An Investigation of Methods for CT Synthesis in MR-only Radiotherapy

Andreasen, Daniel

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

An Investigation of Methods for CT Synthesis in MR-only Radiotherapy

Daniel Andreasen

Kongens Lyngby 2016

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

PHD-2016-428
ISSN: 0909-3192

Summary (English)

In recent years, the interest in using magnetic resonance (MR) imaging in radiotherapy (RT) has increased. This is because MR has a superior soft tissue contrast compared to computed tomography (CT), which makes it a better modality for delineating the target volume (tumor) and possible organs at risk (OARs). In an MR/CT work-flow, independent MR and CT scans are acquired. The target and possible OARs are delineated on the MR and then transferred to CT by aligning the data using a registration. This introduces the risk of systematic registration errors especially in non-rigid body structures, the consequence being a systematic miss of target or increased dose to healthy tissue.

Radiotherapy based on MR as the only modality removes this uncertainty and simplifies the clinical work-flow. However, the information on electron density which is usually contained in the CT must now be derived from the MR. A way to achieve this is to computationally estimate a so-called synthetic CT (sCT) from the MR data, which can then act as a substitute for the CT. This is a challenging task, since no unique relationship between MR and electron density exists.

The goal of this thesis is to develop and investigate the right combination of MR acquisition protocols and computational models for accurate MR-based CT synthesis for use in RT. We investigate different categories of methods for CT synthesis and validate them using clinically relevant quality measures. Specifically, we implement a patch-based multi-atlas method in the brain, which compares favorably to state-of-the-art methods. In our next effort, we substantially improve the speed of the method and apply it in the pelvis, again with promising results. Our final contribution is a voxel-based method, which is developed to

be registration-free and broadly applicable. In initial results, the performance of this method is close to the patch-based.

Summary (Danish)

I de senere år har der været en stigende interesse for at bruge magnetisk resonans (MR) skanninger i stråleterapi. Dette skyldes, at MR har en langt bedre bløddels-kontrast i forhold til computed tomography (CT), hvilket gør det til et bedre modalitet til indtegning af target volumener (tumor) samt organer der skal skånes for stråling. I et MR/CT work-flow, optages uafhængige MR- og CT-skanninger. Indtegninger foretages på MR-skanningen, som her efter registreres med CT-skanningen for at kunne overføre indtegningerne til dosisoptimering. Denne registrering introducerer en risiko for systematiske registreringsfejl, især i ikke-rigide kropsdele. Konsekvensen af en sådan fejl kan i værste tilfælde være en systematisk underdosering af tumoren eller en øget dosis til raskt væv.

Stråleterapi baseret på MR som den eneste modalitet fjerner denne usikkerhed og forenkler det kliniske work-flow. Dog skal oplysningerne om elektrondensitet, som sædvanligvis fås fra CT, nu afledes fra MR. En måde at opnå dette er at estimere en såkaldt syntetisk CT (sCT) skanning ud fra MR-skanningen. Denne skanning kan derefter fungere som erstatning for CT. Det er en særdeles udfordrende opgave at omdanne en MR-skanning til en sCT, da der ikke er et unikt forhold mellem MR og elektrondensitet.

Målet med denne afhandling er at udvikle og undersøge den rette kombination af MR-protokoller og beregningsmodeller til nøjagtig MR-baseret sCT-dannelse til brug i stråleterapi. Vi undersøger forskellige kategorier af metoder til sCT-dannelse og validerer dem ved hjælp af klinisk relevante kvalitetsmål. Specifikt, implementerer vi en patch-baseret multi-atlas metode i hjernen, som klarer sig godt i en sammenligning med *state-of-the-art* metoder. I vores næste indsats forbedrer vi hastigheden af den patch-baserede metode betydeligt og anvender

den i pelvis med lovende resultater. Vores sidste bidrag er en voxel-baseret metode, som er udviklet til at virke uden brug af registreringer og til at være bredt anvendelig. I vores initielle resultater er metoden performance-mæssigt tæt på den patch-baserede.

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in fulfillment of the requirements for acquiring a PhD degree. The thesis was prepared with funding from DTU Compute, Herlev and Gentofte hospital, and a research grant from Varian Medical Systems Inc. It was co-supervised by PhD, DABR Jens M. Edmund and associate professor Koen Van Leemput.

The thesis deals with the creation of synthetic computed tomography images from magnetic resonance images for use in radiotherapy planning based only on magnetic resonance images.

Lyngby, 29-August-2016

Daniel Andreasen

Acknowledgements

First of all, I would like to thank both of my supervisors, Jens M. Edmund and Koen Van Leemput for their guidance, enthusiasm, and company throughout the years. It has been a pleasure working under your supervision.

Also, a great thanks to the radiotherapy research unit at Herlev Hospital for finding much of the funding, which made this PhD project a reality. I want to thank Rasmus H. Hansen for all his help with setting up the MR sequences for our projects, and Annette Kahlen for helping me every time I had to scan strange objects in the scanner. My thanks to Jon Andersen for joining our project and helping with patient recruitment.

Many thanks to professor Bjoern Menze, who hosted me in his image-based biomedical modeling group during my external stay at the Technical University of Munich.

Thanks to my colleagues at the section for image analysis and computer graphics at DTU, especially to my lab mates Oula Puonti, Christian Thode Larsen, and Mikael Agn. It has been great working with you guys as well as drinking the occasional Friday beer together.

Finally, I would like to thank my friends and family for their support. A special thanks to my girlfriend, Minna, whose support has been invaluable in stressful times.

Contents

Chapter 1

Introduction

The field of medical imaging for diagnosis and treatment has seen tremendous advances in the last few decades and has become more and more multi-modal with the routine clinical use of computed tomography (CT), positron emission tomography (PET), and magnetic resonance (MR) imaging among others. The multi-contrast ability of MR imaging makes it one of the most versatile among the modalities. By adjusting the acquisition parameters, one can acquire detailed anatomical images with widely different tissue contrasts as well as functional images. The complementary information provided by multi-contrast MR aids not only in diagnosis and treatment of patients, but also in revealing neurophysiological patterns in both diseased and healthy subjects.

The use of multiple modalities and multi-contrast MR also introduces new challenges. As an example, the total scan time of a patient increases for each additional MR contrast required. This not only increases patient discomfort, but also adds the risk of unwanted changes due to organ motion and deformation between each acquisition. This risk grows even further when the patient has to be moved between two different scanners. To truly benefit from complementary scans acquired on different scanners or at different time points, there is a need to spatially align, or *register*, the different images. Even without motion-induced changes, this is not an easy task and the process can introduce systematic errors in e.g., a treatment situation.

In this thesis, we have approached the above challenges by attempting to make the treatment work-flow in radiotherapy as mono-modal as possible such that the need for accurate multi-modal registrations is eliminated.

1.1 Image synthesis for MR-only radiotherapy

In external beam radiotherapy (RT), the objective is to kill cancer cells by means of high-energy ionizing radiation. The patient is irradiated in such a way as to maximize the amount of dose delivered to the tumor while minimizing the amount of dose given to healthy tissue. This poses a complex optimization problem, where the radiation attenuation properties of the different tissues must be known in order to calculate the three-dimensional (3D) dose distribution in the irradiated tissues. CT contains such information because the intensity in these images is directly related to the electron density of the tissue being imaged. Furthermore, CT has an accurate geometric representation of bone, which is used in combination with a planar X-ray or a cone beam CT (CBCT) to align the patient with respect to the isocenter of the linear accelerator (Linac). For these reasons, CT is being used in both the planning and treatment phase of external beam RT.

In recent years, the interest in using MR imaging in RT has steadily increased. This is because MR has a superior soft tissue contrast compared to CT, which makes it a better modality for delineating the target volume (tumor) and possible organs at risk (OARs). In the MR/CT workflow, independent MR and CT scans are acquired. The target and possible OARs are delineated on the MR and then transferred to CT by aligning the data using a rigid registration. This introduces the risk of systematic registration errors especially in non-rigid body structures, with a systematic miss of target or increased dose to healthy tissue as possible consequences. These errors could be eliminated, if the CT was removed entirely from the RT planning and treatment process. This would in addition simplify the clinical work-flow, leading to reduced health care costs and sparing the patient of the discomfort of having two separate scans made. A way to achieve this is to computationally estimate a so-called synthetic CT (sCT) from the MR data. This is a challenging task because a direct mapping of MR intensities to CT intensities is not possible. This is because the MR signal is not related to electron density but is a relative measure of the proton density and magnetic relaxation properties of tissue. It is critically dependent on the pulse sequence used, i.e., timing and shape of radio frequency pulses, type of gradients and timing of signal detection. In addition, compact bone has a short magnetic relaxation time, which means that in conventional MR imaging little or no signal is detected from compact bone making it indistinguishable from air.

1.2 Goals of the project

The main focus of this PhD project was to develop and investigate the right combination of MR acquisition protocols and computational models for accurate MR-based sCT prediction for use in RT. Furthermore, the aim was also to validate the resulting methods in clinically relevant external beam RT settings.

1.3 Scientific contributions

Peer-reviewed papers included in this thesis

Paper A J. M. Edmund, H. M. Kjer, K. Van Leemput, R. H. Hansen, J. A. Andersen, and D. Andreasen, "A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times," *Phys. Med. Biol. 59(23), 7501 (2014).*

Paper B D. Andreasen, K. Van Leemput, R. H. Hansen, J. A. L. Andersen, and J. M. Edmund, "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain", *Medical Physics 42, 1596–1605 (2015).* Editor's pick Medical Physics April issue.

Paper C D. Andreasen, K. Van Leemput, and J. M. Edmund, "A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis" *Medical Physics, 43, 4742-4752 (2016).*

Paper D D. Andreasen, J. M. Edmund, V. Zografos, B. H. Menze, and K. Van Leemput, "Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features", in *SPIE Medical Imaging (International Society for Optics and Photonics, 2016) pp. 978417–978417.*

Peer-reviewed contributions not included in this thesis

The following contributions were carried out during my PhD studies but are not discussed in this thesis either because of overlap with the above papers or because the theme did not match the thesis.

Paper E J. M. Edmund, D. Andreasen, F. Mahmood, and K. Van Leemput,"Cone beam computed tomography guided treatment delivery and planning verification for magnetic resonance imaging only radiotherapy of the brain", *Acta Oncologica 54, 1496–1500 (2015)*

Abstract A D. Andreasen, K. Van Leemput, R.H. Hansen, J.A.L. Andersen, and J.M. Edmund: "Predicting a Pseudo-CT scan from T1-weighted MR Images using Patches – towards MRI-only based radiotherapy". Accepted for e-poster at ESTRO 33, Vienna, Austria, 2014

Abstract B D. Andreasen, K. Van Leemput, and J.M. Edmund: "Patch-based Generation of a pseudo-CT Scan for MRI-only based Radiotherapy in the Pelvic Region". Accepted for poster presentation at 3rd ESTRO Forum, Barcelona, Spain, 2015.

1.4 Overview of the thesis

The main part of this thesis focuses on the three papers that I am first author of (papers B-D). Paper A has no dedicated chapter, but is referenced throughout the thesis and serves as the basis for much of Chapter 2.

This thesis is structured as follows:

- Chapter 2 provides an overview of the development in approaches to sCT prediction as well as a description of the a priori most promising solutions. I highlight the advantages and limitations of these methods in order to set the stage for our choices of sCT strategies.
- Chapter 3 provides an introduction to a patch-based approach for sCT prediction as well as a summary and discussion of the results presented in paper B.

- Chapter 4 provides an introduction to a speeded up version of the patch-based approach applied in the pelvic region as well as a summary and discussion of the results presented in paper C.
- Chapter 5 highlights issues with intensity normalization in MR scans which can affect the patch-based approach.
- Chapter 6 provides an introduction to a voxel-based sCT method as well as a summary and discussion of the results presented in paper D.
- Chapter 7 summarizes the main findings of this thesis and points out future directions.

Chapter 2

Background

Over the last decade, there has been a tremendous rise in the number of publications dealing with MR-only RT and, especially, in publications describing new methods for sCT prediction.
In this chapter, I provide a short introduction to MR-only RT and its challenges. I establish commonly used measures of sCT quality. Finally, I provide an overview of the current sCT methods and attempt to categorize them according to their underlying principles. Partly based on our results in paper A, I highlight the limitations and challenges with the different categories and propose possible directions to overcome some of these challenges.

2.1 MR-only: the idea and the challenges

In RT treatment planning, MR is increasingly being used in combination with the traditional CT. The main advantage of MR is its superior soft tissue contrast, which improves target and OAR definition in the brain and other sites as compared to CT-based delineations [1–3]. Furthermore, functional imaging using MR is being investigated as an exciting prospect for tumor response evaluation and prediction as well as for so-called dose-painting, where regions in a tumor requiring higher doses are automatically detected [4]. The introduction

of the MR-Linac [1] further establishes the potential of MR in RT with promises of accurate, real-time image guided RT (IGRT) during treatment delivery [5].

Still, the CT is a necessary part of today's RT treatment planning because it contains the crucial information about tissue electron density needed for dose calculation and simulation. Furthermore, reference images based on bone anatomy can easily be generated and used for patient positioning in combination with X-ray images or cone beam CTs (CBCTs) taken at the Linac. In the combined MR/CT workflow, the delineations are done on the MR and then transferred to the CT by spatially aligning the two scans. This is usually done by a manual and/or automatic rigid registration between the CT and MRI scan. This is inherently difficult due to the difference in contrast between the scans; the soft tissue contrast in the MR is not present in the CT, and the bone anatomy shown in the CT is not necessarily well depicted in the MR. In addition, non-rigid organ deformations are potentially happening between the scans in non-brain regions, meaning that a rigid registration is insufficient for aligning them. Even for a relatively rigid structure such as the brain, it has been estimated that a mean registration error of approximately 2 mm is introduced in the process [6, 7]. Similar values have been reported for other body sites such as the prostate [8]. These errors produce a systematic shift in the delineations which, considering the tight margins used in modern RT (< 5 mm), may ultimately lead to a critical target under-dosage and/or an increased dose to the adjacent OARs.

The above observations spurred an interest in eliminating the CT entirely from the treatment chain and performing MR-only radiotherapy. In essence, this requires that a so-called synthetic CT (sCT) (also referred to as a pseudo-CT or substitute CT) can be generated from the MR such that the tissue electron density can be obtained from the synthetic values. This is a complicated task since the voxel intensities in MR are not uniquely related to electron density. On the contrary, the intensities are relative measures of proton density and tissue magnetic relaxation properties, and they vary depending on the equipment and choice of acquisition parameters. Furthermore, with conventional MR sequences, the signal from compact bone is quickly lost due to its rapid transversal signal relaxation time (T_2) in the range 0.5-2 ms [9–11]. This yields an intensity ambiguity between bone and air, even if they are at opposite ends in terms of electron density, see Figure 2.1.

[1] The MR linear accelerator provides MR imaging of the patient during treatment delivery in order to track and respond to anatomical changes/movement.

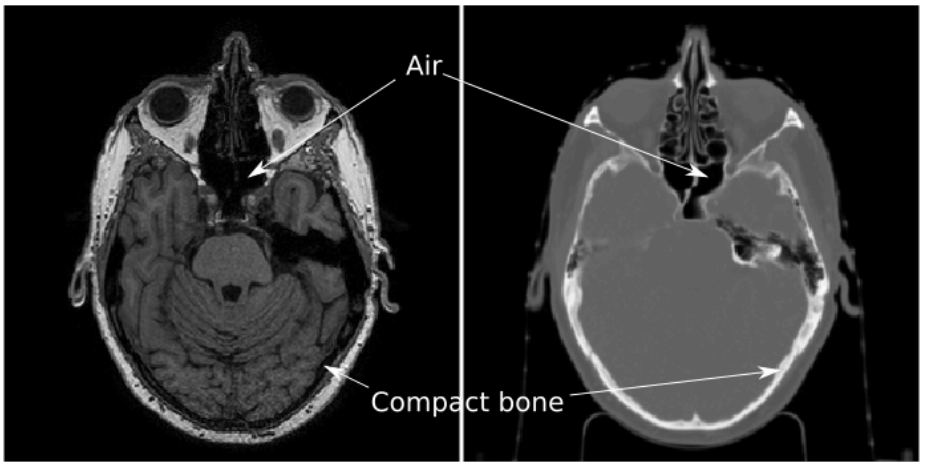


Figure 2.1: With conventional MR sequences, there is an intensity ambiguity between compact bone and air regions. This is shown here for a T_1-weighted MR scan (left) and the corresponding CT scan (right).

2.2 Evaluation of sCT methods

Before describing the sCT methods themselves, I will establish a number of measures to evaluate the quality of an sCT. This is important in order to test a method's feasibility and to benchmark different methods. To do this, the generated sCT must be compared to the ground truth CT, which can be done in numerous ways to highlight different properties of the sCT. Below I describe some of the most common evaluation criteria.

2.2.1 Voxel-wise and geometric measures

Probably the simplest and most commonly used measure is the voxel-wise mean absolute error ($\mathrm{MAE}_{\mathrm{vox}}$) and mean error ($\mathrm{ME}_{\mathrm{vox}}$), defined as

$$\mathrm{MAE}_{\mathrm{vox}} = \frac{1}{N}\sum_{n=1}^{N}|\mathrm{CT}(n) - \mathrm{sCT}(n)|, \tag{2.1}$$

and

$$\mathrm{ME}_{\mathrm{vox}} = \frac{1}{N}\sum_{n=1}^{N}(\mathrm{CT}(n) - \mathrm{sCT}(n)), \tag{2.2}$$

where N is the total number of voxels inside the body outline of the MR, and CT and sCT are ordered as $1 \times N$ vectors indexed by n containing the Hounsfield unit (HU) intensity values of the CT and sCT at corresponding spatial positions. These measures are usually calculated only for voxels inside the body contour of the patient, since the surrounding air often constitutes a relatively large part of the image and is straightforward to predict. Including this region would make the measures dependent on the amount of surrounding air and could result in artificially low errors. The $\mathrm{MAE}_{\mathrm{vox}}$ provides an evaluation of the general quality of the sCT, whereas the $\mathrm{ME}_{\mathrm{vox}}$ can reveal if there is a bias towards under- or overestimating the CT intensity.

Aside from using the electron density information for dose calculation, the CT scan is also used for patient setup at the Linac. One way to do this is by matching X-ray radiographs taken at the Linac with 2D projections generated from the CT, so-called digitally reconstructed radiographs (DRRs). A matching can also be performed based on CBCTs. In essence, this is a matching based on bone anatomy, which means that an sCT should depict bone with a high geometrical accuracy. A way to evaluate the geometry of the predicted bone volume is using the Dice similarity coefficient (DSC) [12] defined for bone as:

$$\mathrm{DSC}_{\mathrm{bone}} = \frac{2\langle V_{bone}^{sCT} \cap V_{bone}^{CT} \rangle}{\langle V_{bone}^{sCT} \rangle + \langle V_{bone}^{CT} \rangle}, \tag{2.3}$$

where V_{bone}^{sCT} and V_{bone}^{CT} are binarized volumetric bone structures in the sCT and the CT, respectively, and $\langle \cdot \rangle$ denotes the number of elements in a structure. The DSC measures the overlap of the two structures and takes on a value between 0 and 1, where 1 means complete overlap and 0 means no overlap. In order to estimate the bone structures, a threshold is usually applied to the CT and sCT in order to segment the bone. The DSC is therefore dependent on the chosen threshold.

2.2.2 Radiologic measure

The pure voxel-wise and geometric error measures reviewed above do not take the radiologic properties of different tissues into account. The relationship between the HU and electron density is piece-wise linear, which means that prediction errors in one part of the HU scale may not have as severe an impact on the converted electron density as in other parts. An example of a verified calibration curve is shown in Figure 2.2. Furthermore, the voxel-wise measures do not account for the fact that during RT treatment, radiation travels through various tissues. Because errors in individual voxels may cancel each other along a path while adding up in the $\mathrm{MAE}_{\mathrm{vox}}$ measure, it may be more relevant to

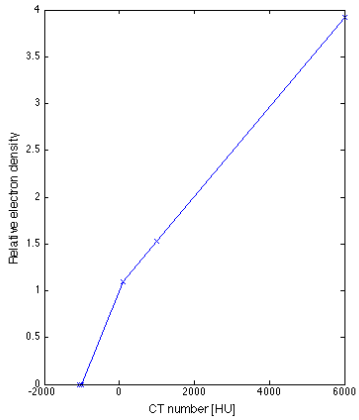


Figure 2.2: Example of a calibration curve for converting HU to electron density used in the treatment planning system.

evaluate the sCT along such simulated radiation paths. The water-equivalent path length (WEPL) provides a means for doing this, while at the same time accounting for the piece-wise linear relationship of the HU and electron density. The WEPL is defined as [13, 14]:

$$\mathrm{WEPL} = \sum_n \Delta l_n \times \rho_n, \tag{2.4}$$

where Δl_n is the physical path length of voxel n, and ρ_n is a radiological scaling factor that depends on the type of radiation and tissue. For MeV photons, it is the electron density relative to water. The value of ρ_n is found from the voxel CT number using a calibration curve similar to the one shown in Figure 2.2. The WEPL can be evaluated along desired paths through the CT and sCT, and mean absolute errors ($\mathrm{MAE}_{\mathrm{WEPL}}$) and mean errors ($\mathrm{ME}_{\mathrm{WEPL}}$) between the paths lengths can be used as quality measure:

$$\mathrm{MAE}_{\mathrm{WEPL}} = \frac{1}{L}\sum_{l=1}^{L} |\mathrm{WEPL}_l^{\mathrm{CT}} - \mathrm{WEPL}_l^{\mathrm{sCT}}|, \tag{2.5}$$

and

$$\mathrm{ME}_{\mathrm{WEPL}} = \frac{1}{L}\sum_{l=1}^{L} (\mathrm{WEPL}_l^{\mathrm{CT}} - \mathrm{WEPL}_l^{\mathrm{sCT}}), \tag{2.6}$$

where L is the number of paths and WEPL_l is the WEPL of the lth path.

2.2.3 Dosimetric measure

Probably the most important measure of sCT quality is the dosimetric accuracy when using the sCT for dose planning instead of the CT. A way to evaluate this, is to create and optimize a treatment plan on the sCT and then transfer the plan to the CT and re-calculate using the same plan parameters and monitor units. This provides a basis for comparing the obtained dose distributions in various ways. Using a so-called cumulative dose-volume histogram (DVH), the 3D dose distribution for a single OAR or tumor volume can be summed up in an interpretable fashion. A cumulative DVH shows the minimum dose, D, that a given percentage of a structure received, see Figure 2.3. By comparing points on the DVHs resulting from the sCT dose calculation and the CT re-calculation, an estimate of the similarity in dose between the two calculations is obtained. Typically used DVH points are those recommended for reporting PTV coverage and/or those recommended as optimization constraints for OARs. Another way to utilize the DVH is to calculate a normalized mean absolute dosimetric error ($\mathrm{nMAE_{dos}}$):

$$\mathrm{nMAE_{dos}} = \frac{1}{M}\sum_{m=1}^{M}\frac{|\mathrm{D}^{CT}_{vol(m)} - \mathrm{D}^{sCT}_{vol(m)}|}{\mathrm{D}_{pre}}, \tag{2.7}$$

where D_{pre} is the prescribed dose, $\mathrm{D}_{vol(m)}$ is the accumulated dose in Gy given to vol(m)% of the volume in either the CT or sCT, and M is the total number of DVH points between $\mathrm{D}_{100\%}$ and $\mathrm{D}_{0\%}$. Similarly, the normalized mean dosimetric error ($\mathrm{nME_{dos}}$) can be calculated as:

$$\mathrm{nME_{dos}} = \frac{1}{M}\sum_{m=1}^{M}\frac{\mathrm{D}^{CT}_{vol(m)} - \mathrm{D}^{sCT}_{vol(m)}}{\mathrm{D}_{pre}}. \tag{2.8}$$

Finally, the combined dosimetric score (CDS) is a combination, which penalizes bias in the errors [11]:

$$\mathrm{CDS} = \mathrm{nMAE_{dos}} + |\mathrm{nME_{dos}}|. \tag{2.9}$$

A downside of evaluating the DVH is that all spatial information is lost, which means that the spatial position of the calculated dose is not accounted for. Two dose matrices can also be quantitatively compared using a γ-index evaluation [15]. To do this, a passing criterion is chosen consisting of the maximum allowed deviation in dose (DD) and the maximum allowed distance to agreement (DTA) of the DD. As an example, a DD/DTA = 3%/3mm criterion means that for a given voxel in one dose matrix, $\mathbf{D_1}$, a dose value within 3% must be found within 3mm of that voxel in a second dose matrix, $\mathbf{D_2}$. Let $\mathbf{r_1}$ and $\mathbf{r_2}$ represent spatial locations in $\mathbf{D_1}$ and $\mathbf{D_2}$, respectively. The γ-index is then defined as [15]:

$$\gamma(\mathbf{r_1}) = \min\{\Gamma(\mathbf{r_1}, \mathbf{r_2})\}\ \forall \mathbf{r_2} \in \mathbf{D_2}, \tag{2.10}$$

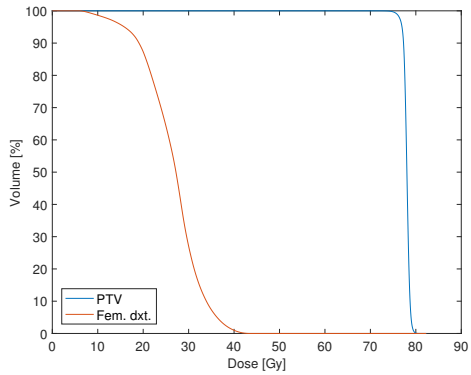


Figure 2.3: Example of a cumulative dose-volume histogram for a PTV (prostate) and an OAR (right femoral head, fem. dxt.). 75 Gy or more was received by 100 % of the PTV, i.e., $D_{100\%} = 75$ Gy, while approximately 30 Gy or more was received by 40 % of the OAR, i.e., $D_{40\%} = 30$ Gy.

with

$$\Gamma(\mathbf{r_1}, \mathbf{r_2}) = \sqrt{\frac{r^2(\mathbf{r_1}, \mathbf{r_2})}{\mathrm{DTA}^2} + \frac{\delta^2(\mathbf{r_1}, \mathbf{r_2})}{\mathrm{DD}^2}}, \tag{2.11}$$

where $r(\mathbf{r_1}, \mathbf{r_2})$ is the distance between $\mathbf{r_1}$ and $\mathbf{r_2}$, and $\delta(\mathbf{r_1}, \mathbf{r_2})$ is the percentage difference in doses observed at $\mathbf{r_1}$ and $\mathbf{r_2}$. This definition means that if $\gamma \leq 1$ the criterion is met, otherwise the voxel fails. The results of a γ-index evaluation can be shown as a γ-index map showing the value of γ at each voxel. This can be combined with reporting the overall pass-rate.

2.2.4 Considerations regarding sCT quality

The evaluation of sCTs comes with some considerable challenges due to the nature of the problem we are trying to solve. One of the issues that MR-only RT tries to address is the registration uncertainty between CT and MR. However, in order to evaluate the quality of an sCT, we need to compare it to the ground truth, which is the CT. This comparison can only be realized if the sCT (and hence the underlying MR) is registered with the CT. Therefore, the uncertainty that we are trying to remove is always included in the evaluation of an sCT. Furthermore, the measured quality of the sCT becomes somewhat dependent on an accurate registration. The dependency is usually ignored when reporting sCT quality as it is difficult to quantify for individual patients. However, it is assumed that a sub-optimal MR-CT registration will negatively bias the sCT quality and as such, reported sCT quality measures can be regarded as worst-case estimates.

Furthermore, it is worth noting that observed differences between the planning CT and MR in e.g., the body outline or organ positions are for the most part no different from differences observed between planning CT and a CBCT of the day.

2.3 sCT methods

There are several different ways to categorize sCT methods. This could be according to body region (brain, pelvis, head and neck, etc.), application (RT or PET/MR attenuation correction), supervised or unsupervised, using special or conventional MR sequences, etc. Although hybrid methods exist, I choose to split the approaches according to their underlying principles, which yields two categories: voxel-based and atlas-based. This is probably the most general categorization and contains all of the above categories.

2.3.1 Voxel-based methods

The idea of the voxel-based approaches is to convert individual MR voxels to their corresponding CT HU value based on a learned criterion such as a regression model or a combination of classification and bulk density assignment [10,11,16–18]. Many approaches in this category use intensity as the main feature, which means that bone/air ambiguities must be solved by e.g., using specialized MR sequences such as the ultrashort echo time (UTE) sequence. Using an unconventional acquisition approach and dual echo times, this sequence is optimized to maximize the amount of signal coming from short T_2 components [19]. An example of dual echo UTE images is shown in Figure 2.4, where the contrast between bone and air is evident.

Aside from assigning a bulk water-equivalent CT number to the entire patient, the simplest voxel-based approaches use manually determined thresholds and morphological operations to label voxels of UTE images according to tissue classes such as soft tissue, bone, and air. A bulk density is then assigned to the classes in order to create the sCT [10, 17, 21]. In paper A, we found that the main drawback of these methods is the need for manual intervention; the difficulty in tuning the relative intensity thresholds and choosing the appropriate bulk density values could place the voxel-wise and geometric accuracy of these methods in the lower end with $DSC_{bone} \approx 0.5$ and $MAE_{vox} \geq 200$ HU in the brain region [11]. More sophisticated methods are supervised, where model parameters are learned in a training phase from pre-acquired and co-registered

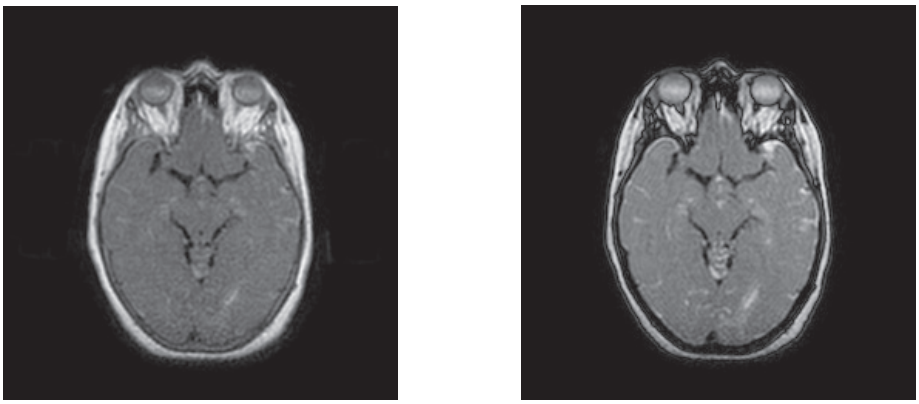
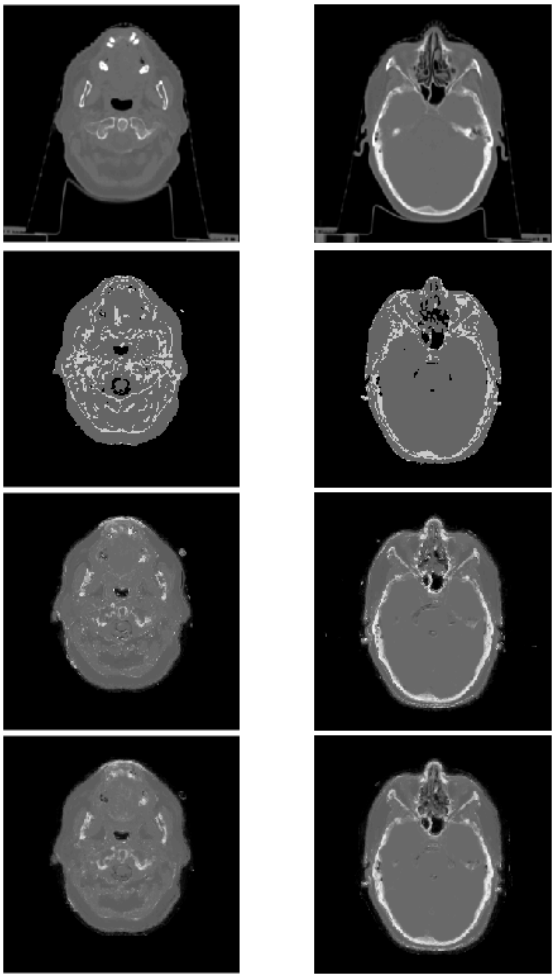


Figure 2.4: Dual echo UTE images. **Left:** Image acquired at $T_E = 0.09\ ms$. **Right:** Image acquired at $T_E = 3.5\ ms$. The signal has been lost from short T_2 components in the second echo image. Images from from [20].

MR and CT data. This automates the sCT generation, and can yield either a continuous-valued or a segmented sCT depending on the chosen model [16, 18, 20]. Once the model parameters have been determined from the training phase, these methods are fast in predicting an sCT and they generally perform better than the above-mentioned threshold-based approaches. In paper A, we found that regression models such as a random forest [20] or a kernel regression model parameterized by a mixture of Gaussians [16] are among the most promising methods in the voxel-based category when using UTE scans of the brain. These methods have $\mathrm{DCS}_{\mathrm{bone}} \geq 0.7$ and $\mathrm{MAE}_{\mathrm{vox}} \approx 130$ HU as well as $< 2\%$ deviation in relevant DVH points in the brain [11]. Examples of different voxel-based sCTs created from UTE scans are shown in Figure 2.5.

The common denominator of the above methods is the need for the non-standard UTE sequence. From a practical point of view, introducing additional MR sequences in the RT work-flow is unappealing since it prolongs the MR scan time of each patient and introduces the risk of PTV and/or OAR movement between the sCT sequence and the delineation sequence. Furthermore, a low signal-to-noise ratio and partial volume effects in the UTE scans means that sCT predictions suffer at tissue interfaces such as in the ear and nasal cavities [10, 18, 22]. To overcome the latter issues, some authors have included spatial information in their models resulting in a decrease in average $\mathrm{MAE}_{\mathrm{vox}}$ from 140 HU to 130 HU [22].

Another class of voxel-based approaches uses conventional MR sequences such as T_1-weighted or T_2-weighted scans. This puts additional demands on the



(a) CT scan

(b) Threshold-based method of [10]. $\mathrm{MAE}_{\mathrm{vox}} = 222$ HU, $\mathrm{DSC}_{\mathrm{bone}} = 0.52$.

(c) Gaussian Mixture Regression method of [16]. $\mathrm{MAE}_{\mathrm{vox}} = 126$ HU, $\mathrm{DSC}_{\mathrm{bone}} = 0.69$.

(d) Random Forest method of [20]. $\mathrm{MAE}_{\mathrm{vox}} = 115$ HU, $\mathrm{DSC}_{\mathrm{bone}} = 0.72$

Figure 2.5: Examples of (a) CT and (b)-(d) sCTs created from UTE scans with different voxel-based approaches. For the supervised methods (c) and (d), training data consisted of pre-aligned MR/CT scans from 3 other patients. The reported quality measures are for the shown patient. Data taken from paper A [11].

model, since the MR intensity ambiguities must be solved in some way. One way to achieve this is to train separate regression models; one for the bone region and one for the remaining regions [23, 24]. This approach, however, requires a segmentation of the bone region, which in the reported approaches, requires a manual delineation. Another way to solve the intensity ambiguities is to extract features from the MR images. These features can then be used in combination with the gray-scale intensity to help distinguish otherwise similar voxels. Aside from spatial features and anatomical knowledge [25], typical features from computer vision describing neighborhood and texture can be used for bone segmentation [26]. A concern with voxel-based methods is that MR voxel intensities can vary between different scanners and different scans even for the same MR sequence and patient. If the methods should be generally applicable, this means that special care should be taken when choosing the input to a voxel-based model. Intensity alone, or features that directly rely on the intensity could yield an approach that is only applicable to certain data or certain scanners.

2.3.2 Atlas-based methods

With conventional atlas-based methods, the sCT prediction relies on a non-linear registration between the patient MR scan to be converted and one or multiple atlases of MR scans with known correspondence to a CT and possibly an organ label-map [27–33]. The MR intensity ambiguities are thus resolved by the spatial information provided by the registration, which means that conventional MR sequences can be used. With a population-based atlas technique, the atlas consists of an average MR scan and a registered average CT. The average scans are formed by registering and averaging a population of MR/CT atlases in a common space prior to observing the patient MR. The sCT is created by non-linearly registering the average MR with the patient MR and applying the same transformation to the average CT, which then provides the sCT estimate [27]. With a multi-atlas technique, the sCT is formed by averaging (or fusing) multiple atlases after a non-linear alignment with the patient MR [28–32, 34]. Each atlas consists of an MR scan and a co-registered CT scan. The procedure is to non-linearly register each atlas MR with the patient MR, after which the same transformations are applied to each atlas CT. The atlas CTs are then fused to provide the sCT estimate. The fusion can be done in a number of ways, e.g., using a local similarity measure to create a locally weighted averaging [28, 30–32]. This gives less weight to regions in the atlases where the non-linear registration was sub-optimal. Figure 2.6 shows an illustration of the multi-atlas approach.

Atlas-based methods generally perform well compared to voxel-based techniques with an $\mathrm{MAE}_{\mathrm{vox}}$ of ≤ 100 HU [30, 35] in the brain. A concern, however, is that

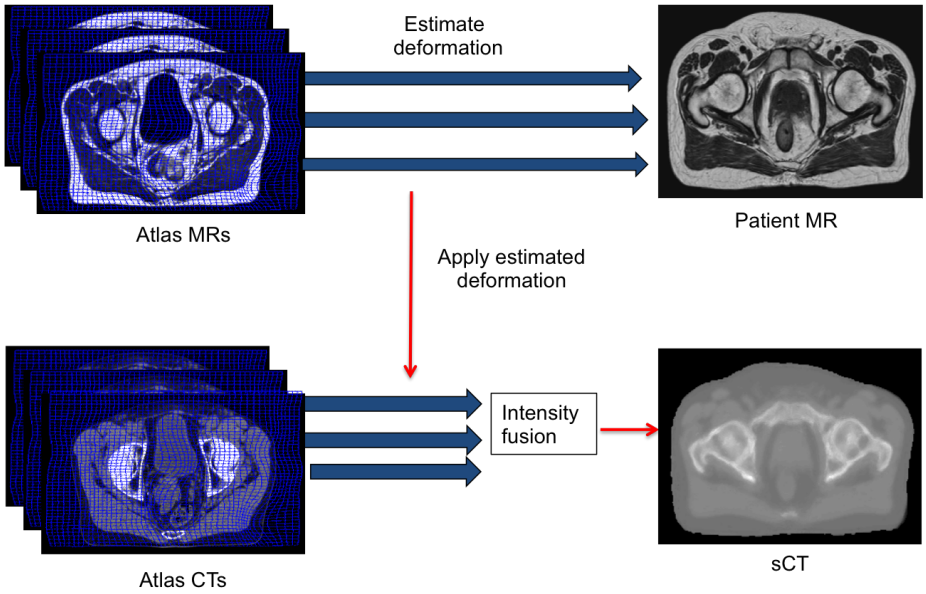


Figure 2.6: Illustration of a multi-atlas approach. The individual atlas MR/CT pairs have been pre-registered. For each atlas MR, a non-linear deformation is estimated that registers the atlas MR with the patient MR. The same deformations are applied to the atlas CTs, which are then fused to provide the sCT estimate.

they depend entirely on the accuracy of the registration; in cases where the patient geometry is very different from the atlas(es), this may pose a problem. The multi-atlas technique remedies this to some degree by using multiple registrations and assuming that at least one of the atlases will be well aligned in a given region of the patient MR. This provides a better sCT estimate than the population-based atlas technique [27, 28]. The improved sCT quality comes at the cost of computation time due to the number of required non-linear registrations between the patient MR and the atlases. Typically between 15-38 patients are used as atlases, meaning that the same number of registrations are required [28, 30, 34].

A way that relies less on the accuracy of the non-linear registration is to perform pattern recognition after aligning the atlases with the patient MR. By defining a rectangular image sub-region (a patch) around each MR voxel and extracting all such patches from the patient MR and registered atlas MRs, a Gaussian process regression model can predict an sCT value based on spatially close and similar atlas MR patches, which have a known correspondence to the atlas

CTs [33]. This improves the prediction accuracy compared to simpler multi-atlas techniques [35]. Some authors, however, point out that the added computational complexity of a pattern recognition step on top of the non-linear registrations may not make it worthwhile [35]. Both the pattern recognition technique and most similarity measures used to create a locally weighted averaging of the atlas CTs rely on an intensity-based similarity measure between the atlas MRs and the patient MR. Similar to the voxel-based approaches, these methods are therefore affected by the relative nature of the MR voxel intensities. An intensity normalization step is usually performed to account for this.

2.4 Considerations regarding treatment site

Most reported techniques for sCT generation have focused on the brain [10, 16–18, 26, 29, 30, 35–38], while relatively fewer attempts have been made for other body sites such as the pelvis [23, 24, 27, 28, 34]. The brain is probably the most straightforward region for CT synthesis. As shown by phantom studies, the limited spatial extent of this region means that geometrical distortions in the MR scans are not a concern as long as the scanner built-in distortion correction is switched on [1,39]. Furthermore, the brain is a relatively rigid structure, which simplifies a registration between the MR and CT scans of the same patient. The assumption of a correct one-to-one spatial correspondence between pre-acquired MR and CT scans forms the basis of all supervised voxel-based approaches as well as atlas-based approaches.

Other body sites can have severe non-rigid deformations between two scans of the same patient. This complicates the intra-patient registration between MR and CT and thus affects the sCT methods. Furthermore, the inter-patient anatomical variation is larger for sites such as the pelvis due to factors like body mass index, patient gender, patient age etc. This mainly affects atlas-based approaches, since they rely on some degree of similarity between the patient and the atlases [27]. Lastly, the greater spatial extent of some body sites may be a concern regarding geometrical distortions in the MR scans. Some authors have reported differences in body contour between pelvic MR and CT scans of the same patients [24, 28]. These differences may also be caused by the aforementioned naturally occuring non-rigid deformations. Other studies suggest that geometrical distortions have a limited magnitude and impact also for larger fields of view [40, 41]

2.5 Conclusion

This chapter has provided an overview of MR-only RT and the different methods for CT synthesis. To sum up, supervised voxel-based approaches are fast at predicting an sCT. However, they require some way of distinguishing bone from air in the MR scans. This can be solved by using UTE scans or by extracting features from conventional MR scans. The UTE scans prolong the patient scan time and are of lower quality than conventional MR scans. The features used are usually not intensity invariant, which may hinder their general applicability across different scanners at different clinics. Atlas-based approaches are appealing since they solve the intensity ambiguities by a non-linear spatial alignment, which means that conventional MR scans can be used. They generally have a higher accuracy than the voxel-based approaches. However, they may be sensitive to anatomical variation between patients and usually require several non-linear registrations, which can be time-consuming.

In the following chapters, I will describe our efforts to address the above issues by: 1) Implementing and validating a patch-based multi-atlas approach for sCT prediction without using non-linear registrations. 2) Speeding up the patch-based pattern recognition step to a clinically acceptable level and testing the developed approach in a challenging non-brain region 3) Implementing a voxel-based method using conventional MR sequences and intensity invariance with a performance close to our atlas-based approach.

Chapter 3

Patch-based generation of sCTs from conventional MR

This chapter introduces the methodology and summarizes the main findings of paper B. In that paper, we introduced a patch-based approach for sCT prediction that did not rely on specialized MR scans, nor on non-linear registrations. Furthermore, we thoroughly compared the method with two state-of-the-art approaches in the voxel-based and atlas-based categories.

3.1 Introduction

In the previous chapter, we briefly mentioned a multi-atlas method utilizing patch-based pattern recognition for improving sCT predictions in regions where non-linear registration was inaccurate [33]. Previously, patch-based methods have seen widespread application in a range of domains such as image denoising, image in-painting, and texture synthesis among others [42–44]. Patch-based segmentation methods have also been proposed in the field of automated MR-based brain and cardiac segmentation with promising results [45–49]. In these approaches, 3D patches (i.e., small cuboidal image subregions) are extracted from the MR and a spatially local search for the most similar patches in a pre-acquired database of labeled MR scans is performed. The known labels

of the resulting database patches are then fused to give the predicted label at each position. An interesting property of some of these methods is that, unlike in [33], they rely only on affine registrations as a rough linear alignment between the database MR scans and the MR to be segmented [45, 47]. The need for an accurate non-linear registration is thus removed and the segmentation is driven mainly by patch similarities.

Based on the promising results of the patch-based approaches for MR segmentation, we chose to investigate the potential of a patch-based method for predicting a continuous-valued sCT of the brain using conventional T_1-weighted MR scans. In particular, we adopted the methodology of [45] and [46], but incorporated the CT HU numbers as label information in the patch model to enable patch-based regression. Furthermore, we investigated the dosimetric properties of the resulting sCT as compared with the CT.

3.2 Patch-based prediction

The assumption behind the patch-based prediction model is that a local similarity in structure and intensity in the MR images of two different patients also translates into a local similarity in their corresponding CT images. With this in mind, we can synthesize an sCT for a new test subject by comparing local regions in the subject MR image with local regions in atlases of MR images with known correspondence to a CT image. When two regions are similar, we can use the known MR-to-CT intensity relationship of the atlas to provide a predicted sCT value. With typical multi-atlas methods, each atlas provides only one weighted candidate for intensity fusion and this candidate is determined entirely by a non-linear registration. The patch-based method, on the other hand, is not limited to using just a single candidate from each atlas patient. Indeed, there may be several patches in one patient that resemble a patch in another, all of which can be used in the intensity fusion. This means that the total number of candidate CT values is increased compared to typical atlas-based methods.

3.2.1 Intensity normalization

The main driver of the patch-based approach is an intensity-based similarity search. Since the tissues in MR scans do not necessarily have a consistent intensity between different scans, a step is needed to rectify this. We use a histogram-matching approach based on intensity landmarks in the MR scans [50]. To normalize the intensities for a group of MR scans, first a simple foreground mask

is created for each scan by thresholding based on their mean intensity. Next, three intensity landmarks are computed from the foreground of each scan: p_1, the minimum intensity, p_2, the intensity at the 99.8^{th} percentile, as well as p_{med}, the median intensity. The intensities of each scan are then scaled to a chosen range (e.g., $[s_1, s_2] = [0, 255]$) and a common landmark, s_{med} is calculated as the average median intensity of the foreground of all scans in the new intensity range. Now for each scan in the original intensity space, the intensities in $[p_1, p_{\text{med}}]$ are linearly scaled to $[s_1, s_{\text{med}}]$ and intensities in $[p_{\text{med}}, p_2]$ are scaled to $[s_{\text{med}}, s_2]$. An illustration of the resulting piece-wise linear scaling segments is shown in Figure 3.1. In Figure 3.2, histograms of T_1-weighted MR scans of five patients acquired on the same scanner are shown before and after normalization.

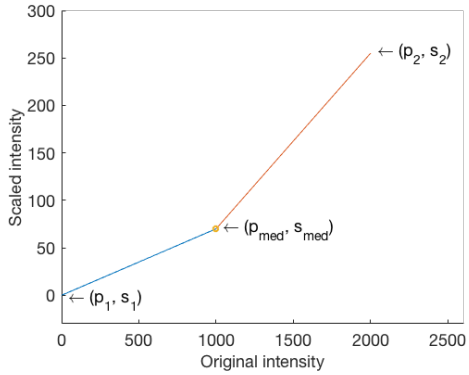


Figure 3.1: Intensities in each MR scan are normalized using individual landmarks $\{p_1, p_{\text{med}}, p_2\}$ and common landmarks $\{s_1, s_{\text{med}}, s_2\}$. Artificial data was use here for illustrative purposes.

3.2.2 Similarity search and intensity fusion

A patch, $P(\mathbf{x})$, is defined as a cube with side length p voxels centered on the spatial location $\mathbf{x} = (x, y, z)$ in an MR image. Similarly, a target value, $T(\mathbf{x})$, is defined for each $P(\mathbf{x})$ as the Hounsfield unit (HU) value at $\mathbf{x}$ in the corresponding aligned CT image. For S patients, patches and corresponding target values are extracted for all positions, $\mathbf{x}$, to create a database of patches, $P_s(\mathbf{x})$, with corresponding target values, $T_s(\mathbf{x})$, where s denotes one of S patients. Using this database, an sCT for a test patient is predicted by extracting patches from his/her MR scan and performing an intensity-based nearest neighbor search in the patch database. For a patch at position $\mathbf{x}' = (x', y', z')$ in the MR of the

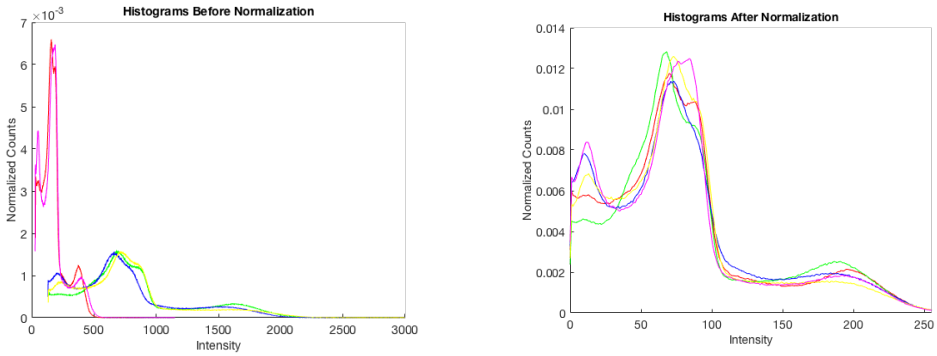


Figure 3.2: Intensity histograms of five T_1-weighted brain MR scans of different patients before and after normalization. Background values have been excluded. Each scan is color-coded differently. The scans were acquired on a 1 T Philips Panorama scanner.

test patient, the similarity measure is the squared L_2-norm defined as:

$$d(s, \mathbf{x}) = ||P(\mathbf{x}') - P_s(\mathbf{x})||_2^2. \tag{3.1}$$

The search then consists of finding the database patch that minimizes $d(s, \mathbf{x})$ and storing the corresponding patient and spatial position:

$$(s^{min}, \mathbf{x}^{min}) = \underset{s,\mathbf{x}}{\operatorname{argmin}}\ d(s, \mathbf{x}). \tag{3.2}$$

The search can be extended to find the K most similar database patches and the K relevant target values are extracted from the database as $T_{s_k^{min}}(\mathbf{x}_k^{min})$ with $k = 1, .., K$. To assign an sCT HU value, a similarity-weighted average is computed:

$$\mathrm{sCT}(\mathbf{x}') = \frac{\sum_k w_k \times T_{s_k^{min}}(\mathbf{x}_k^{min})}{\sum_k w_k}, \tag{3.3}$$

with weights defined as:

$$w_k = \exp\left(\frac{-d(s_k^{min}, \mathbf{x}_k^{min})}{\min_k d(s_k^{min}, \mathbf{x}_k^{min})}\right). \tag{3.4}$$

This ensures that if one patch is more similar than the rest, this patch is weighted highly; conversely, if all patches are equally similar, they are weighted equally.

3.2.3 Search space reduction

The above procedure requires a patch similarity search in the entire database of patches for each patch in the patient MR. Since a typical MR scan consists of millions of voxels, this would be computationally very expensive. A straightforward way to reduce the computational burden is to only consider voxels within the patient and simply assigning a bulk density of -1000 HU (air) to all background voxels. A further way to reduce the search space is to put a spatial constraint on the database patches such that only database patches that are spatially close to $\mathbf{x}'$ will be considered. To infer the spatial relationship between the database MRs and the patient MR, an affine alignment of each database MR and the patient MR is therefore performed. The search is then constrained to only contain database patches in a local cuboidal search volume, V_{search}, of side length v voxels around $\mathbf{x}'$, such that $\mathbf{x} \in V_{search}$ in Equation 3.2. This reduces the number of database patch candidates to $S \times v^3$ for each patch in the patient MR. Finally, highly dissimilar patches can be discarded using a computationally cheaper similarity measure. The structural similarity measure (SSIM) [51] is based on the mean and variance in each patch, and is defined as:

$$\mathrm{SSIM}(\mathbf{x}, s) = \frac{(2\mu_{\mathbf{x}'}\mu_{\mathbf{x},s} + c_1)(2\sigma_{\mathbf{x}'}\sigma_{\mathbf{x},s} + c_2)}{(\mu_{\mathbf{x}'}^2 + \mu_{\mathbf{x},s}^2 + c_1)(\sigma_{\mathbf{x}'}^2 + \sigma_{\mathbf{x},s}^2 + c_2)} \tag{3.5}$$

where $\mu_{\mathbf{x}'}$ and $\sigma_{\mathbf{x}'}$ are the mean an standard deviation of the patch at position $\mathbf{x}'$ in the test patient and $\mu_{\mathbf{x},s}$ and $\sigma_{\mathbf{x},s}$ are the same for a patch at position $\mathbf{x}$ in database patient s. c_1 and c_2 are small constants to avoid infinity. All patches with $\mathrm{SSIM} < 0.95$ are discarded prior to the patch search performed in Equation 3.2. A situation may arise where all patches are discarded in a search volume. In such situations, the sCT value in the affected voxel is flagged as unknown and in post-processing, the voxel is assigned the average sCT value of the closest assigned voxels. For the intensity fusion in Equation 3.3, the $K = 8$ most similar patches are used unless the SSIM only allows a smaller number. An illustration of the patch-based sCT generation is shown in Figure 3.3 for a simplified 2D case.

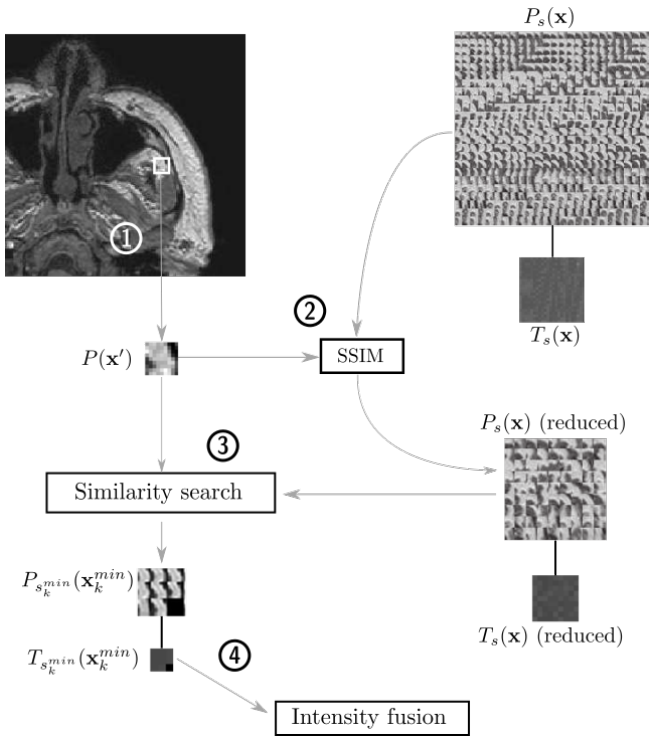


Figure 3.3: 2D patch-based sCT prediction. Step 1: A patch, $P(\mathbf{x}')$, is extracted from the patient MR at position $\mathbf{x}'$. Step 2: The SSIM between MR patches is used to reduce the database of spatially close patches, $P_s(\mathbf{x})$, stored with their corresponding CT value, $T_s(\mathbf{x})$. Step 3: Using Equation 3.2, the $k = 1, .., 8$ most similar database patches, $P_{s_k^{min}}(\mathbf{x}_k^{min})$, and their CT values, $T_{s_k^{min}}(\mathbf{x}_k^{min})$, are found from the remaining database. Step 4: Using Equation 3.3, the CT values, $T_{s_k^{min}}(\mathbf{x}_k^{min})$, are combined to produce the final sCT value at position $\mathbf{x}'$.

3.3 Experiments and results

In this section we summarize the main results presented in paper B [52]. This paper consists of a thorough comparison of the patch-based approach with two state-of-the-art approaches within the voxel-based and multi-atlas-based categories, respectively. The voxel-based method was based on dual echo UTE scans and used a Gaussian mixture regression (GMR) model to learn the relationship between MR and CT intensities [16]. The multi-atlas method used non-linear registrations of T_1-weighted scans and local normalized cross-correlation for the intensity fusion step [30]. We refer to paper B for the exact implementation details, but note that the main parameters for each method were chosen based on a nested cross-validation in order to ensure a fair comparison.

3.3.1 Data

Data consisted of MR and CT scans of five whole brain RT patients, three male and two female, aged 55-82 years. The CT scans were acquired on a Philips Brilliance Big Bore CT with a voxel resolution $0.6 \times 0.6 \times 2$ mm, $512 \times 512 \times 110$ voxels using a standard protocol for brain scans (120 kV, 300 mAs). The MR scans were acquired on a Philips Panorama 1 T open scanner. The sequences were a T_1-weighted 3D Fast Field Echo (FFE), TE/TR = 6.9/25 ms, voxel resolution $0.85 \times 0.85 \times 1.2$ mm, $188 \times 188 \times 152$ voxels and two UTE scans at flip angles 10° and 25° with $TE_1/TE_2/TR = 0.09/3.5/7.1$ ms and an isotropic voxel resolution of 1 mm with 256^3 voxels. The patients were fixed in treatment position during both the MR and CT scanning using thermoplastic masks. Each patient's MR/CT pair was rigidly aligned using mutual information with the default settings as implemented in 3D Slicer [53].

3.3.2 Evaluation

The sCTs were evaluated using a leave-one-out scheme in terms of their voxel-wise, geometric, radiologic, and dosimetric accuracy using the measures described in Chapter 2. For the WEPL evaluation, paths were defined as radial spokes from the center of the brain towards the skin of the patient. For the dosimetric evaluation, two planning target volumes (PTVs) were defined for each patient; one in an anatomically homogeneous region (PTV 1) and one in a challenging heterogeneous region (PTV 2). As shown in Table 3.1 and Figure 3.4, we found that the voxel-based GMR approach had a lower voxel-wise, geometric, and radiologic accuracy compared to the other two methods. This is also clear

from the visual quality of the sCTs shown in Figure 3.5. In all non-dosimetric measures, the patch-based approach had the smallest error, closely followed by the multi-atlas approach. A study containing more patients is probably still needed to determine whether there is a significant difference between the two. In the dosimetric evaluation, PTV2 proved challenging for both the GMR and multi-atlas approach, see Figure 3.6. This can potentially be explained by a lower sCT quality in the nasal cavity of these approaches, since PTV2 was positioned next to these cavities. Currently there is no consensus on the required dosimetric accuracy of an sCT, however, a conservative requirement could be to state that for 95% of patients the dosimetric deviation in PTV coverage should be within 2% [54]. A population-based study is required to establish if this requirement is met; however, the patch-based method passed this criterion for both PTVs for the five patients used here.

	Patch-based	**Multi-atlas**	**GMR**
$\mathrm{MAE}_{\mathrm{vox}}$ [HU]	85 ($\sigma = 14$)	97 ($\sigma = 19$)	148 ($\sigma = 22$)
$\mathrm{ME}_{\mathrm{vox}}$ [HU]	1 ($\sigma = 14$)	-4 ($\sigma = 17$)	22 ($\sigma = 28$)
$\mathrm{DSC}_{\mathrm{bone}}$	0.84 ($\sigma = 0.02$)	0.83 ($\sigma = 0.01$)	0.67 ($\sigma = 0.03$)
$\mathrm{MAE}_{\mathrm{WEPL}}$ [mm]	2.2 ($\sigma = 1.0$)	2.7 ($\sigma = 0.8$)	4.8 ($\sigma = 1.3$)
$\mathrm{ME}_{\mathrm{WEPL}}$ [mm]	0.4 ($\sigma = 1.8$)	-0.6 ($\sigma = 1.9$)	1.1 ($\sigma = 2.1$)

Table 3.1: Voxel-wise, geometric and radiologic error measures. Average value and standard deviation (σ) for five patients are shown.

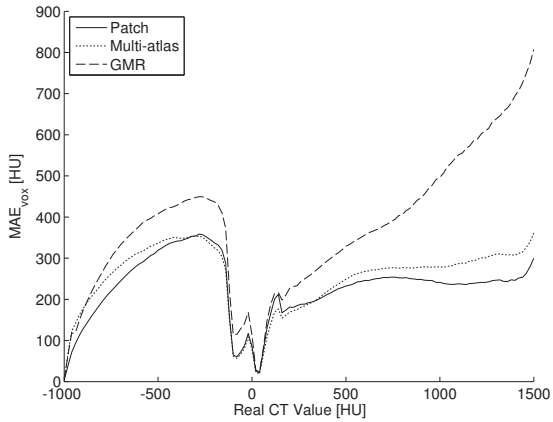


Figure 3.4: $\mathrm{MAE}_{\mathrm{vox}}$ calculated in bins of 20 HU and averaged across the five patients. Figure from [52].

3.3.3 Discussion

With these promising results, we demonstrated that the patch-based approach could be used for sCT prediction without relying on non-linear registrations. A rough affine alignment is computationally simpler and potentially less prone to fail even for regions with large anatomical variation (i.e., non-brain regions). A concern, however, could be that for such regions, an affine registration would be insufficient to bring similar structures into the necessary alignment for the local patch search. A larger search volume would then be needed, which would slow down the method. Computation time for the patch search step is another concern. The time saved by using affine registrations instead of non-linear ones is by far exceeded by the time required for the patch search. As already pointed out in [35], the improvement from a patch search may not be worth-while compared to simpler atlas-based approaches. Using the patch-based approach, we achieved an average prediction time of approximately 15 hours for one sCT, which is too much to enable same-day planning and treatment of e.g., palliative patients.

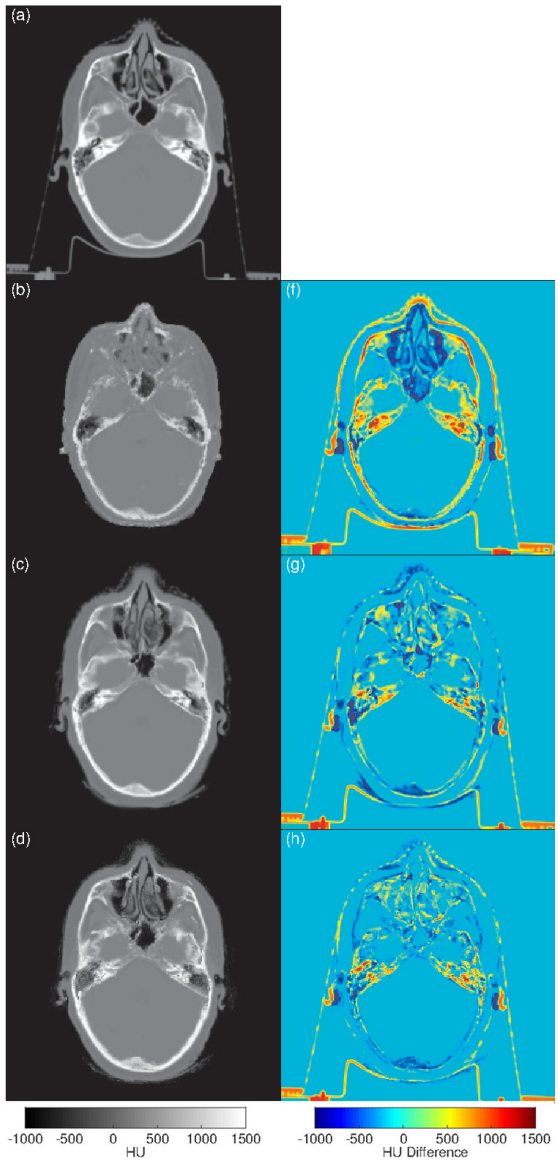


Figure 3.5: Transverse slices for comparison of sCTs with the CT. (a)-(d) shows the CT, the GMR sCT, the multi-atlas sCT and the patch-based sCT, respectively. (f)-(h) shows the difference maps between the CT and the GMR, the multi-atlas and the patch-based sCTs, respectively. Negative values indicate an overestimation of the HU value and positive values indicate an underestimation. Figure from [52].

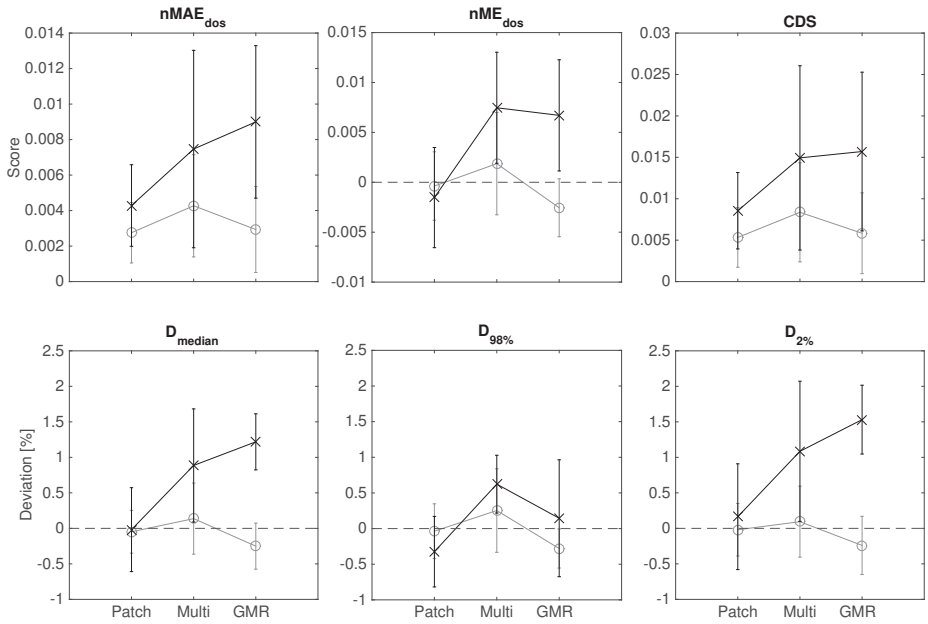


Figure 3.6: Dosimetric errors for the two PTVs. PTV 1: Gray lines and open circles. PTV 2: Black lines and crosses. Dashed line indicates zero. Top row: normalized mean absolute dosimetric error ($\mathrm{nMAE_{dos}}$), normalized mean dosimetric error ($\mathrm{nME_{dos}}$) and combined dosimetric score (CDS). Bottom row: percentage point deviation in $\mathrm{D_{median}}$, $\mathrm{D_{98\%}}$ and $\mathrm{D_{2\%}}$. Average values are shown along with $\pm\sigma$ interval. Figure from [52].

Chapter 4

Speeded up patch-based generation of sCTs in the pelvis

This chapter introduces the methodology and summarizes the main findings of paper C. In that paper, we introduced an approximate nearest neighbor algorithm for speeding up the patch-based approach for sCT prediction. Furthermore, we applied the method in the pelvic region and thoroughly evaluated the resulting sCTs.

4.1 Introduction

In the previous chapter, the patch-based approach showed a promising potential for MR-only RT in the brain region. This served as motivation for also testing the method in other body regions. A downside of the method was the computation time, which made a clinical implementation applicable to all patient groups difficult. The search for similar patches is a nearest-neighbor problem in a high-dimensional vector space, which becomes increasingly demanding as the patch size and number of candidates increase. Due to the high dimension-

ality, a linear search is often the most efficient for solving this problem [55], but this is prohibitively slow for our application. So-called approximate nearest neighbor algorithms try to speed up the search by not guaranteeing to find the exact nearest neighbors but only good approximations. One such approach, specifically designed for patch matching in 2D images, is the PatchMatch (PM) algorithm [56]. Because it is designed to utilize the coherency and spatial structure in images, it uses less memory and is faster than other approximate algorithms such as kd-trees [56, 57]. Like the patch-based approach described in the previous chapter, PM is another example of a computer vision/image editing algorithm that finds usage in medical image analysis. Different authors have successfully adapted PM for patch matching in a multi-atlas setting with MR data [58–60]. For some, the reported speed-up enables near real-time hippocampus segmentation of MR images whilst maintaining a competitive accuracy [60]. With these promising results, the adapted PM algorithm could potentially also speed up sCT prediction.

As described in Chapter 2, non-brain body sites have some intrinsic challenges, which may affect an atlas-based method such as the patch-based approach. After introducing the adapted PM algorithm, we proceed with a description of the changes needed to adapt the patch-based method for the pelvis.

4.2 PatchMatch Algorithm

The original PM algorithm is designed for finding patch correspondences in two 2D images, A and B. It exploits a spatial coherence in the patch correspondences such that if a good match is found for one patch, good matches for its adjacent patches are assumed to be present adjacent to that match. This drastically cuts down on the required number of distance calculations compared to a brute-force search. It consists of 3 main steps: a random initialization, a propagation phase, and a random search phase.

Initialization The algorithm is initialized by randomly assigning a matching patch in B to each patch in A. Let f be a mapping function that for each pixel coordinate (i, j) in A stores the pixel coordinates (i', j') of the current best matching patches in B such that $f(i, j) = (i', j')$. For each pixel coordinate in A, the value of f is initialized by a uniform random sampling of coordinates (i', j') in B.

Propagation After the initialization, a propagation of good matches is carried out. Given a patch, $P(i,j)$, centered on the coordinate (i,j) in A, the idea is that if a good match is found at $f(i-1,j)$ in B for the adjacent patch (one pixel to the left) $P(i-1,j)$ in A, then the patch at $f(i-1,j)+(1,0)$ (one pixel to the right) in B is going to be a good match for $P(i,j)$ in A. By checking for improved candidates at $f(i-1,j)+(1,0)$ and $f(i,j-1)+(0,1)$ in scan order (left to right, top to bottom) in the image, good matches are propagated between adjacent patches.

Random search After propagation, a random search for better candidates is carried out at an exponentially decreasing spatial search radius from the current best matches. Usually this search radius initially spans the entire image and is exponentially decreased a number of times until a pre-defined threshold. The random search is done to avoid that the propagation phase gets stuck in a local minimum.

Iterations The algorithm proceeds by switching between propagation and random search for a fixed number of iterations or until convergence. At even iterations the propagation is done by checking the patches at $f(i+1,j)-(1,0)$ and $f(i,j+1)-(0,1)$ in reverse scan order.

4.2.1 Adaptation for a multi-atlas setting

A number of changes are needed to adapt the PM algorithm for use with volumetric data and multiple atlases. We follow most of the methodology of [58] and [60]. An illustration of the different steps of the adapted PM algorithm are shown in Figure 4.1.

Initialization As the data are volumetric and consist of multiple atlases, a random voxel coordinate (i',j',k') *and* atlas patient, s, for the candidate nearest neighbors must now be chosen such that $f(i,j,k)=(s,i',j',k')$ for each patch in the patient MR. Furthermore, the search volume, V_{search}, used in the patch-based prediction model is used to constrain the initialization. In Figure 4.1, three adjacent patches (red, green, and blue) are initialized with candidate nearest neighbors in three different atlas patients within V_{search} (yellow dashed line).

Propagation The propagation phase is similar to the original PM algorithm except that the third dimension $(i, j, k \pm 1)$ is included in the check for better candidates. In Figure 4.1, propagation of a better candidate to the red patch at (i, j, k) in the patient MR is shown. The blue patch is at $(i, j-1, k)$ and assuming that the nearest neighbor of the blue patch is at $f(i, j-1, k)$ in atlas 1, the patch at $f(i, j-1, k)+(0, 0, 1, 0)$ is checked for a better match to the red patch. Assuming a better match is found, the nearest neighbor of the red patch is now situated in atlas 1 instead of 3. By also checking at $f(i-1, j, k)+(0, 1, 0, 0)$ (the match for the green patch), the nearest neighbor could move to atlas 2.

Random search The random search is constrained by V_{search} and is only performed in the atlas patient of the current nearest neighbor. The tested candidates are chosen randomly within V_{search}, and not within an exponentially decreasing spatial search radius. In this step, candidate nearest neighbors can not jump between atlas patients as in the propagation phase.

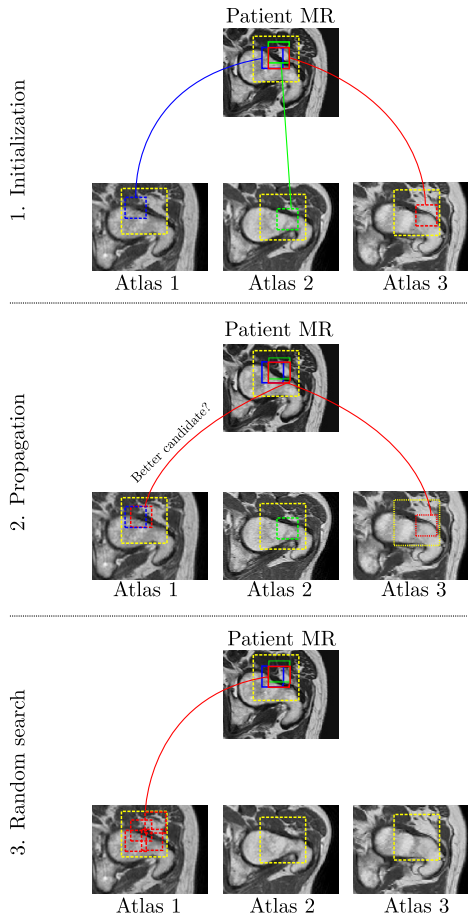


Figure 4.1: The PM algorithm used with multiple atlases. 1. For each patch in the patient MR image (here shown as adjacent red, green, and blue patches), a random initialization of candidate nearest neighbors (shown in same-colored dashes) are assigned by choosing a random atlas and a random position within a constrained neighborhood (shown in yellow dashes). 2. In the propagation phase good matches are propagated between adjacent patches, here shown for the red patch: The blue patch is to *left* of the red patch in the patient MR, so we check if the patch to the *right* of the current nearest neighbor of the blue patch is a better match for the red patch. Similarly, we check in the "above" (green patch) and "in" directions (not shown). 3. A random search for improved candidates of the red patch is carried out in the constrained search area of the atlas of the current nearest neighbor. The process is iteratively going between 2. and 3., in each propagation phase switching the direction (left/above/in or right/below/out) in which adjacent patches are considered.

4.3 Adaptation for the pelvic region

Aside from using an approximate patch search, we also make some other changes to the approach described in paper B in order to adapt the method for the pelvis.

Non-linear intra-subject registration To ensure spatial correspondence between voxels and to account for organ movement between the scans, we non-linearly align each atlas patient's MR scan to his CT scan using Elastix [61]. This is required due to the non-rigidity of the pelvic region, which causes intra-patient differences in the MR and CT scans. These differences are mainly observed in the body outline, bladder shape and positions of air in the intestines. The non-linear registration is performed once to prepare a good training MR/CT database but is not required during sCT prediction.

Atlas pre-selection The patch-based method relies on affine registrations to roughly align the atlases with the patient MR. On our data of ten patients, this worked for the most part without issues, i.e., registering nine atlases to one patient MR. However, there were cases where one or more atlas patients were poorly aligned with the test patient MR after the registration. Misalignment happened frequently for two patients, where the field of view covered a slightly different part of the pelvis than the rest of the patients, e.g., when parts of the femoral heads were missing. The consequence of the misalignment was that the limited patch search volume became situated in wrong regions of the atlas patient. It seemed that when this happened, similar patches could still be found in the wrong region, causing prediction errors. A popular solution used in atlas-based brain segmentation is to introduce an atlas pre-selection step, which leaves out the most dissimilar patients [62]. We employ such an atlas pre-selection based on the squared Euclidean distance between the registered atlas MR and the patient MR, calculated within the body contour of the patient MR after intensity normalization. We keep the five most similar atlases for the remaining steps.

Intensity normalization In paper B, we used a single intensity landmark to split the intensity scaling in two linear segments (Figure 3.1). This landmark was the median intensity and the assumption was that it would robustly correspond to the same tissue type across patients. In order for this assumption to be valid, the relative amounts of each tissue type have to be constant across patients, which was not the case in the pelvis. The fat-to-muscle ratio was not constant and since fat and muscle have distinct intensities in T_1-weighted MR, the median

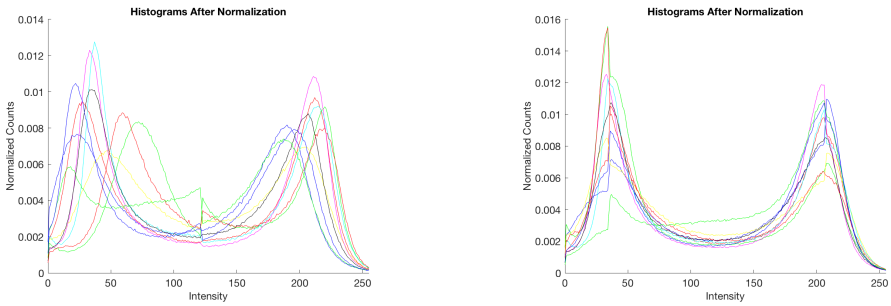


Figure 4.2: Intensity histograms of 10 T_1-weighted pelvic MR scans of different patients after intensity normalization. Left panel shows the normalization using the median landmark. Right panel shows the normalization using two landmarks.

intensity was sensitive to this ratio. Instead of the median intensity we therefore use two landmarks determined as the histogram peaks corresponding to muscle and fat. This provides a more robust intensity normalization as shown in Figure 4.2.

4.4 Experiments and results

In this section we summarize the main results presented in paper C [63]. This paper consists of a thorough evaluation of the approximate patch-based approach applied in the pelvis. Furthermore, a comparison with a simple bulk density sCT approach (called MR_w), where the whole patient was assigned a water-equivalent HU, is carried out.

4.4.1 Data

MR and CT scans of ten prostate RT patients aged 55-82 years were retrospectively obtained. The CT scans were acquired on a Philips Brilliance Big Bore CT using a standard protocol for pelvic scans (120 kVp, 232-503 mAs). The scans were acquired with a voxel spacing of between $0.78 \times 0.78 \times 2$ mm and $1.4 \times 1.4 \times 2$ mm, for an in-plane matrix of 512×512 voxels and 129-199 slices. The MR scans were acquired on a Philips Panorama 1 T open scanner using a bridge body coil with a T_1-weighted sequence, TE/TR = 10/623 ms. The voxel spacing was $0.80 \times 0.80 \times 4$ mm, for an in-plane matrix between 528×528 and

640×640 voxels and 16-24 slices. The approximate duration of the T_1-weighted sequence was 5-7 minutes. The patients were fixed in treatment position during both the MR and CT scanning using the same fixation devices.

4.4.2 Algorithm speed-up

As a first test of the PM algorithm, we ran the algorithm on the same brain data as used in paper B, with the same parameter settings (patch size, search volume size). As shown in Table 4.1, the speed-up of sCT prediction is substantial. On average, the MAE_{vox} was increased by 12 HU, which we find is a worthwhile compromise. We note, however, that the average MAE_{vox} is now equal to what we achieved with the multi-atlas approach in paper B, which is still low compared to many of the approaches described in chapter 2. Whether the same is the case for the dosimetric performance, remains to be investigated. The average prediction time for the pelvis scans was 20.8 minutes.

	Median prediction time	Avg. MAE_{vox}
PM	12.4 min	97 HU
Brute force	16.3 hours	85 HU

Table 4.1: Comparison of the adapted PM algorithm with the brute force search used in paper B. The time needed for affine registrations is not included.

4.4.3 Evaluation

sCTs were predicted for each patient in a leave-one-out fashion, using the remaining nine patients as atlases. In Figure 4.3, examples of the resulting sCTs using the approximate patch-based approach and the MR_w approach are shown. In most regions, the patch-based method predicts the correct HU value as illustrated by the difference maps. The compact bone region is slightly underestimated, illustrating the challenging nature of these voxels.

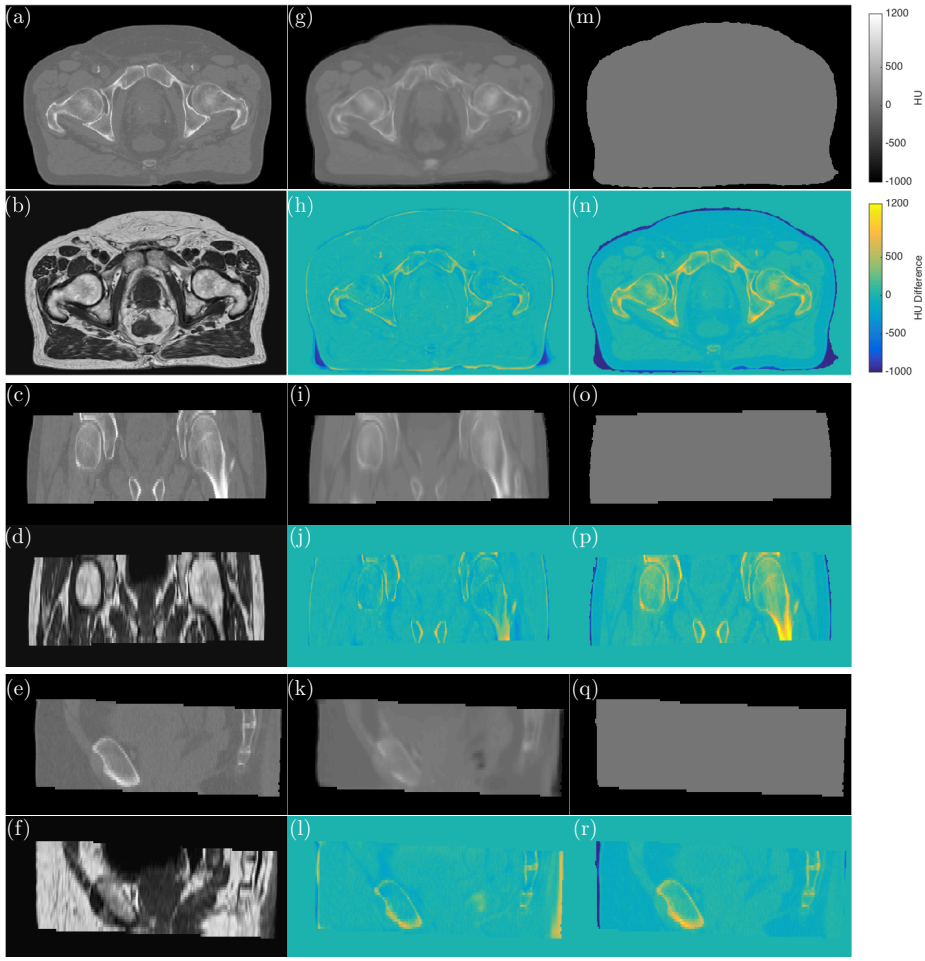


Figure 4.3: Qualitative comparison of CT and sCTs. (a)-(f) show alternately the CT and the MR (non-linearly registered to the CT) that the sCTs are based on, for different views. (g)-(l) show alternately the patch-based sCT and the difference maps between the patch-based sCT and CT. (m)-(r) show alternately the MR_w sCT and the difference maps between the MR_w sCT and CT. The top colorbar relates to the sCTs and the bottom colorbar to the difference maps. A negative HU difference indicates an overestimation of the true HU value.

In Table 4.2, voxel-wise, radiologic and dosimetric (γ-index pass rates) error measures are shown. The patch-based approach performs significantly better than the MR_w method, which should be the case given the simplicity of the MR_w.

Metric	Patch-based	MR_w	p-value
MAE_{vox} [HU]	54 (8)	105 (12)	$< 10^{-7}$
ME_{vox} [HU]	-1 (7)	-16 (11)	0.003
MAE_{WEPL} [mm]	1.2 (0.3)	2.5 (0.3)	$< 10^{-10}$
ME_{WEPL} [mm]	-0.3 (0.7)	0.5 (0.5)	$< 10^{-4}$
Pass rate $\gamma_{2\%/2mm}$	0.99 (0.01)	0.98 (0.01)	0.014
Pass rate $\gamma_{1\%/1mm}$	0.97 (0.02)	0.94 (0.03)	0.019

Table 4.2: Voxel-wise and radiologic error (rows 1-4) as well as the mean 2D γ-index pass rates for all dose points greater than 10% of the prescribed dose (rows 5-6). Average value and standard deviation (σ) for ten patients are shown. In the rightmost column are the results of a paired t-test on the difference between the patch-based and MR_w methods.

In Figure 4.4, the percentage deviation in dose volume histogram (DVH) points relevant to the planning target volume (PTV), rectum, and the femoral heads are shown. The median deviation for the patch-based approach is close to 0% for all the evaluated DVH points. In general, our DVH analysis shows the importance of evaluating not just the high dose PTV region, but also the OARs, which are typically situated in the medium and lower dose regions. It is mostly in these regions that the patch-based sCT strategy demonstrates its advantages, showing a significant difference from the MR_w approach.

4.4.4 Discussion

With the sCTs generated using the approximate patch-based method, we showed promising results for the pelvic region. This region is more challenging than the brain, due to the greater anatomical variation and non-rigid nature of the pelvis. On the other hand, the large amount of soft tissue compared to compact bone in this region may make it more forgiving towards incorrect bone predictions. In our opinion, the MR_w approach performed surprisingly well in the DVH analysis of the PTV as well as in the γ pass rates. The lack of bone predictions did, however, cause significant dosimetric differences in the femoral heads.

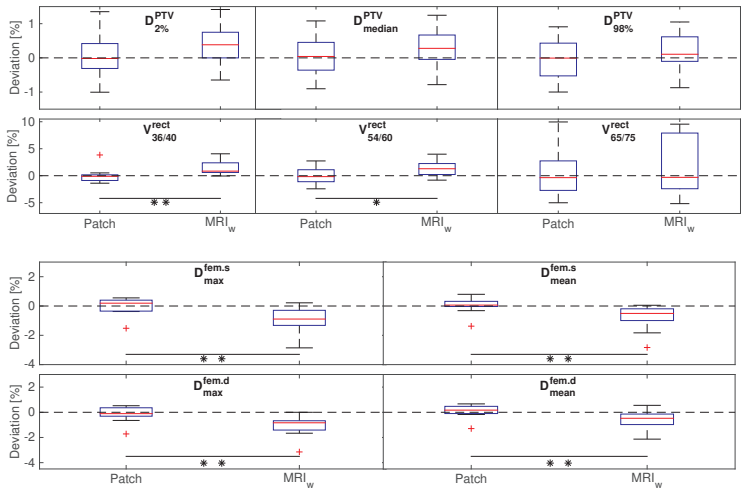


Figure 4.4: Tukey-style boxplots showing the relative deviation in percent for the different methods in relevant DVH points for the PTV and OARs. $**$ indicates a significant difference at the $\alpha = 0.01$ level and $*$ indicates a significant difference at the $\alpha = 0.05$ level. The dashed horizontal line indicates zero deviation. rect: rectum, fem.s: femur sinister, fem.d: femur dexter.

The non-rigidity of the pelvic region poses a substantial challenge, as non-linear registration is required to align the intra-patient MR and CT scans in the atlas database. Since we cyclically use each atlas as test subject in the leave-one-out cross-validation, the accuracy of this registration affects both the evaluation and prediction, as it determines the ground truth relationship between MR and CT and thus also between sCT and CT. The general outcome of the registration was satisfactory, but for some patients a difference between MR and CT was still observed post-registration. It should be noted that the non-linear registrations only need to be performed once in order to create spatial correspondences in the atlases. This is done in a separate offline step. When this is done, the method only relies on affine registrations to predict an sCT. Since it does not affect the algorithm run-time but is important for the sCT prediction, it could potentially be worth the effort to improve the quality of the non-linear registrations when creating the atlases. This could be achieved by a structure-guided registration [28].

Due to variations in the data acquisition, the inter-patient affine registration of atlas MRs with the patient MR failed for some patients. This affected the prediction accuracy of the patch-based approach. An atlas pre-selection step to remove globally dissimilar atlases solved this issue. Ensuring that the exact

same anatomical region is always covered in the scan could potentially remove it altogether. A more robust affine alignment could also be investigated [32].

The above observations regarding the difficulty in registrations show one of the main points in switching to MR-only RT in the first place, which is to avoid introducing a registration-induced uncertainty when transferring delineations from the MR to the CT. Some authors have argued against atlas-based approaches because of the contradiction in trying to remove registration uncertainties by using registrations [64, 65]. The patch-based approach *is* dependent on rough affine alignments of the atlas MRs with the patient MR. It is important to note, however, that unlike the *single* MR to CT registration performed to transfer delineations in the MR/CT treatment planning workflow, the patch-based method relies on *multiple* MR to MR registrations, which only need to roughly align the scans. If one completely fails, atlas pre-selection ensures that the sCT is not necessarily affected. Still, it is theoretically possible to encounter a patient for which all registrations fail, which is the intrinsic weakness of an atlas-based method. Another potential issue with atlas-based approaches (including the patch-based method) is that they are non-parametric. This means that the main computations are done in the prediction step, i.e., when registering the atlases with the patient MR and doing the intensity fusion of the atlas CTs. We have presented a substantially faster patch-based approach than in paper B with a prediction time of approximately 12 minutes in the brain and approximately 20 minutes in the pelvis. This enables same-day planning and treatment of patients. With the introduction of the MR-Linac, even faster prediction times would probably still be preferred to generate an adapted treatment. This could potentially be achieved by a GPU-based implementation.

Chapter 5

MR normalization tests and impact of different MR scanners on sCT prediction

This chapter describes some preliminary tests of the patch-based approach described in chapters 3 and 4 in the scenario that the atlas MR scans are acquired on a different scanner than the patient MR. Furthermore, it includes a test of the method when unexpected anatomical differences are present in the patient MR.

5.1 Introduction

Our patch-based approach relies on finding similar patches by comparing intensity values in the MR scan of the patient and the MR scans in the atlas-database. To enable this comparison, the method includes a normalization step, which attempts to align the intensity histograms of atlas and patient MRs. The method is landmark-based, which means that it attempts to find consistent landmarks that correspond to the same tissue type in the histogram of each MR scan. The landmarks are then aligned by a piece-wise linear scaling of the histograms. For

the brain, we used a single histogram landmark, the median intensity, which was aligned across patients. For the pelvis, we found that the median intensity was not a robust landmark because the relative amounts of fat and muscle varied between patients. This could skew the median value either toward the intensity of fat or the intensity of muscle depending on the patient. We found that the distinct shapes of the histograms in the pelvic region required two landmarks to robustly normalize the intensities. These landmarks corresponded approximately to the intensity of fat and the muscle, respectively, and were found by a simple peak detection algorithm.

Until now, all the MR normalizations were carried out on data from the same MR scanner, a 1 T Philips Panorama. In practice, it would be desirable to have a method that can be applied across scanner platforms, imaging protocols, and disease states. For both the approximate and the exact patch-based method, being scanner-independent requires that the pre-acquired atlas database can be used to convert patient data coming from different MR scanners than the one used to acquire the atlases. For this reason, we decided to test the use of scans from a Panorama scanner as atlases to predict sCTs of the brain on data coming from a 1.5 T Philips Achieva scanner. We also decided to further test the robustness of the normalization procedure by predicting sCTs from Gadolinium (Gd) contrast-enhanced scans acquired on the Achieva scanner, again using the Panorama scans as atlases. This is a test of both the algorithm's robustness to data from a different source as well as to data containing abnormalities (Gd-enhanced tumors), which are not present in the atlas scans.

5.2 Experiments and results

We apply the previously described approximate patch-based method to predict sCTs from the Achieva scans with and without contrast-enhancement.

5.2.1 Data

The atlases we used consisted of four patients scanned at the 1 T Panorama with a 3D T1-weighted sequence (TE/TR = 6.9/25 ms, voxel spacing $0.85 \times 0.85 \times 1.2$ mm, $188 \times 188 \times 152$ voxels). The test patients consisted of four other patients scanned at the 1.5 T Achieva scanner using a 3D T_1-weighted sequence (TE/TR = 4.1/25 ms, voxel spacing $0.4 \times 0.4 \times 0.9$ mm, slice thickness 1.8 mm, $512 \times 512 \times 180$ voxels). In addition, each test patient had a Gd contrast-enhanced scan, highlighting tumors, also acquired on the Achieva. One of the

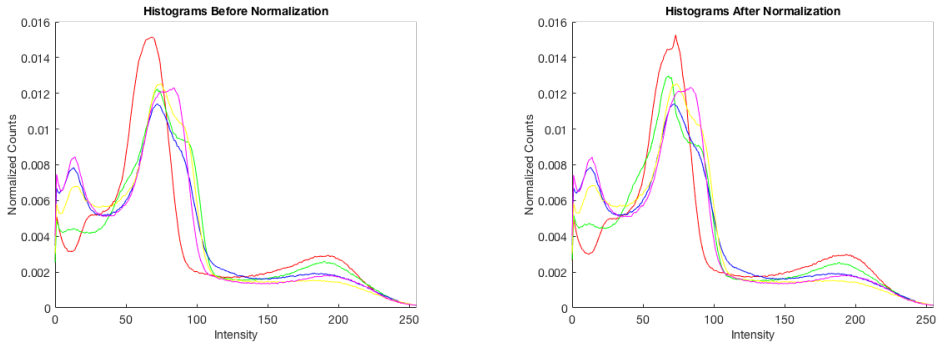


Figure 5.1: Intensity histograms of T_1-weighted brain MR scans of different patients before (left) and after (right) intensity normalization. The red curve is for a scan acquired on an Achieva scanner and the remaining ones are for the atlas scans acquired on a Panorama scanner.

test patients also had a scan on the Panorama, which enabled a direct comparison of the resulting sCTs predicted from the different data sources but with the same atlases. All patients had one corresponding CT scan for comparison, which was acquired on a Philips Big Bore CT scanner using a standard protocol. All MR scans were registered to the CT with a rigid registration using the Elastix toolbox [61].

5.2.2 Normalization and prediction with different scanners

In Figure 5.1, examples of the histograms of one Achieva patient (red curve) and 4 Panorama patients are shown before and after normalization. In the histograms before normalization, the intensities have been scaled to [0, 255] to remove global intensity scaling differences. After normalization using a single landmark, the histograms seem to have a satisfactory alignment. Using the approximate patch-based sCT prediction method of Chapter 4, we achieved an average MAE_{vox} of 117 HU ($\sigma = 17$ HU) for the four patients scanned on the Achieva scanner. For the patient who had both a Panorama and an Achieva scan, the MAE_{vox} was 122 HU and 133 HU, respectively.

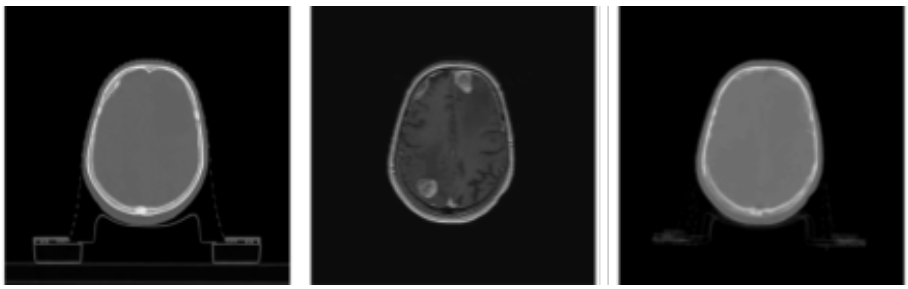


Figure 5.2: Example of Gd-enhanced data. Left: CT scan. Middle: Gd-enhanced T_1-weighted scan. Right: Predicted sCT using the approximate patch-based approach.

5.2.3 Normalization and prediction with different scanners *and* abnormalities

In Figure 5.2, an example of the Gd-enhanced scan is shown along with the predicted sCT and the CT. The Gd-enhanced tumors seem to not visually degrade the sCT quality.

Due to the contrast-enhancement, the histogram of a contrast-enhanced scan differs from a scan without contrast-enhancement. The bright spots in the enhanced tumor regions cause an additional small peak in the high intensity range (see Figure 5.3, left). Similar to the pelvis scans, using the median value as a single landmark in the normalization procedure may therefore not describe the same tissue type in all histograms if the atlases do not contain the contrast-enhancement. Using a simple detection of the largest peak in the histograms seemed to provide a robust landmark contained in all scans (see Figure 5.3, right). This procedure, we believe, could be employed on all T_1-weighted brain scans instead of the median landmark. Using the approximate patch-based approach, the average MAE_{vox} of the sCTs based on the contrast-enhanced scans was 131 HU.

5.2.4 Discussion

In this preliminary study, we tested the approximate patch-based method's robustness to data from a different source as well to data containing previously unseen abnormalities. The results of our normalization tests indicate that as long as the imaging sequence is similar, it is possible to normalize data coming from different scanners to some degree. Using landmarks centered on peaks of

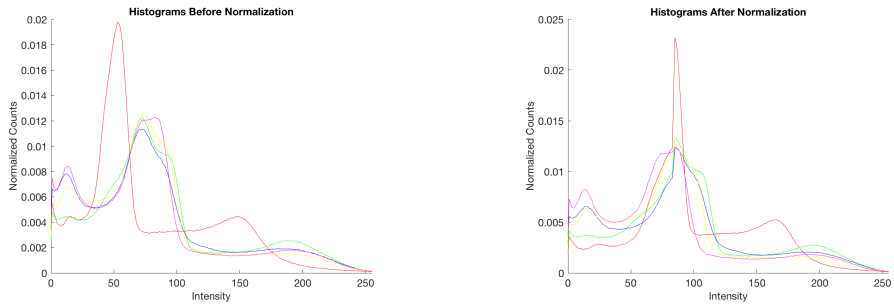


Figure 5.3: Histograms of brain scans from different scanners including contrast. Left panel: Histograms before normalization. The red curve is for a contrast-enhanced scan acquired on the Achieva scanner. Notice the additional peak at an intensity of around 150. The remaining curves are the histograms of the atlas scans acquired on the Panorama. Right panel: Histograms after normalization using a single landmark placed at the most frequently occurring intensity.

the histograms is the most robust as it is not affected by the relative amounts of voxels in different tissue types. Looking at the histograms in Figure 5.1 and Figure 5.3, there are still small differences in their shape. This may be due to natural anatomical variation between patients, which can explain the varying amounts of voxels with a specific intensity. It could also be caused by a sub-optimal normalization. For brain data of five patients scanned on the Panorama, we previously found an average MAE_{vox} of 97 HU using the approximate patch-based algorithm with an atlas of four patients also acquired on the Panorama MR scanner (see Table 4.1). We observe a slightly higher average prediction error of 117 HU for the Achieva scans. For the patient who had both a Panorama scan and an Achieva scan, the Achieva scan also had the highest prediction error.

When adding contrast to the Achieva scans, we saw a further average increase in MAE_{vox} compared to the scans without contrast (131 HU vs. 117 HU). Looking at the histogram normalization in Figure 5.3, the procedure is able to align the largest peaks. However, the shape of the histogram in the high intensity region still differs from the atlas scans (red curve vs. other curves). The contrast enhancement does not only affect tumor regions but also blood vessels and the dura surrounding the brain. This potentially makes it harder to find similar patches in these regions and may explain the higher MAE_{vox}. To minimize motion-related differences, it is favorable to carry out all steps in the MR-only RT chain on one MR scan, i.e., delineation, sCT prediction,

and treatment planning. From the results shown here, one additional MR scan before Gd injection might be necessary for sCT prediction of the brain.

From the current preliminary study of the impact of predicting sCTs based on scans that are different from the atlas scans, we observed that the predictions of the patch-based method seem affected to some degree. However, the relatively small patient group in this study can only serve as an indication of the true performance. An evaluation of the dosimetric impact of the higher MAE_{vox} when predicting from different source scans should also be carried out since this impact may be negligible. It should be noted that the Gd-enhanced scans provide just one example of an abnormality/pathology that affects prediction with the patch-based approach. How other abnormalities affect the method remains to be investigated. We conclude that care should be taken when implementing the method in a clinical setting, where the scans are acquired on a different scanner or with different scan parameters than what is used to acquire the atlases. A calibration procedure could be considered in order to ensure that the normalization and prediction is satisfactory. Furthermore, the impact of abnormalities should be investigated further and QA procedures should be established to ensure that potential outliers in terms of prediction quality are detected.

Chapter 6

Voxel-based sCT generation from conventional MR

This chapter introduces the methodology and summarizes the main findings of paper D. In that paper, we used a voxel-based approach with conventional MR sequences for sCT prediction in the pelvis.

6.1 Introduction

In Chapter 4, we outlined some of the potential issues with an atlas-based method. These included the dependency on accurate registrations and the fact that the main computations are performed during prediction. Although our patch-based approach is an atlas-based method that relies less on accurate registrations and is relatively fast, it does not *remove* these issues entirely. Furthermore, as demonstrated in Chapter 5, the normalization required for the patch search may not always be straightforward and could require individual calibration procedures for different scanners.

A voxel-based parametric model theoretically removes the issues with heavy online computations and registration dependencies. Here, all parameters to predict an sCT are learned in a separate offline training stage. This stage

is usually computationally demanding, but once the parameters are learned, prediction is a matter of applying the parameters to the new data without the need for registrations. This is usually computationally simpler, yielding fast prediction times. As mentioned in Chapter 2, the main challenge with a voxel-based approach is to distinguish between identical compact bone and air voxels in MR. With conventional MR sequences, this must be solved using additional features extracted from the MR. Furthermore, a voxel-based approach will suffer from the same potential issues with cross-scanner MR normalization as the patch-based approach unless special care is taken when choosing the input to the model. Intensity alone, or features that directly rely on the intensity could yield an approach that is only applicable to certain data or certain scanners.

Random forests (RFs) [66] have shown to compare favorably against other prediction models, especially as the dimension of the input data increases [67]. As such, RFs have become a go-to tool in many supervised learning tasks in computer vision. Probably the most well-known application is in Microsoft's Kinect sensor, where RFs are used for real-time human pose estimation [68, 69]. RFs have also shown their versatility for different classification and regression tasks in medical image analysis [70–73]. Furthermore, they have been proposed for predicting one MR contrast from another [74, 75]. This task differs somewhat from CT synthesis because the bone/air ambiguity does not need to be solved; the prediction model only needs to identify and predict the lack of signal in these regions, but is not required to distinguish whether it is caused by bone or air. In paper A, we applied an RF for sCT prediction based on UTE MR scans, achieving competitive results compared to other UTE-based methods [11]. RFs have also been used to segment bone and air in conventional MR images of the brain using PET data, textural and contextual features as input [26].

In paper D, we use RF models to predict sCTs based on conventional T_1-weighted MR scans without using any registrations during CT synthesis. We overcome intensity-ambiguities by using local texture descriptors, spatial features and edge information. Furthermore, inspired by recent works in tumor and organ segmentation [76, 77], we use the concept of auto-context (AC) [78] to iteratively learn and improve context features. We use a combination of classification RFs for learning context features and a regression RF to ultimately predict an sCT. In an attempt to make the method broadly applicable, we use intensity invariant features: the gray-scale intensities are at no stage used directly as input. We apply the model in the pelvic region using the same data as in paper C, i.e., ten prostate patients. In paper D, we compared the method with simple baseline methods since paper C was not yet published. Here, we instead compare the quality measures to the patch-based approach of paper C and discuss the pros and cons of the two methods.

6.2 sCTs using multiple random forests and auto-context features

The RF-based sCT method consists of three main steps:

1. Initial classification.
2. Building auto-context features.
3. Converting features to an sCT estimate.

Below we provide a brief introduction to random forests and describe the steps of the sCT method in detail.

6.2.1 Random forests in short

An RF is an ensemble method consisting of multiple randomized decision trees. By training the decision trees randomly, they individually provide sub-optimal classifications or predictions. An average of the contributions from each decision tree improves the accuracy and generalization error of the RF [66, 79].

A decision tree has an internal structure consisting of nodes. At each internal node, a binary decision function is used to split data to left and right child nodes according to a criterion. The binary decision function takes the following form:

$$h(\boldsymbol{X}_t, \boldsymbol{\beta}_t) \in \{0, 1\} \tag{6.1}$$

where $\boldsymbol{X}_t$ is a matrix of the data reaching node t and $\boldsymbol{\beta}_t$ contains the split parameters at node t. The split is achieved by choosing one or more of the input variables in $\boldsymbol{X}_t$ and then choosing a threshold on the value of those variables (or a linear combination thereof) to decide which data goes left and right. The chosen input variable and threshold is stored in $\boldsymbol{\beta}_t$.

The criterion that decides the threshold often measures the *impurity* of the nodes, which takes on lower values when a node contains many instances of the same class/value and higher values when it contains multiple classes/values in equal amounts. A commonly used impurity measure for classification is the Gini impurity and for regression the variance [80].

The objective is thus to split the data such that the sum of the impurities in the left and right child nodes becomes smaller than the impurity in the parent

node. Using training data, the optimal $\boldsymbol{\beta}_t$s are learned and the tree is grown until all terminal nodes are pure or until a pre-specified stopping criterion is met. At the terminal nodes a prediction is stored based on the training data. For classification, it could be the majority class at the node and for regression it could be the mean value in the node. To predict using the decision tree, test data are propagated down the tree using the learned thresholds. At the terminal nodes, the stored classes/values in each node are assigned to the data that reach those nodes. Randomness can be induced in the decision trees by training them on a random subsets of the data. Furthermore, the node optimization can be randomized by only allowing a random subset of $\boldsymbol{\beta}_t$ to be tested. Below follows a description of our RF-based method.

6.2.2 Step 1: Initial classification

The MR intensity information can be used directly to train a prediction model but this may limit its general applicability. For this reason, we use features that to some degree are independent of linear gray-scale changes.

Local binary patterns We use local binary pattern-like (LBP) features [81] to capture textural information. For each voxel, we do a relational comparison of the mean value of a cuboidal region centered on that voxel with mean values of 26 cuboidal regions positioned on a sphere around that voxel. This results in a 26 digit binary number which is converted to a decimal texture feature (dubbed LBP_μ). We also incorporate a variant of the LBP where the standard deviation of the regions is compared instead (dubbed LBP_σ). We use three LBP_μ and two LBP_σ features with differing region size (RS), 3^3 - 5^3 voxels, and sphere radius (SR), 5 - 13 voxels, to capture information at different scales.

Edge detector As a simple edge detector, we use the standard deviation of a $3\times3\times3$ neighborhood around each voxel, dubbed σ. It should be noted that the standard deviation is invariant to intensity offsets but not to intensity scaling.

Spatial information As spatial features, we use the (x, y) position of each voxel in a patient-dependent coordinate system. The origin is situated in the center-of-mass of the patient and the coordinate axes are scaled to unity at the surface of the patient. This spatial feature takes into account the varying sizes of patients and requires no registration to a common space. Figure 6.1 shows examples of the initial features.

We train a classification RF to segment an MR scan into air, fat, soft tissue and bone classes based on the features described above. The training target segmentations are generated by thresholding the CT scans using the following criteria: air $= [-1000, -200]$ HU, fat $=]-200, 0[$ HU, soft tissue $= [0, 150]$, bone $=]150, \infty[$ HU. For each training patient, we then sample our features at 6×10^5 and 10^5 random positions within the body outline and within the bone volume, respectively. This defines our training set for the classification RF.

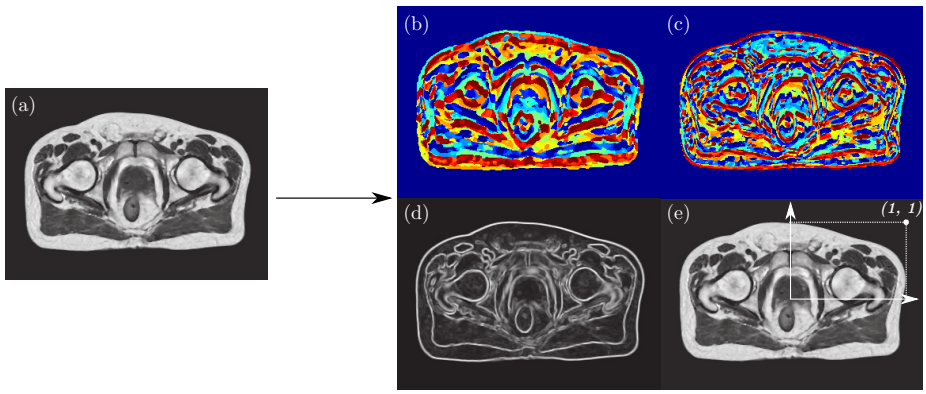


Figure 6.1: Features are extracted from the T_1-weighted MR shown in (a). (b) and (c) shows examples of the LBP_μ ($\text{SR/RS} = 11/5^3$ voxels) and LBP_σ ($\text{SR/RS} = 7/5^3$ voxels), respectively. (d) shows the standard deviation feature. (e) shows the coordinate system used to derive the spatial features; the origin is positioned in the center of mass of the patient and the axes are scaled relative to the outer contour of the patient. Figure from [82].

6.2.3 Step 2: Building auto-context features

We apply the trained initial RF to create a variant of auto-context (AC) features [78] for all training patients, which are then used in combination with the initial features to train an improved classification RF. This process is repeated three times, each time using the most recently trained RF to improve the AC features. The AC features are calculated based on the three probability maps of a voxel belonging to the fat, soft tissue or bone classes. For each voxel, the feature extracts the mean probability of belonging to each class in a small region centered on that voxel, along with the same mean probabilities of six regions chosen at a fixed distance from it in the posterior/anterior, superior/inferior and lateral directions. The AC feature extraction is illustrated in Figure 6.2. In this way, we encode the approximate locations and relations of the tissue classes.

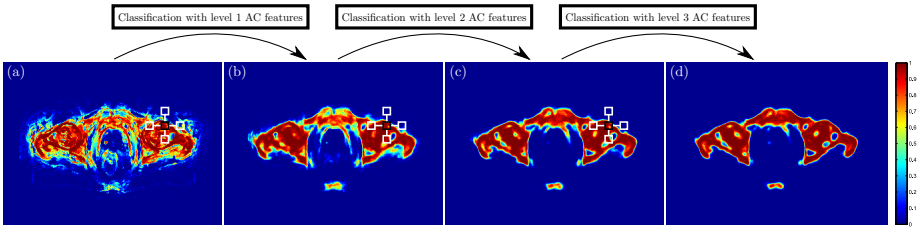


Figure 6.2: Examples of the output probability maps of the bone class for a test patient after sending the data through (a) the first classification RF and (b)-(d) the subsequent classification RFs with improving AC features. The AC feature for a voxel consists of the mean probability of a small region (shown in black) around the voxel along with the mean probability of 6 small regions at a fixed distance in the posterior/anterior, superior/inferior and lateral directions (shown in white for the posterior/anterior and lateral directions). Figure from [82].

6.2.4 Step 3: Converting features to an sCT estimate

As the final step, we train a regression RF based on the initial features and the last level of AC features. We use the non-thresholded CT scans as target values and sample our features from each training patient at 1.3×10^6 and 10^5 random positions within the body outline and bone volume (defined as voxels with HU > 150), respectively.

6.3 Experiments and results

We carry out all model training in a leave-one-out manner using nine patients as training data and applying the model on the tenth. The prediction time for one sCT is approximately 10 minutes. This should be compared to the approximately 20 minutes (excluding registrations) used by the approximate patch-based method on the same data.

In Figure 6.3, transverse slices are shown for the RF-based method and the patch-based method of paper C. As shown in Table 6.1, the RF-based approach is on par with the patch-based approach although the $\mathrm{ME}_{\mathrm{vox}}$ shows a greater tendency for overestimation of the true HU values. Qualitatively, the RF-based sCT is blurrier than the patch-based as seen in Figure 6.3. This can probably be explained by the type of features used, which are mainly derived from small

	RF-based	Patch-based	p-value
MAE_{vox} [HU]	58 ($\sigma = 9$)	54($\sigma = 8$)	0.05
ME_{vox} [HU]	-10 ($\sigma = 10$)	-1 ($\sigma = 7$)	0.001
MAE_{WEPL} [mm]	1.3 ($\sigma = 0.4$)	1.2 ($\sigma = 0.3$)	0.11
ME_{WEPL} [mm]	-0.7 ($\sigma = 0.6$)	-0.3 ($\sigma = 0.7$)	0.01

Table 6.1: The voxel-wise and radiologic errors. Average value and standard deviation (σ) for 10 patients are shown for the two methods along with the p-value from a paired t-test on the difference between the two methods.

box regions. Using more features at the voxel-level could potentially provide a more crisp sCT. The RF-based sCTs seem to have fewer "hot-spots" of underestimation in the compact bone region than the patch-based method, see Figure 6.3 (e) and (f).

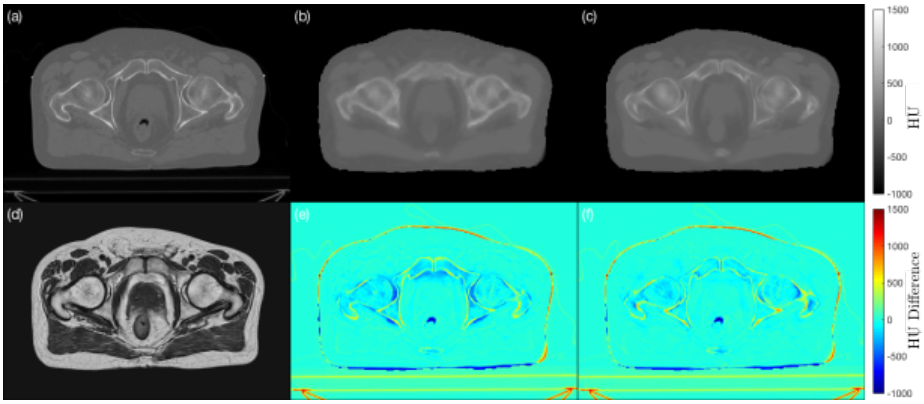


Figure 6.3: Transverse slices of the CT and sCTs for one patient. (a) and (d) show the CT and corresponding T_1-weighted MR. (b) and (c) show the sCTs generated with the RF-based method and the patch-based method, respectively. (e) and (f) show the difference in HU between the CT and the sCTs in (b) and (c), respectively. No colorbar is shown for the MR image.

6.3.1 Discussion

Given the fact that our RF-based method works on conventional MR without any registrations at run-time makes the approach an appealing alternative to atlas-based approaches such as our patch-based method. One of the arguments for using a voxel-based model is the theoretically fast prediction times. Since each tree in an RF is completely independent of the others, RFs have especially demonstrated their speed in fully parallelized and/or GPU-based versions [69, 73]. The resulting prediction time of approximately 10 minutes for an sCT is therefore not impressive, although it is faster than the patch-based approach described in paper C. The CPU-based RF implementation that we used here [83] was not parallelized, which is otherwise key to achieving fast prediction times. The average prediction time of one tree was approximately 1 second. By parallelizing the RF, substantial speed-ups could likely be achieved.

By specifically avoiding intensity-based features, the method could potentially be broadly applicable on data from different scanners. However, this remains to be investigated. In paper B and C, a histogram matching approach was needed to normalize the intensities for the patch search. In this step, assumptions about the intensity frequencies in different tissues must hold. As we saw in Chapter 5, the normalization procedure and resulting sCTs seem affected with data from different scanners. Theoretically, histogram matching should not be necessary with the proposed RF-based approach, although we did apply it in paper D as well. The only reason for this was that it was part of the pre-processing pipeline that we also used in paper C. After realizing this, we ran the algorithm on the same scans without histogram matching. This resulted in an average MAE_{vox} of 63 HU, which is 5 HU more than the average MAE_{vox} when using histogram matching but still almost within the standard deviation. Whether this has a significant dosimetric impact remains to be investigated. We found that a global intensity-scaling of the scans to a common range of $[0, 255]$ was still needed, since the scaling of the intensities affects the local standard deviation features. As future work, we think it would be beneficial to remove/replace these features.

We note that with the rather limited amount of data available to train our models, there is a risk of over-fitting to the data from our scanner. Testing the method's robustness using data from various scanners as well as training models based on more data remains part of our future work. Furthermore, the dosimetric accuracy of the RF-based sCTs remains to be investigated to determine if the quality is indeed as close to the patch-based as this initial investigation suggests.

An interesting property of RFs is the ability to get estimates of the feature importance during training. This can be estimated by calculating the sum of

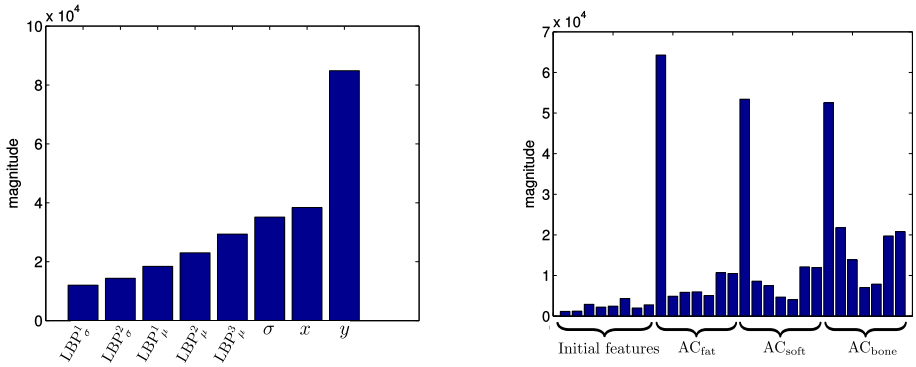


Figure 6.4: Mean decrease in Gini impurity caused by the different variables. Left: The results for the initial classification forest. The superscript in the LBP features increases as the scale of the features increases. Right: The results for the final level of auto-context training. The first feature in each of the AC groups is the mean probability of belonging to the subscripted class, computed in a box centered on each voxel. The remaining six features in each group are the mean probabilities of six neighboring regions.

the decrease in impurity that a given feature causes across all trees and nodes in the RF. A feature that is decreasing impurity more than others is considered more important. This information can be used for feature selection and model reduction. In Figure 6.4, the feature importance is shown as the mean decrease in Gini impurity for the initial classification RF and the final AC classification RF. In the initial classification, the y position is causing the largest decrease in impurity, whereas the LBP_σ features are less important. These could potentially be removed. In the final stage of classification using the AC features, the central AC features of each class dominate, whereas the neighboring AC features are less important. At this stage, the classification from the previous stage is thus the most informative for deciding the splits. In Figure 6.5, the feature importance is shown for the final regression RF. The same pattern is observed as for the classification RF in Figure 6.4. However, the neighboring AC_{bone} features are almost as important as the central AC_{soft}. From the results shown here, it could be interesting to further investigate the role of the current features as well as new features in order to find an optimal model with no redundancy in the inputs.

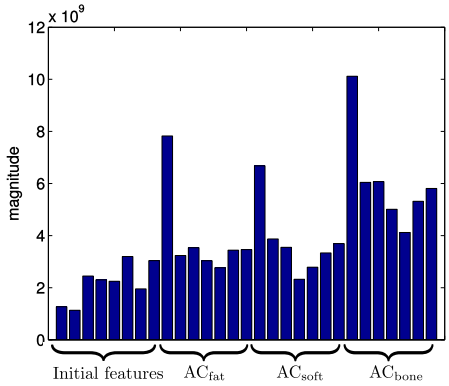


Figure 6.5: Mean decrease in mean squared error caused by the different variables used in training the final regression RF.

Chapter 7

Conclusion and future directions

During the relatively short course of these PhD studies, MR-only radiotherapy and specifically CT synthesis methods have developed rapidly and have gone from being a small niche field to its first commercial implementations [34, 84]. What seems to be missing at the moment is a thorough large-population study, which could establish proper confidence intervals on the dosimetric implications of using MR-only for RT. As an initial step, this could be done parallel to the standard MR/CT work-flow. To my knowledge, such a study is currently beginning in Sweden with the ambitious "Gentle Radiotherapy" initiative [85]. Otherwise, most other studies so far have introduced a novel sCT approach and then performed evaluations on limited data. In that sense, this thesis falls in line with the rest of the field. We have proposed and investigated several different methods for CT synthesis using limited available data. However, we have attempted to always include a benchmark for comparing new and existing methods such that strengths and weaknesses could be illustrated on the basis of the same data.

In an attempt to identify the most promising method, our initial efforts in paper A were aimed at voxel-based approaches in the brain using specialized UTE MR sequences to overcome bone/air ambiguities. We found regression-based methods to be promising candidates, but also realized some of the shortcomings of using specialized MR. These included the added scan time, which was

uncomfortable for the patients. Furthermore, the limited quality of the UTE scans seemed to also put an upper limit on the possible quality of the predicted sCTs. We therefore decided to investigate sCT methods based on conventional MR sequences. These approaches were dominated by atlas-based methods that relied on non-linear registrations to overcome intensity ambiguities. Inspired by work in brain MR segmentation, we proposed a patch-based method, which only required affine registrations of the atlases and we showed a competitive performance compared to voxel-based and more traditional atlas-based methods. The competitive performance did, however, come at the cost of computation time, which yielded a method that could not be used for same-day planning and treatment of patients.

Our next effort was aimed at improving the speed of the algorithm while investigating a different and more challenging treatment site. We applied an approximate patch search algorithm to predict sCTs in the pelvic region with a substantial speed-up compared to our previous implementation. The sCT quality measures were on par with other reported atlas-based methods.

Interestingly, this thesis ended where it started: with a voxel-based approach. However, considerable improvements were incorporated such that the method now worked with conventional MR sequences with broad applicability in mind. The initial results suggested that the method was close in performance to our patch-based approach but theoretically without having many of the potential shortcomings of an atlas-based approach. An investigation of the method on more data is still needed, however. Furthermore, the method could potentially be improved by investigating and improving the input features.

Even though a large-scale feasibility study of MR-only RT is still missing, it should be noted that all small-scale studies published so far show only small dosimetric deviations ($< 2\%$) between using an sCT or a CT [34,35,52,63,65,86]. Many consider these deviations clinically unimportant compared to the possible systematic errors introduced in the MR/CT work-flow [34,65,87]. The promising potential of using synthetic CTs for MR-only RT has therefore been shown multiple times. What is needed next, is to confirm this potential on a larger scale and to develop quality assurance procedures of the sCTs [88].

Bibliography

[1] R Prabhakar, P K Julka, T Ganesh, A Munshi, R C Joshi, and G K Rath. Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. *Jpn. J. Clin. Oncol.*, 37(6):405–11, jun 2007.

[2] Merina Ahmed, Maria Schmidt, Aslam Sohaib, Christine Kong, Kevin Burke, Cheryl Richardson, Marianne Usher, Sinead Brennan, Angela Riddell, Mark Davies, Kate Newbold, Kevin J Harrington, and Christopher M Nutting. The value of magnetic resonance imaging in target volume delineation of base of tongue tumours–a study using flexible surface coils. *Radiother. Oncol.*, 94(2):161–7, feb 2010.

[3] Coen Rasch, Roel Steenbakkers, and Marcel van Herk. Target Definition in Prostate, Head, and Neck. *Semin. Radiat. Oncol.*, 15(3):136–145, jul 2005.

[4] Uulke A. van der Heide, Antonetta C. Houweling, Greetje Groenendaal, Regina G H Beets-Tan, and Philippe Lambin. Functional MRI for radiotherapy dose painting. *Magn. Reson. Imaging*, 30(9):1216–1223, 2012.

[5] Jan J W Lagendijk, Bas W. Raaymakers, and Marco van Vulpen. The Magnetic Resonance Imaging-Linac System. *Semin. Radiat. Oncol.*, 24(3):207–209, 2014.

[6] Giovanni Mauro Cattaneo, Michele Reni, Giovanna Rizzo, Pietro Castellone, Giovanni Luca Ceresoli, Cesare Cozzarini, A. J M Ferreri, Paolo Passoni, and Riccardo Calandrino. Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. *Radiother. Oncol.*, 75(2):217–223, 2005.

[7] K Ulin, MM Urie, and JM Cherlow. Results of a Multi-Institutional Benchmark Test for Cranial CT/ MR Image Registration. *Int. J. Radiat. Oncol.*, 77(5):1584–1589, 2010.

[8] Tufve Nyholm, Morgan Nyberg, Magnus G Karlsson, and Mikael Karlsson. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiat. Oncol.*, 4:54, jan 2009.

[9] Ines L H Reichert, Matthew D Robson, Peter D Gatehouse, Taigang He, Karyn E Chappell, Joanne Holmes, Samia Girgis, and Graeme M Bydder. Magnetic resonance imaging of cortical bone with ultrashort TE pulse sequences. *Magn. Reson. Imaging*, 23(5):611–8, jun 2005.

[10] Vincent Keereman, Yves Fierens, Tom Broux, Yves De Deene, Max Lonneux, and Stefaan Vandenberghe. MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences. *J. Nucl. Med.*, 51(5):812–8, may 2010.

[11] Jens M Edmund, Hans M Kjer, Koen Van Leemput, Rasmus H Hansen, Jon Al Andersen, and Daniel Andreasen. A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times. *Phys. Med. Biol.*, 59(23):7501–19, dec 2014.

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

[13] Shinichiro Mori, George T Y Chen, and Masahiro Endo. Effects of intrafractional motion on water equivalent pathlength in respiratory-gated heavy charged particle beam radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 69(1):308–17, sep 2007.

[14] Christopher M Rank, Christoph Tremmel, Nora Hünemohr, Armin M Nagel, Oliver Jäkel, and Steffen Greilich. MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. *Radiat. Oncol.*, 8:51, jan 2013.

[15] Daniel A. Low, William B. Harms, Sasa Mutic, and James A. Purdy. A technique for the quantitative evaluation of dose distributions. *Med. Phys.*, 25(5):656, may 1998.

[16] Adam Johansson, Mikael Karlsson, and Tufve Nyholm. CT substitute derived from MRI sequences with ultrashort echo time. *Med. Phys.*, 38(5):2708, 2011.

[17] Yannick Berker, Jochen Franke, André Salomon, Moritz Palmowski, Henk C W Donker, Yavuz Temur, Felix M Mottaghy, Christiane Kuhl, David

Izquierdo-Garcia, Zahi a Fayad, Fabian Kiessling, and Volkmar Schulz. MRI-based attenuation correction for hybrid PET/MRI systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence. *J. Nucl. Med.*, 53(5):796–804, may 2012.

[18] Christopher M Rank, Nora Hünemohr, Armin M Nagel, Matthias C Röthke, Oliver Jäkel, and Steffen Greilich. MRI-based simulation of treatment plans for ion radiotherapy in the brain region. *Radiother. Oncol.*, 109(3):414–8, dec 2013.

[19] Jürgen Rahmer, Peter Börnert, Jan Groen, and Clemens Bos. Three-dimensional radial ultrashort echo-time imaging with T2 adapted sampling. *Magn. Reson. Med.*, 55(5):1075–82, may 2006.

[20] Daniel Andreasen. *Creating a pseudo-CT from MRI for MRI-only based Radiation Therapy Planning.* Master's thesis, Technical University of Denmark, 2013.

[21] Ciprian Catana, Andre van der Kouwe, Thomas Benner, Christian J Michel, Michael Hamm, Matthias Fenchel, Bruce Fischl, Bruce Rosen, Matthias Schmand, and a Gregory Sorensen. Toward implementing an MRI-based PET attenuation-correction method for neurologic studies on the MR-PET brain prototype. *J. Nucl. Med.*, 51(9):1431–8, sep 2010.

[22] Adam Johansson, Anders Garpebring, Mikael Karlsson, Thomas Asklund, and Tufve Nyholm. Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information–potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. *Acta Oncol.*, 52(7):1369–73, oct 2013.

[23] Mika Kapanen and Mikko Tenhunen. T1/T2*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncol.*, 52(3):612–8, apr 2013.

[24] Juha Korhonen, Mika Kapanen, Jani Keyriläinen, Tiina Seppälä, and Mikko Tenhunen. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med. Phys.*, 41(1):011704, jan 2014.

[25] G. Wagenknecht, E.R. Kops, L. Tellmann, and H. Herzog. Knowledge-based segmentation of attenuation-relevant regions of the head in T1-weighted MR images for attenuation correction in MR/PET systems. *2009 IEEE Nucl. Sci. Symp. Conf. Rec.*, pages 3338–3343, oct 2009.

[26] Sze-Liang Stanley Chan, Rosalind L. Jeffree, Michael Fay, Stuart Crozier, Zhengyi Yang, Yaniv Gal, and Paul Thomas. Automated Classification of

Bone and Air Volumes for Hybrid PET-MRI Brain Imaging. In *2013 Int. Conf. Digit. Image Comput. Tech. Appl.*, pages 1–8. IEEE, nov 2013.

[27] Jason A Dowling, Jonathan Lambert, Joel Parker, Olivier Salvado, Jurgen Fripp, Anne Capp, Chris Wratten, James W Denham, and Peter B Greer. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 83(1):e5–11, may 2012.

[28] Jason a. Dowling, Jidi Sun, Peter Pichler, David Rivest-Hénault, Soumya Ghose, Haylea Richardson, Chris Wratten, Jarad Martin, Jameen Arm, Leah Best, Shekhar S. Chandra, Jurgen Fripp, Frederick W. Menk, and Peter B. Greer. Automatic substitute CT generation and contouring for MRI-alone external beam radiation therapy from standard MRI sequences. *Int. J. Radiat. Oncol.*, 2015.

[29] J Sjölund, D Forsberg, M Andersson, and H Knutsson. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys. Med. Biol.*, 60(2):825–839, jan 2015.

[30] Ninon Burgos, Manuel Jorge Cardoso, Marc Modat, Stefano Pedemonte, John Dickson, Anna Barnes, John S Duncan, David Atkinson, Simon R Arridge, Brian F Hutton, and Sebastien Ourselin. Attenuation Correction Synthesis for Hybrid PET-MR Scanners. In *Med. image Comput. Comput. Interv. MICCAI 2013*, volume 16, pages 147–154. Springer Berlin Heidelberg, jan 2013.

[31] Ninon Burgos, M. Jorge Cardoso, Kris Thielemans, Marc Modat, John Dickson, Jonathan M. Schott, David Atkinson, Simon R. Arridge, Brian F. Hutton, and Sébastien Ourselin. Multi-contrast attenuation map synthesis for PET/MR scanners: assessment on FDG and Florbetapir PET tracers. *Eur. J. Nucl. Med. Mol. Imaging*, 42(9):1447–1458, aug 2015.

[32] Ninon Burgos, M Jorge Cardoso, Filipa Guerreiro, Catarina Veiga, Marc Modat, Jamie McClelland, Antje-christin Knopf, Shonit Punwani, David Atkinson, Simon R Arridge, Brian F Hutton, and Sébastien Ourselin. Robust CT Synthesis for Radiotherapy Planning: Application to the Head and Neck Region. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Med. Image Comput. Comput. Interv. – MICCAI 2015*, volume 9350 of *Lecture Notes in Computer Science*, pages 476–484. Springer International Publishing, Cham, 2015.

[33] Matthias Hofmann, Florian Steinke, Verena Scheel, Guillaume Charpiat, Jason Farquhar, Philip Aschoff, Michael Brady, Bernhard Schölkopf, and Bernd J Pichler. MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration. *J. Nucl. Med.*, 49(11):1875–83, nov 2008.

[34] Carl Siversson, Fredrik Nordström, Terese Nilsson, Tufve Nyholm, Joakim Jonsson, Adalsteinn Gunnlaugsson, and Lars E. Olsson. Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm. *Med. Phys.*, 42(10):6090–6097, 2015.

[35] Jinsoo Uh, Thomas E Merchant, Yimei Li, Xingyu Li, and Chiaho Hua. MRI-based treatment planning with pseudo CT generated through atlas registration. *Med. Phys.*, 41(5):051711, may 2014.

[36] Madhu Sudhan Reddy Gudur, Wendy Hara, Quynh-Thu Le, Lei Wang, Lei Xing, and Ruijiang Li. A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning. *Phys. Med. Biol.*, 59(21):6595–6606, 2014.

[37] Snehashis Roy. PET Attenuation Correction using Synthetic CT from Ultrashort Echo-time MRI. *J Nucl Med*, 55(12):2071–2077, 2014.

[38] Jens Sjolund, Andreas Eriksson Jarlideni, Mats Andersson, Hans Knutsson, and Hakan Nordstrom. Skull Segmentation in MRI by a Support Vector Machine Combining Local and Global Features. *2014 22nd Int. Conf. Pattern Recognit.*, pages 3274–3279, 2014.

[39] Brian Holch Kristensen, Finn Jørgen Laursen, Vibeke Løgager, Poul Flemming Geertsen, and Anders Krarup-Hansen. Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours. *Radiother. Oncol.*, 87(1):100–9, apr 2008.

[40] Joakim H Jonsson, Magnus G Karlsson, Mikael Karlsson, and Tufve Nyholm. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat. Oncol.*, 5(1):62, jan 2010.

[41] D Pasquier, N Betrouni, M Vermandel, T Lacornerie, E Lartigau, and J Rousseau. MRI alone simulation for conformal radiation therapy of prostate cancer: technical aspects. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, 1:160–163, 2006.

[42] Antoni Buades and Bartomeu Coll. A Non-Local Algorithm for Image Denoising. In *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, volume 2, pages 60–65. IEEE, 2005.

[43] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.*, 13(9):1200–1212, 2004.

[44] Alexei Efros and Wt Freeman. Image Quilting for Texture Synthesis and Transfer. *Proc. 28th Annu. Conf. Comput. Graph. Interact. Tech.*, pages 1–6, 2001.

[45] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D Louis Collins. Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage*, 54(2):940–54, jan 2011.

[46] F Rousseau. A supervised patch-based approach for human brain labeling. *Med. Imaging, IEEE*, 30(10):1852–1862, 2011.

[47] Tong Tong, Robin Wolz, Pierrick Coupé, Joseph V Hajnal, and Daniel Rueckert. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *Neuroimage*, 76:11–23, aug 2013.

[48] Daoqiang Zhang, Qimiao Guo, Guorong Wu, and Dinggang Shen. Sparse Patch-Based Label Fusion for Multi-Atlas Segmentation. In *Multimodal Brain Image Anal.*, number Mv, pages 94–102. Springer Berlin Heidelberg, 2012.

[49] Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters, and Daniel Rueckert. A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images. *IEEE Trans. Med. Imaging*, 32(7):1302–15, jul 2013.

[50] LG Nyul and JK Udupa. Standardizing the MR image intensity scales : making MR intensities have tissue specific meaning. *Med. Imaging*, 3976:496–504, 2000.

[51] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.*, 13(4):600–612, apr 2004.

[52] Daniel Andreasen, Koen Van Leemput, Rasmus H. Hansen, Jon A. L. Andersen, and Jens M. Edmund. Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Med. Phys.*, 42:1596–1605, 2015.

[53] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-CristopheC. Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, John Buatti, StephenR Aylward, JamesV Miller, Steve Pieper, and Ron Kikinis. 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, 11 2012.

[54] Marie E Korsholm, Line W Waring, and Jens M Edmund. A criterion for the reliable use of MRI-only radiotherapy. *Radiat. Oncol.*, 9(1):16, jan 2014.

[55] Marius Muja and David G Lowe. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *Int. Conf. Comput. Vis. Theory Appl. (VISAPP '09)*, pages 1–10, 2009.

[56] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM Trans. Graph.*, 28(3):1, jul 2009.

[57] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, 3(3):209–226, September 1977.

[58] Wenzhe Shi, Jose Caballero, Christian Ledig, Xiahai Zhuang, Wenjia Bai, Kanwal Bhatia, Antonio M Simoes Monteiro de Marvao, Tim Dawes, Declan O'Regan, and Daniel Rueckert. Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch. In *Med. Image Comput. Comput. Interv. – MICCAI 2013*, volume 16, pages 9–16. Springer Berlin Heidelberg, jan 2013.

[59] Wenzhe Shi, Herve Lombaert, Wenjia Bai, Christian Ledig, Xiahai Zhuang, Antonio Marvao, Timothy Dawes, Declan O'Regan, and Daniel Rueckert. Multi-atlas Spectral PatchMatch: Application to Cardiac Image Segmentation. In *Med. Image Comput. Comput. Interv. – MICCAI 2014*, pages 348–355. Springer International Publishing, 2014.

[60] Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis, José V. Manjón, D. Louis Collins, and Pierrick Coupé. An Optimized PatchMatch for multi-scale and multi-feature label fusion. *Neuroimage*, 124:770–782, jan 2016.

[61] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien P W Pluim. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging*, 29(1):196–205, jan 2010.

[62] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[63] Daniel Andreasen, Koen Van Leemput, and Jens M. Edmund. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Med. Phys.*, 43(8):4742–4752, 2016.

[64] Adam Johansson, Mikael Karlsson, Jun Yu, Thomas Asklund, and Tufve Nyholm. Voxel-wise uncertainty in CT substitute derived from MRI. *Med. Phys.*, 39(6):3283–90, jun 2012.

[65] Joakim H Jonsson, Adam Johansson, Karin Söderström, Thomas Asklund, and Tufve Nyholm. Treatment planning of intracranial targets on MRI derived substitute CT data. *Radiother. Oncol.*, 108(1):118–22, jul 2013.

[66] L Breiman. Random forests. *Mach. Learn.*, pages 5–32, 2001.

[67] Rich Caruana, Nikos Karampatziakis, and Ainur Yessenalina. An empirical evaluation of supervised learning in high dimensions. In *Proc. 25th Int. Conf. Mach. Learn. - ICML '08*, pages 96–103, New York, New York, USA, 2008. ACM Press.

[68] Ross Girshick, Jamie Shotton, Pushmeet Kohli, Antonio Criminisi, and Andrew Fitzgibbon. Efficient regression of general-activity human poses from depth images. In *ICCV*. IEEE, October 2011.

[69] Jamie Shotton, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore, Pushmeet Kohli, Antonio Criminisi, Alex Kipman, and Andrew Blake. *Efficient Human Pose Estimation from Single Depth Images*. IEEE, January 2012.

[70] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lanczi, Elizabeth Gerstner, Marc-andre Weber, Tal Arbel, Brian B Avants, Nicholas Ayache, Patricia Buendia, D Louis Collins, Nicolas Cordier, Jason J Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Cagatay Demiralp, Christopher R Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M Iftekharuddin, Raj Jena, Nigel M John, Ender Konukoglu, Danial Lashkari, Jose Antonio Mariz, Raphael Meier, Sergio Pereira, Doina Precup, Stephen J Price, Tammy Riklin Raviv, Syed M S Reza, Michael Ryan, Duygu Sarikaya, Lawrence Schwartz, Hoo-chang Shin, Jamie Shotton, Carlos a Silva, Nuno Sousa, Nagesh K Subbanna, Gabor Szekely, Thomas J Taylor, Owen M Thomas, Nicholas J Tustison, Gozde Unal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging*, 34(10):1993–2024, oct 2015.

[71] Ezequiel Geremia, Olivier Clatz, Bjoern H Menze, Ender Konukoglu, Antonio Criminisi, and Nicholas Ayache. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *Neuroimage*, 57(2):378–390, jul 2011.

[72] Olivier Pauly, Ben Glocker, Antonio Criminisi, Diana Mateus, Axel Martinez Möller, Stephan Nekolla, and Nassir Navab. Fast Multiple Organ

Detection and Localization in Whole-Body MR Dixon Sequences. In *Med. image Comput. Comput. Interv. MICCAI 2011*, volume 14, pages 239–247. Springer Berlin Heidelberg, jan 2011.

[73] Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu. Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In *Med. Comput. Vis. 2010. Recognit. Tech. Appl. Med. Imaging*, pages 106–117. Springer Berlin Heidelberg, 2011.

[74] Min Chen, Amog Jog, Aaron Carass, and Jerry L. Prince. Using image synthesis for multi-channel registration of different image modalities. In Sébastien Ourselin and Martin A. Styner, editors, *Proc SPIE Int Soc Opt Eng.*, volume 136, page 94131Q, mar 2015.

[75] Amod Jog, Aaron Carass, Dzung L Pham, and Jerry L Prince. Random forest FLAIR reconstruction from T1, T2, and PD-weighted MRI. In *2014 IEEE 11th Int. Symp. Biomed. Imaging*, pages 1079–1082. IEEE, apr 2014.

[76] Chunjun Qian, Li Wang, Ambereen Yousuf, Aytekin Oto, and Dinggang Shen. *In Vivo MRI Based Prostate Cancer Identification with Random Forests and Auto-context Model*, pages 314–322. Springer International Publishing, Cham, 2014.

[77] Vasileios Zografos, Alexander Valentinitsch, Markus Rempfler, Federico Tombari, and Bjoern Menze. Hierarchical Multi-Organ Segmentation Without Registration in 3D Abdominal CT Images. In *Med. Comput. Vis. Algorithms Big Data Int. Work. MCV 2015, Held Conjunction with MICCAI 2015, Munich, Ger. Oct. 9, 2015, Revis. Sel. Pap.*, pages 37–46. Springer International Publishing, 2016.

[78] Zhuowen Tu and Xiang Bai. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(10):1744–1757, 2010.

[79] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282 vol.1, Aug 1995.

[80] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classificcation and Regression Trees*. Wadsworth, 1984.

[81] T Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, jul 2002.

[82] Daniel Andreasen, Jens M. Edmund, Vasileios Zografos, Bjoern H. Menze, and Koen Van Leemput. Computed tomography synthesis from magnetic

resonance images in the pelvis using multiple random forests and auto-context features. In Martin A. Styner and Elsa D. Angelini, editors, *Proc. SPIE Med. Imaging*, volume 9784, page 978417. International Society for Optics and Photonics, mar 2016.

[83] Abhishek Jaiantilal. Classification and regression by randomforest-matlab. Available at `http://code.google.com/p/randomforest-matlab`, 2009.

[84] M. Köhler, T. Vaara, M. Van Grootel, R. Hoogeveen, R. Kemppainen, and S. Renisch. Philips mrcat for prostate dose calculations using only mri data. *Philips white paper*, 2015.

[85] Gentle radiotherapy project webpage. `http://www.gentleradiotherapy.se`. Accessed: 2016-08-26.

[86] Jason A. Dowling, Jidi Sun, Peter Pichler, David Rivest-H??nault, Soumya Ghose, Haylea Richardson, Chris Wratten, Jarad Martin, Jameen Arm, Leah Best, Shekhar S. Chandra, Jurgen Fripp, Frederick W. Menk, and Peter B. Greer. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. *Int. J. Radiat. Oncol. Biol. Phys.*, 93(5):1144–1153, 2015.

[87] Joakim H Jonsson, Mohammad M Akhtari, Magnus G Karlsson, Adam Johansson, Thomas Asklund, and Tufve Nyholm. Accuracy of inverse treatment planning on substitute CT images derived from MR data for brain lesions. *Radiat. Oncol.*, 10(1):13, 2015.

[88] Jens M. Edmund, Daniel Andreasen, Faisal Mahmood, and Koen Van Leemput. Cone beam computed tomography guided treatment delivery and planning verification for magnetic resonance imaging only radiotherapy of the brain. *Acta Oncol. (Madr).*, 54(9):1496–1500, 2015.

Appendices

Paper A

A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times.

Jens M Edmund [(1,*)]; Hans M Kjer [(2)]; Koen Van Leemput[(2,3)]; Rasmus H Hansen [(4)]; Jon AL Andersen [(1)] and Daniel Andreasen [(1,2)]

(1) Department of Oncology, Radiotherapy Research Unit, Copenhagen University Hospital, Herlev, Denmark.
(2) Department of applied mathematics and computer science, Technical University of Denmark, Lyngby, Denmark.
(3) Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.
(4) Department of Radiology, Copenhagen University Hospital, Herlev, Denmark.

* Corresponding author, email: jens.edmund@regionh.dk, phone: +45 3868 3170.

Abstract

Introduction

Radiotherapy (RT) based on magnetic resonance imaging (MRI) as the only modality, so-called MRI-only RT, would remove the systematic registration error between MR and computed tomography (CT), and provide co-registered MRI for assessment of treatment response and adaptive RT. Electron densities, however, need to be assigned to the MRI images for dose calculation and patient setup based on digitally reconstructed radiographs (DRRs). Here, we investigate the geometric and dosimetric performance for a number of popular voxel-based methods to generate a so-called pseudo CT (pCT).

Material and methods

Five patients receiving cranial irradiation, each containing a co-registered MRI and CT scan, were included. An ultra short echo time MRI sequence for bone visualization was used. Six methods were investigated for three popular types of voxel-based approaches; 1) Threshold-based segmentation, 2) Bayesian segmentation and 3) statistical regression. Each approach contained two methods. Approach 1 used bulk density assignment of MRI voxels into air, soft tissue and bone based on logical masks and the transverse relaxation time T_2 of the bone. Approach 2 used similar bulk density assignments with Bayesian statistics including or excluding additional spatial information. Approach 3 used a statistical regression correlating MRI voxels with their corresponding CT voxels. A similar photon and proton treatment plan was generated for a target positioned between the nasal cavity and the brainstem for all patients. The CT agreement with the pCT of each method was quantified and compared with the other methods geometrically and dosimetrically using both a number of reported metrics and introducing some novel metrics.

Results

The best geometrical agreement with CT was obtained with the statistical regression methods which performed significantly better than the threshold and Bayesian segmentation methods (excluding spatial information). All methods agreed significantly better with CT than a reference water MRI comparison. The mean dosimetric deviation for photons and protons compared to the CT was about 2 % and highest in the

gradient dose region of the brainstem. Both the threshold based method and the statistical regression methods showed the highest dosimetrical agreement.

Conclusions

Generation of pCTs using statistical regression seems to be the most promising candidate for MRI-only RT of the brain. Further, the total amount of different tissues needs to be taken into account for dosimetric considerations regardless of their correct geometrical position.

Introduction

Magnetic resonance imaging (MRI) has proven advantageous as compared to computed tomography (CT) in the delineation precision of target volume and organs at risks (OARs) in the brain and for other body sites [1-3]. In addition, matched MR and CT images reduce interobserver delineation variability [4]. Therefore, CT is now routinely combined with MRI in the planning of radiotherapy (RT) in many centers. This procedure, however, introduces a systematic error arising from the registration between the two modalities. A mean registration error of about 2 mm has been reported for RT treatment in the cranial region using a stereotactic head frame on the CT [5]. As cranial stereotactic radiation surgery (SRS) is increasingly performed on conventional Linacs in frameless head-and-neck fixation devices [6, 7], registration errors are comparable to those of head and neck. These have been reported to be around 2 mm with a 95% confidence interval of up 5 mm [8]. With the tight margins used in modern IMRT and VMAT plans (< 5 mm) such registration errors could lead to target miss and/or OARs overdose and could prove critical. This is especially true in SRS cases with proton therapy using pencil beam scanning [9, 10].

Removing the CT scan and basing the entire treatment workflow, i.e. simulation, planning and delivery on MRI as the single modality, so-called MRI-only RT, would eliminate this registration error. Also, intrinsically registered functional images such as diffusion weighted MRI could act as a non-invasive guidance for dose escalation volumes within the tumor [11, 12] . The main concerns with MRI-only RT are scanner induced geometrical distortions arising from gradient non-linearity and magnet inhomogeneities, patient induced artifacts such as susceptibility and chemical shift, and, which is the focus of this study, no direct relation between voxel grayscale values and electron density [13]. The latter is needed for dose calculation and bone visualization in the generation of digitally reconstructed radiographs (DRRs) for patient setup verification and delineation guidance of e.g. bone demarcation of lymph nodes. Further, bone density assignment on MRI has proven necessary in attenuation correction of hybrid PET/MRI systems for correct standard uptake volume (SUV) estimation [14].

In general, both atlas- and voxel-based approaches have been pursued for electron density assignment of MR images [15-17]. Here, we will investigate a number of strategies within the voxel-based approach. An MRI sequence which has shown a great potential for voxel-based assignment of electron densities is the dual ultrashort echo times (dUTE) sequence [18, 19]. By fast k-space sampling and a minimum delay in coil transmit-to-receive mode (the so-called tune delay), an acquisition can be obtained immediately after excitation (TE_1) and one at a later time (TE_2). Hence, visualization of contrast for solid-state like structures with a very short T_2 relaxation time such as the cortical bone (between 0.5 ms and 1.7 ms at 1.5 and 0.5 T, respectively) becomes possible [20, 21]. A number of investigators have used the dUTE sequence to make

bulk density assignments based on somewhat flexibly designed thresholds [21-25] or create a so-called pseudo-CT (pCT) using statistical regression [26, 27]. Multiple studies have reported on the dosimetric feasibility of MR bulk density based photon treatment plans for various treatment sites [28-33] but only recently have a few studies emerged in which the CT data is completely absent in the density assignment [34, 35].

Voxel-based segmentation approaches have mostly been investigated individually in the literature with diverging application endpoints in mind. Here, we investigate three categories of approaches which are commonly reported; threshold-based segmentation (hybrid PET/MRI system) [21, 24], Bayesian statistics (brain tissue classification) [37-39] and statistical regression (MRI-only RT) [17, 26, 27].We will investigate and compare the approaches using two methods within each category on the same dataset for radiotherapy purposes by generating pCTs and comparing the geometrical agreement with their corresponding CTs. Further, we investigate the dosimetric impact of pCT dose planning for a similar photon and proton treatment generated on all the patients. Finally, the geometrical and dosimetrical agreement for the methods are compared and discussed.

Material and methods

Imaging

Five patients, three men and two women age 55-82 years, receiving cranial irradiation with palliative intent were included in the study and informed consent was obtained from all patients to acquire extra MR images in addition to their standard treatment. Each patient data contains a co-registered MRI (Philips 1T Panorama HFO) and CT scan (Philips Big Bore CT) with the patient immobilized in a thermoplastic mask both at the CT and MRI scanner. The MR images were obtained with dual flex coils to make room for the patient immobilization devices. Two dUTE scans were acquired at flip angles 10 and 25^{o} each using the minimum and an optimized echo time at TE_1=0.09 and TE_2=3.5 ms, respectively. The two flip angles are introduced to discriminate between T_2 identical tissues with different T_1 properties [15, 26] and are used as input for the statistical regression methods. For the other methods, only the dUTE images at 25^{o} were applied. The TE_2=3.5 ms was found to yield the highest contrast-to-noise ratio for soft tissue versus bone in a bovine knee phantom over a 1.7-6.0 ms investigated interval [36]. Other dUTE MRI parameters were TR=7.1 ms and a FOV and isotropic resolution of 256 and 1 mm, respectively. One dUTE sequence at one flip angle had 512 slices, 256 for each TE (the CT had about half), and had a scan time of about 5 minutes in a radial 3D acquisition mode. In addition, a T1w 3D MRI scan with TE=6.9 ms, TR=25 ms and 1.2 mm resolution was acquired and used for auto-delineation of the brainstem and other OARs. A typical dUTE image can be viewed in Figure 1 at TE_1 (left) and TE_2 (middle). The study was approved by the Capital Regional Ethics Committee (protocol number H-3-2011-107).

Voxel-based approaches

To illustrate the different voxel-based methods, a synthetic patient containing four tissue classes was created from the dUTE images in Figure 1. The four tissue types with distinct attenuation coefficients were: air cavity, adipose and soft tissue, and cortical bone (Figure 1, right). The dUTE intensity histogram of the synthetic patient is illustrated in Figure 2. Here, each tissue is color coded and the optimal TE_2 acquisition time is illustrated when the tissue intensities are separated the most (Figure 2, middle). A bulk or continuous density assigned MRI is collectively referred to as a pseudo-CT (pCT) in the following.

Threshold-based segmentation

Bulk tissue classification based on thresholds has received an increasing interest for attenuation correction in hybrid PET/MRI systems [21, 24]. Figure 3A illustrates a method we label as *R2,* since it is based on the decay constant of the bone ($R2=1/T_2$) [21]. Below a manually set threshold in the TE_1 intensity, all voxels are set to air. If the intensities are above this value and have decayed faster than the decay constant of bone, then voxels are classified as bone and otherwise as soft tissue (soft and adipose). Figure 3B illustrates a method we label as *Logic* based on the Boolean operations performed [23]. An air threshold for the TE_1 intensities is again assigned. If the TE_2 intensities are below a threshold and the TE_1-TE_2 difference intensities are above another threshold (here bone will light up) then voxels are considered as bone. Otherwise, voxels are again assigned to soft tissue(s). For the threshold-based segmentation, the threshold values were manually determined for each individual patient by inspection of the MRI intensities. The decay constant for cortical bone was estimated from the dUTE sequence to be around 0.4 ms^{-1} why this value was used for all the patients.

Bayesian segmentation

Bayesian statistics including spatial information have been investigated for tissue segmentation on MR images of the brain [37-39]. The basic assumption in Figure 3C and D is that the TE_1 and TE_2 voxels come from a mixture of multivariate normal distributions. An expectation maximization (EM) algorithm [40] was used to estimate the parameters of the underlying Gaussian distributions. A posterior probability for belonging to each Gaussian distribution, i.e. tissue class, can then be estimated for every voxel. Finally, voxels are assigned to the tissue class with the highest posterior probability. This describes the method named *Bayes* in Figure 3C which has a known four class distribution. The Markov Random Field, *MRF*, method in Figure 3D further takes the tissue class of neighborhood voxels into account when estimating the posterior probability. From the EM algorithm, the Bayes and MRF methods use an initial value of 7 tissue classes (multiple soft tissue classes were collected to the same HU) for the patients. The neighborhood weighting for MRF was set to 0.7 based on 26 neighboring voxels. At a weighting factor of 0, the MRF method reduces to the Bayes method.

Statistical regression

Statistical regression has shown a great potential for MRI-only RT and PET/MR attenuation correction of the brain [17, 26, 27]. In Figure 3E, Gaussian mixture regression (*GMR*) is visualized in a simplistic way for the TE_2 intensities. Again, the intensities are assumed to origin from a mixture of normal tissue class distributions but now the co-registered CT data are also included, yielding a multidimensional Gaussian mixture model [26]. In this way, a model to perform regression between MRI and CT intensities can be trained with the EM algorithm on multiple patients and a pCT can be estimated from the MRI input data of a patient not included in the training. Random forest regression (*RaF*) can be inspected in Figure 3F. Here, the MRI training data are split at the nodes in a binary decision tree using thresholds to optimize some criteria, e.g. minimize the absolute deviation of the corresponding registered CT data from the mean CT number of the split [41]. Performing the RaF on a random subgroup of the training data and limiting the number of possible splits results in a sub-optimal regression tree (Figure 3F, gray lines). Using a number of regression trees (a forest), the average of the suboptimal regressions improves the generalization error (Figure 3F, thick black line).

In addition to the 4 dUTE images (2 echo times at 2 flip angles), 8 derived filtered images containing dUTE means and standard deviations were included as inputs in the regression training of the patients. The GMR method had 20 initial tissue classes and a *k* means clustering algorithm provided the initial cluster centers as input for the EM algorithm [26]. The RaF method used CART criteria for tree training [42], 80 trees and a minimum terminal node size of 10. Both models included an air mask to reduce the amount of data. The GMR and RaF models were trained with the MRI and CT data of three patients and predicted the pCT of a patient not included in the training using only the MRI input of that patient.

Treatment planning

A photon and proton treatment plan with three equally weighted fields, two lateral opposing and one anterior field, were created for each patient in Eclipse v. 11 (Varian Inc.). The apertures were cropped to the target tumor with a 5 mm margin. In the lateral opposing fields, however, the margin was compromised dorsally due to the brainstem which was either shielded (photons) or given a zero mm margin (protons). Dose calculations were made with the AAA algorithm for the 6 MV photons and the CAPGEN algorithm for the 70-250 MeV protons (CAP general accelerator and the uniform scanning technique). The treatment planning was carried out on each pCT. The plan was then transferred and re-calculated on the real CT using the same planning parameters, i.e. the same monitor units for the photon plans and same beamline settings and normalization value for proton plans. A spherical tumor target with a 3 cm diameter was identically positioned between the nasal cavity and brainstem for all the patients. The tumor was subsequently cropped to the cavity. The brainstem was auto-delineated with iPlan v. 1.5.0 (Brainlab) on the high resolution T1 weighted MRI scan, see Figure 4. In this way, the performance of the pCTs were tested in a worst case scenario in which a large portion of the radiation fields travels through alternating tissues with large density differences before depositing the dose in the tumor (high dose volume) and brainstem (gradient dose volume). A mean absorbed dose of 2 Gy in 30 fractions was prescribed to the tumor volume. For the threshold and Bayesian segmentation methods, soft tissues, bone and air cavities were assigned CT numbers equal to 0, 971 (cranial bone [43]) and -993 HU, respectively. A calculation in which the entire MRI is set to water (MRI_w) was included for comparison. An example of the pCT-CT pair wise DVH comparisons can be seen in Figure 5.

Metrics

A number of metrics were introduced to quantify the geometric and dosimetric performance of the pCTs as compared to the corresponding CTs. The dice similarity coefficient [16, 44] was calculated for soft tissue and bone as

$$DSC_x = \frac{2(V_x^{CT} \cap V_x^{pCT})}{V_x^{CT} + V_x^{pCT}} \quad (1)$$

where V_x is the volume of tissue *x*. The bone volume included all voxels with a CT number > 200 HU giving a density of 1.14 g/cm^3 according to our CT calibration curve which is between the density of cartilage and spongious bone [43]. Soft tissue volumes included voxels with CT numbers > - 100 and ≤ 200 HU giving a minimal density of 0.95 g/cm^3 including most adipose tissue. To further quantify the geometrical agreement between pCT and CT, the mean absolute error, MAE, and mean error, ME, were calculated [17]. They are given by $MAE = \frac{1}{N}\sum_{i=1}^{N}\left|CT_i - pCT_i\right|$ and $ME = \frac{1}{N}\sum_{i=1}^{N} CT_i - pCT_i$ where *N* is the total number

of voxels and *i* is the voxel index. Both quantities have the unit of HU. The percentage of patient voxels classified as air cavity (< -200 HU), bone (> 200 HU) and soft tissue (all additional voxels) was calculated for all pCTs and the percentage point difference as compared to the CT was extracted for each tissue. The cumulative percentage point difference was then calculated. Here, we define the mean absolute percentage deviation as

$$MAPD = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|y_i^{CT} - y_i^{pCT}\right|}{\overline{y^{CT}}} \tag{2}$$

Similar, the mean percentage deviation is defined as

$$MPD = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i^{CT} - y_i^{pCT}}{\overline{y^{CT}}} \tag{3}$$

For geometrical comparisons, *y* is the CT value in HU of voxel *i* and for dosimetrical comparisons *y* is the dose value in Gy of DVH point *i*. $\overline{y^{CT}} = \frac{1}{N}\sum_{i=1}^{N} y_i^{CT}$ is the mean CT value to avoid singularities, especially for geometrical HU values around the value of water. We now define a combined geometrical score as

$$CGS = (\,MAPD + \left|MPD\right|\,) \cdot \frac{1}{M}\sum_{j=1}^{M}(1 - DSC_j) \tag{4}$$

where *M* is the total number of tissue classes and *j* is the tissue class index. In our case, *M* is 2 for soft tissue and bone. The absolute value of MPD is taken to avoid a cancellation with the MAPD value. The normalized summation term of the dice similarity coefficients is approaching zero for perfect agreement (DSC →1) or 1 for maximum disagreement (DSC →0). In this way, the term containing the MAPD and MPD is scaled with a factor between 0 and 1. A good agreement is characterized by a low CGS value. The combined dosimetric score is more simply given as

$$CDS = MAPD + \left|MPD\right| \tag{5}$$

Further, the percentage difference from CT for a number of pCT DVH points was calculated. The tumor DVH points were the near-maximum absorbed dose (D2), the near minimum absorbed dose (D98) and the median absorbed dose (Dmedian) collectively describing the target coverage [45]. For the brainstem, the difference in Dmax was scored as this is commonly used as a DVH constrain in clinical practice [46, 47]. In addition to the CDS score, a DVH-reduction model was introduced to quantify the biological effect of the differences in the entire DVH [46, 48]. This is done through the generalized equivalent uniform dose (gEUD) given by

$$gEUD = \left[\sum_{i=1}^{N} v_i d_i^a\right]^{\frac{1}{a}} \tag{6}$$

where v_i is unitless and represents the *i*'th partial volume receiving dose d_i in Gy. *N* is the total number of partial volumes and *a* is a unitless model parameter specific to the tumor or Brainstem. d_i was converted to

an equivalent dose in 2 Gy fractions (eqd2) using α/β =3 for the brainstem and α/β=10 for the tumor. *a*=7 for the brainstem and *a*=-10 for the tumor [48, 49].

Results

Geometrical performance

The pCT performance in terms of geometrical agreement with the corresponding CT for the different methods is summarized in Table 1. The dice scores (DSC_b, DSC_s), errors (MAE, ME) and prediction deviations (MAPD, MPD) are showing an increasing agreement between pCT and CT with an increasing level of model sophistication and complexity, i.e. from water (top), through the threshold and Bayesian based methods to the statistical regression (bottom). A box plot for the CGS, including its individual components, shows the same trend in Figure 6. A two-tailed pairwise t-test (pooled standard deviation and p adjusted values [50, 51]) between the CGS of the different methods showed a significant ($p<0.05$) lower value for the GMR and RaF methods compared to the R2, Logic and Bayes methods (not MRF). All methods' CGS was significantly lower than that of water. The MAE of the statistical regression methods (GMR and RaF) in Table 1 are in the same range as those reported in the literature (117-176 HU) [26, 27, 35]. The ME showed a systematic underestimation of the HU values as compared to CT for the water and threshold methods and the opposite for the Bayesian based methods. The statistical regression methods showed almost no systematic tendency. In Figure 7, the DRRs of all the methods are visualized for a patient. By inspection of the DRRs, a lower CGS results in a better image quality of the DRR images.

Dosimetrical performance

A box plot for the CDS for the tumor and brainstem, including its individual components, is shown in Figure 8. For all panels, the CDS is highest for the reference MRI water plan reflecting the largest deviation compared to the re-calculation on the CT. This has the highest impact for protons. The MPD for the tumor is in opposite directions for the photon and proton plans in the high dose region (tumor) whereas they are in the same direction in the gradient dose region (brainstem). This is also apparent in Figure 5 for the MRIw panels. For the photon plans, the CDS of the Bayesian statistics methods (Bayes and MRF) seem to perform just as poorly as the MRI water reference calculation. The deviation scored by the CDS is about a factor of 10 smaller for the high dose region as compared to the gradient dose region. The changes in density of the pCTs seem to have largest impact for photons as compared to protons in the high dose region and the opposite in the gradient dose region. Surprisingly, the threshold based methods (R2 and Logic) which have a high CGS value (poor geometrical agreement) show the lowest dosimetric deviation compared to CT for both photons and protons (see discussion). Dose metrics for tumor coverage, brainstem Dmax and biological effect (gEUD) are shown in Figure 9. In general, dose calculations performed on pCTs using protons overestimates the dose compared to CT (deviations below zero) whereas the opposite is the case for photons (deviations above zero). Although the MRI water calculation overall result in the largest deviations and the statistical regression methods in general have the smallest deviations, the trends in data are somewhat heterogeneous. Especially, the threshold based methods seem to give the best agreement with the CT re-calculation and this is most pronounced for photons. Except for water, the mean deviation is within 2 percent for most panels and highest in the gradient dose region of the brainstem especially when considering the deviation of the entire DVH.

Discussion

MRI is increasingly becoming a more integral part of the RT simulation process, and recently, also introduced as an image modality at the accelerator for online treatment adaptation [52]. For online MRI guided systems, a calculation performed on the MRI scan obtained onsite would be needed to get a plan of the day. This calculation could either be made on a density corrected MRI (a pCT) or through deformable registration of the MRI to the planning CT. Such issues together with the need for elimination of CT-MRI registration error raise focus on density corrected MRIs. Here, we have investigated and quantified different voxel-based strategies to obtain a pCT. Both from a geometric and dosimetric point of view, making some form of density correction is always better than assigning all tissue to water at least in the challenging parts of the brain region.

As scored by the CGS, the threshold based methods do not perform well as compared to the other methods although good geometric agreement was reported previously [21, 24, 25]. The dependence of manually set thresholds represents both the strength and weakness of these methods; they can be optimized to a particular dataset but can be hard to reproduce on other datasets. Conversely, the statistical regression methods show the best geometrical agreement and seem to be robust across scanning platforms and patient data [15, 17, 26]. A correlation between geometrical and dosimetrical agreement can be observed for the reference water MRI (poor) and the statistical regression methods (good). This is, however, not the case for the threshold and Bayesian methods. Especially the threshold methods demonstrate a surprisingly good dosimetric agreement which is most pronounced when considering deviations of the entire DVH (DSC and gEUD for tumor and brainstem). Some explanation might be found in column 4-5 of Table 1. Although the threshold based methods have a low voxel-wise geometrical agreement, the deviation in different tissue amount is low and of the same order as the statistical regression methods. In Figure 10, the dosimetric consequence of tissue location and tissue amount is tested for a three field treatment plan similar to the patient plans. From this test, it is clear that tissue amount (Figure 10B, low deviation) rather tissue position (Figure 10C, high deviation) is of dosimetric importance. In other words, it is the photon attenuation and proton stopping power (continuous-slowing-down approximation range [53]) that is of primary concern rather than changes in the fluence spectra entering the different materials. Protons, however, seem to be less forgiving for the latter which can be observed in Figure 9 where the low deviation for the threshold methods is less pronounced as compared to photons. The cumulative error is defined in the caption of Table 1. It corresponds well with the dosimetric agreement for the phantom since it is low (zero) in Figure 10B which has a low dosimetric deviation and poor geometrical agreement (low DSC) and high in Figure 10C which has a high dosimetric deviation and good geometrical agreement (high DSC). This metric seem to be of importance when considering the dosimetric properties of a pCT. The metric, however, is probably a too simplistic measure as both the threshold and statistical regression based methods show a low cumulative error but diverse dosimetric results. One could, for example, imagine the inserts in Figure 9 lying outside the irradiated volume affecting the dice similarity coefficient and cumulative error but leaving the dosimetric deviation unaffected. A measure combining the correct amount of a given tissue and the correct position of it would probably be a better alternative (the MAE of water equivalent path lengths did not resolve the ambiguity in the data).

The overall 2 % dosimetric deviation was set in the context of registration errors by shifting the treatment plans 2 mm in the longitudinally, transverse and vertical direction for all patients. This resulted in an overall displacement vector of 3.4 mm which is between the cited mean of 2 mm and 95% CI of 5 mm [8]. A re-calculation resulted in a Dmedian and D2 deviation between 0.5-1% which is about the same as the pCT deviations shown in Figure 9. Due to the gradient dose regions within field aperture (which is present in intensity modulated treatment plans), however, large differences were observed for D98 (≈18%), gEUD.tumor (≈ 13%), gEUD.brainstem (≈ 21%) and Dmax (16-29%). This illustrates the critical impact of registration errors on clinically relevant dosimetric parameters.

Rank et al. used discriminant analysis in a Gaussian Mixture Model framework to generate pCTs [17]. They reported on the dose differences in the PTV mean dose in a beam arrangement with lateral opposing proton or carbon ion fields [35]. These deviations were within 0.4-1.8 % which is in in good agreement with our results (the Dmedian deviation of the GMR point of upper left panel in Figure 9) although in the lower range. This could probably be explained by the more challenging position of our target behind the nasal cavity compared to their target which was positioned in the posterior part of the brain. Also, the addition of an anterior field through the nasal cavity in our study could add to an increased deviation. They also reported on a very low deviation, about 0.2 %, for a 7 field photon plan for the same target. It is hard to compare this result with the dosimetric photon deviation of this study since the beam arrangement and target position is different. Jonsson et al. generated pCTs using the GMR method [34] . They reported on the dose deviation for D90 of a PTV positioned superior, medial and inferior in the brain, the latter being most similar to our target position. The inferior PTV had the largest D90 deviation of 0.9 % for all the PTVs which is about a percentage point lower than the GMR photon deviations observed in the D98 panel of Figure 9. Although they operated with a three field beam arrangement similar to this study, the anterior field was given only half the weight of the lateral opposing fields. Overall, it seems that we indeed have presented a worst case dosimetric setup and that the methods are likely to perform better in less heterogeneous parts of the brain.

A statistical criterion for the safe use of MRI-only RT was reported to require a deviation less than 2% on the PTV coverage for 95% of the patients compared to a standard CT-based calculation (≈ max|mean deviation ±2 standard deviations|) [28]. Although this study probably has too few patients (low statistical power) for such considerations, the RaF method for both protons and photons seem to be the most promising candidate to fulfill such a criterion when inspecting the error bars (1 standard deviation) in the upper panels of Figure 9. To improve the statistical regression methods, addition of spatial information [27], a water-fat separating MRI mDixon sequence [54] and multiple T_1 and T_2 weighted sequences [17, 26] have been tested. Scored by the MAE, however, this did not improve the CT number agreement significantly; MAE=130 HU (spatial, GMR), MAE=128 HU (mDixon, GMR) and MAE=113 HU (mDixon, RaF). This suggests that focus should be on improving the prediction algorithms rather than the amount of input data.

Conclusions

We investigated the geometrical and dosimetrical performance of pseudo CTs as compared to their corresponding CT for a number of voxel-based algorithms. Overall, pCTs generated from statistical regression methods seem to be the most promising candidates for MRI-only RT of the brain keeping the mean dosimetric deviation within 2 percent and a geometrical deviation to a minimum. The amount of

tissue with distinct absorption properties, however, also seems to be of dosimetric importance regardless of correct position.

Acknowledgements

This work was supported by a research grant from Varian Medical Systems, Inc.

Figures and tables

Figure 1: dUTE MRI images of a patient at times TE_1=0.09 ms (left) and TE_2=3.5 ms (middle). Four tissue classes; air cavity, adipose and soft tissue, and bone, were manually drawn (white lines). A synthetic patient was created containing the four tissues each with the mean and standard deviation from the TE1 image normalized to the mean intensity of adipose tissue and some Gaussian noise (right).

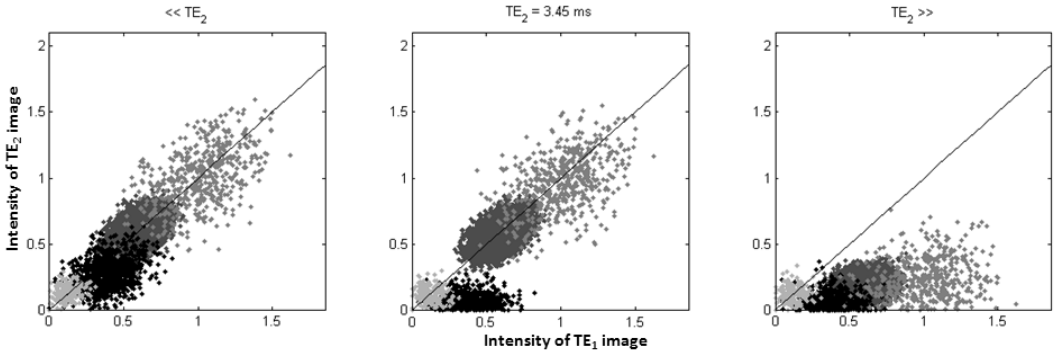


Figure 2: The dUTE intensity histogram of the synthetic patient at different TE_2 acquisition times: 0.35 ms (left), 3.5 ms (middle) and 175 ms (right). The intensities are normalized as in Figure 1. At $TE_2 \approx 3.5$ ms, the tissue classes are separated the most. White gray=air, T_2 constant=0 ms. Light gray=adipose tissue, T_2 constant = 80 ms. Dark gray=soft tissue, T_2 constant = 100 ms. Black=bone, T_2 constant =0.5 ms.

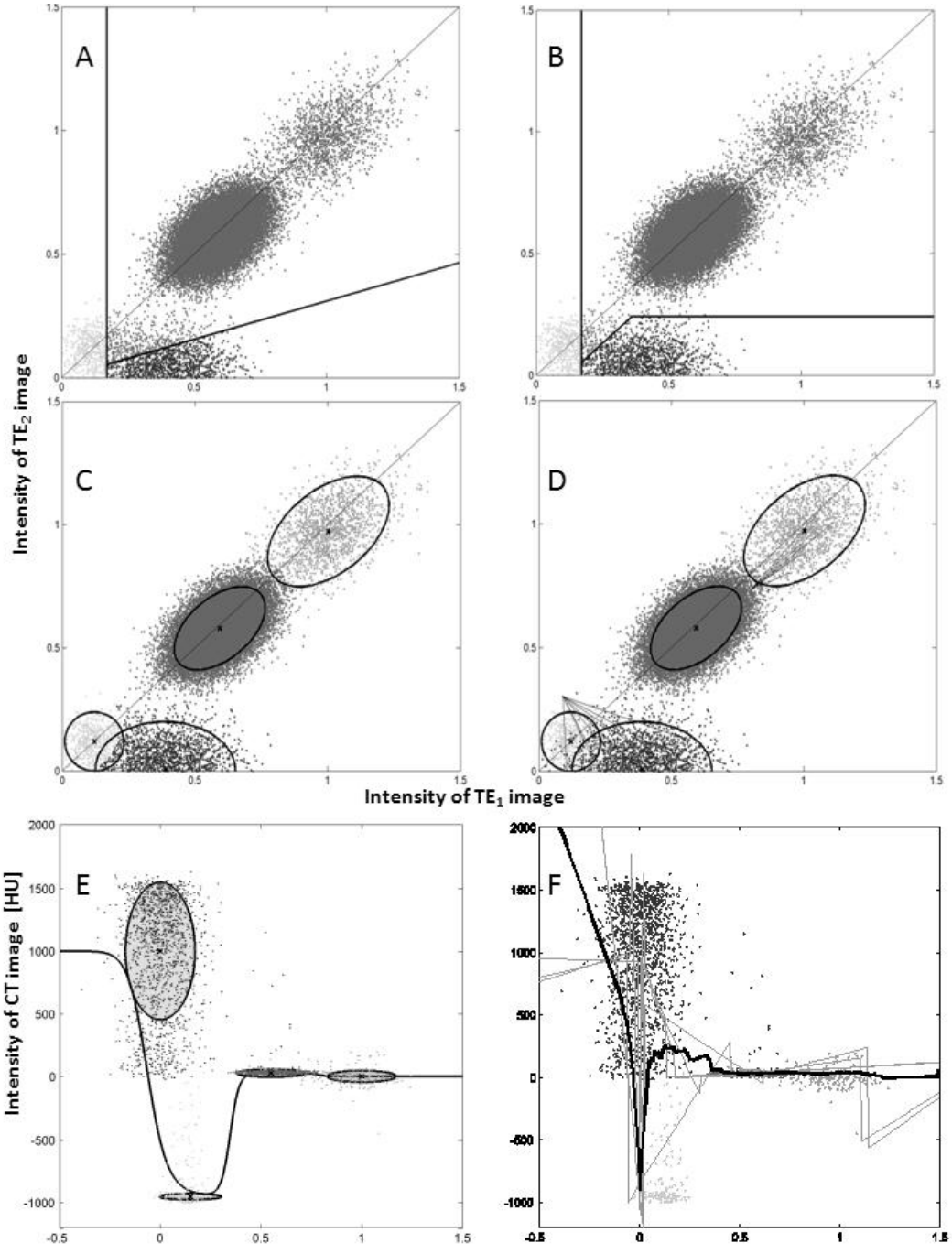

A
B
C
D
E
F
Intensity of TE2 image
Intensity of TE1 image
Intensity of CT image [HU]
Intensity of TE2 image

Figure 3: The different methods used in this study illustrated on the synthetic patient at TE_2=3.5 ms. The MRI intensities are normalized as in Figure 1 and tissues are color coded as in Figure 2. **A** *R2:* bone if decay ≥ 0.35 ms^{-1} and intensity of TE_1 image > 0.1715. **B** *Logic*: Air if intensity of TE_1 image < 0.1715. Bone if intensity of TE_2 image < 0.24 and intensity of TE_1-TE_2 image > 0.12. Otherwise soft tissue. **C** *Bayes*: Trained with 4 classes. Estimated classes are color coded and black ellipses represent the Gaussian distributions. **D** *MRF*: As in C with 8 neighbors with a neighbor weighting factor of 0.4. 2 voxels are shown with thin black lines to the 8 nearest neighbors to indicate a different classification than in the *Bayes* method. **E** *GMR*: MRI registered to the corresponding CT and manually assigned to 4 classes. The model included only the TE_2 images and manually set classes for illustrative purposes. Regression is shown with black line and the original color coding from the synthetic patient. **F** *RaF*: 4 initial classes on the TE_2 image only and color coding as in E. Each tree regression (tree depth=3) is shown with gray lines and the forest (40 trees) regression with black line.

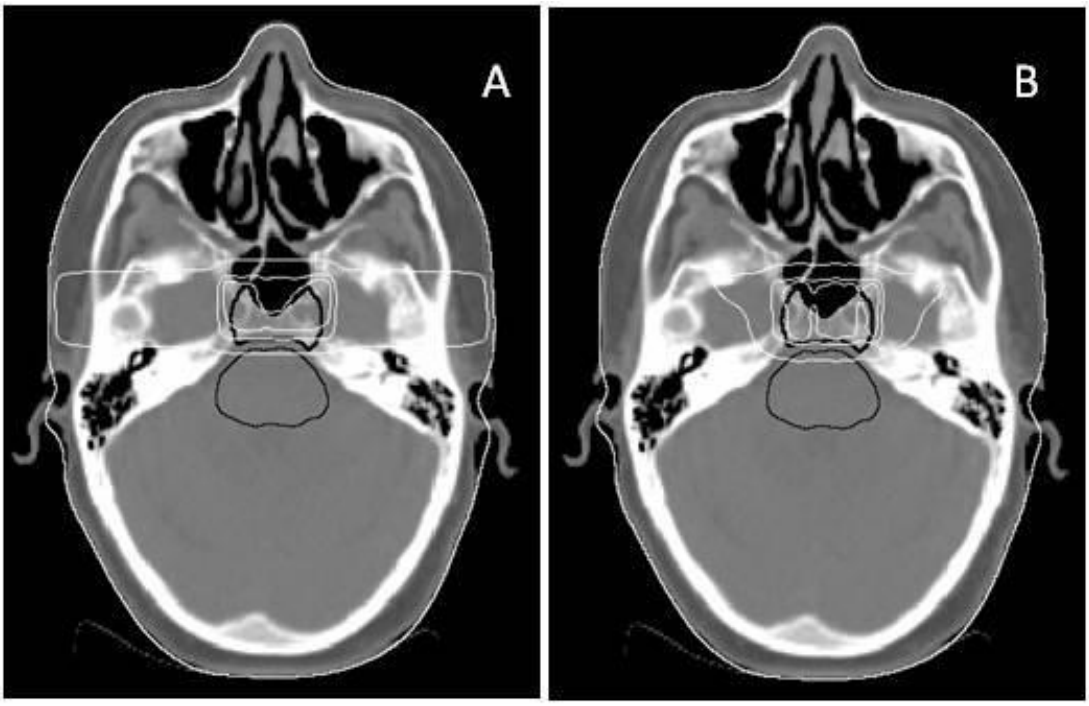


Figure 4: A three field photon (**A**) and proton (**B**) plan. Delineated structures (black lines): tumor (thick) and brainstem (thin). Isodose curves (white): 50% (outer most), 95% and 100% (inner most).

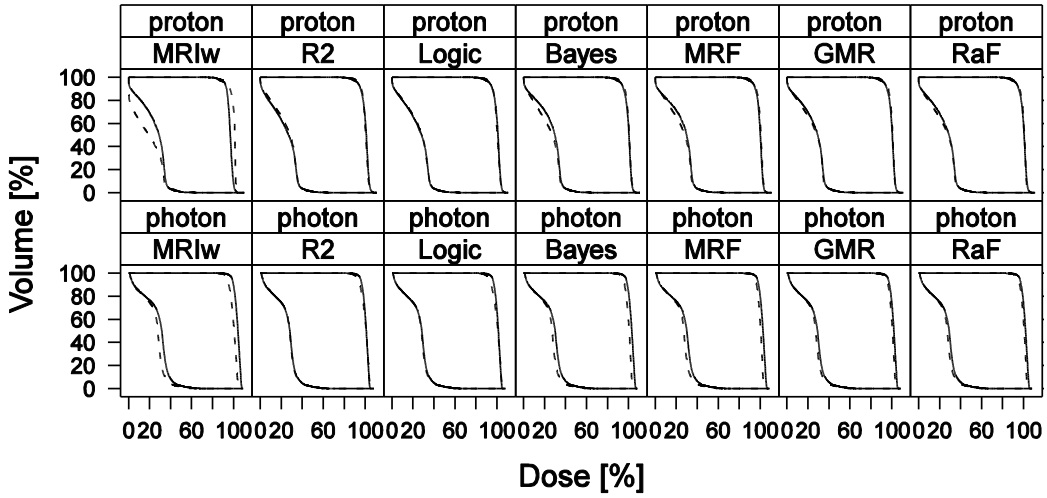


Figure 5: Pair wise comparison of the pCT calculated DVH (dashed line) and the corresponding re-calculated CT DVH (solid line) of a patient. The tumor (rightmost DVH) and brainstem (leftmost DVH) are shown for the proton (top) and photon (bottom) plan.

Method	DSC_b	σ_{DSCb}	DSC_s	σ_{DSCs}	MAE	σ_{MAE}	ME	σ_{ME}	MAPD	σ_{MAPD}	MPD	σ_{MPD}	δ_{cav}	σ_{cav}	δ_{bone}	σ_{bone}	δ_{soft}	σ_{soft}	Δ	σ_{Δ}
MRIw	0.00	0.00	0.85	0.02	213	28	108	34	2.1	0.6	1.0	0.0	-4.0	0.8	-17	2	21	2	27	3
Logic	0.49	0.05	0.87	0.02	226	20	50	36	2.2	0.6	0.4	0.2	0.3	0.4	0	4	-1	4	4	3
R2	0.48	0.03	0.86	0.02	241	29	50	32	2.4	0.6	0.5	0.3	-0.3	0.5	3	2	-3	2	4	3
Bayes	0.53	0.03	0.87	0.02	247	35	-59	29	2.4	0.6	-0.6	0.3	-1.8	0.5	11	3	-9	3	15	4
MRF	0.59	0.04	0.89	0.02	204	36	-27	37	2.0	0.5	-0.3	0.3	-2.0	0.4	3	4	-1	5	5	4
GMR	0.72	0.02	0.90	0.01	136	18	2	22	1.3	0.3	0.0	0.2	-0.7	0.3	3	4	-3	4	5	6
RaF	0.74	0.02	0.91	0.01	128	18	8	24	1.2	0.3	0.0	0.2	-0.5	0.3	3	4	-2	4	4	5

Table 1: Metrics of geometrical agreement between pCT and CT. Average value and standard deviation (σ_x) for the 5 patients are shown. **Column 1**: Dice similarity scores for bone (DSC_b) and soft tissue (DSC_s). **Column 2**: Mean absolute error (MAE) and mean error (ME) in units of HU. **Column 3**: Mean absolute prediction deviation (MAPD) and mean prediction deviation (MPD). **Column 4**: Percent point deviation between relative tissue amount in CT and pCT. **Column 5**: The cumulative error of column 4, $\Delta = \sqrt{\delta_{cav}^2 + \delta_{bone}^2 + \delta_{soft}^2}$.

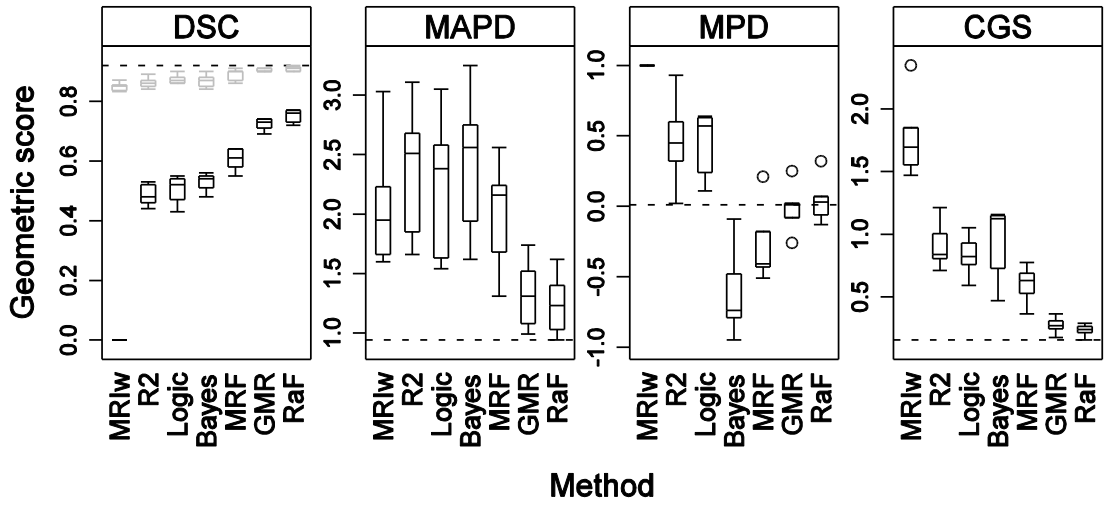


Figure 6: Boxplots of the combined geometrical score (CGS, rightmost) including its individual components for the different methods. Horizontal lines=median values, box=first and third quartiles, vertical lines=1.5 times the interquartile range and points=values > 1.5 times the interquartile range (outliers). For the DSC panel, soft tissue=gray and bone=black. In each panel, the best value, i.e. best agreement, is indicated with a dashed line. MRI_w indicates an MRI with all voxels assigned to the density of water (HU=0).

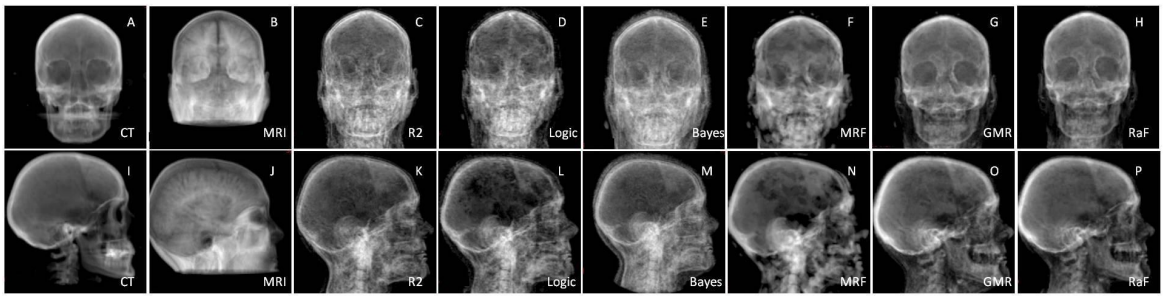


Figure 7: Frontal (top) and lateral (bottom) setup DRRs of a patient. **CT**: A and I, **MRI T1w**: B and J (smaller caudal FOV), **R2**: C and K, **Logic**: D and L, **Bayes**: E and M, **MRF**: F and N, **GMR**: G and O, **RaF**: H and P.

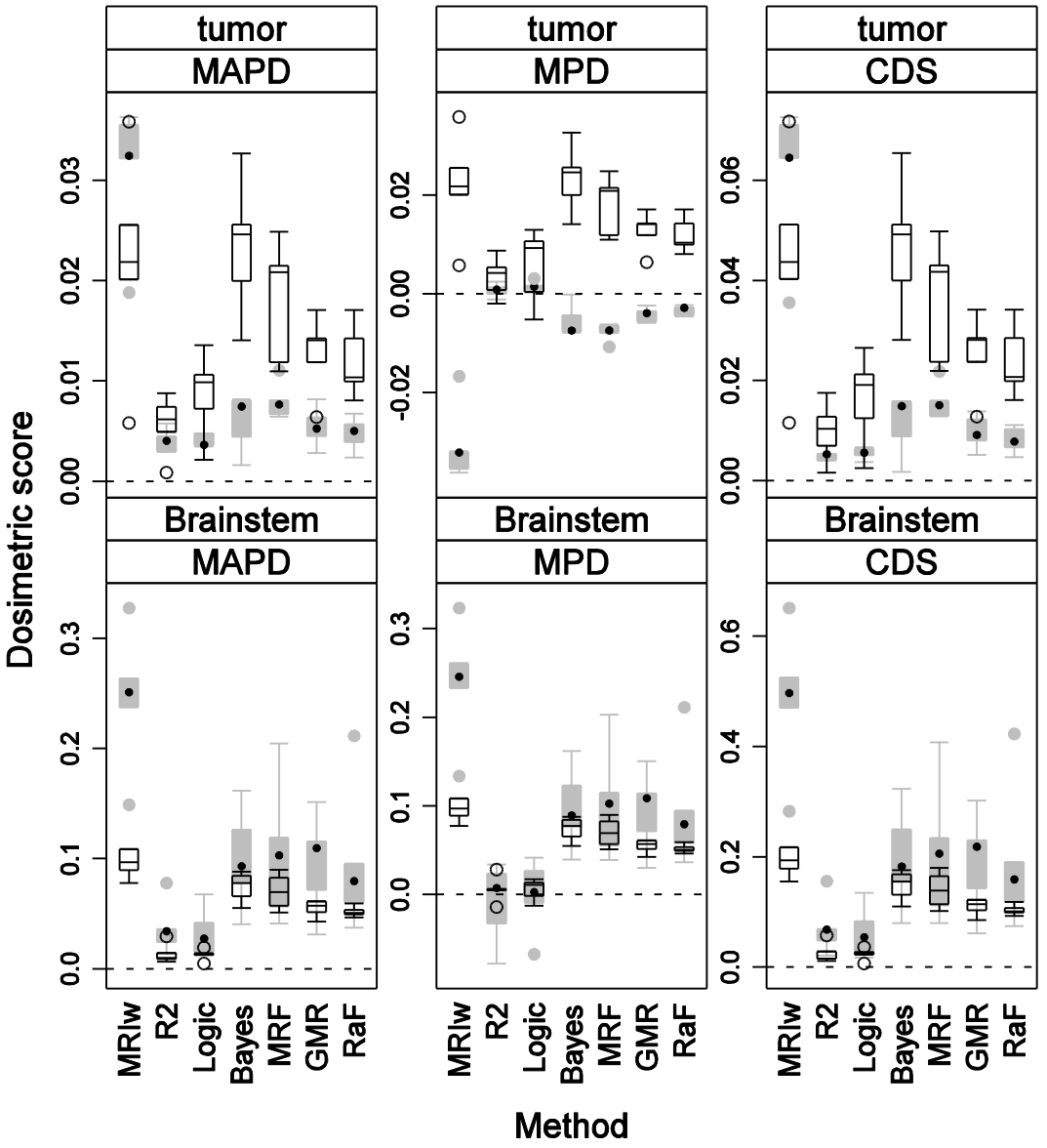


Figure 8: Boxplots of the combined geometrical score (CDS, rightmost) including its individual components for the different methods. Tumor (top) and brainstem (bottom). Photons: Black horizontal line= median value, white box= first and third quartile, black vertical line= 1.5 times the interquartile range and open circle=outliers. Protons: Closed black circle= median value, gray box= first and third quartile, gray vertical line= 1.5 times the interquartile range and closed gray circles=outliers. In each panel, the best value, i.e. best agreement, is indicated with a dashed line.

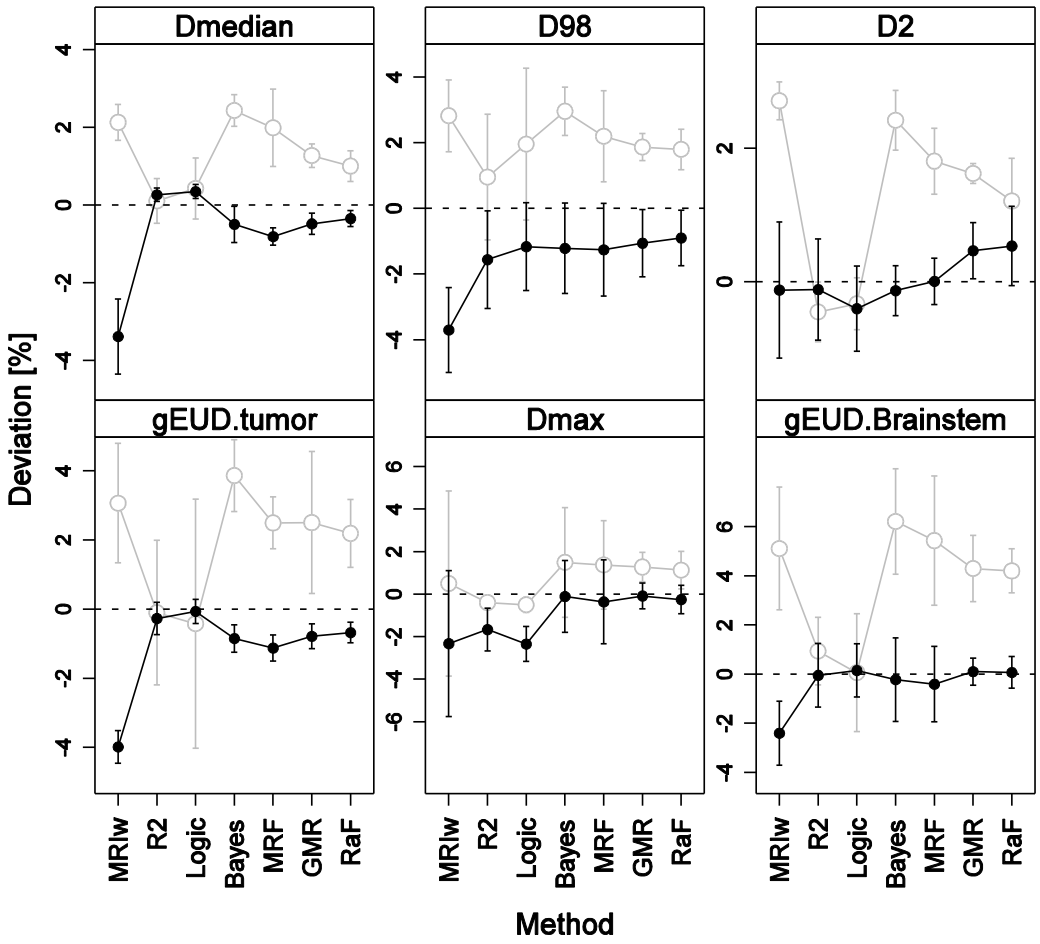


Figure 9: DVH points deviation of the pCT calculated value from the re-calculated CT value in %. Mean values and standard deviations are plotted for photons (gray lines, open circle) and protons (black lines, closed circle). The tree upper panels are the tumor coverage Dmedian (left), D98 (middle) and D2 (right). The three lower panels are gEUD for the tumor (left), Dmax for the brainstem (middle) and gEUD for the brainstem (right). Connected lines are introduced only to illustrate trends in data.

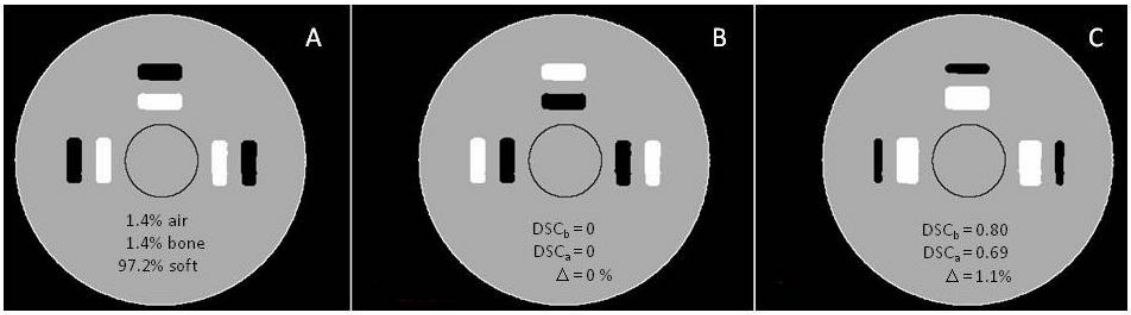


Figure 10: A virtual cylindrical water phantom (0 HU, gray) with air (-1000 HU, black) and bone (1000 HU, white) inserts in the three beam directions. **A:** The reference phantom with a 3 field equally weighted photon and proton plan giving a dose of 2 Gy in 30 fractions to a spherical 5 cm diameter target (black circle). Tissue percentages are given in the figure. The plan is re-calculated on the phantoms in B and C. **B:** Air and bone inserts are interchanged. Dice coefficients and cumulative error are given in the figure. Deviations photons/protons: CDS·10^{-3}=1.6/ 0.2, Dmedian=0.05/0.0%, D98=0.2/0.01%, D2=0.04/-0.01%, Dmax=0%. **C:** Air/bone insert volumes are decreased/increased about 50%. Dice coefficients and cumulative error are given in the figure. Deviations photons/protons: CDS·10^{-3}=45.1/ 5.6, Dmedian=1.2/5.0%, D98=1.6/3.7%, D2=1.4/0.01%, Dmax=1.2/0.3%.

Bibliography

[1] Prabhakar R, Haresh KP, Ganesh T, Joshi RC, Julka PK, Rath GK. Comparison of computed tomography and magnetic resonance based target volume in brain tumors. J Canc Res Ther 2007;3:121.
[2] Fiorentino A, Caivano R, Pedicini P, Fusco V. Clinical target volume definition for glioblastoma radiotherapy planning: magnetic resonance imaging and computed tomography. Clinical & Translational Oncology 2013;15:754-8.
[3] Rasch C, Steenbakkers R, van Herk M. Target definition in prostate, head, and neck. Semin Radiat Oncol 2005;15:136-45.
[4] Cattaneo GM, Reni M, Rizzo G, Castellone P, Ceresoli GL, Cozzarini C, et al. Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR (pre-operative) images on treatment planning CT scans. Radiother Oncol 2005;75:217-23.
[5] Ulin K, Urie MM, Cherlow JM. Results of a Multi-Institutional Benchmark Test for Cranial Ct/Mr Image Registration. International Journal of Radiation Oncology Biology Physics 2010;77:1584-9.
[6] Solberg TD, Medin PM, Mullins J, Li SC. Quality assurance of immobilization and target localization systems for frameless stereotactic cranial and extracranial hypofractionated radiotherapy. International Journal of Radiation Oncology Biology Physics 2008;71:S131-S5.
[7] Minniti G, Scaringi C, Clarke E, Valeriani M, Osti M, Enrici RM. Frameless linac-based stereotactic radiosurgery (SRS) for brain metastases: analysis of patient repositioning using a mask fixation system and clinical outcomes. Radiat Oncol 2011;6.
[8] Daisne JF, Sibomana M, Bol A, Cosnard G, Lonneux M, Gregoire V. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 2003;69:237-45.
[9] Amichetti M, Amelio D, Minniti G. Radiosurgery with photons or protons for benign and malignant tumours of the skull base: a review. Radiat Oncol 2012;7.
[10] Hattangadi JA, Chapman P, Kim D, Bussiere M, Niemierko A, Rowell A, et al. Single Fraction Proton Beam Stereotactic Radiosurgery (PSRS) for Inoperable Cerebral Arteriovenous Malformations (AVMs). International Journal of Radiation Oncology Biology Physics 2012;84:S38-S.

[11] Metcalfe P, Liney GP, Holloway L, Walker A, Barton M, Delaney GP, et al. The Potential for an Enhanced Role For MRI in Radiation-therapy Treatment Planning. Technology in Cancer Research & Treatment 2013;12:429-46.
[12] Lips IM, van der Heide UA, Haustermans K, van Lin ENJT, Pos F, Franken SPG, et al. Single blind randomized Phase III trial to investigate the benefit of a focal lesion ablative microboost in prostate cancer (FLAME-trial): study protocol for a randomized controlled trial. Trials 2011;12.
[13] Nyholm T, Nyberg M, Karlsson MG, Karlsson M. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. Radiat Oncol 2009;4.
[14] Aznar M, Sersar R, Saabye J, Ladefoged C, Andersen F, Rasmussen J, et al. Whole-body PET/MRI: The effect of bone attenuation during MR-based attenuation correction in oncology imaging. Eur J Radiol 2014;(in press).
[15] Johansson A, Karlsson M, Yu J, Asklund T, Nyholm T. Voxel-wise uncertainty in CT substitute derived from MRI. Med Phys 2012;39:3283-90.
[16] Dowling JA, Lambert J, Parker J, Salvado O, Fripp J, Capp A, et al. An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. Int J Radiat Oncol Biol Phys 2012;83:e5-11.
[17] Rank CM, Tremmel C, Hunemohr N, Nagel AM, Jakel O, Greilich S. MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. Radiat Oncol 2013;8:51.
[18] Robson MD, Gatehouse PD, Bydder M, Bydder GM. Magnetic resonance: An introduction to ultrashort TE (UTE) imaging. J Comput Assist Tomogr 2003;27:825-46.
[19] Rahmer J, Blume U, Bornert P. Selective 3D ultrashort TE imaging: comparison of "dual-echo" acquisition and magnetization preparation for improving short-T 2 contrast. MAGMA 2007;20:83-92.
[20] Reichert IL, Robson MD, Gatehouse PD, He T, Chappell KE, Holmes J, et al. Magnetic resonance imaging of cortical bone with ultrashort TE pulse sequences. Magn Reson Imaging 2005;23:611-8.
[21] Keereman V, Fierens Y, Broux T, De Deene Y, Lonneux M, Vandenberghe S. MRI-Based Attenuation Correction for PET/MRI Using Ultrashort Echo Time Sequences. J Nucl Med 2010;51:812-8.
[22] Catana C, van der Kouwe A, Benner T, Michel CJ, Hamm M, Fenchel M, et al. Toward Implementing an MRI-Based PET Attenuation-Correction Method for Neurologic Studies on the MR-PET Brain Prototype. J Nucl Med 2010;51:1431-8.
[23] [Master thesis] Thörnqvist S. Initial step toward MRI-based treatment planning for external radiotherapy. Lund: Lund University; 2009.
[24] Berker Y, Franke J, Salomon A, Palmowski M, Donker HCW, Temur Y, et al. MRI-Based Attenuation Correction for Hybrid PET/MRI Systems: A 4-Class Tissue Segmentation Technique Using a Combined Ultrashort-Echo-Time/Dixon MRI Sequence (vol 53, pg 796, 2012). J Nucl Med 2012;53:1496-.
[25] Edmund JM, Kjer HM, Hansen RH. Auto-segmentation of bone in MRI-only based radiotherapy using ultra short echo time. Radiother Oncol 2012;103 Supplement 1:S75.
[26] Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. Med Phys 2011;38:2708-14.
[27] Johanson A, Garpebring A, Karlsson M, Asklund T, Nyholm T. Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information - potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. Acta Oncol 2013;52:1369-73.
[28] Korsholm ME, Waring LW, Edmund JM. A criterion for the reliable use of MRI-only radiotherapy. Radiat Oncol 2014;9.
[29] Kristensen BH, Laursen FJ, Logager V, Geertsen PF, Krarup-Hansen A. Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours. Radiother Oncol 2008;87:100-9.
[30] Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. Radiat Oncol 2010;5.

[31] Lambert J, Greer PB, Menk F, Patterson J, Parker J, Dahl K, et al. MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. Radiother Oncol 2011;98:330-4.
[32] Prabhakar R, Julka PK, Ganesh T, Munshi A, Joshi RC, Rath GK. Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. Jpn J Clin Oncol 2007;37:405-11.
[33] Stanescu T, Jans HS, Pervez N, Stavrev P, Fallone BG. A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of intracranial lesions. Phys Med Biol 2008;53:3579-93.
[34] Jonsson JH, Johansson A, Soderstrom K, Asklund T, Nyholm T. Treatment planning of intracranial targets on MRI derived substitute CT data. Radiother Oncol 2013;108:118-22.
[35] Rank CM, Hunemohr N, Nagel AM, Rothke MC, Jakel O, Greilich S. MRI-based simulation of treatment plans for ion radiotherapy in the brain region. Radiother Oncol 2013;109:414-8.
[36] Kjer H, Hansen RH, Paulsen RR, Edmund JM. Optimized Acquisition Parameters for MRI Only RT Using Ultrashort Echo Times. International Journal of Radiation Oncology Biology Physics 2012;84:S866-S.
[37] Held K, Kops ER, Krause BJ, Wells WM, Kikinis R, Muller-Gartner HW. Markov random field segmentation of brain MR images. IEEE Trans Med Imaging 1997;16:878-86.
[38] Zhang YY, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 2001;20:45-57.
[39] Van Leemput K, Maes F, Vandermeulen D, Suetens P. Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging 1999;18:897-908.
[40] Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via Em Algorithm. Journal of the Royal Statistical Society Series B-Methodological 1977;39:1-38.
[41] Breiman L. Random forests. Machine Learning 2001;45:5-32.
[42] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees: Wadsworth; 1984.
[43] 46 Ir. Photons, electrons, proton and neutron interaction data for body tissues. Bethesda, Maryland, USA1992.
[44] Dice LR. Measures of the Amount of Ecologic Association between Species. Ecology 1945;26:297-302.
[45] Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). Bethesda2011.
[46] Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of Normal Tissue Complication Probability Models in the Clinic. International Journal of Radiation Oncology Biology Physics 2010;76:S10-S9.
[47] Mayo C, Yorke E, Merchant TE. Radiation Associated Brainstem Injury. International Journal of Radiation Oncology Biology Physics 2010;76:S36-S41.
[48] Gay HA, Niemierko A. A free program for calculating EUD-based NTCP and TCP in external beam radiotherapy. Physica Medica 2007;23:115-25.
[49] Dale RG, Hendry JH, Jones B, Robertson AG, Deehan C, Sinclair JA. Practical methods for compensating for missed treatment days in radiotherapy, with particular reference to head and neck schedules. Clin Oncol 2002;14:382-93.
[50] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria2009.
[51] Crawley MJ. The R book. 1st ed: John Wiley & Sons; 2007.
[52] Hu Y, Green OP, Parikh P, Olsen J, Mutic S. Initial Experience with the ViewRay System - Quality Assurance Testing of the Imaging Component. Med Phys 2012;39:4013-.
[53] ICRU report 49. Stopping Powers and Ranges for Protons and Alpha Particles. Bethesda, MD, USA1993.
[54] Andreasen D, Andersen JL, Hansen RH, Van Leemput K, Edmund JM. The impact of a Dixon sequence in creating a pseudo CT scan from MR images using a Gaussian mixture regression model. 2nd ESTRO Forum; Geneva, Switzerland: ESTRO; 2013.

Paper B

Patch-based Generation of a Pseudo CT from Conventional MRI Sequences for MRI-only Radiotherapy of the Brain

Daniel Andreasen*
Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark and Department of Oncology, Radiotherapy Research Unit, Copenhagen University Hospital, Herlev, Denmark

Koen Van Leemput
Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark and A.A. Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, USA

Rasmus H. Hansen
Department of Radiology, Copenhagen University Hospital, Herlev, Denmark

Jon A.L. Andersen and Jens M. Edmund
Department of Oncology, Radiotherapy Research Unit, Copenhagen University Hospital, Herlev, Denmark

(Dated: February 27, 2015)

Purpose: In radiotherapy (RT) based on magnetic resonance imaging (MRI) as the only modality, the information on electron density must be derived from the MRI scan by creating a so-called pseudo computed tomography (pCT). This is a non-trivial task, since the voxel-intensities in an MRI scan are not uniquely related to electron density. To solve the task, voxel-based or atlas-based models have typically been used. The voxel-based models require a specialized dual ultrashort echo time (dUTE) MRI sequence for bone visualization and the atlas-based models require deformable registrations of conventional MRI scans. In this study, we investigate the potential of a patch-based method for creating a pCT based on conventional T_1-weighted MRI scans without using deformable registrations. We compare this method against two state-of-the-art methods within the voxel-based and atlas-based categories.
Methods: The data consisted of CT and MRI scans of five cranial RT patients. To compare the performance of the different methods, a nested cross validation was done to find optimal model parameters for all the methods. Voxel-wise and geometric evaluations of the pCTs were done. Furthermore, a radiologic evaluation based on water equivalent path lengths was carried out, comparing the upper hemisphere of the head in the pCT and the real CT. Finally, the dosimetric accuracy was tested and compared for a photon treatment plan.
Results: The pCTs produced with the patch-based method had the best voxel-wise, geometric and radiologic agreement with the real CT, closely followed by the atlas-based method. In terms of the dosimetric accuracy, the patch-based method had average deviations of less than 0.5% in measures related to target coverage.
Conclusions: We showed that a patch-based method could generate an accurate pCT based on conventional T_1-weighted MRI sequences and without deformable registrations. In our evaluations, the method performed better than existing voxel-based and atlas-based methods and showed a promising potential for RT of the brain based only on MRI.

I. INTRODUCTION

Magnetic Resonance Imaging (MRI) is increasingly being used in modern radiotherapy (RT) treatment planning in combination with Computed Tomography (CT). The main reason is the superior soft tissue contrast which improves target and organ at risk (OAR) definition in the brain and other sites as compared to CT-based delineations [1–3]. In order to transfer the MRI delineations to the CT, the two scans must be aligned. This is usually done by a manual and/or automatic rigid registration between the CT and MRI scan. It has been estimated that a mean cranial registration error of approximately 2 mm is introduced in this process [4, 5] with similar values for other body sites such as the prostate [6]. These errors produce a systematic shift in the delineations and may ultimately lead to target under-dosage or an increased dose to the adjacent OARs.

Basing the entire RT chain on MRI as the only modality, so-called MRI-only RT, would remove these systematic errors and reduce patient discomfort as well as lower the workload and financial cost. It is, however, non-trivial to exclude the CT, since the MRI images do not contain information about electron density which is needed for accurate dose calculations. Furthermore, with conventional MRI sequences, the signal from cortical bone is weak or non-existing due to its rapid

* Corresponding author email: dana@dtu.dk

transversal signal relaxation time (T_2) in the range 0.5-2 ms [7, 8]. This means that MRI voxels containing bone are indistinguishable from air, and that patient setup based on digitally reconstructed radiographs (DRRs) is unfeasible. In order to solve these problems, electron density assignment must be done based on the MRI, which can then be used for dose calculation and patient setup in an RT setting.

Two different approaches for automatic density assignment have typically been taken to derive a so-called pseudo CT (pCT) from MRI: voxel-based and atlas-based. In the voxel-based approach, a pCT is generated from individual voxel intensities in the MRI scan, either by segmentation and subsequent bulk density assignment [8–11] or using a regression model to predict a continuous-valued pCT [11–14]. Common for these methods is the need for a specialized dual ultrashort echo time (dUTE) MRI sequence that captures the signal from components with a short T_2 relaxation time. This makes bone voxels separable from air in the resulting MRI images [15]. However, adding this non-standard sequence is not only inconvenient, it also causes prediction errors in the pCT scans at tissue interfaces such as in the ear and nasal cavities. This has been attributed to partial volume and susceptibility effects as well as a low signal-to-noise ratio (SNR) of the dUTE scans [8, 16, 17].

Atlas-based methods, on the other hand, estimate pCTs using conventional (non-dUTE) MRI sequences, and must therefore compensate for the bone/air ambiguity in the MRI images. This is achieved using deformable registration of one or multiple atlases of paired MRI/CT scans to the patient MRI scan and then using the warped atlas CT scan(s) as a pCT estimate [18–20]. The pure atlas-based methods rely on a correct deformable registration between atlas and patient MRI. This can be both time consuming (in the case of multiple atlases) and problematic if the patient is dissimilar to the atlas [8]. To correct for deformation errors, the method can be extended with a subsequent step of pattern recognition using Gaussian process regression based on 2D patches [21]. Though this improves robustness, the gain may be marginal compared to the added computational complexity [20].

In the field of automated brain MRI segmentation, patch-based segmentation methods have recently been proposed with promising results [22, 23]. In these approaches, 3D patches (i.e., small cuboidal image subregions) are extracted from the MRI and a spatially local search for the most similar patches in a pre-acquired database of labeled MRI scans is performed. The known labels of the resulting database patches are then fused to give the predicted label at each position. To facilitate the spatially local patch search, most patch-based methods use a rough linear alignment between the database MRI scans and the MRI to be segmented. The need for an accurate deformable registration is thus removed and the segmentation is driven mainly by patch similarities.

Based on these results, in this study we investigate the potential of patch-based methods for predicting a continuous-valued pCT of the brain based on T_1-weighted MRI scans. In particular, we incorporate the CT numbers as label information in the patch model to enable patch-based regression. We compare this method with two state-of-the-art methods in voxel-based and atlas-based pCT prediction, namely (1) Gaussian mixture regression (GMR) based on dUTE scans [12], and (2) multi-atlas information propagation based on T_1-weighted scans [19]. We perform a voxel-wise, geometric and radiologic evaluation as well as a dosimetric evaluation for a photon treatment plan.

II. MATERIALS AND METHODS

A. Imaging

The data used in this study consisted of MRI and CT scans of five whole brain RT patients, three male and two female, aged 55-82 years. The study was approved by the Capital Regional Ethics Committee (protocol number H-3-2011-107). The CT scans were acquired on a Philips Brilliance Big Bore CT with a voxel resolution $0.6 \times 0.6 \times 2$ mm, $512 \times 512 \times 110$ voxels using a standard protocol for brain scans (120 kV, 300 mAs). The MRI scans were acquired on a Philips Panorama 1 T open scanner. The sequences were a T_1-weighted 3D Fast Field Echo (FFE), TE/TR = 6.9/25 ms, voxel resolution $0.85 \times 0.85 \times 1.2$ mm, $188 \times 188 \times 152$ voxels and two dUTE scans at flip angles 10° and 25° with $TE_1/TE_2/TR = 0.09/3.5/7.1$ ms and an isotropic voxel resolution of 1 mm with 256^3 voxels. The scan time of a dUTE scan at one flip angle was approximately 7 minutes. The patients were fixed in treatment position during both the MRI and CT scanning using thermoplastic masks. Informed consent was obtained from all patients prior to acquiring the MRI scans additional to their standard imaging. Each patient's MRI/CT pair was rigidly aligned using mutual information with the default settings as implemented in 3D Slicer [24]. The CT scans were resliced and cropped to match the resolution and field of view of the MRI scans, thus generating dUTE/CT scan pairs at the dUTE scan resolution and T_1-weighted/CT scan pairs at the T_1-weighted scan resolution.

B. Patch-based pCT Prediction

In order to facilitate an intensity-based similarity search, the tissues should have a consistent intensity throughout all T_1-weighted scans. To achieve this, a histogram-matching approach was used with two linear

mapping segments separated at the average median intensity [25].

A patch, $P(x)$, was defined as a cube with side length m voxels centered on the spatial location x in an MRI image. Similarly, a target value, $T(x)$, was defined for each $P(x)$ as the Hounsfield unit (HU) value at x in the corresponding rigidly aligned CT image. For S patients, corresponding patches and target values were extracted for all positions, x, to create a database of patches, $P_s(x)$, with corresponding target values, $T_s(x)$, where s denotes one of S patients. Using this database, a pCT for a test patient was predicted by extracting patches from his/her MRI scan and doing an intensity-based nearest neighbor search in the patch database. For a patch at position y in the MRI of the test patient, the similarity measure was defined as:

$$d(s,x) = ||P(y) - P_s(x)||_2^2, \tag{1}$$

where $d(s,x)$ denotes the squared L_2-norm between $P(y)$ and $P_s(x)$. The search then consisted of finding the database patch that minimized $d(s,x)$ and storing the corresponding patient and spatial position:

$$(s^{min}, x^{min}) = \underset{s,x}{\operatorname{argmin}}\ d(s,x). \tag{2}$$

The search was extended to find the K most similar database patches and the K relevant target values were extracted from the database as $T_{s_k^{min}}(x_k^{min})$ with $k = 1,..,K$. To assign a pCT HU value, a similarity-weighted average was computed:

$$\text{pCT}(y) = \frac{\sum_k w_k \times T_{s_k^{min}}(x_k^{min})}{\sum_k w_k}, \tag{3}$$

with weights defined as:

$$w_k = \exp\left(\frac{-d(s_k^{min}, x_k^{min})}{\min_k d(s_k^{min}, x_k^{min})}\right). \tag{4}$$

This ensured that if one patch was more similar than the rest, this patch would be weighted highly; conversely, if all patches were equally similar, they would be weighted equally.

To ensure that the local similarity search was indeed local, and to limit the number of similarity comparisons required per patch, we applied the search space reduction and patch pre-selection method as described by Coupé *et al.* [22]. An affine alignment of each database MRI and the patient MRI was performed and the search was constrained to only contain database patches in a local cubiodal search volume, V_{search}, of side length v voxels around y, such that $x \in V_{search}$ in Equation 2. The structural similarity measure (SSIM) [26] was used to discard highly dissimilar patches in the local neighborhood. This is a computationally simple similarity measure based on the mean and variance in each patch. All patches with SSIM < 0.95 were discarded prior to the patch search performed in Equation 2. If all patches were discarded in a search volume, the pCT value in the affected voxel was flagged as unknown. In post-processing, such voxels were assigned the average pCT value of the closest assigned voxels. For the intensity fusion in Equation 3, the $K = 8$ most similar patches were used unless the SSIM only allowed a smaller number. An illustration of the patch-based pCT generation is shown in Figure 1 for a simplified 2D case – the actual algorithm works in 3D.

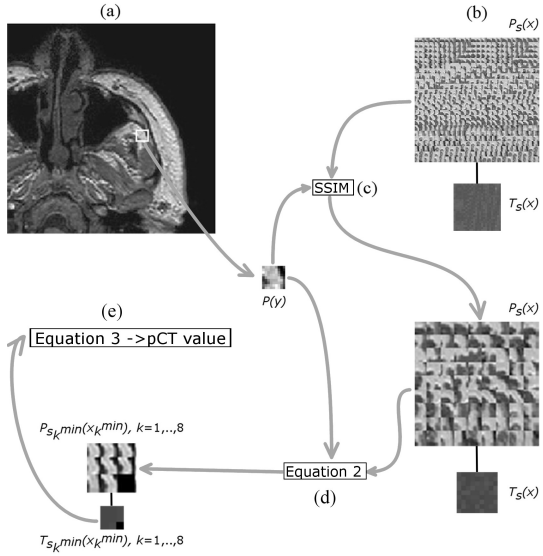


Figure 1. 2D patch-based pCT prediction. (a) A patch, $P(y)$, is extracted from the test MRI at position y. (b) A database of spatially close patches, $P_s(x)$ stored with their corresponding CT value, $T_s(x)$. (c) Using the structural similarity measure (SSIM), highly dissimilar patches are discarded from the database. (d) Using Equation 2, the 8 most similar database patches, $P_{s_k^{min}}(x_k^{min})$, and their CT values, $T_{s_k^{min}}(x_k^{min})$, are found from the remaining database. (e) Using Equation 3, the CT values, $T_{s_k^{min}}(x_k^{min})$, are combined to produce the final pCT value at position y.

C. Comparison of Algorithms: GMR and Multi-atlas

The patch-based method was compared with Gaussian mixture regression based on dUTE scans[12] and multi-atlas information propagation based on T_1-weighted scans[19]; two state-of-the-art techniques for pCT prediction within the voxel-based and atlas-based categories, respectively.

The multi-atlas method consists of deformable reg-

istrations of multiple atlases of MRI/CT pairs to the patient MRI. A multi-resolution B-spline transform in Elastix [27] was used for this purpose. Mutual information with 32 histogram bins was used as a metric and the final control point spacing was set to 5 mm. The local normalized cross correlation (LNCC) between the patient MRI and the atlas MRI was used to determine a voxel-wise ranking of each atlas CT before finally fusing the HU values. A high LNNC means a high degree of local similarity between the patient MRI and the deformed atlas MRI and thus results in a high ranking. Two parameters were involved in these steps; namely a Gaussian kernel width, σ_{gk}, controlling the local neighborhood size involved in the LNCC calculation, and a weight constant, β, controlling the decay of an exponential function used when converting the LNCC ranking to a fusing weight.

For the GMR method, the joint distribution of MRI and CT intensities was estimated as a mixture of multivariate Gaussian distributions using expectation maximization (EM) [28]. Using this model, the expected value of the CT conditioned on newly observed MRI values can be found and used as a pCT prediction. Input images in addition to the dUTE scans were created using mean and standard deviation filters on the dUTE scans. A mask was then created to exclude air surrounding the head from the model training data. 50 independent runs of k-means clustering with k-means++ initialization [29] were performed on the training data to estimate the initial values of the means, covariance matrices and mixing proportions. The values resulting in the minimum total energy were chosen as initialization for the EM algorithm.

D. Model Optimization

In order to ensure a fair comparison between the different methods, a nested cross validation scheme was used to find the optimal parameters for all methods. This meant that for each test patient, leave-one-out cross validation (LOOCV) was performed cyclically on the remaining four patients. In each of these four LOOCV folds, three patients were used to predict pCTs of the fourth using all possible combinations of a manually defined subset of model parameters. For each parameter configuration, the voxel-wise mean absolute error (MAE_{vox}, as defined later in Equation 5) was used to evaluate the pCT. The parameter configuration that resulted in the average best performance across the four folds were then chosen as the optimal for that test patient and was used to predict his/her pCT using the remaining four patients. This scheme is well suited for model comparisons since each test patient is not involved in the parameter optimization, which gives an almost unbiased estimate of the true prediction error[30]. For the patch-based method, the optimal patch side length was found to be $m = 5$ or $m = 7$ voxels (depending on the patient) and the optimal search volume side length was $v = 15$ voxels, yielding a maximum of $15^3 \times 4 = 13500$ database patches to search through for each pCT voxel. Note that the value of K, used in Equation 3 was set empirically to $K = 8$ and therefore not optimized. For the multi-atlas method, parameters for the LNCC ranking scheme were optimized. The Gaussian kernel width, σ_{gk}, was 5–9 voxels depending on the patient, and the weight constant was $\beta = 0.9$ for all patients. Note that the deformable registration parameters were chosen empirically and thus were not included in the nested cross validation. For the GMR method, the number of Gaussians to use in the model was found and set to 20–23, again depending on the patient.

E. Geometric Evaluation

The pCTs were compared in terms of the voxel-wise mean absolute error MAE_{vox} in the head region (excluding surrounding air):

$$\text{MAE}_{\text{vox}} = \frac{1}{N}\sum_{i=1}^{N}|CT(i) - pCT(i)|, \tag{5}$$

where N is the total number of voxels in the head region. To reveal in which tissue regions errors were present, the MAE_{vox} in bins of 20 HU across the HU scale was also calculated. To determine whether the predictions were biased towards an underestimation or overestimation of the real CT number, the mean voxel-wise error was calculated:

$$\text{ME}_{\text{vox}} = \frac{1}{N}\sum_{i=1}^{N}[CT(i) - pCT(i)]. \tag{6}$$

To evaluate the correctness of the pCT bone geometry, the Dice similarity coefficient (DSC)[31] of bone was calculated:

$$\text{DSC} = \frac{2(V_{bone}^{pCT} \cap V_{bone}^{CT})}{V_{bone}^{pCT} + V_{bone}^{CT}}, \tag{7}$$

where V_{bone}^{pCT} and V_{bone}^{CT} are the volumes of bone in the pCT and the real CT, respectively. We defined the bone volume as all voxels with a value > 200 HU giving a density of 1.14 g/cm^3 according to our CT calibration curve. DSC = 1 means complete overlap between the volumes, and DSC = 0 means no overlap.

The MAE_{vox}, ME_{vox} and DSC are pure voxel-wise and geometric measures of the pCT accuracy. To provide a radiologic error measure, an evaluation based on the water equivalent path length (WEPL) was introduced [10, 32]:

$$l' = \sum_i \Delta l_i \times \rho_i, \tag{8}$$

where Δl_i is the physical path length of voxel i, and ρ_i is a radiological scaling factor that depends on the type of

radiation and tissue. For MeV photons, it is the electron density relative to water. The value of ρ_i was found from the voxel CT number using a verified standard calibration curve in the treatment planning system. To compare WEPLs in the CT and pCT, a common point in both was defined at the level of the nasal cavity centered in the head. A sphere with its center in that point was defined, covering the entire head. The WEPL was then calculated in radial spokes from the center point towards the edge of the sphere. When the spokes traversed the edge of the head, the WEPL calculation was terminated so the length of each spoke varied with the patient anatomy. The spokes were defined in spherical coordinates with the center point as origin and with an angular spacing of $2\pi/60$ in both the polar and azimuthal angles. All polar angles between 0.5π and 1.5π were excluded from the calculation, resulting in a volume including only the upper hemisphere of the head, covering most of the brain (see Figure 2). In total this yielded $60 \times 31 = 1860$ spokes. The tissue was sampled every $\Delta l = 0.02$mm along each spoke. The CT number was then found at each sample by trilinear interpolation. To measure the difference in WEPL between the CT and pCT, the mean absolute WEPL error ($\mathrm{MAE_{WEPL}}$) was defined as:

$$\mathrm{MAE_{WEPL}} = \frac{1}{L}\sum_{j=1}^{L} |l_j'^{CT} - l_j'^{pCT}|, \tag{9}$$

where L is the number of spokes and l_j' is the WEPL of the jth spoke. Similarly, the mean WEPL error ($\mathrm{ME_{WEPL}}$) was defined as:

$$\mathrm{ME_{WEPL}} = \frac{1}{L}\sum_{j=1}^{L} l_j'^{CT} - l_j'^{pCT}. \tag{10}$$

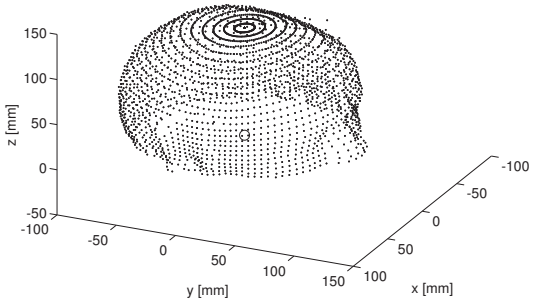


Figure 2. The WEPL was calculated along spokes (not shown) from the center point (open circle) at $(x, y, z) = (0, 0, 0)$ towards all points on the surface of the patient (dots). Along the spokes the CT number was sampled and converted to relative electron density for WEPL calculation.

F. Dosimetric Evaluation

A 6 MV photon treatment plan for two different spherical planning target volumes (PTVs) of 3 cm in diameter was created for all the patients. The first plan used a PTV positioned anterior to the center of the cerebrum (PTV 1, Figure 3) with two 15° wedged lateral opposing fields and one anterior field, all equally weighted. The second plan had a PTV positioned behind the nasal cavity (PTV 2, Figure 3) with four equally weighted fields, two lateral opposing and two anterior/posterior opposing. For all fields, the apertures were cropped to the PTV plus a 0.5 cm margin. The PTVs were chosen to represent theoretically easy and difficult cases, respectively, with PTV 1 positioned in a homogeneous part of the brain and PTV 2 having a large degree of heterogeneity in the tissue composition of its surroundings. The treatment planning was carried out on each patient's pCT in Eclipse v11.0 (Varian Medical Systems Inc., Palo Alto, CA), prescribing a mean dose of 2 Gy in 30 fractions ($\mathrm{D}_{pre} = 60$ Gy) to the PTV and calculating the dose distribution using the analytical anisotropic algorithm (AAA). The plan was then transferred to the real CT and re-calculated using the same plan parameters and monitor units. The PTV dose volume histogram (DVH) was used to evaluate the dosimetric difference between CT and pCT. The percentage point deviation was calculated for the DVH points relevant to PTV coverage, i.e. the near-minimum ($\mathrm{D}_{98\%}$), near-maximum ($\mathrm{D}_{2\%}$) and median (D_{median}) absorbed dose [33]. Furthermore, a normalized mean absolute dosimetric error ($\mathrm{nMAE_{dos}}$) was calculated as:

$$\mathrm{nMAE_{dos}} = \frac{1}{M}\sum_{i=1}^{M} \frac{|\mathrm{D}_{vol(i)}^{CT} - \mathrm{D}_{vol(i)}^{pCT}|}{\mathrm{D}_{pre}} \tag{11}$$

where M is the total number of DVH points between $\mathrm{D}_{100\%}$ and $\mathrm{D}_{0\%}$ in dose increments of 0.1 Gy, and $\mathrm{D}_{vol(i)}$ is the accumulated dose in Gy given to vol(i)% of the volume in either the CT or pCT. Similarly, the normalized mean dosimetric error ($\mathrm{nME_{dos}}$) was calculated as:

$$\mathrm{nME_{dos}} = \frac{1}{M}\sum_{i=1}^{M} \frac{\mathrm{D}_{vol(i)}^{CT} - \mathrm{D}_{vol(i)}^{pCT}}{\mathrm{D}_{pre}} \tag{12}$$

$\mathrm{nMAE_{dos}}$ explains the magnitude of the errors and $\mathrm{nME_{dos}}$ reveals if the errors are biased towards underestimation or overestimation. The combined dosimetric score (CDS) is a combination which penalizes bias in the errors [11]:

$$\mathrm{CDS} = \mathrm{nMAE_{dos}} + |\mathrm{nME_{dos}}| \tag{13}$$

III. RESULTS

In Figure 4, transverse slices of the real CT and predicted pCTs are shown for the different methods. Visually, the results based on $\mathrm{T_1}$-weighted MRI in (c) and (d)

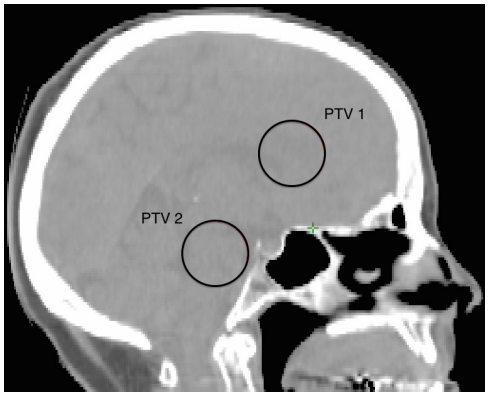


Figure 3. Sagittal view of the two PTVs used in the dosimetric evaluation.

are better than the dUTE-based in (b), especially in the nasal cavities.

Figure 5 shows the MAE_{vox} calculated as a function of the real CT value in bins of 20 HU and averaged for the 5 patients. In general, GMR shows the highest errors in the bone region ($>$ 200 HU), fat region (approximately $[-100; -50]$ HU) and the region between air (-1000 HU) and fat tissue. This HU range is dominated by values of partial volumes effects in the real CT. The patch-based and multi-atlas methods have similar performance in most regions. Table I shows the average voxel-wise, geometric and radiologic errors for the different methods. On average, the patch-based and multi-atlas methods have the lowest MAE_{vox} and highest DSC, with a slight favor for the patch-based method. Looking at the ME_{vox}, the patch-based and multi-atlas methods have average values closer to 0, indicating no consistent pattern in the errors. However, given the magnitude of the standard deviation for GMR, $\mathrm{ME}_{vox} = 0$ is also a plausible value for this method. Looking at the WEPL evaluation in Table I, the patch-based and multi-atlas methods have the smallest $\mathrm{MAE}_{\mathrm{WEPL}}$, with slightly lower values for the patch-based method.

In Figure 6, the results of the dosimetric evaluation are shown. For all methods, the metrics show average smaller errors, i.e. values closer to 0, for PTV 1 compared to PTV 2, illustrating the more challenging position of PTV 2. The difference in performance between the methods is also generally smaller for PTV 1 than for PTV 2. Looking at PTV 1, the nMAE_{dos} is similar for the patch method and GMR, but the nME_{dos} reveals that GMR has a bias towards overestimating the dose. The multi-atlas method has the largest nMAE_{dos}, with a bias (nME_{dos}) towards underestimating the dose. Looking at the coverage of PTV 1 (lower panels), the patch-based method has an average deviation close to 0%. The other methods have deviations $< 0.5\%$. When looking at PTV 2, the average nMAE_{dos} is lowest for the

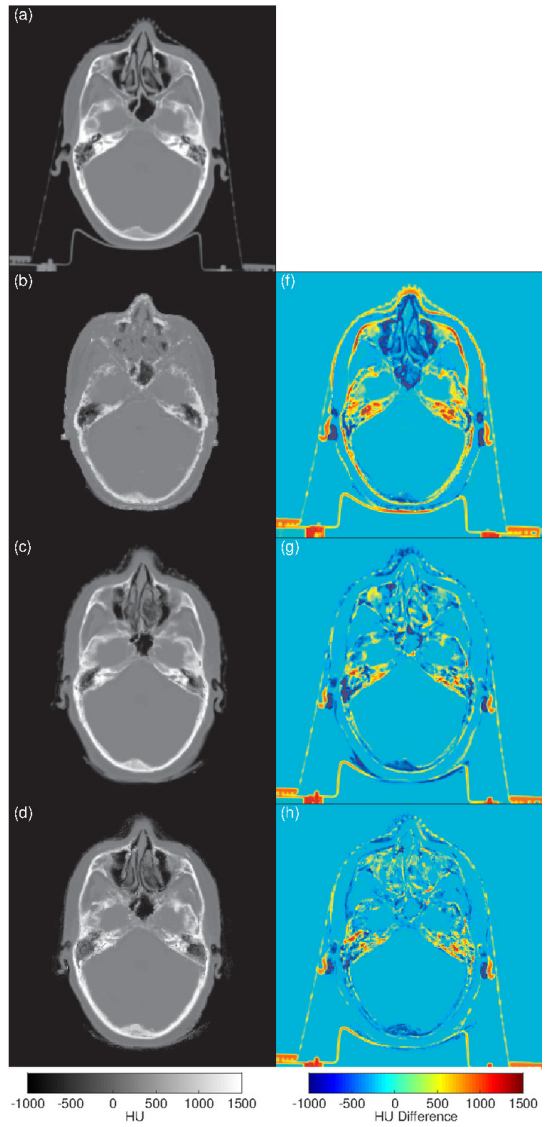


Figure 4. Transverse slices for comparison of pCTs with real CT. (a)-(d) shows the real CT, the GMR pCT, the multi-atlas pCT and the patch-based pCT, respectively. (f)-(h) shows the difference maps between the real CT and the GMR, the multi-atlas and the patch-based pCTs, respectively. Negative values indicate an overestimation of the HU value and positive values indicate an underestimation.

patch-based method, with GMR and multi-atlas having higher values and larger standard deviation. We observe that both the GMR and multi-atlas methods have a bias towards underestimating the dose as seen in the upper

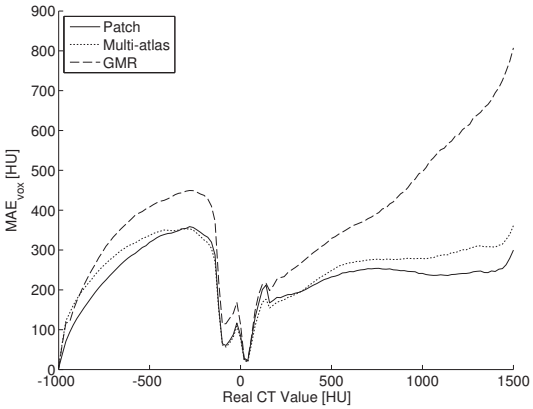


Figure 5. MAE_{vox} calculated in bins of 20 HU and averaged across the five patients.

	Patch-based	**Multi-atlas**	**GMR**
MAE_{vox} [HU]	85 ($\sigma = 14$)	97 ($\sigma = 19$)	148 ($\sigma = 22$)
$\mathrm{ME}_{\mathrm{vox}}$ [HU]	1 ($\sigma = 14$)	-4 ($\sigma = 17$)	22 ($\sigma = 28$)
DSC	0.84 ($\sigma = 0.02$)	0.83 ($\sigma = 0.01$)	0.67 ($\sigma = 0.03$)
$\mathrm{MAE}_{\mathrm{WEPL}}$ [mm]	2.2 ($\sigma = 1.0$)	2.7 ($\sigma = 0.8$)	4.8 ($\sigma = 1.3$)
$\mathrm{ME}_{\mathrm{WEPL}}$ [mm]	0.4 ($\sigma = 1.8$)	-0.6 ($\sigma = 1.9$)	1.1 ($\sigma = 2.1$)

Table I. The voxel-wise and geometric quality measures: Mean absolute voxel-wise error (MAE_{vox}), mean voxel-wise error (ME_{vox}) in HU and Dice similarity coefficient of bone volume (DSC). Radiologic measures: Mean absolute WEPL error ($\mathrm{MAE}_{\mathrm{WEPL}}$) and mean WEPL error ($\mathrm{ME}_{\mathrm{WEPL}}$) of the water equivalent path lengths (in mm). Average value and standard deviation (σ) for the five patients are shown.

middle panel. With regards to the dose coverage (lower panels), the patch-based method has the smallest average deviations, except for $\mathrm{D}_{98\%}$, where GMR has a slightly smaller deviation. On average, multi-atlas has a better performance than GMR in D_{median} and $\mathrm{D}_{2\%}$.

IV. DISCUSSION

In this paper, we evaluated a patch-based method for predicting brain pCTs based on conventional T_1-weighted MRI images. The method required no deformable registrations and was shown to yield comparable or better results than existing methods using Gaussian mixture regression on dUTE scans or multi-atlas information propagation on T_1-weighted scans.

In terms of the MAE_{vox} curves shown in Figure 5, the dUTE-based method showed the largest voxel-wise errors in most tissue regions of the brain. The errors we

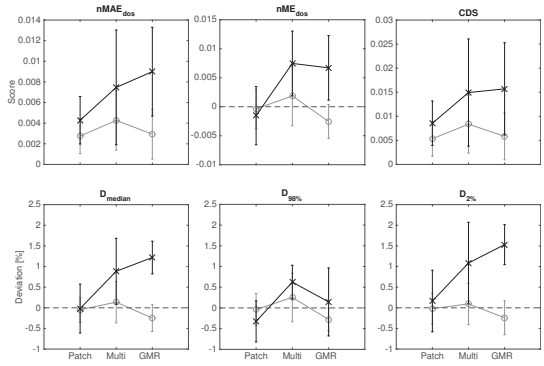


Figure 6. Dosimetric errors for the two PTVs. PTV 1: Gray lines and open circles. PTV 2: Black lines and crosses. Dashed line indicates zero. Top row: normalized mean absolute dosimetric error ($\mathrm{nMAE}_{\mathrm{dos}}$), normalized mean dosimetric error ($\mathrm{nME}_{\mathrm{dos}}$) and combined dosimetric score (CDS). Bottom row: percentage point deviation in D_{median}, $\mathrm{D}_{98\%}$ and $\mathrm{D}_{2\%}$. Average values are shown along with $\pm\sigma$ interval.

observed in the fat region could be caused by the data being recorded on a 1 T scanner where water/fat phase cancellation occurs approximately at 3.5 ms after excitation, i.e. at TE_2 of the dUTE sequence. In a previous study, we tested if adding a Dixon sequence as GMR model input would aid in discriminating fat/water and bone [34]. Though this improved bone predictions, it did not improve predictions in fat voxels. Rank *et al.* reported errors in pCT bone prediction due to fatty-tissue appearing hyper-intense in dUTE scans acquired at 3 T [10]. This suggests that issues in the fat region may be independent of field strength and not due to phase cancellations.

In the bone region, the two methods based on the T_1-weighted MRI had the smallest errors, even though bone is not uniquely defined in terms of intensity in these images. The deformable registration used in the multi-atlas method can compensate for this lack of information, assuming a successful alignment of the atlas MRIs and the patient MRI. In the present study, this assumption seemed to hold, but as implied earlier, this may not always be the case if the anatomical variation is large. Furthermore, the deformable registration introduces another set of parameters to adjust, and it can be hard to find one configuration that provides a successful registration for all patients. Here, we tested a few parameter settings and chose the one that resulted in the average lowest MAE_{vox}. For a more unbiased estimate of the prediction performance of the multi-atlas method, the registration parameters could be included in the nested cross-validation.

The patch-based method, on the other hand, achieved a lower error in the bone region without a deformable registration. Instead, the neighborhood information con-

tained in each patch in combination with the constrained search volume ensured that bone and air patches were not confused.

In terms of the voxel-wise errors and DSC presented in Table I, the patch-based method had the best performance. Johansson *et al.* reported an average MAE_{vox} of 137 HU for five patients using the GMR method [12], which is within the standard deviation of our GMR findings. The GMR results could potentially be improved by adding spatial information to the Gaussian mixture model or using different MRI reconstruction techniques [16, 35]. This was not investigated further in our study. For the multi-atlas method, Burgos *et al.* reported an average MAE_{vox} of 102 HU ($\sigma = 10$) for 7 patients [19], which also agrees with our findings. We included the DSC measure since the bone geometry is important for generating DRRs. As was shown in Figure 5, the patch and multi-atlas methods have the highest prediction accuracy in the bone region, which is also reflected in the DSC results. This suggests that these methods are better suited for DRR generation. In a qualitative evaluation, Jonsson *et al.* described an overall acceptable quality of DRRs generated on the basis of GMR pCTs, except in the nasal cavity and sphenoidal sinuses [36]. From our results, it seems that the pCTs based on T_1-weighted MRI visually have a better quality in this region. However, the quality of the resulting DRRs and their potential for treatment setup was not investigated further in the present study. A factor that could influence the pCT predictions is the non-uniform intensity variations intrinsic to MRI scans. In the present study, we did not apply a correction for this. From a visual inspection, the acquired MRI scans showed only minor non-uniform intensity variations in the imaged volume, which is probably due to the relatively low field strength. For the patch-based method, applying a bias field correction did not improve predictions. It could potentially improve predictions of the other methods, however, and it will be an important step for the patch-based method at higher field strengths and/or larger fields of view.

The WEPL evaluation was introduced to provide a radiologically more relevant error measure than the pure voxel-wise errors. It takes into account the piece-wise linear relationship between the HU and electron density and also provides a simplistic imitation of the treatment simulation were radiation encounters several tissues when sent through the patient. The errors in WEPL are dependent on the position of the center point and the direction of the radial spokes. Here, we chose to evaluate the whole upper hemisphere of the head to produce an average estimate for all possible planning scenarios. Maybe, due to this averaging, the results of the WEPL evaluation did not diverge from the trends observed in the voxel-wise and geometric evaluation.

For the dosimetric evaluation, an easy and a challenging PTV site was chosen, which was reflected in the magnitude of the errors in Figure 6. Overall, the patch-based method had the lowest average errors, following the trend from the other evaluations. Looking at the CDS, GMR performed well and was on par with the patch-based method for PTV 1 and with multi-atlas for PTV 2. We obtained a comparable value of around 0.02 for the CDS of GMR for a target similar to PTV 2 in a previous study [11]. It should be noted that since the CDS only accounts for the magnitude of nME_{dos}, the direction of the error is removed. When averaging across patients, this means that a method with consistent overestimation of the dose can get the same CDS as a method where the direction of error is random. This effect explains why the patch method and GMR have similar CDS for PTV 1, even though the patch method performs better in nME_{dos}.

In $D_{98\%}$, GMR performed well for PTV 2 but with a larger standard deviation. In a dosimetric evaluation of the GMR method, Jonsson *et al.* reported a percentage point deviation of 0.86% in $D_{90\%}$[36] for a target somewhat similar to PTV 2 in the present study. Though $D_{90\%}$ is less sensitive to changes in the DVH shape, it seems that the dosimetric error is of the same order as our findings in $D_{98\%}$.

A statistical criterion for the reliable use of MRI-only states that for 95% of the patients, the maximum uncertainty in the DVH points related to target coverage, should be within 2% [37]. Although a study with more patients is still needed, we note that for PTV 1, all methods fulfilled this criterion. This speaks in favor of a reliable use of MRI-only for targets positioned away from cavities. We also note that the patch-based method fulfilled the criterion for the challenging PTV 2, showing potential for a more general use of MRI-only RT.

From the voxel-wise and radiologic evaluation, we expected that the multi-atlas method would have dosimetric errors closer to those of the patch-based method. This did not seem to be the case, especially for PTV 2, where the multi-atlas and GMR methods were closer in performance. Therefore, even though the voxel-wise and radiologic error may serve as a proxy for the dosimetric performance, they should be accompanied by a dosimetric evaluation to get the full picture. This agrees with our previous findings[11].

Comparing the two methods based on T_1-weighted MRI, the voxel-wise and radiologic differences were small with a slight favor of the patch-based method. In terms of dosimetry, the patch-based method had an average better performance, especially in PTV 2. The advantage of both methods is that they work on any MRI sequence as long as an atlas or database of MRI/CT pairs has been obtained. This means that they can be adapted to the clinical practice without the need for extra sequences and scan time. The patch-based method further has the advantage that it relies only on linear registrations to provide a rough alignment of the database MRI and patient MRI. Linear registrations can be performed faster than deformable ones, which could potentially make the patch-based method faster than the multi-atlas method. In the current Matlab implementation, however, using a brute force search for the most

similar patches, it took approximately 15 hours to predict a pCT with 4 database patients. On the same hardware, the multi-atlas and GMR methods took roughly 36.5 minutes (including deformable registrations) and 6.5 minutes, respectively. None of the implementations were optimized for speed, and especially the patch-based method could be improved substantially by using an approximate nearest neighbor (ANN) patch search algorithm such as OPAL [38]. A subsequent implementation of an ANN algorithm with similarities to OPAL, reduced the pCT prediction time for one patient to 38 minutes with an increase in MAE_{vox} of 9 HU. We believe this can be further improved, but it was not the focus of the present study.

The patch-based method is still dependent on anatomical similarity, even without the deformable registration. Indeed, in smaller regions of the brain in one patient, we found that all patches in the database were discarded due to the SSIM being below the threshold value. These dissimilar regions could be present in the MRI of patients with tumors or other brain abnormalities that are not found in the patch database. This poses a problem for the method. However, based on the SSIM, we were able to produce a pCT with problematic (dissimilar) regions marked for quality assurance. This is in contrast to the multi-atlas approach, where it can be hard to know where the deformable registrations were unsuccessful without manual inspection of each atlas. In the cases where no patches matched, we assigned the average pCT value of the neighboring assigned voxels, which did not seem to affect the accuracy of the pCT. Still, this way of handling the problem may prove too simple in the general case, especially if the dissimilar regions are large or if the tissue type of the neighboring voxels differs from that of the region. Increasing the size of the patient database could potentially alleviate some of this issue, but a focus of our future work is to incorporate a more advanced system for outlier handling.

As mentioned in the introduction, a pattern recognition step using Gaussian process regression (GPR) based on 2D patches has previously been investigated for improving an atlas-based method using deformable registrations [21]. The presented patch-based method shares some similarities with the GPR method since both are examples of so-called kernel smoothers. The difference lies in the way the kernel is defined, which is done explicitly in Equation 4 for the patch-based method and implicitly through a covariance function in the GPR method. Furthermore, the patch-based method uses linear registrations so the assumption of exact spatial correlation between the patient and database is not met. Therefore, the position of a patch does not affect its contribution weight, which is the case in the GPR method. Lastly, the patch-based method adaptively normalizes the kernel for each test patch with the minimum L_2-norm in Equation 4.

An average MAE_{vox} of 100.7 HU was reported with the GPR method on T_1-weighted scans for 17 patients [21], which is close to our results using the multi-atlas or patch-based methods. It was not reported whether this value was calculated for the entire image volume or only for the head region, as in the present study.

Overall, the methods for pCT prediction based on similarities in conventional MRI scans seem promising for MRI-only RT. Inter-patient anatomical variability does introduce an uncertainty in the pCTs, but the higher voxel-wise, geometric and dosimetric accuracy compared to current voxel-based methods may make this a worthwhile compromise. A larger study of the robustness of the patch-based method and the dosimetric uncertainty is still needed and is part of our future work.

V. CONCLUSION

In this study, we showed that a patch-based method could generate a pCT based on a conventional T_1-weighted MRI sequence without using deformable registrations or special dUTE sequences. We demonstrated a competitive performance of the method in several quality measures when compared to state-of-the-art atlas-based and voxel-based methods. In terms of dosimetric accuracy, the patch-based method showed a promising potential for use in MRI-only RT of the brain including PTVs positioned in challenging regions.

ACKNOWLEDGMENTS

This work was supported by a research grant from Varian Medical Systems, Inc.

[1] Prabhakar, R., Haresh, K., Ganesh, T., Joshi, R., Julka, P., Rath, G.: Comparison of computed tomography and magnetic resonance based target volume in brain tumors. Journal of Cancer Research and Therapeutics **3**(2) (2007) 121–123

[2] Ahmed, M., Schmidt, M., Sohaib, A., Kong, C., Burke, K., Richardson, C., Usher, M., Brennan, S., Riddell, A., Davies, M., Newbold, K., Harrington, K.J., Nutting, C.M.: The value of magnetic resonance imaging in target volume delineation of base of tongue tumours – a study using flexible surface coils. Radiotherapy and Oncology **94**(2) (2010) 161 – 167 Selected papers from the 10th Biennial ESTRO Conference on Physics and Radiation Technology for Clinical Radiotherapy.

[3] Rasch, C., Steenbakkers, R., van Herk, M.: Target definition in prostate, head, and neck. Seminars in Radiation Oncology **15**(3) (2005) 136 – 145

[4] Cattaneo, G.M., Reni, M., Rizzo, G., Castellone, P., Ceresoli, G.L., Cozzarini, C., Ferreri, A.J.M., Passoni, P., Calandrino, R.: Target delineation in post-operative radiotherapy of brain gliomas: Interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans. Radiotherapy and Oncology **75**(2) (2005) 217 – 223

[5] Ulin, K., Urie, M.M., Cherlow, J.M.: Results of a multi-institutional benchmark test for cranial CT/MR image registration. International Journal of Radiation Oncology*Biology*Physics **77**(5) (2010) 1584 – 1589

[6] Nyholm, T., Nyberg, M., Karlsson, M., Karlsson, M.: Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. Radiation Oncology **4**(1) (2009) 54

[7] Reichert, I.L., Robson, M.D., Gatehouse, P.D., He, T., Chappell, K.E., Holmes, J., Girgis, S., Bydder, G.M.: Magnetic resonance imaging of cortical bone with ultrashort TE pulse sequences. Magnetic Resonance Imaging **23**(5) (2005) 611 – 618

[8] Keereman, V., Fierens, Y., Broux, T., Deene, Y.D., Lonneux, M., Vandenberghe, S.: MRI-Based Attenuation Correction for PET/MRI Using Ultrashort Echo Time Sequences. J Nucl Med **51** (2010) 812–818

[9] Berker, Y., Franke, J., Salomon, A., Palmowski, M., Donker, H.C., Temur, Y., Mottaghy, F.M., Kuhl, C., Izquierdo-Garcia, D., Fayad, Z.A., Kiessling, F., Schulz, V.: MRI-Based Attenuation Correction for Hybrid PET/MRI Systems: A 4-Class Tissue Segmentation Technique Using a Combined Ultrashort-Echo-Time/Dixon MRI Sequence. Journal of Nuclear Medicine **53**(5) (2012) 796–804

[10] Rank, C.M., Hünemohr, N., Nagel, A.M., Röthke, M.C., Jäkel, O., Greilich, S.: MRI-based simulation of treatment plans for ion radiotherapy in the brain region. Radiotherapy and Oncology **109**(3) (2013) 414 – 418

[11] Edmund, J.M., Kjer, H.M., Van Leemput, K., Hansen, R.H., Andersen, J.A., Andreasen, D.: A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times. Physics in Medicine and Biology (2014) In press.

[12] Johansson, A., Karlsson, M., Nyholm, T.: CT Substitute Derived from MR Sequences with Ultrashort Echo Time. Medical Physics **38**(5) (2011) 2708–2714

[13] Johansson, A., Karlsson, M., Yu, J., Asklund, T., Nyholm, T.: Voxel-wise uncertainty in CT substitute derived from MRI. Med. Phys. **39**(6) (2012) 3283–3290

[14] Andreasen, D.: Creating a pseudo-CT from MRI for MRI-only based radiation therapy planning. Master's thesis, Technical University of Denmark, DTU Compute, E-mail: compute@compute.dtu.dk, Matematiktorvet, Building 303-B, DK-2800 Kgs. Lyngby, Denmark (2013)

[15] Rahmer, J., Blume, U., Börnert, P.: Selective 3D ultrashort TE imaging: comparison of "dual-echo" acquisition and magnetization preparation for improving short-T2 contrast. Magnetic Resonance Materials in Physics, Biology and Medicine **20**(2) (2007) 83–92

[16] Johansson, A., Garpebring, A., Karlsson, M., Asklund, T., Nyholm, T.: Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information – potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. Acta Oncologica **52**(7) (2013) 1369–1373

[17] Rank, C., Tremmel, C., Hunemohr, N., Nagel, A., Jakel, O., Greilich, S.: MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. Radiation Oncology **8**(1) (2013) 51

[18] Dowling, J.A., Lambert, J., Parker, J., Salvado, O., Fripp, J., Capp, A., Wratten, C., Denham, J.W., Greer, P.B.: An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. International journal of radiation oncology, biology, physics **83**(1) (05 2012) e5–e11

[19] Burgos, N., Cardoso, M., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Duncan, J., Atkinson, D., Arridge, S., Hutton, B., Ourselin, S.: Attenuation correction synthesis for hybrid PET-MR scanners. In Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Volume 8149 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2013) 147–154

[20] Uh, J., Merchant, T.E., Li, Y., Li, X., Hua, C.: MRI-based treatment planning with pseudo CT generated through atlas registration. Medical Physics **41**(5) (2014) –

[21] Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schölkopf, B., Pichler, B.J.: MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration. Journal of Nuclear Medicine **49** (2008) 1875–1883

[22] Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. NeuroImage **54**(2) (2011) 940 – 954

[23] Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. Medical Imaging, IEEE Transactions on **30**(10) (Oct 2011) 1852–1862

[24] Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J., Pieper, S., Kikinis, R.: 3D Slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9) (11 2012) 1323–1341

[25] Nyul, L., Udupa, J., Zhang, X.: New variants of a method of MRI scale standardization. Medical Imaging, IEEE Transactions on **19**(2) (Feb 2000) 143–150

[26] Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4) (2004) 600–612

[27] Stefan Klein and Marius Staring and Keelin Murphy and Max A. Viergever and Josien P.W. Pluim: elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1) (January 2010) 196 – 205

[28] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, series B **39**(1) (1977) 1–38

[29] Arthur, D., Vassilvitskii, S.: K-means++: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '07, Philadelphia, PA, USA, Society for Industrial and Applied Mathematics (2007) 1027–1035

[30] Varma, S., Simon, R.: Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics **7**(1) (2006) 91

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

[32] Mori, S., Chen, G.T., Endo, M.: Effects of intrafractional motion on water equivalent pathlength in respiratory-gated heavy charged particle beam radiotherapy. International Journal of Radiation Oncology*Biology*Physics **69**(1) (2007) 308 – 317

[33] International Commission on Radiation Units and Measurements (ICRU), Report 83: 3. Special Considerations Regarding Absorbed-Dose and Dose–Volume Prescribing and Reporting in IMRT. Journal of the ICRU **10**(1) (2010) 27–40

[34] Andreasen, D., Andersen, J.L., Hansen, R.H., Van Leemput, K., Edmund, J.M.: The impact of a Dixon sequence in creating a pseudo CT scan from MR images using a Gaussian mixture regression model. Radiotherapy & Oncology **106, supplement 2** (March 2013) S229

[35] Johansson, A., Garpebring, A., Asklund, T., Nyholm, T.: Ct substitutes derived from mr images reconstructed with parallel imaging. Medical Physics **41**(8) (2014) –

[36] Jonsson, J.H., Johansson, A., Söderström, K., Asklund, T., Nyholm, T.: Treatment planning of intracranial targets on MRI derived substitute CT data. Radiotherapy and Oncology **108**(1) (2013) 118 – 122

[37] Korsholm, M., Waring, L., Edmund, J.: A criterion for the reliable use of MRI-only radiotherapy. Radiation Oncology **9**(1) (2014) 16

[38] Ta, V.T., Giraud, R., Collins, D., Coupé, P.: Optimized patchmatch for near real time and accurate label fusion. In Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. Volume 8675 of Lecture Notes in Computer Science. Springer International Publishing (2014) 105–112

Paper C

A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis

Daniel Andreasen*
Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark and Radiotherapy Research Unit, Department of Oncology, Gentofte and Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark

Koen Van Leemput
Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark and A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129 USA

Jens M. Edmund
Radiotherapy Research Unit, Department of Oncology, Gentofte and Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark
(Dated: February 23, 2017)

Purpose: In radiotherapy (RT) based only on magnetic resonance imaging (MRI), knowledge about tissue electron densities must be derived from the MRI. This can be achieved by converting the MRI scan to a so-called pseudo-computed tomography (pCT). An obstacle is that the voxel intensities in conventional MRI scans are not uniquely related to electron density.
We previously demonstrated that a patch-based method could produce accurate pCTs of the brain using conventional T_1-weighted MRI scans. The method was driven mainly by local patch similarities and relied on simple affine registrations between an atlas database of co-registered MRI/CT scan pairs and the MRI scan to be converted. In this study, we investigate the applicability of the patch-based approach in the pelvis. This region is challenging for a method based on local similarities due to the greater inter-patient variation. We benchmark the method against a baseline pCT strategy where all voxels inside the body contour are assigned a water-equivalent bulk density. Furthermore, we implement a parallelized approximate patch search strategy to speed up the pCT generation time to a more clinically relevant level.
Methods: The data consisted of CT and T_1-weighted MRI scans of 10 prostate patients. pCTs were generated using an approximate patch search algorithm in a leave-one-out fashion and compared with the CT using frequently described metrics such as the voxel-wise mean absolute error (MAE_{vox}) and the deviation in water-equivalent path lengths. Furthermore, the dosimetric accuracy was tested for a volumetric modulated arc therapy (VMAT) plan using dose volume histogram (DVH) point deviations and γ-index analysis.
Results: The patch-based approach had an average MAE_{vox} of 54 HU; median deviations of less than 0.4% in relevant DVH points, and a γ-index pass rate of 0.97 using a 1%/1mm criterion. The patch-based approach showed a significantly better performance than the baseline water pCT in almost all metrics. The approximate patch search strategy was 70x faster than a brute-force search, with an average prediction time of 20.8 minutes.
Conclusion: We showed that a patch-based method based on affine registrations and T_1-weighted MRI could generate accurate pCTs of the pelvis. The main source of differences between pCT and CT was positional changes of air pockets and body outline.

I. INTRODUCTION

The introduction of magnetic resonance imaging (MRI) in external beam radiotherapy (RT) has spurred an interest in removing computed tomography (CT) from the treatment chain; so-called MRI-only RT [1–5]. MRI provides superior soft tissue contrast and improves the precision in delineations [6, 7]. Basing the entire workflow on MRI would additionally eliminate systematic registration uncertainties between MRI and CT [8, 9]. It is non-trivial to exclude the CT from the RT chain, however, since the MRI signal is not uniquely related to the tissue electron density which is needed for accurate dose calculations. Furthermore, with conventional MRI sequences, the signal from the cortical bone is weak or non-existent due to its short transverse relaxation time T_2 [10]. This results in an ambiguity between bone and air voxels, which complicates direct assignment of CT equivalent numbers to the MRI scan and makes patient setup based on bony anatomy difficult.

Using mathematical models based on a-priori knowledge or pre-acquired and correlated MRI and CT scans to predict a so-called pseudo-CT (pCT) from the MRI scan is a promising solution, which has been applied and

* Corresponding author email: dana@dtu.dk

validated in numerous forms for brain pCT generation [1, 11–20]. Relatively fewer attempts, however, have been made to create pCTs for other body parts such as the pelvis [21–25]. Nonetheless, multiple studies have shown the benefit of accounting for bone in MRI-only RT of the pelvis [2, 26, 27], and additionally accounting for the heterogeneity of bones in the pelvis [28].

In general, the methods for pCT estimation can be split into two categories: 1) voxel-based and 2) atlas-based. With the voxel-based methods, a model is used to convert individual MRI voxels to their corresponding CT Hounsfield unit (HU) value, either by regression or a combination of classification and bulk density assignment [1, 12–14, 16]. Some of these methods include a training phase using MRI and CT scans where the relation is known. The drawback of the voxel-based methods is the need to uniquely distinguish air and bone voxels in the MRI. This has previously been addressed by introducing specialized ultrashort echo time (UTE) MRI sequences [1, 13] or by manual delineation of the bone volume and using a separate model for this region [23, 24]. These approaches however, prolong the MRI scan time or put an extra workload on the personnel who have to do the manual bone delineation. Using spatial information or edge position and orientation of the MRI scan can potentially also be used to overcome the intensity ambiguities [29, 30].

With atlas-based methods, the pCT prediction often relies on a non-linear registration between the MRI scan to be converted and one or multiple atlases of MRI scans with known correspondence to the CT [15, 17, 21, 22, 31]. The MRI intensity ambiguities are thus resolved by utilizing the spatial information provided by the registration and specialized MRI sequences are no longer needed. A concern with atlas-based methods is that they depend entirely on the accuracy of the registration; in cases where the patient geometry is very different from the atlas(es), this may pose a problem. When using multiple atlases, an attempt to overcome this is to use a similarity measure such as the local cross-correlation or the sum of squared difference in a patch to create a locally weighted fusion of the atlas CTs [21, 31]. This gives less weight to regions in the atlases where the non-linear registration was sub-optimal. Another approach is to refine the atlas-based prediction using pattern recognition on multiple 2D patches per atlas after the non-linear registration [15].

We previously proposed using a multi-atlas approach based on affine registrations and letting brain pCT prediction be driven mainly by patch similarities between the patient MRI scan and the atlas database of MRI scans [20]. For the brain, this method compared favorably to a multi-atlas method based on non-linear registration and a voxel-based method based on UTE scans. Concerning its applicability to other body sites, as well as its clinical relevance, we identified two possible limitations of the patch-based approach: 1) For sites such as the pelvis where the inter-patient anatomical variation is larger than in the brain, the approach might fail because similar anatomical regions are too far apart to be roughly aligned by an affine transformation. This would cause problems when searching for similar patches in a limited search volume. 2) When using a brute-force patch search strategy, the computational speed benefit of using affine registrations instead of non-linear ones was heavily outweighed by the time used for the patch search. This resulted in a pCT prediction time of around 15 hours, which could hinder its clinical relevance.

In this paper, we evaluate the robustness of the patch-based approach for predicting pCTs of the pelvic region without using non-linear registrations. To adapt the method for the pelvic region, we introduce an atlas preselection step in order to discard highly dissimilar atlas patients prior to the patch search step. This avoids atlas patients where the affine alignment could not bring similar anatomical regions within the limited patch search volume. Furthermore, we use an approximate nearest-neighbor algorithm to dramatically speed up the search for similar patches, making the method more suitable for a clinical setting than our previously proposed method. We evaluate the approach in terms of its geometric, radiologic and dosimetric accuracy using frequently described metrics and we compare the performance with a baseline water pCT where all voxels within the MRI body contour are assigned a water-equivalent bulk density.

II. MATERIALS AND METHODS

A. Imaging

Data consisting of MRI and CT scans of ten prostate RT patients aged 55-82 years were retrospectively obtained. The CT scans were acquired on a Philips Brilliance Big Bore CT using a standard protocol for pelvic scans (120 kVp, 232-503 mAs). The scans were acquired with a voxel spacing of between $0.78 \times 0.78 \times 2$ mm and $1.4 \times 1.4 \times 2$ mm, for an in-plane matrix of 512×512 voxels and 129-199 slices. The MRI scans were acquired on a Philips Panorama 1 T open scanner using a bridge body coil with a T_1-weighted sequence, TE/TR = 10/623 ms. The voxel spacing was $0.80 \times 0.80 \times 4$ mm, for an in-plane matrix between 528×528 and 640×640 voxels and 16-24 slices. The approximate duration of the T_1-weighted sequence was 5-7 minutes. The patients were fixed in treatment position during both the MRI and CT scanning using the same fixation devices.

B. Pre-processing

To give the voxel intensities a consistent tissue meaning in all MRI scans, we applied bias field correction [32] and intensity normalization using a piece-wise linear histogram-matching approach [33]. We found that two distinct peaks corresponding to muscle and fat were consistently present in the MRI intensity histograms and

used these peaks as landmarks for the piece-wise normalization. This differs from the single landmark we used in our previous work on the brain region. To ensure spatial correspondence between voxels and to account for organ movement between the scans, we non-linearly aligned each patient's MRI scan to his CT scan using Elastix [34]. The registration included four steps: (1) To guide the registration to the correct region of the CT, a mask which covered the same anatomy as the MRI was defined for the CT scan. Elastix samples the cost function only within this mask. (2) To roughly position the MRI within the mask region of the CT, an initial manual rigid alignment was carried out. (3) To get a proper linear alignment, an automatic affine registration of the MRI to CT was done. (4) A non-linear registration was applied using a multi-resolution B-spline deformation model with local normalized mutual information as the similarity metric. The parameters used for the different steps of the intra-patient registration are summarized in Table I. These parameters were manually determined by visually inspecting the scans for a satisfactory alignment. The registration resulted in a resampling of the MRI scans to match the CT resolution, leaving the MRI in CT space. To allow a one-to-one comparison with the CT with minimal impact of between-scan motion, all pCTs were predicted from the non-linearly co-registered MRI.

C. Patch-based approach

We used a previously described patch-based method [20] to predict a pCT for each patient using a leave-one-out strategy. For a new test patient, the essential steps consisted of 1) Linearly aligning a pre-acquired database of co-registered atlas MRI/CT pairs with the patient MRI. 2) For each position x in the test patient MRI, extracting a 3D subvolume of voxels (a patch) and performing a search for the K most similar patches using the L2 distance (square root of the sum of squared difference) between patch intensities in a search volume, V, centered on x in the database MRIs. 3) Based on the degree of MRI patch similarity, assigning a weighted average of the corresponding database CT values as the predicted value at position x. The inter-patient alignment required in step 1 was achieved using an affine registration of each atlas MRI to the test patient MRI and then applying the same transformation to the corresponding atlas CT scans. The registration parameters are summarized in Table I. Since the test patient MRI was previously non-linearly co-registered with its corresponding CT, this left all scans (atlases and patient) in the test patient's CT space prior to the subsequent steps.

We made a few modifications to the previously described model for the brain. To discard the globally most dissimilar atlases or atlases where the registration to the test patient scan was sub-optimal, we defined a global similarity between the test patient MRI and each atlas MRI. This has previously been shown to benefit multi-atlas-based brain segmentation methods [43]. After performing step 1, we calculated the global similarity as the intensity-based L2 distance between corresponding voxels in the atlas and test patient MRI. We only used voxels situated inside the MRI body contour of the test patient. The body contour was automatically found in the MRI using Otsu's method [39] and morphological operations. We manually tweaked the parameters for the morphological operations to establish a reasonable baseline and used the same parameters for all scans. We empirically chose to keep only the five most similar atlases for the subsequent local patch search steps. To improve prediction accuracy, we used the whole CT patch instead of just the center voxel of the patch when fusing the corresponding CT values of the most similar MRI patches [35]. This meant that, for each voxel, the found CT patches of neighboring voxels also contributed to the prediction, and that each pCT voxel was an average of $K \times m$ voxel intensities, where m is the number of voxels in a patch.

To speed up the patch search we implemented an approximate nearest neighbor algorithm inspired by the PatchMatch (PM) algorithm [36] and with similarities to the work of Shi *et al.* [37] and Ta *et al.* [38] who also used PM in combination with a database of MRI scans. The PM algorithm was originally proposed for finding patch correspondences in 2D images. It does not guarantee finding the nearest patch (i.e., most similar), but a good approximation, and it drastically cuts down on both the required number of L_2 distance calculations compared to a brute-force search, and the memory requirements compared to other approximate algorithms [36]. It consists of a random initialization of nearest neighbor candidates, followed by an iterative process of propagation of good candidates using adjacent voxels and a random search to avoid local minima (see Barnes *et al.* for details [36]). To use PM for our purpose, we performed the random initialization for each x by picking a random database patient and a random patch within the search volume, V, around x. At even iteration numbers in the propagation phase, we tested for improved candidates at each x in three directions (posterior, superior and right lateral). At uneven iterations, we tested in the opposite directions (anterior, inferior and left lateral). In the random search phase, we searched for improved candidates at random positions inside V in the same database patient as the current best candidate. Contrary to the brute-force approach, the run-time of this multi-atlas PM algorithm is independent of both the number of atlases and the size of V. The accuracy of the algorithm is to some degree dependent on the number of iterations. We found that in practice, the pCT prediction no longer changed considerably after about 15 iterations.

We empirically set the patch and search volume sizes to $m = 7 \times 7 \times 5$ and $V = 13 \times 13 \times 11$ voxels, respectively. We set the number of similar patches to find at each voxel to $K = 8$ and ran the PM algorithm in eight independent parallel sessions.

Registration type:	Intra-patient		Inter-patient
	Affine	Non-linear	Affine
Direction	MRI → CT	MRI → CT	MRI → MRI
Metric	NMI	NMI+BE	NMI
Fixed image mask	✓	✓	
Samples	2000	2048	2048
Resolutions	4	3	4
Max iterations	1000	1500	500
Final B-spline grid spacing [mm]	Not used	$x/y/z$: 8/8/10	Not used

Table I. Summary of registration parameters. NMI: normalized mutual information (32 histogram bins). BE: bending energy penalty. For all registrations, adaptive stochastic gradient descent was used as optimizer. For the affine registration, Gaussian smoothing was applied to achieve multiple resolutions. For the non-linear registration, the B-spline grid was also made coarser at lower resolutions. Inter-patient registration refers to the alignment of atlas MRIs with the test patient MRI.

D. Benchmark method

To benchmark the proposed method, we created an additional pCT by assigning a bulk density to the MRI. We set all voxels within the body outline to 0 HU, thus assuming all tissues to be water (called $\mathrm{MRI_w}$). The body region was found using the approach described in Section II C. This represents the simplest possible way of creating a pCT and any method should outperform this to justify any additional correction steps.

E. Geometric and radiologic evaluation

We evaluated the pCTs in terms of their geometric and radiologic similarity to the CT. A measure of the overall geometric quality of the pCTs is the voxel-wise mean absolute error ($\mathrm{MAE_{vox}}$) within the MRI body outline (excluding surrounding air):

$$\mathrm{MAE_{vox}} = \frac{1}{N}\sum_{i=1}^{N}|\mathrm{CT}(i) - \mathrm{pCT}(i)|, \quad (1)$$

where N is the total number of voxels inside the body outline of the MRI. Similarly the voxel-wise mean error ($\mathrm{ME_{vox}}$) can be calculated:

$$\mathrm{ME_{vox}} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{CT}(i) - \mathrm{pCT}(i)). \quad (2)$$

We also calculated the Pearson correlation coefficient between the CT and pCT for voxels within the MRI body contour.

To get an estimate of the radiologic accuracy of the pCTs, we compared water-equivalent path lengths (WEPL) in the CT and pCTs. The water-equivalent path length is defined as:

$$l' = \sum_i \Delta l_i \times \rho_i, \quad (3)$$

where Δl_i is the physical distance traveled by a photon in voxel i, and ρ_i is the voxel's electron density relative to water, which is relevant for MeV photons. The value of ρ_i is found from the voxel CT number using a verified standard calibration curve in the treatment planning system (Eclipse v11.0, Varian Medical Systems Inc., Palo Alto, CA). We defined the approximate center in the prostate region and calculated l' in 1800 radial spokes towards the surface of the patient in both the CT and the pCT. The mean absolute WEPL error is then defined as:

$$\mathrm{MAE_{WEPL}} = \frac{1}{L}\sum_{j=1}^{L}|l_j'^{CT} - l_j'^{pCT}|, \quad (4)$$

where L is the number of spokes and l'_j is the WEPL of the jth spoke. Again, a similar mean WEPL error is given as:

$$\mathrm{ME_{WEPL}} = \frac{1}{L}\sum_{j=1}^{L}(l_j'^{CT} - l_j'^{pCT}). \quad (5)$$

F. Dosimetric evaluation

A volumetric modulated arc therapy (VMAT) treatment plan was created for each patient and optimized according to our clinical protocol on the pCT. The patients were prescribed a dose of either 78 Gy (six patients) or 70 Gy (four patients) depending on their staging. For both groups, the dose was prescribed in 2 Gy fractions to the prostate. All treatment planning was carried out in Eclipse v11.0 (Varian Medical Systems Inc., Palo Alto, CA) using the analytical anisotropic algorithm (AAA) to calculate the dose distribution. After the plan on the pCT was finished, it was transferred to the CT and recalculated using the same plan parameters and monitor units. Due to the retrospective study design, for some patients the MRI scan did not cover the entire femoral heads or the entire PTV in the inferior/superior direction. In these cases, we cropped the structures to the body outline and extended the pCT by 40 mm in the longitudinal direction by repeating the last slice. The same was done on the CT scan. Note that this extension was only applied in the dosimetric evaluation and not in the geometric and radiologic evaluation.

1. DVH analysis

The dose volume histogram (DVH) was used to evaluate the dosimetric difference between the CT and pCT. The relative change in percent was calculated for the

DVH points relevant to the PTV coverage and the organ at risk (OAR) dose constraints. For the PTV, these were the near-minimum ($D_{98\%}$), near-maximum ($D_{2\%}$) and median (D_{median}) absorbed dose [40]. For the femoral heads they were the mean (D_{mean}) and maximum ($D_{\max}$) absorbed dose. For the rectum, the constraints varied depending on the prescribed dose, but since we were only interested in evaluating the deviations and not the absolute values, we chose to pool the deviations calculated for the different DVH points in the two prescription groups. These were the volumes receiving 65/75 Gy ($V_{65/75}$), 54/60 Gy ($V_{54/60}$) and 36/40 Gy ($V_{36/40}$), where the subscripted values were the constraints for the 70 Gy and the 78 Gy prescription, respectively.

2. *γ-index analysis*

To further evaluate the general accuracy in the pCT dose distribution, we performed an analysis based on the γ-index [41]. This metric evaluates dose differences (DD) and distance-to-agreement (DTA) criteria for each voxel between a test and a reference dose matrix. The result is a pass rate between 0 and 1, where 1 is better. We used the CT-based dose matrix as reference and performed a slice-wise 2D γ-analysis [42] with global normalization and DD/DTA=2%/2mm and 1%/1mm pass criteria, respectively. All voxels with a dose greater than 10% of the prescribed dose were included, as this excludes all voxels outside the patient body [21, 23].

G. Statistical tests

Paired t-tests were performed to test for differences between the pCT strategies in all relevant quality measures. $p \leq 0.05$ was considered significant. Prior to performing the hypothesis testing, the data were tested and found to be approximately normally distributed with constant variances, thereby fulfilling the assumptions of the statistical method.

III. RESULTS

In Figure 1, examples of the generated pCTs are shown along with the corresponding CT and T_1-weighted MRI. In most regions, the patch-based method predicts the correct HU value as illustrated by the difference maps in Figure 1. Some exceptions are the regions near the body outline which are both over- and underestimated and some of the cortical parts of the pelvic bone and femoral heads, which are underestimated. The MRI_w pCT shown in Figure 1 (m), (o) and (q), obviously has errors in the bone region. Furthermore, the method for extracting the MRI body and generating the MRI_w has a consistent error at the body outline causing an overestimation of the HU value. This is probably caused by noise observed in the MRI scan in the surrounding air close to the patient outline.

In Table II, the average MAE_{vox} and ME_{vox} are shown for both pCT approaches. These measures give an indication of the overall geometric quality of the pCTs. The patch-based approach shows a significantly better performance than MRI_w. It is worth noting that even though all cortical bone by default is underestimated by the MRI_w approach, there is an average overestimation of the HU value as indicated by the negative ME_{vox}. This effect is related to the relative amounts of voxels within each tissue. In Table II, the average MAE_{WEPL} and ME_{WEPL} also show a significantly better performance of the patch-based approach compared to the MRI_w approach. The average correlation coefficient between the patch-based pCT and CT was 0.79 ($\sigma = 0.06$). The correlation coefficient is undefined for MRI_w because the standard deviation is zero within the body contour.

Metric	Patch-based	MRI_w	p-value
MAE_{vox} [HU]	54 (8)	105 (12)	$< 10^{-7}$
ME_{vox} [HU]	−1 (7)	−16 (11)	0.003
MAE_{WEPL} [mm]	1.2 (0.3)	2.5 (0.3)	$< 10^{-10}$
ME_{WEPL} [mm]	−0.3 (0.7)	0.5 (0.5)	$< 10^{-4}$
Pass rate $\gamma_{2\%/2mm}$	0.99 (0.01)	0.98 (0.01)	0.014
Pass rate $\gamma_{1\%/1mm}$	0.97 (0.02)	0.94 (0.03)	0.019

Table II. The average mean absolute voxel-wise error (MAE_{vox}) in HU, the average mean absolute WEPL error (MAE_{WEPL}) of the water-equivalent path lengths (in mm). The bottom rows show the mean 2D γ-index pass rates for dose points greater than 10% of the prescribed dose. Average value and standard deviation (σ) for the ten patients are shown. In the rightmost column are the results of a paired t-test on the difference between the patch-based and MRI_w methods

In terms of the γ-index evaluation, Table II shows average pass rates close to 1 for both methods with slightly higher rates for the 2%/2mm criterion than the more strict 1%/1mm criterion. The patch-based approach has a higher pass rate than MRI_w which is also visualized by the γ-index maps in Figure 2. Differences between the patch-based pCT and MRI_w can be observed mainly at the body outline, in the pelvic bone, and the prostate.

The DVH analysis shown in Figure 3 reveals a median deviation on the PTV within ±0.04% and ±0.4% for the patch-based and MRI_w methods, respectively, and standard deviations in the range 0.6-0.7. No significant difference was found between the two methods using a paired t-test ($p > 0.05$). On the femoral heads, the patch-based approach had median deviations within ±0.2% with standard deviations of 0.6-0.7. The MRI_w approach had median deviations within 0.9% with standard deviations of 0.7-0.9. A significant difference was found between the two approaches using a paired t-test ($p < 0.01$).

For the rectum, the patch-based and MRI_w approaches

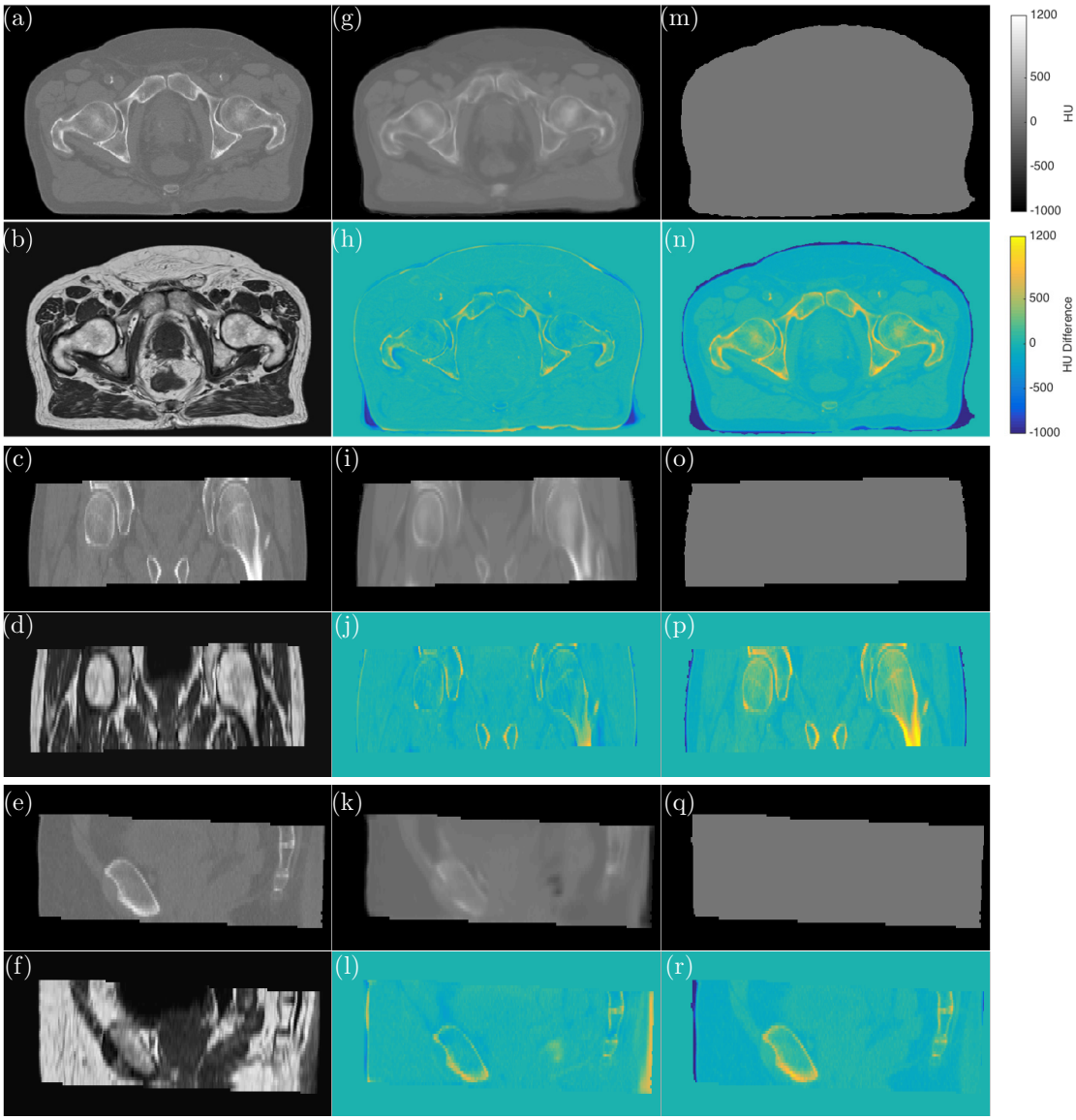


Figure 1. Qualitative comparison of CT and pCTs. (a)-(f) show alternately the CT and the MRI (registered to the CT) that the pCTs are based on, for different views. (g)-(l) show alternately the patch-based pCT and the difference maps between the patch-based pCT and CT. (m)-(r) show alternately the MRI_w pCT and the difference maps between the MRI_w pCT and CT. The top colorbar relates to the pCTs and the bottom colorbar to the difference maps. A negative HU difference indicates an overestimation of the true HU value.

had median deviations within ±0.4% and ±1.3%, respectively. There was a significant difference between the two approaches in $V_{36/40}$ and $V_{54/60}$. Though the median deviations were close to 0% for the patch-based approach, in $V_{65/75}$ we observed a peak deviation of approximately 10% for one patient and a general larger variation in the deviations than the other DVH points. This variation is related to the varying position of air pockets in the rectum between the CT and pCTs.

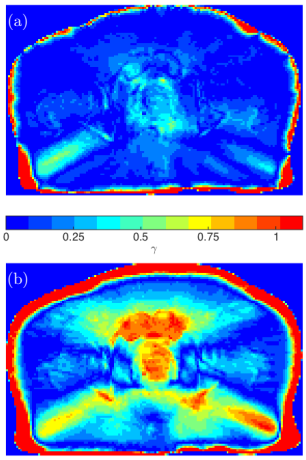


Figure 2. Example of a 1%/1mm criterion γ-index map for one transverse slice of a dose matrix for (a) the patch-based method and (b) the MRI_w method. Pass rates for the shown slice are 0.98 and 0.89 for the patch-based and the MRI_w method, respectively. For the patient shown, the average pass rates are 0.99 and 0.92 for the patch-based and the MRI_w method, respectively.

IV. DISCUSSION

In this paper, we evaluated the robustness of a patch-based method for predicting pCTs based on conventional T_1-weighted MRI images of the male pelvic region. The patch-based approach has previously compared favorably to other multi-atlas and voxel-based approaches in the brain, but for a method based on registration and local patient similarities, the pelvis is more challenging due to a greater anatomical variation between patients. In particular, for a patch-based approach based only on affine registrations, this could affect the search for similar patches since the correct anatomy may not be contained within the local search volume.

In Figure 1, it seems as though the patch-based approach is able to capture most of the correct HU values and geometries of both the bone, muscle and fat regions of the CT. However, some parts of the cortical bone are still underestimated, illustrating the challenging nature of these voxels. Also, since each pCT voxel is the result of an averaging of $K \times m = 8 \times 245$ voxels, there is a general blurring effect at tissue borders.

A. Geometric and radiologic agreement

From the results of the geometric and radiologic evaluation shown in Table II, the patch-based approach performed significantly better than the baseline MRI_w method. With the simplicity of that method, this might come as no surprise. Nevertheless, we believe that including this benchmark is important since it can provide some intuition about the magnitudes of the different error measures, which can otherwise seem quite arbitrary. Many authors have published MAE_{vox} results for brain pCT methods, which were approximately in the range 85−140 HU [1, 11–20]. An MAE_{vox} of 54 HU would therefore seem impressive for the pelvis, but as the MRI_w illustrates, the MAE_{vox} is not comparable across treatment sites. This is mainly caused by differing relative amounts of soft tissue vs. bone and air in different sites.

In Table II, the ME_{vox} revealed a pattern of consistent overestimation of the HU value for MRI_w. This can be explained by the relatively large amount of fat voxels, which become overestimated by assigning a value of 0 HU. The magnitude of the error made in voxels with an HU greater than 0 might be larger than in the fat voxels, e.g., greater than 500 HU for the cortical bone, but the amount of voxels with such an HU value is lower than the amount of fat voxels. This illustrates one of the weaknesses of using a purely voxel-wise and global error metric such as the ME_{vox} or MAE_{vox}. Correct predictions in the bone region might be dosimetrically more important than in soft tissue [2, 26, 27], but since the proportion of these voxels is low, errors do not really show in a voxel-based evaluation. The correlation coefficient might be comparable between different treatment sites, since the standard deviation of the HU values is included. Our result of 0.79 is close to the values reported by Uh *et al.* for the brain when using different atlas-based approaches [18].

As previously mentioned, there was a consistent overestimation of the spatial extent of the MRI body contour by our simple body contour extraction procedure. It should be noted that this caused an overestimation of the spatial extent of the MRI and thus also the MRI_w pCT. The MAE_{vox}, ME_{vox} and correlation coefficient were also calculated within this contour, which always covered the entire CT. For the patch-based pCT, the larger estimated contour could positively affect the relevant quality measures if the air close to the patient is easy to predict. This is not necessarily the case and compared to the size of the patient body, we expect the effect of this to be negligible.

The WEPL-based error metrics give a more informative measure of quality than the voxel-wise metrics, since they evaluate a radiologically and dosimetrically relevant path length independently of the order by which tissues appear along each spoke. We observed an average MAE_{WEPL} of 1.2 and 2.5 mm for the patch-based and MRI_w pCTs, respectively. This gives an indication that the patch-based pCTs are radiologically more accurate than the MRI_w. This is consistent with the larger deviation in DVH points (Figure 3) for the MRI_w as compared to the patch-based pCT. A unique correlation between dose and WEPL would require an equal amount of dosimetric absorption along each radial spoke for the VMAT plans. The dose delivery, however, depends on the opti-

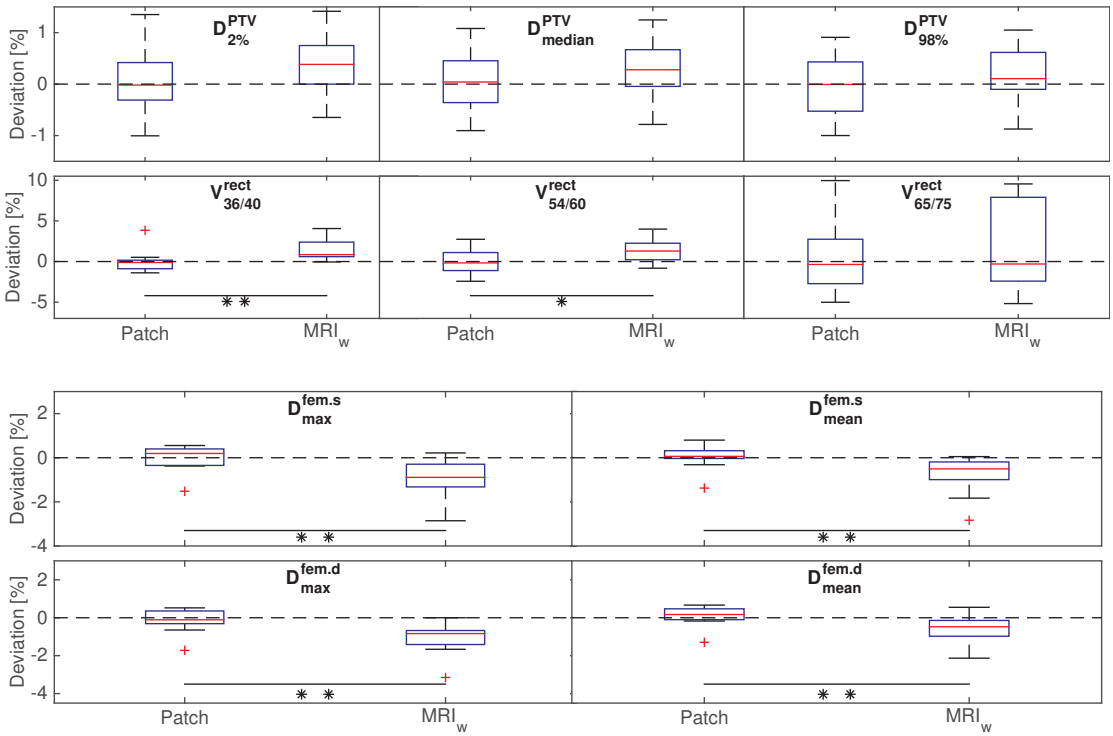


Figure 3. Tukey-style boxplots showing the relative deviation in percent for the different methods in relevant DVH points for the PTV and OARs. ∗∗ indicates a significant difference at the $\alpha = 0.01$ level and ∗ indicates a significant difference at the $\alpha = 0.05$ level. The dashed horizontal line indicates zero deviation. rect: rectum, fem.s: femur sinister, fem.d: femur dexter.

mization objectives which distribute the dose unequally on each spoke depending on the position of organs at risk and PTV.

B. Dosimetric agreement

In the DVH analysis shown in Figure 3, the patch-based approach mostly had median deviations close to 0% and maximum deviations less than ±1.5%. The exception to this was the rectum, where a peak deviation of 10% was observed for one patient in $V_{65/75}$. Looking at the pCT and CT of this patient, the deviation can be explained by rectal air pockets present in the CT but not in the pCT. In general, most patients were affected by air pockets, and these varied in size and position between the CT and MRI scans. This results in a discrepancy between the pCT and CT, which could cause dosimetric differences. Often, an air pocket present in the test patient MRI could not be matched anywhere in the atlas, causing a conversion to soft tissue in the pCT. Other times, air pockets were correctly predicted based on the MRI data, but they were not present in the corresponding CT. In general, air pocket positions will vary between treatment planning and the actual treatment, causing either an under- or over-dosage when present in the treatment field of view (FOV). This issue is not confined to MRI-only RT, and the current mitigation in the CT-based work-flow is often to assign a water-equivalent HU value to those regions. This strategy could also be employed on pCTs and performed automatically by doing the assignment in the CT atlases.

Looking at the femoral head DVH deviations, accounting for bone results in a significantly lower deviation for the patch-based approach than the MRI_w approach. In general, our DVH analysis shows the importance of evaluating not just the high dose PTV region, but also the OARs. It is mostly in the gradient and low-dose regions that a pCT strategy demonstrates its advantages, as shown in Figure 3.

The γ-index analysis showed relatively high average pass rates for both the patch-based approach and MRI_w, indicating high dosimetric agreement. There was a significant difference between the two methods for both the 2%/2mm and 1%/1mm criteria with an average higher pass rate for the patch-based approach than the MRI_w

approach. There is currently no consensus on the required criterion and pass rates for a pCT to be considered of acceptable quality. However, the pass rates for pCT vs. CT planned dose should be higher than those usually considered acceptable when evaluating planned vs. delivered dose. With the rather strict acceptance criteria used here, an average pass rate of 0.97 is probably acceptable, whereas a pass rate of 0.94 could be questioned.

A source of uncertainty in our dosimetric results was the limited FOV for some patients. This could have been avoided by a prospective study design. To account for the scattering component in the dose simulation, we chose to extend the pCTs by repeating the last slice in the inferior/superior direction. The corresponding slices were also repeated in the CT scan. The average MAE_{vox} across the patients in only the repeated slices was 91 HU for the patch-based pCTs. Compared to the overall MAE_{vox} of 54 HU, this indicates that these slices were not that well predicted. For this reason, the repeated slices may have introduced a bias that negatively impacted the dosimetric results. However, we found this compromise better than not being able to account for the scattering component in the dose simulation.

The dosimetric deviations in the current study indicate that the patch-based approach could be used clinically. However, a dosimetric study including more patients would establish a dosimetric deviation confidence interval with more certainty.

C. Comparison with other pCT methods

Overall, our results seem to fall in line with other recently published pCT approaches for the pelvis. Using VMAT treatment on twenty patients and a dual regression model with manual bone delineation, Korhonen *et al.* reported average γ-index pass rates of 0.99 and 0.93 for 2%/2mm and 1%/1mm criteria, respectively [23]. This was for a 2D evaluation of the body region of one transverse slice per patient and was achieved with a correction for body contour differences between CT and pCT. For their multi-atlas-based method with thirty-eight atlas patients, Dowling *et al.* reported median pass rates for thirty-nine patients in the 0.91-0.95 range for a 3D evaluation of the body region and using a 1%/1mm criterion [21]. This was for a seven field IMRT plan. They further reported an average ME_{vox} and MAE_{vox} of 0.6 ($\sigma = 14.7$) and 40.5 ($\sigma = 8.2$), respectively for the body region. For a commercial multi-atlas method with fifteen atlas patients, Siversson *et al.* reported average γ-index pass rates for ten patients between 0.97 and 1.0 using a 2%/1mm criterion and excluding points in the vicinity of the body outline [25]. This was for a VMAT plan. They further reported average deviations in PTV, rectum and femoral heads DVH points close to 0% and an average ME_{vox} and MAE_{vox} of 1.9 ($\sigma = 6.6$) and 36.5 ($\sigma = 4.1$), respectively within the body region. It should be noted that slight differences in the evaluation of the pCTs between the different methods makes it difficult to compare them directly. These differences include whether a rigid or a non-linear registration was used to align the MRI and CT scans prior to the evaluation; whether the dose plan was optimized on the CT or the pCT, and how the γ-index was calculated. Furthermore, the evaluation of an IMRT plan will be affected by the specific field angles chosen, whereas a VMAT plan reflects an average of multiple angles similar to the way the WEPL measure is calculated. Although the proposed approach and the methods mentioned above seem to speak in favor of an atlas-based approach in the pelvis, the choice of an optimal pCT strategy is ongoing research and effort is still being put into making more accurate voxel-based strategies [44]. The main advantage of these approaches is a fast, registration-free prediction, which could make them more robust against inter-patient anatomical variation.

D. Algorithm speed-up

In a previous publication, we showed promising results with the patch-based approach in the brain region. One of the drawbacks was the prediction time, which was around 15 hours for one pCT. To address this issue in this paper, we incorporated an approximate nearest-neighbor algorithm in the pCT approach. To test the consequences of this, we ran the algorithm on the same brain data and hardware set-up (64 bit Linux cluster with 8 quad-core AMD Opteron 8356 2.3 GHz CPUs and a total of 256 GB RAM) as we used previously [20], and achieved on average a 70x speed-up compared to the brute-force approach. This meant that the average prediction time of a brain pCT was approximately 14 minutes. The time required for the affine registrations is not included in this, so roughly 1.5 minutes should be added per patient in the atlas database. Otherwise, the algorithm run-time is independent of the number of patients in the atlas. Using the approximate algorithm, we saw an average increase in the MAE_{vox} of 12 HU ($\sigma = 9$ HU) compared to the brute-force approach, which we find is a worthwhile trade-off. The average prediction time for the pelvic scans used in this study was 20.8 minutes excluding registrations (requiring 1.5-2 minutes per atlas patient). We find this acceptable for a clinical work-flow, as it enables same-day planning and treatment of patients. For a clinical implementation, a possible alternative to a powerful computer cluster could be a single 8-core machine or a cloud-based solution, both of which allow parallel patch-matching. With the introduction of the MRI LINAC, faster prediction times would still be preferred to generate plans of the day. This could potentially be achieved by a GPU-based implementation.

The prediction time of the algorithm is dependent on the number of voxels in the test patient MRI. We tested the algorithm on a patient where we artificially added 34 slices to simulate a 6.8 cm larger FOV in the infe-

rior/superior direction. This resulted in a prediction time of 28 minutes.

There is a random component in the approximate algorithm, meaning that the predictions could vary between different runs on the same patient. To test the severity of this, we predicted 15 pCTs for the same patient with a resulting standard deviation of 0.3 HU in the MAE_{vox}. This is negligible compared to the variation in MAE_{vox} between patients.

E. MRI to CT registration

Aligning the intra-patient MRI and CT scans is an important step, both for evaluating the pCT accuracy and for generating high-quality atlases. This influences the ground truth, which both the evaluation and prediction are based on. Accordingly, if the patient MRI and CT are not well aligned, then neither are the pCT and CT, resulting in evaluation differences not related to the prediction itself. Likewise, misalignments in the atlases will result in less than optimal predictions because database patches will have wrong CT values attached to them. We most commonly observed pre-registration differences in the body outline, bladder filling and positions of air pockets. In the current work, a normalized mutual information-based multi-resolution B-spline approach was used to register each patient's MRI to the CT. From manual inspections, the general outcome was acceptable but for some patients with large deformations between the scans, the registration could still be improved. In most cases, there was a residual difference between the MRI and CT at the body contour. Figures 1(a) and (b) show a severe example of this, where the lower left and right corners of the body in the MRI scan differ from the CT scan. This affected both the patch-based and MRI_w pCTs, causing an over-estimation of the true HU value. In the γ-map in Figure 2, this is also where the fail rate is highest. It might be worthwhile using a structure guided registration [21, 25], which could allow for a less regularized deformation while still ensuring that important structures are kept in alignment. The observed differences in body outline between the MRI and CT also call for a separate investigation of the geometrical integrity of the MRI scans at the regions far from the isocenter. It is hard to say whether these differences were due to geometrical distortions or simply variation in patient position between the scans. Nonetheless, a consequence of a successful intra-patient non-linear registration may have been that any geometric distortions were removed from the MRI and thus not taken into account in the evaluations. In our routine quality assurance of the MRI scanner, we typically observe deviations of less than 2 mm for a region of approximately $300 \times 300 \times 300$ mm around the isocenter. This region is sufficient for delineation purposes but should perhaps be larger for pelvic MRI-only RT to ensure geometrically accurate pCTs. Several groups seem to agree that the amount of distortion in modern scanners is not an issue for MRI-only RT, even for the far-from-isocenter regions covered in pelvic scans [23].

F. Affine registrations

Another important step in the patch-based approach is the affine atlas-to-patient registration, which for the most part worked without issues. However, there were cases where one or more atlas patients were poorly aligned with the test patient MRI after the registration. Misalignment happened frequently for two patients, where the FOV covered a slightly different part of the pelvis than the rest of the patients, i.e., those where parts of the femoral heads were missing. We tested registrations with fewer degrees of freedom (rigid and similarity transforms), but achieved worse results in terms of the MAE_{vox} of the resulting pCT.

The consequence of the misalignment was that similar patches were found in wrong regions of the atlas patient, causing prediction errors. Our solution was to introduce the atlas pre-selection, which left out the most dissimilar patients before the patch search. This step improved the average prediction quality, and serves to illustrate that, for the patch-based approach, the global similarity between the atlas patients and the test patient is still important. This speaks in favor of having a large number of atlases to select from in order to increase the chance of finding a number of highly similar patients for the patch search steps. In the current work, we chose to select only the five most similar atlas patients as the basis for the pCT prediction. This number was empirically chosen, as it resulted in an average lower MAE_{vox} than when all nine atlases were used. In reality, the optimal number of atlas patients to select probably varies depending on how similar the test patient is to the atlases. Finding this optimal number or adding a more flexible atlas pre-selection strategy remains part of our future work. Another future direction could be to improve the robustness of the affine alignment to better handle the alignment of dissimilar patients. In this way, more patients could be included after the atlas pre-selection step.

Using non-linear registration is another way of enforcing similarity between the atlases and the patient. This could potentially eliminate the need for atlas pre-selection and enable a larger number of atlases to be used for prediction. There is a risk, however, that if an affine registration fails, then so will a non-linear one, which typically assumes a fairly accurate affine pre-alignment. Atlas pre-selection would then still be required, and the time needed for registration would increase substantially compared to using affine registration only. Dowling *et al.* showed that, using a combination of a large number of atlases (38), non-linear registration and a weighted fusion of the atlases based on their individual local similarity could produce accurate pCTs [21]. Our results with the patch-based approach using nine patients in the atlas and

affine registrations are close to theirs. In our previous publication, we benchmarked the patch-based approach against a similar multi-atlas approach and showed an equal or slightly better performance of the patch-based approach, using the same atlases with both approaches.

V. CONCLUSION

In this study, we showed that a patch-based multi-atlas approach could generate accurate pCTs of the male pelvic region based on conventional T_1-weighted MRI scans and affine registrations. The main cause of differences between the patch-based pCT and the CT were positional changes of air pockets and body outline which were not directly related to the prediction algorithm itself. The patch-based pCTs achieved a dosimetric accuracy similar to previously published results with a γ pass rate of 0.97 for a 1%/1mm criterion and median DVH point deviations of less than 0.4% from a CT-based dose calculation. The presented approach is a promising candidate for pCT prediction in MRI-only RT of the pelvis.

VI. ACKNOWLEDGMENTS

This work was supported by a research grant from Varian Medical Systems, Inc.

VII. DISCLOSURE OF CONFLICTS OF INTEREST

The authors have no relevant conflicts of interest to disclose.

[1] J. M. Edmund, H. M. Kjer, K. Van Leemput, R. H. Hansen, J. A. Andersen, and D. Andreasen, "A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times." Physics in medicine and biology **59**, 7501–19 (2014).

[2] J. Jonsson, M. Karlsson, M. Karlsson, and T. Nyholm, "Treatment planning using mri data: an analysis of the dose calculation accuracy for different treatment regions," Radiation Oncology **5**, 62 (2010).

[3] J. Lambert, P. B. Greer, F. Menk, J. Patterson, J. Parker, K. Dahl, S. Gupta, A. Capp, C. Wratten, C. Tang, M. Kumar, J. Dowling, S. Hauville, C. Hughes, K. Fisher, P. Lau, J. W. Denham, and O. Salvado, "MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning." Radiotherapy and Oncology **98**, 330–4 (2011).

[4] T. Boettger, T. Nyholm, M. Karlsson, C. Nunna, and J. C. Celi, "Radiation therapy planning and simulation with magnetic resonance images," **6918**, 69181C–69181C–11 (2008).

[5] B. H. Kristensen, F. J. Laursen, V. Løgager, P. F. Geertsen, and A. Krarup-Hansen, "Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours." Radiotherapy and Oncology **87**, 100–9 (2008).

[6] C. Rasch, R. Steenbakkers, and M. van Herk, "Target Definition in Prostate, Head, and Neck," Seminars in Radiation Oncology **15**, 136–145 (2005).

[7] A. Fiorentino, R. Caivano, P. Pedicini, and V. Fusco, "Clinical target volume definition for glioblastoma radiotherapy planning: magnetic resonance imaging and computed tomography," Clinical and Translational Oncology **15**, 754–758 (2013).

[8] K. Ulin, M. Urie, and J. Cherlow, "Results of a Multi-Institutional Benchmark Test for Cranial CT/ MR Image Registration," International Journal of Radiation Oncology **77**, 1584–1589 (2010).

[9] T. Nyholm, M. Nyberg, M. G. Karlsson, and M. Karlsson, "Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments." Radiation oncology (London, England) **4**, 54 (2009).

[10] M. D. Robson, P. D. Gatehouse, M. Bydder, and G. M. Bydder, "Magnetic Resonance: An Introduction to Ultrashort TE (UTE) Imaging," Journal of Computer Assisted Tomography **27**, 825–846 (2003).

[11] J. Edmund, D. Andreasen, F. Mahmood, and K. Van Leemput, "Cone beam computed tomography guided treatment delivery and planning verification for magnetic resonance imaging only radiotherapy of the brain," Acta Oncologica **54**, 1496–1500 (2015).

[12] V. Keereman, Y. Fierens, T. Broux, Y. De Deene, M. Lonneux, and S. Vandenberghe, "MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences." Journal of nuclear medicine : official publication, Society of Nuclear Medicine **51**, 812–8 (2010).

[13] A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MRI sequences with ultrashort echo time," Medical Physics **38**, 2708 (2011).

[14] Y. Berker, J. Franke, A. Salomon, M. Palmowski, H. C. W. Donker, Y. Temur, F. M. Mottaghy, C. Kuhl, D. Izquierdo-Garcia, Z. a. Fayad, F. Kiessling, and V. Schulz, "MRI-based attenuation correction for hybrid PET/MRI systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/Dixon MRI sequence." Journal of nuclear medicine : official publication, Society of Nuclear Medicine **53**, 796–804 (2012).

[15] M. Hofmann, F. Steinke, V. Scheel, G. Charpiat, J. Farquhar, P. Aschoff, M. Brady, B. Schölkopf, and B. J. Pichler, "MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration," Journal of Nuclear Medicine **49**, 1875–1883 (2008).

[16] C. M. Rank, N. Hünemohr, A. M. Nagel, M. C. Röthke, O. Jäkel, and S. Greilich, "MRI-based simulation of treatment plans for ion radiotherapy in the brain region." Radiotherapy and Oncology **109**, 414–8 (2013).

[17] J. Sjölund, D. Forsberg, M. Andersson, and H. Knutsson, "Generating patient specific pseudo-CT of the head from MR using atlas-based regression." Physics in medicine and biology **60**, 825–839 (2015).

[18] J. Uh, T. E. Merchant, Y. Li, X. Li, and C. Hua, "MRI-based treatment planning with pseudo CT generated through atlas registration." Medical physics **41**, 051711 (2014).

[19] M. S. R. Gudur, W. Hara, Q.-T. Le, L. Wang, L. Xing, and R. Li, "A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning," Physics in Medicine and Biology **59**, 6595–6606 (2014).

[20] D. Andreasen, K. Van Leemput, R. H. Hansen, J. A. L. Andersen, and J. M. Edmund, "Patch-based generation of a pseudo ct from conventional mri sequences for mri-only radiotherapy of the brain," Medical Physics **42**, 1596–1605 (2015).

[21] J. A. Dowling, J. Sun, P. Pichler, D. Rivest-Hénault, S. Ghose, H. Richardson, C. Wratten, J. Martin, J. Arm, L. Best, S. S. Chandra, J. Fripp, F. W. Menk, and P. B. Greer, "Automatic substitute {CT} generation and contouring for mri-alone external beam radiation therapy from standard {MRI} sequences," International Journal of Radiation Oncology*Biology*Physics , – (2015).

[22] J. Dowling, J. Lambert, J. Parker, O. Salvado, J. Fripp, A. Capp, C. Wratten, J. W. Denham, and P. B. Greer, "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy." International journal of radiation oncology, biology, physics **83**, e5–11 (2012).

[23] J. Korhonen, M. Kapanen, J. Keyriläinen, T. Seppälä, and M. Tenhunen, "A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer." Medical physics **41**, 011704 (2014).

[24] M. Kapanen and M. Tenhunen, "T1/T2*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning." Acta oncologica (Stockholm, Sweden) **52**, 612–8 (2013).

[25] C. Siversson, F. Nordström, T. Nilsson, T. Nyholm, J. Jonsson, A. Gunnlaugsson, and L. E. Olsson, "Technical Note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm," Medical Physics **42**, 6090–6097 (2015).

[26] M. Korsholm, L. Waring, and J. Edmund, "A criterion for the reliable use of mri-only radiotherapy," Radiation Oncology **9**, 16 (2014).

[27] D. Pasquier, N. Betrouni, M. Vermandel, T. Lacornerie, E. Lartigau, and J. Rousseau, "Mri alone simulation for conformal radiation therapy of prostate cancer: technical aspects," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE* (2006) pp. 160–163.

[28] J. Korhonen, M. Kapanen, J. Keyriläinen, T. Seppälä, L. Tuomikoski, and M. Tenhunen, "Absorbed doses behind bones with MR image-based dose calculations for radiotherapy treatment planning," Medical Physics **40**, 011701 (2013).

[29] S.-L. Chan, Y. Gal, R. Jeffree, M. Fay, P. Thomas, S. Crozier, and Z. Yang, "Automated classification of bone and air volumes for hybrid pet-mri brain imaging," in *Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on* (2013) pp. 1–8.

[30] A. Johansson, A. Garpebring, M. Karlsson, T. Asklund, and T. Nyholm, "Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information–potential application for MR-only radiotherapy and attenuation correction in positron emission tomography." Acta Oncologica (Stockholm, Sweden) **52**, 1369–73 (2013).

[31] N. Burgos, M. Cardoso, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, J. S. Duncan, D. Atkinson, S. R. Arridge, B. F. Hutton, and S. Ourselin, "Attenuation correction synthesis for hybrid PET-MR scanners," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Lecture Notes in Computer Science, Vol. 8149, edited by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Springer Berlin Heidelberg, 2013) pp. 147–154.

[32] N. J. Tustison, B. B. Avants, P. Cook, Y. Zheng, A. Egan, P. Yushkevich, J. C. Gee, *et al.*, "N4itk: improved n3 bias correction," Medical Imaging, IEEE Transactions on **29**, 1310–1320 (2010).

[33] L. Nyul, J. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," Medical Imaging, IEEE Transactions on **19**, 143–150 (2000).

[34] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "elastix: a toolbox for intensity-based medical image registration," IEEE Transactions on Medical Imaging **29**, 196 – 205 (2010).

[35] F. Rousseau, P. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," Medical Imaging, IEEE Transactions on **30**, 1852–1862 (2011).

[36] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," ACM Transactions on Graphics-TOG **28**, 24 (2009).

[37] W. Shi, J. Caballero, C. Ledig, X. Zhuang, W. Bai, K. Bhatia, A. de Marvao, T. Dawes, D. O'Regan, and D. Rueckert, "Cardiac image super-resolution with global correspondence using multi-atlas patchmatch," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Lecture Notes in Computer Science, Vol. 8151, edited by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Springer Berlin Heidelberg, 2013) pp. 9–16.

[38] V.-T. Ta, R. Giraud, D. Collins, and P. Coupé, "Optimized patchmatch for near real time and accurate label fusion," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, Lecture Notes in Computer Science, Vol. 8675, edited by P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe (Springer International Publishing, 2014) pp. 105–112.

[39] N. Otsu, "A threshold selection method from gray-level histograms," Systems, Man and Cybernetics, IEEE Transactions on **9**, 62–66 (1979).

[40] International Commission on Radiation Units and Measurements (ICRU), Report 83, "3. Special Considerations Regarding Absorbed-Dose and Dose–Volume Prescribing and Reporting in IMRT," Journal of the ICRU **10**, 27–40 (2010).

[41] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," Medical Physics **25**, 656–661 (1998).

[42] E. Spezi, D. G. Lewis, and C. W. Smith, "A dicom-rt-based toolbox for the evaluation and verification of radiotherapy plans," Physics in Medicine and Biology **47**, 4223 (2002).

[43] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," NeuroImage **46**, 726 – 738 (2009).

[44] D. Andreasen, J. M. Edmund, V. Zografos, B. H. Menze, and K. Van Leemput, "Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features," in *SPIE Medical Imaging* (International Society for Optics and Photonics, 2016) pp. 978417–978417.

Paper D

Computed Tomography synthesis from Magnetic Resonance images in the pelvis using multiple Random Forests and Auto-Context features

Daniel Andreasen[a,b], Jens M. Edmund[b], Vasileios Zografos[c], Bjoern H. Menze[c], and Koen Van Leemput[a,d]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark
[b]Radiotherapy Research Unit, Department of Oncology, Gentofte and Herlev Hospital, University of Copenhagen, Herlev, Denmark
[c]Department of Computer Science, Technische Universität München, Munich, Germany
[d]A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA

ABSTRACT

In radiotherapy treatment planning that is only based on magnetic resonance imaging (MRI), the electron density information usually obtained from computed tomography (CT) must be derived from the MRI by synthesizing a so-called pseudo CT (pCT). This is a non-trivial task since MRI intensities are neither uniquely nor quantitatively related to electron density. Typical approaches involve either a classification or regression model requiring specialized MRI sequences to solve intensity ambiguities, or an atlas-based model necessitating multiple registrations between atlases and subject scans. In this work, we explore a machine learning approach for creating a pCT of the pelvic region from conventional MRI sequences without using atlases. We use a random forest provided with information about local texture, edges and spatial features derived from the MRI. This helps to solve intensity ambiguities. Furthermore, we use the concept of auto-context by sequentially training a number of classification forests to create and improve context features, which are finally used to train a regression forest for pCT prediction. We evaluate the pCT quality in terms of the voxel-wise error and the radiologic accuracy as measured by water-equivalent path lengths. We compare the performance of our method against two baseline pCT strategies, which either set all MRI voxels in the subject equal to the CT value of water, or in addition transfer the bone volume from the real CT. We show an improved performance compared to both baseline pCTs suggesting that our method may be useful for MRI-only radiotherapy.

Keywords: radiotherapy, magnetic resonance imaging, pseudo CT, CT synthesis, random forest, auto-context

1. INTRODUCTION

In recent years, the interest in eliminating the planning computed tomography (CT) from the workflow in external beam radiotherapy (RT) has increased.[1–4] Instead, the entire workflow would be based on magnetic resonance imaging (MRI), thus eliminating systematic registration uncertainties between the MRI and CT[5,6] and simplifying the treatment chain. It is a difficult task to exclude the CT, since MRI does not contain information on electron density which is needed for dose calculations. Furthermore, the signal from compact bone in conventional MRI is weak or missing due to the majority of short T_2 components.[7] This prohibits 2D-based patient setup using bone anatomy. By synthesizing a so-called pseudo CT (pCT) from the MRI scan, the above issues would be solved, making an MRI-only workflow feasible.

Several approaches for CT synthesis have been proposed and in general, the methods can be split into two categories: 1) voxel-based and 2) atlas-based. With the former, a classification or regression model is trained

Correspondence to Daniel Andreasen, E-mail: dana@dtu.dk

based on individual voxel intensities of pre-acquired MRI/CT pairs from a number of patients.[2,8–10] Other variants in this category include unsupervised methods based on MRI physics knowledge.[11,12] The voxel-based methods generally require specialized ultra short echo time MRI sequences for acquiring the bone signal, or manual delineations to overcome the bone/air intensity-ambiguity. This either prolongs the patient scan time or puts additional workload on the personnel. In recent publications, many variants of methods from the atlas-based category have been proposed.[13–19] Here, pre-acquired and co-registered MRI/CT pairs are used as atlases that describe the relation between MRI and CT. To predict a pCT, the atlas MRIs are first deformably registered with the patient MRI. The estimated deformations are then applied to the atlas CTs and finally the deformed CTs are fused to provide the pCT estimate. These methods rely on conventional MRI sequences and the accuracy is generally high. One difficulty, however, is that the use of deformable registrations can be slow, especially with multiple atlases, and may fail if the subjects are very dissimilar. Using pattern recognition based on patch similarities can remove some of the dependency on an accurate deformable registration[13] and a faster affine registration can even be used instead.[20] Still, the registration is a computationally demanding step, which must be performed for each atlas every time a pCT is predicted. Especially since the trend seems to go towards adding more atlases to the database, i.e., 15-38 patients[16,17,19] this can become a time-consuming task.

Methods using a random forest[21] (RF) have previously been proposed for predicting one MRI contrast from another.[22,23] This task differs somewhat from CT synthesis because the bone/air ambiguity does not need to be solved; the prediction model only needs to identify and predict the lack of signal in these regions, but is not required to distinguish whether it is caused by bone or air. In this work, we use RF models to predict pCTs based on conventional T_1-weighted MRI scans without registrations during pCT synthesis. We overcome intensity-ambiguities by using local texture descriptors, spatial features and edge information. Furthermore, inspired by recent works in tumor and organ segmentation,[24,25] we use the concept of auto-context (AC)[26] to iteratively learn and improve context features. We use a combination of classification RFs for learning context features and a regression RF to ultimately predict a pCT.

By far most work in CT synthesis has been on generating pCTs of the brain, although a few publications have focused on the pelvis.[8,14,16,17] This region is theoretically more challenging due to the greater inter-patient anatomical variability and because of between-scan non-rigid organ movement. We apply our method in this challenging region, and evaluate and compare the resulting pCTs to baseline pCT approaches both in terms of the voxel-wise prediction accuracy and the radiologic error.

2. METHODS

2.1 Data Collection and Pre-processing

MRI and CT scans of 10 prostate RT patients were retrospectively obtained from a Philips 1 T open MR scanner using a body coil and from a Philips Brilliance Big Bore CT scanner. T_1-weighted scans were acquired with a gradient echo sequence (TE/TR = 10/623 ms) and a voxel resolution of $0.8 \times 0.8 \times 4.4$ mm for an in-plane matrix between 528×528 and 640×640 voxels and $16 - 24$ slices. The CT scans were acquired using a standard protocol for pelvic scans (120 kVp, 232-503 mAs) with a voxel resolution between $0.78 \times 0.78 \times 2$ mm and $1.4 \times 1.4 \times 2$ mm, for an in-plane matrix of 512×512 voxels and 129-199 slices. The patients were fixed in treatment position during both the MRI and CT scanning. The T_1-weighted scans were bias field corrected using the N4 algorithm[27] after which the intensities were normalized to a common range using a histogram matching approach.[28] To generate a ground truth relationship between the MRI and CT scans, which accounts for body outline and organ deformations between the scans, the MRI scans were deformably registered and resampled to their corresponding CT using Elastix.[29] For this purpose, a multi-resolution B-spline transformation model and normalized mutual information was used. Note that this registration is carried out before training the RFs and is not a part of the pCT prediction step. Finally, in order to remove noise without severely blurring edges, a simple 2D median filter was applied slice-wise to the MRI scans.

2.2 Initial Feature Extraction

Using the MRI intensity information directly to train a prediction model may limit its general applicability because the absolute MRI intensities are irreproducible and varies between patients, scans and scanners. For this reason, we use features that to some degree are independent of linear gray-scale changes. We use Local

Binary Pattern (LBP) features[30] to capture textural information. For each voxel, we do a relational comparison of the mean value of a cuboidal region centered on that voxel with mean values of 26 cuboidal regions positioned on a sphere around that voxel. This results in a 26 digit binary number which is converted to a decimal texture feature (dubbed LBP_μ). We also incorporate a variant of the LBP where the standard deviation of the regions is compared instead (dubbed LBP_σ). We use three LBP_μ and two LBP_σ features with differing region size (RS), 3^3 - 5^3 voxels, and sphere radius (SR), 5 - 13 voxels, to capture information at different scales. As a simple edge detector, we use the standard deviation of a $3 \times 3 \times 3$ neighborhood around each voxel. As spatial features, we use the (x, y) position of each voxel relative to the center of mass of the patient and the body outline. Figure 1 shows examples of the initial features.

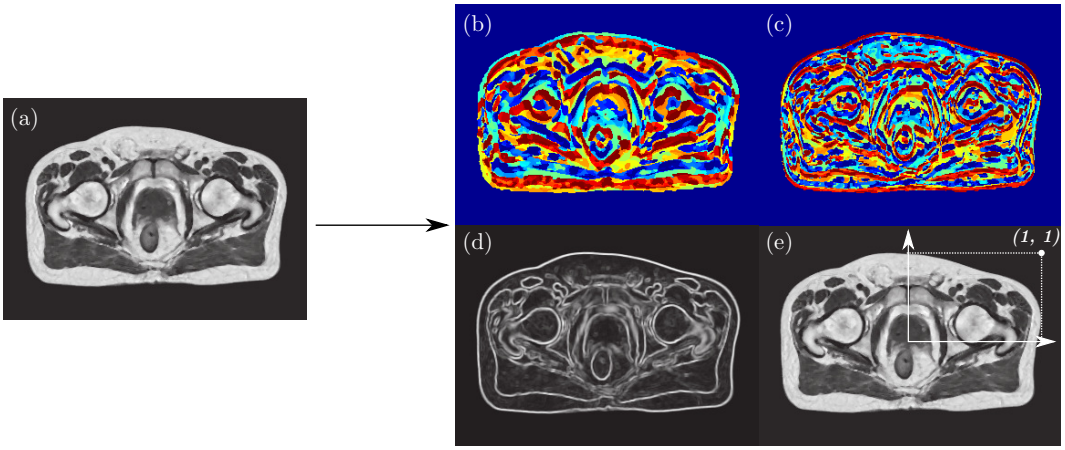


Figure 1. Features are extracted from the T_1-weighted MRI shown in (a). (b) and (c) shows examples of the LBP_μ (SR/RS = $11/5^3$ voxels) and LBP_σ (SR/RS = $7/5^3$ voxels), respectively. (d) shows the standard deviation feature. (e) shows the coordinate system used to derive the spatial features; the origin is positioned in the center of mass of the patient and the axes are scaled relative to the outer contour of the patient.

2.3 Initial Classification RF Training

In the initial step of our approach, we train a classification RF to segment an MRI scan into air, fat, soft tissue and bone classes based on the features described above. We carry out all model training in a leave-one-out manner using nine patients as training data and applying the model on the tenth. The training target segmentations are generated by thresholding the CT scans using the following criteria: air = $[-1000, -200]$ HU, fat = $]-200, 0[$ HU, soft tissue = $[0, 150]$, bone = $]150, \infty[$ HU. For each training patient, we then sample our features at 6×10^5 and 10^5 random positions within the body outline and within the bone volume, respectively. This defines our training set for the classification RF. We use an RF with 500 trees and a minimum terminal leaf size of 50 voxels. Because the bone class is relatively small compared to the others, we use a stratified sampling scheme, meaning that an equal amount of cases from each class were chosen at random during training of each tree. This is achieved by sub-sampling the bigger classes. The parameter settings were chosen as they resulted in a good compromise between the out-of-bag error rate and training time.

2.4 Auto-context Features

In the next step of our approach, we use the initial RF to create a variant of auto-context (AC) features[26] for all training patients, which are then used in combination with the initial features to train an improved classification RF. This process is repeated three times, each time using the most recently trained RF to improve the AC features. The AC features are calculated based on the three probability maps of a voxel belonging to the fat, soft tissue or bone classes. For each voxel, the feature extracts the mean probability of belonging to each class in a small region centered on that voxel, along with the same mean probabilities of six regions chosen at a fixed

distance from it in the posterior/anterior, superior/inferior and lateral directions. The AC feature extraction is illustrated in Figure 2. In this way, we encode the approximate locations and relations of the tissue classes. The AC RFs consist of 150 trees and otherwise the same parameter settings as the initial forest. For each RF at this stage, we sample the training data randomly in the same fashion as the first RF. This means that each forest is not trained on exactly the same training data as in the iteration before. Note however, that we sample with replacement so overlap can occur.

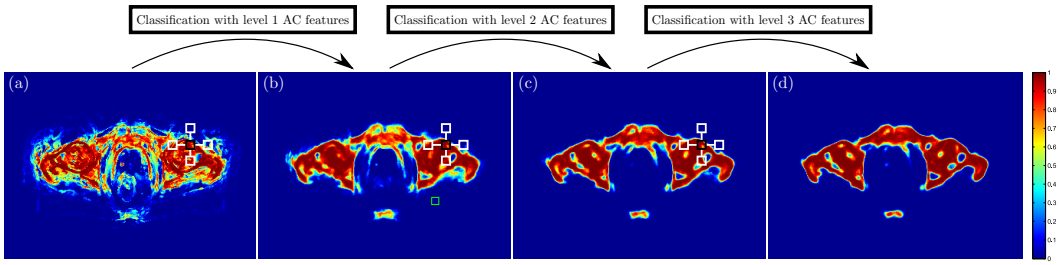


Figure 2. Examples of the output probability maps of the bone class for a test patient after sending the data through (a) the first classification RF and (b)-(d) the subsequent classification RFs with improving AC features. The AC feature for a voxel consists of the mean probability of a small region (shown in black) around the voxel along with the mean probability of 6 small regions at a fixed distance in the posterior/anterior, superior/inferior and lateral directions (shown in white for the posterior/anterior and lateral directions).

2.5 Regression Forest Training

As a final step, we train a regression RF based on the initial features and the last level of AC features. We use the non-thresholded CT scans as target values and sample our features from each training patient at 1.3×10^6 and 10^5 random positions within the body outline and bone volume, respectively. The regression forest consists of 30 trees and a minimum terminal leaf size of 70 voxels.

2.6 Benchmark Methods

As simple benchmark methods, we generate pCTs by assigning all voxels within the body outline of the MRI a bulk density value of 0 Hounsfield units (HU) corresponding to water ($\mathrm{MRI_w}$). Furthermore, we generate a pCT which in addition contains the bone volume transferred from the real CT and assigned a bulk density of the mean HU in that volume ($\mathrm{MRI_{wb}}$). This mimics a method where the bone has been manually (and perfectly) delineated on the MRI.

2.7 Evaluation

Similar to our previous work,[20] we calculate the voxel-wise mean absolute error ($\mathrm{MAE_{vox}}$) defined as:

$$\mathrm{MAE_{vox}} = \frac{1}{N}\sum_{i=1}^{N} |\mathrm{CT}_i - \mathrm{pCT}_i|, \tag{1}$$

where N is the number of voxels inside the body outline, $\mathrm{CT}(i)$ and $\mathrm{pCT}(i)$ are the HU value at voxel i inside the body outline of the real CT or pCT, respectively. Similarly, we calculate the mean error, $\mathrm{ME_{vox}}$. To provide a radiologic error measure that takes into account the attenuation properties of the tissue,[31] we also calculate the mean absolute error in water-equivalent path lengths ($\mathrm{MAE_{WEPL}}$). This is calculated along radial spokes from the center of mass of the patient to the body surface. The WEPL, l', of a spoke is calculated as:

$$l' = \sum_i \Delta l_i \times \rho_i, \tag{2}$$

with Δl_i being the physical path length traveled by a photon in voxel i and ρ_i its electron density relative to water. The value of ρ_i is found from the voxel HU value using a verified standard lookup table in the treatment planning system (Eclipse v11.0, Varian Medical Systems Inc., Palo Alto, CA). We calculate l' in 800 radial spokes towards the body outline in both the real CT and pCT. The $\mathrm{MAE}_{\mathrm{WEPL}}$ is then defined as:

$$\mathrm{MAE}_{\mathrm{WEPL}} = \frac{1}{L}\sum_{j=1}^{L} |l_j'^{CT} - l_j'^{pCT}|, \tag{3}$$

where L is the number of spokes and l_j' is the WEPL of the jth spoke in either the CT or the pCT. Similarly we calculate the mean error in WEPL, $\mathrm{ME}_{\mathrm{WEPL}}$. We perform paired t-tests to test for significant differences in the error metrics between the proposed and the baseline methods. $p < 0.05$ is considered significant.

3. RESULTS

In Figure 3, a visual comparison of the different pCTs and the real CT is shown for a representative transverse slice of one patient. The proposed method is able to capture the challenging bone region but parts of the dense bone is still underestimated as seen in the difference map (Figure 3(f)).

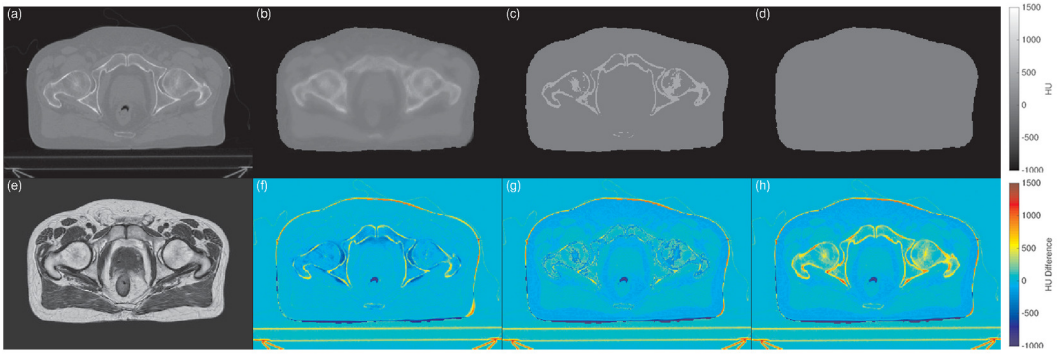


Figure 3. Transverse slices of the real CT and pCTs for one patient. (a) and (e) show the real CT and corresponding T_1-weighted MRI. (b)-(d) show the pCT generated with our method, the $\mathrm{MRI}_{\mathrm{wb}}$ and the $\mathrm{MRI}_{\mathrm{w}}$, respectively. (f)-(g) show the difference in HU between the real CT and the pCTs in (b), (c) and (d), respectively. No colorbar is shown for the MRI image.

Table 1 shows the results of the voxel-wise and radiologic evaluation. The proposed method performs well in comparison to the baseline methods with significantly lower voxel-wise and radiologic absolute errors. The $\mathrm{ME}_{\mathrm{vox}}$ reveals an overall tendency of the proposed method to overestimate the HU value. There is also an average overestimation of the WEPL, revealed by the negative $\mathrm{ME}_{\mathrm{WEPL}}$.

	Proposed	$\mathrm{MRI}_{\mathrm{wb}}$	p-value	$\mathrm{MRI}_{\mathrm{w}}$	p-value
MAE_{vox} [HU]	58 ($\sigma = 9$)	90 ($\sigma = 11$)	$< 10^{-6}$	105 ($\sigma = 12$)	$< 10^{-7}$
ME_{vox} [HU]	-10 ($\sigma = 10$)	-41 ($\sigma = 12$)	$< 10^{-5}$	-16 ($\sigma = 11$)	0.06
$\mathrm{MAE}_{\mathrm{WEPL}}$ [mm]	1.3 ($\sigma = 0.4$)	2.0 ($\sigma = 0.3$)	$< 10^{-3}$	2.5 ($\sigma = 0.3$)	$< 10^{-6}$
$\mathrm{ME}_{\mathrm{WEPL}}$ [mm]	-0.7 ($\sigma = 0.6$)	-0.4 ($\sigma = 0.5$)	0.05	0.6 ($\sigma = 0.6$)	$< 10^{-4}$

Table 1. The voxel-wise $\mathrm{MAE}_{\mathrm{vox}}$ and $\mathrm{ME}_{\mathrm{vox}}$ in HU and the $\mathrm{MAE}_{\mathrm{WEPL}}$ and $\mathrm{ME}_{\mathrm{WEPL}}$ of the water equivalent path lengths (in mm). Average value and standard deviation (σ) for the 10 patients are shown for the different methods along with the p-value from a paired t-test on the difference between the proposed method and each baseline method.

4. DISCUSSION

In this work, we explored a machine learning approach for creating a pCT of the pelvic region based on T_1-weighted MRI. The method required no specialized MRI sequences or atlases to overcome intensity ambiguities in bone/air regions of the MRI. Instead, we extracted simple textural, edge and spatial features from the MRI and trained an RF for tissue classification. As shown in Figure 2(a), the initial classification provides a rough estimate of the positions of each tissue class but some uncertainty and wrong predictions remain. The iterative refinement provided by the AC features in the next levels of RFs converges the classification towards a more correct solution. In Figure 3(b), the visual quality of the pCT comes close to the real CT. There seems to be a blurring effect in the pCT, which might be caused by the LBP and AC features only capturing information at scales larger than the voxels, i.e., from small regions. The blurring causes the bone region to smear into the soft tissue region, which might explain the average overestimation of the HU. Incorporating more features at the voxel scale could potentially remove some of the blurring effect.

The MRI_{wb} method should have a high accuracy in the bone region due to its exact geometry. However, assigning the mean HU value to the whole region means that parts of it are overestimated. On average, 65% of the voxels within the bone region are below the mean bone HU value, which illustrates the difficulties in assigning a single bulk density to segmented regions. Since, in addition, the fat voxels are not accounted for, the ME_{vox} shows an overall tendency for overestimating the HU value.

The proposed method had an average MAE_{vox} of 58 HU, which was significantly lower than the baseline methods. Other recently published methods in the multi-atlas category[16,17] reported average MAE_{vox} and ME_{vox} in the pelvic region of 36.5−40.5 and 0.6−1.9 HU, respectively. They further showed a high dosimetric accuracy of their pCTs. These methods use MRI/CT atlases from up to 38 patients and deformable registration to create pCTs. In contrast, our approach requires no atlases and no registrations at run-time, yielding a less computationally demanding method that is easier to distribute.

We have previously seen that comparing pCT methods based only on simple voxel-wise and radiologic error metrics may not be sufficient.[20] Instead an additional dosimetric evaluation should be carried out, which can highlight if the prediction fails in dosimetrically important regions or in less relevant regions. A dosimetric evaluation remains part of our future work.

Parallel to our work, another group has published a method for CT synthesis with many similarities to ours.[32] Our method differs slightly in the features used and the way auto-context is implemented. Furthermore, their method is still dependent on rigid registrations during pCT prediction to bring their training subjects in alignment with the test patient. They use a so-called structured RF to predict full CT patches instead of just voxels and report an average MAE_{vox} of 48.1 HU for the pelvis of 22 subjects. This is better than our method, but whether it is due to a more advanced RF model or simply because of more training data is hard to say.

There are still multiple potential ways of improving the proposed method. A range of different features could be tested for improving the classification. Furthermore, the LBP and AC features used here were pre-defined in terms of the region sizes and distances between the regions. An often used strategy would be to let the RF randomly pick and learn these parameters. For efficiency, a supervoxel-based strategy could also be employed.[25]

In the current work, we used features that could make the method broadly applicable on data from different scanners. Still, the features that depend on e.g., the standard deviation may describe a specific noise level in the training scans, which may be different in other scanners. Furthermore, with the rather limited amount of data available to train our models, there is a risk of overfitting to the data from our scanner, especially as more levels of auto-context are used. Testing the method's robustness using data from various scanners as well as training models based on more data remains part of our future work.

5. CONCLUSION

We presented a method for creating synthetic CT scans from MRI using random forests and textural, spatial and edge information. Furthermore, we iteratively learned auto-context features to improve the predictive performance of our models. The method worked on conventional T_1-weighted MRI scans and required no registrations at run-time. Our results indicate an improved performance compared to simple bulk density assigned pCTs but further evaluations and comparisons with more advanced methods are still needed.

ACKNOWLEDGMENTS

This work was supported by a research grant from Varian Medical Systems, Inc.

REFERENCES

[1] Kristensen, B. H., Laursen, F. J., Løgager, V., Geertsen, P. F., and Krarup-Hansen, A., "Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours.," *Radiotherapy and Oncology* **87**, 100–9 (4 2008).

[2] Edmund, J. M., Kjer, H. M., Van Leemput, K., Hansen, R. H., Andersen, J. A., and Andreasen, D., "A voxel-based investigation for mri-only radiotherapy of the brain using ultra short echo times.," *Physics in medicine and biology* **59**, 7501–19 (12 2014).

[3] Jonsson, J. H., Karlsson, M. G., Karlsson, M., and Nyholm, T., "Treatment planning using mri data: an analysis of the dose calculation accuracy for different treatment regions," *Radiation Oncology* **5**, 62 (1 2010).

[4] Lambert, J., Greer, P. B., Menk, F., Patterson, J., Parker, J., Dahl, K., Gupta, S., Capp, A., Wratten, C., Tang, C., Kumar, M., Dowling, J., Hauville, S., Hughes, C., Fisher, K., Lau, P., Denham, J. W., and Salvado, O., "Mri-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of mri-based dose planning.," *Radiotherapy and Oncology* **98**, 330–4 (3 2011).

[5] Ulin, K., Urie, M. M., and Cherlow, J. M., "Results of a multi-institutional benchmark test for cranial CT/MR image registration," *International Journal of Radiation Oncology*Biology*Physics* **77**(5), 1584 – 1589 (2010).

[6] Nyholm, T., Nyberg, M., Karlsson, M. G., and Karlsson, M., "Systematisation of spatial uncertainties for comparison between a mr and a ct-based radiotherapy workflow for prostate treatments.," *Radiation Oncology* **4**, 54 (1 2009).

[7] Robson, M. and Gatehouse, P., "Magnetic resonance: An introduction to ultrashort te (ute) imaging," *Journal of computer assisted tomography* **27**(6), 825–846 (2003).

[8] Korhonen, J., Kapanen, M., Keyriläinen, J., Seppälä, T., and Tenhunen, M., "A dual model hu conversion from mri intensity values within and outside of bone segment for mri-based radiotherapy treatment planning of prostate cancer," *Medical Physics* **41**(1), – (2014).

[9] Johansson, A., Karlsson, M., and Nyholm, T., "CT Substitute Derived from MR Sequences with Ultrashort Echo Time," *Medical Physics* **38**(5), 2708–2714 (2011).

[10] Rank, C. M., Hünemohr, N., Nagel, A. M., Röthke, M. C., Jäkel, O., and Greilich, S., "MRI-based simulation of treatment plans for ion radiotherapy in the brain region," *Radiotherapy and Oncology* **109**(3), 414 – 418 (2013).

[11] Keereman, V., Fierens, Y., Broux, T., De Deene, Y., Lonneux, M., and Vandenberghe, S., "Mri-based attenuation correction for pet/mri using ultrashort echo time sequences.," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **51**, 812–8 (5 2010).

[12] Berker, Y., Franke, J., Salomon, A., Palmowski, M., Donker, H. C. W., Temur, Y., Mottaghy, F. M., Kuhl, C., Izquierdo-Garcia, D., Fayad, Z. a., Kiessling, F., and Schulz, V., "Mri-based attenuation correction for hybrid pet/mri systems: a 4-class tissue segmentation technique using a combined ultrashort-echo-time/dixon mri sequence.," *Journal of nuclear medicine* **53**, 796–804 (5 2012).

[13] Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schölkopf, B., and Pichler, B. J., "MRI-Based Attenuation Correction for PET/MRI: A Novel Approach Combining Pattern Recognition and Atlas Registration," *Journal of Nuclear Medicine* **49**, 1875–1883 (2008).

[14] Dowling, J. A., Lambert, J., Parker, J., Salvado, O., Fripp, J., Capp, A., Wratten, C., Denham, J. W., and Greer, P. B., "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy," *International journal of radiation oncology, biology, physics* **83**, e5–e11 (05 2012).

[15] Uh, J., Merchant, T. E., Li, Y., Li, X., and Hua, C., "MRI-based treatment planning with pseudo CT generated through atlas registration," *Medical Physics* **41**(5), – (2014).

[16] Siversson, C., Nordström, F., Nilsson, T., Nyholm, T., Jonsson, J., Gunnlaugsson, A., and Olsson, L. E., "Technical note: Mri only prostate radiotherapy planning using the statistical decomposition algorithm," *Medical Physics* **42**(10), 6090–6097 (2015).

[17] Dowling, J. a., Sun, J., Pichler, P., Rivest-Hénault, D., Ghose, S., Richardson, H., Wratten, C., Martin, J., Arm, J., Best, L., Chandra, S. S., Fripp, J., Menk, F. W., and Greer, P. B., "Automatic substitute ct generation and contouring for mri-alone external beam radiation therapy from standard mri sequences," *International Journal of Radiation Oncology*Biology*Physics* (2015).

[18] Burgos, N., Cardoso, M. J., Modat, M., Pedemonte, S., Dickson, J., Barnes, A., Duncan, J. S., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., "Attenuation correction synthesis for hybrid pet-mr scanners," in [*Medical image computing and computer-assisted intervention: MICCAI 2013*], **16**, 147–154 (1 2013).

[19] Burgos, N., Cardoso, M. J., Guerreiro, F., Veiga, C., Modat, M., McClelland, J., Knopf, A.-c., Punwani, S., Atkinson, D., Arridge, S. R., Hutton, B. F., and Ourselin, S., "Robust ct synthesis for radiotherapy planning: Application to the head and neck region," in [*Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*], Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., eds., *Lecture Notes in Computer Science* **9350**, 476–484, Springer International Publishing (2015).

[20] Andreasen, D., Van Leemput, K., Hansen, R. H., Andersen, J. A., and Edmund, J. M., "Patch-based generation of a pseudo ct from conventional mri sequences for mri-only radiotherapy of the brain," *Medical physics* **42**(4), 1596–1605 (2015).

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

[22] Chen, M., Jog, A., Carass, A., and Prince, J. L., "Using image synthesis for multi-channel registration of different image modalities," in [*Proc SPIE Int Soc Opt Eng.*], Ourselin, S. and Styner, M. A., eds., **136**, 94131Q (3 2015).

[23] Jog, A., Carass, A., Pham, D. L., and Prince, J. L., "Random forest flair reconstruction from t1, t2, and pd-weighted mri," in [*2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*], 1079–1082, IEEE (4 2014).

[24] Qian, C., Wang, L., Yousuf, A., Oto, A., and Shen, D., "In vivo mri based prostate cancer identification with random forests and auto-context model," in [*Machine Learning in Medical Imaging*], Wu, G., Zhang, D., and Zhou, L., eds., *Lecture Notes in Computer Science* **8679**, 314–322, Springer International Publishing (2014).

[25] Zografos, V., Valentinitsch, A., Rempfler, M., Tombari, F., and Menze, B., "Hierarchical multi-organ segmentation without registration in 3d abdominal ct images," *MCV workshop (MICCAI 2015)* (2015).

[26] Tu, Z., "Auto-context and its application to high-level vision tasks," in [*Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*], 1–8, IEEE (2008).

[27] Tustison, N. J., Avants, B. B., Cook, P., Zheng, Y., Egan, A., Yushkevich, P., Gee, J. C., et al., "N4itk: improved n3 bias correction," *Medical Imaging, IEEE Transactions on* **29**(6), 1310–1320 (2010).

[28] Nyul, L., Udupa, J., and Zhang, X., "New variants of a method of MRI scale standardization," *Medical Imaging, IEEE Transactions on* **19**, 143–150 (Feb 2000).

[29] Stefan Klein and Marius Staring and Keelin Murphy and Max A. Viergever and Josien P.W. Pluim, "elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging* **29**, 196 – 205 (January 2010).

[30] Ojala, T., Pietikäinen, M., and Mäenpää, T., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7), 971–987 (2002).

[31] Rank, C. M., Hünemohr, N., Nagel, A. M., Röthke, M. C., Jäkel, O., and Greilich, S., "Mri-based simulation of treatment plans for ion radiotherapy in the brain region.," *Radiotherapy and Oncology* **109**, 414–8 (12 2013).

[32] Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., and Shen, D., "Estimating ct image from mri data using structured random forest and auto-context model," *IEEE Transactions on Medical Imaging* **0062**(c), 1–1 (2015).