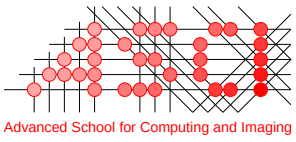
Automatic Patient Modeling for Hyperthermia Treatment Planning of Head and Neck Cancer

Valerio Fortunati

Colophon

This book was typeset by the author using LaTeX2$_\varepsilon$. The main body of the text was set using a 10-points Computer Modern Roman font. All graphics and images were included formatted as Encapsulated PostScript (^{TM}Adobe Systems Incorporated).

Cover design by the author and Daniele Moraldi. The graphics on the cover represents the dualism between technology, in this case automation, and humans.



The research described in this thesis was carried out at the Erasmus MC – University Medical Center Rotterdam (Rotterdam, the Netherlands), under the auspices of the Advanced School for Computing and Imaging (ASCI): dissertation series number 325.

This research is supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO) and partly funded by the Ministry of Economic Affairs (project number 10846).

Financial support for publication of this thesis was kindly provided by the Department of Radiology of Erasmus MC University Medical Center Rotterdam, and the Erasmus University Rotterdam, the Netherlands, the ASCI graduate school, Dutch Technology Foundation STW and Quantib BV.

ISBN: 978-90-6464-867-0

Printed by GVO Drukkers en Vormgevers BV.

Automatic Patient Modeling for Hyperthermia Treatment Planning of Head and Neck Cancer

Automatische patiënt modellering voor hyperthermie planning bij de behandeling van hoofd-hals kanker

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 10 juni 2015 om 13:30 uur

door

Valerio Fortunati

geboren te Rome, Italië

Promotiecommissie

Promotor:	**Prof.dr. W.J. Niessen**
Overige leden:	**Prof.dr. G. C. van Rhoon**
	Prof.dr.ir. J. P. W. Pluim
	Prof.dr. B.W. Raaymakers
Copromotors:	**Dr.ir. T. van Walsum**
	Dr. J.F. Veenland

Contents

Chapter One

Introduction

Always design a thing by considering it in its next larger context a chair in a room, a room in a house, a house in an environment, an environment in a city plan.

— Eliel Saarinen(1873–1950 A.C.)

1.1 Hyperthermia treatment

Head and neck (H&N) cancer is the term used to describe a wide range of malignant tumors originating in the upper airways and swallowing tracts. In 2012, this disease accounted for approximately 5% of all cancers worldwide, with 680,000 new cases diagnosed and 370,000 recorded deaths, resulting as the seventh most common cancer worldwide [1]. The H&N region is often divided in subregions containing different tumor sites, as depicted in Figure 1.1: the oral cavity, paranasal sinus and nasal cavity, salivary glands, nasopharynx, oropharynx, hypopharynx (bottom part of the throat) and larynx. Tumors in the eyes, brain and skin are generally not considered H&N cancer.

Prognosis for patients with advanced head and neck cancer is very poor and treatment of this type of cancer is challenging [25,77]. The standard treatment of advanced H&N tumors is radiochemotherapy which generally results in a 5-year survival rate of 20% to 65% depending on the tumor stage and primary site and is associated with a high treatment toxicity [25]. Severe side effects are often reported, such as loss of swallowing and salivary function leading to difficulties in speaking and eating.

For many tumor sites, adding local hyperthermia treatment to chemotherapy and radiotherapy (RT) has been shown to significantly improve the treatment outcome without increasing treatment toxicity [5,37,57,60,112,124]. Local hyperthermia treatment (HT) consists of elevating the tumor temperature to a temperature between 40° and 44° Celsius. The goal of an hyperthermia treatment is to elevate the temperature of the tumor and the surrounding region at risk to a temperature as high as 43° while sparing the surrounding thermo-sensitive tissues from an excessive thermal dose.

In the H&N region, for tumors located in superficial sites (< 4 cm from outer or inner skin) the effectiveness of hyperthermia has been demonstrated in phase III clinical trials [54, 55, 121]. The findings of these studies are summarized in Table 1.1.

In order to deliver thermal dose in deep (> 4 cm from the skin surface) H&N regions the Hypercollar was developed [90]. The first version of this device, the

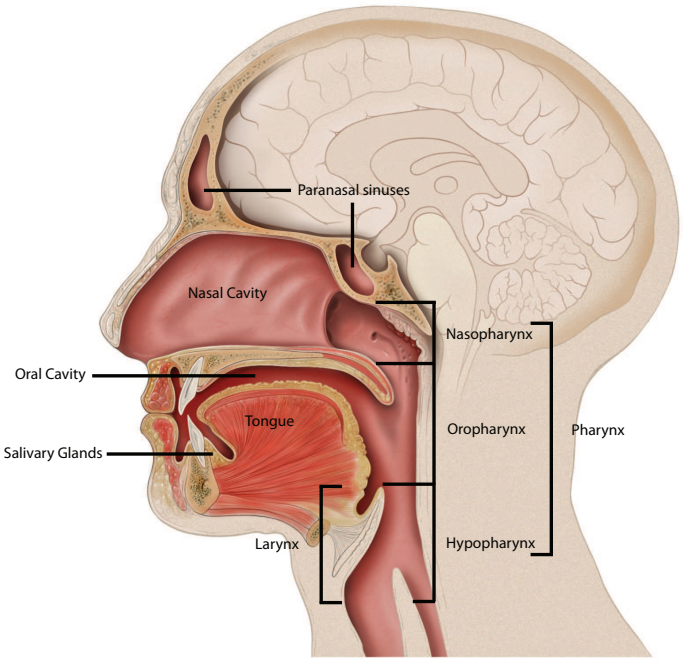


Figure 1.1. Anatomical regions of the head and neck.

Table 1.1. Summary of phase III clinical trial results for hyperthermia treatment of H&N cancer. Abbreviations: CR = Complete response, LC = 5 years local control, OS = 5 years overall survival, RT = radiotherapy, HT = hyperthermia, Loc.Adv. = locally advanced.

Study	Nr. Patients	Tumor	RT only	RT + HT	Toxicity
[121]	41	Loc.Adv.	LC:24%	LC:69%	Similar
			OS:0%	OS:54%	
[55]	56	Loc.Adv.	CR:42%	CR:79%	Similar
[54]	180	Nasopharynx	CR:81%	LC:96%	Similar
		carcinoma	LC:79%	LC:91%	

Hypercollar, and the new version, the Hypercollar 3D [116], are shown in Figure 1.2.

With the technological possibility of highly focused deep irradiation of the neck area [90, 116], the need for patient specific hyperthermia treatment planning (HTP) is increased. HTP is needed for dosimetry, optimization, adaptation, monitoring and evaluation of hyperthermia treatments. Accurate electromagnetic based treatment planning was therefore developed and integrated with adaptive hyperthermia treatment planning [97]. More recently the use of thermal based treatment planning was also investigated [128]. HTP of the H&N is based on 3D anatomical models extracted

from CT images [89,97]. The models are commonly generated by manual delineation of the relevant structures in axial CT slices. It is well known that manual delineation is labor intense (5-8 h for a 3D patient model) and suffers from intra- and interobserver variability. Thus, automation of this process is highly demanded.

Recent literature shows that the higher soft tissue contrast of MR images allows for more accurate delineation of target [2,45,96,117] and normal [126] tissues in the H&N region. Additionally, a better visualization of soft-tissues can potentially facilitate the process of automatic segmentation. The workflow of hyperthermia treatment planning is shown in Figure 1.3 where we highlight the possibility of integrating MR and CT images to improve the patient models. CT images are still required as they are geometrically more accurate than MR and they provide the tissues proton density which is a crucial input for radiotherapy treatment planning.

The automatic construction of these 3D patient models from CT and Magnetic Resonance (MR) images is the main theme of our research. In the following sections we briefly introduce the image processing methods that were developed and evaluated in this thesis, and in Section 1.4 we formulate the research questions and describe the thesis content.

1.2 Image registration

Image registration is the process of spatially aligning two or more images. In this way the image content from different images can be combined. Among the most important applications of image registration we can mention multi-modality image alignment [15] [Chapter 5 and 6 of this thesis], the comparison of follow-up datasets to base-line datasets [109], alignment of images over time [71,83,110] and atlas-based segmentation [68,123] [Chapter 2 and 4 of this thesis].

The registration of two images, a fixed image, defined in the reference coordinate frame, and a moving image which need to be aligned to the reference coordinate frame, is an optimization process. This process is generally performed iteratively by minimizing a cost function, that measures the (dis)similarity of the fixed and the (iteratively deformed) moving image, with respect to the deformation parameters. The components and parameters of this optimization process need to be selected and tuned for the application at hand. The most important components are the optimization algorithm, the cost function and the deformation model.

Different optimization algorithms for registration have been described and compared in [67]. Based on the result of the latter paper, Klein et al. developed the adaptive stochastic gradient descent optimizer [65]. This is the optimization method used in this thesis.

Common intensity-based cost functions are e.g. mean squared differences, normalized cross-correlation and mutual information [80, 94, 115]. The choice of cost function depends on the images that need to be aligned. If the the images have similar/calibrated contrast and intensity values, as for CT images, a mean squared differences similarity is sufficient to obtain a good registration. The normalized cross-correlation is a metric which is more robust to intensity variation and can e.g. be used when aligning contrast and no-contrast CT images. Mutual information-based simi-

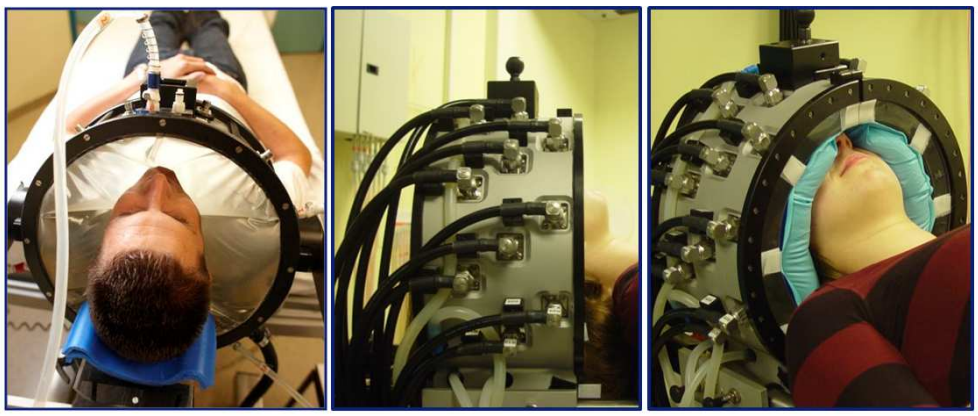


Figure 1.2. The Hypercollar and the Hypercollar 3D

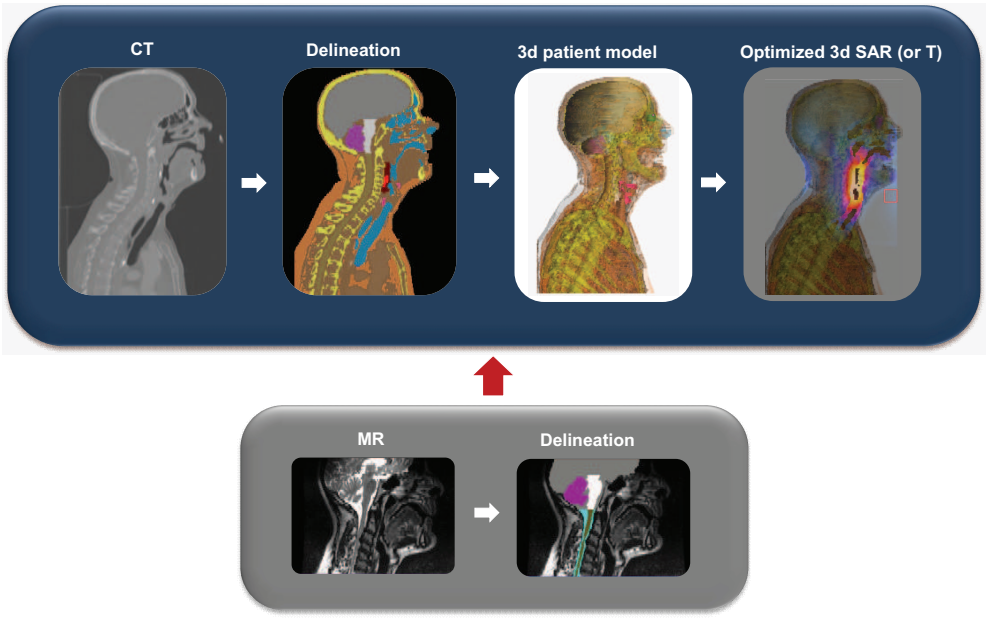


Figure 1.3. Hyperthermia treatment planning workflow and extended workflow when adding MR based information.

larity metrics are derived from information theory and measure the statistical mutual dependence between the images. These measures are robust to image intensity and contrast change and therefore they are commonly applied in multimodal registration.

The deformation parameters which are being optimized during registration describe the spatial transformation between the fixed and moving image. This transformation can be linear, e.g. rigid (translation and rotation) or affine (global scaling, rotation, translation and skewing), or non-linear. When the transformation is non-linear the registration is named non-rigid or deformable. In this thesis we focus on a non-rigid transformation model called Free Form Deformation [100]. This model is based on a parametric description of the transformation using cubic B-Spline polynomials [120]. For more details on image registration, we refer the reader to surveys of image registration that have been published in recent years [62, 108].

In this thesis, image registration was performed using *elastix* (elastix.isi.uu.nl), an open source software based on the ITK toolkit (www.itk.org) that allows to select several registration components and parameters.

1.3 Atlas-based segmentation

The goal of segmentation is to assign a semantic label to the pixels/voxels in the images. In medical image analysis, this procedure generally consists on the identification of tissues, structures or specific geometrical features (e.g. the center line of a vessel) in the image.

In clinical practice, segmentation can be performed manually or (semi-) automatically. Manual segmentation for therapy planning generally consists of manually outlining tissues/structures in a 2D plane (axial, coronal or sagittal) followed by checking the contours in the other two planes. In semi-automatic segmentation methods, the segmentation process only requires a limited input from the user to steer the results in the desired direction. Automatic segmentation methods do not require any user input to perform the segmentation. The benefits of automatic image interpretation and analysis are both in the logistics of therapy planning, and in the quality of the results. Automatic segmentation can relieve clinicians from the labor intensive aspects of their work, allowing them to focus more on other aspects. Furthermore automatic segmentation may increases the consistency and reproducibility of therapy planning.

Automatic segmentation methods differ in the way image information is used, which prior knowledge is incorporated and which kind of constraints are imposed on the segmented geometry and enforced during the segmentation. The most suitable method for a given application depends on the image data, the object imaged, and the type of desired output information.

In many medical applications, discriminating the object in the image using only image dependent features (value, textures, edges) is not sufficient to obtain the desired result. In such cases we can say that there is not a well-defined relationship between the voxel intensity and the label that should be assigned to it. This observation is obvious when we are dealing with anatomical structures rather than tissue types: different structures that are composed of the same tissue type cannot be distinguished from each another by looking only at their intensity values. More knowledge is re-

quired to segment these structures, and anatomical/spatial information needs to be considered.

Methods that can incorporate spatial information in the segmentation process include deformable models [21, 22, 61, 131], active shape and active appearance models, [9, 53, 135] and atlas-based segmentation [6, 48, 68, 132]. Deformable models are contours or surfaces which deform in order to fit the object that has to be segmented. Image-driven (edges, statistics, texture, etc.) forces and regularization forces (e.g. to constrain the result to a smooth curve or surface) are combined to steer the defomable model evolution to segment the object of interest.

Active shape models describe the shape of the object to segment as an average shape and a number of characteristic variations. This models results from the eigenvalues decomposition from a set of example shapes that the object to be segmented can assume. In the active appearance models, both the shape and appearance of the object of interest are captured. Similarly to active shape models, this information is generally represented as an average shape and intensity, and a number of characteristic variations. These models are fitted to the image to extract the object of interest.

Another way to incorporate the information needed to obtain an accurate segmentation is by using an atlas: an image with known segmentation. An atlas can be a manual segmentation of a selected image delineated by an expert clinician but can also be generated merging the information from different manually segmented images [24]. Thus an atlas contains information on the location and shapes of anatomical structures and the spatial relationship between them. Using image registration the information in the atlas can be directly propagated to the image to be segmented (target image). The process of segmenting an image by registering it to an atlas is known as atlas-based segmentation, or registration-based segmentation. In order to capture the deformation between different anatomies the registration must be non-rigid. An overview of atlas-based segmentation techniques is provided in [98].

A natural extension of this concept is multiatlas-based segmentation. In most cases the anatomical information contained in a single atlas is not sufficient to cover all possible anatomical variations. Thus, the atlas-based segmentation of a patient with an anatomy different from the one of the atlas results in an in-accurate segmentation. A solution to this problem is to use more atlases to obtain a collection of possible segmentations. In multiatlas-based segmentation techniques these segmentations are combined to obtain an optimal result [3, 6, 47, 48, 68, 132]. The combination method is generally called atlas fusion. Figure 1.4 shows a scheme of multiatlas-based segmentation, each of the results before the fusion is a single-atlas-based segmentation. It has been demonstrated that using this approach, the combined segmentation achieves an accuracy higher than any of the original segmentations [63, 98].

1.4 Purpose and contents of this thesis

The research presented in this thesis is the result of a close collaboration between the Biomedical Imaging Group Rotterdam (BIGR) and the Hyperthermia Unit (Cancer Institute) of the Erasmus University Medical Center of Rotterdam. This thesis focuses

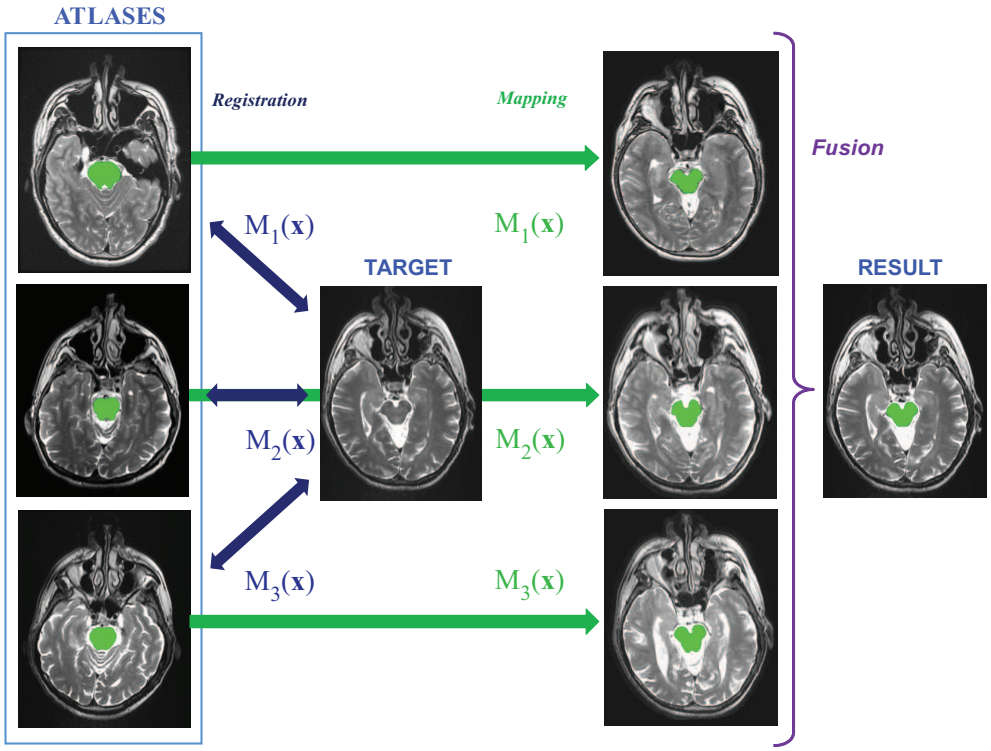


Figure 1.4. Multiatlas-based segmentation. In this example three atlsaes are registered to the target image. Then the label information is transformed to the coordinate frame of the target image. At last the singleatlas segmentation are combined in a so called multiatlas fusion.

on development and evaluation of image processing methods to improve the patient model generation in HTP of the H&N.

In this thesis we investigate automatic methods for 1) the segmentation of CT and MR images of the H&N, and 2) the registration of CT and MR images for the integration of MR into the treatment planning workflow. We introduce automatic segmentation for the generation of 3D patient models aiming to minimize human intervention in the generation of patient models, while maintaining the accuracy of delineation comparable with the current clinical standard. Automatic methods were developed and evaluated for the segmentation of both MR and CT images. In order to evaluate whether the use of these method was advantageous the accuracy of these methods is compared with manual delineations and their interobserver variability.

In Chapter 2 we investigate whether CT images can be segmented automatically with a sufficient accuracy. For this purpose, an atlas-based automatic segmentation

method is developed and evaluated using manual delineations as ground truth. This method combines atlas-based segmentation with an intensity model using graph-cuts. First, we evaluated the advantages of using the proposed approach with respect to multiatlas-based segmentation alone. Next, the influence of different labelers on the method's accuracy was quantified and compared with multiatlas-based segmentation and with the interobserver variability.

To determine whether the automatic segmentation method of Chapter 2 can be introduced in the clinic, in Chapter 3 we compare the impact of manual and automatic segmentation on HTP. The effect of using this method on HT dose distribution and HTP quality is evaluated. In addition the effect of using automatic segmentation over manual delineations was compared to other sources of patient model uncertainties.

In Chapter 4 we investigate improvements to the segmentation method of Chapter 2 applied to MR images. A more advanced atlas fusion approach was used to improve the atlas-based segmentation i.e. using local weighted atlas fusion and local searching technique. We separately evaluate how the different components of the method influence the accuracy of the final segmentation, i.e. using different weighting strategies and parameters and using the intensity model. The accuracy of the method is also compared with the interobserver variability.

To align CT and MR images we evaluate different registration approaches in order to investigate the best strategy to adopt. In Chapter 5 we develop and evaluate a strategy for deformable registration of CT and MR images. In the context of radiotherapy treatment planning rigid registration is widely used for this purpose. The use of rigid registration is possible because the patients are scanned in the same position in CT and MR, using patient specific immobilization equipment. In this chapter, we aim to prove that deformable registration offer a more accurate image alignment than rigid registration even when immobilization equipment is used. Thus, we compared the accuracy of deformable and rigid registration.

Deformable registration can potentially obviate the use of immobilization equipment because of its high flexibility. This aspect is clinically relevant because the immobilization equipment causes severe patient discomfort. In Chapter 6 we investigate whether deformable registration can really obviate the use of immobilization equipment. This is done by optimizing and evaluating deformable registration methods under different acquisition scenarios: with and without the use of immobilization equipment during the acquisition.

Finally, in Chapter 7 we discuss the findings of each chapter, the achievements of our research and future directions of research.

Chapter Two

Tissue Segmentation of Head and Neck CT Images for Treatment Planning: A Multiatlas Approach Combined with Intensity Modeling.

In the twenty-first century, the robot will take the place which slave labor occupied in ancient civilization.

— Nikola Tesla (1856-1943)

Abstract — Hyperthermia treatment of head and neck tumors requires accurate treatment planning, based on 3D patient models that are derived from segmented 3D images. In this chapter, we present and evaluate an automatic segmentation algorithm for CT images of the head and neck.

The proposed method combines anatomical information, based on atlas registration, with local intensity information in a graph cut framework. The method is evaluated with respect to ground truth manual delineation and compared with multiatlas-based segmentation on a dataset of 18 labeled CT images. On a subset of 13 labeled images, the influence of different labelers on the method's accuracy is quantified and compared with the interobserver variability.

The proposed method has a significantly better accuracy than the multiatlas-based method for the eye vitreous humor. For the majority of the tissues (8/11) the segmentation accuracy of the proposed method is approaching the interobserver agreement. Our method showed better robustness to variations in atlas labeling compared with multiatlas segmentation. Moreover the method improved the segmentation reproducibility compared with human observer's segmentations.

The proposed framework provides an accurate automatic segmentation of head and neck tissues in CT images for the generation of 3D patient models, which improves reproducibility, and substantially reduces labor involved in therapy planning.

Based upon: **V Fortunati**, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M Paulides and T van Walsum, "Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling", *Med Phys*, vol. 40, no. 7, pp. 071905, Jul. 2013.

2.1 Introduction

During local hyperthermia treatment (HT), tumor temperature is elevated to $40°C - 44°C$. The effectiveness of radiotherapy (RT) is significantly enhanced when combined with HT [5, 60, 112]. Phase III clinical studies show that hyperthermia is also effective in the head and neck (H&N) region [54,55,121]. Therefore, a site-specific hyperthermia applicator was developed for precise HT of tumors in the entire H&N region using electromagnetic fields. This device, called the Hypercollar, enables deep heating (> 4 cm under the skin) of H&N cancers [8, 89, 90]. Simultaneously, electromagnetic-based hyperthermia treatment planning (HTP) was developed to exploit the specific heat-focusing capabilities of the novel applicator. Recently, HTP-guided adaptive treatment was also developed [20, 36].

HTP requires the computation of patient-specific 3D anatomical models. These models directly affect the settings of the HTP and, improved accuracy of the models directly benefits treatment. In addition, the decision as to whether or not to treat is based on pre-treatment planning. In clinical practice, H&N patient models are generated by manual tissue delineation from CT images [89]. However, the problems with manual delineation are well known, i.e. it is labor intensive, time consuming, and tedious for the operator. Additionally, interobserver and intraobserver variability may affect the reproducibility of treatment conditions. Therefore, there is a great interest in automated procedure for anatomical model generation. However, the generation of accurate anatomical models is particularly challenging due to 1) the complexity and variability of the underlying anatomy, 2) the large number of anatomical structures in a relatively small area, and 3) the low contrast between the relevant tissues in CT images. Moreover the presence of tumors increases the anatomical variability between subjects, i.e. tumor position, volume, and shape vary, and the tissues surrounding the tumor are deformed differently.

The present study aims to develop and evaluate an automatic segmentation algorithm for the tissues relevant in HTP; these are, e.g., the critical tissues, to which the treatment temperature should be restricted, and the cricoid and thyroid cartilage for which di-electrical and thermal parameters differ from the surrounding region. The critical tissues are encephalon (composed of cerebrum, cerebellum and brainstem); spinal cord; [107] eye tissues (consisting of vitreous humor, sclera, cornea, lens and optical nerve); [30] and thyroid gland. Figure 2.1 shows an example of 3D patient model. Since HTP needs a complete 3D patient model all the tissue inside of the skull are labeled as cerebrum.

Atlas-based segmentation methods [98] are popular and are used for H&N CT image segmentation. An atlas is defined as an image to which an expert has assigned a label, corresponding to a tissue, to each voxel. Using non-rigid registration spatial correspondence is obtained between a target image, with undefined segmentation, and the atlas image. Subsequently, its corresponding label image is deformed to the coordinate framework of the target image. Multiple atlas images can be deformed to the target image space and their results can be combined [51, 132] to obtain a more accurate segmentation, also increasing the robustness to anatomical variation.

Segmentation of brainstem, spinal cord and thyroid from H&N CT images has been investigated in the context of RT treatment planning. Atlas-based techniques

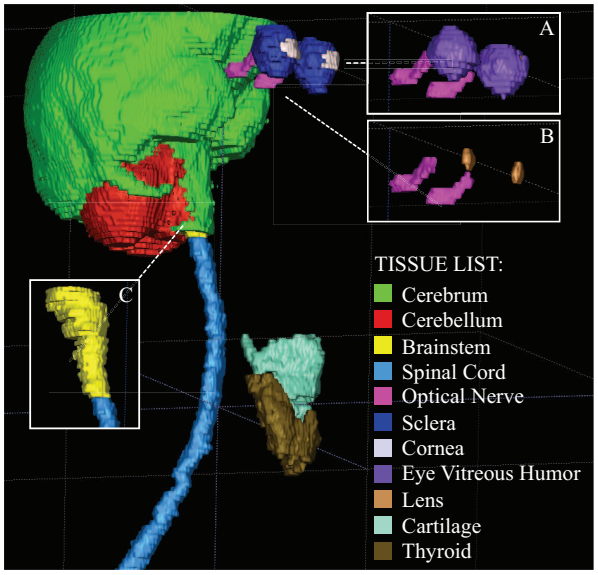


Figure 2.1. 3D anatomical model example. Correspondence between colors and tissue is given in the tissue list. The Boxes A and B show details for the eye tissues (A shows the eye vitreous humor, and B the lens). The box C shows the brainstem.

were adopted for these tissues. For instance, Han et al. used a block matching approach to register regions around the boundaries of the structures of interest, followed by a shape constrained dense deformable registration step to refine alignment of the boundaries [48]. After each atlas was aligned to the target image, STAPLE was used to combine the different atlas results [132]. A tool (ABAS, version 1.1, Elekta-CMS Software) based on the previous method was clinically evaluated by Teguh et al. for combined brainstem and spinal cord segmentation [114]. Gorthi et al. presented and evaluated an active contour-based atlas registration method [46]; in their work the brainstem was modeled by active contours using a level set representation enriched with polarity information. The registration was driven by region-based forces which steer the atlas contour towards regions with a similar mean intensity value. Qazi et al. combined atlas registration, model-based segmentation and voxel classification using multiple features, in a sequential approach, for brainstem segmentation [95]; this approach outperformed all previous methods for brainstem segmentation. Thyroid segmentation was also recently addressed by Chen et al. [23]. Different atlas-based segmentation methods were compared by varying the selection criteria and the atlas combination approach. It was shown that a weighted combination of atlas segmentations using the correlation coefficient as similarity measure outperforms standard combination methods such as majority voting [51] and STAPLE [132].

Segmentation of the eye tissue from CT images was addressed using semi-automatic model-based approaches. Bekes et al. presented a method to segment eyeball, lens and optical nerves [10]. Starting from an initial seed point provided by the user, a

geometrical model was fitted to each of the tissues of interest. A sphere was used to model the eyeball, a cylinder for the optical nerve and an ellipsoid for the lens. The method was evaluated using a STAPLE combination of the automatic results as ground truths, rather than using manual delineations. Later, Cuadra et al., presented segmentation of eyeball and lens [26]; this method is also semi-automatic but differs from the previous approach in that it uses an active contour algorithm guided by region-based forces to segment the lens, and boundary-based forces for the eyeball. The method was evaluated on a set of three patients.

Despite all these approaches, no comprehensive method to segment all the relevant tissues for the HTP in the H&N region is yet available. In addition, automatic methods for the segmentation of cerebrum, cerebellum, eye vitreous humor, sclera, cornea and cartilage from CT images are not yet developed. Moreover, for atlas-based approaches, the influence of using different atlases (i.e. atlases labeled by different observers) on accuracy and robustness has not yet been evaluated.

Generally, atlas-based methods show good performance for H&N CT image segmentation. However for various problems (e.g. brain structures segmentation from MR images) combining an intensity model with multiatlas registration-based segmentation improves accuracy [32, 101, 122, 123]. Intensity information can efficiently be combined with atlas segmentation using a graph [122, 123]. Graph cut approaches give the global optimum in an efficient implementation, are numerically robust, and allow to co-integrate a large range of visual features and constraints [14].

In this chapter we propose such an approach for segmentation of H&N CT images; we employ a graph cut method which combines atlas-based segmentation and intensity modeling. The method is van der Lijn et al. originally proposed for segmentation of the hippocampus [123]. We extended our previous work [33] by including more tissues, a post processing step to fuse each tissue's result, and by providing extensive evaluation.

We aim to minimize human intervention on the generation of patient models, while maintaining the accuracy of delineation comparable with the current clinical standard. The contribution of our work is threefold: 1) we developed a common framework to automatically segment a variety of tissues of different shapes and intensity; 2) we used an intensity model and atlas-based shape prior to address this challenge; and 3) we conducted extensive evaluation which included the effect of interobserver variability in the atlases used on the accuracy and robustness of the method.

2.2 Methods

The segmentation of all the tissues of interest is obtained by decomposing the problem in a set of binary sub-problems, one for each tissue. In this section we describe how each binary problem is solved (from section 2.2.1 to 2.2.4) and how the results are then combined to obtain a single segmented image comprising all tissues (in section 2.2.5).

First the model of the binary problem is provided in section 2.2.1. The segmentation is represented as an energy maximization problem. The energy to maximize is a probability which combines a voxel-wised term and a voxels interaction term which

gives the statistical relationship between neighboring voxels. The voxel-wised term is defined as *association potential* and it is described in section 2.2.2. This term gives the probability to have a foreground or background label at a certain voxel location in the target image, considering the voxels independent to each other. Such probability is obtained from a set of atlas images. Two models are built for this purpose: 1) an *intensity model* and a *spatial prior model*. The *intensity model* is built estimating the histogram of foreground and background regions sampling the entire atlas set. The *spatial prior model* is obtained by registering each atlas image to the target and then combining the registration results in a spatial probability map. The *association potential* combines the shape information contained in the atlases with the appearance of foreground and background regions. The second term of the model is defined as *interaction potential* and it is described in section 2.2.3. This term models the statistical dependency between voxels in the target image and its neighboring voxels given by a neighborhood model. The energy terms described above are combined to built a graph where each node is a voxel of the image. The maximization problem is converted to a minimization by taking the negative logarithm of the energy. Computing the minimum cut of a graph representing the energy function we can find the label configuration which minimizes the energy and so the segmentation of the target image. Figure 2.2 shows a scheme of the method used for each tissue. When a binary result is obtained for all the tissues a post-processing combination step is applied. This process is described in section 2.2.5.

2.2.1 Segmentation model

The problem of labeling a single tissue in a target image can be formulated as an a posteriori maximization problem (MAP) where we aim to estimate the label field $\mathbf{f}$ given the image information $\mathbf{i}$:

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmax}}\, p(\mathbf{f}|\mathbf{i}). \tag{2.1}$$

Assuming a binary segmentation, $\mathbf{f}$ is a vector containing labels $f_m \in 0, 1$ for every voxel m in the set $\mathcal{M}$ of voxels in the image. Vector $\mathbf{i} = \{i_m : m = 1, ..., \mathcal{M}\}$ consists of the intensity values i_m for all voxel locations $\mathcal{M}$. The joint posterior probability $p(\mathbf{f}|\mathbf{i})$ can be simplified assuming that the label f_m conditioned on the image intensity $\mathbf{i}$ depends only on the labels of its neighbors $n \in \mathcal{N}_m$. In this case we can approximate equation 1 as a Discriminative Random Field [72, 73] with one- and two-voxel clique potentials:

$$p(\mathbf{f}|\mathbf{i}) \approx \frac{1}{Z} \exp\left[\sum_{m \in \mathcal{M}} \left(\lambda_1 A(f_m, \mathbf{i}) + \sum_{n \in \mathcal{N}_m} I(f_m, f_n, \mathbf{i})\right)\right]. \tag{2.2}$$

Following the terminology of [72, 73], the association potential is denoted by $A(f_m, \mathbf{i})$ and the interaction potential by $I(f_m, f_n, \mathbf{i})$. Here λ_1 is a parameter that weights the association potential with respect to the interaction potential and Z is a normalization constant which does not influence the maximization procedure and thus it can be disregarded. The association potential is proportional to the log probability that a single voxel is assigned to a foreground or background label, given all intensity

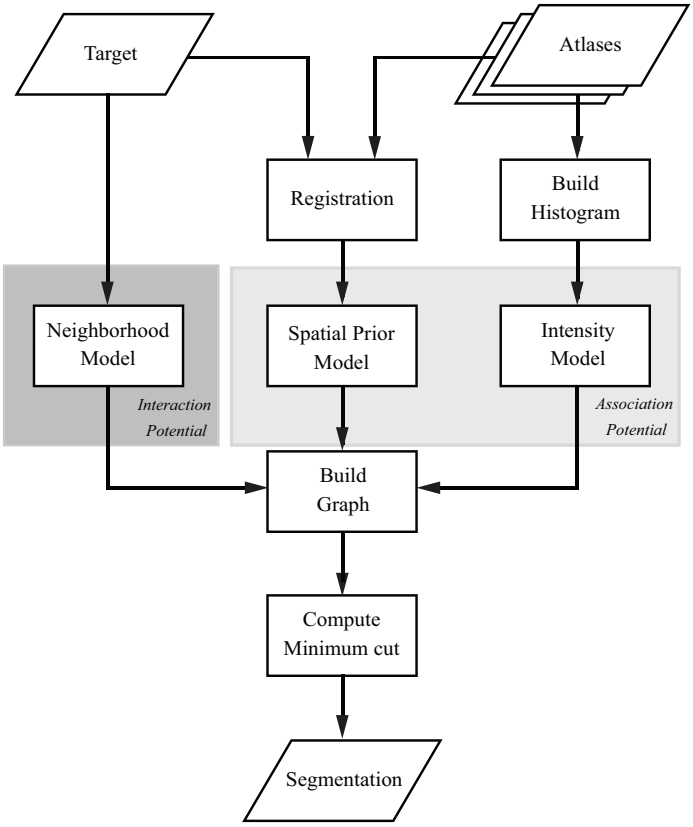


Figure 2.2. Overview of the segmentation method for each tissue.

values of the image. This term is defined for each voxel of the target image and it is based on the statistical model of the tissue shape, given by a probability map obtained by multiple atlas registration, and of tissue appearance which is related to the intensity histogram of the foreground and background regions. The interaction potential models the relation between two neighboring voxels, given their intensities. The set of neighboring voxels $n \in \mathcal{N}_m$ is defined by a neighborhood model. This term is used to promote piece-wise smooth segmentations.

2.2.2 Association Potential

According to Kumar and Herbert [73], the *association potential* has the form:

$$A(f_m, \mathbf{i}) = \ln p_A(f_m | \phi_m(\mathbf{i})). \tag{2.3}$$

The term $p_A(f_m, \phi_m(\mathbf{i}))$ is the association probability function which is used to statistically model the tissue intensity and shape of the tissue to segment (for one voxel).

The function $\phi_m(\mathbf{i})$ associates a feature to a voxel location m taking into account the image intensity $\mathbf{i}$ [73]. In the present study we assume $\phi_m(\mathbf{i}) = i_m$, so our feature is the voxel intensity at each location. The association potential is defined as [122]:

$$p_A(f_m|i_m) = \frac{p_{gp}(f_m) \cdot p_{model}(f_m|i_m)}{p_{gp}(f_m = 0) \cdot p_{model}(f_m = 0|i_m) + p_{gp}(f_m = 1) \cdot p_{model}(f_m = 1|i_m)}, \quad (2.4)$$

where $p_{gp}(f_m)$ is a global prior which sets a threshold on the probability of the foreground voxels in the entire image and $p_{model}(f_m|\mathbf{i})$ is a probability function that contains the spatial and intensity model. The term $p_{gp}(f_m)$ is added since registration is non perfect and therefore the spatial probability map obtained by multiple atlas registration may contain low probability values which occurs especially for small structures such as cornea and sclera. The global prior is defined as:

$$p_{gp}(f_m) = \begin{cases} \alpha, & \text{if } f_m = 0, \\ 1 - \alpha, & \text{if } f_m = 1, \end{cases}, \quad (2.5)$$

with parameter α between 0 and 1. If α has a value of 0.5 the association probability is equal to the model probability $p_{model}(f_m|i_m)$. A value < 0.5 increases the probability of a foreground label for all voxel locations and consequently increases the volume of the segmentation, whereas a larger value decreases the foreground probability.

The model term $p_{model}(f_m|i_m)$ contains the spatial and intensity models built from the atlas set. It can be decomposed using Bayes' rule:

$$p_{model}(f_m|i_m) = \frac{p_{int}(i_m|f_m) \cdot p_s(f_m)}{p(i_m)} = \frac{p_{int}(i_m|f_m) \cdot p_s(f_m)}{p_{int}(i_m|f_m = 0) \cdot p_s(f_m = 0) + p_{int}(i_m|f_m = 1) \cdot p_s(f_m = 1)}, \quad (2.6)$$

where the term $p(i_m)$ is expressed using the total probability theorem. Here $p_s(f_m)$ is a prior term which model the location of the tissue of interest and $p_{int}(i_m|f_m)$ is a prior term which represent the image intensity likelihood of foreground ($f_m = 1$) and background ($f_m = 0$).

2.2.2.1 Intensity model

The probability distribution $p_{int}(i_m|f_m)$ of intensity values at a certain voxel location can be estimated from the atlases $\mathcal{T}$, using a Parzen window estimator with a Gaussian kernel:

$$p(i_m|f_m) = \frac{1}{K} \sum_{k \epsilon K} \frac{1}{h\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{i_m - I_k}{h}\right)^2\right]. \quad (2.7)$$

Here h is the standard deviation of the Gaussian kernel which was empirically set to 50 Hounsfield units (HU) for all tissues, K is the number of samples, and I_k is the intensity of the k-th sample. This model is used for foreground as well as background regions. In other words the intensity model is built by estimating the histogram of

foreground and background regions, sampling the atlas images, and then comparing the intensity i_m of the current location with both the histograms. The histogram of the label (foreground or background) which has an higher number of occurrences for the intensity i_m will have an higher probability.

2.2.2.2 Spatial prior model

The spatial prior term is constructed by non-rigidly registering a set of J atlas images $\mathcal{T}$, in which the target tissue has been manually labeled, to the target image u. The registered labels of the atlas images, are then averaged to create a probability map $p_s(f_m)$:

$$p_s^{(u)}(f_m) = \left(\frac{1}{J} \sum_{t_j \in \mathcal{T}} g_m^{(u;t_j)} \right)^{\lambda_2}, \tag{2.8}$$

where $g_m^{(u;t_j)} \in \{0,1\}$ represents the atlas label of training image t_j registered to the coordinate frame of the target image u and interpolated at voxel location m. The parameter λ_2 determines the balance between the the spatial and intensity models.

A three-step registration procedure is performed. In the first step the images are coarsely aligned using reference points which are automatically detected in the neck of every image using the following procedure: a binary label image which delineates the body of the patient is computed automatically; based on the binary image an axial neck slice is detected as the slice with local minimum area, this slice is located in the upper neck region, just below the hyoid bone; the geometrical center of this slice is selected as reference point. In the second step, a rigid registration is applied using the sum of squared distances as a similarity measure. In the third step, non-rigid registration is performed using a cubic, multiresolution, B-Spline approach, with normalized cross-correlation as similarity measure. For the second and third registration steps Elastix software was used [66]. The parameter files for rigid and non-rigid registration are publicly available in the elastix website (http://elastix.isi.uu.nl/ in the section wiki). The registration parameters were optimized using an independent set of 7 images.

2.2.3 Interaction potential

As the association potential models each voxel independently, the resulting segmentation may be noisy. To obtain a smooth segmentation result, an *interaction potential* is introduced [14]:

$$I(f_m, f_n|\mathbf{i}) = \begin{cases} 0, & \text{if } f_m = f_n, \\ \frac{1}{2} \exp\left(-\frac{(i_m - i_n)^2}{2\xi^2}\right) \cdot \frac{1}{\Delta(m,n)} & \text{if } f_m \neq f_n, \end{cases} \tag{2.9}$$

where $\Delta(m,n)$ is the Euclidean distance between the voxels m and n. This regularizer is commonly used in graph cut segmentation methods [12]. It penalizes difference in labels for neighboring voxels with an intensity difference larger than ξ. In this way homogeneous regions are smoothed. We set ξ equal to the standard deviation of

the foreground image intensity distribution, estimated from the atlas images. The *interaction potential* is based on a 6-voxel 3D neighborhood model for all the tissues.

2.2.4 Graph cut optimization

A MAP solution $\hat{\mathbf{f}}$ can be found by converting Equation 2.1 to an equivalent energy function by taking the negative logarithm and subsequently minimizing the energy. As shown by Kolmogorov et al. [70] the energy function can be globally minimized using the graph cut algorithm [14]. In this work, we used the max-flow algorithm [13]. for optimization.

For every tissue ω, the posterior probability $p(\mathbf{f}, \mathbf{i})$ is completely described by the intensity model p_{int}, the spatial model p_s and a vector of parameters $\theta_\omega = (\alpha, \lambda_1, \lambda_2)$, i.e. the association potential weight λ_1, the spatial model weight λ_2 and the foreground threshold α. The optimal value of the parameters $\hat{\theta}_\omega$ depends on the spatial and intensity model for each tissue, which is not known; however the parameters can be estimated with cross-validation experiments using the manual segmentations of the atlases. This experiment is described in section 2.3.4.1. Once parameters have been chosen, the solution for the tissue ω can be computed as described above. The results for all the tissues are then combined to create a single multi-label image. Details on how the combination is performed are given in section 2.3.2.

2.2.5 Combination of the results

The proposed method is applied to each tissue independently, resulting in a set of binary segmentations. These results are then combined to create a single multi-label image in which each voxel is assigned a label which corresponds to one of the tissues. As each tissue is segmented independently two situations can occur: 1) voxels of neighboring tissues may have multiple labels, or 2) voxels at the boundaries of adjoining tissues can be erroneously assigned to background (there are 'holes' in the segmentation). We deal with these situations applying consecutively the following steps:

1) The label are combined with respect to a tissue priority list. Priority is based on the following ranking: thyroid, cartilages, cornea, lens, eye vitreous humor, sclera, optical nerve, spinal cord, brainstem, cerebellum, and cerebrum. This order has been chosen to give priority to the more internal tissues (relative to their neighboring tissues) with the exception of sclera which despite being a layer around the eye vitreous humor has a lower priority than the humor. This choice was made because the accuracy of sclera segmentation, before the tissues combination, was much lower than the accuracy of eye vitreous humor segmentation. Using this ranking list, after the tissue combination step we have a loss of accuracy in DSC (approx. 5%) only for the sclera.

2) Erroneously assigned voxels ('holes') are detected using binary morphological closing on the result of step 1. The correction is done using a majority vote at minimum distance filter: for each voxel detected as a hole the closer neighbors are checked; the label with the most occurrences within this set of neighbors is assigned to the hole location. When there is a draw more neighbors are included by increasing the

distance to the hole; the process is iterated increasing the distance of the neighbors until a majority vote label is found.

2.3 Experiments

2.3.1 Image data and atlas

Data were collected from patients scheduled for combined RT and HT. The data used in the evaluation study comprised 18 axial, contrast-enhanced CT images. The patients have tumors at different locations: larynx (9), nasopharinx (4), oropharinx (3) and in the base of the neck (2). The in-plane spatial resolution is 1.0 mm (16 sets) and 0.7 mm (2 sets). The slice thickness ranges from 1.5 mm (5 sets), 2.5 mm (11 sets), 3.0 mm (1 set) to 5.0 mm (1 set). The longitudinal field of view depends on the location of the tumor with a typical extension of about 40 cm. In 6 of the 18 cases larynx tumors substantially deformed the thyroid and cartilage regions.

A trained observer (O1) provided manual tissue delineations for all images. In the remainder of this chapter the atlas set labeled by O1 is named $\mathcal{T}_0$. Another two observers (O2 and O3) labeled the same tissues for a subset of 13 images. Thus, we have 13 images delineated by 3 observers; these sets of atlases are named $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$, respectively, where $\mathcal{T}_1$ is a subset of $\mathcal{T}_0$. The observers were medical radiation technology students trained by a radiation oncologist to delineate the structures of interest.

HTP requires a delineation of all tissues in the H&N area. Some of these tissues can be easily segmented automatically, based on HU thresholding; these are muscles, fat, bone, internal air and lungs. An additional 11 tissues were manually annotated because of their dielectric [42] and thermal [82] properties. A tissue is included in the model if its dielectric and thermal properties are different from those of the surrounding tissues. With these criteria the tissues of interest are:

- Encephalon: cerebrum (CE); cerebellum (CB); brainstem (BS)
- Spinal cord (SP)
- Eye tissues: vitreous humor (EH); sclera (SC); cornea (CO); lens (LE); optical nerve (ON)
- Neck tissues: cricoid and thyroid cartilage (CA); thyroid gland (TH)

In the $\mathcal{T}_0$ set, 4 of the 18 patients do not have a thyroid because of the tumor. One of them also has no cartilage as it was surgically removed. For another patient the CT field of view does not include the eyes. Therefore, the number of atlases available for the segmentation varies per tissue: for the enchepalon and spinal cord 18 atlases are available while for the eyes and cartilage 17, and for the thyroid 14 atlases can be used. In the subset of 13 images annotated by the experts, one patient without thyroid is included.

2.3.2 Implementation details

The method proposed by us was compared with a straight-forward atlas-based segmentation approach (AB) using majority voting atlas fusion [51]. Majority voting is a commonly used approach to combine the deformed label images from multiple atlases. For each voxel location, the label which occurs most often in the set of deformed labels from the training set is assigned to the resulting segmentation. In case of equal voting for the same voxel location a priority list is used to chose the final label. The priority is chosen using the same ranking used for the binary results combination (see section 2.2.5) but inverting the priority of eye vitreous humor and sclera. This order has been chosen to give priority to the internal tissues (relative to their neighboring tissues).

In all the experiments the *intensity model* is built by random sampling from the set of training images. When the dataset $\mathcal{T}_0$ is used, 2500 foreground samples are selected randomly from the entire training set for all the tissues, except for the cornea and optical nerve where only 1600 samples could be used due to small size of the tissue. The same number of samples are selected for background and foreground. The background region is defined as a band width up to 20 mm for all tissues, except for eye vitreous humor and optical nerve where the band width is 5 mm. In the third experiment the same number of samples per training image is used when the datasets $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ are used for training.

The average computational time for the graph cut optimization is 34 s per tissue. This results in an average time of 6 min/patient. The average atlas-based majority voting computational time is 45 s/patient. The complete registration procedure comprised rigid and non-rigid steps, takes on average 10 min per atlas. The average segmentation time of a complete dataset (as done in the experiment in section 2.3.4.2) is 3 h/patient where 170 min/patient are taken by the registration. The computational times are computed on a 2.3 GHz quad-core Intel processor, with 12 GB of RAM, running a 64 bit Windows 7 operative system.

2.3.3 Evaluation Measures

For the evaluation the manually-generated label images are used as ground truth. Three evaluation measures are used to compare segmented volumes: 1) the Dice Similarity Coefficient (DSC) [27], 2) the Mean Surface Distance (MSD) [29], and 3) the Hausdorff Surface Distance (HSD) [56].

1) The DSC is defined as the overlap of two segmentation volumes normalized to their total volume. It is defined as:

$$\mathrm{DSC} = 2 \times \frac{|X \cap Y|}{|X| + |Y|} \tag{2.10}$$

where X and Y are 3-D regions represented by binary images and the operator $|\bullet|$ give the number of element in the correspondent set.

2) The MSD measures the average distance between points on the surface of volumes. Given the X and Y volumes and their contour points $x \in X$, and $y \in X$

the MSD is defined as:

$$MSD = \max\{\sum_x \min_y d(x,y), \sum_y \min_x d(x,y)\} \tag{2.11}$$

where $d(x,y)$ denote the Euclidean distance between x and y.

3) The HSD measures the maximum distance between points on the surface of volumes. With the same notation used above, the HSD is defined as:

$$HSD = \max\{\max_x \min_y d(x,y), \max_y \min_x d(x,y)\}. \tag{2.12}$$

To compute the MSD and HSD distances we used the itk (www.itk.org) modules, respectively, itkContourMeanDistanceImageFilter and itkHausdorffDistanceImageFilter.

2.3.4 Experiments

Three experiments are performed. In the first experiment the parameters of the proposed method (atlas-based combined with intensity model, AB+IM) are tuned; in the second experiment the method is evaluated and compared with atlas-based majority voting (AB) to investigate the impact of including an intensity model in the segmentation procedure; in the third experiment the method's robustness is evaluated with respect to variations in the atlases used for training. Moreover the method's accuracy is compared with the interobserver agreement.

2.3.4.1 Parameter optimization

In the first experiment the method parameters θ_ω are tuned using the atlas image set $\mathcal{T}_0$. The optimal parameters $\hat{\theta}_\omega$ are selected by exhaustive search in a pre-defined range of values Θ. The superscript $(u;\mathcal{T})$ is used throughout this chapter whenever it is needed to explicitly specify target image and training set. For every atlas image $t_i \in \mathcal{T}_0 = \{t_1, \ldots, t_{18}\}$ an intensity model $p_{int}^{(t_i;\mathcal{T}_{0,i})}$ and a spatial model $p_s^{(t_i;\mathcal{T}_{0,i})}$ are generated for each tissue ω using the remaining subjects' scans $\mathcal{T}_{0,i} = \mathcal{T}_0 \setminus \{t_i\}$ as training images. Segmentations $\hat{\mathbf{f}}_\omega^{(t_i;\mathcal{T}_{0,i})}(\theta_\omega)$ are computed for all parameters $\theta_\omega \in \Theta_\omega$ and all target images $t_i \in \mathcal{T}_0$. Then the Dice similarity coefficients $\mathrm{DSC}(\hat{\mathbf{f}}_\omega^{(t_i;\mathcal{T}_{0,i})}(\theta_\omega), \mathbf{g}_\omega^{(t_i)})$ between automatic segmentations $\hat{\mathbf{f}}_\omega^{(t_i;\mathcal{T}_{0,i})}(\theta_\omega)$ and manual segmentations $\mathbf{g}_\omega^{(t_i)}$ are determined. The optimal parameters $\hat{\theta}_\omega^{(t_i)}$ are selected, for each tissue and for each test image, as the combination which gave the highest average DSC over all the other images $\mathcal{T}_{0,i}$. The average DSC for each test image t_i is given by:

$$\overline{\mathrm{DSC}}^{(t_i)}(\theta_\omega) = \frac{1}{N-1} \sum_{t_j \in \mathcal{T}_{0,i}} \mathrm{DSC}(\hat{\mathbf{f}}_\omega^{(t_j;\mathcal{T}_{0,j})}(\theta_\omega), \mathbf{g}_\omega^{(t_j)}), \tag{2.13}$$

where N is the number of images in $\mathcal{T}_0$. The optimal parameters are those which maximize this criterion:

$$\hat{\theta}_\omega^{(t_i)} = \operatorname*{argmax}_{\theta_\omega} \overline{\mathrm{DSC}}^{(t_i)}(\theta_\omega). \tag{2.14}$$

In this way the parameters learning of the test image t_i is never based on spatial and intensity models built for the image t_i. Despite the image t_i is used as training image, during parameter tuning, it has a minimum influence on the parameter selection. These parameters are used in all the following experiments.

2.3.4.2 Intensity model evaluation

In the second experiment the proposed (AB+IM) method is compared with the reference (AB) method. Our goal is to verify whether adding an intensity model to the segmentation improves the accuracy with respect to the reference method. Performances of both methods are evaluated by comparison with the manual ground truth available in the set $\mathcal{T}_0$. Using the parameters estimated in the first experiment a leave-one-out evaluation is performed for both AB and AB+IM. The segmentations $\hat{\mathbf{f}}_{\omega}^{(t_i,\mathcal{T}_{0,i})}(\hat{\theta_{\omega}})$ are computed for all tissues and combined (as described in section 2.3.2) resulting in $\hat{\mathbf{f}}^{(t_i,\mathcal{T}_{0,i})}$. Next $SIM_{\omega}(\hat{\mathbf{f}}^{(t_i,\mathcal{T}_{0,i})}), \mathbf{g}^{(t_i)})$, where SIM denotes the similarity measure (DSC, MSD or HSD), with $t_i \in \mathcal{T}_0 = \{t_1, \ldots, t_{18}\}$, are computed for each tissue ω. The AB results $\mathbf{h}^{(t_i,\mathcal{T}_{0,i})}$ are computed for each target t_i using the corresponding training set $\mathcal{T}_{0,i}$. Then the $SIM_{\omega}(\mathbf{h}^{(t_i,\mathcal{T}_{0,i})}, \mathbf{g}^{(t_i)})$ are determined for all the i images and for all the similarity measures.

2.3.4.3 Atlas dependence

The purpose of the third experiment is to evaluate the variation in segmentation results of the AB and the AB+IM methods owing to the choice of atlas set. For this experiment we used the sets of atlases, labeled by three different observers, described in 2.3.1, i.e. $\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$. We define the sets of manual labels as $\mathrm{OB_n} \equiv \{\mathbf{g}^{(t_i^n)} | i = 1, ..., 13, n = 1, 2, 3\}$, where i indexes the image and n indexes the observer and $\mathbf{g}^{(t_i^n)}$ is the manual delineation from observer n of the atlas t_i^n, and $t_i^n \in \mathcal{T}_n = \{t_1^n, \ldots, t_{13}^n\}$.

Results on agreement and accuracy are compared with the interobserver agreement of the three observers. For each tissue we calculated: interobserver agreement (IO-Ag), AB+IM agreement (AB+IM-Ag), AB agreement (AB-Ag), and the accuracy of the AB+IM (AB+IM-Ac) and AB (AB-Ac) methods evaluated and trained using delineation from three observers as ground truth and training sets.

For each pair of observers the agreement is given by the evaluation measure (DSC, MSD or HSD), for each tissue, between corresponding labels in the sets (labels related to the same i-th image). The overall agreement (IO-Ag) for each tissue is estimated as the average value of the evaluation measures obtained by comparing all the observers' delineations with each other.

The results of both methods (AB and AB+IM) are obtained using training images labeled by observer n in a leave-one-out framework, as done in 2.3.4.2. For both methods we calculated, per image, three different results, one for each observer's atlas set ($\mathcal{T}_1$, $\mathcal{T}_2$, and $\mathcal{T}_3$). We define the sets of AB results as $\mathrm{AB_n} \equiv \{\mathbf{h}^{(t_i^n,\mathcal{T}_{n,i})} | i = 1, ..., 13\}$ and the sets of AB+IM results as $\mathrm{AB{+}IM_n} = \{\hat{\mathbf{f}}^{(t_i^n,\mathcal{T}_{n,i})} | i = 1, ..., 13\}$, where $\mathbf{h}^{(t_i^n,\mathcal{T}_{n,i})}$, and $\hat{\mathbf{f}}^{(t_i^n,\mathcal{T}_{n,i})}$ indicates result, for AB and AB+IM respectively, obtained using training atlases labeled by observer n in a leave-one-out framework so that $\mathcal{T}_{n,i} = \mathcal{T}_n \setminus \{t_i^n\}$. Thus, for each image a result is calculated using the other images from the same

Table 2.1. Parameters optimization experiment. Typical parameters for one of the patient. Abbreviations: Par - parameter, CE - cerebrum, CB - cerebellum, BS - brain-stem, SP - spinal cord, ON - optical nerve , EH - eye vitreous humor, SC - sclera, LE - lens, CO - cornea, CA - cartilages, TH - thyroid

Par	CE	CB	BS	SP	ON	EH	SC	LE	CO	CA	TH
λ_1	0.50	0.50	2.00	1.00	3.00	0.25	1.50	1.50	2.00	3.00	1.00
λ_2	3.00	1.00	0.50	2.00	0.40	2.00	0.67	0.33	0.75	0.67	0.50
α	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.40	0.40	0.50

observer's set as atlases. At the end we have three results per image, one for each observer. The parameters as tuned in 2.3.4.1 are used to compute the AB+IM results.

The method (AB+IM or AB) agreement is the agreement of the results obtained training the method with images labeled by different observers. Similar to the IO-Ag we evaluated the method (AB-Ag and AB+IM-Ag) agreement for each tissue as the average measure over the three possible pairwise comparisons within the set of results (AB_n, $AB+IM_n$). So, for every i-th image, and for each method an overall agreement is obtained for both evaluation measures.

The method (AB+IM or AB) accuracy is the average accuracy of the results (AB+IM-Ac or AB-Ac) obtained training the method with images labeled by different observers. The method (AB_n, $AB+IM_n$) accuracy for each observer's atlas set ($\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$), used for training, is evaluated with DSC, MSD and HSD using the three available ground truth delineation sets $\mathbf{g}^{(t_i^n)}$. For every tissue, for each image and for each training set the result is compared with the ground truth delineations resulting in three sets of measures, one for each observer's atlas set used for training. Subsequently, the measures in these sets are averaged to obtain an overall accuracy for each image, and for each evaluation measure (DSC, MSD and HSD).

2.4 Results

2.4.1 Parameter Optimization

Table 2.4 lists a set of optimal parameters $\hat{\theta}_\omega$ obtained following the criterion in Equation 2.14 for one of the images in the dataset $\mathcal{T}_0$.

2.4.2 Intensity Model Evaluation

Figure 2.3 shows the distribution of the DSC, MSD and HSD over all the patient images in $\mathcal{T}_0$ for all tissues, and Table 2.2 lists the DSC, MSD and HSD median values for the AB+IM and the AB methods. The AB+IM method results in a significant improvement of the DSC score compared with AB for all tissues, except for brainstem and spinal cord. The MSD is significantly improved (lower values) for optical nerve, eye vitreous humor, lens and thyroid. The HSD is improved for the brainstem and

eye vitreous humor. The significance of improvement is tested as follows: first, the difference is tested using a two-sided, Wilcoxon signed rank test; next, when significant difference is found, a one-sided Wilcoxon signed rank test is performed to check whether this difference corresponds to an improvement. For both tests $p < 0.05$ is used as threshold for significance. The magnitude of improvement varies between tissues: for optical nerve, sclera, lens, cornea, cartilage and thyroid a DSC improvement (median value > 9%) is shown, whereas for cerebrum, cerebellum and eye vitreous humor a lower improvement was found. The MSD improvement is higher for optical nerve, thyroid and lens tissues than for the eye vitreous humor. The HSD improvement is similar for brainstem and eye vitreous humor. Figure 2.4 presents examples of results of AB+IM and AB (together with the ground truth) that illustrate how the intensity model acts in the segmentation process.

2.4.3 Atlas Dependence

The results of the third experiment are provided in Figure 2.5. The distributions over all images i of the observers agreement, the methods' agreement and accuracy are shown. Statistical tests were performed between method's agreement (AB-Ag and AB+IM-Ag) and the IO-Ag for each tissue. The agreement and accuracy of the proposed method are compared with the agreement and accuracy of the reference method using statistical tests. The method's accuracy (AB-Ac and AB+IM-Ac) was also compared with the IO-Ag. As in the previous experiment, the difference was tested using a two-sided Wilcoxon signed rank test, and improvement is verified using a one-sided test Wilcoxon signed rank test ($p < 0.05$).

2.5 Discussion

We developed and evaluated a method for the automatic tissue segmentation for HTP. It is shown that the combination of intensity model and atlas registration improves the segmentation performance both in accuracy and in robustness to variation in the atlas sets used.

An experiment was performed to optimize parameters. For each tissue of interest, and each test image, optimal parameters were computed. Generally a low variation in optimal parameters over the test images was found for all tissues, except for sclera and cartilage. From the three experiments can be concluded that the spatial model of these tissues is not accurate which is reflected in a higher variation of optimal parameters over test images.

In the second experiment we compared the proposed method with conventional multiatlas-based segmentation to evaluate the effectiveness of the intensity-based classification combined with multiatlas segmentation. We found an improvement in accuracy for DSC for all tissues, except for brainstem and spinal cord. The MSD showed significant improvement for optical nerve, eye vitreous humor, lens, and cartilage when when applying the intensity model. The HSD shows improvement for brainstem, and eye vitreous humor. We observed that the proposed method does not consistently influence the maximum surface error (HSD) which is committed when compared to

Table 2.2. Comparison between the atlas-based method combined with intensity model (AB+IM) and the atlas-based majority voting method (AB). Median DSC and MSD values over images in the set $\mathcal{T}_0$ are provided for all the tissues. Abbreviations: CE - cerebrum, CB - cerebellum, BS - brain-stem, SP - spinal cord, ON - optical nerve , EH - eye vitreous humor, SC - sclera, LE - lens, CO - cornea, CA - cartilages, TH - thyroid

Method	Measure	CE	CB	BS	SP	ON	EH	SC	LE	CO	CA	TH
AB+IM	DSC	0.98	0.92	0.78	0.85	0.62	0.81	0.46	0.67	0.43	0.54	0.68
AB	DSC	0.98	0.90	0.78	0.85	0.42	0.81	0.37	0.41	0.27	0.43	0.52
AB+IM	MSD[mm]	0.75	1.22	2.1	0.81	1.02	1.01	1.09	0.96	1.68	1.81	2.48
AB	MSD[mm]	0.81	1.34	1.89	0.76	1.45	1.01	1.38	1.50	1.67	2.01	3.87
AB+IM	HSD[mm]	18.39	11.44	8.67	6.62	6.63	4.98	6.85	3.73	6.44	13.71	17.52
AB	HSD[mm]	18.45	10.78	9.07	7.19	6.49	5.31	5.93	4.96	7.65	11.75	17.74

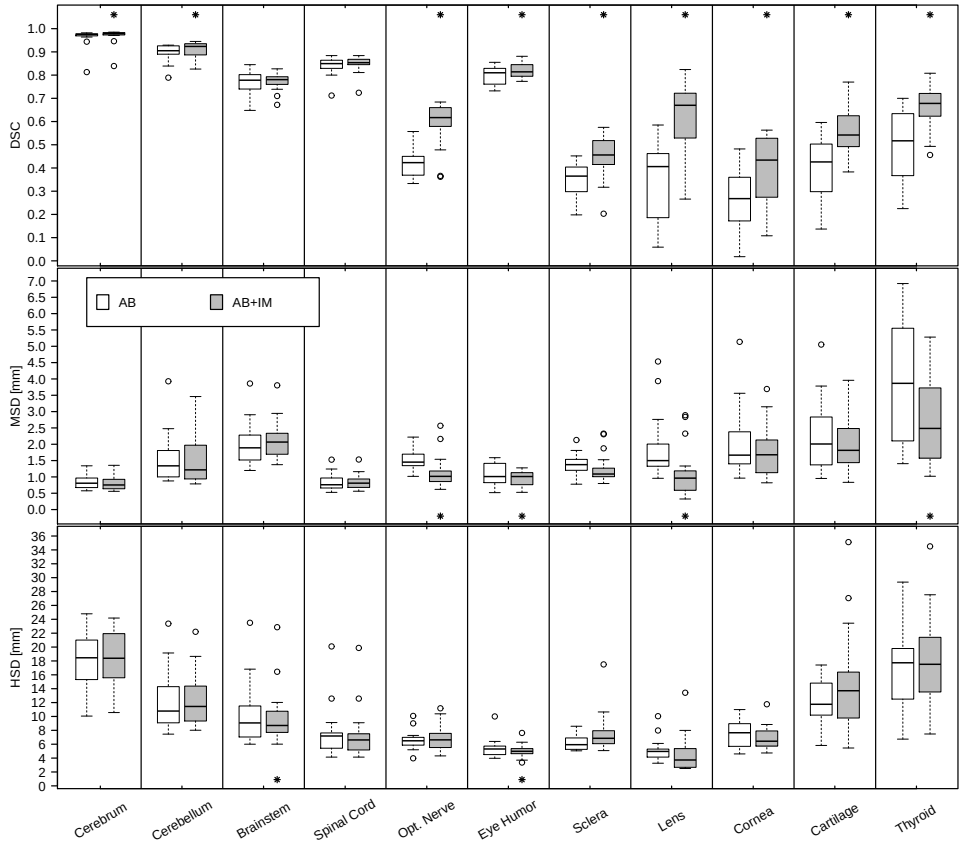


Figure 2.3. Comparison between the atlas-based method combined with intensity model (AB+IM) and the atlas-based majority voting method (AB). Results for DSC, MSD and HSD over the set $\mathcal{T}_0$ are shown for all tissues. The boxplots show confidence intervals, outliers ($\circ$), first and third quartiles, and median values. * indicates significant improvement.

the reference method. The proposed method is more effective in terms of DSC than in terms of MSD and HSD. This can be explained by the fact that the parameter optimization procedure is based on the DSC. The magnitude of improvement varies per tissue: when the atlas-based method is less accurate, the contribution of the intensity model is relatively higher. In these cases the intensity model is weighted more than the spatial model (see Table 2.1, with the exception of the brainstem which has a low spatial weight and no difference between the atlas-based and the proposed method. Therefore the intensity model compensates for an inaccurate spatial model and, thus, for an inaccurate registration.

In the third experiment we evaluated the accuracy and consistency of the proposed method when varying the atlas set used. The accuracy of the proposed method

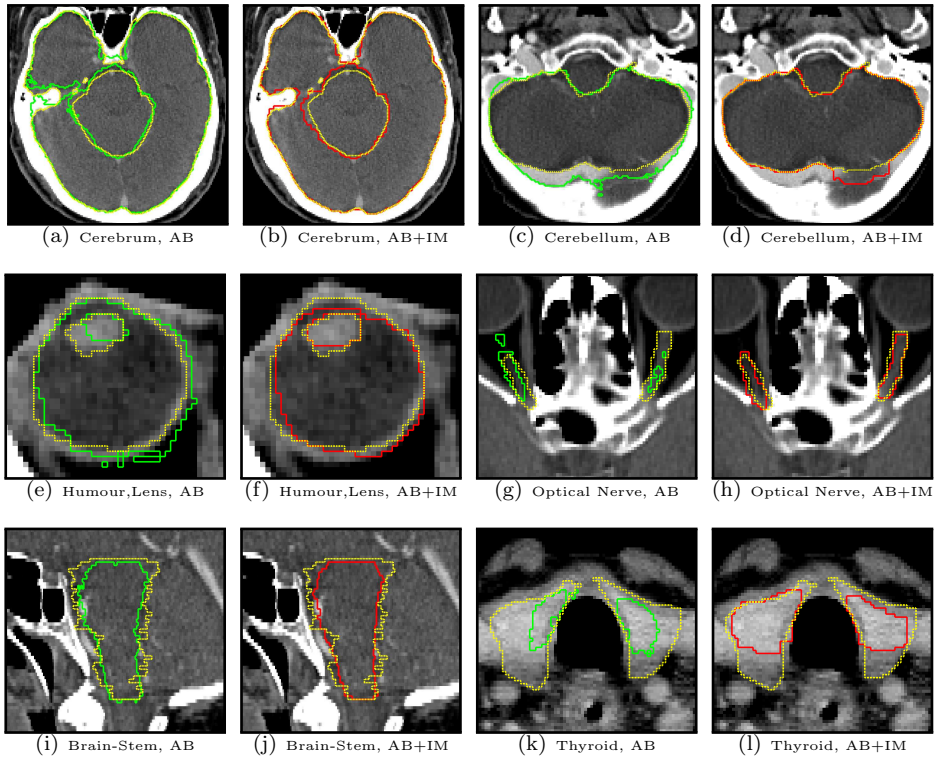


Figure 2.4. Examples of segmentations. Every image pair shows a comparisons between AB (green), AB+IM (red) and manual delineations (yellow) for different tissues. For the cerebrum a discontinuous segmentation (a) is smoothed by the contribution of the intensity model (b). Comparing (c) and (d) with the manual delineation shows that AB+IM has partially corrected the result, but that a region with intensity similar to the cerebellum is still wrongly labeled. For eye vitreous humor, lens (the inner boundary of the humor is the outer of the lens) and optical nerve the positive effect of AB+IM is highlighted. In (i) and (j) the behavior of AB+IM is similar to AB but provides a smoother segmentation. The last two images(k) and (l) show results for the thyroid and demonstrate a better delineation using AB+IM; however, the final boundary is still far from the ground truth.

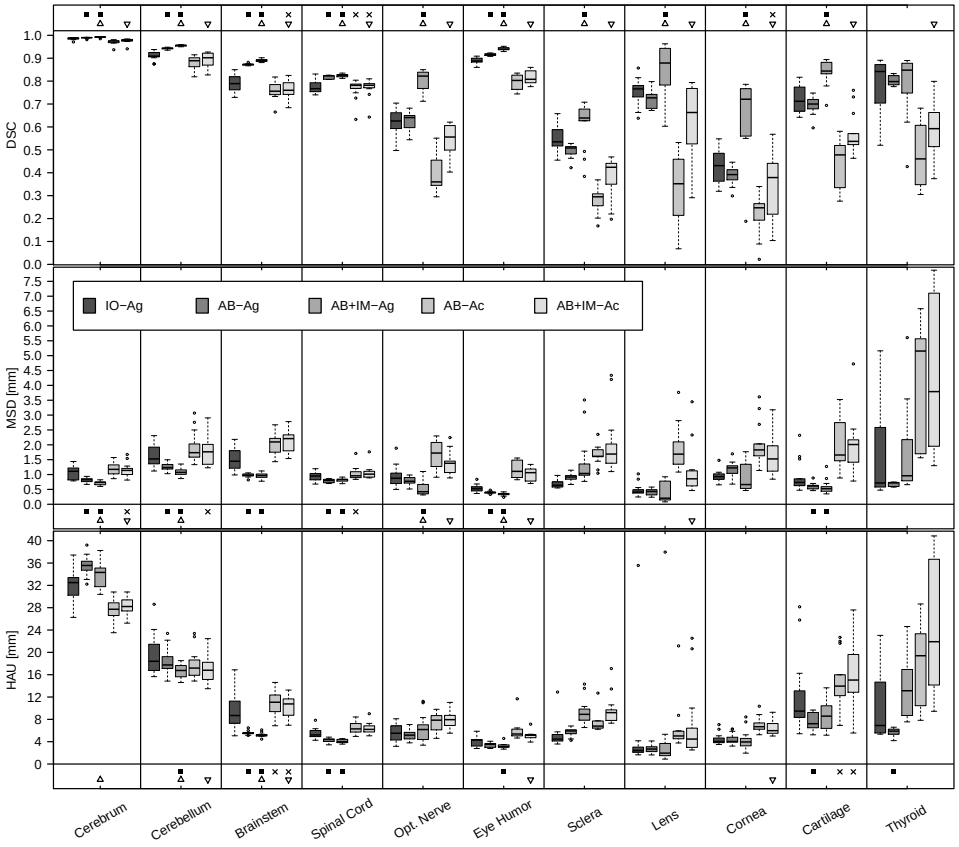


Figure 2.5. Atlas dependence results. Comparison between interobserver agreement, methods' (AB and AB+IM) agreement and accuracy when different atlases are used. Meaning of abbreviations: IO-Ag = interobserver agreement; AB+IM-Ag = atlas-based combined with intensity model agreement; AB-Ag = atlas-based agreement; AB+IM-Ac = atlas-based combined with intensity model accuracy; AB-Ac = atlas-based accuracy. The boxplots report confidence intervals, outliers (∘), first and third quartiles and median values for each distribution named in the legend. ■ indicates significant improvement with respect to IO-Ag. △ indicates significant improvement of AB+IM-Ag with respect to AB-Ag. ▽ indicates significant improvement with respect to AB-Ac. × indicates no significant difference compared with IO-Ag.

approaches the interobserver agreement for 8 out of 11 tissues: cerebrum, cerebellum, brainstem spinal cord, optical nerve, eye vitreous humor, lens and cornea. For the cerebellum, brainstem, spinal cord and cornea we reached the optimum in accuracy, given by the variability of the training set. Similar to the previous experiment, we found an improvement when combining intensity models and atlas-based spatial mod-

els with respect to the atlas-based method. Figure 2.3 shows that for the majority of the tissues the automatic methods are more consistent than the human observers. For the DSC, the proposed method shows a better consistency than the atlas-based segmentation method and the human observers, for all the tissues except for the thyroid. In terms of MSD, the proposed method shows better consistency with respect to the human observers and to the reference method for cerebrum, cerebellum, optical nerve and eyes vitreous humor. It can thus be concluded that the proposed method increases the consistency of results when the atlas label set changes.

The results for cerebrum, optical nerve, eye vitreous humor and lens segmentation show that the integration of intensity model improves the segmentation accuracy and consistency compared with the atlas-based method.

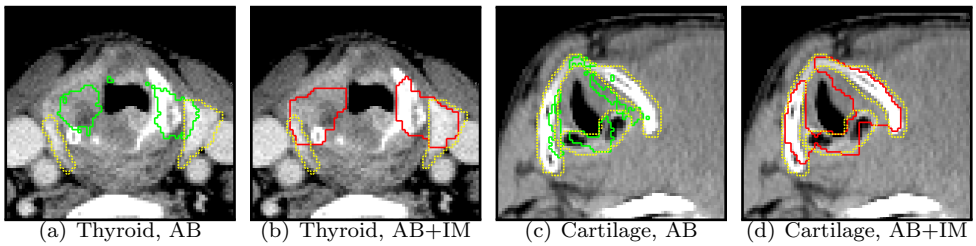
The results for brainstem and cerebellum are affected by artifacts in the manual delineations, with a large interobserver variability for both tissues (Figure 2.3). This variability is caused by the delineation protocol, which is based on axial plane annotation resulting in an irregular shape of the brainstem in sagittal and coronal planes. These artifacts have considerable impact on MSD.

The results for cornea and sclera were poor in terms of DSC, MSD, and HSD; there was a large interobserver variability for both tissues (Figure 2.5). Corneas are hardly identifiable on CT and can be manually delineated only by means of an educated guess; due to its size and shape (thin layer) a consistent delineation of the tissue by different observers is difficult to obtain. This variability limits the method because both spatial and intensity models are built from an inconsistent training set.

The spatial model for cartilage and thyroid, resulting from registration, is not accurate as is demonstrated by the atlas-based accuracy (see Figure 2.3). Due to poor registration accuracy for these two tissues, the final accuracy is far from the interobserver agreement. Figure 2.4 k-l illustrates that in this case the proposed method segmentation is more similar to the ground truth than the reference, but still far from it. The large anatomical differences between the patients, and the presence of tumor, hamper accurate registration in the cartilage and thyroid region. In Figure 2.6 examples of two cases with large tumors are shown. The first case (a-b) is an outlier considering the DSC of the proposed method in the *intensity model evaluation* experiment. The second case shows a result for cartilage with an accuracy close to the median value. Here the intensity model consistently helps to improve the accuracy of the result (AB has a DSC of 0.27 while AB+IM a DSC of 0.54).

The entire automatic segmentation method needs on average 3 hours using 17 atlases for a single patient image. The most demanding procedure with regard to computation is the registration, which requires 170 minutes with serial computation (using a single core processor of 2.3 GHz). However, the process can be accelerated using multi-core parallel computing. The graph cut optimization takes (on average) 7 minutes. Thus, with the current set-up the method is 5 h faster than the manual delineation process, and with the additional advantage that the time required is machine-processing time.

To our knowledge, the most accurate method for brainstem segmentation is the method of Quazi et al. [95] which reported a mean DSC value of 0.91, compared with our median value of 0.78. We think that this difference is due to the atlas set, since for the brainstem the DSC accuracy of our method is statistically no different from the



(a) Thyroid, AB (b) Thyroid, AB+IM (c) Cartilage, AB (d) Cartilage, AB+IM

Figure 2.6. Examples of challenging cases with large tumors. Every image pair shows a comparisons between AB (green), AB+IM (red) and manual delineations (yellow) for different tissues.

interobserver agreement ($p < 0.05$). Spinal cord segmentation results are reported by Han et al. [48] and by Teguh et al. [114] (where the method of Han et al. is validated). The first method reported a median DSC value and a distribution over patients < 0.8 [48], whereas we obtained a median value of 0.85 with a narrow distribution. The second study reported an (average) DSC and (average ± standard deviation) MSD of the combination brainstem/spinal cord of 0.78 and 2.3 ± 1.4 mm respectively [114]. These values are close to the accuracy we obtained for the brainstem (DSC 0.78 and MSD 2.10 mm) but are less than the accuracy for the spinal cord (DSC 0.85 and MSD 0.81 mm).

Eye segmentation was performed for eyeball, optical nerve and lens [10, 26]. Only Cuadra et al. [26] performed the evaluation using manual delineations as ground truth. Comparison of their results (for lens), of a median DSC over 3 patients of 0.769, with our results, of a median DSC over 18 patients of 0.67, shows that we perform approximately 10% worse.

A limitation of the proposed method is the fusion of multiple binary results to solve a multi-label segmentation. By combining the single tissue segmentations in a post-processing step, the spatial and intensity relationship between neighboring tissues are not exploited by the method. An extension of the method in this direction will probably increase the accuracy at tissue boundaries.

From the result of the first experiment it is clear that the atlas information for optical nerve, sclera, lens, cornea, cartilage, and thyroid is relatively poor. Intensity information improves the segmentation, but the atlases are still a limiting factor. The presence of tumors in the dataset increases the anatomical difference in the neck area. Since both methods show low performance in this area we conclude that this is caused by low-registration accuracy. Using a larger atlas set in combination with atlas selection [4, 23, 65] may improve registration accuracy and, thus, the quality of the shape prior. In this way the anatomical variation is better represented by the whole atlas set and, using atlas selection, the set can be adapted to the target anatomy.

We are currently working extending this method to magnetic resonance (MR) images. The superior soft tissue contrast of MR images may improve both the spatial prior model and the intensity model thereby improving the performance of the current set-up.

Use of the current method for HTP is a work in progress. The effect of delineation variability on HTP will be quantified. The specific absorption rate (SAR) distribution [89, 90] will be computed using the proposed method and evaluated in comparison with the actual clinical standard which is based on manual delineations. The results of these comparisons will determine the bounds for the accuracy and reproducibility of the automatic segmentation method, and whether the current approach is sufficient for clinical application.

2.6 Conclusion

We have developed and evaluated a fully automated method for segmentation of a large set of tissues for hyperthermia treatment planning. Using an intensity model combined with multiatlas registration improves the segmentation both in accuracy and consistency compared with atlas registration alone. Our combined atlas registration-graph cut approach enables us to combine the strengths of both approaches: the atlas-based registration gives a suitable spatial prior for a more accurate gradient- and intensity-based graph cut segmentation approach. In this way we prevent inaccurate atlas-based segmentations caused by a deformation field that is constrained by the degrees of freedom of the transformation model, and by an atlas-fusion process that generally does not use intensity information.

The proposed method provides accurate automatic segmentation for the relevant HTP tissues with computational performances compatible with clinical demand. In addition our method increases the segmentation reproducibility with respect to the current standard.

Chapter Three

CT-Based Patient Modeling for Head and Neck Hyperthermia Treatment Planning: Manual versus Automatic Normal Tissue Segmentation

Watch every detail that affects the accuracy of your work.

— Arthur C. Nielsen (1897-1981)

Abstract — To decide whether a recently developed automatic-segmentation algorithm can be introduced in the clinic, we compared the impact of manual- and automatic normal-tissue-segmentation variations on HTP quality.
CT images of seven patients were segmented automatically and manually by four observers, to study interobserver and intraobserver geometrical variation. To determine the impact of this variation on HTP quality, HTP was performed using the automatic and manual segmentation of each observer, for each patient. This impact was compared to other sources of patient model uncertainties, i.e. varying gridsizes and dielectric tissue properties.
Despite geometrical variations, manual and automatic generated 3D patient models resulted in an equal, i.e. 1%, variation in HTP quality. This variation was minor with respect to the total of other sources of patient model uncertainties, i.e. 11.7%.
In conclusion, automatically generated 3D patient models can be introduced in the clinic for H&N HTP.

Based upon: R F Verhaart, **V Fortunati**, G M Verduijn, T van Walsum, J F Veenland and M M Paulides, "CT-Based Patient Modeling for Head and Neck Hyperthermia Treatment Planning: Manual versus Automatic Normal-Tissue-Segmentation", *Radiother Oncol*, vol. 111, no. 1, pp. 158-63, Apr. 2014.

3.1 Introduction

Hyperthermia, i.e. raising tissue temperature to 40–44°C, has been shown to improve clinical outcome when added to radiotherapy or chemotherapy for several tumor sites [38, 59, 124], including the head and neck [5, 54, 55]. We recently developed a hyperthermia applicator to investigate the benefit of deep local heating of head and neck tumors [89, 90]. The clinical use of this device requires hyperthermia treatment planning (HTP) based on electromagnetic simulations for pre-treatment and real-time treatment optimization and tissue dose assessment. Crucial input for HTP are full 3D patient models incorporating all normal tissues and the gross tumor volume (GTV). These models are generated by segmenting tissue regions on computed tomography (CT) images [91]. In Chapter 2 we described an automatic-segmentation algorithm for head and neck HTP that has shown to be accurate, reproducible and substantially reduces operator time. A clinical introduction of the algorithm requires a comparison of the impact of automatic segmentation on the hyperthermia treatment quality with the actual clinical standard, which is based on manual segmentations. Manual segmentations are prone to observer variation, and the patient model influences the accuracy of HTP for deep hyperthermia in the pelvic region [20, 134]. Due to the large number of small tissue regions in the head and neck region, observer variation in tissues segmentation may have a substantial impact on the hyperthermia treatment quality, but, this impact has never been quantified.

CT-based observer variation in tissue segmentation has already been assessed for head and neck radiotherapy treatment planning [16,87]. However, while HTP requires a full 3D patient model, these studies included only a limited number of tissues, and many reported either interobserver or intraobserver variation. But, although not complete, these studies provide an excellent reference to compare the results for separate organs.

Assessments of the exposure by electromagnetic sources of the human body also involve 3D human models. These studies often summarize causes of simulation uncertainties in an uncertainty budget, which includes uncertainties such as variations in dielectric tissue properties and variations in the gridsize [7, 85]. Observer dependent tissue segmentation might also influence the simulated electromagnetic field, however this confounding influence usually is not included in the uncertainty budget.

In this Chapter we report the manual and automatic CT-based segmentation variation for the tissues included in the 3D patient models for head and neck HTP. In addition, we compared the impact of the manual and automatic segmentation variation on the HTP outcomes, i.e. the planned hyperthermia dose and HTP quality. To quantify the importance of segmentation variations, we compare their influence to those of other sources of patient-modeling variation, i.e. gridsize and dielectric tissue-property uncertainties. The decision, whether or not to introduce the automatic-segmentation algorithm into the clinic, can be based on these results.

3.2 Methods and materials

3.2.1 Patient Selection

The analysis presented in this chapter covers 7 of the 34 patients treated with head and neck hyperthermia thus far. The patients were selected to represent the patient population eligible for head and neck hyperthermia, i.e. the patients included were balanced per tumor site, and both small (T2) and large (T4) tumors were included, as show in Table 3.1.

Table 3.1. Patient, tumor and treatment characteristics. HT: hyperthermia, F: female, M: male, r: recurrent, c: clinical, p: pathology, CBK: cyberknife, IMRT: intensity modulated radiation therapy, BT: brachytherapy

Patient	Gender	Age	Tumor Site	TNM classification	HT	Radiotherapy
1	F	38	Nasal Cavity	rT4bN0M0	3x	5x5.5 Gy (CBK)
2	F	58	Thyroid	rT4N1bM0	4x	16x3.13 Gy (IMRT)
3	F	74	Nasopharynx	rT4N0M0	3x	6x6 Gy (CBK)
4	M	51	Oropharynx	cT2N0M0	6x	4, 10x3, 4 Gy (BT)
5	F	59	Oropharynx	cT2N0M0	9x	35x2 Gy (IMRT)
6	M	60	Thyroid	pT4aN1bM0	4x	16x3.13 Gy (IMRT)
7	M	54	Oropharynx	rT2N0M0	1x	6x5.5 Gy (CBK)

3.2.2 Computed Tomography (CT) images

CT scans acquired for radiotherapy treatment planning were used for HTP, leading to advantages in logistics and target region assignment. To make the group of patients for our study as representative as possible, we included CT scans of patients with distinct characteristics to span the entire patient population variability.

CT scans of the patients were obtained using a Somatom Sensation Open (Siemens AG, Erlangen, Germany), except for patient 3, who was scanned with a PQ 5000 (Philips Healthcare, Best, the Netherlands). The slice spacing varied from 1.5 – 2.5 mm and the in-plane resolution varied from 0.7 x 0.7 mm to 1.0 x 1.0 mm, with a scan matrix of 512 x 512. For patients 1–5, an intravenous injection of 100 ml contrast agent (Omnipaque 300, GE Healthcare Inc.) was administered with an injection rate of 1.8 ml/s, and imaging was performed 45 seconds after injection.

3.2.3 Segmentation protocol

All CT slices are segmented into several normal tissues and the target volume. We used the clinical target volume (CTV) as the hyperthermia target volume (HTV), and segmented the GTV in order to assign tumor dielectric tissue properties to this region. The list of segmented tissues was based on the visibility on CT and the dielectric (and thermal) property contrast with adjacent tissues. The brain, spinal cord and eyes were segmented since these are highly thermo-sensitive tissues. Therefore, the thermal dose should be restricted in these tissues.

First, an in-house developed tool (implemented in MevisLab v.2.2.1, MeVis Medical Solutions AG, Bremen, Germany) was used to remove non-patient structures, such as the patient bed and the patients immobilization mask. Second, Hounsfield (HU) thresholds were applied to segment bone (HU: 200 to 3000), muscle (HU: 0 to 200), fat (HU: -300 to 0), lungs and internal air (HU: -1000 to -300) [52]. The lungs were separated from the internal air by applying the threshold only to the slices that contain lung tissue. Third, the automatically segmented bone was manually corrected in case of streak artifacts, and when blood vessels were incorrectly assigned as bone due to the presence of contrast agent. Fourth, the tissues in the brain (cerebrum, cerebellum, brainstem), the spinal cord (myelum), the eyes (sclera, lens, vitreous humor, optical nerve) and the other head and neck tissues (thyroid gland, thyroid and cricoid cartilage) were segmented manually using iSeg (v.3.1, Zurich Med Tech AG, Zurich, Switzerland) and automatically using the method of Chapter 2. Fifth, the HTV and GTV were both manually segmented in Focal (v.4.64, Elekta AB, Stockholm, Sweden) by a head and neck radiation oncologist.

3.2.4 Hyperthermia treatment planning (HTP)

Hyperthermia treatment planning was performed as described by Rijnen et al. [97]. For electromagnetic field simulations, a uniform gridsize of 2 mm was chosen. Dielectric tissue properties, i.e. relative permittivity (ϵ_r), effective conductivity (σ_{eff}) and volume density of mass (ρ) were assigned to each tissue [43, 58]. The commonly used 1g-averaged and 10g-averaged specific absorption rate (SAR) standards [IEEE/IEC62704-1] as calculated in SEMCAD-X (v.14.8.1, Schmid & Partner Engineering AG, Zurich, Switzerland) were used for SAR dosimetry.

For the tissue property sensitivity study, we reduced computational time twelvefold by performing only one simulation, i.e. a simulation with all antennas exited using optimized phase and amplitude settings instead of a simulation per antenna and a subsequent weighted summation of the fields. The validity of this approximation for the sensitivity analysis was verified for three variations (ϵ_r: +6%, σ_{eff}: +6%, ρ: +3%), in which we observed an average error in HTP quality ($|\Delta HTQ|$) of only 0.2% (min – max: 0.02 – 0.6%).

3.2.5 Segmentation evaluation

Three trained medical radiation technologists (observer 1, observer 2, observer 3) and one radiation oncologist (reference) manually segmented per patient the ten tissues, see Figure 3.1 for segmentation examples. The reference segmented the seven patient models only once, while the 3 other observers segmented them twice to investigate intraobserver variation. The manual tissue segmentation of one patient took on average 5-6 hours, all segmentations were done within 4 weeks and the time between first and second segmentation of the same patient varied between 1-15 days. The images were anonymized and supplied in a random order to minimize bias in the manual segmentations. The automatic-segmentation algorithm took on average 1 hour per patient (3.3 GHz Intel Core i7-980 processor, with 24 GB of RAM, running 64 bit Windows 7). Since the segmentation of the radiation oncologist is assumed as most

accurate, variations from that segmentation are reported as segmentation inaccuracies. The observer-reference variations (interobserver variation: reference-observer 1, reference-observer 2, reference-observer 3), the variation between two segmentations per observer (intraobserver variation: observer 1-observer 1, observer 2-observer 2, observer 3-observer 3) and the automatic-reference variations were determined. Observer variation was quantified using the Dice similarity coefficient [27] and the mean surface distance (MSD), see also Section 2.3.3. Since DSC measures the overlap between volumes, the variation is quantified using 1-DSC. We used DSC instead of 1-DSC to compare our results to other studies.

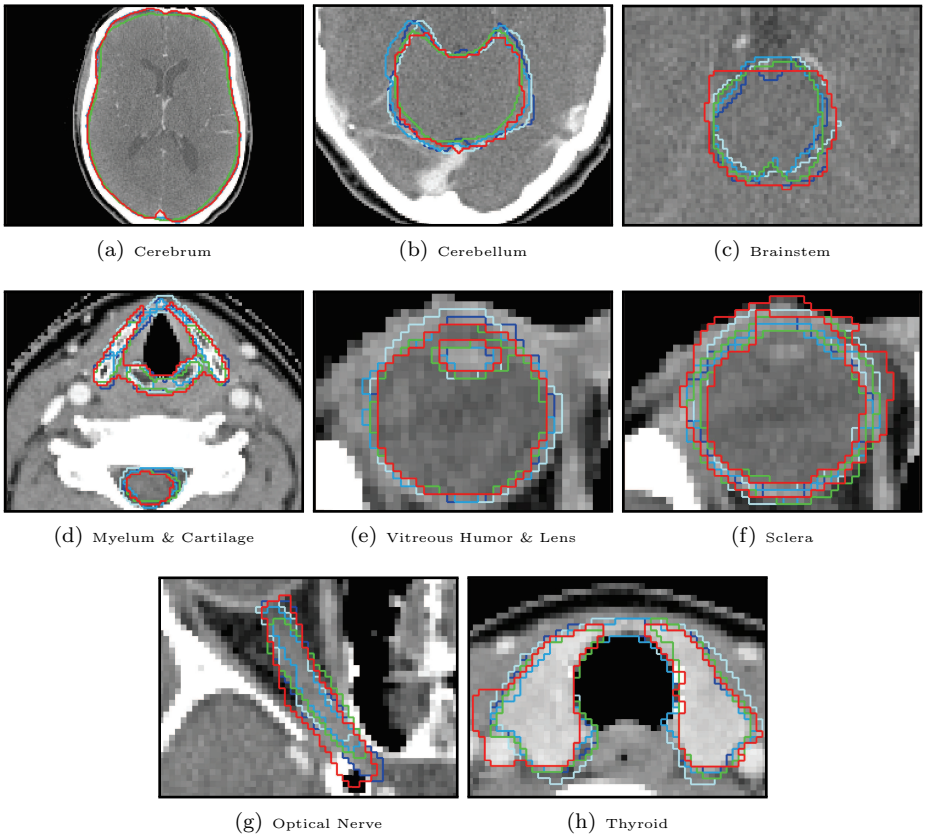


Figure 3.1. The segmentations of observer 1 (dark-blue), observer 2 (light-blue), observer 3 (blue), reference (green) and the auto-segmentation (red). The tissues were segmented on each transverse slice of the image set. For the head and neck HTP, these segmentations are stacked on top of each other to generate the 3D patient model. The 3D SAR dose distribution is calculated on the basis of this model.

3.2.6 Dosimetric evaluation

As temperature dose predictions still come with high uncertainties, we quantified the effect of segmentation variation on the planned hyperthermia dose using the absolute SAR difference expressed in two SAR parameters, i.e. $|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$, and on HTP quality using $|\Delta HTQ|$.

$|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$ are computed per tissue and defined as the SAR difference between two treatment plans, normalized for 1W total input power. Both values were averaged over the tissue volume and observers, and were the mean over the seven patients. Furthermore, the relative difference with respect to the average SAR in the reference is reported. Absolute SAR values were obtained per power level by multiplication with the total input power level [90].

HTQ is computed by using the SAR_{1g} only, and is defined as the average SAR in the first percentile of the SAR in the healthy tissue, i.e. the average over the highest SAR values, divided by the average SAR in the target volume [19]. The HTQ error ($|\Delta HTQ|$) is defined as the absolute difference in HTQ between two treatment plans.

A non-parametric Wilcoxon signed rank test was used to test for statistical significant differences ($p < 0.05$).

3.2.7 Sensitivity analysis

The variation in HTP prediction per observer-dependent segmentation was compared with other sources of 3D patient modeling variation, i.e. gridsize variation and dielectric tissue property uncertainties. The uncertainty is quantified in $|\Delta HTQ|$ to assess the impact on HTP quality and in $|\Delta SAR_{10g}|$ to compare our results to other studies.

For all seven patients, the effect of the gridsize was investigated by increasing gridsize (Δgrid), i.e. from 1 to 5 mm. For each gridsize, $|\Delta HTQ|$ and $|\Delta SAR_{10g}|$ were determined by using the HTP with Δgrid = 1 mm as reference.

Dielectric tissue properties in patients deviate from literature values, e.g. due to age-dependency [92, 93, 103], post-mortem changes [18, 102], and measurement uncertainty [40, 41, 81].The impact of this deviation on HTP quality was determined in seven patients, for fourteen tissues. For each tissue property, three different HTPs were generated, i.e. using average and ± one standard deviation of the literature value. $|\Delta HTQ|$ was determined by using the HTP for the average literature value as reference. The individual uncertainty per tissues was obtained by taking the square-root of the sum of squares of the mean value over seven patients.

The individual standard uncertainty (u_i) of the gridsize, segmentation variation and dielectric properties was obtained by determining the confidence interval from plus to minus one standard deviation. The square-root of the sum of squares was applied to the individual uncertainties to obtain the combined standard uncertainty (u_c). A coverage factor (k) of 2 was used to obtain the expanded uncertainty U leading to the 95% confidence interval [113]. The uncertainties were determined based on the assumption that the applied variations can be described by a Gaussian probability function.

3.3 Results

Table 3.2 shows a complete overview of the results, including the intraobserver variation and the parameters (1-DSC and $|\Delta SAR_{10g}|$). This table shows that the cerebellum, brainstem, spinal cord and vitreous humor could be segmented with a mean 1-DSC of ≤ 0.24 and a mean MSD of ≤ 2.2 mm, leading to a mean $|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$ up to 4.3 mW/kg (4.3%). The cerebrum, sclera, lens, optical nerve, cartilage and thyroid gland were segmented less accurate with a mean 1-DSC of ≥ 0.24 and/or a mean MSD ≥ 2.2 mm, corresponding to a higher mean $|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$ up to 15.1 mW/kg (11.7%). For most tissues, the intraobserver variation caused significantly less segmentation and tissue specific dose variation compared to the interobserver variation. Local tissue SAR differences varied among both SAR parameters used, e.g. the interobserver variation of the lens resulted in a $|\Delta SAR_{1g}|$ of 15 mW/kg (12%) but in a lower $|\Delta SAR_{10g}|$ of 3.5 mW/kg (3.8%).

Considering the manual (interobserver) and automatic segmentation variation per tissue (MSD) and the impact of this variation on simulated hyperthermia dose ($|\Delta SAR_{1g}|$), we found that for cerebellum, spinal cord, optical nerve, cartilage and thyroid gland a non-significant difference between manual and automatic generated 3D models in both segmentation variation and tissue specific dose, while a significant difference was found for cerebrum, brainstem, sclera, lens and vitreous humor (the latter two tissues with respect to segmentation only).

Figure 3.2 shows that simulation accuracy substantially decreased when HTP was based on a gridsize of 3 mm or larger. Because the simulation accuracy was stable for a gridsize in the range of 1.3 – 2.5 mm, we chose a gridsize of 2 mm as trade-off between HTQ error and simulation time.

Figure 3.3 shows the impact of tissue property uncertainties on HTP quality ($|\Delta HTQ|$), for one patient. HTP quality appeared to be most sensitive to inaccuracies in the dielectric tissue properties of muscle, fat, and bone regions.

Table 3.3 summarizes the influences of the patient modeling parameters on $|\Delta HTQ|$, for the seven patients. The $|\Delta HTQ|$ due to interobserver variation (1.0%), intraobserver variation (1.0%) and automatic-segmentation (1.0%) was similar, but substantially smaller that the uncorrelated standard combined uncertainty (11.7%) and the expanded uncertainty (23.4%) for the uncertainties studied.

3.4 Discussion

In this study, we report the CT-based normal-tissue-segmentation variations and the impact of these variations on HTP, i.e. the planned hyperthermia dose and HTP quality, for both manual- and automatic-segmented tissues. For all segmented tissues, the local tissue SAR differences varied among SAR parameters, i.e. $|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$. This makes tissue dose assessment with these parameters difficult, which forms the rationale to investigate the impact on simulated temperatures, as soon as thermal simulations are validated. Although half of the segmented tissues were significantly different after manual or auto-segmentation, the impact on HTP quality was similar, but, this impact was minor compared to the total HTP uncertainty.

Table 3.2. Manual (interobserver and intraobserver) and auto-segmentation variation for the ten selected tissues. The segmentation variation is quantified by the 1- Dice similarity coefficient (1-DSC) and the mean distance (MSD). The difference in simulated hyperthermia dose ($|\Delta SAR_{1g}|$ and $|\Delta SAR_{10g}|$, relative SAR differences normalized at 1 W total input power) in each tissue is shown in mW/Kg. The statistical significant difference between interobserver and intraobserver variation($^{\diamond}$) and between auto-segmentation and interobserver($^{\triangledown}$) or intraobserver(*) variation is determined using a non-parametric Wilcoxon signed rank test. Note that the interobserver and auto-segmentation variation is defined with respect to the reference observer.

	Manual (interobserver)	**Manual (intraobserver)**	**Auto-Segmentation**
	mean, min–max	mean, min–max	mean, min–max
	mean, min–max	mean, min–max	mean, min–max
	mean & mean	mean & mean	mean & mean
Cerebrum			
1-DSC	0.059, 0.052–0.069	$0.010^{\diamond}$, 0.010–0.013	0.058^{*}, 0.050–0.068
MSD [mm]	2.4, 2.0–2.9	$0.43^{\diamond}$, 0.36–0.54	$2.9^{*,\triangledown}$, 2.5–3.2
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.32 (1.2%) & 0.23 (0.90%)	$0.24^{\diamond}$ (0.87%) & 0.18 (0.70%)	$0.44^{*,\triangledown}$ (1.6%) & $0.35^{*,\triangledown}$ (1.4%)
Cerebellum			
1-DSC	0.14, 0.12–0.16	$0.065^{\diamond}$, 0.046–0.086	0.13^{*}, 0.088–0.20
MSD [mm]	2.2, 1.8–2.5	$0.88^{\diamond}$, 0.62–1.2	1.9^{*}, 1.3–3.4
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.92 (3.1%) & 0.60 (2.2%)	$0.63^{\diamond}$ (2.1%) & $0.40^{\diamond}$ (1.5%)	0.94^{*} (3.2%) & 0.57^{*} (2.1%)
Brainstem			
1-DSC	0.22, 0.15–0.27	$0.16^{\diamond}$, 0.15–0.18	0.24^{*}, 0.19–0.32
MSD [mm]	1.7, 1.1–2.4	$1.1^{\diamond}$, 0.92–1.2	$2.2^{*,\triangledown}$, 1.7–3.1
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.87 (6.3%) & 0.59 (3.7%)	0.68 (4.9%) & 0.45 (2.8%)	$1.27^{*,\triangledown}$ (9.3%) & 0.69^{*} (4.3%)
Spinal Cord (Myelum)			
1-DSC	0.21, 0.16–0.27	$0.15^{\diamond}$, 0.13–0.16	0.24^{*}, 0.18–0.34
MSD [mm]	1.0, 0.81–1.3	$0.55^{\diamond}$, 0.51–0.60	1.5^{*}, 0.82–2.4
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.84 (5.7%) & 0.42 (3.6%)	0.43 (2.9%) & 0.18 (1.6%)	0.86^{*} (5.8%) & 0.35 (3.0%)
Sclera			
1-DSC	0.52, 0.44–0.70	$0.37^{\diamond}$, 0.33–0.38	$0.62^{*,\triangledown}$, 0.45–0.76
MSD [mm]	0.87, 0.63–1.8	$0.51^{\diamond}$, 0.45–0.60	$1.8^{*,\triangledown}$, 1.0–3.8
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	3.4 (4.2%) & 1.6 (2.4%)	$1.9^{\diamond}$ (2.4%) & 0.84 (1.3%)	$4.8^{*,\triangledown}$ (6.0%) & 2.1^{*} (3.1%)
Lens			
1-DSC	0.32, 0.24–0.45	$0.20^{\diamond}$, 0.15–0.25	$0.50^{*,\triangledown}$, 0.34–0.64
MSD [mm]	0.66, 0.44–1.2	$0.32^{\diamond}$, 0.23–0.44	$1.6^{*,\triangledown}$, 0.75–2.9
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	15 (12%) & 3.5 (3.8%)	$3.4^{\diamond}$ (2.6%) & 1.3 (1.3%)	15^{*} (12%) & 4.1 (4.4%)
Vitreous Humor			
1-DSC	0.14, 0.11–0.18	$0.086^{\diamond}$, 0.076–0.096	$0.20^{*,\triangledown}$, 0.16–0.28
MSD [mm]	0.70, 0.54–1.1	$0.38^{\diamond}$, 0.35–0.43	$1.2^{*,\triangledown}$, 0.87–1.8
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	3.3 (3.3%) & 2.0 (2.6%)	$1.9^{\diamond}$ (1.9%) & $1.0^{\diamond}$ (1.3%)	4.3^{*} (4.3%) & 2.5 (3.2%)
Optical Nerve			
1-DSC	0.40, 0.34–0.50	$0.30^{\diamond}$, 0.24–0.35	0.42^{*}, 0.37–0.51
MSD [mm]	0.94, 0.62–1.7	$0.57^{\diamond}$, 0.36–0.71	1.0^{*}, 0.76–1.4
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.92 (2.9%) & 0.39 (0.90%)	$0.59^{\diamond}$ (1.9%) & 0.33 (0.80%)	1.2^{*} (3.8%) & 0.63 (1.4%)
Cartilage			
1-DSC	0.41, 0.28–0.54	$0.22^{\diamond}$, 0.12–0.29	$0.51^{*,\triangledown}$, 0.45–0.61
MSD [mm]	4.5, 3.1–6.1	$0.67^{\diamond}$, 0.35–1.2	5.0^{*}, 3.0–6.9
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	5.8 (6.4%) & 3.0 (2.9%)	$1.4^{\diamond}$ (1.6%) & $0.92^{\diamond}$ (0.90%)	6.0^{*} (6.6%) & 2.8^{*} (2.7%)
Thyroid Gland			
1-DSC	0.24, 0.11–0.47	0.16, 0.085–0.29	0.43, 0.20–0.63
MSD [mm]	1.9, 0.54–4.7	0.80, 0.39–1.8	5.1, 1.1–9.3
$\lvert\Delta SAR_{1g}\rvert$&$\lvert\Delta SAR_{10g}\rvert$	0.39 (1.7%) & 0.33 (1.2%)	0.37 (1.6%) & 0.32 (1.2%)	0.68 (3.0%) & 0.47 (1.7%)

Furthermore, a considerable reduction in operator time per patient was achieved when using the automatic segmentation (± 1 hour) instead of the manual segmentation (5-6 hours). Thus, the recently developed automatic-segmentation algorithm can be introduced in the clinic for pre-treatment and real-time treatment optimization.

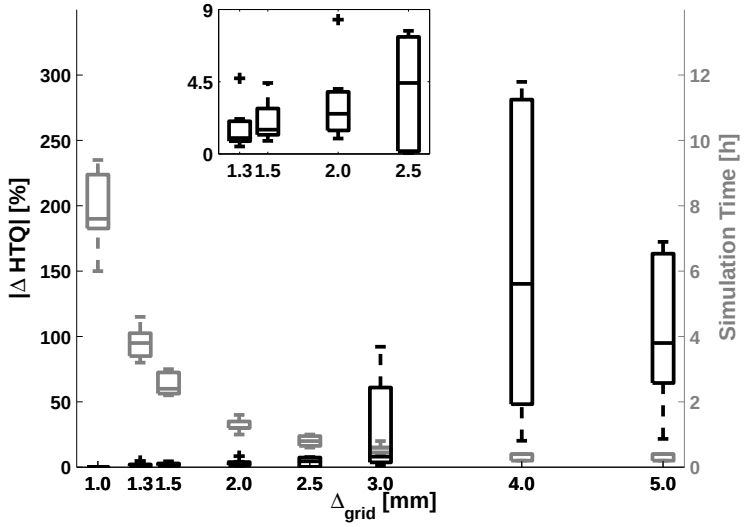


Figure 3.2. The error in hyperthermia-treatment quality parameter (|ΔHTQ|), and simulation time as a function of gridsize (Δgrid). The boxplot indicates the range of the error over seven patients. The insert shows a zoom-in of Δgrid 1.3-2.5 mm. Note that (|ΔHTQ|) is determined using the patient model with Δgrid 1mm as the reference. On each box of the boxplot, the central mark is the median, the edges are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers (99.3%), and outliers are plotted individually(+).

Previous studies that quantified the variations in tissue segmentation for head and neck radiotherapy treatment planning provide an excellent reference for our results. Brouwer et al. [16] reported in an interobserver study with five observers that the thyroid gland can be segmented as accurate as the spinal cord, while we found a difference in accuracy between these two tissues. This difference can be explained by a difference in the segmentation protocol, i.e. we segmented the thyroid and cricoid cartilage as one structure, which may have led to larger segmentation inaccuracies. Recently, Nelms et al. [87] quantified the interobserver agreement for the brainstem, brain and spinal cord. For the brainstem (31 observers) they found a lower DSC (mean ± std) of 0.66 ± 0.17, compared to a DSC of 0.78 ± 0.04 found in our study (Table 3.1). For the spinal cord (29 observers) they found 0.80 ± 0.07 compared to 0.79 ± 0.04 in our study and for the brain (10 observers) they found 0.98 ± 0.01 compared to 0.94 ± 0.01 for the cerebrum in our study. In general, despite some minor differences, the variations in tissue segmentation reported in our study are similar to the results of previous studies.

In the design of this study, we faced the common difficulty of defining HTP quality [91]. Hyperthermia outcome has been correlated to the temperatures, or temperature dose, achieved [37, 105]. However, temperature simulations are unsuitable since they

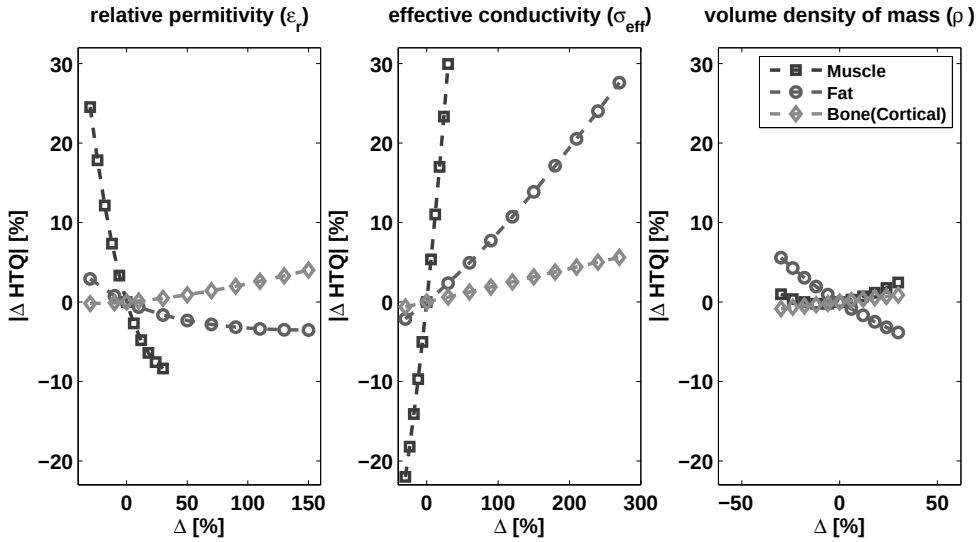


Figure 3.3. The error in hyperthermia-treatment quality parameter (|ΔHTQ|) as a function of the relative tissue property variation Δ, for patient 5. Only the tissues with an individual standard uncertainty > 1.0% are shown.

Table 3.3. Influence of patient modeling parameters on the uncorrelated standard uncertainty (u_c) and the expanded uncertainty (U) as measured with |ΔHTQ| and $|\Delta SAR_{10g}|$. k: coverage factor, , u_i: individual standard uncertainty.

	(u_i) [%]	
	|ΔHTQ|	$\|\Delta SAR_{10g}\|$
Gridsize (Δgrid)	4.0	3.2
Tissue segmentation		
(a) Manual segmentation		
(a_1) Interobserver	1.0	1.5
(a_2) Intraobserver	1.0	1.2
(b) Auto-segmentation	1.0	2.0
Dielectric tissue properties		
(a) Relative permittivity (ϵ_r)	4.1	3.4
(b) Effective conductivity (σ_{eff})	10.0	6.2
(c) Volume density of mass (ρ)	1.0	4.4
u_c (k=1)	11.7	9.3
U (k=2)	23.4	18.7

have not been validated yet for head and neck hyperthermia. In pre-treatment and real-time HTP for head and neck hyperthermia, we therefore optimize the target SAR. For simulated SAR, Paulides et al. verified from treatment data of three patients, on a per-patient basis, a linear relation with measured temperatures (R^2 = 0.59 – 0.94) [89]. For HTQ, Canters et al. [19] showed for deep hyperthermia in the pelvis a linear correlation with simulated steady state temperatures. Hence, in the absence of validated temperature predictions, both parameters: 1) HTQ to quantify HTP quality, and 2) the standardly used averaged-tissue SAR to quantify tissue dose, have potential as surrogate for simulated temperatures.

The dose differences reported in Table 3.2 were all determined for 1 W total input power. An average input power of 400 W is applied during a standard head and neck treatment [89]. Hence, the reported dose values should be multiplied by 400 if the results need to be translated to a standard head and neck treatment. For example, the maximum $|\Delta SAR_{1g}|$ was found in the lens: 0.0151 W/kg, which leads to a $|\Delta SAR_{1g}|$ of 6.0 W/kg at 400W. Note that in this case the maximum SAR_{1g} in the lens is 254.7 W/kg. Future studies are needed to clarify if the resultant temperatures go beyond the defined thermal thresholds.

The reported individual standard uncertainty in $|\Delta SAR_{10g}|$ of the gridsize was estimated as 3.2%, which is lower than the 6.2% (0.26 dB) found by Murbach et al. [85] and the 5.5% found by Bakker et al. [7]. However, the frequency and gridsizes studied (64 MHz and 1.5 – 3 mm [85], 10 – 5600 MHz and 0.7 – 10 mm [7]) differed from our settings, limiting the comparability. Our estimation of the individual standard uncertainty in $|\Delta SAR_{10g}|$ for dielectric tissue parameters (ϵ_r: 3.4%, σ_{eff}: 6.2%) was in agreement with previously reported values (ϵ_r: 2.1% [85], 8.7% [7], σ_{eff}: 3.5% [85], 8.7% [7]). Our individual standard uncertainty in $|\Delta SAR_{10g}|$ for ρ (4.4%) is lower than previously reported values (6.2% [85], 5% [7]). The differences in uncertainty for the dielectric tissue properties and ρ can also be explained by the difference in frequency, but probably also because we employed tissue specific uncertainties while others applied one common uncertainty for all tissues. We expected the uncertainty in segmentation variation and gridsize to be in the same range, since both are changing the 3D patient model. From other studies [7, 85], we know that especially dielectric tissue properties have a high influence, but this had never been shown for head and neck HTP.

3.5 Conclusion

As the uncertainty in HTP quality for manual and automatic segmentation is similar and minor when comparing it to the total HTP uncertainty, we conclude that the recently developed automatic-segmentation algorithm can be introduced in the clinic for pre-treatment and real-time treatment optimization. However, tissue dose assessment, used for safety assessment and pre-treatment and real-time decision making, remains difficult since local tissue SAR estimates vary among different SAR parameters. This forms a rationale for future studies to investigate the impact on simulated temperatures, as soon as thermal simulations are validated.

Chapter Four

Automatic Tissue Segmentation of Head and Neck MR Images for Hyperthermia Treatment Planning

Strive for continuous improvement, instead of perfection.

— Kim Collins (1976)

Abstract — We present here an automatic atlas-based segmentation algorithm for MR images of the head and neck.
Our method combines multiatlas local weighting fusion with intensity modeling. The accuracy of the method was evaluated using a leave-one-out cross validation experiment over a set of 11 patients for which manual delineation were available.
The accuracy of the proposed method was high both in terms of Dice similarity coefficient (DSC) and mean surface distance (MSD) with median DSC higher than 0.8 for all tissues except sclera and median MSD lower than 1 mm for all tissues except the CSF compared to manual delineations. For all tissues, except the spine tissues, the accuracy was approaching the interobserver agreement/variability both in terms of DSC and MSD. The positive effect of adding the intensity modeling to the multiatlas fusion decreased when a more accurate atlas fusion method was used. Using the proposed approach we improved the performance of the approach previously presented for H&N hyperthermia treatment planning making the method making the method suitable for clinical application.

Based upon: **V Fortunati**, R F Verhaart, J F Veenland, W J Niessen M M Paulides and T van Walsum, "Automatic Tissue Segmentation of Head and Neck MR Images for Hyperthermia Treatment Planning.", Major revision in *Phys Med Biol*.

4.1 Introduction

Local hyperthermia treatment (HT) consists of elevating the temperature of tumors between 39° C and 44° C. The effectiveness of radiotherapy (RT) and chemotherapy is significantly enhanced when combined with HT [37, 57, 124]. Phase III clinical studies show that hyperthermia is also effective in the head and neck (H&N) region [54, 55, 121]. In response to these positive results, a site-specific hyperthermia applicator was developed for precise HT of tumors in the entire H&N region using electromagnetic fields. This device, called the HYPERcollar, enables deep heating (> 4 cm under the skin) of H&N cancers [8, 89, 90]. Due to the limited availability of temperature monitoring in-vivo, HT of H&N tumors requires accurate treatment planning (HTP), based on 3D patient models that are derived from segmented 3D images [127]. These segmentations are generally derived by manual delineation of computed tomography (CT) images. However, it is well-known that manual delineation is tedious, labour intense and prone to inter and intraobserver variability. Thus automation of the segmentation process is highly desired. It has been shown that, due to their superior soft tissue contrast, MR images provide better delineation accuracy of target and normal tissues than CT images [2, 45, 96, 117, 126]. Hence, in this chapter we investigate the accuracy of MR based automatic segmentation for some of the tissues needed for HTP. MR-based segmentation need to be registered to the CT coordinate frame in order to obtain a geometrically accurate full 3D patient model [35, 126].

Several methods have been developed in recent years to address the problem of automatic segmentation of H&N CT images for generating anatomical models for radiotherapy [39,48,95] treatment planning. The common aspect of all these methods is that they are based on multiatlas segmentation approaches. In the field of HT planning, we recently published a method to accurately segment all relevant tissue in CT images [34] using atlases and intensity modeling. We also showed that the use of this method leads to a HTP quality comparable to the quality obtained using manually generated delineations [127].

The problem of automatic segmentation using atlases was extensively addressed in the field of brain structure segmentation in MR images [3,6,101,122,130]. Whereas majority voting, being a simple and straightforward way, has been frequently used to combine atlas registration results, more recently multiple authors have shown that local weighting atlas fusion techniques [6, 101, 130] increase the segmentation accuracy. In these approaches, segmentations from multiple atlases are combined using local image similarity metrics to locally weigh the contribution of each segmentation. Among these methods, the approach developed by Wang [130] ranked first in a recent MICCAI grand challenge [75]. Due to the effectiveness of this method we decided to combine local weighting multiatlas fusion with the approach we previously used for CT images [34].

In this chapter, we developed and evaluated an automatic segmentation algorithm for MR images of the head and neck; we used a graph cut method which combines multiatlas-based segmentation, using atlas fusion, and intensity modeling. The method combines the approach described earlier [34] with local weighting atlas fusion [6, 130]. The contribution of our work is twofold: 1) we developed an auto-

matic segmentation method for head and neck MR images that integrates intensity modeling into state-of-the art multiatlas fusion methods; 2) we evaluated various versions of our method on imaging data of 11 patients, and compared the results to the interobserver variability of manual delineations.

4.2 Methods and materials

4.2.1 Image data and atlas

The data used in this chapter was collected from 11 patients with tumors in the H&N region, which were representative for the patient-group eligible for HT. The location of the tumors was the nasopharynx (3 patients) and the oropharynx (8 patients). The data comprised axial, MR T2 weighted (T2w) images acquired on a 1.5 T scanner (Optima MR450w, GE Healthcare, Waukesha, WI) with the following acquisition parameters: Fast Recovery Fast Spin Echo sequence with TE/TR 106.8/7060 ms; in plane resolution of 0.49 mm (5 patients) and 0.68 mm (6 patients); slice thickness of 3 mm; out of plane field of view of 20 cm centered at the tumor location, yielding a slightly different field of view for each patient. For 8 patients, images were acquired in treatment position using a 6 channel surface coil while for the other 3 patients a 24 cnhannel H&N coil was used. The T2w images were corrected for intensity inhomogeneity: for the images acquired with surface coils, a built- in algorithm (Surface Coil Intensity Correction) of the MR scanner was used at the time of acquisition; for the images acquired with H&N coils, the ITK-N4 method [118] was used as a post-processing procedure. In order to compensate for the differences in intensities, all the images were normalized [119]. This normalization involved matching all the intensity histograms to the histogram of one randomly selected image. The code for this normalization is in the ITK library (www.itk.org, itkHistogramMatchingImageFilter.h). We used 256 bins for the histogram computation matching 128 histogram landmarks.

Three trained observers (O1, O2, O3) delineated the tissues in all images. We included the tissues for which the dielectric and thermal properties are different from those of the surrounding tissues. Additionally, tissue delineation had to be feasible on T2w images. With these criteria, the tissues of interest were:

- Encephalon: cerebrum (CE); cerebellum (CB); brainstem (BS).
- Spinal cord: myelum (MY) and cerebro spinal fluid (CSF).
- Eye tissues: vitreous humor (EH); sclera (SC); lens (LE).

The optical nerve was excluded from this study even though it is relevant for HTP. This was done because the optical nerve could not be consistently delineated on T2w images.

4.2.2 Segmentation method

Our segmentation method combines locally weighted multiatlas fusion with voxel-wise classification based on intensity values. Atlas based segmentation consists of the

registration of a target image, with an unknown segmentation, to an atlas for which the segmentation is known. The registration is used to estimate a segmentation by spatially mapping the known segmentation to the target image. In multiatlas segmentation this process is repeated for more than one atlas and an atlas fusion method is used to combine the information of all atlases into a single optimal segmentation [98].

In our method, an atlas fusion method was used to determine a spatial probability map for each tissue. The atlases, in the original coordinate frame, were further used to estimate the intensity distribution of the tissues and their corresponding background regions. This information was used in a graph-cut segmentation algorithm: the probability map was used as a spatial prior and the intensity distributions as an intensity model. A regularization term was also used to ensure a smooth segmentation. Figure 4.1 shows a scheme of the segmentation method.

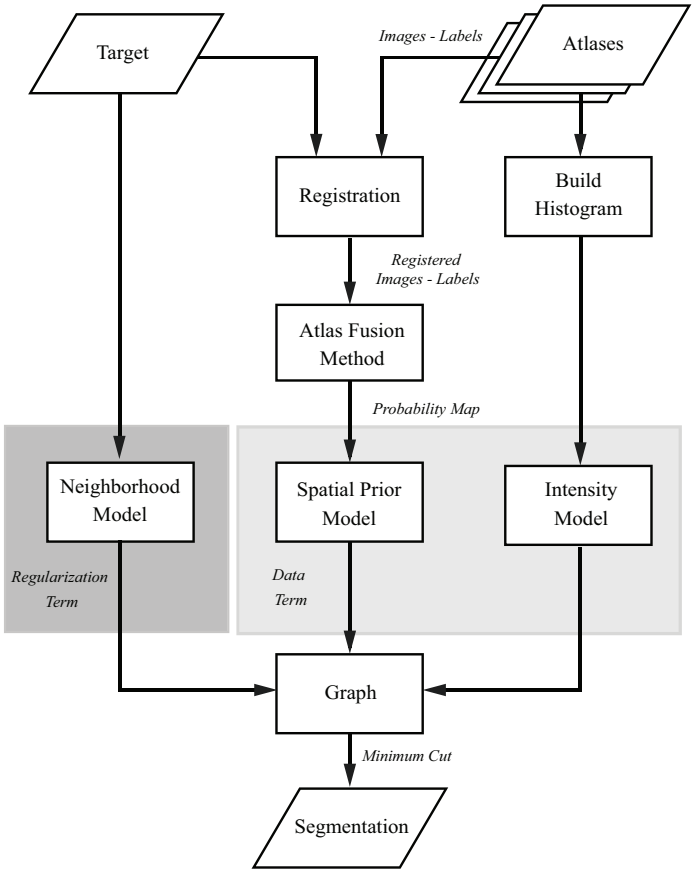


Figure 4.1. Scheme of the segmentation method for each tissue.

For details on how the intensity model was combined with the spatial probability map we refer the reader to our previous work [34]. In the following we focus on

the local weighting atlas fusion technique and its integration into our segmentation framework.

4.2.3 Spatial probability map construction

To construct a spatial prior, first a set of J atlas images was non-rigidly registered to the target image u. Using the registration results for each atlas j and each label l we obtained the segmentation $S_j^l(m)$ and the registered images $I_j(m)$. Fusion was performed with a local weighted voting technique [6,130]. In this method the label at voxel location m was selected as the one with the highest vote $\hat{S}^l(m)$, where $\hat{S}^l(m)$ is computed according to the rule:

$$\hat{S}^l(m) = \sum_{j=0}^{J} \omega_j(m) S_j^l(m), \tag{4.1}$$

where $\omega_j(m)$ is a weighting factor for the location m, which reflects the accuracy of each single atlas segmentation, in the proximity of m. Thus the final vote was calculated by weighting the contribution of each single atlas segmentation with respect to their local segmentation quality.

The weights $\omega_j(m)$ for each location m and atlas j can be estimated using the local image similarity between a patch of the target image $\mathcal{I}_u(\mathcal{N}(m))$ and a patch of atlas j registered to the coordinate frame of the target image u and interpolated at voxel location m, $\mathcal{I}_j(\mathcal{N}(m))$. The patch $\mathcal{N}(m)$ was defined as a 3D cubical region centered at m with width (in voxels) of $2r_p + 1$. We evaluated different similarity metrics (SIM) for local fusion: sum of squared differences (SSD), normalized cross-correlation (NCC) and normalized mutual information (NMI). The weights $\omega_j(m)$ were defined following the work of Artaechevarria et al. [6]:

$$\omega_j(m) = SIM(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m)))^\beta, \tag{4.2}$$

where β is a factor which was positive for NCC and NMI, and negative for SSD in order to make ω a measure of similarity.

Two different advanced weighting techniques have been proposed to improve this generic local weighting framework: 1) using a local search technique to decrease the effects of small registration error; 2) computing the final weight of each atlas considering the joint image similarity between the target image and all atlases. Both approaches were integrated into our method; we discuss them in the next two subsections.

4.2.3.1 Local search technique

The anatomical variability between the atlas and target image makes the registration challenging and prone to errors. This error can be accounted for during atlas fusion [101, 130] by locally searching for the atlas image patch that has the best similarity with the patch of the target image u. Next, the segmentation corresponding to the most similar atlas image patch can be used in the modified voting rule:

$$\hat{S}^l(m) = \sum_{j=0}^{J} \omega_j(\eta_j(m)) S_j^l(\eta_j(m)), \tag{4.3}$$

where $\eta_j(m)$ is the location in the atlas j corresponding to the patch $\mathcal{N}$ that gives the best match between atlas and target image in the local search region $\mathcal{N}_s$ according to the local search:

$$\eta_j(m) = \underset{m' \in \mathcal{N}_s}{\operatorname{argmin}} \|\mathcal{I}_j(\mathcal{N}(m')) - \mathcal{I}_u(\mathcal{N}(m))\|^2, \tag{4.4}$$

$\mathcal{I}_j(\mathcal{N}(m'))$ is the vector of intensities of the j-th atlas within the patch $\mathcal{N}$ centered around m' and $\mathcal{I}_u(\mathcal{N}(m))$ is the vector of intensities of the target image u within the patch $\mathcal{N}$ centered around m. The patch $\mathcal{N}_s(m)$ is defined as a 3D cubical region centered at m with width (in voxels) of $2r_s + 1$.

4.2.3.2 Joint weighted voting

Wang et al. [130] additionally introduced a weighting technique to estimate of the final segmentation taking a possibly correlated error between atlas segmentations into account. In this method the optimal weights are computed by minimizing the total expectation of segmentation error. This expectation is expressed as a matrix containing the joint probabilities of two atlases making a segmentation error at a certain location. Similarly to the general local weighting method this probability is approximated using an intensity similarity between a pair of atlases and the target image in the neighborhood of each voxel. Let $\boldsymbol{\omega}_m = (\omega_0(m), ..., \omega_J(m))$ denote the vector of weights. Following Wang et al. [130] the optimal weights are calculated as:

$$\boldsymbol{\omega}_m^* = \frac{(M_m + \gamma I)^{-1}\mathbf{1}_n}{\mathbf{1}_n^T(M_m + \gamma I)^{-1}\mathbf{1}_n}, \tag{4.5}$$

where γ is a regularization term used to ensure a unique solution for the minimization problem, $\mathbf{1}_n = [1; 1; ....; 1]$ and I is the identity matrix. The weights in Equation 4.5 are normalized, i.e. $\sum_j \omega_m(i) = 1$. The matrix M contains the expected pairwise joint label differences at each location m, which can be estimated using local image similarity, i.e. SSD, NCC and NMI. NCC and the NMI need to be inverted in order to represent an error measure. In case of SSD we defined the matrix M_m according to Wang et al. [130]. For the NCC based joint fusion the matrix was defined as:

$$M_m^{NCC}(i,j) = \left(\sqrt{1 - NCC(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_i(\mathcal{N}(m))}\sqrt{1 - NCC(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m))}\right)^{\beta}, \tag{4.6}$$

where i and j indicate the atlases, β is a positive scaling factor and the value one is used to invert the NCC since $NCC \in [-1, 1]$. Similarly for the NMI based joint fusion we have:

$$M_m^{NMI}(i,j) = \left(\sqrt{3 - NMI(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_i(\mathcal{N}(m))}\sqrt{3 - NMI(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m)))}\right)^{\beta}, \tag{4.7}$$

with the same notation as 4.6. In this case we use the value three to invert the metric since $NMI \in [1, 2]$. Following [130] for all the three metrics, the value of γ was set to approximatively 2% of the maximum value of the similarity metric i.e. 0.1 for SSD and 0.05 for NMI and NCC.

4.2.4 Graph cut segmentation

The multiatlas fusion methods provides a probability map $p_s^l(f_m)$ for each tissue:

$$p_s^l(f_m) = \left(\frac{1}{\sum_j \omega_j(\eta_j(m))} \sum_j S_j^l(\eta_j(m)) \omega_j(\eta_j(m)) \right)^{\lambda_2}, \tag{4.8}$$

where λ_2 is a weighting parameter. This probability map is used as spatial prior in the data terms of the graph cut segmentation algorithm [14]. The other data term used is the intensity model of the tissue which consisted of the intensity distributions of the tissue and its corresponding background region. The intensity distributions is built using the unregistered atlases by a Parzen-window histogram estimation [34]. The intensity model aims to correct for residual registration errors or systematic error in the output segmentation. Two user-defined parameters were used: the weight parameter λ_1 to weigh the contribution of the data terms with respect to the regularization term and the parameter λ_2 to weigh the contribution of the spatial prior term with respect to the intensity model term.

For each tissue l, the graph-cut segmentation depends on a vector of parameters $\theta_l = (\alpha, \lambda_1, \lambda_2)$, i.e. the data term weight λ_1, the spatial model weight λ_2 and the foreground threshold α [34]. This last parameter is introduced to correct for under-segmentation. For each tissue, the optimal value of the parameters $\hat{\theta}_l$ depends on the spatial and intensity model; these parameters are estimated in a cross-validation experiment on the available atlases.

A max-flow algorithm is used to obtain the optimal cut [13] of the graph. The graph cut segmentation is computed separately for each tissue and next all results are combined to create a single multi-label image as described in [34].

4.2.5 Implementation details

Atlas and target images were registered using the opensource registration software elastix [66]. A two-step registration was performed in a multiresolution fashion using mutual information as similarity metric [115] and adaptive stochastic gradient descent optimization [65] with 4000 random samples at each iteration. The first registration step consisted of an affine registration which was done using four resolution levels with 500 iterations per level. The second step consisted of a free form deformation [100], based on third order B-Spline polynomials. In this step, three resolution levels were used and the number of iterations varied from the coarsest to the finest resolution level (500, 1000 and 2000). These parameters were obtained after training on an independent set of 11 T2w images acquired with the same protocol as the segmentation dataset. Twenty two anatomical landmarks [35] were manually placed in the images to assess the registration accuracy and therefore enable the selection of the best parameters.

The intensity model was built by random sampling from the set of training images. Five thousands foreground samples were selected randomly from the entire training set for all tissues, except for the sclera (1500) and for the lenses (700) due to the small volume of the these tissues. In all cases, the same number of samples was selected for

background and foreground where the background region was defined as a band with a width of 5 mm around the foreground.

4.3 Experiments

Three experiments were performed to optimize the method parameters and evaluate its performance. The purpose of the *first experiment* was to evaluate which strategy lead to the most accurate spatial priors. Six strategies were compared: local weighting voting (LW) using 1) SSD, 2) NCC and 3) NMI as image similarity metric; joint weighting voting (JW) using 4) SSD, 5) NCC and 6) NMI as image similarity metric. For each strategy the local similarity patch radius r_p, local search radius r_s, and scaling exponent β were optimized in a leave-one-out cross validation experiment. A pre-defined set of parameters was used for all methods: $r_p \in \{1, 2, 3, 4\}$ voxels; $r_s \in \{0, 1, 2, 3, 4\}$ voxels; $\beta \in \{0.5, 1, 2, ..., 10\}$. The highest average generalized Dice Similarity Coefficient [27] (see also Section 2.3.3) over all images calculated using all tissues except cerebrum and cerebellum (DSC_{Gen}) was used to select the best parameter set. These tissues were excluded since their volumes are much larger than the volume of the other tissues, and thus would dominate the DSC_{Gen}. The DSC measures the normalized overlap between two volumes and the DSC_{Gen} extends the measure to two sets of volumes. For each segmentation the DSC_{Gen} was defined as:

$$\mathrm{DSC}_{Gen} = \frac{2 \sum_{l \in L^*} \hat{S}^l \cap G^l}{\sum_{l \in L^*} \hat{S}^l \cup G^l}, \tag{4.9}$$

where G^l is the ground truth segmentation of tissue l, $\hat{S}^l$ is the automatic segmentation and L^* is the set of all tissue labels excluding cerebrum and cerebellum. Using the optimal parameters, we evaluated the accuracy of the segmentation of each tissue by DSC. For both tuning and evaluation the manual delineations from O1 were used. The best method (using the best metric) was then selected as the one giving the best overall accuracy in DSC. This method was then used as spatial prior in the graph cut segmentation.

Given the optimal method for computing a spatial prior, the purpose of the *second experiment* was to optimize the parameters λ_1, λ_2, and α for the graph cut segmentation. As in [34], this optimization was done in a leave-one-out cross validation experiment consisting on an exhaustive search over a pre-defined set of parameters: $\lambda_1, \lambda_2 \in \{0.125, 0.25, 0.5, 1, 2, 4, 8\}$ and $\alpha \in \{0.4, 0.5\}$. The delineations of observer O1 were used in this experiment. Additionally, the optimal parameters using the MV to compute the spatial prior were calculated.

The purpose of the *third experiment* was to evaluate the overall performance of the combined methods. Specifically we verified the effectiveness of adding the intensity model for the problem at hand, the influence of the spatial prior on the method accuracy, and the accuracy of each method compared to the human observer ability to consistently segment a certain tissue. For this, the optimal version of the methods resulting from the previous two experiments was employed. The following methods were thus compared:

- multiatlas-based majority voting [51].
- multiatlas-based majority voting combined with intensity model (MV+IM) [34].
- best local/joint weighting atlas voting (LW/JW).
- best local/joint weighting atlas voting combined with intensity model (LWIM/JWIM).
- interobserver agreement/variability (IOA/IOV).

For the evaluation the manually-generated label images from the three observers were used as ground truth. The graph-cut method was applied three times using the atlases from each of the observers to built the intensity models and the spatial priors, in a leave-one-out cross-validation scenario. Then the three results were averaged for each patient, to estimate the overall method accuracy. To have a complete quantitative overview of the segmentation accuracy, two evaluation measures were used to compare the segmented volumes: the DSC and the Mean Surface Distance (MSD) [29] (for the definition of these measures see Section 2.3.3).

In the first and third experiment we performed pair-wise comparison between the distribution of accuracies per patients. To assess whether a distribution was higher or lower than another, significance was tested using a one-sided Wilcoxon signed rank test. Significant difference was tested using a two-sided Wilcoxon signed rank test ($p < 0.01$).

4.4 Results

In the first experiment the parameters of LW and JW were optimized for each of the three similarity metrics. Table 4.1 reports the leave-one-out optimal parameters for LW and JW using all metrics.

Table 4.1. Optimal leave-one-out parameters for LW and JW;

Method	Metric	Patient	r_p	r_s	β
LW	SSD	5/6	4	2	$-2/-3$
LW	NCC, NMI	All	4	2	10
JW	SSD/NCC	All	4	2	2/1
JW	NMI	9/2	4	2	1/0.5

For LW, SSD provided a statistically significantly ($p < 0.01$) better accuracy than both NCC and NMI for all tissue except the brainstem, myelum, left lens and humour. NMI provided a statistically significantly ($p < 0.01$) better accuracy than NCC for myleum, right sclera and lens, left sclera and generalized DSC. For all the other tissues SSD never decreased the accuracy.

For JW, SSD provided a statistically significantly ($p < 0.01$) better accuracy than NCC for the brainstem, CSF, and generalized DSC. SSD also provided more accurate

results than NMI: a statistically significantly ($p < 0.01$) better accuracy was found for cerebrum, CSF, right sclera and humour, left humour and generalized DSC. NCC provided a statistically significantly ($p < 0.01$) better accuracy than NMI for cerebrum, cerebellum, CSF, right sclera and humour, left humour and generalized DSC. For all the other tissues the SSD never decreased the accuracy. Table 4.2 shows mean value and standard deviation for all methods and all tissues.

Therefore JW using SSD, which will be referred simply as JW in the following, was considered the most accurate method and thus it was used for the graph cut segmentation.

The graph cut method was optimized two times: first using the spatial prior that results from MV technique (MVIM) and second using the spatial prior that results from JW based on SSD (JWIM). Table 4.1 lists the set of optimal parameters $\hat{\theta}_l$ for MVIM and JWIM, considering all patients (no leave-one-out). This parameter set is suggested for an independent test image.

Figure 4.2 and 4.3 show the results of experiment 3. Figure 4.2 shows the distribution of DSC over all patients for all tissues and the generalized DSC. Figure 4.3 shows the MSD over all patients, for all tissues and the average value over all tissues. The distributions of DSC and MSD over the set of patients between different methods were compared pairwise. The results of statistical analysis are shown in Figure 4.2 (top part) and 4.3(bottom part). In addition, we reported for the accuracy of each method whether there is no statistically significant difference with IOA/IOV.

Figure 4.4 shows examples of segmentation of the same slice using all methods compared in the third experiments for brain and eye tissues. The first two rows exemplifies the added value in accuracy of using intensity modeling. In the last two rows we show that is more sensitive to intensity variations in the image. In the third row a case where the appearance of the sclera is not homogeneous and lead o a segmentation error when using the intensity model. In the fourth row the intensity inhomogeneity hamper the accuracy of all method, however joint fusion is more robust to this artifact.

We calculated the computation times on a 2.3 GHz quad-core Intel processor, with 12 GB of RAM, running a 64 bit Windows 7 operating system running multi-threads implementations for registration and atlas fusion. The complete registration procedure comprising rigid and non-rigid steps took on average 4 minutes per atlas. The average MV computation time was 45 s/patient, and the average LW and JW computation time was 86 min/patient. The average computation time for the graph cut optimization was 10 min/patient for all tissues. Thus the average segmentation time of a complete dataset considering registrations, JW (or LW), and the graph cut was 136 min/patient.

4.5 Discussion

In this chapter we developed and evaluated a multiatlas segmentation method for H&N MR images which combines local weighting multiatlas fusion with intensity modeling using graph cuts. We evaluated which local weighting technique provides the best spatial prior, and this technique was combined with intensity modeling using

Table 4.2. Accuracy of LW and JW evaluated using DSC mean (standard deviation). Abbreviations: = left; R = right; Gen = generalized DSC without cerebrum and cerebellum; The statistically significant higher ($p < 0.01$) distribution, between LW and JW, were printed in bold. The symbol $\star$ indicates whether a distribution was statistically significantly higher than the correspondent, LW or JW, NCC-based method distribution. The symbol + indicates whether a distribution was statistically significantly better than the correspondent, LW or JW, NMI-based method distribution.

ID	LW - SSD	JW - SSD	LW - NCC	JW - NCC	LW - NMI	JW - NMI
Cerebrum	0.971(0.009)$^{\star+}$	0.971(0.010)$^{+}$	0.967(0.013)	**0.969(0.011)**$^{+}$	0.968(0.012)	0.968(0.012)
Cerebellum	0.945(0.013)$^{\star+}$	0.945(0.014)	0.942(0.015)	**0.944(0.014)**$^{+}$	0.942(0.014)	0.942(0.013)
Brainstem	0.933(0.011)	**0.934(0.011)***	0.933(0.013)	0.933(0.012)	0.934(0.012)	0.934(0.012)
Myelum	0.858(0.096)	0.863(0.091)	0.854(0.111)	0.860(0.096)	0.854(0.113)	0.852(0.116)
CSF	0.818(0.075)$^{\star+}$	0.820(0.072)$^{\star+}$	0.803(0.076)	**0.817(0.072)**$^{+}$	0.806(0.076)*	0.811(0.075)
Sclera_R	0.693(0.047)$^{\star+}$	0.687(0.059)$^{+}$	0.660(0.077)	**0.686(0.064)**$^{+}$	0.665(0.075)*	**0.672(0.073)**
Lens_R	0.882(0.047)$^{\star+}$	0.876(0.050)	0.813(0.189)	**0.865(0.071)**	0.841(0.112)*	**0.853(0.090)**
humour_R	0.960(0.014)$^{\star+}$	0.960(0.012)$^{+}$	0.953(0.021)	**0.960(0.014)**$^{+}$	0.955(0.017)	0.955(0.015)
Sclera_L	0.681(0.041)*	0.680(0.046)	0.653(0.062)	**0.680(0.047)**	0.661(0.057)*	**0.670(0.056)**
Lens_L	0.859(0.052)	0.857(0.060)	0.767(0.262)	**0.842(0.086)**	0.787(0.219)	**0.827(0.124)**
humour_L	0.960(0.008)	0.959(0.011)	0.948(0.026)	**0.959(0.011)**$^{+}$	0.951(0.020)	0.952(0.018)
Gen	0.884(0.027)$^{\star+}$	0.886(0.026)*	0.878(0.028)	**0.885(0.027)**$^{+}$	0.880(0.028)*	0.881(0.028)

Table 4.3. Parameters optimization experiment. Optimal parameters considering all patients. Note that these parameters were not used for the evaluation but should be used for new patients not included in the atlas set. Abbreviations: Par = parameter; _L = Left; _R = Right.

Par	CE	CB	BS	CSF	MY	SC_R	EH_R	LE_R	SC_L	EH_L	LE_L
						MVIM					
λ_1	0.25	0.25	0.125	0.25	1.00	8.00	4.00	8.00	8.00	4.00	1.00
λ_2	2.00	1.00	2.00	2.00	2.00	1.00	0.50	1.00	0.50	0.125	2.00
α	0.40	0.50	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.50	0.40
						JWIM					
λ_1	0.25	0.25	0.125	0.50	0.25	8.00	4.00	8.00	8.00	4.00	4.00
λ_2	4.00	4.00	4.00	4.00	8.00	1.00	1.00	4.00	2.0	0.50	4.00
α	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40

graph cut. Next, accuracies of the various automatic methods were compared between each other and with the interobserver agreement/variability of manual delineation.

The accuracy of the proposed method was high both in terms of DSC and of MSD with a median DSC higher than 0.8 for all tissue except sclera, and a median MSD lower than 1 mm for all tissues except the CSF. For all tissues excluding myelum and CSF, the accuracy of the proposed method is approaching the interobserver agreement/variability. The best local weighting technique was the joint weighting approach using SSD as similarity metric and local patch search to account for registration errors. When a more accurate spatial prior was used, adding the intensity model did not substantially improve or decrease the results. However, using the JWIM method the segmentation are smoother than with JW and the computation costs are only slightly increased. Thus the use of JWIM is suggested.

In the first experiment we evaluated which multiatlas fusion approach gave the best spatial prior. We found that using JW a small but consistent improvement was obtained. In addition the use of local path search ($r_s = 2$ in all cases) consistently led to the optimal fusion parameter. Our results are in-line with the original publication proposing this approach and also demonstrate the validity of the method for other similarity metrics, such as NCC and NMI. Although the improvement was small we there was not difference in computation time between the regular and joint fusion approach as observed by Wang et al. [130]. Thus the use of JW gave only advantages. While the optimal patch radii and local search radii were stable over all methods, the optimal β selection showed that for LW a more non-linear scaling of the similarity metric than JW was needed for obtaining the best accuracy (note that for LW using SSD β was negative). In LW a more non-linear scaling is equivalent to a more selective weighting. In JW the weights are not directly dependent on the (scaled) similarity metric, as the relationship between the (scaled) similarity metric of different atlases is also considered in the calculation. Therefore for JW a selective weighting could be obtained even with a less non-linear scaling of the similarity metrics.

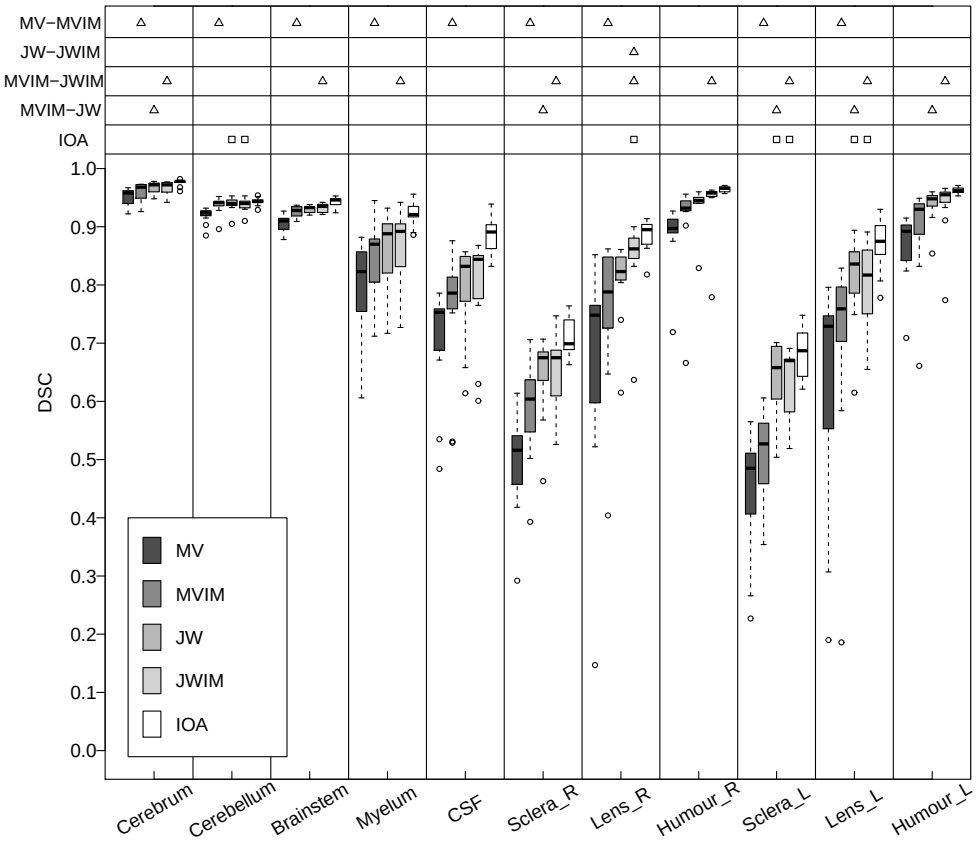


Figure 4.2. Comparison between the accuracy of the four methods and the interobserver agreement (IOA) using DSC. The boxes of the boxplot report first and third quartiles and median values. The whiskers report confidence intervals defined as the most extreme data points (± 1.5 times the inter-quartile range around the box) not considered outliers (∘). Above the boxplot, the results of the statistical analysis are shown. To the left side of this table the two methods under comparison are specified. The symbols △ and ▽ indicate respectively that the distribution of the second method is statistically significantly ($p < 0.01$) higher and lower than of the first method. The symbols □ indicate that there is no significant difference ($p < 0.01$) between the corresponding distribution and the IOA.

In the second experiment we optimized the parameters of the graph cut segmentation method using both MV and JW, with the optimal parameters from the first experiment, to compute the spatial priors. The leave-one-out parameters selection was stable over patients for all tissues. In addition we found that a better spatial prior model (from the JW) consistently led to a higher value (most often twice as high) of λ_2, which could be expected since λ_2 is the relative weight between spatial

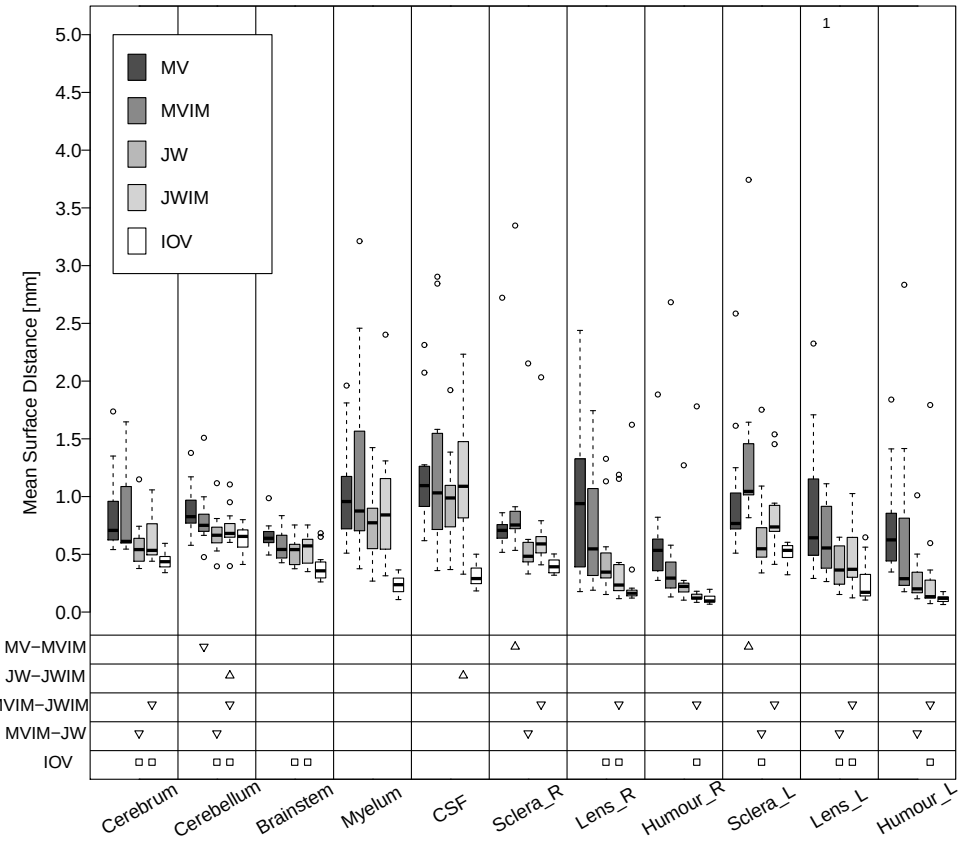


Figure 4.3. Comparison between the accuracy of the four methodsand the interobserver variability (IOV) using MSD. Below the boxplot, the results of the statistical analysis are shown. All symbols have the same meaning as in Figure 4.2 with the difference that we are now reporting the error and not the accuracy, thus lower values are better in this case. For a boxplot, if a number is printed on the top of the plot, it indicates how many outlier with a value higher than 5 mm were found in that distribution.

prior model and intensity model.

In the third experiment, we evaluated four different approaches comparing them to manual reference and to the interobserver agreement/variability. We found that by using a more accurate spatial prior model, the addition of an intensity model became less effective in improving the segmentation accuracy. Using the intensity model in combination with a spatial prior built using MV led to a significant improvement in DSC for all tissues and a slight decrease of accuracy (see MSD) only for the sclera whereas the combination of intensity model (JWIM) and JW significantly improved the accuracy only for the right lens in DSC. For some patients outliers can be observed

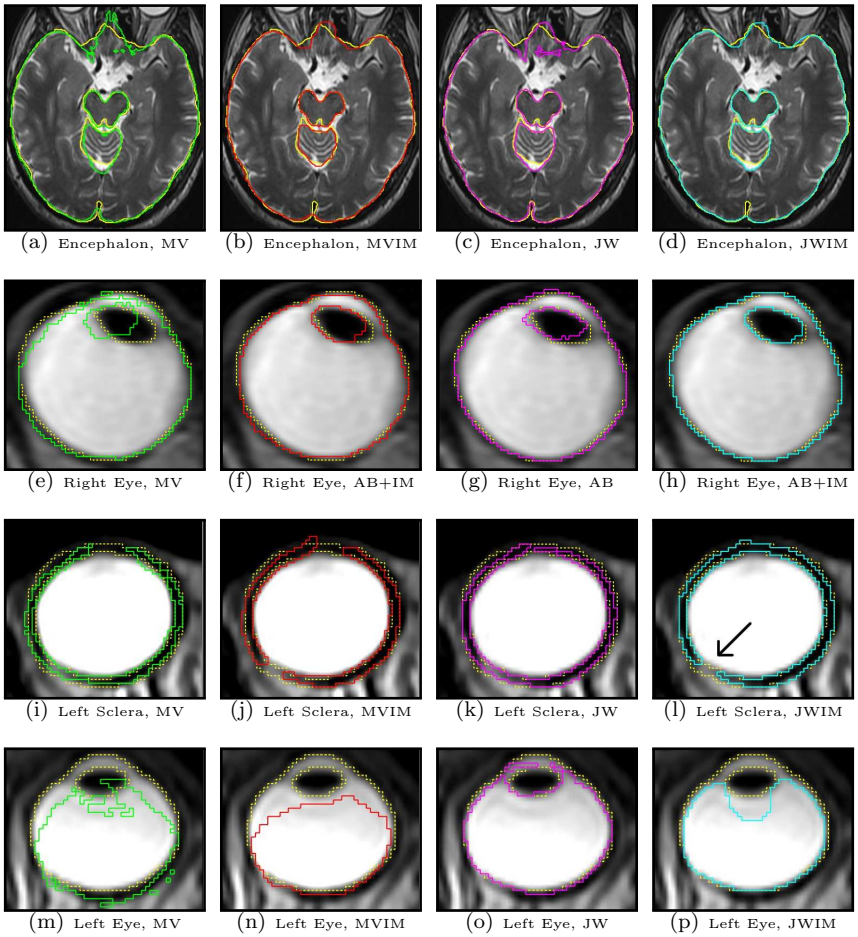


Figure 4.4. Examples of segmentations. Every row shows a comparisons between MV (green), MVIM (red), JW (Cyan), JWIM (Magenta) and manual delineations (yellow) for different tissues. JW and JWIM shows higher accuracy than the other methods. In (b) and (d) the effect of regularization can be seen when comparing to (a) and (c). The arrow in (l) show an example of a location where the appearance of the sclera is not homogeneous.

for the eye tissues both in DSC and MSD. This outlier are caused by signal inhomogeneity due to the use of surface coils (see the last row of Figure 4.4). An improved inhomogeneity correction would prevent these inaccuracies. The MSD shows that for the cerebellum and CSF the use of intensity model led to a small but significant decrease of the segmentation accuracy. However, for the cerebellum this decrease was within the range of IOV. Considering all methods, the worse accuracy for both measures was found at the neck level for the myelum and CSF. This is probably due to limitation of the acquisition coils, i.e. lower signal quality was found at the bottom region of the images which led to intensity inhomogeneity and hampered the accuracy of both spatial prior and intensity model. A better acquisition would prevent this effect. The proposed method, both with and without using the intensity model provided a better accuracy than the MVIM which was previously used for HTP (see Chapter 3). Visual inspection of the segmentation results confirmed that, whereas the JWMI did not improve the accuracy with respect to JW, it provided smoother segmentations, which are generally preferred (e.g. as shown in the first row of Figure 4.4).

The effect of segmentation accuracy on HTP quality was previously examined for CT, in Chapter 3, and MR [126] images. In Chapter 3 it was shown that the effect on treatment quality of using CT-based automatic segmentation was comparable with using manual delineations, thus the method is being used in clinical practice. However manual corrections are currently required for some of the eye tissues. In the second study [126] the HTP quality obtained using patient model based on CT and MR was compared to the HTP quality of patient models based on CT only. It was shown that there was not difference between the quality of MR-based, and CT-based treatment planning using manual delineations. Moreover in this study we reached an accuracy very close to the interobserver agreement/variability such that manual correction might be not required. Therefore, we conclude that the proposed method is ready to be used used in clinic.

In conclusion we developed and evaluated a segmentation method to accurately segment tissue in MR for HTP. We compared and evaluated local weighted voting techniques and combination of the best of these techniques with intensity modeling using graph cut. The proposed method provide a better accuracy than the method currently in use for HTP. The proposed method has an accuracy very close to the interobserver agreement/variability making the method suitable to replace manual delineations in clinical practice.

Chapter Four

Automatic Tissue Segmentation of Head and Neck MR Images for Hyperthermia Treatment Planning

Strive for continuous improvement, instead of perfection.

— Kim Collins (1976)

Abstract — We present here an automatic atlas-based segmentation algorithm for MR images of the head and neck.
Our method combines multiatlas local weighting fusion with intensity modeling. The accuracy of the method was evaluated using a leave-one-out cross validation experiment over a set of 11 patients for which manual delineation were available.
The accuracy of the proposed method was high both in terms of Dice similarity coefficient (DSC) and mean surface distance (MSD) with median DSC higher than 0.8 for all tissues except sclera and median MSD lower than 1 mm for all tissues except the CSF compared to manual delineations. For all tissues, except the spine tissues, the accuracy was approaching the interobserver agreement/variability both in terms of DSC and MSD. The positive effect of adding the intensity modeling to the multiatlas fusion decreased when a more accurate atlas fusion method was used. Using the proposed approach we improved the performance of the approach previously presented for H&N hyperthermia treatment planning making the method making the method suitable for clinical application.

Based upon: **V Fortunati**, R F Verhaart, J F Veenland, W J Niessen M M Paulides and T van Walsum, "Automatic Tissue Segmentation of Head and Neck MR Images for Hyperthermia Treatment Planning.", Major revision in *Phys Med Biol*.

4.1 Introduction

Local hyperthermia treatment (HT) consists of elevating the temperature of tumors between 39° C and 44° C. The effectiveness of radiotherapy (RT) and chemotherapy is significantly enhanced when combined with HT [37, 57, 124]. Phase III clinical studies show that hyperthermia is also effective in the head and neck (H&N) region [54, 55, 121]. In response to these positive results, a site-specific hyperthermia applicator was developed for precise HT of tumors in the entire H&N region using electromagnetic fields. This device, called the HYPERcollar, enables deep heating (> 4 cm under the skin) of H&N cancers [8, 89, 90]. Due to the limited availability of temperature monitoring in-vivo, HT of H&N tumors requires accurate treatment planning (HTP), based on 3D patient models that are derived from segmented 3D images [127]. These segmentations are generally derived by manual delineation of computed tomography (CT) images. However, it is well-known that manual delineation is tedious, labour intense and prone to inter and intraobserver variability. Thus automation of the segmentation process is highly desired. It has been shown that, due to their superior soft tissue contrast, MR images provide better delineation accuracy of target and normal tissues than CT images [2, 45, 96, 117, 126]. Hence, in this chapter we investigate the accuracy of MR based automatic segmentation for some of the tissues needed for HTP. MR-based segmentation need to be registered to the CT coordinate frame in order to obtain a geometrically accurate full 3D patient model [35, 126].

Several methods have been developed in recent years to address the problem of automatic segmentation of H&N CT images for generating anatomical models for radiotherapy [39,48,95] treatment planning. The common aspect of all these methods is that they are based on multiatlas segmentation approaches. In the field of HT planning, we recently published a method to accurately segment all relevant tissue in CT images [34] using atlases and intensity modeling. We also showed that the use of this method leads to a HTP quality comparable to the quality obtained using manually generated delineations [127].

The problem of automatic segmentation using atlases was extensively addressed in the field of brain structure segmentation in MR images [3,6,101,122,130]. Whereas majority voting, being a simple and straightforward way, has been frequently used to combine atlas registration results, more recently multiple authors have shown that local weighting atlas fusion techniques [6, 101, 130] increase the segmentation accuracy. In these approaches, segmentations from multiple atlases are combined using local image similarity metrics to locally weigh the contribution of each segmentation. Among these methods, the approach developed by Wang [130] ranked first in a recent MICCAI grand challenge [75]. Due to the effectiveness of this method we decided to combine local weighting multiatlas fusion with the approach we previously used for CT images [34].

In this chapter, we developed and evaluated an automatic segmentation algorithm for MR images of the head and neck; we used a graph cut method which combines multiatlas-based segmentation, using atlas fusion, and intensity modeling. The method combines the approach described earlier [34] with local weighting atlas fusion [6, 130]. The contribution of our work is twofold: 1) we developed an auto-

matic segmentation method for head and neck MR images that integrates intensity modeling into state-of-the art multiatlas fusion methods; 2) we evaluated various versions of our method on imaging data of 11 patients, and compared the results to the interobserver variability of manual delineations.

4.2 Methods and materials

4.2.1 Image data and atlas

The data used in this chapter was collected from 11 patients with tumors in the H&N region, which were representative for the patient-group eligible for HT. The location of the tumors was the nasopharynx (3 patients) and the oropharynx (8 patients). The data comprised axial, MR T2 weighted (T2w) images acquired on a 1.5 T scanner (Optima MR450w, GE Healthcare, Waukesha, WI) with the following acquisition parameters: Fast Recovery Fast Spin Echo sequence with TE/TR 106.8/7060 ms; in plane resolution of 0.49 mm (5 patients) and 0.68 mm (6 patients); slice thickness of 3 mm; out of plane field of view of 20 cm centered at the tumor location, yielding a slightly different field of view for each patient. For 8 patients, images were acquired in treatment position using a 6 channel surface coil while for the other 3 patients a 24 cnhannel H&N coil was used. The T2w images were corrected for intensity inhomogeneity: for the images acquired with surface coils, a built- in algorithm (Surface Coil Intensity Correction) of the MR scanner was used at the time of acquisition; for the images acquired with H&N coils, the ITK-N4 method [118] was used as a post-processing procedure. In order to compensate for the differences in intensities, all the images were normalized [119]. This normalization involved matching all the intensity histograms to the histogram of one randomly selected image. The code for this normalization is in the ITK library (www.itk.org, itkHistogramMatchingImageFilter.h). We used 256 bins for the histogram computation matching 128 histogram landmarks.

Three trained observers (O1, O2, O3) delineated the tissues in all images. We included the tissues for which the dielectric and thermal properties are different from those of the surrounding tissues. Additionally, tissue delineation had to be feasible on T2w images. With these criteria, the tissues of interest were:

- Encephalon: cerebrum (CE); cerebellum (CB); brainstem (BS).
- Spinal cord: myelum (MY) and cerebro spinal fluid (CSF).
- Eye tissues: vitreous humor (EH); sclera (SC); lens (LE).

The optical nerve was excluded from this study even though it is relevant for HTP. This was done because the optical nerve could not be consistently delineated on T2w images.

4.2.2 Segmentation method

Our segmentation method combines locally weighted multiatlas fusion with voxel-wise classification based on intensity values. Atlas based segmentation consists of the

registration of a target image, with an unknown segmentation, to an atlas for which the segmentation is known. The registration is used to estimate a segmentation by spatially mapping the known segmentation to the target image. In multiatlas segmentation this process is repeated for more than one atlas and an atlas fusion method is used to combine the information of all atlases into a single optimal segmentation [98].

In our method, an atlas fusion method was used to determine a spatial probability map for each tissue. The atlases, in the original coordinate frame, were further used to estimate the intensity distribution of the tissues and their corresponding background regions. This information was used in a graph-cut segmentation algorithm: the probability map was used as a spatial prior and the intensity distributions as an intensity model. A regularization term was also used to ensure a smooth segmentation. Figure 4.1 shows a scheme of the segmentation method.

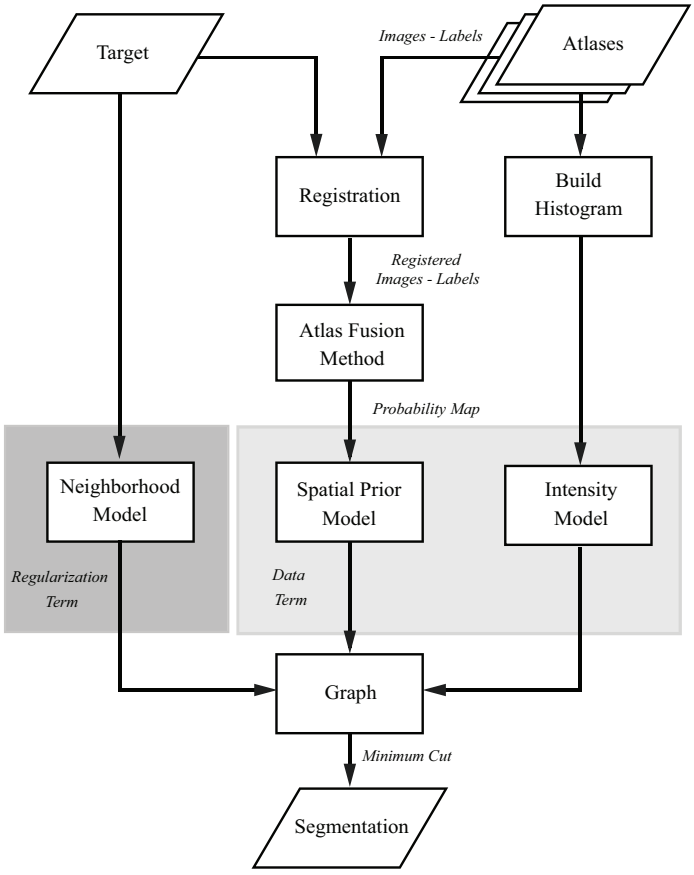


Figure 4.1. Scheme of the segmentation method for each tissue.

For details on how the intensity model was combined with the spatial probability map we refer the reader to our previous work [34]. In the following we focus on

the local weighting atlas fusion technique and its integration into our segmentation framework.

4.2.3 Spatial probability map construction

To construct a spatial prior, first a set of J atlas images was non-rigidly registered to the target image u. Using the registration results for each atlas j and each label l we obtained the segmentation $S_j^l(m)$ and the registered images $I_j(m)$. Fusion was performed with a local weighted voting technique [6,130]. In this method the label at voxel location m was selected as the one with the highest vote $\hat{S}^l(m)$, where $\hat{S}^l(m)$ is computed according to the rule:

$$\hat{S}^l(m) = \sum_{j=0}^{J} \omega_j(m) S_j^l(m), \tag{4.1}$$

where $\omega_j(m)$ is a weighting factor for the location m, which reflects the accuracy of each single atlas segmentation, in the proximity of m. Thus the final vote was calculated by weighting the contribution of each single atlas segmentation with respect to their local segmentation quality.

The weights $\omega_j(m)$ for each location m and atlas j can be estimated using the local image similarity between a patch of the target image $\mathcal{I}_u(\mathcal{N}(m))$ and a patch of atlas j registered to the coordinate frame of the target image u and interpolated at voxel location m, $\mathcal{I}_j(\mathcal{N}(m))$. The patch $\mathcal{N}(m)$ was defined as a 3D cubical region centered at m with width (in voxels) of $2r_p + 1$. We evaluated different similarity metrics (SIM) for local fusion: sum of squared differences (SSD), normalized cross-correlation (NCC) and normalized mutual information (NMI). The weights $\omega_j(m)$ were defined following the work of Artaechevarria et al. [6]:

$$\omega_j(m) = SIM(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m)))^\beta, \tag{4.2}$$

where β is a factor which was positive for NCC and NMI, and negative for SSD in order to make ω a measure of similarity.

Two different advanced weighting techniques have been proposed to improve this generic local weighting framework: 1) using a local search technique to decrease the effects of small registration error; 2) computing the final weight of each atlas considering the joint image similarity between the target image and all atlases. Both approaches were integrated into our method; we discuss them in the next two subsections.

4.2.3.1 Local search technique

The anatomical variability between the atlas and target image makes the registration challenging and prone to errors. This error can be accounted for during atlas fusion [101, 130] by locally searching for the atlas image patch that has the best similarity with the patch of the target image u. Next, the segmentation corresponding to the most similar atlas image patch can be used in the modified voting rule:

$$\hat{S}^l(m) = \sum_{j=0}^{J} \omega_j(\eta_j(m)) S_j^l(\eta_j(m)), \tag{4.3}$$

where $\eta_j(m)$ is the location in the atlas j corresponding to the patch $\mathcal{N}$ that gives the best match between atlas and target image in the local search region $\mathcal{N}_s$ according to the local search:

$$\eta_j(m) = \underset{m' \in \mathcal{N}_s}{\operatorname{argmin}} \|\mathcal{I}_j(\mathcal{N}(m')) - \mathcal{I}_u(\mathcal{N}(m))\|^2, \tag{4.4}$$

$\mathcal{I}_j(\mathcal{N}(m'))$ is the vector of intensities of the j-th atlas within the patch $\mathcal{N}$ centered around m' and $\mathcal{I}_u(\mathcal{N}(m))$ is the vector of intensities of the target image u within the patch $\mathcal{N}$ centered around m. The patch $\mathcal{N}_s(m)$ is defined as a 3D cubical region centered at m with width (in voxels) of $2r_s + 1$.

4.2.3.2 Joint weighted voting

Wang et al. [130] additionally introduced a weighting technique to estimate of the final segmentation taking a possibly correlated error between atlas segmentations into account. In this method the optimal weights are computed by minimizing the total expectation of segmentation error. This expectation is expressed as a matrix containing the joint probabilities of two atlases making a segmentation error at a certain location. Similarly to the general local weighting method this probability is approximated using an intensity similarity between a pair of atlases and the target image in the neighborhood of each voxel. Let $\boldsymbol{\omega}_m = (\omega_0(m), ..., \omega_J(m))$ denote the vector of weights. Following Wang et al. [130] the optimal weights are calculated as:

$$\boldsymbol{\omega}_m^* = \frac{(M_m + \gamma I)^{-1}\mathbf{1}_n}{\mathbf{1}_n^T(M_m + \gamma I)^{-1}\mathbf{1}_n}, \tag{4.5}$$

where γ is a regularization term used to ensure a unique solution for the minimization problem, $\mathbf{1}_n = [1; 1; ....; 1]$ and I is the identity matrix. The weights in Equation 4.5 are normalized, i.e. $\sum_j \omega_m(i) = 1$. The matrix M contains the expected pairwise joint label differences at each location m, which can be estimated using local image similarity, i.e. SSD, NCC and NMI. NCC and the NMI need to be inverted in order to represent an error measure. In case of SSD we defined the matrix M_m according to Wang et al. [130]. For the NCC based joint fusion the matrix was defined as:

$$M_m^{NCC}(i,j) = \left(\sqrt{1 - NCC(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_i(\mathcal{N}(m))}\sqrt{1 - NCC(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m))}\right)^{\beta}, \tag{4.6}$$

where i and j indicate the atlases, β is a positive scaling factor and the value one is used to invert the NCC since $NCC \in [-1, 1]$. Similarly for the NMI based joint fusion we have:

$$M_m^{NMI}(i,j) = \left(\sqrt{3 - NMI(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_i(\mathcal{N}(m))}\sqrt{3 - NMI(\mathcal{I}_u(\mathcal{N}(m)), \mathcal{I}_j(\mathcal{N}(m)))}\right)^{\beta}, \tag{4.7}$$

with the same notation as 4.6. In this case we use the value three to invert the metric since $NMI \in [1, 2]$. Following [130] for all the three metrics, the value of γ was set to approximatively 2% of the maximum value of the similarity metric i.e. 0.1 for SSD and 0.05 for NMI and NCC.

4.2.4 Graph cut segmentation

The multiatlas fusion methods provides a probability map $p_s^l(f_m)$ for each tissue:

$$p_s^l(f_m) = \left(\frac{1}{\sum_j \omega_j(\eta_j(m))} \sum_j S_j^l(\eta_j(m)) \omega_j(\eta_j(m)) \right)^{\lambda_2}, \tag{4.8}$$

where λ_2 is a weighting parameter. This probability map is used as spatial prior in the data terms of the graph cut segmentation algorithm [14]. The other data term used is the intensity model of the tissue which consisted of the intensity distributions of the tissue and its corresponding background region. The intensity distributions is built using the unregistered atlases by a Parzen-window histogram estimation [34]. The intensity model aims to correct for residual registration errors or systematic error in the output segmentation. Two user-defined parameters were used: the weight parameter λ_1 to weigh the contribution of the data terms with respect to the regularization term and the parameter λ_2 to weigh the contribution of the spatial prior term with respect to the intensity model term.

For each tissue l, the graph-cut segmentation depends on a vector of parameters $\theta_l = (\alpha, \lambda_1, \lambda_2)$, i.e. the data term weight λ_1, the spatial model weight λ_2 and the foreground threshold α [34]. This last parameter is introduced to correct for under-segmentation. For each tissue, the optimal value of the parameters $\hat{\theta}_l$ depends on the spatial and intensity model; these parameters are estimated in a cross-validation experiment on the available atlases.

A max-flow algorithm is used to obtain the optimal cut [13] of the graph. The graph cut segmentation is computed separately for each tissue and next all results are combined to create a single multi-label image as described in [34].

4.2.5 Implementation details

Atlas and target images were registered using the opensource registration software elastix [66]. A two-step registration was performed in a multiresolution fashion using mutual information as similarity metric [115] and adaptive stochastic gradient descent optimization [65] with 4000 random samples at each iteration. The first registration step consisted of an affine registration which was done using four resolution levels with 500 iterations per level. The second step consisted of a free form deformation [100], based on third order B-Spline polynomials. In this step, three resolution levels were used and the number of iterations varied from the coarsest to the finest resolution level (500, 1000 and 2000). These parameters were obtained after training on an independent set of 11 T2w images acquired with the same protocol as the segmentation dataset. Twenty two anatomical landmarks [35] were manually placed in the images to assess the registration accuracy and therefore enable the selection of the best parameters.

The intensity model was built by random sampling from the set of training images. Five thousands foreground samples were selected randomly from the entire training set for all tissues, except for the sclera (1500) and for the lenses (700) due to the small volume of the these tissues. In all cases, the same number of samples was selected for

background and foreground where the background region was defined as a band with a width of 5 mm around the foreground.

4.3 Experiments

Three experiments were performed to optimize the method parameters and evaluate its performance. The purpose of the *first experiment* was to evaluate which strategy lead to the most accurate spatial priors. Six strategies were compared: local weighting voting (LW) using 1) SSD, 2) NCC and 3) NMI as image similarity metric; joint weighting voting (JW) using 4) SSD, 5) NCC and 6) NMI as image similarity metric. For each strategy the local similarity patch radius r_p, local search radius r_s, and scaling exponent β were optimized in a leave-one-out cross validation experiment. A pre-defined set of parameters was used for all methods: $r_p \in \{1, 2, 3, 4\}$ voxels; $r_s \in \{0, 1, 2, 3, 4\}$ voxels; $\beta \in \{0.5, 1, 2, ..., 10\}$. The highest average generalized Dice Similarity Coefficient [27] (see also Section 2.3.3) over all images calculated using all tissues except cerebrum and cerebellum (DSC_{Gen}) was used to select the best parameter set. These tissues were excluded since their volumes are much larger than the volume of the other tissues, and thus would dominate the DSC_{Gen}. The DSC measures the normalized overlap between two volumes and the DSC_{Gen} extends the measure to two sets of volumes. For each segmentation the DSC_{Gen} was defined as:

$$\mathrm{DSC}_{Gen} = \frac{2 \sum_{l \in L^*} \hat{S}^l \cap G^l}{\sum_{l \in L^*} \hat{S}^l \cup G^l}, \tag{4.9}$$

where G^l is the ground truth segmentation of tissue l, $\hat{S}^l$ is the automatic segmentation and L^* is the set of all tissue labels excluding cerebrum and cerebellum. Using the optimal parameters, we evaluated the accuracy of the segmentation of each tissue by DSC. For both tuning and evaluation the manual delineations from O1 were used. The best method (using the best metric) was then selected as the one giving the best overall accuracy in DSC. This method was then used as spatial prior in the graph cut segmentation.

Given the optimal method for computing a spatial prior, the purpose of the *second experiment* was to optimize the parameters λ_1, λ_2, and α for the graph cut segmentation. As in [34], this optimization was done in a leave-one-out cross validation experiment consisting on an exhaustive search over a pre-defined set of parameters: $\lambda_1, \lambda_2 \in \{0.125, 0.25, 0.5, 1, 2, 4, 8\}$ and $\alpha \in \{0.4, 0.5\}$. The delineations of observer O1 were used in this experiment. Additionally, the optimal parameters using the MV to compute the spatial prior were calculated.

The purpose of the *third experiment* was to evaluate the overall performance of the combined methods. Specifically we verified the effectiveness of adding the intensity model for the problem at hand, the influence of the spatial prior on the method accuracy, and the accuracy of each method compared to the human observer ability to consistently segment a certain tissue. For this, the optimal version of the methods resulting from the previous two experiments was employed. The following methods were thus compared:

- multiatlas-based majority voting [51].
- multiatlas-based majority voting combined with intensity model (MV+IM) [34].
- best local/joint weighting atlas voting (LW/JW).
- best local/joint weighting atlas voting combined with intensity model (LWIM/JWIM).
- interobserver agreement/variability (IOA/IOV).

For the evaluation the manually-generated label images from the three observers were used as ground truth. The graph-cut method was applied three times using the atlases from each of the observers to built the intensity models and the spatial priors, in a leave-one-out cross-validation scenario. Then the three results were averaged for each patient, to estimate the overall method accuracy. To have a complete quantitative overview of the segmentation accuracy, two evaluation measures were used to compare the segmented volumes: the DSC and the Mean Surface Distance (MSD) [29] (for the definition of these measures see Section 2.3.3).

In the first and third experiment we performed pair-wise comparison between the distribution of accuracies per patients. To assess whether a distribution was higher or lower than another, significance was tested using a one-sided Wilcoxon signed rank test. Significant difference was tested using a two-sided Wilcoxon signed rank test ($p < 0.01$).

4.4 Results

In the first experiment the parameters of LW and JW were optimized for each of the three similarity metrics. Table 4.1 reports the leave-one-out optimal parameters for LW and JW using all metrics.

Table 4.1. Optimal leave-one-out parameters for LW and JW;

Method	Metric	Patient	r_p	r_s	β
LW	SSD	5/6	4	2	$-2/-3$
LW	NCC, NMI	All	4	2	10
JW	SSD/NCC	All	4	2	2/1
JW	NMI	9/2	4	2	1/0.5

For LW, SSD provided a statistically significantly ($p < 0.01$) better accuracy than both NCC and NMI for all tissue except the brainstem, myelum, left lens and humour. NMI provided a statistically significantly ($p < 0.01$) better accuracy than NCC for myleum, right sclera and lens, left sclera and generalized DSC. For all the other tissues SSD never decreased the accuracy.

For JW, SSD provided a statistically significantly ($p < 0.01$) better accuracy than NCC for the brainstem, CSF, and generalized DSC. SSD also provided more accurate

results than NMI: a statistically significantly ($p < 0.01$) better accuracy was found for cerebrum, CSF, right sclera and humour, left humour and generalized DSC. NCC provided a statistically significantly ($p < 0.01$) better accuracy than NMI for cerebrum, cerebellum, CSF, right sclera and humour, left humour and generalized DSC. For all the other tissues the SSD never decreased the accuracy. Table 4.2 shows mean value and standard deviation for all methods and all tissues.

Therefore JW using SSD, which will be referred simply as JW in the following, was considered the most accurate method and thus it was used for the graph cut segmentation.

The graph cut method was optimized two times: first using the spatial prior that results from MV technique (MVIM) and second using the spatial prior that results from JW based on SSD (JWIM). Table 4.1 lists the set of optimal parameters $\hat{\theta}_l$ for MVIM and JWIM, considering all patients (no leave-one-out). This parameter set is suggested for an independent test image.

Figure 4.2 and 4.3 show the results of experiment 3. Figure 4.2 shows the distribution of DSC over all patients for all tissues and the generalized DSC. Figure 4.3 shows the MSD over all patients, for all tissues and the average value over all tissues. The distributions of DSC and MSD over the set of patients between different methods were compared pairwise. The results of statistical analysis are shown in Figure 4.2 (top part) and 4.3(bottom part). In addition, we reported for the accuracy of each method whether there is no statistically significant difference with IOA/IOV.

Figure 4.4 shows examples of segmentation of the same slice using all methods compared in the third experiments for brain and eye tissues. The first two rows exemplifies the added value in accuracy of using intensity modeling. In the last two rows we show that is more sensitive to intensity variations in the image. In the third row a case where the appearance of the sclera is not homogeneous and lead o a segmentation error when using the intensity model. In the fourth row the intensity inhomogeneity hamper the accuracy of all method, however joint fusion is more robust to this artifact.

We calculated the computation times on a 2.3 GHz quad-core Intel processor, with 12 GB of RAM, running a 64 bit Windows 7 operating system running multi-threads implementations for registration and atlas fusion. The complete registration procedure comprising rigid and non-rigid steps took on average 4 minutes per atlas. The average MV computation time was 45 s/patient, and the average LW and JW computation time was 86 min/patient. The average computation time for the graph cut optimization was 10 min/patient for all tissues. Thus the average segmentation time of a complete dataset considering registrations, JW (or LW), and the graph cut was 136 min/patient.

4.5 Discussion

In this chapter we developed and evaluated a multiatlas segmentation method for H&N MR images which combines local weighting multiatlas fusion with intensity modeling using graph cuts. We evaluated which local weighting technique provides the best spatial prior, and this technique was combined with intensity modeling using

Table 4.2. Accuracy of LW and JW evaluated using DSC mean (standard deviation). Abbreviations: = left; R = right; Gen = generalized DSC without cerebrum and cerebellum; The statistically significant higher ($p < 0.01$) distribution, between LW and JW, were printed in bold. The symbol $\star$ indicates whether a distribution was statistically significantly higher than the correspondent, LW or JW, NCC-based method distribution. The symbol + indicates whether a distribution was statistically significantly better than the correspondent, LW or JW, NMI-based method distribution.

ID	LW - SSD	JW - SSD	LW - NCC	JW - NCC	LW - NMI	JW - NMI
Cerebrum	0.971(0.009)$^{\star+}$	0.971(0.010)$^{+}$	0.967(0.013)	**0.969(0.011)**$^{+}$	0.968(0.012)	0.968(0.012)
Cerebellum	0.945(0.013)$^{\star+}$	0.945(0.014)	0.942(0.015)	**0.944(0.014)**$^{+}$	0.942(0.014)	0.942(0.013)
Brainstem	0.933(0.011)	**0.934(0.011)***	0.933(0.013)	0.933(0.012)	0.934(0.012)	0.934(0.012)
Myelum	0.858(0.096)	0.863(0.091)	0.854(0.111)	0.860(0.096)	0.854(0.113)	0.852(0.116)
CSF	0.818(0.075)$^{\star+}$	0.820(0.072)$^{\star+}$	0.803(0.076)	**0.817(0.072)**$^{+}$	0.806(0.076)*	0.811(0.075)
Sclera_R	0.693(0.047)$^{\star+}$	0.687(0.059)$^{+}$	0.660(0.077)	**0.686(0.064)**$^{+}$	0.665(0.075)*	**0.672(0.073)**
Lens_R	0.882(0.047)$^{\star+}$	0.876(0.050)	0.813(0.189)	**0.865(0.071)**	0.841(0.112)*	**0.853(0.090)**
humour_R	0.960(0.014)$^{\star+}$	0.960(0.012)$^{+}$	0.953(0.021)	**0.960(0.014)**$^{+}$	0.955(0.017)	0.955(0.015)
Sclera_L	0.681(0.041)*	0.680(0.046)	0.653(0.062)	**0.680(0.047)**	0.661(0.057)*	**0.670(0.056)**
Lens_L	0.859(0.052)	0.857(0.060)	0.767(0.262)	**0.842(0.086)**	0.787(0.219)	**0.827(0.124)**
humour_L	0.960(0.008)	0.959(0.011)	0.948(0.026)	**0.959(0.011)**$^{+}$	0.951(0.020)	0.952(0.018)
Gen	0.884(0.027)$^{\star+}$	0.886(0.026)*	0.878(0.028)	**0.885(0.027)**$^{+}$	0.880(0.028)*	0.881(0.028)

Table 4.3. Parameters optimization experiment. Optimal parameters considering all patients. Note that these parameters were not used for the evaluation but should be used for new patients not included in the atlas set. Abbreviations: Par = parameter; _L = Left; _R = Right.

Par	CE	CB	BS	CSF	MY	SC_R	EH_R	LE_R	SC_L	EH_L	LE_L
						MVIM					
λ_1	0.25	0.25	0.125	0.25	1.00	8.00	4.00	8.00	8.00	4.00	1.00
λ_2	2.00	1.00	2.00	2.00	2.00	1.00	0.50	1.00	0.50	0.125	2.00
α	0.40	0.50	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.50	0.40
						JWIM					
λ_1	0.25	0.25	0.125	0.50	0.25	8.00	4.00	8.00	8.00	4.00	4.00
λ_2	4.00	4.00	4.00	4.00	8.00	1.00	1.00	4.00	2.0	0.50	4.00
α	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40

graph cut. Next, accuracies of the various automatic methods were compared between each other and with the interobserver agreement/variability of manual delineation.

The accuracy of the proposed method was high both in terms of DSC and of MSD with a median DSC higher than 0.8 for all tissue except sclera, and a median MSD lower than 1 mm for all tissues except the CSF. For all tissues excluding myelum and CSF, the accuracy of the proposed method is approaching the interobserver agreement/variability. The best local weighting technique was the joint weighting approach using SSD as similarity metric and local patch search to account for registration errors. When a more accurate spatial prior was used, adding the intensity model did not substantially improve or decrease the results. However, using the JWIM method the segmentation are smoother than with JW and the computation costs are only slightly increased. Thus the use of JWIM is suggested.

In the first experiment we evaluated which multiatlas fusion approach gave the best spatial prior. We found that using JW a small but consistent improvement was obtained. In addition the use of local path search ($r_s = 2$ in all cases) consistently led to the optimal fusion parameter. Our results are in-line with the original publication proposing this approach and also demonstrate the validity of the method for other similarity metrics, such as NCC and NMI. Although the improvement was small we there was not difference in computation time between the regular and joint fusion approach as observed by Wang et al. [130]. Thus the use of JW gave only advantages. While the optimal patch radii and local search radii were stable over all methods, the optimal β selection showed that for LW a more non-linear scaling of the similarity metric than JW was needed for obtaining the best accuracy (note that for LW using SSD β was negative). In LW a more non-linear scaling is equivalent to a more selective weighting. In JW the weights are not directly dependent on the (scaled) similarity metric, as the relationship between the (scaled) similarity metric of different atlases is also considered in the calculation. Therefore for JW a selective weighting could be obtained even with a less non-linear scaling of the similarity metrics.

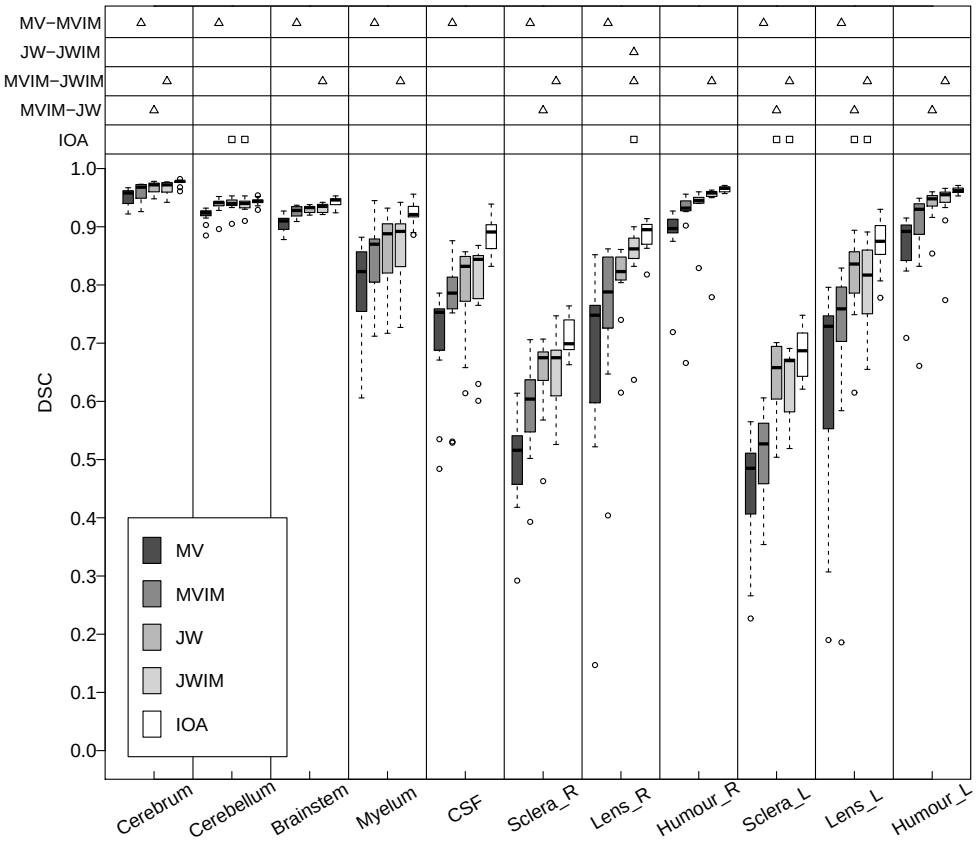


Figure 4.2. Comparison between the accuracy of the four methods and the interobserver agreement (IOA) using DSC. The boxes of the boxplot report first and third quartiles and median values. The whiskers report confidence intervals defined as the most extreme data points (± 1.5 times the inter-quartile range around the box) not considered outliers (∘). Above the boxplot, the results of the statistical analysis are shown. To the left side of this table the two methods under comparison are specified. The symbols △ and ▽ indicate respectively that the distribution of the second method is statistically significantly ($p < 0.01$) higher and lower than of the first method. The symbols □ indicate that there is no significant difference ($p < 0.01$) between the corresponding distribution and the IOA.

In the second experiment we optimized the parameters of the graph cut segmentation method using both MV and JW, with the optimal parameters from the first experiment, to compute the spatial priors. The leave-one-out parameters selection was stable over patients for all tissues. In addition we found that a better spatial prior model (from the JW) consistently led to a higher value (most often twice as high) of λ_2, which could be expected since λ_2 is the relative weight between spatial

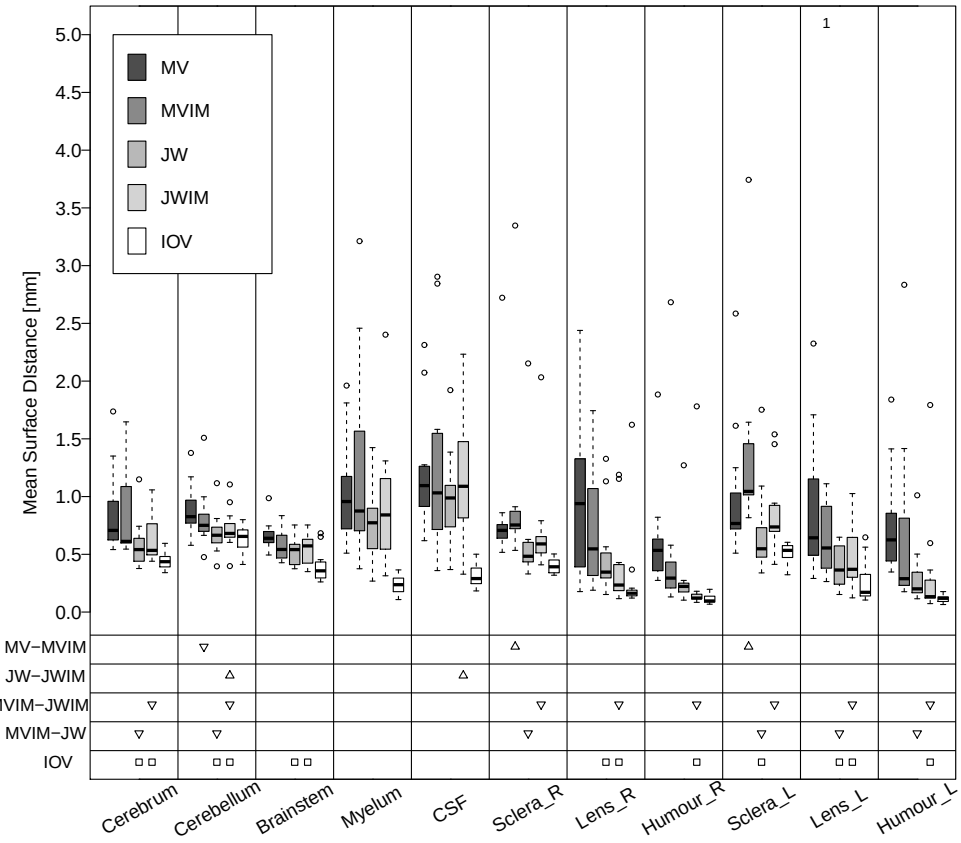


Figure 4.3. Comparison between the accuracy of the four methodsand the interobserver variability (IOV) using MSD. Below the boxplot, the results of the statistical analysis are shown. All symbols have the same meaning as in Figure 4.2 with the difference that we are now reporting the error and not the accuracy, thus lower values are better in this case. For a boxplot, if a number is printed on the top of the plot, it indicates how many outlier with a value higher than 5 mm were found in that distribution.

prior model and intensity model.

In the third experiment, we evaluated four different approaches comparing them to manual reference and to the interobserver agreement/variability. We found that by using a more accurate spatial prior model, the addition of an intensity model became less effective in improving the segmentation accuracy. Using the intensity model in combination with a spatial prior built using MV led to a significant improvement in DSC for all tissues and a slight decrease of accuracy (see MSD) only for the sclera whereas the combination of intensity model (JWIM) and JW significantly improved the accuracy only for the right lens in DSC. For some patients outliers can be observed

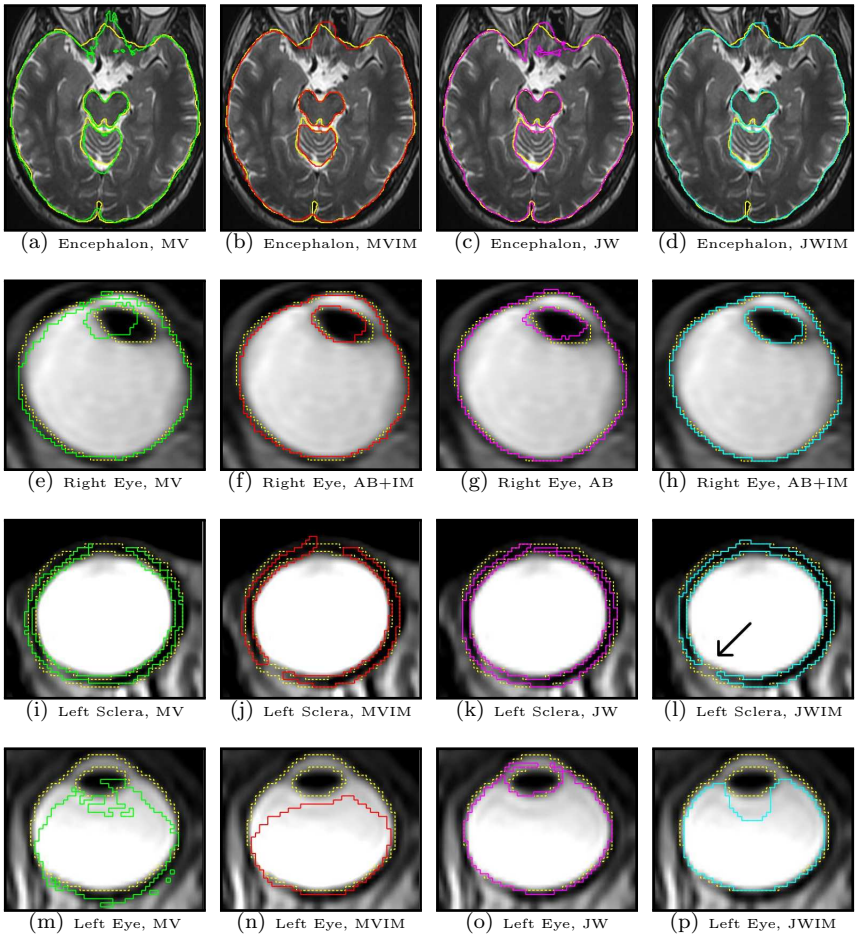


Figure 4.4. Examples of segmentations. Every row shows a comparisons between MV (green), MVIM (red), JW (Cyan), JWIM (Magenta) and manual delineations (yellow) for different tissues. JW and JWIM shows higher accuracy than the other methods. In (b) and (d) the effect of regularization can be seen when comparing to (a) and (c). The arrow in (l) show an example of a location where the appearance of the sclera is not homogeneous.

for the eye tissues both in DSC and MSD. This outlier are caused by signal inhomogeneity due to the use of surface coils (see the last row of Figure 4.4). An improved inhomogeneity correction would prevent these inaccuracies. The MSD shows that for the cerebellum and CSF the use of intensity model led to a small but significant decrease of the segmentation accuracy. However, for the cerebellum this decrease was within the range of IOV. Considering all methods, the worse accuracy for both measures was found at the neck level for the myelum and CSF. This is probably due to limitation of the acquisition coils, i.e. lower signal quality was found at the bottom region of the images which led to intensity inhomogeneity and hampered the accuracy of both spatial prior and intensity model. A better acquisition would prevent this effect. The proposed method, both with and without using the intensity model provided a better accuracy than the MVIM which was previously used for HTP (see Chapter 3). Visual inspection of the segmentation results confirmed that, whereas the JWMI did not improve the accuracy with respect to JW, it provided smoother segmentations, which are generally preferred (e.g. as shown in the first row of Figure 4.4).

The effect of segmentation accuracy on HTP quality was previously examined for CT, in Chapter 3, and MR [126] images. In Chapter 3 it was shown that the effect on treatment quality of using CT-based automatic segmentation was comparable with using manual delineations, thus the method is being used in clinical practice. However manual corrections are currently required for some of the eye tissues. In the second study [126] the HTP quality obtained using patient model based on CT and MR was compared to the HTP quality of patient models based on CT only. It was shown that there was not difference between the quality of MR-based, and CT-based treatment planning using manual delineations. Moreover in this study we reached an accuracy very close to the interobserver agreement/variability such that manual correction might be not required. Therefore, we conclude that the proposed method is ready to be used used in clinic.

In conclusion we developed and evaluated a segmentation method to accurately segment tissue in MR for HTP. We compared and evaluated local weighted voting techniques and combination of the best of these techniques with intensity modeling using graph cut. The proposed method provide a better accuracy than the method currently in use for HTP. The proposed method has an accuracy very close to the interobserver agreement/variability making the method suitable to replace manual delineations in clinical practice.

Chapter Five

Feasibility of Multi-modal Deformable Registration for Head and Neck Tumor Treatment Planning

Measure what is measurable, and make measurable what is not so.

— Galileo Galilei (1564 1642 A.C.)

Abstract — *Purpose:* To investigate the feasibility of using deformable registration in clinical practice to fuse MR and CT images of the H&N for treatment planning.
Method and Materials: A state-of-the-art deformable registration algorithm was optimized, evaluated, and compared with rigid registration. The evaluation was based on manually annotated anatomical landmarks and regions of interest (ROIs) in both modalities. We also developed a multiparametric registration approach, which simultaneously aligns T1- and T2-weighted MR sequences to CT. This was evaluated and compared with single-parametric approaches.
Results: Our results show that deformable registration yielded a better accuracy than rigid registration, without introducing unrealistic deformations. The registration error was lower than the voxel resolution for landmarks-based (median error around 1.6 mm) and ROIs-based evaluation (mean surface distance of approximately 1 mm for all the ROIs excluding the cerebellum). Similar accuracies were obtained for the single and multiparametric approaches.
Conclusions: This study demonstrates that deformable registration of H&N CT and MR images is feasible with a registration error lower than the voxel size.

V Fortunati, R F Verhaart, F Angeloni, A van der Lugt, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "Feasibility of multi-modal deformable registration for head and neck tumor treatment planning", *Int J Radiat Oncol Biol Phys*, vol. 90, no. 1 pp. 85-93, Sep. 2014.

5.1 Introduction

Radiotherapy and hyperthermia treatment planning (RTP and HTP) in the Head and Neck (H&N) region require tissue delineation. This is generally performed on Computed Tomography (CT) images [31,89]. The superior tissue contrast of Magnetic Resonance (MR) images with respect to CT allows improved delineation of target and normal tissues [2,45,96]. CT images still play an important role in treatment planning because they are geometrically accurate and provide electron density information. To exploit the advantages of both CT and MR, the images must be spatially aligned. Because the CT geometry is highly accurate, the MR images needs to be aligned to CT.

Current practice involves rigid registration which requires patient immobilization equipment to limit the deformation between MR and CT scanning sessions. Moreover, geometrical distortion correction algorithms are generally applied to the MR images to enable an accurate alignment with CT using rigid registration only [17, 96, 133]. A quantitative evaluation of the achievable alignment accuracy was performed on phantom images [84, 86] as well as on patient data [133]. Imperfections of immobilization equipment and inaccurate geometrical distortion correction [17] lead to the presence of non-rigid deformation between CT and MR images. These problems can be addressed by using deformable registration eliminating the need for geometrical distortion correction.

Deformable intra-patient registration between CT and MR images was evaluated for several applications in different anatomical regions [15] and in H&N [28, 76]. In both H&N studies no immobilization equipment was used and the evaluation set-up was limited: in the first study [28] the alignment of four patients was evaluated using unspecified anatomical landmarks; in the second study [76] the registration parameters were optimized with respect to the best accuracy obtained over the entire dataset of 12 patients.

The purpose of this study is to investigate the feasibility of deformable registration to fuse MR and CT images of the H&N for RTP and HTP. We optimized and evaluated the optimal settings for a state-of-the-art deformable registration using anatomical landmarks and regions of interest (ROIs). Additionally, a multiparametric registration approach was evaluated and compared with single-parametric approaches. For all approaches we compared rigid and deformable registration results.

5.2 Methods and materials

5.2.1 Data

Twelve axial CT (11 contrast-enhanced), and T1-weighted (T1w) and T2-weighted (T2w) MR images of patients with a H&N tumor were acquired for RTP and HTP. For both modalities, the patient was put in treatment position using a patient-specific immobilization mask and headrest during the acquisition. CT images were acquired using a Siemens scanner (Somatom Sensation Open, Siemens AG, Erlangen, Germany), with a voxel size of $0.98 \times 0.98 \times 2.50$ mm^3, acquisition matrix of 512 x 512, tube voltage of 120 kV, and current-time product of 150 mAs.

Table 5.1. Sequence specifics. *Abbreviations* Seq. = sequence; TE = Echo Time; TR = Repetition Time; Matrix = acquisition matrix; Angle = Flip Angle

Seq.	TE[ms]	TR[ms]	Matrix	Angle[°]
T1w	10.2	465	384×224	90
T2w	106.6	7060	384×224	90

The MR images were acquired on a 1.5 T GE scanner (Optima MR450w, GE Healthcare, Waukesha, WI) using 6 channel surface coils. Eleven out of twelve MR images had a voxel size of $0.68 \times 0.68 \times 3.00$ mm^3 and the twelfth image had a voxel size of $0.49 \times 0.49 \times 3.00$ mm^3. MR sequence parameters are provided in the Table 5.1. The out-of-plane field of view (FOV) was 20 cm centered at the tumor location, yielding a slightly different FOV for each patient. T1w and T2w images were both corrected for intensity inhomogeneity: T1w images were corrected using the N3 method [106], and T2w images using the built in algorithm (Surface Coil Intensity Correction) of the MR scanner.

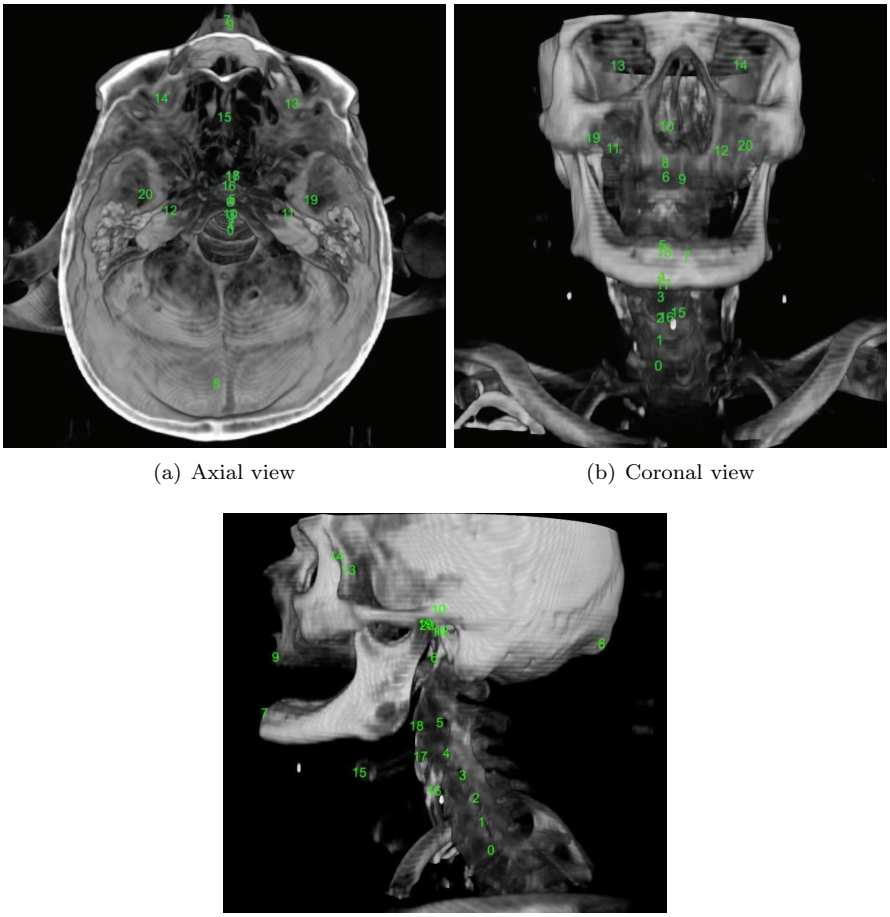
Twenty-one corresponding anatomical landmarks were placed by an expert (F.A.) on CT, T1w, and T2w images. Figure 5.1 shows the the description and anatomical location of the landmarks. Due to differences between the images, not all landmarks could be placed in all patients. The landmarks were annotated twice by the same rater in order to estimate the intraobserver variability. For eight patients, regions of interest (ROIs) were annotated by a trained medical radiation technologist in CT, T1w, and T2w images. These regions are: cerebrum, cerebellum, brainstem, spinal cord, eyeballs, lenses, parotid and submandibular glands For one patient the submandibular glands could not be identified.

5.2.2 Registration method

The MR (T1w and T2w) sequences were aligned with the CT image, and the accuracy of the alignment was quantitatively evaluated. Registration was applied in 3 steps. First an initial alignment was determined by automatic detection of a landmark in the neck (Section 2.2.2.2). Next a rigid registration followed by a deformable registration was performed.

For rigid and deformable registrations we used the opensource software Elastix [66]. For deformable registration we chose a state-of-the-art algorithm [100] optimizing only the most important parameters to minimize the dependency of our results on parameter selection. This algorithm was chosen because of its good performance in the H&N region [76] and in other anatomic sites [64,67,110] and its free availability in the Elastix opensource software [66]. This method showed top-ranking performance in a comparative study for multi-modality interpatient brain images [64] and for lungs intrapatient intramodality CT images registration [110].

The main difference between the original B-Spline registration method [100] and its Elastix implementation is the use of an adaptive stochastic gradient descent optimization [65]. In this approach a set of randomly selected samples is used at each



(a) Axial view (b) Coronal view

(c) Sagittal view

Figure 5.1. Distribution of the anatomical landmarks: 0-5 = Mid point of the inferior endplate along the posterior vertebral margin, from C7 to C2; 6 = Tip of the odontoid process; 7 = Mental protuberance; 8 = External occipital protuberance; 9 = Spina nasalis anterior; 10 = Intersection of the vertebral arteries; 11-12 = The right and left internal carotid artery at the passage through the carotid canal; 13-14 = Emergence of the right and left optical nerve; 15 = Midpoint of the anterior surface of the hyoid bone; 16-18 = Mid point of the inferior end-plate along the anterior vertebral margin, from C4 to C2; 19-20 = Midpoint of the medial surface of right and left mandibular condyle.

iteration to calculate the similarity measure and to estimate the derivative and the step size for the gradient descent optimization, making the optimization procedure adaptive. The stopping criterion for this algorithm is the maximum number of iterations. This and the number of random samples were chosen to be sufficiently high to guarantee registration convergence. Both rigid and deformable registration steps

were done in a multiresolution framework using mutual information (MI) as a similarity measure [78,111,115]. The deformation field in the deformable registration step was modeled with cubic B-Spline polynomials [120]. The number of B-Spline control points used during the registration determines the amount of deformation allowed during the registration. A uniform control-point grid was used for each resolution level: given the finest resolution grid spacing, for each coarser resolution the grid spacing was 2 times larger. A Gaussian pyramid was used taking into account the anisotropy of the images. The smoothing along the z-direction was set two times lower than the smoothing in the other directions, i.e. the smoothing sigmas were chosen as $\sigma_{x,y} = 4, 2, 1$ voxels and $\sigma_z = 2, 1, 1$ voxels. The mutual information similarity metric was calculated using 32 bins. The number of bins was selected following the elastix manual and a previous study [67]. The stopping criterion for the optimization algorithm is the maximum number of iterations which was set to 500 for rigid registration and to 1000 for deformable registration. We used 4000 voxels, randomly sampled for each iteration of the optimization. During registration, fixed and moving image masks were used to exclude background voxels from the registration. These masks were calculated automatically from MR and CT images using intensity thresholding, connected component analysis, and binary morphology operations. The parameters not listed above were set as default by Elastix.

Next to CT-T1w and CT-T2w registration, we investigated the use of both sequences simultaneously to obtain the alignment with CT. This multiparametric approach used a combined similarity measure defined as the weighted average of the single-parametric similarity measure, that is,

$$\mathrm{NMI}(\mathrm{CT},\ \mathrm{T1w},\ \mathrm{T2w}) = \omega \cdot \mathrm{NMI}(\mathrm{CT},\mathrm{T2w}) + (1-\omega) \cdot \mathrm{NMI}(\mathrm{CT},\mathrm{T1w}) \qquad (5.1)$$

where $\omega \in [0,1]$ is a parameter that weighs the contributions from T2w and T1w.

5.2.3 Parameter optimization

Leave-one-out parameter optimization was performed for both rigid and deformable registration: the optimal parameters for each registration approach were found for each patient independently using images of all the other patients. Therefore, the optimal parameters chosen were independent of the particular set of images on which the registration was evaluated, and the performance was not overestimated with respect to our dataset.

The parameter optimization consisted of an exhaustive search over a predefined set of parameters. To account for possible interdependencies between parameters, we evaluated all possible combinations. For the rigid registration the parameters to be optimized were the number of resolution levels (2 or 3) and the multiparametric weight u. For the deformable registration the parameters were the number of resolution levels (2 or 3), the B-Spline controlpoint grid spacing at finest resolution (from 15 to 115 mm in steps of 5 mm), and the multiparametric weight u. For both rigid and deformable cases we evaluated 7 values for $\omega = \{0, 0.17, 0.33, 0.50, 0.67, 0.83, 1\}$. For each patient, the optimal parameters were chosen to give the minimum average Euclidean landmark distance over all the other patients. The optimization criterion for rigid registration was the average (over all other patients) of the median (over all landmarks in a patient)

Euclidean distance between original and transformed landmarks. The median was used because it is less influenced by possible outliers. For an optimal alignment, we thus allow for misalignment of a small set of point because of non-rigid anatomical differences. For the deformable registration we used the average Euclidean distance over all landmarks in a patient, instead of the median, to estimate the alignment error for each patient. In contrast to the rigid case we aim to register all points well, and thus we also take potential outliers into account.

5.2.4 Registration evaluation

The MR to CT images were registered with the single and multiparametric registration, using parameter settings as obtained from the parameter optimization. The registration convergence was assessed for all the approaches. As explained in Section 5.2.2 the similarity metric which is optimized during the registration was an approximation of the similarity metric over the complete image, computed using a limited number of randomly sampled voxels [65]. This approximated measure cannot be used as an absolute similarity measure between the images and so cannot be used to check the registration convergence. Thus, in order to analyze the registration convergence we calculated the absolute similarity metric, using the complete image, for each iteration of the registration.

The accuracy of registration was measured using (1) the Euclidean distance between original and registered landmarks; and (2) the correspondence between original and registered ROIs. In addition, we calculated the determinant of the spatial Jacobian $det|\mathcal{J}|$ of the deformation resulting from deformable registration. The value of $det|\mathcal{J}|$ identifies the amount of local compression ($det|\mathcal{J}| < 1$) or expansion ($det|\mathcal{J}| > 1$) after registration. Realistic deformations are characterized by a conservation of the volume, so by values of $det|\mathcal{J}|$ close to 1.

The Euclidean distance between original (MR) and transformed (CT) landmarks was calculated for each scheme. All schemes were compared with each other. The landmarks were grouped with respect to anatomy in 4 groups: vertebrae, bones, arteries, and optical nerves (Figure 5.2.1). Next, using the results of the registration, we deformed the MR ROIs to CT space. This was done for each of the 8 patients for whom these regions were available. To quantify the correspondence between ROIs, and thus the registration accuracy, we used Dice similarity (DSC) [27], 3-dimensional Mean Surface Distance (MSD) [29],, and 3-dimensional Hausdorff surface distance (HSD) [56] between the regions in CT and the regions deformed from MR to CT. For the definition of these measures see Section 2.3.3.

5.2.5 Statistical analysis

A two-sided Wilcoxon signed rank test ($p < 0.01$ or $p < 0.05$) was used to test significant differences between evaluation measures.

5.3 Results

5.3.1 Parameter optimization

In the parameter optimization experiments, we observed a low variation of registration accuracy when varying the multiparametric weight ω (0.1 mm) and the B-Spline control-point grid spacing (0.1 mm around the optimal value). Figure 5.2 (A) shows the variation of rigid and deformable registration accuracy with respect to ω, averaged over all the patients. Registration accuracy is measured using the metric used for parameter optimization, as described above. For all the approaches, Figure 5.2 (B) shows the accuracy variation with respect the control-point grid spacing at the finest resolution. The other registration parameters are fixed as in the optimal case. The optimal values of control-point grid spacing were 60 mm for CT-T1w, and 85 mm for CT-T2w and the multiparametric approach.

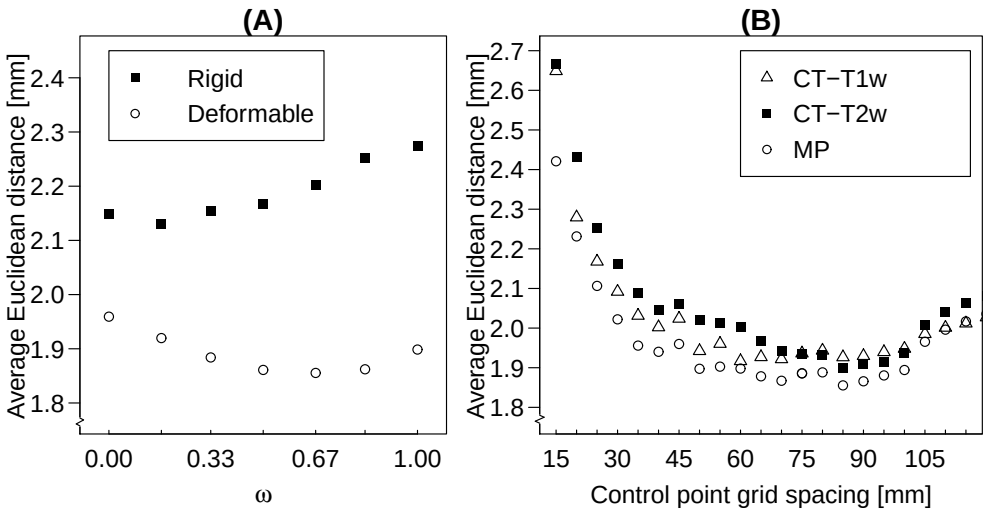


Figure 5.2. (A) Registration accuracy averaged over all patients, as function of multiparametric weight. (B) Registration accuracy averaged over all patients, as function of control point grid spacing. *Abbreviations* MP = multiparametric registration.

Table 5.2 shows the optimal settings per patient, for each registration scheme. The parameters that offered the lower Euclidean distance over the entire cohort are printed in bold. These parameters can be used for a new scan. Table 5.2 shows a minor variation of optimal parameter over patients.

5.3.2 Registration convergence

For each patient and each scheme the registration was executed calculating the MI (as defined in [115]) using all samples of the patient body. For the multiparametric approach, the variation of MI over all registration iterations is shown in Figure 5.3,

Table 5.2. Optimal parameter settings resulting from the leave-one-out optimization. The optimal settings considering the entire cohort are printed in bold. *Abbreviations* Levels = number of resolution levels; Grid = B-Spline control point grid spacing at finest resolution

Registration scheme	Patient ID	Levels	Grid[mm]	ω
Rigid registration				
T1w	**1-12**	**3**	-	-
T2w	**1-12**	**3**	-	-
T1w+T2w	**1-2,4-12**	**2**	-	**0.17**
T1w+T2w	3	3	-	0.5
Deformable registration				
T1w	**1-3,5,10**	**2**	**60**	-
T1w	6	3	85	-
T1w	$4, 7-9, 11-12$	3	70	-
T2w	**1-12**	**2**	**85**	-
T1w+T2w	**1-8,10,12**	**2**	**85**	**0.67**
T1w+T2w	9	2	70	0.67
T1w+T2w	11	3	85	0.67

for the rigid registration, and in Figure 5.4, for the deformable registration. A similar behavior was observed for the CT-T1w and CT-T2w registration scheme. From the figures it is clear that the registration converged for all patients. Since for the rigid registration the convergence was reached before the 100th iteration the maximum number of iteration could be lowered.

5.3.3 Landmark-based registration evaluation

Registration accuracy, with optimal parameters, was evaluated using the Euclidean distance between MR landmarks (placed in T1w or T2w) and CT landmarks. The intraobserver error in landmarks placement was compared with the registration error. Figure 5.5 A shows the distance distribution of all the landmarks for the various registration approaches. In all cases, deformable registration resulted in Euclidean distances statistically significantly lower than when using rigid registration. No statistical difference was found between the registration error obtained using multiparameteric and single-parametric approaches. For all approaches, the distribution of the intraobserver error and of the initial alignment of T1w and T2 was statistically significantly lower ($p < 0.01$) than the registration error.

Table 5.3 shows the results for all the registration schemes. A median landmark alignment of approximatively 1.7 mm was obtained. The median difference between the landmark distance of rigid and deformable registration (not shown in the figures) was 0.39 mm for CT-T1w registration, 0.43 mm for CT-T2w, and 0.45 mm for the multiparametric case. The statistics of $det|\mathcal{J}|$ were computed excluding the background voxels. These values show an overall volume conservation for all registration

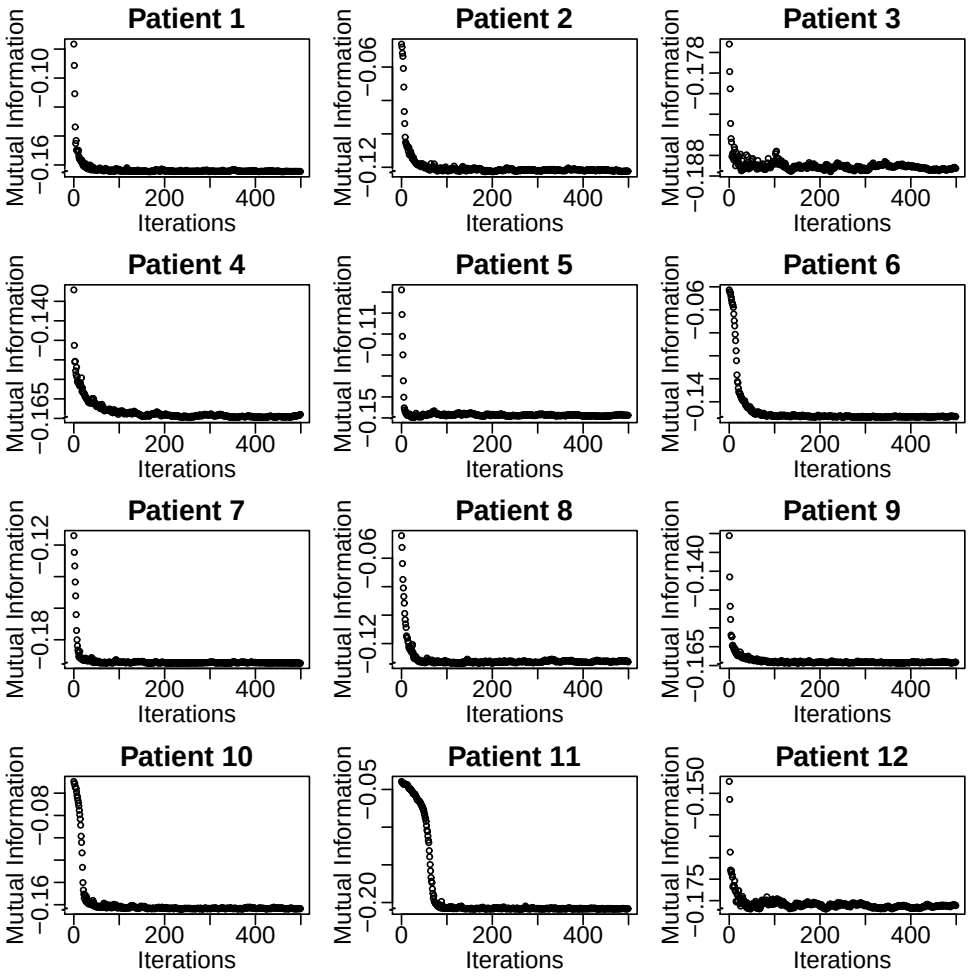


Figure 5.3. Mutual information variation over all iterations for rigid multiparametric registration. The result is shown for the second (finest) resolution level which is more effective during registration in capturing the image characteristics during rigid registration.

approaches. Figure 5.5 shows the distance distribution of multiparametric registration for each landmark group. For vertebrae, bones, and optical nerves, the deformable registration resulted in distribution of Euclidean distances statistically significantly lower than obtained with rigid registration.

The distribution of Euclidean distances over patients and anatomical landmarks is shown in Tables 5.4 and 5.5, respectively for multiparametric rigid and deformable registration. The legend and spatial location of the landmarks is reported in 5.1. Some landmarks are missing because they are not included in the FOV (vertebra landmarks 0 and 1, optical nerve landmarks 10 and 11) or because the lack of contrast in CT

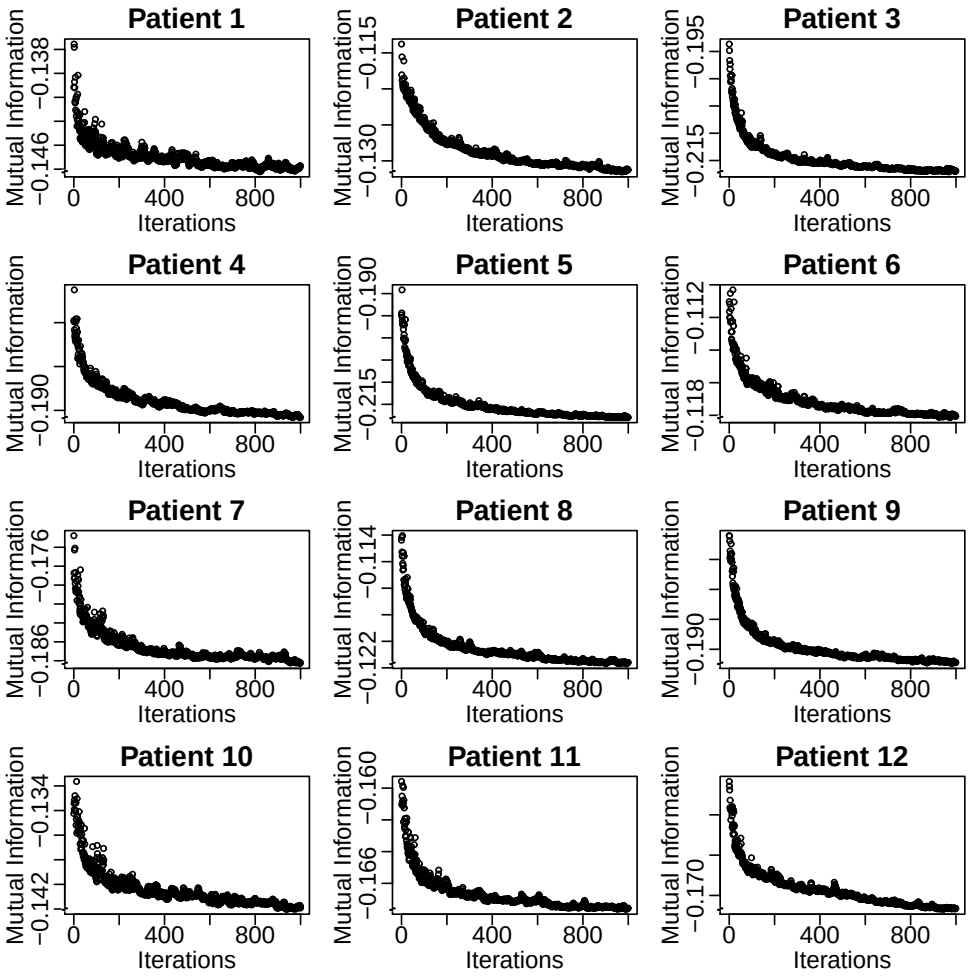


Figure 5.4. Mutual information variation over all iterations for deformable multi-parametric registration. The result is shown for the finest resolution level.

(artery landmarks 11,12 and 13).

Figure 5.5 and Tables 5.4 and 5.5 show that outliers of considerable magnitude were still present after deformable registration. These outliers did not correspond to a particular landmark location or a particular subject. We found similar average values and standard deviations over different landmarks that indicates a spatially uniform error. The only exception was the hyoid bone, for which outliers were found for 6 patients consistently in both rigid and deformable registration. The hyoid was difficult to position for some patients and its position can change considerably between different scans. Thus, for some patients an higher error was obtained.

With deformable registration vertebrae could be aligned better than when using

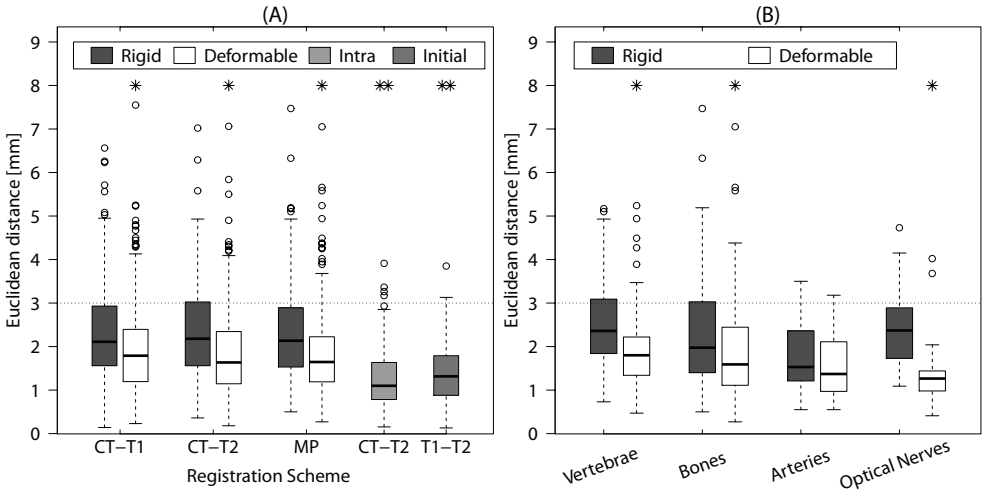


Figure 5.5. (A) Distribution of the distance for all the landmarks after registration. (B) Distribution of the distance for landmark groups for the multiparametric registration. *Abbreviations* Intra = variability in landmarks placement; Initial = alignment without registration; MP = multiparametric. Asterisks ($\star$) indicate which distribution is statistically significantly lower ($p < 0.01$) for each group. Double asterisks ($\star\star$) indicate that the distribution is statistically significantly lower ($p < 0.01$) than all the rigid and deformable distance distributions. The dotted horizontal line indicates the largest voxel size. The boxes of the boxplots report first and third quartiles and median values. Whiskers report confidence intervals, which are defined as the most extreme data points (−1.5 times the interquartile range around the box) not considered outliers (open circles ∘)).

Table 5.3. Results of the landmarks-based evaluation. *Abbreviations* AV = Average value; SD = standard deviation; MD = Median value; IQR = inter quartile range; Min = minimum value; Max = maximum value.

	Landmark distance				Spatial Jacobian			
Registration scheme	AV	SD	MD	IQR	AV	SD	Min	Max
Deformable registration								
T1w	1.96	1.09	1.79	1.20-2.39	1.01	0.04	0.87	1.14
T2w	1.98	1.07	1.64	1.15-2.33	1.02	0.03	0.91	1.10
T1w+T2w	1.88	1.05	1.65	1.19-2.25	1.01	0.03	0.92	1.10
Rigid registration								
T1w	2.30	1.08	2.11	1.56-2.93				
T2w	2.37	1.11	2.18	1.56-3.02				
T1w+T2w	2.34	1.10	2.14	1.53-2.89				

Table 5.4. Error per landmark per patient for **multiparametric rigid registration**. Outliers are printed in bold. *Abbreviations* AV = Average value; SD = standard deviation.

Landmark	Patient ID													
	1	2	3	4	5	6	7	8	9	10	11	12	AV	SD
0	-	-	-	-	-	1.51	2.84	-	-	-	1.88	-	2.08	0.69
1	-	2.17	**3.78**	2.88	-	1.22	1.92	-	-	-	**3.98**	-	2.66	1.09
2	2.19	2.39	2.06	0.73	5.10	1.51	2.01	1.36	2.81	**4.16**	2.85	**4.43**	2.63	1.32
3	1.59	3.31	2.26	2.23	**3.82**	2.81	1.79	2.28	**5.17**	2.61	1.97	1.84	2.64	1.03
4	1.51	**4.03**	1.50	2.05	2.74	3.4	**4.26**	3.25	2.42	3.09	2.42	2.36	2.75	0.89
5	1.32	**3.22**	1.61	3.44	**3.85**	2.62	1.48	2.71	1.98	2.35	2.38	1.45	2.37	0.84
6	2.00	2.24	2.48	2.39	2.3	**4.93**	1.6	3.23	2.43	1.93	2.05	2.72	2.53	0.86
7	1.96	3.00	2.76	1.69	2.31	2.28	1.95	1.82	**4.04**	2.15	0.50	3.03	2.29	0.87
8	2.12	1.09	1.68	**4.04**	2.58	3.25	2.13	**4.15**	1.49	3.67	1.58	3.03	2.57	1.05
9	0.55	2.45	1.39	**6.33**	0.83	2.25	0.84	1.99	2.88	2.28	1.77	1.9	2.12	1.51
10	1.76	0.85	0.69	1.83	2.68	-	0.87	2.73	1.33	0.85	1.04	2.2	1.53	0.76
11	1.53	1.53	2.57	1.21	0.55	-	3.17	2.36	1.42	1.43	2.48	1.53	1.8	0.75
12	0.91	1.89	2.14	1.44	2.45	-	3.5	3.21	1.06	1.79	1.34	2.02	1.98	0.83
13	2.51	-	3.06	2.28	**4.15**	-	2.46	1.11	2.89	1.67	-	2.59	2.52	0.86
14	1.30	-	1.92	2.00	**4.73**	-	1.75	1.73	3.30	1.09	-	2.63	2.27	1.14
15	**3.92**	**5.19**	1.23	**4.83**	1.72	**7.47**	3.33	**4.38**	2.68	**4.41**	1.81	2.71	3.64	1.77
16	1.47	3.45	3.52	2.45	2.76	**4.07**	1.93	3.15	1.53	3.10	2.05	1.19	2.56	0.93
17	1.06	**4.52**	1.98	3.67	1.92	2.20	2.75	3.14	1.74	2.67	1.66	1.21	2.38	1.02
18	0.96	2.89	1.83	2.97	1.24	2.00	3.06	**3.91**	1.51	1.95	2.59	0.98	2.16	0.93
19	1.39	1.72	3.09	1.41	1.14	0.85	**3.78**	2.04	1.76	1.80	1.16	1.47	1.80	0.84
20	0.71	1.20	1.32	**4.00**	1.15	1.10	0.65	1.83	1.08	2.03	1.55	3.17	1.65	1.00
AV	1.62	2.62	2.14	2.69	2.53	2.72	2.29	2.65	2.29	2.37	1.95	2.23		
SD	0.76	1.2	0.8	1.36	1.31	1.68	0.99	0.93	1.07	0.97	0.76	0.86		

Table 5.5. Error per landmark per patient for **multiparametric deformable registration**. Outliers are printed in bold. *Abbreviations* AV = Average value; SD = standard deviation.

Landmark	Patient ID													
	1	2	3	4	5	6	7	8	9	10	11	12	AV	SD
0	-	-	-	-	-	1.61	1.23	-	-	-	3.43	-	2.09	1.18
1	-	2.33	3.29	1.72	-	0.81	0.68	-	-	-	3.18	-	2.00	1.13
2	1.99	1.59	1.32	0.58	**3.89**	2.02	1.60	1.78	1.89	1.56	1.52	2.12	1.82	0.77
3	3.15	1.77	1.05	1.97	2.16	2.20	1.10	1.26	**5.24**	1.40	1.78	1.70	2.07	1.15
4	1.36	1.86	1.10	1.65	1.77	1.34	3.32	1.25	2.54	2.59	1.39	0.78	1.75	0.73
5	2.09	2.91	0.47	2.09	2.37	0.64	1.81	1.17	1.80	1.75	1.19	0.56	1.57	0.77
6	1.32	2.56	1.92	1.89	2.22	**4.49**	1.30	1.18	2.02	1.59	2.07	1.82	2.03	0.87
7	1.27	2.50	1.56	1.33	1.05	1.11	1.53	1.43	0.77	1.42	1.06	2.02	1.42	0.46
8	3.18	2.77	1.89	3.29	3.56	1.19	2.90	1.64	3.17	3.35	0.83	1.80	2.46	0.94
9	1.30	1.57	1.90	**7.05**	0.83	1.32	0.47	1.27	1.48	1.69	0.27	2.26	1.78	1.75
10	0.97	0.84	0.79	1.58	1.53	-	0.64	2.84	0.74	1.18	1.09	1.22	1.22	0.62
11	1.94	1.87	1.10	1.34	1.41	-	2.73	0.78	0.74	1.71	2.24	2.11	1.63	0.62
12	0.55	1.32	1.37	0.81	3.06	-	3.18	2.45	1.24	2.29	2.17	1.56	1.82	0.87
13	2.04	-	1.13	1.88	3.68	-	1.07	0.92	1.16	1.35	-	1.32	1.62	0.86
14	0.93	-	0.69	1.31	**4.02**	-	0.41	0.98	1.22	1.44	-	1.37	1.37	1.05
15	**4.38**	**5.58**	2.23	**3.95**	1.11	**5.66**	1.65	**4.25**	1.71	**4.36**	0.67	2.78	3.19	1.72
16	2.41	2.16	2.85	2.85	**4.27**	**4.94**	1.70	1.68	1.54	2.17	2.15	1.36	2.51	1.10
17	1.16	2.94	1.75	3.41	1.99	3.22	2.07	1.99	1.83	2.80	0.98	1.16	2.11	0.82
18	1.58	1.80	1.37	1.91	0.89	3.13	3.47	3.35	0.69	1.91	1.33	1.27	1.89	0.94
19	1.61	0.53	1.94	1.29	1.45	1.07	3.21	2.38	1.62	1.95	0.90	0.68	1.55	0.76
20	0.94	0.62	1.88	2.80	0.60	1.85	0.88	1.22	1.07	1.56	1.34	2.39	1.43	0.69
AV	1.80	2.08	1.58	2.24	2.20	2.29	1.76	1.78	1.71	2.00	1.56	1.59		
SD	0.95	1.14	0.70	1.43	1.20	1.56	1.00	0.91	1.07	0.80	0.82	0.60		

rigid registration alone (median difference, 0.45 mm). Because the vertebrae are close to the image center we assume that the geometric distortion is minimal in this region, so most of the non-rigid deformation can be attributed to differences in positioning. On the other hand, the alignment of other bones after rigid registration provides an estimation of the geometric distortion because they are more distant from the center. The difference between rigid and deformable distances for the bones (median, 0.4 mm) is a measure for the amount of recovered geometric distortion.

The registration accuracy for each patient is shown in Figure 5.6. For 9 of 12 patients the deformable registration gave a statistically significantly lower Euclidean distance than rigid registration. For the other patients rigid and deformable registration showed similar results.

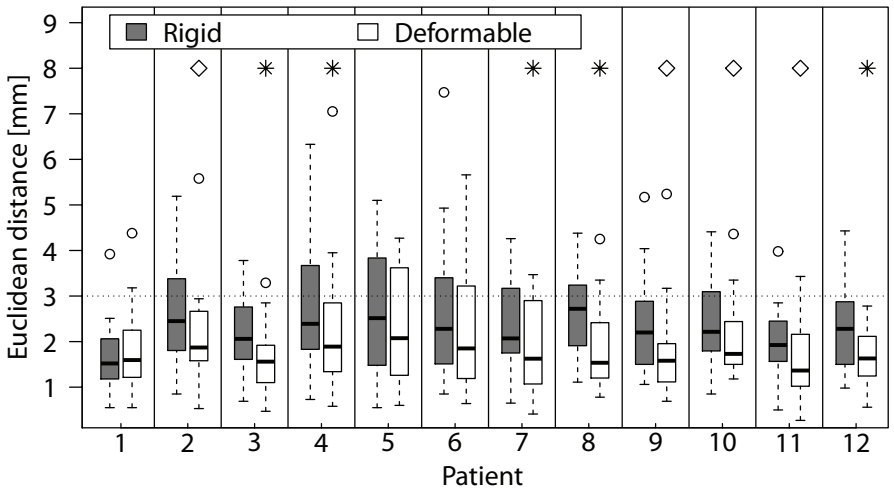


Figure 5.6. Distribution of the distance per patient for all the landmarks after registration. Asterisks ($\star$) and open diamonds ($\diamond$) indicate which distribution is statistically significantly lower ($p < 0.01$ and $p < 0.05$, respectively) for each group. The dotted horizontal line indicates the largest voxel size. Boxplots are defined as in Figure 5.5.

5.3.4 ROIs-based registration evaluation

We calculated the initial relative difference in volume between CT and MR ROIs to evaluate the correspondence betweeen MR and CT delineations. This difference was approximately 5% for cerebrum, brainstem, and parotids; 25% for the spinal cord; -25% for the eyeballs; -30% for the lenses; and 15% for the submandibular glands (the minus indicates that the MR is larger than CT volume). These values indicate that the borders of some tissues had a different appearance in MR and CT.

Because all approaches gave similar performance we only list the results of the multiparametric registration. For rigid and deformable registration, the distribution

of DSC, MSD, and HSD over patients, per region, is reported in Figure 5.7. Considering all the ROIs, the median MSD was approximately 1 mm for all the ROIs excluding cerebellum and the parotids. For 5 of 12 ROIs the deformable registration provided a statistically significantly better accuracy than rigid registration. For the other ROIs rigid and deformable registration provided a similar accuracy.

Figure 5.8 shows examples of ROIs contours drawn originally in T2w and registered using rigid and deformable registration together with CT contours. The figure shows how deformable registration provided a better alignment of contours. The brainstem and cerebellum deformably registered contours appear more accurate than the ground truth.

5.4 Discussion

This study investigated the feasibility of deformable registration of CT and MR images in the H&N region for RTP and HTP. Our evaluation scheme was optimized with respect to the applications (RTP and HTP), landmark and ROI visibility on both images, and volume coverage. Deformable registration resulted in a better accuracy than rigid registration, and the multiparametric and single-parametric approaches had similar performances. Deformable registration provided an accuracy on average higher than reported in previous studies.

Deformable registration accuracy was stable around the optimum of the B-Spline control-point grid spacing, which is the main parameter for the algorithm at hand. This suggests that the algorithm will perform similarly if applied to a new patients image.

Both landmark-based and ROI-based evaluation showed that deformable registration was more accurate than rigid registration. The values of spatial Jacobian determinant indicate a smooth deformation between MR and CT images. These values ensure that unrealistic deformations were not generated by deformable registration. The landmarks and the ROIs alignment together with the smoothness of the deformation demonstrate that the deformable registration provided a better spatial alignment than rigid registration in the entire volume.

For all registration schemes, the registration error was statistically significantly higher than the error in landmark placement and than the initial alignment of T1w and T2w images. These 2 factors therefore had only a limited influence on the evaluation.

The accuracy we obtained for deformable registration is similar to that obtained by Webster et al. [133] and on average better than that of Leibfarth et al. [76] and Du Bois d'Aische et al. [28]. Webster et al. [133] found a distance between CT and MR landmarks of 2.2 – 0.9 mm (mean value – standard deviation) on a set of 6 patients. In their evaluation they assumed the deformation to be rigid and evaluated the accuracy on a small set of landmarks placed in a region with minimal expected geometric distortion. Leibfarth et al. [76] reported an error in alignment of anatomic landmarks of 2.96 – 1.02 mm for 8 patients. Du Bois d'Aische et al. [28] reported a mean error in the range 1.91 – 3.31 mm and variances in the range 0.81 – 1.05 mm for a limited set of 4 patients and for an unspecified set of landmarks. The reason we obtained a lower registration error (1.88 – 1.05 mm) than these approaches may be

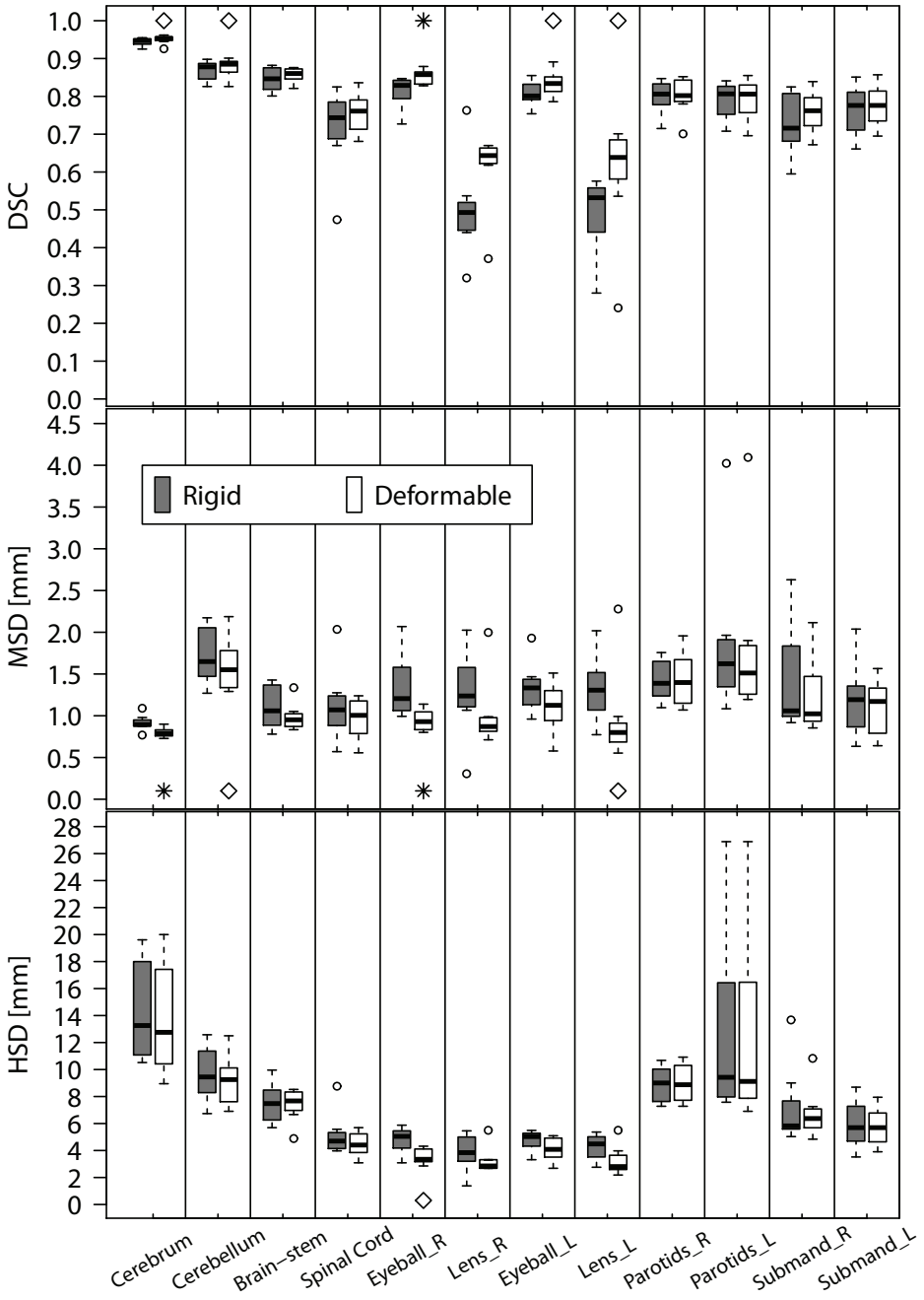


Figure 5.7. Dice similarity (DSC), 3-dimensional mean surface distance (MSD), and 3-dimensional Hausdorff surface distance (HSD) measurements between registered and original ROIs (abscissa). L and R are abbreviations for left and right. Open diamonds ($\diamond$) indicate a statistically significant improvement with $p < 0.05$. Asterisks ($\star$) indicates a statistically significant improvement with $p < 0.05$. Boxplots are defined as in Figure 5.5.

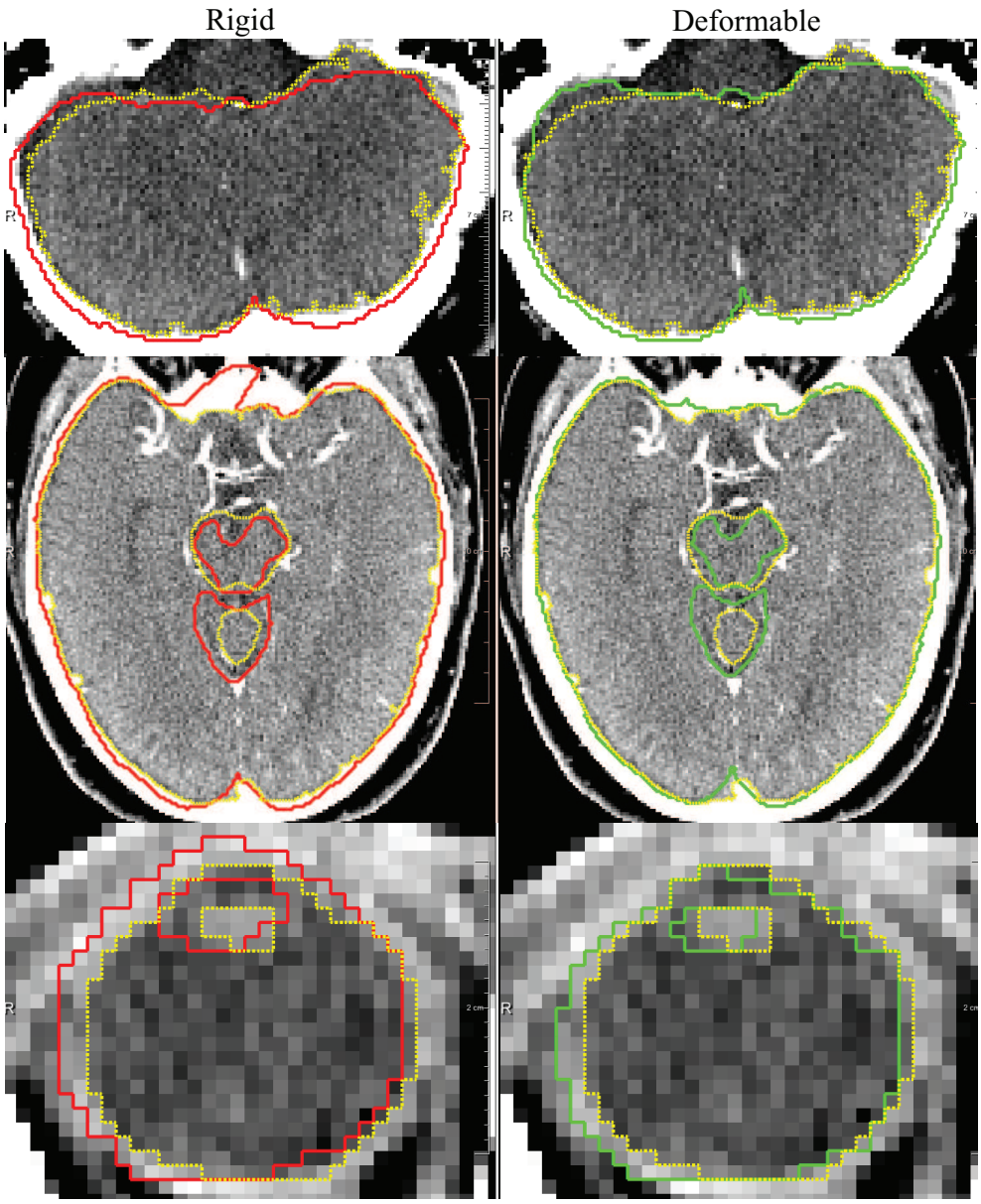


Figure 5.8. Computed tomography axial slices with overlaid region of interest contours. Each image pair shows a comparison between T2w contour registered using rigid (solid red line) or deformable (solid green line) multiparametric registration, and computed tomography contour (dashed yellow). From top to bottom the images show cerebellum; cerebrum, cerebellum, and brainstem; and eyeballs and lenses.

because the patients in our study were scanned with immobilization equipment. Even though deformable registration improves the image alignment, it might not provide sufficient accuracy to avoid patient fixation (see Chapter 6).

The deformable registration significantly improved the accuracy for 5 of 12 ROIs and did not decrease the overall registration accuracy for any of them. This confirms the statement that deformable registration did not introduce unrealistic deformations. Because of better soft-issue contrast, tissue delineation is generally more accurate on MR than on CT, as depicted in Figure 5.8. This fact caused the difference between initial volume of the ROIs for some of the tissues. In addition, intermodality differences in delineation accuracy caused the high HSD for the cerebrum, cerebellum, and glands, both for rigid and deformable registration. Even though deformable registration may improve the image alignment, such differences may mean that this improvement did not influence the evaluation measures for spinal cord, brainstem, and glands. Despite these differences a good alignment of ROIs was found for all tissues, except the lenses.

The value of DSC after deformable registration of the eye lenses was relatively low. One cause is the small volume of the lenses, which decreases the DSC, and another is that their position depends on the direction of the glance, which can be different between scans. We decided to include these tissues in the evaluation because they are important for HTP.

The effect of delineation accuracy on HTP and RTP has been previously investigated, respectively in references [127] and in [88, 129]. The difference in accuracy between rigid and deformable registration may be relevant in terms of local dose distribution both for HTP and RTP. In the case of HTP relevant differences in local dose (up to 12reported for small variation of tissue delineations (i.e. DSC > 0.7 and MHSD < 1 mm) [127]. In the case of RTP variations in target tissue delineation had a significant impact on dosimetry [129]. A small variation in target volume delineation (i.e. DSC > 0.8 and MHSD < 1 mm) yielded significant dose differences (up to 11 Gy). In the organ at risk (OAR) the improvement in accuracy given by deformable registration may be negligible in terms of radiation dose [88]. However, the OAR delineations are also used to calculate the dosimetric statistics per OAR which, therefore, depend on their quality.

The limited use of deformable registration for treatment planning in current practice can be ascribed to the fact that it is not robust to parameter variation and to the presence of unrealistic deformation. Our work is a first step toward solving this issue. A robust parameter optimization scheme made it possible to obtain a low registration error and a smooth deformation. If images are acquired with an acquisition protocol similar to ours, the proposed method can be used reliably in clinical practice. Therefore, these results hold only when immobilization equipment is being used during the CT and MR scans. Because the MI metric is robust to intensity variations, we do not expect the final accuracy of registration to be affected by the use of CT and MR images from different scanners with the same MR field strength. However, the geometric distortion may be more prominent for higher MR field strength than for 1.5 T, and thus the selected parameters may be suboptimal.

In conclusion, we demonstrated the feasibility of deformable registration for RTP and HTP in the H&N. Deformable registration provided overall a significantly higher

accuracy than rigid registration. Our study paves the way for the application of deformable registration in routine clinical practice.

Chapter Six

MRI Integration into Treatment Planning of Head and Neck Tumors: Can Patient Immobilization be Avoided?

There are, in effect, two things, to know and to believe one knows; to know is science; to believe one knows is ignorance

— Hippocrates (460-370 B.C.)

Abstract — To obtain an accurate rigid alignment between magnetic resonance (MR) and computed tomography (CT) images in the head and neck, patient immobilization equipment is applied during image acquisition, which takes time and is uncomfortable. Therefore, we investigated whether immobilization equipment is still needed when applying deformable registration to align MR and CT images.
CT and MR images of 22 patients were aligned using state-of-the-art, rigid and deformable registration algorithms. For 10 patients MR images were acquired in two settings: with no immobilization mask and with a small immobilization mask. For 12 patients the images were acquired using the original radiotherapy immobilization equipment (RT mask). The registration accuracy was evaluated using 21 landmark points, located on CT and MR images. Leave-one-out experiments were performed for evaluation and optimization of parameters of the rigid and deformable registrations.
After registration, the alignment between CT and MR improved significantly when patient immobilization equipment was used: the best alignment was obtained when patients were immobilized with the RT mask. Deformable registration consistently provided a significantly better alignment than rigid registration.
When accurate alignment is required, the use of immobilization equipment is recommended since prevention of movement by use of immobilization equipment provides better alignment than postprocessing using state-of-the-art deformable registration.

Based upon: **V Fortunati**, R F Verhaart, G M Verduijn, A van der Lugt, F Angeloni, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "MRI integration into treatment planning of head and neck tumors: can patient immobilization be avoided?", *Radiother Oncol*, in press.

6.1 Introduction

Radiotherapy and hyperthermia treatment planning (RTP and HTP) in the head and neck (H&N), require 3D patient models, based on tissue delineation. The most common imaging modality used for this application is computed tomography (CT) [31, 91, 127] (Chapter 2) because of its geometrical accuracy and, for RTP, the direct availability of electron density information. However, the superior soft-tissue contrast of magnetic resonance (MR) images with respect to CT has been shown to improve the accuracy with which target and normal tissues can be delineated [2, 45, 96, 117]. To exploit the advantages of both modalities, the images must be aligned. Since it is geometrically more accurate, the CT image is commonly used as reference image to which the MR image is aligned.

In clinical practice, the deformation between CT and MR images is assumed to be rigid since patient immobilization equipment is used during image acquisition to fix the patient in the treatment position [17, 96, 125, 133]. It has been shown that patient fixation improves the accuracy of rigid registration when patient images are acquired in treatment position for both CT and MR [50]. However, the use of such equipment has disadvantages. First, it is associated with patient discomfort as patients have to wear the mask for the entire MR scanning period (up to 1 hour). Second, application-specific flexible coils are required to create a setup that combines immobilization and imaging coils [49]. These flexible coils have a non-uniform signal and reduced signal-to-noise ratio which limit the image quality.

Additionally, patient immobilization is hampered when positron emission tomography (PET) scanning is combined with MR, because the immobilization device causes attenuation of the PET signal which cannot be recovered with signals from conventional MR sequences [79].

Deformable registration might eliminate the need for immobilization equipment. The use of deformable intra-patient registration in multi-modal H&N imaging has recently been investigated by others [28, 76] and by our group (see Chapter 2). In the first two studies, immobilization equipment was not used and, when patients were scanned in free position, in both studies deformable registration was more accurate than rigid registration. The third study showed that when immobilization equipment was used during the MR acquisition, deformable registration provides better image alignment than rigid registration.

This chapter investigates the influence of using patient immobilization equipment during the MR session on the accuracy of the multi-modality registration in the H&N. The main aim is to investigate whether we can avoid the use of immobilization equipment by aligning CT and MR images using deformable registration. Therefore, we optimized and evaluated the optimal settings of a state-of-the-art deformable registration algorithm [66, 100] and compared registration accuracy in various scenarios, with and without immobilization equipment.

6.2 Methods and materials

Two sets of patient data were acquired: set A and set B. Set A consists of CT and axial T2w images of 10 patients with a H&N tumor. The T2w images were acquired using head coils. During the MR acquisition session, these patients were scanned twice in two different positions: 1) without immobilization equipment (set A1), and 2) with a small immobilization mask (set A2), created from the original radiotherapy immobilization mask in order to fit inside standard head coils. Head coils were used because, at the start of our study, flex coils were not available. The small mask was used to fix the patients in the treatment position. Still, patient positioning was different compared to the CT position since the head coils did not allow use of a headrest nor flat table. Set B consists of CT and axial T2-weighted (T2w) MR images for 12 different patients with a H&N tumor acquired using 6-channel flex coils. Set B was acquired in RT position. The patient was fixed in treatment position using a flat table, a headrest and the original radiotherapy immobilization mask (RT mask). The GE Posifix® flat table was used. For all sets, the CT images were acquired once with the same positioning equipment. Set B has also been used in Chapter2).

The MR images of set A were acquired using two 1.5T GE scanners (GE Healthcare, Waukesha, WI, USA): the Optima MR450 (4 patients) and Optima MR450w (6 patients). All image data of set B were acquired using the Optima MR450w. We used a fast recovery fast spin echo sequence with TR/TE 7060/106.8 ms, flip angle 90°, and matrix 384 x 224. All images in set A and one image in set B had a voxel size of $0.49 \times 0.49 \times 3.00$ mm^3. The images of the other 11 patients of set B had a voxel size of $0.68 \times 0.68 \times 3.00$ mm^3. The field of view (FOV) in the craniocaudal direction was 20 cm centered around the tumor location. This yielded a slightly different FOV location for each patient. The set of T2w images acquired using surface coils was corrected for intensity inhomogeneities using the built-in algorithm (Surface Coil Intensity Correction) of the GE MR scanner.

CT images were acquired using a Siemens scanner (Somatom Sensation Open, Siemens AG, Erlangen, Germany), with an acquisition matrix of 512 x 512, and a voxel size of $0.98 \times 0.98 \times 2.50$ mm^3. Tube voltage and current-time product were 120 kV and 150 mAs respectively.

Twenty-one corresponding landmark points were manually placed by an expert on the CT and MR sets (set A1, A2, and B) to evaluate image alignment after registration. Due to differences in the FOV, not all landmarks could be placed in all the images. The landmarks were annotated twice by the same expert in order to estimate the intra-rater variability in both set A (using set A1) and B. The time interval between the two annotations was one month. We used the same landmarks as those used in Chapter 5. See Figure 5.1 for the description and anatomical location of the 21 landmarks used in our evaluation.

For each acquisition scenario (set A1, A2 and B) T2w sequences were aligned to the CT images and the accuracy of the registration was quantitatively evaluated. Registration was performed in three steps. First, an initial alignment was determined by automatic detection of a landmark in the neck (Chapter 2.2.2.2). Next, a rigid registration followed by a deformable registration was performed using the opensource software Elastix (elastix.isi.uu.nl) [66]. Figure 6.1 shows the datasets and

our experimental setup together with pictures of the small and RT mask. Elastix provides a state-of-the-art deformable registration algorithm [66, 100]. This method has shown excellent performance for several applications and anatomical sites [11, 76] (Chapter 2) and top-ranking performance in thoracic CT registration challenge [110] and a comparative study [64]. Other implementations based on the same registration method [100] are available as opensource software packages as well (i.e. NiftyReg [83], IRTK [104]).

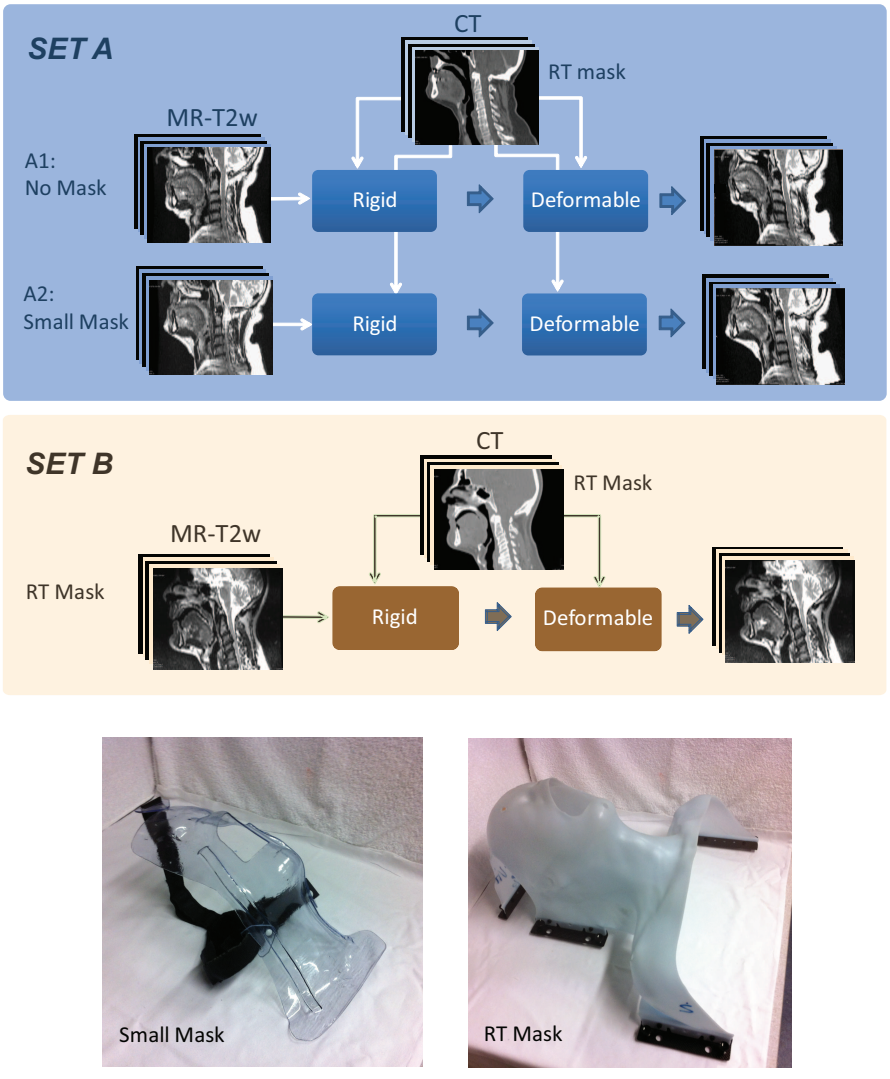


Figure 6.1. Dataset, registration setup and fixation masks.

6.2.1 Registration method

The goal of registration is to find a spatial deformation field $\mathbf{u}(\mathbf{x})$ at each location $\mathbf{x}$ which spatially aligns the moving image $I_M(\mathbf{x}+\mathbf{u}(\mathbf{x}))$ to the fixed image $I_F(\mathbf{x})$. The problem can be formulated as:

$$\hat{\boldsymbol{\mu}} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}}\, \mathcal{C}(\boldsymbol{\mu}; I_F; I_M), \tag{6.1}$$

where the cost function $\mathcal{C}$ is a measure of the image dissimilarity, and $\boldsymbol{\mu}$ is a parameter vector used to represent the deformation field. In this work the deformation was modelled using cubic B-Splines [100] and the cost function was the negated mutual information similarity measure (MI) [78,111,115] combined with a regularization term introduced to penalize undesirable deformations.

$$\mathcal{C}(\boldsymbol{\mu}; I_F; I_M) = -MI(\boldsymbol{\mu}; I_F; I_M) + \lambda \mathcal{R}(\boldsymbol{\mu}), \tag{6.2}$$

where λ is a weighting factor for the regularization term. The MI was computed as in [115] using a joint histogram size of 32 x 32 bins.

Both rigid and deformable registration steps were performed employing a multiresolution strategy. A Gaussian pyramid was used taking into account the anisotropy of the images. The smoothing along the z-direction was set two times lower than the smoothing in the other directions, i.e. the smoothing sigmas were chosen as $\sigma_{x,y} = 4, 2, 1$ voxels and $\sigma_z = 2, 1, 1$ voxels. The number of B-Spline control points used during the registration determines the degrees of freedom of the registration. A uniform control point grid was used for each resolution level: given the finest resolution grid spacing, for each coarser resolution the grid spacing was two times larger.

To solve the minimization problem in Equation 6.1, an adaptive stochastic gradient descent optimization method was used. In this approach a limited number of voxels is used at each iteration to estimate the value of the similarity measure and its derivative [65]. The derivative of the similarity measure is also used at each iteration to adapt the step size of the gradient descent optimization. The stopping criterion of the algorithm is the maximum number of iterations. This number and the number of random samples were chosen to be sufficiently high to guarantee registration convergence, according to [65, 66]. We chose 500 iterations per resolution level for the rigid registration and 2000 iterations per resolution level for the deformable registration. In both cases 4000 samples were used. Compared to Chapter 2 we used more iterations (2000 instead of 1000) for the parameter optimization of deformable registration. This was done in order to consider the possibility of a slower convergence of the optimizer caused by the use of regularization.

During registration, fixed and moving image masks were used to exclude background voxels from the registration. These masks were extracted automatically from MR and CT images using thresholding, connected component analysis, and binary morphology operations.

6.2.2 Parameter optimization and evaluation

For each dataset, a leave-one-out cross-validation was performed for parameter optimization and to evaluate the accuracy of the rigid and deformable registration.

Optimal parameters were found for each patient using an independent training set, i.e. the images of the other patients. In this way, we avoided over-training.

The parameter optimization consisted of an exhaustive search over a predefined set of parameters evaluating all possible combinations of parameters. For the rigid registration the number of resolution levels (2 or 3) was optimized. For the deformable registration the following parameters were optimized: the number of resolution levels (2 or 3), the B-Spline control point grid spacing at finest resolution, and the regularization weight λ. The optimal parameters were chosen as those which gave the best alignment over the training set, given by the lowest average Euclidean distance over all landmarks and all patients.

MR-T2w and CT images were registered for each patient set (A1, A2, B), using optimal parameter settings. The accuracy of registration was measured using the Euclidean distance between the original and registered landmarks. Combined with the accuracy, we evaluated the smoothness of the deformation field. Anatomically plausible deformations should lead to a smooth deformation field between the images. Smoothness was evaluated using the determinant of the spatial Jacobian matrix $J(\mathbf{x})$. The value of this determinant is connected to volume changes caused by the deformation field:

$$det|J(\mathbf{x})| = \begin{cases} \leq 0, & \textit{folding} \text{ effect}, \\ = 1, & \text{volume conservation}, \\ > 1, & \text{volume expansion}, \\ < 1, & \text{volume compression}, \\ >> 1, & \textit{tearing} \text{ effect}. \end{cases} \tag{6.3}$$

6.2.3 Comparison between global and local mutual information

Recently a local version of mutual information was developed [68] and shown to be effective for the problem at hand [76]. Here we compare the accuracy of global mutual information (GMI), presented in this work, with local mutual information (LMI). The same leave-one-out parameter optimization scheme used for GMI was used for LMI with the addition of an extra parameter, the size of the region which is considered for the calculation of LMI. We used cubical regions with 30, 60 and 120 mm size.

In case of LMI the adaptive stochastic gradient descent optimization method cannot be used since it is not possible to correctly estimate the metric gradient at the beginning of each resolution level. Thus a non-adaptive stochastic gradient descent was used [67] with $A = 250$ selected as suggested by Klein et al. in [68] and $\delta = 0.018$ which gave the best performance using one patient.

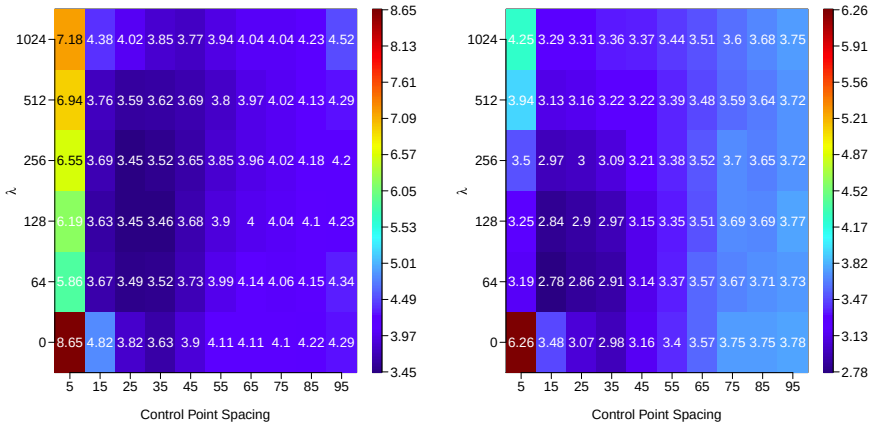
6.2.4 Statistical analysis

A two-sided, Wilcoxon signed rank test was used to test for statistically significant differences between distance measures. For measures corresponding to the same set of patients (set A1 vs. set A2, set A and B vs. the corresponding intra-observer error, and rigid vs. deformable registration) a paired test was used.

6.3 Results

6.3.1 Parameter optimization

Figure 6.2 shows the variation of the registration alignment for set A1, A2, and B, when varying control point grid spacing and regularization weight. The other parameters were fixed, as in the optimal case.



(a) set A1: No Mask (b) set A2: Small Mask

(c) set B: RT Mask

Figure 6.2. Average Euclidean distance over all patients for all combination of control point B-Spline grid spacing and regularization weight λ. The other parameters are fixed.

For both set A1 and A2, for values of the control point grid spacing ≤ 65, the registration accuracy first increases with λ and then, as expected, starts to decrease. For set B the use of regularization had no influence on the registration accuracy for

larger values of grid spacing. The optimal value without using regularization ($\lambda = 0$) was 0.01 mm better than the absolute optimum ($\lambda = 2048$); however this difference was not significant. Since the method is faster and simpler without regularization we chose $\lambda = 0$ as the optimal. Table 6.1 shows the optimal parameter values for each set and patient. In general, the optimal parameter values were stable over patients. In Table 6.1, the optimal parameter values over the entire set which should be used for an independent patient set are printed bold.

Table 6.1. Optimal parameter settings resulting from the leave-one-out optimization. Optimal parameters to use for an independent test set are highlighted in bold. Patient = patient identifier; Levels = number of resolution levels; Grid = B-Spline control point grid spacing at finest resolution;

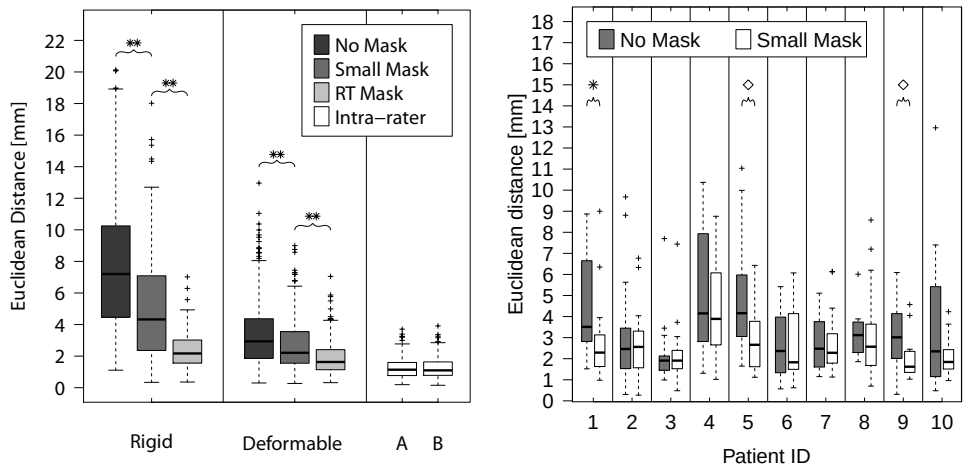
		Rigid registration		
Set.	Patient	Levels	Grid[mm]	λ
set A1	**1, 3 – 10**	**3**	-	-
set A1	2	2	-	-
set A2	**2, 4, 6, 7, 9, 10**	**3**	-	-
set A2	1, 3, 5, 8	2	-	-
set B	**1 – 12**	**3**	-	-
		Deformable registration		
set A1	**1, 2, 5 – 8, 10**	**2**	**25**	**128**
set A1	3, 4, 9	3	25	256
set A2	**1 – 12**	**2**	**15**	**64**
set B	**1 – 5, 7, 9 – 12**	**2**	**85**	-
set B	6	2	95	-
set B	8	3	85	-

6.3.2 Registration evaluation

The alignment after registration for all landmarks and all patients in each set is provided in Figure 6.3 (a) together with the intra-observer variability in landmark placement. Table 6.2 shows data on the alignment.

Figure 6 shows that better patient immobilization leads to better image alignment. Using the RT mask, for both the rigid and deformable registration, we obtained a significantly better alignment ($p < 0.001$) than using the smaller mask or no mask. For both types of registrations, the small immobilization mask provided a better alignment than no mask ($p < 0.001$).

In all cases, deformable registration gave significantly better alignment than rigid registration ($p < 0.001$) and the intra-observer error in landmarks placement was significantly lower ($p < 0.001$) than the registration error obtained for the correspondent set (A and B). We observed that deformable registration had a larger impact on alignment accuracy when there was a greater difference in the patient's position between the scans, i.e. 4.4 mm difference between the rigid and deformable average value for



(a) (b)

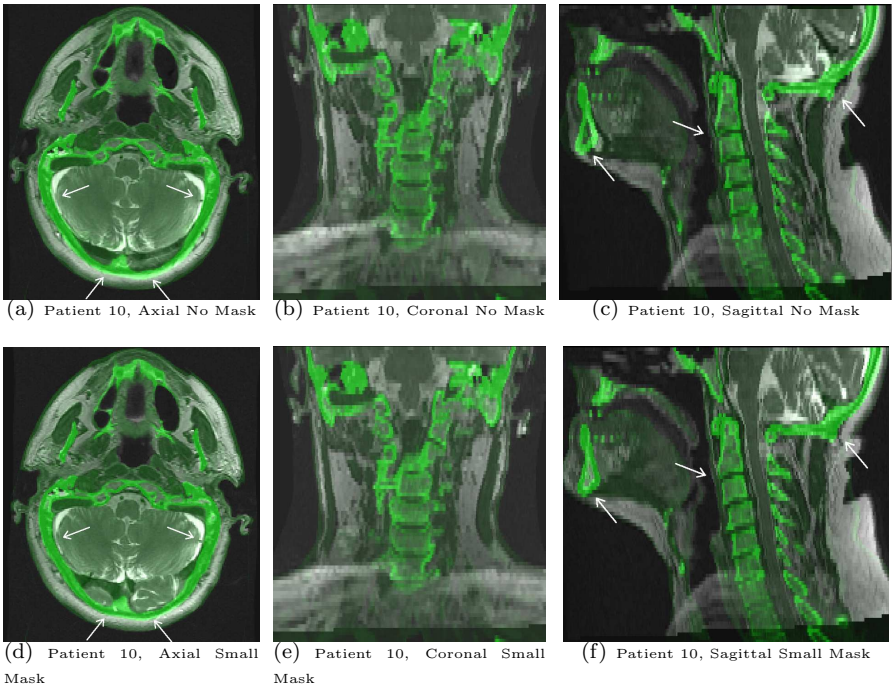
Figure 6.3. Evaluation of registration accuracy using optimal parameter settings. (a) Distribution of rigid and deformable registration accuracy for each set. The intra-observer variability is also provided. (b) Comparison of deformable registration accuracy per patient with and without positioning mask in set A. $\diamond$ indicates a statistical significance with $p < 0.05$, $*$ indicates a statistical significant improvement with $p < 0.01$, and $**$ indicates a statistical significant improvement with $p < 0.001$. The boxes of the boxplots report first and third quartiles, and median values. The whiskers report confidence intervals, which are defined as the most extreme data points (±1.5 times the inter-quartile range around the box) not considered outliers (+).

Table 6.2. Evaluation of registration accuracy using optimal parameters, statistics over all sets. μ = average; σ = standard deviation; IQR = inter quartile range

Rigid registration				
Set	μ[mm]	σ[mm]	Median[mm]	IQR[mm]
set A1	7.93	4.38	7.20	4.47 – 10.25
set A2	4.95	3.09	4.50	2.42 – 6.86
set B	2.36	1.11	2.17	1.56 – 3.02
Deformable registration				
set A1	3.53	2.37	2.93	1.86 – 4.36
set A2	2.73	1.72	2.18	1.56 – 3.49
set B	1.92	1.11	1.63	1.14 – 2.41

set A1, 2.22 mm for set A2, and 0.44 mm for set B. Figure 6 (b) shows comparison of the deformable registration alignment accuracy for set A1 and A2 for each patient. In general a better alignment was obtained for all patients when the small mask was

used (i.e. lower median value and lower distribution for set A2); for three patients this improvement was significant ($p < 0.001$). When the small mask was used, a better alignment of the bones (mandible, vertebrae and skull) could be achieved as shown in Figure 6.4



(a) Patient 10, Axial No Mask (b) Patient 10, Coronal No Mask (c) Patient 10, Sagittal No Mask

(d) Patient 10, Axial Small Mask (e) Patient 10, Coronal Small Mask (f) Patient 10, Sagittal Small Mask

Figure 6.4. Registration results for patient 10 without (a, b, c) and with (d, e, f) positioning mask. Axial, coronal and saggital example slices. The CT image (fixed) is overlaid (in green) on the T2w image. The CT image is used as overlay because of its stronger boundaries. The white arrows point to locations were deformable registration lead to higher improvements.

Table 6.3 shows the average value over all patients of the alignment for each landmark. For the deformable registration, landmarks with lower alignment accuracy can be noticed for the no mask (ID 7, 8, 15, 16, 17, 18) and small mask (15) cases, while the accuracy was spatially uniform when an RT mask was used. The low alignment for 7 and 8 is caused by the wrong alignment of the head position after registration. The low alignment accuracy for 16, 17, 18 in contrast with better alignment of 3, 4, 5 indicates that the vertebrae were not accurately aligned in the no mask case, which is shown in Figure 6.4 (c). The low alignment accuracy of the hyoid (ID 15) bone reflects inaccuracies in the larynx region.

Figure 6.5 shows the distribution over patients of the Jacobian-determinant statistics calculated over the patient body volume. For each set an overall volume conservation was observed (μ close to one and very low σ). Moreover, the range of variability of the Jacobian's determinant excluded the presence of *tearing* and *folding* effect for

Table 6.3. Evaluation of registration accuracy for each landmark for all sets. Accuracy is given by average ± standard deviation.

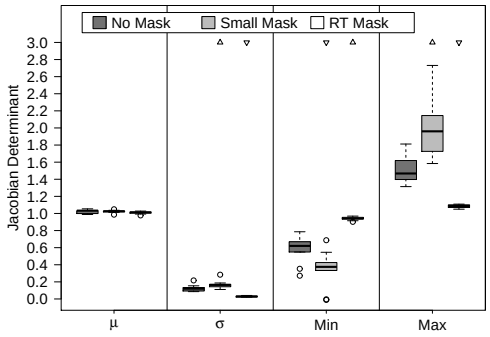
ID	Rigid			Deformable		
	No Mask	Small Mask	RT Mask	No Mask	Small Mask	RT Mask
0	10.26 ± 4.46	7.61 ± 3.63	2.44 ± 0.28	3.24 ± 0.80	2.99 ± 2.25	1.84 ± 1.19
1	8.85 ± 3.55	6.66 ± 2.32	2.90 ± 1.53	3.11 ± 1.30	2.73 ± 1.04	1.99 ± 1.23
2	7.08 ± 3.68	5.21 ± 2.41	2.66 ± 1.31	3.00 ± 2.13	2.72 ± 1.93	2.39 ± 1.33
3	7.51 ± 3.93	5.09 ± 3.68	2.62 ± 0.86	2.97 ± 2.49	2.72 ± 1.42	2.46 ± 1.24
4	7.48 ± 3.44	5.32 ± 2.93	2.54 ± 0.73	3.57 ± 2.44	2.34 ± 0.96	1.93 ± 0.80
5	6.69 ± 3.63	4.18 ± 2.45	2.18 ± 0.82	3.24 ± 1.90	1.99 ± 0.80	1.51 ± 0.69
6	6.47 ± 4.09	4.17 ± 2.70	2.34 ± 0.92	2.63 ± 1.54	2.53 ± 1.56	1.97 ± 0.90
7	9.35 ± 3.84	4.73 ± 3.02	2.18 ± 1.05	4.26 ± 2.39	2.31 ± 1.41	1.33 ± 0.32
8	9.38 ± 5.68	7.82 ± 5.64	2.98 ± 0.81	4.87 ± 3.30	3.06 ± 1.97	2.31 ± 0.81
9	8.37 ± 4.75	4.42 ± 2.37	2.38 ± 1.59	3.79 ± 2.15	3.14 ± 2.11	1.78 ± 1.96
10	7.32 ± 5.34	2.89 ± 2.30	1.62 ± 0.59	2.08 ± 1.23	2.00 ± 1.38	1.53 ± 1.01
11	7.43 ± 5.21	4.19 ± 2.06	2.21 ± 0.81	2.63 ± 1.37	2.47 ± 0.81	1.54 ± 0.80
12	6.93 ± 5.44	3.97 ± 2.39	2.17 ± 0.98	2.02 ± 0.58	2.66 ± 2.10	2.19 ± 0.93
13	8.95 ± 7.25	2.56 ± 1.09	2.65 ± 1.00	2.52 ± 1.22	1.71 ± 0.98	1.75 ± 0.86
14	8.76 ± 5.66	2.59 ± 2.39	2.60 ± 1.18	1.44 ± 1.60	1.53 ± 0.57	1.77 ± 0.97
15	11.50 ± 4.28	9.09 ± 2.94	3.39 ± 1.84	4.65 ± 2.06	5.53 ± 2.32	2.72 ± 1.64
16	8.42 ± 4.14	6.48 ± 4.43	2.79 ± 0.88	5.20 ± 3.42	2.94 ± 2.24	2.75 ± 1.17
17	7.37 ± 3.50	5.31 ± 3.16	2.21 ± 0.94	5.74 ± 3.24	3.21 ± 2.13	2.25 ± 0.82
18	7.49 ± 3.25	4.62 ± 2.51	2.13 ± 0.95	5.07 ± 2.99	2.95 ± 1.63	2.00 ± 1.03
19	6.23 ± 4.65	3.16 ± 1.37	1.87 ± 1.18	2.68 ± 1.65	2.70 ± 1.80	1.77 ± 0.81
20	6.24 ± 4.98	3.35 ± 1.63	1.67 ± 0.93	2.32 ± 1.09	2.49 ± 1.41	1.39 ± 0.75

all patients, with the exception of two patients in set A2. In both cases a slightly negative Jacobian was found (i.e. -0.004 and -0.011) for a very small region (27 and 57 voxels) which indicates *folding*. In both cases these regions corresponded to homogeneous subcutaneous fat located in the back part of the neck.

In Figure 6.5 the symbols $\triangle$ and $\triangledown$ indicate the highest and lowest distributions; these were determined by testing for a significant ($p < 0.05$) difference between each pair of distributions. The smoothest deformation was found when using the RT mask. Use of the small mask resulted in slightly less smoothness of deformation than in the no-mask scenario. This happens because a slightly smaller grid size and a lower regularization were used in the small mask scenario according to the parameter optimization results (see Table 6.1).

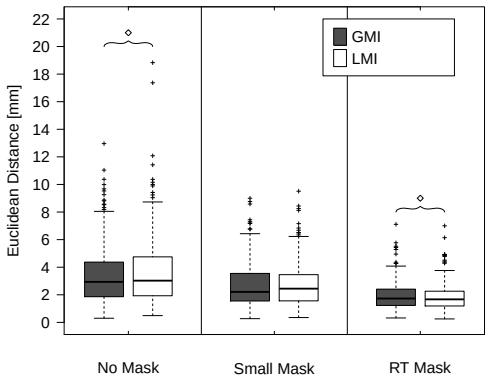
6.3.3 Comparison between global and local mutual information

Figure 6.6 shows the comparison of registration accuracy between global (GMI) and local mutual information (LMI) using optimal parameter settings.



(a)

Figure 6.5. Distribution of Jacobian determinant statistics over patients for the three sets. μ = average; σ = standard deviation; Min = minimum; Max = maximum values. $\triangle$ and $\triangledown$ indicate highest and lowest distributions respectively (significance level of $p < 0.05$). The boxes of the boxplots indicate first and third quartiles, and median values. The whiskers indicate confidence intervals, which are defined as the most extreme data points (±1.5 times the inter-quartile range around the box) not considered as outliers ($\circ$).



(a)

Figure 6.6. Comparison of registration accuracy between global (CMI) and local mutual information (LMI) using optimal parameter settings. $\diamond$ indicates a statistical significance with $p < 0.05$. The boxes of the boxplots report first and third quartiles, and median values. The whiskers report confidence intervals, which are defined as the most extreme data points (±1.5 times the inter-quartile range around the box) not considered outliers ($\circ$).

The results show a small improvement in accuracy when the RT mask was used, a reduction of accuracy when no mask was used and no statistically significant difference

when the small mask was used. The average value of the improvement in the RT mask case was 0.08 mm and the average value of reduction in the no mask case was 0.51 mm (average of the landmark-wise difference).

6.4 Discussion

This study investigated the influence of using patient immobilization equipment on the accuracy of the CT-MR deformable registration in the H&N. A state-of-the-art deformable registration was optimized and evaluated using 21 corresponding anatomical landmarks placed in CT and MR-T2w images. Three datasets in which MR images were acquired with different degrees of immobilization were considered for the evaluation, i.e. no immobilization, or use of a small mask, or of a RT mask. The best image alignment was achieved when the patient was immobilized using the original RT mask and deformable registration was used. In general, a better alignment was achieved when immobilization equipment was used. In agreement with Chapter 5 deformable registration provided a better image alignment than rigid registration. However, to achieve the best level of accuracy we cannot avoid the use of immobilization equipment, even when deformable registration is used.

From the parameter optimization results in Figure 6.1, we conclude that the integration of spatial regularization allows the use of a smaller control point grid size and improves the image alignment, without hampering the smoothness of the deformation field (Figure 6.6). This effect was more important when immobilization equipment was used (as shown in Figure 6.1), i.e. the optimal spacing was 25 mm for set A1, and 15 mm for set A2 and B when $\lambda \neq 0$, while for $\lambda = 0$ the optimal spacing was 35 for set A1 and set A2, and 85 for set B. For the majority of patients the leave-one-out parameters were equal to the optimal parameters calculated using the entire sets; for the remaining patients these parameters ranked as second best. Thus, the optimal parameter settings were robust for patient variations. In addition, the optimal parameter settings were robust to changes in the scanner and coils used for the acquisition (set A was acquired using two different scanners and MR coils).

In the small mask scenario the optimal parameters resulted in a smaller grid size and a lower regularization than the no mask scenario. In case of no mask, more degrees of freedom would be needed to obtain an accurate registration. However this did not happen because in the no mask case the deformation was larger both in amplitude and complexity. Thus a larger grid size resulted optimal to fit the larger deformation. On the other hand, the increased amplitude of the deformation made it harder to fit the higher frequency component as well as in the small mask case.

Set B was used in Chapter 5 to show the feasibility of deformable registration. In this Chapter, this dataset was used as a reference standard. Although it was not possible to make a pairwise comparison between set A and set B, the advantages of using the RT mask are clear from our results.

In set B, the improvement in image alignment between rigid and deformable registration is partly related to the fixation inaccuracy of the RT immobilization equipment. Previous studies have reported the accuracy of patient positioning when using different fixation equipment [44,99]. Using the Posifix fixation setup, a standard

deviation of 2.2 mm was reported at the H&N level [44]. These latter results are in line with the level of accuracy we obtained using rigid registration and the RT mask.

In set A2, for two patients we observed a *folding* effect in two small regions. In both cases, these regions corresponded to subcutaneous fat located at the back of the neck. This effect was caused by the folding of skin and fat due to the different position of the head between CT and MR imaging. Since these regions are peripheral and far from the target and organ at risk, we think this effect is not clinically relevant.

Using B-Spline based deformable registration, Leibfarth et al. [76] reported an alignment of anatomical landmarks of 2.96 ± 1.02 mm for 8 patients. Du Bois d'Aische et al. [28] applied an articulated body motion model to describe the deformation between CT and MR images; the authors reported a mean error of $1.91 - 3.31$ mm and variances of $0.81 - 1.05$ mm for a limited set of four patients. A disadvantage of this method is that it is accurate only in regions close to the spinal cord which generate the motion model. Neither of the latter studies provided details as to which anatomical landmarks were used in the evaluation. We obtained an alignment accuracy slightly lower than these previous studies. However, the value of the final alignment depends on the alignment accuracy obtained after rigid registration, which was lower in our study.

In addition, alignment quantification depends on the anatomical landmarks used for the evaluation, which were not specified in [28, 76]. Furthermore, the approach of [28] not only uses intensity but also vertebrae contours to guide the registration process; thus, for this method, better alignment will be obtained for regions close to the vertebrae.

Leibfarth et al. [76] reported that when using local mutual information (LMI) [68] as similarity metric the most accurate registration was obtained; these authors found a large difference in accuracy between global mutual information (GMI) and LMI (on average 1.82 mm). Our findings are different as we found no relevant improvements in accuracy between LMI and GMI. In the study of Leibfarth et al. [76] GMI might be hampered by the inclusion of background voxels in the metric calculation. In this Chapter, we excluded background voxels also for the computation of GMI by using binary image masks.

The clinical impact of the alignment differences on RTP and HTP can be evaluated by considering the studies [129] and [127]. Voet et al. [129] reported that a small contour distance between different planning target volumes (< 1 mm) led to a relevant underdosage (up to 11 Gy). Thus, since we obtained a mean improvement in alignment accuracy of about 1.5 mm using the RT mask, its use is suggested to achieve an accurate dosimetry. Following Verhaart et al. [127] we can make the same statement with regard to HTP since relevant differences (up to 12%) in local dose (normal tissues) were associated with delineations differences in the order of 1.2 mm.

Three conclusions can be drawn from the present Chapter: 1) deformable registration improves alignment over rigid registration, 2) better immobilization gives better alignment, and 3) deformable registration did not fully compensate for differences in patient positioning. Compared to the current clinical standard, the accuracy obtained without immobilization or with the small mask is insufficient for this application.

Because deformable registration of images acquired using the best patient immobilization equipment provides the most accurate alignment we recommend to use

these settings in clinical practice.

Chapter Seven

Summary and General Discussion

All our knowledge begins with the senses, proceeds then to the understanding, and ends with reason. There is nothing higher than reason.

— Immanuel Kant (1724–1804)

Deep hyperthermia treatment of head and neck (H&N) tumors requires accurate treatment planning, based on 3D patient models that are derived from 3D images. Pre-treatment decision making as well as treatment optimization and adaptations are influenced by the accuracy of these models. In this thesis we investigated automatic methods for 1) the segmentation of CT and MR images of the H&N, and 2) for the registration of CT and MR images to integrate MR into the treatment planning workflow.

In Chapter 2 we presented and evaluated an automatic segmentation algorithm for CT images of the H&N which combines anatomical information, based on atlas registration, with local intensity information in a graph cut framework. In this approach, for each tissue, two models are built for this purpose: an intensity model and a spatial prior model. The intensity model is built by estimating the histogram of foreground and background regions sampling the entire atlas set. The spatial prior model is obtained by registering each atlas image to the target and then combining the registration results in a spatial probability map. A regularization term was also used to promote smooth segmentation. These terms are subsequently used to construct a graph which is solved using graph cut. First we evaluated the advantages of using the proposed approach with respect to multiatlas-based segmentation alone. Next, the influence of different labelers on the method's accuracy was quantified and compared with multiatlas-based segmentation and with the interobserver variability. The method was evaluated with respect to ground truth manual delineation and compared with multiatlas-based segmentation on a dataset of 18 labeled CT images using the Dice similarity coefficient (DSC), the mean surface distance (MSD), and the Hausdorff surface distance (HSD) as evaluation measures. On a subset of 13 labeled images, the influence of different labelers on the method's accuracy was quantified and compared with the interobserver variability. The proposed method had a significantly better score for DSC, MSD and HSD than the multiatlas-based method for the eye vitreous humor. For the DSC, the proposed method performed significantly better for the

segmentation of all the tissues, except brainstem and spinal cord. The MSD showed a significant improvement for optical nerve, eye vitreous humor, lens, and thyroid. For the HSD the proposed method performed significantly better for eye vitreous humor and brainstem. For the majority of the tissues (8/11) the segmentation accuracy of the proposed method approached the interobserver agreement. Moreover our method showed better robustness to variations in atlas labeling compared with multiatlas segmentation. We conclude that the proposed method provides an accurate automatic segmentation of H&N tissues in CT images for the generation of 3D patient models substantially reducing labor involved in therapy planning.

In Chapter 3 we evaluated the impact of using manual and automatic segmentation of Chapter 2 on hyperthermia treatment planning (HTP). To decide whether the automatic segmentation method can be introduced in the clinic, the effect of using this method on HT dose distribution and HTP quality was evaluated. In addition. the effect of using automatic segmentation over manual delineations was compared to other sources of patient model uncertainties, i.e. varying gridsizes and dielectric tissue properties. CT images of seven patients were segmented automatically and manually by four observers, to study inter- and intraobserver geometrical variations. Despite geometrical variations, manual and automatic generated 3D patient models resulted in an equal, i.e. 1%, variation in HTP quality. This variation was minor with respect to the total of other sources of patient model uncertainties, i.e. 11.7%. Furthermore, a considerable reduction in operator time per patient was achieved when using the automatic segmentation (1 hour of processing time + 15 minutes of operator time) instead of the manual segmentation (5-8 hours of operator time). Thus, we conclude that the recently developed automatic segmentation method can be introduced in the clinic for H&N HTP.

In Chapter 4 automatic segmentation of MR images is presented. The use of MR images allows a more accurate tissue delineation for both normal [126] and target [2,45,96,117] tissues. In this chapter, the method presented in Chapter 2 was extended with a more advanced atlas fusion method for deriving the spatial prior term. Different multiatlas locally weighted atlas fusion approaches were investigated using a local search technique for the best matching image region between each atlas and the target image. As in the previous chapter, the intensity model and the regularization term were combined with the spatial prior term in a graph cut framework. The accuracy of the proposed method was high both in terms of Dice similarity coefficient (DSC) and mean surface distance (MSD) with median DSC higher than 0.8 for all tissues except sclera and median MSD lower than 1 mm for all tissues except the CSF, compared to manual delineations. For all tissues, except the spine tissues (myelum and CSF), the accuracy approached the interobserver agreement/variability both in terms of DSC and MSD. The beneficial effect of adding the intensity modelling to the multiatlas fusion approach decreased when a more accurate atlas fusion method was used. Using the proposed approach we improved the performance of the approach presented in Chapter 2 for H&N hyperthermia treatment planning making the method suitable for clinical application.

In Chapter 5, we investigated the feasibility of using deformable registration in clinical practice to align MR and CT images of the H&N for treatment planning. A state-of-the-art deformable registration algorithm was optimized, evaluated, and

compared with rigid registration, which is the current clinical standard. The evaluation was based on 21 anatomical landmarksand regions of interest (ROIs) manually annotated by an expert in both modalities. This evaluation scheme was selected with respect to the applications (RTP and HTP), landmarks and ROIs visibility on both images, and volume coverage. We also developed a multiparametric registration approach, which simultaneously aligns T1- and T2-weighted MR sequences to the CT. This approach was evaluated and compared with single-parametric approaches. Both landmark-based and ROI-based evaluation showed that deformable registration was more accurate than rigid registration without introducing unrealistic deformations. The landmarks and the ROIs alignment together with the smoothness of the deformation demonstrated that the deformable registration provides a better spatial alignment than rigid registration in the entire volume. The registration error was lower than the voxel resolution for landmark-based (median error around 1.6 mm) and ROI-based evaluation (mean surface distance of approximately 1 mm for all the ROIs excluding the cerebellum). Similar accuracies were obtained for the single and multiparametric approaches. We therefore conclude that deformable registration of H&N CT and MR images is feasible with a registration error lower than the current clinical standard.

In Chapter 6 we investigated whether deformable registration of CT and MR images can be used to obviate the use of patient immobilization equipment during image acquisition, which takes time and is uncomfortable for the patient. CT and MR images of 22 patients were aligned using state-of-the-art rigid and deformable registration algorithms. For 10 patients, MR T2-weighted images were acquired in two settings using standard head coils: with no immobilization mask and with a small immobilization mask created from the original radiotherapy immobilization mask in order to fit it inside standard head coils. The small mask was used to fix the patients in the treatment position. Still, patient positioning was different compared to the CT position since the head coils did not allow use of a headrest nor flat table. For 12 patients, the images were acquired using a 6 channel surface coil and the original radiotherapy immobilization equipment i.e. radiotherapy mask, headrest and flat table. The registration accuracy was evaluated using 21 anatomical landmarks, localized on CT and MR images by an expert. Leave-one-out experiments were performed for evaluation and optimization of the rigid and deformable registrations parameters. The best image alignment was achieved when the patient was immobilized using the original RT mask and applying deformable registration (median error of 1.63 mm). In general, a better alignment was achieved when immobilization equipment was used. In agreement with Chapter 5, deformable registration provided a better image alignment than rigid registration. However, to achieve the best level of accuracy we cannot avoid the use of immobilization equipment, even when using deformable registration. Because deformable registration of images acquired using the best patient immobilization equipment provides the most accurate alignment, we recommend to use these settings in clinical practice.

In this thesis, we presented methods for the automatic segmentation of CT and MR images for HTP, and the automatic deformable registration of CT and MR images. The segmentation methods work with accuracies that are close to the interobserver

variations which we shown to have a minor impact on HTP quality. The deformable registration method provides a better alignment accuracy than the current clinical standard without introducing un-realistic deformation. Therefore the methods proposed in this thesis are ready for clinical implementation.

In the first three chapter of this thesis we showed that automatic segmentation of CT and MR images of the H&N is feasible with an overall accuracy comparable with manual delineation. The method described in Chapter 2 was implemented in a software tool which was used to generate full 3D patient model as described in Chapter 3. This tool is currently in use for hyperthermia treatment planning in the clinic. The results of the tool are being verified for accuracy before the generation of patient models. So far, the result of this evaluation is that only minor corrections are needed in particular for the eyes and neck tissues, and in the ears and nose area. The operator time of approximately 5-8h for manual delineations was significantly reduced to approximately 1 hour minutes needed for corrections. This tool can also be extended to incorporate the modification described in Chapter 4 to include the segmentation of MR images and increase the patient model accuracy [126]. In this case, CT images will still be required for the segmentation of some tissues: muscle, fat, bone, cartilage, thyroid, optical nerve. In addition, using MR images more tissues can be included in the patient model i.e. the white and grey matter of the brain and the CSF. It has been shown that HTP quality is significantly influenced by a more detailed segmentation that include these tissues [126].

The method presented in Chapter 4 was not evaluated for CT images. We expect that the positive results obtained for MR images by using locally weighted atlas fusion, will also apply to CT images, as we do not foresee any difficulties or limitations of the model with respect to the image modality. Since CT images have a quantitative intensity value (Hounsfield Unit) the intensity model may be more effective in improving the method accuracy than for the MR images. Thus, this method can be used to improve the CT-based patient models.

Our automatic segmentation method for the H&N was evaluated for tissues relevant for HTP. Most of the tissues segmented are also needed for radiotherapy treatment planning (RTP). Our framework can be easily extended for segmentation of additional tissues relevant for RTP, e.g. parotids and submandibular glands, by adding these to the atlases and optimizing the method parameters for these tissues. The extension may be particularly interesting because of the lack of specific automatic segmentation method for H&N MR images acquired for radiotherapy purposes.

In the second part of the thesis we focused on developing a strategy to obtain accurate registrations of CT and MR images. The registration of CT and anatomical MR scans (T1- or T2-weighted) can also be used to integrate vascular information from contrast enhanced MR into HTP. Since blood perfusion is an important factor in thermal redistribution [74], including a patient-specific vasculature model in the current thermal model implementation [126] may increase the accuracy of thermal simulation in HTP of the H&N [69].

Automatic segmentation of CT and MR images for patient model generation in HTP can be performed with high accuracy, comparable to manual delineation. In addition MR images can be aligned with high accuracy to the CT, using patient im-

mobilization equipment during image acquisition. Therefore we conclude that the integration of automatic MR- and CT-based patient models is ready for clinical application.

Bibliography

Reading furnishes the mind only with materials of knowledge; it is thinking that makes what we read ours.

— John Locke, (1632 - 1704)

[1] "GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and prevalence worldwide in 2012,", `http://globocan.iarc.fr/Pages/fact_sheets_population.aspx`, Accessed 18-01-2015.

[2] M Ahmed, M Schmidt, A Sohaib, CE Kong, K Burke, C Richardson, M Usher, S Brennan, A Riddell, M Davies, K Newbold, KJ Harrington, CM Nutting, "The value of magnetic resonance imaging in target volume delineation of base of tongue tumours–a study using flexible surface coils.", *Radiother Oncol*, vol. 94, pp. 161–7, Mar. 2010.

[3] A Akhondi-Asl & S K Warfield, "Simultaneous truth and performance level estimation through fusion of probabilistic segmentations.", *IEEE Trans Med Imag*, vol. 32, no. 10, pp. 1840–52, Oct. 2013.

[4] P Aljabar, R A Heckemann, A Hammers, J A, V Hajnal, D Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy", *NeuroImage*, vol. 46, no. 3, pp. 726–38, 2009.

[5] M Amichetti, M Romano, L Busana, A Bolner, G Fellin, G Pani, L Tomio, R Valdagni, "Hyperfractionated radiation in combination with local hyperthermia in the treatment of advanced squamous cell carcinoma of the head and neck: a phase I-II study", *Radiother Oncol*, vol. 45, no. 2, pp. 155–8, nov 1997.

[6] X Artaechevarria, A Munoz-Barrutia, C Ortiz-de Solorzano, "Combination strategies in multi-atlas image segmentation: application to brain MR data.", *IEEE Trans Med Imag*, vol. 28, no. 8, pp. 1266–77, Aug. 2009.

[7] J F Bakker, M M Paulides, A Christ, N Kuster, G C van Rhoon, "Assessment of induced SAR in children exposed to electromagnetic plane waves between 10 MHz and 5.6 GHz.", *Phys Med Biol*, vol. 55, no. 11, pp. 3115–30, Jun 2010.

[8] J F Bakker, M M Paulides, A H Westra, H Schippers, G C Van Rhoon, "Design and test of a 434 MHz multi-channel amplifier system for targeted hyperthermia applicators", *Int J Hyperthermia*, vol. 26, no. 2, pp. 158–70, 2010.

[9] A U Batur & M H Hayes, "Adaptive Active Appearance Models", *IEEE Trans Image Proc*, vol. 14, no. 11, pp. 1707–1721, 2005.

[10] G Bekes, E Mate, L G Nyul, A Kuba, M Fidrich, "Geometrical model-based segmentation of the organs of sight on CT images", *Med Phys*, vol. 35, no. 2, pp. 735–43, Feb. 2008.

[11] F F Berendsen, U A van der Heide, T R Langerak, A N T J Kotte, J P W Pluim, "Free-form image registration regularized by a statistical shape model: application to organ segmentation in cervical MR", *Computer Vision and Image Understanding*, vol. 117, pp. 1119–27, Sept. 2013.

[12] Y Boykov & G Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation", *Int J Comput Vision*, vol. 70, no. 2, pp. 109–31, Nov. 2006.

[13] Y Boykov & V Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision", *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no. 9, pp. 1124 –37, Sept. 2004.

[14] Y Boykov, O Veksler, R Zabih, "Fast approximate energy minimization via graph cuts", *IEEE Trans Pattern Anal Mach Intell*, vol. 23, pp. 1222–39, 2001.

[15] K K Brock, "Results of a multi-institution deformable registration accuracy study (MIDRAS).", *Int J Radiat Oncol Biol Phys*, vol. 76, pp. 583–96, Feb. 2010.

[16] C L Brouwer, R J Steenbakkers, A Navran, H P Bijl, "3D Variation in delineation of head and neck organs at risk", *Radiother Oncol*, vol. 7:32, 2012.

[17] J N H Brunt, "Computed tomography-magnetic resonance image registration in radiotherapy treatment planning.", *Clin Oncol*, vol. 22, pp. 688–97, Oct. 2010.

[18] E C Burdette, P G Friederich, R L Seaman, L E Larsen, "In Situ Permittivity of Canine Brain: Regional Variations and Postmortem Changes", *IEEE Trans Microw Theory Techn*, vol. 34, pp. 38–50, 1986.

[19] R A M Canters, M Franckena, M M Paulides, G C Van Rhoon, "Patient positioning in deep hyperthermia: influences of inaccuracies, signal correction possibilities and optimization potential.", *Phys Med Biol*, vol. 54, no. 12, pp. 3923–36, Jun 2009.

[20] R A M Canters, M M Paulides, M F Franckena, J van der Zee, G C van Rhoon, "Implementation of treatment planning in the routine clinical procedure of regional hyperthermia treatment of cervical cancer: an overview and the Rotterdam experience", *Int J Hyperthermia*, vol. 28, no. 6, pp. 570–81, 2012.

[21] V Caselles, R Kimmel, G Sapiro, "Geodesic active contours", *Int J Comput Vis*, vol. 1, no. 22, pp. 1–20, 1997.

[22] T F Chan & L A Vese, "Active Contours Without Edges", *Trans Image Proc*, vol. 10, no. 2, pp. 266–77, 2001.

[23] A Chen, K J Niermann, M A Deeley, B M Dawant, "Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT", *Phys Med Biol*, vol. 57, no. 1, pp. 93–111, Jan. 2012.

[24] O Commowick, S K Warfield, G Malandain, "Using Frankenstein's creature paradigm to build a patient specific atlas", in *MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12, pp. 993–1000, 2009.

[25] B A Conley, "Treatment of Advanced Head and Neck Cancer: What Lessons Have We Learned?", *J Clin Oncol*, vol. 24, no. 7, pp. 1023–5, Jan. 2006.

[26] M B Cuadra, S Gorthi, F I. Karahanoglu, B Paquier, A Pica, H P Do, A Balmer, F Munier, J-P Thiran, "Model-Based Segmentation and Fusion of 3D Computed Tomography and 3D Ultrasound of the Eye for Radiotherapy Planning", in *Computational Vision and Medical Image Processing*, vol. 19 of *Computational Methods in Applied Sciences*, Springer Netherlands, pp. 247–263, 2011.

[27] L R Dice, "Measures of the Amount of Ecologic Association Between Species", *Ecology*, vol. 26, pp. pp. 297–302, July 1945.

[28] A du Bois d'Aische, M de Craene, X Geets, V Grégoire, B Macq, SK Warfield, "Estimation of the deformations induced by articulated bodies: Registration of the spinal column", *Biomed Signal Process Control*, vol. 2, pp. 16–24, Jan. 2007.

[29] M Dubuisson & A K Jain, "A Modified Hausdorff Distance for Object Matching", in *ICPR International Conference on Pattern Recognition*, pp. 566–68, 1994.

[30] J A Elder, "Ocular effects of radiofrequency energy", *Bioelectromagnetics*, vol. Suppl 6, pp. S148–61, 2003.

[31] P M Evans, "Anatomical imaging for radiotherapy.", *Phys Med Biol*, vol. 53, pp. R151–91, June 2008.

[32] B Fischl, D H Salat, E Busa, M Albert, M Dieterich, C Haselgrove, A van der Kouwe, R Killiany, D Kennedy, S Klaveness, A Montillo, N Makris, B Rosen, A M Dale, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain", *Neuron*, vol. 33, no. 3, pp. 341–55, Jan. 2002.

[33] V Fortunati, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M M Paulides, T van Walsum, "Hyperthermia Critical Tissues Automatic Segmentation of Head and Neck CT Images Using Atlas Registration and Graph Cuts", in *IEEE ISBI - International Symposium on Biomedical Imaging: from Nano to Macro*, 2012.

[34] V Fortunati, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M M Paulides, T van Walsum, "Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling", *Med Phys*, vol. 40, no. 7, p. 071905, July 2013.

[35] V Fortunati, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M M Paulides, T van Walsum, "Feasibility of multi-modal deformable registration for head and neck tumor treatment planning", *Int J Radiat Oncol Biol Phys*, vol. 90, pp. 85–93, Sept. 2014.

[36] M Franckena, R A M Canters, F Termorshuizen, J Van Der Zee, G Van Rhoon, "Clinical implementation of hyperthermia treatment planning guided steering: A cross over trial to assess its current contribution to treatment quality", *Int J Hyperthermia*, vol. 26, no. 2, pp. 145–57, 2010.

[37] M Franckena, D Fatehi, M de Bruijne, R A M Canters, Yvette van Norden, Jan Willem Mens, Gerard C van Rhoon, Jacoba van der Zee, "Hyperthermia dose-effect relationship in 420 patients with cervical cancer treated with combined radiotherapy and hyperthermia", *Eur J Cancer*, vol. 45, no. 11, pp. 1969–78, July 2009.

[38] M Franckena, L J A Stalpers, P C M Koper, R G J Wiggenraad, W J Hoogenraad, J D P van Dijk, C C Wrlm-Rodenhuis, J J Jobsen, G C van Rhoon, J van der Zee, "Long-term improvement in treatment outcome after radiotherapy and hyperthermia in locoregionally advanced cervix cancer: an update of the Dutch Deep Hyperthermia Trial.", *Int J Radiat Oncol Biol Phys*, vol. 70, no. 4, pp. 1176–82, Mar 2008.

[39] K D Fritscher, M Peroni, P Zaffino, M F Spadea, R Schubert, G Sharp, "Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours.", *Med Phys*, vol. 41, no. 5, p. 051910, May 2014.

[40] C Gabriel & A Peyman, "Dielectric measurement: error analysis and assessment of uncertainty.", *Phys Med Biol*, vol. 51, no. 23, pp. 6033–46, Dec 2006.

[41] S Gabriel, R W Lau, C Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz.", *Phys Med Biol*, vol. 41, no. 11, pp. 2251–69, Nov 1996.

[42] S Gabriel, R W Lau, C Gabriel, "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues", *Phys Med Biol*, vol. 41, no. 11, pp. 2271–93, Nov. 1996.

[43] S Gabriel, R W Lau, C Gabriel, "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues.", *Phys Med Biol*, vol. 41, no. 11, pp. 2271–93, Nov 1996.

[44] L Gilbeau, M Octave-prignot, T Loncol, L Renard, P Scalliet, V Gre, "Comparison of setup accuracy of three different thermoplastic masks for the treatment of brain and head and neck tumors", *Radiother Oncol*, vol. 58, pp. 155–62, 2001.

[45] A R Gordon, L A Loevner, A Shukla-Dave, R O Redfern, Adina I Sonners, Alex M Kilger, M A Elliott, M Machtay, R S Weber, J D Glickson, D I Rosenthal, "Intraobserver variability in the MR determination of tumor volume in squamous cell carcinoma of the pharynx.", *Am J Neuroradiol*, vol. 25, pp. 1092–8, 2004.

[46] S Gorthi, V Duay, M B Cuadra, U Schick, P-A Tercier, S A Allal, J-P Thiran, "Active Contour-Based Segmentation of Head and Neck with Adaptive Atlas Selection.", in *The MIDAS Journal - Head and Neck Auto-Segmentation Challenge.*, vol. 1, p. 14, 2009.

[47] S Gorthi, V Duay, N Houhou, M B Cuadra, U Schick, M Becker, S A Allal, J-P Thiran, "Segmentation of Head and Neck Lymph Node Regions for Radiotherapy Planning Using Active Contour-Based Atlas Registration", *IEEE J Sel Topics Signal Process*, vol. 3, no. 1, pp. 135–47, Feb. 2009.

[48] X Han, M S Hoogeman, P C Levendag, L S Hibbard, D N Teguh, P Voet, A C Cowen, T K Wolf, "Atlas-based auto-segmentation of head and neck CT images", *MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11, no. Pt 2, pp. 434–41, 2008.

[49] S Hanvey, M Glegg, J Foster, "Magnetic resonance imaging for radiotherapy planning of brain cancer patients using immobilization and surface coils.", *Phys Med Biol*, vol. 54, pp. 5381–94, Sept. 2009.

[50] S Hanvey, M McJury, L M Tho, M Glegg, M Thomson, D Grose, A James, M Rizwanullah, C Paterson, J Foster, "The influence of MRI scan position on patients with oropharyngeal cancer undergoing radical radiotherapy.", *Radiat Oncol*, vol. 8, pp. 129–39, May 2013.

[51] S L Hartmann, M H Parks, P R Martin, B M Dawant, "Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part II, validation on severely atrophied brains", *IEEE Trans Med Imag*, vol. 18, no. 10, pp. 917–26, Oct. 1999.

[52] M Hofer, *CT Teaching Manual: A Systematic Approach to CT Reading*, Thieme, 2007.

[53] C Hoogendoorn, F M Sukno, S Ordás, A F Frangi, "Bilinear Models for Spatio-Temporal Point Distribution Analysis", *International Journal of Computer Vision*, vol. 85, no. 3, pp. 237–252, Feb. 2009.

[54] Y Hua, S Ma, Z Fu, Q Hu, L Wang, Y Piao, "Intracavity hyperthermia in nasopharyngeal cancer: a phase III clinical study", *Int J Hyperthermia*, vol. 27, no. 2, pp. 180–86, 2011.

[55] N G Huilgol, S Gupta, R Dixit, "Chemoradiation with hyperthermia in the treatment of head and neck cancer", *Int J Hyperthermia*, vol. 26, no. 1, pp. 21–5, Feb. 2010.

[56] D P Huttenlocher, W J Rucklidge, G A Klanderman, "Comparing Images Using the Hausdorff Distance", *IEEE Trans Pattern Anal Mach Intell*, vol. 15, pp. 850–63, 1993.

[57] R D Issels, L H Lindner, J Verweij, P Wust, P Reichardt, B-C Schem, S Abdel-Rahman, S Daugaard, C Salat, C-M Wendtner, Z Vujaskovic, R Wessalowski, K-W Jauch, H R Dürr, F Ploner, A Baur-Melnyk, U Mansmann, W Hiddemann, J-Y Blay, P Hohenberger, "Neo-adjuvant chemotherapy alone or with regional hyperthermia for localised high-risk soft-tissue sarcoma: a randomised phase 3 multicentre study.", *Lancet Oncol*, vol. 11, no. 6, pp. 561–70, June 2010.

[58] W T Joines, Y Zhang, C Li, R L Jirtle, "The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz.", *Med Phys*, vol. 21, no. 4, pp. 547–50, Apr 1994.

[59] E L Jones, J R Oleson, L R Prosnitz, T V Samulski, Z Vujaskovic, D Yu, L L Sanders, M W Dewhirst, "Randomized trial of hyperthermia and radiation for superficial tumors.", *J Clin Oncol*, vol. 23, no. 13, pp. 3079–85, May 2005.

[60] H H Kampinga, "Cell biological effects of hyperthermia alone or combined with radiation or drugs: a short introduction to newcomers in the field", *Int J Hyperthermia*, vol. 22, no. 3, pp. 191–6, May 2006.

[61] M Kass, A. Witkin, D Terzopoulos, "Snakes: active contur models", *Int J Comp Vision*, pp. 1633–6, 1988.

[62] F Khalifa, G M Beache, G Gimel, J S Suri, A El-baz, *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, vol. 1, Springer US, Boston, MA, 2011.

[63] J Kittler, M Hatef, R P W Duin, J Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–39, Mar. 1998.

[64] A Klein, J Andersson, B A Ardekani, J Ashburner, B Avants, M-C Chiang, G E Christensen, D L Collins, J Gee, P Hellier, J H Song, M Jenkinson, C Lepage, D Rueckert, P Thompson, T Vercauteren, R P Woods, J J Mann, R V Parsey, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.", *NeuroImage*, vol. 46, pp. 786–802, July 2009.

[65] S Klein, J P W Pluim, M Staring, M A Viergever, "Adaptive Stochastic Gradient Descent Optimisation for Image Registration", *Int J Comput Vision*, vol. 81, pp. 227–39, Aug. 2008.

[66] S Klein, M Staring, K Murphy, M A Viergever, J P W Pluim, "Elastix: a Toolbox for Intensity-Based Medical Image Registration.", *IEEE Trans Med Imag*, vol. 29, pp. 196–205, Jan. 2010.

[67] S Klein, M Staring, J P W Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines.", *IEEE Trans Image Process*, vol. 16, pp. 2879–90, Dec. 2007.

[68] S Klein, U A van der Heide, I M Lips, M van Vulpen, M Staring, J P W Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information", *Med Phys*, vol. 35, pp. 1407–17, 2008a.

[69] H Petra Kok, J Gellermann, C A T van den Berg, P R Stauffer, J W Hand, J Crezee, "Thermal modelling using discrete vasculature for thermal therapy: A review", *Int J Hyperthermia*, vol. 29, no. 4, pp. 336–45, June 2013.

[70] V Kolmogorov & R Zabin, "What energy functions can be minimized via graph cuts?", *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no. 2, pp. 147–59, Feb. 2004.

[71] K Kuczyski, M Siczek, R Stegierski, "Image Registration Tuning for DCE-MRI Breast Imaging", in *Image Processing and Communications Challenges 4*, Ryszard S. Chora (ed.), no. 184 in Advances in Intelligent Systems and Computing, Springer Berlin Heidelberg, pp. 39–46, 2013.

[72] S Kumar & M Hebert, "Discriminative random fields: a discriminative framework for contextual interaction in classification", in *IEEE CV - International Conference on Computer Vision*, vol. 2, pp. 1150 –1157, Oct. 2003.

[73] S Kumar & M Hebert, "Discriminative Random Fields", *Int J Comput Vision*, vol. 68, no. 2, pp. 179–201, June 2006.

[74] J J Lagendijk, P Hofman, J Schipper, "Perfusion analyses in advanced breast carcinoma during hyperthermia", *Int J Hyperthermia*, vol. 4, no. 5, pp. 479–95, Oct. 1988.

[75] B A Landman & S K Warfield, "MICCAI 2012 workshop on multi-atlas labeling", in *MICCAI Grand Challenge and Workshop on Multi-Atlas Labeling Challenge Results.*, pp. 33–42, 2012.

[76] S Leibfarth, D Mönnich, S Welz, C Siegel, N Schwenzer, H Schmidt, D Zips, D Thorwarth, "A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning.", *Acta Oncol*, vol. 52, pp. 1353–9, Oct. 2013.

[77] J-P H Machiels, L F Licitra, R I Haddad, M Tahara, E E W Cohen, "Rationale and design of LUX-Head & Neck 1: a randomised, Phase III trial of afatinib versus methotrexate in patients with recurrent and/or metastatic head and neck squamous cell carcinoma who progressed after platinum-based therapy", *BMC Cancer*, vol. 14, no. 1, p. 473, June 2014.

[78] F Maes, A Collignon, D Vandermeulen, G Marchal, P Suetens, "Multimodality image registration by maximization of mutual information.", *IEEE Trans Med Imag*, vol. 16, pp. 187–98, Apr. 1997.

[79] F Mantlik, M Hofmann, M K Werner, A Sauter, J Kupferschläger, B Schölkopf, B J Pichler, T Beyer, "The effect of patient positioning aids on PET quantification in PET/MR imaging.", *Eur J Nucl Med Mol Imaging*, vol. 38, pp. 920–9, May 2011.

[80] D Mattes, D R Haynor, H Vesselle, T K Lewellen, W Eubank, "PET-CT image registration in the chest using free-form deformations", *IEEE Trans Med Imag*, vol. 22, no. 1, pp. 120–28, Jan. 2003.

[81] R L Mcintosh, "A comprehensive tissue propperties database provided for the thermal assessment of a human at rest", *Biophysical Reviews and Letters*, vol. 5, pp. 129–51, 2010.

[82] R L Mcintosh & V Anderson, "A comprehensive tissue properties database provided for the thermal assessment of a human at rest", *Biophysical Reviews and Letters*, vol. 05, no. 03, pp. 129–51, Sept. 2010.

[83] M Modat & J Mcclelland, "Lung registration using the NiftyReg package", in *MICCAI Workshop: Medical Image Analysis For The Clinic - A Grand Challenge*, pp. 33–42, 2010.

[84] C S Moore, G P Liney, A W Beavis, "Quality assurance of registration of CT and MRI data sets for treatment planning of radiotherapy for head and neck cancers.", *J Appl Clin Med Phys*, vol. 5, pp. 25–35, Jan. 2004.

[85] M Murbach, E Neufeld, W Kainz, K P Pruessmann, N Kuster, "Whole-body and local RF absorption in human models as a function of anatomy and position within 1.5T MR body coil.", *Magn Reson Med*, Feb 2013.

[86] S Mutic, J F Dempsey, W R Bosch, D A Low, R E Drzymala, K S Chao, S M Goddu, P D Cutler, J A Purdy, "Multimodality image registration quality assurance for conformal three-dimensional treatment planning.", *Int J Radiat Oncol Biol Phys*, vol. 51, pp. 255–60, Sept. 2001.

[87] B E Nelms, W A Tom, G Robinson, J Wheeler, "Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer.", *Int J Radiat Oncol Biol Phys*, vol. 82, no. 1, pp. 368–78, Jan 2012.

[88] B E Nelms, W A Tomé, G Robinson, J Wheeler, "Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer.", *Int J Radiat Oncol Biol Phys*, vol. 82, pp. 368–78, Jan. 2012.

[89] M M Paulides, J F Bakker, M Linthorst, J van der Zee, Z Rijnen, E Neufeld, P M T Pattynama, P P Jansen, P C Levendag, G C van Rhoon, "The clinical feasibility of deep hyperthermia treatment in the head and neck: new challenges for positioning and temperature measurement", *Phys Med Biol*, vol. 55, pp. 2465–80, May 2010.

[90] M M Paulides, J F Bakker, E Neufeld, J van der Zee, P P Jansen, P C Levendag, G C van Rhoon, "Winner of the "New Investigator Award" at the European Society of Hyperthermia Oncology Meeting 2007. The HYPERcollar: a novel applicator for hyperthermia in the head and neck", *Int J Hyperthermia*, vol. 23, no. 7, pp. 567–76, Nov. 2007.

[91] M M Paulides, P R Stauffer, E Neufeld, P F Maccarini, A Kyriakou, R A M Canters, C J Diederich, J F Bakker, G C van Rhoon, "Simulation techniques in hyperthermia treatment planning.", *Int J Hyperthermia*, vol. 29, no. 4, pp. 346–57, June 2013.

[92] A Peyman & C Gabriel, "Cole-Cole parameters for the dielectric properties of porcine tissues as a function of age at microwave frequencies.", *Phys Med Biol*, vol. 55, no. 15, pp. N413–9, Aug 2010.

[93] A Peyman, C Gabriel, E H Grant, G Vermeeren, L Martens, "Variation of the dielectric properties of tissues with age: the effect on the values of SAR in children when exposed to walkie-talkie devices.", *Phys Med Biol*, vol. 54, no. 2, pp. 227–41, Jan 2009.

[94] J P W Pluim, J B A Maintz, M A Viergever, "Mutual-information-based registration of medical images: a survey", *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, Aug. 2003.

[95] A A Qazi, V Pekar, J Kim, J Xie, S L Breen, D A Jaffray, "Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach", *Med Phys*, vol. 38, no. 11, pp. 6160–70, Nov. 2011.

[96] C R N Rasch, R J H M Steenbakkers, I Fitton, J C Duppen, Peter J C M Nowak, F A Pameijer, A Eisbruch, J H A M Kaanders, F Paulsen, M van Herk, "Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer.", *Radiat Oncol*, vol. 5, pp. 5–21, Jan. 2010.

[97] Z Rijnen, J F Bakker, R A M Canters, P Togni, G Verduijn, P Levendag, G C van Rhoon, M M Paulides, "Clinical integration of software tool VEDO for adaptive and quantitative application of phased array hyperthermia in the head and neck", *Int J Hyperthermia*, vol. 29, no. 3, pp. 181–93, 2013.

[98] T Rohlfing, R Brandt, R Menzel, D B Russakoff, Jr C R Maurer, "Quo Vadis, Atlas-Based Segmentation?", in *The Handbook of Medical Image Analysis – Volume III: Registration Models*, Jasjit Suri, David L. Wilson, Swamy Laxminarayan (eds.), vol. 3, Kluwer Academic / Plenum Publishers, Ch. 11, pp. 435–486, Aug. 2005.

[99] R L Rotondo, K Sultanem, I Lavoie, J Skelly, L Raymond, "Comparison of repositioning accuracy of two commercially available immobilization systems for treatment of head-and-neck tumors using simulation computed tomography imaging.", *Int J Radiat Oncol Biol Phys*, vol. 70, pp. 1389–96, Apr. 2008.

[100] D Rueckert, L I Sonoda, C Hayes, D L Hill, M O Leach, D J Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images.", *IEEE Trans Med Imag*, vol. 18, pp. 712–21, Aug. 1999.

[101] M R Sabuncu, B T T Yeo, K van Leemput, B Fischl, P Golland, "A generative model for image segmentation based on label fusion", *IEEE Trans Med Imag*, vol. 29, no. 10, pp. 1714–29, Oct. 2010.

[102] G Schmid, G Neubauer, U M Illievich, F Alesch, "Dielectric properties of porcine brain tissue in the transition from life to death at frequencies from 800 to 1900 MHz.", *Bioelectromagnetics*, vol. 24, no. 6, pp. 413–22, Sep 2003.

[103] G Schmid & R Uberbacher, "Age dependence of dielectric properties of bovine brain and ocular tissues in the frequency range of 400 MHz to 18 GHz.", *Phys Med Biol*, vol. 50, no. 19, pp. 4711–20, Oct 2005.

[104] J A Schnabel, D Rueckert, M Quist, J M Blackall, A D Castellano-Smith, T Hartkens, G P Penney, W A Hall, H Liu, C L Truwit, F A Gerritsen, D L G Hill, D J Hawkes, "A Generic Framework for Non-rigid Registration Based on Non-uniform Multi-level Free-Form Deformations", in *MICCAI - Medical Image Computing and Computer-Assisted Intervention*, pp. 573–81, 2001.

[105] M Sherar, F F Liu, M Pintilie, W Levin, J Hunt, R Hill, J Hand, C Vernon, C G van Rhoon, J van der Zee, D G González González, J van Dijk, J Whaley, D Machin, "Relationship between thermal dose and outcome in thermoradiotherapy treatments for superficial recurrences of breast cancer: data from a phase III trial.", *Int J Radiat Oncol Biol Phys*, vol. 39, no. 2, pp. 371–80, Sep 1997.

[106] J G Sled, A P Zijdenbos, A C Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data.", *IEEE Trans Med Imag*, vol. 17, pp. 87–97, Feb. 1998.

[107] P Sminia, J van der Zee, J Wondergem, J Haveman, "Effect of hyperthermia on the central nervous system: a review", *Int J Hyperthermia*, vol. 10, no. 1, pp. 1–30, Feb. 1994.

[108] A Sotiras, C Davatzikos, N Paragios, "Deformable Medical Image Registration: A Survey", *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, July 2013.

[109] M Staring, S Klein, J P W Pluim, "A rigidity penalty term for nonrigid registration", *Med Phys*, vol. 34, no. 11, pp. 4098–108, Nov. 2007.

[110] M Staring, S Klein, J H C Reiber, W J Niessen, B C Stoel, "Pulmonary Image Registration with elastix using a Standard Intensity-Based Algorithm", in *MICCAI Workshop: Medical Image Analysis For The Clinic - A Grand Challenge*, pp. 73–79, 2010.

[111] C Studholme, D L G Hill, D J Hawkes, "An overlap invariant entropy measure of 3D medical image alignment", *Pattern Recogn*, vol. 32, pp. 71–86, Jan. 1999.

[112] T Sugahara, J van der Zee, H H Kampinga, Z Vujaskovic, M Kondo, T Ohnishi, G Li, H J Park, D B Leeper, V Ostapenko, E A Repasky, M Watanabe, C W Song, "Kadota Fund International Forum 2004. Application of thermal stress for the improvement of health, 15-18 June 2004, Awaji Yumebutai International Conference Center, Awaji Island, Hyogo, Japan. Final report", *Int J Hyperthermia*, vol. 24, no. 2, pp. 123–40, Mar. 2008.

[113] B Taylor & C Kuyatt, *Guidelines for evaluating and expressing the uncertainty of NIST measurement results*, NIST Technical Note 1297, 1994.

[114] D N Teguh, P C Levendag, P W J Voet, A Al-Mamgani, X Han, T K Wolf, L S Hibbard, P Nowak, H Akhiat, M L P Dirkx, B J M Heijmen, M S Hoogeman, "Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck", *Int J Radiat Oncol Biol Phys*, vol. 81, no. 4, pp. 950–7, Nov. 2011.

[115] P Thévenaz & M Unser, "Optimization of mutual information for multiresolution image registration.", *IEEE Trans Image Process*, vol. 9, pp. 2083–99, Jan. 2000.

[116] P Togni, Z Rijnen, W C M Numan, R F Verhaart, J F Bakker, G C van Rhoon, M M Paulides, "Electromagnetic redesign of the HYPERcollar applicator: toward improved deep local head-and-neck hyperthermia", *Phys Med Biol*, vol. 58, no. 17, p. 5997, Sept. 2013.

[117] L M Toonkel, K Soila, D Gilbert, J Sheldon, "MRI assisted treatment planning for radiation therapy of the head and neck.", *Magn Reson Imaging*, vol. 6, pp. 315–9, 1988.

[118] N J Tustison, P A Cook, J C Gee, "N4ITK: improved N3 bias correction.", *IEEE Trans Med Imaging*, vol. 29, no. 6, pp. 1310–20, 2010.

[119] J K Udupa, "On Standardizing the MR Image Intensity Scale", *Magn Res in Med*, vol. 42, pp. 1072–81, 1999.

[120] M Unser, "Splines: A perfect fit for signal and image processing", *IEEE Signal Process Mag*, vol. 16, pp. 22–8, 1999.

[121] R Valdagni & M Amichetti, "Report of long-term follow-up in a randomized trial comparing radiation therapy and radiation therapy plus hyperthermia to metastatic lymph nodes in stage IV head and neck patients", *Int J Radiat Oncol Biol Phys*, vol. 28, no. 1, pp. 163–9, Jan. 1994.

[122] F van der Lijn, M de Bruijne, S Klein, T den Heijer, Y Y Hoogendam, A van der Lugt, M M B Breteler, W J Niessen, "Automated brain structure segmentation based on atlas registration and appearance models", *IEEE Trans Med Imag*, vol. 31, no. 2, pp. 276–86, Feb. 2012.

[123] F van der Lijn, T den Heijer, M M B Breteler, W J Niessen, "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts", *NeuroImage*, vol. 43, no. 4, pp. 708–20, Dec. 2008.

[124] J van der Zee, D González González, G C van Rhoon, J D van Dijk, W L van Putten, A A Hart, "Comparison of radiotherapy alone with radiotherapy plus hyperthermia in locally advanced pelvic tumours: a prospective, randomised, multicentre trial. Dutch Deep Hyperthermia Group.", *Lancet*, vol. 355, no. 9210, pp. 1119–25, Apr 2000.

[125] G M Verduijn, L W Bartels, C P J Raaijmakers, C H J Terhaard, F A Pameijer, C A T van den Berg, "Magnetic resonance imaging protocol optimization for delineation of gross tumor volume in hypopharyngeal and laryngeal tumors.", *Int J Radiat Oncol Biol Phys*, vol. 74, pp. 630–6, June 2009.

[126] R F Verhaart, V Fortunati, G M Verduijn, A van der Lugt, T van Walsum, J F Veenland, M M Paulides, "The relevance of MRI for patient modeling in head and neck hyperthermia treatment planning: A comparison of CT and CT-MRI based tissue segmentation on simulated temperature", *Med Phys*, vol. 41, no. 12, p. 123302, Dec. 2014.

[127] R F Verhaart, V Fortunati, G M Verduijn, van Walsum T, J F Veenland, M M Paulides, "CT-based patient modeling for head and neck hyperthermia treatment planning: Manual versus automatic normal-tissue-segmentation.", *Radiother Oncol*, vol. 111, pp. 158–63, Mar. 2014.

[128] R F Verhaart, Z Rijnen, V Fortunati, G M Verduijn, T van Walsum, J F Veenland, M M Paulides, "Temperature simulations in hyperthermia treatment planning of the head and neck region: rigorous optimization of tissue properties", *Strahlenther Onkol*, vol. 190, no. 12, pp. 1117–1124, Nov. 2014.

[129] P1 W J Voet, M L P Dirkx, D N Teguh, M S Hoogeman, P C Levendag, B J M Heijmen, "Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis.", *Radiother Oncol*, vol. 98, pp. 373–7, Mar. 2011.

[130] H Wang, J W Suh, S R Das, "Multi-atlas segmentation with joint label fusion", *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 3, pp. 611–29, June 2012.

[131] X-F Wang, D-S Huang, H Xu, "An efficient local ChanVese model for image segmentation", *Pattern Recognition*, vol. 43, no. 3, pp. 603–18, Mar. 2010.

[132] S K Warfield, K H Zou, W M Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation", *IEEE Trans Med Imag*, vol. 23, no. 7, pp. 903–21, July 2004.

[133] G J Webster, J E Kilgallon, K F Ho, C G Rowbottom, N J Slevin, R I Mackay, "A novel imaging technique for fusion of high-quality immobilised MR images of the head and neck with CT scans for radiotherapy target delineation.", *Brit J Radiol*, vol. 82, pp. 497–503, 2009.

[134] P Wust, J Nadobny, M Seebass, D Stalling, J Gellermann, H C Hege, P Deuflhard, R Felix, "Influence of patient models and numerical methods on predicted power deposition patterns.", *Int J Hyperthermia*, vol. 15, no. 6, pp. 519–40, 1999.

[135] P Zhang & T F Cootes, "Automatic Construction of Parts + Geometry Models for Initializing Groupwise Registration", *Trans Med Imag*, vol. 31, no. 2, pp. 341–58, 2012.

Samenvatting

De behandeling van tumoren door verwarming (hyperthermie) in het hoofd-hals (HH) gebied vereist een nauwkeurige planning. Deze planning is gebaseerd op driedimensionale (3D) modellen van de patint, die zijn afgeleid van beelddata verkregen met bijvoorbeeld MRI en CT. De nauwkeurigheid van deze modellen is van invloed op de beslissingen aangaande de behandeling, en ook op het bepalen van de optimale behandeling. In dit proefschrift onderzoeken we methoden die automatisch 1) de anatomie van het HH gebied segmenteren uit CT- en MR-beelden, en 2) CT- en MR-beelden registreren om MRI in de planning van de behandelingsworkfklow te integreren.

In hoofdstuk 2 presenteerden en evalueerden we een automatisch segmentatie-algoritme voor CT-beelden van het HH gebied. Dit algoritme combineert anatomische informatie, die gebaseerd is op atlasregistratie, met lokale intensiteitsinformatie in een graaf-gebaseerde segmentatiemethode. In deze benadering worden voor elk weefsel twee modellen gebouwd: een intensiteitsmodel en een ruimtelijk model. Het intensiteitsmodel wordt gebouwd op basis van de histogrammen van de intensiteiten van de voorgrond- en de achtergrond-gebieden; de histogrammen worden bepaald door alle atlassen te bemonsteren. Het ruimtelijke model wordt verkregen door elk atlasbeeld te registreren met het beeld van de patint en dan de registratieresultaten te combineren in een ruimtelijk waarschijnlijkheidsbeeld. In het graaf-gebaseerde algorithme wordt ook nog een regularisatie term gebruikt die gladde segmentaties bevordert. Deze drie termen worden vervolgens gebruikt om een graaf te bouwen die wordt opgelost met behulp van de graaf-snede-methode. Eerst hebben we de eigenschappen van de voorgestelde aanpak gevalueerd ten opzichte van de multi-atlas gebaseerde segmentatie. Vervolgens hebben we de invloed van verschillen in de handmatige constructie van de atlassen op de nauwkeurigheid van de methode bepaald en vergeleken met de multi-atlas gebaseerde segmentatie en de interobserver variabiliteit. Als gouden standaard werd een dataset van 18 met de hand gesegmenteerde CT-beelden gebruikt. De Dice similarity coefficint (DSC), de Mean Surface Distance (MSD), en de Hausdorff surface distance (HSD) werden gebruikt als evaluatiematen. Met behulp van dertien gesegmenteerde afbeeldingen werd de invloed van de handmatige atlas-constructie op de nauwkeurigheid van de methode gekwantificeerd en vergeleken met de interobserver variabiliteit. De voorgestelde methode had een statistisch significant betere score voor DSC, MSD en HSD dan de multi-atlas-gebaseerde methode voor het glasachtig lichaam in het oog. Gemeten aan de hand van de DSC presteerde de voorgestelde methode significant beter voor de segmentatie van alle weefsels, behalve

de hersenstam en het ruggenmerg. De MSD toonde een significante verbetering voor de optische zenuw, het glasachtig lichaam, de lens, en de schildklier. Gemeten aan de hand van de HSD werkte de voorgestelde methode significant beter voor het glasachtig lichaam en de hersenstam. Voor de meeste weefsels (8 van de 11) benaderde de segmentatie-nauwkeurigheid van de voorgestelde methode de interobserver overeenkomst. Bovendien kende onze methode een betere robuustheid voor variaties in atlassegmentatie in vergelijking met de multi-atlassegmentatie. We concluderen dat de voorgestelde methode HH weefsels in CT-beelden voldoende nauwkeurig kan segmenteren voor het genereren van 3D patint-modellen, en dat daarmee de inspanning benodigd voor therapieplanning substantieel kan verminderen.

In hoofdstuk 3 hebben we de invloed van het gebruik van handmatige en automatische segmentatie gevalueerd op de behandelingsplanning voor hyperthermie (HTP). Om te beslissen of de automatische segmentatiemethode kan worden ingevoerd in de kliniek, heben we het effect van het gebruik van deze methode op de HT dosisverdeling en HTP-kwaliteit gevalueerd. Het effect van het gebruik van automatische in plaats van handmatige segmentatie werd vergeleken en daarnaast ook de invloed van andere bronnen van onzekerheden in het patint-model, nl. de stapgrootten tijdens de berekening en de dilektrische weefseleigenschappen. CT-beelden van zeven patinten werden automatisch en handmatig gesegmenteerd door vier experts, en hiermee werd de inter- en intraobserver variatie bepaald. Ondanks de verschillen in de segmentaties resulteerde de handmatige en automatische gegenereerde 3D patint-modellen in een vergelijkbare HTP kwaliteit (1 % variatie). Deze variatie was bovendien klein ten opzichte van de totale variatie ten gevolge van andere bronnen van patint-model onzekerheden (11,7%). Ook werd er een aanzienlijke vermindering van de operatortijd per patint bereikt bij automatische segmentatie (1 uur rekentijd + 15 minuten operator-tijd) ten opzichte van handmatige segmentatie (5-8 uur). We kunnen dus concluderen dat de recent ontwikkelde automatische segmentatiemethode kan worden gentroduceerd in de kliniek voor HH HTP.

In hoofdstuk 4 presenteerden we een automatische segmentatiemethode voor MR-beelden. Door gebruik te maken van MR-beelden kan een meer nauwkeurige afbakening van zowel normaal [126] als tumorweefsel [2, 45, 96, 117] verkregen worden. In dit hoofdstuk werd de methode zoals gepresenteerd in hoofdstuk 2 uitgebreid met een meer geavanceerde atlasfusie methode om het ruimtelijk model te verbeteren. Verschillende atlasfusie benaderingen, die gebruik maken van lokale eigenschappen van het beeld, werden vergeleken. Net als in de vorige hoofdstukken werden een intensiteitmodel en een regularisatie component gecombineerd met het ruimtelijke model in een graaf, waarna de segmentatie met behulp van de graaf-snede methode werd bepaald. De voorgestelde methode was erg nauwkeurig, zowel in termen van de Dice similarity coefficint (DSC) als de Mean Surface Distance (MSD) met een mediane DSC hoger dan 0.8 voor alle weefsels behalve de harde oogrok, en een mediane MSD lager dan 1 mm voor alle weefsels behalve de liquor (CSF). Voor alle weefsels met uitzondering van de ruggengraat weefsels (ruggemerg en CSF) benaderde de nauwkeurigheid van de methode de interobserver-overeenstemming / variabiliteit, zowel in termen van DSC als van MSD. Het positieve effect van het toevoegen van de intensiteit modellering daalde wanneer een meer nauwkeurige atlasfusiemethode werd gebruikt. Met behulp van de voorgestelde aanpak hebben we de prestaties van de aanpak gepresenteerd in

hoofdstuk twee voor HH hyperthermie behandelingsplanning verbeterd, en kunnen we concluderen dat de methode voldoende nauwkeurig is voor klinische toepassing.

In hoofdstuk 5 onderzochten we de haalbaarheid van het gebruik van elastische registraties in de klinische praktijk om MR en CT-beelden van het HH gebied zo exact mogelijk spatieel in correspondentie te brengen zodat ze gecombineerd gebruikt kunnen worden in de behandelingsplanning. Een elastische registratie-algoritme werd geoptimaliseerd, gevalueerd en vergeleken met rigide registratie, de huidige klinische standaard. De evaluatie was gebaseerd op 21 anatomische herkenningpunten en aandachtsgebieden (ROIs) die in beide type beelden handmatig werden geplaatst door een expert. Bij de keuze van herkenningspunten en ROIs werd rekening gehouden met beide toepassingen (RTP en HTP), de zichtbaarheid van de herkenningspunten en van de ROIs in beide typen afbeeldingen en de spreiding van de herkenningspunten en de ROIs over het totale volume. We ontwikkelden ook een multiparametrische registratie-aanpak, die tegelijkertijd T1- en T2-gewogen MR sequenties registreert met CT. Deze multiparametrische aanpak werd gevalueerd en vergeleken met de methode die slechts n van de MRI sequenties gebruikt. Uit zowel de herkenningspunt gebaseerde als de ROI-gebaseerde evaluatie bleek dat elastische registratie nauwkeuriger was dan rigide registratie. Ook resulteerde de elastische deformatie niet in onrealistische vervormingen. De goede nauwkeurigheid van de registratie gemeten met de herkenningspunten en met de ROIs gecombineerd met de gladheid van de vervorming toont aan dat de elastische registratie een betere spatile correspondentie geeft dan rigide registratie in het gehele volume. De registratiefout was kleiner dan de voxel resolutie voor de landmark gebaseerde (mediane fout ongeveer 1,6 mm) en de ROI-gebaseerde evaluatie (gemiddelde MSD van ongeveer 1 mm voor alle ROIs exclusief het cerebellum). Vergelijkbare nauwkeurigheden werden verkregen voor de n- en multiparametrische benaderingen. We concluderen daarom dat elastische registratie van HH CT en MR-beelden haalbaar is met een registratiefout kleiner dan de huidige klinische norm .

In hoofdstuk 6 onderzochten we of elastische registratie van CT en MR beelden het gebruik van patint-immobilisatie-apparatuur tijdens het maken van de beelden kan vervangen: het gebruik van de apparatuur kost tijd en is oncomfortabel voor de patint. CT- en MR- beelden van 22 patinten werden geregistreerd met behulp van state-of-the-art rigide en elastische registratie-algoritmes. Voor tien patinten werden T2-gewogen MR-beelden verkregen met standaard hoofd-spoelen in twee opstellingen: zonder immobilisatie-masker en met een kleine immobilisatie-masker, afgeleid van het originele radiotherapie immobilisatie-masker en passend binnen de hoofd-spoel. Het kleine masker werd gebruikt om de patinten in de behandelingspositie te fixeren. Aangezien bij het gebruik van de hoofd-spoel er geen hoofdsteun en geen vlakke tafel gebruikt kon worden was de MRI patint-positionering anders dan bij de CT positionering. Voor 12 patinten werden beelden verkregen met behulp van 6 kanaals oppervlakte-spoelen en met de originele radiotherapie immobilisatie-apparatuur, d.w.z. het radiotherapie masker, de hoofdsteun en een vlakke tafel. De nauwkeurigheid van de registratie werd gevalueerd met behulp van 21 herkenningspunten, die door een expert in de CT- en MR-afbeeldingen aangegeven waren. De evaluatie en optimalisatie van de parameters van de rigide en de elastische registraties werden uitgevoerd met behulp van leave-one-out experimenten. De beste beeldregistratie

werd bereikt wanneer de patint werd gemmobiliseerd met behulp van het oorspronkelijke RT masker en wanneer elastische registratie werd gebruikt (mediaan van registratiefout 1,63 mm). In het algemeen werd een betere registratie verkregen wanneer immobilisatie-apparatuur werd gebruikt. De elastische registratie resulteerde in een betere spatile correspondentie dan de rigide registratie; dit was in overeenstemming met de resultaten van hoofdstuk 5. Het bleek echter ook dat het gebruik van immobilisatie-apparatuur noodzakelijk bleef om de beste nauwkeurigheid te bereiken; elastische registratie kan de fouten die ontstaan door geen gebruik te maken van immobilisatie-apparatuur niet volledig compenseren. Omdat elastische registratie van beelden verkregen met behulp van de beste patint immobilisatie-apparatuur de meest nauwkeurige match geeft raden wij deze combinatie te gebruiken in de klinische praktijk.

In dit proefschrift hebben we methoden voor de automatische segmentatie en de automatische elastische registratie van CT- en MR-beelden voor HTP gepresenteerd en gevalueerd. De segmentatiemethoden werken met een nauwkeurigheid die de interobserver variabiliteit benadert. Ook hebben we aangetoond dat de interobserver variabiliteit een geringe invloed heeft op HTP kwaliteit. De elastische registratiemethode zorgt voor een betere nauwkeurigheid dan de huidige klinische standaard zonder dat er niet-realistische vervormingen ontstaan. Daarom is de methode zoals voorgesteld in dit proefschrift klaar voor klinische implementatie.

In de eerste drie hoofdstukken van dit proefschrift laten we zien dat de automatische segmentatie van CT en MR beelden van het HH gebied mogelijk is met een nauwkeurigheid vergelijkbaar met handmatig intekenen. De methode beschreven in hoofdstuk 2 werd gebruikt om volledige 3D-patint-modellen te genereren uit CT, zoals beschreven in hoofdstuk 3. Deze methode is momenteel in gebruik voor HTP in de kliniek, waarbij de resultaten van deze methode nog altijd geverifieerd worden door een expert. Tot nu toe is gebleken dat alleen kleine correcties nodig zijn, met name voor de ogen, nek weefsels en bij de oren en het gebied bij de neus. De arbeidstijd van ca. 5-8 uur voor handmatig intekenen werd significant teruggebracht tot maximaal 1 uur voor correcties. De methode kan worden uitgebreid met de aanpassingen zoals die in hoofdstuk 4 beschreven, zodat de segmentatie van MR-beelden meegenomen kan worden om de nauwkeurigheid van het patint-model te verhogen [126]. CT-beelden blijven nodig voor de segmentatie van bepaalde weefsels, zoals spier, vet, bot, kraakbeen, de schildklier, en de optische zenuw. Met behulp van MRI kunnen meer weefsels, zoals de witte en grijze stof van het hersenweefsel en de liquor, worden opgenomen in het patint-model. Het is aangetoond dat HTP-kwaliteit significant benvloed wordt door een meer gedetailleerde segmentatie [126].

De methode die we in hoofdstuk 4 presenteerden is nog niet toegepast op CT-beelden. Wij denken dat de positieve resultaten verkregen op de MR-beelden door het gebruik van lokaal gewogen atlasfusie, ook zullen gelden voor CT-beelden: we voorzien geen moeilijkheden of beperkingen van het model met betrekking tot de afbeeldingsmodaliteit. Integendeel, juist omdat CT een kwantitatieve intensiteitsschaal heeft (Hounsfield Unit) is het te verwachten dat het intensiteitsmodel effectiever is in het verbeteren van de nauwkeurigheid bij CT- dan bij MR-beelden. Deze methode kan gebruikt worden om de CT-gebaseerde patint-modellen te verbeteren. Onze me-

thode voor automatische segmentatie van het HH gebied werd gevalueerd met behulp van weefsels relevant voor HTP. De meeste van de gesegmenteerde weefsels zijn ook relevant voor radiotherapie behandeling planning (RTP). Onze benadering kan eenvoudig worden uitgebreid met de segmentatie van additionele weefsels relevant voor RTP, bijv. de oorspeekselklier en de submandibulaire klieren, door de atlassen uit te breiden met deze weefsels en de methode-parameters voor deze weefsels te optimaliseren. Deze uitbreiding kan met name interessant zijn in verband met het ontbreken van specifieke automatische segmentatiemethodes voor HH beelden die verkregen zijn voor radiotherapie doeleinden.

In het tweede deel van het proefschrift hebben we ons gericht op het ontwikkelen van een strategie voor het verkrijgen van een nauwkeurige registratie van CT- en MR- beelden. De registratie van CT- en anatomische MR-scans (T1 of T2-gewogen) kan ook worden gebruikt om informatie over de bloedvaten, verkregen uit contrast MR, te integreren in HTP. Aangezien de doorbloeding een belangrijke factor is voor de temperatuur (her)verdeling [74], kan het toevoegen van een patint specifiek vaatmodel aan het huidige thermisch model [126] de nauwkeurigheid van de thermische simulatie van HTP in het HH gebied verhogen [69].

Concluderend kan de automatische segmentatie van CT- en MR-beelden voor het genereren van patint-modellen in HTP worden uitgevoerd met een hoge nauwkeurigheid, die vergelijkbaar is met de nauwkeurigheid van handmatig intekenen. Daarnaast kunnen MR-beelden met CT worden geregistreerd met een hoge nauwkeurigheid als gebruik gemaakt wordt van patint immobilisatie-apparatuur tijdens het verkrijgen van de beelden. Wij concluderen daarom dat het gebruik van automatische MR- en CT-gebaseerde patint-modellen voor HTP klaar is voor klinische toepassing.

PhD Portfolio

Research

- PhD period: 2011-2014
- Departments of Radiology and Medical Informatics.
- ASCI research school.

In-Depth Courses:

- MRI in Radiotherapy, ASCI, 2011.
- Advanced Pattern Recognition, ASCI, 2011.
- International Computer Vision (ICV) Summer School, 2011.
- European Institute for Biomedical Imaging Research (EIBIR) Summer School, 2011.
- Knowledge Driven Image Segmentation, ASCI, 2012.
- Measuring Features, ASCI, 2012.
- Presentation Course, EMC, 2012.
- Biomedical English Writing Course, EMC, 2013.

International conferences (poster/presentation):

- IEEE International Symposium on Biomedical Imaging, Barcelona (ISBI), 2012 (Poster).
- Annual Meeting of the European Society for Hyperthermic Oncology (ESHO), 2013 (Poster).
- Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), 2013 (Presentation).
- Annual Meeting of the European Society for Radiotherapy & Oncology (ESTRO), 2014 (Presentation).

Other:

- Atendance at the workshop on open source medical image analysis software (ISBI), 2012.
- Presentations (3x) at the annual meeting of Cardiovascular Risk Management by Advanced Medical Image Analysis (CARISMA), 2012-2014.
- Presentation at the Medical Imaging Symposium for PhD students (MISP), 2014.
- Attendance at the Medical Imaging Symposium for PhD students (MISP), 2011,2013,2014.
- Presentation at Dutch Biomedical Engineering conference (BME), 2013.
- Presentations (3x) at the department seminar of the Biomedical Imaging Group Rotterdam, 2011-2014.
- Presentations (4x) at the Medical Informatics Research Lunch 2011-2014.
- Attendance at Erasmus MC PhD Day 2014.
- Reviewing for Computer Vision and Image Understanding and for the British Journal of Radiology.

Publications

Publications in International Journals:

- **V Fortunati**, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "Tissue segmentation of head and neck CT images for treatment planning: a multi-atlas approach combined with intensity modeling", *Med Phys*, vol. 40, no. 7, pp. 071905, Jul. 2013.

- R F Verhaart, **V Fortunati**, G M Verduijn, T van Walsum, J F Veenland and M M Paulides, "CT-Based Patient Modelling for Head and Neck Hyperthermia Treatment Planning: Manual versus Automatic Normal-Tissue-Segmentation", *Radiother Oncol*, vol. 111, no. 1, pp. 158-63, Apr. 2014.

- **V Fortunati**, R F Verhaart, F Angeloni, A van der Lugt, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "Feasibility of multi-modal deformable registration for head and neck tumor treatment planning", *Int J Radiat Oncol Biol Phys*, vol. 90, no. 1 pp. 85-93, Sep. 2014.

- R F Verhaart, Z Rijnen, **V Fortunati**, G M Verduijn, J F Veenland, T van Walsum and M M Paulides, "Temperature simulations in head and neck hyperthermia treatment planning: rigorous tissue property optimization", *Strahlenther Onkol*, vol. 190, no. 12, pp. 1117-24, Nov. 2014.

- R F Verhaart, **V Fortunati**, G M Verduijn, A van der Lugt, T van Walsum, J F Veenland and M M Paulides, "The relevance of MRI for patient modeling in head and neck hyperthermia treatment planning: A comparison of CT and CT-MRI based tissue segmentation on simulated temperature", *Med Phys*, vol. 41, no. 12, pp. 123302, Dec. 2014.

- F Adibzadeh, R F Verhaart, G M Verduijn, **V Fortunati**, Z Rijnen, G C van Rhoon and M M Paulides, "Association of acute adverse effects with high local SAR induced in the brain from prolonged RF head and neck hyperthermia", *Phys Med Biol*, vol. 60, no. 3, pp. 995-1006, Feb. 2015.

- A S Korsager, **V Fortunati**, F van der Lijn, J Carl, W J Niessen, J Ostergaard and T van Walsum, "The use of atlas registration and graph cuts for prostate segmentation in magnetic resonance images", *Med Phys*, vol. 42, pp. 1614, Mar. 2015.

- **V Fortunati**, R F Verhaart, G M Verduijn, A van der Lugt, F Angeloni, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "MRI integration into treatment planning of head and neck tumors: can patient immobilization be avoided?", *Radiother Oncol*, in press.

Publications in International Conference Proceedings:

- **V Fortunati**, R F Verhaart, F van der Lijn, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "Hyperthermia Critical Tissues Automatic Segmrntation of Head and Neck CT images using atlas registration and Graph Cuts" *Proceedings of International Symposium on Biomedical Imaging: from Nano to Macro*, 2012.

Conference Abstracts:

- R F Verhaart, **V Fortunati**, T van Walsum, J F Veenland, A van der Lugt, G C van Rhoon, W J Niessen, M M Paulides. "Accurate Patient Modeling for Planning-Guided Deep Head and Neck Hyperthermia: Variation in Tissue Delineation", *Molecular Medicine Day*, 2012.

- **V Fortunati**, R F Verhaart, W J Niessen, J F Veenland, M M Paulides and T van Walsum, Critical Tissues Segmentation of Head and Neck CT Images for Hyperthermia Treatment Planning, *Dutch Bio-Medical Engineering Conference (BME)*, 2013.

- R F Verhaart, **V Fortunati**, J Veenland, T van Walsum, W J Niessen, G van Rhoon, M M Paulides. "Sensitivity of head and neck hyperthermia treatment planning to observer variations in 3D patient model generation" *Dutch Bio-Medical Engineering Conference (BME)*, 2013.

- **V Fortunati**, R F Verhaart, W J Niessen, G C Van Rhoon, J F Veenland, M M Paulides and T van Walsum, "Automatic 3D patient model generation for head and neck hyperthermia treatment planning", *Annual Meeting of the European Society for Hyperthermic Oncology (ESHO)*, 2013.

- R F Verhaart, G M Verduijn, **V Fortunati**, G C van Rhoon, J F Veenland, M M Paulides. "Importance of Bone Segmentation for head and neck hyperthermia treatment planning" *Annual Meeting of the European Society for Hyperthermic Oncology (ESHO)*, 2013.

- **V Fortunati**, R F Verhaart, F Angeloni, W J Niessen, A van der Lugt, M M Paulides and T van Walsum, "Rigid and non-rigid registration for integrated MR and CT-based head and neck radiotherapy and hyperthermia treatment planning", *Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, 2013

- **V Fortunati**, R F Verhaart, F Angeloni, A van der Lugt, W J Niessen, J F Veenland, M M Paulides and T van Walsum, "MRI integration into treatment

planning of the head and neck: can we avoid patient immobilization?", *Annual meeting of the European society for radiotherapy and oncology (ESTRO)*, 2014.

- R F Verhaart, **V Fortunati**, F Adibzadeh, G M Verduijn, J F Veenland, T van Walsum, M M Paulides. "Patient modelling in head and neck hyperthermia treatment planning: is CT-based auto-segmentation sufficient?", *Annual meeting of the European society for radiotherapy and oncology (ESTRO)*, 2014.
- R F Verhaart, Z Rijnen, **V Fortunati**, G M Verduijn, T van Walsum, J F Veenland, M M Paulides, "Towards temperature-based hyperthermia treatment planning: a technique for tissue property optimization and its robustness for probe-location", *Annual Meeting of the European Society for Hyperthermic Oncology (ESHO)*, 2014.

Acknowledgment

Concordia parvae res crescunt discordia maximae dilabuntur - (where there is harmony, small things grow; where there is disharmony, very great things decay.)

— Sallust (86–34 B.C.)

About four years ago I started my Ph.D. trajectory as a researcher at Erasmus MC, University Medical Center Rotterdam (EMC). During these years the Biomedical Imaging Group Rotterdam (BIGR) was my second home. I would like to thank all the BIGR members, and the people of the departments of Medical Informatics and Radiology that made my working days pleasant and enjoyable.

I would like to acknowledge my promotor Prof.dr. Wiro Niessen and my co-promotors Dr.ir Theo van Walsum and Dr. Jifke Veenland, for offering me a Ph.D. position in this nest of knowledge and experience that is the BIGR group. Wiro you are the best possible *boss* that any human mind could imagine. You motivated me to improve as a scientist and as a person, always smiling and full of curiosity. Fortunately I can still enjoy your scientific guidance at Quantib. Theo and Jifke, I was lucky to have you both as my daily supervisors. Your expertise and your different points of view made my research complete. Your patience and dedication as much as your competence made me honored every day of being your student.

The inner committee members Prof.dr.G.C. van Rhoon, Prof.dr.ir. J.P.W. Pluim, Prof.dr. B.W. Raaymakers; the plenary committee members Prof.dr. A. van der Lugt and Prof.dr. B.J.M. Heijmen; the expert M.M. Paulides; thank you for accepting to be in my doctoral committee and attend to my defense.

My Ph.D. project was carried out in close collaboration with the hyperthermia Unit, part of the Radiotherapy Department of the EMC. A large contribution on this thesis was given by Dr.ir Maarten Paulides and Ir. René Verhaart who is also my paranymph. Maarten, thank you for your very valuable contribution and precious suggestions that enhanced the link between my work and clinical hyerthermia. René we started our Ph.D. together and we collaborated in almost everything we have done during these years. I thank you for your availability, for your pragmatism and for pushing me to get things done. I wish you all the best with finishing your thesis.

A special thank goes also to my other paranymph Ihor Smal who helped me a lot with everything. First as my roommate than as my neighbor. Whatever issue I

had, Ihor could always help me find a solution. It was a pleasure walking with you everyday to EMC for the last two yeas and I miss that. I love discussing (and arguing) with you about all sort of things, from science, to religion, to the ingredients of Pizza.

Dr.ir. Stefan Klein thank you a lot for the huge help and guidance you gave me with all the registration related issues I found during my research. Marcel, Henry, Hakim, Fedde, Pierre, Andres, Jyotirmoy, Karin, Esther, Esben, Coert, Reinhard, Renske, Adrienne, Dirk, Marius, Marco, Francesco, Guillaume, Adria, Diego, Gerardo, Hortense, Oleh, Azadeh, Marleen, Eric, Jean-Marie, Wyke, Carolyn, Eugene, Luu, Annegreet, Gennady, Miroslav, Emilie, Wei, Gijs, Rozanna, Rebecca, and Anouk thank you all for the great time we had during the department outings, drinks, parties, Ph.D. dinners, concerts, conferences and the every day life. Thanks also go to all the staff of department of Medical Informatics and Radiology especially to Desiree de Jong and Petra 't Hart-Assems for taking care of all the bureaucratic issues during my appointment at EMC. Desiree, thank you also for your help with the printing of the thesis and the organization of the defense.

Vorrei ringraziare tutta la mia famiglia ed i miei amici per il loro affetto e supporto. Grazie a mia moglie Valentina e mio figlio Enea, siete la cosa piú importante che quest vita mi abbia dato. Grazie ai miei genitori Fabio e Noemi per avermi sostenuto in tutte le mie scelte. Grazie a Daniele che ha disegnato la copertina di questo volume ed ha tutti i miei amici di Roma. Infine grazie a tutti gli amici Italiani in Olanda con cui ho condiviso momenti meravigliosi in questi anni nel paese di tulipani, mulini a vento, e bitterballen.

Valerio Fortunati
Rotterdam, April 2015

Curriculum Vitae

Valerio Fortunati was born in Rome in 1983. After high school he has started studying Electronic Engineering at the university of Roma Tre in Rome. He has taken his bachelor's degree in 2007 presenting a project about biometrical secure recognition of users through dynamic signature detection.

After that he followed the same study path as an MSc student and he graduated in 2010. He started to improve his knowledge about image processing during his master's project called "Texture synthesis using local descriptors". This project has been developed at the University of Groningen(RuG) in the Netherlands during an abroad study period of 8 months.

He worked from January 2011 to December 2014 as a Ph.D. student at the Biomedical Imaging Group Rotterdam involved in the project: "Hyperthermia treatment planning for head and neck cancer: improving patient models using advanced MRI characteristics". His work consisted of the development of automatic methods for patient-models construction from CT and MR images for Hyperthermia treatment planning.

Since January 2015 he is working at Quantib B.V. as research and development engineer.