UNIVERSITY OF NICE - SOPHIA ANTIPOLIS DOCTORAL SCHOOL STIC SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

PHD THESIS

to obtain the title of

PhD of Science

of the University of Nice - Sophia Antipolis Specialty : CONTROL, SIGNAL AND IMAGE PROCESSING

> Defended by Ezequiel GEREMIA

Spatial Random Forests for Brain Lesions Segmentation in MRIs and Model-based Tumor Cell Extrapolation

Thesis Advisor: Nicholas AYACHE prepared at Inria Sophia Antipolis, ASCLEPIOS Team defended on January 30th, 2013

Jury :

Reviewers:	Guido Gerig	-	University of Utah
	Koen VAN LEEMPUT	-	Technical University of Denmark
Advisor :	Nicholas Ayache	-	INRIA
Co- $Advisor$:	Hervé Delingette	-	INRIA
President :	Gábor Székely	-	ETH Zürich
Examinators :	Antonio Criminisi	-	Microsoft Research Cambridge
Invited :	Björn Menze	-	ETH Zürich

Financial support

This work was carried out thanks to the support of the MedYMA advanced grant 291080 funded by the European Research Council (ERC).

Aknowledgements

I would like to warmly thank Nicholas Ayache for giving me the opportunity to work with him. Thank you for all your support and patience throughout these four years.

I thank Olivier Clatz, Hervé Delingette and Ender Konukoglu for their inspiring company. Your enthusiasm and precious advices have been a great asset for finding my way in the medical image community. Many thanks to Antonio Criminisi who received me at Microsoft Research. Thanks for enlightening the path with your didactic tricks and for making me feel at home in a foreign land. A huge thanks to Bjoern Menze for investing untold time and energy at my side. I am infinitely grateful for your constant effort in pushing my work forward even in times I doubted it. I also thank Marc-André Weber and Bram Stieltjes from the German Cancer Research Center (DKFZ) for sharing their databases.

I express my gratitude to the members of the jury committee for attending my PhD defence and the very interesting discussion we had after. Thank you Guido Gerig and Koen Van Leemput for accepting to be the reviewers of this thesis. Thank you for the time and attention you have devoted to this thesis and also for your valuable comments. I would also like to thank Gábor Székely for accepting to preside the jury, for his interest and for his encouraging comments.

Isabelle Strobant, thank you for all your help, support and friendly chats.

During my thesis, I had the chance to collaborate with exceptional people: Erin Stretton, Islem Rekik and Ján Margeta. Thanks for your friendship and all the transcendental things I learnt at your side. Adityo, it has been an honor sharing your office. Thanks for being there, my dear friend.

Thanks to all my comrades at Asclepios and MSR with whom I shared great moments and happy times. Thanks to the soccer team for the playful matches. Thanks to my solitary companions in the forest.

A huge hug to all my friends for the countless adventures. I kindly thank all my family in Argentina and cherish that invisible link that binds us. Thanks to my close family, Silvia, Roberto, Luciano, Giuliano and Carla for your love and for being a constant inspiration in my life. I tenderly thank you, Tamara, for fill my life with happiness.

All the best in your lives!

Contents

1	Intr	oduct	ion	1
	1.1	Conte	xt	1
	1.2	Contr	ibutions	2
	1.3	Relate	ed work	3
2	Spa	tial D	ecision Forests for Lesion Segmentation	
	in N	Aulti-O	Channel MRIs of the Brain	5
	2.1	Introd	luction	6
	2.2	Mater	rials	7
		2.2.1	MICCAI Grand Challenge 2008 dataset	7
		2.2.2	Data preprocessing	9
		2.2.3	Notations	10
	2.3	Metho	ds	10
		2.3.1	Context-rich decision forest	10
		2.3.2	Visual features	13
	2.4	Exper	iments and results	14
		2.4.1	Results on the public Grand Challenge dataset	15
		2.4.2	Results on the private Grand Challenge dataset	15
	2.5	Discus	ssion	16
		2.5.1	Interpreting segmentation results	16
		2.5.2	Influence of preprocessing	17
		2.5.3	Influence of forest parameters	17
		2.5.4	Analysis of feature relevance	21
	2.6	Concl	$usion \ldots \ldots$	26
3	Bra	in tun	nor cell density estimation from multi-modal MR images	
Ŭ	base	ed on a	a synthetic tumor growth model	37
	3.1	Introd	luction	37
	3.2	Learn	ing to estimate tissue cell density from synthetic training data	39
		3.2.1	Generative tumor simulation model	39
		3.2.2	Regression forests for estimating tissue cell density	41
	3.3	Exper	riments	43
		3.3.1	Experiment 1: Estimating cell density in synthetic cases	43
		3.3.2	Experiment 2: Segmenting tumors in clinical images	44
	3.4	Discus	ssion	44
	3.5	Concl	usions	54

4	\mathbf{Spa}	tially Adaptive Random Forests for Classification Problems in	
	Med	dical Imaging	61
	4.1	Introduction	61
	4.2	Data representation	63
		4.2.1 Multi-scale image tree	63
		4.2.2 Visual features	63
		4.2.3 Ground truth	64
	4.3	Spatially adaptive random forest	64
		4.3.1 Training data	67
		4.3.2 Decision node representation	67
		4.3.3 Training	67
		4.3.4 Prediction	68
	4.4	Experiments and results	68
	4.5	Discussion and conclusion	69
F	Com	alusians and Dependenting	71
Э	Con E 1	Conclusions and Perspectives	71
	0.1 ജ.റ		71 72
	0.2	Perspectives	10
Α	\mathbf{List}	of Publications	75
в	Lav	ered spatio-temporal forests for left ventricle segmentation from	
2	4D	cardiac MRI data	77
	B 1	Introduction	77
	B.2	Lavered spatio-temporal decision forests	78
		B.2.1 Strategy to learn from spatio-temporal data	79
		B 2 2 Features	79
		B 2 3 Data preprocessing	81
	B 3	First layer: Decision forests for image intensity standardization and	01
	D.0	position normalization	81
		B 3 1 Intensity standardization	82
		B 3 2 Orientation normalization	82
	B 4	Second layer: Learning to segment with the shape	83
	D.1	B 4.1 Using voyel coordinates	83
		B 4 2 Transforming the volumes back	84
	B 5	Results	84
	B.6	Conclusions	86
	D.0		00
С	Pre	dicting the Location of Glioma Recurrence After a Resection	
	Sur	gery	87
	C.1	Introduction	87
	C.2	Materials and Method	88
	C.3	Results	94
	C 1	Discussion and Conclusion	06

C.5 Acknowledgments	96
Bibliography	97

Contents

1.1	Context	1
1.2	Contributions	2
1.3	Related work	3

1.1 Context

How to make the most out of medical images to support diagnosis in the best possible way? Clinicians are confronted to this question every time they analyse an X-Ray, or browse a volumetric computed tomography (CT) or magnetic resonance (MR) scans. Current clinical protocols rely on extensive imaging modalities to drive the diagnosis. Because of the large size of acquired images, clinicians spend a long time parsing them before reaching the region of interest. Computer scientists and clinicians are collaborating to develop ergonomic tools that automatically extract the valuable information out of the image. This thesis is a humble contribution towards this goal.

The focus is set on the segmentation of brain lesions in multi-sequence MR images (MRIs). More specifically, we will focus on multiple sclerosis (MS) lesions, low and high grade gliomas. For these pathologies, however, visual assessment has important limitations. In general, this result from the fact that the underlying bio-physiological mechanisms, and its mapping to MR intensities, are not fully understood. However, MRIs play a critical role to localize lesions and monitor their evolution during treatment.

In clinical datasets, lesions are often delineated by clinicians for these reasons. However, annotating brain lesions is a tedious task which sometimes requires to make arbitrary decisions, even when using semi-automatic methods. Unfortunately, resulting segmentations are reliable but contain an unsuitable expert-driven bias. In the last years, the merged efforts of computer science and medical imaging communities opened the way to a novel approach for segmentation: supervised methods. Supervised methods make use of already exhaustive expert-annotated datasets to train classifiers to automatically solve task-specific segmentation problems. In this work, we focus on random decision forests.

Random forests are discriminative classifiers. They generate a hierarchical representation of the training data which is optimized for testing. Additionally, they provide the huge advantage of being highly parallelizable which lead to number of real-time applications. Unlike state-of-the-art classifiers, such as SVMs, the hierarchical compression of the training data is easy to interpret and provides the user with informative uncertainty measures on the segmentation results.

1.2 Contributions

In this thesis, we investigate three different problems related to brain lesions segmentation in MRIs using the powerful random forest framework. They are developed in chapters 2, 3 and 4, respectively:

- Spatial random forests for MS lesions segmentation [Geremia 2010, Geremia 2011]. First, we focus on segmentation of brain lesions which is an essential task to diagnosis, prognosis and therapy planning. A contextaware random forest is designed for the automatic multi-class segmentation of MS lesions in MR images. It uses multi-channel MRIs, prior knowledge on tissue classes, symmetrical and long-range spatial context to discriminate lesions from background. Quantitative evaluation was carried out on publicly available labelled datasets from the MICCAI 2008 MS Lesion Segmentation Challenge. It demonstrated state-of-the-art performance.
- Multi-variate regression forests for glioma cell density extrapolation [Geremia 2012b]. Then, we investigate the promising perspective of estimating the brain tumor cell density from MRIs. A generative-discriminative framework is presented to learn the latent and clinically unavailable tumor cell density from model-based estimations associated with synthetic MRIs. The generative model is a state-of-the-art publicly available biophysiological tumor growth simulator. The discriminative model builds on multi-variate regression random forests to estimate the voxel-wise distribution of the tumor cell density from input MRIs. The method was evaluated on a large dataset of synthetic cases and a reduced set of clinical cases from the German Center for Cancer Research (DKFZ) with promising results.
- Spatially Adaptive Random Forests for Classification Problems in Medical Imaging. Finally, we present the "Spatially Adaptive Random Forest" (SARF) which merges the benefits of multi-scale and random forest methods. Thanks to multi-scale data representation, the computation effort focuses on challenging regions of large medical volumes rather than uniformly processing the whole image. SARF is demonstrated in the context of multi-class gliomas segmentation in multi-modal MR images. Quantitative evaluation was carried out on publicly available labelled datasets from the MICCAI 2012 BRATS Challenge and demonstrated top segmentation results.

Recently, two articles inspired on the spatial random forest approach were published as part of the MICCAI 2012 Workshop on Multimodal Brain Tumor Segmentation Challenge [Menze 2012, Geremia 2012a]. This work lead to co-write a chapter with Darko Zikic on MS lesions and glioma segmentation [Criminisi 2013]. Chapter 4 was submitted for review to the ISBI conference [Geremia 2013b].

In the appendices, we present additional fruitful collaborations carried out during the time of the PhD but which go beyond the scope of this thesis. Appendice B reports a work by Ján Margeta et al. [Margeta 2011] on an adaptation of the spatial random forests for the segmentation of the left ventricle of the heart in 4D MRIs. Appendice C presents another work by Erin Stretton et al. [Stretton 2012] on how to predict glioma recurrence after resection surgery.

1.3 Related work

The motivation of this work arises from an inspirational bibliography which is overviewed here.

A large variety of methods were proposed for MS lesion. Generative methods were proposed consisting in a tissue classification by means of an expectation maximization (EM) algorithm. The EM algorithm can be modified to be robust against lesion affected regions, its outcome is then parsed in order to detect outliers which, in this case, coincide with MS lesions [Van Leemput 2001]. Another approach consists in adding to the EM a partial volume model between tissue classes and combining it with a Mahalanobis thresholding which highlights the lesions [Dugas-Phocion 2004]. Morphological postprocessing on resulting regions of interest was shown to improve the classification performance [Souplet 2008]. For multi-class glioma segmentation, a segmentation framework based on outlier detection was proposed to discriminate the tumor core from the edema, only using the T2-weighted MRI [Prastawa 2004]. More recently, segmentation approaches based on supervised random forests were applied to brain tissue segmentation in MRIs [Yi 2009], and delineation of the my-ocardium in real-time 3D echocardiography [Lempitsky 2009].

In the case of gliomas, segmentation results, however, give limited insight into the underlying bio-physiological tumor growth process. Indeed, it is widely accepted in the oncological community that tumor cells diffuse in the brain tissue. The long tails of the tumor cell density is invisible in MRIs. Clinicians speculate on this when they delineate radiotherapy margins 2 cm beyond the visible margins of MRIs. A recent method presented a way of modelling the invisible tumor cell distribution from a segmentation of the the edema surrounding the glioma [Konukoglu 2010b]. In parallel, a generative model was proposed to brain tumor appearance in MRIs from simulated tumor cell distributions [Prastawa 2009]. This approach provides useful training data to learn the tumor cell density distribution directly from the MRIs. This was made possible by the rise of multi-variate random forests which were also being applied to organ localization in CT scans [Criminisi 2009].

Many works described the random forest framework from a more methodological perspective [Criminisi 2011b]. These highlight the ability of the random forests to partition the feature space with respect to a task-specific optimality criterium. A similar partition can be driven directly from the image by using efficient aggregation algorithms [Corso 2008] or supervoxel formulations [Achanta 2012]. The spatial partition of the MRIs can then be used to feed a multi-scale discriminative classifier [Akselrod-Ballin 2006]. This type of approach demonstrated significant improvement over the state of the art [dos Santos 2012].

CHAPTER 2 Spatial Decision Forests for Lesion Segmentation in Multi-Channel MRIs of the Brain

Contents

2.1	Intro	oduction	6
2.2	\mathbf{Mat}	erials	7
	2.2.1	MICCAI Grand Challenge 2008 dataset	7
	2.2.2	Data preprocessing	9
	2.2.3	Notations	10
2.3	Met	hods	10
	2.3.1	Context-rich decision forest	10
	2.3.2	Visual features	13
2.4	Exp	eriments and results	14
	2.4.1	Results on the public Grand Challenge dataset \ldots	15
	2.4.2	Results on the private Grand Challenge dataset	15
2.5	\mathbf{Disc}	ussion	16
	2.5.1	Interpreting segmentation results	16
	2.5.2	Influence of preprocessing	17
	2.5.3	Influence of forest parameters	17
	2.5.4	Analysis of feature relevance	21
2.6	Con	clusion	26

Chapter 2 focuses on segmentation which is an essential task to diagnosis, prognosis and therapy planning of brain lesions. A general and efficient framework for automatic segmentation is presented with applications to MS lesions, low grade and high grade gliomas in MR images. It builds on supervised learning of discriminative random decision forests to provide a voxel-wise probabilistic classification of image volumes. Interestingly, a ranking of the most discriminative features used during classification can be derived. The method uses multi-channel MR intensities (T1, T1C, T2, Flair), knowledge on tissue classes and long-range spatial context to discriminate lesions from background. A symmetry feature is introduced accounting for the fact that some lesions tend to develop in an asymmetric way. Quantitative evaluation of the proposed method carried out on publicly available labeled datasets demonstrates state of the art performance.

2.1 Introduction

Multiple Sclerosis (MS) is a chronic, inflammatory and demyelinating disease that primarily affects the white matter of the central nervous system. Automatic detection and segmentation of MS lesions can help diagnosis and patient follow-up. It offers an attractive alternative to manual segmentation which remains a timeconsuming task and suffers from intra- and inter-expert variability. MS lesions, however, show a high variability in appearance and shape which makes automatic segmentation a challenging task. MS lesions lack common intensity and texture characteristics, their shapes are variable and their location within the white matter varies across patients.

A variety of methods have been proposed for the automatic segmentation of MS lesions. For instance, in [Anbeek 2004] and [Admiraal-Behloul 2005], the authors propose to segment white matter signal abnormalities by using an intensity-based k-nearest neighbors method with spatial prior and a fuzzy inference system, respectively. A similar classifier combined with a template-driven segmentation was proposed in [Wu 2006] to segment MS lesions into three different subtypes (enhancing lesions, T1 black holes, T2 hyperintense lesions). A false positive reduction based on a rule-based method, a level set method and a support vector machine classifier is presented in [Yamamoto 2010] along with a multiple-gray level thresholding technique.

Generative methods were proposed consisting in a tissue classification by means of an expectation maximization (EM) algorithm. For instance, the method presented in [Datta 2006] aims at segmenting and quantifying black holes among MS lesions. The EM algorithm can be modified to be robust against lesion affected regions, its outcome is then parsed in order to detect outliers which, in this case, coincide with MS lesions [Van Leemput 2001]. Another approach consists in adding to the EM a partial volume model between tissue classes and combining it with a Mahalanobis thresholding which highlights the lesions [Dugas-Phocion 2004]. Morphological postprocessing on resulting regions of interest was shown to improve the classification performance [Souplet 2008]. In [Freifeld 2009], a constrained Gaussian mixture model is proposed, with no spatial prior, to capture the tissue spatial layout. MS lesions are detected as outliers and then grouped in an additional tissue class. Final delineation is performed using probability-based curve evolution. Multi-scale segmentation can be combined with discriminative classification to take into account regional properties [Akselrod-Ballin 2006]. Beyond the information introduced via the spatial prior atlases, these methods are limited in their ability to take advantage of long-range spatial context in the classification task.

To overcome this shortcoming, we propose the use of an ensemble of discriminative classifiers. Our algorithm builds on the random decision forest framework which has multiple applications in bioinformatics [Menze 2009], and, for example, more recently also in the image processing community [Andres 2008, Yi 2009, Criminisi 2010b]. Adding spatial and multi-channel features to this classifier proved effective in object recognition [Shotton 2009], brain tissue segmentation in MR images [Yi 2009], myocardium delineation in 3D echocardiography [Lempitsky 2009] and organ localization in CT volumes [Criminisi 2010b].

Applying multi-channel and context-rich random forest classification to the MS lesion segmentation problem is novel, to our knowledge. The presented classifier also exploits a specific discriminative symmetry feature which stems from the assumption that the healthy brain is approximately symmetric with respect to the mid-sagittal plane and that MS lesions tend to develop in asymmetric ways. We then show how the forest combines the most discriminative channels for the task of MS lesion segmentation.

2.2 Materials

This section describes the data, algorithms and notations which are referred to in the rest of the article.

2.2.1 MICCAI Grand Challenge 2008 dataset

The results in this article rely on a strong evaluation effort. This section presents the MICCAI¹ Grand Challenge 2008 datasets, which is the largest dataset publicly available, and explains the way our method is compared against the winner of the challenge [Souplet 2008].

2.2.1.1 Presentation

The MICCAI Grand Challenge 2008 [Styner 2008] aims at evaluating and comparing algorithms in an independent and standardized way for the task of MS lesion segmentation. The organizers make publicly available two datasets through their website. A dataset of labeled MR images which can be used to train a segmentation algorithm, and an unlabeled dataset on which the algorithm should be tested. The website offers to quantitatively evaluate the segmentation results on the unlabeled dataset using the associated private ground truth database, and to publish the resulting scores. This project is an original initiative to provide an unbiased comparison between MS lesions segmentation algorithms. In the rest of the article, the dataset for which labels are publicly available will be referred to as *public* dataset, whereas the dataset for which data is not available will be referred to as *private* dataset.

2.2.1.2 Data

The public dataset contains 20 cases, 10 from the Children's Hospital in Boston (CHB) and 10 from the University of North Carolina (UNC), which are labeled by a CHB expert rater. The private dataset contains 25 cases, 15 from CHB and 10 from

¹MICCAI is the annual international conference on Medical Image Computing and Computer Assisted Intervention.

Name	Definition	Unit	Best	Worse
TNR	$\frac{TN}{FP+TN}$	%	100	0
TPR	$rac{TP}{TP+FN}$	%	100	0
FPR	$\frac{FP}{FP+TN}$	%	0	100
PPV	$\frac{TP}{TP + FP}$	%	100	0
VO	$rac{Vol(Seg\cap GT)}{Vol(Seg\cup GT)}$	%	0	100
VD	$\frac{Vol(Seg) - Vol(GT)}{Vol(GT)}$	%	0	$<\infty$
	$\sum_{u \in \partial(GT)} \min_{u \in \partial(GT)} d(u, v) + \sum_{u \in \partial(Seg)} \min_{u \in \partial(Seg)} d(u, v)$			
SD	$\frac{v \in \partial(Seg)}{card(Seg \cup GT)}$	mm	0	$<\infty$

Chapter 2. Spatial Decision Forests for Lesion Segmentation in Multi-Channel MRIs of the Brain

Table 2.1: The evaluation metrics true negative rate (TNR), true positive rate (TPR), false positive rate (FPR) and positive predictive value (PPV) are defined using the following notations: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The volume overlap (VO) and the relative absolute volume difference (VD) evaluates the differences between the segmentation (Seg) and the ground truth (GT)by computing their volume (Vol). The average symmetric surface distance (SD) measures how close the segmentation and the ground truth are from each other using the Euclidean distance d on the set of boundary voxels noted ∂ . The best, respectively worse, column contains the metric score of the perfect segmentation, respectively of a completely-off segmentation.

UNC. The private dataset was annotated by a single expert rater at CHB and jointly by 2 expert raters UNC. For each case, the centers provided 3 MR volumes: a T1-weighted image, a T2-weighted image and a Flair image. These were co-registered and sampled to fit the isotropic $0.5 \times 0.5 \times 0.5 \text{ }mm^3$ resolution.

2.2.1.3 Evaluation

Quantitative evaluation is carried out on the private dataset using a set of known metrics defined in [Styner 2008] and summed up in Table 2.1. The two full sets of expert segmentations were used as reference for method comparison.

2.2.1.4 Top-ranked methods

The challenge results highlight three top-ranked methods reflecting three different approaches to the task of MS lesion segmentation. A k-nearest neighbor classification of brain tissue relying on spatial location and intensity value was proposed in [Anbeek 2008]. This method provides a voxel-wise probabilistic classification of MS lesions. Alternatively, the iterative method proposed in [Shiee 2008, Shiee 2010] jointly performs brain tissue classification and MS lesion segmentation by combining statistical and topological atlases. Finally, in [Souplet 2008], the authors show that a global threshold on the Flair MR sequence, infered using an EM brain tissue classification, suffices to detect most MS lesions. The final segmentation is then

constrained to appear in the white matter by applying morphological operations.

The method proposed in [Souplet 2008] won the MICCAI MS Segmentation Challenge 2008. It will be referred to as *winner* method in the rest of the article. The segmentation results on public and private datasets were made available by the authors and will be used as reference.

2.2.2 Data preprocessing

We sub-sample and crop the images so that they all have the same size, $159 \times 207 \times 79$ voxels, and the same resolution, $1 \times 1 \times 2 \ mm^3$. Sub-sampling and cropping intend to reduce the time spent learning the classifier. RF acquisition field inhomogeneities are corrected [Prima 2001] and inter-subject intensity variations are normalized [Rey 2002]. The images are then aligned on the mid-sagittal plane [Prima 2002]. Spatial prior is added by registering the MNI atlas [Evans 1993] to the anatomical images, each voxel of the atlas providing the probability of belonging to the white matter (WM), the grey matter (GM) and the cerebro-spinal fluid (CSF) (cf. Figure 2.1).



Figure 2.1: Case CHB07 from the public Grand Challenge dataset. From top to bottom: three axial slices of the same patient. From left to right: preprocessed T1-weighted (I_{T1}) , T2-weighted (I_{T2}) and Flair MR images (I_{Flair}) , the associated ground truth GT and the registered white matter atlas (P_{WM}) .

2.2.3 Notations

The multi-channel aspect of the method presented in this article requires to carefully define and name each channel. MR images from the Grand Challenge dataset will be noted I_s where the index $s \in \{T1, T2, Flair\}$ stands for an MR sequence. Registered spatial priors will be noted P_t where the index $t \in \{WM, GM, CSF\}$ stands for a brain tissue class.

Although having different semantics, anatomical images and spatial priors will be treated under the unified term *signal channel* and denoted $C \in \{I_{T1}, I_{T2}, I_{Flair}, P_{WM}, P_{GM}, P_{CSF}\}$.

The data consists of a collection of voxel samples $\mathbf{v} = (\mathbf{x}, \mathbf{C})$ where $\mathbf{x} = (x, y, z)$ is the spatial position of the voxel and where

$$\mathbf{C} = (I_{T1}, I_{T2}, I_{Flair}, P_{WM}, P_{GM}, P_{CSF})$$

is the multi-channel image the voxel belongs to.

2.3 Methods

This section describes our adaptation of the random decision forests to the segmentation of MS lesions and illustrates the visual features employed.

2.3.1 Context-rich decision forest

Our detection and segmentation problem can be formalized as a binary classification of voxel samples into either background or lesions. This classification problem is addressed by a supervised method: discriminative random decision forest, an ensemble classifier using decision trees as base classifiers. Decision trees are discriminative classifiers which are known to suffer from over-fitting. A random decision forest [Amit 1997] achieves better generalization by growing an ensemble of many independent decision trees on a random subset of the training data and by randomizing the features made available to each node during training [Breiman 2001].

2.3.1.1 Forest training

The training data consists of a set of labeled voxels $\mathcal{T} = \{\mathbf{v}_k, Y(\mathbf{v}_k)\}$ where the label $Y(\mathbf{v}_k)$ is given by an expert. When asked to classify a new image, the classifier aims to assign every voxel \mathbf{v} in the volume a label $y(\mathbf{v})$. In our case, $y(\mathbf{v}) \in \{0, 1\}$, 1 for lesion and 0 for background.

The forest has T components with t indexing each tree. During training, all observations (voxels) \mathbf{v}_k are pushed through each of the trees. Each internal node applies a binary test [Shotton 2009, Yi 2009, Lempitsky 2009, Criminisi 2010b] as follows:

$$t^{\tau_{low},\tau_{up},\theta}(\mathbf{v}_k) = \begin{cases} true, & \text{if } \tau_{low} \le \theta(\mathbf{v}_k) < \tau_{up} \\ false, & \text{otherwise} \end{cases}$$

where θ is a function identifying the visual feature extracted at position \mathbf{x}_k . There are several ways of defining θ , either as a local intensity-based average, local spatial prior or context-rich cue. These are investigated in more detail in the next section. The value of the extracted visual feature is thresholded by τ_{low} and τ_{up} . The voxel \mathbf{v}_k is then sent to one of the two child nodes based on the outcome of this test.

During training, each node p is optimized using the partition of the training data \mathcal{T}_p it receives as input. At the end of the training process, each node p is assigned to the optimal binary test $t^{\lambda_p^*}$, where $\lambda_p^* = (\tau_{low}^*, \tau_{up}^*, \theta^*)_p$. The optimality criterion is the information gain, denoted IG, as defined in [Quinlan 1993]

$$IG(\lambda, \mathcal{T}_p) = H(\mathcal{T}_p) - H(\mathcal{T}_p|(t^{\lambda}(\mathbf{v}_k))_k)$$

where $\mathcal{T}_p \subset \mathcal{T}$ and where H denotes the entropy. More precisely, the term $H(\mathcal{T}_p|(t^{\lambda}(\mathbf{v}_k))_k)$ measures the error made when approximating the expert labeling Y by the binary test t^{λ} . The optimal parameter λ_p^* maximizes the information gain

$$\lambda_p^* = \arg\max_{\lambda} IG(\lambda, \mathcal{T}_p)$$

for node p. As a result, the optimal binary test is the test discriminating lesion from background voxels such as maximizing the information gain.

Only a randomly sampled subset Θ of the feature space is available at each node for optimization, while the threshold space is uniformly discretized. The optimal $\lambda^* = (\tau_{low}^*, \tau_{up}^*, \theta^*)$ is found by exhaustive search jointly over the feature and threshold space. Random sampling of the features leads to increased inter-node and inter-tree variability which improves generalization [Breiman 2001].

Trees are grown to a maximum depth D. At the node level, a leaf node is generated when the information gain is below a minimal value IG_{min} .

As a result of the training process, each leaf node l_t of every tree t receives a partition \mathcal{T}_{l_t} of the training data. The partition \mathcal{T}_{l_t} can be divided into two sets respectively containing background and lesion voxels and defined as

$$\mathcal{T}_{l_t}^b = \{ (\mathbf{v}, Y(\mathbf{v})) \in \mathcal{T}_{l_t} | Y(\mathbf{v}) = b \}$$

where $b \in \{0, 1\}$ stands for the background and lesion class, respectively. Subsequently, the following empirical posterior probability is defined

$$p_{l_t}(Y(\mathbf{v}) = b) = \frac{|\mathcal{T}_{l_t}^b|}{|\mathcal{T}_{l_t}|}$$

and stored at the leaf node.

Figure 2.2 illustrates how the decision trees partition the data in the feature space and how resulting probabilities are stored in leaf nodes.

Chapter 2. Spatial Decision Forests for Lesion Segmentation in Multi-Channel MRIs of the Brain



Figure 2.2: Decision trees encode feature space partitions. (a) A decision tree of depth D = 2 is considered in this example. Decision node 1 and leaf nodes 5 and 6 are colored to track the partitions of the training data in the feature space. The black cross stands for an unseen sample (voxel) which is classified while propagated down the tree. (b) A zoom on node 2 shows that its binary test, denoted by $t^{\tau_{low,2},\tau_{up,2},\theta_2}$, is optimized over a partition of the training data, denoted by $T_2 = \{\mathbf{v}_{k,2}, Y(v_{k,2})\}$. The leaf node 6 encloses the class distribution of the set of voxels reaching it during training. Two classes are considered background and lesion. (c) The dots stand for the training voxels and are colored according to their class. The black cross denotes a voxel from an unseen volume considered for prediction. Every decision node in the forest applies an axis-aligned feature test. Here we focus on decision nodes 0 and 2 using features θ_0 and θ_2 , respectively.

2.3.1.2 Prediction

12

When applied to a new test volume $\mathcal{T}_{test} = {\mathbf{v}_k}$, each voxel \mathbf{v}_k is propagated through all the trees by successive application of the relevant binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, posteriors $p_{l_t}(Y(\mathbf{v}) = b)$ are gathered in order to compute the final posterior probability defined as follows:

$$p(y(\mathbf{v}) = b) = \frac{1}{T} \sum_{t=1}^{T} p_{l_t}(Y(\mathbf{v}) = b)$$

which is a mean over all the trees in the forest. This probability may be thresholded at a fixed value $\tau_{posterior}$ if a binary segmentation is required.

A posterior map P_b is obtained by applying the same prediction procedure to all voxels. Thus for every voxel \mathbf{v} , $P_b(\mathbf{v}) = p(y(\mathbf{v}) = b)$ is the posterior probability of belonging to the class b. This probability map can be thresholded at a fixed



Figure 2.3: Posterior maps learned from two distinct synthetic training sets. In both cases, the training data consists of two classes, green and red, and is used to learn a large forest, here T = 350. The posterior map is obtained by classifying a dense grid in the feature space and is then overlayed with the associated training data (shown as points). Larger opacities indicate larger probability of a pixel belonging to a class while uncertain regions are indicated by less saturated colours. The white line plots the locus of points for which $p(y(\mathbf{v}) = green) = p(y(\mathbf{v}) = red)$. We observe that 1) forest posteriors mimic the maximum-margin behaviour, 2) uncertainty increases when moving away from the training data.

value to obtain a segmentation. Choosing $\tau_{posterior} = 0.5$ is equivalent to looking for $b^* = \arg \max_b P_b(\mathbf{v})$.

Probability maps learned by random forests exhibit a maximum-margin like behavior, which is known to be a property of support vector machines (cf. Figure 2.3). In addition random forests provide a confidence measure for a voxel to belong to a given class. Along with their ability to rank the most discriminative features (cf. Section 2.5.4), these properties motivate the use of random forests for the task of MS lesion segmentation.

2.3.2 Visual features

Two kinds of visual features are computed:

1) local features:

$$\theta_C^{loc}(\mathbf{v}) = C(\mathbf{x})$$

where C is an intensity or a prior channel;

2) context-rich features comparing the voxel of interest with distant regions.

The first context-rich feature compares the local voxel value in channel C_1 with the mean value in channel C_2 over two 3D boxes R_1 and R_2 within an extended neighborhood:

$$\theta_{C_1,C_2,R_1,R_2}^{cont}(\mathbf{v}) = C_1(\mathbf{x}) - \frac{1}{Vol(R_1)} \sum_{\mathbf{x}' \in R_1} C_2(\mathbf{x}') - \frac{1}{Vol(R_2)} \sum_{\mathbf{x}' \in R_2} C_2(\mathbf{x}')$$

where C_1 and C_2 are both intensity or prior channels. The regions R_1 and R_2 are sampled randomly in a large neighborhood of the voxel **v** (cf. Figure 2.4). The sum over these regions is efficiently computed using integral volume processing [Shotton 2009].

The second context-rich feature compares the voxel of interest at \mathbf{x} with its symmetric counterpart with respect to the mid-sagittal plane, noted $S(\mathbf{x})$:

$$\theta_C^{sym}(\mathbf{v}) = C(\mathbf{x}) - C \circ S(\mathbf{x})$$

where C is an intensity channel. Instead of comparing with the exact symmetric $S(\mathbf{x})$ of the voxel, we consider, respectively, its 6, 26 and 32 neighbors in a sphere S (cf. Figure 2.4), centered on $S(\mathbf{x})$. We obtain a softer version of the symmetric feature which reads:

$$\theta_{C,S}^{sym}(\mathbf{v}) = \min_{\mathbf{x}' \in S} \{ C(\mathbf{x}) - C(\mathbf{x}') \}$$

which loosens the hard symmetric constrain.

When encoded in additional channels, new precomputed visual features can be taken into account in a straightforward way [Yi 2009].



Figure 2.4: **2D view of context-rich features.** (a) A context-rich feature depicting two regions R_1 and R_2 with constant offset relatively to **x**. (b-d) Three examples of randomly sampled features in an extended neighborhood. (e) The symmetric feature with respect to the mid-sagittal plane. (f) The hard symmetric constraint. (g-i) The soft symmetry feature considering neighboring voxels in a sphere of increasing radius. See text for details.

2.4 Experiments and results

Results presented in this section aim at evaluating the segmentation results and comparing the context-rich random forest approach to the challenge winner [Souplet 2008]. Experiments described here are discussed in Section 2.5.

Exhaustive segmentation results are available for both public and private datasets under the following url:

ftp://ftp-sop.inria.fr/asclepios/Published-Material/Ezequiel.Geremia/

2.4.1 Results on the public Grand Challenge dataset

For quantitative evaluation, the 20 available cases from the public dataset are classified and compared to the winner method [Souplet 2008]. A three-fold cross-validation is carried out on this dataset: the forest is trained on $\frac{2}{3}$ of the cases and tested on the other $\frac{1}{3}$, this operation is repeated 3 times in order to collect test errors for each case.

The binary classification is evaluated using two measures, true positive rate (TPR) and positive predictive value (PPV), both equal 1 for perfect segmentation (cf. Table 2.1).

Forest parameters are fixed to the following values: number of random regions $|\Theta| \simeq 950$, number of trees T = 30, tree depth D = 20, lower bound for the information gain $IG_{min} = 10^{-5}$, posterior threshold $\tau_{posterior} = 0.5$. Parameters T and D are set here to maximum values, Section 2.5.3 explains how these parameters can be optimized in order to improve segmentation results.

Tables 2.4 and 2.2 reports extensive results allowing comparison on every case of the Grand Challenge public dataset. It shows that the learned context-rich random forest achieves better TPR in all cases, and better PPV in 70% of the cases. Computed *p*-values for the pair-sample *t*-test show that these improvements are significative for both TPR ($p = 1.3 \cdot 10^{-7}$) and PPV (p = 0.0041) scores.

Metric [%]	Challenge winner [Souplet 2008]	Context-rich RF	RI [%]	<i>p</i> -value
TPR	19.21 ± 13.68	39.39 ± 18.40	105	$1.3\cdot10^{-7}$
PPV	29.55 ± 16.26	39.78 ± 20.19	35	0.0041

Table 2.2: Comparison of context-rich random forests with the challenge winner method on the public dataset. Relative improvement over the challenge winner algorithm [Souplet 2008], defined as $RI = (score_{RF} - score_{winner})/score_{winner}$, are significant for both TPR ($p = 1.3 \cdot 10^{-7}$) and PPV (p = 0.0041) scores. Significant improvements over the challenge winner algorithm are highlighted in bold.

2.4.2 Results on the private Grand Challenge dataset

A context-rich random forest was learned on the whole public dataset from the MS Lesion Challenge, i.e. 20 labeled cases. Forest parameters are fixed to the following values: number of random regions $|\Theta| \simeq 950$, number of trees T = 30, tree depth D = 20, lower bound for the information gain $IG_{min} = 10^{-5}$, posterior threshold $\tau_{posterior} = 0.5$. Considerations that lead to these parameter values are detailed in Section 2.5.3.

Chapter 2. Spatial Decision Forests for Lesion Segmentation in Multi-Channel MRIs of the Brain

The Grand Challenge 2008 website carried out a complementary and independent evaluation of our algorithm on the previously unseen private dataset. The results, reported in Tables 2.3 to 2.6, confirm a significant improvement over the challenge winner algorithm [Souplet 2008]. The presented spatial random forest achieves, on average, slightly larger true positive (TPR), which is beneficial (cf. Table 2.1), and comparable false positive (FPR) rates but lower volume difference (VD) and surface distance (SD) values (cf. Table 2.3). Pair-sample *p*-values were computed for the *t*-test on the private dataset. Results show significant improvement over the winner method [Souplet 2008] on SD $(p = 4.2 \cdot 10^{-6})$ for the CHB rater, and on SD $(p = 6.1 \cdot 10^{-3})$ for the UNC rater.

Rater	Metric [%]	Challenge winner [Souplet 2008]	Context-rich RF	RI [%]	<i>p</i> -value
	VD	86.48 ± 104.9	52.94 ± 28.63	-38.7	0.094
СНВ	SD	8.20 ± 10.89	5.27 ± 9.54	-35.7	$4.2\cdot 10^{-6}$
	TPR	57.45 ± 23.22	58.08 ± 20.03	+1.0	0.90
	FPR	68.97 ± 19.38	70.01 ± 16.32	+1.5	0.70
UNC	VD	55.76 ± 31.81	50.56 ± 41.41	-9.4	0.66
	SD	7.4 ± 8.28	5.6 ± 6.67	-24.3	$6.1\cdot 10^{-3}$
	TPR	49.34 ± 15.77	51.35 ± 19.98	+3.9	0.54
	FPR	76.18 ± 17.07	76.81 ± 11.70	+0.1	0.83

Table 2.3: Average results computed by the Grand Challenge on the private dataset. The relative improvement on the winner algorithm [Souplet 2008] on the private dataset is defined as follows $RI = (score_{RF} - score_{winner})/score_{winner}$. The RI and the p-values a and is reported on top for each metric of the Grand Challenge, associated *p*-values are reported below. Independent quantitative evaluation confirms improvement on the winner algorithm [Souplet 2008]. Boldface highlights significant improvements. The spatial random forest achieves, on average, slightly larger true positive (TPR), which is beneficial and comparable false positive (FPR) rates but lower volume difference (VD) and surface distance (SD) values.

2.5 Discussion

2.5.1 Interpreting segmentation results

Quantitative evaluation of segmentation results, for both public (cf. Tables 2.2 and 2.4) and private (cf. Tables 2.3 to 2.6) datasets, shows significant improvement on the challenge winner method [Souplet 2008]. Although segmentation results include most MS lesions delineated by the expert, we observe that some MS lesions are missing. Missed MS lesions are located in specific locations which are not represented in the training data, e.g. in the corpus callosum (cf. Figure 2.5, slice 38). This is a limitation of the supervised approach. In this very case, however, the posterior map highlights the missed lesion in the corpus callosum as belonging to the lesion class with high uncertainty. Low confidence (or high uncertainty) reflects the incorrect spatial prior inferred from an incomplete training set. Indeed, in the training set, there is no example of MS lesions appearing in the corpus callosum.

On the contrary, the random forest is able to detect suspicious regions with high certainty. Suspicious regions are visually very similar to MS lesions and widely represented in the training data, but they are not delineated by the expert, e.g. the left frontal lobe lesion again in Figure 2.5, slice 38. The appearance model and spatial prior implicitly learned from the training data points out that hyper-intense regions in the FLAIR MR sequence which lay in the white matter (cf. Section 2.5.4) can be considered as MS lesions with high confidence.

Recent histopathological studies have shown that grey matter regions are also heavily affected by the MS disease [Geurts 2008]. In our case, the public dataset does not show any MS lesion in the grey matter of the brain. Subsequently, the decision forest learns that MS lesions preferably appear in the white matter. Adding new cases showing grey matter MS lesions in the training set would allow the forest to automatically adapt the segmentation to include this kind of lesions. This observation stresses the necessity of gathering large and heterogeneous datasets for training purposes.

When focusing on quantitative measures, we observe that cases UNC01 and UNC06 from the public dataset show surprisingly low scores (cf. Tables 2.2 and 2.4). The labels by the CHB expert for these two cases are abnormal: the ground truth is mirrored with respect to the anatomical images. This may be considered as a label error and explains the low scores for these two specific cases. The Grand Challenge website confirmed this observation and subsequently corrected the online database. We also observe that learning on the whole public dataset and testing on the private dataset (cf. Table 2.3) produces better average results than the three-fold cross-validation carried out on the public dataset (cf. Tables 2.2 and 2.4). Again this illustrates the benefit of learning the classifier on large enough datasets capturing better the variability of the data.

2.5.2 Influence of preprocessing

Data normalization is critical. Spatial normalization affects the evaluation of context-rich features, θ^{cont} , which are sensitive to rotation. Moreover, all intensity-based features require inter-case intensity normalization.

The trees are generated in parallel on 30 nodes and gathered to form the forest. Cropping and sub-sampling the training images aims at reducing, by a factor larger than 10, the execution time spent to learn a single tree. On IBM e325 dual-Opterons 246 at a maximum frequency of 2Ghz, learning a tree on 20 sub-sampled images and with parameters fixed in Section 2.4.2 on a single CPU takes, on average, 8 hours.

2.5.3 Influence of forest parameters

The number of the trees and their depth, respectively denoted T and D, characterize the generalization power and the complexity of the non-parametric model learned by the forest. This section aims at understanding the contribution of each of these



Figure 2.5: Segmenting Case CHB05 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior}))$ with $\tau_{posterior} = 0.5$). Segmentation results show that most of lesions are detected. Although some lesions are not detected, e.g. peri-ventricular lesion in slice 38, they appear enhanced in the posterior map. Moreover the segmentations of slices 38 and 42 show peri-ventricular regions, visually very similar to MS lesions, but not delineated in the ground truth.



Figure 2.6: Segmenting case CHB01 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior})$ with $\tau_{posterior} = 0.5$).



Figure 2.7: Segmenting case CHB07 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior})$ with $\tau_{posterior} = 0.5$).

meta-parameters.

A 3-fold cross-validation on the public dataset is carried out for each parameter combination. Segmentation results are evaluated for each combination using two different metrics: the area under the receiver operating characteristic (ROC) curve and the area under the precision-recall curve. The ROC curve plots TPR vs. FPRscores computed on the test data for every value of $\tau_{posterior} \in [0, 1]$. The precisionrecall curve plots PPV vs. TPR scores computed on the test data for every value of $\tau_{posterior} \in [0, 1]$. Results are reported in Figure 2.8.

We observe that 1) for a fixed depth, increasing the number of trees leads to better generalization; 2) for a fixed number of trees, low depth values lead to underfitting while high depth values lead to overfitting; 3) overfitting is reduced by increasing the number of trees.

This analysis was carried out a posteriori. Tuning the meta-parameters of the forest on the training data is not a valid practice. Using out-of-bag samples for forest parametrization is indeed preferable. Due to the fact that little training data is available for the MS lesion class, available labeled data was exclusively used to train the forest. From this perspective, the forest parameters were set to arbitrary but high enough values to avoid under- and overfitting: T = 30 and D = 20.

Forest parameters were indeed selected in a safety-area with respect to underand overfitting. The safety-area corresponds to a sufficiently flat region in the evolution of the areas under the ROC and the precision-recall curve. As shown in Figure 2.9, increasing the number of trees tends to benefit the generalization power of the classifier. We also observe that the performance of the classifier stabilizes for large enough forests.

2.5.4 Analysis of feature relevance

During training, features considered for node optimization form a large and heterogeneous set (cf. Section 2.3.2). Unlike other classifiers, random forests provide an elegant way of ranking these features according to their discriminative power. In this section, we aim at better understanding which are the most discriminative channels and visual cues (local, context-rich or symmetric) used in the classification process.

2.5.4.1 Most discriminative visual features

The first approach consists in counting the nodes in which a given feature was selected. We observe that local features were selected in 24% of the nodes, context-rich features were selected in 71% of the nodes whereas symmetry features were selected in 5% of the nodes (cf. Figure 2.10). In this case, no distinction is made as for the depth at which a given feature was selected.

Context-rich features exhibit high variability (900 of them are randomly sampled at every node). This variability combined with their ability to highlight regions which differ from their neighborhood explains why they were chosen. Together



Figure 2.8: Influence of forest parameters on segmentation results. Both curves were plotted using mean results from a 3-fold cross validation on the public dataset. Left: the figure shows the influence of forest parameters on the area under the precision-recall curve. Right: the figure shows the influence of forest parameters on the area under the ROC curve. The ideal classifier would ensure area under the curve to be equal to 1 for both curves. We observe that 1) for a fixed depth, increasing the number of trees leads to better generalization; 2) for a fixed number of trees, low depth values lead to underfitting while high values lead to overfitting; 3) overfitting vanishes by increasing the number of trees.

with local features, context-rich features learn a multi-channel appearance model conditioned by tissue spatial priors. Symmetry features are under-represented in the forest and thus prove to be the least discriminative ones. This is due to the fact that a large proportion of peri-ventricular MS lesions tend to develop in a symmetric way. Nevertheless, symmetric features appear in top levels of the tree (up to third level) which indicates that they provide an alternative to local and context-rich features when these two fail.

A finer estimation of the feature importance consists in weighting the counting process. For a given feature, instead of only counting the nodes in which it appears, we also take into account the proportion of lesion voxels it helps discriminating: "the larger the proportion of lesion voxels it helps to discriminate, the larger the weight of the feature". This leads us to define for a fixed depth value d, the importance of a given feature type, denoted (IFT), as:

$$IFT(\alpha) = \frac{1}{|\mathcal{T}^1|} \sum_{p} |\mathcal{T}_p^1| \cdot \chi_\alpha(\theta_p) \quad \alpha \in \{loc, cont, sym\}$$
(2.1)

where α denotes a feature type, p indices the nodes in layer d, \mathcal{T}^1 is the training set of lesion voxels which partition \mathcal{T}_p^1 reached node p, and χ is the indicator function such that:

$$\chi_{\alpha}(\theta_p) = \begin{cases} 1, & \text{if } \theta_p \text{ is of type } \alpha \\ 0, & \text{otherwise} \end{cases}$$
(2.2)

The feature importance evaluates to 21.1~% for local features, 76.6~% for context-rich



Figure 2.9: Influence of the number of trees on segmentation results. Both curves were plotted using mean results from a 3-fold cross validation on the public dataset. Top: the figure shows the influence of the number of trees on the area under the precision-recall curve. Bottom: the figure shows the influence of the number of trees on the area under the ROC curve. We observe that, for a fixed depth D = 14, increasing the number of trees improves generalization as stated in [Breiman 2001]. The increase in performance stabilizes around the value T = 30.

features and 2.3 % for symmetry features. Results are comparable to those obtained only by counting the features in the forest, but the real advantage of this measure is to allow us to draw depth-by-depth feature importance analysis in a normalized way.

The feature importance as a function of the depth of the tree is reported in Figure 2.11. Presented results are averaged values over a forest containing T = 30 trees. Again, we observe that context-rich features are predominantly selected as the most discriminative, which confirms the trend reported in Figure 2.10. However, as shown in Figure 2.11, the preponderance of context-rich features is not uniform throughout the tree. Indeed, local features are the most discriminative in layers 0 and 2. A careful analysis of selected channels helps understanding why local features are selected in the top layers of the tree (cf. Section 2.5.4.2).

The selected context-rich features show high variability. More specifically, the long-range regions are distributed all over the neighborhood. Depth-by-depth analysis does not show any specific pattern in the position of the regions with respect to the origin voxel. In addition, the volume of the regions also show high variability. The observed heterogeneity of selected context-rich features aims at coping with the variability of MS lesions (shape, location and size).

The symmetry feature is under-represented in the forest. Its discriminative power is thus very low compared to local and context-rich features. This observation induces two complementary interpretations to explain why symmetry features are the least significant: 1) most of MS lesions appear in peri-ventricular regions and in a symmetric way, 2) most of MS lesions can be clearly identified by their signature across MR sequences and their relative position in the white matter of the brain. However, in deeper layers of the trees, the symmetry feature is more significant and tends to classify ambiguous asymmetrical regions. When looking into the selected features, we also notice that the hard symmetric constraint is preferred over the loose symmetric constrain (cf. Section 2.3.2). Indeed, the feature importance evaluates to 1.6% for the hard symmetric feature, and to 0.7% for the loose symmetric feature. Moreover, in the rare cases where the loose constrain is selected, the 6neighbors version predominates (cf. Section 2.3.2). This observation supports the idea that considering brain hemispheres as symmetric is an accurate approximation in our specific setting (cf. Sections 2.2 and 2.3).

2.5.4.2 Most discriminative channels

The second approach focuses on the depth at which a given feature was selected. For every tree in the forest, the root node always applies a test on the FLAIR sequence (θ_{FLAIR}^{loc}) . It means that out of all available features, containing local, context-rich and symmetry multi-channel features, θ_{FLAIR}^{loc} was found to be the most discriminative. At the second level of the tree, a context-rich feature on spatial priors $(\theta_{WM,GM}^{cont})$ appears to be the most discriminative over all trees in the forest. It aims at discarding all voxels which do not belong to the white matter.

The optimal decision sequence found while training the context-rich forest can



Figure 2.10: Ranking features according to the proportion of nodes in which they appear. Context-rich features are selected in 71% of the nodes, local features are selected in 24% of the nodes whereas symmetry features are selected in 5% of the nodes.



Figure 2.11: **Type of feature selected by layer of the tree.** For a fixed depth, the red circle stands for the importance of the context-rich feature (θ^{cont}), while the green circle stands for the importance of the local feature (θ^{loc}). For clarity, symmetry features (θ^{sym}) are omitted as they are under-represented in the forest. The blue line monitors the proportion of training samples of the lesion class which do not reside in leaf nodes, for each layer of the tree. We observe that context-rich features are predominantly selected as the most discriminative ones except in layers 0 and 2.





Figure 2.12: Combination of features and channels learned by the forest to discriminate MS lesions. The first layer of all trees in the forest performs a threshold on the FLAIR MR sequence. The second one discards all voxels which do not belong to the white matter. The posterior map is obtained by using a forest with trees of depth 2 and thus highlights hyper-intense FLAIR voxels which lie in peri-ventricular regions.

thus be thought as a threshold on the FLAIR MR sequence followed by an intersection with the white matter mask (cf. Figure 2.12). Interestingly, this sequence matches the first and second steps of the pipeline proposed by the challenge winner method [Souplet 2008]. Note that in our case, it is automatically generated during the training process. Deeper layers in the trees, then, refine the segmentation of MS lesions by applying more accurate decisions.

The feature importance (cf. Equation 2.1) can be extended in a straightforward way to be parametrized not only by the type of feature (local, context-rich, symmetric) but also by the channel. When globally looking at the selected channels (cf. Figure 2.13), we notice that their importance varies throughout the tree: first layers, as mentioned before, favor detection of bright spots in the white matter by successively testing the FLAIR MR sequence, spatial priors on WM and GM tissues and finally testing on the T2 MR sequence; deeper layers take into account other modalities to adjust the segmentation.

2.6 Conclusion

 $\mathbf{26}$

We demonstrated the power of the RF formulation applied to the difficult task of MS lesion segmentation in multi-channel MR images. We presented three kinds of 3D features based on multi-channel intensity, prior and context-rich information. Those features are part of a context-rich random decision forest classifier which demonstrated improved results on one of the state of the art algorithms on the public MS challenge dataset. In addition, the random decision forest framework provided a means to automatically select the most discriminative features to achieve the best possible segmentation. Future work could include the use of more sophisticated


Figure 2.13: Channel importance as a function of the depth of the tree. Plots draw the channel importance drawn as a function of the depth of the tree for both local (top) and long-range features (bottom). For a fixed depth, only the most discriminative channel is depicted. Note how successive layers of the tree test complementary channels: the first layer performs a local test on the FLAIR MR sequence in order to detect bright spots, the second one discards all voxels which do not belong to the white matter by using context-rich information over the WM and GM channels. Note that a large spectrum of available channels is tested throughout the tree.

features to reduce even further the preprocessing requirements. One could also explore the application of our approach to the segmentation of brain tumors in multi-sequence MR images of the brain. Finally, one could investigate an extension of the proposed approach to larger multi-class problems in order to try to simultaneously segment brain tissues (WM, GM, CSF) along with MS lesions.

Appendix: Exhaustive results on public and private datasets

In this section, we provides exhaustive quantitative results on the public and private datasets in Tables 2.4, 2.5 and 2.6. In Figures 2.14, 2.15, 2.16 and 2.17, we also provide additional visual evidence of the performance of the presented algorithm.

Patient	Ch. v	vinner	Context-rich RF		
	[Souple	et 2008]			
(public)	TPR	PPV	TPR	PPV	
CHB01	21.14	40.79	48.24	63.99	
CHB02	17.53	28.71	43.44	62.17	
CHB03	16.42	20.71	21.37	56.17	
CHB04	11.58	54.74	30.24	77.32	
CHB05	21.60	41.97	39.64	51.25	
CHB06	12.38	45.76	31.04	50.32	
CHB07	12.42	38.06	39.73	53.88	
CHB08	12.17	54.45	45.84	64.87	
CHB09	2.03	17.49	22.95	27.09	
CHB10	4.68	17.02	22.40	38.20	
UNC01	0.05	0.03	1.16	1.00	
UNC02	36.43	38.65	47.81	35.25	
UNC03	11.12	15.90	23.73	34.95	
UNC04	37.47	53.05	53.24	37.49	
UNC05	37.80	7.61	56.00	18.60	
UNC06	8.07	8.81	14.25	7.81	
UNC07	56.70	17.26	75.79	15.57	
UNC08	26.97	29.82	51.74	31.05	
UNC09	15.68	43.00	66.23	35.59	
UNC10	22.00	27.24	52.96	33.09	

Table 2.4: Quantitative results on the public dataset comparing the challenge winner [Souplet 2008] to the context-rich random forest we propose. Boldface highlights improvements over the challenge winner.

CHB rater								
Patient	Ch. winner ([Souplet 2008])				Context-rich random forests			
(private)	VD	SD	TPR	FPR	VD	SD	TPR	FPR
CHB01	57.6	2.6	83.9	35.3	40.7	2.6	48.4	75.3
CHB02	71.3	2.3	63.2	41.3	59.7	1.5	47.4	45.0
CHB03	75.3	12.2	33.3	90.4	79.1	6.6	46.7	87.2
CHB04	76.1	3.8	72.2	50.0	65.5	2.1	66.7	66.0
CHB05	74.0	3.2	69.6	71.0	53.6	1.6	65.2	68.6
CHB06	68.6	6.6	27.3	90.6	77.3	2.3	27.3	58.2
CHB07	77.3	4.8	55.3	74.0	40.3	1.5	57.9	62.9
CHB08	66.3	2.3	61.8	62.9	31.7	1.0	76.5	61.8
CHB09	63.1	2.6	35.8	55.4	44.5	2.3	24.1	54.9
CHB10	49.4	3.9	79.3	76.4	22.6	1.4	75.9	73.2
CHB01	72.9	3.6	51.7	81.7	43.4	1.2	41.4	76.6
CHB12	86.1	6.8	15.4	77.7	49.7	1.3	46.2	64.1
CHB13	69.3	4.6	76.2	64.6	52.2	1.4	85.7	56.5
CHB15	62.9	1.9	57.4	49.7	35.7	1.3	46.8	49.6
UNC01	36.5	5.3	53.1	69.2	55.6	4.4	56.3	77.6
UNC02	87.8	3.5	40.9	31.9	64.6	2.1	36.4	35.6
UNC03	30.5	1.3	55.9	40.6	41.7	2.7	27.9	54.6
UNC04	85.5	7.1	48.1	86.4	43.8	2.1	66.7	83.1
UNC05	5.3	6.4	78.3	77.9	18.1	2.4	78.3	74.5
UNC06	15.5	19.4	62.5	86.9	30.2	14.1	62.5	88.3
UNC07	26.2	6.4	80.0	71.5	134.3	4.9	76.7	85.6
UNC08	28.9	10.6	94.4	76.1	0.7	4.9	83.3	88.1
UNC09	263.9	53.0	0.0	100.0	75.7	46.7	100.0	98.4
UNC10	525.2	22.5	83.3	93.8	110.1	14.3	50.0	94.5

Table 2.5: Quantitative results computed by the Grand Challenge on the private dataset using the CHB rater as reference. Boldface highlights improvements over the challenge winner [Souplet 2008].

UNC rater								
Patient	Ch. winner ([Souplet 2008])				Context-rich random forests			
(private)	VD	SD	TPR	FPR	VD	SD	TPR	FPR
CHB01	70.3	3.9	52.0	27.6	58.5	5.1	26.7	74.2
CHB02	32.7	6.2	63.6	84.1	5.5	4.8	50.0	83.8
CHB03	48.9	11.6	50.0	87.7	56.7	3.0	71.4	84.6
CHB04	50.2	5.2	72.7	81.0	28.3	3.8	100.0	84.5
CHB05	36.8	8.8	48.1	92.5	144.7	9.3	48.2	92.7
CHB06	69.9	6.7	16.7	93.0	78.3	2.7	27.8	60.4
CHB07	62.7	5.5	48.3	85.6	1.9	3.1	58.3	77.0
CHB08	49.6	2.1	81.5	74.8	2.0	1.3	92.6	77.9
CHB09	56.3	3.6	34.9	69.6	34.2	4.0	30.9	70.8
CHB10	3.4	7.4	68.4	93.4	58.4	5.1	73.7	87.4
CHB01	16.3	6.0	47.7	91.4	74.9	6.1	38.6	92.8
CHB12	86.0	7.1	12.0	81.7	49.4	1.6	42.2	64.8
CHB13	50.0	5.4	50.0	85.4	22.0	6.3	80.0	87.0
CHB15	71.9	2.6	53.4	44.6	51.2	2.5	45.2	53.6
UNC01	56.6	4.8	51.2	75.7	6.4	4.5	55.8	74.1
UNC02	7.7	4.1	54.4	72.5	168.5	4.1	55.9	79.7
UNC03	46.2	1.5	53.5	49.4	54.9	3.5	26.1	57.9
UNC04	90.6	8.6	34.2	86.4	63.8	2.5	55.3	76.7
UNC05	53.4	6.0	52.4	75.2	63.8	3.5	42.9	66.7
UNC06	81.1	7.0	41.4	71.4	84.4	6.3	41.4	60.6
UNC07	45.6	2.6	55.7	58.3	1.0	3.3	52.5	74.3
UNC08	21.1	3.9	63.8	65.2	38.4	2.9	48.9	79.9
UNC09	157.9	44.0	33.3	96.6	24.5	35.3	33.3	98.4
UNC10	73.0	13.0	45.0	85.2	41.9	9.9	35.0	83.6

Table 2.6: Quantitative results computed by the Grand Challenge on the private dataset using the UNC rater as reference. Boldface highlights improvements over the challenge winner [Souplet 2008].



Figure 2.14: Segmenting case CHB02 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior}))$ with $\tau_{posterior} = 0.5$).



Figure 2.15: Segmenting case CHB08 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior}))$ with $\tau_{posterior} = 0.5$).



Figure 2.16: Segmenting case UNC02 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior}))$ with $\tau_{posterior} = 0.5$).



Figure 2.17: Segmenting case UNC04 from the public Grand Challenge dataset. From left to right: preprocessed T1-weighted (I_{T1}) and Flair MR images (I_{Flair}) overlayed with the associated ground truth GT, the posterior map $Posterior = (P_{lesion}(\mathbf{v}_k))_k$ and the Flair sequence overlayed with the segmentation $(Seg = (Posterior \ge \tau_{posterior}))$ with $\tau_{posterior} = 0.5$).

CHAPTER 3

Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model

Contents

3.1 Intr	3.1 Introduction				
3.2 Lea	rning to estimate tissue cell density from synthetic				
trai	ning data	39			
3.2.1	Generative tumor simulation model $\ldots \ldots \ldots \ldots \ldots$	39			
3.2.2	Regression forests for estimating tissue cell density	41			
3.3 Exp	eriments	43			
3.3.1	Experiment 1: Estimating cell density in synthetic cases	43			
3.3.2	Experiment 2: Segmenting tumors in clinical images	44			
3.4 Disc	cussion	44			
3.5 Con	clusions	54			

Chapter 3 investigates the promising perspective of estimating the brain tumor cell density from MR images. A generative-discriminative framework is presented to learn model-based estimations of missing cues, such as the tumor cell density which is very difficult to obtain. The generative model is a validated and publicly available biophysiological tumor growth simulator. It outputs synthetic multi-modal MR images, tissue class annotation and the tumor cell density for which clinical ground truth is very difficult to obtain. The discriminative model builds on multivariate regression random forests to estimate the voxel-wise distribution of tumor cell density from input MR images. The random forest is learnt on a large dataset of 500 synthetic cases and their associated ground truth generated by the brain tumor simulator. On real clinical cases from the low-grade glioma DKFZ dataset, the method provides realistic tumor cell density estimations closely related to the multimodal image information. A binary tumor segmentation is derived from the tumor cell density distribution and compared against the state of the art as a consistency check. Quantitative evaluation on the low-grade glioma DKFZ dataset demonstrates tissue class accuracy comparable with the state of the art with the added benefit of providing the latent tumor cell density.

3.1 Introduction

Brain tumors are complex patho-physiological processes representing a series of pathological changes to brain tissue [Angelini 2012]. Increasing effort is invested

Chapter 3. Brain tumor cell density estimation from multi-modal MR 38 images based on a synthetic tumor growth model

in modelling the underlying biological processes involved in brain tumor growth [Cristini 2010, Deisboeck 2010]. As brain tumors show a large variety of different appearances in multi-modal clinical images, the accurate diagnosis and analysis of these images remains a significant challenge. We show in the example of gliomas, the most frequent brain tumor [Angelini 2007], how a generative patho-physiological model of tumor growth can be used in conjunction with a discriminative tumor recognition algorithm, based on random regression forests. Applied to real data the random forest is capable of predicting the precise location of the tumor and its substructures. In addition, our model can also infer the spatial distribution of (unobservable) latent physiological features such as tumor cell densities, thus avoiding the need for expensive patho-physiological model inversion [Menze 2011].

Generative probabilistic segmentation models of spatial tissue distribution and appearance proved to generalize well to previously unseen images [Prastawa 2004, Zou 2002, Menze 2010, Riklin-Raviv 2010]. In [Prastawa 2004], tumors are modeled as outliers relative to the expected appearance of healthy tissue following a related approach for MS lesion detection [Van Leemput 2001]. Other methods [Zou 2002, Menze 2010] provide explicit models for the tumor class. For instance, [Menze 2010] builds a tumor appearance model for channel specific segmentation of the tumor, combining a tissue appearance model with a latent tumor class prior from [Riklin-Raviv 2010]. Tumor growth models (*e.g.* reaction-diffusion models) have been used repeatedly to improve image registration [Cabezas 2011] and, hence, atlas-based tissue segmentation [Gooya 2011b]. Similarly, [Zacharaki 2009] relies on a bio-mechanical tumor growth model to estimate brain tissue loss and displacement. Generative approaches require a detailed formal description of the image generation process and may need considerable modifications when applied to slightly different tasks. These approaches also tend to be computationally inefficient.

In contrast, discriminative techniques focus on modeling the difference between *e.g.* a lesion and healthy tissues, directly [Lee 2008, Wels 2008, Corso 2008]. A number of recent techniques based on decision tree ensembles have shown strong generalization capabilities and computational efficiency, even when applied to large data sets [Criminisi 2010b, Montillo 2011, Gray 2011]. In [Geremia 2011], for example, a *classification* forest is used for segmenting multiple sclerosis lesions using long-range spatial features. In [Wels 2008], the authors derived a constrained minimization problem suitable for min-cut optimization that incorporates an observation model provided by a discriminative Probabilistic Boosting Trees classifier into the process of segmentation. For multi-modal brain lesion segmentation, [Corso 2008] propose a hierarchical segmentation framework by weighted aggregation with generic local image features. Unfortunately, fully supervised discriminative approaches may require large expert-annotated training sets. Obtaining such data is often prohibitive in many clinical applications.

This paper proposes a new way of combining the best of the generative and discriminative world. We use a generative model of glioma [Prastawa 2009] to synthesize a large set of heterogeneous MR images complete with their ground truth annotations. Such images are then used to train a multi-variate *regression* forest

3.2. Learning to estimate tissue cell density from synthetic training data

tumor predictor [Geremia 2011, Criminisi 2011b]. Thus given a previously unseen image the forest can perform an efficient, per-voxel estimation of both tumor infiltration density and tissue type. The general idea of training a discriminative predictor (a classifier or a regressor) on a large collection of synthetic training data is inspired by the recent success of the Microsoft Kinect for XBox 360 system [Shotton 2011]. This approach has great potential in different domains and especially for medical applications where obtaining extensive expert-labelled is nearly impossible.

3.2 Learning to estimate tissue cell density from synthetic training data

This section describes the two basic steps of our algorithm: i) synthesizing heterogeneous MR images showing tumors, and ii) training a tumor detector which works on *real* patient images.

3.2.1 Generative tumor simulation model

The automatic generation of our synthetic training dataset relies on the publicly available brain tumor simulator presented in [Prastawa 2009]. It builds on an anisotropic glioma growth model [Clatz 2005b] with extensions to model the induced mass-effect and the accumulation of the contrast agents in both blood vessels and active tumor regions. Then, multi-sequence MR images are synthesized using characteristic image textures for healthy and pathological tissue classes (cf. Fig. 3.1 and 3.2).

We generate synthetic pathological cases with varying tumor location, tumor count, levels of tumor expansion and extent of edema. The resulting synthetic cases successfully reproduce mass-effect, contrast enhancement and infiltration patterns similar to what observed in the real cases. The synthetic dataset contains 740 synthetic cases. It includes a large variability of brain tumors ranging from very diffusive tumors, showing a large edema-infiltration pattern without necrotic core, to bulky tumors with a large necrotic core surrounded by an enhanced vascularization pattern. For each case, the simulation provides four MR sequences (cf. Fig. 3.1 and 3.2) which offer different views of the same underlying tumor density distribution.

This synthetic ground truth provides a diverse view of the pathological process including mass-effect and infiltration, but also very detailed annotations for the healthy structures of the brain. The ground truth consists of voxel-wise annotations on the data that are: white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), edema, necrotic tumor core, active tumor rim and blood vessels. Unlike binary annotations which provide a mask for each tissue class, the ground truth consists of a continuous scalar map for each tissue class. Each scalar map provides, for every voxel in the volume, the density of every tissue class.

Chapter 3. Brain tumor cell density estimation from multi-modal MR 40 images based on a synthetic tumor growth model



Figure 3.1: Synthetic MR images. From left to right: T1, T1+Gad, T2, and FLAIR MR images. They show a bulky tumor characterized by a large necrotic and a surrounding vascularization pattern.



Figure 3.2: Synthetic MR images. From left to right: T1, T1+Gad, T2, and FLAIR MR images. They show a very infiltrating tumor characterized by the extent of the edema.

3.2.2 Regression forests for estimating tissue cell density

Problem setting We adapt the regression forest framework similar to the one of [Criminisi 2010b] to train an estimator of tissue cell densities from visual cues in the multi-channel MR images. For each voxel \mathbf{v} , the ground truth provides the density $R_c(\mathbf{v}) \in [0, 1]$ of each tissue class $c \in \mathcal{C}$. The density distribution R is normalized so that it satisfies $\sum_{c \in \mathcal{C}} R_c(\mathbf{v}) = 1$ in every voxel \mathbf{v} .

Feature representation To calculate the local image features – both during training and for predictions – we sub-sample or interpolate all images to $1 \times 1 \times 2 \text{ mm}^3$ resolution. We perform a skull-stripping and an intensity normalization [Coltuc 2006] so that real MR images match the intensity distribution of synthetic MR sequences. Then image features are calculated for each voxel **v**. Features include local multi-channel intensity (T1, T1+Gad, T2, Flair) as well as long-range displaced box features such as in [Geremia 2011]. In addition we also incorporate symmetry features, calculated after estimating the mid-sagittal plane [Prima 2002]. In total, every voxel is associated with a 213-long vector of feature responses.

Regression forest training The forest consists of T trees indexed by t. During training observations of all voxels \mathbf{v} are pushed through each of the trees. Each internal node p applies a binary test $t_p = \tau_{low} \leq \theta(\mathbf{v}) < \tau_{up}$ implementing a double thresholding (τ_{low}, τ_{up}) of the visual feature $\theta(\mathbf{v})$ evaluated at voxel \mathbf{v} . The voxel \mathbf{v} is then sent to one of the two child nodes based on the outcome of this test. As a result, each node p receives a partition of the input training data $\mathcal{T}_p = \{\mathbf{v}, R(\mathbf{v})\}_p$ composed of a voxel **v** and a vector $R(\mathbf{v}) \in [0,1]^{|\mathcal{C}|}$ storing the cell density value for each tissue class. We model the resulting distribution via a multi-variate Gaussian $\mathcal{N}_p(\mu_p, \Gamma_p)$ where μ_p and Γ_p are the mean and covariance matrix of all $R(\mathbf{v}) \in \mathcal{T}_p$, respectively. During training, the parameters (τ_{low}, τ_{up}) of the node test function and the employed visual feature θ are optimized to maximize the information gain. We define the information gain $IG(t_p)$ to measure the quality of the test func-tion t_p which splits \mathcal{T}_p into \mathcal{T}_p^{left} and \mathcal{T}_p^{right} . The information gain is defined as $IG(t_p) = -\sum_{k \in \{left, right\}} \omega_k log \rho_k$ with $\omega = |\mathcal{T}_p^k| / |\mathcal{T}_p|$ and $\rho_k = max|eig(\Gamma_k)|$ where eiq denotes all matrix eigenvalues. In contrast to the more conventional information gain used in [Criminisi 2010b], our formulation gives a robust estimate of the dispersion. Our definition of the information gain focuses on the direction showing maximum dispersion, i.e. ρ_k , and is robust against the missing information on tissue classes.

Indeed, the information gain presented in [Criminisi 2010b] models the dispersions as $|\Gamma_k|$ which would evaluate to 0 in our case, because $\sum_{c \in \mathcal{C}} R_c(\mathbf{v}) = 1$. A possibility to solve this problem would be to reduce the dimension of the vector $R(\mathbf{v})$ so that it only contains independent observations. However, this solution will still provide a null dispersion measure in the case a tissue class is missing from the input partition \mathcal{T}_p . Our definition of the information gain addresses this issue.

At each node p, the optimal test $t_p^* = \arg \max_{\Lambda} IG(t_p)$ is found by exhaustive

search over a random subset of the feature space $\Lambda = \{\tau_{low}, \tau_{up}, \theta\}$. Maximizing the information gain encourages minimizing ρ_p , thus decreasing the prediction error when approximating \mathcal{T}_p with \mathcal{N}_p . The trees are grown to a maximum depth D, as long as $|\mathcal{T}_p| > 100$.

After training, the random forest embeds a hierarchical piece-wise Gaussian model which captures the multi-modality of the training data. In fact, each leaf node l_t of every tree t stores the Gaussian distribution \mathcal{N}_{l_t} associated with the partition of the training data arrived at that leaf \mathcal{T}_{l_t} .

Note that the employed Gaussian distributions have an infinite support which exceeds the definition domain of cell density values [0, 1]. The employed random regression forest approximates the multi-variate distribution R by a piece-wise Gaussian distribution \hat{R} . In the previous description, "Gaussian distribution" should be seen as a synonym for the conjunction of a mean and a covariance. Indeed, we do not use any of the Gaussian distribution's properties nor evaluate it during testing. However, we recall that it is possible to re-parametrize the domain [0, 1] to $[-\inf, +\inf]$, e.g. using *logit*, and thus learn "true" Gaussian distributions without changing the algorithm.

Regression forest prediction When applied to a previously unseen test volume $\mathcal{T}_{test} = \{\mathbf{v}\}$, each voxel \mathbf{v} is propagated through all the trained trees by successive application of the relevant binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, estimated cell densities $r_t(\mathbf{v}) = \mu_{l_t}$ are averaged together to compute the forest tissue cell density estimation $r(\mathbf{v}) = (\sum_{t \in [1..T]} r_t(\mathbf{v}))/T$. Note that in each leaf l_t we maintain an estimate of the confidence Γ_{l_t} associated to the cell density estimation μ_{l_t} .

Active learning and volume subsampling In order to make the training phase computationally tractable, we subsample the training volumes. We constrain the sampling to embed a balanced representation of all tissue classes in the training set. We use active learning to improve the performance of the classifier by increasing the sampling frequence in challenging regions.

Previously, we splitted the synthetic dataset in three disjoint subsets: two training sets, denoted \mathcal{T}_{train}^1 and \mathcal{T}_{train}^2 , and one test set denoted \mathcal{T}_{test} . These different subsets contain $|\mathcal{T}_{train}^1| = |\mathcal{T}_{train}^2| = 250$ cases and $|\mathcal{T}_{test}| = 240$ cases, respectively.

First, we train a forest \mathcal{F}^1 (T = 80 trees of depth D = 20) on a random, but balanced, subset of voxels extracted from \mathcal{T}^1_{train} . Then, we use the forest \mathcal{F}^1 to predict the tissue cell density at a voxel level for all the cases in the datset \mathcal{T}^2_{train} . The results show high error margins in regions showing high appearance ambiguities, e.g. regions around the GM-WM boundary (cf. Fig. 3.3 and 3.4) which strongly ressembles edema. The forest \mathcal{F}^1 fails at discriminating the GM-WM boundary from edema because the bounday regions were sampled in a very sparse way from \mathcal{T}^1_{train} . In order to address this issue, we propose to use a fully automatic sampling method based on active learning. The sampling method consists of creating a new training set from the regions in which the forest \mathcal{F}^1 made significant errors. The error made by the prediction with respect to the ground truth is computed using the norm of the difference $||R(\mathbf{v}) - r(\mathbf{v})||$. A fixed proportion (50%) of the voxels with highest error values in \mathcal{T}_{train}^2 are added to the original random sampling to train a new forest \mathcal{F}^2 (T = 80trees of depth D = 20).

Finally, we merge the obtained forests into a larger one $\mathcal{F} = \{\mathcal{F}^1, \mathcal{F}^2\}$, and use it to predict the tissue cell densities on the synthetic cases from \mathcal{T}_{test} . Results are reported in the next section.

3.3 Experiments

We conducted two main experiments. First, as a proof of concept, we tested how well the learned forest reproduces the tissue cell densities in the synthetic model. In a second experiment we applied our method to real, previously unseen, clinical images and measured accuracy by comparing the detected and ground truth tumor outlines.

We evaluate the predictions for every test case using two complementary metrics: a segmentation metric and a robust regression metric. The segmentation metric compares binary versions of the physiological maps, independently normalized for each tissue class. The binary masks are obtained by thresholding the prediction and the ground truth at the same value. Then, we evaluate the true postive rate TPR = TP/(TP + FN), the false positive rate FPR = FP/(FP + TN) and the positive predictive value PPV = TP/(TP + FP), where TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. Finally, we compute the area under the ROC and the one precision-recall curves to measure how well the prediction fits the ground truth.

The robust regression metric evaluates the estimation error between the predicted continuous map and the ground truth. For every tissue class c, we compute the mean over the voxels v of the estimation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$. In order to avoid artificial decrease of the mean error, we make this metric robust by only considering regions of the physiological map showing at least 1 $\%_0$ signal in either the prediction or the ground truth.

In both experiments, we used the same forest containing T = 160 trees of depth D = 20 trained on 500 synthetic cases. The values of these meta-parameters were tuned by training and testing on a different synthetic set.

3.3.1 Experiment 1: Estimating cell density in synthetic cases

We tested the random forest on a previously unseen synthetic dataset with 240 cases. Results (Fig. 3.3 and 3.4) show a good qualitative match between predicted and ground truth physiological maps. As a segmentation metric we calculate the true and false positive rates as well as the positive predictive value for each possible threshold jointly on r and R and summarize it through ROC and precision-recall

Chapter 3. Brain tumor cell density estimation from multi-modal MR 44 images based on a synthetic tumor growth model

curves. For every tissue class c, we also compute the mean approximation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$ (integrating over voxels with > .001 tumor cell density for tumor classes). Results in Fig. 3.3, 3.4, 3.5, 3.6 and show excellent results for WM, GM, CSF. The predicted tumor cell density is in good agreement with ground truth, although a systematic bias leads to a slightly larger variance in the error metric (cf. Fig. 3.7). This effect is stable over all data sets and we can correct for it.

The synthetic training and testing tests were generated using a single BrainWeb phantom. As a result, there is very little anatomical variability in the healthy regions of the brain (WM, GM, CSF) which is a limitation of the generated synthetic dataset. However, some variability can be observed in the regions affected by the mass effect induced by the tumor growth simulation. Moreover, the image synthesis step produces different noise mapping and bias field for each image resulting in heterogeneous intensity patterns within the same brain tissue region which makes segmentation and regression more challenging.

3.3.2 Experiment 2: Segmenting tumors in clinical images

We tested the same random forest on 14 clinical cases showing low and high grade glioma (cf. Fig. 3.9, Fig. 3.10, 3.11, 3.12) with T1, T1+Gad, T2 and FLAIR images. None of the clinical cases was used during training. Training was done exclusively on synthetic images. The manually-obtained ground truth consists of a binary tumor mask delineating the tumor+edema region. We calculated the same tumor outline from the predicted continuous physiological masks as done for the synthetic model [Prastawa 2009]. Segmentation results (Fig. 3.8) are in excellent agreement with a state-of-the-art unsupervised multimodal brain tumor segmentation method that also outperformed standard EM segmentation in an earlier study [Menze 2010]. Note that the method presented in [Menze 2010] significantly outperformed [Prastawa 2004]. Interestingly, in a qualitative evaluation (cf. Fig. 3.9 and 3.10), the tumor cell density map shows smooth transition between the active rim of the tumor (red) and the edema (green).

3.4 Discussion

In Section 3.3.1, we showed that the regression random forest successfully tackles the multi-regression problem on the synthetic cases. Interestingly, the regression random forest succeeded in recovering the tumor cell density from the image information. As a result, we obtained a collection of continuous cell density maps retrieving the semantics behind the MR images.

In Section 3.3.2, we show that our method achieves state of the art brain tumor segmentation with the added benefit of providing a tumor cell estimate which can be used to initialize and, hence, to speed up more expensive model inversion schemes, such as [Menze 2011]. We observed that for high grade gliomas the cell densities associated with the necrotic core and the active rim were mixed. This is due to the



Figure 3.3: Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels including the active tumor rim (orange), and necrotic core (yellow).



Figure 3.4: Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels including the active tumor rim (orange), and necrotic core (yellow).



Figure 3.5: Prediction of the cell densities along a section of the tumor. Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively, compared to the ground truth (dotted line).



Figure 3.6: **Prediction of the cell densities along a section of the tumor.** Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively, compared to the ground truth (dotted line).



Figure 3.7: Evaluation of the predictions on the synthetic dataset for each cell density map. Each row represents a tissue class: WM, GM, CSF, edema, necrotic core, blood vessels including the active tumor rim, respectively. Top, from left to right: area under the precision-recall curve, and area under the ROC curve. Each point of the ROC and precision-recall curves is built by thresholding the prediction and the ground truth at the same value. Bottom, from left to right: estimation of the mean prediction error, and the dice score. The ground truth and the prediction density maps were thresholded at the same value, i.e. 0.3.

Chapter 3. Brain tumor cell density estimation from multi-modal MR 50 images based on a synthetic tumor growth model



Figure 3.8: Evaluation of the predictions on the clinical dataset. Box plots of the area under the ROC curve (left), under the precision-recall curve (right), and the dice score. Comparison of the proposed method (G-RF) with the method presented in [Menze 2010].



Figure 3.9: Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).



Figure 3.10: Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).



Figure 3.11: **Prediction of the cell densities along a section of the tumor.** Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively.



Figure 3.12: Prediction of the cell densities along a section of the tumor. Top, from left to right: T1+Gadolinium, FLAIR image, the intensity profile along the section (yellow). Bottom, from left to right: prediction of the cell density for the edema, necrotic core and active rim, respectively.

Chapter 3. Brain tumor cell density estimation from multi-modal MR 54 images based on a synthetic tumor growth model

fact that the intensity normalization step only partially matches the appearance of real tumors to the appearance of synthetic tumors. We expect that updating the simulator to generate images that match a specific scanner and scanner protocol would significantly improve the results of our method.

3.5 Conclusions

In this chapter, we presented a new generative-discriminative algorithm for the automatic detection of glioma tumors in multi-modal MR brain images. A regression forest model was trained on multiple synthetically-generated labelled images. Then the system demonstrated to work accurately on previously unseen synthetic cases. It showed promising results on real patient images which led to state of the art tumor segmentation results. Our algorithm can estimate continuous tissue cell densities both for healthy tissues (WM, GM, CSF) as well as tumoral ones.

In the future, the employed generative-discriminative model can be extended to include other imaging modalities such as DTI and MR Spectroscopic (MRS) images. MRS images give a complementary macroscopic view on the biological processes underlying tumor growth. It is of great interest at the time of predicting the tumor cell density tails invisible from MR images.

The idea of combining complex image synthesis models with data-driven inference is of a general nature and can be extended to other diseases where obtaining expert labels is laborious or expensive or even impossible.

Appendix: Additional qualitative results on synthetic and real cases

In this section, we illustrate obtained results with a serie of additional synthetic (cf. Fig. 3.13, 3.14 and 3.15) and real cases (cf. Fig. 3.16, 3.17 and 3.18).



Figure 3.13: Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels including the active tumor rim (orange), and necrotic core (yellow).



Figure 3.14: Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels including the active tumor rim (orange), and necrotic core (yellow).



Figure 3.15: Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels including the active tumor rim (orange), and necrotic core (yellow).



Figure 3.16: Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).



Figure 3.17: Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).



Figure 3.18: Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).

CHAPTER 4 Spatially Adaptive Random Forests for Classification Problems in Medical Imaging

Contents

4.1 Introduction		••••	61		
4.2 Data representation		••••	63		
4.2.1 Multi-scale image tree \ldots \ldots \ldots \ldots \ldots			63		
$4.2.2 \text{Visual features} \dots \dots \dots \dots \dots \dots \dots \dots \dots $			63		
4.2.3 Ground truth			64		
4.3 Spatially adaptive random forest		• • • •	64		
4.3.1 Training data			67		
4.3.2 Decision node representation			67		
4.3.3 Training			67		
4.3.4 Prediction			68		
4.4 Experiments and results 68					
4.5 Discussion and conclusion		• • • •	69		

Chapter 4 shows how multi-scale image parsing can benefit previously presented classification and regression methods. A general and efficient supervised discriminative algorithm is presented to merge the benefits of multi-scale approaches and random forest methods. "Spatially Adaptive Random Forests" (SARF) find the optimal image sampling associated to a given classification or regression task. Thanks to multi-scale data representation, the computation effort focuses on challenging regions rather than uniformly processing the whole image. SARF is applied to multi-class glioma segmentation from multi-modal MR images. The experiments show state of the art results on the publicly available MICCAI 2012 BRATS dataset. Unlike competing methods, SARF finds the optimal observation scale using a coarse-to-fine strategy, and hence avoids time consuming parsing of the image at the voxel level which is often unnecessary.

4.1 Introduction

Medical imaging protocols produce large amounts of multi-modal volumetric images. The large size of the datasets contributes to the success of supervised discriminative methods for semantic label extraction. Automatic classification of semantically relevant structures in medical images is challenging due to (a) the large size of data volumes, and (b) the severe class overlap in the feature space. Our study focuses on multi-scale segmentation methods which address these issues.

Using multi-scale image representations, relying for example on spectral representations [Cour 2005], helps speeding up segmentation algorithms and apply them to data sets that were previously considered to be of prohibitive size. Interestingly, adaptive multi-resolution hierarchies have also been shown to efficiently encode image information for compression and rendering [Lefebvre 2007]. In medical applications, recent work focused on learning hierarchical anatomical representations explicitly from expert annotations. Alternatively, a generative model can be learnt from expert labelled ground truth and incorporated to a multi-level affinity-based segmentation [Corso 2008]. The approach in [Wolz 2012] builds on a hierarchical registration and weighting scheme to retrieve organ-specific atlases at different scales. Although, these methods can be adapted to take into account additional image channels, they exclude the use of high dimensional features.

Hierarchical representations can be integrated into discriminative supervised learning algorithms. For instance, boosting weak classifiers that are hierarchically trained on different scales showed to significantly reduce training time [dos Santos 2012]. A similar approach was proposed for the segmentation of multiple sclerosis lesions in multi-channel MRIs using random forests [Akselrod-Ballin 2009]. In the latter, segmentations from multiple scales are merged to increase the robustness of the algorithm. In both methods, the image is relabelled at coarse scales to discard regions containing heterogeneous label distributions. As a result, the classifiers miss critical coarse-scale cues which penalizes the performance of the final segmentation. Other supervised approaches build on context-rich random forests for segmentation of brain lesions in MRIs [Geremia 2011, Zikic 2012]. Still, their excellent performance requires a careful tuning of class weights during training, and is highly sensitive to the spatial sub-sampling of the training data for the different classes.

We present the novel "Spatially Adaptive Random Forest" (SARF) to address these shortcomings. SARF is a supervised learning algorithm which aims at automatic semantic label extraction in multi-modal medical images. It builds on discriminative random forests, an efficient multi-scale 3D image representation, and structured labelling. SARF learns the optimal image sampling associated to the segmentation task from the training data. The ground truth, which is provided at the voxel level, is extrapolated to coarse levels by using label histograms. During both training and testing, the algorithm quickly handles background regions of the image, and focuses on more challenging ones to refine the segmentation. This is made possible by adding a scale transition condition to the random forest algorithm.

We demonstrate SARF in the context of multi-class glioma segmentation in multi-modal MR images. SARF ranked in the top three when applied to the publicly available MICCAI 2012 BRATS Challenge dataset.
4.2 Data representation

We derive a hierarchical data representation to efficiently browse the image, and its associated ground truth, at different scales. Fig. 4.1 illustrates the visual results obtained for a representative case in the BRATS glioma dataset.

4.2.1 Multi-scale image tree

A multi-scale hierarchical tree is presented for encoding multi-channel volumetric images. It builds on the volumetric counterpart of SLIC superpixels [Achanta 2012] to iteratively generate a compressed representation of the image at different scales. Previous SLIC implementations aimed at 2D RGB images and 3D mono-channel volumes. In this work, the SLIC algorithm was adapted to multi-channel 3D volumes. It takes into account intensity information from all available channels to extract connected and homogeneous supervoxels. Similarly to spatial trees presented in [Siskind 2007], each layer of the final tree embeds a different scale of the image. The generation of the multi-scale data representation is fast and does not require any parameter tuning. It inherits the strengths of the SLIC algorithm which showed outstanding performance compared to the state-of-the-art. For instance in [Achanta 2012], the authors report the segmentation of a 2D image of size 2048×1536 in 14.94s on an Intel Dual Core 2.26 GHz processor with 2 GB RAM. In our case, using a sequential Matlab implementation, a volume of size $81 \times 98 \times 83$ is segmented in 1min on an IBM e325, dual-Opterons 246 at a maximum frequency of 2Ghz, using a single CPU.

Multi-channel images are defined as the vectorial map $\mathbf{I}: \Omega \subset \mathbb{R}^3 \to \mathbb{R}^{|\Gamma|}$ where Γ is the set of image channels. Thus, every voxel is associated a spatial position $\mathbf{x} \in \Omega$ and a multi-channel intensity vector $\mathbf{I}(\mathbf{x})$. A spatial partition is a set of disjoint supervoxels $\mathcal{P}_{\mathbf{I}}^k = \{v_n^k\} \subset \mathcal{P}(\Omega)$ partitioning the image domain Ω at scale k. As presented in [Achanta 2012], the generation of $\mathcal{P}_{\mathbf{I}}^k$ is exclusively based on the image information \mathbf{I} . At the finest scale, $\mathcal{P}_{\mathbf{I}}^1$ is the set of singletons formed by the individual image voxels. We adapt the SLIC algorithm to recursively cluster supervoxels from $\mathcal{P}_{\mathbf{I}}^k$ into a coarser partition noted $\mathcal{P}_{\mathbf{I}}^{k+1}$. In practice, a scalar value s^k controls the maximum size of supervoxels at scale k.

This procedure is repeated for an increasing sequence of supervoxel sizes $\{s^k\}_{k \leq K}$ until the whole image is contained in a single supervoxel. In the rest of the article, we set K = 7 and $\{s^k\}_{k \leq 7}$ to a fixed set of scalars. The resulting sequence $\{\mathcal{P}_{\mathbf{I}}^k\}$ is encoded in the layers of a multi-resolution tree noted $\mathcal{M}_{\mathbf{I}} = \{\mathcal{M}_{\mathbf{I}}^k\}$. At layer k > 1, $\mathcal{M}_{\mathbf{I}}^k$ maps every supervoxel $v \in \mathcal{P}_{\mathbf{I}}^k$ to a set of disjoint finer supervoxels $\mathcal{M}_{\mathbf{I}}^k(v) = \{w_i\} \subset \mathcal{P}_{\mathbf{I}}^{k-1}$ satisfying $v = \bigcup_i w_i$. Inter-scale relationships are illustrated in Fig. 4.2.

4.2.2 Visual features

Supervoxels provide a convenient primitive from which to compute local image features. A visual feature at scale k is a map defined by $\theta^k : \mathcal{P}^k_{\mathbf{I}} \to \mathbb{R}$. Arbi-

trary large amounts of task-specific features can be derived in a straightforward way from the multi-scale representation of the image [Geremia 2011, Zikic 2012, Akselrod-Ballin 2009].

Here, we provide three examples of possible features. A local feature $\theta_{med}^{k,\gamma}$ which maps the supervoxel v to the median intensity in channel γ of the voxels it contains, noted $\theta_{med}^{k,\gamma}(v) = \mu_{\frac{1}{2}} \{ \mathbf{I}_{\gamma}(\mathbf{x}) | \mathbf{x} \in v \}$. A prior feature $\theta_{prior}^{k,\delta} = \theta_{med}^{k,\delta}$ where the channel δ maps the spatial distribution of healthy tissues including white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF). The prior feature is obtained by affinely registering the MNI atlases on each patient. A long-range feature defined by $\theta_{sym}^{k,\gamma}(v) = \max\{S \circ \mathbf{I}_{\gamma}(\mathbf{x}) | \mathbf{x} \in v\}$, where S is a reflection of \mathbb{R}^3 . The symmetry feature was specifically designed for the detection of abnormal regions in brain MRIs which are often asymmetrical with respect to the mid-sagittal plane.

4.2.3 Ground truth

At the voxel level, the ground truth associated to the image \mathbf{I} is defined by $G_{\mathbf{I}}$: $\Omega \to \mathcal{C}$, each voxel \mathbf{x} being associated a class label $G_{\mathbf{I}}(\mathbf{x}) \in \mathcal{C}$. Here, we consider the tissue classes $\mathcal{C} = \{back, edema, core\}$ standing for background healthy brain, edema and tumor core, respectively. The generalization of $G_{\mathbf{I}}$ to coarser scales reads $\mathcal{G}_{\mathbf{I}} = \{\mathcal{G}_{\mathbf{I}}^k\}$ where $\mathcal{G}_{\mathbf{I}}^k : \mathcal{P}_{\mathbf{I}}^k \to \mathcal{C}$.

In previous work [dos Santos 2012, Akselrod-Ballin 2009], each supervoxel v was affected the class label $\mathcal{G}_{\mathbf{I}}^{k}(v) = c \in \mathcal{C}$ satisfying $|\{\mathbf{x} \in v \mid G_{\mathbf{I}}(\mathbf{x}) = c\}| > \tau_{hom}$. When such label did not exist, the supervoxel was removed from the training set. The threshold $\tau_{hom} = 70$ or 80% aims at selecting homogeneous supervoxels, while discarding those showing severe label mixture. In multi-class segmentation, this means discarding challenging, but often critical, image regions, and thus indirectly penalizing the prediction performance.

To address this flaw, we introduce an unambiguous labelling function $\mathcal{H}_{\mathbf{I}}^{k}$ with values in $\mathbb{N}^{|\mathcal{C}|}$. The histogram $\mathcal{H}_{\mathbf{I}}^{k}(v) = (h[c])_{c\in\mathcal{C}} = (|\{\mathbf{x} \in v \mid G_{\mathbf{I}}(\mathbf{x}) = c\}|)_{c\in\mathcal{C}}$ counts the class label occurrences in the supervoxel v. Consequently at scale k, the ground truth is defined as $\mathcal{G}_{\mathbf{I}}^{k}(v) = \arg \max_{c\in\mathcal{C}} h[c]$. Unlike [dos Santos 2012, Akselrod-Ballin 2009], our labelling method $\mathcal{H}_{\mathbf{I}} = {\mathcal{H}_{\mathbf{I}}^{k}}$ keeps track of the class mixture in every supervoxel. In Section 4.3.3, we explain how this is integrated to the random forest framework to help refining the segmentation in challenging image regions.

4.3 Spatially adaptive random forest

The random forest framework [Breiman 2001] is extended to benefit from the presented multi-scale image representation. In the following, we provide a sound formulation of SARF applicable to the general problem of multi-class image segmentation. Fig. 4.3 provides a schematic illustration on the multi-scale aspect of the tree and feature space.



Figure 4.1: **Data representation.** Top: the multi-channel MRI noted I, including (T1, T1+gadolinium, T2, FLAIR) overlayed with the expert annotations for the edema (yellow) and the tumor core (red). Bottom: the affinely registered MNI atlas (WM, GM, CSF), and the image partitions at three different scales ($\mathcal{P}_{I}^{6}, \mathcal{P}_{I}^{4}, \mathcal{P}_{I}^{2}$). The arrows point out the necrotic core (NC), the active proliferative rim (AR) of the tumor and the surrounding edema (Ed) in the intermediate supervoxel decomposition of the image.



Figure 4.2: **Partition refinement.** Top: a schematic illustration of supervoxel refinement at three different scales. Bottom: the associated image partitions considered $(\mathcal{P}_{I}^{6}, \mathcal{P}_{I}^{4}, \mathcal{P}_{I}^{2})$.

4.3.1 Training data

The SARF is an ensemble of trees, each processing the multi-scale data hierarchy from coarse to fine. During training, the data entering the root node of each tree consists of all supervoxels $v_j \in \bigcup_n \mathcal{P}_{\mathbf{I}_n}^K$, where *n* indexes the case, considered at the coarsest scale *K*. In the following, \mathcal{M} , \mathcal{G} and \mathcal{H} , denote the extensions to the dataset $\{\mathbf{I}_n\}$ of $\mathcal{M}_{\mathbf{I}}$, $\mathcal{G}_{\mathbf{I}}$, and $\mathcal{H}_{\mathbf{I}}$, respectively. Every supervoxel v_j is associated the class label $c_j = \mathcal{G}^k(v_j)$, and the label histogram $h_j = \mathcal{H}^k(v_j)$. The resulting training data reads $\mathcal{T} = \{(v_j, c_j, h_j)\}$ with $j \in J$.

4.3.2 Decision node representation

Each internal node p applies a binary test $t_{J_p}^{k,\theta,\tau}$ to the data it receives $\mathcal{T}_p^k = \{(v_j, h_j)\}_{j \in J_p}$. The binary test is defined by $t_p^{k,\theta,\tau}(v_j) = \theta^k(v_j) > \tau$. It is parametrized by the scale k, the type of visual feature θ and the threshold τ . Based on the outcome of this test, \mathcal{T}_p^k is splitted into $\mathcal{T}_{L(p)}^k$ and $\mathcal{T}_{R(p)}^k$, which are propagated to the left and right child node receptively.

During training, every node p stores the scale k, the optimal parameters θ and τ used to split the data. Additionally, it saves the label distribution $d_{\mathcal{T}_p^k} = (|\{j \in J_p \mid c_j = c\}|)_{c \in \mathcal{C}}$, and the class mixture $h_{\mathcal{T}_p^k} = \sum_{j \in J_p} h_j$ considered on \mathcal{T}_p^k .

4.3.3 Training

For each node p, the parameter $\lambda = (\theta, \tau)$ is optimized using the input training data \mathcal{T}_p^k . The optimality criterium is the information gain defined as $IG(\lambda, \mathcal{T}_p^k) = H(d_{\mathcal{T}_p^k}) - \sum_{B \in \{L,R\}} w_B \cdot H(d_{\mathcal{T}_{B(p)}^k})$, where $w_B = |\mathcal{T}_{B(p)}^k| / |\mathcal{T}_p^k|$. The entropy H satisfies $H(d_{\mathcal{T}_p^k}) = -\sum_{c \in \mathcal{C}} P(c) \cdot \log P(c)$ with $P(c) = d_{\mathcal{T}_p^k}[c] / \sum_{c' \in \mathcal{C}} d_{\mathcal{T}_p^k}[c']$. The optimal parameters satisfy $\lambda_p^* = \arg \max_{\lambda} IG(\lambda, \mathcal{T}_p^k)$.

This procedure is repeated recursively for the derived nodes. The tree is grown down to scale k = 1 until every leaf node p is pure, i.e. $H(\mathcal{T}_p^1) = 0$. Unlike previous work [Akselrod-Ballin 2009, Geremia 2011, Zikic 2012], we introduce spatial refinement in the random forest framework to capture fine structures. Indeed, when the supervoxels are too large to properly describe annotated image regions, the scale k is decremented. Formally, this occurs when $H(d_{\mathcal{T}_p^k}) = 0$ and $H(h_{\mathcal{T}_p^1}) \neq 0$. In this case, $\mathcal{T}_p^k = \{(v_j, h_j)\}_{j \in J_p}$ is replaced by its decomposition into finer supervoxels $\mathcal{T}_p^{k-1} = \{(w_i, c'_i, h'_i)\}_{i \in I_p}$ where $w_i \in \bigcup_j \mathcal{M}^k(v_j), c'_i = \mathcal{G}^{k-1}(w_j)$, and $h'_i = \mathcal{H}^{k-1}(w_j)$. The node is then optimized at the scale k - 1.

We solve the optimization problem by exhaustive search over a random set of thresholds and the whole set of feature $\{\theta_{med}, \theta_{prior}, \theta_{sym}\}$. To further decorrelate the weak classifiers, we train each tree with a different partition of the same image. This is done by randomizing the initial position of the seeds in the SLIC algorithm [Achanta 2012].



Chapter 4. Spatially Adaptive Random Forests for Classification

Figure 4.3: Schematic representation of the SARF algorithm. Right: Each tree of the forest successively refines the scale of observation. Left: at each node, the feature space is partitioned by considering the supervoxels at a fixed scale. The class labels of coarse supervoxels (here blue and red) are defined by majority voting of finer scales.

4.3.4Prediction

When applied to an unseen test volume \mathbf{I}_{test} , a different image partition $\mathcal{P}_{\mathbf{I}_{test}}^{K,t}$ is computed for every tree $t \in [1..T]$. In every tree, each node p applies the binary test t^{k,θ^*,τ^*} to the input data, after having refined it to scale k, if necessary. As a result, for every voxel $v \in \mathcal{P}_{\mathbf{I}_{test}}^{1,t}$ there is a tree-specific sequence of super-voxels $\{v_t^k\}_{k \in [k_{min},K]}$ of decreasing size such that $v_t^k \in \mathcal{P}_{\mathbf{I}_{test}}^{k,t}$. The finest supervoxel $w_t = v_t^{k_{min}}$ reaches the leaf node p_t , and is affected the associated posterior class distribution $d_t(w_t) = d_{\mathcal{T}_{p_t}^k} / \sum_{c \in \mathcal{C}} d_{\mathcal{T}_{p_t}^k}[c] \in \mathcal{L}$. For every voxel $v \in \mathcal{P}_{\mathbf{I}_{test}}^1$, posteriors from all trees are averaged to from the forest posterior such that $d_f(v) = \sum_t d_t(w_t)/T$. Finally, the predicted class label affected to v is $\widehat{G}_{\mathbf{I}_{test}}(v) = \arg \max_{c \in \mathcal{C}} d_f(v)[c]$. A morphological closure is applied to the necrotic core class as only postprocessing.

4.4Experiments and results

We demonstrate SARF in the specific context of multi-class glioma segmentation in multi-modal MRIs. For training purposes, we rely on the publicly available MIC-CAI BRATS Challenge 2012 dataset which contains 80 cases. Brain tumor image data used in this work were obtained from the MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation (http://www.imm.dtu.dk/projects/BRATS2012)



Figure 4.4: Challenge results and scale map. The image displays the FLAIR and T1+gadolinium images overlayed with the predicted tumor core + edema contour in yellow. The right most image is the scale map, displaying for each voxel the finest scale used to classify it. Dark blue stands for the finest scales while red stands for the coarsest ones.

organized by B. Menze, A. Jakab, S. Bauer, M. Reyes, M. Prastawa, and K. Van Leemput. The challenge database contains fully anonymized images from the following institutions: ETH Zurich, University of Bern, University of Debrecen, and University of Utah. Our method was then applied to a test dataset of 30 cases which ground truth was kept secret. After independent evaluation of the results by the challenge website, SARF ranked third among eight state-of-the-art methods. Obtained results are illustrated in Fig. 4.4 and 4.5.

4.5 Discussion and conclusion

The presented SARF framework shows promising results still limited by the small size of the employed feature set. Indeed, it would greatly benefit from richer visual features designed for glioma segmentation, and subsequent regularization as used in top-ranked methods [Zikic 2012, Bauer 2011].

Fig. 4.4 shows that the SARF focuses on challenging image regions by processing them at finer scales. Interestingly, SARF automatically finds the average optimal scale used to segment gliomas, here k = 3. This supports the fact that parsing the image at the voxel level is often unnecessary. These findings arise from the novelty of the multi-scale classification approach implemented in SARF, and result in significantly reducing training and testing times.



Figure 4.5: Challenge qualitative results on four different cases. Top: the figure displays the FLAIR and T1+gadolinium (T1C) images overlayed with the predicted edema (yellow) and tumor core (red). Bottom: the Dice scores of the top-ranked methods for the edema and tumor core classes

Contents

5.1	Conclusions	71
5.2	Perspectives	73

Random Forest is an example of a tool that is useful in doing analyses of scientific data. But the cleverest algorithms are no substitute for human intelligence and knowledge of the data in the problem. Take the output of random forests not as absolute truth, but as smart computer generated guesses that may be helpful in leading to a deeper understanding of the problem. Leo Breiman and Adele Cutler ¹

5.1 Conclusions

This chapter goes over the partial conclusions of previous chapters and brings them together to answer the introductory question: *How to make the most out of medical images to support diagnosis in the best possible way*?

- Spatial random forests for MS lesions segmentation. We demonstrated the power of the RF formulation applied to the difficult task of MS lesion segmentation in multi-channel MR images. We presented three kinds of 3D features based on multi-channel intensity, prior and context-rich information. Those features are part of a context-rich random decision forest classifier which demonstrated improved results on one of the state of the art algorithms on the public MS challenge dataset. In addition, the random decision forest framework provided a means to automatically select the most discriminative features to achieve the best possible segmentation. We provided a thorough methodology to estimate the optimal forests meta-parameters (cf. Section 2.5.4). Future work could include the use of more sophisticated features to reduce even further the preprocessing requirements. One could also explore the application of our approach to the segmentation of brain tumors in multi-sequence MR images of the brain. Finally, one could investigate an extension of the proposed approach to larger multi-class problems in order to try to simultaneously segment brain tissues (WM, GM, CSF) along with MS lesions.
- Multi-variate regression forests for glioma cell density extrapolation. This paper presented a new generative-discriminative algorithm for the

 $^{^{1}}http:www.stat.berkeley.edu \widetilde{b}reiman Random Forests cc_philosophy.htm$

automatic detection of glioma tumors in multi-modal MR brain images. A regression forest model was trained on multiple synthetically-generated labelled images. Then the system demonstrated to work accurately on previously unseen synthetic cases. It showed promising results on real patient images which led to state of the art tumor segmentation results. Our algorithm can estimate continuous tissue cell densities both for healthy tissues (WM, GM, CSF) as well as tumoral ones. The idea of combining complex image synthesis models with data-driven inference is of a general nature and can be extended to other diseases where obtaining expert labels is laborious or expensive or even impossible.

• Spatially Adaptive Random Forests for Classification Problems in Medical Imaging. The presented SARF framework focuses on challenging image regions by processing them at finer scales. Interestingly, SARF automatically finds the average optimal scale used to segment gliomas. This supports the fact that parsing the image at the voxel level is often unnecessary. These findings arise from the novelty of the multi-scale classification approach implemented in SARF, and result in significantly reducing training and testing times.

Our C++ implementation of the random forest (cf. Chapter 2) achieved segmentation of MR scans $(159 \times 207 \times 79 \text{ voxels})$ in less than 1min 30s for each tree. This performance holds for the IBM e325, dual-Opterons 246 at a maximum frequency of 2Ghz, using a single CPU. The best performance was achieved by generating and evaluating the random forest in a depth-first fashion. Note that each tree of the forest was trained and tested using a cluster computing approach.

Random forests offer an efficient tool to process MRIs and demonstrated excellent performance when applied to challenging segmentation tasks. More importantly, they provide an elegant way of ranking the visual information MRIs contain. Interestingly, in the case of MS lesions segmentation, the algorithm emphasizes the importance of the FLAIR image which is the preferred MR sequence for glioma segmentation in current clinical practice. Spatially Adaptive Random Forest is able to retrieve the optimal spatial scale to perform a specific task. This is made possible by automatically recognizing characteristic visual features and successively refining the observation scale.

With these complementary approaches, we proposed a method to segment brain lesions in MRIs which mimics the human visual search process. A human observer would, indeed, parse the image from coarse shapes to fine details, successively focusing on interesting image regions. Because they mimic a natural and intuitive process, the resulting hierarchical classifiers are easy to interpret and to tune. This insight is essential to improve our understanding of the data and to tackle new challenges.

5.2 Perspectives

The work presented in this thesis is a humble step contribution to the development of reliable tools to support clinical diagnosis. Here, we detail promising research directions towards this goal:

- Intensity normalization. The visual features feeding the random forest framework presented in this thesis are highly dependent on the MR intensity signal. Unlike in CT images, MR intensity is not calibrated on any physiological value. As a result, the exact same tissue can have very different intensity responses depending on the employed scanner or on the parametrization of the sequence used. In a supervised setting where tissues are classified according to their intensity, it is critical to normalize intensity in images originating from different scanners. In this thesis, we experimented three different methods [Rey 2002, Coltuc 2006, Margeta 2011] with convincing results for scans acquired in the same data center. However, calibrating MR images coming from different data centers showed to be a far more challenging task and would benefit from further investigation.
- Multi-scale random forests with spatial context. Spatially adaptive random forests provide a general formulation for multi-scale classification of large image volumes. The hierarchical supervoxel image decomposition provides, without any computational overhead, a fast way to compute geodesic trees in the image space [Criminisi 2011a]. Long-range features [Geremia 2011, Zikic 2012] and normalized intensity-based features [Akselrod-Ballin 2009] associated to geodesic proximity [Criminisi 2010a] might positively impact segmentation results by reducing the dependance on intensity calibration. Recent works [Kim 2011, Jagadeesh 2012] pointed out to the benefits of considering powerful high-order data representations: hypergraphs. We believe that these improvements can lead to a very competitive supervised algorithm for the general purpose of image segmentation.
- Learning complex tasks from realistic generative models. In the future, the employed generative-discriminative model can be extended to include other imaging modalities such as DTI and MR Spectroscopic (MRS) images. MRS images give a complementary macroscopic view on the biological processes underlying tumor progression. They provide low resolution cues on infiltrated regions which are invisible from MR sequences. This information can be used to estimate the invisible tumor cell density beyond the tumor margins. This approach would be highly relevant in the context of radiotherapy, which aims at targeting cancerous tumor cells infiltrated beyond the tumor margins. We also generate information about the location of major blood vessels which is critical for surgical planning.
- Semantic segmentation and cell density estimation for glioma grade characterization. Semantic labelling of gliomas and cell density estimation is

a preliminary step in the way of glioma characterization. Indeed, they provide case-specific estimations of the cell density and the tissue class for every voxel in the image. This information can be used to initialize and, hence, to speed up more expensive model inversion schemes, such as [Menze 2011]. It can also be used to initialize complex simulation algorithms [Bresch 2010] which require the tumor cell distribution map as input.

Book Chapter

• [Geremia 2013a] E. Geremia, D. Zikic, O. Clatz, B. H. Menze, B. Glocker, E. Konukoglu, J. Shotton, O. M. Thomas, S. J. Price, T. Das, R. Jena, N. Ayache and A. Criminisi. Classification Forests for Semantic Segmentation of Brain Lesions in Multi-Channel MRI. In Antonio Criminisi and Jamie Shotton, editeurs, Decision Forests in Computer Vision and Medical Image Analysis. Springer, 2013. To appear.

Journal Paper

• [Geremia 2011] E. Geremia, O. Clatz, B. H. Menze, E. Konukoglu, A. Criminisi and N. Ayache. Spatial decision forests for MS lesion segmentation in multichannel magnetic resonance images. NeuroImage, vol. 57, no. 2, pages 378-390, 2011.

Conference Articles

- [Geremia 2010] E. Geremia, B. H. Menze, O. Clatz, E. Konukoglu, A. Criminisi and N. Ayache. Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images. In Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), LNCS, Beijing, China, September 2010. Springer.
- [Margeta 2011] J. Margeta, E. Geremia, A. Criminisi and N. Ayache. Layered Spatio-temporal Forests for Left Ventricle Segmentation from 4D Cardiac MRI Data. In Proceedings of STACOM Workshop at MICCAI 2011. Springer, September 2011.
- [Geremia 2012b] E. Geremia, B. H. Menze, M. Prastawa, M.-A. Weber, A. Criminisi and N. Ayache. Brain tumor cell density estimation from multimodal MR images based on a synthetic tumor growth model. In MICCAI Workshop on Medical Computer Vision, LNCS. Springer, October 2012.
- [Geremia 2012a] E. Geremia, B. H. Menze and N. Ayache. Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images. In MICCAI Challenge on Multimodal Brain Tumor Segmentation, October 2012.
- [Menze 2012] B. H. Menze, E. Geremia, N. Ayache and G. Szekely. Segmenting glioma in multi-modal images using a generative-discriminative model for

brain lesion segmentation. In MICCAI Challenge on Multimodal Brain Tumor Segmentation, October 2012.

- [Stretton 2012] E. Stretton, E. Mandonnet, E. Geremia, B. H. Menze, H. Delingette and N. Ayache. Predicting the Location of Glioma Recurrence After a Resection Surgery. In Proceedings of 2nd International MICCAI Workshop on Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data (STIA'12), LNCS, Nice, October 2011. Springer.
- [Geremia 2013b] E. Geremia, B. H. Menze and N. Ayache. Spatially Adaptive Random Forests for Classification Problems in Medical Imaging. In ISBI. IEEE, 2013. submitted.

In this appendix, we present a new method for fully automatic left ventricle segmentation from 4D cardiac MR datasets. To deal with the diverse dataset, we propose a machine learning approach using two layers of spatio-temporal decision forests with almost no assumptions on the data nor explicitly specifying the segmentation rules. We introduce 4D spatio-temporal features to classification with decision forests and propose a method for context aware MR intensity standardization and image alignment. The second layer is then used for the final image segmentation. We present our first results on the STACOM LV Segmentation Challenge 2011 validation datasets. This work was published in [Margeta 2011].

B.1 Introduction

The left ventricle plays a fundamental role in circulation of oxygenated blood to the body. To assess its function, several indicators are often calculated in clinical practice. Many of these are based on ventricular volume and mass measurements at reference cardiac phases. To calculate these an accurate delineation of the myocardium and the cavity is necessary. To remove the bias and variance of manual segmentation, and obtain reproducible measurements, an automatic segmentation technique is desirable.

Compared to computed tomography (CT), cardiac magnetic resonance imaging (cMRI) offers superior temporal resolution, soft tissue contrast, no ionizing radiation, and a vast flexibility in image acquisition characteristics. As a disadvantage, cMRI scans often yield significantly lower resolution in the plane orthogonal to the plane of acquisition, the images can suffer from magnetic field inhomogeneities and respiration artifacts can manifest as slice shifts. Moreover, the lack of standard units (compared to the Hounsfield scale in CT) makes it difficult to directly apply most of the intensity based segmentation techniques.

Motivated by the success of Lempitsky et al. [Lempitsky 2009] in myocardium segmentation from 3D ultrasound sequences in near real time and Geremia et al.[Geremia 2011] for multiple sclerosis lesion segmentation, we propose a fully automated voxel-wise segmentation method based on decision forests (DF) with no assumptions on shape, appearance, motion (except for periodicity and temporal ordering) or knowledge about the cardiac phase of the images in the sequence. The left ventricle segmentation problem is defined as the classification of voxels into myocardium and background.

Instead of robustly registering to an atlas [Shi 2011], building a model [Lu 2011]

or running a highly specialized segmentation algorithm we leave the learning algorithm to automatically decide the relevant features for solving the segmentation problem using the provided ground-truth only. In principle, any pathology can be learnt once a similar example is represented within the training dataset. The previously used decision forests [Lempitsky 2009][Geremia 2011] rely on features that work best when image intensities and orientations are very similar. To tackle the highly variable dataset, we propose a layered learning approach, where the output of each layer serves a different purpose. The first layer is used to prepare the data for a more semantically meaningful and accurate segmentation task in the second layer.

The main contributions of this paper are: a method to use decision forests to solve the MR intensity standardization problem (Section B.3.1) and, similarly, perform a context sensitive rigid registration (Section B.3.2) to align all images to a reference pose. We also suggest a way to introduce temporal dimension into the currently used 3D random features (Section B.2.2). Using the intensity standardized and pose normalized images, which we add spatial information to, we then train a second forest layer (Section B.4). This helps the trees to automatically build their own latent shape representation.

B.1.0.1 Dataset.

STACOM 2011 LV segmentation challenge data [Fonseca 2011] were divided into two sets. Training set (100 3D+t short axis (SA) volumes with manually delineated myocardia at each cardiac phase) and validations sets (5×20 3D+t SA volumes with no delineation provided).

This dataset clearly shows the anatomical variability of heart shape and appearance and some of the main issues of cMRI mentioned above.

B.2 Layered spatio-temporal decision forests

Decision forests are an ensemble supervised learning method consisting of a set of binary decision trees. The training set contains a set of feature measurements and associated labels (myocardium/background) for each of the voxels in the set.

The trees are built in a top-down fashion, from the root, down to the leaves. At each node, local features and a randomly sampled subset of context-rich features are considered for feature selection. Random sampling of the features leads to increased inter-node and inter-tree variability and improved generalization. Each feature θ can be regarded as a binary decision (in our case $\tau_l < \theta < \tau_h$) that splits the original set into two disjoint subsets. The trees then select the most discriminative features for each split such that the information gain is maximized. The data division then recursively continues until a significant part of the voxels at the node belongs to a single class or no significant information gain can be obtained by further splitting. The node then becomes a leaf. The averaged class distributions of all the leaves in the forest reached by the voxel then represent the posterior probabilities of it belonging to either the myocardium or the background. See Geremia et al. [Geremia 2011] for more details.

B.2.1 Strategy to learn from spatio-temporal data

In our approach, we serially train two layers of decision forests, each with the aim to learn to segment, but using slightly modified training data and features. Training with all the 3D+t data was not feasible within the time limits of the challenge, therefore a reduced strategy was designed. This strategy is repeated for each tree:

- 1. Select a random subset of k 4D volumes from the whole training set
- 2. Randomly choose a reference 3D frame I^c for each selected 4D volume
- 3. Select two frames I^{c-o} , I^{c+o} with a fixed offset o on both sides from the reference cardiac image I^c (with periodic wrapping at sequence boundaries)
- 4. Train the tree using a set of k triplets (I^c, I^{c-o}, I^{c+o})

To reduce the computational time, the size of the subset for each tree was set to k = 15, and only one fixed offset o = 4 is currently used. The choice of o was made such that the motion between the selected frames is significant even when more stable cardiac phases (end systole or end diastole) are selected as the reference frame and that almost a half of the cardiac cycle could be covered.

B.2.2 Features

We use several features families to generate the random feature pool operating on the triplets of frames. Their overview can be seen on Figure B.2.2).

B.2.2.1 Local features.

Proposed in [Geremia 2011] as an average of intensities in the vicinity of the tested voxel to deal with noise in magnetic resonance imaging:

$$\theta_{I^c}^{loc}(x) = \theta_{I^c}^{loc}([x, y, z]) = \sum_{x'=x-1}^{x' \le x+1} \sum_{y'=y-1}^{y' \le y+1} \sum_{z'=z-1}^{z' \le z+1} I^c([x', y', z'])$$
(B.1)

Although these features are not intensity invariant, they can still quite well reject some highly improbable intensities.

B.2.2.2 Context rich features.

Defined also in [Geremia 2011], for multichannel MR acquisitions as a difference between the local source image intensity I^S and box averages of remote regions in image I^R :

$$\theta_{I^{S},I^{R}}^{CR}(x) = I^{S}(x) - \frac{1}{Vol(R_{1})} \sum_{x' \in R_{1}} I^{R}(x') - \frac{1}{Vol(R_{2})} \sum_{x' \in R_{2}} I^{R}(x')$$
(B.2)

Appendix B. Layered spatio-temporal forests for left ventricle segmentation from 4D cardiac MRI data



Figure B.1: Illustration of image based features extracted from the images. a) Local features $(3 \times 3 \times 3 \text{ box} \text{ average } S \text{ around the source voxel in the current frame } I^c)$ [Geremia 2011]. b) Context rich features [Geremia 2011] measuring the difference between source box average S and the sum of remote region averages R1 and R2. c) Components x,y,z of voxel coordinates as features[Lempitsky 2009]. d) Spatio-temporal context rich features with the current frame as the source image and offset frame $I^{c\pm o}$ as the remote.

The 3D regions R_1 and R_2 are randomly sampled in a large neighborhood around the origin voxel. These capture strong contrast changes and long-range intensity relationships. In our case we define context-rich features as $\theta_{Ic,Ic}^{CR}(x)$.

B.2.2.3 Spatio-temporal context rich features.

The domain of the moving heart can be coarsely extracted by just thresholding the temporal difference magnitude of the image. We propose to exploit this wealth of information and extend the previous context-rich features into the temporal domain by comparing the "current" 3D frame I^c and another frame offset from c by $\pm o$. The temporal context-rich features can be defined as $\theta_{I^c}^{TCR1} = \theta_{I^c,I^{c+o}}^{CR}(x)$ and $\theta_{I^c}^{TCR1} = \theta_{I^c,I^{c+o}}^{CR}(x)$.

Similarly, we measure the differences between the symmetrically offset frames contained in the triplet as $\theta_{Ic}^{TCR2}(x) = \theta_{Ic+o,Ic-o}^{CR}(x)$ and $\theta_{Ic}^{TCR2}(x) = \theta_{Ic-o,Ic+o}^{CR}(x)$. These spatio-temporal features can be seen as an approximation of a temporal differentiation around the center frame. Note that we use both +o and -o to keep some symmetry of the remote region distribution.

B.2.2.4 Voxel coordinates.

Finally, as in [Lempitsky 2009], we can insert absolute voxel coordinates: $\theta_C^X(x) = x_x$, $\theta_C^Y(x) = x_y$, $\theta_C^Z(x) = x_z$ into the feature pool. However, not until these coordinates have a strong anatomical meaning. This happens later, in the second forest layer when the images are reoriented into the standard pose.

B.3. First layer: Decision forests for image intensity standardization and position normalization



Figure B.2: Short (top) and long (bottom) axis views of the posterior probabilities after the first layer. Brighter value means higher probability.

B.2.3 Data preprocessing

To use fast evaluation of previously defined features based on integral images [Viola 2001], it is necessary to have consistent spacing. Therefore, all the volumes were resampled to one of the most common spatial spacings in the dataset (1.56, 1.56, 7.42mm) and temporal sequence length (20 frames).

Intensity ranges of the images were all linearly rescaled to a fixed range. Similarly to Nyúl et al. [Nyúl 1999], we clamp intensities beyond the 99.8 percentile as they usually do not convey much useful information.

B.3 First layer: Decision forests for image intensity standardization and position normalization

Following the above mentioned training subset selection strategy we can train the first layer of the forests. This is done directly on the images after intensity rescaling i.e. images are brought into the same intensity range but have their original poses. Although short axis scans are often acquired close to a position where the ventricular ring is centered, slice orientation is chosen manually during the acquisition, and precise alignment cannot be guaranteed. Therefore we skip the usage of absolute voxel coordinate features at this step.

Several authors (e.g. [Shi 2011]) have proposed to use Haar like features to detect the heart and crop the heart region. Images can be then registered using the cropped volumes. This removes the influence of background structures and improves the success rate for the registration. However, an extraction of the cropped region will not be necessary to perform a robust registration in our case. We train the first layer of the forests on a rather general scenario, to end up with at least a very rough classification performance (see Figure B.2). As we show in the next two sections, using the rough posterior probability map of a tissue belonging to a ventricle this performance can be already good enough for ventricle detection, intensity standardization and alignment onto a reference orientation without any

prior knowledge of the data apart from the ground-truth.

B.3.1 Intensity standardization

MR intensity value differences of the same tissue are significant not only between scanners and acquisition protocols [Shah 2010] but also for the same follow-up patients [Nyúl 1999]. Therefore good intensity standardization is crucial for any intensity based segmentation algorithm. The variance in median intensities of the myocardia between different cases in the STACOM training set is quite large. There is no unique mode and the distribution is fairly spread in the whole intensity range (0, 65535). Median myocardial intensities span range (1954, 36430), with standard deviation of 5956 and inter-quantile range 7663). This is a serious problem for any intensity based segmentation method.

Many of the intensity standardization algorithms [Bergeest 2008] used today are based on the methods of Nyúl et al. [Nyúl 1999][Nyúl 2000] and the alignment of histogram based landmarks (e.g. modes, percentiles or statistics of homogeneous connected regions) by rescaling image intensities with a piecewise linear mapping. Many of these methods do work reasonably well for brain images where the white matter is clearly the most dominant tissue. In cMRI, the largest homogeneous regions would most of the time belong to the lungs, liver or cavities, rather than the myocardium.

However, from the rough image first layer classification we already obtain some information about the strength of the belief in the foreground and background object. We propose to remap the source image intensities by a piecewise linear function such that the weighted median (as median is more robust to outliers than the mean) M_{source}^c of the images is transformed to a reference value M_{ref} . The weighted median is defined as follows:

$$M_{source}^{c} = \arg\min_{\mu} \sum_{x \in I^{c}} w(x) . |I^{c}(x) - \mu|$$
(B.3)

Where x is the voxel iterator and w(x) are the weights (first layer posterior probabilities). We avoid sorting all volume intensities by approximating the weighted median with the weighted version of the P^2 algorithm [Jain 1985][Egloff 2005]. This algorithm dynamically approximates the cumulative probability density function with a piece-wise quadratic polynomial by adjusting positions of just five markers as the weighted samples are streamed in. Each of these markers are associated with their position, percentile and an intensity value corresponding to that percentile. The positions are updated such that they correspond to the sum of weights of samples whose intensity value is smaller than the value the markers hold.

B.3.2 Orientation normalization

In the approach of Lempitsky et al. [Lempitsky 2009] voxel absolute coordinates are used as features directly. This choice cannot be justified without aligning the

images onto a reference pose. Moreover, features we use for classification are not rotation invariant. Therefore if all the volumes could be registered to have the same orientation, the classification would certainly benefit from it. The interpatient cardiac registration is generally a difficult problem due to the high variability in the thoracic cage. Shi et al. [Shi 2011] do first learning based heart detection and then apply a locally affine registration method which they claim to be robust for large differences.

A robust learning based linear inter-patient organ registration was proposed by Konukoglu et al.[Konukoglu 2011]. Here, each organ is represented with a smooth probability map fit to the bounding boxes obtained as a result of a regression forest. Then, registration of these probability maps is performed.

This sigmoid representation is however rather limiting since it disregards the orientation that we would like to correct for. Without any assumptions on the shape of the distribution, we propose to rigidly align the myocardium enhanced first layer posterior probability maps instead. For this step we propose to use a fast and robust rigid block matching registration technique [Ourselin 2000]. The reference we used was chosen randomly among the images where the apex was at least partially closed. A better choice of the reference, is currently out of scope of this paper. However, an algorithm similar to Hoogendoorn et al.[Hoogendoorn 2010] or a generative technique similar to [Iglesias 2011] could be used.

To reduce the computational time, only probability maps of frames from the middle of the sequence are used to estimate the intensity and pose transformations. The same transformations are then applied for all the frames and ground truths in the sequence which will be needed to train the second layer.

B.4 Second layer: Learning to segment with the shape

B.4.1 Using voxel coordinates

Once the images are registered to a reference volume, the voxel coordinates start to encode spatial relationships with respect to the reference coordinate frame and the coordinate features can be now included in training of the second decision forest layer. Moreover, if the intensity standardization step succeeds, the intensities have more tissue specific meaning (at least for the myocardium).

Thanks to the incorporation of coordinate based features, the tree can completely automatically learn its own latent representation of the possible set of shapes, regularize the classification, and help to remove objects far away from the ventricle. However, this step strongly relies on the success of the previous registration step. Currently, only one reference image is used. Registration to multiple targets should therefore improve robustness and alleviate this problem.



Appendix B. Layered spatio-temporal forests for left ventricle

Figure B.3: Short (top) and long (bottom) axis views on the posterior probabilities after the second layer and segmentation results (isocontour of the probability map at 0.5).

B.4.2 Transforming the volumes back

After the classification is done in the reference space, the posterior probability maps can be transformed back to the original reference frame and resampled accordingly. This shows the advantage of a soft classification technique where the final binary mask is obtained by thresholding the transformed non-integer posterior map, thus avoiding some of the interpolation artifacts.

B.5 Results

Here we show the preliminary results of our method. The forest parameters for the first layer were fixed as follows: 20 trees with depth 20 each. To train each tree, 15 triplets of frames were randomly selected from different volumes of the training set (91 volumes in total). For the second layer: 27 trees each with depth 20. For each tree 12 triplets were randomly selected from different volumes of the training set (91 volumes in total). This leads to usage of only 8% triplets from the whole training set. Hence, there is a vast reserve in utilisation of the training data and setting optimal forest sizes. These parameters were chosen rather empirically to fit into the time limits of the challenge.

The following results were obtained after blind evaluation of our classifications on 90 previously unseen test volumes i.e. 25415 slices from the validation dataset

	sensitivity	specificity	accuracy	PPV	NPV	dice	jaccard
	$\frac{TP}{TP+FN}$	$\frac{TN}{FP+TN}$	$\frac{TP+TN}{P+N}$	$\frac{TP}{TP+FP}$	$\frac{TN}{TN+FN}$	$\frac{2 A \cap B }{ A + B }$	$\frac{ A \cap B }{ A \cup B }$
mean	0.6857	0.9897	0.9861	0.4791	0.9962	0.5045	0.3730
median	0.8099	0.9907	0.9875	0.5234	0.9978	0.5995	0.4281
σ	0.3137	0.0077	0.0077	0.2069	0.0046	0.2571	0.2098

Table B.1: Statistics on the per-slice measures of our segmentation results on 90 volumes from the validation dataset calculated from the entire slices with no region of interest specified. The basal and apical slices contribute to the large differences between the mean and median values and also contribute to the higher variance.

by the STACOM LV segmentation workshop organisers (See Table B.1).

In most of the cases, the algorithm was able to correctly identify the left ventricle myocardium (with median specificity of 0.81). This was possible without the need to explicitly define the segmentation rules and problem specific assumptions (e.g. circularity of the myocardium or cavity contrast). It was also not neccessary to include additional information into the training set (e.g. mitral valve plane position or manual segmentation of a frame in the sequence) nor to rely on a robust non-rigid registration technique.

All the measures were calculated per-slice. This way of calculating the measures caused some of them (specificity, accuracy and NPV) to reach high values but also to have less explicative power since the number of the background voxels (TN) dominates the expression. Some of these measures (sensitivity and PPV) strongly penalize any voxel misclassifications in the apical and basal areas where the slices contain only very few true myocardial voxels. Performance of our algorithm is currently rather mediocre at basal and apical slices (with median specificity as low as 0.23 at the apex). This is partly due to limited feature evaluation at image borders and the pose standardization step, where voxels at boundaries can get transformed out of the classified volume. The poor performance at these regions results in increased variance of the measures and helps to explain the significant differences between mean and the median values of the measures.

Compared to the state of the art algorithms for left ventricle segmentation, slightly lower segmentation performance was achieved. It should be noted that the classification is run independently for each voxel. No smoothness, connectivity nor temporal consistency constraints are enforced to demonstrate the performance of the pure machine learning approach. Therefore, isolated segmentation islets or holes in the resulting binary segmentation can occur as a result of misclassification. However, thanks to the coordinate features, most of the voxels far from the myocardium are usually well discarded and also the solution becomes more regular as a result of the latent cardiac shape model built by the forests. In the soft classification, the holes are represented as a drop in the segmentation confidence but rarely fall to zero. This information could be easily considered in a subsequent postprocessing step to further improve the segmentation. However, adding these was not the goal of this





Figure B.4: a) Automatically calculated volume curve from patient DET0026701 during a single cardiac cycle with detected end systole (ES) and end diastole (ED) frames at the volume maximum and minimum respectively. b) Long axis crosssection through the binarized segmentations at ED and ES.

paper.

B.6 Conclusions

We aimed to present a fully automatic machine learning based algorithm for left ventricle segmentation with no explicit definition of task specific segmentation rules, model creation, user interaction nor post-processing. The algorithm learnt to automatically select the most discriminative features for the task using the ground-truth only. The only assumptions we make is that the motion of the object to be segmented is periodic for the construction of frame triplets and that the tissue intensity mapping between two different cases can be roughly approximated by a piecewise linear function. We also introduced a machine learning based intensity standardization method that allows to do tissue specific remapping of intensities and obtain a more CT like behaviour.

Finally, using a curvature-based iterative hole filling algorithm [Krishnan 2009] on the binarized segmentation, we could automatically calculate volumetric measurements and detect the main cardiac phases as the volume curve extremas (see Figure B.4).

Acknowledgements.

This work was partly supported by Microsoft Research through its PhD Scholarship Programme. We used data and infrastructure made available through the Cardiac Atlas Project (www.cardiacatlas.org) [Fonseca 2011].

In this appendix, we propose a method for estimating the location of glioma recurrence after surgical resection. This method consists of a pipeline including the registration of images at different time points, the estimation of the tumor infiltration map, and the prediction of tumor regrowth using a reaction-diffusion model. A data set acquired on a patient with a low-grade glioma and post surgery MRIs is considered to evaluate the accuracy of the estimated recurrence locations found using our method. We observed good agreement in tumor volume prediction and qualitative matching in regrowth locations. Therefore, the proposed method seems adequate for modeling low-grade glioma recurrence. This tool could help clinicians anticipate tumor regrowth and better characterize the radiologically non-visible infiltrative extent of the tumor. Such information could pave the way for model-based personalization of treatment planning in a near future. This work was published in [Stretton 2012].

C.1 Introduction

Glioma surgical resection has shown to be a critical therapeutic modality and is usually the first type of therapy given to patients. Resections are part of a standard treatment that has demonstrated increased patients' survival time [Sanai 2008]. However, gliomas are a diffuse, infiltrative and resilient form of brain cancer. Most low-grade glioma patients have a tumor recurrence after the first tumor resection. The tumor tends to reoccur most often immediately adjacent to the site of resection despite how extensive the resection [Sawaya 1999]. Treatment then includes a second surgery, chemotherapy or radiation therapy, and there is no consensus regarding the best option in this setting. We present a biomathematical tool that would estimate the radiologically non-visible part of the tumor from a longitudinal set of images. Such virtual imaging could potentially guide the clinician in the decision making process (intuitively, surgery should be prefered for a tumor without a large non-visible extent, i.e., the so called "bulky" tumors).

Mathematicians and computer scientists have proposed various methods to tackle portions of this problem [Cobzas 2009, Gooya 2011a, Harpold 2007, Hogea 2008, Konukoglu 2010a, Menze 2011, Swanson 2007, Zacharaki 2009]. Clatz et al. [Clatz 2005a] and Jbabdi et al. [Jbabdi 2005] proposed DTIs construction methods that estimate the tumor cell diffusion in white matter based on water diffusion in white matter. Konukoglu et al. [Konukoglu 2010a] built upon these models to personalize a tumor growth model to estimate the product of $d_{w,g} * \rho$ (tumor diffusion in white and gray matter multiplied by the tumor proliferation rate). These

Appendix C. Predicting the Location of Glioma Recurrence After a Resection Surgery

models allow us to reasonably capture the progression of a tumor for a given patient before a resection or therapy.

The latest work on modeling glioma regrowth following brain tumor surgery was by Swanson et al. [Swanson 2007]. They developed a 3D model of tumor growth accounting for the heterogeneity of brain tissue. In a post mortem study, they investigated the effectiveness of using different types of brain resections. However, their model was limited to personalization using patient T1 Gad and T2 MRIs, without taking into account the anisotropy in white matter fiber tracts visible in diffusion tensor imaging (DTI). In addition, they ran their simulations on âvirtual controlsâ instead of on patient data.

The pipeline approach that we present in this paper introduces several new features. First, the 3-D simulation results from using our pipeline estimates the most likely location of tumor progression after surgery since tumors do not typically grow at the same rate in all directions. Tumors grow faster in the white matter than in the gray matter of the brain [Giese 2001]. Therefore, simulating future tumor growth would be very helpful for therapy planning. Second, it estimates the profile of the tumor regrowth, thus informing about the radiologically nonvisible extent of the tumor. Third, the simulation results from using our pipeline helps to differentiate hyper-intense voxels between scarring tissue, edema or tumor recurrence. The areas bordering the resection cavity could be flagged as high and low risk of tumor recurrence areas. Our problem requires solving complex registration problems between pre-op and post-op, combining a tail extrapolation algorithm (to estimate the invisible part of the tumor) with a tumor progression algorithm (to predict future extension). To our knowledge, modeling tumor recurrence after a brain tumor resection using a patient DTI and patient data has not been done before.

This paper is organized into four sections. In Section 2, we describe a method for estimating the location of glioma recurrence. In Section 3, we present the results of our experiments, which show that this method is feasible. In Section 4, we discuss these results and future work.

C.2 Materials and Method

88

The proposed method for estimating the location of glioma recurrence after a resection consists of several interconnected steps. The first step entails segmenting the images. The second step consists of a sequence of registrations. The third step is estimating the tumor's infiltration tail on the date of surgery, and the fourth step uses a simulation method to predict the location of tumor regrowth at future time instances. The fifth step allows us to tell if the tumor is a bulky or diffuse tumor. Both the tail extrapolation algorithm and the prediction algorithm use the same model framework. We tested the proposed approach on data from a clinical study.



Figure C.1: Day -3 DTI: (1) anisotropic white matter tensors and (2) isotropic gray matter tensors. Region of red box in Figure C.2.

Model Framework. Tracqui et al. [Tracqui 1995] proposed using reactiondiffusion-based growth models in the form of the Fisher Kolmogorov equation (FK):

$$\frac{\partial u}{\partial t} = \underbrace{\nabla \cdot (D(\mathbf{x})\nabla u)}_{\text{Diffusion Term}} + \underbrace{\rho \cdot u \cdot (1-u)}_{\text{Logistic Reaction Term}}; \underbrace{\eta_{\partial \omega} \cdot (D\nabla u) = 0}_{\text{Boundary Condition}}$$
(C.1)

where u is the tumor cell density, D is the diffusion tensor for tumor cells using the tumor diffusion tensor construction method described below, ρ is the proliferation rate and $\eta_{\partial\omega}$ are the normal directions of the boundaries of the brain surface.

To use this framework, we need a tensor image constructed from a DTI to form $D(\mathbf{x})$, an estimate on parameter values (d_w, d_g, ρ) , and segmentations of several areas of the brain. d_w and d_g are scalars that multiply the diffusion tensors.

There are several tensor construction methods that have been proposed to model anisotropic diffusion [Clatz 2005a, Jbabdi 2005]. We used a tensor construction method, proposed by Clatz et al. [Clatz 2005a], that uses global scaling on the DTI,

$$D(\mathbf{x}) = \begin{cases} d_g I & \text{if } \mathbf{x} \text{ is in grey matter} \\ d_w D_{water} & \text{if } \mathbf{x} \text{ is in white matter} \end{cases}$$
(C.2)

where $D(\mathbf{x})$ is the inhomogeneous diffusion term, which takes into account that tumor cells are thought to move faster along anisotropic white matter fiber tracts, estimated by $d_w D_{water}$, than in isotropic gray matter $d_g I$. D_{water} is the water diffusion tensor in the brain measured by the DTI and I is the identity matrix which can be seen as an isotropic diffusion tensor (see Figure C.1).

Estimating the parameter values is detailed in the Optimizing d_w and ρ Algorithm paragraph.

Interval from	MRIs		
Surgery Date in Days	Modalities		
-49	Flair		
-3	DTI		
-1	T1 & Flair		
+1	Flair		
+74	T1 & Flair		
+172	T1 & Flair		

Appendix C. Predicting the Location of Glioma Recurrence After a Resection Surgery

Table C.1: Patient MRI acquisition dates.

Assumption. Gliomas appear as hyper-intense voxels in Flair MRIs in which edema appears bright. We assume that where there is edema, there is 20% or more tumor cell density threshold of visibility. In reality, there might be edema without tumor cells close by and vice versa. Tracqui et al. [Tracqui 1995] proposed 40% maximal tumor cell density to be visible in T2 MRIs, Konukoglu et al. [Konukoglu 2010a] used Tracqui's value, and Swanson et al. [Swanson 2007] used a value of 2%. Menze et al. [Menze 2011] suggested the maximal tumor cell density that is visible in Flair MRIs to be 9.5%. We chose the tumor cell density threshold of visibility value as 20% because it is an intermediate value in literature for T2 MRIs, which includes Flair. Currently, Flair is the imaging modality that shows the most glioma tumor cell density threshold of visibility extents, although distinction with scar tissue or edema is not possible with this sequence.

Data. Our data consists of a current patient with a supra-complete resection and long post-operation (post-op) follow-ups, complements of our collaborating neuro-surgeon with informed consent from the patient. It is difficult to acquire this type of longitudinal data, particularly due to limited availability of DTIs. For this reason, the pipeline was only tested on 1 data set. However, this is the first time this data set is being used for research and is not the same data set used by Konukoglu et al. [Konukoglu 2010a] and Clatz et al. [Clatz 2005a]. This patient had MRIs acquired on three different dates before surgery and three dates after surgery (see Figure C.1). The tumor, resection cavity and tumor regrowth for all of the dates were segmented by the neurosurgeon from Flair MRIs.

The voxel size of our MRIs range from $0.5 \ge 0.5 \ge 2.0 \ mm^3$ to $0.5 \ge 0.5 \ge 5.5 \ mm^3$. All images were re-sampled to be $1 \ge 1 \ge 1 \ mm^3$ by resampling the baseline using an in house tool and then registering all images (see Figure C.3) to the baseline.

Segmentation. The areas of the brain that need to be segmented to clearly define their boundaries are the white and gray matters, the cerebrospinal fluid (CSF) and the tumor at each time point.

The segmentations of white and gray matter are used to mark out the inhomogeneous tissue boundaries used by d_w and d_q . The CSF segmentation is used to



Figure C.2: Patient registered Flair MRIs axial views. (a) Hyper-intense region that was considered tumor before resection on Day 0.Red box displays the area shown in Figure C.1. (b) Distortion in the resection cavity. Hyper-intense regions were considered scar tissue or hemorrhages caused by surgery for this date. (c) and (d) exhibits hyper-intense regions which could be scar tissue, hemorrhages or tumor recurrence. It can be seen with these images that it is not possible to classify these hyper-intense regions and decipher if the tumor recurrence is a bulky or diffuse-type recurring tumor.

define the no flux boundary conditions of the model, i.e., tumor cells cannot enter these masks. To create these segmentations we thresholded white matter and brain parenchyma (white matter + gray matter) probability maps from MNI 152 (Atlas) [Fonov 2009] into binary masks, which recovered all of the necessary sulci structure and separated lobes. This conversion was achieved with the help of the neurosurgeon, who decided the best threshold values for the CSF, gray and white matter probability maps.

The tumor segmentations can be used for three purposes. First, a tumor segmentation is used as the starting boundary where the tumor growth simulation begins. Second, two tumor segmentations at two different time points can be used in a minimization algorithm to find the FK parameters: d_w , d_g , and ρ . Third, the following acquisition time point tumor segmentations are used to validate that the simulation results, which were grown from the first time point tumor segmentation, were reasonable.

Registration. The registration sequence employed has several interrelated steps (see Figure C.3). The most important part of our registration pipeline is the method we use for nonlinear registration of the images, where there exists no one-to-one correspondence between both images due to the tumor resection or growth. The non-linear deformation between the pre-op images and the post-op images can be assessed with the ventricles swelling and brain tissue shifting position after surgery (even several months after surgery). The idea of the nonlinear registration algorithm employed is to use local confidence weights and to model pathological regions with zero confidence. Lamecker et al. [Lamecker 2010] added this algorithm as an extension to the efficient and publicly available diffeomorphic demons registration framework. The algorithm requires a mask to cover the areas that cannot be matched between the images (i.e., resection cavity plus tumor volume). This mask volume is excluded from the registration. An inpainting step is used to estimate the

Appendix C. Predicting the Location of Glioma Recurrence After a 92 Resection Surgery



Figure C.3: Registration pipeline where all images are registered to the baseline. R stands for rigid, A for affine and NL for non-linear registration using a mask. In the non-linear registration we used an inpainting step to register the voxels covered by the mask.

registration in the areas covered by the mask.

The registration sequence can be divided into two parts: (1) registration of the atlas-based white matter and brain segmentations (segmentation details are in the Segmentation paragraph), and (2) registration of all patient images and segmentations to the baseline MRI. This registration sequence is depicted in Figure C.3 and the results can be seen in Figure C.2.

Registering the atlas-based white matter and brain segmentations required two steps. First, we non-linearly registered the Atlas T1 MRI to the patient's baseline MRI (Day -1 T1 re-sampled). Lastly, we applied this displacement field transformation to the white matter and brain binary segmentations.

Registering all of the patient images and segmentations involved three main steps. First, we rigidly registered all pre-op MRIs to the baseline. Next, we removed the skull and did histogram matching on the post-op T1 MRIs before non-linearly registering them to the baseline (the manually segmented mask consisted of the combined pre-op Day -1 tumor and post-op Day +74 resection cavity). Finally, we applied the transformations found registering the post-op T1 MRIs to the post-op Flair MRIs and segmentations.

Tail Extrapolation Algorithm. The third step in the method we are proposing uses the FK equation and the tensor construction method. Konukoglu et al. [Konukoglu 2010a] proposed a static model to overcome the problem of estimating the tumor infiltration tail by extrapolating the tumor invasion margins. The non-linear reaction term in Equation 1 is linearized around u = 0 and the tail distribution is shown to be asymptotically described by a Hamilton-Jacobi equation of the tumor cell density function. Using a Fast Marching method, an efficient algorithm was proposed that estimates the tumor cell invasion profile outside the visible boundaries in MRIs. For this step, we used the neurosurgeon's Day -1 rigidly registered tumor segmentation, the non-linearly registered white matter and brain segmentations, the rigidly registered DTI, and parameters d_w , d_g and ρ (see Optimizing d_w and ρ Algorithm paragraph for FK parameter choice). As the initial condition to this model, we make the assumption of 20% tumor cell density threshold of visibility (see Assumption paragraph). **Prediction Algorithm.** For the fourth step in our method we used this estimated tumor infiltration tail as the initial condition to the FK model (developed by Clatz [Clatz 2005a] and Konukoglu et al. [Konukoglu 2010a]) to simulate the location and predicted tumor cell density of recurrence for a given date [Clatz 2005a, Konukoglu 2010a]. This is done by propagating u by the time defined from MRI acquisition dates. We used two acquisition dates before surgery and one after surgery (-49, -1 and +74) to predict where the tumor will grow at the 4th acquisition date (2nd after surgery). The d_w, d_g and ρ parameters that were estimated with the first three acquisition times were used.

Optimizing d_w and ρ **Algorithm.** The fifth step in our method was determining which d_w and ρ fit each particular patient's data (personalization) since previous algorithms were only able to estimate the velocity constant ($v = 2\sqrt{\rho d_w}$), but not d_w and ρ separately [Konukoglu 2010a, Menze 2011]. For example, if d_w/ρ is low, the tumor is said to be bulky (not very infiltrative); where as if d_w/ρ is high, the tumor is said to be diffuse. We created a tool to sweep through the physically feasible values, proposed by Harpold et al. [Harpold 2007], of d_w and ρ keeping $d_w * \rho$ constant. There are two parts to this process: find v, and solve for d_w and ρ .

First, finding v can be done in two different ways. Konukoglu et al. [Konukoglu 2010a] proposed a minimization method for estimating the FK parameters: $d_w^* \rho$ (differential speed), d_g , T_0 (initial tumor start date). However, for this patient, the tumor does not visibly change volume or shape between Day -49 and Day -1 (possibly due to an overestimation of the tumor extent at Day -49, which was performed soon after a generalized seizure). We used the second way of finding v, which was to assume that the diameter velocity of the tumor was 4 mm/year, which was proposed by Mandonnet et al. in [Mandonnet 2003] for low grade glioma tumor growth.

Then, to solve for d_w and ρ , we swept through the possible parameter values of d_w (4 to 10 $mm^2/year$) and ρ (0.4 to 1.0 1/year), keeping v constant at 4 mm/year, iterating through steps 2 and 3 of our method. We started the Tail Extrapolation Algorithm from Flair segmentation Day -1 with resection cavity removed from the image to compare with Flair segmentation Day + 74. We found the value of $d_w = 6$ $mm^2/year$ and $\rho = 0.667 \, 1/year$ to be the most appropriate by qualitative analysis. These values of d_w and ρ were used to predict Flair segmentation Day + 172 and the results are discussed in the Results section (also see Figure C.4).

The parameters that determine the shape of the tumor, which are perceptible only locally in white matter, are the tensor construction method and the ratio d_w/d_g , where $d_w/d_g = 1$ is isotropic growth and $d_w/d_g = 100$ is highly anisotropic growth. There are two ways of finding d_g : using a minimization algorithm, such as the one proposed by Konukoglu et al. [Konukoglu 2010a], or sweeping through the possible values of d_g , once you have found d_w and ρ , by iterating through steps 2 and 3 of our method. Since we were not able to use a minimization algorithm on this patient's data, due to a likely seizure-induced overestimation of real tumor size at



Figure C.4: For Day +74, there is an agreement in volume for 3 of the segmentations. For Day +172, both the neurosurgeon's estimation of tumor growth and the FK prediction match in volume.

first MRI, we swept through the values of d_g (1 to 6 $mm^2/year$). We found $d_g = 1 mm^2/year$ to be the most appropriate value by qualitative analysis.

C.3 Results

94

In this paper, we have proposed a method to predict where tumor regrowth will occur for glioma resection patients.

Quantitatively, Figure C.4 shows the chronological progression of the patient's possible tumor regrowth. Due to the large amount of brain shift plus tumor evolution in the post-op MRIs, the non-linear registration compresses and stretches the tissues surrounding the resection cavity. For this reason we believe matching volumes and not surfaces is reasonable. Using overlap measures would imply to perform voxel to voxel comparison between pre-op and post-op images. This is a very challenging registration problem due to the large deformations caused by the tumor removal. As the registration errors are still large in those areas, we chose to compare the tumor segmentation and prediction by using global measures (volumes) rather than local measures like overlap. We use two criteria for evaluating our method's accuracy: the neurosurgeon's segmentations and hyper-intense signal segmentations. The hyperintense signal segmentation shows all the voxels that could possibly be tumor due to their intensity in the image. Results show that for Day +74, there is a good volume agreement for 3 of the segmentations. For Day +172, both the neurosurgeon's estimation of tumor growth and the FK simulated prediction (thresholded at 20%tumor infiltration) match in volume. This demonstrates that the model includes the visible part of the tumor in its prediction, but also flags areas which are not visible with current MRI technology (shown in magenta). Additionally, the agreement in Neurosurgeon's Tumor Segmentation



-Hyper-Intense Thresholded Voxels -

Figure C.5: Compare this figure with Figure C.1 and C.2. (a) Shows the estimated tumor infiltration tail in yellow. This tail cannot be distinguished with current MRI technology. In (b) and (c), the simulated tumor regrowth predictions are shown in yellow. Observe that the hyper-intense regions do not exactly cover the same regions as were flagged by the neurosurgeon (blue). However, both areas are covered by the prediction of 1% tumor infiltration.

volume between the neurosurgeon's segmentation and the model's results signifies that the tumor location outlined by the neurosurgeon is not a simple function of signal intensity.

Qualitatively, we show in Figure C.5 that our model provides a reasonable estimate of the tumor infiltration tail after resection. Figures C.2 and C.1 display the same axial slice and should be used to aid interpretation of this figure. In Figure C.5(a) we show the estimated tumor infiltration tail that cannot be seen in MRI images (step 2 of our method). In Figure C.5(b) and (c), the predicted tumor regrowth is displayed in yellow (step 3 of our model). We can see from Figure C.5(a)that the tumor tail (1% tumor infiltration) was not removed with the brain resection. This tail was the seed of regrowth, which is evident in the Day +74 and Day +172 MRIs. If we compare Figure C.5(b) and (c) with Figure C.1, we can see that the patient's white matter tensors, which are bordering the resection cavity, are anisotropic. These tensor's shape were a large contributor to dictating the speed and direction in which the tumor was simulated to grow. The green lines outline the hyper-intense voxels in the Flair MRIs. These regions could be scarring and/or edema caused by surgery and/or tumor recurrence cell density above or equal to 20% of maximal cell density. The blue line was classified by the neurosurgeon as possible tumor. Observe that the hyper-intense regions do not exactly cover the same regions as were flagged by the neurosurgeon. However, both areas are covered by the FK simulated prediction at the tumor cell density threshold of visibility value of 1%.

Depending on the size and resolution of the image, the automated process of registration, estimating the tumor infiltration map and simulating future tumor regrowth sites for one future time instance can take about 20 hours on a single CPU running at 2.2 GHz. The main time-consuming step is the non-linear registration

Neurosurgeon's Cavity Segmentation

with inpainting.

C.4 Discussion and Conclusion

We presented an approach to predict tumor regrowth after a brain tumor resection. We used a novel pipeline combining image registration with a static model for estimating the tumor infiltration tail and a dynamic simulation model for predicting future tumor regrowth. Our results show that prediction is possible for future tumor regrowth using a reaction-diffusion-type model that employs a patient DTI.

The non-linear registration step that we employed was key in making our method possible. Other non-linear registration methods, such as demons (without extensions) or pyramidal block-matching algorithms that use masks, were not able to deal with the resection cavity to tumor registration. The non-linear registration step that we used was designed to work with an atlas to patient registration in the presence of pathologies in the patient image. Although it worked quite well for the tumor resection application, we could improve the registration results if we extended this algorithm to use more specific prior information for resection images.

In the future, we intend to study more glioma resection patients having regrowth after surgery using this method. We will study all of the parameter interactions of our method, as well as explore using other tensor construction techniques for the tail extrapolation algorithm and prediction algorithm parts of our method, e.g. Jbabdi et al. [Jbabdi 2005]. Since glioma growth modeling is patient-specific, we intend to improve our method and validate it using a large patient data set. This data set will help us analyze the best way to improve the registration, minimization of parameters and investigate if the tumor growth rate stays constant after a tumor resection, as seen previously among numerous patients. With a large number of patients studied, we will develop a method to predict more precisely these parameters separately, prior to a glioma resection. This will enable the model to more precisely predict where the tumor could reoccur after surgery.

C.5 Acknowledgments

This work was partially supported by the Care4me ITEA2 project and ERC MedYMA. A big thank you to Maxime Sermesant for his support, and also to Gregoire Malandain and Jatin Relan.

96

Bibliography

- [Achanta 2012] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua and Sabine Süsstrunk. SLIC Superpixels Compared to State-ofthe-Art Superpixel Methods. IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pages 2274–2282, 2012. (Cited on pages 4, 63 and 67.)
- [Admiraal-Behloul 2005] F. Admiraal-Behloul, D.M.J. van den Heuvel, H. Olofsen, M.J.P. van Osch, J. van der Grond, M.A. van Buchem and J.H.C. Reiber. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage, vol. 28, no. 3, pages 607 - 617, 2005. (Cited on page 6.)
- [Akselrod-Ballin 2006] Ayelet Akselrod-Ballin, Meirav Galun, Ronen Basri, Achi Brandt, Moshe John Gomori, Massimo Filippi and Paula Valsasina. An Integrated Segmentation and Classification Approach Applied to Multiple Sclerosis Analysis. In CVPR '06: IEEE, pages 1122–1129, 2006. (Cited on pages 4 and 6.)
- [Akselrod-Ballin 2009] Ayelet Akselrod-Ballin, Meirav Galun, Moshe John Gomori, Massimo Filippi, Paula Valsasina, Ronen Basri and Achi Brandt. Automatic Segmentation and Classification of Multiple Sclerosis in Multichannel MRI. IEEE Trans. Biomed. Engineering, vol. 56, no. 10, pages 2461–2469, 2009. (Cited on pages 62, 64, 67 and 73.)
- [Amit 1997] Yali Amit and Donald Geman. Shape Quantization And Recognition With Randomized Trees. Neural Computation, vol. 9, no. 7, pages 1545–1588, 1997. (Cited on page 10.)
- [Anbeek 2004] P. Anbeek, K.L. Vincken, M.J.P. van Osch, R.H.C. Bisschops and J. van der Grond. Probabilistic Segmentation of White Matter Lesions in MR Imaging. NeuroImage, vol. 21, no. 3, pages 1037–1044, March 2004. (Cited on page 6.)
- [Anbeek 2008] P. Anbeek, K. Vincken and M. Viergever. Automated MS-Lesion Segmentation by K-Nearest Neighbor Classification. In The MIDAS Journal -MS Lesion Segmentation (MICCAI 2008 Workshop), 2008. (Cited on page 8.)
- [Andres 2008] Björn Andres, Ullrich Köthe, Moritz Helmstaedter, Winfried Denk and Fred A. Hamprecht. Segmentation of SBFSEM Volume Data of Neural Tissue by Hierarchical Classification. In DAGM-Symposium, pages 142–152, 2008. (Cited on page 6.)
- [Angelini 2007] Elsa D. Angelini, Olivier Clatz, Emmanuel M, Ender Konukoglu, Laurent Capelle and Hugues Duffau. *Glioma Dynamics and Computational*

Models: A Review of Segmentation, Registration, and In Silico Growth Algorithms and their Clinical Applications. Cur Med Imag Rev, vol. 3, pages 262–276, 2007. (Cited on page 38.)

- [Angelini 2012] Elsa D. Angelini, Julie Delon, Alpha Boubacar Bah, Laurent Capelle and Emmanuel Mandonnet. Differential MRI analysis for quantification of low grade glioma growth. Medical Image Analysis, vol. 16, pages 114-126, 2012. (Cited on page 37.)
- [Bauer 2011] Stefan Bauer, Lutz-Peter Nolte and Mauricio Reyes. Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. In MICCAI (3), pages 354–361, 2011. (Cited on pages 69 and 70.)
- [Bergeest 2008] J.P. Bergeest and F. Florian Jäger. A Comparison of Five Methods for Signal Intensity Standardization in MRI. In Thomas Tolxdorff, Jürgen Braun, Thomas M. Deserno, Alexander Horsch, Heinz Handels, Hans-Peter Meinzer and W. Brauer, editeurs, Bildverarbeitung für die Medizin 2008, pages 36-40. Springer, 2008. (Cited on page 82.)
- [Breiman 2001] Leo Breiman. Random Forests. Machine Learning, vol. 45, no. 1, pages 5–32, 2001. (Cited on pages 10, 11, 23 and 64.)
- [Bresch 2010] Didier Bresch, Thierry Colin, Emmanuel Grenier, Benjamin Ribba and Olivier Saut. Computational modeling of solid tumor growth: the avascular stage. SIAM Journal on Scientific Computing, vol. 32, no. 4, pages 2321–2344, 2010. (Cited on page 74.)
- [Cabezas 2011] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. Comp Meth Prog Biomed, vol. 104, pages 158–164, 2011. (Cited on page 38.)
- [Clatz 2005a] O. Clatz, M. Sermesant, P.Y. Bondiau, H. Delingette, S.K. Warfield, G. Malandain and N. Ayache. *Realistic simulation of the 3-D growth of brain* tumors in MR images coupling diffusion with biomechanical deformation. Medical Imaging, IEEE Transactions on, vol. 24, no. 10, pages 1334–1346, 2005. (Cited on pages 87, 89, 90 and 93.)
- [Clatz 2005b] Olivier Clatz, Maxime Sermesant, Pierre-Yves Bondiau, Hervé Delingette, Simon K. Warfield, Grégoire Malandain and Nicholas Ayache. Realistic Simulation of the 3D Growth of Brain Tumors in MR Images Coupling Diffusion with Mass Effect. IEEE Transactions on Medical Imaging, vol. 24, pages 1334-1346, 2005. (Cited on page 39.)
- [Cobzas 2009] D. Cobzas, P. Mosayebi, A. Murtha and M. Jagersand. Tumor invasion margin on the riemannian space of brain fibers. Medical Image Com-
puting and Computer-Assisted Intervention-MICCAI 2009, pages 531-539, 2009. (Cited on page 87.)

- [Coltuc 2006] D. Coltuc, P. Bolon and J.-M. Chassery. *Exact histogram specification*. IEEE TIP, vol. 15, pages 1143 –1152, 2006. (Cited on pages 41 and 73.)
- [Corso 2008] Jason J. Corso, Eitan Sharon, S. Dube, Suzie El-Saden, Usha Sinha and Alan L. Yuille. Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification. IEEE Transactions on Medical Imaging, vol. 27, pages 629–640, 2008. (Cited on pages 4, 38 and 62.)
- [Cour 2005] Timothée Cour, Florence Bénézit and Jianbo Shi. Spectral Segmentation with Multiscale Graph Decomposition. In CVPR (2), pages 1124–1131, 2005. (Cited on page 62.)
- [Criminisi 2009] Antonio Criminisi, Jamie Shotton, and Stefano Bucciarelli. Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In MICCAI workshop on Probabilistic Models for Medical Image Analysis (MICCAI-PMMIA), 2009. (Cited on page 3.)
- [Criminisi 2010a] Antonio Criminisi, Toby Sharp, Carsten Rother and Patrick Pérez. Geodesic image and video editing. ACM Trans. Graph., vol. 29, no. 5, page 134, 2010. (Cited on page 73.)
- [Criminisi 2010b] Antonio Criminisi, Jamie Shotton, Duncan P. Robertson and Ender Konukoglu. Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In Proc MICCAI-MCV, LNCS 6533, pages 106–117. Springer, 2010. (Cited on pages 6, 7, 10, 38 and 41.)
- [Criminisi 2011a] Antonio Criminisi, Toby Sharp and Patrick Perez. Geodesic Forests for Image Editing. Rapport technique, Microsoft, 2011. (Cited on page 73.)
- [Criminisi 2011b] Antonio Criminisi, Jamie Shotton and Ender Konukoglu. Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Rapport technique, Microsoft, 2011. (Cited on pages 3 and 39.)
- [Criminisi 2013] Antonio Criminisi and Jamie Shotton. Decision forests in computer vision and medical image analysis. Springer, 2013. (Cited on page 3.)
- [Cristini 2010] Vittorio Cristini and John Lowengrub. Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach, pages 185–205. Cambridge University Press, 2010. (Cited on page 38.)
- [Datta 2006] Sushmita Datta, Balasrinivasa Rao Sajja, Renjie He, Jerry S. Wolinsky, Rakesh K. Gupta and Ponnada A. Narayana. Segmentation and quantification of black holes in multiple sclerosis. NeuroImage, vol. 29, no. 2, pages 467 - 474, 2006. (Cited on page 6.)

- [Deisboeck 2010] Thomas S. Deisboeck and Georgio S. Stamatakos. Multiscale cancer modeling, pages 359–406. CRC Press, 2010. (Cited on page 38.)
- [dos Santos 2012] Jefersson Alex dos Santos, Philippe Henri Gosselin, Sylvie Philipp-Foliguet, Ricardo da Silva Torres and Alexandre X. Falcão. Multiscale Classification of Remote Sensing Images. IEEE T. Geoscience and Remote Sensing, vol. 50, no. 10, pages 3764–3775, 2012. (Cited on pages 4, 62 and 64.)
- [Dugas-Phocion 2004] Guillaume Dugas-Phocion, Miguel Ángel González Ballester, Grégoire Malandain, Nicholas Ayache, Christine Lebrun, Stéphane Chanalet and Caroline Bensa. *Hierarchical Segmentation of Multiple Sclerosis Lesions* in Multi-Sequence MRI. In ISBI, pages 157–160. IEEE, 2004. (Cited on pages 3 and 6.)
- [Egloff 2005] Daniel Egloff. Weighted P2 quantile, Boost Accumulators 1.46 (www.boost.org), 2005. (Cited on page 82.)
- [Evans 1993] A. C. Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly and T. M. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. In IEEE-Nuclear Science Symposium and Medical Imaging Conference, pages 1813–1817, 1993. (Cited on page 9.)
- [Fonov 2009] VS Fonov, AC Evans, RC McKinstry, CR Almli and DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage, vol. 47, pages S102–S102, 2009. (Cited on page 91.)
- [Fonseca 2011] CG Fonseca, M Backhaus, DA Bluemke, RD Britten, JD Chung, BR Cowan, ID Dinov, JP Finn, PJ Hunter, AH Kadish, DC Lee, JAC Lima, P Medrano-Gracia, K Shivkumar, A Suinesiaputra, W Tao and AA Young. The Cardiac Atlas Project- an Imaging Database for Computational Modeling and Statistical Atlases of the Heart. Bioinformatics, vol. 27, no. 16, pages 2288–2295, 2011. (Cited on pages 78 and 86.)
- [Freifeld 2009] Oren Freifeld, Hayit Greenspan and Jacob Goldberger. Multiple sclerosis lesion detection using constrained GMM and curve evolution. J. of Biomed. Imaging, vol. 2009, pages 1–13, 2009. (Cited on page 6.)
- [Geremia 2010] Ezequiel Geremia, Bjoern H. Menze, Olivier Clatz, Ender Konukoglu, Antonio Criminisi and Nicholas Ayache. Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images. In Medical Image Computing and Computer-Assisted Intervention (MICCAI'10), LNCS, Beijing, China, September 2010. Springer. (Cited on pages 2 and 75.)
- [Geremia 2011] Ezequiel Geremia, Olivier Clatz, Bjoern H. Menze, Ender Konukoglu, Antonio Criminisi and Nicholas Ayache. Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel Magnetic Resonance Images.

NeuroImage, vol. 57, pages 378–90, 2011. (Cited on pages 2, 38, 39, 41, 62, 64, 67, 73, 75, 77, 78, 79 and 80.)

- [Geremia 2012a] Ezequiel Geremia, Bjoern H. Menze and Nicholas Ayache. Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images. In MICCAI Challenge on Multimodal Brain Tumor Segmentation, October 2012. (Cited on pages 2 and 75.)
- [Geremia 2012b] Ezequiel Geremia, Bjoern H. Menze, Marcel Prastawa, Marc-André Weber, Antonio Criminisi and Nicholas Ayache. Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model. In MICCAI Workshop on Medical Computer Vision, LNCS. Springer, October 2012. (Cited on pages 2 and 75.)
- [Geremia 2013a] E. Geremia, D. Zikic, O. Clatz, B. H. Menze, B. Glocker, E. Konukoglu, J. Shotton, O. M. Thomas, S. J. Price, T. Das, R. Jena, N. Ayache and A. Criminisi. *Classification Forests for Semantic Segmentation of Brain Lesions in Multi-Channel MRI*. In Antonio Criminisi and Jamie Shotton, editeurs, Decision Forests in Computer Vision and Medical Image Analysis. Springer, 2013. (Cited on page 75.)
- [Geremia 2013b] Ezequiel Geremia, Bjoern H. Menze and Nicholas Ayache. Spatially Adaptive Random Forests for Classification Problems in Medical Imaging. In ISBI. IEEE, 2013. (Cited on pages 3 and 76.)
- [Geurts 2008] Jeroen J. Geurts and Frederik Barkhof. Grey matter pathology in multiple sclerosis. Lancet neurology, vol. 7, no. 9, pages 841–851, September 2008. (Cited on page 17.)
- [Giese 2001] A. Giese and M. Westphal. Treatment of malignant glioma: a problem beyond the margins of resection. Journal of cancer research and clinical oncology, vol. 127, no. 4, pages 217–225, 2001. (Cited on page 88.)
- [Gooya 2011a] A. Gooya, G. Biros and C. Davatzikos. Deformable registration of glioma images using em algorithm and diffusion reaction modeling. IEEE Trans Med Imaging, vol. 30, no. 2, pages 375–390, Feb 2011. (Cited on page 87.)
- [Gooya 2011b] Ali Gooya, Kilian M. Pohl, Michel Bilello, George Biros and Christos Davatzikos. Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model. In Proc MICCAI, LNCS 6891, pages 532-540. Springer, 2011. (Cited on page 38.)
- [Gray 2011] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers and D. Rueckert. Random Forest-Based Manifold Learning for Classification of Imaging Data in Dementia. In Proc MICCAI-MLMI, LNCS 7009, pages 159–166. Springer, 2011. (Cited on page 38.)

- [Hamamci 2012] Andac Hamamci, Nadir Kucuk, Kutlay Karaman, Kayihan Engin and Gözde B. Ünal. Tumor-Cut: Segmentation of Brain Tumors on Contrast Enhanced MR Images for Radiosurgery Applications. IEEE Trans. Med. Imaging, vol. 31, no. 3, pages 790–804, 2012. (Cited on page 70.)
- [Harpold 2007] H.L.P. Harpold, E.C. Alvord Jr and K.R. Swanson. The evolution of mathematical modeling of glioma proliferation and invasion. Journal of Neuropathology & Experimental Neurology, vol. 66, no. 1, page 1, 2007. (Cited on pages 87 and 93.)
- [Hogea 2008] C. Hogea, C. Davatzikos and G. Biros. An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. J. of Math. Bio., vol. 56, no. 6, pages 793-825, 2008. (Cited on page 87.)
- [Hoogendoorn 2010] Corne Hoogendoorn, Tristan Whitmarsh, Nicolas Duchateau, Federico M. Sukno, Mathieu De Craene and Alejandro F. Frangi. A groupwise mutual information metric for cost efficient selection of a suitable reference in cardiac computational atlas construction. In SPIE Medical Imaging, volume 7962, pages 76231R-76231R-9, 2010. (Cited on page 83.)
- [Iglesias 2011] J.E. Iglesias, Ender Konukoglu, Albert Montillo, Zhuowen Tu and Antonio Criminisi. Combining Generative and Discriminative Models for Semantic Segmentation of CT Scans via Active Learning. In Gábor Székely and Horst K. Hahn, editeurs, Information Processing in Medical Imaging, volume 6801 of LNCS. Springer, 2011. (Cited on page 83.)
- [Jagadeesh 2012] Vignesh Jagadeesh, Min-Chi Shih, B. S. Manjunath and Kenneth Rose. Scalable Tracing of Electron Micrographs by Fusing Top Down and Bottom Up Cues Using Hypergraph Diffusion. In MICCAI (3), pages 321– 328, 2012. (Cited on page 73.)
- [Jain 1985] Raj Jain and Imrich Chlamtac. The P2 algorithm for dynamic calculation of quantiles and histograms without storing observations. Communications of the ACM, vol. 28, no. 10, pages 1076–1085, October 1985. (Cited on page 82.)
- [Jbabdi 2005] S. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K.R. Swanson, M. Pélégrini-Issac, R. Guillevin and H. Benali. Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging. MRM, vol. 54, no. 3, pages 616-624, 2005. (Cited on pages 87, 89 and 96.)
- [Kim 2011] Sungwoong Kim, Sebastian Nowozin, Pushmeet Kohli and Chang Dong Yoo. Higher-Order Correlation Clustering for Image Segmentation. In NIPS, pages 1530–1538, 2011. (Cited on page 73.)
- [Konukoglu 2010a] E. Konukoglu, O. Clatz, B.H. Menze, B. Stieltjes, M.A. Weber, E. Mandonnet, H. Delingette and N. Ayache. *Image guided personalization*

of reaction-diffusion type tumor growth models using modified anisotropic eikonal equations. Medical Imaging, IEEE Transactions on, vol. 29, no. 1, pages 77–95, 2010. (Cited on pages 87, 90, 92 and 93.)

- [Konukoglu 2010b] Ender Konukoglu, Olivier Clatz, Pierre-Yves Bondiau, Hervé Delingette and Nicholas Ayache. Extrapolating Glioma Invasion Margin in Brain Magnetic Resonance Images: Suggesting New Irradiation Margins. Medical Image Analysis, vol. 14, no. 2, pages 111–125, 2010. (Cited on page 3.)
- [Konukoglu 2011] Ender Konukoglu, Antonio Criminisi, Sayan Pathak, Duncan Robertson, Steve White, David Haynor and Khan Siddiqui. Robust linear registration of CT images using random regression forests. In SPIE Medical Imaging, volume 7962, pages 79621X-79621X-8, 2011. (Cited on page 83.)
- [Krishnan 2009] Karthik Krishnan, Luis Ibanez, Wesley Turner and Ricardo Avila. Algorithms, architecture, validation of an open source toolkit for segmenting CT lung lesions. In Matthew Brown, Marleen de Bruijne, Bram van Ginneken, Atilla Kiraly, Jan-Martin Kuhnigk, Cristian Lorenz, Jamie R. McClelland, Kensaku Mori, Anthony Reeves and Joseph M. Reinhardt, editeurs, MICCAI Workshop on Pulmonary Image Analysis, pages 365–375. CreateSpace, 2009. (Cited on page 86.)
- [Lamecker 2010] H. Lamecker, X. Pennecet al. Atlas to image-with-tumor registration based on demons and deformation inpainting. In Proc. MICCAI Workshop on Computational Imaging Biomarkers for Tumors-From Qualitative to Quantitative, CIBT. Citeseer, 2010. (Cited on page 91.)
- [Lee 2008] Chi-Hoon Lee, Shaojun Wang, Albert Murtha, Matthew R. G. Brown and Russell Greiner. Segmenting Brain Tumors Using Pseudo-Conditional Random Fields. In Proc MICCAI, LNCS 5242, pages 359–366. Springer, 2008. (Cited on page 38.)
- [Lefebvre 2007] Sylvain Lefebvre and Hugues Hoppe. Compressed Random-Access Trees for Spatially Coherent Data. In Rendering Techniques, pages 339–349, 2007. (Cited on page 62.)
- [Lempitsky 2009] Victor S. Lempitsky, Michael Verhoek, J. Alison Noble and Andrew Blake. Random Forest Classification for Automatic Delineation of Myocardium in Real-Time 3D Echocardiography. In FIMH, LNCS 5528, pages 447-456. Springer, 2009. (Cited on pages 3, 7, 10, 77, 78, 80 and 82.)
- [Lu 2011] Xiaoguang Lu, Yang Wang, Bogdan Georgescu, Arne Littman and Dorin Comaniciu. Automatic Delineation of Left and Right Ventricles in Cardiac MRI Sequences Using a Joint Ventricular Model. In Dimitris Metaxas and Leon Axel, editeurs, Functional Imaging and Modeling of the Heart, volume 6666 of LNCS, pages 250-258. Springer, 2011. (Cited on page 77.)

- [Mandonnet 2003] E. Mandonnet, J.Y. Delattre, M.L. Tanguy, K.R. Swanson, A.F. Carpentier, H. Duffau, P. Cornu, R. Van Effenterre, E.C. Alvord Jr and L. Capelle. Continuous growth of mean tumor diameter in a subset of grade II gliomas. Annals of neurology, vol. 53, no. 4, pages 524–528, 2003. (Cited on page 93.)
- [Margeta 2011] Jan Margeta, Ezequiel Geremia, Antonio Criminisi and Nicholas Ayache. Layered Spatio-temporal Forests for Left Ventricle Segmentation from 4D Cardiac MRI Data. In Proceedings of STACOM Workshop at MICCAI 2011. Springer, September 2011. (Cited on pages 3, 73, 75 and 77.)
- [Menze 2009] B H Menze, B M Kelm, R Masuch, U Himmelreich, W Petrich and F A Hamprecht. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinformatics, vol. 10, page 213, 2009. (Cited on page 6.)
- [Menze 2010] Bjoern H Menze, Koen Van Leemput, Danial Lashkari, Marc-André Weber, Nicholas Ayache and Polina Golland. A Generative Model for Brain Tumor Segmentation in Multi-Modal Images. In Proc MICCAI, 6362, pages 151-159. Springer, 2010. (Cited on pages 38, 44 and 50.)
- [Menze 2011] Bjoern H Menze, Koen Van Leemput, Antti Honkela, Ender Konukoglu, Marc-Andre Weber, Nicholas Ayache and Polina Golland. A Generative Approach for Image-Based Modeling of Tumor Growth. In Imagebased modeling of tumor growth in patients with glioma. Springer, 2011. (Cited on pages 38, 44, 74, 87, 90 and 93.)
- [Menze 2012] Bjoern H. Menze, Ezequiel Geremia, Nicholas Ayache and Gabor Szekely. Segmenting glioma in multi-modal images using a generativediscriminative model for brain lesion segmentation. In MICCAI Challenge on Multimodal Brain Tumor Segmentation, October 2012. (Cited on pages 2 and 75.)
- [Montillo 2011] Albert Montillo, Jamie Shotton, John M. Winn, Juan Eugenio Iglesias, Dimitris N. Metaxas and Antonio Criminisi. Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images. In Proc IPMI, pages 184–196, 2011. (Cited on page 38.)
- [Nyúl 1999] L G Nyúl and J K Udupa. On standardizing the MR image intensity scale. Magnetic resonance in medicine, vol. 42, no. 6, pages 1072-81, December 1999. (Cited on pages 81 and 82.)
- [Nyúl 2000] László G. Nyúl, Jayaram K. Udupa and Xuan Zhang. New Variants of a Method of MRI Scale Standardization. IEEE Transactions on Medical Imaging, vol. 19, no. 2, pages 143–150, 2000. (Cited on page 82.)
- [Ourselin 2000] S. Ourselin, A. Roche, S. Prima and N. Ayache. Block matching: A general framework to improve robustness of rigid registration of medical

images. In Scott Delp, Anthony DiGoia and Branislav Jaramaz, editeurs, Medical Image Computing and Computer-Assisted Intervention 2000, volume 1935 of *LNCS*. Springer, 2000. (Cited on page 83.)

- [Prastawa 2004] Marcel Prastawa, Elizabeth Bullitt, Sean Ho and Guido Gerig. A brain tumor segmentation framework based on outlier detection. Medical Image Analysis, vol. 8, pages 275–283, 2004. (Cited on pages 3, 38 and 44.)
- [Prastawa 2009] Marcel Prastawa, Elizabeth Bullitt and Guido Gerig. Simulation of brain tumors in MR images for evaluation of segmentation efficacy. Medical Image Analysis, vol. 13, pages 297–311, 2009. (Cited on pages 3, 38, 39 and 44.)
- [Prima 2001] Sylvain Prima, Nicholas Ayache, Tom Barrick and Neil Roberts. Maximum Likelihood Estimation of the Bias Field in MR Brain Images: Investigating Different Modelings of the Imaging Process. In MICCAI, LNCS 2208, pages 811-819. Springer, 2001. (Cited on page 9.)
- [Prima 2002] Sylvain Prima, Sébastien Ourselin and Nicholas Ayache. Computation of the Mid-Sagittal Plane in 3D Brain Images. IEEE Trans. Med. Imaging, vol. 21, no. 2, pages 122–138, 2002. (Cited on pages 9 and 41.)
- [Quinlan 1993] J. Ross Quinlan. C4.5: Programs for machine learning. Morgan Kaufmann, 1993. (Cited on page 11.)
- [Rey 2002] David Rey. Détection et quantification de processus évolutifs dans des images médicales tridimensionnelles : application à la sclérose en plaques. Thèse de sciences, Université de Nice Sophia-Antipolis, October 2002. (in French). (Cited on pages 9 and 73.)
- [Riklin-Raviv 2010] Tammy Riklin-Raviv, Koen Van Leemput, Bjoern H. Menze, William M. Wells and Polina Golland. Segmentation of image ensembles via latent atlases. Medical Image Analysis, vol. 14, pages 654–665, 2010. (Cited on page 38.)
- [Sanai 2008] N. Sanai and M.S. Berger. Glioma extent of resection and its impact on patient outcome. Neurosurgery, vol. 62, no. 4, page 753, 2008. (Cited on page 87.)
- [Sawaya 1999] R. Sawaya. Extent of resection in malignant gliomas: a critical summary. Journal of Neuro-Oncology, vol. 42, no. 3, pages 303-305, 1999. (Cited on page 87.)
- [Shah 2010] Mohak Shah, Yiming Xiao, Nagesh Subbanna, Simon Francis, Douglas L Arnold, D Louis Collins and Tal Arbel. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. Medical image analysis, vol. 15, no. 2, pages 267–282, December 2010. (Cited on page 82.)

- [Shi 2011] Wenzhe Shi, Xiahai Zhuang, Haiyan Wang, Simon Duckett, Declan Oregan, Philip Edwards, Sebastien Ourselin and Daniel Rueckert. Automatic Segmentation of Different Pathologies from Cardiac Cine MRI Using Registration and Multiple Component EM Estimation. In Dimitris Metaxas and Leon Axel, editeurs, Functional Imaging and Modeling of the Heart, volume 6666 of LNCS, pages 163–170. Springer, 2011. (Cited on pages 77, 81 and 83.)
- [Shiee 2008] N. Shiee, P. Bazin and D. Pham. Multiple Sclerosis Lesion Segmentation Using Statistical and Topological Atlases. In The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop), 2008. (Cited on page 8.)
- [Shiee 2010] Navid Shiee, Pierre-Louis Bazin, Arzu Ozturk, Daniel S. Reich, Peter A. Calabresi and Dzung L. Pham. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. NeuroImage, vol. 49, no. 2, pages 1524 - 1535, 2010. (Cited on page 8.)
- [Shotton 2009] Jamie Shotton, John M. Winn, Carsten Rother and Antonio Criminisi. TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context. Int. J. Comp. Vision, vol. 81, no. 1, pages 2–23, 2009. (Cited on pages 6, 10 and 14.)
- [Shotton 2011] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake. *Real-time human pose recognition in parts from single depth images*. In Proc CVPR, pages 1297–1304, 2011. (Cited on page 39.)
- [Siskind 2007] Jeffrey Mark Siskind, James Sherman Jr., Ilya Pollak, Mary P. Harper and Charles A. Bouman. Spatial Random Tree Grammars for Modeling Hierarchal Structure in Images with Regions of Arbitrary Shape. IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 9, pages 1504–1519, 2007. (Cited on page 63.)
- [Souplet 2008] Jean-Christophe Souplet, Christine Lebrun, Nicholas Ayache and Grégoire Malandain. An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. In The MIDAS Journal - MS Lesion Segmentation (MIC-CAI 2008 Workshop), 2008. (Cited on pages 3, 6, 7, 8, 9, 15, 16, 26, 29, 30 and 31.)
- [Stretton 2012] Erin Stretton, Emmanuel Mandonnet, Ezequiel Geremia, Bjoern H. Menze, Herve Delingette and Nicholas Ayache. Predicting the Location of Glioma Recurrence After a Resection Surgery. In Proceedings of 2nd International MICCAI Workshop on Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data (STIA'12), LNCS, Nice, October 2012. Springer. (Cited on pages 3, 76 and 87.)
- [Styner 2008] M. Styner, J. Lee, B. Chin, M.S. Chin, O. Commowick, H. Tran, S. Markovic-Plese, V. Jewells and S.K. Warfield. 3D Segmentation in the

Clinic: A Grand Challenge II: MS Lesion Segmentation. In MIDAS Journal, pages 1–5, Sep 2008. (Cited on pages 7 and 8.)

- [Swanson 2007] KR Swanson, RC Rostomily and EC Alvord. A mathematical modelling tool for predicting survival of individual patients following resection of glioblastoma: a proof of principle. British journal of cancer, vol. 98, no. 1, pages 113-119, 2007. (Cited on pages 87, 88 and 90.)
- [Tracqui 1995] P. Tracqui, GC Cruywagen, DE Woodward, GT Bartoo, JD Murray and EC Alvord Jr. A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. Cell Proliferation, vol. 28, no. 1, pages 17-31, 1995. (Cited on pages 89 and 90.)
- [Van Leemput 2001] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, Alan C. F. Colchester and Paul Suetens. Automated Segmentation of Multiple Sclerosis Lesions by Model Outlier Detection. IEEE Trans. Med. Imaging, vol. 20, no. 8, pages 677–688, 2001. (Cited on pages 3, 6 and 38.)
- [Viola 2001] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, pages I-511-I-518. IEEE Computer Society, 2001. (Cited on page 81.)
- [Wels 2008] Michael Wels, Gustavo Carneiro, Alexander Aplas, Martin Huber, Joachim Hornegger and Dorin Comaniciu. A Discriminative Model-Constrained Graph Cuts Approach to Fully Automated Pediatric Brain Tumor Segmentation in 3-D MRI. In Proc MICCAI, pages 67–75, 2008. (Cited on page 38.)
- [Wolz 2012] Robin Wolz, Chengwen Chu, Kazunari Misawa, Kensaku Mori and Daniel Rueckert. Multi-organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases. In MICCAI (1), pages 10–17, 2012. (Cited on page 62.)
- [Wu 2006] Y. Wu, S. Warfield, I. Tan, W. Wells III, D. Meier, R. van Schijndel, F. Barkhof and C. Guttmann. Automated Segmentation of Multiple Sclerosis Lesion Subtypes with Multichannel MRI. Neuroimage, vol. 32, no. 3, pages 1205-1215, 09 2006. (Cited on page 6.)
- [Yamamoto 2010] Daisuke Yamamoto, Hidetaka Arimura, Shingo Kakeda, Taiki Magome, Yasuo Yamashita, Fukai Toyofuku, Masafumi Ohki, Yoshiharu Higashida and Yukunori Korogi. Computer-aided detection of multiple sclerosis lesions in brain magnetic resonance images: False positive reduction scheme consisted of rule-based, level set method, and support vector machine. Computerized Medical Imaging and Graphics, vol. 34, no. 5, pages 404 – 413, 2010. (Cited on page 6.)

- [Yi 2009] Zhao Yi, Antonio Criminisi, Jamie Shotton and Andrew Blake. Discriminative, Semantic Segmentation of Brain Tissue in MR Images. LNCS 5762, pages 558-565. Springer, 2009. (Cited on pages 3, 6, 7, 10 and 14.)
- [Zacharaki 2009] Evangelia I. Zacharaki, Cosmina S. Hogea, Dinggang Shen, George Biros and Christos Davatzikos. Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. NeuroImage, vol. 46, pages 762-774, 2009. (Cited on pages 38 and 87.)
- [Zikic 2012] Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, C. Demiralp, Jamie Shotton, O. M. Thomas, T. Das, R. Jena and S. J. Price. Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR. In MICCAI (3), pages 369–376, 2012. (Cited on pages 62, 64, 67, 69, 70 and 73.)
- [Zou 2002] Kelly H. Zou, William M. Wells III, Michael Kaus, Ron Kikinis, Ferenc A. Jolesz and Simon K. Warfield. Statistical Validation of Automated Probabilistic Segmentation against Composite Latent Expert Ground Truth in MR Imaging of Brain Tumors. In Proc MICCAI, pages 315–322, 2002. (Cited on page 38.)

Random Decision Forests for Brain Lesions Segmentation in MRIs and Model-based Tumor Cell Extrapolation

Abstract: The large size of the datasets produced by medical imaging protocols contributes to the success of supervised discriminative methods for semantic labelling of images. Our study makes use of a general and efficient emerging framework, discriminative random forests, for the detection of brain lesions in multi-modal magnetic resonance images (MRIs). The contribution is three-fold. First, we focus on segmentation of brain lesions which is an essential task to diagnosis, prognosis and therapy planning. A context-aware random forest is designed for the automatic multi-class segmentation of MS lesions in MR images. It uses multi-channel MRIs, prior knowledge on tissue classes, symmetrical and long-range spatial context to discriminate lesions from background. Then, we investigate the promising perspective of estimating the brain tumor cell density from MRIs. A generative-discriminative framework is presented to learn the latent and clinically unavailable tumor cell density from model-based estimations associated with synthetic MRIs. The generative model is a validated and publicly available biophysiological tumor growth simulator. The discriminative model builds on multi-variate regression random forests to estimate the voxel-wise distribution of tumor cell density from input MRIs. Finally, we present the aSpatially Adaptive Random Forestsâ which merge the benefits of multi-scale and random forest methods and apply it to previously cited classification and regression settings. Quantitative evaluation of the proposed methods are carried out on publicly available labeled datasets and demonstrate state of the art performance.

 ${\bf Keywords:}\ {\bf machine}\ {\bf learning,\ classification,\ regression,\ random\ forest,\ segmentation,\ MRI$