THE UNIVERSITY *of* EDINBURGH

Textural Features for Bladder Cancer Definition on CT Images

Hanqing Liao

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
November 6, 2012

Abstract

Genitourinary cancer refers to the presence of tumours in the genital or urinary organs such as bladder, kidney and prostate. In 2008 the worldwide incidence of bladder cancer was 382,600 with a mortality of 150,282. Radiotherapy is one of the main treatment choices for genitourinary cancer where accurate delineation of the gross tumour volume (GTV) on computed tomography (CT) images is crucial for the success of this treatment. Limited CT resolution and contrast in soft tissue organs make this difficult and has led to significant inter- and intra- clinical variability in defining the extent of the GTV, especially at the junctions of different organs. In addition the introduction of new imaging techniques and modalities has significantly increased the number of the medical images that require contouring. More advanced image processing is required to help reduce contouring variability and assist in handling the increased volume of data.

In this thesis image analysis methodologies were used to extract low-level features such as entropy, moment and correlation from radiotherapy planning CT images. These distinctive features were identified and used for defining the GTV and to implement a fully-automatic contouring system. The first key contribution is to demonstrate that second-order statistics from co-occurrence matrices (GTSDM) give higher accuracy in classifying soft tissue regions of interest (ROIs) into GTV and non-GTV. Loadings of the principal components (PCs) of the GTSDM features were found to be consistent over different patients. Exhaustive feature selection suggested that entropies and correlations produced consistently larger areas under receiver operating characteristic (AUROC) curves than first-order features.

The second significant contribution is to demonstrate that in the bladder-prostate junction, where the largest inter-clinical variability is observed, the second-order principal entropy from stationery wavelet denoised CT images (DPE) increased the saliency of the bladder prostate junction. As a result thresholding of the DPE produced good agreement between gold standard clinical contours and those produced by this approach with Dice coefficients.

The third contribution is to implement a fully automatic and reproducible system for bladder cancer GTV auto-contouring based on classifying second-order statistics. The Dice similarity coefficients (DSCs) were employed to evaluate the automatic contours. It was found that in the mid-range of the bladder the automatic contours are accurate, but in the inferior and superior ends of bladder automatic contours were more likely to have small DSCs with clinical contours, which reconcile with the fact of clinical variability in defining GTVs. A novel male bladder probability atlas was constructed based on the clinical contours and volume estimation from the classification results. Registration of the classification results with this probabilistic atlas consistently increases the DSCs of the inferior slices.

Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself, unless otherwise explicitly declared, was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at The University of Edinburgh.

Hanqing Liao

Acknowledgements

First of all I wish to thank my supervisors Professor Stephen McLaughlin and Dr. William Nailon whose patient, consistent steering of my research and thoughtful advices and delicate proof readings helped me enormously to form the integrity of this thesis. I would also like to thank Dr. Robin Steel for his collaboration on the atlas-based registration part of this thesis and Dr. Duncan McLaren for his generous and rigorous clinical reviews of my experiment results.

Secondly I should thank my family, especially my mother Dr. Rong Liao, for funding most part of my research for four years without which I would never have had the opportunity to start this PhD project. I should owe gratitude again to Professor Stephen McLaughlin who provided me with an annuity during the course of my research to cover part of the cost. Moreover, I should express my gratefulness to Professor Stephen Warrington, Dr. Rebecca Liu and Mr. Ashley Evans who occasionally provided me with some causal jobs of an industrious context, which covered part of my daily cost and broadened my eyesight significantly.

I should thank the faculties of both School of the Engineering, University of Edinburgh and Department of Oncology Physics, Western General Hospital for all the technical and data support. Affiliations should be given to Mr. Kun Cheng, Mr. Dean Montgomery, Mr. Yang Feng and Miss Yuan Tian, with whom I finally felt that there is a team in working towards the fully automatic radiotherapy planning which not only I belonged to, but also contributed to. I would thank Mr Mael Jambou, Mr Hugo Sanchez and Miss Yuanwei He for the cooperation of containing all the miserable encounters in my final stage of PhD. Special affiliations should be assigned to Dr Yigui Ma for all the happiest, sweetest memory in my PhD years.

Finally I wish to give special, sincere gratefulness to those who I met in every academic conferences, seminars, workshop who gave brilliant and inspiring ideas to advance my understanding of the scientific research.

Contents

List of figures

List of tables

Acronyms and Abbreviations

2D	Two-dimensional
3D	Three-dimensional
3D CRT	Three-dimensional conformal radiotherapy
AUROC	Area under receiver operating characteristic
BPJ	Bladder-prostate junction
CT	Computed tomography
CTV	Clinical target volume
DPE	Second-order principal entropies from the denoised images
DSC	Dice similarity coefficient
DWT	Discrete wavelet transform
FOE	First-order entropy
FOS	First-order Statistics
GE	General Electric Co.
GLRLM	Gray-level run-lengths matrix
GTSDM	Gray-tone spatial dependence matrix, or co-occurrence matrix
GTV	Gross tumour volume
ICRU	International Commission on Radiation Units and Measurement
IGRT	Image guided radiotherapy
IMRT	Intensity modulated radiotherapy
LRT	Log-likelihood ratio test
MRI	Magnetic resonance imaging
NBC	naive Bayes classifier
OAR	Organs at risk
PC	Principal component
PCA	Principal component analysis
PE	Second-order principal entropies
PET/CT	Positron emission tomography / computed tomography
PTV	Planning target volume

ll

ROC	Receiver operating characteristics
ROI	Region of interest
STD	Standard deviation
SVD	Singular value decomposition ll
SWT	Stationary wavelet transform
TNM system	Tumour, node, metastasis system for cancer staging
TUR	Transurethral resection

Nomenclature

μ	Mean
ϕ	Scaling function
ψ	Wavelet function
σ	Standard deviation
θ	Angle
$\mathbf{C}_{\mathbf{d},\theta}$	GTSDM
$C_{d,\theta}(i,j)$	GTSDM elements: number of co-occurrences
C	Correlation
f	Features
$\mathbf{I}$	Image, either 2D or 3D, depending on the context
$I(x,y)$	2D Image as a function of spatial index x and y
$I(x,y,z)$	3D Image as a function of spatial index x, y and z
$i(x,y)$	Intensity of a pixel in image $\mathbf{I}$
H	Entropy
$p(i)$	First-order distribution function of image intensity i
$p(i,j)$	Joint distribution function of intensities of two pixels i and$j$
$\mathbf{R}_\theta$	GLRLM
$r(i,l)$	GLRLM element: number of runs l of intensity i

Chapter 1
Introduction

1.1 Introduction

The advent of digital imaging, especially three-dimensional (3D) imaging to medical practice offers a huge advantage over traditional film-based imaging methods. With the help of computing power and sophisticated algorithms digital medical images can be efficiently acquired, visualised, analysed, stored, replicated and communicated. The flexibility and the ability to process these digital medical images using advanced mathematical techniques provides invaluable information for medical decisions at a significantly reduced cost compared to traditional film-based images.

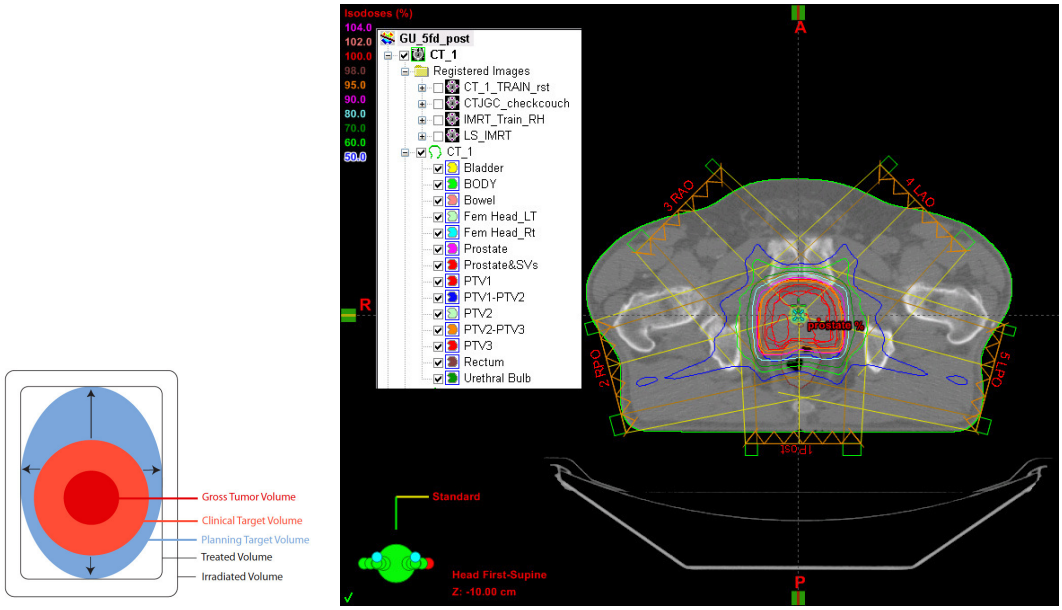
One crucial branch of medical image processing is auto-segmentation/classification. The purpose of segmentation is to extract the shape of a desired semantic region(s) such as organs from a medical image automatically or semi-automatically. Classification on the other hand assigns the semantic meaning to the segmented region(s). By leveraging the advanced computation power successful segmentation system can increase the quality and reduce the time consumption for anatomy delineation. In general auto-segmentation/classification is a very difficult task because image interpretation is highly subjective: a considerable amount of clinical knowledge is needed to interpret medical images. Thanks to the effort of scientific research, successful segmentation/classification applications have gradually become standard in medical practice[8]. However, significant challenges still remain, of which the automatic segmentation/classification of the human pelvic organs in the treatment of genitourinary cancer by radiotherapy planning, which is the focus of this thesis.

1.2 Context

Genitourinary cancer refers to the presence of malignant tumours in the genitourinary organs such as kidney, bladder and prostate. In 2008 bladder cancer had an annual worldwide incidence of 382,600 cases with a mortality of 150,282 cases. It is more common in men than

women (ratio 3:1) and has a particularly high incidence in North America, Western Europe, Africa and Middle East[9]. In 2004 it was reported that the five year survival rate was 60% to 80% [10]. Among all the treatment options external beam radiotherapy can deliver a very accurate radiation dose to the tumour site and is therefore very popular in bladder cancer treatment. Successful radiotherapy calls for precise tumour localisation. Medical imaging is intensively used for this task because it provides good insight of human anatomy. To promote the international agreement on prescribing, recording and reporting of the radiotherapy target volume the International Commission on Radiation Units and Measurement (ICRU) issued a series of reports for different modalities[1, 11, 12], which is illustrated in Figure 1.1.

However two major problems exist when localising tumour using medical images: one is that due to limited resolution the saliency of the cancerous structures may not be sufficient, therefore the extent of tumour is difficult to identify and agree on. In the ICRU reports the primary tumour site is defined as gross tumour volume (GTV) which is the demonstrable malignant growth in the image. The sub-clinical growth, which contains all GTV and some invisible cancerous structures with a substantial amount of cancer cells as a margin of GTV, is defined as clinical target volume (CTV). Both GTV and CTV are anatomical concepts which need to be defined independent of the treatment methods. Presently computed tomography (CT) is the standard image modality for tumour contouring because it has consistent geometry, provides electron density information for dose calculation, and high contrast bone tissues as landmarks[13]. These properties are not available in other image modalities such as magnetic resonance imaging (MRI). However the soft tissue contrast in CT images is poor, plus cancerous structures are enormously variable among the population, observing demonstrable, consistent and widely-accepted features in the image for GTV and CTV definition can be very subjective. GTV and CTV definition is therefore highly expertise and experience demanding. As a result significant inter- and intra-clinical variability is reported in the literature: contours drawn by different clinicians may be significantly different while the contours drawn by the same clinician are not reproducible[14–16]. For genitourinary cancer Figure 1.2 shows three CT images from [2] at different positions of bladder: superior (C), mid (A) and inferior (B). It can be seen that the soft tissue contrast is indeed limited, (B) and (C) in which the clinical contours have significant variability at the superior and inferior ends of the bladder. Imprecise definition of GTV and CTV will lead to the propagation of geometric error to the next steps of radiotherapy, which will compromise the outcome[2]. To reduce the inter- and intra-clinical variability distinctive features must be found on the image using image analysis techniques.



(a) (b)

Figure 1.1: *Left: Illustration of GTV, CTV, PTV and OAR from ICRU Report 50[1]. Right: a radiotherapy planing system in clinical practice: contours and radiation beams.*

Another problem is that patient movement will change the tumour location instantly. In standard clinical practice there is usually one week between CT scan and radiation treatment. During this period the position of the tumour may change. To cope with this ICRU recommended in its reports that the planning target volume (PTV) to be all the possible positions of the tumour which contains CTV and a motion margin, as is illustrated in Figure 1.1[1, 11, 12]. The PTV is a geometric concept and can be larger than necessary. The recent development of image guided radiotherapy (IGRT) makes it possible to acquire CT images at time of treatment to account for the target position and volume to reduce the margin caused by movement. IGRT can give a more accurate dose delivery only if the target or radiation can be accurately defined during the treatment. However the large number of images that require contouring as well as the need for the presence of a clinician at the time of treatment make manual contouring for IGRT impractical. Automatic segmentation has the potential of shortening the time required for contouring and improving the accuracy and precision of dose delivery.

For these reasons in 2007 it is asserted that developing a fast and fully automatic contouring system should be given a high priority in the medical image analysis field[17]. In this thesis medical image segmentation/classification methodologies, particularly texture analysis, are proposed to build an automatic, objective and accurate GTV contouring system for bladder

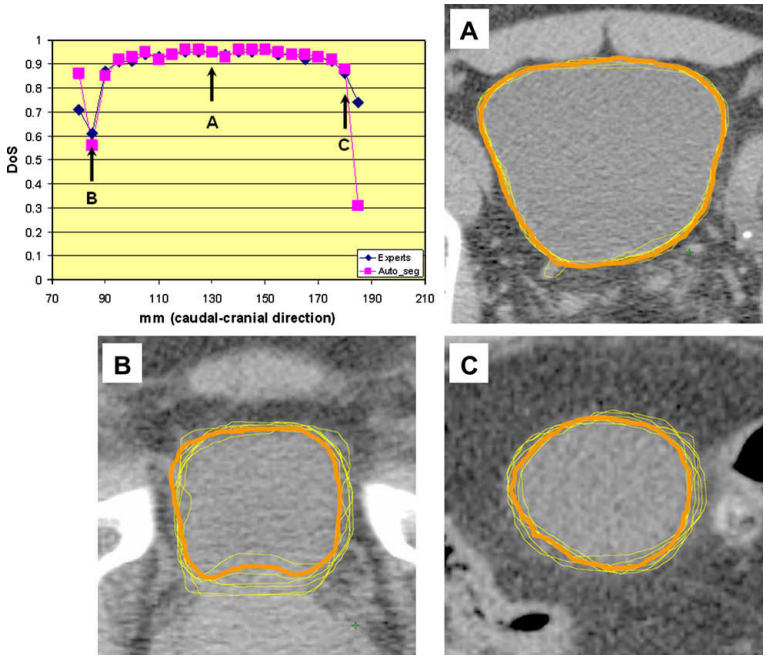


Figure 1.2: *Three CT images with different clinical contours from [2] at different positions of bladder: (A) mid, (B) inferior and (C) superior. Clinical contours show high variability in the (C) superior and (B) inferior range.*

cancer treatment planning.

1.3 Medical Image Segmentation/Classification: An Overview

A two-dimensional (2D) image can be considered as a two-variable function $I(x, y)$ with respect to the spatial index x and y, which maps a scene of the physical world into a plane. A stack of images may form a three-dimensional (3D) function $I(x, y, z)$, or a 3D image. Traditional film-based images can be considered continuous. Because digital imaging systems engage computers to process the discrete, quantised images, the efficiency of analysing and processing images has been greatly improved. In addition many new applications such as remote diagnosis have arisen.

1.3.1 Digital Medical Images: Fundamental Terminology

For digital medical images, the minimal resolvable unit in 2D is called a **pixel**. The size of real physical world that one pixel represents is called the **resolution**, the spatial indexes are called **co-ordination system** and the functional value of the pixel is called the **intensity**, or **gray**

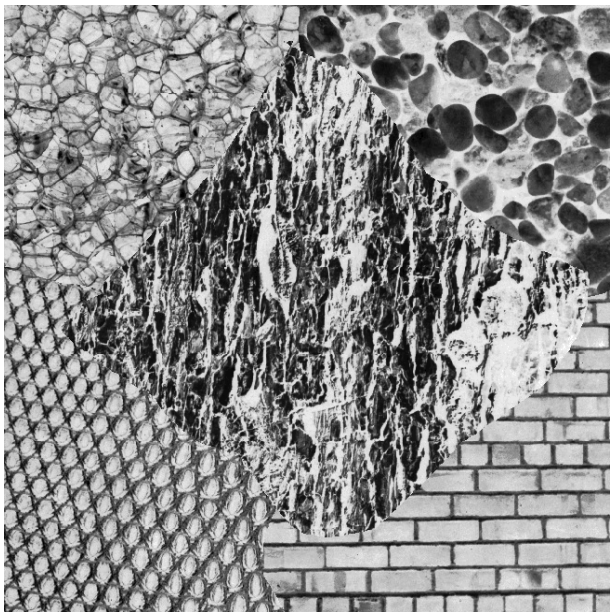


Figure 1.3: *Five Brodatz textures from [3]. The regions are heterogeneous and as a result, edges do not correspond to the semantic region boundaries.*

level, which are used interchangeably. To assist in the understanding of medical images image analysis and processing methodologies have been developed for image enhancement, feature extraction, segmentation/classification and modelling by applying mathematical operations on the image. A group of pixels can form a semantic region, which is a subset of the image $S \subseteq I$ representing one semantic entity. In medical image segmentation the entity may be an organ, the GTV, or any region of interest (ROI). The border of two semantic regions is named a **boundary**, which defines the extent of the semantic regions.

1.3.2 Boundary Definition: Different Perspectives

The boundary may correspond to locations of significant intensity changes, or **edges**. For example in Figure 1.2 (A) the boundary of bladder corresponds to the intensity change between bladder and surrounding fat tissues. This kind of boundary is relatively easy to define. However when two semantic regions are both homogeneous and the intensity difference between them is very limited, there is no edge information for defining the boundary. This is the case in soft tissue organ junction. For example shown in Figure 1.2 (B) illustrates the boundary of the bladder and prostate where there is little intensity difference[18]. This makes tumour volume definition very difficult and causes significant inter- and intra-clinical variability.

However, when the regions are heterogeneous, simple statistics such as the mean and variance may be similar while edges may appear within the regions which do not correspond to the semantic boundary. This is often the case when two textural regions are adjacent. Common examples of texture are given by Brodatz [3], however a mathematical definition of the texture does not exist. For example Figure 1.3 shows an image with five different Brodatz textures. In which the edges do not correspond to semantic region boundaries. *Texture analysis* is a set of methodologies for characterizing and differentiating different textures using both low-level and high-level operations.

1.4 Objective: Towards an Automatic GTV Contouring System

Shape and texture are two of the most important properties in a medical image. Traditionally soft tissues in the pelvis region are considered homogeneous; as a result clinicians must use their experience to infer the extent of the GTV. In the literature, the majority of existing automatic segmentation algorithms use the anatomical knowledge and/or build shape models to bypass the difficulty of the boundary definition. This set of knowledge-shape based methodologies is successful in defining the extent of organs. When the boundary is weak however, the performance of this approach is poor. Moreover, because each cancer is unique to each patient, using only prior shape information of cancer volume may not be the best approach. Distinctive features which suggest clear boundary definition of the tumour are highly desirable. In this thesis, the soft tissue regions are considered heterogeneous. Texture analysis was therefore applied to exploit the spatial structure of a region, which intuitively contains more information than statistics based on stand-alone pixels.

This thesis focuses on using medical image processing methodologies, especially texture analysis to advance knowledge and technology towards a reproducible, highly accurate and fully automated GTV contouring for radiotherapy planning of genitourinary cancer. Novel statistical low-level features are extracted from local ROIs in CT images. Furthermore, the saliency and usefulness of these novel features in distinguishing GTV from non-GTV soft tissues need to be examined, which can potentially be used as a quantitative measurements to differentiate GTV from non-GTV on CT images. These novel features are integrated into the existing machine learning algorithms to show the improvement to the existing methods.

1.5 Contributions

The contribution of this thesis lies in three major aspects. First of all the statistical textural features were extracted from the local soft tissue ROIs in pelvic regions of bladder cancer patients. Compared to first-order statistics (FOS), textural features can distinguish GTV from non-GTV with higher accuracy, which was measured by area under receiver operating characteristic curve (AUROC). Both the Bayesian modelling and likelihood function modelling were used to show that two principal features yielded a consistent increase of AUROC.

The second contribution is to demonstrate that principal textural features are meaningful, in that they can increase the saliency of GTV. It was further speculated that the textural feature extraction after stationary wavelet denoising can further increase the saliency. By analysing the noise on CT images using a water phantom, the standard deviations of the noise on CT images were derived, which were later used as parameters in the soft-thresholding denoising algorithm. The principal textural features from denoised images were indeed showed better saliency.

The third contribution is to use the Bayesian classification algorithm and other post-processing algorithms to build a fully automatic GTV contouring system. The volumes of GTV obtained by the automatic algorithm were compared with the volumes of GTV obtained by the clinician; the Dice similarity coefficient (DSC) was used to characterize the overlap of clinical contours and automatic contours. Results showed that in the mid range of bladder the automatic contours yielded large DSCs with clinical contours but in the superior and inferior range the overlap of the DSCs were small. In order to improve the classification the bladder shape information was formed into a probabilistic atlas by rigid registration, and Bayesian inference was used to combine the probabilistic atlas and the probability from textural features for better classification.

In summary, this work is of inter-disciplinary importance and may break some new ground in the methodology of cancer definition on medical images. By giving more distinctive low-level features more accurate high-level models can provide better automatic contours, which advances the technology of automatic radiotherapy planning applications. Also, if any consistent features can be demonstrated with a high degree of differentiability in CT images, further studies can be carried out to obtain correspondence between image properties and pathological facts. This will thus contribute to the understanding of the characteristics of the tumour visual properties as shown on CT images, and provide quantitative textural assessment as an objective descriptor to reduce the inter- and intra-clinical variability. Last but not least, statistical textural

features correspond to the physical interaction between X-Ray and tissues. Finding distinctive textural features for GTV definition may lead to further research on why the textural features yielded higher distinction, which is a potential advance in knowledge.

1.6 Outline of the Thesis

This chapter gives the overview of the definition, importance and difficulties of the bladder cancer automatic contouring problem. The fundamental concepts in image analysis which are essential to understanding the thesis are also introduced.

In Chapter 2 the background of cancer automatic contouring problem is described. As CT images are considered a standard in defining the target volume of the radiotherapy, the development and properties of CT imaging are introduced in Section 2.2. Basic anatomy is introduced in Section 2.3 with both anatomical atlas illustrations and CT images. In Section 2.4 the diagnostic procedure and main treatment choices for bladder cancer are described, with an emphasis on the value and limit of CT imaging in diagnosing bladder cancer. In Section 2.5 the current research and practice in external beam radiotherapy are discussed, and the importance of automatic tumour contouring is detailed.

Chapter 3 gives a systematic review of the medical image processing and analysis methodologies relevant to this thesis. In Section 3.2 low-level feature extraction in the spatial domain are described, which focuses on the statistical methodologies. In Section 3.3 feature extraction techniques in the transform domain, especially the wavelet domain, are addressed. Section 3.4 covers the major existing high-level image segmentation/classification methodologies, which include registration-based, classification-based, morphology-based, shape-model-based and knowledge based segmentation. Section 3.5 discusses the most significant automatic segmentation/classification systems proposed in the literature which is usually a combination of several methodologies. The key to a successful segmentation system is to find consistent and distinctive low-level features in the image, while the high-level methodologies optimise the results that the low-level features can provide. However for the genitourinary cancer segmentation problems the distinctive features are yet to be defined. This is one of the major contributions of this work.

In Chapter 4 a set of statistical textural features are demonstrated as consistently yielding higher classification accuracy than image intensities and simple first-order statistics. In Section 4.3 it

is demonstrated that the intensities as features are insufficient to define the cancer regions. However, the variation in soft tissue regions is larger than the variation in water phantom (pure noise), which suggested that the texture may offer extra information. In Section 4.4 GTSDM features are proposed to quantise the texture information, and principal component analysis (PCA) is used to reduce the number of features. It is shown that the principal GTSDM features yield better classification performance than first order statistics, with the loadings of principal components (PCs) tested by a leave-one-out cross-validation approach. In Section 4.5 an exhaustive feature selection scheme is developed to examine the classification accuracy based on statistical features. The naive Bayes classifier (NBC) is used as the induction algorithm while the area under receiver operating characteristic (AUROC) curve was used to evaluate the classification accuracy. Results suggest that the AUROCs from the PCs representing entropies and correlations were consistently larger than AUROCs from first-order statistics.

In Chapter 5 the principal entropies from GTSDM are applied to define the extent of GTV in the bladder-prostate junction, which is one of the most difficult problems in genitourinary cancer contouring due to poor soft tissue contrast. It is first demonstrated in Section 5.3 and 5.4 that the second-order principal entropies (PE) yield significant saliency to distinguish bladder from prostate. Moreover in Section 5.5 stationary wavelet denoising was applied to some difficult cases and the second-order principal entropies were extracted from the denoised images (DPE). Results suggest that denoising can adapt to the structures in the soft tissues and therefore DPE offers a higher saliency than PE. Simple thresholding was applied to the DPE maps to extract GTV contours. Clinical reviews are conducted to evaluate these contours.

In Chapter 6 a fully automatic system is implemented based on classifying GTSDM features. Two different schemes: high sensitivity and high specificity schemes were used in Section 6.2. Percentage volume error and Dice similarity coefficient (DSC) were both employed to evaluate the automatic contours. Results suggest that the high specificity scheme produced better contours. For one patient the automatic contours were clinically evaluated. The proposed texture analysis method gave high performance in the mid-region of bladder in axial plane, while in the inferior parts of bladder, or the bladder-prostate junction, the proposed statistical textural features cannot distinguish the bladder from the prostate, leading to poor automatic contours, which confirmed the difficulty of defining GTV extent in the inferior range of bladder. In Section 6.4 the posterior probability of ROIs being GTV inferred from textural features were combined in a Bayesian model with a probabilistic atlas generated by registering clinical con-

tours to show increased classification of ROIs in inferior part of bladder. Chapter 7 concludes this thesis with summary, conclusion and discussions on the future direction for this work.

Chapter 2
Background

2.1 Introduction

Genitourinary cancer treatment is a sophisticated and complicated process. A significant amount of collaboration is needed among clinical oncologists, physicists and engineers, all of whom have sophisticated knowledge and experience in their respective field of expertise to deliver the optimal treatment outcome. This chapter provides the essential background information for understanding this thesis.

A solid understanding of pelvic anatomy is required to assist in the diagnose of cancerous growth in this region by examining the shapes and texture of the objects in medical images. Upon diagnosis, pathological grading and staging need to be carried out to provide evidence for choosing the most efficient treatment procedure for that particular cancer case. Currently the most popular treatment choices include surgery, radiotherapy, chemotherapy and biological therapy.

Thanks to advances in technology, especially the introduction of intensity modulated radiotherapy (IMRT) and image guided radiotherapy (IGRT), external beam radiotherapy provides a very accurate dose delivery to the PTV while sparing the OAR. It is therefore a very popular choice in genitourinary cancer treatment. The radiotherapy procedure comprises of contouring, which defines the physical extent of tumour(s) as the target; dose calculation, which maximises the dose delivery to the tumour(s) while sparing OAR; simulation and irradiation, which delivers the radiation dose. Advances in optimization algorithms significantly increases the dose calculation accuracy and efficiency, which enables the automatic dose calculation, while the IMRT and IGRT ensure accurate and efficient physical dose delivery [17]. However, the tumour volume definition is predominantly carried out manually by experienced oncologists, because of the poor contrast and resolution of the medical image, especially CT images. Manual contouring suffers from several significant draw-backs, including inter- and intra-clinical variability. It is not reproducible and it is time consuming. The radiotherapy community have stated that the issue of developing automatic contouring methodologies should be given a high priority [17].

In this chapter first CT imaging is described in Section 2.2, with the resolution, contrast and noise described. In Section 2.3 the anatomy of the human pelvis is illustrated by atlas and CT images. In Section 2.4 facts on basic bladder cancer diagnosis and treatment choices are discussed. Cancer staging accuracy based on medical images is reviewed to show the value and constraints of the imaging modalities used. In Section 2.5 the external beam radiotherapy procedure is described and discussed, with an emphasis on the role of CT images in planning, or more specifically target definition. A discussion is given in Section 2.5.4 to further discuss the value of auto-contouring, and Section 2.6 summarises this chapter.

2.2 CT Imaging

2.2.1 The Geometry of CT Images

One important property of CT images is the consistent geometry. That is, they have less spatial distortion, consequently the volume obtained from CT images is accurate, while other modalities such as MRI suffer from spatial distortion. For radiotherapy planning the consistent geometry is essential for the accuracy of target definition. The geometric consistency of CT images is from both the data acquisition and the image reconstruction. The x-ray source covers all body width with the fan-beam or cone-beam geometry at the image acquisition. After data acquisition, numerical computations are carried out to reconstruct images from collected projection data [19].

2.2.2 CT Numbers

In this thesis, the term "CT numbers" refer to the gray-levels on the original CT images while the word "intensity" and "gray-levels" are used interchangeably to indicate the gray-levels on processed images. CT numbers are measured in Hounsfield units (H) to measure the x-ray attenuation after going through materials. A change of one Hounsfield unit corresponds to a change of 0.1% of the attenuation coefficient of water [20]:

$$H = \frac{\mu(tissue) - \mu(water)}{\mu(water)} \times 1000 \tag{2.1}$$

The CT number of water is defined as zero: $H(water) = 0$. All major manufacturers claim

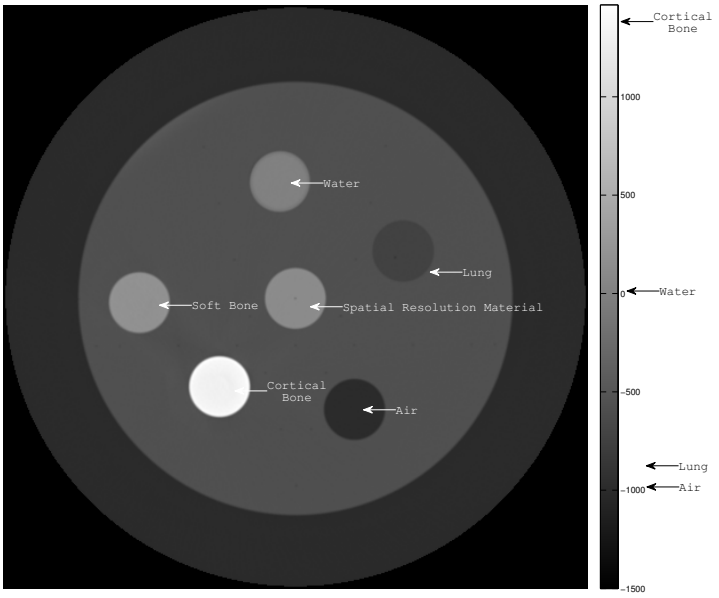


Figure 2.1: *Six-material Calibration containing air, water, soft bone, cortical bone, lung and a spatial resolution material. The mean of CT numbers of each material is indicated by arrows, while the variance should be controlled within a tight interval.*

essentially equivalent performance figures, so that the CT numbers produced by different equipment are consistent and tightly controlled. The control can be achieved by scanning a calibration phantom containing six different materials: air, water, soft bone, cortical bone and lung as shown in Figure 2.1. The mean and variance of the CT numbers within each material should fall in a tight interval. For example air has the average CT number of $H(air) = -1000$ while cortical bone has an average CT number of $H(bone) = 1000$. The consistency of the CT numbers also serve as a very good basis for inter- and intra-patient image analysis and comparison.

The characteristics of CT bring many advantages for radiotherapy planning. Firstly, the bones on CT images are bright and easy to identify, so they can be used as rigid landmarks. Moreover, since the CT numbers account for the attenuation of the x-ray, it can be easily used to derive the electron density information for accurate radiation dose calculation [21, 22]. Other modalities such as ultrasound or MR do not provide such density information. However, CT images contain limited soft tissue contrast, which has resulted in significant inter- and intra-clinician variability being reported for delineation of the GTV and organs at risk [14–16]. In pelvis organ segmentation, the lack of soft tissue contrast is a very significant source of such variability, which will be addressed in Section 2.3.

2.2.3 Noise on CT Images

CT images suffer from non-stationary, non-Gaussian and correlated noise [23, 24]. Noise may be from either a physical process such as beam hardening, electronic noise, or computational process such as round-off errors in the reconstruction process. It is reported that the radiation dose from the x-ray source has the most significant influence on the noise level, since the major cause of noise is the statistical inaccuracy of a limited number of x-ray quanta received by the detector array [23]. Increasing the x-ray dose will in general improve the signal to noise ratio, but meanwhile it will cause more damage to the patient. One solution to this is to adopt advanced denoising algorithms that maintain image quality whilst minimizing dose.

Denoising methods vary when different body parts are scanned. The purpose of denoising is to reduce the amount of random variations and artefacts, making the image smooth whilst preserving edges. For measuring noise the calibration phantom shown in Figure 2.1 is scanned. The smoothness of a homogeneous region is measured by the standard deviation within the region, while the sharpness of the edge can be measured by the modulated transfer function (MTF) derived from the spatial resolution material [20].

One particular "noise", called structural "noise" in [23], is of particular interest. It refers to the density variations in the object being imaged which may affect the diagnosis. While artificial shadows may cause false diagnosis, it is not always correct to consider the variation in a homogeneous ROI as noise. The spatial variation may form important texture yielding diagnostic information. The key purpose of this study is to examine the statistical properties to see if they are useful to define the GTV for bladder cancer radiotherapy.

2.3 Anatomy of the Pelvis

In this section, the basic pelvic anatomy relating to genitourinary cancer is illustrated. Anatomical atlases are used to indicate the positions of the organs while the CT images corresponding to the atlases are shown to indicate the resolution and contrast of different organs in CT images. The purpose of this is to identify factors which cause difficulty for the automatic contouring task. The most important organs under consideration are the bladder, prostate and rectum.

Because cancer can grow beyond organ boundaries, in genitourinary cancer treatment usually one main organ with its immediately adjacent cancerous structures may be considered as GTV,

whereas the healthy organs close to the GTV are OARs. For example in bladder cancer radiotherapy planning, due to the limited resolution the whole bladder is usually defined as GTV which may sometimes include part of prostate, whereas the rectum is considered OAR and should be protected from radiation if possible. Since soft tissues are of poor contrast in CT images, identifying the cancer tissue automatically requires distinctive features. This is especially difficult at the bladder-prostate junctions where little intensity information indicates the boundary of the cancer growth.

2.3.1 Coordinate System

The superior-inferior, anterior-posterior and left-right axes systems, together with three planes: axial, sagittal and coronal plane are most commonly used coordinate systems in anatomical work. Figure 2.2 offers an illustration from [4] indicating the axes system and the three planes. The axial plane expands from head (superior) to foot (inferior); the sagittal plane runs from one side of the body (say, left) to the other side (right) and the coronal plane spreads from the back (posterior) to the chest (anterior). For multi-slice CT images, the axial cross-sectional scans can first be stacked together to form a 3D volumetric representation of the body, then re-sliced into another arbitrary plane for investigation. For CT imaging of pelvic regions, multiple cross-sectional 2D images are taken in the axial plane. By stacking up all the axial images a 3D image can be formed. Re-slicing of the 3D image in sagittal or coronal planes brings different insight and may be helpful in defining target volume.

2.3.2 Bladder

The bladder is a distensible hollow organ in the lower abdomen that is used for storing urine. In the bladder wall there are three layers of tissues [25]:

Inner: also called lining, comprising transitional cells, which stretch when the bladder expanded. It is reported that 90% of bladder cancer originates from the malignant growth of the transitional cells [25].

Middle: comprising strong muscle tissues for distension.

Outer: perivesical fat, fibrous tissue and blood vessels which cover, support and protect bladder.

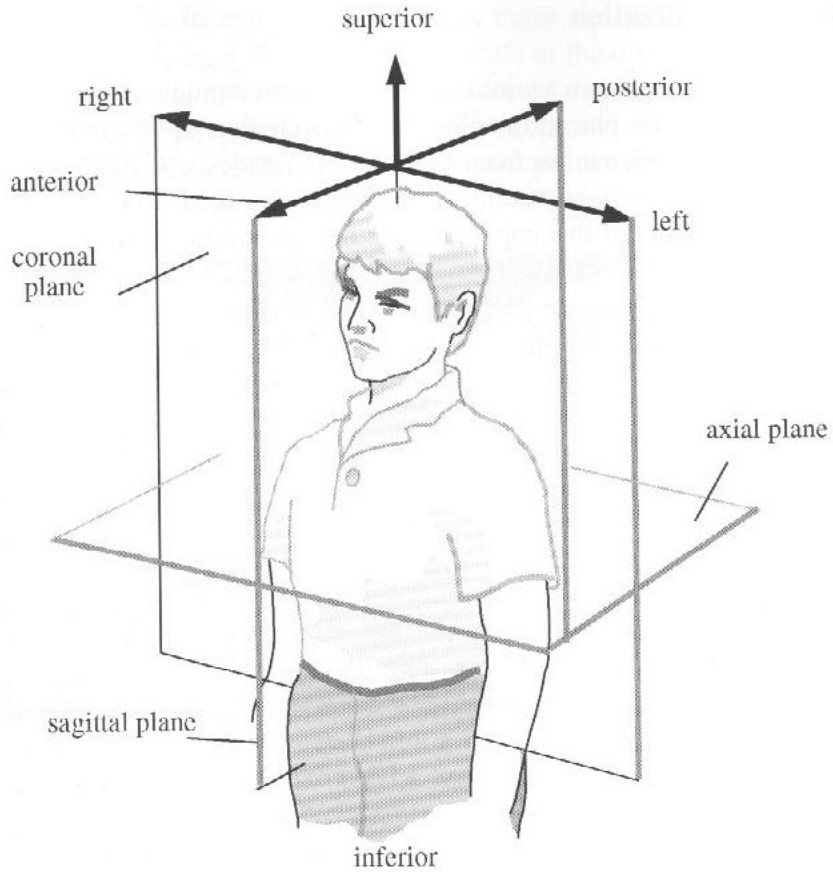


Figure 2.2: *The coordinate system for anatomy structure from [4]. There are three pairs of directions: superior-inferior (head to feet), anterior-posterior (chest to back) and left-right. Three planes as indicated: Axial, sagittal and coronal. The images can be stacked together to form a 3D image and the 3D image may then be re-sampled into arbitrary planes.*

The anatomy of bladder differs from male to female. Figure 2.3 shows the male (a) and female (b) pelvic anatomy in the sagittal plane [5]. For males, peritoneum covers the superior half of the bladder. The two-gland seminal vesicles are posteroinferior to the mid range of bladder. The inferior part of the bladder is immediately adjacent to the prostate. Figure 2.4 (a) shows the axial anatomy of the male urinary bladder at the urethral orifice level from [5]. Figure 2.4 (b) is a CT image at the corresponding level where the seminal vesicles attach to the bladder. The three layers of tissues cannot be resolved in CT images, which limits the diagnostic accuracy.

In female however the bladder is situated adjacent to the uterus and vagina. The left of Figure 2.4 (c) cites the sectional anatomy of the female urinary bladder from [5] while the right of the

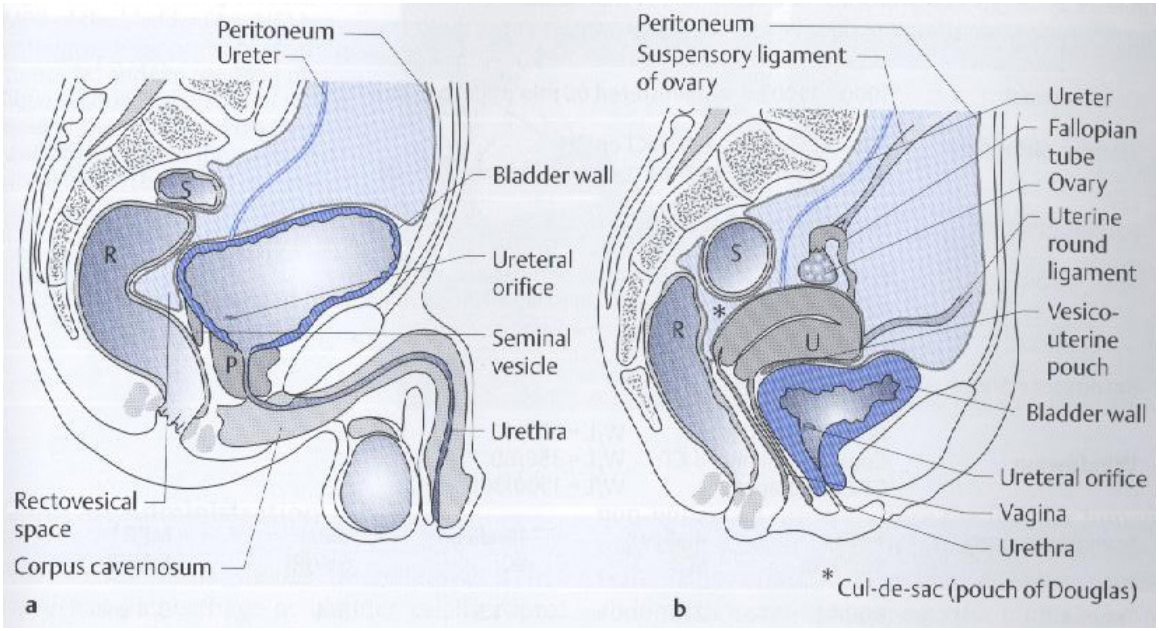


Figure 2.3: *Pelvic anatomy in sagittal plane from [5]: (a): male while (b): female. For male the inferior base of the bladder is adjacent to prostate (P). Rectum (R) is directly posterior to the bladder. For female the bladder is directly inferior to the uterus (U). In superior range, the cervix is directly posterior to the bladder while in the inferior range, the vagina is directly posterior to the bladder. Rectum situates posterior to the vagina.*

Figure 2.4 (d) shows a CT image at the same level. Soft tissue organs such as cervix and vagina have similar intensity to the bladder.

2.3.3 Prostate

Prostate is a doughnut-shaped organ comprised of five lobes: two lateral, one anterior, one middle and one posterior [26], which appears only in male anatomy. Urethra goes through the prostate. The base of the prostate is attached to the inferior end of the bladder, and there is no fat plane between the two organs. The anatomy of bladder-prostate junction is illustrated by Figure 2.5 (a) [27], while Figure 2.5 (b) shows a CT image corresponding to the same level. It can be seen that in CT images due to limited contrast it is difficult to define the boundary of the two organs such as bladder and prostate; it is even more difficult to define the extent of any cancerous growth.

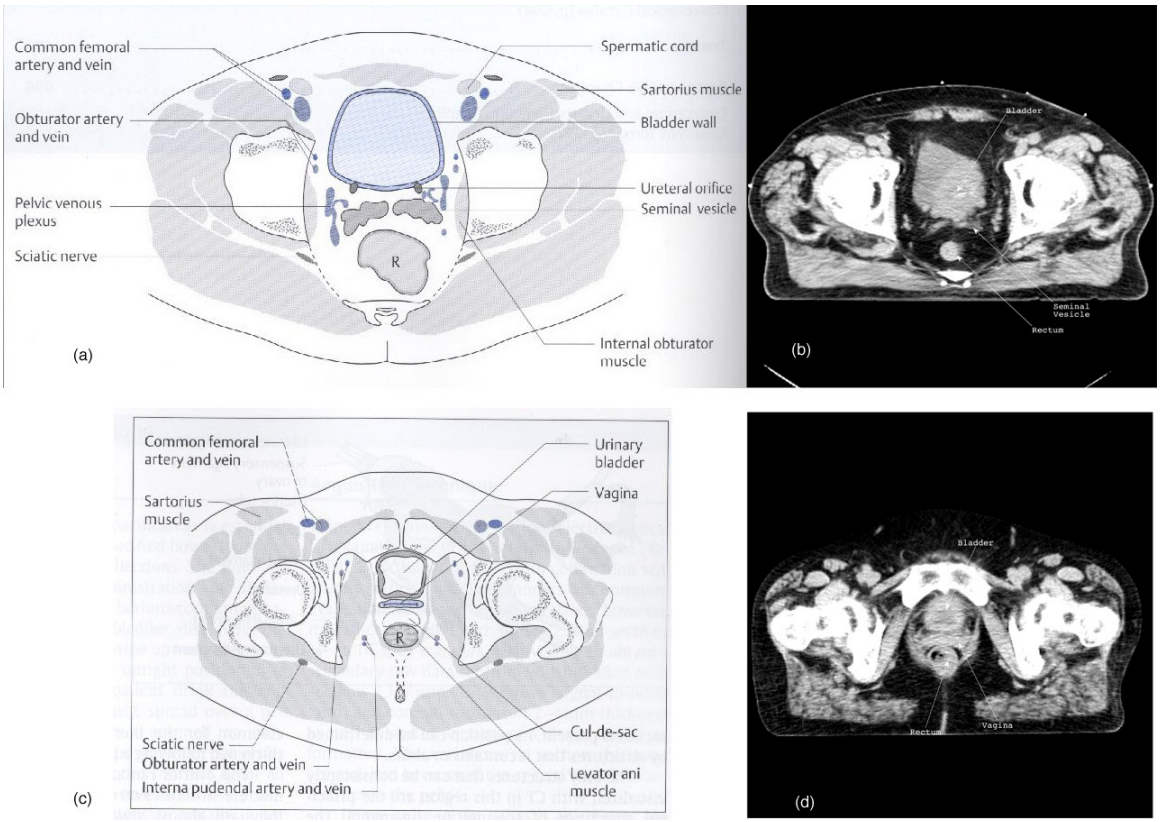


Figure 2.4: *(a): The axial anatomical atlas of the male urinary bladder at the urethral orifice level from [5]; (b): The CT image from a male bladder cancer patient at the level corresponding to (a); (c): The axial anatomical atlas of the female urinary bladder from [5]; (d): The CT image from a female bladder cancer patient at the level corresponding to (d).*

2.3.4 Rectum

Rectum is the last section of the gastrointestinal tract, which is typically 20cm. It is posterior to bladder. Due to the randomness of its filling and movement the shape and position of rectum varies considerably in different CT scans. Its texture is heterogeneous.

The physical movement and distension of the rectum is very frequent, which may change the position of the GTV and OAR, leading to potentially less-than-optimal treatment outcome in radiotherapy [28]. Unlike respiration movement which is regular and predicable, the movement of the rectum is irregular and unpredictable. To take all possible motion into account, the PTV is defined which comprises a larger volume. If accurate information of tumour site can be obtained, the PTV definition may be significantly reduced. This is the merit of IGRT which is discussed in Section 2.5.

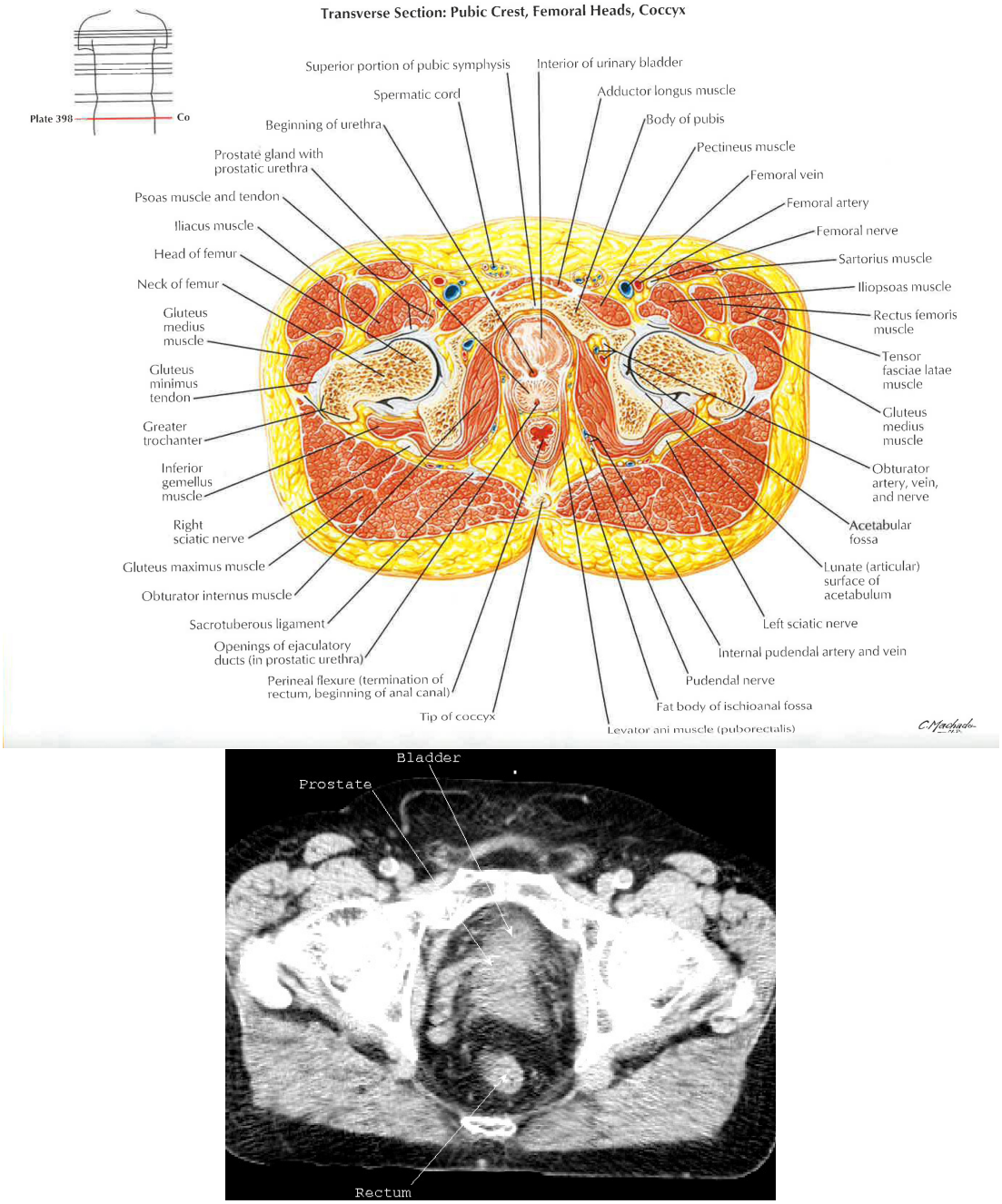


Figure 2.5: *(a): The atlas of bladder-prostate junction from [27]. (b): The CT image corresponds to the position of the atlas. On the CT image it can be seen that the soft tissue resolution and contrast is limited, and the boundary of bladder and prostate cannot be seen clearly. To define the extent of tumour growth requires considerable experience, and is subject to significant inter- and intra-clinical variability.*

2.4 Treatment Procedure of Bladder Cancer

The procedure for treating genitourinary cancer is: diagnosis, grading, staging, planning and dose delivery. In this section the main steps and the major choices in these steps are introduced based on the descriptions in [25, 29]. The emphasis is on the role of medical imaging in this procedure and on the accuracy and limit of each imaging modality.

2.4.1 Diagnosis

The appearance of blood in urine, or other symptoms such as abnormal urinary activity may lead to clinical examination. For bladder cancer diagnosis the urine test, cystoscopy and/or biopsy may be used. The urine test can be conducted under a microscope to determine whether or not cancer cells are present. Cystoscopy uses a long, thin and flexible tube with a camera and the light source inserted through the urethra to look directly into the bladder for evidence of cancer. If the presence of cancer is suspected, a biopsy may be operated, which takes samples of tissue from the bladder with the cystoscopy for pathological investigation. The pathological examination of the sample tissue is the only sure way to diagnose whether the growth is benign or malignant.

2.4.2 Grading

Once the cancer is diagnosed the grading is carried out to compare the cancer cells with ordinary healthy cells. The less similar the cancer cells to the healthy cells the more invasive they are, and the faster they grow. The treatment options will take into account the tumour grade. Both diagnosis and grading are based mainly on pathology.

2.4.3 Staging

After diagnosis staging of cancer is carried out to determine the extent of the cancerous growth: whether it has remained primarily in local area or invaded nearby tissues such as muscle, or whether it has spread to other organs. Currently the most popular staging system is the TNM system (tumour, node, metastasis). The T category cancer is still in the local area of primary tumour, N category cancer indicates enlarged lymph nodes while M category cancer refers to remote metastasis. T category can be further classified into five stages: Ta, T1, T2, T3, and T4.

For T2, T3 and T4, there are two sub-stage: a and b. The descriptions are given as follows [10]:

Ta: The tumour is within a small surface area of the inner layer of bladder.

T1: Tumour invades the subepithelial connective tissue as shown in Figure 2.6.

T2: Tumour invades muscle. a: inner half; b: outer half.

T3: Tumour invades perivesical fat as shown in Figure 2.6. a: microscopically; b: macroscopically.

T4: Tumour invades the adjacent organs: prostate, uterus, vagina, pelvic wall and/or abdominal wall. a: prostate, uterus, vagina; b: pelvic wall, abdominal wall.

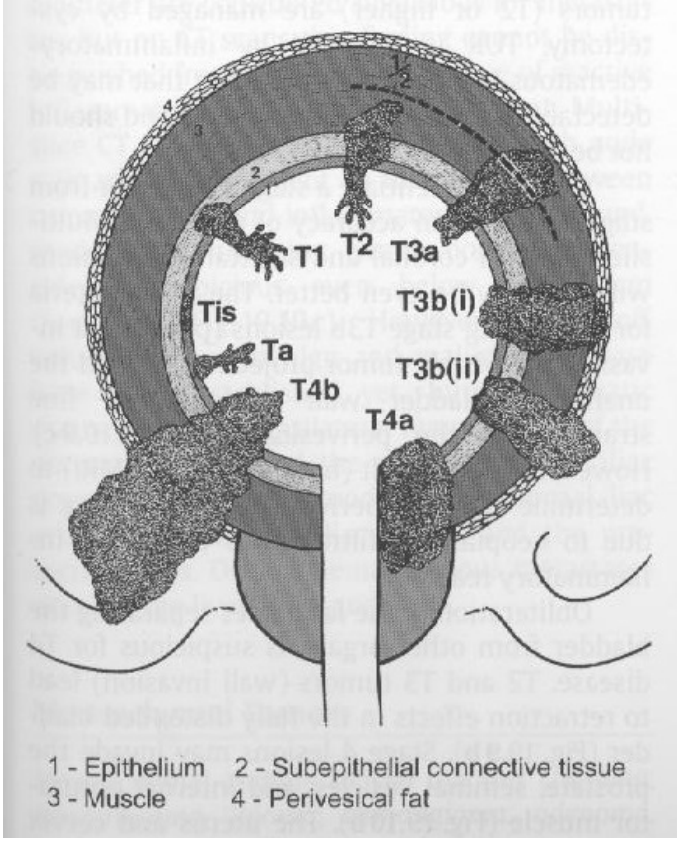


Figure 2.6: *T category of TNM transitional cell malignant growth code in bladder wall from [6]. The tumour is invasive, which can grow without boundary into healthy tissues. Depending on its extent of growth different codes apply as descriptors.*

The corresponding illustration is given in Figure 2.6 [6]. The N (node) category stands for lymph node enlargement while the M (metastasis) category denotes the distant metastasis of cancer cells.

Trans-urethral resection (TUR) is used to acquire a localised specimen for cancer staging. Current medical imaging modalities do not provide resolution high enough to distinguish superficial, early stage cancer such as Ta, T1 and T2 cancer. Therefore for early-stage cancer, which is confined to the local tumour site, clinical staging is superior to medical imaging staging [30]. However for invasive cancer, the insight provided by clinical evaluation is limited, and the extent of the tumour cannot be accurately defined. In this case medical imaging will give better performance.

Medical imaging provides invaluable insight into human anatomy, and is intensively used for cancer staging. Traditionally the diagnostic precision of the imaging modality is described by the sensitivity and specificity of the diagnosis by examining images from different modalities. Sensitivity is the true positive probability ratio defined by:

$$sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{2.2}$$

while specificity measuring the true negative fraction is defined by:

$$specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \tag{2.3}$$

The most commonly used modalities for staging are ultrasound, CT, and MRI. Different imaging modalities offer different image qualities in terms of resolutions, tissue contrasts and geometries, at different costs. The details of the use of imaging for cancer staging is summarized below:

Ultrasound: the resolution and contrast are the poorest of the three modalities. It can show the spread of cancer to other organs, i.e. ultrasound imaging can only detect N category or M category cancer. It is also the cheapest modality of the three.

CT: the soft tissue contrast is higher than ultrasound imaging yet not as clear as MRI. Layers of the bladder wall cannot be distinguished on CT images and therefore cannot distinguish T1 and T2 and T3 cancer. Can differentiate the growth through the bladder wall, i.e. T3a and T3b cancer with high specificity [31]. However, whether the cause of the growth is inflammation or cancer cannot be differentiated from the image property, leading to poor sensitivity [32]. CT is less expensive than MRI but more expensive than

ultrasound imaging;

MRI: the soft tissue contrast on MRI is higher than on CT. Yet MRI cannot resolve superficial tumours in muscle layers. Moreover it also suffers from the same problem of low sensitivity. No gross dissimilarity is reported in the staging accuracy of CT and MRI [32, 33]. The cost of MRI is the highest of the three modalities.

Given these factors, the CT image is now used as the main modality for staging [34]. This study will focus on defining the the T category, local primary tumour volume.

2.4.4 Choices of Cancer Treatment

The options available for the treatment of the cancer are: surgery, radiotherapy, chemotherapy and biological therapy, or a combination of these [25]. The decision of therapy is based on the type of the cancer, the grading and the staging results.

Surgery includes TUR, partial and radical cystectomy. For approximately 75% superficial bladder cancer, TUR is carried out to remove the tumour by inserting the a resectoscope into the bladder via urethra, and operating with a resecting loop until healthy muscle can be observed. For more advanced cancer radical cystectomy is used as a standard in the America. It is a major operation which removes the entire bladder, nearby lymph nodes, and any potentially affected organs: prostate, seminal vesicles for male, or urethra, anterior vaginal wall, cervix for female [25, 29]. It is likely that the cancer will recur after surgery, so chemotherapy and radiotherapy may be used after the surgery to further control the growth of cancer cell.

Radiotherapy uses radioactive rays to kill cancer cells by damaging its DNA. Recent advances in radiotherapy make it capable to deliver highly concentrated dose to the tumour cite while sparing the healthy tissue. As a result, radiotherapy becomes a standard for bladder cancer treatment in many countries. Since this study focuses on the contouring of target volume for radiotherapy, the procedure of radiotherapy will be illustrated in detail in Section 2.5.

Chemotherapy uses drugs to kill cancer cells. It remains a controversial in treating bladder cancer and is not considered as a standard [29].

Biological therapy refers to enhancing the immune system of the body to eliminate cancer cells. For example, liquid containing weakened bacteria may be applied to the bladder weeks after TUR to aid the immune system in killing cancer cells [25].

2.5 External Beam Radiotherapy

The most frequently applied radiation is external beam radiotherapy, which uses a linear accelerator to deliver a high dose of radiation to the tumour volume. Modern radiotherapy comprises planning, simulation and radiation delivery, and quality control [1]. In planning the extent of tumour(s) is defined as the radiation target (GTV, CTV and PTV); after the target is defined a dose delivery calculation will be made based on the purpose of the treatment, for example curative or palliative. After the plan is finalised a simulation is carried out to verify the radiation setup, and the dose will subsequently be delivered according to the plan. After the physical dose delivery quality control needs to be implemented to assure that the outcome of the treatment meets the expectation. Since the thesis studies image analysis for GTV definition, in this section radiotherapy planning is detailed, with an emphasis on the impacts and challenges brought by the recent introduction of IMRT and IGRT.

2.5.1 Target Localization

Presently target localisation is carried out manually by experienced clinicians who contour the structures of interest on medical images [17]. The term "contouring" refers to the definition of the tumour as the radiation target and the organs at risk that must be avoided. As is mentioned in Section 1.2, in Report 50 ICRU suggested the targets of contouring [1] to be: gross tumour volume (GTV), clinical target volume (CTV), planning target volume (PTV) and organs at risk (OAR) which are described below:

GTV : The palpable/demonstrable extent and location of the malignant tumour. It is the primary tumour judged by purely anatomical and biological factors independent of treatment modalities. If surgery is conducted to remove the tumour prior to radiotherapy no GTV exists when contouring.

CTV : The volume containing both GTV and sub-clinical microscopic malignant growth. Because the microscopic growth information cannot be obtained from medical imaging, it is particularly difficult to define the CTV. According to this definition CTV should not take tumour movement into account.

PTV : The static, geometrical volume considering all possible geometrical variations and inaccuracies to make sure that CTV actually absorbed the prescribed dose.

OAR : The healthy tissue that must be avoided or at least the dose minimised as much as possible.

The contouring process defines the tumour volume to irradiate. At present because CT images give undistorted geometry, electron density information for radiation dose calculation and easy-to-identify landmarks for accurate set-up, CT images are used as a standard in bladder cancer contouring.

Both GTV contouring and cancer staging are used to define the extent of the cancer growth, and are mainly based on CT images. However they are used for different purposes: staging for deciding the treatment choice, while GTV contouring for radiotherapy target definition. This leads to several differences: staging is a qualitative description of the cancer growth. When taking images for staging purposes, a patient requires a full bladder with either urine or contrast agent to boost the visual characteristics of the tumour. The GTV contouring on the other hand quantitatively define the exact volume of primary tumour, which calls for high geometric accuracy. The distension of the bladder will reduce such accuracy. Therefore when taking images for contouring purpose, a patient is supposed to have an empty bladder. The reported accuracy of staging however can be considered as an indicator of the contouring. However without the help of contrast agent, the poor resolution and contrast of soft tissue on CT images make the contouring a very challenging task. As is shown previously, organs such as bladder, seminal vesicles, prostate and rectum have very similar gray levels. The extent of the GTV is therefore difficult to identify, which causes the inter- and intra-clinical variability. In order to reduce such variability, distinctive features providing objective assessment of the extent of the tumour growth could be very helpful.

2.5.2 Dose Calculation and IMRT

After defining the radiation target whereby a maximum dose of radiation is delivered to the PTV whilst sparing the OAR. The most popular radiotherapy is the 3D conformal radiotherapy (3D CRT), especially the IMRT, which enables very accurate dose delivery [35, 36]. The 3D CRT distribute the prescribed dose conforming as close as possible to the 3D anatomical structure. This is achieved by the advanced treatment planning algorithms as well as the radiation equipment. IMRT is an approach to delivering conformal radiotherapy. By modulating the fluences of the radiation, each beam of X-ray may have non-uniform intensity. As a result

the distribution of the dose delivery can be greatly improved. With IMRT it is even possible to irradiate a concave shape, which is not possible by a conventional means. The high accuracy of IMRT requires accurate target volume definition to avoid the injury of healthy tissue whilst maximising the dose to the tumour. As is mentioned in Section 2.3 during the radiotherapy planning period patient movement and internal organ movement changes the position of the organs. As a result the PTV must take this possible movement into account by including a margin, which reduces the accuracy improvement brought along by IMRT.

2.5.3 Simulation and Dose Delivery and IGRT

Recently IGRT was introduced to reduce the uncertainty of PTV position during the course of the radiotherapy. The basic idea of IGRT is to acquire CT images shortly before and during radiation delivery to instantly define/adapt to the PTV. This can lead to a much smaller PTV definition than traditional methods. However, it is obviously impractical to have clinicians available for contouring during the radiation delivery. Therefore a key component of a successful IGRT scheme is to utilise the imaging information to re-plan, and therefore optimise, a treatment.

2.5.4 Discussion: New Challenges Brought by IMRT and IGRT

IMRT can give very accurate dose delivery to the physical site which is desirable in improving the accuracy and precision. However the increased accuracy also suggests that IMRT is more sensitive to position errors than traditional 2D or 3D dose delivery methods [37]. Uncertainties in PTV position may lead to missing the radiation target, and therefore the failure of the treatment. If a larger-than necessary volume is included in PTV definition, the merit of the IMRT is compromised. IGRT can work with IMRT to reduce the position uncertainties. Coupled with other motion management methods IGRT can give more accurate target definition during the course of the treatment.

Contouring is carried out by experienced oncologists. The subjectiveness of human experts introduces significant inter- and intra-clinical variabilities of contouring [14–16]. The reproducibility of the contour is also a problem. In order to decrease the inter- and intra-clinical variability, more distinctive features from the CT images are desirable in order to demonstrate the objective extent of the tumour. Image analysis algorithms, especially low-level feature ex-

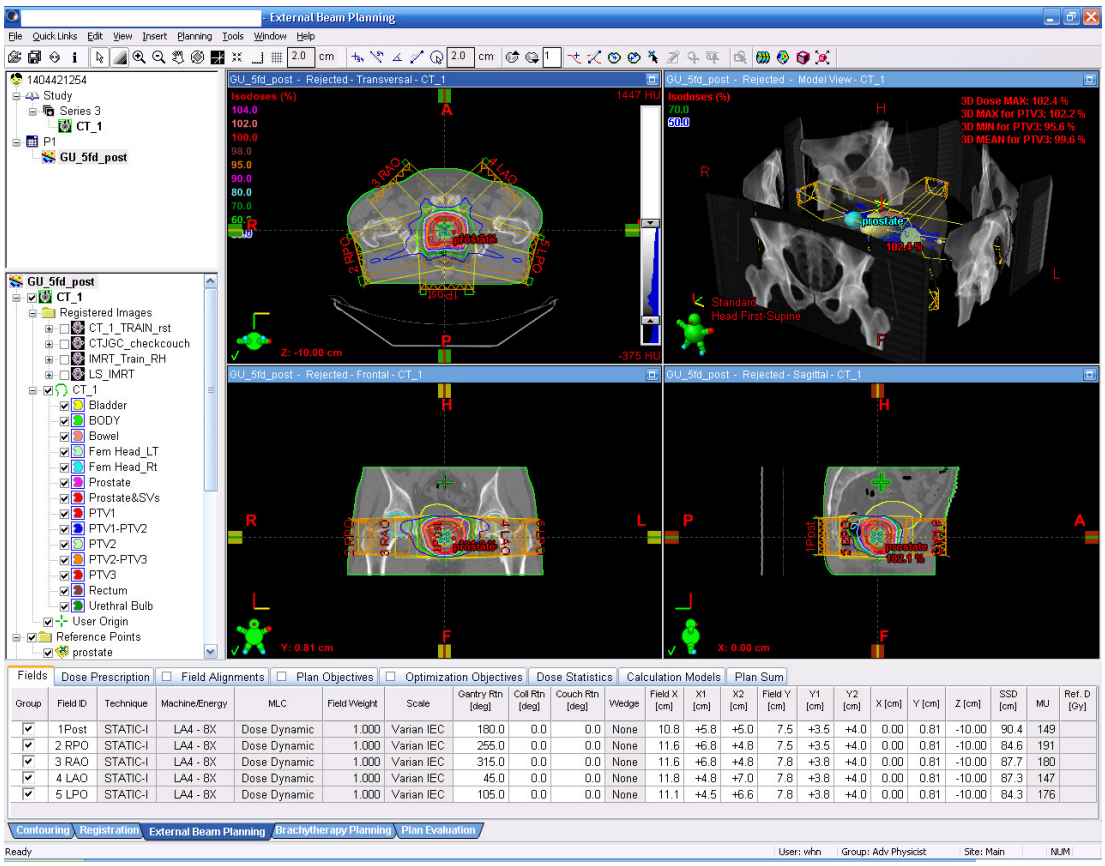


Figure 2.7: *A 3D radiotherapy planning application for IMRT. The contoured CT pelvic images in three planes: axial (upper left), coronal (lower left) and sagittal (lower right) as well as a 3D visualization of the organs involved. from different angles focus on the PTV while sparing the OAR. For traditional 3D radiotherapy the intensity of each beam is uniform and as a result, the accuracy is limited. IMRT uses non-uniform, modulated fluences of rays (photon, electron or proton) to irradiate the target volume(s), and as a result the resolution of dose delivery is significantly increased. The accuracy of the IMRT makes target delineation having more important because it can influence the outcome of the treatment.*

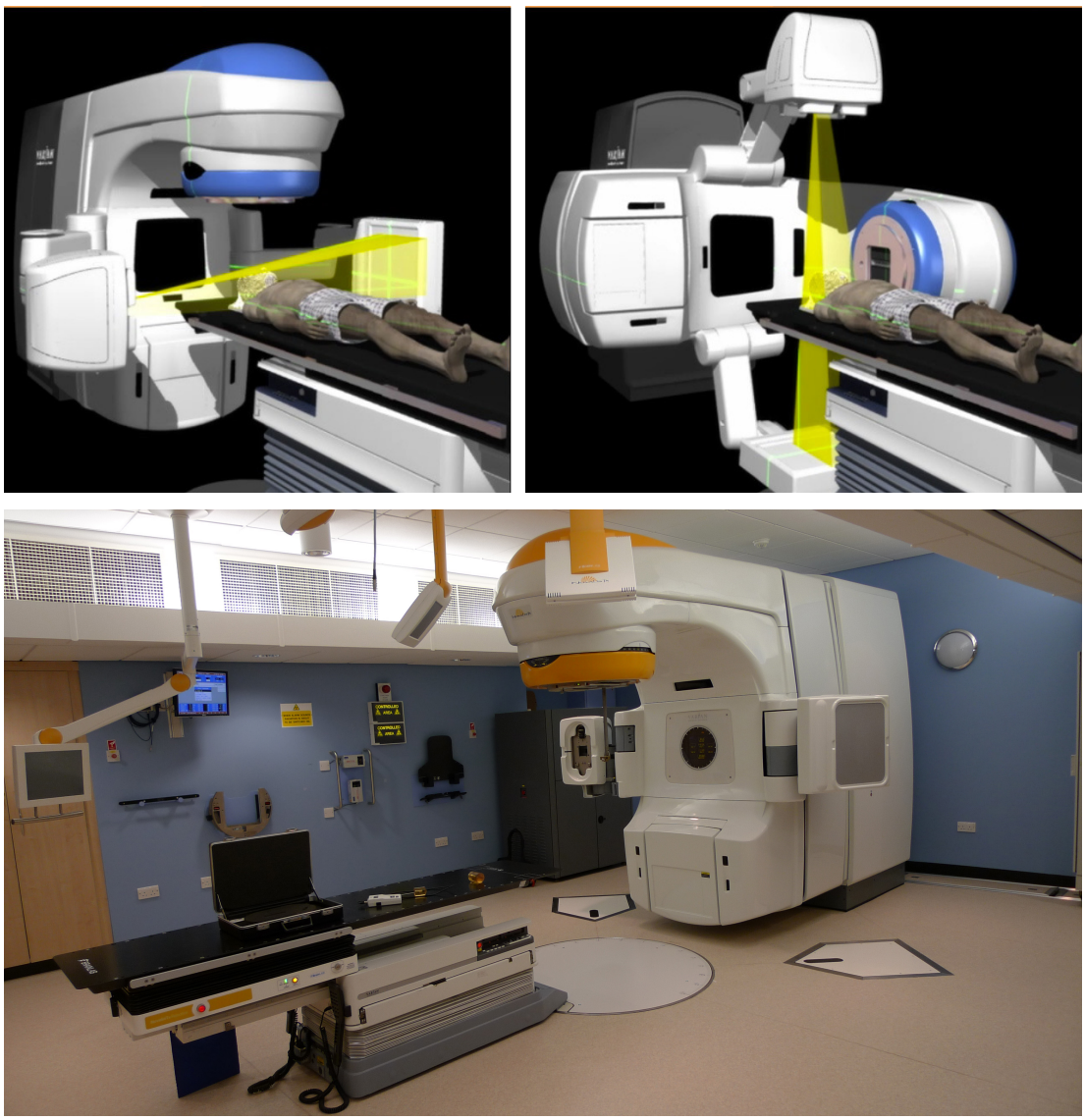


Figure 2.8: *Image guided radiotherapy (IGRT). Upper left: A set of cone beam CT imaging is operating during the radiation to guide the radiation. Upper right: radiation beam for dose delivery. Lower: a linear accelerator with IGRT for radiation delivery in clinical practice. The PTV takes into account of all possible tumour locations due to patient movement. Due to the patient movement at the time of treatment, the tumour volume may move to another location. PTV has to include all possible tumour sites, leading to a larger-than-optimal PTV definition. For IGRT (right PTV), By using imaging as a guidance during the treatment time, the GTV and CTV can be defined instantly and as a result, the accuracy of PTV definition is increased.*

traction can be applied to fulfil this task.

Another problem is, as the number of images produced by the multi-slice CT scanner is large and increasing, the contouring is very time-consuming. For example, research conducted by University of California, San Francisco (UCSF) on the planning/treatment time indicates that for complex head and neck cancer contouring, it takes two hours per case, which remains almost unchanged from 2002 up to 2007 [17]. This is partly because no algorithm can yield satisfactory contours: during the treatment planning period, lots of factors such as patient movement or internal organ movement may change the position of the tumour. By including all geometrical variations of the volume in the PTV, the irradiated volume is larger than optimal.

One possible solution, the image guided radiotherapy (IGRT), is proposed to reduce the inaccuracy of PTV due to the location variation of the tumour, which uses the medical imaging as a guidance/feed-back at the time of irradiation to locate the tumour site. By confining the PTV to a smaller focal volume during treatment, the geometrical accuracy of PTV can be improved. However, since it is not possible for clinicians to be present at every treatment, automatic contouring is highly desired.

2.6 Summary

In this chapter the medical context is given to address the importance and difficulties of developing automatic contouring algorithms for radiotherapy planning. The advantages and disadvantages of CT imaging for radiotherapy planning are described. Anatomy of human pelvis is illustrated, with an emphasis on the bladder-prostate junction. The choices of cancer treatments are also listed. The procedure of external beam radiotherapy is detailed. It can be inferred that distinctive and objective features can help defining the GTV and reducing the inter- and intra-clinical variability, decreasing the time consumption for treatment planning which not only has significant influence on the productivity of the radiotherapy planning, but has great potential to be part of IGRT process to increase the accuracy of PTV definition.

Chapter 3
Methodology and Literature Review

3.1 Introduction

In Chapter 2 the importance of and difficulty in developing algorithms for automatic GTV contouring in radiotherapy planning was addressed. One of the most significant constraints in developing an automatic pelvis organ segmentation algorithm on CT images is the limited resolution and contrast in soft tissue organs, as is illustrated in Section 2.3. In this chapter the major methodologies of medical image processing are addressed, and the existing system which uses synthesised multiple image processing methodologies proposed in the literature for pelvis organ segmentation/classification is reviewed.

Depending on the level of abstraction, the segmentation/classification methodologies can be divided into **low-level feature extraction** and **high-level modelling**. Low-level feature extraction explores characteristic features, such as statistics, gradients, etc. to describe the properties of the images. All the descriptive features are called low-level features. High-level modelling on the other hand assumes that the features are generated by a model, which is controlled by several parameters. The model can be considered as an abstraction of the image properties based on the the low-level features, therefore the parameters of the model are called high-level features.

In the existing literature a lot of effort has gone into developing algorithms which enable automatic pelvic organ contouring. Most existing algorithms are based on high-level modelling which incorporate organ shape information into simple low-level features to form an optimisation problem, which usually converges to a local minimum, meaning that given shape priors and the low-level features the models can give sub-optimal contours. These models make a trade-off between the prior knowledge and the low-level feature observations, which is essentially an educated and balanced guess. In the literature high similarity between clinical contours and model-generated automatic contours is reported [2, 38–40]. However the optimisation can only be as good as the prior knowledge and the low-level features. Given the argument that

significant inter- and intra-clinical variability exists among experienced oncologists, the prior knowledge used by the shape model cannot be considered as fully persuasive.

Yet relatively little attention has been given to extracting significantly distinctive low-level features which may serve as a new basis of the contouring. The core of this thesis is to investigate the textural features from CT images and their applications in automatic contouring. In this chapter the major methodologies for organ segmentation are classified into low-level feature extraction and high-level modelling and are reviewed separately. Section 3.2 reviews the low-level feature extraction methodologies in the spatial domain. The statistical features are emphasised. Section 3.3 discusses the low-level features in the transform domain, especially in the wavelet-scale domain. Subsequently in Section 3.4 the high-level modelling methodologies are categorised into classification based, knowledge based, morphology based, registration based and model based approaches. In implementing a systematic segmentation application, multiple methodologies are often jointly used, the major automatic systems for pelvis structure contouring proposed in the literature are reviewed in Section 3.5. A discussion is given in Section 3.6 to confirm the importance of the low-level features in pelvis structure segmentation problems. Section 3.7 summarise this chapter.

3.2 Low-Level Feature Extraction in Spatial Domain

Feature extraction is usually the first step in image analysis. Without distinctive features, the best result an algorithm can produce only amounts to a best-effort guess. The most heavily used features are mainly from two categories: statistical features and edge/gradient features. In this section the statistical features are described. Statistical features were heavily studied around late 1970s to characterise textures. The statistical feature extraction methodology first builds the distribution of images. Depending on how many pixels are involved in building the distribution, features are categorised into first-order features from gray-level histogram, second-order features from co-occurrence matrices which involves the joint distribution of two pixels, and higher-order features from run-length matrices with the joint distribution of more than two pixels. Statistics such as moments, entropies and/or correlations are extracted from the distribution as features.

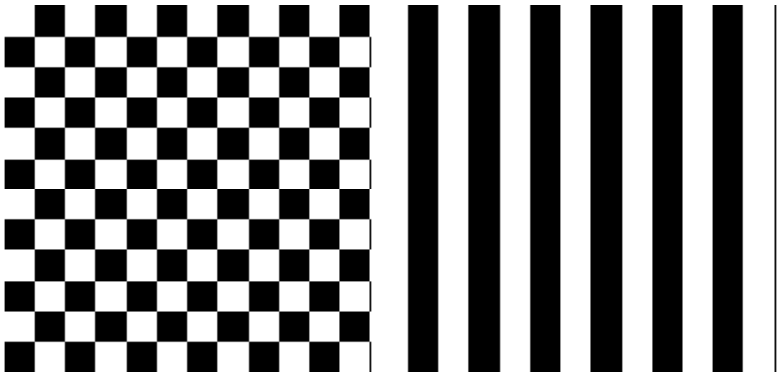


Figure 3.1: *Two different textures with the same first order distribution:* $p(0) = p(1) = 0.5$*. In this case the first-order statistics cannot differentiate the two images.*

3.2.1 First-order Distribution

Suppose $i(x, y)$ is the gray-level of one pixel from an image $\mathbf{I}$ with indexes x and y. The distribution of intensities $i(x, y)$ is described by $p(i)$. Because $p(i)$ involves only single pixels, it is referred to as a first-order distribution. Consider an image as a set of random variables, given one single implementation $\mathbf{I}$ it is very difficult to obtain the true distribution of each random variable. However, a crude estimation can be carried out by assuming that the random variable generating the image is ergodic (which means the statistical property of the signal can be completely characterised by one sufficiently large sample process, see [41] for detailed explanation). By using the intensities of all pixels in an image an estimation of $p(i)$ can be derived. There are parametric and non-parametric methods to estimate $p(i)$ [42, p. 67]; the easiest and most widely applied method is to calculate normalised histogram. First-order distribution is easy to implement and fast to calculate and is therefore widely used as the low-level feature.

However, one obvious drawback of the first-order distribution for image processing is that it does not take into account the spatial relation of different pixels. For example in Figure 3.1 two images are shown: they both are half-black, half-white. Their first-order distribution are identical. However the left texture is a zebra-like pattern whereas the right one is a check-box pattern: they are not of the same texture. When cases like in Figure 3.1 occur, it is insufficient to use first-order distribution. Second-order and higher-order distributions need to be considered.

3.2.2 Second-order Distributions

The term second-order distribution refers to the joint distribution of two random variables $p(i, j)$. The gray-tone spatial dependence matrix (GTSDM) as is originally proposed by Haralick *et al.* [43] for texture analysis. According to Haralick *et al.* , the co-occurrence $c_{d,\theta}(i, j)$ is defined as the number of pixel pairs with gray-levels of two pixels i, j with distance d from each other in direction $\theta = 0°, 45°, 90°, 135°$. It can be used as an estimation of the joint distribution of two pixels at a particular direction and distance (hence the name second-order distribution) $p(i, j)$:

$$\mathbf{p}_{d,\theta} = \frac{\mathbf{C}_{d,\theta}}{n^2}, \tag{3.1}$$

where n is the size of the image. The element $p_{d,\theta}(i, j)$ corresponds to the co-occurrence number $C_{d,\theta}(i, j)$. If the co-occurrence matrices with the same d but different θ are averaged, the $\mathbf{p}_{d,\theta}$ reduces to $\mathbf{p}_d(i, j)$ which provide 2D directionally invariant estimation of the joint distribution of two pixels with distance d.

Second order statistics are very widely used in cancer classification. Kale et al. [44] used co-occurrence matrices in three spatial dimensions and one spectral dimension to detect and classify breast cancer in contrast enhanced MR images. Yu et al. [45] applied texture analysis to fluoro-deoxy-glucose (FDG) positron emission tomography/computed tomography (PET/CT) images to classify different regions of interest (ROI) for head and neck cancer radiation treatment planning and reported a high accuracy from the area under the receiver operating characteristics (ROC) curves. Philips et al. [46] explored three dimensional (3-D) directional invariance co-occurrence matrices to classify liver and non-liver volumes in CT images, suggesting that texture classifiers can be used to determine possible tumour volume. Nailon et al. [7, 47] discussed the importance of applying texture analysis for bladder cancer radiotherapy planning on CT images, and showed that it is possible to obtain good distinction between bladder, rectum and a control region containing multiple pathology.

3.2.3 Higher-order Distributions

Higher-order statistics refers to the statistics derived from a joint-distribution higher than second-order of pixel gray levels. The gray-level run-lengths matrices (GLRLM) method as origi-

nally proposed by Galloway [48], also for texture analysis, is a way of quantifying higher-order statistics by counting the number of adjacent pixels in different directions, often in $\theta = 0°, 45°, 90°, 135°$ with the same gray level (termed as the length of a run). Galloway originally defined five features from GLRLM. In subsequent publications several other moments have been defined [48–51]. The entropy however, is not explicitly defined to the GLRLM. However the concept of entropy in terms of average of logarithmic probability can be applied directly to the GLRLM to calculate the entropy. Nailon *et al.* [7] also studied the classification power of the GLRLM, reporting an inferior performance to the GTSDM.

In this thesis three main types of statistics: moments, entropies and correlations are discussed.

3.2.4 Moments

The moments measure the bias of the distribution from one centre, usually the origin or mean. For example the mean, energy, variance, skewness and kurtosis are the most well-known moments. Moreover geometric moments invariant are proposed in the literature as rotation-invariant features, and is used in registration algorithms [52]. Second-order moments from GTSDM such as angular second moment or inverse difference moment are given in Appendix A.

3.2.5 Entropies

The information entropy originally proposed by Claude E. Shannon [53] on the other hand quantifies the degree of non-uniformity of a distribution into "bits". For an image $\mathbf{I}(x, y)$:

$$H = -\sum p(i) \log_2 p(i), \tag{3.2}$$

where i is the image intensity and $p(i)$ is the first-order distribution of i. Image saliency corresponds to the local complexity which can be quantised by entropy [54]. By calculating the entropy from a local region of interest (ROI), the saliency in the feature domain may increase. In a first-order distribution the data are assumed random and the spatial relationship of pixels ignored.

Second-order entropies can be extracted from GTSDM. Features obtained from joint distributions take account of the spatial distribution of the image and therefore usually provide more discriminatory power. In [43] the sum entropy, second-order entropy and difference entropy

were defined:

$$H_{sum} = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log_2\{p_{x+y}(i)\}, \tag{3.3}$$

$$H(i,j) = -\sum_{i}\sum_{j} p(i,j) \log_2\{p(i,j)\}, \tag{3.4}$$

$$H_{dif} == -\sum_{i=0}^{N_g} p_{x-y}(i) \log_2\{p_{x-y}(i)\}, \tag{3.5}$$

3.2.6 Correlations

Correlation is usually carried out between two vectors of random variables. It is well-known that the correlation and power spectrum are Fourier transform pairs, so correlation is a measure of local frequency: the longer the the vector, the more frequency resolution can be obtained from the correlations. In statistical applications with a limited amount of data, such as an ROI of small size, there is a trade-off between the statistical consistency and the length of the vector.

3.3 Wavelet Transform: Applications in Image Processing

Besides the spatial domain, images can be transformed into other domains, that is, be represented by another set of bases. The purpose of a transform is mainly to facilitate the separation of the useful part of the signal in the transformed domain, making a sparse representation of the signal to facilitate the analysis. For example in communication applications the Fourier transform is most heavily used to project a time domain signal into the frequency domain, so that the signal energy can be confined to specified frequency.

In image classification and segmentation the multi-resolution decomposition, or wavelet transform, is the most commonly used transform to derive sparse representation. The wavelet theory was the focus of mathematical/engineering study in late 1980s, which lead to the compressed representation of low-rank textures: after several key publications by Donoho and his co-authors [55–57]. It is found that for a large class of signals their wavelet-domain representation is sparse, and as a result the wavelet transform is widely applied in image compression. In this section a brief description is given to address two aspects of the wavelet transform: the scaling/wavelet function and the feature extraction methods in scale/wavelet domain. Moreover in Appendix B the theory on the optimality of wavelet transform in image denoising is

formulated.

3.3.1 Wavelet Transform

The construction of a wavelet basis is beyond this thesis; interested readers are referred to [58, 59]. In short discrete wavelet decomposition can be implemented by filtering the image with orthogonal filter bank banks followed by downsampling the signal, as is illustrated in Figure 3.2. By coupled design of the filter responses, the signal of finite energy: $I(t) \in \mathbf{L^2}(\mathbb{R})$ can be split into two orthogonal subspaces: the approximation $\mathbf{A}_1$ and the detail $\mathbf{W}_1$. The decomposition can be applied iteratively to the approximation so that: $\mathbf{A}_{s-1} = \mathbf{A}_s \oplus \mathbf{W}_s$, where s is called level of multi-resolution.

For approximation subspace $\mathbf{A}_s$, the basis is called scaling function, which is a dilated and translated version of a mother scaling function ϕ:

$$\phi_{s,n}(t) = \frac{1}{\sqrt{2^s}}\phi(\frac{t-n}{2^s}), s \in \mathbb{Z}, n \in \mathbb{Z} \tag{3.6}$$

By projecting the signal onto the scaling function at level i, the coefficients can be derived:

$$a_s =< I, \phi_s > \tag{3.7}$$

where $<,>$ is the inner product. c_s can be considered as the low frequency components at level s. The level 0 approximation is defined to be the original signal: $I = c_0$.

In the subspace $\mathbf{W}_s$ the basis is called a wavelet function by dilating and translating a mother wavelet function ψ:

$$\psi_{s,n}(t) = \frac{1}{\sqrt{2^s}}\psi(\frac{t-n}{2^s}), s \in \mathbb{Z}, n \in \mathbb{Z} \tag{3.8}$$

The wavelet coefficients can also be derived by projecting the signal onto the wavelet function:

$$w_s =< I, \psi_s > \tag{3.9}$$

w_s represents the high frequency components.

Because digital images are of finite energy:

$$\mathbf{I} \subset \mathbf{L^2}(\mathbf{Z}) \tag{3.10}$$

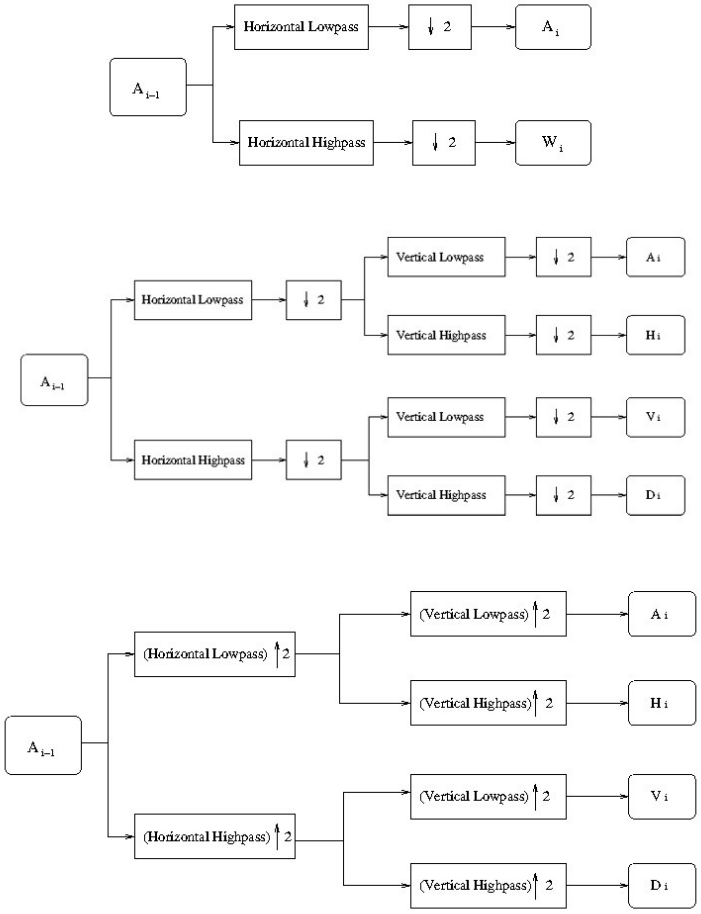


Figure 3.2: Upper*: flow chart for 1D discrete wavelet decomposition, where the symbol* $\uparrow$ *stands for upsampling while* $\downarrow$ *stands for downsampling. a filter bank with two complimentary filters (one lowpass filter coupled with one highpass filter) followed by downsampling operation.* **Middle***: flow chart for 2D discrete wavelet decomposition (DWT): By carefully designing the lowpass and highpass filters, the decomposition can split the approximation in level* $s-1$ *into four mutually orthogonal subspaces: the approximation* $\mathbf{A_s}$ *subspace in level* s *which contains low frequency components, and the horizontal* $\mathbf{H_s}$*, vertical* $\mathbf{V_s}$ *and diagonal* $\mathbf{D_s}$ *subspace from highpass filters which contains the high frequency components.* **Lower***: flow chart for stationary wavelet decomposition. The same process is carried out to filter the image. The difference lies in that the down sampling operation is eliminated, and the filter responded are upsampled. This lead to a redundant representation.*

The discrete wavelet is a set of bases in $\mathbf{L^2(Z)}$ on which the image $\mathbf{I}$ can be projected. In 2D there are four subspaces: the level 1 approximation: $\mathbf{A_1}$ contains low frequency component, while level 1 horizontal $\mathbf{H_1}$, vertical $\mathbf{V_1}$ and diagonal $\mathbf{D_1}$. For non-redundant discrete wavelet decomposition, the four subspaces are orthogonal. The filtering-downsampling operation can be applied iteratively to the approximation, and the s^{th} iteration splits the level $s-1$ approximation $\mathbf{A_{s-1}}$ into level s horizontal $\mathbf{H_s}$, vertical $\mathbf{V_s}$ and diagonal $\mathbf{D_s}$ subspaces. Define $\mathbf{A_0} = \mathbf{I}$ and $\oplus$ is the direct sum of vector spaces[58], thus:

$$\mathbf{A_{i-1}} = \mathbf{A_s} \oplus \mathbf{H_s} \oplus \mathbf{V_s} \oplus \mathbf{D_s} \tag{3.11}$$

For convenience the symbol: $\mathbf{A_{s-1}} = [\mathbf{A_s}, \mathbf{H_s}, \mathbf{V_s}, \mathbf{D_s}]$ is used to represent the decomposition.

Some of the most important properties of wavelet are listed below:

Compactness: unlike sinusoid bases which extend from minus infinity to plus infinity, the wavelets can be of finite duration, or termed "of compact support". This is very useful in image analysis: it can give localised frequency description.

Smoothness: the wavelet basis $\mathbf{\Psi}$ with v vanishing moments is orthogonal to the polynomials of degree $0, 1...v-1$ which means it can separate piecewise smooth structures from highly oscillating components whose order may above $v-1$. The corresponding scaling function $\mathbf{\Phi}$ however can locally approximate the polynomial of degree $0, 1...v-1$ perfectly.

Shift-variance: For orthogonal, complete and non-redundant wavelet transform, the shift-invariant property is no longer valid because of the downsampling. While this does not affect the compression applications, it does cause problems in classification. In recent years a lot of research efforts has gone into examining the classification power of the simple FOS in the wavelet domain, with good results reported[60–62]. More advanced algorithms examine the correlation of the wavelet coefficients in different channels [63].

3.3.2 Feature extraction in Scale/Wavelet Domain

Feature extraction after wavelet transform or filter bank operation is reported to be very successful. For characterising texture, Unser and Eden proposed to extract local statistics such as moments after wavelet transform, and reported that the wavelet transform provides superior segmentation results to the principal component transform (Karhunen-Loeve transform) [64].

Models were usually used to process the transformed image for classification/segmentation. Different kinds of wavelets/filter banks associated with different models were reported [65, 66]. The statistical relation across different frequency channels was also studied [67].

However the success of statistical features in the orthogonal wavelet domain is limited. First of all as the downsampling step is shift-variant, it is very likely that two identical textures with a spatial translation have very different wavelet representation, leading to the failure of further classification. Moreover, unlike man-made signals, in a natural image the ROI may have very limited resolution. By downsampling the size of the ROI may become too small to have statistical meaning. In order to avoid these problems filter banks without downsampling, that is redundant wavelet transform, is more commonly applied to extract features.

3.3.2.1 Gradients and Edges

In a homogeneous image, the edge in terms of large magnitude of spatial gradients suggests the boundary of different semantic regions. The edge is often spatially connected and sparse. However the boundary of texture images cannot be defined by edges, since the intensity variation is very rapid and stochastic, the density of the edge is also high in texture images. Early effort has been seen to characterise the density of edges in textured images as features to locate the boundary of textural regions by using morphological operations [68]. The density and directionality of the edges in textured images can be characterised by co-occurrence matrices [69]. Recently the concept of textural gradient has received a lot of attention. One use of this is to recover the 3D gradients of a textured surface by examining its 2D image properties. Clerc and Mallat formulated a texture gradient equation and proposed a 2D warplet transform to estimate the gradient [70]. The other is to derive a homogeneous gradient field to represent the variation of textural images for segmentation. Hill *et al.* proposed a complex non-decimated wavelet based estimation of the texture gradient, which can be jointly used with watershed algorithm to segment the textured images [71]. The application of the texture gradient for medical image segmentation is not yet known.

3.4 High-level Classification/Segmentation for Medical Images

After the desired low-level features are extracted from the images, the high level segmentation/classification methodologies need to be applied to yield meaningful semantic regions.

Moreover, the high-level modelling usually employs optimisation algorithms which can bring out a local optimal solution that the low-level features offer. In this section, the five major methodologies for medical image segmentation/ classification are reviewed: classification based approaches, knowledge based approaches, morphology based approaches, registration based approaches and shape model based approaches. Although these methodologies are distinctive from one another, they can be jointly used in a synthesised segmentation/classification system, which will be detailed in Section 3.5.

3.4.1 Classification

The classification in general is a very large topic for the pattern recognition community. Here only some key developments and algorithms are reviewed. All the classifiers mentioned in this section can be found in [42]. According to whether the classification criteria needs to be learnt from training data, algorithms can be divided into supervised learning and unsupervised learning. According to whether the clustering of the data involves using models such as normal distribution model, algorithms can be divided into model-based and non-model-based methods. Early classification algorithms includes linear discriminant model, k-means clustering, fussy-C-means clustering and Bayesian classification, later development involves non-linear classifiers such as neural network and support vector machine (SVM).

The classification based approaches usually split the images into small ROIs. By classifying the features from the ROIs into different categories, the images can be segmented into different semantic regions. The ROI can either be defined as a particular geometric shape such as a square/rectangular region, or be generated from other segmentation algorithm such as watershed [72]. After defining the ROI, features (may either be low-level features or high-level features) are extracted from the ROI and fed into the classifier.

The learning process however may very likely be subject to over-fitting. Over-fitting refers to such situation, that if a classifier is trained to build a decision boundary which aggressively reduces the classification error for training set, this decision boundary may in fact not be a good one to separate the feature space: the random nature of the data leads to inconsistent distributions in training set and testing set. Over-fitting commonly occurrs when there are limited number of feature observations but a large number of different kinds of features. As a result, the classifier may give very good classification results for training set, but for testing set the classification results are poor.

To cope with over-fitting, a common strategy is cross-validation. The tenet of cross validation is to use as much information as possible to make the prediction, but each data can only appear in either the training set or the test set, not both [73]. In this study, as the image sets contained repeated and registered scans, intra-patient images were similar, so they were not allowed to appear in both training and test sets. The leave-one-out cross validation was used. At each iteration data from 6 patients were used as training set, and data from the remaining patient was examined as unknown test set.

3.4.1.1 Bayesian Classifier

The Bayesian analysis methodology, instead of making a point estimation, infers the **posterior distributions** of different classes of data from the training set by using the formula:

$$\begin{aligned} p(C_i|z_m(1), z_m(2)...z_m(n)) &\propto p(C_i)p(z_m(1), z_m(2)...z_m(n)|C_i) \\ &\propto p(C_i)\prod_{j=1}^{n} p(z_m(j)|C_i) \\ &\propto \prod_{j=1}^{n} p(C_i|z_m(j))/p(C_i) \end{aligned} \tag{3.12}$$

It assigns a posterior probability for each test data belonging to one particular class. One particular advantage in Bayesian analysis is, the results of different measurements can easily be quantised into posterior probabilities and combined into a new posterior probability.

3.4.1.2 Feature Selection

One problem associated with classification is that there may be lots of irrelevant features. According to Bayesian probability theory, more information input will only benefit the classification. However, it is reported in the literature that because the true distribution of the data is usually unknown and the best hypothesis is difficult to justify, under a less-than-optimal distribution hypothesis, usually Gaussian, features of poor discriminating power will deteriorate the classification [74].

Moreover, when the number of features input into these distance based classifiers is very large, the observed data become more and more sparse in the feature space, making it difficult to find decision boundaries in feature space. On the other hand the Euclidean distance became a

less distinctive measurement among different classes of data. This problem is called "curse of dimensionality" which deteriorates the classification performance. In order to yield better classification, the dimension of the feature space needs to be controlled. To tackle these difficulties, The feature reduction/selection algorithms are therefore developed to bring down the number of features input into the classifier.

Principal Component Analysis PCA is a well-established method to map the features into a linear sub-space, so that the correlation in the second-order sense between any two dimensions in that sub-space is minimal, while the variations of features are maximally maintained [75]. For a feature matrix $\mathbf{F}$ whose m columns are the different features and n rows are the observed features from different ROIs. The new feature matrix $\mathbf{Z}$ in the principal domain is:

$$\mathbf{Z} = \mathbf{F}\mathbf{V}' \tag{3.13}$$

where the transform coefficient matrix $\mathbf{V}'$, or termed *loadings*, is derived from the singular value decomposition:

$$\mathbf{F} = \mathbf{U}\mathbf{S}\mathbf{V}', \mathbf{U}'\mathbf{U} = \mathbf{I_n}, \mathbf{V}\mathbf{V}' = \mathbf{I_m}, \tag{3.14}$$

and $\mathbf{S}$ is a diagonal matrix. Since the covariance matrix of $\mathbf{F}$ can be decomposed as:

$$\mathbf{\Sigma_m} = \mathbf{F}'\mathbf{F} = (\mathbf{U}\mathbf{S}\mathbf{V}')'(\mathbf{U}\mathbf{S}\mathbf{V}') = \mathbf{V}\mathbf{S^2}\mathbf{V}', \tag{3.15}$$

The eigenvalues in $\mathbf{S^2}$ are the magnitudes of variances in the corresponding principal subspaces and are usually sorted in descent order. The new variables, i.e. the columns of $\mathbf{Z}$, are called principal components (PCs). It can be seen that if only a subspace of dimension $k < p$ is used to approximate the original features, the first k PCs can maximally retain variation in the features and are therefore the best linear approximations.

3.4.2 Morphology

Morphology based approaches exploit the spatial properties of low-level features to define the shape on the images, without assuming the high-level knowledge. Mathematical morphology, which extracts low-level shape features from the image by using different shaped elements, is a very powerful tool for image post-processing[76]. In medical image segmentation, morphological segmentation can segment images into different ROIs of different shapes, but cannot determine the semantic meaning of the ROIs whether the shapes represent organs, tumours, etc. The semantic meaning can either be specified manually by a user or be assigned by another classifier after segmentation. The most popular morphological segmentation algorithms are the region growing and the watershed.

3.4.2.1 Region Growing

The region growing algorithm usually comprises of two main steps: initial seed selection and a termination criterion. The initial seed can either be automatically found or manually specified. The criteria for example can grow the region until the termination criterion is met. When the edge is not significant, the seeded regions are likely to leak into other regions, leading to the failure of the segmentation. In the literature region growing is applied to find the bladder and prostate automatically. As the edge is very weak in bladder-prostate junction, leakage will occur [77, 78] and the bladder-prostate boundary cannot be defined. Mazonakis et al. [77] proposed the region growing technique to segment prostate, bladder and rectum for prostate cancer patients. In their study it was reported that segmentation leakage may occur when two adjacent structures have similar gray-levels, and a manually drawn region was applied as a hard constraint. Shi et al. [78] proposed a hybrid scheme including mean shift clustering, region-growing and rolling ball algorithms to segment the bladder volume. Sensitivity, specificity and Hausdorff distance were used as the accuracy measurement, and the results were promising. In order to confine the growing, marks need to be put manually on the image as the indicator of the boundary.

3.4.2.2 Watershed

Another very popular morphological segmentation algorithm is watershed segmentation. The watershed algorithm first transforms the data into a gradient representation, finding the local

minimals to define the catchment basins; A flood filling is then applied to the gradient representation. When water from two basins are met, the boundary is defined. Watershed is very popular in medical image segmentation [79, 80]. The draw-back is it is the over-segmentation due to the presence of the noise. In this case either assign markers on the image to prevent over-segmentation [81], or use the clustering algorithm to re-group the over-classified image [80]. Recently for textured image the texture gradient is often used with watershed for segmentation [71], yet to the author's best knowledge the concept of textural gradient has not yet been applied to medical image analysis.

3.4.3 Registration

The term registration refers to transforming the image coordinates to find correspondence of maximum similarity between two images [82]. The image whose coordinate system is put through transformation is called the **target image** while the other image without spatial transform is called the **reference image**. Registration is usually modelled as an optimisation problem with similarity described by a target function and physical constraints described by mathematical formula. Registration is an extremely important subject and is widely applied in image analysis, especially medical image analysis. Registration algorithms can be classified according to different principles. Based on the physical property of the transform it can be classified as rigid registration and non-rigid registration; based on the input to the optimisation algorithm it can be classified as intensity-based registration and feature-based registration.

3.4.3.1 Rigid Registration

In mathematics rigid motion refers to the motion of an object in which any two points remain at the same relative distance. In algebra this can be implemented by multiplying a matrix representing the object by an orthogonal matrix for rotation, or adding a constant for translation [82, 83]. In physical world there is no such rigid motion in strict sense: subtle relative positional displacements occur whenever a physical motion is involved. However for medical context bones can be considered rigid objects and their motion can be modelled by rigid motion. When used for registration the algorithm rotates and translates the target image to find the optimal match with the reference image. Its advantage is that it is relatively simple to understand and easy to implement, but for soft tissues the rigid motion assumption no longer holds because they can distent or deform instantly. In that case non-rigid/deformable registration is needed.

3.4.3.2 Non-rigid Registration

Non-rigid motion refers to the motion of an object in which there are at least two points change their relative distance. In algebra this is characterised by multiplying an object matrix by an arbitrary non-orthogonal matrix. The transform matrix space is extremely large, some of them do not represent meaningful physical shape. A lot of attention has been given in the literature to reducing the transform matrix space down to a feasible size for computation. Since deformable transform itself is not used in this thesis, interested readers are recommended to see [82] for details.

3.4.3.3 Segmentation using Atlas Registration

In image processing literature, manually delineated anatomical structures are usually referred as an atlas to standardise the general anatomy [84], for example, "Atlas of Human Anatomy" [27] is a good illustrative reference of anatomical structures. Atlases are intensively studied to standardise brain anatomy. In 1988 a brain atlas was carefully documented by establishing coordinate systems, reference planes and landmarks [85]. It serves as the basis for brain atlas studies. This kind of atlas is useful in distinguishing regions near the landmarks and structures of small inter-patient variability, however it does not count for the population variation. Since then lots of studies have focused on building a computerised atlas instead of paper atlas to count for the population variability of the brain. Mazziotta *et al.* addressed the rationale for describing the brain structure by a probabilistic atlas, i.e. to characterise the population variability by probability [86]. Guimond constructed an average shape atlas from manually segmented MR images by first applying affine registration to align data set to a single reference image, then applying deformable registration to obtain the transformation [87]. The transformations obtained from all data sets were averaged to obtain an average transformation, upon which the average shape is constructed. It was reported that the average atlas will converge to the centroid of the MR image set. Lorenzen *et al.* worked on building a probabilistic atlas from multi-modality images. In order to register different modalities, for example, CT and MR, the Bayesian inference was first applied to convert the gray levels of images into probabilities that a voxel belonging to a particular category of tissue, and registering the probabilities by using mutual information criterion to describe the similarity [88, 89].

Van Leemput on the other hand, reviewed the major methodology for constructing the brain atlas, and posed two problems for current methodology [90]. Firstly, the limited number of

training data to construct the probabilistic atlas may not be a valid representation of the general population; besides, the optimality of degree of deformation used for registering different data are difficult to justify: too small a deformation would fail to align different structures, while too large a deformation may weaken the accountability of population variability. After the construction of brain atlas, it can be used to segment structures. For example Bricq *et al.* applied the probabilistic atlas as a prior to the hidden Markov chain model to segment the brain lesion [91].

Similar methodologies have been adopted to construct atlas for the abdominal organs. The focus would primarily be to enable automatic target segmentation. For example Park and his co-authors built atlases of four abdominal organs: the liver, two kidneys and the spinal cord [92], and used the atlas as a 3D mask in later study to registration-based lesion segmentation in liver [93].

For the pelvic organ segmentation problem, as mentioned before, the variability of the bladder and rectum is significant and the atlas cannot be used for segmentation. For prostate however, the shape is relatively consistent among population and as a result, efforts have been seen to build the prostate atlas to assist the segmentation. Also the concept of multi-atlas segmentation was proposed to find the most similar atlases from a database of images to match the target image for segmentation. Klein *et al.* used the mutual information as the similarity measurement to non-rigidly registering manually segmented MR prostate images to the target image[94, 95]. A threshold is applied to select the prostate atlases with high similarity. After that the atlases were averaged to obtain the prostate contours.

3.4.4 Shape Modelling

Shape modelling refers to finding a parameterised function, usually a set of x-y coordinates interpolated by various basis functions, to describe a geometric shape in the image. Similar to registration it is also usually an optimisation problem with a target function and some constraints. However unlike registration whose target function measures similarity of two images, the target function of a shape model balances the prior knowledge of the shape and the features on the target image.

Early development of shape models such as active contours are proposed to evolve the initial contour (a set of points on the image). It is done by minimising the objective function taking

into account of internal energy of the contour (spline between points) and the external energy of the shape constraint[96]. More advanced techniques, for example the deformable shape models, use a parameter-controlled shape to define the energy function while use anatomical knowledge as constraints in the optimisation problem. Chen et al. [39] proposed a point-based active shape model to segment the prostate and rectum for prostate cancer treatment planning. For readers interested in the application of shape models in medical image segmentation in general, Heimann and Meinzer give a very good review [97].

It is important to note here that while the two terms "modelling" and "registration" have clearly different meanings, in segmentation algorithms they are often jointly used. For example registration can be used to find the correspondences in training data to build the shape model, or to initially map the model to the target image, whereas deformable models can also be used to register images. As a result, the model based approaches may have lots of registration operations while the registration based approaches can also employ models. The key difference between model based approaches and registration based approaches is, while the registration based approaches strive to find a pixel-to-pixel correspondence between the training images and the target images, the model based approaches in general does not require any such correspondence and using the extent of the structures on reference images to derive the corresponding structure on target image.

The application of the shape model based segmentation are successful in segmenting organs, which show consistent shapes in population. However, the use of shape to fit tumour extent is an ill-defined problem. Without distinctive low-level features, the best shape model would be a best guess.

3.4.5 Knowledge Based Approaches

The natural way of doing medical image segmentation is to exploit the knowledge of human anatomy. To fulfil some very difficult segmentation task, first some easy-to-identify landmarks can be found. On CT images for example, the pelvis bone is of high CT number while air in the rectum is of low CT number, both of them are easy to identify. After segmenting these easy-to-identify structures, the anatomical knowledge can be used to set up a set of relative position rules to find the desired organ based on the identified structure as reference: bladder should always lie within the pelvis bones, or the only organ in pelvis region which may contain air is the rectum (assuming no air is used as a negative contrast agent in the bladder). A solver

can be implemented to derive the target region by following the set of positional rules which lead the segmentation algorithm from the landmarks to the target region. This approach is also referred to as rule-based approach.

In the literature, Archip *et al.* used this approach to segment spinal cord on CT images by defining and using a set of position relations of different organs [98]. For pelvic organ segmentation problem, commercial software from Varian Medical Systems(Varian Medical Systems iLab, Baden, Switzerland). In their recent publication proposes a rule-based solution to delineate bladder, prostate and rectum [99]. This is achieved by first identifying pelvic bones as landmarks, then applying different rules to segment different organs. For example, in order to extract rectum, the CT numbers corresponding to air were extracted within a predefined region were thresholded, and the convex hull containing all these air regions is taken as the minimum rectum contour. Anatomical rules can also be formed to constrain the position of bladder-prostate junction for prostate segmentation.

3.5 Applications of Image Segmentation / Classification: a Review

3.5.1 Performance Evaluation: Qualitative and Quantitative Methods

Before going further to review more approaches, the methods used for evaluating the automatic contours are discussed here. In a subsequent study, research group from Varian Co. carried out clinical examinations to evaluate contours generated the rule-based method proposed in [99] both qualitatively and quantitatively [2]. The accuracy of automatic contours is described both qualitatively and quantitatively. For qualitative evaluation, automatic contours were presented to clinicians, who gave a description, such as “excellent”, “good”, “acceptable” or “not acceptable” to indicate their degree of support. The quantitative evaluations usually use clinical contours as good standard, and employ quantitative descriptions, such as DSC, to indicate the difference between the automatic contours and clinical contours. This comparison is based on the fact that clinical contours have a high degree of agreement. For structures with high inter-observer variability, the role of clinical contours as gold standard is weakened. However, the average clinical contour can still be used as a reference to which the automatic contours are compared to give a suggestion of performance.

In quantitative examination, six clinicians were required to delineate bladder, prostate and rectum independently to set a gold standard to which the automatic contours are compared. The

overlap of contours was quantified by DSC, which is defined as two times the overlapped area over the sum of two areas[100]:

$$D = \frac{2(A \cap B)}{A \cup B} \tag{3.16}$$

where A is the area of the clinical contour, while B is the area of the automatic contour. Interesting results show that in the bladder-prostate junction the average DSC of clinical contour for bladder and prostate were between 0.5 and 0.6, indicating large inter-clinical variability. In mid range of bladder however, clinical and automatic contours have the DSCs between 0.8 and 0.9, indicating good agreement. These result suggests that toward the bladder-prostate junction, if no strong evidence is shown to bring down the inter-clinical variability, any algorithms claiming an automatic bladder/prostate contour(s) in bladder-prostate junction level on CT images yielding a DSC over 0.6 with clinical contour would amount to mimicking the style of particular clinician(s), and the validity of automatic contours is difficult to justify. In mid range however, high quality auto-segmentation is currently possible. In order to bring down the inter-clinical variability, more distinct low-level features are highly desired.

3.5.2 Medical Image Classification Applications

In medical image segmentation research, the distance based classifiers together with the feature reduction / selection algorithms are jointly studied, and proposed as a system to contain the task. Nailon *et al.* used the k-means clustering together with a sequential forward selection algorithm to classify the statistical textural features from three specified ROIs on CT images: GTV, OAR and a control region in between the two, and achieved visually-distinguishable clustering results [7, 47]. Yu *et al.* also applied the k-means clustering to examine the textural features from ROIs of head and neck cancer on PET/CT data, the resulting sensitivity and specificity were 89% and 99%, respectively [45]. The features calculated in wavelet domain can also be used for classifying ultrasound images [60], CT images [61], MRI [101]. Although the optimality of the classification performance is difficult to justify, the systems are reported of good performance. Recently the context features are proposed to classify the prostate in CT images by Shen and his co-authors [40, 102, 103] for contouring.

Neural networks are also very widely applied to classify medical images in both spatial domain [104, 105] and wavelet domain [60] to classify ROIs. Generally the training cost of the neural network can be considerable, but because the neural network learning is an optimisation process, the optimality of the classification results can be argued. There are also some applications

which uses the SVM to classify brain MRI [106, 107], and prostate ultrasound images [108]. The application of kernel based analysis remains open in medical image analysis.

3.6 Discussions

Intensive works was carried out to introduce the automatic pelvis segmentation application in clinical practice. However, there are two major problems associated with current solution for atlas based segmentation. First of all, as most of the application are based heavily on the shape and intensity information from previous cases rather than derived from the images under consideration, they have not yet provided meaningful evidence of the extent of different organs based on the current image, which is essential for eliminating the inter- and intra- observer variability. If the methodology is based on previous knowledge and indistinguishable intensities of two organs such as in the case of bladder-prostate junction, the best efforts for segmentation amount to an educated guess. In order to assist doctors to reduce the observer variability, meaningful structures must be derived from the images currently under consideration to show clear structures for clinical reference.

3.7 Summary

In this chapter the major methodologies for extracting features from the images, and classifying / segmenting the medical images into semantic regions have reviewed. It can be seen that the majority of the application so far for pelvis segmentation are based on the shape and first order features. For the bladder cancer delineation problem, the extent of the tumour is not clearly distinguishable using the first order statistics. As a result, there are no satisfactory features defined to address this problem. In work presented in later chapters, the second- and higher-order statistical features, originally used to characterise the image textures, will be used to extract more distinctive features.

Chapter 4
Statistical Texture Analysis for GTV Classification on CT Images

4.1 Introduction

In Chapter 3 the major methodologies for medical image low-level feature extraction and high-level segmentation/classification were reviewed. Traditional pelvis organ segmentation methodologies were based on simple low level features such as pixel intensities or first-order statistics from the first order histograms. In this chapter second-order statistical features were used as low level features for classifying soft tissue ROIs. A set of 59 stacks of (3D) CT images from eight bladder cancer patients were examined. The detailed description of data set was given in Section 4.2. In all of the subsequent experiments different subsets of the data were examined.

Three main issues were addressed: how to extract GTSDM features, how to reduce/select the features, and how to evaluate the classification performance of the features. In Section 4.3 the intensities and simple FOS were shown insufficient in distinguishing the GTV from non-GTV soft tissue ROIs. By applying PCA to GTSDM features, the principal components (PCs) can give a clear distinction between three selected anatomical regions: bladder as GTV, rectum as OAR and control regions in between. The consistency of PCA across image sets from different patients tested using leave-one-out cross-validation showed that the the six largest PCs accounted for over 95% of the variations in original feature space and their loadings were consistent. The loading matrix can be fixed and applied to different data sets.

Third in Section 4.5 a wrapper feature selection system was implemented to select a subset of PCs which gave the highest classification performance from the six most significant PCs. The naive Bayes classifier (NBC) was used as the induction algorithm, AUROC is adopted as the criterion and an exhaustive search was conducted. Leave-one-out cross validation was employed for consistency and the results suggest that two or three PCs can give a much larger AUROC than FOS can. The most frequently selected PCs were PC 1 which comprised mainly

an average of four entropy features and PC 3 which contains mainly an average of two correlation features. This also indicates that the correlations and entropies offer two uncorrelated degrees of freedom in the feature space.

4.2 Data Set Definition

In this work a set of CT images from six male and two female bladder cancer patients were studied. The patients were treated with a radiation dose of 52.5 Gy (20 fractions/4 weeks) at the Edinburgh Cancer Centre (ECC). The age range of the patients was 63 to 81 years with a mean of 74 years. Each patient received a CT scan with a presumed empty bladder. CT scanning was performed twice weekly during the course of treatment to define the GTV for treatment, OAR, and to determine the optimum dose delivery plan. For each patient five to eight scans were taken, resulting in a total of 59 stacks sets of repeated scans. Repeated scans were registered against bony anatomy using the Advantage Fusion software package (v.1.15; GE medical Systems).

Each set contained 64 to 79 CT images, of which 6 to 14 slices were contoured by an experienced clinician. One male patient was scanned using a 5mm slice thickness, and the other seven patients were scanned using a 3mm slice thickness (1GE Hipeed Fx/i, GE medical systems). The resulting resolution of the CT images was approximately 1mm in the axial plane with a 2048 gray level dynamic range. All the images are scaled to the interval $[0, 2048]$ before study. The size of each slice was 512×512. A sub-image of 176×128 region was sufficiently large to contain the bladder (GTV) and rectum (OAR), and was therefore chosen by using bones as landmarks from each image. For example, Figure 4.1 illustrates nine sub-images from two male patients (first and second rows) and one female patient (third row) in the superior (first column), mid (second column) and inferior (third column) range of bladder. In later text, unless otherwise specified, the word "image" and the symbol $I(x, y)$ refer to the 176×128 sub-image.

4.3 Low-level Features for GTV characterization

4.3.1 Intensities

Soft tissues have an average CT number close to water, because the data was scaled to lie in the range $[0, 2048]$ before analysis was conducted, the mean of the intensities in GTV was

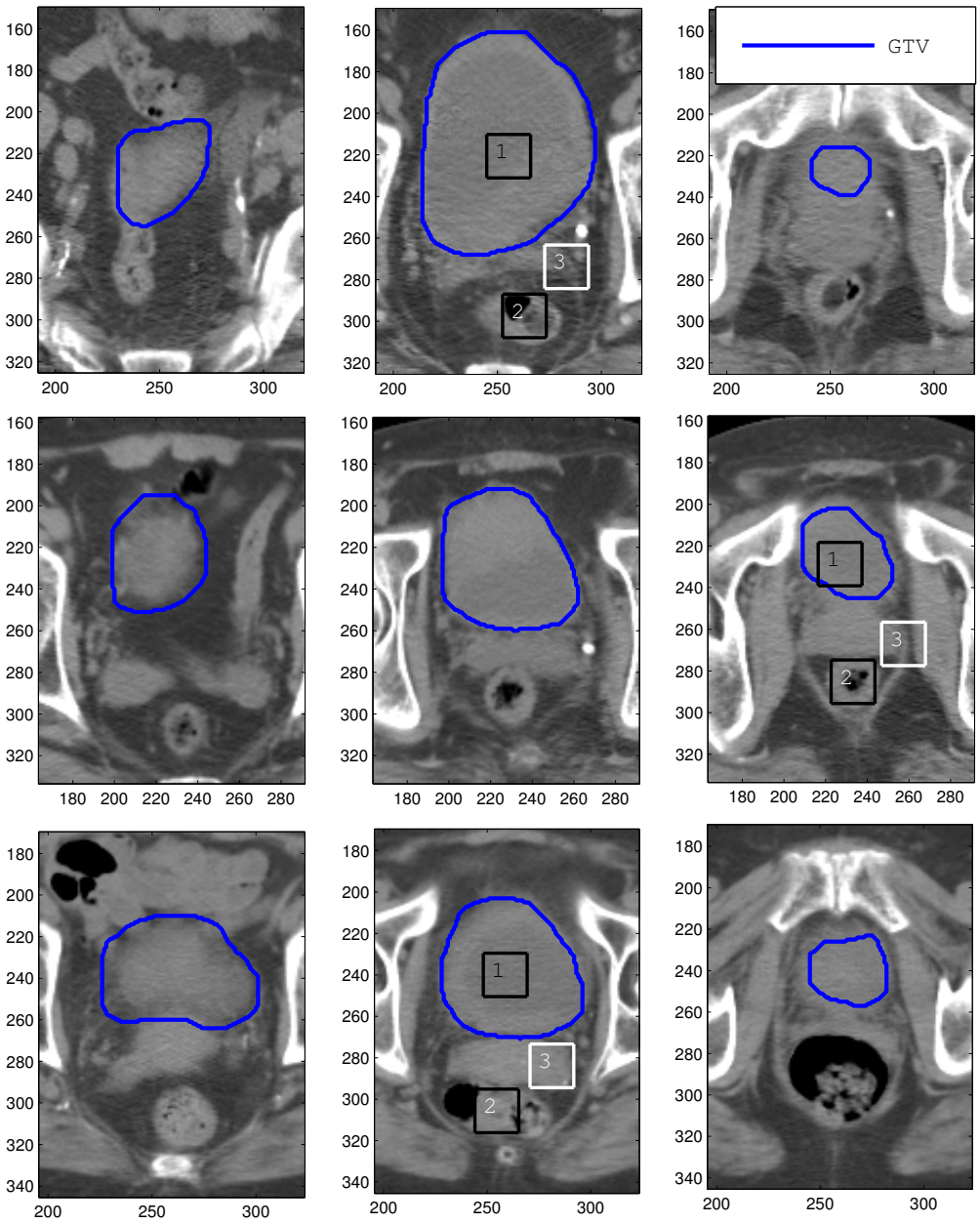


Figure 4.1: *Nine CT images in the pelvis region from three different patients. Different rows are from different patient. Region 1 is within the bladder (GTV). Region 2 is the rectum (OAR) and Region 3 is a control region in between containing multiple pathology. GTV contours are shown in solid line. The first, second and third columns are superior, mid and inferior parts of bladder.*

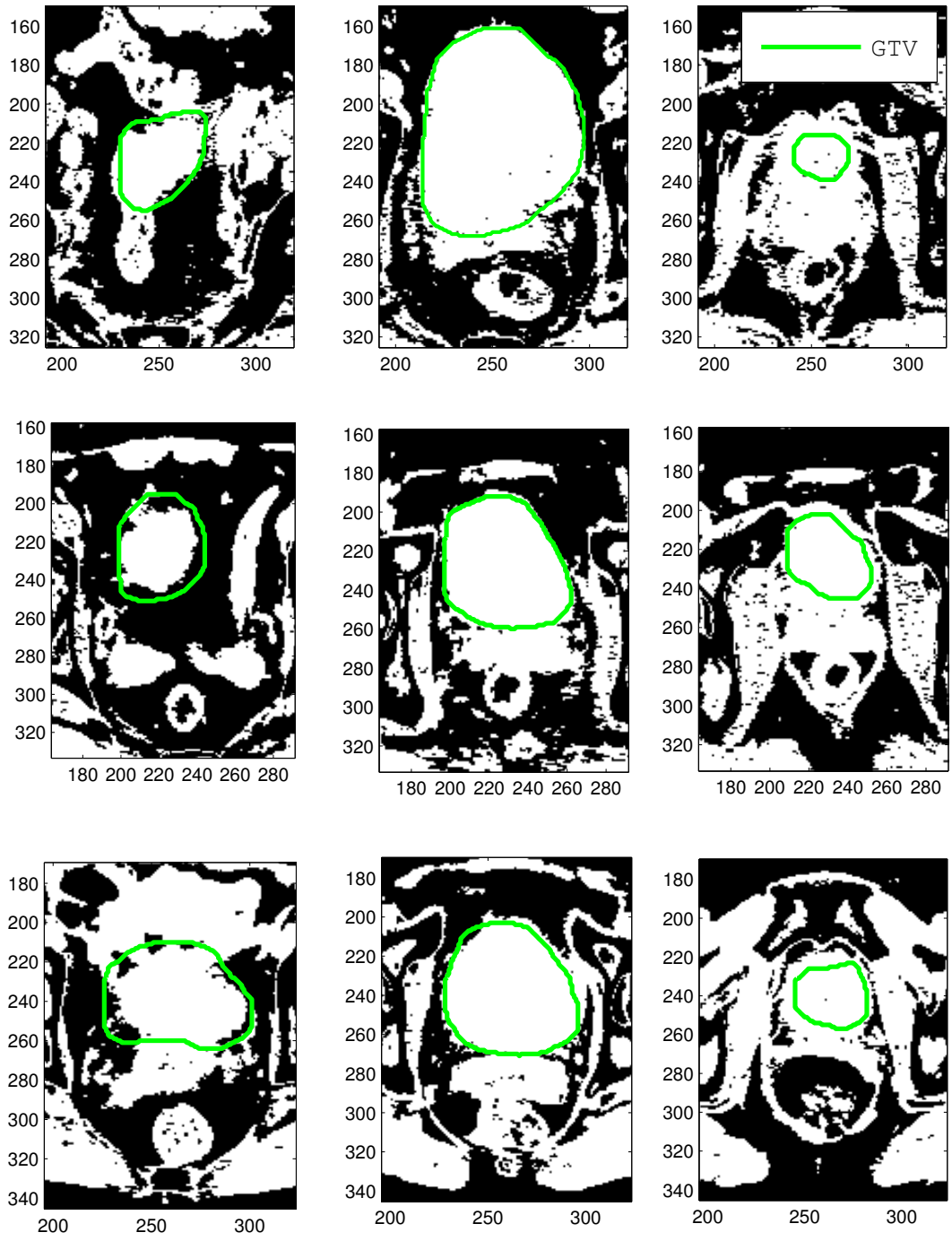
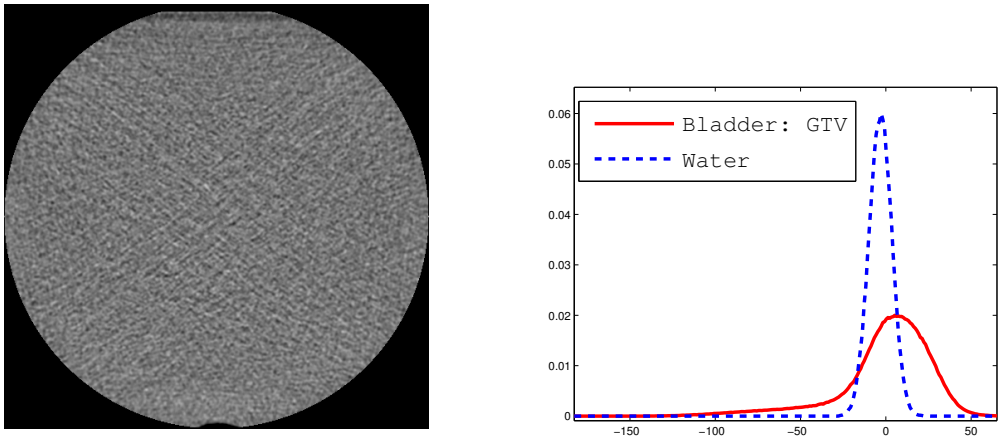


Figure 4.2: *Thresholded soft-tissue regions from images in Figure 4.1 by applying thresholds in Equation 4.1. GTV contours are shown in solid line. A significant number of OAR regions are at the same intensity interval as the GTV.*



(a) Water Phantom

(b) Histograms of intensities in GTVs and Water Phantom

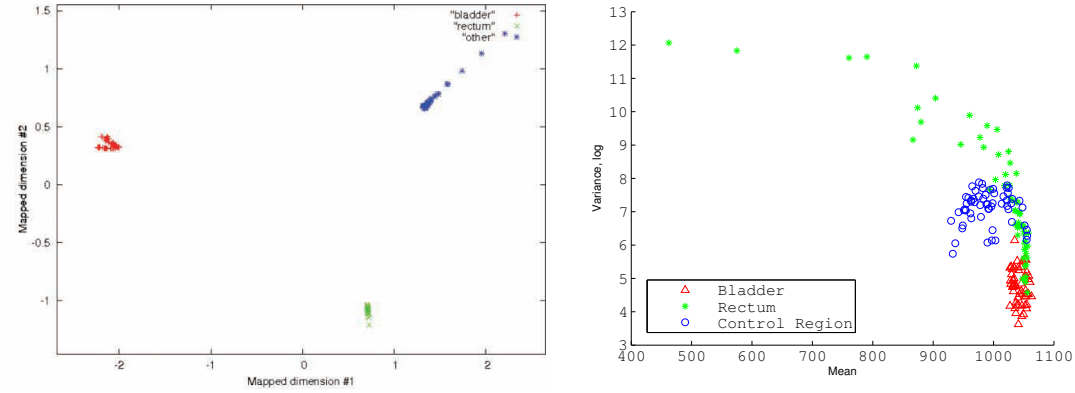
Figure 4.3: *4.3(a): A water phantom used for estimating the standard deviation of noise at the intensity level of zero. 4.3(b): The normalized histograms of intensities in water phantom (dashed line) and GTVs (solid line). The variance of intensities in GTVs are significantly larger than that in water phantom.*

$\varepsilon_{GTV} = 1039$. The solid line in Figure 4.3(b) plots the normalized histograms of the bladder within clinical GTV, subtracting ε_{GTV} from the intensities. The STD of the intensities within clinical GTV is $\sigma_{GTV} = 27.43$, and 94.39% of pixels in GTV have their intensity confined in the interval $[\varepsilon_{GTV} - 2\sigma, \varepsilon_{GTV} + 2\sigma]$. If a thresholding operation T is applied to the CT images such that:

$$T(I(x,y)) = \begin{cases} 1 & \text{if } i(x,y) \in [\varepsilon_{GTV} - 2\sigma, \varepsilon_{GTV} + 2\sigma] \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

Figure 4.2 shows the binary masks obtained by this threshold on the corresponding images in Figure 4.1. A significant amount of soft tissues in pelvis regions have intensities within the interval $[\varepsilon_{GTV} - 2\sigma, \varepsilon_{GTV} + 2\sigma]$. This suggests that the intensity alone is insufficient to distinguish GTV from non-GTV.

However, the intensity variation in GTV is significantly larger than the variation of pure noise. To show this, the normalized histogram of a water phantom is plotted in dashed line in Figure 4.3(b). Water phantom is considered CT neutral and should comprise only noise. The STD of the water phantom is $\sigma_{water} = 6.83$. The richness of intensity variation in GTV suggests that statistical features may offer important information for characterizing the GTV.



(a) (b)

Figure 4.4: *(a). The means and log-variances of the three ROIs from 59 radiotherapy planning CT images. Significant overlap can be observed, so different ROIs cannot be distinguished by Means and Variances. (b). Scatter plot of the images of the three GTSDM features which gave the largest distinction: angular second moment, correlation and variance in 2D domain projected by Sammon plot from [7]. Visual distinction can be seen between the projected features from different classes of ROIs.*

4.3.2 FOS and GTSDM Features

A previous study by Nailon *et al.* reported that of all the FOS, GTSDM features, GLRLM features and fractal features, three GTSDM features gave sufficient distinction to segment three ROIs on CT images: bladder (GTV), rectum (OAR) and a control region in between[7, 47]. In their study one image with demonstrable cancer characteristics was selected from each of the data set for study by an experienced oncologist, on which the three ROIs were specified. For example in Figure 4.1 the three ROIs were marked by square boxes on the selected images. Figure 4.4(a) shows the three GTSDM features which gave the largest distinction: angular second moment, correlation and variance in 2D domain projected by a Sammon plot. Clear distinction can be seen between the projected features from different classes of ROIs.

The FOS however were insufficient to distinguish the three classes of ROIs. For example the means and variances of the same ROIs are scattered in Figure 4.4(b). There were significant overlaps among the FOS features. Two major factors cause the overlaps: the homogeneity of soft tissues makes inter-class variation small, where as the heterogeneity of the rectum due to the random appearance of air and filling makes intra-class variation (especially for rectum) large. Based on these results, the attention of this chapter is focused on using the GTSDM

features to characterize different classes of soft tissues.

4.4 Feature Reduction using PCA

Successful classification algorithms are based on low-level features which have small intra-class variation and large inter-class variation, as is the case in Figure 4.4(a). Features with little variation suggest poor classification performance. It is not difficult to see that the 14 GTSDM features defined in [43] are likely to be correlated. As a result in the 14 dimensional feature space, some sub-spaces may have limited variation, and can therefore be discarded. To establish the distinguishing power of the GTSDM features, PCA is employed to find a consistent and efficient representation of the most significant GTSDM features.

4.4.1 PCA of GTSDM features from Selected Anatomical ROIs

The GTSDM features from the images used in [7] were examined, with two major improvements made. The first was that the size of the ROIs was fixed, which may significantly influence the textural features. Previously the size of ROIs was chosen to be the smallest square region containing the rectum. As the sizes of the rectum varies considerably in different CT slices the selection of the size of ROI was found to be very subjective, which is not suitable for a fully automatic setup. The size of the ROI is a trade-off between accuracy and statistical consistency: a small ROI is desired because too large a ROI is likely to include other unwanted tissues, leading to undesired heterogeneity. However a larger ROI is likely to yield better statistical consistency.

In this study the size of the ROI was fixed to be 20×20 area. The size of the boxes shown in Figure 4.1 is 20×20. For 14 features from GTSDM with distance $d = 1$, Figure 4.5(a) illustrates the amplitude of the eigenvalues of $\mathbf{\Sigma_{14}}$. It can be seen that there are five significant degrees of freedom in the principal features, but according to Figure 4.5(b), the two most significant PC, $z_{14}(1)$ and $z_{14}(2)$, will not give a satisfactory visual distinction. The bladder and control regions can be separated reasonably well, however the rectum cannot be separated from the other two classes. This suggests that a negative consequence of fixing the size of ROIs is that the robustness of the features has deteriorated due to the lack of variation.

To address this problem more features from co-occurrence matrices with different distances were extracted from the ROIs. A total of 56 features from GTSDM with $d = 1, 2, 3, 4$ were

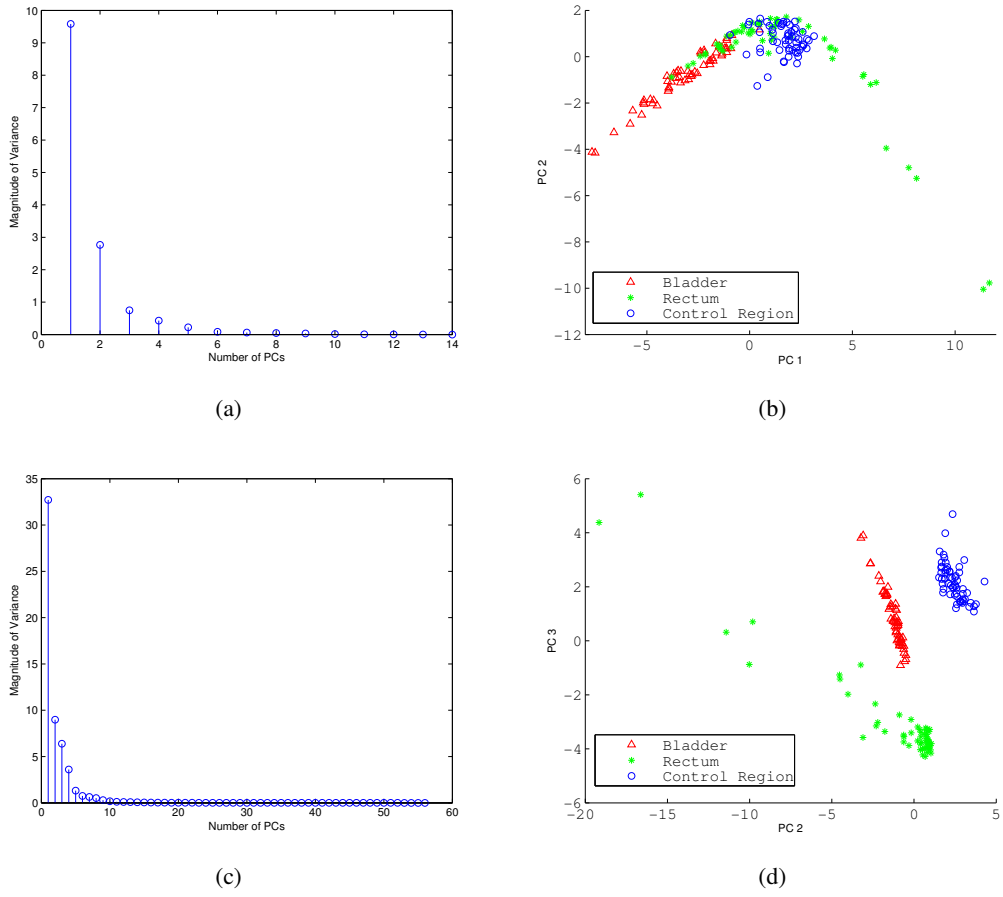


Figure 4.5: *(a): The Eigenvalues of the covariance matrix* $\mathbf{\Sigma_{14}}$ *of features from GTSDM with* $d = 1$*, Representing the variance in each principal subspace. Five significant degree-of-freedom can be observed. (b): Visualizing PC 1 against PC 2: ROI still cannot be distinguished intuitively. (c): The Eigenvalues of the covariance matrix* $\mathbf{\Sigma_{56}}$ *from GTSDM with* $d = 1, 2, 3, 4$*, Representing the variances in each principal subspace. Six significant degree-of-freedom can be observed, Comparing with Figure 4.5(a), the amplitude of variance in each PC subspace increases significantly. (d): Visualizing PC 2 against PC 3: ROI can be clearly distinguished. This Figure was published in [109]*

extracted from the ROIs for PCA. Figure 4.5(c) illustrates the magnitude of the eigenvalues of $\mathbf{\Sigma_{56}}$, where it is shown that the degrees of freedom increases and the magnitude of variance in each principal subspace also increases significantly. Figure 4.5(d) provides a visualization of $z_{56}(2)$ and $z_{56}(3)$ where it can be seen intuitively that the three classes of ROIs can be distinguished by the two PCs. This confirms the conclusion in [7] that the joint distributions of two pixels separated by $d = 1, 2, 3, 4$ from each other ($p(x, y, d)$) contains important information for characterizing the GTV, which can indeed compensate for the lack of variation introduced by fixing the size of the ROI.

4.4.2 PCA of GTSDM Features from Arbitrary Soft Tissue ROIs in Pelvis Region

After identifying that the PCs of GTSDM features have a good distinguishing power, this low-level feature extraction methodology was applied to all soft tissue regions. In this sub-section images from seven patients scanned with 3mm thickness, described in Section 4.2 were used for study. Images from patient scanned with a 5mm thickness were excluded. A total of 51 stacks of repeated scans from five male and two female patients. In order to form ROIs thresholding was applied, according to Equation 4.1, to obtain soft tissues I_{soft} as shown in Figure 4.2. After applying the thresholding 94.36% of GTV were included in the thresholding results:

$$I_{soft} = T(I(x, y)), such\ that : p(I_{GTV} \subset I_{soft}) > 0.9436 \tag{4.2}$$

To reduce the computational required the centres of the ROIs I_c were spaced at intervals of 4 pixels:

$$I_c = I(4m, 4n) \cap I_{soft}, m, n \in \mathbf{N} \tag{4.3}$$

Thus a total of 334,706 soft tissue ROIs were produced. Of all these ROIs there were 124,298 whose centres lie in GTV contour, which is 37.17%. The size of the ROIs remained at 20×20. However the consistency of the PCA approach must be tested before the classifier can be applied to the PCs, since PCA is calculated by SVD of the feature matrix and its consistency therefore depends on the uniformity of features[110, 111]. The consistency problem can be stated as follows: for a feature matrix $\mathbf{F}$ of n observations and m columns, find the consistent

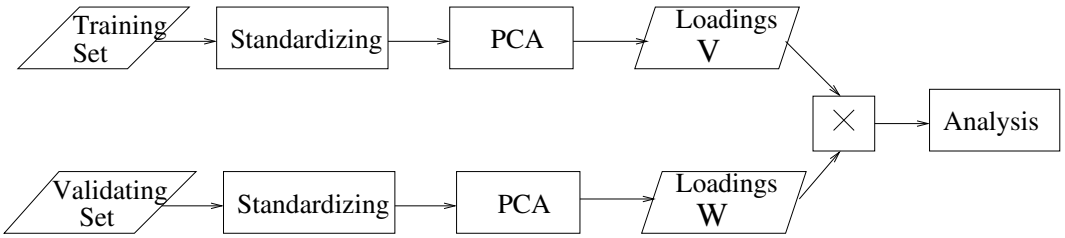


Figure 4.6: *Block diagram for cross-validating the consistency of PCA. Features are divided into two non-intersecting subsets: a training set and a testing set. The testing set comprises all GTSDM features from one particular patient to represent an unknown new observation while the training set contains all GTSDM features from the rest of the patients to stand for the known observations. The inner product of two coefficient matrices is eigen-decomposed. The fact that the absolute values of all eigen-values are close to 1 indicates that the two coefficient matrices are in the same linear space.*

dimension $k < m$ and the corresponding loading matrix $V(1, ..., k)$ so that the k largest PCs: $\mathbf{Z}(1, ..., k) = \mathbf{FV}'(1, ..., k)$ explains over 95% of the variation of the original features. This comprises of two issues. One is the dimension of the PCs k, which can be observed from the magnitude of the k largest elements of $\mathbf{S}$. The second column in Table 4.2 suggests that for the images of 3mm thickness from seven different patients that $k = 6$ largest PCs explained over 95% of the variation.

The other is the consistency of the loadings$\mathbf{V}(1, ..., 6)$, which needs to be tested by cross validation. The purpose of finding a consistent $\mathbf{V}(1, ..., 6)$ is that for a new feature matrix $\mathbf{F}$ from a known patient the transform can be directly applied. Cross-validation is carried out using a leave-one-out scheme. The flow-chart of the cross validation is given in Figure 4.6. The testing set comprises all GTSDM features from one particular patient to represent a new observation $\mathbf{F_{te}}$ while the training set contains all GTSDM features from the rest six patients as the known observation, until all features are tested. Before PCA each feature in both $\mathbf{F_{te}}$ and $\mathbf{F_{tr}}$ were re-scaled into zero mean and unit variance. This operation is called *standardization*[42, p. 567]. After PCA the elements of the inner product of two loading matrices: $\mathbf{W}$ from training set and $\mathbf{V}$ from testing set were analyzed to find the corresponding vectors in the two coefficient matrices. In the subsequent discussion, the symbol $\mathbf{w_j}$ stands for the jth PC in the loading matrix from training features while $\mathbf{v_i}$ represent the ith PC in the loading matrix from test features.

From basic linear algebra the elements of the inner products of two transform matrices: $\mathbf{\Sigma_{WV}} = \mathbf{W}' \cdot \mathbf{V}$ give the cosines of angles between the corresponding vectors in $\mathbf{V_A}$ and $\mathbf{V_B}$. If the absolute value of a particular element $\mathbf{\Sigma}(ij)$ is close to 1, then the ith column of $\mathbf{W}$ is nearly

PV	$\mathbf{v_1}$		$\mathbf{v_2}$		$\mathbf{v_3}$		$\mathbf{v_4}$		$\mathbf{v_5}$		$\mathbf{v_6}$	
	$\mathbf{w_j}$	θ	$\mathbf{w_j}$	θ	$\mathbf{w_j}$	θ	$\mathbf{w_j}$	θ	$\mathbf{w_j}$	θ	$\mathbf{w_j}$	θ
1	$\mathbf{w_1}$	3.17°	$\mathbf{w_2}$	3.06°	$\mathbf{w_3}$	5.96°	$\mathbf{w_5}$	41.83°	$\mathbf{w_4}$	47.96°	$\mathbf{w_6}$	21.10°
2	$\mathbf{w_1}$	2.64°	$\mathbf{w_2}$	2.88°	$\mathbf{w_3}$	4.20°	$\mathbf{w_4}$	14.14°	$\mathbf{w_5}$	21.36°	$\mathbf{w_6}$	24.81°
3	$\mathbf{w_1}$	3.24°	$\mathbf{w_2}$	2.36°	$\mathbf{w_3}$	7.15°	$\mathbf{w_4}$	9.90°	$\mathbf{w_5}$	5.81°	$\mathbf{w_6}$	11.30°
4	$\mathbf{w_1}$	4.77°	$\mathbf{w_2}$	8.75°	$\mathbf{w_3}$	9.38°	$\mathbf{w_5}$	25.47°	$\mathbf{w_6}$	29.95°	$\mathbf{w_4}$	37.52°
5	$\mathbf{w_1}$	4.93°	$\mathbf{w_2}$	4.26°	$\mathbf{w_3}$	8.44°	$\mathbf{w_4}$	9.97°	$\mathbf{w_5}$	6.46°	$\mathbf{w_6}$	13.44°
6	$\mathbf{w_1}$	2.83°	$\mathbf{w_2}$	4.45°	$\mathbf{w_3}$	6.52°	$\mathbf{w_4}$	9.03°	$\mathbf{w_5}$	12.33°	$\mathbf{w_6}$	13.24°
7	$\mathbf{w_1}$	1.71°	$\mathbf{w_2}$	3.92°	$\mathbf{w_3}$	8.82°	$\mathbf{w_4}$	15.40°	$\mathbf{w_5}$	19.08°	$\mathbf{w_6}$	14.68°

Table 4.1: *Testing the consistency of the loadings of PCs by cross-validation. The Loadings of training set and testing set formed two vectors. The minimal angle between the ith loading vector $\mathbf{v_i}$ from testing set and a loading vector $\mathbf{w_j}$ from training set was calculated and listed. It can be seen that for PC 1, 2, 3 the loadings were very consistent.*

orthogonal to the jth column of $\mathbf{V}$. Further, if for the first r columns of $\mathbf{W}$ there are corresponding columns in $\mathbf{V}$ with $\mathbf{\Sigma(i, j)}$ close to 1, the two transform matrices are consistent[112].

Table 4.1 shows the smallest angle between two PCs $\mathbf{w_j}$ and $\mathbf{v_i}$ for $i, j = 1, ..., 6$. For $i = 1, 2, 3$ the loading vector $\mathbf{v_i}$ has a very small angle with $\mathbf{w_i}$, indicating the loadings of the first three PCs are consistent. For $i = 4, 5, 6$ the loading vector $\mathbf{v_i}$ has a unique corresponding loading vector $\mathbf{w_j}, j \in \{4, 5, 6\}$, indicating the loadings of the 4,5,6 PCs are not as consistent.

Furthermore if the absolute values of the eigenvalues of $\mathbf{\Sigma_{WV}}$ are close to 1, then the two matrices $\mathbf{W}$ and $\mathbf{V}$ are indeed in the same linear space. The third to fifth column in Table 4.2 record the minimum, maximum and average absolute eigenvalues. It can be seen that they are very close to 1, which suggests $\mathbf{W}$ and $\mathbf{V}$ are in the same linear space. This suggests that the results in Table 4.1 indicate that the loadings of the 4,5,6 largest PCs are not as consistent as the loadings of the 1,2,3 largest PCs.

After demonstrating that $k = 6$ PCs are sufficient to represent the original 56 GTSDM features and the loadings are indeed consistent, the PCA was applied to all the features to obtain a loading $\mathbf{V}_{GTSDM}$, The six largest PCs: $\mathbf{Z}(1, ...6) = \mathbf{F}\mathbf{V}_{GTSDM}(1, ..., 6)$ were then used as the input to the feature selection algorithm.

P	Variation Explained	Min $\|\lambda\|$	Max $\|\lambda\|$	Average $\|\lambda\|$
1	96.96%	0.9685	0.9996	0.9879
2	96.39%	0.8640	0.9989	0.9727
3	96.47%	0.9945	0.9994	0.9969
4	96.49%	0.9716	0.9978	0.9887
5	95.58%	0.9786	0.9970	0.9846
6	96.28%	0.9927	0.9996	0.9967
7	96.60%	0.9940	0.9996	0.9961

Table 4.2: *The patient number (P) as testing set. PCA is performed on 56 GTSDM and the percentage of variation explained by the six largest PCs are given in the second column. The loadings of the six largest PCs from training set* $[\mathbf{v_1}, ..., \mathbf{v_6}]$ *are then multiplied with the loadings of the six largest PCs from validating sets* $[\mathbf{w1}, ..., \mathbf{w_6}]$*. The absolute values of eigenvalues* λ *of* $[\mathbf{v_1}, ..., \mathbf{v_6}]' \cdot [\mathbf{w1}, ..., \mathbf{w_6}]$ *were studied. For different pairs of training-validating sets, the minimum, maximum and average of* $|\lambda|$ *are close to 1, indicating the loadings from different sets are in the same subspace.*

4.5 Wrapper Feature Selection

After projecting the features into the six-dimensional principal domain the focus is then to find a subspace of the principal domain in which the features show the best distinction of the two classes: GTV and non-GTV. If feature selection is carried out using the induction algorithm (classifier) as a black box, by controlling the input to the induction algorithm and observing the output accuracy a wrapper feature selection can be implemented to select a subset of features which give the highest classification accuracy. Three key components are need to implement the selection: a classifier as an induction algorithm, a criterion to evaluate the classification performance, and a searching strategy, which is detailed in the following paragraphs.

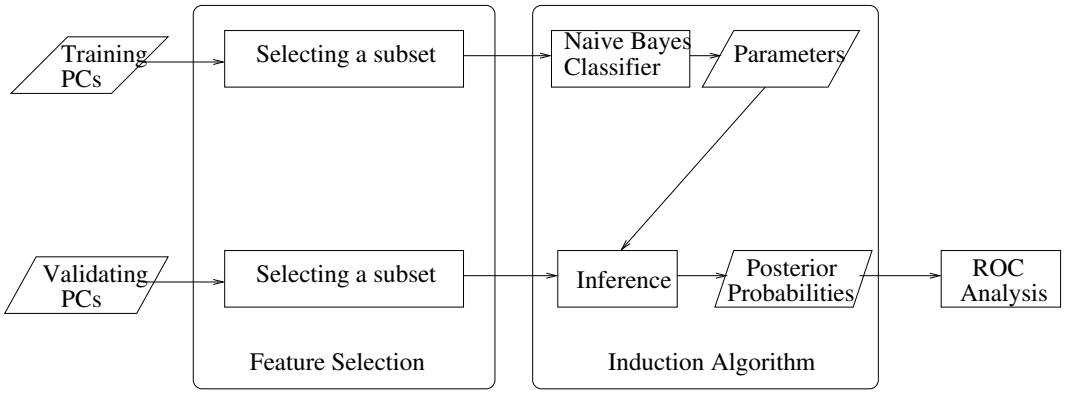


Figure 4.7: *The block diagram of the wrapper feature selection algorithm used for selecting the PCs of highest classification power. Cross validation is carried out to examine the consistency of the classification.*

4.5.1 Induction Algorithm: Naive Bayes Classifier (NBC)

In this work NBC was used as the induction algorithm. This is because Bayesian analysis is a well-established methodology and is simple to implement, and has potential to include different features into the classifier by combining posterior probabilities. The naive Bayes classifier is based on the assumption that the features are conditionally independent given the target classes T_i[113, 114]:

$$p(z_1, z_2...z_n|T_i) = \prod_{j=1}^{n} p(z_j|T_i) \tag{4.4}$$

It is reported that even if the strong conditional independence does not hold, the classifier can still give good results as long as in the experiment the posterior probabilities differ significantly[42, p. 380]. The posterior probability $p(T_i|z_1, z_2...z_k)$ under the conditional independence assumption can be derived by:

$$\begin{aligned} p(T_i|z_1, z_2...z_k) &\propto p(T_i)p(z_1, z_2...z_k|T_i) \\ &\propto p(T_i)\prod_{j=1}^{k} p(z_j|T_i) \\ &\propto \prod_{j=1}^{k} p(T_i|z_j)/p(T_i) \end{aligned} \tag{4.5}$$

Both the prior probability and the conditional probability $p(T_i|z_j)$ can be estimated from the training data. After the posterior probabilities are derived, a log-likelihood ratio (LR) can be calculated for each observation:

$$LR = \ln(p(z_m(1), z_m(2)...z_m(n)|T_0)) - \ln(p(z_m(1), z_m(2)...z_m(n)|T_1)) \tag{4.6}$$

However, in general because parameters of the conditional probabilities need to be estimated, over-fitting may occur, therefore cross validation is also necessary. Over-fitting means that the parameters may be not consistent for all data; as a result the parameters estimated from training data will characterize and classify the training data very well, but for new data the classification

results may be poor. Over-fitting occurs most often when there are far more parameters in the conditional probability model(s) than the data available. This is not the case here: there are far more ROIs than the parameters in the model. However, in order to test the consistency of the classification the PCs were again divided into a training set and validating set in a leave-one-out scheme for cross validation. A subset of all PCs from one patient was used to train the NBC; then the trained NBC was applied to the same subset of PCs in the validating set comprising PCs from the remaining patients. The classification accuracy of the validating set was used for comparison among subsets of features. The accuracy in training set was ignored to avoid potential over fitting.

By thresholding the LR, the ROIs can be classified as either as T_0 (non-GTV) or T_1 (GTV), which is called log-likelihood ratio test (LRT). The variation of thresholds reflects the trade-off between sensitivity and specificity. Traditionally sensitivity and specificity are reported separately as a means to measure classification accuracy. In ICRU Report 79[115], the ROC analysis is proposed to establish the sensitivity and specificity as the accuracy measurement for diagnosis. ROC analysis however can also be used here to measure classification accuracy.

4.5.2 Criterion: ROC analysis

As sensitivity and specificity are not independent, using a single pair to characterize the classification performance would introduce reader reporting bias[115]. By varying the threshold and plotting the obtained $(1 - specificity)$ against $sensitivity$ a ROC curve can be formed. In [115] ICRU recommended using AUROC to measure the diagnosis accuracy of different imaging modalities since it is free of bias from different choice of thresholds. A larger AUROC suggests a higher accuracy.

ROC analysis is considered the best-developed statistical tool for describing medical tests measured on continuous scales and is popular in radiology[116]. Andrews[117] used ROC curve analysis to optimize the radiation dose delivered to control treatment-induced radiation injury. Obuchowski[118] presented the application of ROC to detect if disease is truly present and used the AUROC to measure accuracy. Recently medical image classification studies reported that textural features may yield larger AUROC than intensities on images of different modalities. Yao *et al.* used the wavelet transform to establish the textural features of breast MR images. The temporal sequence of features were classified by support vector machine, the results yield larger AUROC and partial AUROC with $sensitivity \in [0.9, 1]$ than results from intensity

sequence[101]. Yu *et al.* also used AUROC to evaluation the discrimination performance of normal and abnormal head and neck cancer tissues on FDG PET/CT images[45].

In this work the ROC analysis was adopted to measure the classification power of different categories of features. By varying the threshold applied to LR different pairs of sensitivity and specificity can be obtained. Trapezoidal numerical integration was used to calculate the AUROC. It takes the curvature between two adjacent $(1 - specificity)$ - $sensitivity$ pair as linear and calculate the AUROC, which will serve as a global under-estimator of the true AUROC. In this study a total of 100 different thresholds were applied so that the ROC is smooth and the estimation error by linear approximation is negligible.

4.5.3 Searching Strategy

In general an exhaustive search is too computationally demanding for high dimensional features, however since PCA reduced the feature dimension to $k = 6$, the number of possible combinations was 53. This made the exhaustive search of all combinations of PCs reasonably efficient.

PV	Max AUROC from PCs	PCs Selected	AUROC from FOS
1	0.9176	1, 3	0.7339
2	0.9346	1, 2, 3	0.7552
3	0.9277	1, 5	0.6947
4	0.8891	1, 3	0.5887
5	0.8779	1, 3	0.6678
6	0.8816	1, 3, 5	0.6362
7	0.9230	1, 3, 5	0.7072

Table 4.3: *AUROCs using Cross-validation and Feature Selection*

4.5.4 Results and Discussions

In this study all 53 subsets of the six PCs were used as the input into the cross-validating NBC classifier to obtain 53 AUROCs of the validating set. The largest AUROCs and their corresponding subsets of PCs are given in the second and third column respectively in Table 4.3. In general the AUROC from feature selection is not the optimum AUROC that the 56 original GTSDM features can produce, but given it is consistent and taking most variation of the original features into account, it can be considered as near optimal. For comparison the means

and variances were used as these features went though the same cross-validating classification to obtain the AUROC for ROIs of each patient, which is given in the fourth column in Table 4.3. It can be seen that the PCs of GTSDM features offer much larger and more consistent AUROCs than the FOS.

The loadings of PC 1, 2, 3 and 5 are illustrated in Figure 4.8. The two most frequently selected PCs are PC 1 and PC 3. Closer examine of the loadings reveal that PC 1 comprises of mainly the average of entropy features (sum entropy, entropy, difference entropy and information measurements of correlation) from co-occurrence matrices with $d = 1, 2, 3, 4$, while PC 3 contains mainly the correlations (correlation, inverse difference moment and maximal correlation coefficient) with distance $d = 1, 2, 3, 4$. This is indicative of a very important property of the features: entropies and correlations give two uncorrelated degrees of freedom in measuring the second-order distribution of ROIs as well as high classification accuracy. This is very important in increasing the computational efficiency.

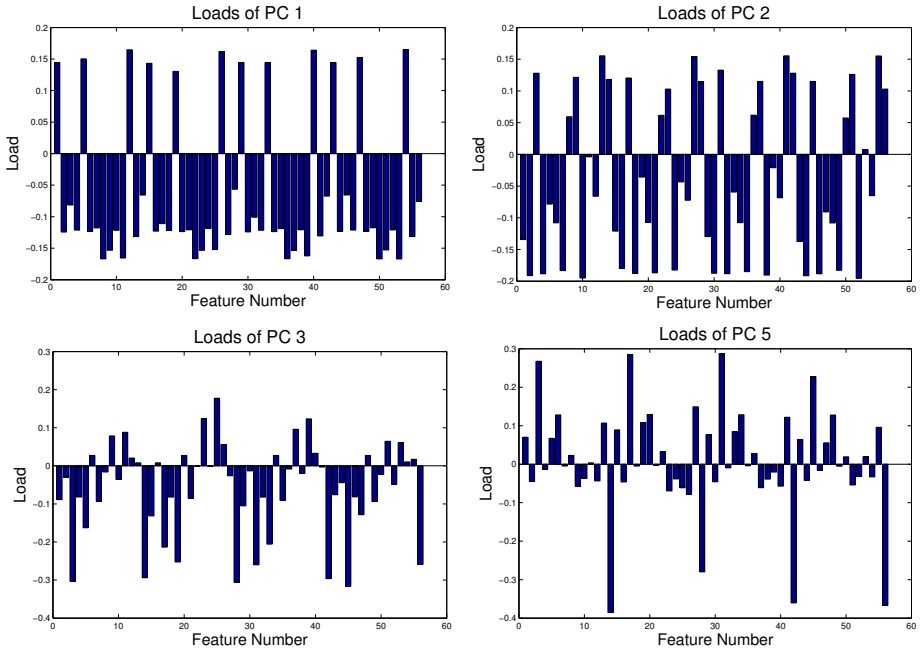


Figure 4.8: *Bar Charts: Loadings of PC 1, 2, 3, 5.*

The largest AUROCs for different patients using PCs however shows variation. For example Patients 4, 5 and 6 had consistently smaller AUROCs than the other patients, whereas Patient 2 had the largest AUROCs. The first row of Figure 4.1 shows three images from patient 2, the second and third row show three images from patient 4 and 5 respectively. From the images it

can be seen that the size of bladder of patient 2 is significantly larger than the other two patients; as a result the resolution of the bladder is also better. So strictly speaking the consistency of the AUROC depends on the resolution of the bladder in the image which is not strictly consistent. The variability of anatomy may change the distributions of the PCs which leads to the variation in classification performance. The better the bladder is resolved in the image, the larger the AUROC is. For all the seven patients used in this study, the classification accuracy was high enough to generate automatic contours.

4.6 Maximum Likelihood: Distribution of Selected GTSDM Features

After demonstrating that the PCs can offer high classification performance using Bayesian inference, a robust and efficient subset of features for classification without resorting to prior knowledge is desired. By observing the loadings of PC 1, a new feature $F1$ can be formed by calculating the difference of Feature 8: *sum entropy* and Feature 12: *entropy measurement of correlations* over $d = 1, 2, 3, 4$:

$$F1 = \sum_{i=1}^{4} (F_8(d = i) - F_{12}(d = i)) \tag{4.7}$$

Another feature $F3$ corresponding to PC 3 can just be the sum of Feature 3: *correlation* over $d = 1, 2, 3, 4$:

$$F2 = \sum_{i=1}^{4} (F_3(d = i)) \tag{4.8}$$

The F1 and F2 are scattered in Figure 4.9. The red dots indicate the features from ROIs within GTV while the blue dots indicate the features from ROIs of non-GTV. $F1$ represents the difference of two homogeneity measurements. From the scatter plot it can be seen as d is varied the GTVs had the difference of two entropies represented by $F1$ much smaller than the non-GTVs, which implies that statistically *the GTVs are more homogeneous than non-GTVs*. $F3$ alone shows limited separation power; however it assists the $F1$. For understanding the distribution

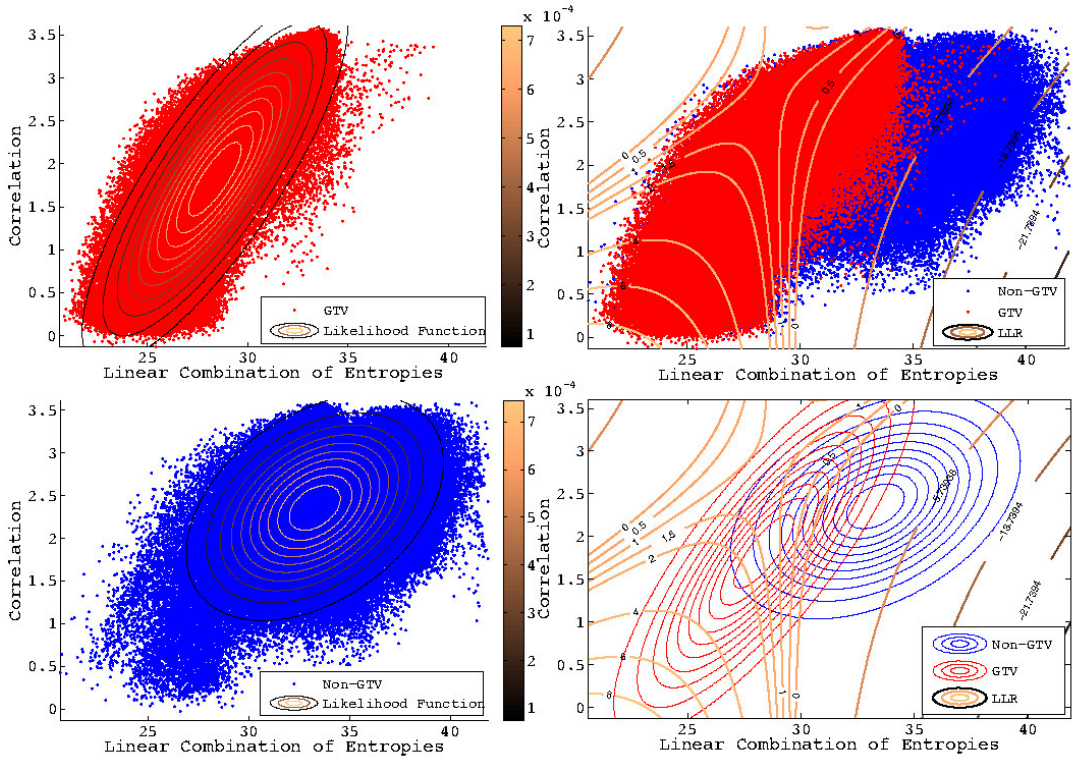


Figure 4.9: *Scatter-plot and Contour-plot: Joint-distribution of PC1 (Linear Combination of Entropies) and PC3 (Correlation). Upper Left: Scatter-plot of PC 1 and PC 3 from GTV ROIs and contour-plot by fitting Gaussian likelihood function using ML. Lower Left: Scatter-plot of PC 1 and PC 3 from GTV ROIs and contour-plot by fitting Gaussian likelihood function using ML. Upper Right: Scatter-plot of PC1 and PC3 from both GTV (red) and non-GTV(blue) ROIs. Entropy features from non-GTV ROIs were higher in general. The bronze contours stands for decision boundaries using ML criterion. Lower Right: Contour-plot by fitting Gaussian likelihood functions using ML: GTV (red), non-GTV (blue) and decision boundaries (bronze).*

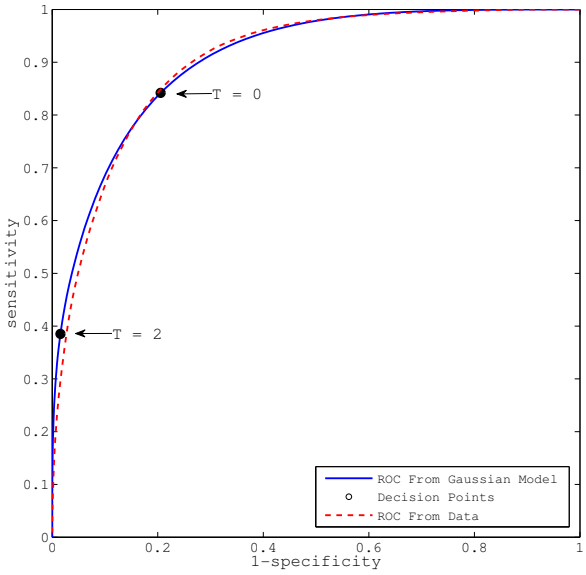


Figure 4.10: *Two ROC curves by 1. ML tests of PCs (dashed red line) 2. ML tests of Gaussian likelihood functions (solid blue line). Two ROC curves showed high degree of agreement.*

of the features, the conditional probability, or likelihood function can be estimated:

$$p(F1, F2|T_i) \sim N(\mu_i, \mathbf{\Sigma}_i) \tag{4.9}$$

where $N(\mu_i, \mathbf{\Sigma}_i)$ is the two dimensional Gaussian distribution with mean vector μ_i and covariance matrix $\mathbf{\Sigma}_i$. Assign the class T_0 to be non-GTV while T_1 to be GTV. In Figure 4.9 the two likelihood functions were plotted over the two classes of features. Subsequently a LRT can be formed using the two likelihood functions:

$$LR = \ln(p(F1, F2|T_1)) - \ln(p(F1, F2|T_0)) = \ln(N(\mu_1, \mathbf{\Sigma}_1)) - \ln(N(\mu_0, \mathbf{\Sigma}_0)) \tag{4.10}$$

Note that the LRT in Equation 4.10 is different from the LRT by NBC in Equation 4.6, since the LRT in Equation 4.10 does not account for prior probability. The thick bronze lines in the second column of Figure 4.9 indicates the thresholds derived from the LRT. A theoretical ROC curve can also be derived by integrating the true positive, false positive, true negative and false

negative probabilities. The solid blue line in Figure 4.10 plots the ROC curve from Gaussian likelihood functions. The AUROC was 0.9072. If the same thresholds, as shown in thick bronze line, were applied to the data, an empirical ROC curve can be obtained. The dashed red line in Figure 4.10 plots the ROC curve from thresholding the data. The AUROC was 0.9044. It can be seen that the two ROCs match very well, indicating that the Gaussian likelihood ratio function can explain the features very well. By fixing the threshold of LRT a particular threshold can be selected from all the thick bronze lines shown in Figure 4.9 to separate the feature space into "GTV" and "non-GTV". In the next chapter the schemes for thresholding the features and the post-processing are discussed.

4.7 Conclusions

In this chapter the homogeneous regions of soft tissues in pelvic regions were considered textural, and the local statistics were extracted from the ROIs for characterization. For clinically selected three regions of interest: bladder, rectum and a control region from each of the 59 CT images, the FOS cannot distinguish the three; it is demonstrated that two PCs from 56 GTSDM features can offer non-overlapping distinction between the three regions. Further, the ROIs of all soft tissues in the pelvic region were used for feature extraction. The purpose was to examine statistically whether principal GTSDM features can give consistently better classification power. The accuracy of classification was represented by AUROC while the consistency was tested by cross-validation. Results suggest that the GTSDM features can consistently offer larger a AUROC than FOS can, and the loadings of the PCs were consistent too. Finally by using a joint-Gaussian distribution to model two PCs: PC 1 and PC2, the AUROC from the model reconciled with the AUROC from real data.

From these results it can be concluded that the GTSDM features is superior than FOS when used for distinguishing GTV from other non-GTV soft tissue ROIs. The local textural pattern from those seemingly homogeneous ROIs contained significant amount of useful information for GTV definition. The performance of the GTSDM features in distinguishing GTV from non-GTV was consistent despite of the large anatomical difference among different patients.

While the increase classification power implies that GTSDM features is of better distinguishing power in defining GTV, it is not intuitively clear how and why this works. In the next chapter, it will be shown that the principal entropy increased the saliency of the GTV area.

Chapter 5
Entropy and Wavelet Denoising for Enhancing the Bladder-Prostate Junction

5.1 Introduction

In Chapter 4 both FOS and GTSDM features were proposed to classify soft tissue ROI into GTV and non-GTV. The results suggest that classification based on GTSDM features can give larger AUROC than classification based on FOS. GTSDM features take spatial relations of pixels into account, so they offered a larger variation in feature space. It is desirable to see how this increased variation affected the saliency of the very subtle boundaries between the GTV and non-GTV in soft tissues. In this chapter the principal second-order entropies (PE) denoted as H_p, which were demonstrated to comprise largest principal component of all features, were extracted from co-occurrence matrices of local ROIs to obtain increased saliency in BPJ. While it was reviewed in Chapter 3 that high-level modelling is very successful in bringing prior information to produce automatic contours highly similar to clinical contours, the majority of the studies mentioned that increasing image saliency will improve results[38, 119]. More importantly, because cancer may grow without a definitive shape identifying salient features to describe the GTV is more important than any prior information based approaches.

The contribution of this chapter is first in demonstrating that the PEs actually increased the saliency of the BPJ and explain the behaviour of PE. Wavelet-denoising was also applied to CT images before feature extraction to further improve the saliency of the BPJ. It was found that the principal entropy from denoised CT image (DPE) can further improve the saliency. Correspondence between the correlations C with the changes of entropies, $H_p - H$ and $H_d - H_p$, was found to demonstrate the merit of this denoising-before-extraction method: GTVs correspond to low-entropy regions and the improved saliency was because boundary and homogeneous regions were treated differently. GTV contours can be formed by thresholding DPEs. Dice similarity coefficients (DSC) and clinical review were carried out to evaluate the thresholded contours.

5.2 Data Acquisition and Pre-processing

In this chapter the subset of CT images from five male patients in Chapter 4 was studied. The setup and parameters were exactly the same. From each stack of CT images one slice at a fixed position was chosen manually for analysis. Six BPJ images of size 176×128 are shown in Figure 5.1: in the first row the three images are from the same patient at different times during the course of treatment; in the second row the three images are from three other patients. The inter- and intra-patient anatomy difference is significant. Thresholding based only on intensity produced the soft-tissue contour shown in green in Figure 5.10. This again demonstrates that the intensity is insufficient for defining the GTV. However as the GTV in T category cancer were mostly soft tissues, the soft-tissue contour was used to define the centres of ROIs for analysis: the centres of the ROIs I_c were chosen to be every 4 pixels in the soft-tissue area I_{soft}: $I_c = I(4m, 4n) \cap I_{soft}, m, n \in \mathbf{N}$.

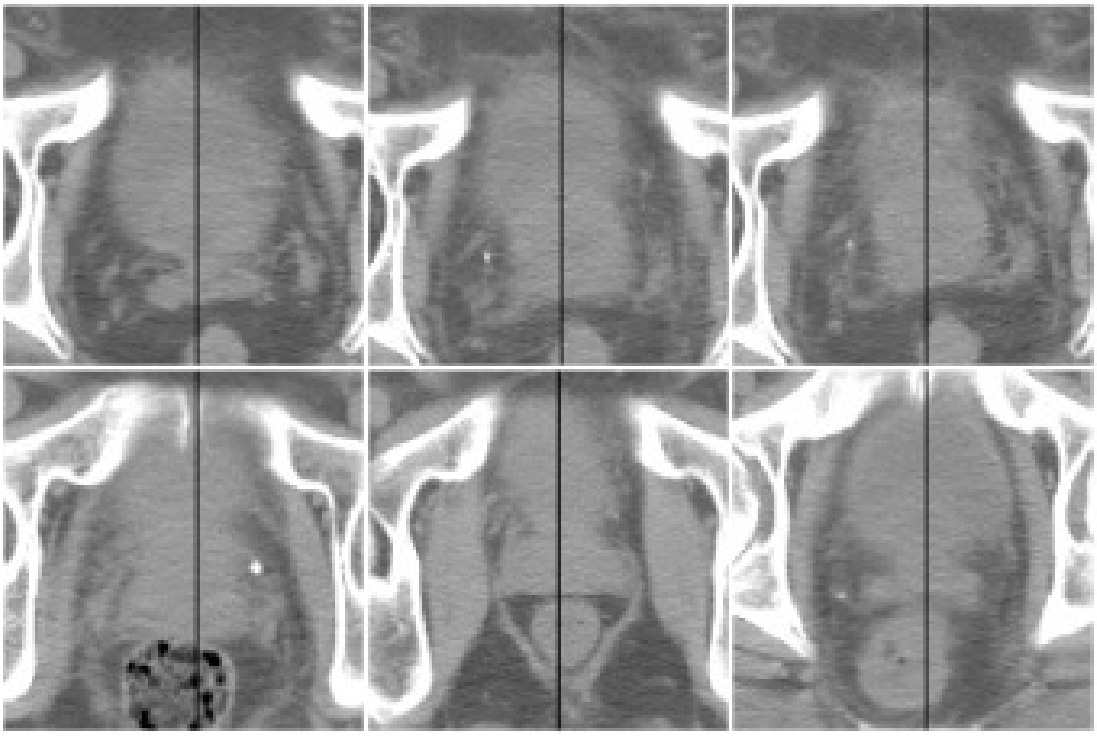


Figure 5.1: *Six CT images of bladder-prostate junction from four patients. The first row is from three images sets of one patient, the second row shows images from three other patients. Clinical GTV contours were plotted in blue. The outer green contours were produced by thresholding the image intensities. One line profile at the black vertical line stretching from top to bottom on each image was scaled into* $[0, 15]$ *and plotted in Figure 5.9.*

To further illustrate the difficulty of the GTV contouring problem in BPJ, one vertical line profile from each image at the black-bar marked position is scaled into the interval $[0, 15]$ and is plotted in red line in Figure 5.9. The boundary of bladder and prostate is very subtle and as a result the extent of GTV is difficult to determine. This corresponds to the report that the average overlap between inter- and intra-clinician contours in BPJ was between 50% and 60%in [2] .

5.3 First-order Shannon Entropy for Characterizing Image Saliency

The information entropy is used for measuring the uniformity of a distribution by the unit "bit". For an image $\mathbf{I}(x, y)$:

$$H = -\sum p(i) \log_2 p(i) \tag{5.1}$$

where i is the image intensity and $p(i)$ is the first-order distribution of i. First-order distribution considered all the data are different implements of the same random variable (*r. v.*) so that the spatial relation of pixels are not considered. In a highly complex or textured local ROI, a wider range of intensities are likely to occur and the entropy will thus be higher than that in a homogeneous ROI.

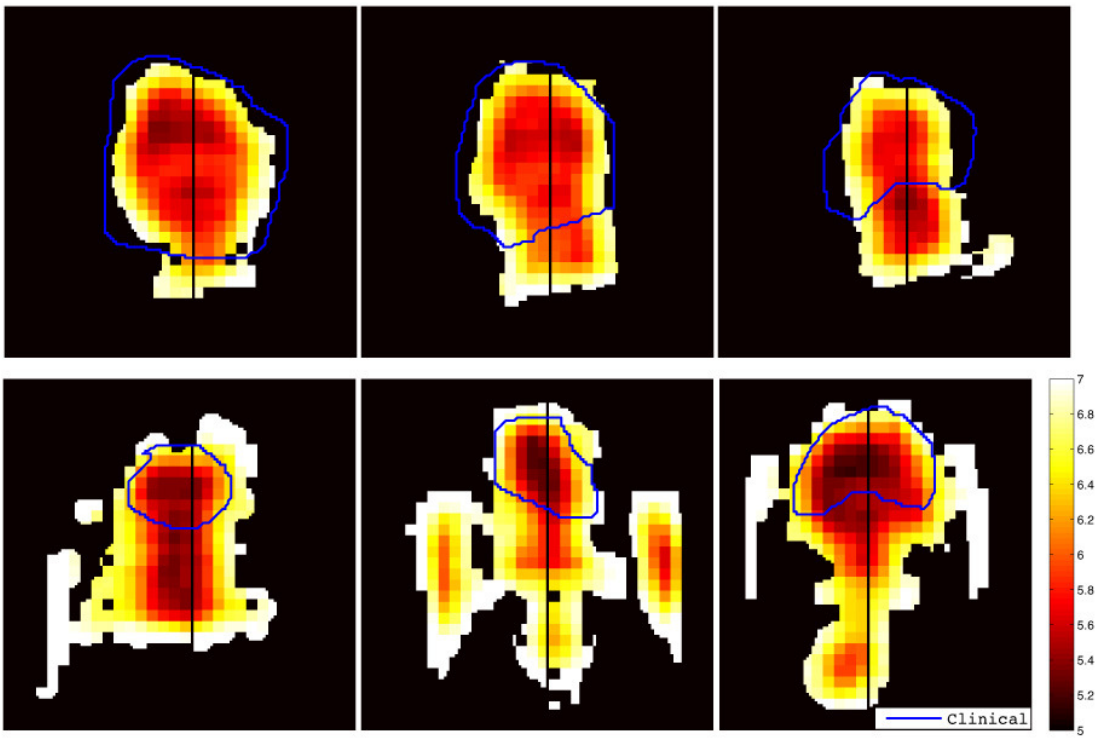


Figure 5.2: *Feature maps of BPJ: Shannon entropies. Note that the saliency is still low.*

The size and resolution of the ROI, referred to as scales, is crucial for the entropy of the ROI: for homogeneous ROIs the entropy is steady across scales whereas for heterogeneous ROIs there will be a narrow range of scales in which the entropy feature is large[54]. In medical image feature extraction, the size of the ROI is always an important issue. Clinically, a large ROI is more likely to include multiple types of tissues, forming a textured region, while a small ROI is more likely to be uniformly composed and therefore be homogeneous. The purpose of this study was to examine the bladder-prostate junction which is highly homogeneous, since a large ROI would chosen introduce unnecessary textural composition of tissues, a smaller ROI was preferred.

5.4 Second-order Entropies

Second-order entropy can be extracted from four co-occurrence matrices $p(i,j)$ of pixel distances $d = 1, 2, 3, 4$. These were used to investigate whether a higher saliency is obtained from second- rather than first-order entropies. As is demonstrated in Chapter 4 second-order entropy offered the largest variation. The sum entropy, second-order entropy and difference entropy defined in [43]:

$$H_{sum} = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log_2\{p_{x+y}(i)\}, \tag{5.2}$$

$$H(i,j) = -\sum_{i}\sum_{j} p(i,j) \log_2\{p(i,j)\}, \tag{5.3}$$

$$H_{dif} == -\sum_{i=0}^{N_g} p_{x-y}(i) \log_2\{p_{x-y}(i)\}, \tag{5.4}$$

In this study the 12 second-order entropies were first standardized into the interval $[0, 1]$ then projected onto the principal domain, based on the results in Chapter 4. The most significant principal components, explained 91% of the total variations. The second-order entropies were thus linearly combined according to their loadings to form PE H_p.

5.4.1 Correlations

In statistical applications with a limited amount of data, such as an ROI of fixed size, there is a trade-off between the statistical consistency and the length of the vector. Correlation from

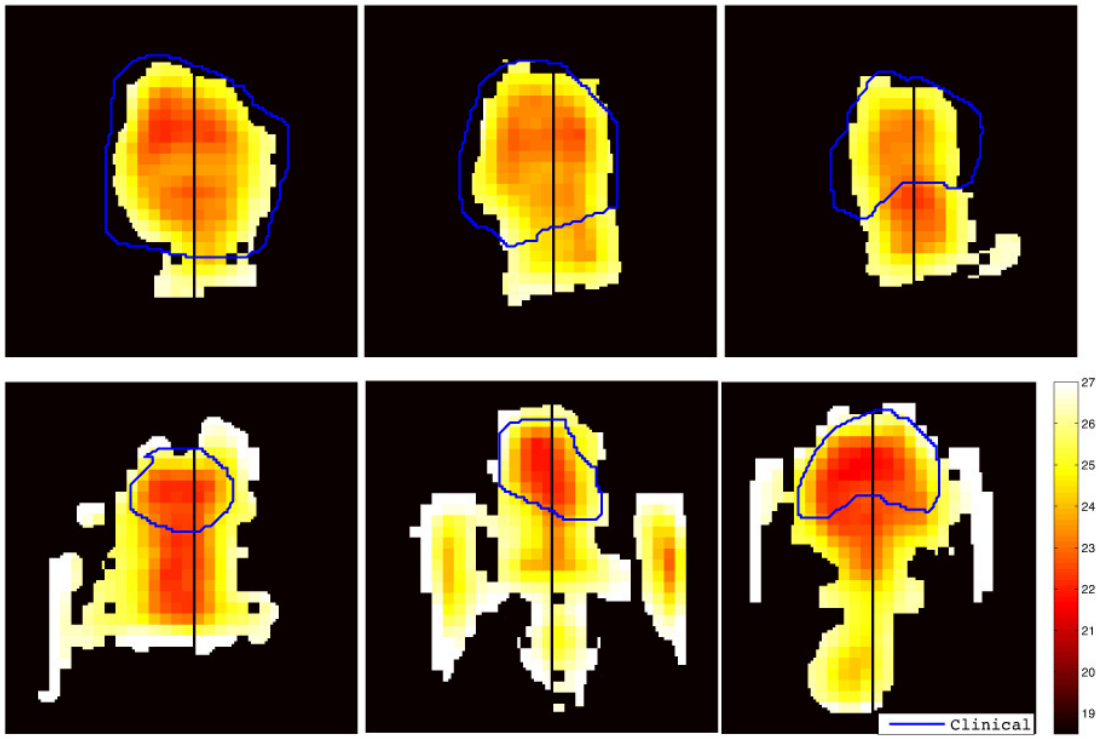


Figure 5.3: *Feature Maps of BPJ: PE. The Saliency increased significantly. However for some images it is still difficult to recognize the boundary of bladder and prostate.*

co-occurrence matrices is a projection from one scalar pixel to another scalar pixel:

$$C = \frac{\sum_i \sum_j p(i,j)(ij) - \mu_x \mu_y}{\sigma_x \sigma_y}. \tag{5.5}$$

This definition of correlation maximized statistical consistency but was a crude descriptor of frequency. Intuitively, low correlation indicates absence of structure in the signal whereas high correlation indicated the presence of a certain local structure, which may be understood as a particular spatial frequency. As will be shown later, the existence of differing correlations in the homogeneous BPJ soft tissues was key to the success of enhancing the saliency by extracting entropy features. The four correlations on the other hand were simply averaged to form a new estimation of correlation.

5.5 Second-order Entropies from Wavelet-Denoised Images

Although second-order entropies can increase the power of single features, in very homogeneous images their power is limited. It is intuitive that the noise on CT images increases the

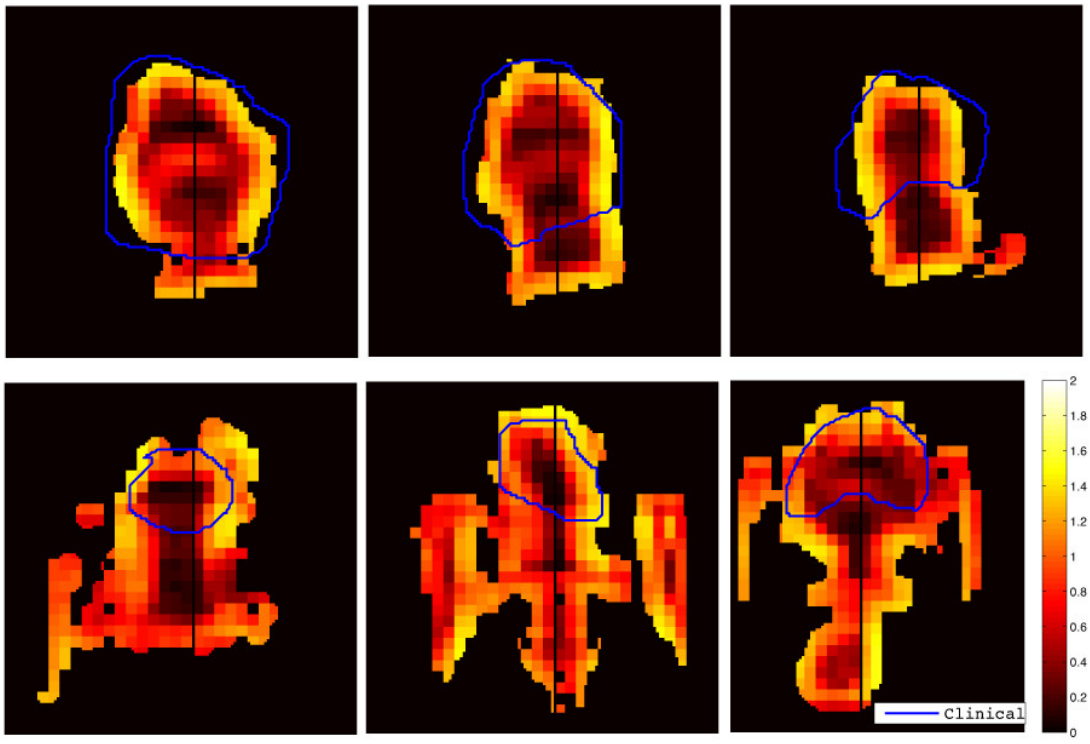


Figure 5.4: *Feature Maps of BPJ: Correlations. More structures can be observed from the correlation maps than from original soft-tissue intensities, however the variations are limited.*

non-uniformity of the ROI and as a result the distinguishing power of the entropy features decreases. Because the correlations reveal that the spatial frequency may differ across the BPJ soft tissues, it was hypothesized that by reducing the noise whilst preserving structures of high correlation that the distinguishing power of the principal entropy would be increased.

5.5.1 Analyzing Noise in CT Images Using Water Phantom

In this sub-section the water phantom shown in Section 4.3 was used to provide samples of CT noise. For comparison an image of independent and identically distributed (*i. i. d.*) Gaussian noise of the same mean and STD with the water phantom was generated. Both the water phantom and the Gaussian noise image were decomposed by Daubechies 2 stationary wavelets to level 3. This produces three levels horizontal, vertical and diagonal subbands plus a level 3 approximation of the image:

$$\mathbf{I} = \mathbf{A_0} = [\mathbf{A_3}, \mathbf{H_3}, \mathbf{V_3}, \mathbf{D_3} ... \mathbf{H_1}, \mathbf{V_1}, \mathbf{D_1}] \tag{5.6}$$

In order to estimate the STD of the noise in scale domain, the water phantom was transformed with Daubechies 2 wavelet up to scale 4. By applying Daubechies 2 wavelet iteratively to the approximation A_s in scale s, the low frequency, twice-differentiable components in A_s are as the approximation A_{s+1} in scale $s+1$; other high frequency components are in the horizontal, vertical and diagonal wavelet components: $H_{s+1}, V_{s+1}, D_{s+1}$ respectively. The means and STDs were estimated from 64 non-overlaping regions of size 30-by-30 from a 240-by-240 square region in the centre as shown in Figure 5.5 to avoid the pseudo-Gibbs phenomenon in the boundary.

For each subband the coefficients in the centre of a 240×240 square as indicated in Figure 5.5 (a) were used for estimating the STD of the coefficients in that subband in order to avoid the effect of the sharp edge. The square region was divided into 64 non-overlapping sub-squares of the size 30-by-30. The STDs of the 64 sub-squares from subbands of *i. i. d.* Gaussian noise is scattered with blue dots in Figure 5.6 (a) while the STDs of the 64 sub-squares from subbands of water phantom is scattered with blue dots in Figure 5.6 (b).

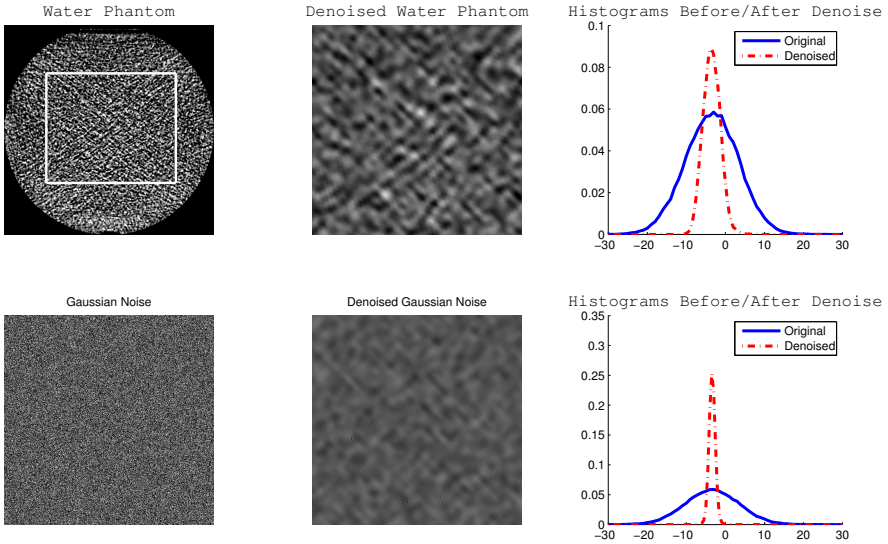


Figure 5.5: *The water phantom/Gaussian noise of the same means and variances before and after denoising, and the corresponding histograms. Gaussian noise got a larger reduction in variance because it did not have structures.*

For *i. i. d.* Gaussian noise, the STDs of the regions were scattered against scales in Figure 5.6 (a). The STDs of wavelet coefficients from Gaussian noise regions stayed at almost the same level across different scales. This corresponds to the theory that in i.i.d. Gaussian noise where there is no correlation in adjacent pixels.

For water phantom, the STDs of the regions were also scattered against scales in Figure 5.6 (b). It can be seen that the STD of the wavelet coefficients from water phantom regions increased as the scale became larger, suggesting that the noise has spatial correlation among pixels.

In order to eliminate the noise in water phantom, the mean of STDs in each scale is used as the threshold. The solid line in the right of Figure 5.5 is the histogram of the CT numbers of the water phantom. The mean of the CT numbers is $\mu_o = -3.38$ and the standard deviation (STD) is $\sigma_o = 6.83$. The gray levels are within tight interval, and the distribution is almost symmetric, centring at the mean value. The dash line in the right of Figure 5.5 plots the histogram of the grey levels of the denoised water phantom. For the FOS, the mean became $\mu_{den} = -3.39$ while the STD became $\sigma_{den} = 2.78$, which asserts that the mean remained unchanged while STD got significantly attenuated.

The denoised images are shown in the middle of Figure 5.5 while the histogram is plotted in dash line in the left of Figure 5.5. Comparing the histograms before and after denoising, it can be seen that both extreme high intensity and extreme low intensity has deteriorated. As the wavelet coefficients generally represent the high-frequency oscillations, it can be asserted that in water phantom image, both the extreme high intensity and extreme low-intensity pixels are the high frequency oscillations in wavelet space, and therefore are eliminated. After denoising the mean of the gray level was -3.38 while the STD is 2.78, significantly smaller than before denoising.

Moreover, it can be seen that both high intensity end and low intensity end of the histogram are attenuated while the density in the centre has grown significantly. As the thresholding is applied to the wavelet coefficients corresponding to high frequency fluctuation whereas the approximation in scale 5 corresponding to low frequency component is not changed, a general conclusion can be made that on water phantom, the extreme-high and extreme-low intensities which got eliminated is of high frequency up to scale 4.

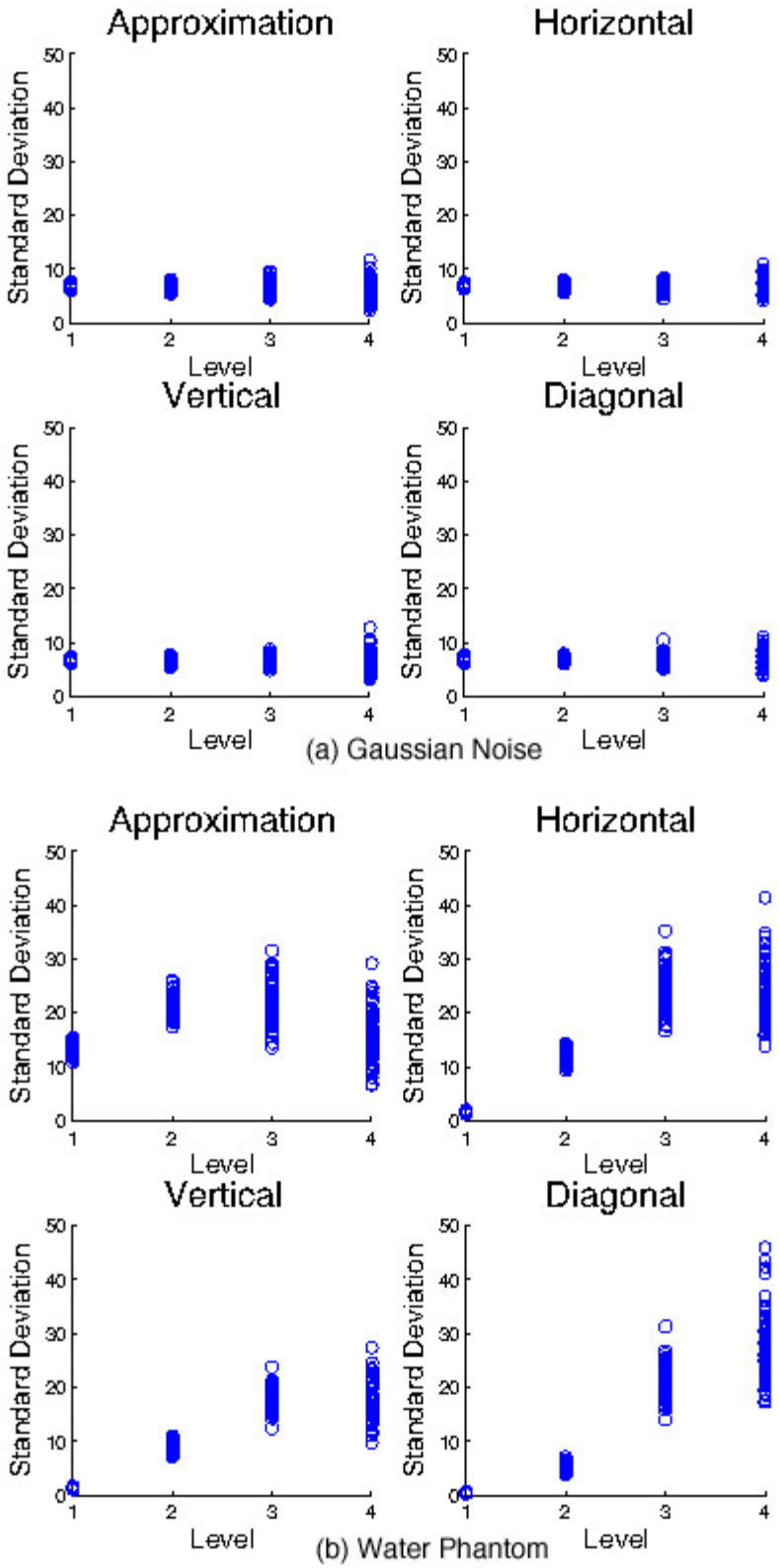


Figure 5.6: *STDs of vertical stationary wavelet components across scales. It is clear that the noise in water phantom is different from Gaussian noise.*

5.5.2 Second-order Entropies from Denoised BPJ Images and Contour Formation

After deriving the parameters (noise STD in each sub-band) for denoising CT images, the soft-thresholding in the stationary wavelet domain was carried out and the second-order entropies were extracted from ROIs to form DPE maps. Figure 5.7 shows the derived DPE map. Saliency of the GTV further increased to an extent that a simple threshold can be used to extract the low-entropy areas. The threshold selection is ad-hoc, however it can be found really easily.

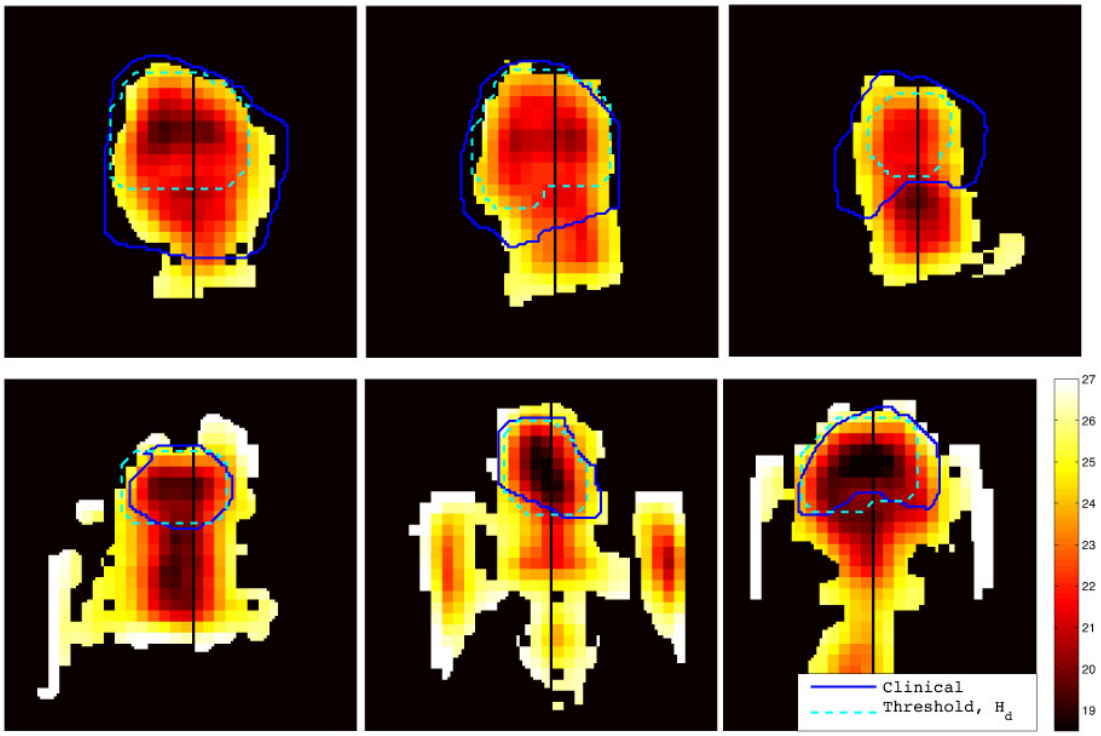


Figure 5.7: *Feature Map of BPJ: DPE. GTVs corresponded to the low-DPE areas. Saliency increased significantly and a simple thresholding can be applied to extract GTV contours.*

The GTVs in all of the examined cases correspond to areas of low entropy. The saliency of the denoised PE map was sufficiently high that a simple thresholding method could be applied to extract the low entropy ROIs as GTVs. The heterogeneity of boundary regions led to consistently smaller contours. As a result the contours were morphologically dilated by 10-pixels which is half the size of an ROI. The thresholded contours are shown Figure 5.10 (cyan).

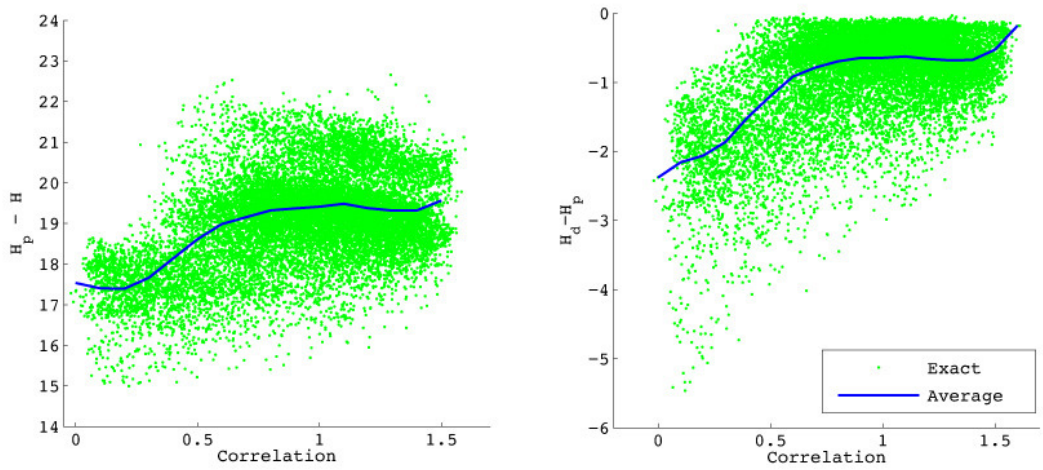


Figure 5.8: *Left: $H_p - H$ is in positive relation with C. Right: $H_d - H$ is in positive relation with C.*

5.5.3 Thresholded Contours: Evaluation Using DSC

A high degree of agreement can be observed between the clinical contours and thresholded contours. In all of the 36 cases studied, the mean DSC was 0.76 and the standard deviation (STD) was 0.13. More specifically 17 cases had a DSC above 0.8, which demonstrated high correspondence; 17 cases had a DSC between 0.8 and 0.6, which is above the inter-clinical variability threshold; 2 cases had a DSC below 0.6. Note that thresholding is the simplest method of segmentation thus more advanced models would improve the DSC further.

5.5.4 Clinical Evaluation of Difficult Cases

Six of the contours with the lowest DSCs are from images of one patient, three of the images are shown in the first row of Figure 5.10. In these cases the thresholded contours included only the bladder, whereas the clinical GTVs included full or part of prostate, resulting in small DSCs. To further evaluate the thresholded contours, all eight BPJ images from that patient with both clinical and thresholded contours were randomly presented to an experienced oncologist for grading and comment. The quality of the contours was described by “excellent”, “good”, “acceptable” and “not acceptable”. Of all eight clinical contours, four were given “excellent”, three were “good” and one “acceptable, while for thresholded contours, one was “excellent”, two were “good”, two were “acceptable” and three were “not acceptable”. This was because the extent of the cancer in these cases meant that part of the prostate was included in the GTV.

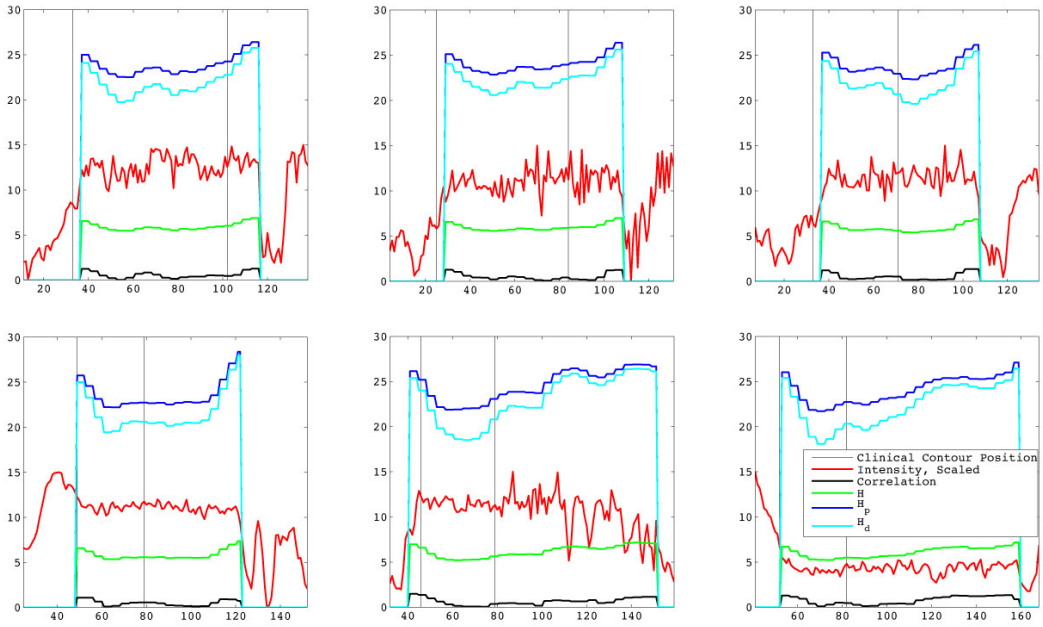


Figure 5.9: *Line Profiles of: intensity (red), correlation (black), entropy (green), PE (blue) and DPE (cyan) from the black-line marked positions. The black bars indicate the position of clinical contours.*

5.6 Discussions

5.6.1 Increase of the Saliency in Features

The saliency of the BPJ increased significantly in the DPE map. To see this clearly, Figure 5.9 shows a line profile of: original image intensity (red), correlation (black), entropy (green), PE (blue) and DPE (cyan) from the black-line marked positions in Figure 5.10 of each of the six images are plotted. The vertical black bars in Figure 5.9 shows the position of clinical contours. It is obvious that the original image intensity is very noisy and shows little structure; the entropy is smooth but there is a lack of variation.

The correlation, despite the limited variation, shows very clear structures. This suggests that the smoothness in different local ROIs of the soft tissues in BPJ is different, which implies two important results: firstly, by calculating the PE the changes of entropy compared to first-order entropy $H_p - H$ in different ROIs may act according to C; secondly, the wavelet may be able to distinguish the structures on the BPJ, and therefore $H_d - H_p$ may act according to C. Both of the two hypotheses were confirmed by Figure 5.8, which demonstrates that both $H_p - H$ and $H_d - Hp$ have a positive correlation: the higher the correlation the more probable the rise in

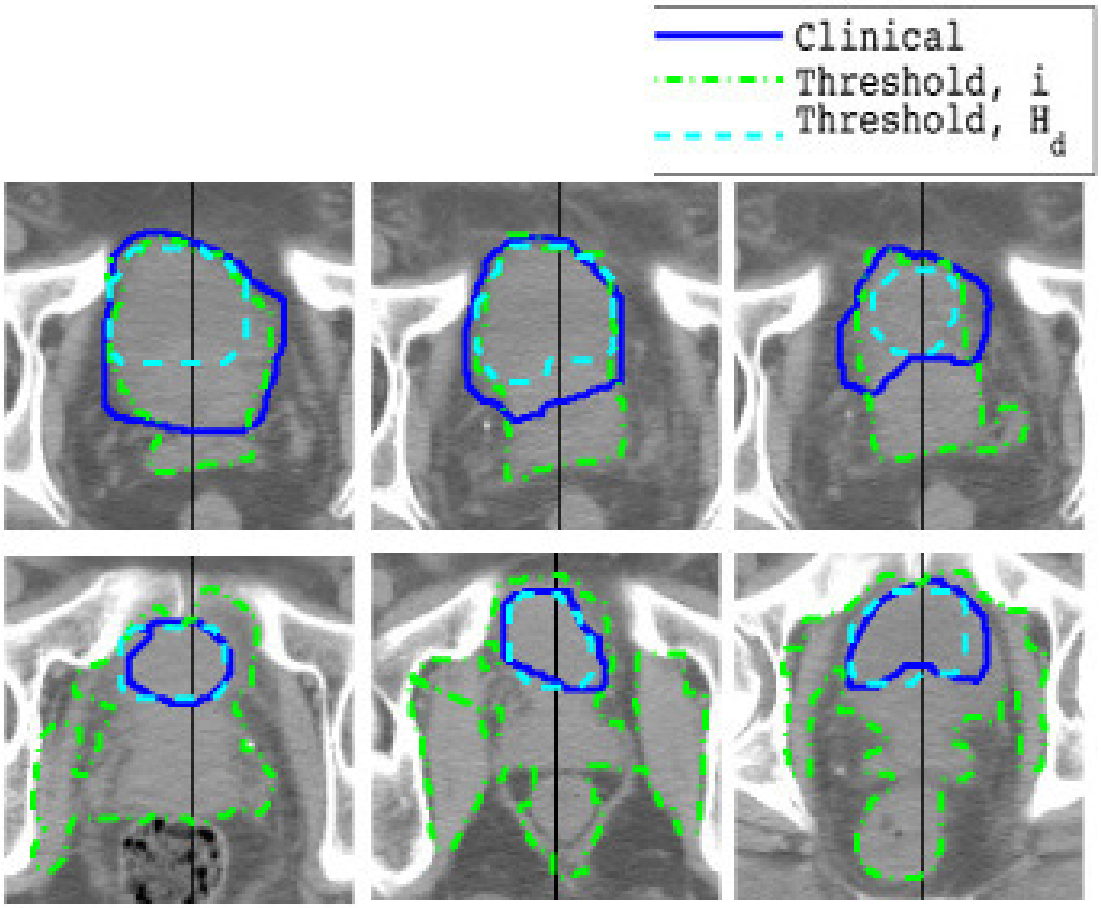


Figure 5.10: *Six CT images of bladder-prostate junction corresponding to images in Figure 5.1. The first row is from three images sets of one patient, the second row shows images from three other patients. Clinical GTV contours were plotted in blue. The outer green contours were produced by thresholding the image intensities. One line profile at the black vertical line stretching from top to bottom on each saliency map was plotted in Figure 5.9.*

PE and because the PE can treat ROIs of different structures differently the saliency increases. Similarly, the reduction in DPE over PE was because of the relation to the correlation: the difference $H_d - H_p$ is plotted against C_p in the right of Figure 5.8. Soft thresholding of wavelet coefficients reduced the entropy more significantly when the ROI had a low correlation.

5.7 Conclusions and Future Works

In this chapter the shannon entropy, PE, correlation and DPE were calculated for characterizing the very poorly resulted bladder-prostate junctions in CT images. As is demonstrated in Chapter

4, the increased variation in PE made the saliency of GTV much better than it can be seen from the original images. It is thus speculated that it is because in the homogeneous regions there are some useful low-frequency components which is essential to define the GTV, while there are some high-frequency nuisance which deteriorated the definition. The stationary wavelet was thus employed to decompose the images. Soft-thresholding scheme is applied to the wavelet coefficients to reduce the high-frequency components. The GTV in DPE maps were shown of better saliency. The low entropy regions were extracted by a simple thresholding, and they corresponded well with the clinical contours in most cases. However when clinical contours included part of the prostate, GTV did not simply the low-entropy area.

It can thus be concluded that the DPE can offer higher saliency in defining GTV in BPJ. It can further be inferred that some of the high frequency components deteriorated the GTV definition: by soft-thresholding the wavelet coefficients such effects were controlled.

In a future study the same algorithm will be applied to the prostate-rectum junction. Moreover pathological evidence need to be sought to prove relationship between DPE maps and the extent of cancer.

Chapter 6
GTV Auto-contouring: Combining Texture Analysis, Probabilistic Atlas and Anatomical Knowledge

6.1 Introduction

In Chapter 4, the second-order statistical features were shown to have higher accuracy than the FOS in classifying GTV and non-GTV soft tissue ROIs in terms of AUROC, indicating that information on the spatial relationship of pixels will make a positive contribution to characterizing the soft tissues. In Chapter 5, the principal components of second-order features, PE and PC were demonstrated to be able to increase the saliency of the bladder-prostate junctions, which has potential to reduce the inter- and intra-clinical variability. In this chapter a fully automatic bladder cancer auto-contouring algorithm is developed based on classifying ROIs according to the posterior probabilities derived from textural features. Subsequently textural posterior probabilities were fused with a probabilistic atlas to derive a more robust classification.

With the knowledge from previous chapters two new features were built as analogies of the PC 1 and PC 3 in Section 4.6. Two typical thresholds on the ROC curve: a high sensitivity one and a high specificity one were used together with image processing algorithms as two different schemes to form automatic contours in Section 6.2. This completed a fully automatic and reproducible automatic GTV contouring system based on GTSDM features and the incorporation of geometric properties.

In Section 6.3 three evaluation methodologies were applied to examine the quality of the automatic contours. The volumes of automatic contours were first compared to the volumes of clinical contours. Subsequently to describe the overlap of the automatic contours and clinical contours the DSCs were computed for each axial CT image. Both the volumes and DSCs suggests that the high specificity scheme yielded automatic contours more similar to clinical contours. For the images of the mid-range bladder automatic contours showed a high degree of

agreement with the clinical contours, however for the superior and inferior ends the automatic contours a low degree of agreement with the clinical contours was observed. This corresponds to the clinical difficulty defined in [2] to contour the superior and inferior ends of the bladder due to the limited resolution and contrast.

The contribution of this chapter lies in using the Bayesian probability inference to combine the texture and shape information of the organ for a more statistically consistent estimation of the GTV. The naive Bayes classifier in Section 6.2 uses only textural information. It does not take into account the bladder shape information. This is far from the medical practice in which clinicians base their contours on their knowledge of anatomy. Shape and position of organs are also very important and the majority of existing pelvic organ segmentation algorithms are based on shape models.In the literature most of studies circumvent this issue by using shape prior information of prostate to make an educated guess. Bladder shape is not used because of its varying nature. However, because classification results from textural posterior probabilities were of high accuracy in distinguishing bladder position and volume, a probabilistic atlas of the bladder is first proposed to confine the extent of the bladder.

In Section 6.4 a total of 36 stacks sets from five male patients described in Section 4.2 were considered. A probabilistic atlas was constructed by rigid registration of clinical GTVs. It is shown this probabilistic atlas is of high sensitivity because it can cover all areas of the clinical contour. The automatic GTV from classifying textural posterior probabilities was registered to the atlas. In order to combine the texture and shape information Bayesian probability inference was used to combine the posterior probabilities from both shape and texture properties to improve the bladder-prostate segmentation accuracy. Moreover by combining the texture and atlas posterior probabilities under a Bayesian analysis framework, the ROC analysis suggests that the average AUROC from the combined probabilities can be still larger than any of the two posterior used alone.

6.2 Classification to Form Automatic Contours

In Chapter 4 it was demonstrated that two PCs representing entropies and correlations offered improved classification accuracy. In this section the automatic contours were generated from the classification results. Three steps are required. First the threshold of the LRT must be fixed to produce classification results. Two schemes for selecting thresholds were compared: a high

sensitivity one and a high specificity one. Image processing methodologies were applied to correct errors by exploiting the geometry of the anatomy. Finally the low-level edges of the processed classification results were used as the automatic contours.

6.2.1 Threshold Selection

The threshold selection problem is a trade-off between sensitivity and specificity: a large threshold would give high specificity while a small threshold would give high sensitivity. In diagnosis system high sensitivity is usually preferred. However as the geometry of the anatomy can help in correcting errors by classification, it would in fact also be interesting to use a high specificity threshold. In this section two thresholds based on a log-likelihood function defined in Equation 4.6 were examined: $T_{sen} = 0$ for high sensitivity and $T_{spe} = 2$ for high specificity. Both of the two points are marked on the ROC curve in Figure 4.10. Specifically, when $T_{sen} = 0$ was applied to the LRT, the sensitivity from likelihood functions was 0.8419 while the specificity was 0.7945; when $T_{spe} = 2$ was used, the sensitivity was 0.3852 while the specificity was 0.9845. The classification results from thresholding were put back to the 4×4 area defined in Equation 4.3 to form 3D binary masks representing the GTV.

6.2.2 Post-processing and Automatic Contour Formation

For all results presented so far the ROI classification was based purely on features from local textural properties, irrespective of the geometric properties of the image: the classifier was judging whether a pixel belonged to GTV by looking at a 20×20 neighbourhood around the pixel.

For classification using $T_{sen} = 0$, the isolated false positive errors can be eliminated by morphological opening. As mathematical morphology such as erosion and dilation, opening and closing was considered basic operations, it is not detailed here. For more information on this topic, [76] gives a very good description. The edge of the processed images was used as the automatic contours. Nine automatic contours for the images shown in Figure 4.1 are plotted on CT images in Figure 6.1. For the images from Patient 2 shown in the first row, the bladder is resolved well. The threshold was conservative because the GTV boundaries are tighter than the clinical contours. However for the images from patient 4 and 6, shown in the second and third row, the resolution of the bladder is poorer, and in these images there are significant volumes

of healthy tissues that have very similar texture with GTV. As a result there were lots of false positive ratio was high, which indicates the threshold was aggressive. Since the false positive cannot easily be corrected automatically, high sensitivity thresholding is not a good choice.

For high specificity classification $T_{spe} = 2$, the green contours in Figure 6.2 plot the maximum 3D six-connected GTV in each image. Compared to the automatic contours in Figure 6.1 it can be seen that heterogeneous regions are further excluded by the high specificity threshold, which confirms the argument in Section 4.6 that the GTSDM features corresponded to the homogeneity of the ROIs. Because the ROIs along the bladder-fat boundary are heterogeneous, they were classified as non-GTV. The boundary of false negative classification can thus be compensated by a 10-pixel dilation, which is half the size of an ROI. The automatic contours after dilation are also shown in Figure 6.2. Intuitively the high-specificity scheme associated with the post-processing algorithms gives better automatic contours than the high sensitivity scheme.

6.3 Automatic Contour Evaluation

The two feature classification schemes and the associated post-processing produced two sets of fully automatic contours. In this section three evaluations are presented to examine the quality of the two sets of automatic contours. First the difference between automatic contours and clinical contours was calculated. Moreover the DSCs are used to compare the similarity between the automatic contours and clinical contours. The DSCs are plotted against the normalized positions of the CT images to find the performance limit of the automatic contours. Finally an experienced clinician was invited to evaluate the automatic contours.

6.3.1 Volumes of the Automatic Contours

Figure 6.3(a) scatters the automatic GTVs from high sensitivity thresholding against clinical GTVs in the unit of cubic centimetre (CC). There are consistent volume bias associated with different patients.

$$\frac{1}{51}\sum_{i=1}^{51}|V_C(i) - V_{sen}(i)| = 24.33(CC) \tag{6.1}$$

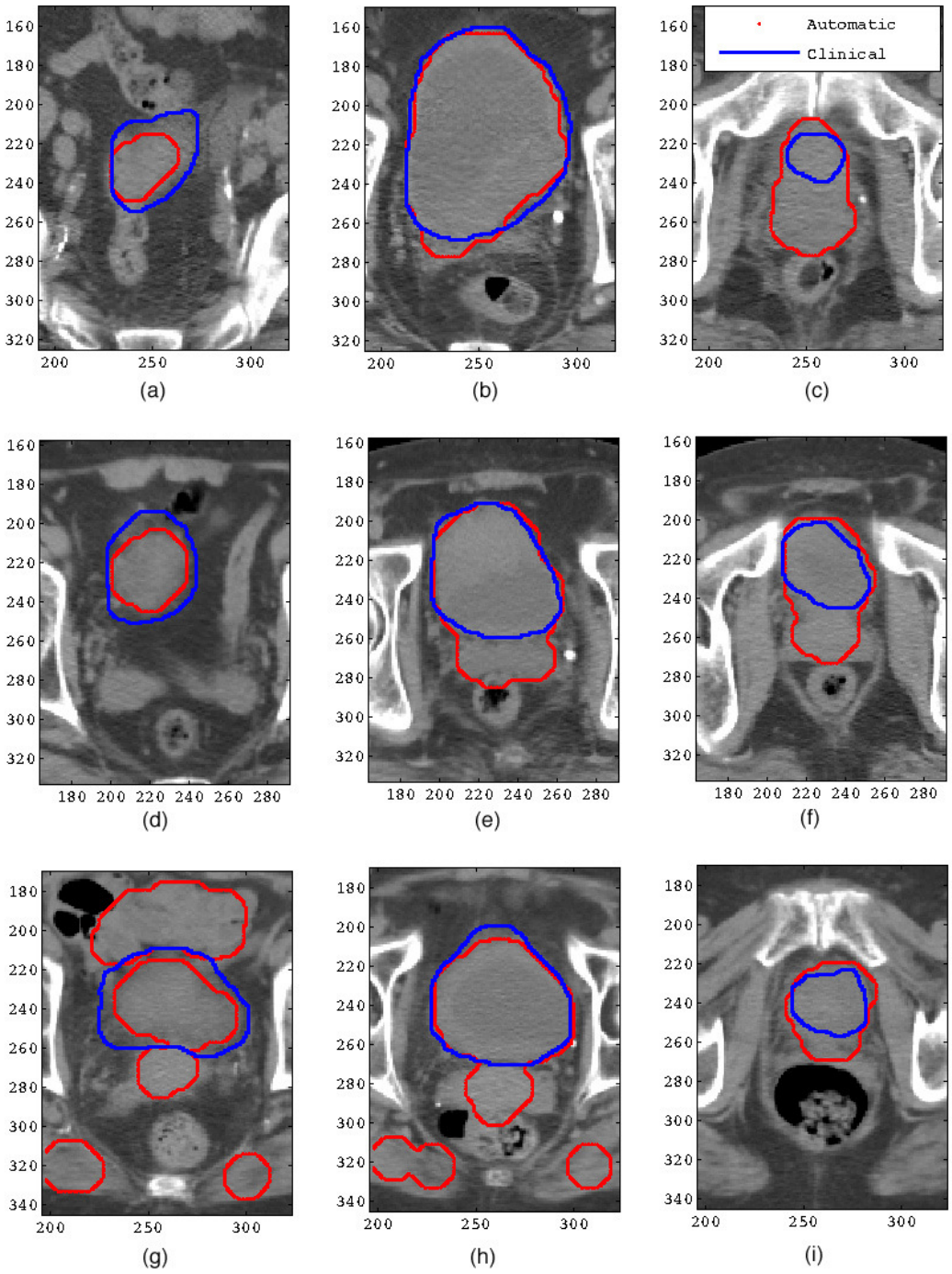


Figure 6.1: *Nine CT images from three different patients. Row 1 and Row 2 are from male patients while Row 3 is from a female patient. Column 1 is the superior part of bladder. Bladder is very insignificant and a lot of other soft tissues are of the same intensity. Column 2 is the mid part in which bladder resolves the best. Column 3 is the inferior part. Automatic contours by high sensitivity scheme followed by opening are shown. It can be seen that the automatic contours are best in mid range. Because of the High sensitivity arrangement significant non-GTV structures were contoured, which cannot be easily corrected.*

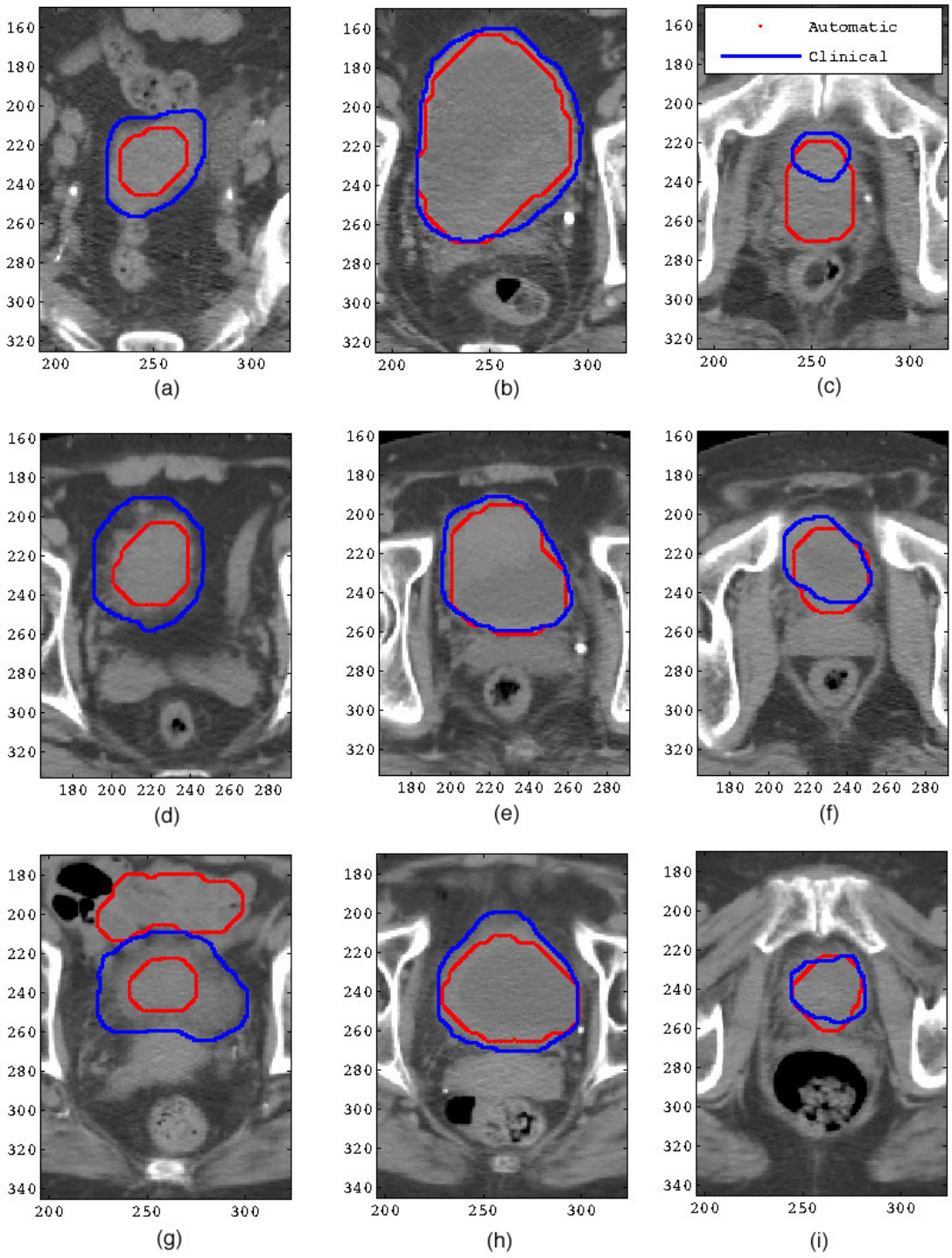
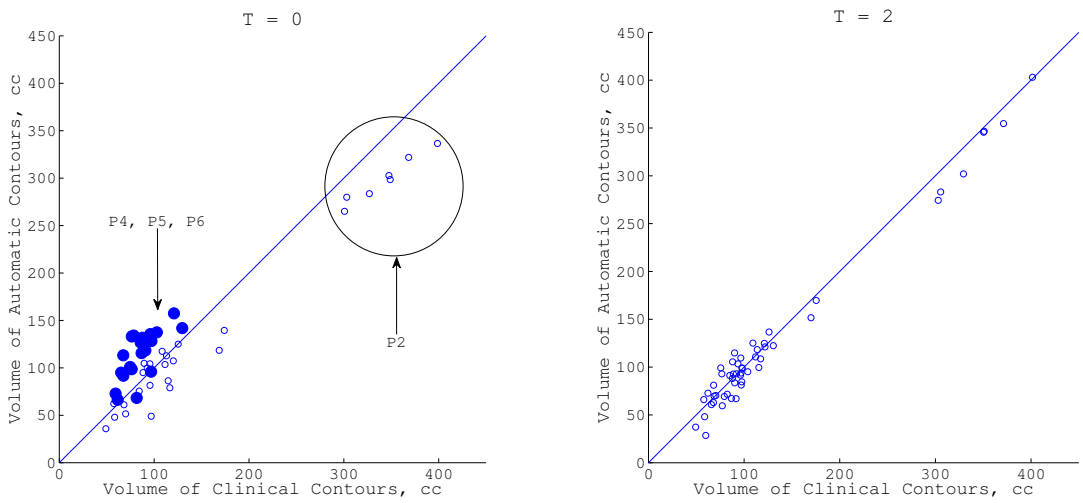


Figure 6.2: *Automatic contours from high specificity thresholding scheme after opening and dilation. A lot of false positive classifications in high sensitivity scheme shown in Figure 6.1 were eliminated, however the boundaries of GTV were too conservative. As a result a 10-pixel dilation was applied to each contour.*



(a) (b)

Figure 6.3: *Comparison of Volumes: Clinical GTV VS Automatic GTV.*

and the average difference percentage:

$$\frac{1}{51}\sum_{i=1}^{51}(|V_C - V_{sen}|/|V_C|) = 20.69\% \tag{6.2}$$

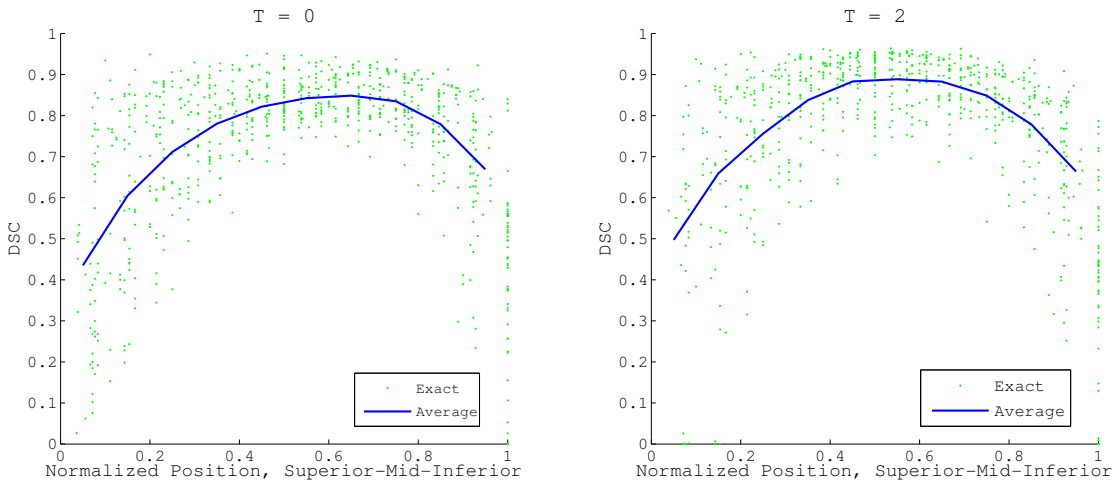
This phenomenon arised from the LRT in Equation 4.10 where the prior probability was not included. As the volume of the GTV may vary from patient to patient, the prior probability of the soft tissue being GTV was correspondingly changed. For the high specificity thresholding scheme Figure 6.3(b) scatters the volumes of automatic contours and the volumes of clinical contours. Compared to the volumes from high sensitivity scheme, the differences of the high specificity contours with clinical contours are less significant. The average absolute difference:

$$\frac{1}{51}\sum_{i=1}^{5}1|V_C(i) - V_{spe}(i)| = 10.76(CC) \tag{6.3}$$

and the average difference percentage:

$$\frac{1}{51}\sum_{i=1}^{5}1(|V_C(i) - V_{spe}(i)|/|V_C(i)|) = 10.93\% \tag{6.4}$$

The volume measurements indicate that the automatic contours from the high specificity scheme have more consistent volumes than the clinical contours.



(a) (b)

Figure 6.4: *Exact and average of DSCs of clinical contours and automatic contours across normalized positions.*

6.3.2 DSC Evaluation of Automatic Contours

DSCs was again used to describe the overlap of automatic contours with the clinical contours. The purpose of this sub-section is to reveal the relationship between classification accuracy and the position of the CT images in the axial plane.

Because the sizes of the bladders varied considerably for different patients, the positions of the images were normalized to the interval of $[0, 1]$. The first superior slice was in the position of 0 while the last inferior slice was in the position of 1, with all other slices in between. The superior slices therefore referred to the slices with the normalized positions in $[0, 0.25]$, the mid-range slices in $(0.25, 0.75]$ while the inferior slices in $(0.75, 1]$. Note this was just a coarse measure of the image positions.

For each CT image a DSC was calculated. The exact DSCs are scattered in Figure 6.4(a) for high sensitivity scheme and Figure 6.4(b) for high specificity scheme against the normalized positions. To show the variability, the positions were divided into ten intervals: $[0, 0.1)$, $[0.1, 0.2)$, *etc.* The mean of the DSCs in each interval was calculated and plotted against the positions in Figure 6.4(a) for high sensitivity scheme and in Figure 6.4(b) for a high specificity scheme.

To further examine the automatic contours, the means and standard deviation (STD) of DSCs in the superior, mid and inferior range are listed in Tables 6.1. The mean DSCs by high sensi-

Location	High Sensitivity Scheme		High Specificity Scheme	
	μ	σ	μ	σ
Superior	0.5763	0.2288	0.6464	0.2396
Mid	0.8160	0.0858	0.8643	0.0850
Inferior	0.6727	0.2145	0.6731	0.2228

Table 6.1: *Mean and variance of DSCs between automatic and clinical contours in different positions of male patients.*

tivity scheme were found to be consistently smaller than the mean DSCs of the high specificity scheme. For both schemes in the mid-range the automatic contours showed high agreement with the clinical contours. This is comparable to the performance reported by other researchers in auto-segmentation of the genitourinary organs. In particular, in a qualitative and quantitative study by six clinicians on the use of the Varian SmartSegmentation software (Varian Medical Systems, Inc., Palo Alto, CA, USA) to segment the prostate, bladder and rectum of 39 patients. It was also noted by Husykens *et al.* that agreement between the clinicians and the auto-segmentation result was higher in the middle of the bladder, with agreement typically in the 80% to 90% range. This was followed by dips in the cranio-caudal extremes as shown in Figure 1.2 [2]. In our work this fall-off in performance at the cranio-caudal extremes (superior, inferior regions) was also observed. This is shown in Figure 6.4(a) and Figure 6.4(b).

6.3.3 Clinical Evaluation

A subset of the automatic contours containing eight stacks of images from one male patient was presented to the oncologist for opinion. The set-up is that both clinical and automatic contours were plotted on the image by different colours, say red and blue, but the colour is inconsistent within one stack of image so that the clinician cannot infer the origin of the contours by colour. The evaluation results were included in Appendix C.

For superior range it turned out that the automatic contours, because they were more conservative and extended very close to the boundary of the bladder and the fat tissues, they were considered as a preferred contour to the "too-wide" clinical contours. For mid-range both clinical and automatic contours received very good remarks. For the inferior range however, the clinical contours were preferred, which was in accordance with the DSC evaluation, and confirmed the assertion that GTV definition in bladder-prostate junction is very difficult.

6.4 Refining the Classification Results by Probabilistic Atlas

The purpose of using a probabilistic atlas is to quantify the anatomical constraints into posterior probabilities to improve the results from texture analysis, especially for inferior range of bladder. The form of posterior probability is chosen because it allows the texture and atlas to give conditionally independent assessments of whether an ROI should be classified as GTV. By combining the posterior probabilities, the results may be improved.

6.4.1 Probabilistic Atlas Construction using Rigid Registration

In order to form a probabilistic atlas for a test set, all the 3-D clinical bladder masks in the training set were rigidly registered together. This was done by first randomly selecting one of the K oncologist segmented bladder binary masks as a reference, then uniformly scaling the volumes of the other K 1 3-D bladder masks to equal the reference mask, and rigidly registering the K 1 masks one-at-a-time to the reference mask. The six-degree-of-freedom rigid registration was implemented by a co-researcher, Dr Robin Steel, using the Insight Toolkit (ITK) and each attempted to maximize the overlap volume between the two masks. The red volume in Figure 6.5 represents a probabilistic atlas by registration. After a match was found, the co-registered masks were normalized to form a probabilistic atlas. Thus where all masks overlapped, the probabilistic atlas would have value 1, falling progressively to 0 where no overlap of masks existed.

6.4.2 Registering Texture Classification Results to Probabilistic Atlas

The textural classification results obtained in Sub-section 6.2.2 can be registered to the probabilistic atlas constructed in Sub-section 6.4.1. As before the atlas voxel spacing was first uniformly scaled so that it was as if the atlas had been constructed from K masks each with the same volume as the texture classification result. Rigid registration was carried out to maximize the volume integral of the atlas value over the region of the texture classification result. Figure 6.5 shows the registration results: the red mask is the probabilistic atlas, while the green mask is the texture classification results. The probabilistic atlas from the training set was smooth and in accordance with the classification results. After the registration the probabilistic atlas was re-sampled in the axial plan in accordance with the texture classification results. The probability from the atlas is considered a posterior probability: p GTV Atlas . Up to this step, for

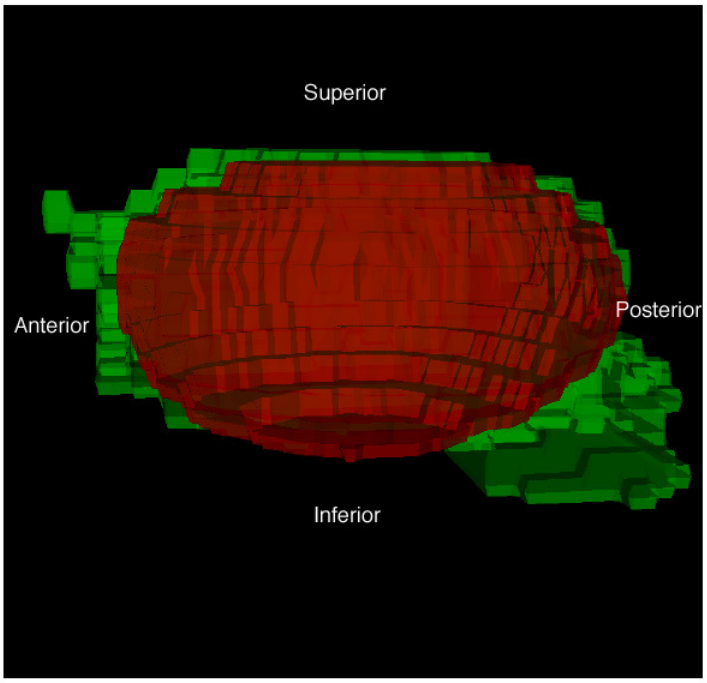


Figure 6.5: *The 3D registration of the texture classification results to the probabilistic atlas. Red: probabilistic atlas constructed by registering 3D GTVs of training data. Green: 3D GTV by classifying textural features from the testing image data. The anterior, posterior, superior and inferior ends were shown. It can be seen that towards the inferior end of the bladder false positive classification led to the inclusion of wrong regions (the prostate), and the probabilistic atlas was able to correct the false positive classification.*

a particular ROI, there were two posterior probabilities to describe how likely it is within the GTV: p GTV Texture from texture analysis, and p GTV Atlas as described in this section. The posterior probability p GTV Atlas can be thresholded to yield contours. Six atlas contours are plotted in solid line in Figure 6.6. It can be seen that there are much less false negative classifications from atlas than from texture analysis. For example in upper middle image in Figure 4, the contour from texture analysis had no overlap with the clinical contour. The atlas contour however, included part of the clinically contoured area, suggesting a high probability that the area may be bladder. In our data, the clinical contours are at least partly overlapped with the atlas contours, while there are eight non-overlapping cases by texture analysis.

The problem of false positive classification however, was much more difficult. It can be seen from Figure 6.6 that atlas contours consistently give false positive classifications. This confirms the assertion that atlas alone is insufficient for defining GTV. Moreover the malignant growth of

cancer does not have a defined shape, therefore using the atlas from the mere previous knowledge without judging the test image may not be valid even if the atlas may give a large overlap to the clinical contours. In order to take the objective measurements of the test image into account, other techniques such as texture analysis must be used together with the atlas for GTV classification. One point to note here is that the texture analysis and probabilistic atlas tended to give different false positive classifications as can be seen from Figure 6.6. This diversity is a very important property since it allows false positive classifications to cancel when probabilities are combined.

6.4.3 Combining Texture and Atlas Posterior Probabilities using a Naive Bayes Model

Bayesian analysis describes different measurements by using the same form: posterior probabilities. One particular virtue of the posterior probability is that different measurements can be easily combined to yield a more complicated measurement. In previous sections it was shown that both texture properties and shape properties can be quantized into posterior probabilities and be used for classifying GTV. In this section, the two probabilities are combined by using the naive Bayes model to infer a combined probability: .

In order to use both the texture and atlas information for classification, the posterior probability p GTV Texture Atlas needs to be inferred for each ROI. This is done by again applying the naive Bayes model: $P\{GTV|Texture\}$, $P\{GTV|Atlas\}$, $P\{GTV|Texture, Atlas\}$, $P\{GTV\}$. This implies that to obtain the combined posterior $P\{GTV|Texture, Atlas\}$ is equal to the multiplication of the two posterior probabilities: $P\{GTV|Texture\}$ and $P\{GTV|Atlas\}$ then divided by the prior $P\{GTV\}$.

The conditional independence again may not strictly hold, but the combination by multiplication is intuitively expected to improve both the false positive and false negative classifications. After obtaining the three posterior probabilities: $P\{GTV|Texture\}$, $P\{GTV|Atlas\}$ and $P\{GTV|Texture, Atlas\}$.

6.4.4 Evaluations using ROC Analysis

In this sub-section the ROC methodology is employed to evaluate the classification performance of the posterior probabilities. The ROC analysis is performed only with soft tissue ROI rather

than the whole image. This is firstly because the GTV area is small compared to the image area, leading to an overwhelmingly large true negative fraction. In addition, in our data set all the clinical contours are applied to the soft tissue. By specifying all soft tissue ROI to be the GTV, there will be no false negative fractions, indicating no GTV region will be excluded if the soft tissue regions be used as the basis population. For inferior range of male patients, the AUROC by using different features are summarized in Table 6.2.

Patient	Texture	Atlas	Atlas with Texture
1	0.67	0.81	0.79
2	0.86	0.86	0.90
3	0.86	0.89	0.91
4	0.79	0.79	0.83
7	0.78	0.75	0.77
Average	0.79	0.82	0.84
STD	0.078	0.056	0.057

Table 6.2: *AUROCs of different methods for ROIs from inferior slices of the male patients*

6.4.5 Evaluation using DSCs

Figure 6.6 shows six cases from five different patients. It can be seen that the combination of the texture and atlas posterior probabilities can partly correct both the false negative and the false positive classifications. Three points can be made here. Firstly, the contours from combined probabilities may even contain the regions which neither the texture nor the atlas contours contain, as is shown in the upper right image in Figure 6.6. Secondly, in lower middle and lower left images in Figure 6.6, the contour from combined probabilities excluded the overlapped part of the texture contour and the atlas contour. This means bringing the texture and atlas information together is more sophisticated than a simple overlap of contours. The classification benefited from the divergence of texture and atlas posterior probabilities. However, when there were limited divergence in false positive classifications, the error cannot be corrected. For example, in upper middle image in Figure 4 the prostate is still partly included.

6.5 Conclusion

In this chapter a fully automatic, reproducible bladder cancer GTV contouring system was implemented by classifying the selected GTSDM features. The approach was found to be sensitive

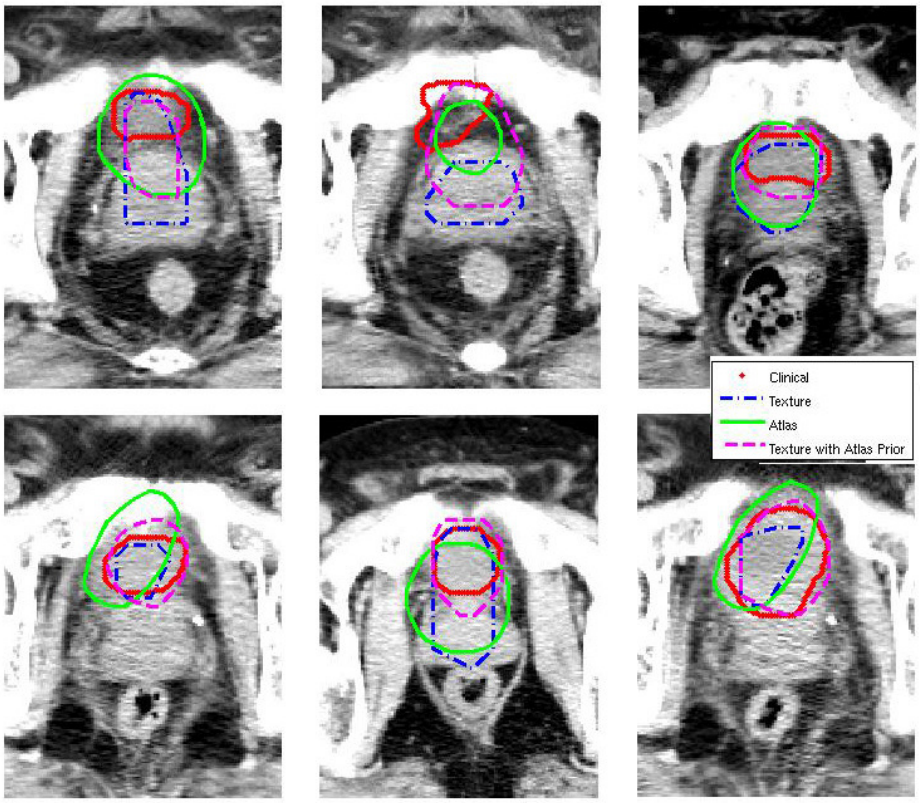


Figure 6.6: *Six automatic contours by combining textural and atlas probabilities.*

to the anatomical location of the CT, being most accurate in the nucleus of the bladder and less accurate in the superior and inferior regions. This is consistent with the problems encountered by clinicians in contouring the GTV and organs at risk in these regions. Rigid registration was used to construct a probabilistic atlas. Posterior probabilities derived from textural features and probabilistic atlases are combined under the Bayesian framework to improve the accuracy of automatic definition of the GTV.

From these results it can be concluded that by using a very simple classification scheme, objective, fully automatic and clinically promising contours can be generated. This is because of the increased distinguishing power from GTSDM features. By combining the textural and shape information, the resulting automatic contours in the inferior slices can be consistently more similar to the clinical contours.

In future research, the probabilistic atlas can be more specifically classified according to the staging of the disease to offer more accurate shape information. An entropy atlas can also be constructed by registering the entropy maps as an alternative way to combine the textural and shape information.

Chapter 7
Conclusions

In this thesis the automatic contouring of bladder cancer GTV is studied. Texture analysis is proposed to extract the textural information from the poorly resoluted/contrasted, seemingly homogeneous pelvic regions in CT images to improve the GTV definition. This research can potentially reduce the inter- and intra-clinical variability in defining the GTV, and make the GTV localization fully automatic to facilitate the IGRT, resulting in a more accurate radiation dose delivery. Chapter 1 described the context of the research, stated the objective and approach of this thesis. Texture as an alternative to the traditional gray-level for defining the GTV is proposed.

Chapter 2 further gave four major background information of the thesis. First of all the CT imaging is introduced, the gray-level in the unit of Hounsfield is described, and the intensity range of different tissues are specified. Noise in CT images is also discussed. Secondly the anatomy of human pelvis is illustrated. The coordinate system is established, and the relative position of bladder, prostate and rectum is detailed. Thirdly the bladder cancer diagnosis, staging and treatment choices are described, with an emphasis on the role of imaging in the staging. Finally the currently most popular external radiation therapy, IMRT and IGRT are discussed to show the importance of and difficulty in the radiation target localization. Reliable automatic contouring of pelvic organs are highly desired to increase the dose delivery accuracy.

Chapter 3 reviews the major methodologies of image auto segmentation/classification. The methodologies are divided into low-level feature extraction and high-level modelling. The statistical features in spatial domain are first described. Subsequently the wavelet transform and the corresponding feature extraction applications are formulated. Wavelet decomposition is an important tool for image analysis because of its compact support, but the shift-variant property makes the feature extraction in transform domain challenging. In high-level modelling, the classification algorithms is most relevant to this thesis. The over-fitting problem is asserted, and the cross-validation is considered to rule out the possibility of over-fitting. The curse of dimensionality is also discussed, and the principal component analysis is formulated as feature

reduction method. Then morphological segmentation, registration and shape modelling are reviewed; the applications of these methods to segment medical images, especially pelvic images are also reviewed.

Chapter 4 begins with the feature extraction from three selected regions in CT images: bladder, rectum and a control region. It is shown that the FOS is insufficient to distinguish the three regions, however the richness of variation in the soft tissue ROIs implies more advanced techniques can be applied to extract more distinctive features. Then the GTSDM features from four co-occurrence matrices of different distances are fed into the PCA, and the results suggest that two PCs are sufficient to distinguish the bladder, rectum and the control regions. In order to test the distinguishing power of the GTSDM features as well as their consistency, the feature selection is conducted, and the results suggest that PC 1 standing for entropies and PC 3 standing for correlations form two uncorrelated degrees of freedom, and they jointly yield larger AUROC.

Chapter 5 shows that the increased variation in PE leads to a better GTV saliency in BPJ. However for some very homogeneous regions the saliency is still not high enough. The wavelet denoising is applied to reduce the amount of high frequency oscillation, and the DPE offers still higher saliency than PE.

In Chapter 6 a fully automatic bladder cancer GTV contouring system is implemented based on the classification of textural features. It is shown that the quality of the automatic contours is related to the position of the image: for superior and mid range the automatic contours are satisfying, yet for the inferior range the contours are poor, which reconciles with the fact that GTV definition in bladder-prostate junction is very difficult. A probabilistic atlas is thus constructed to incorporate shape information with the textural information to derive a more consistent GTV contour in the inferior range of bladder.

From these results it can be concluded that the textural features can increase the saliency of the GTV in bladder cancer definition, and therefore give a higher classification accuracy. Leveraging this increased accuracy, more accurate automatic contours can be derived, and potentially the inter- and intra-observer variability can be reduced. Incorporated with the shape information, textural features are also useful for a more consistent definition of GTV.

In the future, the pathological evidence can be gathered to correspond the textural features with the cancer pathology to consolidate their relation. This is very important to get clinicians agreed on the indication provided by low-level features, and thus bring down the inter- and intra-

observer variability. Moreover a textural feature based atlas can also be constructed according to the staging of the bladder cancer for further research. This textural atlas may be superior to the intensity based atlas because of the higher saliency. Finally, more research is required to leverage multiple imaging modalities, such as isotropic multi-slice CT, MRI and PET, and extract and infuse features from all those images for a more accurate GTV definition.

Appendix A
Statistical Textural Features

A.1 First-Order Statistics:

The most commonly used first-order statistics: mean, variance and entropy

Mean:

$$\mu = \sum i \cdot p(i) \tag{A.1}$$

Variance:

$$\sigma = \sum (i - \mu)^2 \cdot p(i) \tag{A.2}$$

First-order Entropy:

$$H = \sum p(i) \cdot \log_2 p(i) \tag{A.3}$$

A.2 Second-Order Statistics from Co-occurrence Matrix

Fourteen features from co-occurrence matrix $p(i, j)$ as is defined in [43].

p_x and p_y are marginal distributions. whereas p_{x+y} is the distribution of sum of two pixels:

$$p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{N_g} \sum_{j=1}^{N_g} p(i, j), k = 2, 3, ..., 2N_g \tag{A.4}$$

similarly p_{x-y} is the distribution of difference of two pixels:

$$p_{x-y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^{N_g} \sum_{j=1}^{N_g} p(i, j), k = 0, 1, ..., N_g - 1 \tag{A.5}$$

N_g is the number of grey levels in the image.

Angular Second Moment:

$$f_1 = \sum_i \sum_j p(i,j)^2 \tag{A.6}$$

Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \{ \sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} p(i,j) \} \tag{A.7}$$

Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij - \mu_i \mu_j)}{\sigma_i \sigma_j} \tag{A.8}$$

Sum of Squares:

$$f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j) \tag{A.9}$$

$$f_1 = \sum_i \sum_j p(i,j)^2 \tag{A.10}$$

Inverse Difference Moment:

$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j) \tag{A.11}$$

Sum Average:

$$f_6 = \sum_{i=2}^{2N_g} i \cdot p_{x+y}(i) \tag{A.12}$$

Sum Variance:

$$f_7 = \sum_{i=2}^{2N_g} (i - f_6) p_{x+y}(i) \tag{A.13}$$

Sum Entropy:

$$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log_2\{p_{x+y}(i)\} \tag{A.14}$$

Entropy:

$$f_9 = -\sum_i \sum_j p(i,j) \log_2\{p(i,j)\} \tag{A.15}$$

Difference Variance:

$$f_{10} = \text{variance of } p_{x-y} \tag{A.16}$$

Difference Entropy:

$$f_{11} = -\sum_{i=0}^{N_g} p_{x-y}(i) \log_2\{p_{x-y}(i)\} \tag{A.17}$$

Information Measurements of Correlation:

$$f_{12} = \frac{H_{XY} - H_{XY1}}{max\{H_X, H_Y\}} \tag{A.18}$$

$$f_{13} = (1 - exp[-2(H_{XY2} - H_{XY})])^{\frac{1}{2}} \tag{A.19}$$

$$H_{XY} = f_9,$$

H_X and H_Y are entropies of p_x and p_y

$$H_{XY1} = -\sum_i \sum_j p(i,j) \log_2\{p_x(i)p_y(j)\}$$

$$H_{XY2} = -\sum_i \sum_j p_x(i)p_y(j) \log_2\{p_x(i)p_y(j)\}$$

Maximal Correlation Coefficient:

$$f_{14} = (\text{Second largest eigenvalue of Q})^{\frac{1}{2}} \tag{A.20}$$

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(k)p_y(k)}$$

Appendix B
Wavelet Denoising: Theory and Heuristics

B.1 Oracle Projection: Theoretical Optimal

In this sub-section the diagonal estimation problem for denoising is reviewed based on descriptions in [59]. The aim is to give the heuristic of using redundant wavelet for noise suppression in this thesis. Assume that a discrete signal f of length N is contaminated by an *i. i. d.* Gaussian noise w $N(0, \sigma)$ independent of the signal, the data can be represented by:

$$x[n] = f[n] + w[n] \tag{B.1}$$

In order to recover the signal f, the noisy data is projected into a set of vectors $\mathbf{b}$ as a transform:

$$X[m] =< x, b_m >, F[m] =< f, b_m >, W[m] =< w, b_m > \tag{B.2}$$

When the set of vectors $\mathbf{b}$ is an orthonormal basis, a diagonal estimator can be formed by multiplying the transformed data $X[m]$ with a coefficient $a_m(X[m])$ which is a function of the data, then applying the inverse transform:

$$\tilde{\mathbf{f}} = \sum_{m=0}^{N-1} a_m(X[m]) \cdot X[m] \cdot b_m \tag{B.3}$$

In a simple case, if $a_m \in 0, 1$ and an oracle is given to tell the magnitude of the unknown target function f in the transformed domain $|F|$, then a_m can become $a_m(F[m])$:

$$a_m = \begin{cases} 1 & \text{if } |F_m| > \sigma \\ 0 & \text{otherwise} \end{cases} \tag{B.4}$$

Suppose there are M non-zero terms. The MSE of this oracle projection scheme is:

$$MSE_{op} = \sum_{m|a_m=0} |F[m]|^2 + M\sigma^2 \tag{B.5}$$

the first term is the approximation error by setting $a_m = 0$ while $|F_m| \neq 0$, and the second term is the noise effects in the M remaining terms. Intuitively, if a small number of M terms in the transformed domain can give small approximation error, the denoising is very efficient. That is, the heuristic of the diagonal estimator is that the signal f is sparse in the basis $\mathbf{b}$ while the noise is dense in $\mathbf{b}$.

The MSE_{op} is the optimal in projection denoising in the basis $\mathbf{b}$ since it adopts the oracle to know acquire the unknown information of F, which is in general not available. So MSE_{op} is a benchmark to which other methods are compared, but in general it is not achievable.

B.2 Denoising by Soft Thresholding in Wavelet Domain: a Simple yet Near-Optimal Estimator

The soft-thresholding of wavelet coefficients was shown by Donoho to keep the piecewise polynomial components almost intact while removing dense noise in the images [56]. Let $\mathbf{\Psi}$ be the wavelet basis, and $\mathbf{W}$ be the wavelet coefficients of the image: $\mathbf{W}_s =< \mathbf{I}, \mathbf{\Psi}_s >$, where s is the scale. The wavelet basis $\mathbf{\Psi}$ with v vanishing moments is orthogonal to the polynomials of order $0, 1...v-1$ which means it can separate piecewise smooth structures from highly oscillating components whose order may above $v-1$. Soft thresholding t is applied to the wavelet coefficients as:

$$W' = \begin{cases} sgn(W)(|W| - |t|) & \text{if } |W| > |t| \\ 0 & \text{otherwise} \end{cases} \tag{B.6}$$

This approach treats smooth structures and oscillating noise differently, which is the merit of using the wavelet denoising method. Donoho and Johnstone proved in [55] proved that for the model in B.1 that for $N > 4$ the MSE of soft thresholding $T = \sigma(2logN)^{\frac{1}{2}}$ is:

$$MSE_{sth} \leq (2logN + 1)(\sigma^2 + MSE_{op})\text{as indicated} \tag{B.7}$$

and the factor $2logN$ is optimal among all achievable diagonal estimators in a basis $\mathbf{b}$. Moreover, in a later publication Donoho also proved the soft thresholding estimator $\tilde{f}$ is at least as smooth as f with a high probability[56].

B.3 Redundant Wavelet Noise Analysis

However, artifacts may be a problem with this approach. Since the non-redundant discrete wavelet transform (DWT) is shift variant any direct modification of the wavelet coefficients may lead to pseudo-Gibbs phenomenon. In this study a redundant wavelet transform called the stationary wavelet transform (SWT) [120] was used for denoising. The SWT, instead of downsampling the filtered image, upsamples the wavelet in each iteration. Removing downsampling results in strict shift invariance.

Further, for a redundant frame $\mathbf{d}$ with $P > N$ vectors and the frame bound A and B[59]:

$$A\|f\|^2 \leq \sum_{p=1}^{P} | < f, d_p > |^2 \leq B\|f\|^2 \tag{B.8}$$

for any $P > 4$, the MSE of soft thresholding of the redundant frame coefficients can achieve is:

$$MSE_{sth} \leq frac(2logN + 1)A(\sigma^2 + MSE_{op}) \tag{B.9}$$

This indicates that the soft thresholding of the redundant frame behaves as an averaging of A estimators in orthogonal bases, which often decreases the MSE. As a result the SWT-denoised image is smoother than the DWT-denoised image.

Appendix C
Clinical Review

EMBS.

Description/assessment.

Duncan.

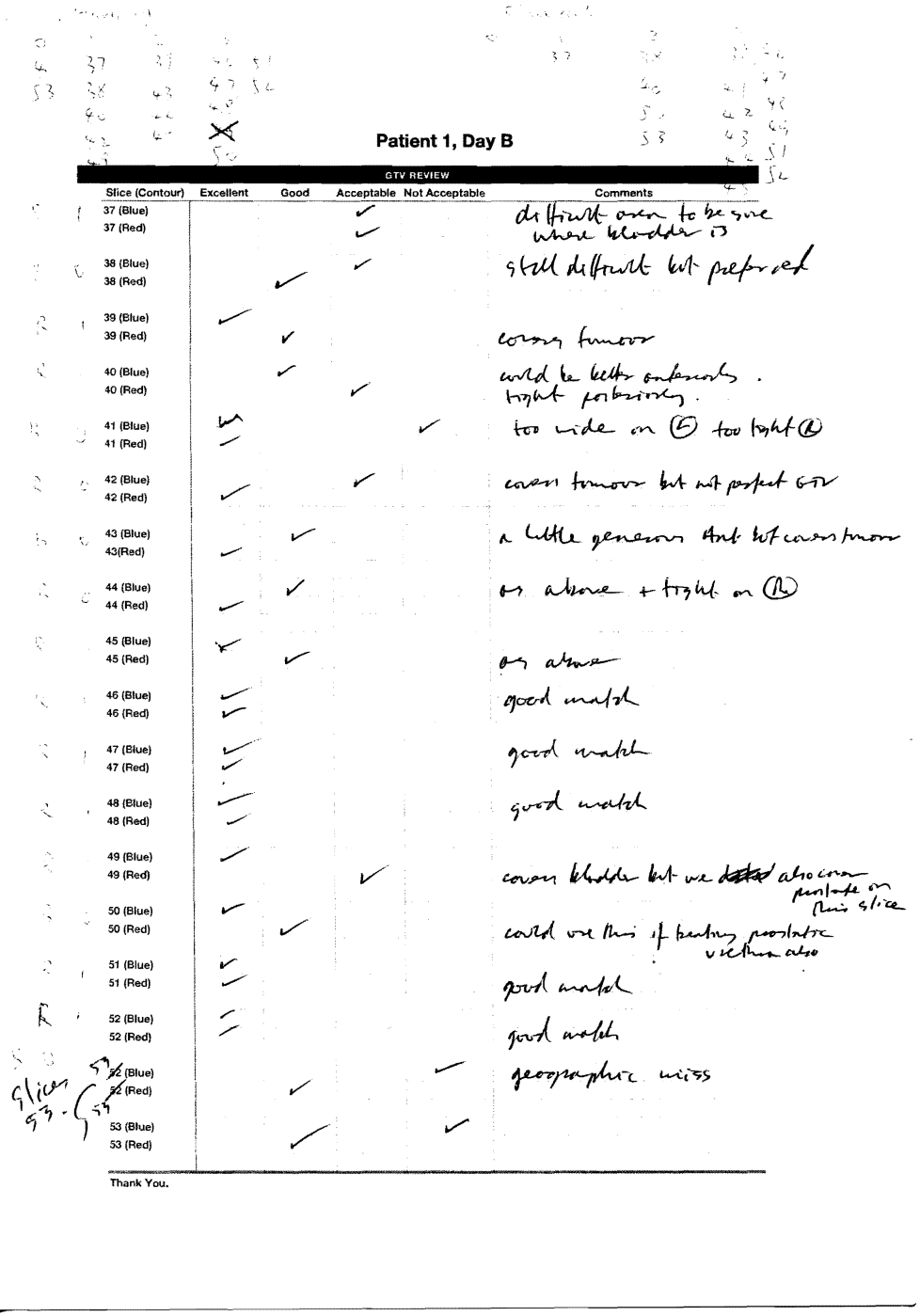
123456.

Patient 1, Day A

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
40 (Blue)	✓				
40 (Red)				✓	too small
41 (Blue)					
41 (Red)	✓			✓	post commissure too tight
42 (Blue)	✓		✓		a little tight right lat wall
42 (Red)					
43 (Blue)		✓		✓	too tight @ head girdle could have been better clearance
43(Red)					
44 (Blue)	✓				} good match
44 (Red)	✓				
45 (Blue)	✓				} good match
45 (Red)	✓				
46 (Blue)	✓				} good match
46 (Red)	✓				
47 (Blue)	✓				} good match
47 (Red)	✓				
48 (Blue)	✓				
48 (Red)		✓			a little wide on GTV
49 (Blue)	✓				
49 (Red)				✓	too tight - can preclude bowel into bladder
50 (Blue)	✓				
50 (Red)			✓		treats prostate + bladder - a little large
51 (Blue)				✓	too small
51 (Red)	✓				
52 (Blue)	✓				
52 (Red)				✓	geographic miss

Thank You.



Patient 1, Day B

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
37 (Blue)			✓		difficult area to be sure where bladder is
37 (Red)			✓		
38 (Blue)			✓		still difficult but preferred
38 (Red)		✓			
39 (Blue)	✓				
39 (Red)		✓			covers tumour
40 (Blue)		✓			would be better anteriorly. tight posteriorly.
40 (Red)			✓		
41 (Blue)	✓			✓	too wide on (E) too tight (A)
41 (Red)	✓				
42 (Blue)			✓		covers tumour but not perfect GTV
42 (Red)	✓				
43 (Blue)		✓			a little generous but but covers tumour
43(Red)	✓				
44 (Blue)		✓			as above + tight on (A)
44 (Red)	✓				
45 (Blue)	✓				
45 (Red)		✓			as above
46 (Blue)	✓				good match
46 (Red)	✓				
47 (Blue)	✓				good match
47 (Red)	✓				
48 (Blue)	✓				good match
48 (Red)	✓				
49 (Blue)	✓				
49 (Red)			✓		covers bladder but we ~~dated~~ also cover prostate on this slice
50 (Blue)	✓				
50 (Red)		✓			could use this if treating prostate + urethra also
51 (Blue)	✓				
51 (Red)	✓				good match
52 (Blue)	✓				
52 (Red)	✓				good match
53 (Blue)				✓	geographic miss
53 (Red)		✓			
				✓	

slices 53 - (53)

Thank You.

Patient 1, Day C

GTV REVIEW

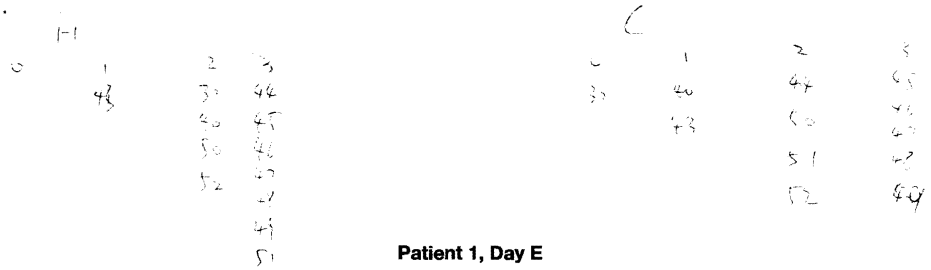
Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
40 (Blue)		✓	✓		a bit tight - post
40 (Red)					
41 (Blue)	✓				good match
41 (Red)	✓				
42 (Blue)			✓		
42 (Red)	✓				
43 (Blue)		✓			too tight on (12) pretty good match
43(Red)	✓				
44 (Blue)	✓				pretty good match
44 (Red)					
45 (Blue)	✓				good match
45 (Red)	✓				
46 (Blue)	✓				good match
46 (Red)	✓				
47 (Blue)		✓			still acceptable
47 (Red)	✓				
48 (Blue)		✓			too far ant + post: a little generous on (12)
48 (Red)		✓			
49 (Blue)	✓				probably a little small
49 (Red)		✓			
50 (Blue)		✓			
50 (Red)	✓				
51 (Blue)		✓			more difficult clinically to decide GTV
51 (Red)	✓				

Thank You.

Patient 1, Day D

GTV REVIEW					
Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
39 (Blue)		✓			difficult area for clinicians also
39 (Red)			✓		
40 (Blue)		✓			
40 (Red)				✓	too tight post
41 (Blue)	✓				
41 (Red)	✓				acceptable
42 (Blue)	✓				
42 (Red)				✓	too tight on (R)
43 (Blue)	✓				
43(Red)	✓				good match
44 (Blue)	✓				
44 (Red)	✓				good match - generous Ant
45 (Blue)	✓				
45 (Red)	✓				good match
46 (Blue)	✓				
46 (Red)		✓			generous Ant
47 (Blue)	✓				
47 (Red)				✓	too generous Ant
48 (Blue)			✓		
48 (Red)		✓			
49 (Blue)				✓	prostate bladder interface a problem
49 (Red)		✓			
50 (Blue)				✓	too small
50 (Red)		✓			
51 (Blue)			✓		neither one is perfect!
51 (Red)			✓		
52 (Blue)			✓		not great
52 (Red)				✓	too small

Thank You.



Patient 1, Day E

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
39 (Blue)		✓			
39 (Red)				✓	post too tight
40 (Blue)			✓		a bit tight - (R) side
40 (Red)		✓			
41 (Blue)					
41 (Red)		✓			) neither outline perfect
42 (Blue)			✓		
42 (Red)					
43 (Blue)		✓	✓		tight on (L)
43(Red)			✓		tight on (R)
44 (Blue)	✓				
44 (Red)		✓			
45 (Blue)	✓				improving extra ant coverage but all tumor covered
45 (Red)	✓				
46 (Blue)	✓				good match
46 (Red)	✓				
47 (Blue)	✓				generous ant but tumor covered
47 (Red)	✓				
48 (Blue)	✓				good match both ou
48 (Red)	✓				
49 (Blue)	✓				good coverage
49 (Red)	✓				
50 (Blue)		✓			bladder only
50 (Red)		✓			bladder + prostate - clinical choice of GTV
51 (Blue)	✓				
51 (Red)		✓			a little small
52 (Blue)		✓			a little small
52 (Red)		✓			a little big.

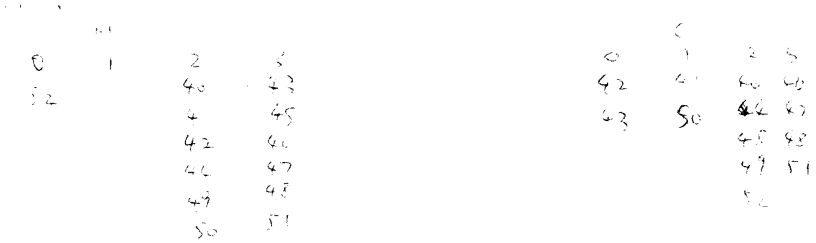
Thank You.

Patient 1, Day F

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
40 (Blue)					
40 (Red)		✓	✓		better coverage
41 (Blue)		✓			both a little high
41 (Red)		✓			Ant.
42 (Blue)			✓		too high (IL) side.
42 (Red)		✓			
43 (Blue)					
43(Red)			✓		shell a bit high
44 (Blue)		✓			
44 (Red)		✓			Improving
45 (Blue)		✓			
45 (Red)		✓			an above bit high-on (IL)
46 (Blue)		✓			
46 (Red)		✓			covers GTV
47 (Blue)		✓			
47 (Red)		✓			covers tumor
48 (Blue)		✓			
48 (Red)		✓			both cover tumor
49 (Blue)			✓		too small
49 (Red)			✓		too large.
50 (Blue)		✓			
50 (Red)		✓			too large
51 (Blue)			✓		
51 (Red)			✓		too small.

Thank You.



Patient 1, Day G

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
40 (Blue)		✓			both o/l but both could be longer.
40 (Red)		✓			
41 (Blue)		✓			
41 (Red)			✓		a bit too small
42 (Blue)				✓	too tight on (R)
42 (Red)		✓			
43 (Blue)	✓				
43(Red)			✓	✓	too tight on (R)
44 (Blue)		✓			both could be a bit more accurate
44 (Red)		✓			
45 (Blue)		✓			
45 (Red)	✓				better tumor coverage
46 (Blue)	✓				
46 (Red)	✓				pretty good match
47 (Blue)					
47 (Red)	✓				
48 (Blue)	✓				pretty good match a little loose but o/l
48 (Red)	✓				
49 (Blue)		✓			same problem. blue = bladder
49 (Red)		✓			red = bladder + prostate = what we o/l more.
50 (Blue)			✓		cover prostate
50 (Red)		✓			
51 (Blue)	✓				cover tumor
51 (Red)	✓				
52 (Blue)		✓			
52 (Red)				✓	Geographic miss.

Thank You.

Patient 1, Day H

GTV REVIEW

Slice (Contour)	Excellent	Good	Acceptable	Not Acceptable	Comments
40 (Blue)		✓			perfectly on coverage
40 (Red)	✓				
41 (Blue)	✓				
41 (Red)	✓				good coverage
42 (Blue)	✓				
42 (Red)	✓				good coverage
43 (Blue)	✓				
43(Red)	✓				good coverage
44 (Blue)		✓			a little tight- (R) Cut
44 (Red)	✓				
45 (Blue)		✓			both a little tight on (R)
45 (Red)		✓			
46 (Blue)	✓				good match
46 (Red)	✓				
47 (Blue)	✓				good match
47 (Red)	✓				
48 (Blue)	✓				good match
48 (Red)	✓				
49 (Blue)		✓			tight Ant
49 (Red)	✓				
50 (Blue)	✓				
50 (Red)			✓		quite large
51 (Blue)			✓		
51 (Red)	✓				cover OK
52 (Blue)		✓			
52 (Red)				✓	miss

Thank You.

Appendix D
Publications

- Hanqing Liao, William H. Nailon, Duncan B. McLaren and Steve McLaughlin, “Classification of Bladder Cancer on Radiotherapy Planning CT Images Using Textural Features,” in European Signal Processing Conference, 2010. EUSIPCO ’10, pp 118-122. 2010

CLASSIFICATION OF BLADDER CANCER ON RADIOTHERAPY PLANNING CT IMAGES USING TEXTURAL FEATURES

Hanqing Liao[1], *William H. Nailon*[1], *Duncan B. McLaren*[2] *and Steve McLaughlin*[1]

[1]Institute of Digital Communication, University of Edinburgh
AGB, King's Buildings, EH8 9UL, Edinburgh, UK
phone: + (44) 131 650 5659, email: h.liao@ed.ac.uk
[2]Directorate of Clinical Oncology, Western General Hospital
Crewe Road South, EH4 2XU, Edinburgh, UK phone: + (44) 1315373560, fax: + (44) 1315371092

ABSTRACT

Highly reliable classification of anatomical regions is an important step in the delineation of the gross tumour volume (GTV) in computed tomography (CT) images during radiotherapy planning. In this study pixel-based statistics such as mean and variance were insufficient for classifying the bladder, rectum and a control region. Statistical texture analysis were used to extract features from gray-tone spatial dependence matrices (GTSDM). The features were de-correlated and reduced using principal component analysis (PCA), and the principal components (PC) were classified by a naive Bayes classifier (NBC). The results suggests that the three most significant PC of the 56 features from GTSDM with distances $d = 1, 2, 3, 4$ give the highest average correct classification percentage.

1. INTRODUCTION

Accurate delineation of the GTV on CT images is vital in cancer radiotherapy planning to limit the radiation damage to normal tissue to maximize the dose to cancerous tissue. The International Commission on Radiation Units and measurements (ICRU)[1] defines the GTV as "the gross palpable or visible/demonstrable extent and location of the malignant growth." This is based on "purely anatomic-topographic and biological considerations without regard to technical factors of treatment." In the treatment of cancer by radiotherapy, CT images are used for treatment planning because they offer significant advantages over other imaging modalities such as magnetic resonance imaging (MRI). Firstly, CT images show superior consistent geometry, that is, they have less spatial distortion, consequently the volume obtained from CT images is more accurate, which is crucial for radiotherapy planning. Secondly, electron density information can easily be derived from CT images for accurate dose calculation. Thirdly, bones appears bright and in the CT images, which is important for identifying rigid landmarks and verifying set-up accuracy[20].

However, the soft tissue contrast in CT images is relatively poor compared to MR. Determination of the GTV in CT images thus demands significant clinical-experience, and is extremely time-consuming, which leads to numerous problems. Firstly, significant inter- and intra-clinical variability of GTV has been reported in literature[21]. Secondly, because treatment periods are long, many factors such as patient movement may change the position of the GTV, resulting in less than optimal treatment. Furthermore, the widespread introduction of multi-slice CT places significant pressure on clinicians because the number of images that require outlining increases significantly. In light of these issues, there is a need for a reliable, objective method to assist clinicians contouring on CT images. Several methods, such as region growing, thresholding, Markov random field (MRF) models, classifying etc. have been proposed in the literature to find the GTV automatically, and Pham et al.[15] present a review of medical image segmentation methods.

Recently texture analysis methods have been reported as offering good classification performance, which will be reviewed in Section 2. This paper extends the work by Nailon et al.[14] by investigating different feature reduction and classification strategies and modifications to the texture analysis algorithms. Texture analysis is a set of computer image processing methods aimed at extracting the information required to represent textures as textural features. In this paper, texture analysis is used to find textural features that are similar among anatomical regions with similar pathology and distinct between different anatomical regions. Figure 1(a) shows a typical CT image from a bladder cancer patient. There are three regions of interest (ROI): bladder, rectum and a control region containing multiple pathology. For radiotherapy planning, the focus is on the delivery of as much dose to the tumour as possible while limiting the dose to surrounding organs. It would therefore be desirable to classify the GTV automatically with high degree of accuracy and reliability. To this end three distinct ROI from the bladder, rectum and a control region containing multiple pathology were investigated.

Two advances to the previous approach have been made. In the first, 56 features from four co-occurrence matrices with distance $d = 1, 2, 3, 4$ were used. In addition, the PCA was used to de-correlate the obtained features. The j most significant PC are then classified by NBC. The results show that the three most significant principal features can offer high correct rate classification of bladder, rectum and control region. Secondly the cross-validation experiments are conducted so that N different images are used as training set while thre rest images are used as testing set. The average correct classification percentage are high, which suggests the method is reliable. This approach has the potential to be used as part of an algorithm for assisting clinicians delineate the GTV.

The paper is organized as follows. In Section 2 the co-occurrence matrices texture analysis method is reviewed. Section 3-5 describes the methods of GTSDM, PCA and NBC. Section 6 illustrates the results and discusses the textures at different scales on CT images: the macro-textures

and micro-textures, and how they affect the GTSDM. Section 7 concludes the paper and discusses the future work.

2. LITERATURE REVIEW: GTSDM

Famous examples of texture images have been given by Brodatz[3], however there is no accurate definition on texture. The purpose of statistical texture analysis is to describe the characteristics of textured images by features, which can be used for classification. Haralick et al. proposed a method called gray-tone spatial dependence matrices (GTSDM), or co-occurrence matrices, to classify different textures[7]. This method characterizes texture by exploring the statistical properties of the spatial dependency of a pixel with its neighbours. Features from the GTSDM are reported to have high texture classification performance in comparative studies of different texture analysis methods[17, 19, 2]. Moreover, the GTSDM are also reported successful in classifying sonar [8] and radar[4, 18, 11, 5] images. The GTSDM method is the most heavily studied texture classification approach.

In medical image processing, Hamilton et al. [6] used features from the GTSDM approach to identify focal areas of colorectal dysplasia from a background of histologically normal tissue and reported an accuracy of 86% for the training data set and 83% for the large histological scene split into smaller component images. Koss et al. [12] applied the GTSDM method to an abdominal CT image to segment 7 different organs, and reported a successful percentage of 79 - 100%. Nailon et al. studied CT images of genitourinary cancer [14], and report features from GTSDM showed the best performance in classifying bladder and rectum regions. Philips et al. [16] form and examine 3-D liver CT images and reported a variation in accuracy from 84.663% to 89.459% by changing the directions of the GTSDM.

3. GTSDM CALCULATION AND FEATURE EXTRACTION

The texture analysis using GTSDM is used to characterize different ROI. According to [7], the co-occurrence $c_i j$ is defined as a function of gray-levels i, j of two pixels with distance d from each other in direction $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. In this study it is found that there are no significant difference between GTSDM with different θ, so the four GTSDM with different θ were averaged for statistical consistency. In order to characterize the GTSDM, 14 statistical features defined in [7] were extracted from GTSDM with distance d. It is also found that different d offered additional information for classification, so in this experiment $d = 1, 2, 3, 4$ were used, and totally 56 features were extracted. In the following paragraphs, the features will be denoted as $\mathbf{x_m(k)}$, m is the total number of features for one ROI,.

4. FEATURE REDUCTION USING PCA

One problem with the statistical features defined in [7] is they may be correlated with each other. While auto-feature selection algorithms are proposed in the literatures[14], the de-correlation problem has not received much attention. In this study PCA is used to map the features into a linear subspace with minimum correlation in second-order sense. For the features $\mathbf{x_m}$ with covariance matrix $\mathbf{\Sigma}$, the the kth principal component (PC) is given by $z_k = \alpha_\mathbf{k}' \mathbf{x_m}$, where $\alpha_\mathbf{k}$ is the eigenvector corresponding to the kth largest eigenvalue of $\mathbf{\Sigma}$, and the variance of the PC z_k is the kth largest eigenvalue λ_k of $\mathbf{\Sigma}$[10]. It can thus be inferred that PCA can maximally retain variation present in the dataset while reducing the number of features. In this study, the features were first normalized to the interval $[0, 1]$ to avoid pick up features large in number to be PC. The PC of $\mathbf{x_m}$ is denoted as $\mathbf{z_m}$, and the kth most significant PC is $\mathbf{z_m(k)}$.

5. NAIVE BAYES CLASSIFICATION

NBC is based on a simple assumption that the features are conditionally independent given the target classes C_i[13, 9]:

$$P(z_m(1), z_m(2)...z_m(n)|C_i) = \prod_{j=1}^{n} P(z_m(j)|C_i) \qquad (1)$$

The PCA can maximally de-correlate the features to meet the conditional independent assumption of NBC. In existing literature, Yu et al.[22] reported good performance using PCA-NBC jointly to classify aerial images. In this study, the PCA-NBC method is employed to evaluate the classification performance of statistical textural features from different ROI. In order to do this, the posterior probability $P(C_i|z_m(1), z_m(2)...z_m(n))$ is required. This can be calculated as follows:

$$\begin{aligned} p(C_i|z_m(1), z_m(2)...z_m(n)) &\propto p(C_i)p(z_m(1), z_m(2)...z_m(n)|C_i) \\ &\propto p(C_i)\prod_{j=1}^{n} p(z_m(j)|C_i) \\ &\propto \prod_{j=1}^{n} p(C_i|z_m(j))/p(C_i) \end{aligned} \qquad (2)$$

If for $j = 1...n$, $z_m(j) \in C_i$ is assumed to have a Gaussian distribution, the distributions in (2) can be inferred from a training set. New data can then be classified by maximum a posteriori (MAP) criterion.

6. RESULTS AND DISCUSSIONS

In this study 59 CT images acquired on 8 bladder cancer patients in different days during the treatment were examined. Images were scaled so that all pixels have positive gray levels. The centers of ROI in each image were given by experienced clinician, and for each ROI a 20-by-20-pixel area is examined. The means and variances of different ROI are shown in Figure 1(b). It can be seen that the means and variances of the three ROI have significant overlap, thus they cannot yield satisfactory classification result.

For 14 features from GTSDM with distance $d = 1$, Figure 2(a) illustrates the amplitude of the eigenvalues of $\mathbf{\Sigma_{14}}$. It can be seen that there are five significant degrees of freedom in principal features, but according to Figure 2(b), the two most significant PC, $z_{14}(1)$ and $z_{14}(2)$, still cannot give a satisfactory classification result, since the distances between PC of different ROI are not large enough. Then 56 features from GTSDM with $d = 1, 2, 3 and 4$ were used for PCA, Figure 3(a) illustrates the amplitude of the eigenvalues of $\mathbf{\Sigma_{56}}$, it is shown that the degrees of freedom increase, and the amplitude of variance in each PC subspace also increases. More-

over, Figure 3(b) gives a visualization of $z_{56}(2)$ and $z_{56}(3)$. it can be seen intuitively that the three ROI can be distinguished by the two PC.

The classification performance of different number of PC from GTSDM with $d = 1,2,3,4$ were examined by using NBC. First, j most significant PC, i.e. $z_{56}(1)...z_{56}(j)$ were used for classification. Then the PC set with j variables and 59 records was randomly divided into the training group containing N records and the testing group with the rest records. By assuming all PC are conditionally independent and Gaussianly distributed, the prior probability $p(C_i)$ and the conditional distribution $p(C_i|z_{56}(k)), k = 1,...j$ can be determined from the training set, and posterior probability $p(C_{bladder}|z_{56}(1),z_{56}(2)...z_{56}(j))$, $p(C_{rectum}|z_{56}(1),z_{56}(2)...z_{56}(j))$ and $p(C_{control}|z_{56}(1),z_{56}(2)...z_{56}(j))$ can thus be calculated for each testing record. Decisions can be made by MAP criterion. For each *j-N* setup, 500 cross validation experiments were conducted to evaluate the classification performance by using different training sets. The average correct classification rate are shown in Table 1. Significant classification performance can be achieved by using PC of statistical features from GTSDM with $d = 1,2,3,4$. This further substantiates the assertion that the texture of different ROI contains important information for high-accuracy classification.

Table 1: Correct Classification Rate (%) using the j most significant PC and N Training Images

N	j = 1	j = 2	j = 3	j = 4	j = 5	j = 6
3	77	83	88	86	83	82
6	82	92	95	95	93	93
9	83	93	96	95	94	94
12	83	93	96	95	94	94
15	83	92	95	94	94	93
18	82	91	93	92	92	92

The highest classification rate occurred when $j = 3, N = 9$. By increasing the number of PC used for classification from 1 to 3, the correct percentage also increased. However, when more than 3 PC were used, the correct percentage dropped. According to the Bayesian probability rule, more variables should always increase the classification performance, but in most machine learning cases, it is impossible to know the underlying distribution of the variables, the assumed distribution will be invalid if too many irrelevant variables are used. For PCA, classification power will drop from $z_m(1)$ to $z_m(n)$ as the variances within the principal subspace decreases, therefore it is crucial to find the threshold j so that $z_m(1),...z_m(j)$ offer the best classification performance.

7. CONCLUSIONS AND FUTURE WORK

In this study the importance of GTV auto classification has been addressed. One difficulty encountered is that the contrast of the soft tissue in CT images is poor, and the means and variances are insufficient for characterization of different ROI for classification. Texture analysis was used to classify different anatomical ROI on CT images. For 59 CT images from bladder cancer patients, high correct percentage classification is achieved by using NBC to classify three most significant PC of 56 features from GTSDM with $d = 1,2,3,4$. As the GTSDM with different d describe spatial structure of the texture, the result proves that there is significant information in the texture of ROI for classification. In future work, the proposed method will be applied to the whole CT image to try to identify the ROI by features.

The proposed method follows the track of feature extraction - feature reduction process. However, the fact that only three PC were used suggests the GTSDM method requires more computational power than necessary. As PC represent the underlying latent variables, it is more promising to find the underlying latent variables directly to reduce the computational complexity. One possible approach is to define new features to characterize the GTSDM specially for ROI classification. This will also be examined in future work.

REFERENCES

[1] Prescribing, recording, and reporting photon beam therapy. Technical Report 50, Interational Commition on Radiation Units and Measurements, 7910 Woodmont Avenue, Bethesda, Maryland, 20814, USA, September 1993.

[2] J. Berry, J.R. and J. Goutsias. A comparative study of matrix measures for maximum likelihood texture classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(1):252–261, Jan/Feb 1991.

[3] P. Brodatz. *Textures: A Photographic Album for Artists and Designers*. Dover Publications Inc., 2000.

[4] L. Bruzzone, P. Pellegretti, and F. Roli. An experimental analysis of the use of grey level co-occurrence statistics for sar-image classification. *Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. 'Quantitative Remote Sensing for Science and Applications', International*, 2:1431–1433 vol.2, 10-14 1995.

[5] J. Carr and F. de Miranda. The semivariogram in comparison to the co-occurrence matrix for classification of image texture. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(6):1945–1952, Nov 1998.

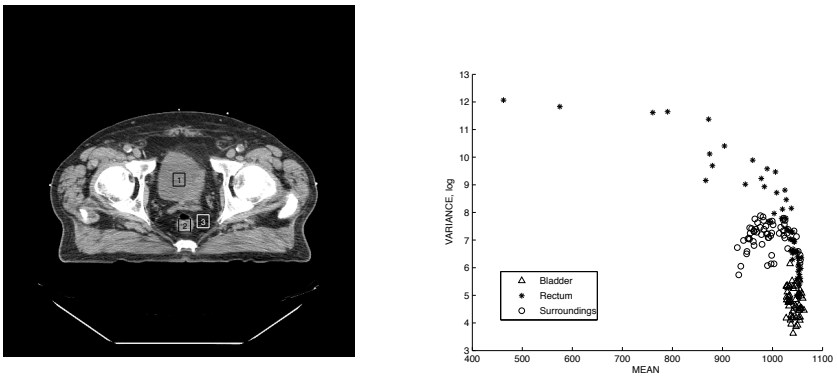
[6] P. W. Hamilton, P. H. Bartels, D. Thompson, N. H. Anderson, R. Montironi, and J. M. Sloan. Automated location of dysplastic fields in colorectal histology using image texture analysis. *The Journal of Pathology*, 182(1):68–75, 1997.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3(6):610–621, Nov. 1973.

[8] K. Imen, R. Fablet, J.-M. Boucher, and J.-M. Augustin. Statistical discrimination of seabed textures in sonar images using co-occurrence statistics. *Oceans 2005 - Europe*, 1:605–610 Vol. 1, June 2005.

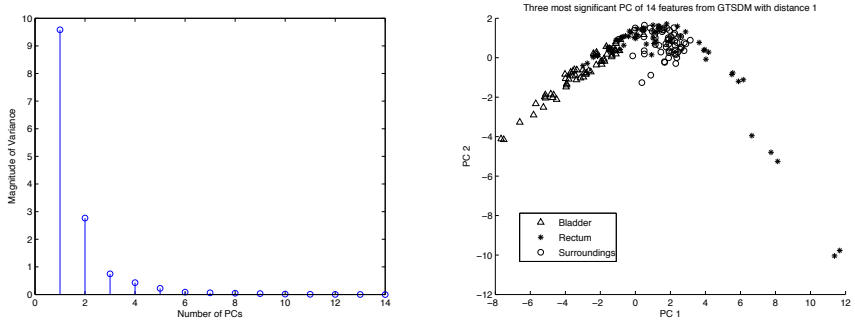
[9] W. Q. A. M. S.-A. M. Islam, M.J. Investigating the performance of naive- bayes classifiers and k- nearest neighbor classifiers. In *Convergence Information Technology, 2007. International Conference on*, pages 1541 –1546, Nov. 2007.

[10] I.T.Jolliffe. *Principal Component Analysis*. Springer, second edition edition, 2002.



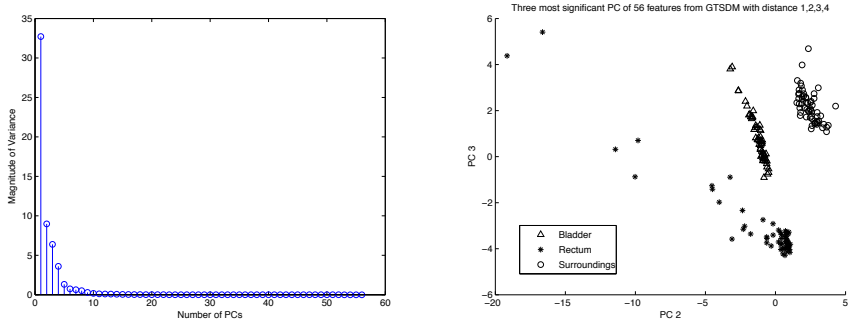
(a) (b)

Figure 1: Left (a): A typical CT Image for Radiotherapy Planning containing three Regions of Interest (ROI): 1. Bladder, 2. Rectum and 3. a control region. Right (b): The Means and Variances (log) of the Three ROI from 59 radiotherapy planning CT images. Significant overlap can be observed, so Different ROI cannot be distinguished by Means and Variances.



(a) (b)

Figure 2: Left (a): The Eigenvalues of the Covariance Matrix $\mathbf{\Sigma_{14}}$ features from GTSDM with $d = 1$, Representing the Variance in each principal subspace. Five Significant Degree-of-Freedom can be observed. Right (b): Visualizing PC 1 against PC 2: ROI still cannot be distinguished intuitively.



(a) (b)

Figure 3: Left (a): The Eigenvalues of the Covariance Matrix $\mathbf{\Sigma_{56}}$ from GTSDM with $d = 1, 2, 3, 4$, Representing the Variances in each principal subspace. Six Significant Degree-of-Freedom can be observed, Comparing with Figure 2(a), the Amplitude of Variance in Each PC subspace increases significantly. Right (b): Visualizing PC 2 against PC 3: ROI can be distinguished intuitively. Further Classification Performances using Different Number of PC are Described in Table 1.

[11] U. Kandaswamy, D. Adjeroh, and M. Lee. Efficient texture analysis of sar imagery. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(9):2075–2083, Sept. 2005.

[12] J. Koss, F. Newman, T. Johnson, and D. Kirch. Abdominal organ segmentation using texture transforms and a hopfield neural network. *Medical Imaging, IEEE Transactions on*, 18(7):640–648, July 1999.

[13] S. L. Martinez-Arroyo, M. Learning an optimal naive bayes classifier. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 958 –958, 0-0 2006.

[14] W. Nailon, A. Redpath, and D. McLaren. Texture analysis of 3d bladder cancer ct images for improving radiotherapy planning. *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 652–655, May 2008.

[15] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. In *Annual Review of Biomedical Engineering*, volume 2, pages 315–338. 2000.

[16] C. Philips, D. Li, D. Raicu, and J. Furst. Directional invariance of co-occurrence matrices within the liver. *Biocomputation, Bioinformatics, and Biomedical Technologies, 2008. BIOTECHNO '08. International Conference on*, pages 29–34, 29 2008-July 5 2008.

[17] C. H. Richard Conners. A theoretical comparison of texture algorithm. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.PAMI-2(3), May 1980.

[18] L.-K. Soh and C. Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(2):780–795, Mar 1999.

[19] J. S. Weszka and A. Rosenfeld. A comparative study of texture measures for terrain classification. *NASA STI/Recon Technical Report N*, 76:13470–+, Mar. 1975.

[20] R. Williams, J and I. Thwaites, D, editors. *Radiotherapy Physics: in Practice*. Oxford University Press, second edition edition, 2000.

[21] M. Yamamoto, Y. Nagata, K. Okajima, T. Ishigaki, R. Murata, T. Mizowaki, M. Kokubo, and M. Hiraoka. Differences in target outline delineation from ct scans of brain tumours using different methods and different observers. *Radiotherapy and Oncology*, 50(2):151 – 156, 1999.

[22] X. Yu, Z. Zheng, L. Li, and Z. Ye. Texture classification of aerial image based on pca-nbc. In L. Zhang, J. Zhang, and M. Liao, editors, *Proceedings of SPIE, the International Society for Optical Engineering*, volume 6043, page 60432G. SPIE, 2005.

References

[1] ICRU, "Prescribing, recording, and reporting photon beam therapy," Tech. Rep. 50, Interational Commition on Radiation Units and Measurements, Sept 1993.

[2] D. P. Huyskens, P. Maingon, L. Vanuytsel, V. Remouchamps, T. Roques, B. Dubray, B. Haas, P. Kunz, T. Coradi, R. Bhlman, R. Reddick, A. V. Esch, and E. Salamon, "A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer," *Radiotherapy and Oncology*, vol. 90, no. 3, pp. 337 – 345, 2009.

[3] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. Dover Publications Inc., 2000.

[4] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. SPIE Press, 2003.

[5] M. Prokop, *Spiral and Multislice Computed Tomography of the Body*. Georg Thieme Verlag, 2003.

[6] C. W. L.H. Sobin, ed., *TNM Classification of Malignant Tumours*. John Wiley & Sons, 6 ed., 2002.

[7] W. H. Nailon, A. T. Redpath, and D. B. McLaren, "Characterisation of radiotherapy planning volumes using textural analysis," *Acta Oncologica*, vol. 47, 2008.

[8] I. N. Bankman, ed., *Handbook of Medical Image Processing and Analysis*. Academic Press, 2009.

[9] International Agency for Research on Cancer, "GLOBOCAN 2008. Cancer incidence, mortality and prevalence worldwide," 2008.

[10] J. N. Eble, G. Sauter, J. I. Epstein, and I. A. Sesterhenn, eds., *Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs*. International Agency for Research on Cancer (IARC) Press, 2004.

[11] ICRU, "Prescribing, recording, and reporting electron beam therapy," Tech. Rep. 71, Interational Commition on Radiation Units and Measurements, Jun 2004.

[12] ICRU, "Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT)," Tech. Rep. 83, Interational Commition on Radiation Units and Measurements, 2010.

[13] J. R. Williams and D. I. Thwaites, eds., *Radiotherapy Physics: in Practice*. Oxford University Press, second edition ed., 2000.

[14] P. Tai, J. V. Dyk, E. Yu, J. Battista, L. Stitt, and T. Coad, "Variability of target volume delineation in cervical esophageal cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 42, no. 2, pp. 277 – 288, 1998.

[15] M. Yamamoto, Y. Nagata, K. Okajima, T. Ishigaki, R. Murata, T. Mizowaki, M. Kokubo, and M. Hiraoka, "Differences in target outline delineation from ct scans of brain tumours using different methods and different observers," *Radiotherapy and Oncology*, vol. 50, no. 2, pp. 151 – 156, 1999.

[16] J. S. Cooper, S. K. Mukherji, A. Y. Toledano, C. Beldon, I. M. Schmalfuss, R. Amdur, S. Sailer, L. A. Loevner, P. Kousouboris, K. K. Ang, J. Cormack, and J. Sicksevalu, "An evaluation of the variability of tumor-shape definition derived by experienced observers from ct images of supraglottic carcinomas," *International journal of radiation oncology, biology, physics*, vol. volume 67, pp. Pages 972–975, Mar 2007.

[17] J. Meyer, ed., *IMRT, IGRT, SBRT: Advances in the Treatment Planning and Delivery of Radiotherapy*. Karger, 1st ed., 2007.

[18] R. Susomboon, D. Raicu, J. Furst, and D. Channin, "Automatic single-organ segmentation in computed tomography images," pp. 1081–1086, Dec. 2006.

[19] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. SPIE Press, 2003.

[20] H. H. Barrett and W. Swindell, *Radiological Imaging*. Academic Press, revised edition ed., 1981. ISBN 0-12-079603-1.

[21] N. Matsufuji, H. Tomura, Y. Futami, H. Yamashita, A. Higashi, S. Minohara, M. Endo, and T. Kanai, "Relationship between ct number and electron density, scatter angle and nuclear reaction for hadron-therapy treatment planning," *Physics in Medicine and Biology*, vol. 43, pp. 3261–3275, 1998.

[22] J. R. Williams and D. I. Thwaites, eds., *Radiotherapy Physics: in Practice*. Oxford University Press, second edition ed., 2000.

[23] K. M. Hanson, *Radiology of the Skull and Brain*, vol. 5. C. V. Mosby, 1981.

[24] A. Borsdorf, R. Raupach, T. Flohr, and J. Hornegger, "Wavelet based noise reduction in CT-images using correlation analysis," *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1685 –1703, Dec. 2008.

[25] U. D. o. H. National Institute of Health and H. Services, "What you need to know about bladder cancer."

[26] M.Prokop, *Spiral and Multislice Computed Tomography of the Body*. Georg Thieme Verlag, 2003.

[27] F. H. Netter, *Atlas of Human Anatomy*. Saunders Elsevier, 5 ed., 2011.

[28] M. Lotfi, M. Bagheri, M. Mosleh-Shirazi, R. Faghihi, and M. Baradaran-Ghahfarokhi, "Evaluation of the changes in the shape and location of the prostate and pelvic organs due to bladder filling and rectal distension," *Iranian Red Crescent Medical Journal*, vol. 13, 2011.

[29] R. Govindan and M. Arquette, eds., *The Washington Manual of Oncology*. Lippincott Williams & Wilkins, 2002.

[30] J. O. Barentsz, G. J. Jager, J. A. Witjes, and J. H. J. Ruijs, "Primary staging of urinary bladder carcinoma: the role of MRI and a comparison with CT," *European Radiology*, vol. 6, pp. 129–133, 1996.

[31] M.Prokop, *Spiral and Multislice Computed Tomography of the Body*. Georg Thieme Verlag, 2003.

[32] A. Macvicar, "Bladder cancer staging," *BJU International*, vol. 86, pp. 111–122, 2000.

[33] A. Tekes, I. Kamel, K. Imam, G. Szarf, M. Schoenberg, K. Nasir, R. Thompson, and D. Bluemke, "Dynamic MRI of Bladder Cancer: Evaluation of Staging Accuracy," *Am. J. Roentgenol.*, vol. 184, pp. 121–127, Jan. 2005.

[34] P. Price, K. Sikora, and T. Illidge, eds., *Treatment of Cancer*. Hodder Arnold, 5 ed., 2008.

[35] S. Webb, "The physical basis of imrt and inverse planning," *The British Journal of Radiology*, vol. 76, pp. 678–689, 2003.

[36] F. M. Khan, ed., *The Physics of Radiation Therapy*. Lippincott Williams & Wilkins, 3 ed., 2003.

[37] J. Meyer, ed., *IMRT, IGRT, SBRT: Advances in the Treatment Planning and Delivery of Radiotherapy*. Karger, 2nd ed., 2011.

[38] M. J. Costa, H. Delingette, S. Novellas, and N. Ayache, "Automatic segmentation of bladder and prostate using coupled 3d deformable models," in *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention - Volume Part I*, MICCAI'07, (Berlin, Heidelberg), pp. 252–260, Springer-Verlag, 2007.

[39] R. J. R. Siqi Chen, D. Michael Lovelock, "Segmenting the prostate and rectum in CT imagery using anatomical constraints," *Medical image analysis*, Jul 2010.

[40] W. Li, S. Liao, Q. Feng, W. Chen, and D. Shen, "Learning image context for segmentation of prostate in ct-guided radiotherapy," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, vol. 6893, pp. 570–578, 2011.

[41] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Dover Publications, 1994.

[42] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[43] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 3, pp. 610–621, Nov. 1973.

[44] M. Kale, B. Clymer, R. Koch, J. Heverhagen, S. Sammet, R. Stevens, and M. Knopp, "Multispectral co-occurrence with three random variables in dynamic contrast enhanced magnetic resonance imaging of breast cancer," *Medical Imaging, IEEE Transactions on*, vol. 27, pp. 1425–1431, Oct. 2008.

[45] H. Yu, C. Caldwell, K. Mah, and D. Mozeg, "Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning," *Medical Imaging, IEEE Transactions on*, vol. 28, pp. 374–383, Mar 2009.

[46] C. Philips, D. Li, D. Raicu, and J. Furst, "Directional invariance of co-occurrence matrices within the liver," *Biocomputation, Bioinformatics, and Biomedical Technologies, 2008. BIOTECHNO '08. International Conference on*, pp. 29–34, Jul 2008.

[47] W. Nailon, A. Redpath, and D. McLaren, "Texture analysis of 3D bladder cancer CT images for improving radiotherapy planning," *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 652–655, May 2008.

[48] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer Graphics and Image Processing*, vol. 4, no. 2, pp. 172 – 179, 1975.

[49] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern Recogn. Lett.*, vol. 11, no. 6, pp. 415–420, 1990.

[50] X. Tang, "Texture information in run-length matrices," *Image Processing, IEEE Transactions on*, vol. 7, pp. 1602–1609, Nov 1998.

[51] B. V. Dasarathy and E. B. Holder, "Image characterizations based on joint gray level-run length distributions," *Pattern Recogn. Lett.*, vol. 12, no. 8, pp. 497–502, 1991.

[52] D. Shen and C. Davatzikos, "HAMMER: Hierarchical attribute matching mechanism for elastic registration," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 1421 –1439, Nov. 2002.

[53] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[54] B. B. Francisco Escolano, Pablo Suau, *Information Theory in Computer Vision and Pattern Recognition*. Springer, 2009.

[55] D. L. DONOHO and J. M. JOHNSTONE, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[56] D. Donoho, "De-noising by soft-thresholding," *Information Theory, IEEE Transactions on*, vol. 41, pp. 613 –627, May 1995.

[57] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *Information Theory, IEEE Transactions on*, vol. 45, pp. 2225 –2238, Nov 1999.

[58] T. N. Gilbert Strang, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1997.

[59] S. Mallat, *A Wavelet Tour of Signal Processing, the Sparse Way*. Academic Press, 2009.

[60] A. Mojsilovic, M. Popovic, and D. Sevic, "Classification of the ultrasound liver images with the 2N 1-D wavelet transform," in *Image Processing, 1996. Proceedings., International Conference on*, vol. 1, pp. 367 –370 vol.1, Sept 1996.

[61] A. Depeursinge, D. Sage, A. Hidki, A. Platon, P.-A. Poletti, M. Unser, and H. Muller, "Lung tissue classification using wavelet frames," pp. 6259–6262, Aug. 2007.

[62] H. Man and R. Duan, "Moment features in directional subband domain for rotation invariant texture classification," pp. 1–4, Nov 2005.

[63] X. Yin, B. W.-H. Ng, B. Ferguson, S. P. Mickan, and D. Abbott, "2-D wavelet segmentation in 3-D T-ray tomography," *Sensors Journal, IEEE*, vol. 7, pp. 342–343, Mar 2007.

[64] M. Unser and M. Eden, "Multiresolution feature extraction and selection for texture segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, pp. 717 –728, Jul 1989.

[65] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 13, pp. 99 –113, Feb 1991.

[66] D. Dunn and W. Higgins, "Optimal gabor filters for texture segmentation," *Image Processing, IEEE Transactions on*, vol. 4, pp. 947 –964, Jul 1995.

[67] K. Huang and S. Aviyente, "Wavelet feature selection for image classification," *Image Processing, IEEE Transactions on*, vol. 17, pp. 1709 –1720, Sept. 2008.

[68] P. Palmer and M. Petrou, "Locating boundaries of textured regions," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 35, pp. 1367 –1371, Sept 1997.

[69] C. R. Dyer, T.-H. Hong, and A. Rosenfeld, "Texture classification using gray level cooccurence based on edge maxima," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 10, pp. 158 –163, Mar 1980.

[70] M. Clerc and S. Mallat, "The texture gradient equation for recovering shape from texture," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 536 –549, Apr 2002.

[71] P. Hill, C. Canagarajah, and D. Bull, "Image segmentation using a texture gradient based watershed transform," *Image Processing, IEEE Transactions on*, vol. 12, pp. 1618 – 1633, Dec. 2003.

[72] S. Wang, X. Ma, X. Zhang, and L. Jiao, "Watershed-based textural image segmentation," in *Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on*, pp. 312 –315, Dec 2007.

[73] H. T. Eastment and W. J. Krzanowski, "Cross-validatory choice of the number of components from a principal component analysis," *Technometrics*, vol. 24, no. 1, pp. pp. 73–77, 1982.

[74] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, pp. 273–324, December 1997.

[75] I.T.Jolliffe, *Principal Component Analysis*. Springer, second edition ed., 2002.

[76] F. Y. Shih, *Image Processing and Pattern Recognition*. IEEE Press, 2010.

[77] M. Mazonakis, J. Damilakis, and H. Varveris, "Image segmentation in treatment planning for prostate cancer using the region growing technique," *British Journal of Radiology*, vol. 74, pp. 243–249, 2001.

[78] F. Shi, J. Yang, and Y.-m. Zhu, "Automatic segmentation of bladder in CT images," *Journal of Zhejiang University - Science A*, vol. 10, pp. 239–246, 2009. 10.1631/jzus.A0820157.

[79] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. Warfield, "Improved watershed transform for medical image segmentation using prior information," *Medical Imaging, IEEE Transactions on*, vol. 23, pp. 447 –458, Apr 2004.

[80] Y. Zhang, S. Wu, G. Yu, and D. Wang, "A hybrid image segmentation approach using watershed transform and FCM," in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, vol. 4, pp. 2 –6, Aug. 2007.

[81] H. Zhu, B. Zhang, A. Song, and W. Zhang, "An improved method to reduce oversegmentation of watershed transformation and its application in the contour extraction of brain image," in *Dependable, Autonomic and Secure Computing, 2009. DASC '09. Eighth IEEE International Conference on*, pp. 407 –412, Dec. 2009.

[82] I. N. Bankman, ed., *Handbook of Medical Image Processing and Analysis*. Academic Press, 2009.

[83] M. Artin, *Algebra*. No. 109-144, Prentice Hall, 1991.

[84] I. N. Bankman, ed., *Handbook of Medical Image Processing and Analysis*. Academic Press, 2009.

[85] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System : an Approach to Cerebral Imaging*. Thieme-Stratton Corp, 1988.

[86] J. Mazziotta, A. Toga, A. Evans, P. Fox, and L. J., "A probabilistic atlas of the human brain: Theory and rationale for its development," *The International Consortium for Brain Mapping*, vol. 2, pp. 89–101, 1995.

[87] A. Guimond, J. Meunier, and J.-P. Thirion, "Average brain models: A convergence study," *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 192 – 210, 2000.

[88] P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi, "Multi-modal image set registration and atlas formation," *Medical Image Analysis*, vol. 10, no. 3, pp. 440 – 451, 2006. Special Issue on The Second International Workshop on Biomedical Image Registration (WBIR'03).

[89] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "Dramms: Deformable registration via attribute matching and mutual-saliency weighting," *Medical Image Analysis*, vol. 15, no. 4, pp. 622 – 639, 2011. ¡ce:title¿Special section on IPMI 2009¡/ce:title¿.

[90] K. Van Leemput, "Probabilistic brain atlas encoding using Bayesian inference," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (R. Larsen, M. Nielsen, and J. Sporring, eds.), vol. 4190 of *Lecture Notes in Computer Science*, pp. 704–711, Springer Berlin / Heidelberg, 2006.

[91] S. Bricq, C. Collet, and J.-P. Armspach, "Lesions detection on 3D brain MRI using trimmmed likelihood estimator and probabilistic atlas," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 93 –96, May 2008.

[92] H. Park, P. Bland, and C. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 483 –492, Apr 2003.

[93] C. Meyer, H. Park, J. Balter, and P. Bland, "Method for quantifying volumetric lesion change in interval liver CT examinations," *Medical Imaging, IEEE Transactions on*, vol. 22, pp. 776 –781, Jun 2003.

[94] S. Klein, U. van der Heide, B. Raaymakers, A. Kotte, M. Staring, and J. Pluim, "Segmentation of the prostate in mr images by atlas matching," in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pp. 1300 –1303, Apr 2007.

[95] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and P. J. P., "Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information," *Medical Physics*, vol. 35, pp. 1407–1416, 2008.

[96] A. Blake and M. Isard, *Active contours : the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer, 1998.

[97] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543 – 563, 2009.

[98] N. Archip, P.-J. Erard, M. Egmont-Petersen, J.-M. Haefliger, and J.-F. Germond, "A knowledge-based approach to automatic detection of the spinal cord in CT images," *Medical Imaging, IEEE Transactions on*, vol. 21, pp. 1504 –1516, Dec. 2002.

[99] B. Haas, T. Coradi, M. Scholz, P. Kunz, M. Huber, U. Oppitz, L. André, V. Lengkeek, D. Huyskens, A. van Esch, and R. Reddick, "Automatic segmentation of thoracic and pelvic ct images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies," *Physics in Medicine and Biology*, vol. 53, no. 6, p. 1751, 2008.

[100] G. Wagenknecht, E. Kops, J. Kaffanke, L. Tellmann, F. Mottaghy, M. Piroth, and H. Herzog, "Ct-based evaluation of segmented head regions for attenuation correction in mr-pet systems," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pp. 2793 –2797, 30 2010-nov. 6 2010.

[101] J. Yao, J. Chen, and C. Chow, "Breast tumor analysis in dynamic contrast enhanced MRI using texture features and wavelet transform," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 3, pp. 94–100, Feb. 2009.

[102] W. Li, S. Liao, Q. Feng, W. Chen, and D. Shen, "Learning image context for segmentation of the prostate in ct-guided radiotherapy," *PHYSICS IN MEDICINE AND BIOLOGY*, 2012.

[103] S. Liao and D. Shen, "A feature-based learning framework for accurate prostate localization in ct images," *Image Processing, IEEE Transactions on*, vol. 21, pp. 3546 –3559, aug. 2012.

[104] J. Koss, F. Newman, T. Johnson, and D. Kirch, "Abdominal organ segmentation using texture transforms and a hopfield neural network," *Medical Imaging, IEEE Transactions on*, vol. 18, pp. 640–648, Jul 1999.

[105] M. Sasikala, N. Kumaravel, and L. Subhashini, "Automatic tumor segmentation using optimal texture features," pp. 1–4, Jul 2006.

[106] B. Zhang, W. Zhu, H. Zhu, A. Song, and W. Zhang, "An SVM based automatic segmentation method for brain magnetic resonance image series," in *Ubiquitous Intelligence Computing and 7th International Conference on Autonomic Trusted Computing (UIC/ATC), 2010 7th International Conference on*, pp. 375 –379, Oct. 2010.

[107] N. Abdullah, U. Ngah, and S. Aziz, "Image classification of brain MRI using support vector machine," in *Imaging Systems and Techniques (IST), 2011 IEEE International Conference on*, pp. 242 –247, May 2011.

[108] Y. Zhan and D. Shen, "Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method," *Medical Imaging, IEEE Transactions on*, vol. 25, pp. 256–272, Mar 2006.

[109] 18th European Signal Processing Conference (EUSIPCO-2010), *Classification of Bladder Cancer on Radiotherapy Planning CT Images Using Textural Features*, Aug. 2010.

[110] H. T. Eastment and W. J. Krzanowski, "Cross-validatory choice of the number of components from a principal component analysis," *Technometrics*, vol. 24, no. 1, pp. pp. 73–77, 1982.

[111] W. J. Krzanowski, "Cross-validation in principal component analysis," *Biometrics*, vol. 43, no. 3, pp. pp. 575–584, 1987.

[112] W. J. Krzanowski, "Between-groups comparison of principal components," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. pp. 703–707, 1979.

[113] S. L. Martinez-Arroyo, M., "Learning an optimal naive Bayes classifier," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, pp. 958 –958, 0-0 2006.

[114] M. Islam, Q. Wu, M. Ahmadi, and M. Sid-Ahmed, "Investigating the performance of naive Bayes classifiers and K- nearest neighbor classifiers," pp. 1541 –1546, Nov. 2007.

[115] ICRU, "Receiver operating characteristic analysis in medical imaging," *Journal of ICRU*, vol. 8, Apr 2008.

[116] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press, 2003.

[117] J. R. Andrews, "Benefit, risk, and optimization by ROC analysis in cancer radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 11, no. 8, pp. 1557 – 1562, 1985.

[118] N. A. Obuchowski, "Receiver operating characteristic curves and their use in radiology," *Radiology*, vol. 229, pp. 3–8, Oct 2003.

[119] S. Bond and J. Brady, "Non-rigid registration for colorectal cancer MR images," in *Computer Vision for Biomedical Image Applications* (Y. Liu, T. Jiang, and C. Zhang, eds.), vol. 3765 of *Lecture Notes in Computer Science*, pp. 399–408, Springer Berlin / Heidelberg, 2005.

[120] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," pp. 281–300, Springer-Verlag, 1995.