high dimensional space [172, 175]. However, utilizing dimensionality reduction methods to learn a low dimensional space may add computational complexity to the retrieval algorithm.

Linear dimensionality reduction methods, such as Principal Component Analysis (PCA), attempt to find a low dimensional space that is a linear projection of the high dimensional space. Hence linear dimensionality reduction methods only preserve linear relationships between images [168, 171]. Semi-supervised learning (SSL) methods, such as Linear Discriminant Analysis, have been proposed to take into account semantic information such as partial class labels when learning a low dimensional projection in order to co-localize semantically similar images [173, 180, 184]. Comaniciu et. al. [168] utilized a weighted sum of image descriptors, where weights were determined by maximizing an objective function, to retrieve images corresponding to different hemotologic malignancies. This approach is equivalent to a linear dimensionality reduction method as only linear relationships between images are preserved during retrieval. Zheng et. al. [171] utilized multi-dimensional scaling, a linear dimensionality reduction scheme, to compute a low dimensional space in which image retrieval could be performed for a set of histopathology images taken from different anatomical regions (e.g. spleen, prostate, colon, etc.). However, these methods assume that a linear projection of the high dimensional space will best preserve relationships between images.



(a) (b) (c) (d)

Figure 4.11: (a) $3D$ Swiss Roll dataset comprising 2000 samples belonging to two classes (red, blue). The arrow displays the direction of greatest variance along the manifold. (b) $2D$ low dimensional embedding space found via Graph Embedding. Note that the two classes cluster on different regions of the low dimensional embedding space. (c) $2D$ low dimensional embedding space found via semi-supervised Graph Embedding. Note that the two classes are more separated than for Graph Embedding. (d) $2D$ low dimensional embedding space found via Graph Embedding (closed points) and OSE (open points).

Manifold learning schemes attempt to find a low dimensional embedding space which preserves the manifold structure of the image descriptors in the high dimensional space. Hence manifold learning methods attempt to preserve the non-linear relationships between image descriptors [155, 165, 166]. Graph Embedding [155], a specific instance of a manifold learning scheme, attempts to model the manifold structure using local, pairwise relationships between image descriptors in the high dimensional space thereby preserving these relationships between images in the low dimensional space. Recent work has demonstrated that manifold learning schemes, such as Graph Embedding, may result in low dimensional spaces better suited for CBIR when image similarity is defined by a non-linear manifold in the high dimensional space [172,175,177,183]. Semi-supervised manifold learning methods, which utilize SSL in conjunction with manifold learning, attempt to learn a low dimensional embedding space such that semantic, non-linear relationships between images in the high dimensional space are preserved [185]. To our knowledge no CBIR systems for histopathology have leveraged semi-supervised manifold learning. However, CBIR systems for color photography [183, 186] have been proposed which leverage such methods.

Figure 4.11 demonstrates the ability of Graph Embedding to preserve non-linear relationships between samples in the case of a synthetic Swiss Roll dataset. Figure

4.11(a) shows a synthetic Swiss Roll dataset consisting of 2000 samples described by a $3D$ space, the arrow demonstrating the direction of greatest variance along the manifold. In this example Graph Embedding is able to find a low dimensional space ($2D$) which preserves the underlying structure of the dataset as evidenced by the planar $2D$ embedding space shown in Figure 4.11(b). Figure 4.11(c) shows the results of semi-supervised Graph Embedding for the Swiss Roll. Note that for SSGraph Embedding (Figure 4.11(c)) samples from two classes (blue, red) have a larger separation compared to Graph Embedding (Figure 4.11(b)).

Despite the advantages of manifold learning, only a few papers have attempted to use manifold learning in conjunction with CBIR of medical imagery [172, 175, 177], due to its computational cost. A computationally expensive eigenvalue decomposition must be calculated for every new query image [187, 188]. Hence there is a need to develop manifold learning schemes which are more computationally efficient and do not require a eigenvalue decomposition for each new query image. Algorithms have been developed to avoid recomputing the eigenvalue decomposition for out-of-sample images, but have not previously been evaluated in the context of CBIR for medical imagery [187, 188].

Locality Preserving Projections attempts to approximate the low dimensional embedding space found by manifold learning as a linear combination of image descriptors in the high dimensional space [187]. Locality Preserving Projections is reliant on a linear combination of the image descriptors accurately modeling relationships between images, and hence accurately modeling relationships in the low dimensional space. If the low dimensional space found via manifold learning is not approximately linear Locality Preserving Projections will not correctly estimate the low dimensional space. Alternatively, out-of-sample extrapolation (OSE) [188] attempts to determine the location (or embedding) of a new query image in the low dimensional space as a weighted sum of the low dimensional embeddings already calculated for a set of images. In the context of a CBIR system, the calculated embeddings would correspond to the embedding location for the database images. Unlike Locality Preserving Projections, the non-linear relationships between images are preserved, and hence OSE may be better able to resolve differences between images belonging to different classes. Figure 4.11(d)

shows the results of OSE for the Swiss Roll dataset where samples which were projected into the low dimensional space via OSE are represented by open points.

We have developed OSE-SSL algorithm, which represents a novel combination of SSL and OSE, designed specifically to be computationally tractable. The novel integration of these two methods involves projecting never-before seen images into a low dimensional embedding space that takes into account semantic information (class label information). Hence OSE-SSL (a) integrates known label information to learn a low dimensional embedding space and (b) overcomes the out-of-sample problem. We demonstrate the use of OSE-SSL in the context of CBIR applications. Figure 4.12 illustrates a flowchart of our OSE-SSL CBIR system. The CBIR system is characterized by (1) offline database construction where SSL is applied to quantitative image descriptors for a set of database images to obtain a low dimensional embedding space and (2) online image retrieval where OSE is used to compute the embedding location of a never before seen query image. Offline database construction consists of (a) extracting image descriptors for all database images, and (b) applying SSL to determine the low dimensional embedding space for images contained within the database. Once offline database construction has been completed online image retrieval is then performed efficiently utilizing OSE. Online image retrieval consists of (c) extracting image descriptors from the query image, (d) OSE of the query image into the low dimensional embedding space, and (e) ranking image similarity in the low dimensional embedding space.

OSE-SSL confers several advantages to a CBIR system. Firstly, OSE-SSL efficiently calculates embedding locations for images not contained in the database, such as query images, by removing the need to recompute the eigenvalue decomposition. Secondly, OSE allows for non-linear relationships between images to be appropriately modeled when learning the low dimensional embedding space. Thirdly, OSE-SSL allows for the utilization of semantic information when calculating the low dimensional embedding space. Leveraging semantic information allows for construction of an low dimensional embedding space where semantically similar images are near each other while semantically dissimilar images are farther away.

We demonstrate OSE-SSL in the context of the ESD framework (Section 4.2). We

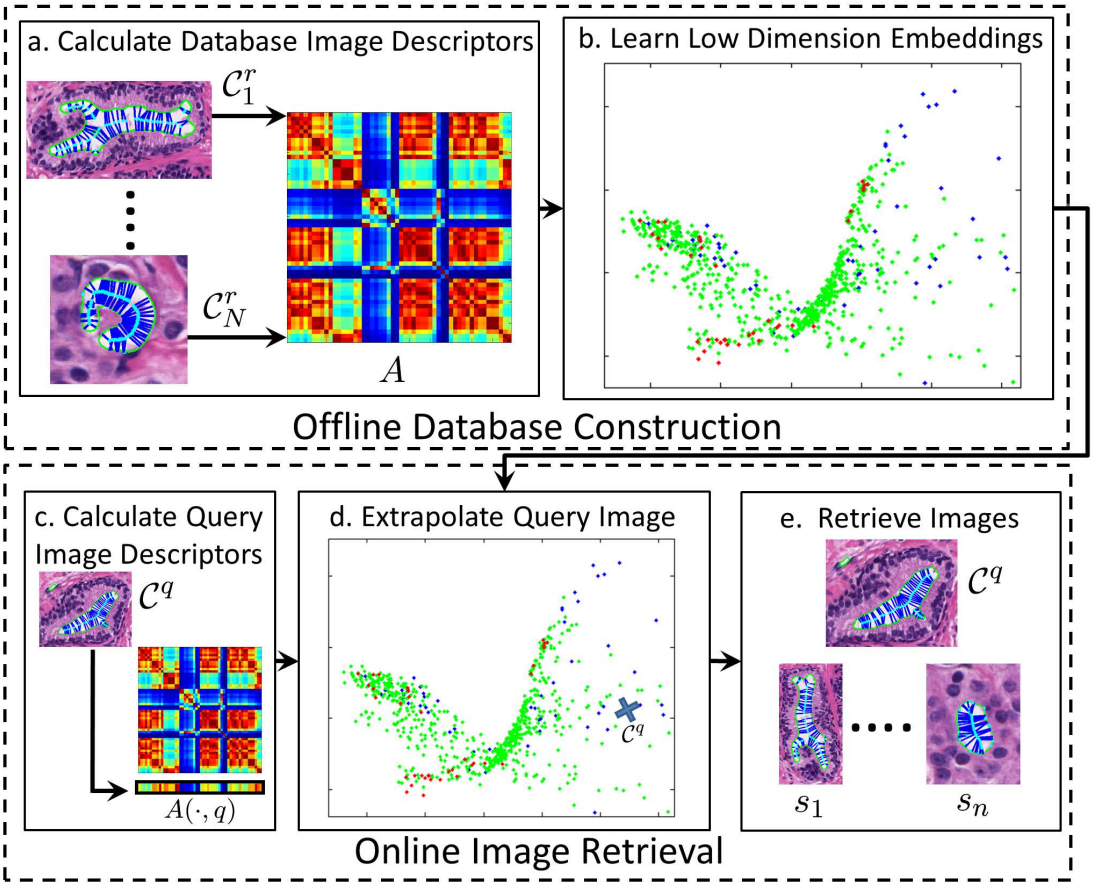


Figure 4.12: A flowchart of the OSE-SSL CBIR system. The system has an offline database construction phase (top) and an online retrieval phase (bottom). Database construction consists of (a) obtaining a set of N repository images ($\mathbf{C} = [\mathcal{C}_1^r, \ldots, \mathcal{C}_N^r]$) and extract image features, represented by the dissimilarity matrix A. (b) Performing SSL to learn the low dimensional embedding space which optimally describes similarity between images in $\mathbf{C}$. Retrieval of images most similar to a query image $\mathcal{C}^q$ is then performed via (c) extracting image features from $\mathcal{C}^q$, represented by $A(\cdot, q)$. (d) OSE of $\mathcal{C}^q$ into the low dimensional embedding space. (e) Image retrieval of the n most similar images ($s_1, \ldots, s_n$) to $\mathcal{C}^q$ according to Euclidean distance in the low dimensional embedding space.

evaluate our system on two datasets (a) synthetic MPEG-7 dataset [189] and (b) a digitized prostate histopathology dataset. The MPEG-7 dataset consists of synthetic black and white silhouette images, hence only morphologic similarity is relevant for image retrieval. The MPEG-7 dataset was selected to demonstrate that ESDs are able to accurately retrieve images according to morphology. The prostate histopathology dataset was chosen due to the challenges in accurately distinguishing between intermediate Gleason grades [45];

4.3.2 Out-of-Sample Extrapolation Utilizing Semi-Supervised Manifold Learning (OSE-SSL)

Notation

Table 4.5 displays the notation used to explain the methodology of OSE-SSL. A database of N images is defined by $\mathbf{C} = [\mathcal{C}_1^r, \ldots, \mathcal{C}_N^r]$. r denotes that the image is contained in $\mathbf{C}$, to contrast with $\mathcal{C}^q$ where q denotes a query image not contained in database. Each image in the database has a corresponding label defined by $\mathbf{L} = [l_1, \ldots, l_N]$. Every label $l_i \in \mathbf{L}$ takes on a discrete value $l_i \in \{1, 2, \ldots, Z\}$ where $\mathbf{C}$ contains images belonging to Z classes.

Symbol	Definition	Symbol	Definition
$\mathbf{C}$	Image database	N	Number of images in $\mathbf{C}$
$\mathcal{C}_i^r$	ith image in $\mathbf{C}$	$\mathcal{C}^q$	User-selected query image
$\phi(\mathcal{C}_i^r, \mathcal{C}_j^r)$	Dissimilarity function between $\mathcal{C}_i^r$ and $\mathcal{C}_j^r$	$\mathbf{y}^r$	Embeddings in the low dimensional space $\mathbb{R}^d$
$\mathbf{L}$	Image label information	A	Dissimilarity matrix for $\mathbf{C}$
W	Similarity matrix for $\mathbf{C}$	y_i^r	Embedding location for $\mathcal{C}_i^r$
y^q	Embedding location for $\mathcal{C}^q$	$\mathbb{D}_{OS}$	Distance metric in $\mathbb{R}^d$

Table 4.5: Notation used to describe OSE-SSL.

For two images $\mathcal{C}_i^r \in \mathbf{C}$ and $\mathcal{C}_j^r \in \mathbf{C}, j \neq i$ we define pairwise dissimilarity as $A(i,j) = \phi(\mathcal{C}_i^r, \mathcal{C}_j^r)$. The function $\phi(\cdot,\cdot)$ can represent any dissimilarity function such that if $\phi(\mathcal{C}_i^r, \mathcal{C}_j^r) > \phi(\mathcal{C}_i^r, \mathcal{C}_k^r)$ then it follows that $\mathcal{C}_i^r$ and $\mathcal{C}_j^r$ are more dissimilar than $\mathcal{C}_i^r$ and $\mathcal{C}_k^r$. The function $\phi(\mathcal{C}_i^r, \mathcal{C}_j^r)$ is evaluated over all $i, j \in \{1, \ldots, N\}, j \neq i$ to obtain A.

A is an $N \times N$ matrix representing pairwise dissimilarity between all images contained in $\mathbf{C}$.

Review of Manifold Learning

Graph Embedding

The goal of Graph Embedding is to determine a set of low dimensional embedding locations $\mathbf{y} \in \mathbb{R}^n$ that preserves the relationships between images in $\mathbf{C} \in \mathbb{R}^N$ where $n \ll N$. Graph Embedding determines $\mathbf{y}$ by modeling the similarity between images according to a similarity matrix W. Given the dissimilarity matrix A described in Section 4.3.2, W is found by $W(i,j) = e^{-A(i,j)/\gamma}$, where γ is a user selected scaling parameter. $\mathbf{y}$ is then found by minimizing the pairwise reconstruction error defined as,

$$\mathbf{y} = \underset{\mathbf{y}}{\operatorname{argmin}} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} ||y_i - y_j||^2 W(i,j) \right], \tag{4.22}$$

where $||\cdot||$ denotes the L2-norm. An image $\mathcal{C}_i^r$ is associated with the embedding location y_i. Belkin *et. al.* [190] demonstrated that Equation 4.22 is equivalent to the following eigenvalue decomposition,

$$(D - W)\mathbf{y} = \lambda D \mathbf{y}, \tag{4.23}$$

where D is a diagonal matrix defined as $D(i,i) = \sum_{j=1}^{N} W(i,j)$. The largest d eigenvalues in λ correspond to the n eigenvectors $\mathbf{y}$ which are defined as the n dimensional embedding locations. $\mathbf{y}$ correspond to the projection of the matrix W into $\mathbb{R}^n$ such that the pairwise similarity between the elements in W, and hence the pairwise similarity between images, are preserved. Furthermore the eigenvectors $\mathbf{y}$ are orthonormal, hence, each additional eigenvector (or dimension) provides independent information on the image similarity in W.

Semi-Supervised Manifold Learning (SSL)

For $\mathbf{C}$ let a corresponding set of known labels be defined as $\mathbf{L}^r \subset \mathbf{L}$ where $\mathbf{L}^r = [l_1, \ldots, l_N]$. Note that N^l, the number of labels, is less than N, the number of database

images, as we assume that some labels may be unknown for images contained in $\mathbf{C}$. A similarity matrix W^r is constructed by altering elements in W, as defined in Section 4.3.2, according to $\mathbf{L}^r$. Images which correspond to the same class have higher values in W^r compared to W, while images which correspond to different classes have lower values in W^r compared to W. For images where no label information is known the values in W^r and W are equivalent. W^r is calculated as,

$$W^r(i,j) = \begin{cases} \omega(1+\omega) & \text{if } l_i = l_j, \\ \omega(1-\omega) & \text{if } l_i \neq l_j, \\ \omega & \text{otherwise,} \end{cases} \tag{4.24}$$

where $\omega = W(i,j)$. The "otherwise" case corresponds to instances where label information is unknown for either l_i or l_j. Once the similarity matrix W^r has been calculated, the eigenvalue decomposition described by Equation 4.23 is performed on W^r to obtain $\mathbf{y}^r$.

By altering W^r according to Equation 4.24, images belonging to the same class (i. e. $l_i = l_j$) will be close together in the low dimensional embedding space. Images belong to different classes (i. e. $l_i \neq l_j$) will be farther apart in the low dimensional embedding space. Images where class information is unknown (i.e. l_i or l_j are undefined) will be near images determined to be similar, in terms of $\phi(\cdot,\cdot)$, regardless of class.

Out-of-Sample Extrapolation (OSE)

OSE uses $\mathbf{y}$ determined from $\mathbf{C}$ to extrapolate y^q for $\mathcal{C}^q$. Assuming that $\mathbf{y}$ accurately describes the non-linear relationships in $\mathbf{C}$, which should be the case when $\mathbf{C}$ is sufficiently large, OSE is able to accurately determine y^q [191, 192].

OSE is divided into three steps,

1. Manifold Learning: A set of low dimensional embeddings $\mathbf{y}$ are learned by performing Graph Embedding on $\mathbf{C}$ as described in Section 4.3.2.

2. Query Image Descriptor Calculation: Pairwise dissimilarity $A(i,q)$ is calculated between $\mathcal{C}^q$ and every image contained in $\mathbf{C}$ as described in Section 4.3.2. $W(i,q)$ is calculated from $A(i,q)$ as in Section 4.3.2.

3. Query Sample Extrapolation: The embedding location y^q for $\mathcal{C}^q$ is extrapolated via,

$$y_k^q = \frac{1}{\lambda_k} \sum_{i=1}^{N} y_{i,k} W(i,q), \tag{4.25}$$

where $k \in \{1, \ldots, n\}$ is the kth embedding dimension corresponding to the kth smallest eigenvalue λ_k.

Intuitively, OSE calculates y^q as a weighted sum of the database embeddings $y_i : i \in \{1, \ldots, N\}$ where weights are based on image similarity described by $W(i,q)$.

Out-of-Sample Extrapolation for Semi-Supervised Manifold Learning

OSE-SSL is a novel combination of the previously described SSL and OSE algorithms that projects never-before seen images into a low dimensional embedding space that incorporates semantic information. OSE-SSL calculates $\mathbf{y}^r$ for $\mathbf{C}$ such that (a) the image class labels $\mathbf{L}^r$ are taken into account and (b) image similarity is optimally represented by $\mathbf{y}^r$. After $\mathbf{y}^r$ have been calculated for $\mathbf{C}$, a new never before seen image $\mathcal{C}^q$ can be extrapolated into the low dimensional space to obtain $\mathbf{y}^q$. OSE-SSL calculates the embedding $\mathbf{y}^q$ in a computationally efficient manner.

Our novel methodology for OSE-SSL can be divided into an offline *ConstructOSE-SSL* algorithm and an online *ApplyOSE-SSL* algorithm both of which are described in detail in Section 4.3.2. Section 4.3.2 describes the application of OSE-SSL to CBIR. Finally, the computational complexity of OSE-SSL is described in Section 4.3.2.

OSE-SSL Algorithm

The algorithm for OSE-SSL is divided into two parts, (1) *ConstructOSE-SSL* which is an offline computationally intensive algorithm to learn $\mathbf{y}^r$ that only needs to be performed once for $\mathbf{C}$ and (2) *ApplyOSE-SSL* which is an online algorithm to extrapolate y^q for $\mathcal{C}^q$. The combination of these two algorithms results in a low dimensional representation for both $\mathbf{C}$ and $\mathcal{C}^q$.

The *ConstructOSE-SSL* algorithm takes into account only images contained in the database $\mathbf{C}$ and the corresponding semantic information $\mathbf{L}^r$. *ConstructOSE-SSL* is

describe in Algorithm 2.

Algorithm 2 ConstructOSE-SSL

Input: $\mathbf{C}$, $\mathbf{L}^r$
Output: λ^r , $\mathbf{y}^r$

1: Find $A(i,j) = \phi(\mathcal{C}_i^r, \mathcal{C}_j^r)\ \forall i,j \in \{1,\ldots,N\}$.
2: Find W^r by Equation 4.24.
3: Find λ^r , $\mathbf{y}^r$ by Equation 4.23.

Once the eigenvalues λ^r and the embedding locations $\mathbf{y}^r$ have been computed, extrapolation of $\mathcal{C}^q$ into the low dimensional embedding space can be performed via the *ApplyOSE-SSL* algorithm. *ApplyOSE-SSL* is describe in Algorithm 3.

Algorithm 3 ApplyOSE-SSL

Input: C^q, λ^r, $\mathbf{y}^r$
Output: y^q

1: Find $A(i,q) = \phi(\mathcal{C}_i^r, \mathcal{C}^q)$ for all $i \in \{1,\ldots,N\}$.
2: Calculate $W(i,q) = e^{-A(i,q)/\sigma}$.
3: Find y^q by Equation 4.25.

These two algorithms in combination allow for a low dimensional embedding space to be found for $\mathbf{C}$ and $\mathcal{C}^q$.

Application to Image Retrieval

The goal of a CBIR system is to retrieve n images in $\mathbf{C}$ which are most similar to $\mathcal{C}^q$. The application of OSE-SSL to a CBIR system can be applied to learn the metric $\mathbb{D}_{OS}(\mathcal{C}_i^r, \mathcal{C}^q)$ where $\mathbb{D}_{OS}(\mathcal{C}_i^r, \mathcal{C}^q)$ is defined such that smaller values correspond to more similar images.

Offline database construction is an important precursor to image retrieval and is performed using the algorithm *ConstructOSE-SSL*. Online retrieval of the most similar images in $\mathbf{C}$ is performed by the algorithm *RetrieveOSE-SSL*. *RetrieveOSE-SSL* is describe in Algorithm 4.

Algorithm 4 RetrieveOSE-SSL

Input: C^q, $\mathbf{y}^r$
Output: $\mathcal{C}^r$

1: Extrapolation of y^q for $\mathcal{C}^q$ via *ApplyOSE-SSL*.
2: Calculate similarity between $\mathbf{C}$ and $\mathcal{C}^q$ by,

$$\mathbb{D}_{OS}(\mathcal{C}_i^r, \mathcal{C}^q) = ||y_i^r - y^q||. \tag{4.26}$$

3: Sort $\mathbb{D}_{OS}$ from smallest to largest value to give $\mathbf{s}$.
4: Return $\mathcal{C}^r$ corresponding to the smallest n values in $\mathbf{s}$.

OSE-SSL Computational Complexity

To analyze the computational complexity of our novel OSE-SSL algorithm we consider *ConstructOSE-SSL* and *ApplyOSE-SSL* separately. *ConstructOSE-SSL* is a SSL algorithm applied to $\mathbf{C}$. SSL has a computational complexity of $O(N^3)$ due to the eigenvalue decomposition in Equation 4.23 which is the rate limiting step [193]. However as *ConstructOSE-SSL* is utilized only to learn a low dimensional representation of $\mathbf{C}$ it is performed offline prior to image retrieval. *ApplyOSE-SSL* learns y^q for $\mathcal{C}^q$ and hence must be performed online. The computational complexity of OSE is $O(N)$ due to the weighted summation in Equation 4.25 [193].

4.3.3 Experimental Design and Results for Evaluation of Out-of-Sample Extrapolation of Semi-Supervised Learning

We evaluated our *RetrieveOSE-SSL* algorithm on two datasets described in Table 4.6. The synthetic MPEG-7 dataset as described in Section 4.3.3 was selected to demonstrate *RetrieveOSE-SSL* can accurately retrieve images according to morphology. The prostate histopathology dataset as described in Section 4.3.3 was selected to demonstrate the application of *RetrieveOSE-SSL* to retrieving images according to morphology in a relevant clinical context, that of the Gleason grading system. The use of these datasets to determine $\mathbf{C}$ and $\mathcal{C}^q$ and perform CBIR is described in Section 4.3.3. Evaluation measures used to determine the accuracy of *RetrieveOSE-SSL* are described in Section 4.3.3. All code was implemented in MatLab® 2012b and run on a computer with a 3.0 GHz Xeon Quad-Core processor and 16 GB of RAM.

Dataset	Database Size (N)	Number of Classes (Z)	Description
MPEG-7	1400	70	Synthetic silhouette images of various objects.
Prostate	888	3	Prostate histopathology images each containing one prostate gland.

Table 4.6: Description of the datasets used for evaluation OSE-SSL.

MPEG-7 Data Description

The MPEG-7 Core Experiment CE Shape-1 part B was selected as a synthetic dataset to evaluate CBIR. This dataset contains 1400 silhouette images divided equally into 70 classes (20 objects per class) [189]. The dissimilarity metric $\phi(\mathcal{C}_i^r, \mathcal{C}_j^r)$ is calculated by extracting Explicit Shape Descriptors (ESDs) for the silhouette in each image (Section 4.2).

Prostate Histopathology Data Description

Prostate tissue biopsy cores were obtained from 58 patient studies. Each tissue biopsy was stained with Hemotoxylin and Eosin (H & E) and digitized using a ScanScope CS^{TM} whole-slide scanning system at .25 μm per pixel (40$\times$ optical magnification). An expert pathologist selected regions of interests (ROIs) on the digitized biopsy image, for a total of 102 ROIs. The expert pathologist then classified each ROI as benign (BE) (24 ROIs), Gleason grade 3 (G3) (67 ROIs), or Gleason grade 4 (G4) (11 ROIs). Every gland contained within each ROI was segmented by a human expert to obtain lumen and nuclear boundaries, the human expert was blinded to the Gleason grade for all glands. Glands which did not contain either a nuclear or lumen boundary, or where the contour was not fully contained within the ROI were removed from the study, resulting in a total of 888 glands. Glands were distributed across the three classes: Benign ($N = 93$), Gleason grade 3 ($N = 748$), and Gleason grade 4 ($N = 47$).

Dissimilarity between prostate histopathology images is determined according to morphologic similarity between prostate glands on each image. The function $\phi(\mathcal{C}_i^r, \mathcal{C}_j^r)$ is calculated as described in Section 4.2.

Database Construction

For each dataset in Table 4.6 a query image $\mathcal{C}^q$ was selected such that each image in the dataset was selected once. $\mathbf{C}$ was constructed by randomly selecting N images from the dataset in such a way as to always maintain class balance (i.e. stratified sampling of database images). Additionally, the query image $\mathcal{C}^q$ was always excluded from $\mathbf{C}$. Construction of $\mathbf{L}^r$ was performed by randomly selecting N^l labels from the images in $\mathbf{C}$ in such a way as to maintain class balance. Additionally for all experiments $N^l \leq N$, so that the total number of known labels were always less than N. The ability of *RetrieveOSE-SSL* to return images belonging to the same class as $\mathcal{C}^q$ was evaluated as described in Section 4.3.3.

Measure	Description
Silhouette Index (SI)	$\eta^{SI} = \sum_{i=1}^{N} \frac{\nu(i)-\psi(i)}{\max[\psi(i),\nu(i)]}$ where $\psi(i) = \sum_{j,l_j=l_i} \|\|\tilde{y}_i - \tilde{y}_j\|\|$ and $\nu(i) = \sum_{j,l_j \neq l_i} \|\|\tilde{y}_i - \tilde{y}_j\|\|$
Area under the precision recall curve (AUPRC)	Area generated by plotting $p(\alpha)$ versus $r(\alpha)$ where $p(\alpha) = \frac{\Phi(\alpha)}{\alpha}$ and $r(\alpha) = \frac{\Phi(\alpha)}{\Phi(N)}$. $\Phi(\alpha)$ denotes the number of relevant objects in the closest α points.
Bull's Eye	$B = \frac{\Phi(2\times\Phi(N))}{\Phi(N)}$ where $\Phi(N)$ denotes the number of relevant objects in the database.

Table 4.7: Evaluation measures to compare CBIR systems.

Evaluation Measures

OSE-SSL was evaluated on (a) Silhouette Index (SI) of $\mathbf{y}$, a measure of how well images cluster according to class [194], and (b) area under the precision-recall curve (AUPRC) of *RetrieveOSE-SSL*, a description of the behavior of an image retrieval system in terms of how many and in what order relevant images are returned. The synthetic MPEG-7 dataset was evaluated using the Bull's Eye measure, a description of how many of the top returned images are in same class as $\mathcal{C}^q$. Table 4.7 describes all evaluation measures.

Distance Metric	Description
High dimensional metric ($\mathbb{D}_H$)	$\mathbb{D}_H(\mathcal{C}_i^r, \mathcal{C}^q) = \sqrt{\sum_{j=1}^{N}(A(i,j) - A(q,j))^2}$.
Linear low dimensional metric ($\mathbb{D}_{PCA}$)	$\mathbb{D}_{PCA}(\mathcal{C}_i^r, \mathcal{C}^q) = \|\|x_i^r - x^q\|\|_2$, where $x_i^r \in \mathbf{x}$ is obtained by taking the top b principle components from applying PCA to A.
Graph Embedding low dimensional metric ($\mathbb{D}_{GE}$)	$\mathbb{D}_{GE}(\mathcal{C}_i^r, \mathcal{C}^q) = \|\|y_i^r - y^q\|\|_2$, where y_i^r, y^q are obtained by applying Graph Embedding to A as described in Equation 4.23. Note this is equivalent $\mathbb{D}_{OS}$ where $\mathcal{C}^q$ is contained in $\mathbf{C}$.

Table 4.8: Comparative distance metrics utilized to define alternative image similarity measures.

Evaluation Measure	Distance Metric			
	$\mathbb{D}_H$	$\mathbb{D}_{PCA}$	$\mathbb{D}_{GE}$	$\mathbb{D}_{OS}$
Area under the precision recall curve	0.05 ± 0.01	0.023 ± 0.018	0.164 ± 0.130	$\mathbf{0.171 \pm 0.148}$
p-value	8.2×10^{-4}	3.4×10^{-4}	-	0.213
Bull's Eye	69.63 ± 18.83	69.63 ± 18.83	77.46 ± 12.76	$\mathbf{79.74 \pm 9.75}$
p-value	6.5×10^{-3}	6.5×10^{-3}	-	0.56

Table 4.9: Area under the precision recall curve, Silhouette Index and Bull's Eye values for Experiment 1. Values for the best performing metric are bolded. p-values are reported for a Student's t-test to evaluate whether the distance metric $\mathbb{D}_{GE}$ outperformed the distance metrics ($\mathbb{D}_H$, $\mathbb{D}_{PCA}$, or $\mathbb{D}_{OS}$). The null hypothesis is $\mathbb{D}_{GE}$ and the compared distance metric are equivalent.

Experiment 1: Distance Metric for Synthetic MPEG-7 Database

We compared the ability of $\mathbb{D}_{OS}$ to retrieve relevant images with respect to three other distance metrics, $\mathbb{D}_H$, $\mathbb{D}_{PCA}$ and $\mathbb{D}_{GE}$, discussed in Table 4.8. $\mathbb{D}_{GE}$ is a special case of $\mathbb{D}_{OS}$ where $\mathcal{C}^q$ is contained in $\mathbf{C}$ (equivalent to $N^l = 0.0$ and $N = 1.0$), hence, y^q is calculated using Equation 2 for both Graph Embedding and OSE-SSL. For $\mathbb{D}_{OS}$ some labels are known ($N^l = 0.5$) and not all images are contained in the database ($N = 0.9$).

Table 4.9 reports Bull's Eye and area under the precision recall curve for each distance metric. Both $\mathbb{D}_{OS}$ and $\mathbb{D}_{GE}$ outperform $\mathbb{D}_H$ and $\mathbb{D}_{PCA}$ in terms of area under the precision recall curve and Bull's Eye score. A Student's t-test was performed with

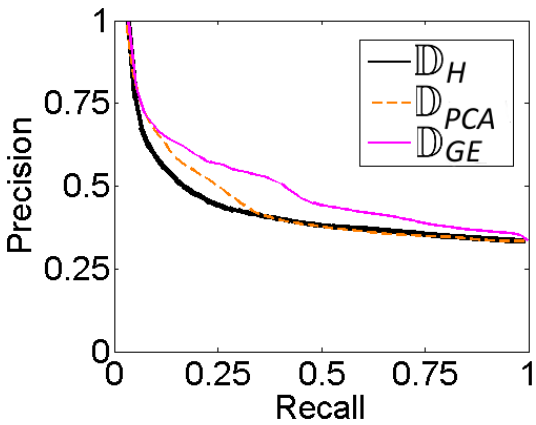


Figure 4.13: Precision-recall curves for Experiment 2 showing retrieval for the metrics: $\mathbb{D}_H$ (black), $\mathbb{D}_{PCA}$ (orange), and $\mathbb{D}_{GE}$ (pink). The precision-recall curves for $\mathbb{D}_H$ and $\mathbb{D}_{PCA}$ perform similarly while $\mathbb{D}_{GE}$ outperforms both.

the null hypothesis was that there were no differences in performance between $\mathbb{D}_{GE}$ and the other distance metric compared. The differences between $\mathbb{D}_{GE}$ and either $\mathbb{D}_H$ or $\mathbb{D}_{PCA}$ are statistically significant ($p < 0.05$) and the null hypothesis is rejected. The differences between $\mathbb{D}_{GE}$ and $\mathbb{D}_{OS}$ are not statistically significant. Additionally, $\mathbb{D}_{OS}$ is comparable to several state of the art shape modeling approaches including curvature scale space and shape context, methods which had Bull's Eye scores in the range of $75 - 80\%$ [119, 151, 195].

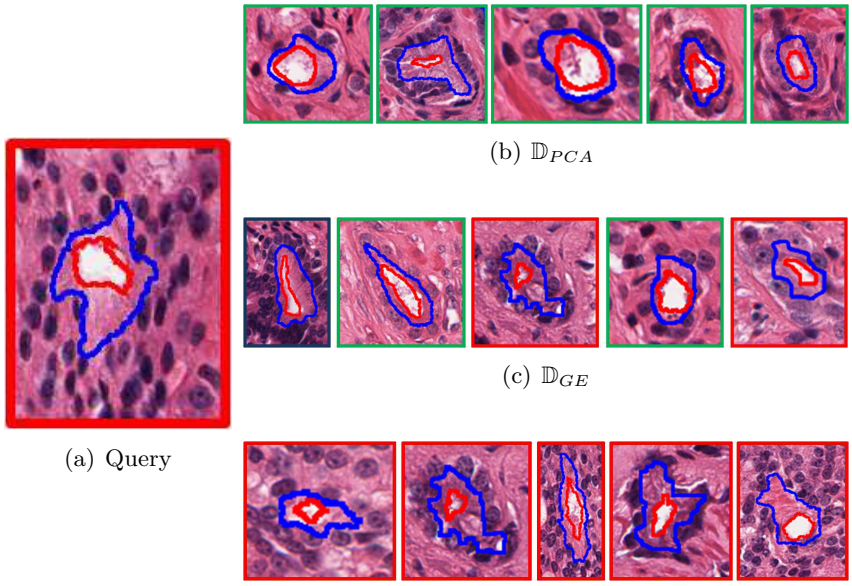
Experiment 2: Distance Metric for Prostate Histopathology Database

In this experiment we evaluated the ability of $\mathbb{D}_{OS}$ to retrieve relevant images for the prostate histopathology dataset. Three other distance metrics, $\mathbb{D}_H$, $\mathbb{D}_{PCA}$, and $\mathbb{D}_{GE}$, discussed in Table 4.8 were used for comparison. $\mathbb{D}_{GE}$ is a special case of $\mathbb{D}_{OS}$ where $\mathcal{C}^q$ is contained in $\mathbf{C}$ (equivalent to $N^l = 0.0$ and $N = 1.0$), hence, y^q is calculated using Equation 2 for both Graph Embedding and OSE-SSL. For $\mathbb{D}_{OS}$ some labels are known ($N^l = 0.5$) and not all images are contained in the database ($N = 0.9$)

In Table 4.10 we report Silhouette Index and area under the precision recall curve for the prostate histopathology database. $\mathbb{D}_{GE}$ results in a higher area under the precision recall curve and Silhouette Index compared to $\mathbb{D}_H$ or $\mathbb{D}_{PCA}$, and these differences are statistically significant ($p < 0.05$). Additionally, increase in Silhouette Index and area under the precision recall curve for $\mathbb{D}_{OS}(N = 0.9, N^l = 0.5)$ over $\mathbb{D}_{GE}$ is statistically significant. In all cases the null hypothesis is that there are no differences between $\mathbb{D}_{GE}$ and the compared distance metrics.

Evaluation Measure	Distance Metric			
	$\mathbb{D}_H$	$\mathbb{D}_{PCA}$	$\mathbb{D}_{GE}$	$\mathbb{D}_{OS}$
Area under the precision recall curve	0.42 ± 0.01	0.44 ± 0.01	0.50 ± 0.01	$\mathbf{0.53 \pm 0.03}$
p-value	9.8×10^{-23}	5.52×10^{-13}	-	6.1×10^{-3}
Silhouette Index	-0.06 ± 0.02	-0.10 ± 0.02	0.08 ± 0.03	$\mathbf{0.14 \pm 0.12}$
p-value	2.01×10^{-14}	2.01×10^{-15}	-	2.86×10^{-6}

Table 4.10: area under the precision recall curve and Silhouette Index values for Experiment 2. Values for the best performing metric are bolded. p-values are reported for a paired Student's t-test to evaluate whether the distance metric $\mathbb{D}_{GE}$ outperformed a comparative distance metric ($\mathbb{D}_H$, $\mathbb{D}_{PCA}$, or $\mathbb{OS}$). The null hypothesis is $\mathbb{D}_{GE}$ and the compared distance metric are equivalent.



(a) Query

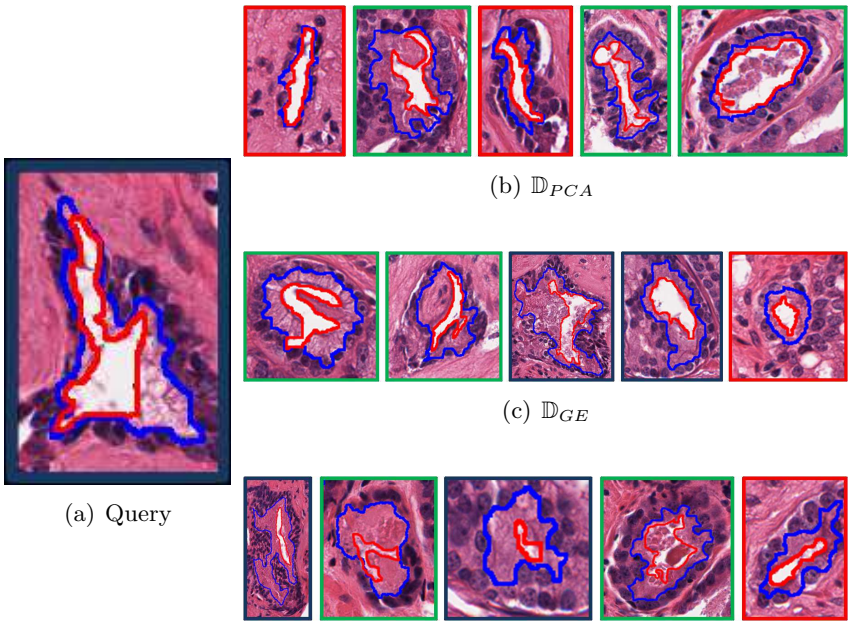
(b) $\mathbb{D}_{PCA}$

(c) $\mathbb{D}_{GE}$

(d) $\mathbb{D}_{OS}(N = 0.9, N^l = 0.5)$

Figure 4.14: (a) Gleason grade 4 query image and top 5 images retrieved (left to right) by (b) $\mathbb{D}_{PCA}$, (c) $\mathbb{D}_{GE}$, and (d) $\mathbb{D}_{OS}$. Retrieved images belonging to the same class as the query image (Gleason grade 4) are outlined in red while those belonging to Gleason grade 3 are in green, and Benign are in blue.

fGr



(a) Query

(b) $\mathbb{D}_{PCA}$

(c) $\mathbb{D}_{GE}$

(d) $\mathbb{D}_{OS}(N = 0.9, N^l = 0.5)$

Figure 4.15: (a) Benign query image and top 5 images retrieved (left to right) by (b) $\mathbb{D}_{PCA}$, (c) $\mathbb{D}_{GE}$, and (d) $\mathbb{D}_{OS}$. Retrieved images belonging to the same class as the query image are outlined in blue (benign) while those belonging to Gleason grade 3 are in green, and Gleason grade 4 are in red.

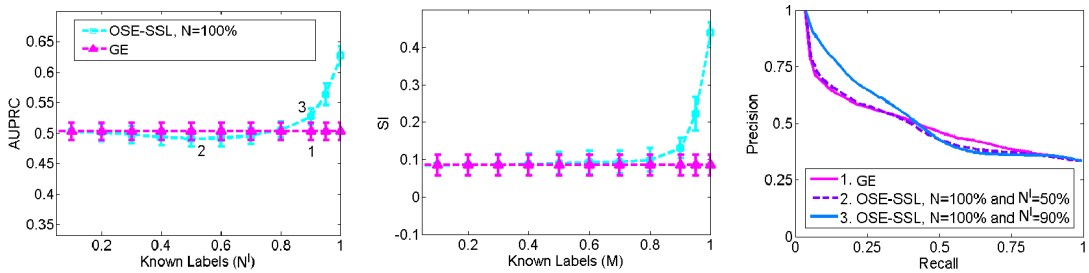
The top 5 retrieved images are shown for $\mathcal{C}^q$ of an example Gleason grade 4 gland in Figure 4.14, where $\mathbb{D}_{OS}(N = 0.9, N^l = 0.5)$ was able to retrieve only glands belonging to the same class as $\mathcal{C}_q$. $\mathbb{D}_{GE}$ also outperforms $\mathbb{D}_{PCA}$ for $\mathcal{C}_q$. Figure 4.15 displays a particularly hard to classify $\mathcal{C}^q$ of a benign gland, and the corresponding top 5 images retrieved. Further evaluation of this gland showed that due to its small size compared to other benign glands, $\Phi(\cdot, \cdot)$ often resulted in a higher than expected dissimilarity between this gland and other BE glands, resulting in retrieving glands belonging to other classes. While $\mathbb{D}_{PCA}$ did not retrieve only glands belonging to the same class in this example $\mathbb{D}_{OS}$ and $\mathbb{D}_{GE}$ were able to retrieve glands belonging to the same class, with $\mathbb{D}_{OS}$ ranking glands belonging to the same class higher compared to $\mathbb{D}_{GE}$.

Experiment 3: Parameter Sensitivity

In this experiment we evaluated the ability of $\mathbb{D}_{OS}$ to retrieve relevant images for the prostate histopathology dataset under for a range of parameter conditions. For $\mathbb{D}_{OS}$ there are two parameters which may be selected by the user, N the number of images contained in $\mathbf{C}$ and N^l the number of labels known for $\mathbf{C}$. Parameters N^l and N were evaluated independently (Sections 4.3.3 and 4.3.3) by holding the parameter not under consideration constant. The defaults for the parameter not under consideration were $N = 1.0$ and $N^l = 0.0$, as already mentioned when $N = 1.0$ and $N^l = 0.0$ the distance metrics $\mathbb{D}_{OS}$ and $\mathbb{D}_{GE}$ are equivalent. The parameters N^l and N were also evaluated together (Section 4.3.3) to explore the synergistic effects of N^l and N on image retrieval.

Effect of Known Label Size

We hypothesized that adding label information via SSL would improve the ability of the low dimensional embedding space to distinguish between images belonging to different classes. Figure 4.16 displays the Silhouette Index and area under the precision recall curve values of the baseline case of no labels (pink) and SSL by varying the number of known labels (light blue). Adding label information improved Silhouette Index and area under the precision recall curve for large N^l.



(a) (b) (c)

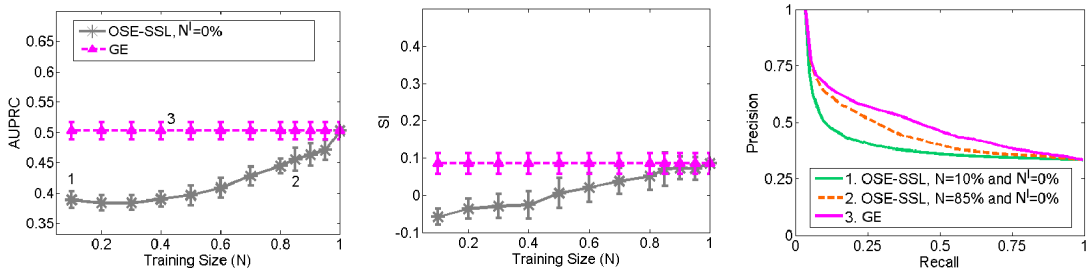
Figure 4.16: Effects of increasing the known labels (N^l) on the prostate histopatholgy database for (a) area under the precision recall curve and (b) Silhouette Index in the low dimensional space obtained via OSE-SSL. The X axis reflects increasing the size of the known labels (N^l) as a function of the percentage of the training set size. The pink line corresponds to the baseline case of $N^l = 0.0$. (c) Three example precision-recall curves for the area under the precision recall curve values are illustrated in (a).

Effect of Database Size

We hypothesized for OSE-SSL small N would be unable to uncover the underlying structure in the database and result in embeddings which are less than optimal. As shown in Figure 4.17, for $N < 0.9$ OSE was unable to accurately extrapolate embeddings. However, for $N \geq 0.9$ there are no statistically significant differences (p-value> 0.05) between embeddings found via OSE and recomputing the eigenvalue decomposition of the similarity matrix (i.e. embeddings found via Graph Embedding).

Relationship between Database Size and Known Label Size

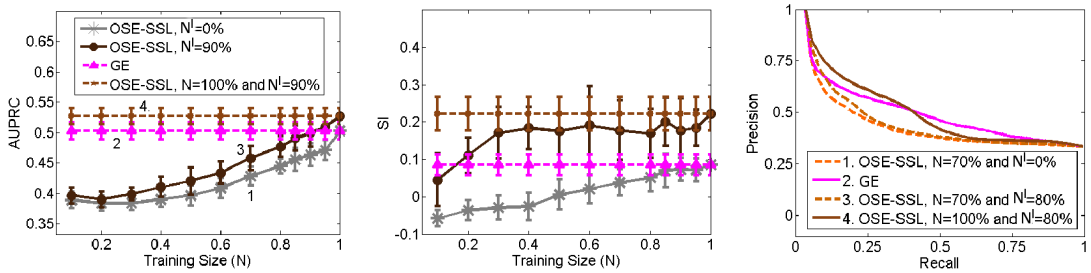
The relationships between the SSL and OSE components of the OSE-SSL were evaluated. Increasing the known labels (N^l) necessitates a concomitant increase in database size (N) to appropriately model the embedding space. This trend is shown in Figure 4.18 where for $N^l = 0.0$ a training set size of $N = 0.9$ is able to appropriately extrapolate embeddings. However, when $N^l = 0.85$ a training set size of $N = 1.0$ is required to appropriately extrapolate embeddings (i.e. Graph Embedding must be utilized to learn the embeddings). In this database, N is not sufficiently high to capture the underlying structure if N^l is increased. Despite not having a large enough N to capture the underlying image structure increasing N^l does result in better area under the precision



(a) (b) (c)

Figure 4.17: Effects of increasing the training set size (N) on the prostate histopatholgy database for (a) area under the precision recall curve and (b) Silhouette Index in the low dimensional space obtained via OSE-SSL. The X axis reflects increasing the size of the training set (N) as a function of the percentage of the total dataset size. The pink line corresponds to the baseline case of $N = 1.0$. (c) Three example precision-recall curves for the area under the precision recall curve values are illustrated in (a).

recall curve and Silhouette Index measures even for small N.



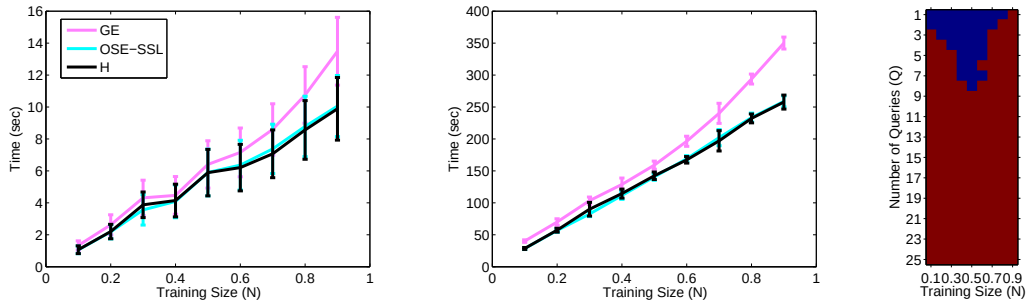
(a) (b) (c)

Figure 4.18: Effects of increasing the training set size (N) in conjunction with the known label (N^l) on the prostate histopatholgy database for (a) area under the precision recall curve and (b) Silhouette Index in the low dimensional space obtained via OSE-SSL. The X axis reflects increasing the size of the training set (N) as a function of the percentage of the total dataset size. Different lines (0 and 0.9 are shown) reflect increasing the size of known labels as a function of the training set size. (c) Four example precision-recall curves for the area under the precision recall curve values are illustrated in (a).

Experiment 4: Computational Time

In this experiment we evaluated the time to retrieve images used the three distance metrics: $\mathbb{D}_H$, $\mathbb{D}_{GE}$, and $\mathbb{D}_{OS}$ to retrieve relevant images for the prostate histopathology dataset using a range of training database sizes (N) and number of query images (N^q).

As shown in Figure 4.19 $\mathbb{D}_H$ and $\mathbb{D}_{OS}$ are able to retrieve images most similar to a query in approximately the same amount of time while $\mathbb{D}_{GE}$ requires more time to perform an equivalent retrieval. Figure 4.19 (c) displays under what conditions the time increases in retrieval for $\mathbb{D}_{GE}$ are statistically significant (red). For larger N and larger N^q, $\mathbb{D}_{GE}$ takes a statistically significant amount of time longer, the higher the values for N and N^q the more pronounced this effect is. The increase in time for $\mathbb{D}_{GE}$ is due to two factors (a) $\mathbb{D}_{GE}$ requires more pairwise comparisons between $\mathcal{C}^q$ and the images contained in $\mathbf{C}$ and (b) $\mathbb{D}_{GE}$ requires a computationally expensive eigenvalue decomposition to compute y^q, the low dimensional embedding for the query image.



(a) (b) (c)

Figure 4.19: Time to retrieve database images most similar to a set of query images using three distance metrics: $\mathbb{D}_H$ (H), $\mathbb{D}_{GE}$ (Graph Embedding), and $\mathbb{D}_{OS}$ (OSE-SSL). The effects of training set size (N) and number of query images (N^q) were evaluated. Retrieval time for (a) $N^q = 1$ and (b) $N^q = 25$ are shown, note the different y-axis scaling for each plot to better highlight the difference between the compared algorithms. In all cases $\mathbb{D}_H$ and $\mathbb{D}_{OS}$ performed the retrieval in similar amounts of time. In contrast $\mathbb{D}_{GE}$ typically required more time to perform retrievals than either $\mathbb{D}_H$ or $\mathbb{D}_{OS}$. (c) Visual representation of when the differences in retrieval time for $\mathbb{D}_{GE}$ and $\mathbb{D}_{OS}$ are statistically significant ($p < 0.01$, red) or not statistically significant ($p > 0.01$, blue).

4.3.4 Concluding Remarks

In this section a novel combination of out-of-sample extrapolation with semi-supervised manifold learning (OSE-SSL) was described. OSE-SSL refines relationships between images in the low dimensional embedding space according to semantic information via SSL and then utilizes OSE to project never before seen images into the low dimensional

space learned via SSL. We have demonstrated the application of OSE-SSL for content-based image retrieval (CBIR) of prostate histpathology. Image similarity within our CBIR framework is defined using ESDs. ESDs are leveraged to determine similarity between images, and then apply the OSE-SSL algorithm to retrieve images which are most similar in a computationally efficient manner.

CBIR for histopathology, has as histopathology images require many image descriptors to accurately describe the large amounts of complex data present. Retrieval directly within the high dimensional feature space for histopathology images is difficult. Manifold learning can be leveraged to find a low dimensional representation where image similarity calculation and retrieval can be performed accurately and efficiently.

Our OSE-SSL CBIR algorithm was evaluated for a prostate histopathology database containing 888 glands and the synthetic MPEG-7 dataset. OSE-SSL outperformed image retrieval in the high dimensional space as well as in a low dimension space found by Principle Component Analysis (PCA). We demonstrated that for each database OSE-SSL was able to accurately retrieve images utilized a low dimensional embedding space found via SSL on a training database that was smaller compared to the full dataset. For the MPEG-7 dataset OSE-SSL is able to retrieve images with a Bull's eye value of $79.74 \pm 9.75\%$. For the prostate histopathology dataset $N^l = 0.85$ of the dataset, or 754 images, was required to achieve retrieval rates comparable to those achieved by performing an eigenvalue decomposition for each new query image. Finally, incorporating known label information was able to improve retrieval rates.

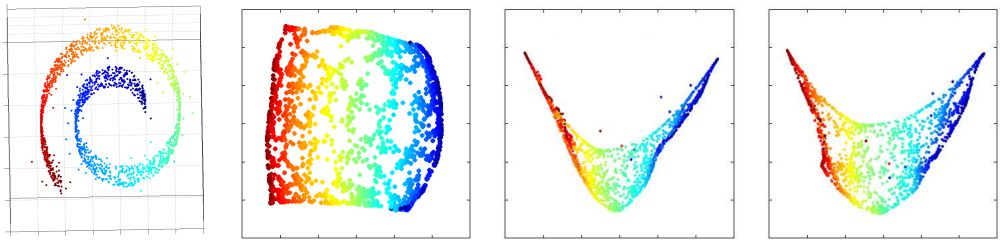
The current work is limited in that CBIR was performed on a per patch basis, where multiple patches are defined over a single slide. However, pathologists typically utilize the whole slide to determine Gleason grade. Additionally, in this work we have leveraged only gland morphology to determine similarity between image patches. However, pathologists typically evaluate Gleason grade using the morphology and arrangement of glands and nuclei [44]. Future work will involve incorporating our gland based retrieval into a whole slide similarity metric, which will be capable of retrieving whole slides which contain similar image characteristics, likely including measures of nuclei arrangement [167] and nuclei morphology [50].

The current work is also limited by the fact that all 58 patients had prostate tissue biopsy cores acquired at a single institution. Therefore, the dataset used in this work may be more homogeneous, in terms of tissue staining and digitization of the slides, compared to a dataset of prostate histopathology images acquired across several institutions. While these differences between institutions will likely affect pre-processing steps such as automated segmentation, in this work we have limited the effects of a homogeneous dataset by relying on manual segmentation. The variability in gland morphology is independent of institution, as gland morphology is a function of disease grade. Future work will evaluate the presented methodology on a larger patient cohort acquired across institutions.

Additionally, the current work only evaluated morphologic features (ESDs) of glands present on prostate histopathology. Previous work has shown that texture [172, 175] and nuclear architecture [169, 172, 175] are also able to provide accurate image retrieval of prostate histopathology. The OSE-SSL algorithm is not limited to ESDs, hence, alternative dissimilarity measures that combine ESDs with other features derived from the prostate histopathology images can be implemented within our CBIR framework. Future work will evaluate the best way to implement a dissimilarity measure that combines multiple image features.

4.4 Statistical Shape Model for Manifold Learning (SSMM) for Robust Explicit Shape Descriptors

As previously described Explicit Shape Descriptors (ESD) use Graph Embedding, a manifold learning algorithm, to take the high dimensional feature space A, obtained by calculating differences between MASMs, and reduce A to the lower dimensional space $\mathbf{y}$ which is better suited to classification. However there are several manifold learning algorithms to calculate a low dimensional embedding space that can preserve subtle relationships between samples in the high dimensional space [155, 165, 166]. Manifold learning schemes tend to be sensitive to the dataset considered, and changes in the dataset may cause changes to the learned manifold [196].



(a) (b) (c) (d)

Figure 4.20: (a) Original $3D$ Swiss Roll dataset with Gaussian noise added to 2% of samples in the datasest. (b) $2D$ manifold $\mathcal{M}$ in the absence of noise. This manifold structure best preserves the relationships between samples in the original high dimensional space. (c) Manifold $\hat{\mathcal{M}}$ found by applying Graph Embedding to a dataset containing noise and (d) the manifold $\tilde{\mathcal{M}}$ found by regularization of $\hat{\mathcal{M}}$ using the SSMM.

To formalized the problem, manifold learning algorithms find a low dimensional manifold representation $\mathcal{M}$ from a dataset of images $\mathbf{C}$ which preserves relationships between images in $\mathbf{C}$. Most manifold learning algorithms assume that $\mathbf{C}$ is contained in a high dimensional space $\mathbb{R}^D$ [155, 165, 166, 190]. Additionally, manifold learning techniques assume that $\mathbf{C}$ is densely clustered and concentrated within a small region of $\mathbb{R}^D$. An example of dense clustering can be seen in Figure 4.20(a) which shows an example of the synthetic Swiss Roll dataset. In $\mathbb{R}^3$ the samples cluster along a $2D$ planar structure.

Manifold learning algorithms first construct a dissimilarity matrix A which quantifies dissimilarity between samples in $\mathbb{R}^D$ [155, 165, 166]. For a dataset $\mathbf{c}$ containing N samples, A is a $N \times N$ dissimilarity matrix defined such that $A(\mathcal{C}_i, \mathcal{C}_j)$ quantifies the differences between the samples $\mathcal{C}_i, \mathcal{C}_j \in \mathbf{C}$. Typically $A(\mathcal{C}_i, \mathcal{C}_j) = \phi(\mathcal{C}_i, \mathcal{C}_j)$ where $\phi(\cdot, \cdot)$ is a dissimilarity measure (e.g. heat kernel [106, 197], geodesic distance [166], ESDs as described in Section 4.2) which is dataset and feature set dependent. Manifold learning algorithms then calculate $\mathcal{M}$ to best preserve the relationships in A. Manifold learning techniques preserve relationships in A differently, some methods such as Local Linear Embedding (LLE) [165] try to preserve the local neighborhood relationships between samples. Isomaps [166] and Graph Embedding (GE) [155] find the best embedding space to preserve the global structure of A, albeit with different underlying algorithms.

Manifold learning schemes tend to be sensitive to the dataset considered, and changes in the dataset may cause changes to the learned manifold [196]. Consider a sample $\mathcal{C}_i \in \mathbf{C}$ perturbed by some error ϵ; the new location for $\mathcal{C}_i$ would be $\hat{\mathcal{C}}_i = \mathcal{C}_i + \epsilon$. A would have to be altered such that $\hat{A}(\hat{\mathcal{C}}_i, \mathcal{C}_j) = \phi(\hat{\mathcal{C}}_i, \mathcal{C}_j)$ for all $\mathcal{C}_j$ contained in $\mathbf{C}$, resulting in changes to $2N - 2$ elements in A. The manifold $\hat{\mathcal{M}}$ learned from $\hat{A}$ will reflect those changes. Hence even a small change in $\mathbf{C}$ may cause large changes to $\mathcal{M}$. Figure 4.21 demonstrates this phenomenon for the prostate histology dataset comprising 888 glands. Two manifolds were generated by applying the ESDs to 90% of samples in the dataset (800 glands) such that for each manifold a different set of 88 samples were excluded. Each manifold has a distinct structure evident by (a) changes in the planar structure of the manifold and (b) changes in object-class relationships on the manifold, displayed as color differences between manifolds.

Consider a large dataset $\mathbf{C}$ from which the manifold $\mathcal{M}$ is generated. In the absence of knowing the true manifold, $\mathcal{M}$ is the best manifold representation to capture the relationships between samples in the dataset. If we consider a subset $\hat{\mathbf{C}} \subset \mathbf{C}$ then $\hat{\mathbf{C}}$ can be used to create an alternative manifold $\hat{\mathcal{M}}$ which approximates $\mathcal{M}$. Manifold regularization constrains the structure of $\hat{\mathcal{M}}$ giving a better approximation of $\mathcal{M}$ and hence resulting in a better representation of the relationships between samples in $\hat{\mathbf{C}}$.

To overcome the sensitivity of manifold learning algorithms to noise we present a statistical shape model of manifolds (SSMMs) to perform manifold regularization. SSMM merges the theory of ensemble learning [198] with statistical shape models (SSMs) [199]. The theory behind ensemble learning is that an ensemble of weak classifiers has higher accuracy compared to any single weak classifier [200]. Similarly, consensus clustering takes a ensemble of weak clusterings of a dataset, obtained by applying an unstable clustering method such as K-mean clustering to a dataset multiple times, and combines the ensemble to generate a strong clustering of the dataset [201]. Viswanath *et. al.* [54] demonstrated that consensus embedding, obtained by generating an ensemble of manifolds from a single dataset, produced a more stable low dimensional manifold compared to any single manifold in the ensemble. In this work we hypothesize an ensemble of manifolds will have a more accurate representation of the manifold shape

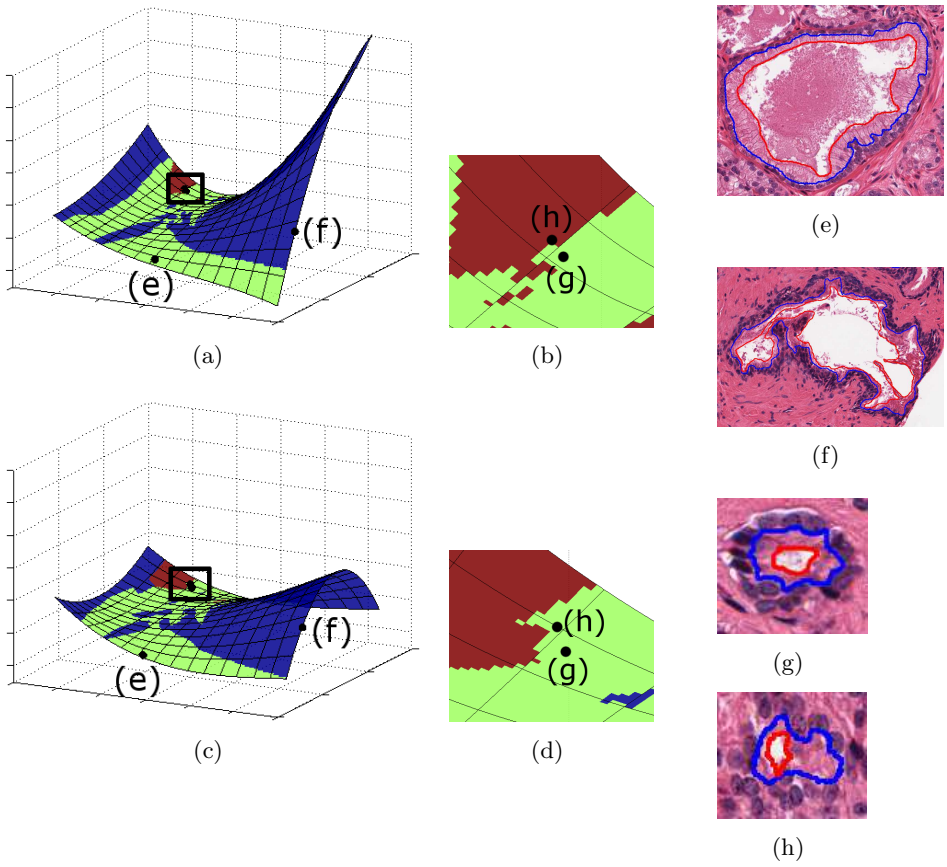


Figure 4.21: (a), (c) Two manifolds $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_2$ generated by performing Graph Embedding, a manifold learning scheme, on quantitative morphologic features extracted from 800 prostate histopathology glands. The manifolds $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_2$ were generated from two distinct datasets $\mathbf{C}_1$ and $\mathbf{C}_2$ such that 88 glands excluded from either $\mathbf{C}_1$ or $\mathbf{C}_2$. (b), (d) Manifold region enclosed by the black box in (a) and (c) respectively. Representative glands for (f) benign (benign), (e), (g) Gleason grade 3, and (h) Gleason grade 4 (G4) classes. A classifier trained in the reduced dimensional space allows for assignment of a single class to each region on the manifold, such that blue regions correspond to benign, green regions correspond to Gleason grade 3, and red regions correspond to Gleason grade 4. Differences between the manifolds can be seen in changes in global structure as well as class-object relationships on the manifold, which are evident by changes in region color. In the case of (h) a representative Gleason grade 4 gland, in one manifold ((c)) the gland was incorrectly projected on to the Gleason grade 3 class region.

that any single manifold. The concept of SSMMs is that an ensemble of manifolds can be modeled with a SSM. SSMs have been proposed to model shape variation in anatomical structures [199]. In much the same way, we utilize a SSM to model which manifold shapes are statistically most likely to occur. The SSMM describes the maximum likelihood estimate of the manifold shape and primary modes of variation for a series of different manifolds constructed by randomly selecting a subset of samples from a dataset. For a new, related dataset, the resulting manifold can be constrained to only the range of shapes dictated by the SSMM. Hence every sample on the new manifold is spatially and locally constrained to within 2 standard deviations of its location on the average manifold shape.

The SSMM can be utilized in several ways. (1) Regions on a new, related manifold which deviate from the SSMM can be identified. By identifying these regions, meaningful differences between the dataset and the SSMM may be determined. (2) Noisy samples on a manifold can be identified based on their deviation from the SSMM. Removing these samples from the dataset may result in a more accurate low dimensional manifold, and hence improve classification accuracy. (3) A classifier can be trained on the SSMM which would allow for (a) classifier decision boundaries to be applied to a new, related manifold without retraining the classifier or (b) new, related samples could be projection onto the SSMM. The projection of newly acquired samples onto a previously calculated manifold can be performed by out-of-sample extrapolation (OSE) [188].

4.4.1 Previous Work in Manifold Regularization

Manifold learning is well known to be sensitive to the dataset considered, as well as noise and outliers contained within a dataset [196, 197]. Perturbations in the manifold structure may reduce classification performance in the low dimensional embedding space as object-class relationships may be obscured. Manifold regularization techniques have been proposed which impose additional constraints on manifold learning to better preserve object-class relationships in the low dimensional space. For instance, Chang *et. al.* [196] proposed a weighted manifold learning scheme, where outlier samples were

assigned low weights, to reduce the effect outliers have on learning the manifold. Other manifold regularizers perform local smoothing on the learned manifold [202]. Manifold regularization techniques may add a smoothness constraint into the manifold learning algorithm [197, 203]. All of these methods over smooth the manifold, as they reduce the effects of outliers which including meaningful information as well as noise.

Another type of regularization learns a consensus embedding from a set of manifolds. Hou *et. al.* [204] learned a set of manifolds by obtaining multiple views for each sample in the dataset and then generated a consensus manifold across the views. Other consensus embedding schemes have varied the parameters or samples considered to find a manifold set, and then generated a consensus embedding from the set [205, 206]. These methods rely on the manifolds in the set being independent, which may not be a valid assumption when generating manifold sets across manifold learning parameters. Additionally, relationships between samples across the individual manifolds are not taken into account when determining a consensus embedding .

4.4.2 Brief Overview and Novel Contributions of Statistical Shape Model of Manifolds (SSMM)

A flowchart of the proposed SSMM methodology is displayed in Figure 4.22. Table 4.11 list the notation used throughout the paper. To construct the SSMM we (1) generate a set of manifolds $\mathbb{M}$ for a dataset $\mathbf{C}$. For this task we divide the dataset $\mathbf{C}$ into K folds, and then generate $\mathbb{M}$ using a leave-one-fold-out scheme. (2) As manifolds in $\mathbb{M}$ will be misaligned, primarily due to rotational and translational differences, a Procrustes based registration scheme aligns all the manifolds in $\mathbb{M}$. (3) Calculate the maximum likelihood estimator and primary modes of variation for $\mathbb{M}$. Once constructed the SSMM constrains a new manifold instance $\dot{\mathcal{M}}$ of related datasets to only those shapes statistically most likely to occur resulting in the regularized manifold $\tilde{\mathcal{M}}$. In this work we demonstrate that the SSMM can (a) determine noisy samples by identifying samples which deviate from the SSMM, and (b) accurately perform OSE of newly acquired samples onto a manifold constrained by the SSMM.

The novel contributions of the SSMM are:

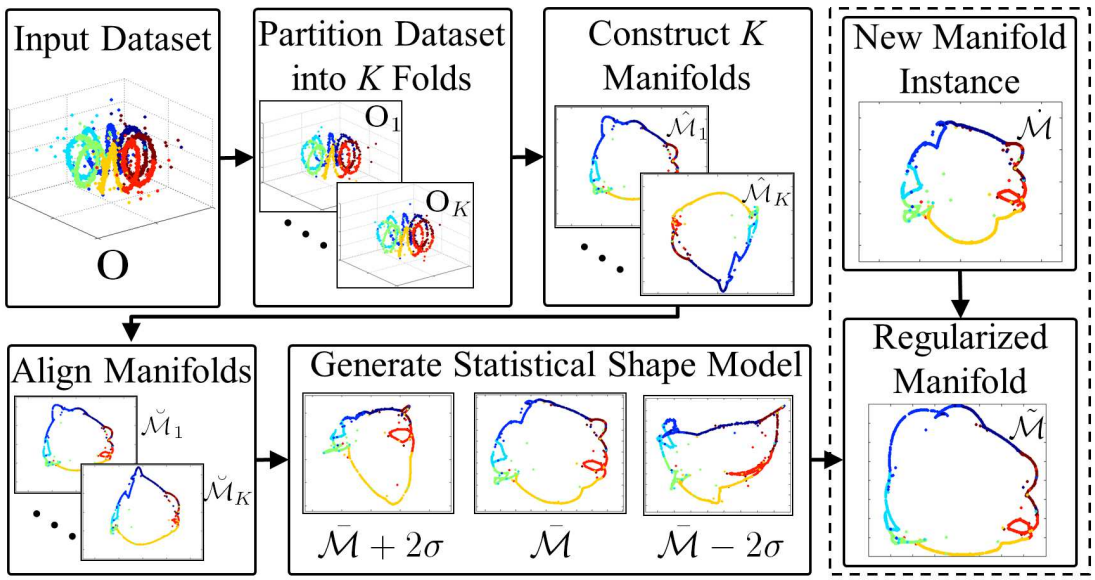


Figure 4.22: Flowchart which describes the construction of the SSMM and its application to manifold regularization for the synthetic Helix dataset. SSMM construction consists of dividing the dataset $\mathbf{C}$ into K folds, denoted as $\{\mathbf{C}_1, \ldots, \mathbf{C}_K\}$. The K folds of $\mathbf{C}$ are utilized to find the manifold set $\mathbb{M} = \{\hat{\mathcal{M}}_1, \ldots, \hat{\mathcal{M}}_K\}$. The manifolds in $\mathbb{M}$ are then aligned via Procrustes based registration scheme resulting in the aligned manifold set $\breve{\mathbb{M}} = \{\breve{\mathcal{M}}_1, \ldots, \breve{\mathcal{M}}_K\}$. The SSMM finds the maximum likelihood estimator ($\bar{\mathcal{M}}$) and primary modes of variation for $\breve{\mathbb{M}}$. Shown are the modes of variation corresponding to the statistical extremes of the model $\bar{\mathcal{M}} - 2\sigma$ and $\bar{\mathcal{M}} + 2\sigma$. Given a new manifold instance $\dot{\mathcal{M}}$ the SSMM constrains the structure to only those statistically likely to occur ($\bar{\mathcal{M}} \pm 2\sigma$). This results in the regularized manifold $\tilde{\mathcal{M}}$ which is a better approximation of the underlying relationships in $\mathbf{C}$ than any constituent manifold in $\mathbb{M}$. For the synthetic Helix dataset shown in this flowchart the ideal manifold is a $2D$ circular structure.

Symbol	Description	Symbol	Description
$\mathbf{C}$	Image database	$\|\|\cdot\|\|$	L2 norm
$\mathcal{C}_i, \mathcal{C}_j$	Images contained in $\mathbf{C}$	$\breve{\mathbb{M}}$	Aligned manifold set
$\mathbf{C}_k$	kth fold of $\mathbf{C}$ for $k \in \{1, \ldots, K\}$	$\bar{\mathcal{M}}$	Maximum likelihood estimate of $\breve{\mathbb{M}}$
$\mathbb{R}^D$	High dimensional feature space	V	Primary modes of variation for $\breve{\mathbb{M}}$
$\phi(\cdot, \cdot)$	Dissimilarity measure	$\dot{\mathcal{M}}$	New manifold instance
A	Dissimilarity matrix defined as $\phi(\mathcal{C}_i, \mathcal{C}_j)$ evaluated for all $\mathcal{C}_i, \mathcal{C}_j \in \mathbf{c}$	$\breve{\mathcal{M}}$	Manifold constrained via the SSMM
γ	Graph Embedding scaling term	O	New image instance
$\mathbb{R}^n$	Low dimensional embedding space	O^e	Samples which contain noise in O
$\mathbb{M}$	Ensemble manifold set	O^c	Samples which do not contain noise in O
$\hat{\mathcal{M}}_k$	kth manifold in $\mathbb{M}$	τ	Threshold to determine sample deviation from the SSMM
$\hat{\mathbf{y}}_k$	Embedding locations on $\hat{\mathcal{M}}_k$	$\mathcal{M}^c$	Manifold generated from O^c
$y_{i,k}$	Embedding location for $\mathcal{C}_i$ on $\hat{\mathcal{M}}_k$	O^{te}	Testing samples not contained in O
	Transformation to align $\hat{\mathcal{M}}_b$ to $\hat{\mathcal{M}}_a$	$\mathcal{M}^{te,c}$	Manifold with samples in O^{te} projected onto $\mathcal{M}^c$.

Table 4.11: Notation used to describe Statistical Shape Model of Manifolds (SSMM).

- A computerized decision support system which utilizes a SSMM based on the morphologic features of glands on prostate histopathology to automatically distinguish between Gleason grades 3 and 4.
- A novel combination of SSMs and ensemble learning theory to generate a more accurate manifold representation of $\mathbf{C}$.
- A novel method to generate $\mathbb{M}$, an set of K manifolds, containing all samples in $\mathbf{C}$.
- A novel manifold registration to align all manifolds in $\mathbb{M}$. As each sample $\mathcal{C}_i \in \mathbf{C}$ has a corresponding embedding location $y_{i,k}$ on the manifold $\mathcal{M}_k$, the registration algorithm minimizes the differences between $y_{i,k}$ for all $k \in \{1, \ldots, K\}$ and all $\mathcal{C}_i \in \mathbf{C}$ via Procrustes registration [207].

4.4.3 Statistical Shape Model of Manifolds Theory

In this section we prove theoretically that SSMMs are appropriate to determine the maximum likelihood estimator of a manifold shape. Specifically, we prove that constructing a SSMM from a set of manifolds is guaranteed to represent the underlying manifold structure at least as well as any manifold contained in the set. To perform the theoretical analysis we utilize the theory of ensemble learning [198].

A dataset of N images is defined as $\mathbf{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_N\}$. A sample $\mathcal{C}_i \in \mathbf{C}$ is defined as a point in a D-dimensional space $\mathbb{R}^D$.

Definition 1. *A true manifold $\mathcal{M} \in \mathbb{R}^n$ is defined by a set of N true embedding locations $\mathcal{M} = \{y_1, \ldots, y_N\}$. Each true embedding location $y_i \in \mathbb{R}^n$ corresponds to a sample $\mathcal{C}_i \in \mathbb{R}^D$ where $n << D$.*

Definition 2. *A manifold $\hat{\mathcal{M}}$ estimates $\mathcal{M}$ by a set of N embedding locations $\hat{\mathcal{M}} = \{\hat{y}_1, \ldots, \hat{y}_N\}$. Each embedding location $\hat{y}_i \in \mathbb{R}^n$ corresponds to a sample $\mathcal{C}_i \in \mathbb{R}^D$ where $n << D$.*

Definition 3. *The manifold $\hat{\mathcal{M}}$ approximates $\mathcal{M}$ with an error $\epsilon_{\hat{\mathcal{M}}}$ given as,*

$$\epsilon_{\hat{\mathcal{M}}} = E_i\big(\hat{y}_i - y_i\big)^2, \tag{4.27}$$

where E_i is the maximum likelihood expectation over $i = \{1, \ldots, N\}$.

Proposition 1. *Given a set of K independent, identically distributed manifolds $\mathbb{M} = \{\hat{\mathcal{M}}_1, \ldots, \hat{\mathcal{M}}_K\}$, a manifold $\bar{\mathcal{M}}$ exists such that $\bar{\mathcal{M}} \to \mathcal{M}$ as $K \to \infty$.*

Proof. Each estimated manifold $\hat{\mathcal{M}}_k$ is defined by the embedding locations $\hat{y}_{i,k} \in \hat{\mathcal{M}}_k : i \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}$. Assuming that each embedding location $\hat{y}_{i,k} \in \hat{\mathcal{M}}_k : i \in \{1, \ldots, N\}$, $k \in \{1, \ldots, K\}$ is normally distributed about y_i, the Central Limit Theorem states,

$$\lim_{K \to \infty} \Bigg(E_k(\hat{y}_{i,k}) - y_i \Bigg) = 0, \tag{4.28}$$

where E_k is the maximum likelihood expectation over $k = \{1, \ldots, K\}$. Therefore $\bar{\mathcal{M}}$ exists and is defined $\bar{\mathcal{M}} = E_k(\hat{y}_{i,k}) : i \in \{1, \ldots, N\}$.

The error between $\bar{\mathcal{M}}$ and $\mathcal{M}$ is defined as (similar to Equation 4.27),

$$\epsilon_{\bar{\mathcal{M}}} = E_i\big(\bar{y}_i - y_i\big)^2, \tag{4.29}$$

where $\bar{xy}_i = E_k(\hat{y}_{i,k})$. From Equation 4.27 the error over K embeddings is given as,

$$\epsilon_{K,\hat{\mathcal{M}}} = E_k\big[\epsilon_{\hat{\mathcal{M}}}\big] = E_k\big[E_i\big(\hat{y}_{i,k} - y_i\big)^2\big]. \tag{4.30}$$

Proposition 2. *Given K independent, identically distributed manifolds, $\hat{\mathcal{M}}_k \in \mathbb{M}$, $\epsilon_{K,\hat{\mathcal{M}}} \geq \epsilon_{\bar{\mathcal{M}}}$.*

Proof. Comparing Equation 4.30 and Equation 4.29 in a manner analogous to Bagging [198] gives the proof. In Bagging, an ensemble classifier is constructed from a set of weak classifiers. Similarly, the $\bar{\mathcal{M}}$ obtained from the SSMM can be viewed as an ensemble embedding constructed from a set of weak embeddings $\mathbb{M}$. Hence the result follows.

4.4.4 Construction of Statistical Shape Manifold Model (SSMM)

In this section we present our methodology for constructing a SSMM. Concepts discuss in Section 4.3.2 are leveraged for construction of the SSMM. We then describe our novel

K fold algorithm to calculate $\mathbb{M}$ in Section 4.4.4. Section 4.4.4 presents the Procrustes based registration of manifolds in $\mathbb{M}$. In Section 4.4.4 we discuss SSMM construction and Section 4.4.4 describes the fitting of $\dot{\mathcal{M}}$ to the SSMM. Finally we present two novel algorithms for (a) noise identification and removal in Section 4.4.4 and (b) OSE of newly acquired samples onto the SSMM in Section 4.4.4.

Construction of the Manifold Set

A set of K manifolds $\mathbb{M} = \{\hat{\mathcal{M}}_1, \ldots, \hat{\mathcal{M}}_K\}$ are obtained from a dataset of N images defined as $\mathbf{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_N\}$. $\mathbb{M}$ is generated utilizing a K fold scheme via the following steps:

1. Samples in $\mathbf{C}$ are randomly divided into K equal partitions such that $\mathbf{C} = \{\mathbf{C}_1 \cup \ldots \cup \mathbf{C}_K\}$.

2. Testing and training sets are obtained via a leave one fold out scheme. A testing set is defined as $\mathbf{C}_k^t = \mathbf{O}_k : k \in \{1, \ldots, K\}$ and the corresponding training set is defined as $\mathbf{C}_k^r \cup \mathbf{C}_k^t = \mathbf{C}$.

3. Each training set $\mathbf{C}_k^r$ is utilized to find $\hat{\mathbf{y}}_k$ which defines $\hat{\mathcal{M}}_k$ via GE as described in Section 4.3.2. The samples in $\mathbf{C}_k^r$ are then used to determine the structure of the manifold $\hat{\mathcal{M}}_k$.

4. Each test set $\mathbf{C}_k^t$ is extrapolated into the manifold $\hat{\mathcal{M}}_k$ to determine $\tilde{\mathbf{y}}_k$ via out-of-sample extrapolation as described in Section 4.3.2.

5. Training and testing sets are combined to determine $\mathbf{y}_k = \{\hat{\mathbf{y}}_k, \tilde{\mathbf{y}}_k\}$. This combination allows for point correspondence between manifolds in $\mathbb{M}$ to be estimated.

In this work $K = 10$ was chosen to construct $\mathbb{M}$, and the steps described above were performed 5 times for a total of 50 constituent manifolds in $\mathbb{M}$. Graph Embedding was chosen for experiments showcased in this work, but it is worth noting any manifold learning [155, 165, 166, 190] algorithm can be used to construct $\mathbb{M}$.

Manifold Alignment via Procrustes Based Registration

Manifolds contained in $\mathbb{M}$ may not align, the algorithm for manifold learning preserves pairwise relationships between samples but may not preserve the global relationship of samples in the low dimensional embedding space. Procrustes registration is applied to align all manifolds in $\mathbb{M}$ [207]. Procrustes registration can be performed since there are point correspondences between all manifolds in $\mathbb{M}$ as each sample in $\mathbf{O}$ has a location on every manifold in $\mathbb{M}$.

A reference manifold $\hat{\mathcal{M}}_a : a \in \{1, \ldots, K\}$ is randomly selected. All other manifolds, $\hat{\mathcal{M}}_b : b \neq a$ are registered to $\hat{\mathcal{M}}_a$ by minimizing,

$$\breve{\mathcal{M}}_b = \min_{T^{\mathcal{M}}} \left(\sum_i ||y_{i,a} - T^{\mathcal{M}}(y_{i,b})|| \right), \tag{4.31}$$

where $y_{i,a}$ is a embedding location in $\mathcal{M}_a$ for an image $\mathcal{C}_i$ and $y_{i,b}$ is a embedding location in $\mathcal{M}_b$ for an image $\mathcal{C}_i$. The transform $T^{\mathcal{M}}$ selected was a rigid transform, to take into account scale and rotational differences between $\hat{\mathcal{M}}_a$ and $\hat{\mathcal{M}}_b$. $|| \cdot ||$ denotes the L2-norm. Registration is performed for all $\hat{\mathcal{M}} \in \mathbb{M}$ to obtain the aligned set of manifolds $\breve{\mathbb{M}}$.

Statistical Shape Manifold Model (SSMM)

Once all manifolds are aligned the statistical properties of the manifold set can be determined. The SSMM is defined via the mean and principal modes of variation for $\breve{\mathbb{M}}$. The mean of $\breve{\mathbb{M}}$ is calculated by,

$$\bar{\mathcal{M}} = \tfrac{1}{K} \sum_k \breve{y}_{i,k} : \forall \breve{y}_{i,k} \in \breve{\mathcal{M}}_k. \tag{4.32}$$

The principal modes of variation for the manifold defined as V are obtained by performing PCA on $\breve{\mathbb{M}}$ [199].

Only the V corresponding to the top 95% of variance in the sample locations $y_{i,k}$ for all $k \in \{1, \ldots, K\}$ are retrained to constrain the SSMM to those shapes within 2 standard deviations of the mean shape.

Constraining a New Manifold Instance to the SSMM

A new manifold $\dot{\mathcal{M}}$ is obtained by applying Graph Embedding to $\mathbf{C}$. $\dot{\mathcal{M}}$ is constrained to only likely shapes as defined by the SSMM obtained in Section 4.4.4.

$$\tilde{\mathcal{M}} = T^r_{K+1}\big(\bar{\mathcal{M}} + V * \upsilon\big), \tag{4.33}$$

where υ controls the shape of $\tilde{\mathcal{M}}$ and T^r_{K+1} is a rigid transformation between the SSMM and $\dot{\mathcal{M}}$. υ is found via a linear least squares fit between the SSMM and $\dot{\mathcal{M}}$ and is constrained to $\bar{\mathcal{M}} \pm 2\sigma$ to limit the SSMM to only those shapes statistically most likely to occur [199].

Application of SSMM to Identify Noisy Samples

The SSMM can aid in the identification of samples which contain noise. The algorithm *FilterManifold* assumes samples which contain noise are those samples which deviate most from the SSMM.

A dataset contains N samples defined as $O = \{o_1, \ldots, o_N\}$. Algorithm 5 can be used to identify the samples which contain noise O^n and the samples which do not contain noise O^c within O given a user defined threshold τ.

The value assigned to τ is dataset specific as sample variation across datasets may vary. In this work τ was chosen such that 5% of the samples in the dataset were excluded.

Algorithm 5 FilterManifold

Input: O, τ
Output: $\mathcal{M}^c$

1: Obtain $\tilde{\mathcal{M}}$ from O via application of the SSMM.
2: Obtain $\hat{\mathcal{M}}$ from O by GE (Eq. 4.23).
3: Calculate $e(o_i) = ||\hat{y}_i - \tilde{y}_i||$.
4: Obtain $O^n = o_i : o_i \in O, e(o_i) \geq \tau$.
5: Obtain $O^c : O^c \cap O^n = \emptyset$.
6: Obtain $\mathcal{M}^c$ for O^c via Graph Embedding (Eq. 4.23)

Application of SSMM to OSE

The SSMM can be utilized for robust OSE, by generating a more accurate manifold representation of a dataset. The algorithm *OSE-SSMM* demonstrates how the SSMM can be used for this purpose.

A dataset O is divided into training samples O^{tr} and testing samples O^{te} such that $O^{tr} \cap O^{te} = \emptyset$. To find a set of testing embeddings $\mathcal{M}^{te,c}$ for a filtered manifold we apply algorithm 6

Algorithm 6 OSE-SSMM

Input: O_{tr}, O_{te}, τ
Output: $\mathcal{M}^{te,c}$

1: Obtain $\mathcal{M}^{tr,c}$ for O^{tr} via *FilterManifold*.
2: Obtain $\mathcal{M}^{te,c}$ for O^{te} via NM (Eq. 4.25) with $\mathcal{M}_c^{tr}$ as the training manifold.

4.4.5 Experimental Design and Results for Evaluation of Statistical Shape Model of Manifolds (SSMM)

Dataset Description

Dataset	Sample Size	Dissimilarity Measure
Synthetic Swiss Roll [166]	3000	$A(\mathcal{C}_i, \mathcal{C}_j) \begin{cases} \|\|\mathcal{C}_i - \mathcal{C}_j\|\| & \text{if } \|\|\mathcal{C}_i - \mathcal{C}_j\|\| < \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases}$ $\mathcal{N}$ is a neighborhood parameter.
Synthetic Helix [166]	3000	$A(\mathcal{C}_i, \mathcal{C}_j) \begin{cases} \|\|\mathcal{C}_i - \mathcal{C}_j\|\| & \text{if } \|\|\mathcal{C}_i - \mathcal{C}_j\|\| < \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases}$ $\mathcal{N}$ is a neighborhood parameter.
Prostate Histology	888 (58 patients)	ESDs (Section 4.2)

Table 4.12: Description of datasets and their dissimilarity measures.

Synthetic Datasets

Two synthetic datasets, Swiss Roll and Helix, described in Table 4.12 were utilized to demonstrate the application of SSMM to manifold regularization. The Swiss Roll is a

$2D$ planar manifold divided into two classes which exists in a $3D$ space. The Helix is a $1D$ circular manifold divided into six classes which exists in a $3D$ space. The benefit of both datasets is that the high dimensional $3D$ space and the low dimensional $2D$ embedding space may be visualized.

Gaussian noise was added to 5% of samples within each dataset where the standard deviation of the noise was set equal to 15% of the standard deviation of samples in the dataset. The dissimilarity measures for both datasets are reported in Table 4.12.

Prostate Histopathology

Prostate needle core tissue biopsies were obtained from 58 patients. Biopsies were stained with H & E and digitized at 40× optical magnification using an Aeperio scanner. An expert pathologist selected regions of interest (ROIs) on each biopsy. In total 120 ROIs were selected across. Each ROI was assigned a Gleason grade of either benign, Gleason grade 3, or grade 4. All glands contained within each ROI were manually segmented to obtain a total of 888 glands from benign ($N = 93$), Gleason grade 3 ($N = 748$), and Gleason grade 4 ($N = 47$) ROIs. For this set of experiments only Gleason grade 3 and Gleason grade 4 glands were considered during classification. DBS was the dissimilarity measure utilized to quantify morphologic differences between glands [2].

Evaluation Measures

Silhouette Index

Silhouette Index is a measure of how well samples cluster by class label [194] with 1 corresponding to perfect clustering by class and -1 corresponding to no clustering by class. Silhouette Index is calculated as, $\eta^{SI} = \sum_{i=1}^{N} \frac{\nu(i)-\psi(i)}{\max[\psi(i),\nu(i)]}$ where $\psi(i) = \sum_{j,l_j=l_i} ||\tilde{y}_i - \tilde{y}_j||$ and $\nu(i) = \sum_{j,l_j\neq l_i} ||\tilde{y}_i - \tilde{y}_j||$.

Area Under the Receiver Operator Characteristic (ROC) Curve

A probabilistic boosting tree classifier [208] was trained and evaluated using a 5×2 cross validation scheme [209]. For each of the 5 runs, the dataset was divided into 2 folds

such that all samples from a single patient were contained in the same fold and all folds maintained class balance. The probabilistic boosting tree classifier assigns a probability value to each sample of belonging to the positive class. Altering the threshold level of the probablistic boosting tree classifier allows for the construction of a ROC Curve. For each ROC Curve the area under the curve is calculated.

Experiment 1: Application of SSMM to Filtered Manifold Learning

For each dataset O in Table 4.12, a manifold $\mathcal{M}$ was calculated from O using GE as described in Section 4.3.2. Similarly a filtered manifold $\mathcal{M}^c$ was found by *FilterManifold* as described in Section 4.4.4. The measures described in Section 4.4.5 were used to evaluate $\mathcal{M}$ and $\mathcal{M}^c$. A Student's t-test was calculated to determine the statistical significance between $\mathcal{M}$ and $\mathcal{M}^c$ for each evaluation measure described in Section 4.4.5.

Experimental results for all datasets are reported in Table 4.13. Across all datasets $\mathcal{M}^c$ outperforms $\mathcal{M}$ in terms of Silhouette Index and area under the ROC curve. In the prostate histology dataset these increases in Silhouette Index and area under the ROC curve were statistically significant ($p \leq 0.1$). Hence $\mathcal{M}^c$ is better able to preserve object-class relationships in the datasets evaluated.

For the synthetic datasets, changes in Silhouette Index and area under the ROC curve are not always statistically significant. However, as may be noted in Figure 4.20 (d) $\mathcal{M}^c$ is a closer approximation to the true embedding (Figure 4.20 (b)) than compared to $\mathcal{M}$ (Figure 4.20 (c)). In Figure 4.20 the samples are colored according to their location on the true embedding to aid in visualization.

Experiment 2: Application of SSMM to Filtered OSE

For each dataset O in Table 4.12, a training set O^{tr} and a testing set O^{te} were defined so that O^{te} is 10% of O and $O^{tr} \cup O^{te} = \emptyset$. O^{tr} and O^{te} were used to construct an original manifold $\mathcal{M}^{te}$ and filtered manifold $\mathcal{M}^{te,c}$. $\mathcal{M}^{te}$ is generated by applying GE as described in Section 4.3.2 and then applying out-of-sample extrapolation as described in Section 4.3.2 to O^{te} where $\mathcal{M}^{tr}$ is the training manifold. The filtered manifold $\mathcal{M}^{te,c}$ is calculated by *OSE-SSMM* as described in Section 4.4.4. The measures described in

(a)

Dataset	$\mathcal{M}$	$\mathcal{M}^c$	*p-value*
Swiss Roll	0.56 ± 0.01	$\mathbf{0.57 \pm 0.03}$	0.063
Helix	0.44 ± 0.05	$\mathbf{0.47 \pm 0.02}$	0.138
Prostate	0.02 ± 0.01	$\mathbf{0.05 \pm 0.03}$	0.032

(b)

Dataset	$\mathcal{M}$	$\mathcal{M}^c$	*p-value*
Swiss Roll	0.876 ± 0.067	$\mathbf{0.935 \pm 0.065}$	0.071
Helix	0.995 ± 0.002	$\mathbf{0.996 \pm 0.002}$	0.240
Prostate	0.779 ± 0.075	$\mathbf{0.832 \pm 0.048}$	0.073

Table 4.13: (a) Silhouette Index and (b) area under the ROC curve are reported for $\mathcal{M}$ and $\mathcal{M}^c$. The best value for each dataset is bolded. p-values are reported for a Student's t-test comparing $\mathcal{M}$ and $\mathcal{M}^c$.

Section 4.4.5 were used to evaluate $\mathcal{M}^{te}$ and $\mathcal{M}^{te,c}$. A Student's t-test was calculated to determine the statistical significance between $\mathcal{M}^{te}$ and $\mathcal{M}^{te,c}$ for each evaluation measure described in Section 4.4.5.

Experimental results for all datasets are reported in Table 4.4.5. For the histopathology dataset $\mathcal{M}^{te,c}$ outperforms $\mathcal{M}^{te}$ in terms of Silhouette Index and area under the ROC curve. The synthetic datasets, the Swiss Roll and Helix, do not show a significant improvement in performance.

(a)

Dataset	$\mathcal{M}^{te}$	$\mathcal{M}^{te,c}$	*p-value*
Swiss Roll	0.57 ± 0.01	$\mathbf{0.58 \pm 0.01}$	0.061
Helix	0.47 ± 0.01	0.47 ± 0.01	0.77
Prostate	-0.04 ± 0.01	$\mathbf{-0.02 \pm 0.02}$	0.005

(b)

Dataset	$\mathcal{M}^{te}$	$\mathcal{M}^{te,c}$	*p-value*
Swiss Roll	0.997 ± 0.003	$\mathbf{0.999 \pm 0.002}$	0.102
Helix	0.994 ± 0.002	$\mathbf{0.996 \pm 0.002}$	0.089
Prostate	0.779 ± 0.054	$\mathbf{0.834 \pm 0.051}$	0.032

Table 4.14: (a) Silhouette Index and (b) area under the ROC curve are reported for $\mathcal{M}^{te}$ and $\mathcal{M}^{te,c}$. The best value for each dataset is bolded. p-values are reported for a Student's t-test comparing $\mathcal{M}^{te}$ and $\mathcal{M}^{te,c}$.

4.4.6 Concluding Remarks on Statistical Shape Model of Manifolds (SSMM)

We presented a statistical shape model of manifolds (SSMM) which is a novel integration of statistical shape models (SSMs) with ensemble learning for regularizing low dimensional data representations of high dimensional spaces. New, related manifolds may then be constrained by the SSMM to only those shapes statistically most likely to occur.

The SSMM may be utilized for several applications including (a) identification of noisy samples based on their deviation from the SSMM. Removing these samples from the dataset may result in higher area under the receiver operator characteristic (ROC) curve. (b) A classifier could be trained on the SSMM allowing for (i) classifier decision boundaries to be applied to a new related manifold without retraining the classifier or (ii) new, related samples to be classified by projection of the samples onto the SSMM. (c) identification of regions on a new, related manifold which deviate the SSMM. Identifying these regions may aid in determining meaningful differences between the dataset and SSMM.

To construct the SSMM we (1) generate a set of manifolds $\mathbb{M}$ for a database of images $\mathbf{C}$, (2) align manifolds in $\mathbb{M}$, and (3) calculate the maximum likelihood estimate of the manifold shape and its primary modes of variation. The SSMM allows for constraining a new, related manifold instance to only those shapes statistically most likely to occur.

We have demonstrated in this work that SSMM can improve area under the ROC curve in the context of Gleason grading of prostate histopathology utilized quantitative morphologic features of glands. For the dataset considered, the tissue samples corresponded to either Gleason grade 3 or grade 4. Improvements in area under the ROC curve via the SSMM were demonstrated for two applications: (1) We demonstrated that outlier samples within a manifold can be identified as those samples which deviate from the SSMM via *FilterManifold*. Removal of outlier samples increased area under the ROC curve and Silhouette Index. (2) We demonstrated via *OSE-SSMM* that manifold regularization by the SSMM improves Silhouette Index and area under the ROC

curve when performing out-of-sample extrapolation on never before seen samples onto the SSMM.

In future work we intend to explore the ability of the SSMM to identify regions of a new, related manifold which deviate from the SSMM. These regions will then be further investigated to determine subtle difference between the dataset and the SSMM. Secondly, we plan on investigating the effects of dataset size on the SSMM by evaluating how accurately the mean manifold shape and primary modes of variation of the manifold shape are represented for SSMMs trained on different dataset sizes.

Chapter 5

Concluding Remarks and Future Work

The current gold standard for prostate cancer detection, transrectal ultrasound (TRUS)-guided biopsy, misses half of all prostate cancer nodules [11]. Due to the low sensitivity of TRUS-guided biopsy, one in three patients will undergo a repeat biopsy to confirm their negative biopsy [210]. Hence, there is a clinical need to improve the targeting of needle biopsy to those regions suspicious for prostate cancer, and thereby reduce the number of false negative biopsies.

In this work, a framework for constructing a computerized decision support system for fused Magnetic Resonance Imaging (MRI)-TRUS was developed to address the clinical need for improving biopsy. A computerized decision support system that leverages the fused MRI-TRUS imagery can aid in (1) detecting prostate cancer and (2) distinguishing aggressive from non-aggressive prostate cancer. A fused MRI-TRUS computerized decision support system will allow clinicians to acquire biopsy samples from those regions most likely to contain aggressive prostate cancer and may ultimately reduce the number of false negative biopsies.

Currently, no fused MRI-TRUS computerized decision support systems exist. Several studies have shown that computerized decision support systems for MRI [25,38–40] and TRUS [34–37] alone have utility in detecting prostate cancer. MRI computerized decision support systems are limited in guiding needle biopsy because MRI-guided biopsy has long procedure times, is expensive due to the need for specialized equipment and technicians, and is stressful for many patients. TRUS computer decision support systems are more clinically feasible as TRUS-guided biopsy is routine. However TRUS-based systems are typically less accurate than MRI-based systems [34–37].

Recent work from our lab has demonstrated that combining information across

modalities increases the performance of computerized decisions support compared to any single constituent modality [54]. Hence, our work developing a fused MRI-TRUS computerized decision support system is based on the premise that combining information from MRI and TRUS will result in more accurate prostate cancer detection than either modality alone. Such a combination will enable needle biopsy to be guided by acoustic and anatomic information from TRUS along with the structural, function, and metabolic information from MRI.

In order to enable the construction of a computerized decision support system for fused MRI-TRUS we developed several sophisticated image analysis tools. These tools can be used to (a) determine the spatial location of aggressive prostate cancer on whole mount histopathology (WMH), (b) map the location of aggressive prostate cancer from WMH onto MRI, and (c) fuse MRI with delineated aggressive prostate cancer spatial extent (determine from WMH) with TRUS. Aggressive prostate cancer spatial extent mapped onto fused MRI-TRUS imagery enables training of a computerized decisions support system. The novel algorithms developed to perform the aforementioned tasks are comprised of:

1. Multi-attribute probabilistic prostate elastic registration (MAPPER) to register MRI and TRUS images of the prostate.

2. Prostalign to semi-automatically align WMH to MRI images of the prostate.

3. Explicit Shape Descriptors (ESDs) to quantify prostate cancer aggressiveness on histopathology by leveraging morphologic features of prostate glands.

Below a description of the contributions and results of each algorithm described in greater detail.

MAPPER was presented in Chapter 2 to spatially align prostate MRI and TRUS automatically during biopsy. Specifically, MAPPER uses a semi-automated segmentation algorithm on MRI to delineate the prostate. The delineated prostate on MRI is then aligned to a probabilistic map of prostate location on TRUS enabling registration of prostate MRI onto TRUS. In comparison state-of-the-art MRI-TRUS registration

algorithms rely on manual intervention during biopsy to either delineate the prostate or select corresponding fiducials on MRI and TRUS [30, 60, 62–67]. Requiring manual interaction to align MRI and TRUS results in increased procedure times and patient discomfort. MAPPER represents a significant advance over the current state-of-the-art approaches in that (1) it requires no user interaction during the biopsy procedure to determine the location of the prostate on TRUS and (2) although MAPPER requires minimal manual interaction segmenting the prostate on MRI this step can be preformed prior to the biopsy procedure and, hence, will not increase biopsy procedure time. MAPPER was evaluated on 13 patient studies from two datasets – Dataset 1 had 6 studies with a side-firing TRUS probe and 1.5 T surface coil MRI, Dataset 2 had 7 studies with a volumetric end-firing TRUS probe and 3.0 T endorectal coil MRI. MAPPER has a root mean square error (RMSE) for expertly selected fiducials (consisting of the urethra, calcifications, and the centroids of cancer nodules) of 3.36 ± 1.10 mm for Dataset 1 and 3.14 ± 0.75 mm for Dataset 2. Comparatively, state-of-the-art MRI-TRUS fusion methods that require manual intervention on ultrasound report RMSE in the range of 3.06-2.07 mm.

One limitation of MAPPER is the use of B-Splines with an imposed regularization constraint to ensure the underlying deformation in the prostate was smoothly varying. Other transformations such as Finite Element Models (FEM), which allow for explicit modeling of tissue physics, could also potentially be used to constrain the registration to only physically plausible transformations of the prostate [65]. Another limitation is the use of the Multi-Feature Appearance (MFA) model of prostate appearance on MRI previously described in Toth and Madabhushi [74]. The accuracy of MAPPER is reliant on an accurate segmentation of the prostate on MRI, hence, developing a more accurate segmentation scheme may improve the accuracy of MAPPER.

Prostalign was presented in Chapter 3 to spatially align WMH onto MRI. Several previous methods to align WMH onto MRI have been presented that require obtaining additional *ex vivo* imaging modalities [90, 91], implanting fiducials [89, 92], or creating patient specific molds [98]. These methods are difficult to implement in a wider clinical setting because they are time consuming, expensive, and may require additional imaging

equipment. While there are a few methods that can directly register WMH to *in vivo* MRI [93, 94], they are often not robust to large deformations in the prostate or tissue loss. Hence we developed Prostalign, a method to manual select corresponding fiducials on WMH and MRI to drive a thin-plate spline registration. We demonstrated accurate alignment between the prostate on WMH, MRI, and TRUS for one patient study.

In Chapter 4 ESDs were presented to assess prostate cancer aggressiveness utilizing gland morphology. Pathologists typically assess prostate cancer aggressiveness with the Gleason grading system, a qualitatively ranking determined by differences in the arrangement and morphology of the nuclei and glands within a tissue sample [42, 44]. Several methods to assess Gleason grade have been presented that utilize tissue texture [46, 47, 101, 102], the arrangement of nuclei and glands [48, 102], nuclei morphology [49, 50], and gland morphology [102]. However, many of these systems have difficulty distinguishing between intermediate Gleason grades 3 and 4, a clinical important task as Gleason grades 4 and 5 are considered more aggressive, while lower Gleason grades (≤ 3) are considered less aggressive. Gland morphology is often an important cue in distinguishing between intermediate Gleason grades 3 and 4, hence, ESDs were developed to quantitatively describe gland morphology. ESDs are calculated by: (a) representing the shape of each gland using a medial axis shape model (MASM), (b) registering MASMs using a novel diffeomorphic based similarity (DBS) measure, (c) determining parameter correspondence between registered MASMs, (d) extracting a low dimensional representation of morphologic features utilizing the non-linear dimensionality reduction scheme Graph Embedding, (e) classifying the morphologic features using a Support Vector Machine (SVM). ESDs are a concise set of features that are capable of distinguishing between intermediate Gleason grades with a maximum accuracy of 89% for 888 prostate glands acquired from 58 patient needle core biopsies.

ESDs for Gleason grading are dependent on glands being present to classify the tissue. The highest Gleason grade 5 is characterized by the absence of glands, hence, ESDs will be unable to classify tissue belonging to Gleason grade 5. In future work we will integrate the ESD features with other types of histologic image attributes, such as the shape and arrangement of nuclei [111] or texture [48, 167], to build classification

tools for the highest Gleason grade of prostate histopathology. Additionally a classifier that incorporates ESDs with other features may be better able to classify tissue that contains glands with atypical shape attributes.

When considering all the three novel algorithms presented in this work – MAPPER for MRI-TRUS fusion, Prostalign for MRI-WMH alignment, and ESDs for assessment of prostate cancer aggressiveness – the spatial extent of aggressive prostate cancer can be learned from the WMH and mapped onto fused MRI-TRUS imagery. Thist enables the aggressive prostate cancer location to be mapped on fused MRI-TRUS imagery, and ultimately can lead to the development of fused MRI-TRUS classifiers for identifying aggressive prostate cancer. However, one limitation of the current methodology is error propagation; namely, that errors delineating aggressive prostate cancer on WMH will be made more inaccurate if there are errors spatially aligning WMH onto MRI. Furthermore these errors will be further exacerbated by errors in registration of MRI onto TRUS, ultimately leading to inaccurate delineation of cancer spatial extent when training the computerized decision support system. Hence, in future work it is important to experimentally evaluate the accuracy of the entire combined workflow, in addition to the experimental evaluate of the individual components as performed in this work. Additionally, in this work validation of the individual components was done utilizing expert annotations (Gleason grade for ESDs and corresponding landmarks for MAPPER), which are well documented to be subject to intra- and inter-observer variability [45, 61]. A more robust and long term analysis of our methodology assessing (a) the number of cancer positive needle biopsy cores obtained above blinded sextant biopsy and (b) patient outcomes is need evaluate the methods presented in this work.

Finally, future work is need to train and validate the computerized decision support system. To train the computerized decision support system features on MRI and TRUS will need to be extracted; such features may include texture features [25,34–36,39], raw RF features [37], and functional parameters [25,38]. The ground truth spatial extent of prostate cancer can then be utilized to train a classification scheme [41] Due to the large number of features that may be useful for classification, machine learning methods to determine features useful for classification and combine them in an intelligent manner

may be leveraged to improve classifier performance [25, 54]. Ultimately, such a system will need to be validate in a clinical trail to assess the ability to obtain cancer positive cores from regions the computerized decision support system identifies as being likely to contain aggressive prostate cancer.

References

[1] R. Sparks, R. Toth, J. Chappelow, G. Xiao, and A. Madabhushi. An integrated framework for analyzing three-dimensional shape differences: Evaluating prostate morphometry. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1081–1084, 2010.

[2] R. Sparks and A. Madabhushi. Novel morphometric based classification via diffeomorphic based shape representation using manifold learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 6363, pages 658–665, 2010.

[3] R. Sparks and A. Madabhushi. Out-of-sample extrapolation using semi-supervised manifold learning (OSE-SSL): Content-based image retrieval for prostate histology grading. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 734–737, 2011.

[4] R. Sparks and A. Madabhushi. Gleason grading of prostate histology utilizing manifold regularization via statistical shape model of manifolds. In *SPIE Medical Imaging*, volume 8315, page 83151J, 2012.

[5] R. Sparks, B. N. Bloch, E. Feleppa, D. Barratt, and A. Madabhushi. Fully automated prostate magnetic resonance imaging and transrectal ultrasound fusion via a probabilistic registration metric. In *SPIE Medical Imaging*, volume 8671, pages 86710A–86710A–14, 2013.

[6] R. Sparks and A. Madabhushi. Explicit shape descriptors: Novel morphologic features for histopathology classification. *Medical Image Analysis*, 17(8):997 – 1009, 2013.

[7] R. Sparks and A. Madabhushi. Statistical shape model for manifold regularization: Gleason grading of prostate histology. *Computer Vision and Image Understanding*, 117(9):1138 – 1146, 2013.

[8] K. K. Hodge, J. E. McNeal, M. K. Terris, and T. A. Stamey. Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. *The Journal of urology*, 142(1):71–75, 1989.

[9] E. J. Halpern and S. E. Strup. Using gray-scale and color and power doppler sonography to detect prostatic cancer. *American Journal of Roentgenology*, 174(3):623–627, 2000.

[10] B. Spajic, H. Eupic, D. Tomas, G. Stimac, B. Kruslin, and O. Kraus. The incidence of hyperechoic prostate cancer in transrectal ultrasound-guided biopsy specimens. *Urology*, 70(4):734–737, 2007.

[11] 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? *Journal of Urology*, 166(5):1679–1683, 2001.

[12] M. Truong, J. A. Slezak, C. P. Lin, V. Iremashvili, M. Sado, A. A. Razmaria, G. Leverson, M. S. Soloway, S. E. Eggener, E. J. Abel, T. M. Downs, and D. F. Jarrard. Development and multi-institutional validation of an upgrading risk tool for gleason 6 prostate cancer. *Cancer*, 119(22):3992–4002, 2013.

[13] A. C. Loch, A. Bannowsky, L. Baeurle, B. Grabski, B. Knig, G. Flier, O. Schmitz-Krause, U. Paul, and T. Loch. Technical and anatomical essentials for transrectal ultrasound of the prostate. *World Journal of Urology*, 25(4):361–366, 2007.

[14] W. J. Ellis and M. K. Brawer. Repeat prostate needle biopsy: Who needs it? *Journal of Urology*, 153(5):1496–1498, 1995.

[15] D. W. Keetch, W. J. Catalona, and D. S. Smith. Serial prostatic biopsies in men with persistently elevated serum prostate specific antigen values. *Journal of Urology*, 151(6):1571–1574, 1994.

[16] D. Cheng and C. M. Tempany. MR imaging of the prostate and bladder. *Seminars in Ultrasound, CT, and MRI*, 19(1):67 – 89, 1998.

[17] G. J. M. Parker, J. Suckling, S. F. Tanner, A. R. Padhani, P. B. Revell, J. E. Husband, and M. O. Leach. Probing tumor microvascularity by measurement, analysis and display of contrast agent uptake kinetics. *Journal of Magnetic Resonance Imaging*, 7(3):564–574, 1997.

[18] E. M. Charles-Edwards and N. M. deSouza. Diffusion-weighted magnetic resonance imaging and its application to cancer. *Cancer Imaging*, 6:135–143, 2006.

[19] J. Kurhanewicz and D. B. Vigneron. Advances in MR spectroscopy of the prostate. *Magnetic Resonance Imaging Clinics of North America*, 16(4):697–710, 2008.

[20] M. A. Haider, T. H. van der Kwast, J. Tanguay, A. J. Evans, A.-T. Hashmi, G. Lockwood, and J. Trachtenberg. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *American Journal of Roentgenology*, 189(2):323–328, 2007.

[21] S. Viswanath, B. N. Bloch, M. Rosen, J. Chappelow, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur, and A. Madabhushi. Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 Tesla MRI. In *SPIE Medical Imaging*, volume 7260, Miami, Florida, 2009.

[22] P. Tiwari, M. Rosen, and A. Madabhushi. A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). *Medical Physics*, 36(9):3927–3939, 2009.

[23] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR in Biomedicine*, 25(4):607–619, 2012.

[24] S. Viswanath, B. N. Bloch, J. Chappelow, P. Patel, N. Rofsky, R. Lenkinski, E. Genega, and A. Madabhushi. Enhanced multi-protocol analysis via intelligent supervised embedding (EMPrAvISE): detecting prostate cancer on multiparametric MRI. In *SPIE Medical Imaging*, volume 7963, page 79630U, 2011.

[25] Pallavi Tiwari, John Kurhanewicz, Mark Rosen, and Anant Madabhushi. Semi supervised multi kernel (sesmik) graph embedding: Identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. In T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2010*, volume 6363 of *Lecture Notes in Computer Science*, pages 666–673. Springer Berlin Heidelberg, 2010.

[26] D. Beyersdorff, M. Taupitz, B. Winkelmann, T. Fischer, S. Lenk, S. A. Loening, and B. Hamm. Patients with a history of elevated prostate-specific antigen levels and negative transrectal US-guided quadrant or sextant biopsy results: value of MR imaging. *Radiology*, 224(3):701–706, 2002.

[27] K. M. Pondman, J. J. Fütterer, B. ten Haken, L. J. Schultze Kool, J. A. Witjes, T. Hambrock, K. J. Macura, and J. O. Barentsz. MR-guided biopsy of the prostate: an overview of techniques and a systemic review. *European Urology*, 54:517–527, 2008.

[28] Labanaris, A. P. and Engelhard, K. and Zugor, V. and Nutzel, R. and Kuhn, R. Prostate cancer detection using an extended prostate biopsy schema in combination with additional targeted cores from suspicious images in conventional and functional endorectal magnetic resonance imaging of the prostate. *Prostate Cancer and Prostatic Diseases*, 13(1):65–70, 2010.

[29] B. A. Hadaschik, T. H. Kuru, C. Tulea, P. Rieker, I. V. Popeneciu, T. Simpfendörfer, J. Huber, P. Zogal, D. Teber, S. Pahernik, M. Roethke, P. Zamecnik, W. Roth, G. Sakas, H. P. Schlemmer, and M. Hohenfellner. A novel stereotactic prostate biopsy system integrating pre-interventional magnetic resonance imaging and live ultrasound fusion. *The Journal of Urology*, 186(6):2214 – 2220, 2011.

[30] S. Natarajan, L. S. Marks, D. J. A. Margolis, J. Huang, M. L. Macairan, P. Lieu, and A. Fenster. Clinical application of a 3D ultrasound-guided prostate biopsy system. *Urologic Oncology: Seminars and Original Investigations*, 29(3):334 – 342, 2011.

[31] E. Rud, E. Baco, and H. B. Eggesbo. Mri and ultrasound-guided prostate biopsy using soft image fusion. *Anticancer Research*, 32(8):3383–3389, 2012.

[32] L. Schimmöller, M. Quentin, C. Arsov, R.S. Lanzman, A. Hiester, R. Rabenalt, G. Antoch, P. Albers, and D. Blondin. Inter-reader agreement of the esur score for prostate mri using in-bore mri-guided biopsies as the reference standard. *European Radiology*, 2013.

[33] A. B. Rosenkrantz, R. P. Lim, M. Haghighi, M. B. Somberg, J. S. Babb, and S. S. Taneja. Comparison of interreader reproducibility of the prostate imaging reporting and data system and likert scales for evaluation of multiparametric prostate mri. *American Journal of Roentgenology*, 201(4):W612–W618, September 2013.

[34] O. Basset, Z. Sun, J. L. Mestas, and G. Gimenez. Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrices. *Ultrasonic Imaging*, 15:218 – 237, 1993.

[35] U. Scheipers, H. Ermert, H. J. Sommerfeld, M. Garcia-Schrmann, T. Senge, and S. Philippou. Ultrasonic multifeature tissue characterization for prostate diagnostics. *Ultrasound in Medicine Biology*, 29(8):1137 – 1149, 2003.

[36] A. L. Huynen, R. J. B. Giesen, J. J. M. C. H. de la Rosette, R. G. Aarnink, F. M. J. Debruyne, and H. Wijkstra. Analysis of ultrasonographic prostate images for the detection of prostatic carcinoma: The automated urologic diagnostic expert system. *Ultrasound in Medicine & Biology*, 20(1):1 – 10, 1994.

[37] E. J. Feleppa, C. R. Porter, J. Ketterling, P. Lee, S. Dasgupta, S. Urban, and A. Kalisz. Recent developments in tissue-type imaging (tti) for planning and monitoring treatment of prostate cancer. *Ultrasonic Imaging*, 26(3):163–172, 2004.

[38] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. *Physics in Medicine and Biology*, 57(6):1527, 2012.

[39] Langer, D. L. and van der Kwast, T. H. and Evans, A. J. and Trachtenberg, J. and Wilson, B. C. and Haider, M. A. Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of Magnetic Resonance Imaging*, 30(2):327–334, 2009.

[40] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto. Quantitative analysis of multiparametric prostate mr images: Differentiation between prostate cancer and normal tissue and correlation with gleason scorea computer-aided diagnosis development study. *Radiology*, 267(3):787–796, 2013.

[41] J. Chappelow, S. Viswanath, J. Monaco, M. Rosen, J. Tomaszewski, M. Feldman, and A. Madabhushi. Improving supervised classification accuracy using non-rigid multimodal image registration: detecting prostate cancer. In *SPIE Medical Imaging*, volume 6915, pages 69150V–69150V–12, 2008.

[42] D. F. Gleason. Classification of prostatic carcinomas. *Cancer Chemotherapy Reports*, 50:125–128, 1966.

[43] P. A. Ganz, J. M. Barry, W. Burke, N. F. Col, P. S. Corso, E. Dodson, M. E. Hammond, B. A. Kogan, C. F. Lynch, L. Newcomer, E. J. Seifter, J. A. Tooze, K. Viswanath, and H. Wessells. National institutes of health state-of-the-science conference: Role of active surveillance in the management of men with localized prostate cancer. *Annals of Internal Medicine*, 156(8):591–595, 2012.

[44] J. I. Epstein, W. C. Allsbrook, M. B. Amin, and L. L. Egevad. The 2005 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *American Journal of Surgical Pathology*, 29(9):1228–1242, 2005.

[45] W. C. Allsbrook, K. A. Mangold, M. H. Johnson, R. B. Lane, C. G. Lane, and J. I. Epstein. Interobserver reproducibility of Gleason grading of prostatic carcinoma: General pathologist. *Human Pathology*, 32(1):81–88, 2001.

[46] P. W. Huang and C. H. Lee. Automatic classification for pathological prostate images based on fractal analysis. *IEEE Transactions on Medical Imaging*, 28(7):1037–1050, 2009.

[47] M. D. DiFranco, G. OHurley, E. W. Kay, R. W. G. Watson, and P. Cunningham. Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. *Computerized Medical Imaging and Graphics*, 35(7):629–645, 2011.

[48] S. Doyle, J. Monaco, M. Feldman, J. Tomaszewski, and A. Madabhushi. An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC Bioinformatics*, 12(1):424, 2011.

[49] R. W. Veltri, C. Marlow, M. A. Khan, M. C. Miller, J. I. Epstein, and A. W. Partin. Significant variations in nuclear structure occur between and within gleason grading patterns 3, 4, and 5 determined by digital image analysis. *The Prostate*, 67(11):1202–1210, 2007.

[50] S. Ali and A. Madabhushi. Active contour for overlap resolution using watershed based initialization (ACOReW): Applications to histopathology. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 614–617, 2011.

[51] S. Ali, R. Veltri, J. A. Epstein, C. Christudass, and A. Madabhushi. Cell cluster graph for prediction of biochemical recurrence in prostate cancer patients from tissue microarrays. In *SPIE Medical Imaging*, volume 8676, pages 86760H–86760H–11, 2013.

[52] G. Lee, R. Sparks, S. Ali, A. Madabhushi, M. D. Feldman, S. R. Master, N. Shih, and J. E. Tomaszewski. Co-occurring gland tensors in localized cluster graphs: Quantitative histomorphometry for predicting biochemical recurrence for intermediate grade prostate cancer. In *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pages 113–116, 2013.

[53] G. Lee, S. Ali, R. Veltri, J. I. Epstein, C. Christudass, and A. Madabhushi. Cell orientation entropy (core): Predicting biochemical recurrence from prostate cancer tissue microarrays. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*, volume 8151 of *Lecture Notes in Computer Science*, pages 396–403. Springer Berlin Heidelberg, 2013.

[54] S. Viswanath and A. Madabhushi. Consensus embedding: theory, algorithms and application to segmentation and classification of biomedical data. *BMC Bioinformatics*, 13(1):26, 2012.

[55] Asha Singanamalli, Rachel Sparks, Mirabela Rusu, Natalie Shih, Amy Ziober, John Tomaszewski, Mark Rosen, Michael Feldman, and Anant Madabhushi. Identifying in vivo dce mri parameters correlated with ex vivo quantitative microvessel architecture: A radiohistomorphometric approach. In *SPIE Medical Imaging*, volume 8676, pages 867604–867604–14, 2013.

[56] A. M. D. 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: A Cancer Journal for Clinicians*, 60(2):70–98, 2010.

[57] R. J. Babaian, A. Toi, K. Kamoi, P. Troncoso, J. Sweet, R. Evans, D. Johnston, and M. Chen. A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy. *The Journal of Urology*, 163(1):152 – 157, 2000.

[58] A. G. Anastasiadis, M. P. Lichy, U. Nagele, M. A. Kuczyk, A. S. Merseburger, J. Hennenlotter, S. Corvin, K. D. Sievert, C. D. Claussen, A. Stenzl, and H. P. Schlemmer. MRI-guided biopsy of the prostate increases diagnostic performance in men with elevated or increasing PSA levels after previous negative TRUS biopsies. *European Urology*, 50:738–749, 2006.

[59] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187 –198, 1997.

[60] J. Mitra, Z. Kato, R. Martí, A. Oliver, X. Lladó, D. Sidibé, S. Ghose, J. C. Vilanova, J. Comet, and F. Meriaudeau. A spline-based non-linear diffeomorphism for multimodal prostate registration. *Medical Image Analysis*, 2012.

[61] D. Liu, N. Usmani, S. Ghosh, W. Kamal, J. Pedersen, N. Pervez, D. Yee, B. Danielson, A. Murtha, J. Amanie, and R. S. Sloboda. Comparison of prostate volume, shape, and contouring variability determined from preimplant magnetic resonance and transrectal ultrasound images. *Brachytherapy*, 11(4):284 – 291, 2012.

[62] 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. *Computer Aided Surgery*, 13(5):255–264, 2008.

[63] I. Kaplan, N. E. Oldenburg, P. Meskell, M. Blake, P. Church, and E. J. Holupka. Real time MRI-ultrasound image guided stereotactic prostate biopsy. *Magnetic Resonance Imaging*, 20:295–299, 2002.

[64] R. Narayanan, J. Kurhanewicz, K. Shinohara, E. D. Crawford, A. Simoneau, and J. S. Suri. MRI-ultrasound registration for targeted prostate biopsy. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 991 –994, 2009.

[65] Y. Hu, H. U. Ahmed, Z. Taylor, C. Allen, M. Emberton, D. Hawkes, and D. Barratt. MR to ultrasound registration for image-guided prostate interventions. *Medical Image Analysis*, 2011.

[66] C. Reynier, J. Troccaz, P. Fourneret, A. Dusserre, C. Gay-Jeune, J. L. Descotes, M. Bolla, and J. Y. Giraud. Mri/trus data fusion for prostate brachytherapy. preliminary results. *Medical Physics*, 31(6):1568–1575, 2004.

[67] V. V. Karnik, A. Fenster, J. Bax, D. W. Cool, L. Gardi, I. Gyacskov, C. Romagnoli, and A. D. Ward. Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy. *Medical Physics*, 37(2):802–813, 2010.

[68] G. J. Bubley, B. N. Bloch, C. Vazquez, E. Genega, E. Holupka, N. Rofsky, and I. Kaplan. Accuracy of endorectal magnetic resonance/transrectal ultrasound fusion for detection of prostate cancer during brachytherapy. *Urology*, 81(6):1284–1290, 2013.

[69] P. A. Pinto, P. H. Chung, A. R. Rastinehad, A. A. Baccala, J. Kruecker, C. J. Benjamin, S. Xu, P. Yan, S. Kadoury, C. Chua, J. K. Locklin, B. Turkbey, J. H. Shih, S. P. Gates, C. Buckner, G. Bratslavsky, W. M. Linehan, N. D. Glossop, P. L. Choyke, and B. J. Wood. Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging. *The Journal of Urology*, 186(4):1281 – 1285, 2011.

[70] N. K. Yerram, D. Volkin, B. Turkbey, J. Nix, A. N. Hoang, S. Vourganti, G. N. Gupta, W. M. Linehan, P. L. Choyke, B. J. Wood, and P. A. Pinto. Low suspicion lesions on multiparametric magnetic resonance imaging predict for the absence of high-risk prostate cancer. *BJU International*, 110(11b):E783–E788, 2012.

[71] V. Daanen, J. Gastaldo, J. Y. Giraud, P. Fourneret, J. L. Descotes, M. Bolla, D. Collomb, and J. Troccaz. MRI/TRUS data fusion for brachytherapy. *International Journal of Medical Robotics and Computer Assisted Surgery*, 2:256–261, 2006.

[72] Geoffrey A. Sonn, Shyam Natarajan, Daniel J.A. Margolis, Malu MacAiran, Patricia Lieu, Jiaoti Huang, Frederick J. Dorey, and Leonard S. Marks. Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. *The Journal of Urology*, 189(1):86 – 92, 2013.

[73] L. Dickinson, Y. Hu, H. U. Ahmed, C. Allen, A. P. Kirkham, M. Emberton, and D. Barratt. Image-directed, tissue-preserving focal therapy of prostate cancer: a feasibility study of a novel deformable magnetic resonance-ultrasound (mr-us) registration system. *BJU International*, 112(5):594–601, 2013.

[74] R. Toth and A. Madabhushi. Multi-feature landmark free active appearance models: Application to prostate MRI segmentation. *IEEE Transactions on Medical Imaging*, 38(8):1638–1650, Aug 2012.

[75] Y. Zhan and D. Shen. Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. *IEEE Transactions on Medical Imaging*, 25(3):256–272, 2006.

[76] S. Ghose, A. Oliver, R. Mart, X. Llad, J. Freixenet, J. C. Vilanova, and F. Meriaudeau. Texture guided active appearance model propagation for prostate segmentation. In Anant Madabhushi, Jason Dowling, Pingkun Yan, Aaron Fenster, Purang Abolmaesumi, and Nobuhiko Hata, editors, *Prostate Cancer Imaging. Computer-Aided Diagnosis, Prognosis, and Intervention*, volume 6367 of *Lecture Notes in Computer Science*, pages 111–120. Springer Berlin Heidelberg, 2010.

[77] F. A. Cosío. Automatic initialization of an active shape model of the prostate. *Medical Image Analysis*, 12(4):469 – 483, 2008.

[78] C. B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Transactions on Sonics and Ultrasonics*, 25(1):1–6, 1978.

[79] M. D. Yacoub, J. E. V. Bautista, and L. Guerra de Rezende Guedes. On higher order statistics of the Nakagami-m distribution. *IEEE Transactions on Vehicular Technology*, 48(3):790–794, 1999.

[80] S. D. Pye, S. R. Wild, and W. N. McDicken. Adaptive time gain compensation for ultrasonic imaging. *Ultrasound in Medicine & Biology*, 18(2):205 – 212, 1992.

[81] G. Xiao, M. Brady, J. A. Noble, and Y. Zhang. Segmentation of ultrasound b-mode images with intensity inhomogeneity correction. *IEEE Transactions on Medical Imaging*, 21(1):48 –57, 2002.

[82] E. A. Ashton and K. J. Parker. Multiple resolution Bayesian segmentation of ultrasound images. *Ultrasonic Imaging*, 17(4):291–304, 1995.

[83] M. S. Cohen, R. M. Dubois, and M. M. Zeineh. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Human Brain Mapping*, 10(4):204–211, 2000.

[84] J. A. Greenwood and D. Durand. Aids for fitting the gamma distribution by maximum likelihood. *Technometrics*, 2(1):pp. 55–65, 1960.

[85] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. In *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, pages 14–19, 1990.

[86] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712 –721, 1999.

[87] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker. Engineering and algorithm design for an image processing api: A technical report on itk - the insight toolkit. In ed. Westwood, editor, *Proc. of Medicine Meets Virtual Reality*, pages 586–592. IOS Press Amsterdam, Jan 2002.

[88] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007.

[89] A. D. Ward, C. Crukley, C. A. McKenzie, J. Montreuil, E. Gibson, C. Romagnoli, J. A. Gomez, M. Moussa, J. Chin, G. Bauman, and A. Fenster. Prostate: Registration of digital histopathologic images to in vivo mr images acquired by using endorectal receive coil. *Radiology*, 263(3):856–864, 2012.

[90] H. Park, M. R. Piert, A. Khan, R. Shah, H. Hussain, J. Siddiqui, T. L. Chenevert, and C. R. Meyer. Registration methodology for histological sections and in vivo imaging of human prostate. *Academic Radiology*, 15(8):1027 – 1039, 2008.

[91] C. Orczyk, H. Rusinek, A. B. Rosenkrantz, A. Mikheev, F. M. Deng, J. Melamed, and S. S. Taneja. Preliminary experience with a novel method of three-dimensional co-registration of prostate cancer digital histology and in vivo multiparametric {MRI}. *Clinical Radiology*, 68(12):e652 – e658, 2013.

[92] E. Gibson, C. Crukley, M. Gaed, J. A. Gmez, M. Moussa, J. L. Chin, G. S. Bauman, A. Fenster, and A. D. Ward. Registration of prostate histology images to ex vivo mr images via strand-shaped fiducials. *Journal of Magnetic Resonance Imaging*, 36(6):1402–1412, 2012.

[93] J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, and A. Madabhushi. Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information. *Medical Physics*, 38(4):2005–2018, 2011.

[94] Y. Zhan, Y. Ou, M. Feldman, J. Tomaszeweski, C. Davatzikos, and D. Shen. Registering histologic and {MR} images of prostate for image-based cancer detection. *Academic Radiology*, 14(11):1367 – 1381, 2007.

[95] A. P. Moskalik, P. L. Carson, J. M. Rubin, J. B. Fowlkes, K. Wojno, and R. L. Bree. 3d registration of ultrasound with histology in the prostate. In *Ultrasonics Symposium, 1997. Proceedings., 1997 IEEE*, volume 2, pages 1397–1400 vol.2, 1997.

[96] L. S. Taylor, B. C. Porter, G. Nadasdy, P. A. di SantAgnese, D. Pasternack, Z. Wu, R. B. Baggs, D. J. Rubens, and K. J. Parker. Three-dimensional registration of prostate images from histology and ultrasound. *Ultrasound in Medicine & Biology*, 30(2):161 – 168, 2004.

[97] Stefan G. Schalk, Tamerlan A. Saidov, Hessel Wijkstra, and Massimo Mischi. 3d registration of histology and ultrasound data for validation of prostate cancer imaging. In *Proc. SPIE*, volume 8669, pages 86692T–86692T–10, 2013.

[98] B. Turkbey, H. Mani, V. Shah, A. R. Rastinehad, M. Bernardo, T. Pohida, Y. Pang, D. Daar, C. Benjamin, Y. L. McKinney, H. Trivedi, C. Chua, G. Bratslavsky, J. H. Shih, W. M. Linehan, M. J. Merino, P. L. Choyke, and P. A. Pinto. Multiparametric 3t prostate magnetic resonance imaging to detect cancer: Histopathological correlation using prostatectomy specimens processed in

customized magnetic resonance imaging based molds. *The Journal of Urology*, 186(5):1818 – 1824, 2011.

[99] F. L. Bookstein. Principle warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Learning*, 11(6):567–585, 1989.

[100] A. Madabhushi, S. Agner, A. Basavanhally, S. Doyle, and G. Lee. Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Computerized Medical Imaging and Graphics*, 35(78):506 – 514, 2011.

[101] K. Jafari-Khouzani and H. Soltanian-Zadeh. Multiwavelet grading of pathological images of prostate. *IEEE Transactions on Biomedical Engineering*, 50(6):697–704, 2003.

[102] A. Tabesh, M. Teverovskiy, H. Pang, V. Kumar, D. Verbel, A. Kotsianti, and O. Saidi. Multifeature prostate cancer diagnosis and Gleason grading of histological images. *IEEE Transactions on Medical Imaging*, 26(10):1366–1378, 2007.

[103] A. Golugula, G. Lee, S. R. Master, M. D. Feldman, J. E. Tomaszewski, D. W. Speicher, and A. Madabhushi. Supervised regularized canonical correlation analysis: Integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC Bioinformatics*, 12(1):483, 2011.

[104] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2 edition, 2001.

[105] T. Li, C. Zhang, and M. Ogihara. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, 20(15):2429–2437, 2004.

[106] S. Naik, S. Doyle, A. Madabhushi, J. Tomaszeweski, and M. Feldman. Gland segmentation and gleason grading of prostate histology by integrating low-, high-level and domain specific information. In *Workshop on Microscopic Image Analysis with Applications in Biology*, 2007.

[107] K. A. Jobst, A. D. Smith, M. Szatmari, M. M. Esiri, A. Jaskowski, N. Hindley, B. McDonald, and A. J. Molyneux. Rapidly progressing atrophy of medial temporal lobe in Alzheimer's disease. *Lancet*, 343(8901):829–830, 1994.

[108] R. M. Rangayyan and T. M. Nguyen. Fractal analysis of contours of breast masses in mammograms. *Journal of Digital Imaging*, 20(3):223–237, 2007.

[109] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H. S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero, F. Eustache, O. Colliot, and the Alzheimer's Disease Neuroimaging Initiative. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486, 2009.

[110] K. Gorczowski, M. Styner, J. Y. Jeong, J. S. Marron, J. Piven, H. C. Hazlett, S. M. Pizer, and G. Gerig. Multi-object analysis of volume, pose, and shape using

statistical discrimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):652–661, 2010.

[111] S. Doyle, M. Feldman, N. Shih, J. Tomaszewski, and A. Madabhushi. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. *BMC Bioinformatics*, 13(1):282, 2012.

[112] J. C. L. Looi, M. D. Macfarlane, M. Walterfang, M. Styner, D. Velakoulis, J. Ltt, D. van Westen, and C. Nilsson. Morphometric analysis of subcortical structures in progressive supranuclear palsy: In vivo evidence of neostriatal and mesencephalic atrophy. *Psychiatry Research: Neuroimaging*, 194(2):163 – 175, 2011.

[113] D. Pachauri, C. Hinrichs, M.K. Chung, S. C. Johnson, and V. Singh. Topology-based kernels with application to inference problems in alzheimer's disease. *IEEE Transactions on Medical Imaging*, 30(10):1760 –1770, 2011.

[114] J. P. Monaco, J. E. Tomaszewski, M. D. Feldman, I. Hagemann, M. Moradi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Medial Image Analysis*, 14(4):617–629, 2010.

[115] H. Georgiou, M. Mavroforakis, N. Dimitropoulos, D. Cavouras, and S. Theodoridis. Multi-scaled morphological features for the characterization of mammographic masses using statistical classification schemes. *Artificial Intelligence in Medicine*, 41(1):39–55, 2007.

[116] W. N. Street, W. H. Wolberg, and O. L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. *IS&T/SPIE 1999 International Symposium on Electronic Imaging:Science and Technology*, 1905:861–870, 1993.

[117] G. P. Liney, M. Sreenivas, P. Gibbs, R. Garcia-Alvarez, and L. W. Turnbull. Breast lesion analysis of shape technique: Semiautomated vs. manual morphological description. *Journal of Magnetic Resonance Imaging*, 23(4):493–498, 2006.

[118] R. R. Bailey and M. Srinath. Orthogonal moment features for use with parametric and non-parametric classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):389 –399, 1996.

[119] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24):509–522, 2002.

[120] H. Blum. A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form*, 19(5):362–380, 1967.

[121] C. Brechböhler, G. Gerig, and O. Kübler. Parametrization of closed surfaces for 3-D shape description. *Computer Vision and Image Understanding*, 61(2):154 – 170, 1995.

[122] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. Use of active shape models for locating structure in medical images. *Image and Vision Computing*, 12(6):355–365, 1994.

[123] H. Kauppinen, T. Seppanen, and M. Pietikainen. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):201 –207, 1995.

[124] F. Mokhtarian and S. Abbasi. Shape similarity retrieval under affine transforms. *Pattern Recognition*, 35:31–41, 2002.

[125] E. Persoon and K. S. Fu. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(3):170–179, 1977.

[126] M. Reuter, F. E. Wolter, and N. Peinecke. LaplaceBeltrami spectra as Shape-DNA of surfaces and solids. *Computer-Aided Design*, 38(4):342 – 366, 2006.

[127] S. Rueda, J. K. Udupa, and L. Bai. Shape modeling via local curvature scale. *Pattern Recognition Letters*, 31(4):324 – 336, 2010.

[128] Y. Shinagawa, T. L. Kunii, and Y. L. Kergosien. Surface coding based on Morse theory. *IEEE Computer Graphics and Applications*, 11(5):66 –78, 1991.

[129] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

[130] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.

[131] K. Sjöstrand, E. Rostrup, C. Ryberg, R. Larsen, C. Studholme, H. Baezner, J. Ferro, F. Fazekas, L. Pantoni, D. Inzitari, G. W. Aldemar, and LADIS Study Group. Sparse decomposition and modeling of anatomical shape variation. *IEEE Transactions on Medical Imaging*, 26(12):1625–1635, 2007.

[132] D. A. Alcantara, O. Carmichael, W. Harcourt-Smith, K. Sterner, S. R. Frost, R. Dutton, P. Thompson, E. Delson, and N. Amenta. Exploration of shape variation using localized components analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(8):1510–1516, 2009.

[133] A. Sajjanhar, G. Lu, D. Zhang, J. Hou, and Y. P. Chen. Spherical harmonics and distance transform for image representation and retrieval. In *Proc. Intelligent Data Engineering and Automated Learning*, volume 5788, pages 309–316, 2009.

[134] O. van Kaick, G. Hamarneh, A. D. Ward, M. Schweitzer, and H. Zhang. Learning fourier descriptors for computer-aided diagnosis of the supraspinatus. *Academic Radiology*, 17(8):1040 – 1049, 2010.

[135] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18(10):851 –865, 1999.

[136] P. A. Yushkevich, H. Zhang, and J. C. Gee. Continuous medial representation for anatomical structures. *IEEE Transactions on Medical Imaging*, 25(12):1547–1564, 2006.

[137] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

[138] A. Lauric, E. Miller, S. Frisken, and A. M. Malek. Automated detection of intracranial aneurysms based on parent vessel 3d analysis. *Medical Image Analysis*, 14(2):149 – 159, 2010.

[139] Y. Xu, G. Liang, G. Hu, Y. Yang, J. Geng, and P. K. Saha. Quantification of coronary arterial stenoses in cta using fuzzy distance transform. *Computerized Medical Imaging and Graphics*, 36(1):11 – 24, 2012.

[140] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *International Journal of Computer Vision*, 35(1):13–32, 1999.

[141] M. S. Hassouna and A. A. Farag. Variational curve skeletons using gradient vector flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2257 –2274, dec. 2009.

[142] C. Arcelli, G. S. di Baja, and L. Serino. Distance-driven skeletonization in voxel images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):709 –720, april 2011.

[143] X. Bai, L. J. Latecki, and W. Y. Liu. Skeleton pruning by contour partitioning with discrete curve evolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3):449–462, 2007.

[144] D. Macrini, S. Dickinson, D. Fleet, and K. Siddiqi. Bone graphs: Medial shape parsing and abstraction. *Computer Vision and Image Understanding*, 115(7):1044 – 1061, 2011.

[145] A. D. Ward and G. Hamarneh. The groupwise medial axis transform for fuzzy skeletonization and pruning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1084 –1096, june 2010.

[146] M. C. Chang and B. B. Kimia. Measuring 3D shape similarity by graph-based matching of the medial scaffolds. *Computer Vision and Image Understanding*, 115(5):707 – 720, 2011.

[147] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:550–571, 2004.

[148] X. Bai and L. J. Latecki. Path similarity skeleton graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1282–1292, 2008.

[149] W. B. Goh. Strategies for shape matching using skeletons. *Computer Vision and Image Understanding*, 110(3):326 – 345, 2008.

[150] C. D. Ruberto. Recognition of shapes by attributed skeletal graphs. *Pattern Recognition*, 37(1):21–31, 2004.

[151] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu. Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Learning*, 32(5):861–874, 2010.

[152] A. Egozi, Y. Keller, and H. Guterman. Improving shape retrieval by spectral matching and meta similarity. *IEEE Transactions on Image Processing*, 19(5):1319–1327, 2010.

[153] X. Qian, H. D. Tagare, R. K. Fulbright, R. Long, and S. Antani. Optimal embedding for shape indexing in medical image databases. *Medical Image Analysis*, 14(3):243–254, 2010.

[154] K. Nasreddine, A. Benzinou, and R. Fablet. Variational shape matching for shape classification and retrieval. *Pattern Recognition Letters*, 31(12):1650 – 1657, 2010.

[155] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Learning*, 22(8):888–905, 2000.

[156] H. Guo, A. Rangarajan, and S. Joshi. Diffeomorphic point matching. In N. Paragios, Y. Chen, and O. Faugeras, editors, *Handbook of Mathematical Models in Computer Vision*, pages 205–219. Springer US, 2006.

[157] K. Rose, E. Gurewitz, and G. Fox. A deterministic annealing approach to clustering. *Pattern Recognition Letters*, 11(9):589–594, 1990.

[158] C. Twining, S. Marsland, and C. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. In *Proc. British Machine Vision Conference (BMVC) 2002*, volume 2, pages 847–856, 2002.

[159] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[160] W. R. Franklin and A. H. Barr. Faster calculation of superquadric shapes. *IEEE Computer Graphics and Applications*, 1(3):41–47, 1981.

[161] S. C. Agner, S. Soman, E. Libfeld, M. McDonald, K. Thomas, S. Englander, M. A. Rosen, D. Chin, J. Nosher, and A. Madabhushi. Textural kinetics: A novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification. *Journal of Digital Imaging*, 24(3):446–463, 2011.

[162] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850 –863, 1993.

[163] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[164] J. L. Rodgers and W. A. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[165] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[166] J. Tenenbaum, V. de Silvia, and J. Langford. A global framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[167] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi. A boosted bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies. *IEEE Transactions on Biomedical Engineering*, 59(5):1205–1218, 2012.

[168] D. Comaniciu, P. Meer, and D. J. Foran. Image-guided decision support system for pathology. *Machine Vision and Applications*, 11:213–224, 1999.

[169] A. W. Wetzel, R. Crowley, S. Kim, R. Dawson, L. Zheng, Y. M. Joo, Y. Yagi, J. Gilbertson, C. Gadd, D. W. Deerfield, and M. J. Becich. Evaluation of prostate tumor grades by content-based image retrieval. In Robert J. Mericsko, editor, *27th AIPR Workshop: Advances in Computer-Assisted Recognition*, volume 3584, pages 244–252. SPIE, 1999.

[170] H. L. Tang, R. Hanka, and H. H. S. Ip. Histological image retrieval based on semantic content analysis. *Information Technology in Biomedicine, IEEE Transactions on*, 7(1):26 –36, 2003.

[171] L. Zheng, A. W. Wetzel, J. Gilbertson, and M. J. Becich. Design and analysis of a content-based pathology image retrieval system. *Information Technology in Biomedicine, IEEE Transactions on*, 7(4):249 –255, 2003.

[172] S. Doyle, M. Hwang, S. Naik, M. Feldman, J. Tomaszewwski, and A. Madabhushi. Using manifold learning for content-based image retrieval of prostate histopathology. In *Workshop on Content-Based Image Retrieval for Biomedical Image Archives (in conjunction with MICCAI)*, pages 53–62, 2007.

[173] B. Lessmann, T. W. Nattkemper, V. H. Hans, and A. Degenhard. A method for linking computed image features to histological semantics in neuropathology. *Journal of Biomedical Informatics*, 40(6):631 – 641, 2007.

[174] F. Yu and H. H. S. Ip. Semantic content analysis and annotation of histological images. *Computers in Biology and Medicine*, 38(6):635 – 649, 2008.

[175] S. Naik, S. Doyle, A. Basavanally, S. Ganesan, M. D. Feldman, J. E. Tomaszeweski, and A. Madabhushi. A boosted distance metric: Application to content based image retrieval and classification of digitized histopathology. In *Proc. SPIE*, volume 7260, page 72603F, 2009.

[176] W. Yang, S. Zhang, Y. Chen, W. Li, and Y. Chen. Shape symmetry analysis of breast tumors on ultrasound images. *Computers in Biology and Medicine*, 39(3):231–238, 2009.

[177] J. C. Caicedo, F. A. Gonzlez, and E. Romero. Content-based histopathology image retrieval using a kernel-based semantic annotation framework. *Journal of Biomedical Informatics*, 44(4):519 – 528, 2011.

[178] A. Madabhushi. Digital pathology image analysis: Oppurtunities and challenges. *Imaging in Medicine*, 1:7–10, 2009.

[179] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick. Assert: A physician-in-the-loop content-based retrieval system for hrct image databases. *Computer Vision and Image Understanding*, 75(12):111 – 132, 1999.

[180] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick. A similarity learning approach to content-based image retrieval: application to digital mammography. *Medical Imaging, IEEE Transactions on*, 23(10):1233 – 1244, oct. 2004.

[181] R. Bellman. Curse of dimensionality. *Adaptive control processes: a guided tour. Princeton, NJ*, 1961.

[182] G. Lee, C. Rodriguez, and A. Madabhushi. Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(3):368–384, 2008.

[183] J. He, M. Li, H. J. Zhang, H. Tong, and C. Zhang. Generalized manifold-ranking-based image retrieval. *IEEE Transactions on Image Processing*, 15(10):3170–3177, 2006.

[184] W. Bian and D. Tao. Biased discriminant euclidean embedding for content-based image retrieval. *IEEE Transactions on Image Processing*, 19(2):545 –554, 2010.

[185] H. Zhao. Combining labeled and unlabeled data with graph embedding. *Neurocomputing*, 69(16-18):2385 – 2389, 2006.

[186] X. He. Laplacian regularized d-optimal design for active learning and its application to image retrieval. *IEEE Transactions on Image Processing*, 19(1):254–262, 2010.

[187] X. He and P. Niyogi. Locality preserving projections. In *Neural Information Processing Systems*, 2003.

[188] Y. Bengio, J. F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering. In *In Advances in Neural Information Processing Systems*, pages 177–184, 2003.

[189] M. Bober. MPEG-7 visual shape descriptors. *IEEE Trans. Circuits Systems*, 2001.

[190] M. Belkin and P. Niyogi. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural Compuation*, 15:1373–1396, 2003.

[191] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):214–225, 2004.

[192] M. Ouimet and Y. Bengio. Greedy spectral embedding. In *AISTATS*, pages 261–268, 2005.

[193] K. Zhang and J. T. Kwok. Density-weighted Nyström method for computing large kernel eigensystems. *Neural Computation*, 21:121–146, 2009.

[194] P. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational Applied Mathematics*, 20:53–65, 1987.

[195] H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):286 –299, feb. 2007.

[196] H. Chang and D. Yeung. Robust locally linear embedding. *Pattern Recognition*, 39(6):1053 – 1065, 2006.

[197] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.

[198] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[199] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models their training and application. *Computer Vision & Image Understanding*, 61:38–59, 1995.

[200] T. G. Dietterich. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*, pages 1–15. Springer-Verlag, 2000.

[201] A. L. N. Fred and A. K. Jain. Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:835–850, 2005.

[202] J. H. Park, Z. Zhang, H. Zha, and R. Kasturi. Local smoothing for manifold learning. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–452 – II–459 Vol.2, 2004.

[203] B. Geng, C. Xu, D. Tao, L. Yang, and X. S. Hua. Ensemble manifold regularization. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2396–2402, 2009.

[204] C. Hou, C. Zhang, Y. Wu, and F. Nie. Multiple view semi-supervised dimensionality reduction. *Pattern Recognition*, 43(3):720–730, 2010.

[205] J. Jia, X. Xiao, B. Liu, and L. Jiao. Bagging-based spectral clustering ensemble selection. *Pattern Recognition Letters*, 32:1456–1467, 2011.

[206] P. Tiwari, M. Rosen, and A. Madabhushi. Consensus-locally linear embedding (c-lle): Application to prostate cancer detection on magnetic resonance spectroscopy. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*, volume 5242, pages 330–338, 2008.

[207] C. Wang and S. Mahadevan. Manifold alignment using procrustes analysis. In *Proceedings of the 25th international conference on Machine learning*, pages 1120–1127, 2008.

[208] Z. Tu. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1589 –1596, 2005.

[209] T. G. Dietterich. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923, 1998.

[210] H. G. Welch, E. S. Fisher, D. J. Gottlieb, and M. J. Barry. Detection of prostate cancer via biopsy in the medicareseer population during the psa era. *Journal of the National Cancer Institute*, 99(18):1395–1400, 2007.