ACCURATE SEGMENTATION OF CT PELVIC ORGANS VIA INCREMENTAL CASCADE LEARNING AND REGRESSION-BASED DEFORMABLE MODELS

Yaozong Gao

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science in the University of North Carolina at Chapel Hill.

Chapel Hill
2016

Approved by:

Dinggang Shen

Marc Niethammer

Yiqiang Zhan

Stephen M. Pizer

Jun Lian

ABSTRACT

YAOZONG GAO: ACCURATE SEGMENTATION OF CT PELVIC ORGANS VIA INCREMENTAL CASCADE LEARNING AND REGRESSION-BASED DEFORMABLE MODELS.
(Under the direction of Dinggang Shen.)

Accurate segmentation of male pelvic organs from computed tomography (CT) images is important in image guided radiotherapy (IGRT) of prostate cancer. The efficacy of radiation treatment highly depends on the segmentation accuracy of planning and treatment CT images. Clinically manual delineation is still generally performed in most hospitals. However, it is time consuming and suffers large inter-operator variability due to the low tissue contrast of CT images. To reduce the manual efforts and improve the consistency of segmentation, it is desirable to develop an automatic method for rapid and accurate segmentation of pelvic organs from planning and treatment CT images.

This dissertation marries machine learning and medical image analysis for addressing two fundamental yet challenging segmentation problems in image guided radiotherapy of prostate cancer.

- **Planning-CT Segmentation.** Deformable models are popular methods for planning-CT segmentation. However, they are well known to be sensitive to initialization and ineffective in segmenting organs with complex shapes. To address these limitations, this dissertation investigates a novel deformable model named regression-based deformable model (RDM). Instead of locally deforming the shape model, in RDM the deformation at each model point is explicitly estimated from local image appearance and used to guide deformable segmentation. As the estimated deformation can be long-distance and is spatially adaptive to each model point, RDM is insensitive to initialization and more flexible than conventional deformable models. These properties render it very suitable for CT pelvic organ segmentation, where initialization is difficult to get and organs may have complex shapes.

- **Treatment-CT Segmentation.** Most existing methods have two limitations when they are applied to treatment-CT segmentation. First, they have a limited accuracy because they overlook the availability of patient-specific data in the IGRT workflow. Second, they are time consuming and may take minutes or even longer for segmentation. To improve both accuracy and efficiency, this dissertation combines incremental learning with anatomical landmark detection for fast localization of the prostate in treatment CT images. Specifically, cascade classifiers are learned from a population to automatically detect several anatomical landmarks in the image. Based on these landmarks, the prostate is quickly localized by aligning and then fusing previous segmented prostate shapes of the same patient. To improve the performance of landmark detection, a novel learning scheme named "incremental learning with selective memory" is proposed to personalize the population-based cascade classifiers to the patient under treatment. Extensive experiments on a large dataset show that the proposed method achieves comparable accuracy to the state of the art methods while substantially reducing runtime from minutes to just 4 seconds.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my advisor Dr. Dinggang Shen for his continuous support of my PhD study during the past five years. I am grateful to Dinggang for introducing me to the field of machine learning and medical image analysis and for improving and refining my writing skills. I am also grateful to him for giving me freedom to explore my own ideas and pursue various projects without objection. His guidance, encouragement and support are always priceless to me.

I would also like to thank my committee members: Dr. Yiqiang Zhan for his insightful advice on this research and his guidance on my career development; Dr. Marc Niethammar for offering me the opportunity to study at University of North Carolina and encouraging me during the five years; Dr. Stephen Pizer for his valuable feedback to this research and considerable effort on reading and revising the thesis draft; Dr. Jun Lian for his expertise in image guided radiotherapy and serving as my tutor in imaging techniques.

I would like to thank the co-authors and collaborators of my various publications, including Dr. Shu (Billy) Liao, Dr. Li Wang, Dr. Feng Shi, Dr. Guorong Wu, Dr. Yanrong Guo and all the people who ever worked in the IDEA lab of BRIC. In particular, Billy taught me a lot in the first year of my PhD study; Li helped extend my work on random forest to brain applications and achieved exciting results on several segmentation challenges; Guorong provided invaluable discussions on anatomical landmark detection and brain image registration.

I would like to thank my fellow students in the department of computer science, who helped me get through these years. To name a few, Yi Hong, Tian Cao, Xiao Yang, Ilwoo Lyu, etc. And special thanks to Yu Meng and Dong Nie, who work with me in the same lab. Without you, I may feel lonely working as a research assistant outside the computer science department.

I am also grateful to all members of our Dota team including Dr. Feng Shi, Dr. Li Wang, Dr. Jian Cheng, Dr. Rui Min, Dr. Jun Zhang, Dr. Liye Wang and Dr. Yinghuan Shi. I won't forget the nights we fought together in the Chateau Apartments 524. Without you, my PhD life won't be this fascinating.

Last, but no the least, I would like to thank my family members. My parents, Mr. Guosheng Gao and Mrs. Yuee Xu, have been financially and mentally supportive in every decision I have made. My wife, Jingbo Chen, sacrificed her promising career and came to a world with a different language in order to stay with me. My daughter, little Ariel, has been wonderfully understanding during my dissertation and job hunting process. I wouldn't have reached this far without their constant and unconditionally support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AP	Apex Center
AT	Anterior Point
ASD	Average Surface Distance
ASM	Active Shape Model
BS	Base Center
CDM	Classification based Deformable Model
CT	Computed Tomography
DSC	Dice Similarity Coefficient
FOV	Field of View
IGRT	Image-guided Radiotherapy
IL	Incremental Learning
ILSM	Incremental Learning with Selective Memory
LF	Left Lateral Point
MIX	Mixture Learning with Patient-specific and Population Data
MR	Magnetic Resonance
PC	Prostate Center
PCA	Principal Component Analysis
POP	Population Learning
PPAT	Pure Patient-specific Learning
PPV	Positive Predictive Value

PT	Posterior Point
SEN	Sensitivity
SBRT	Stereotactic Body Radiation Therapy
RANSAC	Random Sample Consensus
RDM	Regression-based Deformable Model
RT	Right Lateral Point
ToC	Table of Contents

CHAPTER 1 : INTRODUCTION

1.1 Image Guided Radiotherapy

Prostate cancer is a common type of cancer in American men. It is also the second leading cause of cancer death in American men [American Cancer Society, 2015]. When a patient is diagnosed with prostate cancer in the early stage, image guided radiotherapy (IGRT) is usually recommended as one of the effective treatments for prostate cancer. IGRT consists of a planning stage followed by a treatment stage, as illustrated in fig. 1.1. *In the planning stage*, a computed tomography (CT) scan called a planning CT is acquired from the patient. Radiation oncologists then delineate the target (the prostate and sometimes the seminal vesicles) and nearby organs at risk on the CT scan; frequently this is done manually. Based on the organ delineations, a treatment plan is designed with the goal of delivering the prescribed dose to the target volume while sparing nearby healthy organs such as the bladder, rectum and femoral heads. *The treatment stage* typically lasts about several weeks with typically one treatment per day. To account for daily prostate motions, a CT scan called a treatment CT can be acquired inside the radiotherapy vault at each treatment day right before the radiation therapy. Since the treatment CT captures a present snapshot of the patient's anatomy, the patient can be set up so that radiation can be aimed at the targeted area as planned. In addition, if the change of anatomy is significant, radiation oncology staff can adapt the treatment plan to optimize the distribution of radiation dose to effectively treat *current* anatomy of the prostate and avoid neighboring normal organs. Consequently,

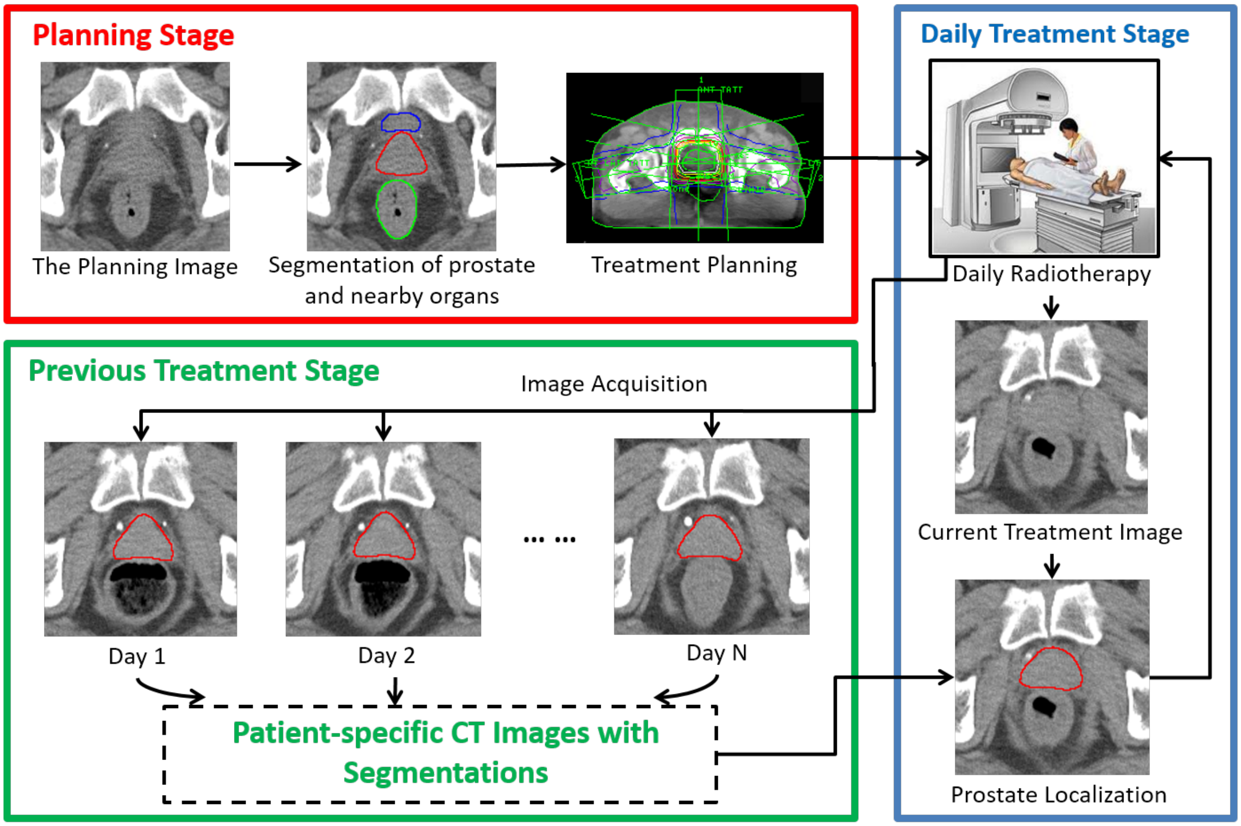


Figure 1.1: Illustration of image guided radiotherapy

IGRT increases the probability of tumor control and reduces the possibility of side effects. [Xing et al., 2006, Dawson and Jaffray, 2007].

There are two segmentation problems in the IGRT, *planning-CT segmentation* and *treatment-CT segmentation*. The efficacy of IGRT depends on the accuracy of both segmentations.

- *Planning-CT segmentation* aims to accurately segment the target (prostate) and nearby pelvic organs from CT images. In this dissertation, besides the prostate, the pelvic organs of interest include the bladder, rectum and two femoral heads. As the segmentations of these pelvic organs are used for treatment planning, their segmentation accuracy is critical for IGRT.

- *Treatment-CT segmentation* aims to accurately and quickly localize the prostate in daily treatment CT images. The segmentation can be used for three purposes. 1) The segmentation can be used to guide radiation treatment. Based on the prostate segmentation, the treatment plan can be aligned from the planning image space to the current treatment image space for precisely targeting the current anatomy of the prostate. 2) The segmentation can be used to calculate the dose accumulation. By deformably registering daily treatment images to the planning image space based on the segmented structures, the accumulation of radiation dose over the past treatment period can be calculated in the pelvic region. This dose accumulation provides feedbacks in the adaptive radiotherapy that can be used to modify the treatment plan for improving radiation treatment at follow-up fractions. 3) The segmentation can be used to tell whether a significant change of anatomy happens and whether a re-optimization of treatment plan is necessary. In this dissertation, the first purpose is the main focus of treatment-CT segmentation. As the segmentation is used to guide the daily treatment, besides accuracy, efficiency is also important for treatment-CT segmentation. If an algorithm is computationally expensive, the anatomical structures in the area of interest may have changed during the computation, which could invalidate the purpose of segmentation.

In most clinical practices, manual delineation is usually adopted in both planning and treatment-CT segmentation. However, it is often a time-consuming and labor-intensive process, which typically takes 25-35 minutes for an experienced radiation oncologist to delineate the target (prostate) and four major pelvic organs at risk. Moreover, manual delineation often suffers large inter-operator variability [Foskey et al., 2005, Lay et al., 2013]. Therefore, it is clinically desirable to develop a robust, accurate and automatic algorithm for planning-CT and treatment-CT segmentation.

The following sections are organized as follows. Section 1.2 presents the challenges in developing automatic methods for segmenting male pelvic organs in CT images. Section 1.3 and section 1.4 summarize the existing methods for planning-CT segmentation and treatment-CT segmentation, respectively. Their limitations are also discussed. Section 1.5 presents the contributions of this dissertation in both planning and treatment-CT segmentation. Section 1.6 gives a brief overview of the remaining chapters. The summary of this chapter is given in section 1.7.

1.2 Challenges of Automatic Segmentation

It is generally difficult to automatically segment male pelvic organs from CT images due to three challenges as illustrated in fig. 1.2. 1) Certain parts of pelvic organ boundaries exhibit low contrast in CT images, such as the prostate boundaries, the rectum boundaries and the touching boundaries between the bladder and the prostate. 2) The shapes of the bladder and rectum are highly variable. They can change significantly across patients and between CTs of one patient himself due to different amounts of urine in the bladder and bowel gas in the rectum. 3) Not only is the shape of the rectum variable due to the bowel gas, but also is the appearance of the rectum.

Besides the above challenges, pelvic CT images often have large diversity. For example, 1) pelvic CT images are often acquired with different fields of view, which causes substantial variation of volume dimensions and organ positions; 2) contrast agents may be injected into some patients before image acquisition, which may partially or fully brighten the bladder in CT images; 3) fiducial markers or the catheter may be implanted into patients, thus changing

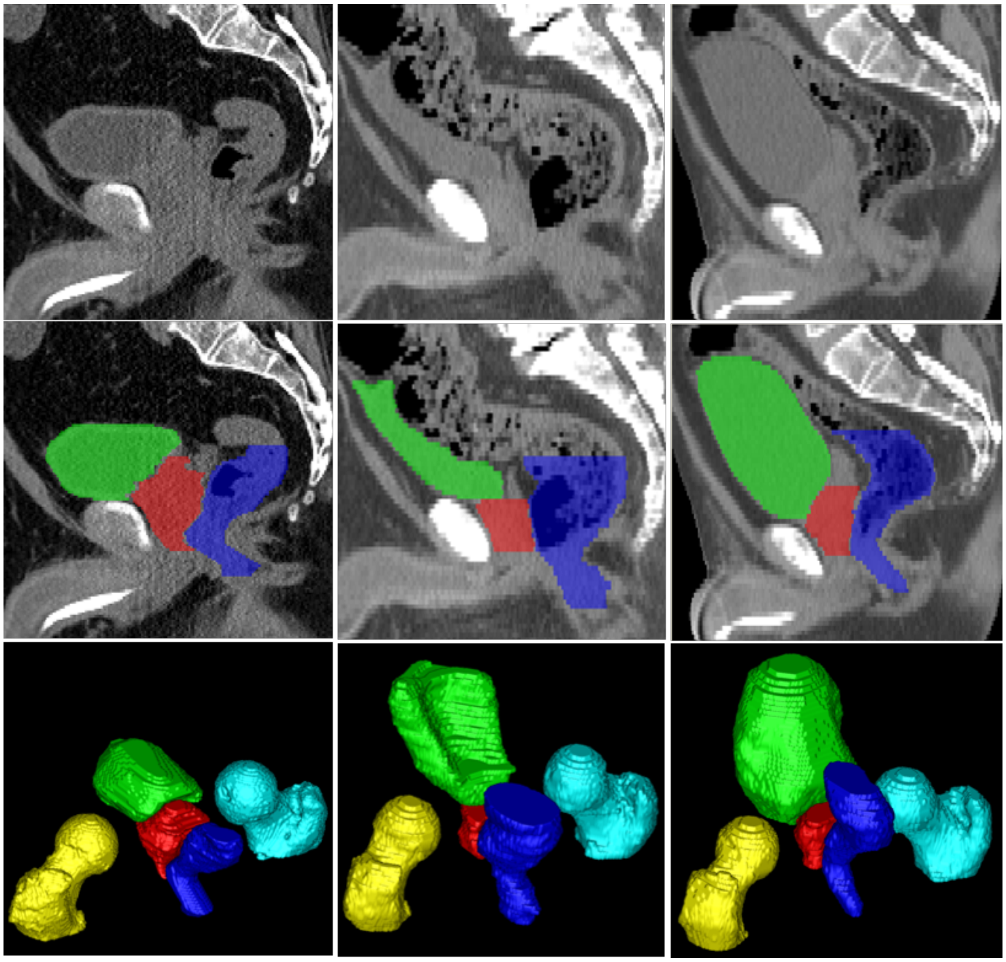


Figure 1.2: Typical CT scan slices and their pelvic organ segmentations. The three columns indicate images from three patients. The first, second, and third rows show, respectively, a sagittal CT slice, the same slice overlaid with segmentations, and a 3D view of segmentations of each patient. Red: Prostate; Green: Bladder; Blue: Rectum; Yellow: Left femoral head; Cyan: Right femoral head.

the textures of major pelvic organs. The diversity of pelvic CT images, in addition to the anatomical variations, further complicates the segmentation of male pelvic organs from CT images.

1.3 Planning-CT Segmentation

The segmentation of male pelvic organs from planning CT images has been investigated for a long time. Most developed methods fall into the category of deformable model based segmentation. Few methods were based on deformable image registration. The reasons are attributed to the diversity of pelvic CT images and large anatomical variation. As shown in fig. 1.2, large anatomical variations cause significant differences in shapes and appearances of pelvic organs across subjects. These make registration between CT images of different subjects very difficult. In addition, CT images of different subjects may be acquired with different fields of view, with/without contrast agent, and with/without metal implants. These differences make the correspondence detection challenging even in image registration. In contrast, deformable models suffer less from these problems, once they are well initialized. They can potentially overcome problems caused by image noise and artifacts by imposing the global shape constraint during segmentation. Besides, deformable models are usually more efficient than deformable image registration, since most of them only operate on the organ boundary instead of the entire image domain. These reasons make deformable models popular in planning-CT segmentation.

1.3.1 Previous Work

In most deformable models, a shape model is iteratively deformed toward the organ boundary by maximizing an objective function, which typically consists of an image matching term and a shape matching term. The image matching term measures how well the image appearance around the shape model matches the expectation learned from segmented training images, and the shape matching term measures how plausible the shape model is in terms of local smoothness and global shape. Various image and shape matching terms have been proposed in the literature to improve the accuracy and robustness of deformable models in CT pelvic organ segmentation. For example, [Freedman et al., 2005] proposed to match the intensity distribution inside the organ for prostate segmentation. To consider the spatial relationship between nearby organs, many methods imposed additional constraints in the shape matching term to improve the robustness of segmentation. For example, [Rousson et al., 2005] proposed a Bayesian formulation that considers the non-overlapping constraint to segment the prostate and rectum. [Pizer et al., 2005] proposed a medial shape model named *M-reps* for joint segmentation of the prostate, bladder and rectum. [Costa et al., 2007] proposed coupled 3D deformable models to segment the prostate and bladder by considering an asymmetric non-overlapping constraint. [Chen et al., 2011] incorporated anatomical constraints from pelvic bones into a Bayesian framework for jointly segmenting the prostate and rectum. Instead of using intensity/gradient to define the image matching term, recently machine learning techniques have been proposed to characterize the matching of organ boundary. For example, [Lu et al., 2012] detected the boundaries of pelvic organs by using probabilistic boosting trees together with a Jensen-Shannon divergence-based measurement. To improve the robustness of deformable model initialization, [Lay et al., 2013] proposed to learn global image context for fast localization of pelvic organs. The initialized shape model is then refined iteratively by a discriminantive learned boundary detector, similar to the way done in [Lu et al., 2012].

Beside the above published work, deformable models have also been implemented in commercial software for planning-CT segmentation. For example, Morphormics Inc. developed a tool called *mxStructure* that automatically contours the pelvic organs for treatment planning. According to [Pizer, 2016], the segmentation tool is based on a skeleton shape model and an undescribed appearance model; the tool can accurately segment the pelvic organs within one minute. Due to its accuracy and speediness, mxStructure has already been deployed in many hospitals to assist treatment planning. The Phillips company developed a treatment planning system called *Pinnacle* [Koninklijke Philips N.V., 2016] that provides a module for automatic segmentation of pelvic organs in CT images. The method implemented in the module is based on deformable models that use image gradients to drive shape models onto organ boundaries. If interested, readers may refer to [Chaney and Pizer, 2016] for detailed descriptions of the use of deformable models in commercial products.

1.3.2 Limitations of Previous Work

In deformable models, the image matching term is often defined using intensity profile, gradient profile, regional histogram of intensity and regional histogram of gradient. Besides, the quantitle function was also used in the literature [Broadhurst et al., 2006]. Compared to the regional histogram, the quantitle function shows better properties for statistical analysis, such as for principal component analysis. All these definitions of the image matching term work well for segmenting organs with distinctive intensity patterns and clear boundaries. However, their performance is limited in the segmentation of CT pelvic organs, because 1) the intensity distributions of different pelvic organs can be similar, and 2) the boundaries of pelvic organs are unclear. While these limitations can be overcome by using a classifier to learn discriminative boundary patterns of pelvic organs [Lu et al., 2012, Lay et al., 2013], the existing deformable models still suffer several intrinsic problems, which make them ineffective in the planning-CT segmentation.

- **Initialization.** Deformable models are sensitive to initialization. In deformable models, the shape model is deformed locally around the initialization, and large deformations are often penalized in the objective function. As a result, the performance of existing deformable models highly depend on the position, size and shape of the initialized model. A good initialization often leads to a good segmentation accuracy.

However, in the case of pelvic organ segmentation, where inter-subject anatomical variations are large, it is often difficult to obtain a good initialization. Therefore, the performance of existing deformable models is limited.

- **Tubular Organs.** Deformable models have difficulty in segmenting tubular organs, such as the rectum. The contributing reasons are 1) initialization and 2) local search range. First, it is hard to initialize a shape model for tubular organs due to their large shape variation. Second, it is tricky to find an appropriate value for local search range during shape deformation. A small local search range prevents sufficient deformations for an initialized shape model to attach the organ boundary, while a large local search range can easily cause mesh folding or shrinkage because the left boundary of shape model may find high boundary responses on the right tube wall.

To overcome the aforementioned limitations, it is necessary to propose a new deformation mechanism for deformable models that is robust to arbitrary initialization and flexible to segment tubular organs.

1.4 Treatment-CT Segmentation

The above methods for planning-CT segmentation can be directly applied for treatment-CT segmentation. However, their performance is limited compared to methods specially designed for treatment-CT segmentation. The major difference between planning- and treatment-CT segmentation lies in the availability of patient-specific data. Specifically, when segmenting a new treatment CT image, there exist an already segmented planning CT image and previous treatment CT images of the same patient (fig. 1.1). Since intra-patient shape and appearance variations are less pronounced than inter-patient variations, an algorithm can exploit these additional data to learn patient-specific characteristics and improve the segmentation accuracy.

It is noteworthy that instead of acquiring a treatment CT image the prostate location at each treatment day can also be accurately approximated using implanted markers. Such marker tracking techniques, using either radiofrequency or 2D planar images, can reduce the radiation dose to the patient from CT imaging. However, the lack of volumetric image data prevents calculation of delivered dose on structures as desired in adaptive radiotherapy. To overcome this problem, [Lee et al., 2010] proposed to estimate the treatment image by mapping planning image data to the treatment space via the deformation field estimated using the implanted markers. They showed that the calculated dose histograms using the estimated images are close to those using real treatment images. However, this technique is still under research and has not been widely adopted in the clinic. Besides, markers need to be implanted inside the prostate, which may cause complications such as urinary tract infection [Shinohara and Roach, 2007]. Therefore, this dissertation still considers the conventional scenario where a treatment image is acquired for localizing the prostate at each treatment day.

1.4.1 Previous Work

Most methods developed for treatment-CT segmentation fall into the category of either deformable registration or voxel-wise labeling. In the literature, few research [Feng et al., 2009] adopted deformable models for treatment-CT segmentation. The major reason comes from the ease of registration between CT images of the same subject. By registering the previous images of the same subject to a new treatment image, the segmentions on the previous images can also be aligned onto the new treatment image space and then used for prostate localization. Besides, the registration also benefits voxel-wise labeling, which can be performed only in a small region around the roughly localized prostate. It not only increases computational efficiency but also allows classifiers to be specifically trained for labeling voxels near the prostate, thus improving labeling accuracy. In the following paragraphs, existing methods based on deformable registration and voxel-wise labeling are respectively discussed.

- **Deformable registration** has been investigated in the medical image analysis community for many years as a way to align the corresponding structures between two images. It is popular in treatment-CT segmentation. For example, [Foskey et al., 2005] proposed to localize the prostate by deformably registering the segmented planning CT of the same patient to the current treatment image. To address the challenge caused by the bowel gas, they designed a deflation method to explicitly eliminate the bowel gas before registration. [Lu et al., 2011] proposed to integrate deformable segmentation and registration into a single framework for segmenting pelvic organs in treatment CT images. In their framework, the segmentation module considers not only the average organ intensity and shape prior but also the segmentation likelihood derived by registering the segmentation from the planning image to the treatment image. Similarly, their registration module matches not only the image intensity but also the tentative organ segmentation result. The segmentation and registration steps are alternately conducted until convergence. Besides using intensity-based registration, several methods learned informative features to guide registration. For example, [Liao and Shen, 2012] proposed to select informative voxels and features from patient-specific image data and use them to guide deformable registration. During the feature selection step, salient regions but irrelevant to the prostate localization, such as the region filled with the bowel gas, are automatically filtered out, which makes registration more robust. To account for registration errors, [Liao et al., 2013] further proposed a patch-based label fusion framework, which uses sparse representation to identify similar voxels from warped CT images of the same patient and propagates only their labels for prostate localization.

- **Voxel-wise labeling** labels each voxel in the image based on local image appearance. It learns a strong classifier to distinguish voxels inside the target organ (positives) from those outside (negatives), according to the segmented training images. Once learned, the classifier is applied voxel-wisely to produce an organ likelihood map for a testing image, where the target organ is enhanced and can be easily segmented by either thresholding or simple segmentation methods. For example, [Li et al., 2012] proposed to utilize context information to iteratively refine the voxel-wise labeling result. Then, a level set was used to segment the prostate from the labeling map. [Gao et al., 2012a] proposed a sparse representation based classifier and employed multi-atlas labeling for prostate segmentation. To utilize valuable information from manual interactions, [Shi et al., 2013] proposed a semi-supervised learning framework that learns a classifier by integrating information from both manual interactions and previous segmented image data.

1.4.2 Limitations of Previous Work

Different from planning-CT segmentation that is conducted offline, if the segmentation is to be used to affect the current radiation treatment, treatment-CT segmentation has to be performed online when a patient is on the treatment bed awaiting his current treatment. Therefore, treatment-CT segmentation demands higher segmentation efficiency than planning-CT segmentation. However, the existing methods are too slow to meet this need. For example, in deformable registration based methods it typically takes minutes or even longer to register an atlas to a treatment CT image. In the case when multiple atlases are used, the time for registration would be even longer. In voxel-wise labeling, efficiency is often limited by using a complex classifier [Gao et al., 2012a] or performing iterative classification refinement [Li et al., 2012]. The long localization procedure makes the existing methods inapplicable to image guided radiotherapy, although they are still useful in the adaptive radiotherapy where the segmentation can be conducted offline.

Besides expensive computations, voxel-wise labeling methods [Li et al., 2012, Gao et al., 2012a, Shi et al., 2013] suffer another limitation. They require at least three patient-specific images manually segmented for learning a classifier. This requirement imposes additional burdens on radiation oncologists, as manual segmentation is time consuming (10 minutes for the prostate). Moreover, there may be no sufficient patient-specific data available, especially in the beginning treatment day when only one planning image is available.

To overcome the above limitations, it is necessary to develop a fast segmentation method that meets the following requirements. The method should rely on little manual delineation and be robust when the amount of patient-specific data is limited, such as in the beginning of treatment days.

1.5 Thesis

Thesis: *Deformable models benefit in accuracy from explicitly learning deformations from image appearance. Landmarks can be utilized for fast and accurate segmentation of treatment CTs by effectively combining limited patient-specific data with massive population data in the* *cascade learning framework.*

This dissertation investigates solutions to address the limitations of existing methods in planning-CT and treatment-CT segmentations, and it proposes new algorithms for accurate and efficient segmentation of male pelvic organs from CT images. The dissertation focuses on improving not only the accuracy of segmentation algorithms but also the processing efficiency. In particular, two specific aims are proposed.

- **Specific Aim 1 (Planning-CT segmentation).** *The accuracy of deformable models can be improved by explicitly learning deformations from image appearance.* Many conventional deformable models rely on local search to drive shape models onto organ boundaries. This deformation mechanism makes them sensitive to initialization and also limits their flexibility to segment tubular organs, such as the rectum. Since image appearance is informative to the anatomical location in the image, the direction and distance from any voxel to the boundary of target organ can be potentially predicted based on the appearance of local image patch. This information can be used as deformation to guide the move of shape model toward the organ boundary, which could effectively address the limitations of conventional deformable models.

- **Specific Aim 2 (Treatment-CT segmentation).** *Landmarks can be utilized for fast and accurate segmentation of treatment CTs by effectively combining limited patient-specific data with massive population data in the cascade learning framework.* Most conventional methods overlook the availability of patient-specific data in the treatment-CT segmentation, which limits their performance. By exploiting the patient-specific data, the existing methods obtain high segmentation accuracy at the expense of high computational complexity and much manual effort for annotation. To improve efficiency and reduce effort of manual annotation, anatomical landmarks can be used for segmentation, since they are efficient to detect and easy to annotate. In landmark-based segmentation, anatomical landmarks are first detected on a new image and then used to guide the registration between existing segmented images and the new image. After registration, the existing segmentations are aligned onto the new image space, where the final segmentation is obtained by a label fusion method, such as the majority voting. The efficiency and accuracy of landmark-based segmentation depend on the landmark detection. While it is efficient to detect landmarks using the cascade learning framework, the accuracy could be limited if the landmark detectors are learned from massive population data because of large inter-patient anatomical variation. On the other hand, it is also infeasible to learn from limited patient-specific data; doing so tends to suffer from the overfitting problem. To this end, an effective strategy should be explored to combine limited patient-specific data with massive population data in the cascade learning framework.

In order to support the thesis and the above two specific aims, the detailed contributions of this dissertation include

- A novel deformable model, namely a *regression-based deformable model*, is proposed to hierarchically deform a shape model onto the target organ boundary based on an explicitly learned deformation field; **(Aim 1)**
- An *auto-context model* is adopted to iteratively refine the predicted deformation field by gradually incorporating the neighborhood prediction information; **(Aim 1)**
- A *multitask random forest* is proposed to learn the deformation from local image ap-

pearance by coupling deformation regression and organ classification in a common random forest; **(Aim 1)**

- A *multi-resolution strategy* is adopted to segment multiple pelvic organs from CT images, where the coarse-level deformation fields are jointly estimated for all organs to consider their spatial relationship and where the fine-level deformation fields are separately estimated for each organ to make the respective prediction models specific; **(Aim 1)**

- *Extensive experiments* on a large prostate CT dataset ($>$ 300 patients) show that the proposed method can accurately segment the prostate, bladder, rectum and two femoral heads from planning CT images and that it outperforms many existing methods in this task; **(Aim 1)**

- The *cascade learning framework* is adapted to address the problem of unbalanced training samples in the classification-based landmark detection. It can efficiently localize a landmark in a 3D medical image volume within one second using a multi-resolution implementation; **(Aim 2)**

- An incremental learning scheme, namely *incremental learning with selective memory*, is proposed to update the existing landmark detector learned from massive population data with limited patient-specific data. It can be used to personalize the population-based landmark detectors to a specific patient; **(Aim 2)**

- A schematic illustration is provided to explain *the mechanism* behind incremental learning with selective memory; **(Aim 2)**

- *Random sample consensus (RANSAC)* is used to align the previous segmentations of the same patient onto the target treatment image by considering the possibility of mis-detected landmarks; **(Aim 2)**

- *Extensive experiments* on a large prostate CT dataset ($>$ 400 treatment CT images) show that the proposed method is able to accurately localize the prostate in treatment CTs within 4 seconds; the method satisfies the accuracy and efficiency requirement of IGRT. **(Aim 2)**

1.6 Overview of Chapters

The remaining chapters of this dissertation are organized as follows.

- Chapter 2 presents the background of related techniques and evaluation metrics used in the dissertation. The techniques include random forests and deformable models. Under random forests, the basics and mathematical notations are presented, followed by the application of random forests to multi-class classification and multi-variate regression problems. Under deformable models, the active shape model is elaborated, as it is closely related to the proposed regression-based deformable model. Finally, several evaluation metrics are given. They are used to evaluate the proposed segmentation methods and compare them with other existing methods.

- Chapter 3 presents regression-based deformable models (RDM) for planning-CT segmentation. Different from conventional deformable models, RDM explicitly learns a deformation field to guide deformable segmentation. The learned deformation field is able to overcome the sensitivity of deformable models to initialization, and also it is able to improve their flexibility to segment tubular organs. To learn a reliable deformation field, two techniques are presented. First, *an auto-context model* is proposed to iteratively refine the estimated deformation field by exploiting local structured information. Second, *a multitask random forest* is proposed to couple deformation regression with organ classification. Compared to a random forest trained only for deformation regression, the multitask random forest is able to improve the robustness of deformation field estimation by exploiting information from organ classification. Extensive experimental results are given to evaluate each design of the segmentation method and to show the superior performance of the proposed method over several existing methods.

- Chapter 4 presents a fast landmark-based approach for treatment-CT segmentation. To efficiently detect a landmark, the detection problem is formulated as a binary classification problem, where positives and negatives are voxels near and far away from the annotated landmark, respectively. To handle the highly imbalanced training samples (i.e., limited positives and unlimited negatives), a cascade learning framework is presented to gradually separate negatives from positives. Due to large inter-patient anatomical variations, the classic population-based cascade learning doesn't perform well. To improve its performance, a novel learning scheme, namely incremental learning with selective memory (ILSM), is proposed to update cascade classifiers learned from a population with limited patient-specific data. Extensive experiments show the effectiveness of ILSM over other learning schemes. Comparing with existing methods, ILSM reduces runtime to seconds while maintaining competitive segmentation accuracy.

- Chapter 5 concludes the dissertation and discusses the limitations of the proposed methods as well as future work. In the conclusion, the methods proposed in this dissertation for planning-CT and treatment-CT segmentation are briefly summarized. Their limitations are also discussed. In the future work, interesting future directions are discussed, which include several potential strategies to improve the proposed methods.

1.7 Summary

Planning-CT and treatment-CT segmentation plays important roles in IGRT. While much effort has been devoted to solving them, the existing methods suffer either limited segmentation accuracy or high runtime complexity. This chapter reviewed the existing methods for planning-CT and treatment-CT segmentation and discussed their limitations. In particular, popular deformable model based methods were reviewed for planning-CT segmentation. Their sensitivity to initialization and inflexibility to segment tubular organs were discussed. In treatment-CT segmentation, this chapter summarized popular methods from deformable registration to voxel-wise labeling. Most methods are computationally expensive, which hinders their practical use in IGRT. Besides, voxel-wise labeling methods require sufficient patient-specific data for training, which may not be feasible in the beginning treatment days. To overcome these limitations, this chapter sketched the solutions proposed in the dissertation. Specifically, in planning-CT segmentation, a novel deformable model (RDM) was proposed to address the limitations of conventional deformable models by explicitly learning deformation from image data. In the treatment-CT segmentation, a landmark-based approach was proposed for fast prostate localization. To improve the accuracy of landmark detection, a novel learning scheme (ILSM) was proposed to gradually update population landmark detectors with patient-specific data collected during radiotherapy. Finally, this chapter, as the first chapter, provided a high-level overview of the remaining chapters.

CHAPTER 2 : BACKGROUND

2.1 Random Forests

Random forests are general machine learning methods for supervised learning, e.g., classification and regression. Random forests are popular in computer vision and medical image analysis [Criminisi and Shotton, 2013] due to their high efficiency and good scalability. A random forest consists of multiple binary decision trees. Each decision tree consists of two types of nodes: split node and leaf node. Fig. 2.1 gives the visualization of the two types of nodes in a binary decision tree. The split nodes are interior nodes in a decision tree. Each of them is associated with a binary split function, which routes a given sample either to its left or right child node based on a tuple of descriptors. These descriptors are called features in the machine learning field. The leaf nodes are terminal nodes in a decision tree. Each of them stores the information of training samples routed to it. This information is used for prediction of a new testing sample.

The following subsections are organized as follows. Sections 2.1.1 and 2.1.2 introduce the training and application of random forests, respectively. Section 2.1.3 presents the use of random forests for multiclass classification and multivariate regression. Finally, section 2.1.4 briefly discusses the advantages of random forests over other machine learning methods in the field of image analysis.

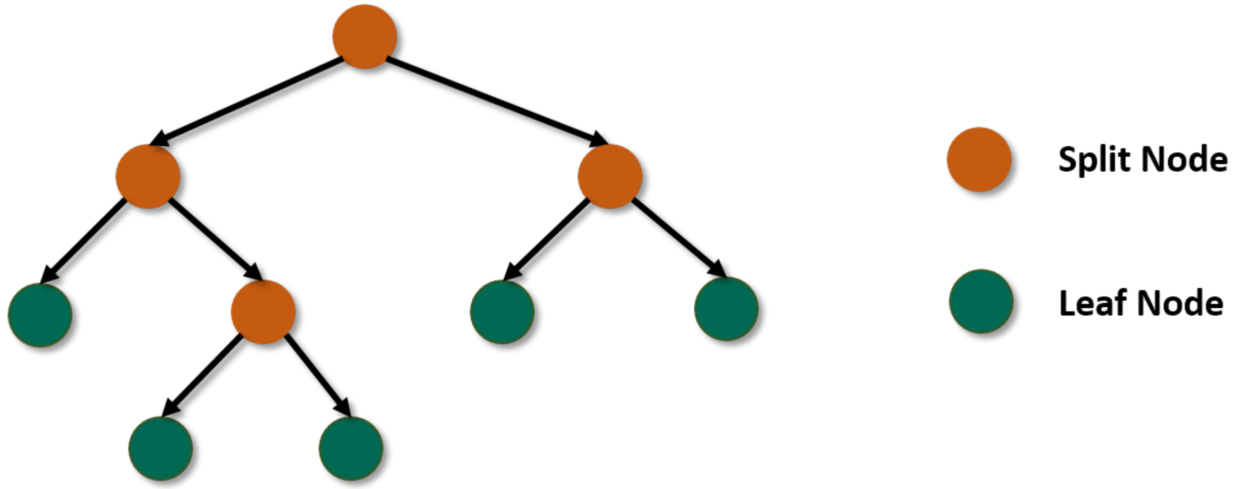


Figure 2.1: A decision tree in a random forest.

2.1.1 Training of Random Forests

As an ensemble model, each decision tree in a random forest is trained independently. To increase the diversity of decision trees, a different training subset is randomly sampled with replacement from the entire training set to train each tree. Studies [Breiman, 2001, Liu et al., 2005] show that high diversity prevents overfitting and usually leads to lower generalization error.

Decision Tree Prediction. Given a sampled training set, a decision tree is trained recursively starting with the root node. Each node learns the optimal split function that separates the arrival training set into two subsets by maximizing the purity of each split subset. Mathematically, the optimal split function is found by maximizing the following objective function:

$$\arg\max_{\phi\in\Phi} \frac{1}{|\mathcal{S}_{\mathrm{L}}|}\mathcal{P}(\mathcal{S}_{\mathrm{L}}) + \frac{1}{|\mathcal{S}_{\mathrm{R}}|}\mathcal{P}(\mathcal{S}_{\mathrm{R}}), \tag{2.1}$$

$$\mathcal{S}_{\mathrm{L}} = \{s \in \mathcal{S} | f(s|\phi) = 0\}, \qquad \mathcal{S}_{\mathrm{R}} = \{s \in \mathcal{S} | f(s|\phi) = 1\}, \tag{2.2}$$

where $\mathcal{S}$ is the training set arriving at this node, $\mathcal{S}_{\text{L}}$ and $\mathcal{S}_{\text{R}}$ are the subsets split to the left and right child nodes, respectively, $\mathcal{P}(.)$ calculates the purity of a training set, $f(s|\phi)$ is a binary split function with parameters ϕ, and Φ is a candidate parameter set. In the classic random forests, a decision stump is used as the split function due to its efficiency. Mathematically, a decision stump is formulated as $f(s|\phi) = s(i) > t$, where i is a feature index, t is a threshold, and $s(i)$ extracts the i-th feature of sample s.

To solve eq. 2.1, a random set of split functions is first generated, e.g., by randomizing feature index i and threshold t, and then exhaustive search is used to find the optimal split function in the random set that maximizes eq. 2.1. Afterwards, the training set $\mathcal{S}$ is divided into two subsets $\mathcal{S}_{\text{L}}$ and $\mathcal{S}_{\text{R}}$ according to the learned split function, and then the same procedure is performed to further split each subset into smaller ones with more purity. This recursive process stops when 1) the decision tree reaches a predefined maximum tree depth, or 2) the training set $\mathcal{S}$ is too small to be split, or 3) the purity within a node is above a threshold. When a split stops, the corresponding node is made as a leaf node, and the training set that arrives the node is stored there. Practically, it is memory inefficient to store training samples in the leaf nodes. Therefore, only the task-specific statistics of a training set are stored.

2.1.2 Application of Random Forests

Given a target sample, the prediction of each decision tree in a random forest is independent. The final prediction of a random forest is the average prediction over all decision trees.

Decision Tree Application. A target sample s is first pushed to the root node of a decision tree, and then it is guided to a leaf node by the split function associated with each split node.

Specifically, the testing sample is routed to the left child node if $f(s|\phi) = 0$ and to the right child node if $f(s|\phi) = 1$. When a leaf node is reached, the task-specific statistics stored in it are retrieved for prediction.

2.1.3 Random Forest Classification and Regression

Random forests are general to many supervised learning tasks [Criminisi et al., 2011]. To adapt a random forest to a specific task, the purity function $\mathcal{P}(.)$ and the task-specific statistics need to be defined. In this subsection, multiclass classification and multivariate regression are taken as two examples to show how random forests can be used to solve general supervised learning problems.

Multiclass Classification. Classification is the prediction of a discrete variable, called the label, from a tuple of features. In classification the purity function $\mathcal{P}(.)$ is defined based on the labels of training samples. It encourages each decision tree to partition a training set into subsets with the same label. Therefore, the purity function in the classification measures the label consistency of a training set. It can be quantified by the negative entropy $\mathcal{E}$.

$$\mathcal{E}(\mathcal{S}) = \sum_c p_c \log p_c, \tag{2.3}$$

where p_c denotes the percentage of training samples with the label c in a training set $\mathcal{S}$. The larger $\mathcal{E}(\mathcal{S})$, the purer the training set $\mathcal{S}$ is in terms of the class label.

In the classification task each leaf node stores *a label distribution* of training samples that arrive at this node. This information can be used to infer the class likelihood of a testing sample when it arrives at one leaf node. In the testing stage, the label distribution output by each tree is first averaged and then normalized to unit sum. After normalization, each entry in the label distribution indicates the likelihood of the testing sample belonging to one class. Finally, the label of the testing sample is determined as the class with the maximum likelihood.

Multivariate Regression. Regression is the prediction of a continuous variable, called the regression target, from a tuple of features. The regression target can be a scalar or a vector. In regression the purity function $\mathcal{P}(.)$ is defined based on the regression targets of the training samples. It encourages each decision tree to partition a training set into subsets with similar regression targets. Therefore, the purity function in regression measures the consistency of regression targets in a training set. It can be quantified by the summation of negative variances at different dimensions of regression target.

$$\mathcal{V}(\mathcal{S}) = -\sum_{k} v_k, \tag{2.4}$$

where v_k measures the variance of regression targets at the k-th dimension in a training set $\mathcal{S}$. The larger $\mathcal{V}(\mathcal{S})$ is, the purer the training set $\mathcal{S}$ is in terms of the regression targets.

In the regression task each leaf node stores *the average regression target* of training samples that arrive at this node. In the testing stage, when a testing sample arrives at a leaf node of one decision tree, the average regression target is retrieved from the leaf node to serve as the prediction output for the decision tree. Given a group of decision trees in a random forest, the prediction of a forest is the averaged output across all decision trees.

2.1.4 Advantages

Random forests have many advantages over other machine learning methods for supervised learning. A few of them are listed below.

Efficiency

Most machine learning models (e.g., the support vector machine) decouples feature extraction from model prediction. Regardless of the importance, all features have to be computed before the learned model can be applied for prediction. This makes the testing time linear with the feature dimension. As the feature dimension is often large in real applications (e.g., thousands), efficiency becomes a concern for most machine learning methods. In the field of image analysis, this concern is aggravated if the prediction is performed on the voxel level.

Random forests are efficient for handling data with high dimensional features because the testing time depends only on the tree depth, which is often much smaller than the feature dimension. More importantly, features can be computed on a need basis. Because a testing sample traverses only one path from the root node to a leaf node, only features along the path need to be computed. So although the entire decision tree may take thousands of features, the maximum number of features needed in a prediction is equal to the tree depth, which is often a small number. By computing only these necessary features, random forests significantly save time for feature extraction. Besides, each decision tree can be evaluated independently; hence, random forests fit the parallel processing infrastructure. Moreover, the prediction of random forests involves only floating-point comparisons to decide which path to take; doing so makes it even faster than a simple linear model that relies on floating-point multiplications.

Scalability

Scalability is an important factor when choosing a machine learning algorithm for image analysis because each voxel is a training sample and potentially millions of training samples can be collected for training. To fit the massive training samples well, the learned model should have a reasonably large model complexity. However, a large model complexity often means a high computational complexity for most learning models.

Random forests scale well with massive training samples. As each leaf node outputs a unique prediction, the model complexity of a random forest can be approximated by the number of leaves in a decision tree. Since the number of leaves grows exponentially with the tree depth, a random forest with a limited tree depth (e.g., 20-40) is often sufficient to fit millions of training samples with a negligible impact on the runtime efficiency.

Nonlinearity

Prediction with image data often involves learning a highly nonlinear mapping from image features to either a discrete variable in the classification or a continuous variable in the regression. Due to this nature of nonlinearity, linear models often do not work well with image data. While the kernel tricks [Shawe-Taylor and Cristianini, 2004] can often be used to adapt linear models for nonlinear predictions, they don't scale well when a training dataset is large.

Compared to other nonlinear models, random forests are more adaptive to individual testing samples. They utilize different sets of features for predictions of different testing samples; doing so makes random forests more flexible to fit training data than many other nonlinear models that use the same set of features for all testing samples. This increases the flexibility of random forests in the training and also increases their performance in the testing.

Convenience

Random forests are convenient to use in practice. Different from many learning algorithms, random forests don't require the input features to be normalized. This property makes it easy for random forests to integrate features from multiple sources. The reason for not requiring feature normalization is because each split node uses only one feature and each feature is used independent from others.

In addition, random forests have a limited number of parameters to tune, and the performance is quite robust to the choice of parameters due to the combination of multiple independent models. By averaging the prediction results from independent decision trees, the variance of random forests is reduced and so is the risk of overfitting.

2.2 Deformable Models

Deformable models can be classified as either parametric deformable models [Kass et al., 1988, Cohen, 1991, Cootes et al., 1995] or non-parametric deformable models (geometric deformable models) [Caselles et al., 1993, Chan and Vese, 2001]. In the former case, a shape is often represented by a set of boundary points (landmarks). Parametric deformable models allow a direct interaction with boundary points and thus are often faster than non-parametric ones [Xu et al., 2000, Assley and Chellakkon, 2014]. However, it is difficult for them to handle topology changes during deformations. In the non-parametric case, a shape can be implicitly represented by the zero level set of a higher-dimensional scalar function. Such deformable models can naturally handle the topology changes, but they are often slower than parametric ones. Among various deformable models, this chapter focuses on a particular parametric deformable model called the "active shape model" (ASM) [Cootes et al., 1995] as it is closely related to the regression-based deformable model proposed in chapter 3. Discussions of other deformable models are beyond the scope of this dissertation.

ASM is popular in the field of medical image segmentation. In ASM a shape is represented by a collection of points. The segmentation is conducted by iteratively deforming a shape toward the boundary of target object. Different from other deformable models [Kass et al., 1988], which impose only a local smoothness constraint on the deformed shape, the deformation of ASM is constrained in a global shape space. The global shape constraint increases the robustness of segmentation and makes ASM a useful tool for organ segmentation in noisy and low-contrast medical images.

The ASM algorithm has two stages: a training stage and an application stage. The training stage is detailed in section 2.2.1, where a statistical shape space is learned from segmented training images. The application stage is described in section 2.2.2, which shows how the mean shape is iteratively deformed to fit the boundary of target object under the constraint from the learned shape space. Finally, section 2.2.3 discusses the limitations of ASM.

Since this dissertation focuses on 3D image segmentation, the concepts and algorithm of ASM are described in the 3D space. Readers interested in 2D segmentation may refer to [Cootes et al., 1995]. Besides, readers should be aware that there are many variants of ASM published in the literature. This dissertation by no means provides an exhaustive literature review of all ASM variants. The ASM algorithm described in the following sections is only a particular implementation of ASM in order to give readers a flavor of how ASM is used for segmentation.

2.2.1 Training of the Active Shape Model

The training part of the ASM algorithm aims to learn a statistical shape space that captures the mean and variation of the shapes of target object from segmented training images. Two steps are performed to learn a shape space: 1) building shape correspondence across subjects and 2) principal component analysis of correspondent shapes. Both steps are elaborated below.

Shape Correspondence

A shape in the 3D ASM is often represented by a triangle mesh. Two shapes are in correspondence if 1) they have the same number of vertices and 2) vertices with the same index correspond to roughly the same anatomical location. In the 2D ASM the shape correspondence is often built by manually annotating landmarks along the boundary of target object. However, manual annotation is burdensome for building shape correspondence in the 3D ASM, as a triangle mesh often consists of thousands of vertices. To reduce the manual efforts, an automatic procedure is necessary for building the shape correspondence in the 3D space. While there are sophisticated methods out there, e.g., those based on entropy or description length [Davies et al., 2001, Davies et al., 2010, Cates et al., 2007], the following paragraphs describe a simple method for building the shape correspondence.

The method starts with constructing a reference shape and then registers it to individual shapes for building the shape correspondence across subjects.

- **Reference Shape.** Given a set of binary segmentation images, the mean segmentation image is first constructed in three steps: 1) a template image space is defined. This can be done by simply selecting an arbitrary binary segmentation image. However, doing so may introduce biases. To overcome potential biases, the Fréchet mean image [Joshi et al., 2004, Fletcher et al., 2009] can be computed and used as the template image space; 2) all binary segmentation images are linearly aligned onto the template image space using a similarity transform; 3) all aligned binary segmentation images are voxel-wisely averaged to form a mean segmentation image. To construct the reference shape, the marching cubes algorithm [Lorensen and Cline, 1987] is first adopted to extract a dense mesh from the mean segmentation image. Afterwards, mesh decimation and remeshing are alternately performed to reduce the number of vertices to a manageable size (e.g., 1000-2000) while keeping all vertices evenly distributed on the surface. The final triangle mesh after mesh decimation is used as the reference shape.

- **Surface Registration.** To build the shape correspondence across subjects, the dense triangle mesh of each subject is first extracted from its segmentation image using the marching cubes algorithm. Then, the reference shape is non-rigidly registered to each dense mesh by a robust surface registration algorithm [Myronenko and Song, 2010]. Since all registered reference shapes come from a single source and fit individual shapes well, these shapes are in correspondence and can be used for learning a statistical shape space.

Principal Component Analysis

Given shapes that are in correspondence, they are first co-aligned into a common space by generalized Procrustes analysis [Gower, 1975]. Then, the mean shape is computed as the vertex-wise average of aligned shapes; it is subtracted from each aligned shape; and principal component analysis (PCA) is adopted to compute the variation modes by an eigen-decomposition of the covariance matrix.

$$\mathbb{C} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T, \tag{2.5}$$

$$\{c_k, \mathbf{e}_k\} = \mathrm{eign}(\mathbb{C}), \tag{2.6}$$

where N is the number of shapes, $\mathbf{u}_i$ is the i-th aligned shape, $\bar{\mathbf{u}}$ is the sample mean shape, $\mathbb{C}$ is the covariance matrix, and c_k and $\mathbf{e}_k$ are the k-th eigen-value and eigen-vector of the covariance matrix $\mathbb{C}$. The output of principal component analysis gives a statistical shape space described by a multivariate Gaussian distribution with the mean $\bar{\mathbf{u}}$ and the variation modes $\{c_k, \mathbf{e}_k\}$. In practice, only the K eigen-modes with the largest eigen-values are preserved to make the shape space compact. Although eigen-values can be affected by noise, a common practice to select K is still based on the eigen-values. Specifically, K is often selected as the minimum number of eigen-modes that account for the majority of shape variations, i.e., $\min K,\ \text{s.t.}\ (\sum_{k=1}^{K} c_k / \sum_k c_k) > \epsilon$, where ϵ is often chosen as a value close to 100%, such as 90%.

2.2.2 Application of the Active Shape Model

Given a target image to be segmented, the mean shape is first initialized in the image space. The initialization provides the position, rotation and scaling of the target object in the target image. Although automatic initialization methods exist, typically based on landmarks or registration, most of them are not universal to different applications. As a result, manual initialization is still heavily used in practice. After the mean shape is initialized, it is iteratively deformed toward the boundary of target object until convergence.

Each iteration involves two steps: 1) vertex-wise local deformation and 2) global refinement by the statistical shape space.

Vertex-wise Local Deformation

As shown in fig. 2.2, each vertex of the shape locally searches along its normal direction and finds a position most likely to be the object boundary. Then, the vertex is deformed to the boundary position.

There are many ways to characterize the object boundary. A simple way is to use the image gradient magnitude based on the assumption that voxels with large gradient magnitudes are more likely to be on the object boundary than those with small gradient magnitudes. While this strategy works well for objects with clear boundaries, it fails notably if the object boundary is indistinct, such as the boundaries of pelvic organs in CT images. Recently there is a trend that detects the object boundary by learning a classifier based on local image features. Since the patterns of object boundary are learned from multiple local image features, it is often more effective than using simple gradient magnitudes.

Global Refinement by the Shape Space

As each vertex deforms independently in the previous step, the deformed shape is likely to be unsmooth and implausible in terms of the global shape. To overcome this problem, ASM relies on the learned shape space to refine the shape after independent vertex-wise deformation. Given a deformed shape, it is first linearly aligned to the mean shape by a similarity transform. Then, the aligned shape is refined by finding the closest shape in the shape space. Finally, the closest shape is transformed back to the image space as the refined shape. With this global refinement the deformation is always constrained in the learned

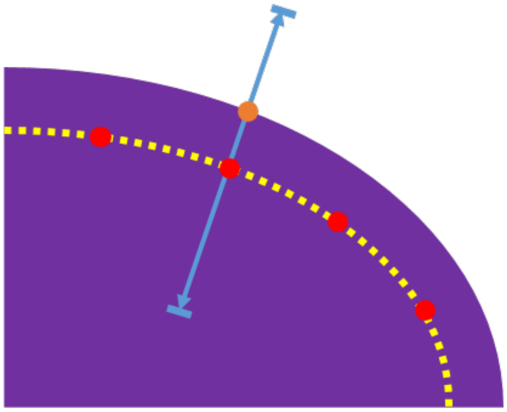


Figure 2.2: Illustration of vertex deformation in the active shape model. The purple area is the target object; the yellow dashed line indicates the current location of shape model; the red points are the boundary points (vertices) on the shape model; the blue line shows the normal direction of one vertex; the orange point shows the position along the normal direction with the maximal boundary response.

shape space, and the final segmentation is plausible and looks like those observed in the training set. The following mathematical equations explain how to find the closest shape $\hat{\mathbf{u}}_t$ in the shape space for a given shape $\mathbf{u}_t$.

$$\alpha_k = (\mathbf{u}_t - \bar{\mathbf{u}})^T \mathbf{e}_k, \quad \gamma = \max(1, \frac{1}{T} \sum_k \frac{\alpha_k^2}{c_k}), \tag{2.7}$$

$$\hat{\mathbf{u}}_t = \bar{\mathbf{u}} + \frac{1}{\gamma} \sum_k \alpha_k \mathbf{e}_k, \tag{2.8}$$

where α_k is the k-th coefficient of shape $\mathbf{u}_t$ mapped into the shape space, $\bar{\mathbf{u}}$ is the mean shape, c_k and $\mathbf{e}_k$ are the k-th eigen-value and eigen-vector of the shape space, T is a predefined parameter that determines the size of the shape space in terms of Mahalanobis distance, and γ is a scaling factor that rescales $\mathbf{u}_t$ into the shape space if it is outside.

2.2.3 Limitations

With a shape space the deformed shape is always constrained in a plausible shape set learned from training data. This advantage makes ASM well suited for medical image segmentation, where image appearance may be unreliable due to noise and artifacts. However, ASM has several limitations that should be addressed before it can be a powerful tool for CT pelvic organ segmentation.

Sensitivity to initialization

Because each vertex is deformed locally, the performance of ASM relies on a good initialization. If the shape model is not initialized close to the object boundary, local search won't be able to find the boundary. However, it can be tricky to automatically and robustly initialize the shape model. In the CT pelvic organ segmentation, the difficulties come from two aspects: 1) it is challenging to accurately detect the position, orientation and size of pelvic organs in CT images due to low contrast and large inter-subject anatomical variation; 2) the mean shape can be dramatically different from individual shapes to segment. This difference renders the mean shape initialization ineffective for organs with large shape variations, such as the rectum.

Inflexibility to Segment Tubular Organs

There are two issues when ASM is applied to segment tubular organs, such as the rectum.

- **Search Range.** To specify how far a vertex can search along the normal direction for the object boundary, a local search range needs to be specified in ASM. While it is often easy to specify it for ellipsoid-like organs, it is challenging for tubular and thin organs, such as the rectum. A small local search range is insufficient to find the boundary while a large local search range can cause mesh folding or shrinkage as the vertices on the left tube wall may find the boundary location on the right tube wall. Ideally the local search range should be spatially adaptive. If a vertex is close to the object boundary, its search range should be small. If a vertex is far away from the object boundary, its search range should be large.

- **Shape Space.** There are three challenges when using a shape space for tubular organ segmentation: 1) the variation of tubular shapes are mostly nonlinear, e.g., twisting and bending, while the PCA shape space captures only linear variations [Cootes et al., 1995]; 2) due to large variations of tubular organs, the number of training data is often limited to sufficiently describe these variations; 3) the shape distribution of tubular organs doesn't necessarily follow the Gaussian assumption of the PCA shape space.

To overcome these limitations, it is necessary to 1) change the deformation mechanism from local to non-local, thus making deformable models insensitive to initialization; 2) adapt the deformation of each vertex based on its distance to the object boundary, thus addressing the issue of search range; 3) increase the robustness of shape deformation and reduce its dependency on the shape space, as the PCA shape space may not be suitable to describe the shape statistics of tubular organs.

2.3 Segmentation Evaluation

The manual segmentation is often used as a gold standard to assess the quality of automatic segmentation. This section introduces four quantitative metrics used in the dissertation for evaluating the proposed segmentation methods and comparing them with other existing methods.

- **Dice Similarity Coefficient (DSC).** DSC measures the overlap ratio between an automatic segmentation and a manual segmentation. It ranges from 0% to 100%. 0% indicates the worst segmentation and 100% indicates the best segmentation. Mathematically DSC is defined as the equation below.

$$\mathrm{DSC} = \frac{\|\mathrm{Vol}_{\mathrm{gt}} \cap \mathrm{Vol}_{\mathrm{auto}}\|}{(\|\mathrm{Vol}_{\mathrm{gt}}\| + \|\mathrm{Vol}_{\mathrm{auto}}\|)/2}, \tag{2.9}$$

  where $\mathrm{Vol}_{\mathrm{gt}}$ and $\mathrm{Vol}_{\mathrm{auto}}$ are the voxel sets of manually labeled and automatically segmented objects, respectively. The values of DSC vary strongly with both object size and shape. It is more difficult for automatic segmentation methods to achieve high values of DSC on objects with small sizes and elongated shapes than those with large sizes and sphere-like shapes.

- *Average Surface Distance (ASD).* ASD measures the average distance between the surfaces of an automatic segmentation and a manual segmentation. Mathematically it is defined as the equation below.

$$\mathrm{ASD} = \frac{1}{2}\left(\underset{a \in \mathrm{Vol}_{\mathrm{gt}}}{\mathrm{mean}} \; \min_{b \in \mathrm{Vol}_{\mathrm{auto}}} d(a,b) + \underset{a \in \mathrm{Vol}_{\mathrm{auto}}}{\mathrm{mean}} \; \min_{b \in \mathrm{Vol}_{\mathrm{gt}}} d(a,b)\right), \tag{2.10}$$

  where $d(a, b)$ is the Euclidean distance between voxels a and b measured in millimeters. Compared to the values of DSC, the values of ASD are more sensitive to strong local variations. For example, an undesired thin spike doesn't necessarily take up much volume and thus may not affect the values of DSC much. However, it causes a significant change on the boundary distance and thus would greatly increase the values of ASD.

Among various surface distance metrics, ASD is only one of them. Besides ASD, another commonly used surface distance metric is Hausdorff distance. Hausdorff distance has a similar definition with ASD except that Hausdorff distance computes the 90 percentile of surface distances instead of taking the mean as done in eq. 2.10. Therefore, Hausdorff distance is more sensitive to local variations than ASD. However, as ASD is more frequently used in the literature of CT pelvic organ segmentation than Hausdorff distance, this dissertation reports only the values of ASD for the purpose of comparison.

- *Sensitivity (SEN) and Positive Predictive Value (PPV).* SEN measures the percentage of a manual segmentation that overlaps with an automatic segmentation, and PPV measures the percentage of an automatic segmentation that overlaps with a manual segmentation. These two metrics are informative to over-segmentation and under-segmentation. In the case of over-segmentation, SEN is high and PPV is low. In the case of under-segmentation, SEN is low and PPV is high. Mathematically they are defined as the equations below.

$$\mathrm{SEN} = \frac{\|\mathrm{Vol}_{\mathrm{gt}} \cap \mathrm{Vol}_{\mathrm{auto}}\|}{\|\mathrm{Vol}_{\mathrm{gt}}\|}, \tag{2.11}$$

$$\mathrm{PPV} = \frac{\|\mathrm{Vol}_{\mathrm{gt}} \cap \mathrm{Vol}_{\mathrm{auto}}\|}{\|\mathrm{Vol}_{\mathrm{auto}}\|}. \tag{2.12}$$

2.4 Summary

This chapter presented the necessary background for understanding the rest chapters. Section 2.1 introduced random forests as a general method for supervised learning. It was shown that random forests can be naturally used for multiclass classification and multivariate regression. As an efficient, scalable and non-linear learning model, random forests fit well in the field of image analysis. Section 2.2 introduced a popular deformable model called the active shape model (ASM). In ASM a statistical shape space is learned from a training set and used to constrain the shape deformation. The shape space improves the robustness of deformable segmentation in the presence of image noise and artifacts. Besides, the limitations of ASM were also discussed in the application of CT pelvic organ segmentation. Finally, section 2.3 presented several quantitative metrics for evaluating an automatic segmentation algorithm.

CHAPTER 3 : LEARNING DEFORMATIONS FOR PLANNING-CT SEGMENTATION

As mentioned in section 1.3.2, conventional deformable models are sensitive to initialization and ineffective for segmenting tubular organs. These limitations make them not well suited for CT pelvic organ segmentation, where robust initialization of deformable models is difficult and organs may have tubular shapes, e.g., the rectum. To overcome these limitations, this chapter investigates a novel deformable model named *"regression-based deformable model"* (RDM) [1] to segment male pelvic organs from CT images; these organs include the prostate, bladder, rectum and two femoral heads. In RDM, a deformation field toward an organ boundary is predicted from an intensity image by a regression model. It is used to explicitly guide a deformable model for segmentation. Compared to conventional deformable models, the estimated deformation field in RDM provides non-local deformations. Guided by these deformations, RDMs are insensitive to initialization. Moreover, as deformations become spatially adaptive, RDMs are more flexible than conventional deformable models to segment tubular organs. These properties render RDMs appealing for CT pelvic organ segmentation.

To accurately and robustly estimate the deformation field for a RDM, this chapter investigates two novel machine learning techniques as briefly summarized below and detailed in sections 3.1 and 3.2, respectively.

[1] This work was published in IEEE Transactions on medical imaging [Gao et al., 2016]. This chapter uses parts of text descriptions and figures from the published paper.

- **Auto-context Model.** In conventional voxel-wise prediction, the deformation at one voxel is independently predicted without considering those of its neighborhood. As deformations in the spatial neighborhood are highly correlated, the independent estimation often results in a noisy and spatially inconsistent deformation field. To improve the prediction accuracy, *an auto-context model* is adopted to iteratively refine the deformation field by considering not only local image appearance but also predicted deformations at neighboring voxels. A synthetic experiment shows that the auto-context model captures the structured information in the spatial neighborhood. This information is useful to suppress prediction noise and improve the spatial consistency of deformation field.

- **Multitask Random Forest.** The auto-context model improves the deformation field by exploiting the information of estimated deformations from the spatial neighborhood. However, if a majority of deformations are mis-predicted in the spatial neighborhood, the auto-context refinement would lead to wrong predictions. To relieve this problem, *a multitask random forest* is proposed to jointly learn deformation regression and organ classification in a single random forest. It has two advantages compared to the standard random forest. 1) Through joint learning the multitask random forest is forced to exploit the commonality between related tasks; doing so is helpful to reduce the risk of overfitting. 2) By integrating the multitask random forest with the auto-context model, the information output from these two tasks can be exchanged during the iterative refinement procedure. As the information from the two tasks is complementary, the mis-predictions in the estimated deformation field can be potentially corrected by exploiting the information from organ classification and vice versa. Therefore, the multitask random forest improves the robustness of deformation field estimation.

With the above techniques, a deformation field can be predicted for a target image and used to guide deformable segmentation. However, it is risky to deform the shape model directly using the estimated deformation field because of potential mis-predictions. To further improve the robustness of deformation, two strategies are proposed in sections 3.3 and 3.4, respectively. Section 3.3 proposes a hierarchical deformation strategy where the shape deformation is highly constrained in the beginning and gradually relaxed as the shape model approaches the object boundary. Section 3.4 investigates a multi-resolution segmentation framework where multiple organs are jointly segmented in the coarse resolutions, and their segmentations are separately refined in the fine resolutions.

Fig. 3.1 shows the flowchart of the proposed method for planning-CT segmentation, which consists of three major components: for deformation field estimation, 1) the auto-context model; 2) the multitask random forest; and 3) for organ segmentation, regression-based hierarchical deformation. Each step will be detailed in the following sections.

3.1 Deformation Regression and Auto-context

This section covers the details of the auto-context model for deformation field estimation. Specifically, it first defines the task of deformation regression and describes a conventional method for deformation field estimation. Then, the limitation of the conventional method is discussed, and the auto-context model is introduced as an iterative solution to overcome this limitation. Finally, a synthetic experiment is presented to explain the reason behind the strong performance of the auto-context model.

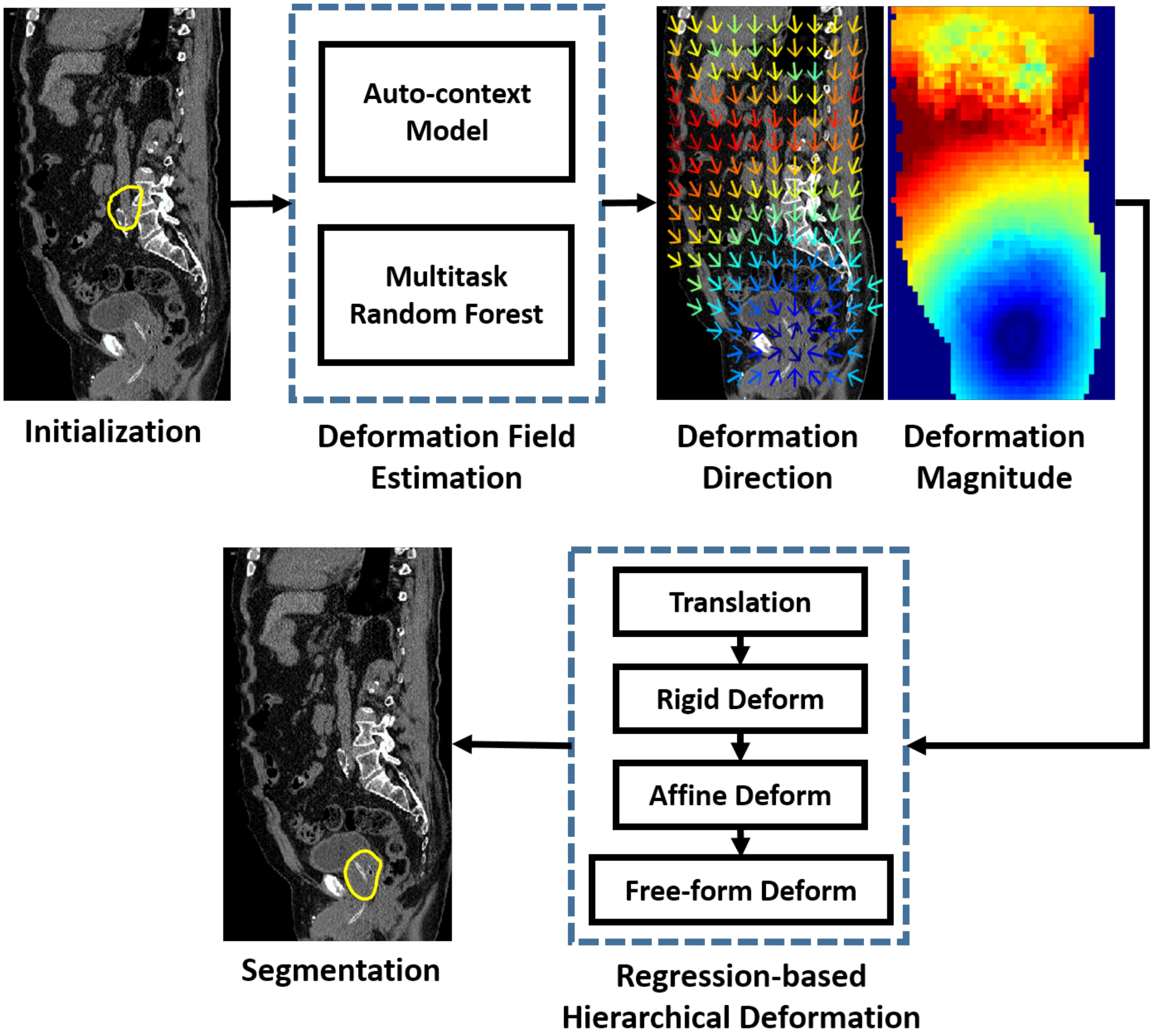


Figure 3.1: The flowchart of regression-based deformable model. The yellow contours in the first and second rows indicate the initialized deformable model and the final segmentation, respectively. In the images of deformation direction and deformation magnitude, color indicates the magnitude of estimated deformation. The colder the color is, the smaller the magnitude is.

3.1.1 Deformation and Deformation Regression

In RDM, as illustrated in fig. 3.2, *the deformation* at one voxel is defined as the 3D displacement vector from this voxel to the nearest voxel on the organ boundary. Given a testing image with an unknown organ boundary, the deformation at any image location needs to be predicted based on local image appearance. *Deformation regression* aims to learn a mapping from local appearance features to the deformation based on a set of training images, where manual contours of the target organ are available. At runtime, the learned mapping is applied to predicting the deformations in the testing image, where no manual contour is available.

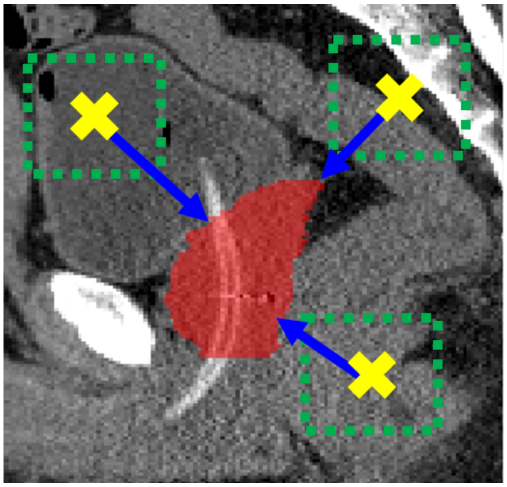


Figure 3.2: Illustration of deformations at several voxel positions. The blue arrows denote the deformations at several voxel positions (yellow crosses) toward the organ boundary (red). The green dashed boxes indicate the local image patches centered at these voxels, where appearance features are extracted.

To learn such a mapping, random forests are used in this work due to their many advantages, such as efficiency and non-linearity as discussed in section 2.1.4. In the training stage, given training images with contoured organ boundaries, a set of voxels are randomly sampled from training images near the organ boundary. Each voxel is represented by local appearance features (i.e., Haar-like features) and associated with the ground truth deformation calculated from the manual contour. These voxels serve as training samples to a random forest, and the random forest is able to learn a regression model that predicts the deformation at any voxel based on local appearance features. The learned random forest is named as "*regression forest*", since it is specifically trained for deformation regression.

Given a learned regression forest, the deformation field of a testing image can be estimated by independently predicting the deformation at each voxel location. However, such an approach ignores the fact that deformations at neighboring voxels are highly correlated. As a result, the estimated deformation field is often noisy and spatially inconsistent, as shown in the first row of fig. 3.3.

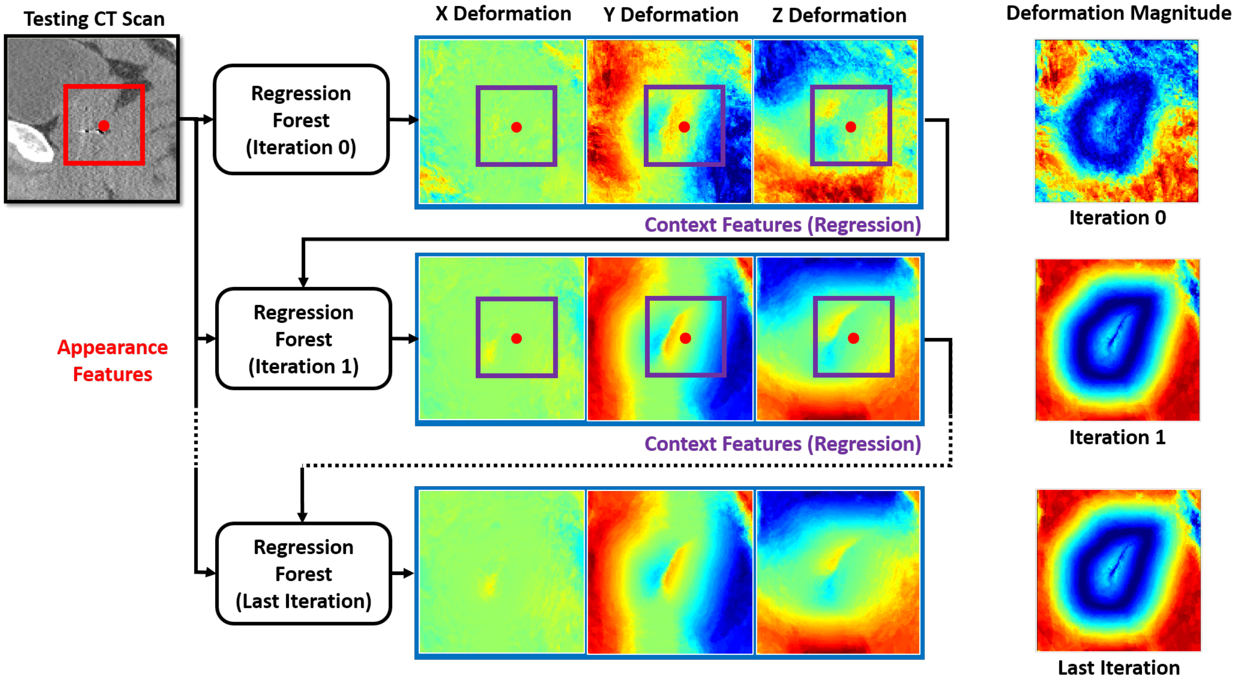

Testing CT Scan
X Deformation
Y Deformation
Z Deformation
Deformation Magnitude
Regression Forest (Iteration 0)
Iteration 0
Context Features (Regression)
Appearance Features
Regression Forest (Iteration 1)
Iteration 1
Context Features (Regression)
Regression Forest (Last Iteration)
Last Iteration

Figure 3.3: Flowchart of the auto-context model with the regression forest for deformation regression. The red point indicates a voxel. The red rectangle is a local patch of this voxel in the CT image where appearance features are extracted. Purple rectangles are local patches of this voxel in the deformation fields where context features are extracted.

3.1.2 Auto-context Model

To overcome this drawback, deformations predicted at neighboring voxels need to be considered during the voxel-wise estimation of a deformation field. In this work, the auto-context model [Tu and Bai, 2010] is used for this purpose.

The auto-context model was originally proposed in [Tu and Bai, 2010] as an iterative approach for refining the likelihood map from voxel-wise classification. The idea is to consider not only local image appearance but also neighboring classification results during voxel-wise classification. By combining these two pieces of information, the auto-context model is shown to be effective in improving classification results. It is not difficult to see that the same idea can also be borrowed to refine the deformation field from voxel-wise regression. To be specific, the following paragraphs describe the training and testing of the auto-context model when it is applied to refining the deformation field.

- **Auto-context Training.** The training of the auto-context model typically takes several iterations, e.g., 2-3 iterations. A regressor (e.g., regression forest) is trained at each iteration. *In the first iteration*, appearance features (i.e., Haar-like features) are extracted from CT image to train the first regressor. Once the first regressor is trained, it is applied back to each training image to generate a tentative deformation field.

  In the second iteration, the features of each training voxel consist of not only appearance features from the the CT image but also Haar-like features extracted from the tentatively estimated deformation field. The latter features are called "context features" because they capture the context information, i.e., predicted deformation information in the spatial neighborhood. With the introduction of new features, a second regressor is trained. As the second regressor considers not only the CT appearance but also

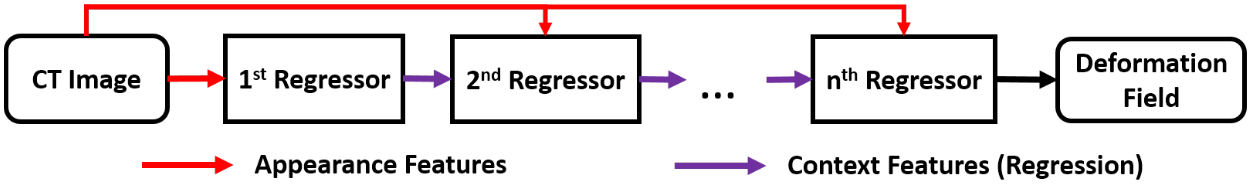


Figure 3.4: Schematic diagram of auto-context with n iterations.

estimated deformations at neighboring voxels, it often leads to a better deformation field, as shown in the second row of fig. 3.3.

Given a refined deformation field by the second regressor, the context features are updated and can be used together with the appearance features to train the third classifier. The same procedure is repeated until the final iteration is reached. Because each iteration involves voxel-wise classification of all training images, more iterations take longer training time. In practice, 2-3 iterations are often used.

- **Auto-context Application.** Given a testing CT image, the learned regressors are sequentially applied, as illustrated in fig. 3.3. *In the first iteration*, the first learned regressor is applied voxel-wise on the testing image to generate a deformation field by using only appearance features. *In the second iteration*, the second learned regressor is used to predict a new deformation field by combining appearance features from the CT image with context features from the deformation field estimated in the previous iteration. This procedure is repeated until all learned regressors have been applied. The deformation field output by the last regressor is the output of the auto-context model. Fig. 3.4 provides the schematic diagram.

By considering predicted deformation information in the spatial neighborhood, the auto-context model suppresses prediction noise and improves the spatial consistency of the deformation field compared to conventional methods that consider only local image appearance (comparing the first and third rows of fig. 3.3).

3.1.3 Understanding the Auto-context Model

The auto-context model differs from conventional voxel-wise prediction methods only in the existence of context features. In this section, it is shown that context features capture neighborhood structured information from training images. This structured information can be enforced by the auto-context model in the prediction map of a testing image.

To justify this statement, a synthetic experiment was designed under voxel-wise classification. In this experiment, given a binary training image that represents a shape, a sequence of random forest classifiers was trained in the same manner as the auto-context model. Then, the learned classifiers were sequentially applied to a testing image with a different shape. The hypothesis is that if the classifiers learn the neighborhood structured information of the training shape the testing shape would generally evolve to be the training shape under the iterative classification. Fig. 3.5 gives the results for three cases, where the training shapes are the sphere, prostate and bladder, respectively. It can be seen that the testing shapes refined by the auto-context model eventually become almost identical to the respective training shapes. This observation indicates that the structured information learned from training images can be enforced in the testing image by the auto-context model. A large number of iterations was used in this experiment to facilitate large shape refinements for the purpose of demonstration. In practice, the refinement won't be this great since only 2-3 iterations are often used.

This experiment shows that by extracting context features from classification maps, the auto-context model learns structured label information in the spatial neighborhood. Sim-

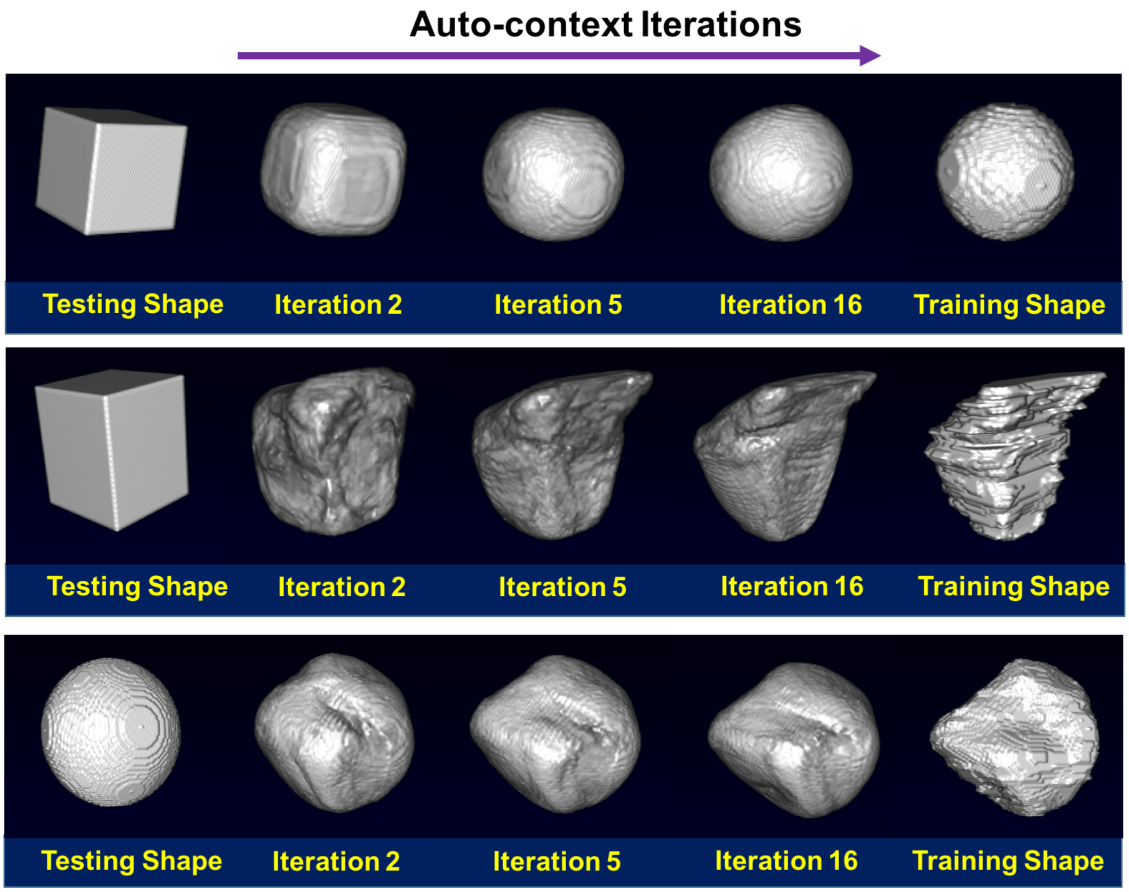


Figure 3.5: Shape refinements by the auto-context model (AC). Upper: sphere. Middle: prostate. Lower: bladder.

ilarly, if the context features were extracted from estimated deformation fields, the auto-context model would learn structured deformation information. The learned structured information is the key ingredient that makes the auto-context model outperform conventional voxel-wise prediction models.

3.2 Multitask Random Forest

The auto-context model improves the deformation field estimation by integrating estimated deformation information from the spatial neighborhood. It works well if mispredictions in the spatial neighborhood are minor. However, if a majority of the deformations are mis-predicted in the spatial neighborhood, the auto-context model would lead to wrong refinements. To relieve this problem, a *multitask random forest*[2] is proposed to replace the standard regression forest used in the auto-context model. In the multitask random forest, deformation regression is jointly learned with another task, i.e., organ classification, in a single random forest. *Organ classification* refers to the classification task that uses local appearance features to distinguish voxels inside the organ from those outside.

Compared to the standard regression forest, the multitask random forest has two advantages: 1) through joint learning the multitask random forest is forced to exploit the commonality between deformation regression and organ classification. The exploited commonality is helpful to reduce the risk of overfitting; 2) by integrating the multitask random forest with the auto-context model, organ classification provides additional context features, i.e., estimated class information in the spatial neighborhood. They are useful to improve the estimated deformation field near the organ boundary when deformations are not well predicted from local appearance features.

3.2.1 Mathematical Definition

To adapt a random forest for multitask learning, the purity of a training set $\mathcal{S}$ is modified as follows in order to consider multiple tasks in the learning process.

$$\mathcal{P}(\mathcal{S}) = \sum_i w_i \frac{\mathcal{P}^i(\mathcal{S})}{\mathcal{Z}^i}, \tag{3.1}$$

[2] The name of multitask random forest comes from multitask learning, where multiple tasks are jointly learned using a shared representation/model. [Caruana, 1997]

where w_i is the weight for the i-th task, $\mathcal{P}^i(\mathcal{S})$ is the purity definition for the i-th task, and $\{\mathcal{Z}^i\}$ are coefficients that normalize the purity across different tasks. For each task, $\mathcal{Z}^i$ is defined as the task-specific purity of the entire training set, i.e., the purity at the root node. In this work since only two tasks, i.e., deformation regression and organ classification, are considered, the purity definition can be further specialized as follows:

$$\mathcal{P}(\mathcal{S}) = w\frac{\mathcal{V}^{\text{DR}}(\mathcal{S})}{\mathcal{Z}^{\text{DR}}} + (1-w)\frac{\mathcal{E}^{\text{OC}}(\mathcal{S})}{\mathcal{Z}^{\text{OC}}}, \tag{3.2}$$

where $w \in [0,1]$ is the weight coefficient, $\mathcal{V}^{\text{DR}}$ is the purity definition for deformation regression that measures the variation of deformations in a training set $\mathcal{S}$, and $\mathcal{E}^{\text{DR}}$ is the purity definition for organ classification that measures the consistency of class labels in a training set $\mathcal{S}$. Their mathematical definitions are given below:

$$\mathcal{V}^{\text{DR}}(\mathcal{S}) = -\frac{1}{|\mathcal{S}|}\text{tr}\left(\sum_{\mathbf{x}\in\mathcal{S}}(\mathbf{d}_{\mathbf{x}} - \bar{\mathbf{d}})(\mathbf{d}_{\mathbf{x}} - \bar{\mathbf{d}})^T\right), \tag{3.3}$$

$$\mathcal{E}^{\text{OC}}(\mathcal{S}) = p_+\text{log}p_+ + p_-\text{log}p_-, \tag{3.4}$$

where tr is the trace operator, $\mathbf{d}_{\mathbf{x}}$ is the deformation at a training voxel $\mathbf{x}$, $\bar{\mathbf{d}}$ is the mean deformation in the training set $\mathcal{S}$, and p_+ and p_- are the percentages of positive and negative training voxels in the training set $\mathcal{S}$, respectively. Positive voxels are voxels inside the organ while negative voxels are those outside the organ. In the multitask random forest, each leaf stores not only the average deformation of training samples (voxels) that arrive at it but also the label distribution of those training samples. Therefore, the multitask random forest is able to simultaneously predict the deformation and class label of a testing voxel. Through voxel-wise prediction, the multitask random forest is able to produce both deformation field and organ likelihood map.

3.2.2 A Better Model for Deformation Regression

As deformation regression and organ classification are jointly learned in a single random forest, the multitask random forest is optimized to select common features and thresholds that are informative to both tasks. As pointed in [Caruana, 1997], sharing the same model among related tasks could improve the generalization. Moreover, it is found in my work that joint learning of deformation regression and organ classification clarifies *the ambiguity* that exists in the training of the regression forest.

Fig. 3.6 illustrates the ambiguity where two voxels (yellow crosses), i.e., one inside and one outside the organ, have the same deformation toward the organ boundary but have different image appearances. In the regression forest, which is trained only for deformation regression, the training of random forest tries to find image features that group these two voxels in the same leaf node, since they have the same deformation. However, due to dramatic appearance difference, generally it is infeasible to find such features. In the end, the random forest may find meaningless features that happen to well fit the training set but cannot generalize well in the testing. Thus, a risk of overfitting is imposed.

On the other hand, the multitask random forest considers not only the deformation but also the class label during splitting (eq. 3.2). It can well separate these two voxels into different leaf nodes, since the two voxels have different class labels, i.e., one is outside the organ while the other is inside. Therefore, the ambiguity that exists in the regression forest can be well resolved in the multitask random forest by exploiting the class label information from organ classification. The risk of overfitting is reduced and the generalization is improved.

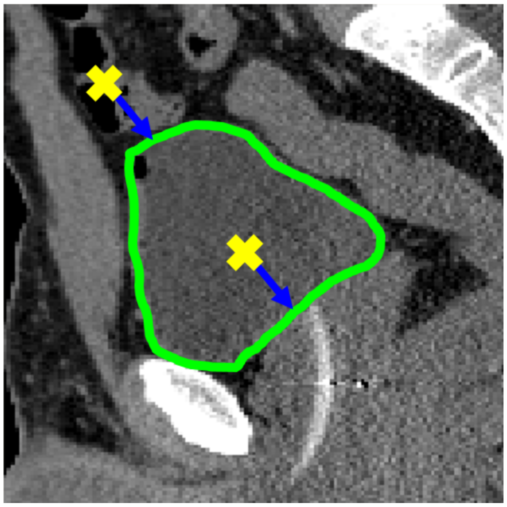


Figure 3.6: Ambiguity in the training of random forest when deformation is used as the only supervised guidance for splitting. The green contour indicates the bladder boundary. Yellow crosses indicate two voxels with the same deformation toward the organ boundary but with different image appearances.

3.2.3 Integration with the Auto-context Model

Different from the regression forest, which outputs only the deformation field, the multitask random forest produces both a deformation field and an organ likelihood map in a single pass. Therefore, context features can be extracted from both prediction maps and used in the auto-context model to refine the deformation field. Fig. 3.7 shows the flowchart of the auto-context model with the multitask random forest. Compared to the auto-context model with the regression forest (fig. 3.3), additional context features are extracted from the organ likelihood map (blue rectangle). These features capture the estimated class information in the spatial neighborhood. They provide little information to the refinement of the deformation field far away from the organ boundary, since the organ likelihood map is homogeneous in those faraway regions (fig. 3.7), e.g., it is either pure black if outside the organ or pure white if inside the organ. However, near the organ boundary the estimated class information provides cues about the boundary location. It can be used to estimate deformations and thus is useful to deformation field refinement, especially when deformations are not well predicted from local appearance features at the first iteration of the auto-context model.

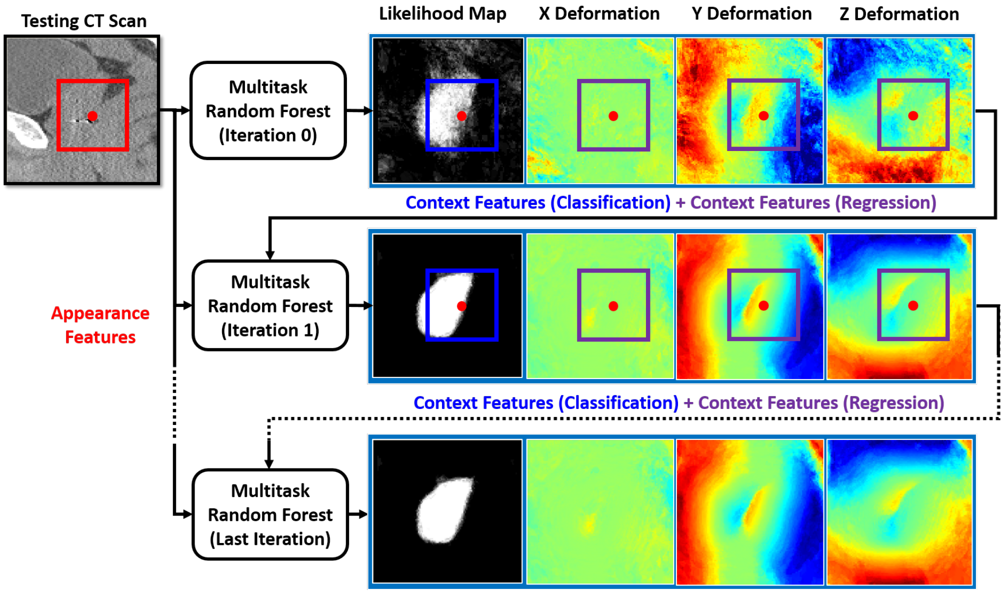

Testing CT Scan
Likelihood Map
X Deformation
Y Deformation
Z Deformation
Multitask Random Forest (Iteration 0)
Context Features (Classification) + Context Features (Regression)
Appearance Features
Multitask Random Forest (Iteration 1)
Context Features (Classification) + Context Features (Regression)
Multitask Random Forest (Last Iteration)

Figure 3.7: Flowchart of the auto-context model with the multitask random forest. The red point indicates a voxel. The red, blue and purple rectangles are the local patches of this voxel on the CT image, estimated organ likelihood maps and deformation fields, respectively.

Fig. 3.8 gives a typical example to illustrate the importance of estimated class information in deformation field refinement. As shown in fig. 3.8(a), if many deformations are mispredicted in a small region at the first iteration of the auto-context model (red rectangle), the auto-context model is not able to correct them by using context features only from estimated deformation field, since deformations predicted in the neighborhood are also inaccurate. As a result, the deformation field is often generated with missing parts of organ boundaries (fig. 3.8(a)). This problem is called "*the missing boundary problem*", which is common if the auto-context model is used with the regression forest.

In contrast, the multitask random forest produces an additional organ likelihood map that provides complementary information to deformation regression. As shown in fig. 3.8(b),

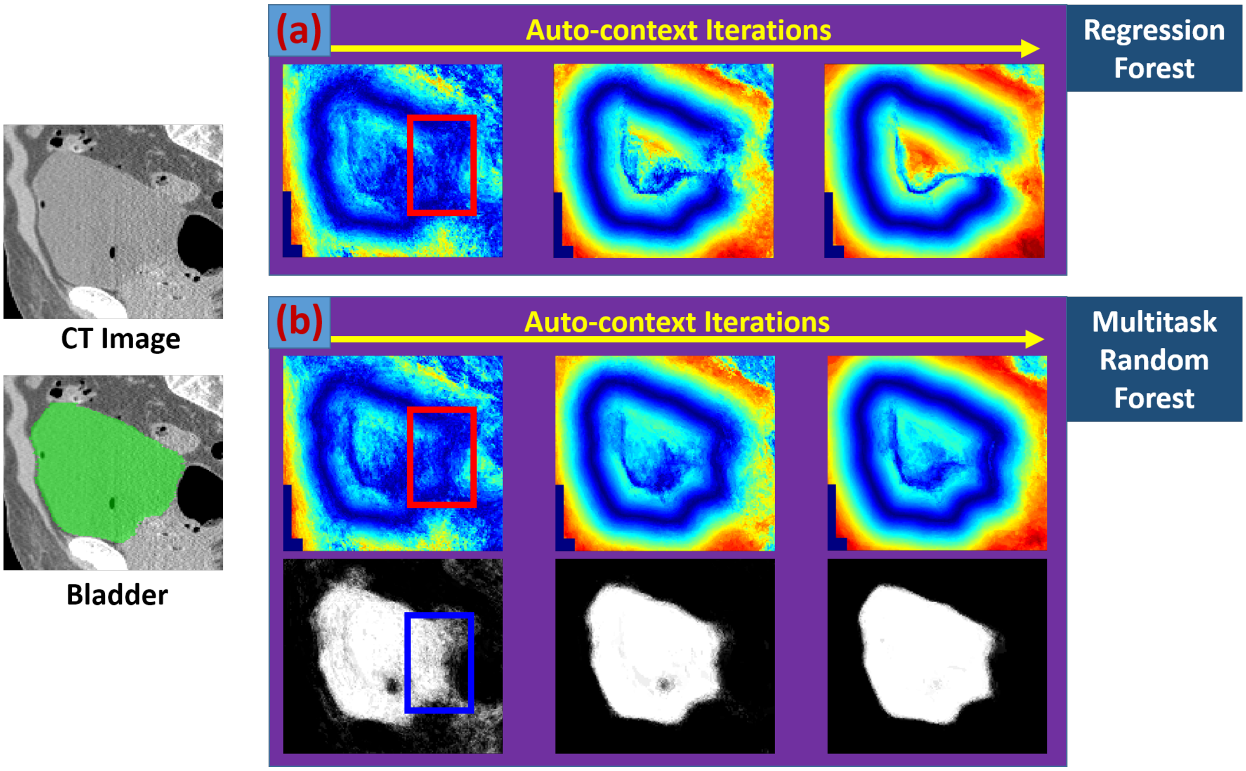


Figure 3.8: The missing boundary problem. Red rectangles indicate the image region where deformations are not well predicted. The blue rectangle shows the same region as the red one on the organ likelihood map.

when deformations are not well predicted near the organ boundary (red rectangle), the estimated class information from the organ likelihood map provides additional cues about the location of the organ boundary (blue rectangle). This information is helpful to estimate deformations. By adding it into the auto-context model, the robustness of deformation field estimation is improved, since now the auto-context refinement considers not only estimated deformations in the neighborhood but also the boundary location provided by organ classification. Under the collaboration of the two types of context features, i.e., one from the deformation field and one from the organ likelihood map, the multitask random forest is able to address the missing boundary problem suffered by the regression forest, as illustrated in the rightmost image of fig. 3.8(b).

3.3 Regression-based Deformable Models

A deformation field estimated by the auto-context model with the multitask random forest can be used to guide a deformable model for organ segmentation. However, due to potential mis-predictions, it is risky to freely deform the shape model using the estimated deformation field. To improve the robustness of segmentation, a hierarchical deformation strategy is proposed, as depicted in alg. 3.1. To start a segmentation, the mean shape model that is calculated as the average of all training shapes is first initialized on the center of a testing image (fig. 3.1). During deformable segmentation, initially the shape model is only allowed to translate under the guidance from the estimated deformation field. Once it is well positioned, it is allowed to rigidly rotate so as to estimate the orientation of the shape model. Afterwards, the deformation is further relaxed to the affine transformation so as to estimate the scaling and shearing parameters of the shape model. Finally, the shape model is freely deformed under the guidance from the deformation field. The jointly estimated likelihood map from the final iteration of the auto-context model is not used in the segmentation stage. In this work organ classification is used only to improve the generalization of deformation regression.

Compared to conventional deformable models, regression-based deformable models (RDMs) have two major advantages.

- **Robustness to Initialization.** Different from conventional deformable models, the shape model in the RDM is no longer deformed locally around the initialization. The estimated deformation field provides a non-local deformation to each vertex of the shape model. As a result, even if shape models are initialized far away from the target organ, they can still be rapidly deformed to the correct position under the guidance

Algorithm 3.1 Regression-based Hierarchical Deformation

Input: I - testing CT scan, $\mathcal{D}$ - learned multitask random forest,
$\mathcal{M}_{\text{init}}$ - initialized shape model

Output: $\mathcal{M}$ - the final segmentation

Notation: `World2Voxel`$(I, \mathbf{p})$ outputs voxel coordinate of vertex $\mathbf{p}$ on the image I, $\mathcal{D}(I, \mathbf{x})$ returns the 3D deformation at voxel $\mathbf{x}$ on the image I, and $\Theta(K)$ denotes valid transform matrix set of transform type K.

```
function DEFORM($I$, $\mathcal{D}$, $\mathcal{M}$, $\mathcal{K}$)                                    ▷ Subroutine
    for $Iteration \leftarrow 1$ to $MaxIteration$ do
        $\mathcal{M}_{\text{deform}} = \mathcal{M}$
        for all vertex $\mathbf{p} \in \mathcal{M}_{\text{deform}}$ do
            $\mathbf{x} = \texttt{World2Voxel}(I, \mathbf{p})$
            $\mathbf{p} \leftarrow \mathbf{p} + \mathcal{D}(I, \mathbf{x})$
        end for
        if $K \in \{\text{Translation}, \text{Rigid}, \text{Affine}\}$ then
            Estimate transform matrix $T \in \mathbb{R}^{4\times4}$:
                $\arg\min_T \|T(\mathcal{M}) - \mathcal{M}_{\text{deform}}\|^2,$ s.t., $T \in \Theta(K)$
            $\mathcal{M} = T(\mathcal{M})$
        else
            $\mathcal{M} = \text{SmoothSurface}(\mathcal{M}_{\text{deform}})$
            $\mathcal{M} = \text{RemeshSurface}(\mathcal{M})$
        end if
    end for
    return $\mathcal{M}$
end function
function HIERARCHICALDEFORM($I$,$\mathcal{D}$,$\mathcal{M}_{\text{init}}$)                           ▷ Main routine
    $\mathcal{M}$ = DEFORM($I$, $\mathcal{D}$, $\mathcal{M}_{\text{init}}$, “Translation”)
    $\mathcal{M}$ = DEFORM($I$, $\mathcal{D}$, $\mathcal{M}$,     “Rigid”)
    $\mathcal{M}$ = DEFORM($I$, $\mathcal{D}$, $\mathcal{M}$,     “Affine”)
    $\mathcal{M}$ = DEFORM($I$, $\mathcal{D}$, $\mathcal{M}$,     “FreeForm”)
    return $\mathcal{M}$
end function
```

from the estimated deformation field. In the case of CT pelvic organ segmentation, it is sufficient for RDMs to work if the mean shape model is initialized in the image center. However, it is almost impossible for conventional deformable models to work under the same initialization. RDMs are much more robust to initialization than conventional deformable models.

- **Adaptive Deformation Parameters.** Explicitly predicting the deformation of each vertex eliminates the necessity of specifying and tuning many deformation parameters. For example, *search range* is an important parameter in the active shape model (ASM). It specifies how far a vertex of the shape model locally searches the organ boundary. In the segmentation of tubular organs, it can be tricky to set up this parameter. A small search range may not be large enough to deform the shape model onto the organ boundary while a large search range may cause mesh folding or shrinkage. By explicitly predicting the deformation of each vertex, RDMs eliminate the necessity to specify this parameter.

  Besides, *the deformation direction and step size* of each vertex are now adaptively determined in the RDM for optimally driving the shape model onto the target organ. This adaptivity distinguishes RDMs from conventional deformable models (e.g., ASM) that use a fixed deformation direction (e.g., normal direction) and step size, and it also increases the flexibility of deformable models to segment organs with complex shapes.

  In addition, the flexibility of deformable models is also increased by *the spatially varying deformations* estimated in the deformation field. The spatially varying deformations allow each vertex of the shape model to have dramatically different deformations. For vertices close to the organ boundary, the estimated deformations tend to be small, thus encouraging detailed boundary refinement. For vertices far away from the organ boundary, the estimated deformations tend to be large for rapidly driving these vertices close to the boundary.

In summary, RDMs provide non-local deformations that make them insensitive to initialization. Besides, many important yet ad-hoc designs, such as search range, deformation direction and step size, are either eliminated or automatically determined for each vertex according to the learned multitask random forest. These properties make RDMs appealing for CT pelvic organ segmentation where 1) reliable initializations are difficult to get and 2) organs may have complex shapes (e.g., the rectum).

3.4 Multi-resolution Segmentation

To further improve the efficiency and robustness of the proposed method, the segmentation is conducted in multi-resolution. I found that experimentally four resolutions are the best choice.

Training. In each of the two coarsest resolutions, one multitask random forest is trained jointly for all five organs. Specifically, instead of predicting a deformation to a single organ, the joint multitask random forest predicts a concatenated deformation that consists of deformations to all the five organs; instead of predicting the likelihood being inside a single organ, the joint multitask random forest predicts the likelihoods being inside different organs. Joint estimation of deformations to multiple organs is beneficial to take into account that the spatial relationship among them and is thus helpful to improve the robustness of deformation field estimation in the two coarsest resolutions.

In the two finest resolutions, one multitask random forest is trained separately for each organ. Compared to the joint random forests in the coarsest resolutions, learning an individual random forest captures specific appearance characteristics of each organ. These organ-specific random forests are more effective for detailed boundary refinement once the shape models are close to their respective organ boundaries after being driven by the coarse-level joint random forests.

Testing. The testing image is first downsampled to the coarsest resolution (voxel size $8 \times 8 \times 8$ mm^3) where rough segmentations of the five organs are rapidly obtained. These segmentations serve as good initializations for the next finer resolution. The segmentation is then sequentially performed across different resolutions until it reaches the finest resolution (voxel size $1 \times 1 \times 1$ mm^3) where the final segmentation is obtained.

The benefits of the multi-resolution strategy are straightforward. Instead of predicting the deformation field for the whole image in the fine resolution, now only a sub-region of the deformation field around the initialization (given by the previous coarser resolution) has to be estimated, thus significantly improving the efficiency. Apart from the efficiency, the robustness also benefits from the joint estimation of multiple deformation fields in the two coarsest resolutions, as the spatial relationship among different organs is implicitly considered during the joint deformation regression.

3.5 Experimental Results

The experimental data consists of 313 CT scans from 313 prostate cancer patients, where 35 of the 313 CT scans are enhanced by the injection of a contrast agent. These scans were collected from the North Carolina Cancer Hospital. The image size of a CT scan is $512 \times 512 \times (61 \sim 508)$. The in-plane resolution ranges from 0.938 mm to 1.365 mm, and the slice thickness ranges from 1 mm to 3 mm. Five pelvic organs including the prostate, bladder, rectum and two femoral heads were manually contoured by two experienced radiation oncologists. Then interpretations were averaged as the consensus. These contours serve as ground truth in the experiments.

The following subsections are organized as follows.

- Section 3.5.1 describes the details of 3D Haar-like features that are used as both appearance features and context features in the multitask random forest.
- Section 3.5.2 provides the parameter setting, sensitivity analysis of parameters, and computational time of the proposed method.
- Section 3.5.3 evaluates the auto-context model in deformation field estimation by comparing it to independent voxel-wise prediction.
- Section 3.5.4 compares the multitask random forest with the regression forest for deformation field estimation.
- Section 3.5.5 compares the multitask random forest with the classification forest for deformation field estimation.
- Section 3.5.6 compares RDMs with conventional classification-based deformable models to show the importance of the estimated deformation field in guiding deformable segmentation.
- Section 3.5.7 compares RDMs with other existing methods for CT pelvic organ segmentation.

3.5.1 3D Haar-like Features

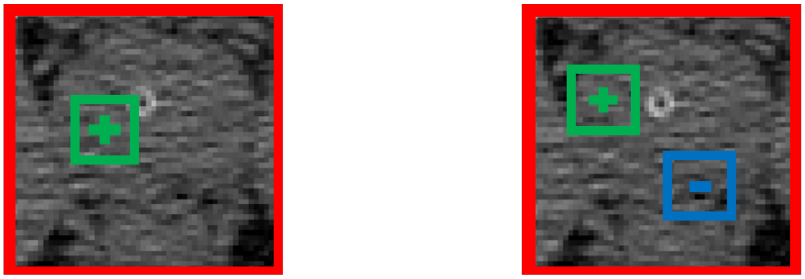
In this work, 3D Haar-like features were used as both appearance features and context features. The appearance features were 3D Haar-like features extracted from local patches in the CT image, and the context features were 3D Haar-like features extracted from local patches in the estimated deformation field or organ likelihood map. The main reason for adopting 3D Haar-like features is their high computational efficiency. 3D Haar-like features can be efficiently computed using the integral image [Viola and Jones, 2004].

As illustrated in fig. 3.9, two types of Haar-like features were considered: 1) one-block Haar-like features that compute the average intensity at one location within the local patch, and 2) two-block Haar-like features that compute the average intensity difference between two locations within the local patch. Their mathematical definitions can be formulated as follows:

$$f(I_{\mathbf{x}}|\mathbf{c}_1, s_1, \mathbf{c}_2, s_2) = \frac{1}{(2s_1+1)^3} \sum_{\|\mathbf{y}-\mathbf{c}_1\| \le s_1} I_{\mathbf{x}}(\mathbf{y}) - \frac{\lambda}{(2s_2+1)^3} \sum_{\|\mathbf{y}-\mathbf{c}_2\| \le s_2} I_{\mathbf{x}}(\mathbf{y}), \tag{3.5}$$

where $I_{\mathbf{x}}$ denotes a local patch centered at voxel $\mathbf{x}$, $f(I_{\mathbf{x}}|\mathbf{c}_1, s_1, \mathbf{c}_2, s_2)$ denotes one Haar-like feature with parameters $\{\mathbf{c}_1, s_1, \mathbf{c}_2, s_2\}$, where $\mathbf{c}_1 \in \mathbb{R}^3$ and s_1 are the center and size of the positive block, respectively, and $\mathbf{c}_2 \in \mathbb{R}^3$ and s_2 are the center and size of the negative block, respectively. $\lambda \in \{0, 1\}$ is a switch between the two types of Haar-like features. When $\lambda = 0$, eq. 3.5 uses one-block Haar-like features. When $\lambda = 1$, eq. 3.5 uses two-block Haar-like features.

In the training stage, each binary tree of a random forest was trained independently. For each tree, a bunch of Haar-like features was generated by uniformly and randomly



(a) One-block Haar-like feature

(b) Two-block Haar-like feature

Figure 3.9: Two types of Haar-like features used in the multitask random forest. Left: one-block Haar-like feature. Right: two-block Haar-like feature. The red, green and blue rectangles denote the local patch, the positive block and the negative block, respectively.

sampling parameters of the Haar-like features, i.e., $\{\mathbf{c}_1, s_1, \mathbf{c}_2, s_2, \lambda\}$, under the constraint that positive and negative blocks should stay inside the local patch. These random Haar-like features were used as the feature representation for each training sample (voxel). The reason for using different feature representations for different trees is to increase the diversity of random forests. As an ensemble model, the performance of random forests benefits from the diversity of binary decision trees. The reason for using random features is closely related to the built-in feature selection mechanism of random forests. Unlike other prediction models, e.g., the support vector machine [Vapnik, 1995, Chang and Lin, 2011], random forests select the optimal feature set during the learning process. Uninformative Haar-like features that are not related to deformation regression and organ classification will not be selected in the split nodes. So random forests are robust to the inclusion of uninformative features.

3.5.2 Parameter Setting & Computational Time

Random Forest Parameters. The number of binary decision trees was 10. The training sets of different trees were different but might overlap since they were randomly drawn from the training images. The maximum tree depth was 100. The numbers of random features and candidate thresholds in each node as set in the training stage were 1000 and 100, respectively. The minimum number of training samples in each leaf node was 8. In this application trees typically stopped at the depth of 50. These random forest parameters have been widely evaluated in many applications [Wang et al., 2015] [Gao and Shen, 2015] [Zhang et al., 2015] [Huynh et al., 2015]. In general, the performance of random forests increases with increase of tree number, tree depth and number of random features. The performance is not sensitive to the number of thresholds and the minimum number of training samples in each leaf node. However, the increase of tree number, tree depth and number of random features could significantly increase both training and testing time. As a compromise, the above parameter setting was used in the experiments.

Multi-resolution Parameters. The multi-resolution parameters are quite standard. The number of resolutions was 4. The spacings of four resolutions were 1 mm, 2 mm, 4 mm and 8 mm, respectively.

Other Parameters. Two iterations were used in the auto-context model. Section 3.5.3 evaluates the segmentation accuracy with respect to the number of auto-context iterations. It was found that two iterations are often sufficient for convergence. Block sizes s_1 and s_2 in eq. 3.5 were randomly picked from the set $\{3, 5\}$ to improve the robustness of Haar-like features to the random CT noise. The maximum number of deformation iterations was 20, and the weight w between organ classification and deformation regression in eq. 3.2 was 0.5.

Sensitivity. Fig. 3.10 (a) gives the sensitivity analysis of the weight w in eq. 3.2. It can be seen that the segmentation accuracy doesn't vary much between $w = 0.25$ and $w = 0.75$. However, when $w = 1$ the performance drops notably because the multitask random forest degrades to the regression forest. In this figure the performance with respect to $w = 0$

was not plotted because pure organ classification generally doesn't produce deformation fields for RDMs. In addition, fig. 3.10 (b) gives the sensitivity analysis of the number of deformation iterations. It can be seen that the performance converges after 20 iterations. So the maximum number of deformation iterations was chosen to be 20.

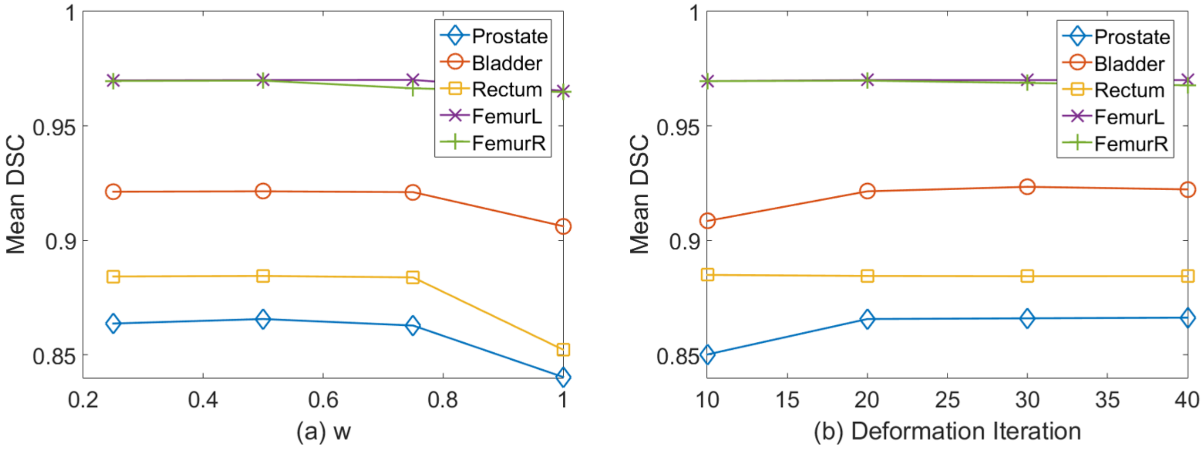


Figure 3.10: Sensitivity of the segmentation accuracy to the weight w between deformation regression and organ classification in eq. 3.2 (left) and to the number of deformation iterations (right). Higher DSC is better.

Runtime. It takes about 1.8 mins for the proposed method to segment five pelvic organs on a laptop with an Intel i7-4710HQ CPU and 16 GB memory. OpenMP was used with 4 threads for parallel computation. The training time of the proposed method is about 3-4 hours for each tree with 1884000 training samples extracted from 157 training images (i.e., 12000 training samples per training image).

3.5.3 Auto-context Model

This subsection evaluates the contribution of the auto-context model for refining the deformation fields. Table 3.1 shows the segmentation accuracies obtained without the auto-context model, with 1 iteration and 2 iterations of auto-context, respectively. It can be seen that the use of the auto-context model greatly boosts the segmentation accuracy, especially in the first iteration. With an additional iteration, the segmentation accuracy further improves. However, the improvement is not as great as the first one. Considering the computational efficiency, 2 iterations were used in this work.

Table 3.1: Segmentation accuracies (DSC) of five pelvic organs at different auto-context (AC) iterations. Bold numbers indicate the best performance.

DSC (%)	No AC	1-iteration AC	2-iteration AC
Prostate	82.8 ± 12.9	86.0 ± 4.3	$\mathbf{86.6 \pm 4.1}$
Bladder	87.9 ± 14.2	91.8 ± 5.8	$\mathbf{92.1 \pm 4.7}$
Rectum	84.4 ± 7.3	87.0 ± 4.8	$\mathbf{88.4 \pm 4.8}$
FemurL	89.5 ± 19.5	95.4 ± 1.3	$\mathbf{97.0 \pm 1.5}$
FemurR	87.7 ± 24.0	95.4 ± 2.3	$\mathbf{97.0 \pm 1.5}$

3.5.4 Multitask Random Forest versus Regression Forest

To show the advantages of multitask random forest in deformation field estimation, it was compared to the regression forest in this subsection. For a fair comparison, the same types of features, parameter settings and the auto-context model were used in both methods.

Fig. 3.11 shows a qualitative comparison between deformation fields estimated by the regression forest and the multitask random forest. As mentioned in section 3.2.3, the regression forest suffers the missing boundary problem, i.e., deformation fields may be generated with missing parts of organ boundaries due to mis-predictions of the majority voxels in a small region. With such deformation fields, deformable models will be misled, resulting in poor segmentation results. In contrast, with the help of organ classification the multitask random forest generates more accurate deformation fields. As shown in fig. 3.11, the problem of missing boundaries is well addressed by the multitask random forest.

Table. 3.2 presents a quantitative comparison between the regression forest and the

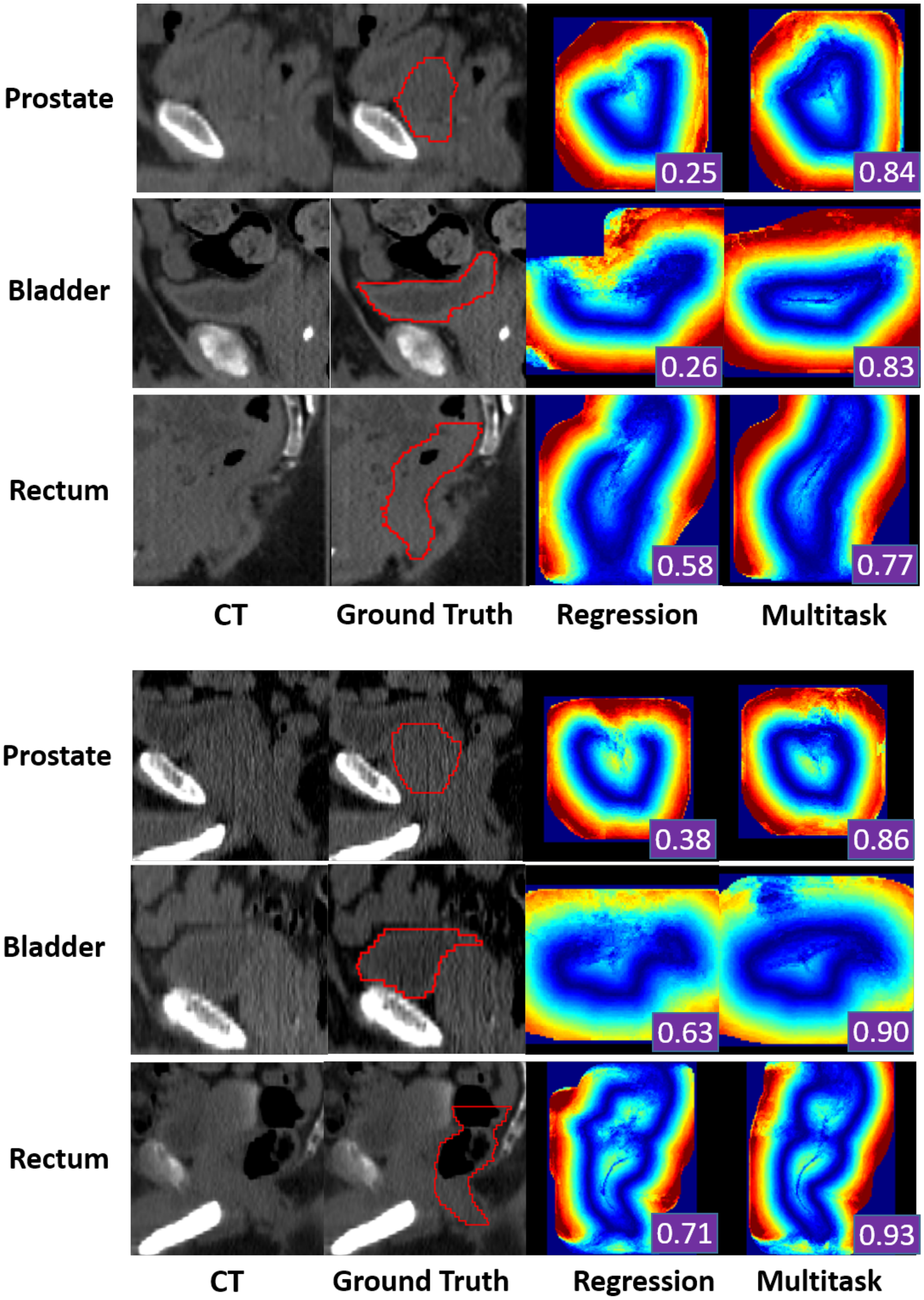


Figure 3.11: Qualitative comparison of deformation fields predicted by the regression forest and the multitask random forest. The red contours are the ground-truth segmentation manually contoured by radiation oncologists. The segmentation accuracy (DSC) obtained by using each estimated deformation field is shown as a white number in the right-bottom of each color-coded image.

multitask random forest for guiding deformable segmentation. The p values calculated from paired t-tests show that the multitask random forest obtained significantly better results than the regression forest in segmenting the prostate, bladder and rectum. For the left and right femoral heads, which have high contrast in CT images, both methods performed equally well. Their slight difference was caused by one failure case of the regression forest.

Table 3.2: Quantitative comparison of segmentation accuracies (DSC) obtained by the regression forest (Regression) and the multitask random forest (Multitask). p-values were computed by paired t-tests. Bold numbers indicate the better performance.

DSC (%)	Regression	Multitask	p value
Prostate	84.0 ± 12.6	$\mathbf{86.6 \pm 4.1}$	$< 10^{-5}$
Bladder	90.6 ± 8.6	$\mathbf{92.1 \pm 4.7}$	$< 10^{-4}$
Rectum	85.2 ± 6.6	$\mathbf{88.4 \pm 4.8}$	$< 10^{-5}$
FemurL	96.5 ± 5.6	$\mathbf{97.0 \pm 1.5}$	0.13
FemurR	96.5 ± 5.8	$\mathbf{97.0 \pm 1.5}$	0.12

3.5.5 Multitask Random Forest versus Classification Forest

In this subsection, the multitask random forest was compared with the classification forest for deformation field estimation. Different from the multitask random forest that directly estimates the deformation field from the intensity image, the classification forest first predicts an organ likelihood map from the intensity image. Then, the deformation field is generated from the organ likelihood map by thresholding and distance transformation. For a fair comparison, the same types of features, parameter settings and the auto-context model were used for both methods. Fig. 3.12 shows qualitative comparisons of classification map and deformation field between the classification forest and the multitask random forest. Compared to the classification forest, it has several advantages to use the multitask random forest for deformation field estimation.

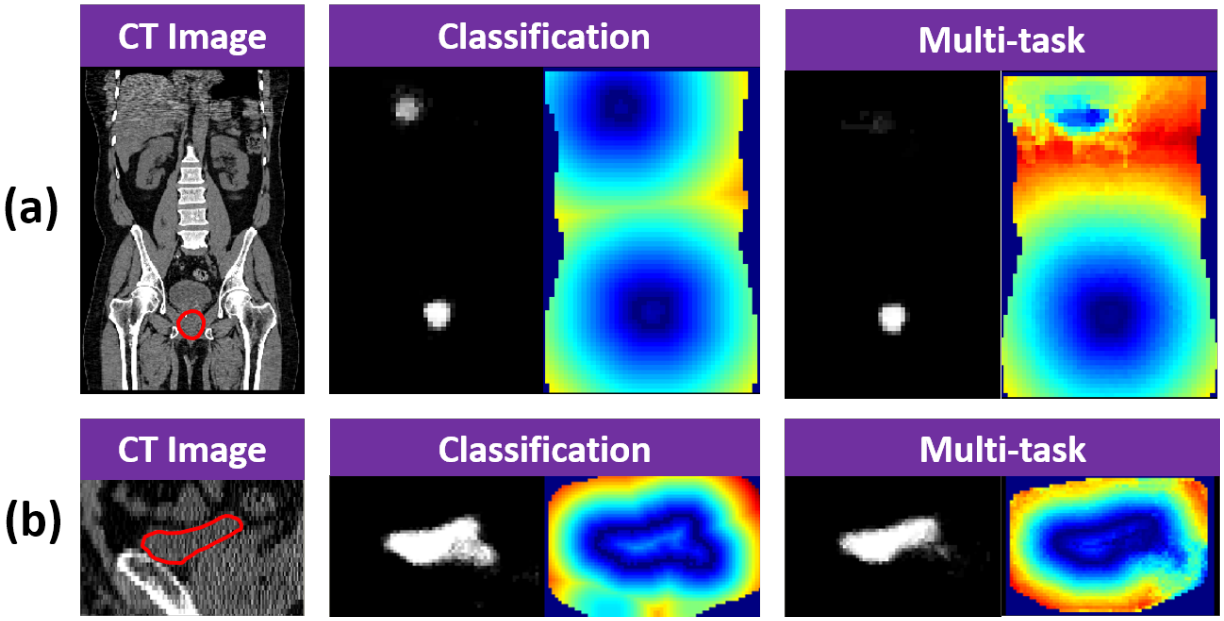


Figure 3.12: Qualitative comparisons of deformation fields estimated by the classification forest and the multitask random forest. (a) and (b) are two typical cases. The red contours are manual segmentations delineated by radiation oncologists.

- In the multitask random forest, as shown in section 3.5.4, the deformation field is improved by exploiting the estimated class information from organ classification. Similarly, the organ likelihood map is also improved by exploiting the estimated deformation information from deformation regression. By joint learning of these two tasks in the multitask random forest and exchanging their predicted information during the auto-context iterations, the complementary information from one task improves the performance of the other task. As seen from both fig. 3.12 (a) and (b), the multitask random forest produces better organ classification maps than the classification forest.

- The classification errors can be easily propagated if the deformation field is derived from the classification map, especially when there are multiple positive responses in the classification map. Fig. 3.12 (a) provides a typical example where mis-classifications in a small region ruin half of the deformation field. In contrast, if the deformation field is voxel-wisely predicted by the multitask random forest, the mis-predictions are restricted only in local regions and will not be propagated.

- As illustrated in fig. 3.13, near the organ boundary deformation fields predicted by the multitask random forest are often smoother than those generated from classification maps because the binarization of classification maps by simple thresholding often produces zigzag organ boundaries.

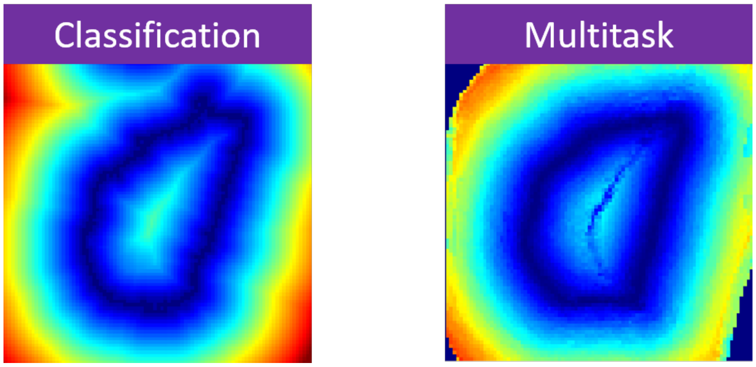


Figure 3.13: Left: deformation field derived from the classification map. Right: deformation field estimated by the multitask random forest.

Table 3.3 presents a quantitative comparison between the classification forest and the multitask random forest for deformation field estimation. The p values calculated from paired t-tests show that the multitask random forest is significantly better than the classification forest for deformation field estimation and guiding regression-based deformable segmentation.

3.5.6 Comparison with Conventional Deformable Models

To show the effectiveness of RDMs, they were compared with conventional classification-based deformable models (CDMs) via modified active shape models. Unlike RDMs, CDMs require good initializations to work well. Once the shape model (3D mesh) is well initialized,

Table 3.3: Quantitative comparison of segmentation accuracies (DSC) obtained by the classification forest (Classification) and the multitask random forest (Multitask). p-values were computed by paired t-tests. Bold numbers indicate the better performance.

DSC (%)	Classification	Multitask	p value
Prostate	85.6 ± 4.2	$\mathbf{86.6 \pm 4.1}$	$< 10^{-3}$
Bladder	90.9 ± 5.2	$\mathbf{92.1 \pm 4.7}$	$< 10^{-5}$
Rectum	86.5 ± 5.2	$\mathbf{88.4 \pm 4.8}$	$< 10^{-5}$
FemurL	96.1 ± 1.4	$\mathbf{97.0 \pm 1.5}$	$< 10^{-5}$
FemurR	96.1 ± 1.4	$\mathbf{97.0 \pm 1.5}$	$< 10^{-5}$

every vertex on the shape model independently deforms along its normal direction to a position with the maximal boundary response. After an one-step deformation of all vertices, the entire shape model is often smoothed or regularized by a shape space (e.g., through PCA) before the next round of deformation. These two steps alternate until convergence or reaching the maximum number of iterations.

In this experiment, a random forest classifier was used to classify every voxel in a testing image into either "organ" or "background". The gradient on the obtained organ likelihood map was used as the boundary response to guide CDMs. After one-step deformation, mesh smoothing and remeshing were used to regularize the shape model. This step is the same as RDMs. For a fair comparison, the random forest classifier used the same types of Haar-like features and the auto-context model as those in RDMs.

Two initialization methods have been tested for CDMs.

- **Box-based initialization.** The regression-based anatomy detection method [Criminisi et al., 2013] was utilized to automatically detect the bounding box of the target organ. Based on the detected box, the mean shape was initialized on the box center and further scaled to fit the box size. After initialization, the shape model deformed in the same way as described above.

- **Multi-resolution strategy.** The mean shape model was initialized to the classification mass center in the coarsest resolution. Once initialized the shape model deformed on the organ likelihood map until convergence. Afterwards, the deformed shape model was used as an initialization to the next finer resolution. The deformation was hierarchically performed until it meets the finest resolution. The multi-resolution parameters were the same with those described in section 3.5.2.

Table 3.4 shows the segmentation accuracies obtained by RDMs and CDMs with the two initialization strategies, respectively. Because CDMs rely on local search to deform, the parameter of search range is critical to segmentation. To optimize the performance of CDMs, the search range of each organ was manually searched from 10 to 35 mm with a step size of 5 mm. From the results listed in table 3.4, it can be seen that CDMs perform reasonably well for organs with rigid shapes and stable positions, such as the prostate and femoral heads, although their performance is still inferior to RDMs'. However, they fail notably when segmenting organs with highly variable shapes, such as the bladder and rectum.

Table 3.4: Quantitative comparison (DSC) between classification-based and regression-based deformable models. Bold numbers indicate the best performance.

DSC (%)	Classification		Regression
	Box	Multi-resolution	
Prostate	83.7 ± 12.3	83.3 ± 12.0	$\mathbf{86.6 \pm 4.1}$
Bladder	73.1 ± 32.4	87.1 ± 20.0	$\mathbf{92.1 \pm 4.7}$
Rectum	53.9 ± 26.9	57.3 ± 33.6	$\mathbf{88.4 \pm 4.8}$
FemurL	95.6 ± 4.6	95.9 ± 7.8	$\mathbf{97.0 \pm 1.5}$
FemurR	95.6 ± 4.1	96.4 ± 5.6	$\mathbf{97.0 \pm 1.5}$

The main reason for those failures is that initialization is demanded by conventional deformable models. However, a good initialization is often difficult to obtain for flexible organs such as the bladder and rectum. Fig. 3.14 presents several typical bounding-box-based initializations for illustration. It can be seen that it is challenging to accurately detect the bounding box of the bladder due to dramatic changes of bladder sizes and positions across subjects. For the rectum initialization, it is even more challenging. As shown in the right panel of fig. 3.14, although the detected bounding boxes (green) are reasonably good, the initialized shapes (green) are still far from the true organ boundaries (red) because of the dissimilarity between the mean rectum shape and individual rectum shapes. The highly variable shapes make the bounding-box based initialization less effective in initializing the rectum compared to other organs, such as the prostate and femoral heads that have relatively stable shapes. The same challenges also apply to the multi-resolution initialization strategy.

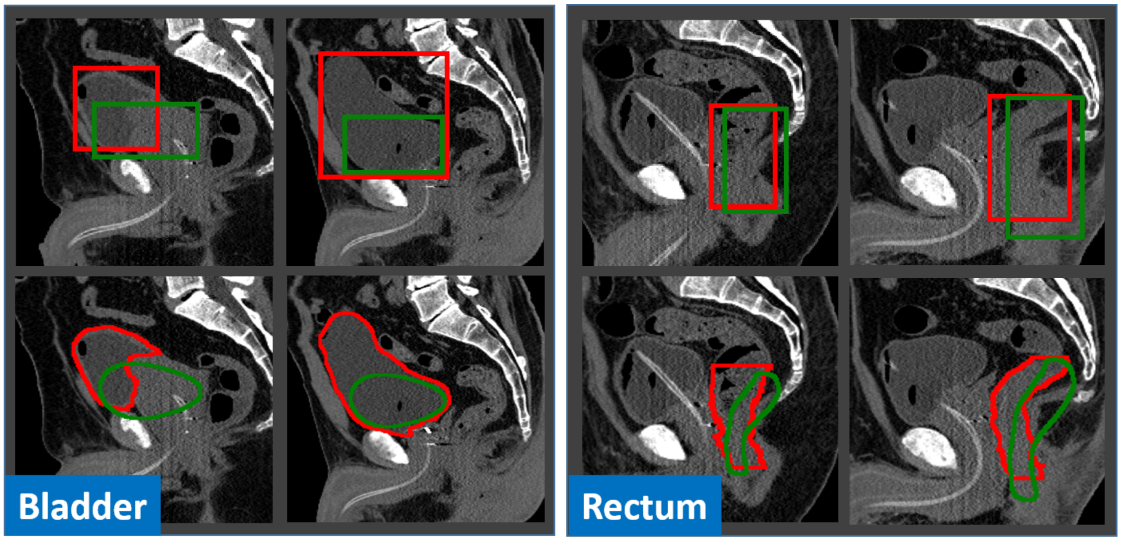


Figure 3.14: Typical cases of bounding-box-based initialization (Left: bladder; Right: rectum). The second row shows the initialized shapes according to the detected bounding boxes in the first row. The red and green contours indicate the ground-truth and the results obtained by anatomy detection, respectively.

Besides the initialization, conventional deformable models (e.g., ASM) still faces another challenge when they are applied to segmenting the rectum. That is the difficulty in determining the search range. Due to the tubular structure of the rectum, large search ranges would easily cause mesh folding as vertices of left rectum wall may find high boundary responses from the right rectum wall and vice versa. On the other hand, small search ranges are insufficient to drive the deformable model onto organ boundaries if the shape model is not well initialized. These two contradictory factors make it infeasible to find a compromise search range. This also explains why the segmentation accuracy of the rectum by CDMs is much lower compared to that of other organs.

In contrast to conventional deformable models, RDMs are guided by deformation fields that provide non-local external forces to overcome the sensitivity of deformable models to initialization. Because of this fact RDMs do not require a model initialization step that is often critical in most deformable segmentation methods. This characteristic renders RDMs suitable for segmenting organs that are difficult to initialize, such as the bladder and rectum. Additionally, the deformation direction and step size of each vertex are optimally predicted during the deformation according to the underlying image appearance. This feature makes RDMs appealing to segment organs with complex shapes, such as the rectum, where the conventional deformation strategies (e.g., normal deformation direction and fixed step size) do not work. All these factors contribute to the success of RDMs in CT pelvic organ segmentation.

3.5.7 Comparison with Other Segmentation Methods

Finally, this subsection compares the proposed method with several existing methods for CT pelvic organ segmentation. Because different methods segment different subsets of the five pelvic organs and use different metrics to measure their performance, the comparisons with other works are separated into multiple tables (tables 3.5 - 3.8). The results show that the proposed method is evaluated on the largest CT dataset and also achieves the best seg-

Table 3.5: Comparison with other existing works based on average surface distance (ASD). Bold numbers indicate the best performance.

ASD (mm)	Lay et al.	Lu et al.	Proposed
Testing dataset	45	188	313
Prostate	3.57 ± 2.01	2.37 ± 0.89	$\mathbf{1.77 \pm 0.66}$
Bladder	3.08 ± 2.25	2.81 ± 1.86	$\mathbf{1.37 \pm 0.82}$
Rectum	3.97 ± 1.43	4.23 ± 1.46	$\mathbf{1.38 \pm 0.75}$
FemurL	1.90 ± 1.18	N/A	$\mathbf{0.49 \pm 0.25}$
FemurR	1.88 ± 0.78	N/A	$\mathbf{0.49 \pm 0.22}$

Table 3.6: Comparison with other existing works based on Dice similarity coefficient (DSC). Bold numbers indicate the best performance.

DSC	Martinez	Proposed
Testing dataset	86	313
Prostate	0.87 ± 0.07	$\mathbf{0.87 \pm 0.04}$
Bladder	0.89 ± 0.08	$\mathbf{0.92 \pm 0.05}$
Rectum	0.82 ± 0.06	$\mathbf{0.88 \pm 0.05}$

mentation accuracy except that the positive predictive value (PPV) of the proposed method in the prostate segmentation is slightly lower than Chen [Chen et al., 2011]. However, the sensitivity (SEN) of the proposed method is much higher than theirs. Besides, a careful scrutiny on these tables reveals that the improvement of the proposed method over other existing methods is larger on the bladder and rectum than other organs. This is mainly because it is difficult for conventional deformable models to segment organs that are difficult to initialize and with highly variable shapes. However, with the introduction of non-local deformations, adaptive deformation direction and step size, and spatially varying deformations, these limitations can be effectively addressed by RDMs.

It is worth noting that most existing works use either sophisticated methods for model initialization [Lay et al., 2013][Lu et al., 2012][Martínez et al., 2014] or rely on shape priors [Lay et al., 2013][Freedman et al., 2005][Costa et al., 2007][Chen et al., 2011][Lu et al.,

Table 3.7: Comparison with other existing works based on MEDIAN sensitivity (SEN) and positive predictive value (PPV). Bold numbers indicate the best performance.

Median		Freedman et al.	Chen et al.	Proposed
Testing dataset		48	185	313
Prostate	SEN	0.83	0.84	**0.90**
	PPV	0.85	**0.87**	0.86
Rectum	SEN	0.74	0.71	**0.91**
	PPV	0.85	0.76	**0.89**

Table 3.8: Comparison with other existing works based on MEAN sensitivity (SEN) and positive predictive value (PPV). Bold numbers indicate the best performance.

Mean		Rousson et al.	Costa et al.	Proposed
Testing dataset		16	16	313
Prostate	SEN	0.84	0.81	**0.88**
	PPV	0.79	**0.85**	**0.85**
Bladder	SEN	N/A	0.75	**0.94**
	PPV	N/A	0.80	**0.92**

2012][Martínez et al., 2014][Rousson et al., 2005] to regularize the segmentation. In contrast, the proposed method uses a fairly simple initialization method (i.e., initialize the mean shape model in the image center), and it does not rely on shape priors (e.g., PCA shape analysis). It is interesting to observe that even with this setup the proposed method still results in more accurate outcomes when compared to previous methods. This demonstrates the robustness of the proposed method to initialization and the effectiveness of the proposed method in CT pelvic organ segmentation.

3.6 Summary

Planning-CT segmentation plays an important role in the image guided radiotherapy of prostate cancer. The goal of planning-CT segmentation is to automatically and accurately segment the prostate, bladder, rectum and two femoral heads from planning CT images. The segmentations can be used for dose planning that designs the direction and dose of each radiation beam to accurately deliver the prescribed dose on the prostate while sparing nearby healthy tissues. Therefore, the treatment efficacy heavily relies on the segmentation accuracy.

Conventional deformable models are sensitive to initialization and are not flexible to segment organs with tubular shapes. These problems limit their performance in CT pelvic organ segmentation. To overcome them, this chapter proposed explicitly learning a deformation field to guide deformable segmentation. To reliably estimate the deformation field from local image appearance, two novel techniques were proposed. *First*, an auto-context model was adopted to iteratively refine the estimated deformation field. By considering predicted deformations in the spatial neighborhood, the auto-context model suppresses prediction noise and improves the spatial consistency of deformation field compared to independent voxel-wise prediction. *Second*, a multitask random forest was proposed. It couples deformation regression with organ classification in a single random forest. By joint learning of these two tasks in the auto-context model, the predicted information from each task can be used as context features to help the other task. Compared to the random forest trained only for deformation regression, the multitask random forest improves the robustness of deformation field estimation. *Finally*, a regression-based deformable model was proposed to hierarchically deform the shape model based on the estimated deformation field. It effectively addresses the limitations of conventional deformable models by using non-local deformations and adaptive deformation parameters.

Extensive experiments on a large pelvic CT dataset showed that the proposed method is effective in CT pelvic organ segmentation. Compared to the regression forest, the multi-task random forest can overcome the problem of missing boundary and yields more accurate deformation fields. Compared to conventional classification-based deformable models, explicitly learning a deformation field helps deformable models overcome the sensitivity to initialization and increase their flexibility to segment tubular organs. Compared to other existing methods, the proposed method exhibited very competitive segmentation performance, especially for the bladder and rectum that are difficult to robustly initialize.

CHAPTER 4 : INCREMENTAL LEARNING FOR TREATMENT-CT SEGMENTATION

As shown in fig. 1.1, image guided radiotherapy (IGRT) consists of a planning stage followed by a treatment stage. In *the planning stage*, a planning CT is acquired from the patient, and major pelvic organs are segmented for treatment planning. In *the treatment stage*, a treatment CT can be acquired right before the radiation therapy at each treatment day (e.g., in adaptive radiation therapy). Since the treatment CT captures a present snapshot of the patient's anatomy, radiation oncologists are able to adapt the treatment plan to precisely target radiation dose to the *current* position of tumors and avoid neighboring healthy tissues. Consequently, IGRT increases the probability of tumor control and typically shortens radiation therapy schedules [Xing et al., 2006, Dawson and Jaffray, 2007]. In order to adapt the treatment plan in an online fashion, it is important to localize the prostate in the daily treatment images fast and accurately. Thus, an automatic prostate localization algorithm would be a valuable asset in IGRT.

However, prostate localization in treatment CTs (treatment-CT segmentation) is challenging for three reasons. First, unlike the planning CT, the treatment CTs are typically acquired with low dose protocols in order to reduce unnecessary radiation exposure to patients during treatment. As a result, the image contrast to noise ratio of treatment CT is relatively lower compared to the planning CT. Second, due to the existence of bowel gas and filling, the image appearance of treatment CTs can change dramatically. Third, unpredictable daily prostate motion [Liu et al., 2010] further complicates the treatment-CT segmentation.

Although many methods have been proposed for prostate localization/segmentation, their accuracy is limited as they overlook a remarkable opportunity for treatment-CT segmentation that is inherent in the IGRT workflow. In fact, at each treatment day several CT scans of the patient may have already been acquired and segmented in the planning day and previous treatment days. If the prostate appearance characteristics of this *specific* patient can be learned from these *patient-specific* images, an algorithm could exploit this information to localize the prostate much more effectively.

To this end, a novel learning scheme, namely incremental learning with selective memory (ILSM) [1], is proposed in this chapter for fast and accurate localization of the prostate in treatment CTs. Compared with previous prostate localization methods, the contributions of this work are two-fold: 1) by leveraging the large amount of population data (CT images of other patients) and the very limited amount of patient-specific data, ILSM is able to learn patient-specific characteristics from *only one image of the patient* and apply the learned model to the localization of beginning treatment CTs; 2) the proposed method can obtain comparable (if not better) localization accuracy to the state-of-the-art methods while substantially reducing the runtime to 4 seconds. Extensive validations show that the proposed method satisfies both accuracy and efficiency requirements in the IGRT workflow. Also, compared to previous methods [Li et al., 2012, Liao et al., 2013, Gao et al., 2012b] that require manual annotation of the entire prostate on the patient-specific training images, ILSM needs only the annotations of seven prostate anatomical landmarks, thereby significantly

[1] This work was published in IEEE Transactions on medical imaging [Gao et al., 2014]. This chapter uses parts of text descriptions and figures from the published paper.

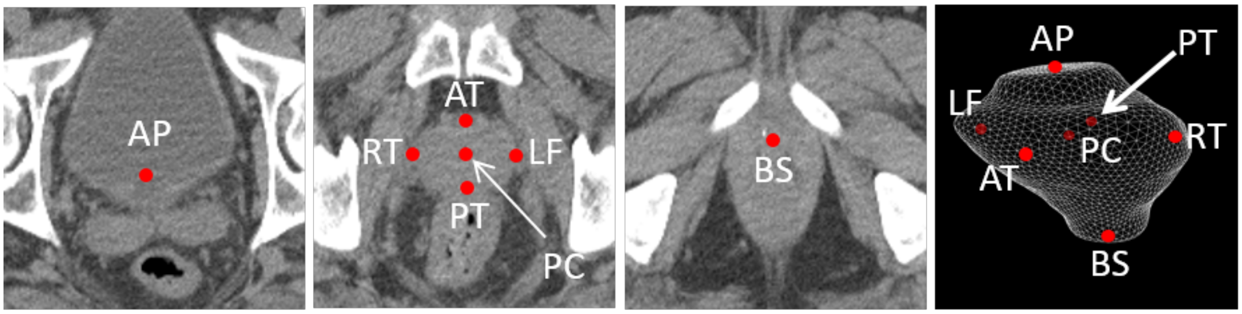


Figure 4.1: Seven prostate anatomical landmarks: apex center (AP), prostate center (PC), right lateral point (RT), left lateral point (LF), posterior point (PT), anterior point (AT) and base center (BS).

reducing the labor required for manual annotations.

To leverage both population and patient-specific data, the learning framework (ILSM) starts with learning a population-based discriminative appearance model. This model is then "personalized" according to the appearance information from CT images of the specific patient under treatment. Instead of either preserving or discarding *all* knowledge learned from the population, the proposed method *selectively* inherits the part of population-based knowledge that is in accordance with the current patient and at the same time *incrementally learns* the patient-specific characteristics. This is where the name "incremental learning with selective memory" comes from. Once the population-based discriminative appearance model is personalized, it can be used to detect distinctive anatomical landmarks in new treatment images of the same patient for fast prostate localization. Compared with traditional learning schemes, such as pure patient-specific learning, population-based learning, and mixture learning with patient-specific and population data, ILSM exhibits better capability to capture the patient-specific characteristics embedded in the data.

The proposed method aims to localize the prostate in daily treatment images via learning a set of local discriminant appearance models. Specifically, these models are used as anatomy detectors to detect distinctive prostate anatomical landmarks as shown in fig. 4.1. Based on the detected landmarks, multiple patient-specific shape atlases (i.e., prostate shapes segmented in planning and previous treatment images) can be aligned onto the treatment image space by random sample consensus (RANSAC) [Fischler and Bolles, 1981]. Finally, majority voting is adopted to fuse the labels from different shape atlases.

As shown in fig. 4.2, the proposed method consists of three components: 1) cascade learning for constructing population-based anatomy detectors, 2) incremental learning with selective memory for personalizing anatomy detectors to individual patients, and 3) multi-atlas RANSAC for localizing the prostate in the treatment CT. Each step will be detailed in the following sections.

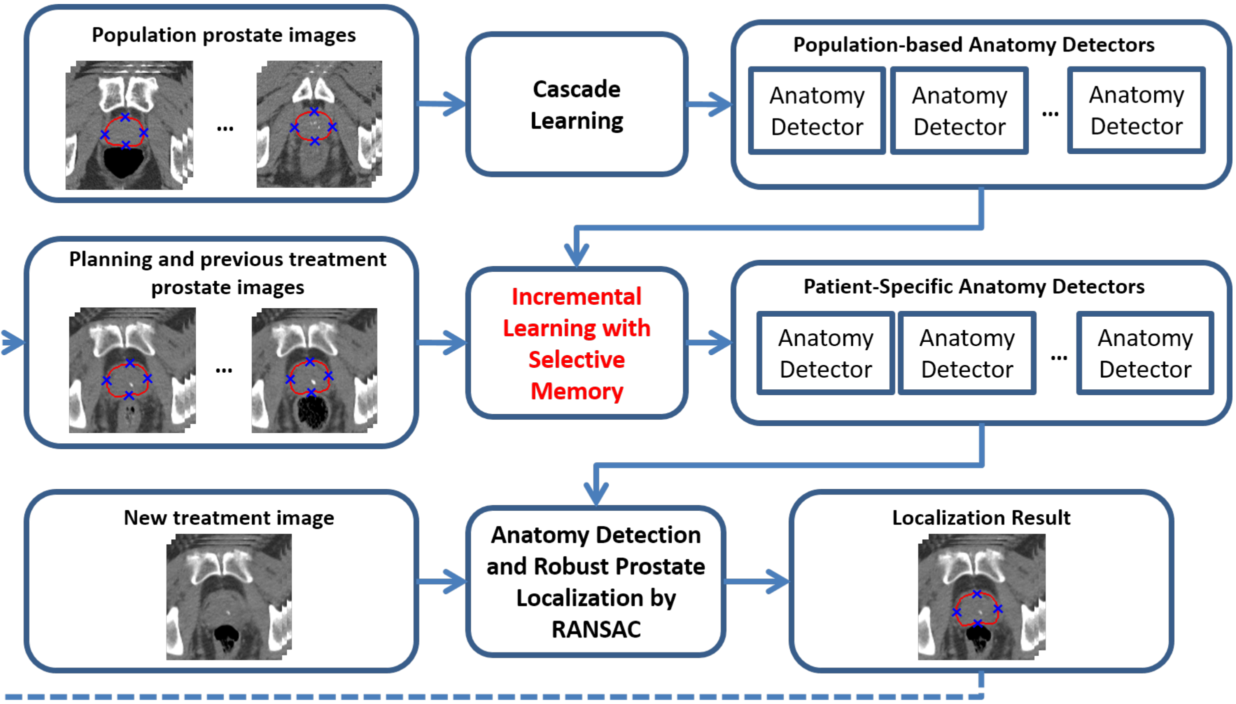

Population prostate images
...
Cascade Learning
Population-based Anatomy Detectors
Anatomy Detector
Anatomy Detector
...
Anatomy Detector
Planning and previous treatment prostate images
...
Incremental Learning with Selective Memory
Patient-Specific Anatomy Detectors
Anatomy Detector
Anatomy Detector
...
Anatomy Detector
New treatment image
Anatomy Detection and Robust Prostate Localization by RANSAC
Localization Result

Figure 4.2: The flowchart of the proposed method to localize the prostate in treatment CT images

4.1 Cascade Learning for Anatomy Detection

The proposed prostate localization method relies on several anatomical landmarks of the prostate. Inspired by Viola's face detection work [Viola and Jones, 2004], this work uses a learning-based detection method that formulates landmark detection as a classification problem. Specifically, for each image, the voxel of the specific landmark is positive and all others are negatives. In the training stage, a cascade learning framework is employed to learn a sequence of classifiers to gradually separate negatives from positives (fig. 4.3). Compared to learning only a single classifier, cascade learning has shown better classification accuracy and runtime efficiency [Viola and Jones, 2004][Zhan et al., 2011b]. Mathematically, cascade learning can be formulated as

Input: Positive voxel set $X_{\mathcal{P}}$, negative voxel set $X_{\mathcal{N}}$, and label set $\mathcal{L} = \{+1, -1\}$

Classifier: $C(x) : \mathbb{F}(x) \to \mathcal{L}$, $\mathbb{F}(x)$ denotes the appearance features of a voxel x

Initial Set: $X_0 = X_{\mathcal{P}} \cup X_{\mathcal{N}}$

Objective: Optimize $C_k, k = 1, 2, \cdots, K$, such that

$$X_0 \supseteq X_1 \supseteq \cdots \supseteq X_k \supseteq \cdots \supseteq X_K,\ X_K \supseteq X_{\mathcal{P}}, \text{ and } \|X_K \cap X_{\mathcal{N}}\| \leq \tau \|X_{\mathcal{P}}\|$$

where $X_k = \{x | x \in X_{k-1} \text{ and } C_k(x) = +1\}$, and τ controls the tolerance ratio of false positives.

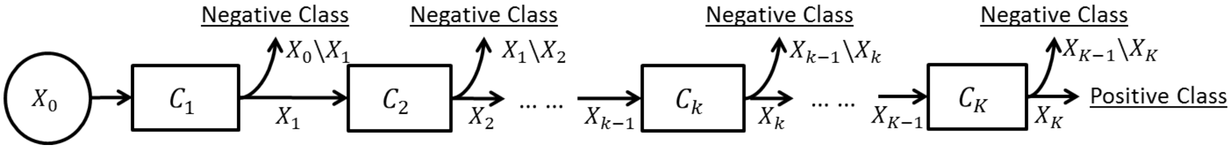


Figure 4.3: Illustration of cascade learning.

The cascade classifiers $C_k, k = 1, 2, \cdots, K$, are optimized sequentially. As shown in eq. 4.1 below, C_k is optimized to minimize the false positives left over by the previous $k - 1$

classifiers.

$$C_k = \arg\min_{C} \|\{x|x \in X_{k-1} \cap X_{\mathcal{N}} \text{ and } C(x) = +1\}\| \;\; \text{s.t.} \;\; \forall x \in X_{\mathcal{P}}, C(x) = +1 \qquad (4.1)$$

where $\|.\|$ denotes the cardinality of a set. It is worth noting that the constraint in eq. 4.1 can be simply satisfied by adjusting the threshold of classifier C_k [Viola and Jones, 2004] to make sure that all positive training samples are correctly classified. This cascade learning framework is general to any image feature and classifier. Extended Haar wavelets [Zhan et al., 2011a, Zhan et al., 2008] and the Adaboost [Viola and Jones, 2004] classifier are employed in this study.

Once the cascade classifiers $\{C_k(x)\}$ are learned, they have captured the appearance characteristics of the specific anatomical landmark. Given a testing image, the learned cascade is applied to each voxel. The voxel with the highest classification score after going through the entire cascade is selected as the detected landmark. To increase the efficiency and robustness of the detection procedure, a multi-scale scheme is further adopted. Specifically, the detected landmark in the coarse resolution serves as the initialization for landmark detection in a following finer resolution in which the landmark is only searched in a local neighborhood centered by the initialization. In this way, the search space is largely reduced and the detection procedure is more robust to local minima.

4.2 Incremental Learning with Selective Memory

4.2.1 Motivation

Using cascade learning, one can learn anatomy detectors from training images of different patients (*population-based learning*). However, since intra-patient anatomy variations are much less pronounced than inter-patient variations (fig. 4.4), patient-specific appearance information available in the IGRT workflow should be exploited in order to improve the detection accuracy for an individual patient. Unfortunately, the number of patient-specific images is often very limited, especially at the beginning of IGRT. To overcome this problem, one may apply random spatial/intensity transformations to produce more "synthetic" training samples with larger variability. However, these artificially created transformations may not capture the real intra-patient variations, e.g., the uncertainty of bowel gas and filling (fig. 4.4). As a result, cascade learning using only patient-specific data (*pure patient-specific learning*) often suffers from overfitting. One can also mix population and patient-specific images for training (*mixture learning*). However, since patient-specific images are the "minority" in the training samples, detectors trained by mixed samples might not capture patient-specific characteristics very well. To address this problem, a new learning scheme named ILSM is proposed to combine the general information in the population images with the personal information in the patient-specific images. Specifically, population-based anatomy detectors serve as an initial appearance model and are subsequently "personalized" by the limited patient-specific data. In particular, ILSM uses *backward pruning* to discard obsolete population appearance information and *forward learning* to incorporate the online-learned patient-specific appearance characteristics.

4.2.2 Notations

Denote $D^{\text{pop}} = \{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{pop}}\}$ as a population-based anatomy detector (learned as outlined in section 4.1) that contains a cascade of classifiers. $X_{\mathcal{P}}^{\text{pat}}$ and $X_{\mathcal{N}}^{\text{pat}}$ are positives and negatives from the patient-specific training images, respectively. $D(x)$ denotes the class label (landmark vs non-landmark) of voxel x predicted by detector D.

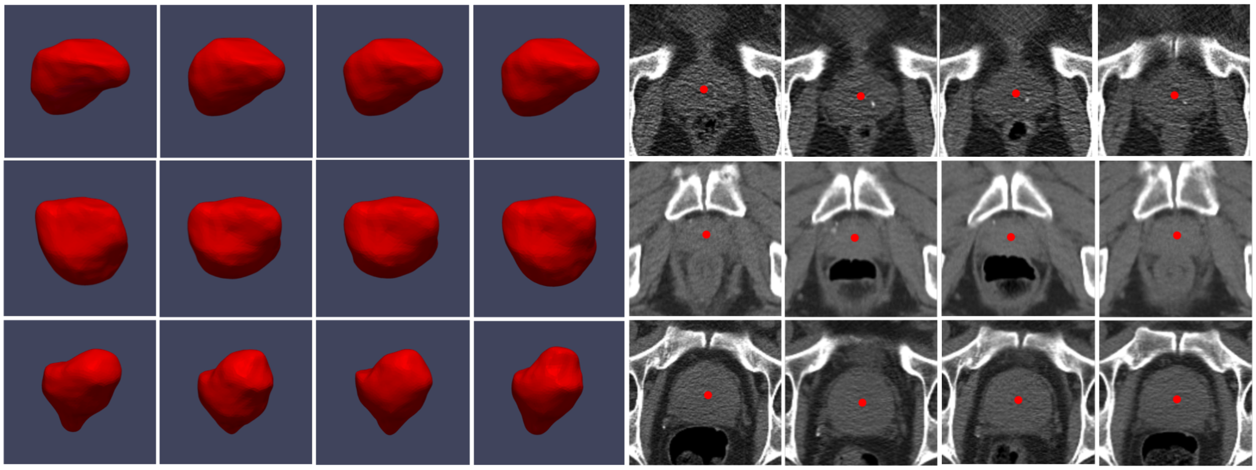


Figure 4.4: Inter- and intra-patient prostate shape and appearance variations. The red points denote the prostate center. Each row represents prostate shapes and images from the same patient.

4.2.3 Backward Pruning

The general appearance model learned from a population is not necessarily applicable to the specific patient. More specifically, the anatomical landmarks in the patient-specific images (positives) may be classified as negatives by the population-based anatomy detectors, i.e., $\exists k \in \{1, 2, \cdots, K^{\text{pop}}\}, \exists x \in X_{\mathcal{P}}^{\text{pat}}, C_k^{\text{pop}}(x) = -1$. In order to discard these parts of the population appearance model that do not fit the patient-specific characteristics, *backward pruning* is proposed to tailor the population-based detector. As shown in alg. 4.1, in backward pruning, the cascade is pruned from the last level until all patient-specific positives successfully pass through the cascade. This is equivalent to searching for the maximum number of cascade levels that could be preserved from the population-based anatomy detector (eq. 4.2).

$$K^{\text{bk}} = \max\{k | C_i^{\text{pop}}(x) = +1, \forall i \leq k, \forall x \in X_{\mathcal{P}}^{\text{pat}}\} \tag{4.2}$$

Algorithm 4.1 Backward pruning algorithm.

```
Input: $D^{\text{pop}} = \{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{pop}}\}$ - the population-based detector
       $X_{\mathcal{P}}^{\text{pat}}$ - patient-specific positive samples
Output: $D^{\text{bk}}$ - the tailored population-based detector
Init: $k = K^{\text{pop}}$, $D^{\text{bk}} = D^{\text{pop}}$.
while $\exists x \in X_{\mathcal{P}}^{\text{pat}} : D^{\text{bk}}(x) = -1$ do
    $D^{\text{bk}} = D^{\text{bk}} \backslash C_k^{\text{pop}}$
    $k = k - 1$
end while
$K^{\text{bk}} = k$
return $D^{\text{bk}} = \{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{bk}}\}$
```

4.2.4 Forward Learning

Once the population cascade has been tailored, the remaining cascade of classifiers encodes the population appearance information that is consistent with the patient-specific characteristics. Yet until now no real patient-specific information has been incorporated into the cascade. More specifically, false positives might exist in the patient-specific samples, i.e., $\exists x \in X_{\mathcal{N}}^{\text{pat}}, \forall k \leq K^{\text{bk}}, C_k^{\text{pop}}(x) = +1$. In the forward learning stage, the remaining cascade from the backward pruning algorithm is used as an initialization, and the cascade learning is re-applied to eliminate the patient-specific false positives left over by the previously inherited population classifiers. As shown in alg. 4.2, a greedy strategy is adopted to sequentially optimize a set of additional patient-specific classifiers $\{C_k^{\text{pat}}, k = 1, 2, \cdots, K^{\text{pat}}\}$.

After backward pruning and forward learning, the personalized anatomy detector includes two groups of classifiers (fig. 4.5). While $\{C_k^{\text{pat}}, k = 1, 2, \cdots, K^{\text{pat}}\}$ encode patient-specific characteristics, $\{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{bk}}\}$ contain population information that is applicable to this specific patient. This information effectively remedies the limited variability from the small number of patient-specific training images.

Algorithm 4.2 Forward learning algorithm.

```
Input: $D^{\text{bk}} = \{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{bk}}\}$ - the tailored population-based detector
       $X_{\mathcal{P}}^{\text{pat}}$ - patient-specific positive samples
       $X_{\mathcal{N}}^{\text{pat}}$ - patient-specific negative samples
Output: $D^{\text{pat}}$ - the patient-specific detector
Init: $k = 1$, $D^{\text{pat}} = D^{\text{bk}}$, $X_0 = \{x | x \in X_{\mathcal{N}}^{\text{pat}} \cup X_{\mathcal{P}}^{\text{pat}}, D^{\text{bk}}(x) = +1\}$
while $\|X_{k-1} \cap X_{\mathcal{N}}^{\text{pat}}\| > \tau \|X_{\mathcal{P}}^{\text{pat}}\|$ do
    Train the classifier by minimizing eq. 4.3 below

        $C_k^{\text{pat}} = \arg\min_{C} \|\{x | x \in X_{k-1} \cap X_{\mathcal{N}}^{\text{pat}}, C(x) = +1\}\|$
        $s.t.\ \forall x \in X_{\mathcal{P}}^{\text{pat}}, C(x) = +1$        (4.3)

    $X_k = \{x | x \in X_{k-1}, C_k^{\text{pat}}(x) = +1\}$
    $D^{\text{pat}} = D^{\text{pat}} \cup C_k^{\text{pat}}$
    $k = k + 1$
end while
$K^{\text{pat}} = k - 1$
return $D^{\text{pat}} = \{C_k^{\text{pop}}, k = 1, 2, \cdots, K^{\text{bk}}\} \cup \{C_k^{\text{pat}}, k = 1, 2, \cdots, K^{\text{pat}}\}$
```

$\|.\|$ *denotes the cardinality of a set.* τ *is the parameter controlling the tolerance ratio of false positives.*

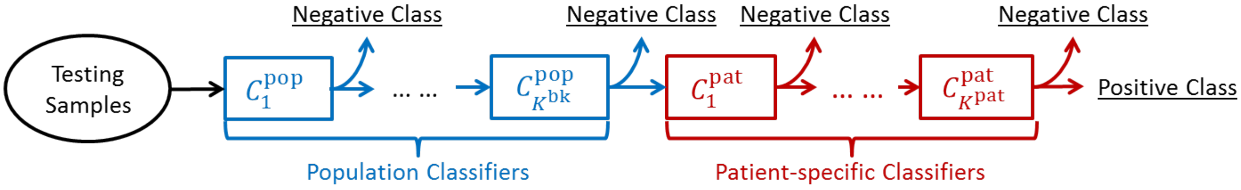


Figure 4.5: Incrementally learned anatomy detector.

4.2.5 Insight of ILSM

In fact, pure patient-specific learning (PPAT) and traditional incremental learning (IL) can also be employed to incorporate the patient-specific information. It is interesting to compare ILSM with PPAT and IL. PPAT only uses patient-specific data for training. In other words, it completely discards all knowledge learned from the population, which is known as "catastrophic forgetting" [Polikar et al., 2001]. The method is prone to overfitting if the patient-specific data is very limited. On the other hand, IL aims to gradually adapt the clas-

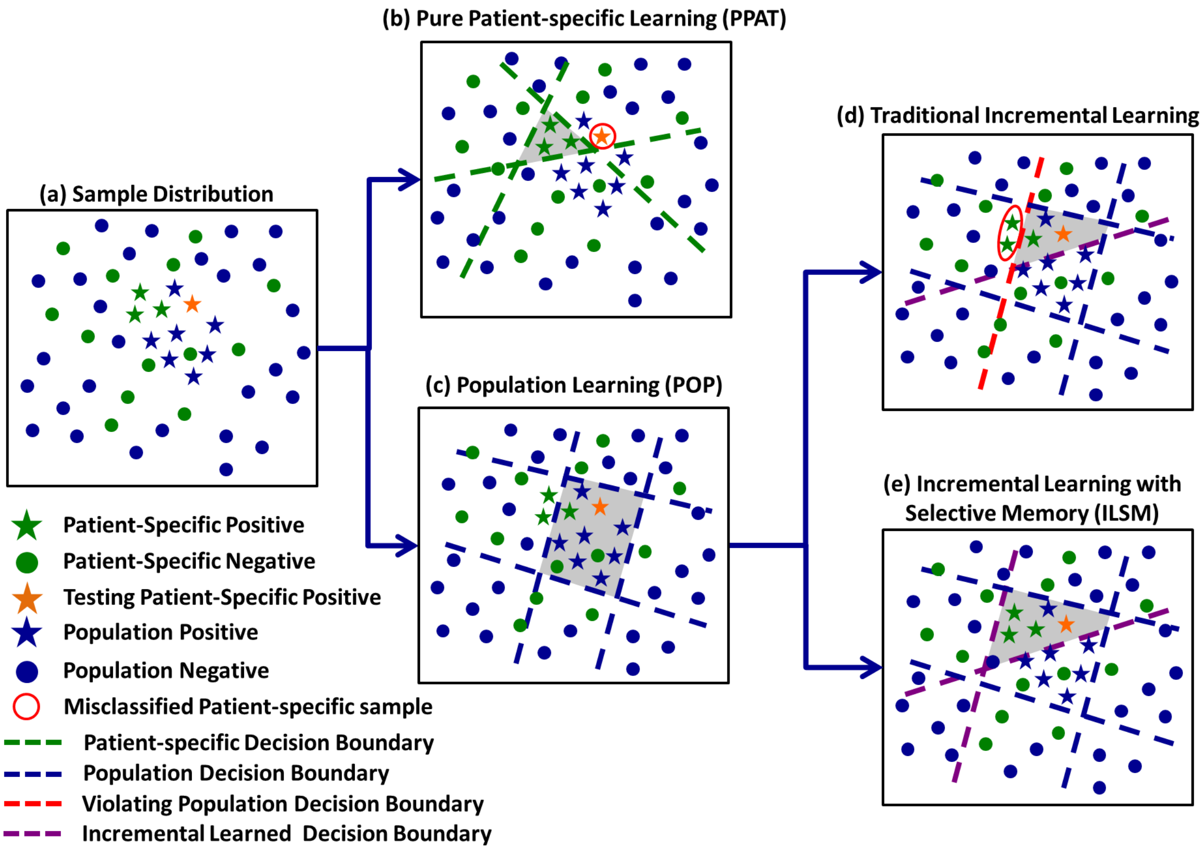

(a) Sample Distribution
(b) Pure Patient-specific Learning (PPAT)
(c) Population Learning (POP)
(d) Traditional Incremental Learning (IL)
(e) Incremental Learning with Selective Memory (ILSM)
Patient-Specific Positive
Patient-Specific Negative
Testing Patient-Specific Positive
Population Positive
Population Negative
Misclassified Patient-specific sample
Patient-specific Decision Boundary
Population Decision Boundary
Violating Population Decision Boundary
Incremental Learned Decision Boundary

Figure 4.6: A schematic illustration of differences among PPAT, IL and ILSM.

sifiers with new data. It assumes that the previously learned knowledge is always applicable for the new incoming data and tries to "remember" all of them. Consequently, the incrementally learned patient-specific knowledge can be impaired by incompatible population-based knowledge. In fact, in the context of cascade learning, IL can be regarded as the proposed method without backward pruning. In contrast to PPAT and IL, ILSM aims to "selectively" remember the subset of pre-learned knowledge consistent with the characteristics in the new data. ILSM's "selective memory" helps to overcome the limitations of the other two methods.

Fig. 4.6 schematically explains the differences among PPAT, IL, and ILSM from the perspective of decision boundary refinement. Fig. 4.6(a) shows the sample distribution in a 2D feature space. Stars and circles represent positive and negative samples, respectively.

Blue stars/circles are population training samples, and green ones denote patient-specific samples. The orange star is a testing sample.

As shown in fig. 4.6(b), since PPAT only uses patient-specific samples (stars/circles in green), the generated decision boundaries (green lines) closely encompass the positive patient-specific training samples (green stars). These decision boundaries might overfit the very limited number of patient-specific samples. As a result, a testing sample (the orange star in fig. 4.6(b)), which has slight differences from these training samples, is mis-classified.

IL derives the decision boundaries in two steps. First, as shown in fig. 4.6(c), it learns the decision boundaries using population samples (blue stars/circles). Second, these boundaries are adapted to accommodate patient-specific samples. For example, in fig. 4.6(d), an additional purple line is generated to separate patient-specific positives (green stars) and negatives (green circles). Since IL aims to preserve all pre-learned population-based boundaries (blue and red lines), some patient-specific data (circled in red in fig. 4.6(d)) are still mis-classified due to the "unforgettable" decision boundary (the red line in fig. 4.6(d)).

Similar to IL, ILSM also starts from a population-based learning (fig. 4.6(c)). However, in adapting the decision boundaries to patient-specific samples, it is able to "forget" some pre-learned knowledge that is not applicable to the patient-specific data. Specifically, the obsolete decision boundary (red line in fig. 4.6(d)) can be discarded in the "backward prunning" step of ILSM. Hence, ILSM can correctly classify all patient-specific data (fig. 4.6(e)). In addition, by re-using some applicable population-based decision boundaries (blue lines), the overfitting risks are also highly reduced. In this way, ILSM can address the limitations of both PPAT and IL.

In fact, ILSM can be considered as a more general learning framework of which IL and PPAT are just two special cases. In alg. 4.1, if all positive samples from patient-specific images can be correctly classified by D^{pop}, the backward pruning will stop at the first place, i.e., $K^{\text{bk}} = K^{\text{pop}}$ (alg. 4.1). The learned patient-specific detector will then preserve all population characteristics, which is the same as IL. At the other extreme, if the population-based detector is completely incompatible with patient-specific samples, the backward pruning will not stop until $D^{\text{bk}} = \emptyset$ (alg. 4.2), which means all population-based classifiers will be discarded. In such cases, the forward learning will start from scratch with patient-specific samples and ILSM becomes equivalent to PPAT. In practice, this situation rarely happens. A manual check of trained detectors shows that all of these cases happened when sufficient patient-specific images (≥ 5) had already been collected. In such situation, PPAT is capable to obtain similar performance with ILSM.

4.3 Robust Prostate Localization by RANSAC

Once the population-based anatomy detectors are "personalized" by ILSM, they are used to detect the corresponding prostate anatomical landmarks (fig. 4.1) in new treatment images. Based on the detected landmarks, any patient-specific prostate shape model (e.g., the prostate shape delineated in the planning stage) can be aligned onto the treatment image space for fast localization. For robust performance against wrongly detected landmarks, the RANSAC algorithm [Fischler and Bolles, 1981] is used to estimate the optimal transformation that fits the shape model onto the detected landmarks (alg. 4.3). Considering the limited number of anatomical landmarks (seven) as well as in the interest of computational efficiency, rigid transformation is used in this work.

One can simply align the planning prostate shape onto the treatment image for localization, which is referred as single-atlas RANSAC. However, due to the daily shape variations

Algorithm 4.3 Robust Surface Transformation by RANSAC

```
Definition: $\mathbb{N} = 7$ - number of anatomical landmarks
Input: $p_k, k = 1, 2, \cdots, \mathbb{N}$ - landmarks in one patient-specific training image $I_{\text{pat}}$
       $m_k, k = 1, 2, \cdots, \mathbb{N}$ - detected landmarks in the treatment image $I_{\text{treat}}$
       $\mathcal{M}$ - minimum number of landmarks required for transformation estimation
       $\eta$ - threshold to determine whether a landmark agrees on the transformation
Output: $\mathrm{T}^{\text{opt}}$ - optimal transformation between prostate shapes in $I_{\text{pat}}$ and $I_{\text{treat}}$
Init: $\mathrm{T}^{\text{opt}} = \text{nil}$, $\mathcal{E}^{\text{opt}} = \text{infinity}$
for each landmark subset $S$ of $\{1, 2, \cdots, \mathbb{N}\}$ with $\|S\| \geq \mathcal{M}$ do
    $\mathrm{T}_S^{\text{maybe}} = \arg\min_{\mathrm{T}} \sum_{k \in S} \|m_k - \mathrm{T}p_k\|_2$
    for any $k$ not in $S$ do
        if $\|m_k - \mathrm{T}_S^{\text{maybe}} p_k\|_2 < \eta$ then
            add $k$ into $S$
        end if
    end for
    $\mathrm{T}_S^{\text{opt}} = \arg\min_{\mathrm{T}} \sum_{k \in S} \|m_k - \mathrm{T}p_k\|_2$, $\mathcal{E}_S^{\text{opt}} = \sum_{k \in S} \|m_k - \mathrm{T}_S^{\text{opt}} p_k\|_2$
    if $\mathcal{E}_S^{\text{opt}} < \mathcal{E}^{\text{opt}}$ then
        $\mathcal{E}^{\text{opt}} = \mathcal{E}_S^{\text{opt}}$, $\mathrm{T}^{\text{opt}} = \mathrm{T}_S^{\text{opt}}$
    end if
end for
return $\mathrm{T}^{\text{opt}}$
```

under radiotherapy, the performance of using a single shape model is usually limited. To overcome this limitation, a multi-atlas RANSAC is proposed for robust prostate localization. Instead of using a single shape model, this work uses all patient-specific shape models available in both planning and previous treatment stages for multi-atlas labeling of a new treatment image. In other words, each patient-specific shape model is treated as a shape atlas. Once anatomical landmarks are detected in the new treatment image, all available shape atlases can be independently aligned onto the new treamtent image space by RANSAC (alg. 4.3). Then, majority voting is adopted to fuse the labels from different shape atlases. Thus, by integrating all patient-specific shape information into a multi-atlas scheme, the localization procedure is more robust to daily shape variations than single-atlas RANSAC. Fig. 4.7 illustrates the multi-atlas model fitting process.

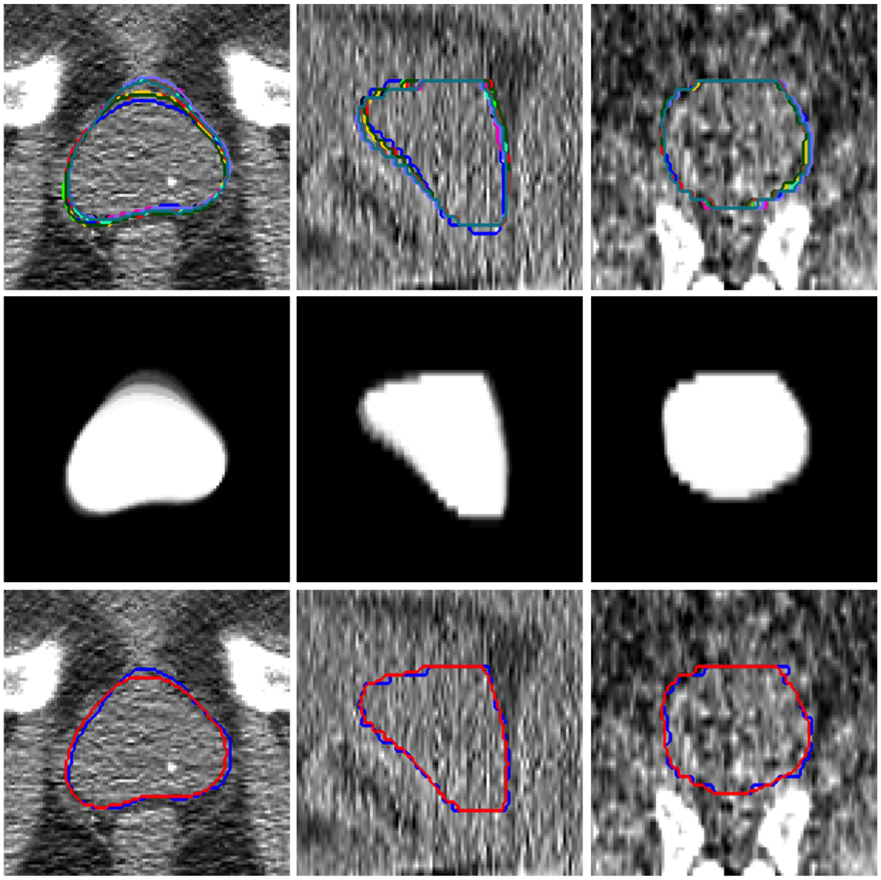


Figure 4.7: Illustration of multi-atlas RANSAC. The first row shows aligned patient-specific shape models (denoted by different colors) in a new treatment CT. The second row shows the prostate likelihood map by averaging all aligned prostate masks. The third row shows the final segmentation (red contours) overlaid by the ground truth (blue contours). Three columns are the middle slices in transverse, sagittal and coronal views, respectively.

4.4 Experimental Results

Extensive experiments have been conducted and summarized in this section for evaluating the performance of the proposed method and comparing it with other alternative methods. Specifically, this section is organized as detailed below.

- Section 4.4.1 presents the description of two image datasets used in the experiments.

- Section 4.4.2 describes the accuracy and efficiency requirements of IGRT.

- Section 4.4.3 discusses the parameter selection and experimental setting.
- Section 4.4.4 analyzes the number of cascade classifiers pruned and appended in the backward pruning and forward learning stages, respectively.
- Section 4.4.5 evaluates the proposed method by comparing it with numerous alternative methods.
- Section 4.4.6 reports the performance of the proposed method in terms of localization accuracy, robustness to unsupervised annotation, generalization, sensitivity to landmark selection, temporal accuracy, and speed.

4.4.1 Data Description

The experimental data consists of two datasets acquired at the University of North Carolina Cancer Hospital. The first dataset consists of 25 patients with 349 images in total. The planning images in this dataset were scanned by a Siemens Somatom CT scanner with the field of view (FOV) 50 cm. The treatment images in this dataset were scanned by a Siemens Somatom CT-on-rails scanner with FOV 40 cm. The second dataset consists of 7 patients with 129 images in total. The planning images in the second dataset were acquired from a Philips Big bore scanner with FOV 60 cm. The treatment images in the seond dataset were acquired using the same machine and FOV as in the first dataset. For convenience the first and second datasets are respectively named as dataset A and dataset B in the rest of chapter. In both datasets every patient has one planning CT scan and $8 \sim 20$ treatment CT scans. The prostates in all CT images have been manually contoured by an experienced expert to serve as the gold standards. Table 4.1 lists other information of the two datasets, e.g., spacing and image size.

Table 4.1: Description of two CT Prostate datasets.

	Dataset A	Dataset B
Planning resolution (mm)	$0.98 \times 0.98 \times 3$	$1.24 \times 1.24 \times 3$
Treatment resolution (mm)	$0.98 \times 0.98 \times 3$	
Image size	$512 \times 512 \times 30 \sim 120$	
Number of patients	25	7
Number of images	349	129

4.4.2 Accuracy and Efficiency Requirement for IGRT

As indicated by an experienced clinician, a localization algorithm with average surface distance less than 3 mm, DSC greater than 0.80 and runtime less than 2 minutes would be acceptable for standard conventional radiation therapy. For stereotactic body radiation therapy (SBRT), which delivers much higher dose per fraction (800 cGy) than conventional radiation therapy, it is desirable to track the intra-fraction prostate motion during the radiation treatment to reduce the chances of missing the target. Thus, a localization algorithm with a higher efficiency is often required. According to a clinician whom I talked to in the North Carolina Cancer Hospital, in order to track the intra-fraction prostate motion in SBRT, the time for the entire prostate localization procedure should be kept within 1 min including the time for review and manual adjustment. Since the time for manual adjustment heavily depends on the segmentation quality, it is difficult to give a quantitative acceptable threshold for the algorithm speed. In principle, a faster algorithm would save more time for better quality control and in the meanwhile minimize the discomfort of the patients when they are fixed in the treatment bed.

4.4.3 Parameter and Experimental Setting

Three scales (coarse, middle, and fine) were used in the population-based learning. Table 4.2 lists the training parameters of landmark detection at different scales that will be elaborated in the following paragraphs.

Table 4.2: Training parameters for multi-scale landmark detection.

Scale	Spacing (mm)	W (mm)	d_n (mm)
Coarse	4	80	400
Middle	2	40	200
Fine	1	20	100

In the training of each cascade, positive training samples $X_{\mathcal{P}}$ were the voxels annotated as landmarks. The negative training sample set $X_{\mathcal{N}}$ consisted of all voxels whose distances are within d_n from the annotated landmarks. At every cascade level k, if $\|X_{\mathcal{P}}\| \,/\, \|X_{k-1} \cap X_{\mathcal{N}}\| < \tau$, a portion of the negatives was randomly sampled from $X_{k-1} \cap X_{\mathcal{N}}$ such that the positive/negative ratio is equal to τ (in this work, $\tau = 1/5$). Otherwise, if $\|X_{\mathcal{P}}\| \,/\, \|X_{k-1} \cap X_{\mathcal{N}}\| \geq \tau$, all samples in $X_{k-1} \cap X_{\mathcal{N}}$ were used as negative samples. τ was also used as a relative threshold for stopping cascade learning and forward learning when the false positive / positive ratio is less than τ. In this way, the positive/negative ratio was restricted between τ and $1/\tau$ at every cascade level, thus avoiding the problems caused by the imbalanced training dataset.

Each training voxel was represented by a set of extended Haar wavelet features [Zhan et al., 2011a] that were computed by convolving the Haar-like kernels with the intensity image. The Haar-like kernels were generated by scaling the predefined Haar-like templates. Each Haar-like template consists of one or more 3D rectangle functions with different polarities:

$$H(x) = \sum_{i=1}^{Z} b_i R(x - a_i) \tag{4.4}$$

$$R(x) = \begin{cases} 1 \quad , \quad \|x\|_\infty \leq 1 \\ 0 \quad , \quad \|x\|_\infty > 1 \end{cases} \tag{4.5}$$

where $Z \in \{1, 2\}$ is the number of 3D rectangle functions, and $b_i \in \{-1, 1\}$ and a_i are the polarity and translation of the i-th 3D rectangle function, respectively. Fig. 4.8 shows the 14 Haar-like templates used in this work. These templates were chosen in order to capture the intensities and intensity differences at different locations and directions within a local patch. The coefficients for scaling the Haar-like templates were 3 and 5. For each training voxel, the extended Haar wavelet features were computed in its $W \times W \times W$ local neighborhood. Then, all these computed features were concatenated to form a patch-based feature representation for the voxel. The training parameters are listed in table 4.2. In the cascade learning step, the Adaboost classifier was employed as cascade classifier. The training of Adaboost classifier stopped when 20 weak classifiers were obtained.

In the multi-atlas RANSAC, the minimum number of landmarks $\mathcal{M}$ required for transformation estimation was set to 3 since only a 3D rigid transformation needs to be estimated. The threshold η for determining the landmark agreement was set to 5 mm. In the remainder of this section, all results from ILSM were generated with the same parameter setting.

Five-fold cross validation was used to evaluate the proposed method on dataset A. Specifically, the population-based detectors of one fold were trained using CT scans from the other four folds. For each fold, about 250 CTs were used in the training of population-based detectors. For the experiments on dataset B, the population-based detectors were trained using all CT scans in dataset A. In this way, the generalization of the proposed method can be

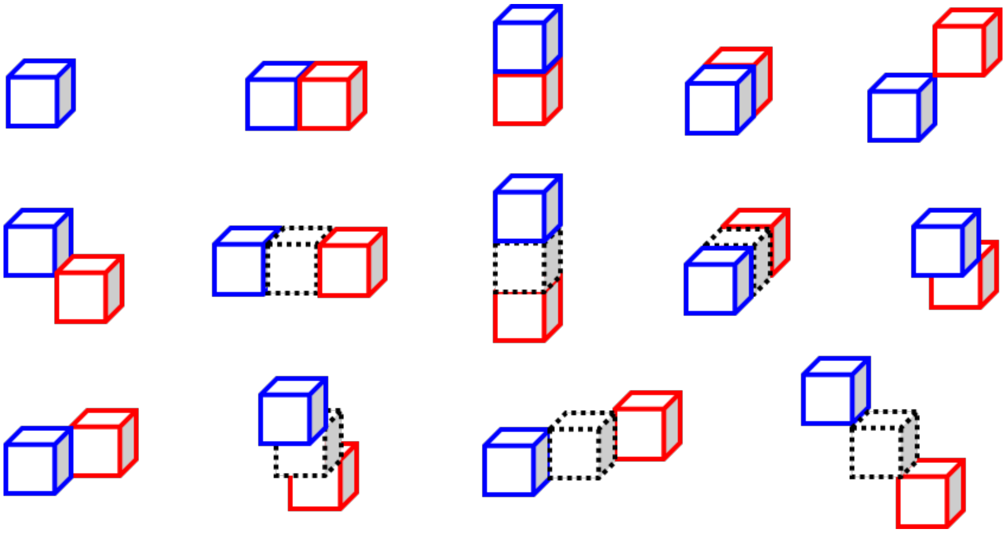


Figure 4.8: 14 Haar templates. Blue and red cubes are 3D rectangle functions with positive and negative polarities, respectively. Cubes with dashed borders are the empty areas which are shown only for the purpose of visualization.

validated by applying the detectors learned from dataset A to dataset B, which was acquired with a different scanner and protocol.

To emulate the real clinical setting, for prostate localization in treatment day $N + 1$, the previous N treatment images and the planning image were used as patient-specific training data (fig. 1.1). It was found from the experiments that, when N reached 4, there was negligible accuracy gained from performing additional ILSMs. Therefore, after treatment day 4, ILSM was not performed to further refine the patient-specific landmark detectors. Instead, the existing detectors were directly adopted for prostate localization. If not explicitly mentioned, all the reported performances of ILSM were computed using *up to* 5 patient-specific training images (4 treatment images + 1 planning image).

4.4.4 Number of Cascade Classifiers

To gain an insight on the number of cascade classifiers remaining after backward pruning or appended by forward learning, the statistics of K_{pop}, K_{bk} and K_{pat} are summarized in table 4.3. To recap, K_{pop} is the number of classifiers in the population-based cascade, K_{bk} is the number of classifiers after backward pruning, and K_{pat} is the number of classifiers appended by forward learning. As is seen, the majority of population cascade classifiers in the coarse and middle scale were retained after backward pruning (i.e., K_{bk} is close to K_{pop}). However, when it comes to the fine scale, many population cascade classifiers were discarded. The reason for this may be related to the fact that individual differences are embodied in the fine scale but not evident in the coarse and middle scales. Finally, for the number of patient-specific cascade classifiers appended in the forward learning stage, experimental results show that usually 2-3 classifiers are sufficient.

Table 4.3: Statistics of numbers of cascade classifiers.

Scale	K_{pop}	K_{bk}	K_{pat}
Coarse	13.1 ± 1.2	12.6 ± 3.4	2.1 ± 1.3
Middle	13.5 ± 0.6	12.7 ± 2.8	2.5 ± 1.6
Fine	15.5 ± 2.1	6.3 ± 2.8	2.9 ± 0.1

4.4.5 Comparison Studies

This subsection compares the proposed method with several alternative methods for prostate localization in treatment CT images. These methods include 1) four traditional learning-based schemes for anatomy detection, 2) single-atlas RANSAC using only the prostate shape from the planning image, 3) a traditional registration method based on pelvic bones, 4) several published methods developed on the same dataset, and 5) other published methods developed on different datasets.

Comparison with Traditional Learning-based Approaches

To illustrate the effectiveness of the proposed learning framework, ILSM was compared with four other learning-based approaches on dataset A. All of these methods localize the prostate through learning-based anatomy detection with the same features, classifiers and cascade framework (as described in section 4.1). Their differences lie in the training images and learning strategies that are shown in table 4.4. To increase the variability of limited patient-specific data, each patient-specific training image was rotated from -30 to 30 degrees with a step size 5 degrees.

Table 4.4: Differences between ILSM and four learning-based methods. (POP: population-based learning; PPAT: pure patient-specific learning; MIX: population and patient-specific mixture learning; IL: incremental learning without backward pruning; ILSM: proposed incremental learning with selective memory.)

		POP	PPAT	MIX	IL	ILSM
Training Images	Population	✓		✓	✓	✓
	Patient-specific		✓	✓	✓	✓
Learning Strategies	Cascade Learning	✓	✓	✓	✓	✓
	Backward Pruning					✓
	Forward Learning				✓	✓

Table 4.5 compares the four learning-based approaches with ILSM on landmark detection errors. To exclude the influence from other components, the reported landmark detection error was directly measured without using RANSAC for outlier detection and correction. It can be seen that ILSM outperformed the other four learning-based approaches on all seven anatomical landmarks. In order to better interpret the landmark detection accuracies of the proposed method, an experiment was further conducted to assess the inter-operator annotation variability on CT prostate landmarks. Specifically, four different operators were asked to independently annotate the seven antomical landmarks on 19 CT scans of one patient.

Then, the differences of annotated landmarks were calculated between any pair of operators. Finally, all pair-wise differences were averaged to obtain the inter-operator annotation variability (listed in table 4.6). From table 4.6, it can be seen that on average ILSM achieved comparable (if not better) accuracy to the inter-operator annotation variability, exhibiting better mean error but slightly worse standard deviation.

Table 4.5: Quantitative comparisons on landmark detection error (mm) between ILSM and four learning-based methods on dataset A. Landmark errors reported here are calculated without using RANSAC for outlier detection and correction. The last row shows the p-values of paired t-tests when comparing landmark errors of four learning-based methods with that of ILSM.

	POP	PPAT	MIX	IL	ILSM
PC	6.69 ± 3.65	4.89 ± 5.64	6.03 ± 3.03	5.87 ± 4.01	$\mathbf{4.73 \pm 2.69}$
RT	7.85 ± 8.44	6.09 ± 9.00	5.72 ± 4.04	6.33 ± 4.82	$\mathbf{3.76 \pm 2.80}$
LF	6.89 ± 4.63	5.39 ± 7.61	5.61 ± 3.63	5.90 ± 4.54	$\mathbf{3.69 \pm 2.69}$
PT	7.04 ± 5.04	8.66 ± 13.75	6.18 ± 4.76	6.74 ± 5.05	$\mathbf{4.78 \pm 4.90}$
AT	6.60 ± 4.97	4.54 ± 5.06	5.38 ± 4.55	5.68 ± 4.97	$\mathbf{3.54 \pm 2.19}$
BS	6.12 ± 2.97	5.63 ± 7.44	6.63 ± 3.98	5.61 ± 2.94	$\mathbf{4.68 \pm 2.71}$
AP	10.42 ± 6.03	8.94 ± 16.07	8.77 ± 5.00	9.50 ± 7.17	$\mathbf{6.28 \pm 4.60}$
Average	7.37 ± 5.52	6.31 ± 10.13	6.33 ± 4.32	6.52 ± 5.09	$\mathbf{4.49 \pm 3.49}$
p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	n/a

Table 4.6: Quantitative comparison between landmark detection error (mm) of ILSM and inter-rater annotation variability on 19 treatment scans of one patient. Landmark errors reported here are calculated without using RANSAC for outlier detection and correction. The p-value reported here is computed by a paired t-test.

	PC	RT	LF	PT	AT	BS	AP	Average	p-value
ILSM	4.72 ± 1.42	3.03 ± 1.75	3.17 ± 1.61	2.45 ± 1.00	3.24 ± 1.28	5.57 ± 1.98	7.18 ± 4.17	4.20 ± 2.65	n/a
Inter-rater	4.50 ± 1.22	5.25 ± 1.27	5.77 ± 1.49	5.71 ± 2.85	4.44 ± 3.09	4.63 ± 1.32	4.44 ± 1.05	4.96 ± 2.00	0.01

Table 4.7 compares the four learning-based approaches with ILSM on overlap ratios (DSC). To exclude the influence of multi-atlas RANSAC, only a single shape atlas (i.e., the

planning prostate shape) was used for localization. Here, “Acceptance” denotes acceptance rate. It is the percentage of images where an algorithm performs with a higher accuracy than inter-operator variability (DSC = 0.81) [Foskey et al., 2005]. According to an experienced clinician, these results can be accepted without manual editing. It can be seen that ILSM achieved the best localization accuracy among all methods. Not surprisingly, by utilizing patient-specific information, all three methods (i.e., PPAT, MIX and IL) outperformed POP. However, their performances were still inferior to ILSM, which shows the effectiveness of ILSM in combining both population and patient-specific characteristics.

Table 4.7: Quantitative comparisons on overlap ratios (DSC) between ILSM and four learning-based methods in dataset A. (S) and (M) indicate single-atlas and multi-atlas RANSAC, respectively. The p-values in the last row are calculated between four learning-based methods and ILSM by paired t-tests.

	POP (S)	PPAT (S)	MIX (S)	IL (S)	ILSM (S)	ILSM (M)
Mean DSC	0.81 ±0.10	0.84 ±0.15	0.83 ±0.09	0.83 ±0.09	0.87 ±0.06	0.88 ±0.06
Acceptance	66%	85%	74%	77%	90%	91%
p-value	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	n/a

Fig. 4.9 shows the differences in localization accuracy between ILSM and PPAT with respect to the number of patient-specific training images. It can be seen that when the number of patient-specific training images was limited (< 3), the performance of PPAT was very poor even with artificial transformations (e.g., rotation) to increase the variability in training samples. This was especially the case when only one patient-specific training image was used. Due to the limited patient-specific patterns observed, PPAT suffered from severe overfitting and resulted in high failure rates for some patients. In such cases, ILSM significantly outperformed PPAT by $40\% - 70\%$ DSC as shown in fig. 4.9(a). The main reason why simple artificial transformations (e.g., rotation) failed to improve the performance of PPAT is that generally they cannot well capture intra-patient anatomical appearance vari-

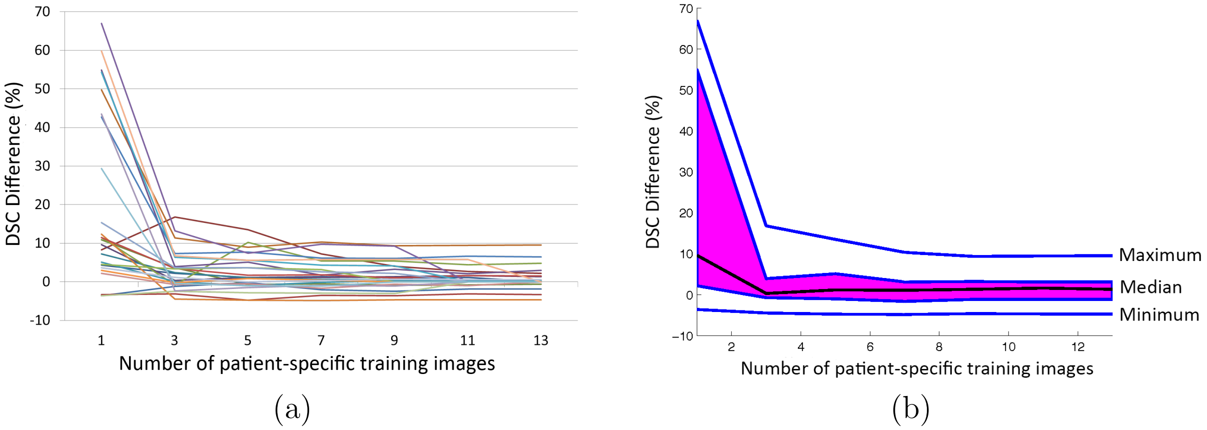


Figure 4.9: The function boxplot [Sun and Genton, 2011] of DSC difference curves between ILSM and PPAT for convergence analysis on dataset A. Each DSC difference curve is a function of DSC difference between ILSM and PPAT with respect to the number of patient-specific training images. (a) shows 25 DSC difference curves each of which corresponds to one patient. (b) shows the function boxplot of 25 curves in (a). The black curve in (b) corresponds to the median curve, the magenta area covers the central 50% of the curves, and two outmost blue curves are the extreme maximum and minimum curve, respectively.

ations such as bowel gas and filling. This also explains why previous *pure* patient-specific learning algorithms [Li et al., 2012, Liao et al., 2013, Gao et al., 2012a] often start with three patient-specific training images. By leveraging both population and patient-specific data, ILSM can achieve DSC 0.85 ± 0.06 on the first two treatment images using only a single planning CT as patient-specific training data while in the same setting PPAT only obtained DSC 0.79 ± 0.15. As the number of patient-specific training images increased, the performance difference between ILSM and PPAT gradually decreased. Ideally when sufficient patient-specific data is collected, the performance of ILSM and PPAT should converge. However, by using up to 13 patient-specific training images, ILSM was still slightly better than PPAT (1.5% DSC difference), which implies the effectiveness of the general appearance characteristics learned from the population.

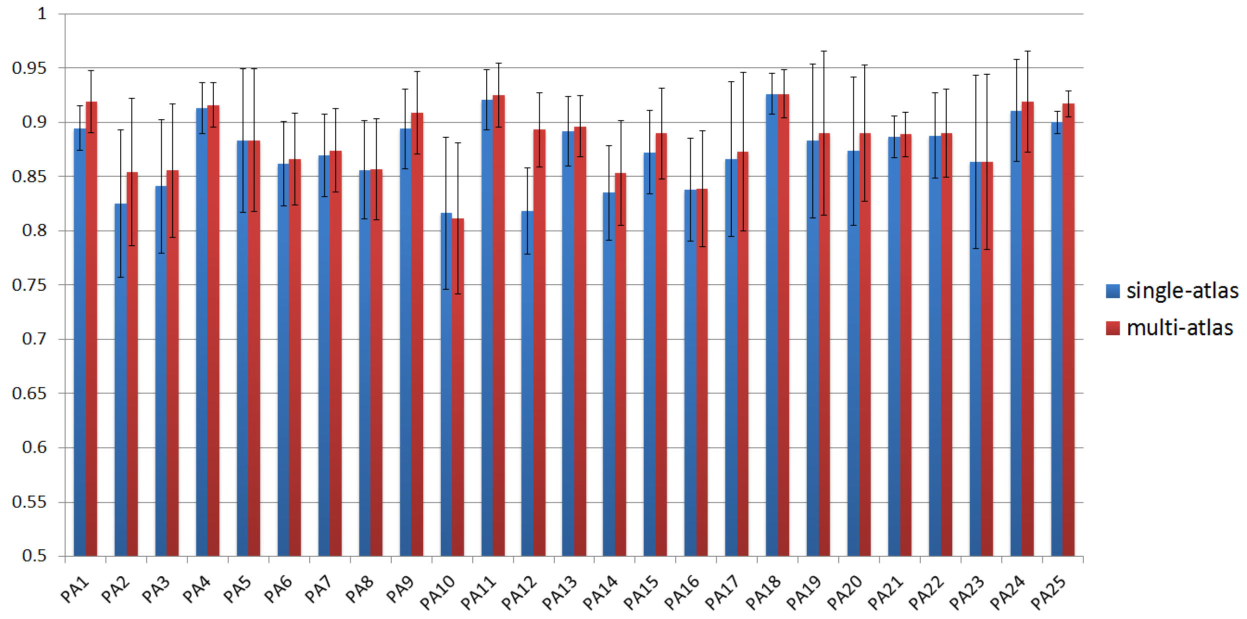


Figure 4.10: Comparison between single-atlas and multi-atlas RANSAC on dataset A using overlap ratio (DSC).

Comparison with Single-atlas RANSAC

Fig. 4.10 shows the average DSCs of all 25 patients in dataset A with single-atlas RANSAC and multi-atlas RANSAC. For single-atlas RANSAC, the planning prostate shape was used as the shape atlas. For multi-atlas RANSAC, shape atlases consisted of not only the planning prostate shape but also previously segmented prostate shapes of the patient. It can be seen that in almost all patients multi-atlas RANSAC achieved better localization accuracy than single-atlas RANSAC. Table 4.7 also compares single-atlas and multi-atlas RANSAC on average DSC and acceptance rate. It shows that the localization accuracy of ILSM can be further boosted by using multi-atlas RANSAC (1% improvement on both average DSC and acceptance rate).

Comparison with Traditional Bone Alignment

Bone alignment is usually adopted as a standard preprocessing step in many prostate localization methods [Foskey et al., 2005, Li et al., 2012, Liao et al., 2013, Gao et al., 2012a]. The basic idea is to register the current treatment CT scan with the previous one of the same patient by aligning the pelvic bones. The prostate mask in the previous CT can thereby be transformed to the current treatment CT. In the bone alignment, the pelvic bones in two CT scans are first segmented by thresholding. Based on the segmented binary bone images, the optimal rigid transformation is estimated and used to align two scans. Since the prostate is very close to the pelvic bone, bone alignment often achieves satisfactory overlap ratios of the prostate. For a fair comparison, the same multi-atlas scheme was adopted as described in section 4.3 to evaluate the performance gain of the proposed method over bone alignment. The FLIRT toolkit [Fischer and Modersitzki, 2003] was used for bone alignment as in the previous methods [Foskey et al., 2005, Li et al., 2012, Liao et al., 2013, Gao et al., 2012a]. Fig. 4.11 visually shows the overlapping degree of the prostate after bone alignment for 12 typical patients. The DSC obtained by bone alignment on this dataset is 0.78 ± 0.12, which is significantly lower than the DSC achieved by the proposed method (0.89 ± 0.06). In addition, bone alignment takes more computational time than the proposed method. To align two CT scans of image size $512 \times 512 \times 60$, bone alignment typically takes 5 minutes while the proposed method takes only 4 seconds on the same image size.

To consider local intensity information around the prostate in the alignment procedure, an experiment was further conducted to compare a local intensity-based rigid registration method with the proposed method. In the former method, bone alignment was first performed to align a previous CT scan with the current treatment CT based on the pelvic bone.

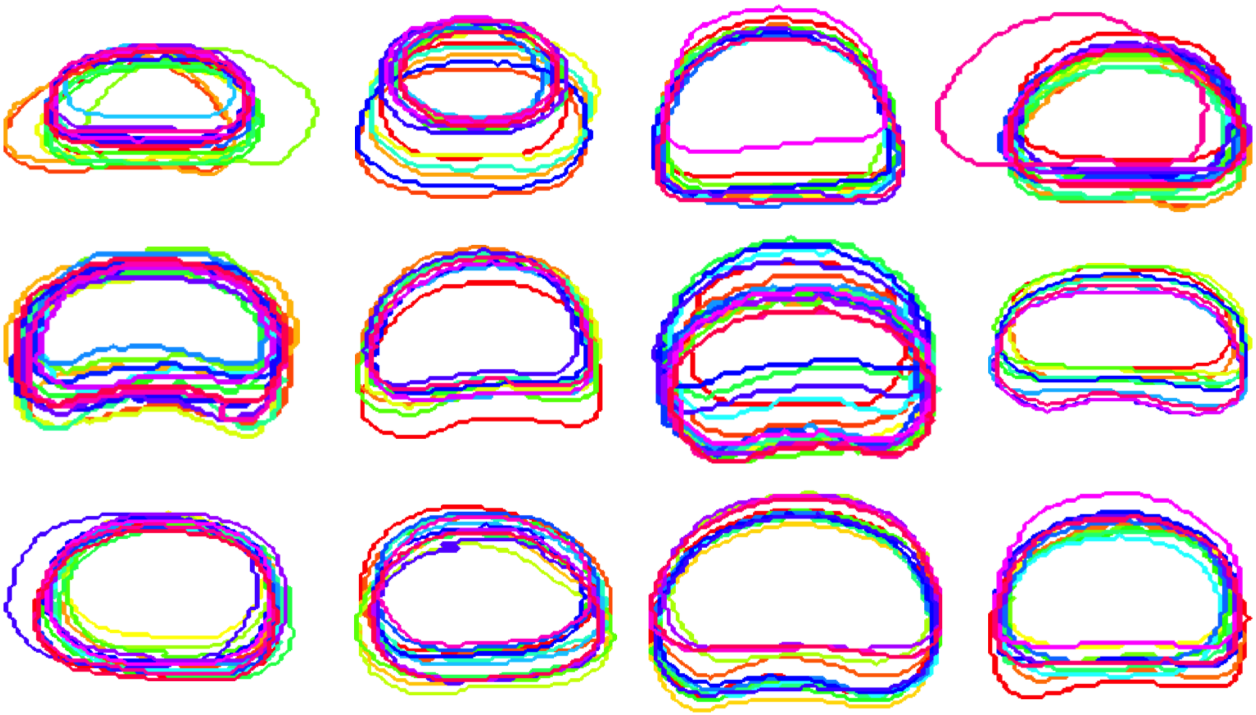


Figure 4.11: Prostate contours of treatment CTs from 12 typical patients after bone alignment. The contours in the figure are from the middle transversal slices of the prostates after bone alignment.

Then, a tight bounding box was determined using the prostate mask of the previous CT scan. Based on the determined bounding box, the two CT scans were further registered using an intensity-based rigid registration method as implemented in FLIRT by using correlation ratio as the cost function. Finally, given the estimated rigid transformation, the prostate mask in the previous CT scan was transformed onto the current treatment CT for localization. Following the same multi-atlas scheme (section 4.3), it was found that compared to bone alignment local intensity-based rigid registration improved the localization accuracy from mean DSC 0.78 to 0.80. However, the standard deviation of DSC also increased from 0.12 to 0.14 due to some failure cases caused by the bad initialization of bone alignment. In contrast, the proposed method achieved much higher accuracy (0.89 ± 0.06) with faster localization speed (4 seconds).

Comparison with CT prostate localization methods on the same dataset

The proposed method can achieve localization accuracy at DSC 0.89 ± 0.06 and average surface distance 1.72 ± 1.00 mm on 446 treatment CT scans of 32 patients. Table 4.8 quantatively compares the performance of the method with five other state-of-the-art methods on the same dataset. These methods respectively employ a deformable model [Feng et al., 2009], registration [Liao and Shen, 2012], multi-atlas based segmentation [Liao et al., 2013] and classification [Li et al., 2012, Gao et al., 2012a] to localize the prostate on treatment CTs. Because the quantitative measurements of compared methods were not available for each testing image, the statistical significances of differences between compared methods and the proposed method are not listed in table 4.8.

It can be seen from table 4.8 that the proposed method achieves comparable accuracy to the state-of-the-art methods while substantially reducing the localization time to just 4 seconds. This fast localization speed helps overcome the limitation of previous localization methods: if the prostate unexpectedly moves during the long localization procedure, their method has to be performed again. It is also worth noting that previous methods [Li et al., 2012, Liao et al., 2013, Gao et al., 2012a] require at least three patient-specific training images for initialization due to the nature of pure patient-specific learning, which indicates that such methods cannot be adopted to segment the first two treatment CTs. By effectively combining both population and patient-specific information, even with only one planning CT, the proposed method can still achieve reasonably accurate localization results on the first two treatment CTs (DSC 0.85 ± 0.06).

Table 4.8: Quantitative comparison with other CT prostate localization methods on the same dataset (DSC: Dice similarity coefficient, ASD: Average surface distance, Sen.: Sensitivity, PPV.: Positive predictive value).

Method	Deformable	Registration	Multi-atlas	Classification		ILSM
	Feng et al.	Liao et al.	Liao et al.	Li et al.	Gao et al.	
Automaticity	Fully	Fully	Fully	Fully	Semi	Fully
Mean DSC	0.89	0.90	0.91	0.91	0.91	0.89
Mean ASD	2.08	1.08	0.97	1.40	1.24	1.72
Median Sen.	n/a	0.89	0.90	0.90	0.92	0.89
Median PPV.	n/a	0.89	0.92	0.90	0.92	0.92
Speed (sec.)	96	228	156	180	600	4

Comparison with CT prostate localization methods on the different datasets

Table 4.9 lists the performance of other CT prostate localization methods for reference. Due to the fact that neither their data nor the source codes of these methods are publicly available, only the numbers reported in their publications are cited. Based on the reported numbers, it can be seen that the proposed method has been evaluated on the largest dataset and achieved the best localization accuracy.

Table 4.9: Comparison with other CT prostate localization methods on different datasets for reference (DSC: Dice Similarity Coefficient, Sen.: Sensitivity, PPV.: Positive Predictive Value).

Method	Deformable Models		Registration	ILSM
	Costa et al.	Chen et al.	Foskey et al.	
image #	16	185	65	446
patient #	n/a	13	5	32
Mean DSC	n/a	n/a	0.84	0.89
Median Sen.	0.79	0.84	n/a	0.89
Median PPV.	0.86	0.87	n/a	0.92
Speed (sec.)	n/a	60	750	4

4.4.6 Algorithm Performance

This subsection reports the performance of the proposed algorithm in terms of localization accuracy, robustness to unsupervised annotation, generalization, sensitivity to landmark selection, temporal accuracy, and speed.

Localization Accuracy

Table 4.10 shows the localization accuracy of the proposed method on dataset A and dataset B. It can be seen that the proposed method is able to achieve more consistent and accurate localizations (DSC 0.89 ± 0.06) than inter-operator variability (DSC 0.81 ± 0.06) [Foskey et al., 2005]. This indicates that the proposed method in fact well satisfies the accuracy requirement of IGRT and can be adopted in the clinical setting. To assess the worst and optimal accuracy of the proposed method, two further experiments were conducted. The first experiment detected only one anatomical landmark (prostate center) and used only one shape atlas (planning prostate shape) for localization. The performance in this setting is regarded as the worst performance that the proposed method can get. The second experiment localized the prostate using manually annotated landmarks and multiple shape atlases (prostate shapes in planning and previous treatment images). This accuracy indicates the optimal performance of the proposed method. Table 4.11 lists the worst and optimal accuracy on different quantitative measures. It should be noted that the only difference between the optimal accuracy (shown in table 4.11) and the reported accuracy of the proposed method (shown in table 4.10) is in the landmark localization. The optimal accuracy was calculated using the manually annotated landmarks, and the performance of the proposed method was obtained using automatically detected landmarks. By comparing the two, it can be seen that the performance of the proposed method is quite close to the optimality, which indicates that accurate landmark detection results can be achieved by using ILSM. On the other hand, by using only one anatomical landmark and a single shape atlas, the localization accuracy is still comparable to the inter-operator variability (DSC 0.81 ± 0.06), which shows the effectiveness of ILSM in CT prostate localization.

Table 4.10: Localization accuracy of ILSM on two datasets (DSC: Dice Similarity Coefficient, ASD: Average Surface Distance, Sen: Sensitivity, PPV: Positive Predictive Value).

	DSC	ASD (mm)	Sen.	PPV.
Dataset A	0.88 ± 0.06	1.89 ± 0.98	0.87 ± 0.06	0.89 ± 0.06
Dataset B	0.91 ± 0.05	1.27 ± 0.90	0.88 ± 0.05	0.93 ± 0.06
All	0.89 ± 0.06	1.72 ± 1.00	0.88 ± 0.06	0.90 ± 0.06

Table 4.11: Worst and optimal accuracy of ILSM in CT prostate localization. The reported values are calculated on both dataset A and dataset B.

	DSC	ASD (mm)	Sen.	PPV.
Worst	0.92 ± 0.03	1.00 ± 0.60	0.91 ± 0.05	0.94 ± 0.03
Optimal	0.81 ± 0.09	3.01 ± 2.01	0.80 ± 0.10	0.83 ± 0.10

Robustness to unsupervised annotation

As shown in fig. 4.2, in order to incorporate patient-specific characteristics, ILSM requires annotations in planning and previous treatment images. Annotations in planning images are always provided by physicians. Afterwards, there are two ways to obtain annotations in treatment images. 1) *Supervised annotation.* In this scenario, detectors trained by planning and previous treatment images are applied to localize the landmarks in current treatment images. The detection results need to be reviewed and corrected by physicians before being used to train detectors for the next treatment days. 2) *Unsupervised annotation.* The auto-detected results are considered as ground truth and used to train detectors for the next treatment days without manual review/corrections. Although the first scenario guarantees all training data are correctly annotated, the second scenario has the advantage of less manual operations (i.e., no manual operation, except the annotation in the planning CT) as long as the uncorrected annotation errors do not significantly degrade the localization accuracy.

ILSM was validated in both scenarios on dataset A. To simulate the supervised annotation, the manually annotated landmarks were directly used as the corrected landmarks for training. Compared with the average DSC of $0.88 \pm 0.06\%$ achieved using supervised annotation, the proposed method can achieve average DSC $0.85 \pm 0.06\%$ using unsupervised annotation. This is still more accurate than the inter-operator variability ($0.81 \pm 0.06\%$). Therefore, if some specific IGRT workflows require very few manual operations, the proposed method can be employed in the unsupervised annotation mode yet still with acceptable accuracy.

It is worth noting that compared with previous methods [Li et al., 2012, Liao et al., 2013, Gao et al., 2012a] that require precise manual segmentation of the entire prostate in the training treatment images, the proposed method requires only the annotations of at most seven anatomical landmarks, which dramatically reduces physicians' efforts on manual annotation. To be precise, the annotation time of an experienced radiation oncologist was recorded on the 19 treatment scans of one patient. It takes 11.7 ± 2.5 minutes to manually segment the entire prostate while it takes only 1.2 ± 0.3 minutes to annotate seven anatomical landmarks. If the proposed method is used to automatically detect the seven landmarks and radiation oncologists are only asked to verify and edit the detected landmarks, the landmark annotation time can be further reduced to 8.3 ± 1.3 seconds.

Generalization

A learning-based algorithm has to have good generalization in order to be applied on data from various institutions and scanners. To evaluate the generalization of ILSM, the localization algorithm was tested on dataset B, which was acquired under a different scanner from that of dataset A. All CT scans of dataset A (349 scans) were used to train the population-based landmark detectors for dataset B. The localization accuracy is shown in table 4.10, which indicates the good generalization of the proposed method. This is mainly due to the "selective memory" nature of ILSM. Even when the population landmark detectors are trained using a dataset with slightly different scanning protocols, after "personalization" by ILSM, the portion of the population-based appearance knowledge that is not in accordance with the current patient-specific characteristics will be discarded. By preserving only the applicable knowledge learned from population data, the generalization of the learned detectors is improved. Table 4.10 summarizes the overall performance of the proposed method on total 32 patients.

Sensitivity to Landmark Selection

To assess the sensitivity of the proposed method to landmark selection, the performance of the proposed algorithm was tested by alternately excluding one of the seven landmarks. Table 4.12 lists the DSCs of the proposed method on dataset A by excluding any of the seven landmarks. Overall, the performance is quite consistent no matter which subset of six landmarks is picked. Further, it is surprising to see that by excluding any of the landmarks in {PC,BS,AP}, the localization accuracy can actually be increased compared to the performance of DSC 0.88 ± 0.06 obtained by using all landmarks. The reason for this can be

Table 4.12: Sensitivity to landmark selection as measured by DSC. The table below shows the localization accuracies of six landmarks by excluding any of the seven landmarks used in the paper. The reported values are computed on dataset A.

Excluded	PC	RT	LF	PT
DSC	0.89 ± 0.05	0.88 ± 0.06	0.88 ± 0.06	0.88 ± 0.06
Excluded	AT	BS	AP	
DSC	0.88 ± 0.05	0.89 ± 0.05	0.89 ± 0.05	

inferred from table 4.5. That is compared with other landmarks the landmark detections of PC, BS and AP are less accurate due to the indistinct image appearance in those regions. Therefore, removing any of them helps improve the overall localization accuracy.

Temporal Analysis of Localization Accuracy

Fig. 4.12 shows the localization accuracy curve with respect to the number of patient-specific training images used. Not surprisingly, the localization accuracy of the proposed method increases as more patient-specific training data (i.e., image and shape) is available. The most significant improvement happens when the number of patient-specific training images increases from 1 to 3. As the number of patient-specific training images increases to 5, the localization accuracy levels off, which indicates that after the 4-th treatment day the patient-specific landmark detectors are sufficiently accurate, and thus there is no need to do additional incremental learning. That is, the existing landmark detectors can be directly applied to localize the prostate in the future treatment images. Considering that the period of standard radiation treatment typically takes 35 days, the ILSM procedure is only needed in the first 4 treatment days (about $4/35 \approx 11\%$ fraction of the entire treatment course).

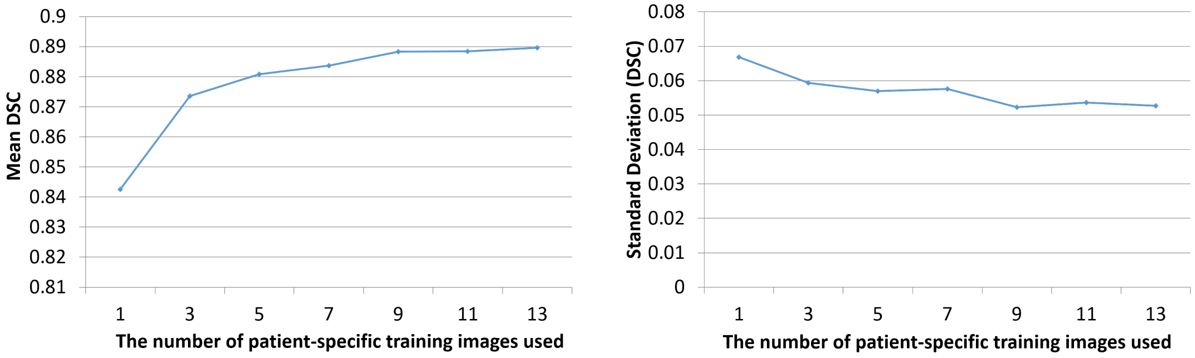


Figure 4.12: Temporal analysis of localization accuracy on dataset A.

Speed

The typical runtime for the proposed method to localize the prostate is around 4 seconds (on an Intel Q6600 2.4GHz desktop with 4 GB memory), which is almost real-time compared to previous methods. Thanks to incremental learning, the training time is reduced from 3-4 hours (traditional population-based training) to 30 minutes per landmark detector. Each landmark detector is independent and thus can be trained in parallel. It is also worth noting that the incremental learning process can be completed overnight before the treatment day. Therefore, the learning step does not take any additional time when the patients are receiving treatment.

4.4.7 Experiment Summary

In summary, the experiments show the following:

- Compared to traditional learning schemes, ILSM shows better landmark detection and prostate localization accuracy.

- Compared to other state-of-the-art methods, 1) the proposed method can achieve comparable accuracy with much faster speed; 2) the proposed method can be applied onto any treatment day of radiotherapy since it is still reasonably accurate (DSC ~ 0.85) even with only one patient-specific training image (i.e., the planning CT); 3) the proposed method requires only annotations of seven anatomical landmarks, thus significantly reducing physicians' manual efforts (from 11 mins to 1 min).

- Validated on 446 treatment CTs, average DSC 0.89 ± 0.06 was achieved in 4 seconds, which indicates that the proposed method is well-suited for the accuracy and speed requirements of IGRT.

4.5 Summary

Different from planning-CT segmentation, the IGRT workflow provides extra patient-specific data that could be utilized in treatment-CT segmentation. However, as the patient-specific data is very limited, conventional learning-based approaches are prone to overfitting. To address this issue, a novel learning scheme, namely incremental learning with selective memory (ILSM), was proposed in this chapter. By leveraging the large amount of population data and the limited amount of patient-specific data, ILSM takes both generalization and specificity into account when learning discriminant anatomy detectors. The learned detectors can accurately and efficiently localize the anatomical landmarks in new treatment CT images. After the landmarks are localized, a multi-atlas RANSAC algorithm was proposed to align the prostate shapes from the previous scans of the patient to the current treatment image for robust prostate localization. Validated on a large dataset (446 CT scans), ILSM shows comparable accuracy (DSC 0.89 ± 0.06) to the state-of-the-art methods while significantly reducing the runtime to 4 seconds. Moreover, in comparisons with traditional learning-based schemes (e.g., population learning, pure patient-specific learning, and mixture learning with population and patient-specific data), ILSM shows better capability to capture patient-specific appearance characteristics from limited patient-specific data.

CHAPTER 5 : SUMMARY, DISCUSSION AND FUTURE WORK

5.1 Summary

Automatic segmentation of pelvic CT images plays an important role in image guided radiotherapy of prostate cancer. It saves radiation oncologists' time for manual contouring and also improves the contouring consistency compared to human experts. In the planning stage, the automatic segmentation can be used for treatment planning. In particular, it provides the target region (prostate) where the prescribed dose should be delivered, and it also provides the regions of nearby healthy tissues where the radiation should be avoided. In the treatment stage, the daily segmentation of the prostate captures the current position of the prostate. It can be used to transform the treatment plan from the planning image space to the treatment image space for daily radiotherapy. Therefore, the accuracy of automatic segmentation is critical for the efficacy of radiation treatment.

The difficulty of planning-CT segmentation comes from low contrast and dramatic appearance and shape variations of pelvic organs across subjects. Deformable models (e.g., the active shape model) are often used to segment pelvic organs from CT images, because they effectively combine low-level appearance cues with high-level shape information in the segmentation. However, deformable models are sensitive to initialization and not flexible to segment organs with tubular shapes (e.g., the rectum). These drawbacks limit their performance in the planning-CT segmentation. To overcome these limitations, chapter 3 proposed explicitly learning a deformation field for guiding deformable segmentation. Two techniques were proposed to robustly and accurately estimate the deformation fields.

- An auto-context model was proposed to iteratively refine the deformation field by taking into account the predicted deformations in a spatial neighborhood. It was shown that the auto-context model captures the structured information, which is helpful to suppress prediction noise and improve spatial consistency of estimated deformation fields.

- A multitask random forest was proposed to estimate the deformation field jointly with the organ classification map. Through joint learning of deformation regression and organ classification in a single random forest, the multitask random forest improves the robustness of deformation field estimation by exploiting information from organ classification.

The proposed method was validated on a large dataset of pelvic CT images. The extensive experiments showed that 1) the auto-context model improves the accuracy of deformation field estimation; 2) the multitask random forest is better than the standard regression forest for deformation field estimation; 3) the regression-based deformable models are insensitive to initialization and flexible to segment tubular organs, and thus better suited for planning-CT segmentation than conventional deformable models.

While the proposed method for planning-CT segmentation can be directly used to localize the prostate in daily treatment images, its accuracy would be limited due to the neglect of patient-specific data available in the planning and previous treatment days. To fully exploit the limited patient-specific data, chapter 4 proposed an incremental learning algorithm, namely incremental learning with selective memory (ILSM), to update the population-learned anatomy detectors. The anatomy detectors are learned by the conventional cascade learning framework and used to detect distinctive prostate landmarks in the treatment images. Afterwards, the detected landmarks can be used to localize the prostate by aligning the segmented prostates from the planning and previous treatment images onto the current treatment image. In ILSM two steps were used to update the anatomy detectors. The backward pruning step discards the obsolete information from population-learned cascade classifiers that are incompatible with the patient-specific data. The forward learning step learns additional patient-specific cascade classifiers that fit the personalized characteristics. After updating the cascade classifiers, the anatomy detectors consist of both general information from a population and specific information from the patient data.

Extensive experiments showed that 1) ILSM is better suited than other learning schemes for anatomical landmark detection in the treatment images; 2) the localization accuracy is improved by using multiple segmented prostate shapes from previous days; 3) ILSM is able to achieve competitive localization accuracy to the state-of-the-art methods, but with significant speedup.

The contributions of this dissertation are as follows:

1) *A novel deformable model, namely a regression-based deformable model, is proposed to hierarchically deform a shape model onto the target organ boundary based on an explicitly learned deformation field.*

The regression-based deformable model (RDM) was presented in section 3.3. In RDM a deformation field is predicted from image data and used to guide deformable segmentation. The deformation field in RDM is different from the one used in pair-wise image registration. In RDM each voxel in the deformation field is associated with a displacement vector that points from this voxel to the nearest voxel on the organ boundary. The displacement vector tells the distance and direction from a voxel to the organ boundary, and it can be used to guide the deformation of each vertex on the shape model toward the organ boundary. To increase the robustness of regression-based deformation, a hierarchical deformation strategy was proposed to gradually relax the freedom of deformation as the shape model approaches the organ boundary. The experiments in sections 3.5.6 and 3.5.7 showed that 1) RDM is less sensitive to initialization than conventional deformable models; It can achieve accurate segmentation results of CT pelvic organs by initializing the mean shape in the image center; 2) RDM is more flexible in segmenting tubular organs (e.g., the rectum) than several other methods under comparison.

2) *An auto-context model is adopted to iteratively refine the predicted deformation field by gradually incorporating the neighborhood prediction information.*

The conventional voxel-wise prediction doesn't take into account the correlations of deformations in the spatial neighborhood. As a result, the predicted deformation field is often noisy and spatially inconsistent. To overcome this problem, section 3.1 proposed using an auto-context model to iteratively predict the deformation field. The auto-context model is built upon the conventional voxel-wise prediction. It adds additional iterations to refine the deformation field. In the first iteration, the auto-context model predicts the deformation of each voxel independently based on image appearance, which is the same with the conventional voxel-wise prediction. In the later iterations, the auto-context model learns additional deformation regressors by extracting not only appearance features from the target image but also context features from the intermediate deformation field predicted in the previous iteration. Since context features capture the deformation information in the neighborhood, the auto-context model is able to utilize this information to reduce prediction noise and improve the spatial consistency of the predicted deformation field. Experiments in section 3.5.3 showed that the auto-context model leads to a big improvement in segmentation accuracy compared to conventional voxel-wise prediction.

3) *A multitask random forest is proposed to learn the deformation from local image appearance by coupling deformation regression and organ classification in a common random forest.*

The multi-task random forest was presented in section 3.2. It shares a similar idea with many multitask learning algorithms. Learning multiple related tasks using the same representation/model tends to improve the generalization of both tasks. In this dissertation the multi-task random forest combines the learning of a deformation regressor and an organ classifier in a single random forest. Embedding such a multi-task random forest in the auto-context model allows the prediction information from deformation regression and organ classification to be exchanged during the iterative refinement. Experimental results in sections 3.5.4 and 3.5.5 showed that the multi-task random forest improves both classification and regression performance compared with standard classification and regression forests.

4) *A multi-resolution strategy is adopted to segment multiple pelvic organs from CT images, where the coarse-level deformation fields are jointly estimated for all organs to consider their spatial relationship and where the fine-level deformation fields are separately estimated for each organ to make the respective prediction models specific.*

The multi-resolution strategy for deformation field estimation was presented in section 3.4. The prediction models at different resolutions are trained with different tradeoffs between accuracy and robustness. In the coarse resolutions the robustness is of the first priority and the accuracy is secondary. The robustness of the prediction model is improved from two aspects: 1) the training samples are drawn from the entire image domain and 2) deformation regressors of different organs are jointly trained to exploit the inter-organ spatial relationship. In the fine resolutions, the accuracy of the prediction model is critical. Since deformable models are assumed to be close to target organ boundaries in the fine resolutions, the accuracy can be improved by 1) taking training samples only around the target organ boundaries and 2) training one prediction model for each organ. Although the multi-resolution strategy was not quantitatively evaluated in the experiments, empirically I observed that the multi-resolution strategy improves the robustness and accuracy of regression-based deformable segmentation. Besides, the efficiency of RDM is also improved by jointly estimating the deformation fields of different organs in the coarse resolutions.

5) *Extensive experiments on a large prostate CT dataset ($>$ 300 patients) show that the proposed method can accurately segment the prostate, bladder, rectum and two femoral heads from planning CT images and that it outperforms many existing methods in this task.*

To the best of my knowledge, this is the largest dataset ever reported in the literature for evaluating automatic segmentation methods for CT pelvic organs. In the experiments I compared the multi-task random forest with the standard random forest (e.g., classification and regression forests) for deformation field estimation. I showed that joint regression and classification can improve the organ classification and deformation regression and that it can also overcome the missing boundary problem suffered by the standard regression forest. Besides, I also compared regression-based deformable models with a variant of the active shape model with different initialization strategies. The results showed that learning a deformation field to guide deformable segmentation can largely reduce the sensitivity of deformable models to initialization and improve the accuracy for segmenting organs with complex shapes, such as the rectum. The segmentation accuracy evaluated over CT scans from 300 patients showed that the proposed method can achieve quite accurate results for CT pelvic organ segmentation.

6) *The cascade learning framework is adapted to address the problem of unbalanced training samples in the classification-based landmark detection. It can efficiently localize a landmark in a 3D medical image volume within one second using a multi-resolution implementation.*

The cascade learning was presented in section 4.1. The idea of cascade learning was originally proposed in face detection in natural images. In this dissertation I borrowed the idea to address the unbalanced training problem existing in the classification-based landmark detection. Specifically, in the classification-based approach positive training samples are voxels near the annotated landmarks in the training images, and negative training samples are voxels elsewhere. To separate the limited positive samples with unlimited negative samples, a cascade of classifiers are sequentially learned in a boosting manner to gradually filter out the negative samples from positive samples. With an extension to 3D Haar-like features and a multi-resolution implementation, I showed that the cascade learning can be used to accurately and efficiently detect anatomical landmarks in 3D medical images.

7) *An incremental learning scheme, namely incremental learning with selective memory, is proposed to update the existing landmark detector learned from massive population data with limited patient-specific data. It can be used to personalize the population-based landmark detectors to a specific patient.*

Incremental learning with selective memory (ILSM) was presented in section 4.2. It can personalize a population-based landmark detector into a patient-specific one by selectively discarding population-based classifiers that are inapplicable to the patient and meanwhile incrementally learning patient-specific classifiers that fit the appearance characteristics of CT scans of the patient. The difference between conventional incremental learning algorithms and ILSM is that ILSM introduces a backward pruning step to discard classifiers that are inapplicable to the patient. As validated in section 4.4.5, this backward pruning step greatly improves the performance of landmark detection and segmentation compared with conventional incremental learning without backward pruning. ILSM provides a new way to combine limited patient-specific data with massive population data. Compared to a direct combination of both data, ILSM offers two advantages: 1) ILSM avoids re-training from scratch, thereby reducing the training time; 2) ILSM avoids the limited patient-specific data being overwhelmed in the massive population data; hence it can better capture the patient-specific appearance characteristics.

8) *A schematic illustration is provided to explain the mechanism behind incremental learning with selective memory.*

In section 4.2.5 a schematic illustration was given for understanding the incremental learning with selective memory (ILSM). The illustration showed how ILSM changes the decision boundaries of population-based cascade classifiers in order to accomodate the limited patient-specific data. The illustration also compared ILSM with two other learning schemes, pure patient-specific learning and conventional incremental learning, in the way that decision boundaries are formed during the learning process. Different from pure patient-specific learning (PPAT), which generates decision boundaries tightly around the limited patient-specific data, ILSM derives parts of decision boundaries from the population data. These population-based decision boundaries increase the generalization of classifiers when the number of patient-specific training data is limited. On the other hand, different from conventional incremental learning (IL), which assumes that the decision boundaries learned from the population data are all correct, ILSM discards some inapplicable population-based decision boundaries in order to better fit the patient-specific data. By selectively deriving decision boundaries from population-based classifiers, ILSM improves the performance of landmark detection over PPAT and IL.

9) *Random sample consensus (RANSAC) is used to align the previous segmentations of the same patient onto the target treatment image by considering the possibility of mis-detected landmarks*

There are always situations when landmarks are detected in wrong positions. To make prostate localization robust to wrongly detected landmarks, section 4.3 presented a robust model fitting algorithm - RANSAC, which is well known in computer vision but less explored in the field of medical image analysis. I used RANSAC to fit a previous prostate shape onto the detected landmarks. Because the number of detected landmarks are more than the minimum number of landmarks required for estimating the rigid transformation, the RANSAC fitting process is tolerant of a few landmarks being wrongly detected. To further improve the localization accuracy, I borrowed the idea from multi-atlas based segmentation and aligned not only the prostate shape in the planning image but also the prostate shapes in the previous treatment images onto the target image for localization.

10) *Extensive experiments on a large prostate CT dataset ($>$ 400 treatment CT images) show that the proposed method is able to accurately localize the prostate in treatment CTs within 4 seconds; the method satisfies the accuracy and efficiency requirement of IGRT.*

Extensive experiments were conducted to evaluate the proposed method on a large CT dataset for treatment-CT segmentation. All the experiments were presented in section 4.4. In the experiments I evaluated ILSM by comparing it with other learning schemes, such as population-based learning, pure patient-specific learning, mixture learning and conventional incremental learning, on both landmark detection accuracy and localization accuracy. I also compared the proposed method with other popular methods on treatment-CT segmentation, such as registration-based methods and classification-based methods. All the comparisons showed that ILSM is a good fit for prostate localization in treatment CT images in terms of both accuracy and speed. Besides these comparison experiments, other experiments were also conducted to evaluate the performance of ILSM using only one landmark, the robustness to unsupervised annotations, the sensitivity to landmark selection and the temporal localization accuracy. These experiments offer a thorough understanding of the performance of ILSM under different real-world settings.

Thesis. *Deformable models benefit in accuracy from explicitly learning deformations from image appearance. Landmarks can be utilized for fast and accurate segmentation of treatment CTs by effectively combining limited patient-specific data with massive population data in the cascade learning framework.*

This dissertation showed that learning a deformation field to guide deformable segmentation overcomes two limitations of conventional deformable models, i.e., sensitivity to initialization and inflexibility to segment organs with complex shapes and shape variations. The deformation field is useful in situations when it is difficult to initialize deformable models or the target organ has large shape variations. The problem of CT pelvic organ segmentation falls exactly into these situations. Specifically, the bladder and rectum have variable shapes and are thus difficult for initialization. And the rectum has complex tubular shapes, which bring challenges to conventional deformable models. A global deformation field eliminates the pain of initialization as deformations toward the target organ boundaries are available at every image location. Even if the shape model is initialized far away, it can still be deformed onto the boundaries following the guidance from the deformation field. In addition to initialization, the flexbility of deformable models is also increased thanks to the deformation field. Based on the location, the deformation at each vertex can be dramatically different. Deformations are large if a vertex is far from the boundary and small if it is close. The spatially adaptive deformations make it easy to deform the mean shape onto the rectum boundary even if the mean shape is quite different from the target rectum shape to be segmented.

To improve the efficiency of prostate localization in treatment CTs, this dissertation proposed a landmark-based segmentation method. To address the limitation of insufficient patient-specific data in the beginning of radiotherapy, an incremental learning algorithm was proposed to update population-based landmark detector with limited patient-specific data. This algorithm can effectively combine massive population data with limited patient-specific data. It also addresses the overfitting problem if the landmark detector is trained with only limited patient-specific data. The experimental results showed that patient-specific data is critical for high segmentation accuracy of the prostate in treatment CT images. By combining massive population data with limited patient-specific data, the accuracy and robustness of prostate segmentation is improved especially in the beginning treatment days when the patient-specific data is limited. The idea of incremental learning provides a new way to combine population and patient-specific data. It is applicable not only in daily treatment-CT segmentation but also in other applications where longitudinal images from the same patient are acquired in an ordered time manner.

5.2 Discussion

Cone beam CT. In this dissertation the daily treatment images were acquired on a CT-on-rail scanner. Nowadays instead of acquiring a CT image, more and more hospitals are acquiring a cone beam CT image (CBCT) at each treatment day in order to reduce the imaging radiation to the patient. As the proposed method is purely data-driven, it would be relatively easy to extend the method to localize the prostate in CBCT images without any change of implementation. However, the CBCT images are often noisier than CT images, and motion artifacts are more severe in CBCT images than CT images. Thus it is still a question that whether the proposed method can work well in CBCT prostate localization. If the appearance patterns of a landmark are consistent across different treatment days in CBCT, I think it is very likely that the proposed method would perform well in CBCT images. However, if the noise and artifacts make the appearance patterns of the same landmark different across treatment days, it will downgrade the performance when the method is applied to CBCT images. Thus, there is still work that needs to be done for evaluating the proposed method in CBCT images.

Metal Artifacts. The proposed methods work well in the presence of mild metal artifacts, such as those caused by fiducial markers. However, they fail in the presence of severe metal artifacts, such as those caused by an implanted hip prosthesis. While metal deletion techniques [Boas and Fleischmann, 2012] can be used to reduce such image artifacts in the reconstruction stage, they are not universally available. Therefore, it is still desirable to address severe metal artifacts in the post-processing stage after image reconstruction. Recent research studies on sparse representation and matrix recovery have shown that corrupted images can be well recovered under mild assumptions, such as assuming the low rank property of image matrix. These techniques can be potentially useful to recover CT images from severe metal artifacts. By adopting them as a pre-processing step, it may be feasible to directly apply the proposed methods for organ segmentation in CT images that are contaminated by severe metal artifacts.

Applicability to Other Problems. Although the methods discussed in this dissertation are proposed for segmenting pelvic organs in CT images, they are applicable to other general problems as well. For example, the regression-based deformable model (RDM) can be applied to most segmentation problems in medical images. It is particularly suitable for segmenting organs with complex shapes and organs that are difficult for initialization. For example, we can use RDM for segmenting the cervix in female CT images and for localizing regions of interest in magnetic resonance (MR) brain images. The landmark detection method is general and can be applied to detect any anatomical landmark in medical images. In this dissertation landmarks are used for segmentation. But they can be also used for other purposes, such as registration and disease diagnosis. The incremental learning algorithm can be used in other applications where images of the same patient are acquired longitudinally. In such a context, patient-specific images acquired in the past can be used in the incremental learning algorithm to improve the segmentation of future images from the patient.

Limitations. The regression-based deformable model (RDM) utilizes context information for segmentation. It works well if the context information is reliable across subjects. In situations where the context information is variable, RDM will fail. A good example is tumor segmentation. In brain MR images, tumors can appear almost everywhere in the brain. The uncertainty of tumor positions makes context information useless. In such a case, it is impossible to learn a regressor that is able to predict the deformation from an antomical structure to the tumor since the relative position of any anatomical structure to the tumor can change. So RDM is not good for tumor segmentation. Similarly, RDM is not good for general segmentation problems in natural images where context information is variable and not useful.

Landmark-based segmentation is very suitable for prostate localization in treatment CT images because in most cases the prostate shape changes little across different treatment days. However, due to the bowel gas, the prostate shape may change notably. In such a situation, the landmark-based approach may not work well since none of the existing prostate shapes is similar to the target prostate shape. Similarly, the landmark-based approach doesn't work well for segmentation of the bladder and the rectum in treatment CT images because of their large shape variations. In summary, if the shape variation of the target organ is large, landmark-based approaches may not get accurate segmentation because a few landmarks are insufficient to describe the entire shape. More sophosticated methods need to be adopted afterwards to further refine the landmark-based segmentation.

5.3 Future Work

The following lists a few directions that may be interesting to explore.

Feature Learning. In this dissertation only 3D Haar-like features were used. Although they are very efficient to compute, these simple features have limitations.

- Haar-like features are not rotation invariant. They do not work well when the testing image is rotated by certain angles that are unseen in the training set. This situation could happen when the patient setup is not well performed or the patient cannot lie flat on his back due to a severe disease, e.g., spinal stenosis.

- Haar-like features are too simple to capture complex texture features. The texture features may be important if the proposed methods are applied to MR images that contain richer texture than CT images.

To address the limitations of Haar-like features, recent advances in feature learning can be potentially borrowed to learn informative features adaptive to the image data at hand. For example, convolutional neural networks may be adopted to learn image features and a regression model together for deformation estimation. It is expected that the performance of the proposed methods can be improved by using data-specific and task-specific features.

Multi-modality. Due to the low tissue contrast of CT images, MR images are increasingly used in the clinic for image guided radiotherapy. Since MR images provide better tissue contrast than CT images, MR segmentation can be used to guide CT segmentation. Specifically, automatic segmentation methods can be first applied to extract contours of pelvic organs from MR images. Then, these MR contours can be used to construct patient-specific shape models for CT segmentation. With a better shape constraint, the robustness and accuracy of CT segmentation can be improved.

Landmark Selection. The selection of landmarks is important in landmark-based segmentation. Nevertheless, landmarks are often selected manually by finding distinctive boundary locations on the target organ. For organs with stable shapes, such as the prostate, it is easy to select useful landmarks for localization and segmentation. However, it may not be intuitive to find those landmarks for organs with large shape variation, such as the bladder and rectum. Therefore, it is desirable to develop an unsupervised learning method for automatic landmark discovery. A good landmark for organ localization should have 1) distinctive image appearance compared to surrounding image locations and 2) consistent image appearances across different subjects.

BIBLIOGRAPHY

[American Cancer Society, 2015] American Cancer Society (2015). Cancer facts & figures 2015.

[Assley and Chellakkon, 2014] Assley, P. S. B. S. and Chellakkon, H. S. (2014). A comparative study on medical image segmentation methods. *Applied Medical Informatics*, 34(1):31–45.

[Boas and Fleischmann, 2012] Boas, F. E. and Fleischmann, D. (2012). Ct artifacts: causes and reduction techniques. *Imaging in Medicine*, 4(2):229–240.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[Broadhurst et al., 2006] Broadhurst, R. E., Stough, J., Pizer, S. M., and Chaney, E. L. (2006). A statistical appearance model based on intensity quantile histograms. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pages 422–425.

[Caruana, 1997] Caruana, R. (1997). Multitask learning. *Machine Learning*, 28(1):41–75.

[Caselles et al., 1993] Caselles, V., Catté, F., Coll, T., and Dibos, F. (1993). A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31.

[Cates et al., 2007] Cates, J., Fletcher, P. T., Styner, M., Shenton, M., and Whitaker, R. (2007). *Shape Modeling and Analysis with Entropy-Based Particle Systems*, pages 333–345. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Chan and Vese, 2001] Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277.

[Chaney and Pizer, 2016] Chaney, E. L. and Pizer, S. M. (2016). *Image Processing in Radiation Therapy*, chapter Deformable Shape Models for Image Segmentation. CRC Press.

[Chang and Lin, 2011] Chang, C.-C. and Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27.

[Chen et al., 2011] Chen, S., Lovelock, D. M., and Radke, R. J. (2011). Segmenting the prostate and rectum in ct imagery using anatomical constraints. *Medical Image Analysis*, 15(1):1–11.

[Cohen, 1991] Cohen, L. D. (1991). On active contour models and balloons. *CVGIP: Image Underst.*, 53(2):211–218.

[Cootes et al., 1995] Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models&mdash;their training and application. *Comput. Vis. Image Underst.*, 61(1):38–59.

[Costa et al., 2007] Costa, M. J., Delingette, H., Novellas, S., and Ayache, N. (2007). Automatic segmentation of bladder and prostate using coupled 3d deformable models. *Med Image Comput Comput Assist Interv*, 10(Pt 1):252–60.

[Criminisi et al., 2013] Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., and Siddiqui, K. (2013). Regression forests for efficient anatomy detection and localization in computed tomography scans. *Medical Image Analysis (MedIA)*.

[Criminisi and Shotton, 2013] Criminisi, A. and Shotton, J. (2013). *Decision Forests for Computer Vision and Medical Image Analysis*. Springer Publishing Company, Incorporated.

[Criminisi et al., 2011] Criminisi, A., Shotton, J., and Konukoglu, E. (2011). Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical Report MSR-TR-2011-114, Microsoft Research.

[Davies et al., 2001] Davies, R. H., Cootes, T. F., and Taylor, C. J. (2001). *Information Processing in Medical Imaging: 17th International Conference, IPMI 2001 Davis, CA, USA, June 18–22, 2001 Proceedings*, chapter A Minimum Description Length Approach to Statistical Shape Modelling, pages 50–63. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Davies et al., 2010] Davies, R. H., Twining, C. J., Cootes, T. F., and Taylor, C. J. (2010). Building 3-d statistical shape models by direct optimization. *IEEE Transactions on Medical Imaging*, 29(4):961–981.

[Dawson and Jaffray, 2007] Dawson, L. and Jaffray, D. (2007). Advances in image-guided radiation therapy. *Journal of Clinical Oncology*, 25(8):938–946.

[Feng et al., 2009] Feng, Q., Foskey, M., Tang, S., Chen, W., and Shen, D. (2009). Segmenting ct prostate images using population and patient-specific statistics for radiotherapy.

[Fischer and Modersitzki, 2003] Fischer, B. and Modersitzki, J. (2003). *FLIRT: A Flexible Image Registration Toolbox Biomedical Image Registration*, volume 2717 of *Lecture Notes in Computer Science*, pages 261–270. Springer Berlin / Heidelberg.

[Fischler and Bolles, 1981] Fischler, M. and Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.

[Fletcher et al., 2009] Fletcher, P. T., Venkatasubramanian, S., and Joshi, S. (2009). The geometric median on riemannian manifolds with application to robust atlas estimation. *NeuroImage*, 45(1, Supplement 1):S143 – S152. Mathematics in Brain Imaging.

[Foskey et al., 2005] Foskey, M., Davis, B., Goyal, L., Chang, S., Chaney, E., Strehl, N., Tomei, S., Rosenman, J., and Joshi, S. (2005). Large deformation three-dimensional image registration in image-guided radiation therapy. *Phys Med Biol*, 50(24):5869.

[Freedman et al., 2005] Freedman, D., Radke, R. J., Tao, Z., Yongwon, J., Lovelock, D. M., and Chen, G. T. Y. (2005). Model-based segmentation of medical imagery by matching distributions. *Medical Imaging, IEEE Transactions on*, 24(3):281–292.

[Gao et al., 2012a] Gao, Y., Liao, S., and Shen, D. (2012a). Prostate segmentation by sparse representation based classification. *Medical Physics*, 39(10):6372–6387.

[Gao et al., 2012b] Gao, Y., Liao, S., and Shen, D. (2012b). Prostate segmentation by sparse representation based classification. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, volume 7512 of *Lecture Notes in Computer Science*, pages 451–458. Springer Berlin Heidelberg.

[Gao et al., 2016] Gao, Y., Shao, Y., Lian, J., Wang, A. Z., Chen, R. C., and Shen, D. (2016). Accurate segmentation of ct male pelvic organs via regression-based deformable models and multi-task random forests. *IEEE Transactions on Medical Imaging*, 35(6):1532–1543.

[Gao and Shen, 2015] Gao, Y. and Shen, D. (2015). Collaborative regression-based anatomical landmark detection. *Physics in Medicine and Biology*, 60(24):9377.

[Gao et al., 2014] Gao, Y., Zhan, Y., and Shen, D. (2014). Incremental learning with selective memory (ilsm): Towards fast prostate localization for image guided radiotherapy. *Medical Imaging, IEEE Transactions on*, 33(2):518–534.

[Gower, 1975] Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1):33–51.

[Huynh et al., 2015] Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., and Shen, D. (2015). Estimating ct image from mri data using structured random forest and auto-context model. *Medical Imaging, IEEE Transactions on*, PP(99):1–1.

[Joshi et al., 2004] Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23, Supplement 1:S151 – S160. Mathematics in Brain Imaging.

[Kass et al., 1988] Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.

[Koninklijke Philips N.V., 2016] Koninklijke Philips N.V. (2004-2016). Phillips pinnacle3 model-based segmentation. URL: http://www.usa.philips.com/healthcare/product/HCNOCTN139/pinnacle3-model-based-segmentation.

[Lay et al., 2013] Lay, N., Birkbeck, N., Zhang, J., and Zhou, S. (2013). Rapid multi-organ segmentation using context integration and discriminative models. In Gee, J., Joshi, S., Pohl, K., Wells, W., and Zllei, L., editors, *Information Processing in Medical Imaging*, volume 7917 of *Lecture Notes in Computer Science*, pages 450–462. Springer Berlin Heidelberg.

[Lee et al., 2010] Lee, H.-P., Foskey, M., Levy, J., Saboo, R., and Chaney, E. (2010). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III*, chapter Image Estimation from Marker Locations for Dose Calculation in Prostate Radiation Therapy, pages 335–342. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Li et al., 2012] Li, W., Liao, S., Feng, Q., Chen, W., and Shen, D. (2012). Learning image context for segmentation of the prostate in ct-guided radiotherapy. *Physics in Medicine and Biology*, 57(5):1283–1308.

[Liao et al., 2013] Liao, S., Gao, Y., Lian, J., and Shen, D. (2013). Sparse patch-based label propagation for accurate prostate localization in ct images. *Medical Imaging, IEEE Transactions on*, 32(2):419–434.

[Liao and Shen, 2012] Liao, S. and Shen, D. (2012). A feature-based learning framework for accurate prostate localization in ct images. *Image Processing, IEEE Transactions on*, 21(8):3546–3559.

[Liu et al., 2005] Liu, F. T., Ting, K. M., and Fan, W. (2005). *Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, 2005. Proceedings*, chapter Maximizing Tree Diversity by Building Complete-Random Decision Trees, pages 605–610. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Liu et al., 2010] Liu, W., Qian, J., Hancock, S. L., Xing, L., and Luxton, G. (2010). Clinical development of a failure detection-based online repositioning strategy for prostate imrt—experiments, simulation, and dosimetry study. *Medical Physics*, 37(10):5287–5297.

[Lorensen and Cline, 1987] Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169.

[Lu et al., 2011] Lu, C., Chelikani, S., Papademetris, X., Knisely, J. P., Milosevic, M. F., Chen, Z., Jaffray, D. A., Staib, L. H., and Duncan, J. S. (2011). An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotherapy. *Medical Image Analysis*, 15(5):772 – 785. Special Issue on the 2010 Conference on Medical Image Computing and Computer-Assisted Intervention.

[Lu et al., 2012] Lu, C., Zheng, Y., Birkbeck, N., Zhang, J., Kohlberger, T., Tietjen, C., Boettger, T., Duncan, J., and Zhou, S. (2012). Precise segmentation of multiple organs in ct volumes using learning-based approach and information theory. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, volume 7511 of *Lecture Notes in Computer Science*, pages 462–469. Springer Berlin Heidelberg.

[Martínez et al., 2014] Martínez, F., Romero, E., Dréan, G., Simon, A., Haigron, P., de Crevoisier, R., and Acosta, O. (2014). Segmentation of pelvic structures for planning ct using a geometrical shape model tuned by a multi-scale edge detector. *Physics in Medicine and Biology*, 59(6):1471.

[Myronenko and Song, 2010] Myronenko, A. and Song, X. (2010). Point set registration: Coherent point drift. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2262–2275.

[Pizer et al., 2005] Pizer, S., Fletcher, P., Joshi, S., Gash, A., Stough, J., Thall, A., Tracton, G., and Chaney, E. (2005). A method and software for segmentation of anatomic object emsembles by deformable m-reps. *Medical Physics*, 32(5):1335–1345.

[Pizer, 2016] Pizer, S. M. (2016). Personal Communication.

[Polikar et al., 2001] Polikar, R., Upda, L., Upda, S. S., and Honavar, V. (2001). Learn++: an incremental learning algorithm for supervised neural networks. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 31(4):497–508.

[Rousson et al., 2005] Rousson, M., Khamene, A., Diallo, M., Celi, J., and Sauer, F. (2005). Constrained surface evolutions for prostate and bladder segmentation in ct images. In Liu, Y., Jiang, T., and Zhang, C., editors, *Computer Vision for Biomedical Image Applications*, volume 3765 of *Lecture Notes in Computer Science*, pages 251–260. Springer Berlin Heidelberg.

[Shawe-Taylor and Cristianini, 2004] Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York, NY, USA.

[Shi et al., 2013] Shi, Y., Liao, S., Gao, Y., Zhang, D., Gao, Y., and Shen, D. (2013). Prostate segmentation in ct images via spatial-constrained transductive lasso. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 2227–2234.

[Shinohara and Roach, 2007] Shinohara, K. and Roach, Mack, I. (2007). Technique for implantation of fiducial markers in the prostate. *Urology*, 71(2):196–200.

[Sun and Genton, 2011] Sun, Y. and Genton, M. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

[Tu and Bai, 2010] Tu, Z. and Bai, X. (2010). Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Trans Pattern Anal Mach Intell*, 32(10):1744–57.

[Vapnik, 1995] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA.

[Viola and Jones, 2004] Viola, P. and Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154.

[Wang et al., 2015] Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J. H., Lin, W., and Shen, D. (2015). Links: Learning-based multi-source integration framework for segmentation of infant brain images. *NeuroImage*, 108:160 – 172.

[Xing et al., 2006] Xing, L., Thorndyke, B., Schreibmann, E., Yang, Y., Li, T.-F., Kim, G.-Y., Luxton, G., and Koong, A. (2006). Overview of image-guided radiation therapy. *Medical Dosimetry*, 31(2):91–112.

[Xu et al., 2000] Xu, C., Pham, D. L., and Prince, J. L. (2000). Image segmentation using deformable models. In *Handbook of Medical Imaging. Vol.2 Medical Image Processing and Analysis*, pages 129–174. Bellingham, WA: SPIE.

[Zhan et al., 2011a] Zhan, Y., Dewan, M., Harder, M., Krishnan, A., and Zhou, X. S. (2011a). Robust automatic knee mr slice positioning through redundant and hierarchical anatomy detection. *Medical Imaging, IEEE Transactions on*, 30(12):14.

[Zhan et al., 2011b] Zhan, Y., Dewan, M., and Zhou, X. S. (2011b). Auto-alignment of knee mr scout scans through redundant, adaptive and hierarchical anatomy detection. In *Proceedings of the 22Nd International Conference on Information Processing in Medical Imaging*, IPMI'11, pages 111–122, Berlin, Heidelberg. Springer-Verlag.

[Zhan et al., 2008] Zhan, Y., Zhou, X., Peng, Z., and Krishnan, A. (2008). *Active Scheduling of Organ Detection and Segmentation in Whole-Body Medical Images*, volume 5241 of *Lecture Notes in Computer Science*, chapter 38, pages 313–321. Springer Berlin Heidelberg.

[Zhang et al., 2015] Zhang, J., Gao, Y., Wang, L., Tang, Z., Xia, J., and Shen, D. (2015). Automatic craniomaxillofacial landmark digitization via segmentation-guided partially-joint regression forest model and multi-scale statistical features. *Biomedical Engineering, IEEE Transactions on*, PP(99):1–1.