HAL
open science

Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2012

Bjoern Menze, Andras Jakab, Stefan Bauer, Mauricio Reyes, Marcel Prastawa, Koen van Leemput

To cite this version:

Bjoern Menze, Andras Jakab, Stefan Bauer, Mauricio Reyes, Marcel Prastawa, et al.. Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2012. Bjoern Menze and Andras Jakab and Mauricio Reyes and Stefan Bauer and Marcel Prastawa and Koen Van Leemput. MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), Oct 2012, Nice, France. MICCAI, pp.77, 2012. hal-00912935

HAL Id: hal-00912935

https://inria.hal.science/hal-00912935v1

Submitted on 11 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MICCAI 2012 Challenge on

Multimodal Brain Tumor Segmentation

Proceedings of MICCAI-BRATS 2012
October 1st, Nice, France

Preface

Because of their unpredictable appearance and shape, segmenting brain tumors from multi-modal imaging data is one of the most challenging tasks in medical image analysis. Although many different segmentation strategies have been proposed in the literature, it is hard to compare existing methods because the validation datasets that are used differ widely in terms of input data (structural MR contrasts; perfusion or diffusion data; ...), the type of lesion (primary or secondary tumors; solid or infiltratively growing), and the state of the disease (pre- or post-treatment).

In order to gauge the current state-of-the-art in automated brain tumor segmentation and compare between different methods, we are organizing a **Multimodal Brain Tumor Segmentation** (BRATS) challenge that is held in conjunction with the 15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2012) on October 1st, 2012 in Nice, France.

For this purpose, we are making available a large dataset of brain tumor MR scans in which the tumor and edema regions have been manually delineated. In addition, we also provide realistically generated synthetic brain tumor datasets for which the ground truth segmentation is known. All images show low- or high-grade glioma cases.

Participating teams downloaded the training data for algorithmic tweaking and tuning. The teams then evaluated their segmentation performance on the training data, and submitted a short paper describing the results and the segmentation method that was used that were subsequently reviewed by the organizing committee. A total of 12 submissions were accepted for the final challenge. The corresponding short papers describing training results and the methodological approaches are summarized in this volume.

On the day of the challenge itself, an independent set of test scans is made available and analyzed on the spot by each team, after which the methods are ranked according to their performance. The challenge concludes with a round-table discussion of the obtained results as well as invited talks by clinical experts.

In the weeks following the challenge participating teams will be invited to contribute to a joint paper describing and summarizing the challenge outcome, which we will then submit to a high-impact journal in the field.

Bjoern Menze, Andras Jakab, Stefan Bauer,

Mauricio Reyes, Marcel Prastawa, Koen Van Leemput

August 2012

Challenge Program

Organizers

Bjoern Menze (ETH Zurich, INRIA)
Andras Jakab (ETH Zurich, University of Debrecen)
Stefan Bauer (University of Bern)
Mauricio Reyes (University of Bern)
Marcel Prastawa (University of Utah)
Koen Van Leemput (Harvard Medical School, Technical University of Denmark)

Invited Speakers

Prof. Roland Wiest, Bern University Hospital
Imaging workflow in daily clinical routine - benefits and pitfalls of advanced brain tumor imaging

Dr. Andras Jakab, University of Debrecen Medical and Health Science Center
Multimodal tumor analysis in clinical routine (TBF)

Prof. Jan Unkelbach, Massachusetts General Hospital, Harvard Medical School
Automated tumor segmentation for radiotherapy planning (TBF)

Contributions

Context-sensitive Classification Forests for Segmentation of Brain Tumor Tissues
D. Zikic, B. Glocker, E. Konukoglu, J. Shotton, A. Criminisi, D. H. Ye, C. Demiralp, O. M. Thomas, T. Das, R. Jena, and S. J. Price ..... *page 1*

Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization
S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L.-P. Nolte, and M. Reyes ..... *page 10*

Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images
E. Geremia, B. H. Menze, and N. Ayache ..... *page 14*

Multimodal Brain Tumor Segmentation Using the "Tumor-cut" Method on the BraTS Dataset
A. Hamamci and G. Unal ..... *page 19*

Brain Tumor Segmentation Based on GMM and Active Contour Method with a Model-aware Edge Map
L. Zhao, W. Wu, and J.J. Corso ..... *page 24*

Probabilistic Gabor and Markov Random Fields Segmentation of Brain Tumours in MRI Volumes
N. K. Subbanna and T. Arbel ..... *page 28*

Hybrid Clustering and Logistic Regression for Multi-Modal Brain Tumor Segmentation
H.-C. Shin ..... *page 32*

Hierarchical Random Walker for Multimodal Brain Tumor Segmentation
Y. Xiao and J. Hu ..... *page 36*

Automatic Brain Tumor Segmentation Based on a Coupled Global-Local Intensity Bayesian Model
X. Tomas-Fernandez, and S. K. Warfield ..... *page 41*

Segmenting Glioma in Multi-Modal Images Using a Generative Model for Brain Lesion Segmentation
B. H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland ..... *page 49*

Segmenting Glioma in Multi-Modal Images Using a Generative-Discriminative Model for Brain Lesion Segmentation
B. H. Menze, E. Geremia, N. Ayache, and G. Szekely ..... *page 56*

Multi-modal Brain Tumor Segmentation via Latent Atlases
T. Riklin Raviv, K. Van Leemput, and B. H. Menze ..... *page 64*

Context-sensitive Classification Forests for Segmentation of Brain Tumor Tissues

D. Zikic[1], B. Glocker[1], E. Konukoglu[1], J. Shotton[1], A. Criminisi[1],
D. H. Ye[2], C. Demiralp[3],
O. M. Thomas[4,5], T. Das[4], R. Jena[4], S. J. Price[4,6]

[1]Microsoft Research Cambridge, UK
[2]Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[3]Brown University, Providence, RI, USA
[4]Cambridge University Hospitals, Cambridge, UK
[5]Department of Radiology, Cambridge University, UK
[6]Department of Clinical Neurosciences, Cambridge University, UK

Abstract. We describe our submission to the Brain Tumor Segmentation Challenge (BraTS) at MICCAI 2012, which is based on our method for tissue-specific segmentation of high-grade brain tumors [3].
The main idea is to cast the segmentation as a classification task, and use the discriminative power of context information. We realize this idea by equipping a classification forest (CF) with spatially non-local features to represent the data, and by providing the CF with initial probability estimates for the single tissue classes as additional input (along-side the MRI channels). The initial probabilities are patient-specific, and computed at test time based on a learned model of intensity. Through the combination of the initial probabilities and the non-local features, our approach is able to capture the context information for each data point. Our method is fully automatic, with segmentation run times in the range of 1-2 minutes per patient. We evaluate the submission by cross-validation on the real and synthetic, high- and low-grade tumor BraTS data sets.

1 Introduction

This BraTS submission is based on our work presented in [3]. We approach the segmentation of the tumor tissues as a classification problem, where each point in the brain is assigned a certain tissue class. The basic building block of our approach is a standard classification forest (CF), which is a discriminative multi-class classification method. Classification forests allow us to describe brain points to be classified by very high-dimensional features, which are able to capture information about the spatial context. These features are based on the multi-channel intensities and are spatially non-local. Furthermore, we augment the input data to the classification forest with initial tissue probabilities, which are estimated as posterior probabilities resulting from a generative intensity-based model, parametrized by Guassian Mixture models (GMM). Together with the

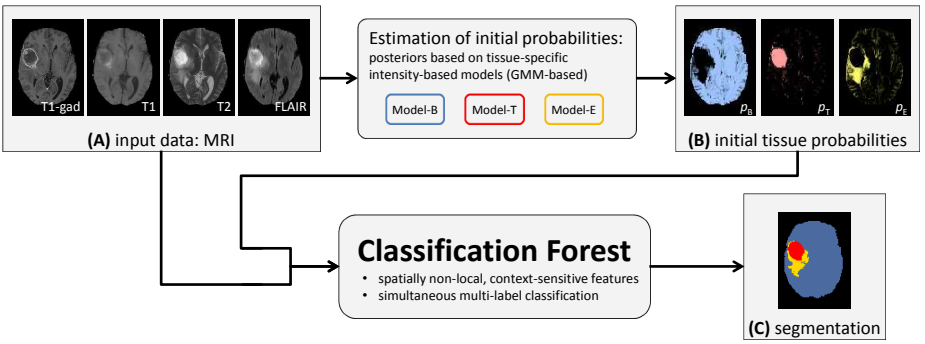


Fig. 1: Schematic Method Overview: Based on the input data **(A)**, we first roughly estimate the initial probabilities for the single tissues **(B)**, based on the local intensity information alone. In a second step, we combine the initial probabilities **(B)** with the input data from **(A)**, resulting in a higher-dimensional multi-channel input for the classification forest. The forest computes the segmentation **(C)** by a simultaneous multi-label classification, based on non-local and context-sensitive features.

context-sensitive features, the initial probabilities as additional input increase the amount of context information and thus improve the classification results.

In this paper, we focus on describing our BraTS submission. For more details on motivation for our approach and relation to previous work, please see [3].

2 Method: Context-sensitive Classification Forests

An overview of our approach is given in Figure 1. We use a standard classification forest [1], based on spatially non-local features, and combine it with initial probability estimates for the individual tissue classes. The initial tissue probabilities are based on local intensity information alone. They are estimated with a parametric GMM-based model, as described in Section 2.1. The initial probabilities are then used as additional input channels for the forest, together with the MR image data I.

In Section 2.2 we give a brief description of classification forests. The types of the context-sensitive features are described in Section 2.3.

We classify three classes $\mathcal{C} = \{\mathrm{B}, \mathrm{T}, \mathrm{E}\}$ for background (B), tumor (T), and edema (E). The MR input data is denoted by $I = (I_{\mathrm{T1C}}, I_{\mathrm{T1}}, I_{\mathrm{T2}}, I_{\mathrm{FLAIR}})$.

2.1 Estimating Initial Tissue Probabilities

As the first step of our approach, we estimate the initial class probabilities for a given patient based on the intensity representation in the MRI input data.

The initial probabilities are computed as posterior probabilities based on the likelihoods obtained by training a set of GMMs on the training data. For each

class $c \in \mathcal{C}$, we train a single GMM, which captures the likelihood $p_{\text{lik}}(\mathbf{i}|c)$ of the multi-dimensional intensity $\mathbf{i} \in \mathbb{R}^4$ for the class c. With the trained likelihood p_{lik}, for a given test patient data set I, the GMM-based posterior probability $p^{\text{GMM}}(c|\mathbf{p})$ for the class c is estimated for each point $\mathbf{p} \in \mathbb{R}^3$ by

$$p^{\text{GMM}}(c|\mathbf{p}) = \frac{p_{\text{lik}}(I(\mathbf{p})|c)\, p(c)}{\sum_{c_j} p_{\text{lik}}(I(\mathbf{p})|c_j)\, p(c_j)} \ , \tag{1}$$

with $p(c)$ denoting the prior probability for the class c, computed as a normalized empirical histogram. We can now use the posterior probabilities directly as input for the classification forests, in addition to the multi-channel MR data I. So now, with $p_c^{\text{GMM}}(\mathbf{p}) := p^{\text{GMM}}(c|\mathbf{p})$, our data for one patient consists of the following channels

$$C = (I_{\text{T1-gad}}, I_{\text{T1}}, I_{\text{T2}}, I_{\text{FLAIR}}, p_{\text{AC}}^{\text{GMM}}, p_{\text{NC}}^{\text{GMM}}, p_{\text{E}}^{\text{GMM}}, p_{\text{B}}^{\text{GMM}}) \ . \tag{2}$$

For simplicity, we will denote single channels by C_j.

2.2 Classification Forests

We employ a classification forest (CF) to determine a class $c \in \mathcal{C}$ for a given spatial input point $\mathbf{p} \in \Omega$ from a spatial domain Ω of the patient. Our classification forest operates on the representation of a spatial point $\mathbf{p}$ by a corresponding feature vector $x(\mathbf{p}, C)$, which is based on spatially non-local information from the channels C. CFs are ensembles of (binary) classification trees, indexed and referred to by $t \in [1, T]$. As a supervised method, CFs operate in two stages: training and testing.

During training, each tree t learns a weak class predictor $p_t(c|x(\mathbf{p}, C))$. The input training data set is $\{(x(\mathbf{p}, C^{(k)}), c^{(k)}(\mathbf{p})) : \mathbf{p} \in \Omega^{(k)}\}$, that is, the feature representations of all spatial points $\mathbf{p} \in \Omega^{(k)}$, in all training patient data sets k, and the corresponding manual labels $c^{(k)}(\mathbf{p})$.

To simplify notation, we will refer to a data point at $\mathbf{p}$ by its feature representation x. The set of all data points shall be X.

In a classification tree, each node i contains a set of training examples X_i, and a class predictor $p_t^i(c|x)$, which is the probability corresponding to the fraction of points with class c in X_i (normalized empirical histogram). Starting with the complete training data set X at the root, the training is performed by successively splitting the training examples at every node based on their feature representation, and assigning the partitions X_{L} and X_{R} to the left and right child node. At each node, a number of splits along randomly chosen dimensions of the feature space is considered, and the one maximizing the Information Gain is applied (i.e., an axis-aligned hyperplane is used in the split function). Tree growing is stopped at a certain tree depth D.

At testing, a data point x to be classified is pushed through each tree t, by applying the learned split functions. Upon arriving at a leaf node l, the leaf probability is used as the tree probability, i.e. $p_t(c|x) = p_t^l(c|x)$. The overall probability

is computed as the average of tree probabilities, i.e. $p(c|x) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|x)$. The actual class estimate $\hat{c}$ is chosen as the class with the highest probability, i.e. $\hat{c} = \arg\max_c p(c|x)$.

For more details on classification forests, see for example [1].

2.3 Context-sensitive Feature Types

We employ three features types, which are intensity-based and parametrized. Features of these types describe a point to be labeled based on its non-local neighborhood, such that they are context-sensitive. The first two of these feature types are quite generic, while the third one is designed with the intuition of detecting structure changes. We denote the parametrized feature types by $x^{\text{type}}_{\text{params}}$. Each combination of type and parameter settings generates one dimension in the feature space, that is $x_i = x^{\text{type}_i}_{\text{params}_i}$. Theoretically, the number of possible combinations of type and parameter settings is infinite, and even with exhaustive discrete sampling it remains substantial. In practice, a certain predefined number d' of combinations of feature types and parameter settings is randomly drawn for training. In our experiments, we use $d' = 2000$.

We use the following notation: Again, $\mathbf{p}$ is a spatial point, to be assigned a class, and C_j is an input channel. $R^s_j(\mathbf{p})$ denotes an $\mathbf{p}$-centered and axis aligned 3D cuboid region in C_j with edge lengths $\mathbf{l} = (l_x, l_y, l_z)$, and $\mathbf{u} \in \mathbb{R}^3$ is an offset vector.

- **Feature Type 1:** measures the intensity difference between $\mathbf{p}$ in a channel C_{j_1} and an offset point $\mathbf{p} + \mathbf{u}$ in a channel C_{j_2}

$$x^{\text{t1}}_{j_1,j_2,\mathbf{u}}(\mathbf{p}, C) = C_{j_1}(\mathbf{p}) - C_{j_2}(\mathbf{p} + \mathbf{u}) \ . \tag{3}$$

- **Feature Type 2:** measures the difference between intensity means of a cuboid around $\mathbf{p}$ in C_{j_1}, and around an offset point $\mathbf{p} + \mathbf{u}$ in C_{j_2}

$$x^{\text{t2}}_{j_1,j_2,\mathbf{l}_1,\mathbf{l}_2,\mathbf{u}}(\mathbf{p}, C) = \mu(R^{\mathbf{l}_1}_{j_1}(\mathbf{p})) - \mu(R^{\mathbf{l}_2}_{j_2}(\mathbf{p} + \mathbf{u})) \ . \tag{4}$$

- **Feature Type 3:** captures the intensity range along a 3D line between $\mathbf{p}$ and $\mathbf{p}+\mathbf{u}$ in one channel. This type is designed with the intuition that structure changes can yield a large intensity change, e.g. NC being dark and AC bright in T1-gad.

$$x^{\text{t3}}_{j,\mathbf{u}}(\mathbf{p}, C) = \max_{\lambda}(C_j(\mathbf{p} + \lambda\mathbf{u})) - \min_{\lambda}(C_j(\mathbf{p} + \lambda\mathbf{u})) \quad \text{with} \quad \lambda \in [0, 1] \ . \tag{5}$$

In the experiments, the types and parameters are drawn uniformly. The offsets u_i originate from the range $[0, 20]$mm, and the cuboid lengths l_i from $[0, 40]$mm.

Dice score	High-grade (real)		Low-grade (real)		High-grade (synth)		Low-grade (synth)	
	Edema	Tumor	Edema	Tumor	Edema	Tumor	Edema	Tumor
mean	0.70	0.71	0.44	0.62	0.65	0.90	0.55	0.71
std. dev.	0.09	0.24	0.18	0.27	0.27	0.05	0.23	0.20
median	0.70	0.78	0.44	0.74	0.76	0.92	0.65	0.78

Table 1: Evaluation summary. The Dice scores are computed by the online evaluation tool provided by the organizers of the BraTS challenge.

3 Evaluation

We evaluate our approach on the real and synthetic data from the BraTS challenge. Both real and synthetic examples contain separate high-grade (HG) and low-grade (LG) data sets. This results in 4 data sets (Real-HG, Real-LG, Synth-HG, Synth-LG). For each of these data sets, we perform the evaluation independently, i.e., we use only the data from one data set for the training and the testing for this data set.

In terms of sizes, Real-HG contains 20 patients, Synth-LG has 10 patients, and the two synthetic data sets contain 25 patients each. For the real data sets, we test our approach on each patient by leave-one-out cross-validation, meaning that for each patient, the training is performed on all other images from the data set, excluding the tested image itself. For the synthetic images, we perform a leave-5-out cross-validation.

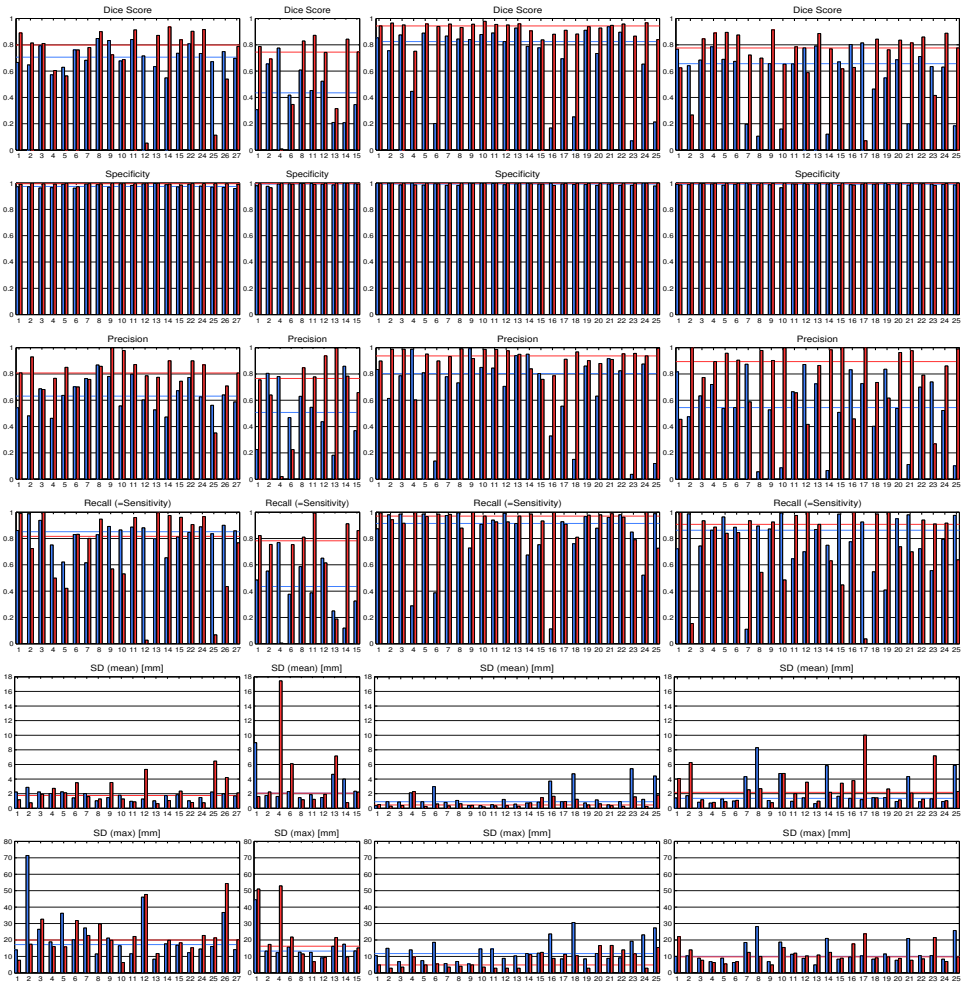
Pre-processing. We apply bias-field normalization by the ITK N3 implementation from [2]. Then, we align the mean intensities of the images within each channel by a global multiplicative factor. For speed reasons, we run the evaluation on a down-sampled version of the input images, with isotropic spatial resolution of 2mm. The computed segmentations are up-sampled back to 1mm for the evaluation.

Settings. In all tests, we employ forests with $T = 40$ trees of depth $D = 20$.

Runtime. Our segmentation method is fully automatic, with segmentation run times in the range of 1-2 minutes per patient. The training of one tree takes approximately 20 minutes on a single desktop PC.

Results. We evaluated our segmentations by the BraTS online evaluation tool, and we summarize the results for the Dice score in Table 1.

Overall, the results indicate a higher segmentation quality for the high-grade tumors than for the low-grade cases, and a better performance on the synthetic data than the real data set.



(a) Real-HG (b)Real-LG (c) Synth-HG (d) Synth-LG

Fig. 2: Per patient evaluation for the four BraTS data sets (Real-HG, Real-LG, Synth-HG, Synth-LG). We show the results for edema (blue) and tumor tissue (red) per patient, and indicate the respective median results with the horizontal lines. We report the following measures: Dice, Specificity, Precision, Recall(=Sensitivity), Mean Surface Distance (SD), and Maximal SD.

Further Evaluation. Furthermore, we reproduce most of the BraTS measures (except Kappa) by our own evaluation in Figure 2. It can be seen in Figure 2, that the *Specificity* is not a very discriminative measure in this application. Therefore, we rather evaluate *Precision*, which is similar in nature, but does

not take the background class into account (TN), and is thus more sensitive to errors.

In order to obtain a better understanding of the data and the performance of our method we perform three further measurements.

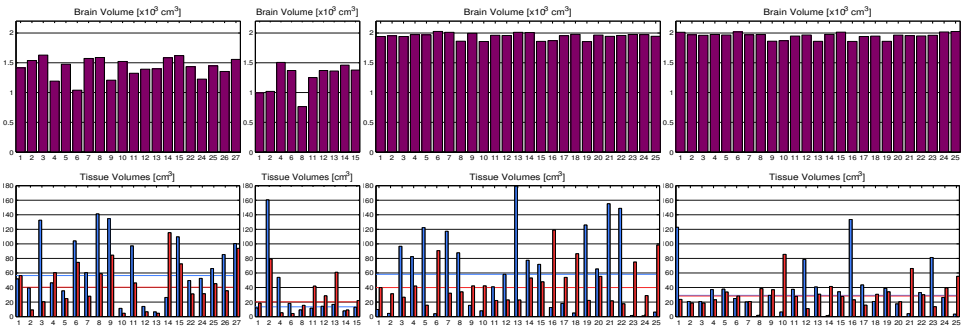
1. In Figure 3, we measure the volumes of the brain, and the edema and tumor tissues for the individual patients. This is done in order to be able to evaluate how target volumes influence the segmentation quality.
2. In Figure 4, we report the results for the basic types of classification outcomes, i.e. true positives (TP), false positives (FP), and false negatives (FN). It is interesting to note the correlation of the TP values with the tissue volumes (cf. Fig. 3). Also, it seems that for edema, the error of our method consists of more FP estimates (wrongly labeled as edema) than FN estimates (wrongly not labeled as edema), i.e. it performs an over-segmentation.
3. In Figure 5, we report additional measures, which might have an application-specific relevance. We compute the overall *Error*, i.e. the volume of all misclassified points $\mathrm{FN} + \mathrm{FP}$, and the corresponding relative version, which relates the error to the target volume T, i.e. $(\mathrm{FN} + \mathrm{FP})/\mathrm{T}$. Also, we compute the absolute and the relative *Volume Error* $|\mathrm{T} - (\mathrm{TP} + \mathrm{FP})|$, and $|\mathrm{T} - (\mathrm{TP} + \mathrm{FP})|/\mathrm{T}$, which indicate the potential performance for volumetric measurements. The volume error is less sensitive than the error measure, since it does not require an overlap of segmentations but only that the estimated volume is correct (volume error can be expressed as $|\mathrm{FN} - \mathrm{FP}|$).

Acknowledgments

S. J. Price is funded by a Clinician Scientist Award from the National Institute for Health Research (NIHR). O. M. Thomas is a Clinical Lecturer supported by the NIHR Cambridge Biomedical Research Centre.

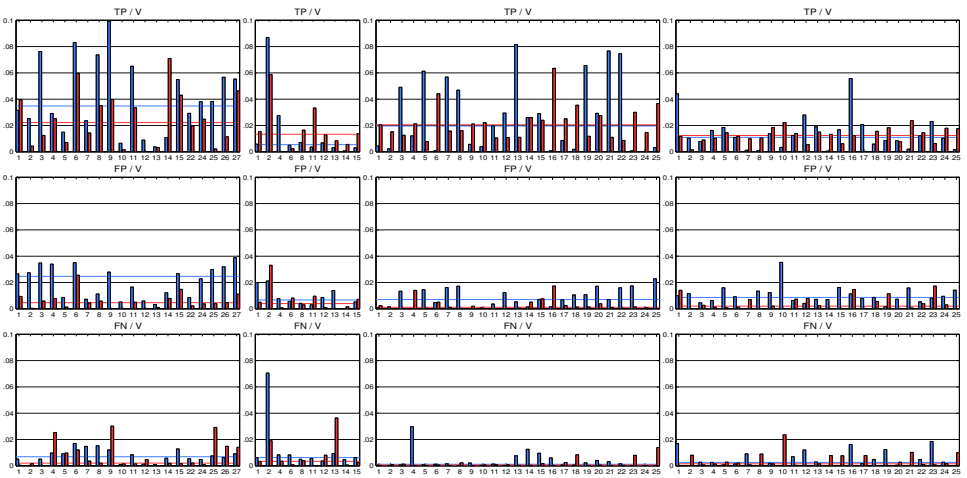
References

1. A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2-3), 2012.
2. N. Tustison and J. Gee. N4ITK: Nick's N3 ITK implementation for MRI bias field correction. *The Insight Journal*, 2010.
3. D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and Price S. J. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In *Proc. Medical Image Computing and Computer Assisted Intervention*, 2012.



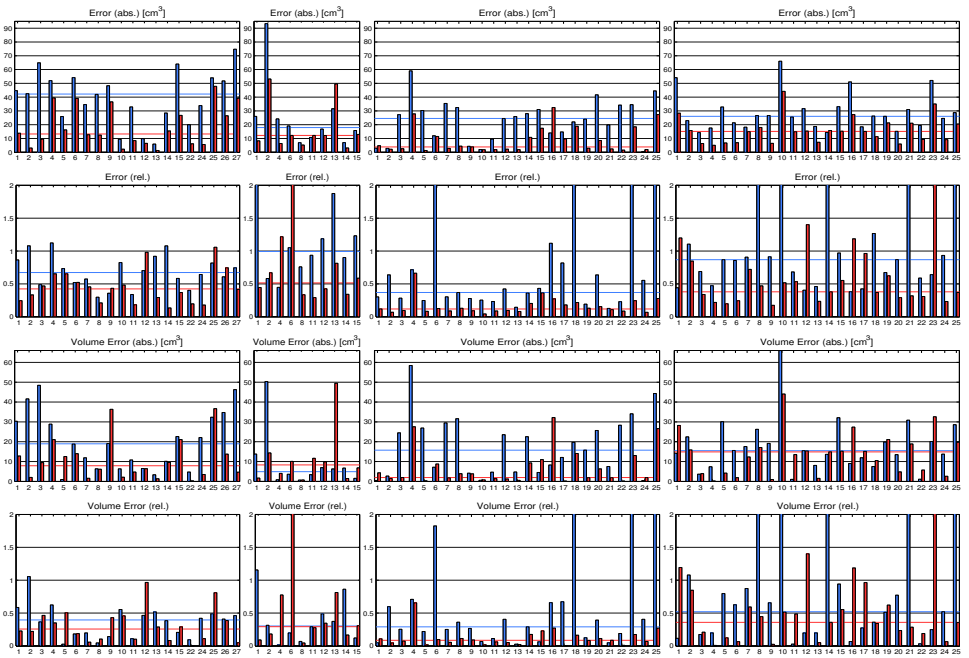
(e) Real-HG (f) Real-LG (g) Synth-HG (h) Synth-LG

Fig. 3: Volume statistics of the BraTS data sets. We compute the brain volumes (top row), and the volumes of the edema (blue) and tumor (red) tissues per patient.



(a) Real-HG (b) Real-LG (c) Synth-HG (d) Synth-LG

Fig. 4: We report the values of true positives (TP), false positives (FP), and false negatives (FN), for edema (blue), and tumor (red) tissues. To make the values comparable, we report them as percentage of the patient brain volume (V). Again, horizontal lines represent median values. It is interesting to note the correlation of the TP values with the tissue volumes (cf. Fig. 3). Also, it seems that for edema, the error of our method consists of more FP estimates (wrongly labeled as edema) than FN estimates (wrongly not labeled as edema), i.e. it performs an over-segmentation.



(a) Real-HG (b)Real-LG (c) Synth-HG (d) Synth-LG

Fig. 5: We further evaluate additional measures which might have application-specific relevance. Again, we have blue=edema, red=tumor, and horizontal line=median. In the two top rows, we compute the *Error*, i.e. the volume of all misclassified points $\mathrm{FN}+\mathrm{FP}$, and the relative version, which relates the error to the target volume T, i.e. $(\mathrm{FN}+\mathrm{FP})/\mathrm{T}$. In the bottom two rows, we compute the absolute and the relative *Volume Error* $|\mathrm{T}-(\mathrm{TP}+\mathrm{FP})|$, and $|\mathrm{T}-(\mathrm{TP}+\mathrm{FP})|/\mathrm{T}$.

Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization

Stefan Bauer[1], Thomas Fejes[1,2], Johannes Slotboom[2], Roland Wiest[2], Lutz-P. Nolte[1], and Mauricio Reyes[1]

[1] Institute for Surgical Technology and Biomechanics, University of Bern
[2] Inselspital, Bern University Hospital, Switzerland
stefan.bauer@istb.unibe.ch

Abstract. We propose a fully automatic method for brain tumor segmentation, which integrates random forest classification with hierarchical conditional random field regularization in an energy minimization scheme. It has been evaluated on the BRATS2012 dataset, which contains low- and high-grade gliomas from simulated and real-patient images. The method achieved convincing results (average Dice coefficient: 0.73 and 0.59 for tumor and edema respectively) within a reasonably fast computation time (approximately 4 to 12 minutes).

1 Introduction

Fast and accurate segmentation of brain tumor images is an important but difficult task in many clinical applications. In recent years, a number of different automatic approaches have been proposed [1], but despite significant intra- and inter-rater variabilities and the large time consumption of manual segmentation, none of the automatic approaches is in routine clinical use yet. However, with the anticipated shift from diameter-based criteria to volume-based criteria in neuroradiological brain tumor assessment, this is likely to change in the future.

We are presenting a fully automatic method for brain tumor segmentation, which is based on classification with integrated hierarchical regularization. Not only does it offer to separate healthy from pathologic tissues, but it also subcategorizes the healthy tissues into CSF, WM, GM and the pathologic tissues into necrotic, active and edema compartment.

2 Methods

The general idea is based on a previous approach presented in [2]. After preprocessing (denoising, bias-field correction, rescaling and histogram matching) [6], the segmentation task is modeled as an energy minimization problem in a conditional random field (CRF) [8] formulation. The energy consists of the sum of the singleton potentials in the first term and the pairwise potentials in the second

term of equation (1). The expression is minimized using [7] in a hierarchical way similar to [2].

$$E = \sum_i V(y_i, \mathbf{x}_i) + \sum_{ij} W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) \tag{1}$$

The singleton potentials $V(y_i, \mathbf{x}_i)$ are computed according to equation (2), where y_i is the label output from the classifier, $\mathbf{x}_i$ is the feature vector and δ is the Kronecker-δ function.

$$V(y_i, \mathbf{x}_i) = p(y_i|\mathbf{x_i}) \cdot (1 - \delta(\tilde{y}_i, y_i)) \tag{2}$$

In contrast to our previous approach, here we make use of random forests [4], [3] as a classifier instead of support vector machines (SVM). Random forests are ensembles of decision trees, which are randomly different. Training on each decision tree is performed by optimizing the parameters of a split function at every tree node via maximizing the information gain when splitting the training data. For testing, the feature vector is pushed through each tree, applying a test at each split node until a leaf node is reached. The label posterior is calculated by averaging the posteriors of the leave nodes from all trees $p(y_i|\mathbf{x}_i) = 1/T \cdot \sum_t^T p_t(y_i|\mathbf{x}_i)$. Compared to SVMs, random forests have the advantage of being able to naturally handle multi-class problems and they provide a probabilistic output instead of hard label separations [5]. We use the probabilistic output for the weighting factor $p(y_i|\mathbf{x}_i)$ in equation (2), in order to control the degree of spatial regularization based on the posterior probability of each voxel label. A 28-dimensional feature vector is used for the classifier, which combines the intensities in each modality with the first-order textures (mean, variance, skewness, kurtosis, energy, entropy) computed from local patches around every voxel in each modality.

We have also developed an improved way to compute the pairwise potentials $W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j)$, which account for the spatial regularization. In equation (3) $w_s(i, j)$ is a weighting function, which depends on the voxel spacing in each dimension. The term $(1 - \delta(y_i, y_j))$ penalizes different labels of adjacent voxels, while the intensity term $\exp\left(\frac{\mathrm{PCD}(\mathbf{x}_i - \mathbf{x}_j)}{2 \cdot \bar{x}}\right)$ regulates the degree of smoothing based on the local intensity variation, where PCD is a pseudo-Chebyshev distance and $\bar{x}$ is a generalized mean intensity. $D_{pq}(y_i, y_j)$ allows us to incorporate prior knowledge by penalizing different tissue adjancencies individually.

$$W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) = w_s(i, j) \cdot (1 - \delta(y_i, y_j)) \cdot \exp\left(\frac{\mathrm{PCD}(\mathbf{x}_i - \mathbf{x}_j)}{2 \cdot \bar{x}}\right) \cdot D_{pq}(y_i, y_j) \tag{3}$$

3 Results

The performance of the proposed method has been evaluated on the BRATS2012 dataset [3] using 5-fold cross-validation. The BRATS2012 dataset contains skull-

[3] http://www2.imm.dtu.dk/projects/BRATS2012/

stripped multimodal MR images (T_1, $T_{1contrast}$, T_2, Flair) of 80 low- and high-grade gliomas from simulations and real patient cases (1mm isotropic resolution). In order to be compatible with the BRATS ground truth, our "necrotic" and "active" labels were combined to form the "core" label, the "edema" label was unmodified and all other labels were ignored.

Quantitative results for different overlap and surface distance metrics, which were obtained using the BRATS2012 online evaluation tool, are detailed in table 1 and exemplary image results are shown in figure 1. Computation time for the segmentation ranged from 4 to 12 minutes depending on the size of the dataset.

We also compared the proposed approach to our previous method [2] which used SVMs as a classifier instead of random forests and which had a less sophisticated regularization. With the new method, the computation time could be reduced by more than a factor of two and the accuracy measured by the Dice coefficient was also improved.

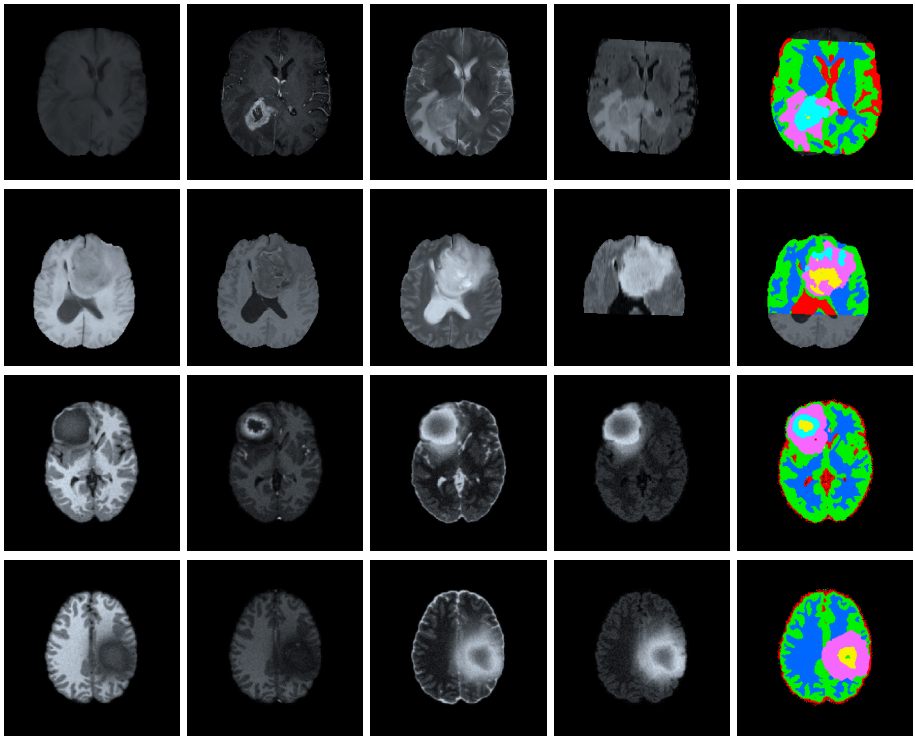


Fig. 1. Exemplary image results shown on one axial slice for a high-grade glioma patient (first row), a low-grade glioma patient (second row), a simulated high-grade glioma dataset (third row) and a simulated low-grade glioma dataset (last row). Each row shows from left to right: T_1, $T_{1contrast}$, T_2, Flair image and the label map obtained from the automatic segmentation (color code: red=CSF, green=GM, blue=WM, yellow=necrotic, turquoise=active, pink=edema.)

Table 1. Quantitative results from the BRATS2012 online evaluation tool. HG stands for high-grade, LG for low-grade and Sim for the simulated glioma datasets. The metrics in the table from left to right are: Dice, Jaccard, sensitivity, specificity, average distance, Hausdorff distance, Cohen's kappa.

		Dice	Jaccard	Sens.	Spec.	AD [mm]	HD [mm]	Kappa
HG	edema	0.61±0.15	0.45±0.15	1.0±0.0	0.56±0.15	5.0±5.3	60±31	0.32±0.25
	tumor	0.62±0.27	0.50±0.25	1.0±0.0	0.59±0.31	6.3±7.8	69±25	
LG	edema	0.35±0.18	0.23±0.13	1.0±0.0	0.49±0.23	10.4±9.2	69±28	0.07±0.23
	tumor	0.49±0.26	0.36±0.24	1.0±0.0	0.49±0.28	5.4±3.8	53±32	
Sim-HG	edema	0.68±0.26	0.56±0.26	1.0±0.0	0.90±0.07	1.3±0.7	12±6	0.67±0.13
	tumor	0.90±0.06	0.81±0.09	1.0±0.0	0.91±0.08	1.5±1.7	16±10	
Sim-LG	edema	0.57±0.24	0.44±0.22	1.0±0.0	0.84±0.17	1.6±0.9	10±6	0.38±0.18
	tumor	0.74±0.10	0.59±0.12	1.0±0.0	0.77±0.18	2.6±1.1	16±5	
All	edema	0.59±0.24	0.45±0.23	1.0±0.0	0.75±0.22	3.5±5.1	30±31	0.42±0.27
	tumor	0.73±0.22	0.61±0.23	1.0±0.0	0.73±0.26	3.6±4.6	34±29	

4 Discussion and Conclusion

We have presented a method for fully automatic segmentation of brain tumors, which achieves convincing results within a reasonable computation time on clinical and simulated multimodal MR images. Thanks to the ability of the approach to delineate subcompartments of healthy and pathologic tissues, it can have a significant impact in clinical applications, especially tumor volumetry. To evaluate this more thoroughly, a prototype of the method is currently being integrated into the neuroradiology workflow at Inselspital, Bern University Hospital.

References

1. Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma Dynamics and Computational Models: A Review of Segmentation, Registration, and In Silico Growth Algorithms and their Clinical Applications. Current Medical Imaging Reviews 3(4) (2007)
2. Bauer, S., Nolte, L.P., Reyes, M.: Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: MICCAI. LNCS, vol. 14. Springer, Toronto (2011)
3. Bochkanov, S., Bystritsky, V.: ALGLIB, `www.alglib.net`
4. Breiman, L.: Random forests. Machine Learning 45(1) (2001)
5. Criminisi, A., Shotton, J., Konukoglu, E.: Decision Forests for Classification , Regression , Density Estimation , Manifold Learning and Semi-Supervised Learning. Tech. rep., Microsoft Research (2011)
6. Ibanez, L., Schroeder, W., Ng, L., Cates, J., Others: The ITK software guide (2003)
7. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. Computer Vision and Image Understanding 112(1) (2008)
8. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: ICML Proceedings. Citeseer (2001)

Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images

E. Geremia[1], B. H. Menze[1,2], N. Ayache[1]

[1] Asclepios Research Project, INRIA Sophia-Antipolis, France.
[2] Computer Science and Artificial Intelligence Laboratory, MIT, USA.

Abstract. A fully automatic algorithm is presented for the automatic segmentation of gliomas in 3D MR images. It builds on the discriminative random decision forest framework to provide a voxel-wise probabilistic classification of the volume. Our method uses multi-channel MR intensities (T1, T1C, T2, Flair), spatial prior and long-range comparisons with 3D regions to discriminate lesions. A symmetry feature is introduced accounting for the fact that gliomas tend to develop in an asymmetric way. Quantitative evaluation of the data is carried out on publicly available labeled cases from the BRATS Segmentation Challenge 2012 dataset and demonstrates improved results over the state of the art.

1 Materials and methods

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

1.1 Dataset

To calculate the local image features – both during training and for predictions – we performed an intensity normalization [1]. For each data group (i.e. BRATS_HG and BRATS_LG), we fitted the intensity histogram of each sequence (T1, T1C, T2 and FLAIR) to a reference case. Then image features are calculated for each voxel $\mathbf{v}$. Features include local multi-channel intensity (T1, T1C, T2, Flair) as well as long-range displaced box features such as in [2]. In addition we also incorporate symmetry features, calculated after estimating the mid-sagittal plane [3]. In total, every voxel is associated with a 412−long vector of feature responses.

We will adhere to the following notation: the data consists of a collection of voxel samples $\mathbf{v} = (\mathbf{x}, \mathbf{C})$, each characterized by a position $\mathbf{x} = (x, y, z)$ and associated with a list of signal channels $\mathbf{C}$. Signal channels $\mathbf{C} = (\mathbf{I}, \mathbf{P})$ include multi-sequence MR images $\mathbf{I} = (I_{T1}, I_{T1C}, I_{T2}, I_{Flair})$ and spatial priors $\mathbf{P} = (P_{WM}, P_{GM}, P_{CSF})$. Anatomical images and spatial priors, although having different semantics, can be treated under the unified term "signal channel". We account for noise in MR images by averaging values over a 3^3 voxels box centered on $\mathbf{x}$, such an average is noted $C_c(\mathbf{x})$, e.g. $C_c = I_{Flair}$ or P_{GM}.

1.2 Context-rich decision forest

Our detection and segmentation problem can be formalized as a multi-class classification of voxel samples into either background, edema or tumor core. This classification problem is addressed by a supervised method: discriminative random decision forest, an ensemble learner using decision trees as base learners. Decision trees are discriminative classifiers which are known to suffer from overfitting. A random decision forest [4] 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 [5].

Forest training. The forest has T components with t indexing each tree. The training data consists in 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, $y(\mathbf{v}) \in \{0, 1, 2\}$, 2 for the tumor core, 1 for edema and 0 for background.

During training, all observations $\mathbf{v}_k$ are pushed through each of the trees. Each internal node applies a binary test [6–9] as follows:

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

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. Training the classifier means selecting the most discriminative binary test for each node by optimizing over $(\tau_{low}, \tau_{up}, \theta)$ in order to maximize the information gain on the input data partition [10], noted $\mathcal{T}_p$, defined as follows: $IG_{\tau_{low},\tau_{up},\theta}(\mathcal{T}_p) = H(\mathcal{T}_p) - H(\mathcal{T}_p|\{t^{\tau_{low},\tau_{up},\theta}(\mathbf{v}_k)\})$ where $\mathcal{T}_p \subset \mathcal{T}$, H stands for the entropy.

Only a randomly sampled subset Θ of the feature space is available for internal node optimization, while the threshold space is uniformly discretized. The optimal $(\tau^*_{low}, \tau^*_{up}, \theta^*)$ is selected 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. Nodes are grown to a maximum depth D. Another stopping criterion is to stop growing a node when too few training points reach it, i.e. when the information gain is below a minimal value IG_{min}.

As a result of the training process, each leaf node l of every tree t receives a partition $\mathcal{T}_{l_t}$ of the training data. The following empirical posterior probability is then stored at the leaf $p_{l_t}(Y(\mathbf{v}) = b) = |\{(\mathbf{v}, Y(\mathbf{v})) \in \mathcal{T}_{l_t} | Y(\mathbf{v}) = b\}|/|\mathcal{T}_{l_t}|$ where $b \in \{0, 1\}$ denotes the background or lesion class, respectively.

Prediction. When applied to a new test data $\mathcal{T}_{test} = \{\mathbf{v}_k\}$, each voxel $\mathbf{v}_k$ is propagated through all the trees by successive application of the relevant

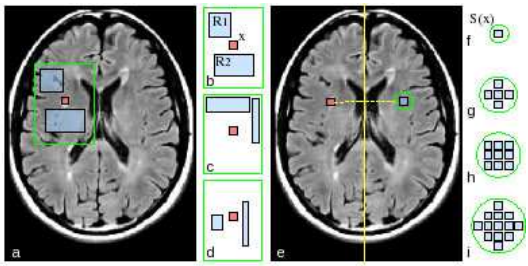


Fig. 1. 2D view of context-rich features. (a) A context-rich feature depicting two regions R_1 and R_2 with constant offset relatively to $\mathbf{x}$. (b-d) Three examples of randomly sampled features in an extended neighborhood. (e) The symmetric feature with respect to the mid-sagittal plane. (f) The hard symmetric constraint. (g-i) The soft symmetry feature considering neighboring voxels in a sphere of increasing radius. See text for details.

binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, posteriors $p_{l_t}(Y(\mathbf{v}) = c)$ are gathered in order to compute the final posterior probability defined as follows: $p(y(\mathbf{v}) = c) = \frac{1}{T}\sum_{t=1}^{T} p_{l_t}(Y(\mathbf{v}) = c)$. The voxel $\mathbf{v}_k$ is affected the class $c \in \{0, 1, 2\}$ which satisfies $c = \arg\max_c p(y(\mathbf{v}) = c)$. For each class, the largest connected component is selected to be the final segmentation.

1.3 Visual features

In this section, two kinds of visual features are computed: 1) local features: $\theta_c^{loc}(\mathbf{v}) = C_c(\mathbf{x})$ where c indexes an intensity or a prior channel; 2) context-rich features comparing the voxel of interest with distant regions . The first context-rich feature looks for relevant 3D regions R_1 and R_2 to compare within an extended neighborhood: $\theta^{cont}_{c_1,c_2,R_1,R_2}(\mathbf{v}) = C_{c_1}(\mathbf{x}) - \frac{1}{vol(R_1 \cup R_2)} \sum_{\mathbf{x}' \in R_1 \cup R_2} C_{c_2}(\mathbf{x}')$ where c_1 and c_2 are two signal channels. The regions R_1 and R_2 are sampled randomly in a large neighborhood of the voxel $\mathbf{v}$ (cf. Fig. 1). The sum over these regions is efficiently computed using integral volume processing [6]. 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 $S(\mathbf{x})$: $\theta_c^{sym}(\mathbf{v}) = C_c(\mathbf{x}) - C_c \circ S(\mathbf{x})$ 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 $\mathcal{S}$ (cf. Fig. 1), centered on $S(\mathbf{x})$. We obtain a softer version of the symmetric feature which reads: $\theta^{sym}_{c,\mathcal{S}}(\mathbf{v}) = \min_{\mathbf{x}' \in \mathcal{S}}\{C_c(\mathbf{x}) - C_c(\mathbf{x}')\}$.

2 Results

In our experiments, forest parameters are fixed to the following values: number of random regions per node $|\Theta| \simeq 100$, number of trees $T = 30$, tree depth $D = 20$, lower bound for the information gain $IG_{min} = 10^{-5}$. These values were chosen based on prior parameter optimization on synthetic data (SimBRATS_HG and SimBRATS_LG) and worked well for real data too.

Table 1. Segmentation of high grade gliomas in the BRATS dataset. Dice, TPR and PPV are reported for the segmentation of the edema only, the core only and the whole tumor.

	Edema			**Core**			**Tumor**		
Patient	Dice	TPR	PPV	Dice	TPR	PPV	Dice	TPR	PPV
HG01	0.46	0.72	0.34	0.74	0.77	0.71	0.65	0.84	0.53
HG02	0.58	0.97	0.41	0.65	0.51	0.89	0.61	0.93	0.46
HG03	0.70	0.88	0.58	0.79	0.99	0.65	0.76	0.95	0.63
HG04	0.43	0.69	0.31	0.45	0.36	0.59	0.78	0.91	0.69
HG05	0.49	0.60	0.41	0.39	0.25	0.92	0.54	0.49	0.61
HG06	0.61	0.77	0.51	0.75	0.69	0.82	0.75	0.84	0.68
HG07	0.63	0.68	0.58	0.76	0.63	0.96	0.70	0.70	0.70
HG08	0.73	0.78	0.69	0.63	0.65	0.62	0.84	0.89	0.80
HG09	0.80	0.81	0.77	0.69	0.55	0.93	0.84	0.79	0.90
HG10	0.00	0.00	0.00	0.80	0.69	0.96	0.09	0.20	0.05
HG11	0.69	0.78	0.61	0.81	0.87	0.76	0.83	0.92	0.75
HG12	0.67	0.88	0.54	0.00	0.00	0.00	0.86	0.91	0.81
HG13	0.49	0.85	0.35	0.92	0.98	0.87	0.66	0.96	0.51
HG14	0.33	0.81	0.20	0.47	0.31	0.92	0.84	0.84	0.84
HG15	0.67	0.83	0.57	0.83	0.76	0.91	0.78	0.86	0.71
HG22	0.63	0.90	0.49	0.51	0.36	0.86	0.69	0.77	0.62
HG24	0.52	0.83	0.37	0.67	0.53	0.91	0.57	0.74	0.47
HG25	0.51	0.57	0.46	0.05	0.02	0.95	0.55	0.48	0.64
HG26	0.66	0.57	0.80	0.03	0.02	0.07	0.57	0.45	0.77
HG27	0.57	0.93	0.41	0.57	0.41	0.98	0.74	0.85	0.65
mean	**0.56**	**0.74**	**0.47**	**0.58**	**0.52**	**0.76**	**0.68**	**0.77**	**0.64**
std	**0.17**	**0.21**	**0.19**	**0.27**	**0.30**	**0.28**	**0.18**	**0.20**	**0.18**

For quantitative evaluation, 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 three times in order to collect test errors for each case. Note that the random forest is trained on the preprocessed data. Prediction on a single image lasts for approximately 10 minutes.

The binary classification is evaluated using two measures, true positive rate (TPR) and positive predictive value (PPV), both equal 1 for perfect segmentation. Formally, $Dice = \frac{TP}{FP+2 \cdot TP+FN}$, $TPR = \frac{TP}{TP+FN}$ and $PPV = \frac{TP}{TP+FP}$ where TP counts the number of true positive voxels in the classification compared to the ground truth, FP the false positives, FN the false negatives.

Aknowledgments

This work was partially supported by the European Research Council through the ERC Advance Grant MedYMA on Biophysical Modelling and Analysis of Dynamic Medical Images.

Table 2. Segmentation of low grade gliomas in the BRATS dataset. Dice, TPR and PPV are reported for the segmentation of the edema only, the core only and the whole tumor.

	Edema			**Core**			**Tumor**		
Patient	Dice	TPR	PPV	Dice	TPR	PPV	Dice	TPR	PPV
LG01	0.00	0.00	0.00	0.83	0.92	0.76	0.71	0.67	0.77
LG02	0.43	0.35	0.56	0.32	0.23	0.49	0.70	0.55	0.96
LG04	0.46	0.35	0.66	0.05	0.16	0.03	0.62	0.62	0.62
LG06	0.45	0.41	0.48	0.18	0.99	0.10	0.49	0.87	0.34
LG08	0.30	0.29	0.32	0.44	0.37	0.55	0.71	0.63	0.81
LG11	0.21	0.46	0.13	0.14	0.24	0.10	0.47	0.86	0.32
LG12	0.26	0.52	0.17	0.00	0.00	0.00	0.49	0.62	0.40
LG13	0.22	0.27	0.18	0.00	0.00	0.00	0.42	0.32	0.61
LG14	0.19	0.20	0.19	0.00	0.00	0.00	0.34	0.47	0.27
LG15	0.34	0.34	0.34	0.00	0.00	0.00	0.22	0.29	0.18
mean	**0.29**	**0.32**	**0.30**	**0.20**	**0.29**	**0.20**	**0.52**	**0.59**	**0.53**
std	**0.14**	**0.15**	**0.21**	**0.27**	**0.37**	**0.28**	**0.17**	**0.19**	**0.26**

References

1. Coltuc, D., Bolon, P., Chassery, J.M.: Exact histogram specification. IEEE TIP **15** (2006) 1143 –1152
2. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images. NeuroImage **57** (2011) 378–90
3. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3d brain images. IEEE Trans. Med. Imaging **21**(2) (2002) 122–138
4. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation **9**(7) (1997) 1545–1588
5. Breiman, L.: Random forests. Machine Learning **45**(1) (2001) 5–32
6. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. Int. J. Comp. Vision **81**(1) (2009) 2–23
7. Yi, Z., Criminisi, A., Shotton, J., Blake, A.: Discriminative, semantic segmentation of brain tissue in MR images. LNCS 5762, Springer (2009) 558–565
8. Lempitsky, V.S., Verhoek, M., Noble, J.A., Blake, A.: Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In: FIMH. LNCS 5528, Springer (2009) 447–456
9. Criminisi, A., Shotton, J., Bucciarelli, S.: 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)
10. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993)

Multimodal Brain Tumor Segmentation Using The "Tumor-cut" Method on The BraTS Dataset

Andac Hamamci, Gozde Unal

Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
gozdeunal@sabanciuniv.edu

Abstract. In this paper, the tumor segmentation method used is described and the experimental results obtained are reported for the "BraTS 2012 - Multimodal Brain Tumor Segmentation Challenge" of MICCAI'12. "Tumor-cut" method, presented in [1] is adapted to multi-modal data to include edema segmentation. The method is semi-automatic, requiring the user to draw the maximum diameter of the tumor, which takes about a minute user-interaction time per case. The typical run-time for each case is around 10-20 minutes depending on the size of the tumor. Overall Dice overlap with the expert segmentation is 0.36 ± 0.25 for the edema/infiltration and 0.69 ± 0.20 for the tumor region.

1 Introduction

In our "Tumor-cut" paper, the semi-supervised tumor segmentation method is described, in detail [1]. This method, specifically targets gross tumor volume (GTV) of the brain tumors on the contrast enhanced T1-weighted MR images (T1C). Here, we extend our method to multi-modal MRI (T1C + FLAIR) to include edema segmentation, and also evaluated on low-grade non-enhanced tumors. Although, we segment the necrotic areas of the high-grade tumors, quantitative necrotic segmentation results are not reported in this paper due to the lack of the ground truth labels of necrotic regions.

2 Methods

The main steps of the "Tumor-cut" method, presented in [1], is given in Fig. 1. The weight of the regularizer of the level-set is set to 0.2, which is determined in [1] by experiments on both real and synthetic data. The same algorithm is applied on FLAIR volumes to segment the clinical tumor volume ($CTV = GTV + Edema$). This time, the user is asked to draw the maximum diameter of the edema region visible on FLAIR images. The main differences observed between GTV (on T1C) and CTV (on FLAIR) is that the sphericity of the edema area is lower and there occurs more curvatures. For the FLAIR segmentation, to allow results with more curvatures, a four times lower regularizer weight is set arbitrarily, at 0.05. The user-input is gathered by two different maximum diameter lines, drawn separately, one for FLAIR and one for T1C volumes. The resulting maps are combined simply by assigning tumor labels using T1C segmentation and assigning edema label to the difference area of the FLAIR segmentation minus the T1C segmentation.

$$V_{Edema} = \{x \in V_{FLAIR} | x \notin V_{T1C}\}$$

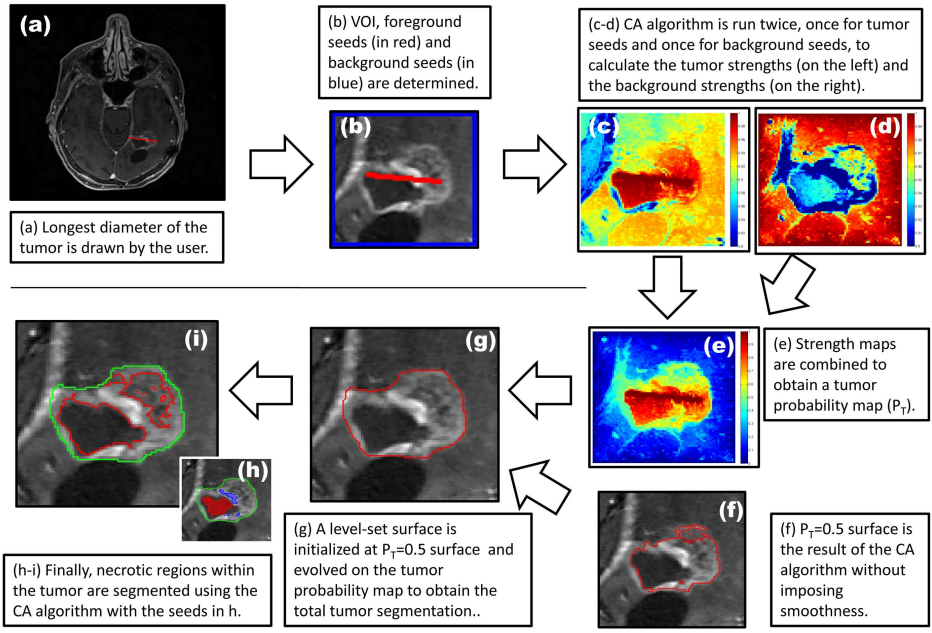


Fig. 1: Flow diagram, which shows the main steps of the "Tumor-cut" algorithm [1].

Table 1: Overall results obtained on the BraTS dataset.

Dice Overlap Edema / Tumor	Jaccard Score Edema / Tumor	Specificity Edema / Tumor	Sensitivity Edema / Tumor	Cohen's Kappa
$0.37 \pm 0.25/0.69 \pm 0.19$	$0.25 \pm 0.20/0.56 \pm 0.23$	$0.99 \pm 0.01/1.00 \pm 0.00$	$0.54 \pm 0.33/0.86 \pm 0.19$	0.27 ± 0.23

3 Results

The method is implemented on Matlab environment, running on a windows 7 workstation, using mex files for core algorithms. For each case, user interaction takes about a minute and typical run-time for each case is around 10-20 minutes, depending on the size of the tumor. The dataset is downloaded from Kitware/Midas web site and the online system provided by the "Virtual Skeleton Database" is used for the evaluation. The Dice overlap results of the runs on each case of the BraTS data set is tabulated in Table 2. For each subset, including high grade and low grade, simulated and patient data, the Dice overlap scores obtained are given as bar charts in Figs 3-6. The overall Dice overlap, Jaccard scores, Sensitivity/Specificity and Cohen's Kappa results with the standard deviations are reported in Table 1.

4 Discussion and Conclusions

Originally we limited the scope of the "Tumor-cut" algorithm to the contrast enhanced gross tumor volumes, which corresponds to the active parts of the high grade tumors in Table 2. For the high-grade cases, the results obtained on the patient dataset (0.73)

PATIENT DATA			
Case	**Edema/Active**	**Case**	**Edema/Active**
LG-1	0.42 / 0.83	*HG-1*	0.71 / 0.91
LG-2	0.34 / 0.55	*HG-2*	0.65 / 0.63
LG-4	0.51 / 0.61	*HG-3*	0.84 / 0.95
LG-6	0.64 / 0.71	*HG-4*	0.19 / 0.75
LG-8	0.20 / 0.78	*HG-5*	0.56 / 0.33
LG-11	0.28 / 0.88	*HG-6*	0.52 / 0.54
LG-12	0.53 / 0.81	*HG-7*	0.64 / 0.76
LG-13	0.31 / 0.48	*HG-8*	0.66 / 0.94
LG-14	0.14 / 0.73	*HG-9*	0.75 / 0.75
LG-15	0.42 / 0.75	*HG-10*	0.44 / 0.92
		HG-11	0.85 / 0.89
		HG-12	0.30 / 0.49
		HG-13	0.74 / 0.95
		HG-14	0.49 / 0.89
		HG-15	0.57 / 0.90
		HG-22	0.73 / 0.54
		HG-24	0.66 / 0.92
		HG-25	0.19 / 0.36
		HG-26	0.30 / 0.61
		HG-27	0.48 / 0.56
Average	**0.38 / 0.71**	**Average**	**0.56 / 0.73**

SIMULATED DATA			
Case	**Edema/Active**	**Case**	**Edema/Active**
Sim-LG-1	0.23 / 0.28	*Sim-HG-1*	0.53 / 0.97
Sim-LG-2	0.13 / 0.53	*Sim-HG-2*	0.46 / 0.95
Sim-LG-3	0.08 / 0.50	*Sim-HG-3*	0.67 / 0.78
Sim-LG-4	0.20 / 0.53	*Sim-HG-4*	0.40 / 0.66
Sim-LG-5	0.10 / 0.52	*Sim-HG-5*	0.67 / 0.70
Sim-LG-6	0.11 / 0.64	*Sim-HG-6*	0.08 / 0.92
Sim-LG-7	0.28 / 0.62	*Sim-HG-7*	0.69 / 0.90
Sim-LG-8	0.04 / 0.78	*Sim-HG-8*	0.56 / 0.89
Sim-LG-9	0.06 / 0.54	*Sim-HG-9*	0.67 / 0.93
Sim-LG-10	0.03 / 0.74	*Sim-HG-10*	0.54 / 0.98
Sim-LG-11	0.13 / 0.49	*Sim-HG-11*	0.39 / 0.92
Sim-LG-12	0.24 / 0.24	*Sim-HG-12*	0.71 / 0.78
Sim-LG-13	0.04 / 0.42	*Sim-HG-13*	0.67 / 0.70
Sim-LG-14	0 / 0.70	*Sim-HG-14*	0.62 / 0.66
Sim-LG-15	0.14 / 0.51	*Sim-HG-15*	0.16 / 0.59
Sim-LG-16	0.30 / 0.29	*Sim-HG-16*	0.24 / 0.93
Sim-LG-17	0.10 / 0.38	*Sim-HG-17*	0.18 / 0.59
Sim-LG-18	0.15 / 0.69	*Sim-HG-18*	0.05 / 0.89
Sim-LG-19	0.33 / 0.60	*Sim-HG-19*	0.75 / 0.71
Sim-LG-20	0.06 / 0.60	*Sim-HG-20*	0.26 / 0.51
Sim-LG-21	0.04 / 0.83	*Sim-HG-21*	0.71 / 0.75
Sim-LG-22	0.20 / 0.57	*Sim-HG-22*	0.62 / 0.69
Sim-LG-23	0.30 / 0.22	*Sim-HG-23*	0.02 / 0.71
Sim-LG-24	0.08 / 0.66	*Sim-HG-24*	0.09 / 0.97
Sim-LG-25	0.03 / 0.81	*Sim-HG-25*	0.08 / 0.90
Average	**0.14 / 0.55**	**Average**	**0.43 / 0.80**

Fig. 2: Table of the Dice overlap results obtained on each case.

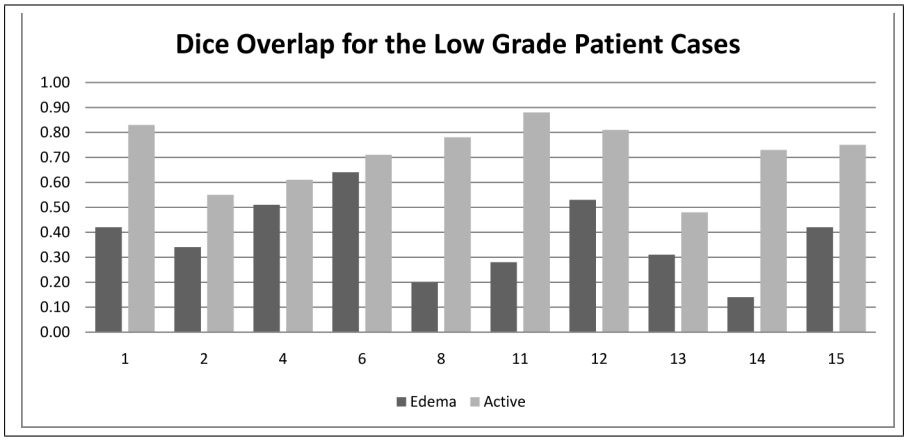


Fig. 3: Dice overlap results obtained on each case of the low-grade patient subset.

and the simulated dataset (0.80) are consistent with the results reported in [1] (0.80 and 0.83). Because the edema region is calculated by substraction of the two maps, the overlap scores for edema is not independent of the scores for the tumor. For the low-grade simulated data, low performance in some cases is mostly due to the low number of voxels labeled as edema -comparing to patient cases- in the ground truth segmentation, which causes low overlap scores.

Acknowledgement. This work was partially supported by TUBA-GEBIP (Turkish Academy of Sciences) and EU FP7 Grant No: PIRG03-GA-2008-231052.

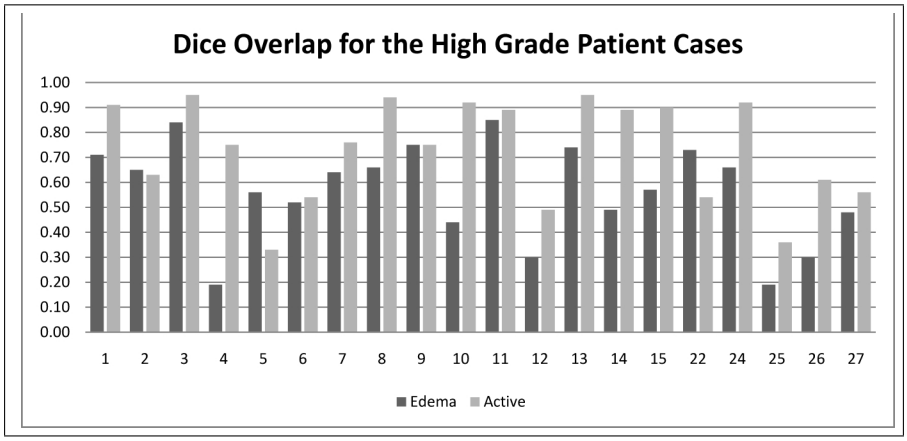


Fig. 4: Dice overlap results obtained on each case of the high-grade patient subset.

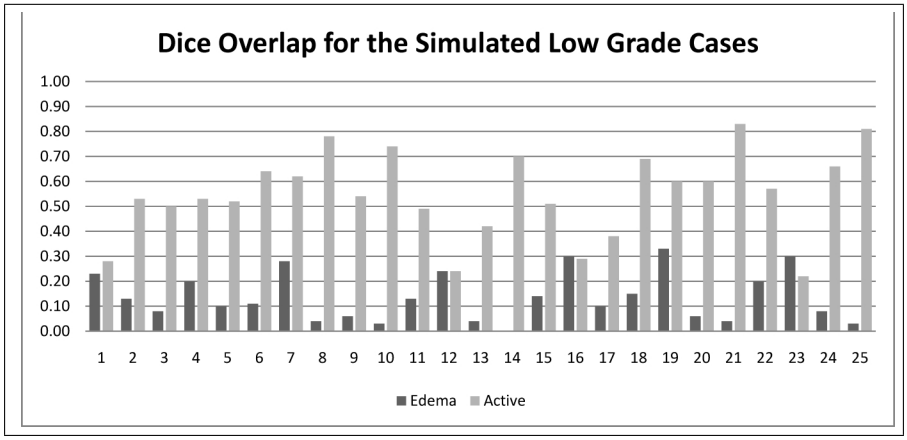


Fig. 5: Dice overlap results obtained on each case of the simulated low-grade subset.

References

1. Hamamci, A., Kucuk, N., Karaman, K., Engin, K., Unal, G.: Tumor-cut: Segmentation of brain tumors on contrast enhanced mr images for radiosurgery applications. Medical Imaging, IEEE Transactions on **31**(3) (march 2012) 790 –804

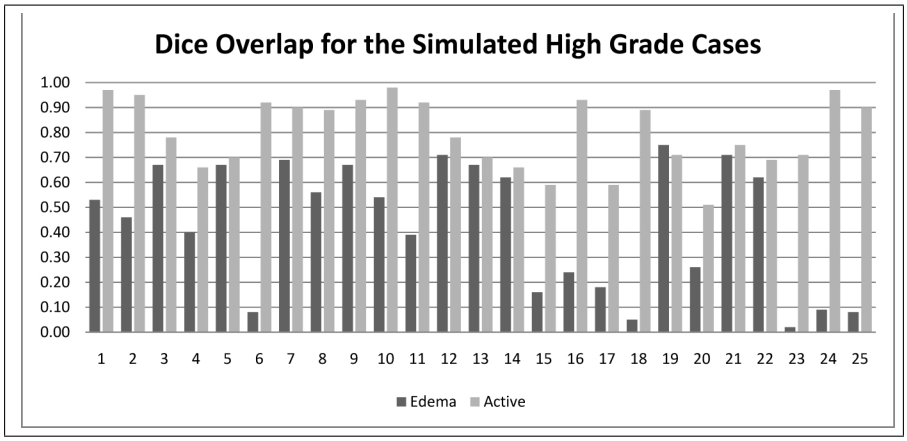


Fig. 6: Dice overlap results obtained on each case of the simulated high-grade subset.

Brain tumor segmentation based on GMM and active contour method with a model-aware edge map

Liang Zhao[1], Wei Wu[1,2], Jason J.Corso[1]

1 Department of Computer Science and Engineering,Unveristy at buffalo SUNY at Buffalo,NY,14260,USA
2 School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China

(lzhao6@buffalo.edu) (wwu25@buffalo.edu) (jcorso@buffalo.edu)

We present a method for automatic segmentation of heterogeneous brain tumors. Our method will take about 30 minutes to process one volume in Matlab.

1 Theory background

Our method combines the model of gray distribution of pixels (Gaussian Mixture Models, GMM) with the edge information between two difference classes of tissue in the brain. High detection precision can be achieved. The core of our method is based on the following three models.

1.1 Gaussian Mixture Model(GMM)

We model five classes of data in brain: brain white matter, brain gray matter, brain csf, tumor and edema.

Denote the parameters of Gaussian component [1] $\varphi_i = \{\phi_i, \mu_i, \Sigma_i\}$, where μ_i is mean of vector and Σ_i is the covariance matrix. The ϕ_i parameter is called the mixing coefficient and describes the relative weight of component i in the complete model. The complete model can be written $\psi = \{k, \varphi_i, \cdots, \varphi_k\}$, where k is the number of component in the data. A mixture model on d-dimension data x is written:

$$P(x;\Psi) = \sum_{i=1}^{k} \phi_i \, p(x;\mu_i,\Sigma_i)$$

$$= \sum_{i=1}^{k} \phi_i \frac{\exp\left(-\frac{1}{2}(x-\mu_i)^T|\Sigma_i|^{-1}(x-\mu_i)\right)}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \quad (1)$$

The standard expectation-maximization algorithm is used to estimate the parameters of each class mixture model in a maximum likelihood formulation.

1.2 Probability distance model between two different tissues

We define a likelihood function $P(\{s_u\}|\{m_u\})$ for the probability of the observed statistics $\{s_u\}$ conditioned on model variables $\{m_u\}$ of pixels $\{u\}$. Denote b_1= white matter, b_2=gray matter, b_3=csf, b_4=edema, b_5 =tumor, According to the characteristic brain MRI images , we assume

$$p = (m_2|s_1, s_2 - s_1, m_1) = p(m_2|s_2 - s_1, m_1) \qquad (2)$$

To characterize an edge between normal brain tissues and abnormal brain tissues (edema) we can deduce the following term (the edge between edema and tumor can be derived similarly).

$$P(m_1 = \text{normal brain}, m_2 = \text{abnormal brain}|s_1, s_2)$$

$$= \sum_{i=1}^{3}\sum_{j=4}^{5} p(\left(m_1 = b_i, m_2 = b_j\right|s_1, s_2)$$

$$= \sum_{i=1}^{3}\sum_{j=4}^{5} p(\left(m_2 = b_j\right|s_1, s_2 - s_1, m_1 = b_{i,})p(m_1 = b_i|s_1)$$

$$= \sum_{i=1}^{3}(\sum_{j=4}^{5} p(m_2 = b_j|s_2 - s_1, m_1 = b_{i,}))\frac{p(s_1|m_1 = b_i)p(m_1 = b_i)}{\sum_{k=1}^{5} p(s_1|m_1 = b_k)p(m_1 = b_k)}$$

$$= \sum_{i=1}^{3}\frac{\sum_{j=4}^{5} p(s_2 - s_1|m_2 = b_j, m_1 = b_i)p(m_2 = b_j|m_1 = b_i)}{\sum_{j=1}^{5} p(s_{2-} s_1|m_j = b_j, m_1 = b_i)p(m_j = b_j)}\frac{p(s_1|m_1 = b_i)p(m_1 = b_i)}{\sum_{k=1}^{5} p(s_1|m_1 = b_k)p(m_1 = b_k)}$$

(3)

1.3 Active contour model

The traditional deformable active contour model [2] is a curve X(S) = [x(s),y(s)],s $\in$[0,1], that move within the image to minimize an energy function. The curve dynamically changes the shape of an initial contour in response to internal and external forces. The internal forces provide the smoothness of the contour. While the external forces push the curve more toward the desired features, such as boundaries. The object contour is extracted when the energy function is minimized. The energy is defined as

$$E = \int_0^1 \frac{1}{2}\left[\alpha|x'(s)|^2 + \beta\left|x''(s)\right|^2 + E_{ext}x(s)\right) ds \qquad (4)$$

where, $x'(s)$ and $x''(s)$ are first and second derivatives of $x(s)$with respect to s. The parameter α controls the tension of the curve and β controls its rigidity. E_{ext} is the

external energy which is calculated from the image data. To minimize the energy function, the snake must satisfy the Euler equation:

$$\alpha x''(s) - \beta x''''(s) - \nabla E_{ext} = 0 \tag{5}$$

According gradient vector flow snake was proposed by Xu and Prince [3], we define a new static external force field called GVF field:

$$F_{ext} = v(x,y) = [u(x,y), v(x,y)] \tag{6}$$

where u and v are the grey changes on x-axis and y-axis of the image ly. F_{ext} can be computed by minimizing the following energy function:

$$E = \iint \mu\left(u_x^2 + u_y^2 + v_x^2 + v_x^2\right) + |\nabla f|^2 \, |v - \nabla f|^2 dxdy$$

where, $u_x, u_{y,} v_x, v_y$ are derivative of x-axis and y-axis respectively. f(x,y) is the edge map (using probability distance model between two different tissues). μ is a regularization parameter governing the tradeoff between the first term and the second term in the formula [4].

2 Algorithm implementation

2.1 Acquiring rough localization information of tumor and edema

On the training data, we run Expectation Maximization for the GMM on each class of data with free means, variances, and mixing proportions. These means are labeled and saved for the testing data.

For the testing data, firstly, we use Graph-based Segmentation to get some super-voxels in the volumes [5]. Secondly, every voxel can be classified by EM for GMM with fixed means, and labeled according to the labeled means. Thirdly, every super-pixel can be classified by using maxing vote's method. Finally, we chose some special super-pixels whose 50% pixels are almost tumor or edema.

2.2 Seeking edges of between different tissues based on probability distance model

We compute the probability distance for every pixel according to (3). A model-aware edge map can be got. The pixel-pair-class likelihoods, $p\left(s_2 - s_1 | m_2 = b_j, m_1 = b_i\right)$ are computed against GMM.

2.3 Acquiring high precision boundary based on snake method

Based on the rough location information achieved in step (1), we can get an initial contour of the object. And combining with the edges from step (2), precise boundaries

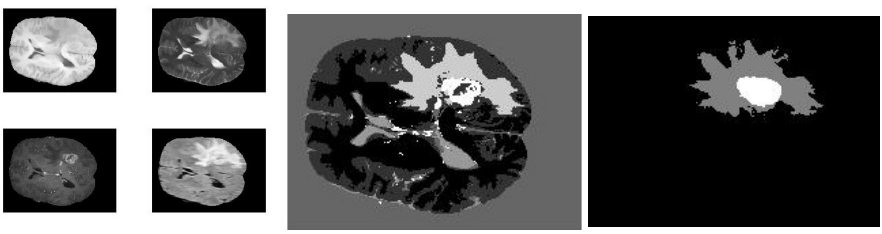
between tumor and edema, and edema and normal brain tissues are located using active contour method.

3 Experiment results

We only process the true data, the whole result can be found on the web. We pick some data and list them as table 1.The figures are raw data, segmentation result with GMM with fixed means and segmentation result with model-aware Snake.

Table 1. Segmentation performance on part of the training data

Subjcet	Average Dist 1	Average Dist 2	Dice 1	Dice 2	Hausdroff Dist 1	Hausdroff Dist 2	Cohen's Kappa	Sensitivity 1	Sensitivity 2	Specificity 1
averages	17.573	0	0.348	0.307	131.772	114.957	0.156	0.367	0.327	0.999
BRATS_LG0008	0	0	0.199	0.83	193.891	180.73	0.16	0.281	0.935	0.998
BRATS_LG0002	123.01	0	0.602	0.322	134.994	143.156	0.353	0.51	0.246	0.997
BRATS_HG0015	0	0	0.716	0.875	0	141.032	0.623	0.572	0.81	1
BRATS_HG0008	0	0	0.728	0.823	113.214	125.41	0.602	0.601	0.751	0.999
BRATS_HG0006	0	0	0.329	0.786	176.529	0	0.348	0.208	0.81	0.999
BRATS_HG0003	0	0	0.699	0.912	146.58	130.088	0.554	0.544	0.92	1
BRATS_HG0002	0	0	0.834	0.537	157.197	84.94	0.341	0.93	0.756	0.998



4 Reference

1. Jason J. Corso, Eitan Sharon, Shishir Dube, Usha Sinha and Alan Yuille, "Efficient multi-level brain tumor segmentation with integrated Bayesian model classification", IEEE trans. *Med,Imag.*,vol.27, no.2, pp.629-640, 2008.
2. Yezzi. A., Kichenassamy.S, Kumar.A, Olver.P, and Tannenbaum.A, "A geometric snake model for segmentation of medical imagery", IEEE trans.*Med,Imag.*,vol.16,no.2,pp.199-209,1997.
3. C. Xu and J.L. Prince, "Snakes, shapes, and gradient vector flow", IEEE Trans. *on Image Processing,* vol.7,no.3, pp. 359-369, 1998.
4. A. Rajendran, R. Dhanasekaran, "Brain tumor segmentation on MRI brain images with fuzzy clustering and GVF snake model", INT J COMPUT COMMUN, Vol.7, No. 3, pp. 530-539, 2012.
5. Xu, C. and Corso, J. J.. "Evaluation of Supervoxel Methods for Early Video Processing" CVPR 2012.

Probabilistic Gabor and Markov Random Fields Segmentation of Brain Tumours in MRI Volumes

N. K. Subbanna and T. Arbel

Centre for Intelligent Machines, McGill University, Montreal, Quebec, Canada

Abstract. In this paper, we present a fully automated technique two stage technique for segmenting brain tumours from multispectral human brain magnetic resonance images (MRIs). From the training volumes, we model the brain tumour, oedema and the other healthy brain tissues using their combined space characteristics. Our segmentation technique works on a combination of Bayesian classification of the Gabor decomposition of the brain MRI volumes to produce an initial classification of brain tumours, along with the other classes. We follow our initial classification with a Markov Random Field (MRF) classification of the Bayesian output to resolve local inhomogeneities, and impose a smoothing constraint. Our results show a Dice similarity coefficient of 0.668 for the brain tumours and 0.56 for the oedema.

1 Introduction

Brain tumours are a serious health problem, and it is estimated that roughly 100,000 people are diagnosed with brain tumours every years. One of the primary diagnostic and treatment evaluation tools for brain tumours is the magnetic resonance image (MRI) of the brain. A reliable method for segmenting brain tumours would be very useful. However, brain tumours, owing to their extreme diversity of shape, size, type of tumour, etc., present a serious challenge to segmentation techniques. Given the importance of the problem, over the years, there have been a large number of techniques attempted to segment brain tumours automatically. Some of the more important techniques include multilevel segmentation by Bayesian weighted aggregation [1], knowledge based fuzzy techniques [2], and atlas based classification [3]. Wavelet based decompositions are attractive since they are good at capturing large textures of the kind found in brain tumours effectively, and it is unsurprising that there are a few attempts to employ wavelets. One of the more prominent is the wavelet decomposition used in conjunction with support vector machines [4]. In this paper, however, we build on this technique by constructing models not for just the tumours and the oedema, but also for the healthy tissues. We, then, utilise the natural ability of the combined space features to capture the existing patterns to train the machine to recognise the patterns of the tumours, and distinguish it from the other healthy tissues, and provide us with an initial classification. From this initial classification, we then use Markov Random Fields (MRFs) to capture the local label homogeneities and also eliminate false positives that occur due to spurious tumour textures that may arise in other parts of the brain.

2 Gabor Bayesian Classification

2.1 Training

Our goal is to correctly identify the oedema and the active tumour. During the initial training phase, we first register a tissue atlas to the training volumes obtain the healthy tissues, Grey Matter (GM), White Matter (WM), and Cerebro-Spinal fluid (CSF). We superimpose the tumour and oedema maps provided by the experts to obtain all the classes in the training volumes.

We decompose the training volumes into their constituent Gabor filter bank outputs. The input volumes are the MRI intensity volumes in the four modalities, viz, T1, T2, T1c and FLAIR, so at each voxel, we have a four dimensional vector $\mathbf{I}_i = (I_i^{T1}, I_i^{T1c}, I_i^{FLAIR}, I_i^{T2})$. Each image is decomposed to its filter bank output using multiwindow Gabor transforms of the form suggested by [5]. The filter bank outputs are obtained by convolving each modality volume with the Gabor filter bank, which is obtained using the equation

$$h(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left(-\frac{1}{2}\left[\frac{x^2}{{\sigma_x}^2} + \frac{y^2}{{\sigma_y}^2}\right]\right) cos(2\pi u_0 x) \tag{1}$$

where σ_x and σ_y are the spreads in the x and y directions and u_0 is the modulating frequency. In our case, we choose 4 orientations between 0 and π radians and 5 frequencies. Each chosen frequency is an octave of the previous to ensure that the entire spectrum is covered. We model each class as a Gaussian mixture model, and an 8 component Gaussian mixture suffices to model the different classes in the combined space. We model the Gabor coefficients of all classes, including the tumour and the oedema, using Gaussian mixture models.

2.2 Classification

Once the test volume is obtained, it is decomposed into its Gabor filter bank outputs using eqn. (1). The class of each voxel is obtained using Bayesian classification, which is given by

$$P(C_i \mid \mathbf{I}_i^G) \propto P(\mathbf{I}_i^g \mid C_i)P(C_i), \tag{2}$$

where C is a random variable that can take the value of the 5 classes, and $\mathbf{I}_i^G = I_i^0, I_i^1, \ldots, I_i^{R-1}$ is the set of R Gabor coefficients of the particular voxel i. It is our experience that the active tumours are quite correctly determined by the Gabor Bayesian technique, but there are often false positive oedema segmentations in regions that mimic the presence of oedema.

3 Markov Random Field Classification

The first Bayesian classification results in tumour candidates. We refine this classification by building an MRF based model. We focus on both the intensities

of the voxels and the intensity differences as contrasts are much more consistent. The MRF model is based on the intensity of the voxel, the spatial intensity differences and the class of neighbouring voxels. This can be written as

$$P(C_i \mid \mathbf{I}_i, \mathbf{I}_{N_i}) = P(\mathbf{I}_i \mid C_i)P(C_i) \sum_{C_{N_i}=0}^{M-1} P(\mathbf{\Delta I}_{N_i} \mid C_i, C_{N_i})P(C_{N_i} \mid C_i) \quad (3)$$

where C_{N_i} are the classes of the neighbours of i, $\mathbf{\Delta I}_{N-I} = \mathbf{I}_{N_i} - \mathbf{I}_i$.

3.1 Training

Here, we build the intensity distributions of the classes and model them using multivariate Gaussian models. For the neighbourhood, we consider an 8 neighbourhood around the voxel in axial plane and the corresponding voxels in the slices above and below, and build the distributions of each pair, triplet and quadriplet of classes that have an edge or vertex in common in the defined neighbourhood using multivariate Gaussian models. This allows us to model all neighbourhood relations completely in both a mathematical and a practical sense. We use the initial Gabor classification as the prior with the oedema probabilities falling sharply away from the tumour for the second phase.

3.2 MRF Classification

We need to compute $P(\mathbf{C} \mid \mathbf{I})$ where $\mathbf{C}$ is a configuration of the labels of all the voxels in the volume and $\mathbf{I}$ is the set of intensities across all the modalities for all the voxels in the configuration. A sound method of computing $P(\mathbf{C} \mid \mathbf{I})$ is by considering the problem as an MRF, which suggests that all class labels are dependent only on their local neighbourhood. Using eqn. (3), we can obtain the energy funcion for the configuration of labels in the volume with

$$\begin{aligned} U(\mathbf{C}) = &\sum_{i=0}^{Z-1} (\mathbf{I}_i - \mu_{C_i})^T \Sigma_{C_i}^{-1} (\mathbf{I}_i - \mu_{C_i}) \\ &+ \sum_{N_i} (\Delta \mathbf{I}_{N_i} - \mu_{C_{N_i}, C_i})^T \Sigma_{C_{N_i}, C_i}^{-1} (\Delta \mathbf{I}_{N_i} - \mu_{C_{N_i}, C_i}) + \alpha m(C_{N_i}, C_i), \quad (4) \end{aligned}$$

where $\Delta I_{N_i} = \mathbf{I}_{N_i} - \mathbf{I}_i$, $m(C_{N_i}, C_i) = 1$ if $C_{N_i} = C_i$, Z is the total number of voxels, and 0 otherwise, and α is the weighting coefficient vector. To maximise $P(\mathbf{C})$, we use iterated conditional modes (ICM) [6] to minimise $U(\mathbf{C})$, where $\mathbf{C}_{min} = \text{argmin}_{C \in \mathcal{F}} U(\mathbf{C})$, and $\mathcal{F}$ is the set of all possible label configurations.

4 Results

In Fig. 1, we compare the results of the two slices where our results are compared against those of the experts' segmentation. In both cases, it can be seen that our results are comparable to the experts' comparison.

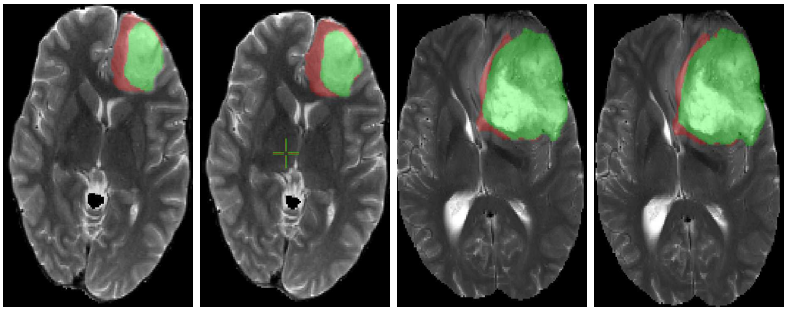


Fig. 1. (a) (b) (c) (d)
(a) Expert labelling of slice 81 of the volume LG0015 and (b) its corresponding algorithmic labelling. (c) Similarly, the expert labelling of slice 93 of the volume HG0015 and (d) its corresponding labelling by the algorithm. As may be seen, visually, our algorithm's performance is very close to the experts' evaluation.

Quantitatively, we train our algorithm on 29 volumes given and test it on the remaining one in a leave one out fashion. We get a Dice similarity coefficient of 0.561 ± 0.118 for the oedema and 0.668 ± 0.126 for the active tumour when we compare our segmentation against those of experts.

References

1. J. J. Corso, et. al., "Efficient Multilevel Brain Tumour Segmentatin with Integrated Bayesian Model Classification", IEEE Trans. Med. Imag., Vol. 27(5), pp. 629-640.
2. M. C. Clark et. al., "Automatic Tumour Segmentation using Knowledge Based Techniques", IEEE Trans. Med. Imag., Vol. 17(2), pp. 187-201.
3. M. Kaus et. al., "Adaptive Template Moderated Brain Tumour Segmentation in MRI", Bildverarbeitung fur die Medizin, pp. 102-106.
4. G. Farias et. al., "Brain Tumour Diagnosis with Wavelets and Support Vector Machines", in Proc. 3rd Int. Conf. Intell. Systems and Knowledge Engg., 2008.
5. A. K. Jain, and F. Farrokhnia, "Unsupervised Texture Segmentation Using Gabor filters" Pattern Recognition, Vol. 24(12), pp. 1167-1186, 1991.
6. R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification", John Wiley and Sons, 2000.

Hybrid Clustering and Logistic Regression for Multi-Modal Brain Tumor Segmentation

Hoo-Chang Shin

CRUK and EPSRC Cancer Imaging Centre
Institute of Cancer Research and Royal Marsden NHS Foundation Trust
Sutton, United Kingdom
{hoo.shin}@icr.ac.uk

Abstract. Tumor is an abnormal tissue type, therefore it is hard to be identified by some classical classification methods. It was tried to find a non-linear decision boundary to classify tumor and edema by a joint approach of hybrid clustering and logistic regression.

Keywords: Sparse dictionary learning, K-means clustering, Logistic Regression

1 Introduction

Classifying tumor is challenging as it represents a collection of some abnormal tissue types which results in not enough labeled training dataset to "learn" about the different characteristics of an unseen tumor. In this situation, clustering would be a viable approach, as it divides a given dataset into a number of sub-groups without requiring the labeled training dataset. While the more classical clustering methods such as k-means and Expectation Maximization (EM) produce good clustering result, they just divide a given dataset into a number of sub-groups, such that there is no "learning" process involved that a learned knowledge can be applied to an unseen dataset.

Sparse dictionary learning [1, 2] has been applied to a number of wide range of disciplines such as signal reconstruction for medical image acquisition [3], image denoising [4], object recognition [5], and medical image segmentation [6]. By representing a given dataset by a combination of some learned dictionary's basis vector sets, the dataset can be clustered. This approach, often referred as sparse coding, was applied for object recognition [7, 8] and multi-modal medical image segmentation [9].

It was noticed that edema is already quite well classified by this method, whereas tumor is not. There are different types of tumor in the dataset, therefore there is no clear pattern in the images of different modalities for tumor. Logistic regression was applied to find a non-linear decision boundary to classify these different types of tumors from a more normal tissues.This is combined with volume-wise k-means clustering method to segment a cluster of tumor-like region.

2 Methods

2.1 Segmentation of Edema by Sparse Coding

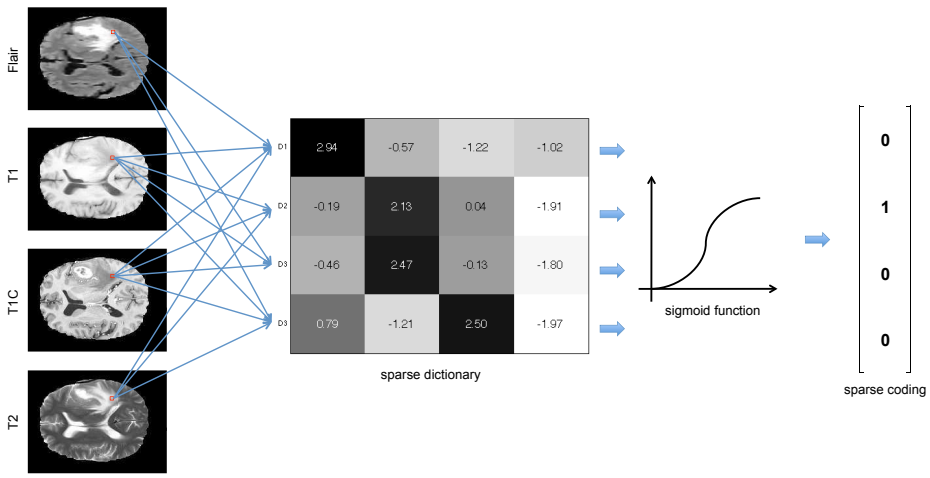


Fig. 1. Sparse coding of a slice in a volume with its multi-modal MR images by a 4×4 size sparse dictionary.

A sparse dictionary for the behavior of the image intensities in the given multiple image modalities (Flair, T1, T1C, and T2) is learned by a sparse autoencoder [10, 11]. The 4×4 sparse basis dictionary is shown in Fig. 1, where each column represents a dictionary entry. The image intensities in a pixel position of different image modalities are convolved with each of the dictionary entry, where the values become a binary digit after being applied a logistic sigmoid function.

Different combinations of dictionary entries represent the characteristics of different tissue types, which results in different binary digit numbers. $15(= 2^4 \quad 1)$ different types of tissue characteristics can be captured by sparse coding with a dictionary of 4×4 size. The hyper-parameters in sparse dictionary learning, such as the size of the dictionary and the sparsity constraints, are chosen by cross-validation.

The visualization of the 15 different tissue types identified by the sparse coding is shown in Fig. 2 with the ground truth of edema and tumor for the slice. It is noticeable that edema is already quite well classified by the bright blue region where it contains a region of a tumor as well. The segmentation performance for edema by sparse coding in F1-score $(2 \cdot (\mathit{precision} \cdot \mathit{recall})/(\mathit{precision} + \mathit{recall}))$ on 10 cross-validation set was 0.54, which outperformed the other methods tried for edema segmentation (logistic regression: 0.246, neural network[1]: 0.14).

[1] neural network classification on image intensities of each voxel in the

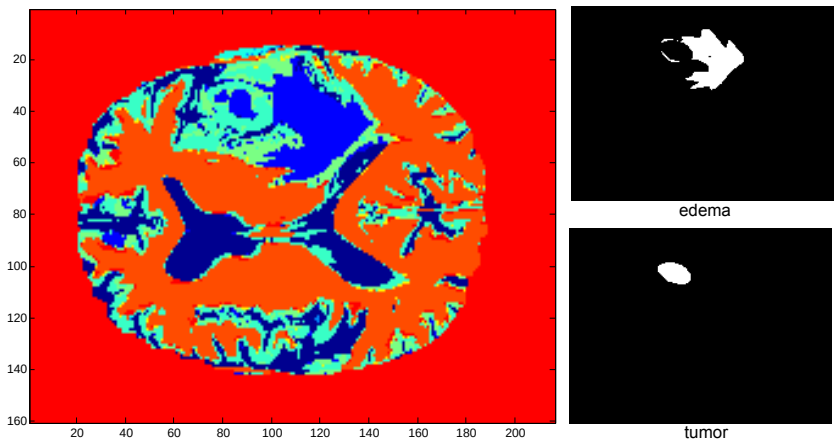


Fig. 2. Tissue type clustering result of an image slice by sparse coding with 4×4 size dictionary.

2.2 Segmentation of Tumor by Logistic Regression and K-means Clustering

Whereas edema was well classified by the sparse coding, tumor was not, probably because tumor is a highly heterogeneous tissue type or just an abnormal region. In the training dataset, tumor in some volumes were well observed by Flair images, whereas in some other volume it was better observed by T1-Contrast images, but no obvious pattern could be found unlike the other tissue types or edema. Therefore, it was tried to find a non-linear decision boundary to separate these "abnormal" tumor from a more "normal" tissue types with logistic regression. Logistic regression models the conditional probability of given feature $\mathbf{x}$ belonging to a class $y \in \{0, 1\}$ as the logistic sigmoid function $(1/(1 + e^{\mathbf{x}}))$. Second order polynomial feature with two combination of the four image modalities were used

$$\mathbf{x} = \{x_1x_2, x_1^2x_2, x_1x_2^2, x_1^2x_2^2, x_1x_3, \cdots, x_1^2x_4^2, x_3x_4, x_3^2x_4, x_3x_4^2, x_3^2x_4^2\}$$

, where x_1, x_2, x_3, x_4 represent the image intensities of each image modality. The F1-score for tumor classification with this method was 0.246 on a cross-validation dataset with 10 volumes, outperforming the other classification methods for tumor classification (sparse coding: 0.0031, neural network: 0.0001).

Another insight for the tumor segmentation was, that a tumor is usually surrounded by an edema. Consequently, if the region found by the logistic regression is inside the already segmented edema region, it is regarded as well classified. When the region segmented by logistic regression is outside of the edema region, and then it is regarded as segmentation failure. In this case, a multi-modal k-means clustering was applied to capture a cluster of a region within edema but has a different characteristic than edema.

3 Results

The average segmentation performance evaluated are shown in Table 1. No human input is required during the segmentation process, and the average segmentation time for a single patient volume was about 1.8 minutes.

Table 1. Performance evaluation results

Av.Dist.1	Av.Dist.2	Dice.1	Dice.2	Haus.1	Haus.2	Cohen's	Sens.1	Sens.2	Spec.1	Spec.2
6.526	15.478	0.391	0.3	37.883	94.282	0.144	0.511	0.416	0.992	0.995

References

1. Lewicki, M.S. and Sejnowski, T.J.: Learning overcomplete representations. Neural Computation, 12, 337–365 (2000)
2. Kreutz-Delgado, K., Murray, J.F., Rao, B.D., Engan, K., Lee, T.W., Sejnowski, T.J.: Dictionary learning algorithms for sparse representation. Neural Computation, 15, 349–396 (2003)
3. Ravishankar, S., Bresler, Y.: MR image reconstruction from highly undersampled k-space data by dictionary learning, IEEE Transactions on Medical Imaging, 30, 1028–1041 (2011)
4. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries, IEEE Transactions on Image Processing, 15, 3736–3745 (2006)
5. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31, 210–227 (2009)
6. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D., Zhou, X.: Deformable segmentation via sparse shape representation. In Proceedings of MICCAI 2011, 451–458 (2011)
7. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification, In Proceedings of CVPR 2009, 1794–1801, (2009)
8. Boureau, Y.L., Bach, F., LeCun, Y., Ponce, J.: Learning mid-level features for recognition, In Proceedings of CVPR 2011, 2559–2566, (2010)
9. Shin, H.C., Orton, M., Collins, D.J., Doran, S., Leach, M.O.: Autoencoder in Time-Series Analysis for Unsupervised Tissues Characterisation in a Large Unlabelled Medical Image Dataset, In Proceedings of ICMLA 2011, 259–264 (2011)
10. Marc'Aurelio, R., Boureau, L., and LeCun, Y.: Sparse Feature Lerning for Deep Belief Networks, Advances in Neural Information Processing Systems, vol. 20 (2007)
11. Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P.: Exploring strategies for training deep neural networks, The Journal of Machine Learning Research, 10, 1-40 (2009)

Hierarchical Random Walker for Multimodal Brain Tumor Segmentation

Yang Xiao, Jie Hu

School of biomedical engineering, Southern Medical University, Guangzhou, China
genq@163.com

Abstract. In this text, a Random Walker (RW) based method is proposed for brain tumor MR images segmentation with interaction. Not only the method is designed to achieve a final segmentation result, but also it is a convenient tool for users to modify their results iteratively. To overcome the shortcoming of typical RW algorithm, we extend RW to feature space for soft clustering, and then carry out pixel-wise segmentation in image space. Proposed method is performed on multimodal brain MR images, including T2-weighted, contrast enhanced T1-weighted, and FLAIR sequences.

Keywords: image segmentation, brain tumor, Random Walker, interactive, hierarchical

1 Introduction

Segmentation of brain tumor image is very important. Because the inter-patient variation of tumor shape, position, texture, and size is large, a robust fully-automatic method is difficult to design. In addition, intervention is needed when automatic result is evaluated and validated. Therefore, we focus on interactive algorithms.

Since the Graph Cuts (GC) algorithm is brought to attention [1], a family of interactive segmentation methods are proposed and of interest for recent decade, such as Random Walker (RW) [2] and Shortest Path (SP) [3], which provide another chance to modify the segmentation results when automatic implementation is unavailable or incorrect, especially on medical images. In [4], these popular interactive algorithms are unified into a common optimization framework. However, RW is distinguished from others due to its particular soft segmentation. In next section, RW will be discussed, and our improvement of typical algorithm and its application for multimodal brain tumor MR will be illustrated.

2 Method

2.1 Typical Random Walker

The RW algorithm is formulated on a weighted graph and optimized by solving a Dirichlet problem. The objective function is given by

$$\min_{x} \sum_{e_{i,j} \in E} w_{i,j}(x_i - x_j)^2, \quad \text{s.t. } x(F) = 1,\ x(B) = 0, \tag{1}$$

where x is the target soft label, and w is similarity measure between pixel i and j, and the constraints are set by interaction.
Furthermore, similarity measure w is often defined as

$$w_{i,j} = \exp(-\beta(\nabla I_{i,j})^2), \tag{2}$$

where $\nabla I_{i,j}$ denotes absolute distance in feature space between pixel i and j, and β is a parameter given by designer.

2.2 RW in feature space

Similarly, the discrete feature space is defined as an edge-weighted graph, and the objective function of clustering is given by

$$\min_{\mu} \sum_{e_{i,j} \in E} s_{i,j}(\mu_i - \mu_j)^2, \quad \text{s.t. } \mu(F) = 1,\ \mu(B) = 0, \tag{3}$$

where μ represents clustering label, as well as s is similarity measure of distribution density between sample i and j in feature space, which is defined by

$$s_{i,j} = \exp(-\alpha(\nabla P_{i,j})^2), \tag{4}$$

where $\nabla P_{i,j}$ is defined as variation of distribution density in feature space, and α is a parameter given by designer.
In typical RW, similarity measure is defined by its negative correlativity to absolute distance in feature space, which neglects prior label information provided by user. To mine the prior information from the seeds as much as possible, we extend RW to feature space as a soft semi-supervised clustering. When the clustering labels are obtained, we will re-define the similarity measure as in (2) by relative distance between labels as follow,

$$w_{i,j} = \exp(-\beta(\mu_i - \mu_j)^2). \tag{5}$$

2.3 Proposed Method

The whole workflow of proposed method is illustrated in Fig. 1. In the pre-processing step, T2, CET1, and FLAIR sequences are filtered by Gaussian kernel. To initialize the hierarchical RW, the user has to select two groups of seeds which locate tumor and edema respectively, and the interface for interaction, where the user can input seeds in images of any modal, is shown in Fig. 2.

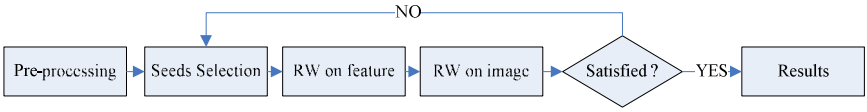


Fig. 1. Flow diagram of proposed method.

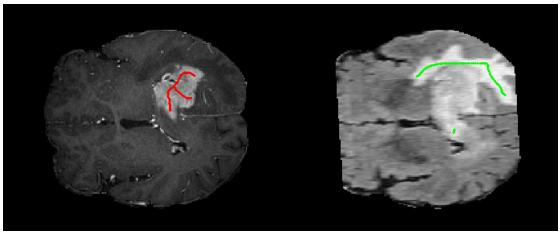


Fig. 2. Interface for selecting seeds.

After the initialization, the core algorithm of proposed method, which is hierarchical RW, is performed, and then results are displayed in the interface for user's iterative modification.

RW is carried out in T2-FLAIR intensity space, as described in last sub-section, to identify the abnormal (tumor and edema), and in T2-T1C intensity space to identify the tumor respectively. In Fig. 3, the joint distribution density and results of clustering is shown.

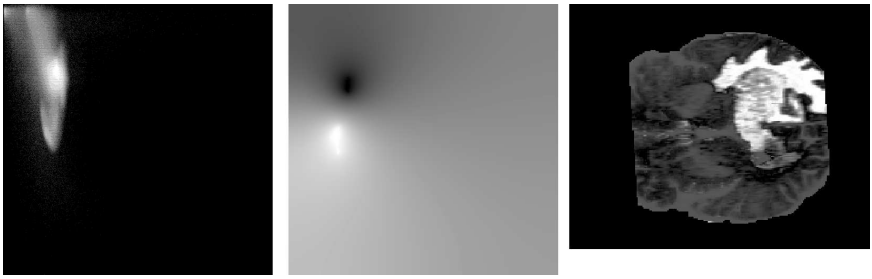


Fig. 3. From left to right: density in T2-FLAIR space; soft clustering labels; and images labeled by clustering result.

The labels of clustering are utilized to calculate similarity measure of pixel-pair according to (5), and typical RW in image space is employed to obtain segmentation labels, as shown in Fig. 4. If the segmentation is not accepted, the user could iteratively revise the result until satisfied.

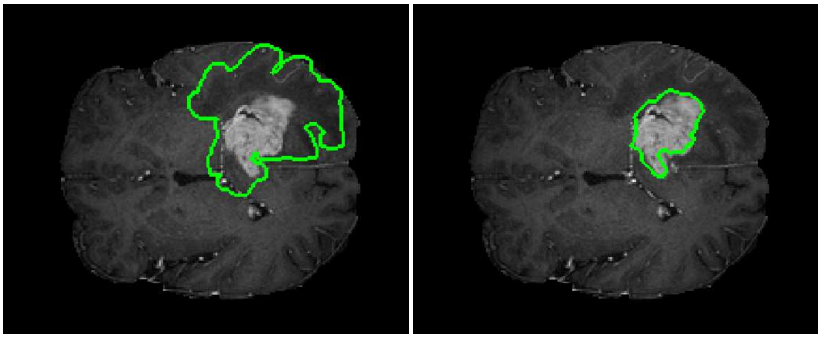


Fig. 4. Illustration of segmentation results in some slice: edema (left) and tumor (right).

3 Results

The proposed method was applied to 80 sets of brain tumor MR images, of which 12 groups failed due to high space complexity of RW. With enough interaction, a user could gain any segmentation result by an interactive segmentation algorithm, so we evaluate the results acquired from single interaction without any modification. The performance of our method is illustrated in Table 1 and Table 2 where 7 segmentation performance metrics are introduced.

Table 1. The mean of evaluation metrics.

Mean

		Dice		Jaccard		Specificity		Sensitivity		Avg Dist.		Hausdorff Dist.		Kappa
Edema	Tumor	0.25	0.53	0.17	0.41	0.82	0.82	0.39	0.65	8.24	4.91	37.38	30.89	0.18

Table 2. The standard deviation of evaluation metrics.

Standard Deviation

		Dice		Jaccard		Specificity		Sensitivity		Avg Dist.		Hausdorff Dist.		Kappa
Edema	Tumor	0.21	0.15	0.17	0.16	0.00	0.00	0.22	0.15	25.97	18.16	51.21	51.48	0.18

4 Discussion and Conclusion

We presented a hierarchical RW method which assists user to locate tumor and edema interactively in dozens of seconds. The improved algorithm performed better than typical one, while the feature space is multi-dimensional.

The RW is optimized by solving a large sparse equation, which needs adequate memory resource, and when the size of image is huge, a technique for linear equation is necessary.

References:

1. Boykov, Y.Y. and M.P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. in Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. 2001.
2. Grady, L., Random walks for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 2006. 28(11): p. 1768-1783.
3. Xue, B. and G. Sapiro. A Geodesic Framework for Fast Interactive Image and Video Segmentation and Matting. in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. 2007.
4. Couprie, C., et al., Power Watershed: A Unifying Graph-Based Optimization Framework. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011. 33(7): p. 1384-1399.

Automatic Brain Tumor Segmentation based on a Coupled Global-Local Intensity Bayesian Model

Xavier Tomas-Fernandez[1] and Simon K. Warfield[1]

Children's Hospital Boston, Boston MA 02115 , USA,
Xavier.Tomas-Fernadez@childrens.harvard.edu,
http://crl.med.harvard.edu/

Abstract. Localizing and quantifying the volume of brain tumor from magnetic resonance images is a key task for the analysis of brain cancer. Based in the well established global gaussian mixture model of brain tissue segmentation, we present a novel algorithm for brain tissue segmentation and brain tumor detection. We propose a tissue model which combines the patient global intensity model with a population local intensity model derived from an aligned reference of healthy subjects. We used the Expectation-Maximization algorithm to estimate the parameters which maximize the tissue maximum a posteriory probabilities. Brain tumors were modeled as outliers with respect to our coupled global/local intensity model. Brain tumor segmentation was validated using the 30 glioma patients scans from the training dataset from MICCAI BRATS 2012 challenge.

Keywords: Magnetic Resonance Imaging, Segmentation, Brain Tumor, Bayesian, Outliers

1 Introduction

Quantifying the volume of a brain tumor is a key indicator of tumor progression. The standard technique to obtain an accurate segmentation is through manual labeling by experienced technicians. However, this manual process is not only very time consuming, but prone to inherent inter- and intra-rater variability. Therefore, there is a need for automated methods for brain tumor segmentation that can analyse large amounts of data in a reproducible way and correlates well with expert analyses.

Several approaches to brain tumor segmentation used intensity based statistical classification algorithms. Kaus et al [3] used the adaptive template-moderated classification algorithm [9] to segment the MR image into four healthy tissue classes and tumor. Their technique proceeded as an iterative sequence of spatially varying classification and nonlinear registration. Prastawa et al. [7] detect outliers based on refined intensity distributions for healthy brain tissue initially derived from a registered probabilistic atlas.

Intensity based classification relies on contrast between tissue types in feature space and adequate signal compared to image noise. Statistical classification identifies an optimal boundary, in feature space, between tissue types, and the separability of two tissue classes is related to the overlap between classes in feature space. Brain tumor MRI intensity feature space overlap with those of brain healthy tissues making difficult to identify a decision boundary, resulting in increased tissue misclassification.

In this paper, we present a novel algorithm for brain tissue segmentation and brain tumor detection which extents the well established global Bayesian model of brain tissue segmentation that has been proposed previously in several works (i.e. [10]). We propose a novel tissue model which combines a patient global intensity model with a population local intensity model derived from a reference of healthy subjects. We reformulate the challenge of detecting brain tumors by modeling them as outliers with respect to our coupled global/local intensity model derived from the target subject and an aligned healthy reference population.

2 Methods

In this section, we first introduce a local intensity prior probability estimated from a reference healthy population. This local tissue intensity prior and a global Bayesian model of brain tissue intensities will be the foundation of our coupled local and global Bayesian tissue model which we introduce in Section 2.2. Following, the procedure to detect brain tumors, will be described in Section 2.3.

2.1 Local Reference Population Bayesian Intensity Tissue Model

Consider a reference population $\mathbf{P}$ formed by R healthy subjects aligned to the subject of interest. Each reference subject is composed by a multispectral grayscale MRI $\mathbf{V}$ (i.e. T1w, T2w and FLAIR) and the corresponding tissue segmentation $\mathbf{L}$ (i.e. GM, WM and CSF), thus $\mathbf{P} = \{\mathbf{V}, \mathbf{L}\} = \{\mathbf{V}_1, ..., \mathbf{V}_R; \mathbf{L}_1, ..., \mathbf{L}_R\}$. Each reference grayscale MRI $\mathbf{V_r} = \{\mathbf{V}_{r1}, ..., \mathbf{V}_{rN}\}$ is formed by a finite set of N voxels with $\mathbf{V}_{ri} \epsilon \mathbb{R}^m$. Also each reference tissue segmentation $\mathbf{L_r} = \{\mathbf{L}_{r1}, ..., \mathbf{L}_{rN}\}$ is formed by a finite set of N voxels with $\mathbf{L}_{ri} = \mathbf{e}_k = \{l_{ri1}, ..., l_{riK}\}$ is a K-dimensional vector with each component l_{rik} being 1 or 0 according whether $\mathbf{v}_{ri}$ did or did not arise from the kth class.

At each voxel i, the reference population $\mathbf{P}$ intensity distribution will be modeled as a Gaussian Mixture Model parametrized by $\boldsymbol{\xi}_i = \{\boldsymbol{\pi}_{\mathbf{P}i}; \boldsymbol{\mu}_{\mathbf{P}i}, \boldsymbol{\Sigma}_{\mathbf{P}i}\}$. Where $\boldsymbol{\pi}_{\mathbf{P}i}$, $\boldsymbol{\mu}_{\mathbf{P}i}$ and $\boldsymbol{\Sigma}_{\mathbf{P}i}$ are respectively the population tissue mixture vector, the population mean intensity vector and the population intensity covariance matrix at voxel i. Because $\{\mathbf{V}, \mathbf{L}\}$ are observed variables, $\boldsymbol{\xi}_i$ can be derived using the following expressions:

$$\pi_{Pik} = \frac{1}{R} \sum_{j=1}^{R} p(L_{ij} = e_k) \tag{1}$$

$$\boldsymbol{\mu}_{Pik} = \frac{\sum_{j=1}^{R} \mathbf{V}_{ij} p(L_{ij} = e_k)}{\sum_{j=1}^{R} p(L_{ij} = e_k)} \tag{2}$$

$$\boldsymbol{\Sigma}_{Pik} = \frac{\sum_{j=1}^{R} (\mathbf{V}_{ij} - \boldsymbol{\mu}_{Pik})^T (\mathbf{V}_{ij} - \boldsymbol{\mu}_{Pik}) p(L_{ij} = e_k)}{\sum_{j=1}^{R} p(L_{ij} = e_k)} \tag{3}$$

where $p(L_{ij} = e_k)$ is the probability of voxel i of the j_{th} reference subject belong to tissue k given by $\mathbf{L}_j$.

2.2 Coupling Global and Local Models

Consider that the observed data $\mathbf{Y}$ and hidden labels $\mathbf{Z}$ are described by a parametric conditional probability density function defined by $\{\boldsymbol{\psi}, \boldsymbol{\xi}\}$, where $\boldsymbol{\psi}$ parametrize a within subject global GMM and $\boldsymbol{\xi}$ parametrize a local GMM derived from an aligned reference of R healthy subjects $\mathbf{P}$.

Segmenting the observed data $\mathbf{Y}$ implies the estimation of parameters $\{\boldsymbol{\psi}, \boldsymbol{\xi}\}$.

$$\begin{aligned} log\, L_C(\boldsymbol{\psi}, \boldsymbol{\xi}) &= log\,(f(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\psi}, \boldsymbol{\xi})) \\ &= log\left(\prod_{i=1}^{N} \sum_{k=1}^{K} f(\mathbf{Z}_i = e_k|\boldsymbol{\psi}_k, \boldsymbol{\xi}_{ik}) f(\mathbf{Y}_i|\mathbf{Z}_i = e_k, \boldsymbol{\psi}_k, \boldsymbol{\xi}_{ik})\right) \end{aligned} \tag{4}$$

From Section 2.1, the local intensity model is just dependent on the aligned reference population $\mathbf{P} = \{\mathbf{V}, \mathbf{L}\}$ which is observable. Thus, the parameter $\boldsymbol{\xi}$ will be constant. We can rewrite the complete log-likelihood as:

$$\begin{aligned} log\, L_C(\boldsymbol{\psi}, \boldsymbol{\xi}) &= \log\,(f(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\psi}, \boldsymbol{\xi})) = \log\,(f(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\psi}) f(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\xi})) \\ &= \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} f(\mathbf{Z}_i = \mathbf{e}_k|\boldsymbol{\psi}_k) f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_k, \boldsymbol{\psi}_k)\right) \\ &+ \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} f(\mathbf{Z}_i = \mathbf{e}_k|\boldsymbol{\xi}_{ik}) f(\mathbf{Y}_i|\mathbf{Z}_i = e_k, \boldsymbol{\xi}_{ik})\right) \end{aligned}$$

where $\mathbf{e}_k$ is a K-dimensional vector with each component being 1 or 0 according whether $\mathbf{Y}_i$ did or did not arise from the kth class. Thus:

$$\begin{aligned} log\, L_C(\boldsymbol{\psi}, \boldsymbol{\xi}) &= \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ij} \left(\log\left(f(\mathbf{Z}_i = \mathbf{e}_k|\boldsymbol{\psi}_k) + f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_k, \boldsymbol{\psi}_k)\right)\right) \\ &+ \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ij} \left(\log\left(f(\mathbf{Z}_i = \mathbf{e}_k and|\boldsymbol{\xi}_{ik}) + f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_k, \boldsymbol{\xi}_{ik})\right)\right) \end{aligned} \tag{5}$$

Because the underlying tissue segmentation $\mathbf{Z}$ is unknown, the EM algorithm will be used to find the parameters that maximize the complete log-likelihood.

E-Step: Because $\boldsymbol{\xi}$ is constant, the E-step of the coupled model requires the computation of the conditional expectation of $\log\left(L_C(\boldsymbol{\psi},\boldsymbol{\xi})\right)$ given $\mathbf{Y}$, using $\boldsymbol{\psi}^{(m)}$ for $\boldsymbol{\psi}$.

$$Q(\boldsymbol{\psi},\boldsymbol{\xi};\boldsymbol{\psi}^{(m)},\boldsymbol{\xi}^{(m)}) \sim Q(\boldsymbol{\psi};\boldsymbol{\psi}^{(m)}) = E_{\boldsymbol{\psi}^{(\mathbf{m})}}\{\log L_C(\boldsymbol{\psi},\boldsymbol{\xi})|\mathbf{Y}\}$$

The complete log-likelihood is linear in the hidden labels z_{ij}. The E-step requires the calculation of the current conditional expectation of $\mathbf{Z}_i$ given the observed data $\mathbf{Y}$:

$$\begin{aligned} E_{\psi^{(\mathbf{m})}}(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y},\boldsymbol{\xi}_{ij}) &= f(\mathbf{Z}_i = \mathbf{e}_j|\mathbf{Y}_i,\boldsymbol{\xi}_{ij},\boldsymbol{\psi}_j^{(m)}) \\ &= \frac{f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_j,\boldsymbol{\xi}_{ij},\boldsymbol{\psi}_j^{(m)})f(\mathbf{Z}_i = \mathbf{e}_j|\boldsymbol{\xi}_{ij},\boldsymbol{\psi}_j^{(m)})}{\sum_{k=1}^{K} f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_k,\boldsymbol{\xi}_{ik},\boldsymbol{\psi}_k^{(m)})f(\mathbf{Z}_i = \mathbf{e}_k|\boldsymbol{\xi}_{ik},\boldsymbol{\psi}_k^{(m)})} \end{aligned} \tag{6}$$

Note that:

$$f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_j,\boldsymbol{\xi}_{ij},\boldsymbol{\psi}_j^{(m)}) \sim f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_j,\boldsymbol{\psi}_j^{(m)})f(\mathbf{Y}_i|\mathbf{Z}_i = \mathbf{e}_j,\boldsymbol{\xi}_{ij})$$

$$f(\mathbf{Z}_i = \mathbf{e}_j|\boldsymbol{\xi}_{ij},\boldsymbol{\psi}_j^{(m)}) \sim f(\mathbf{Z}_i = \mathbf{e}_j|\boldsymbol{\psi}_j^{(m)})f(\mathbf{Z}_i = \mathbf{e}_j|\boldsymbol{\xi}_{ij})$$

M-Step: Because the local reference population model parameter $\boldsymbol{\xi}$ is constant, the Maximization step will consist in the maximization of $Q(\boldsymbol{\psi},\boldsymbol{\xi};\boldsymbol{\psi}^{(m)},\boldsymbol{\xi}^{(m)})$ with respect to $\boldsymbol{\psi}$, which results in the following update equations:

$$f(\mathbf{Z}_i = \mathbf{e}_k|\boldsymbol{\psi}_k^{(m+1)}) = \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y}_i,\boldsymbol{\psi}^{(m)},\boldsymbol{\xi}_{ik}) \tag{7}$$

$$\boldsymbol{\mu}_k^{(m+1)} = \frac{\sum_{i=1}^{N}\mathbf{Y}_i\, f(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y},\boldsymbol{\psi}^{(m)})}{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y},\boldsymbol{\psi}^{(m)})} \tag{8}$$

$$\boldsymbol{\Sigma}_k^{(m+1)} = \frac{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y},\boldsymbol{\psi}^{(m)})(\mathbf{Y}_i - \boldsymbol{\mu}_k^{(m)})^T(\mathbf{Y}_i - \boldsymbol{\mu}_k^{(m)})}{\sum_{i=1}^{N} f(\mathbf{Z}_i = \mathbf{e}_k|\mathbf{Y},\boldsymbol{\psi}^{(m)})} \tag{9}$$

2.3 Brain Tumor Detection

Intuitively from (5), the local intensity model downweights the likelihood of those voxels with an abnormal intensity given the reference population. Because brain tumors show an abnormal intensity level compared to a healthy subject in the same location, we assume that tumors correspond with brain areas with low likelihood, making feasible brain tumor detection as outliers toward our coupled global/local intensity model. To extract the voxel outliers and to estimate the parameters of the different brain tissues in a robust way, we used the Trimmed Likelihood Estimator (TLE).

The TLE was proposed by [6] as a modification of the Maximum Likelihood Estimator in order to be robust to the presence of outliers. Using the TLE, the

complete log-likelihood (5) can be expressed as:

$$\log L_C(\boldsymbol{\psi}, \boldsymbol{\xi}) \;=\; \log \left(\prod_{i=1}^{h} f(Y_{\nu(i)}, Z_{\nu(i)} | \boldsymbol{\psi}, \boldsymbol{\xi}_i) \right)$$

with $\nu = \{\nu(1), ..., \nu(N)\}$ being the corresponding permutation of indices sorted by their probability $f(Y_{\nu(i)}, Z_{\nu(i)} | \boldsymbol{\psi}, \boldsymbol{\xi}_i)$, and h is the trimming parameter corresponding to the percentage of values included in the parameter estimation.

The trimming proportion h will define a set of outlier voxels $\mathcal{C}$ which among others will encompass the brain tumors. Because the set of outlier voxels $\mathcal{C}$ contained no only brain tumor but also false positives, a Graph-Cuts algorithm [2] was used to segment $\mathcal{C}$ into tumors or false positives.

3 Results

We validated our brain tumor segmentation algorithm using the clinical training data provided in the MICCAI 2012 brain tumor segmentation (BRATS) challenge.

3.1 Reference Population

We collected data from 15 volunteers on a 3T clinical MR scanner from GE Medical Systems (Waukesha, WI, USA) using an 8-channel receiver head coil and three different pulse sequences: a T1-weighted MPRAGE (Magnetization Prepared Rapid Acquisition Gradient Echo) sequence; a T2-weighted scan from an FSE (Fast Spin Echo) sequence; and a FLAIR scan, also run with an FSE sequence. We acquired the T1w sequence axially; the T2w and FLAIR sequences were sagitally acquired. All sequences were acquired with a matrix size of 256x256 and a field of view of 28 cm. Slice thickness was 1.3 mm for the T1w-MPRAGE sequence; 1 mm for the T2w-FSE sequence; and 2 mm for the FLAIR-FSE sequence. The MPRAGE parameters were TR 10/TE 6 ms with a flip angle of 8. For the FSE, the paramenters were TR 3000/TE 140 ms with an echo train length of 15.

After image acquisition, we aligned the T2w and FLAIR images to the T1w scan. Next, we re-oriented these MR images to an axial orientation. Last, a trained expert manually segmented the intra-cranial volume, CSF, GM and WM tissues [5].

To achieve accurate alignment between healthy volunteers and a gliobastoma patient we used a nonlinear registration algorithm proposed by [1], which although not intrinsic to our method, was selected because it is robust in the presence of WM lesions [8]. We note, however that other non-linear registration approaches are applicable with our technique [4].

Since the intensity levels of the subject of interest and the reference population need to be in a comparable range, we used a linear transformation to find a match between the median intensity of each modality (of each reference subject) and those found in the scans of the subject of interest.

3.2 MICCAI BRATS 2012 Training Data

MICCAI brain tumor segmentation challenge training data consists of multi-contrast MR scans of 30 glioma patients (both low-grade and high-grade, and both with and without resection) along with expert annotations for "active tumor" and "edema". For each patient, T1, T2, FLAIR, and post-Gadolinium T1 MR images are available. All volumes were linearly co-registered to the T1 contrast image, skull stripped, and interpolated to 1mm isotropic resolution.

Each gliobastoma patient was segmented using the post-Gadolinium T1, T2 and FLAIR MR scans. Algorithm parameters were fixed to neighborhood radius $\mathcal{R} = 1$ and trimming percentage $h = 10\%$. The estimated global-local outliers were classified as either tumor or edema using the graph-cuts algorithm described in Section 2.3. An example of the resulting segmentation can be found in Figure 1.

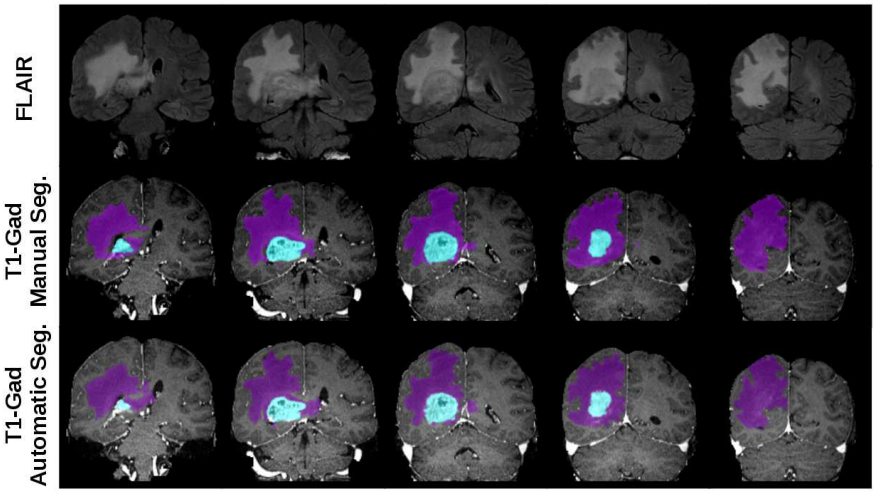


Fig. 1. Example of gliobastoma segmentation achieved by our combined global-local model. Top, FLAIR coronal slices of High grade gliobastoma Case0003; Middle, tumor manual segmentation provided by the challenge organizers; Bottom, automatic tumor segmentation achieved by our combined global/local model.

The resulting gliobastoma segmentations were uploaded to the MICCAI BRATS 2012 challenge[1] for evaluation. The online evaluation tool computed the dice score, jaccard score, specificity, sensitivity, average closest distance and hausdorff distance for both edema and active tumor. Results achieved by our combined global-local model can be found in Table 1.

[1] https://vsd.bfh.ch/WebSite/BRATS2012

	Dice	Jaccard	Specificity	Sensitivity	Avg. Dist.	Hausdorff Dist.
Tumor	0.43±0.29	0.32±0.25	1.00±0.01	0.47±0.32	6.10±12.24	26.82±18.26
Edema	0.55±0.28	0.43±0.27	1.00±0.00	0.58±0.31	6.83±12.83	32.49±22.58

Table 1. MICCAI BRATS 2012 training data results $\mathcal{R} = 0$ and $h = 20\%$.

4 Discussion

Preliminary results achieved on the clinical training data, show how our combined model allows a highly specific tumor segmentation. However, in some subjects, the heterogeneous intensity distribution of active tumors across all MRI modalities provided in the database, our algorithms fails to classify accurately outlier voxels into active tumor and edema.

Overall, we proposed a novel fully automatic algorithm for segmenting MR images of gliobastoma patients. This algorithm uses an approach that combines global and local tissue intensity models derived from an aligned set of healthy reference subjects. Specifically, tumors are segmented as outliers within the combined local/global intensity Gaussian Mixture Model (GMM). A graph-cuts algorithm is used to classify the estimated coupled model outlier voxels into active tumor and edema.

References

1. Commowick, O.: Creation et Utilisation d'Atlas Anatomiques Numeriques pour la Radiotherapie (Design and Use of Anatomical Atlases for Radiotherapy). Ph.D. thesis, Universite de Nice-Sophia Antipolis (2007)
2. Greig, D., Porteous, B.: Exact maximum a posteriori estimation for binary images. of the Royal Statistical Society. Series 51, 271–279 (1989)
3. Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.a., Kikinis, R.: Automated segmentation of MR images of brain tumors. Radiology 218(2), 586–91 (Feb 2001)
4. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46(3), 786–802 (Jul 2009)
5. Makris, N., Meyer, J.W., Bates, J.F., Yeterian, E.H., Kennedy, D.N., Caviness, V.S.: MRI-Based topographic parcellation of human cerebral white matter and nuclei II. Rationale and applications with systematics of cerebral connectivity. NeuroImage 9(1), 18–45 (Jan 1999)
6. Neykov, N., Filzmoser, P., Dimova, R., Neytchev, P.: Robust fitting of mixtures using the trimmed likelihood estimator. Computational Statistics & Data Analysis 52(1), 299–308 (Sep 2007)
7. Prastawa, M., Bullitt, E., Ho, S.: A brain tumor segmentation framework based on outlier detection. Medical Image Analysis 8(3), 275–283 (2004)

8. Suarez, R.O., Commowick, O., Prabhu, S.P., Warfield, S.K.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. NeuroImage 59(4), 3690–3700 (Nov 2011)
9. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. Medical image analysis 4(1), 43–55 (Mar 2000)
10. Wells, W.M., Grimson, W.L., Kikinis, R., Jolesz, F.a.: Adaptive segmentation of MRI data. IEEE transactions on medical imaging 15(4), 429–42 (Jan 1996)

Segmenting Glioma in Multi-Modal Images using a Generative Model for Brain Lesion Segmentation

Bjoern H. Menze[1,2], Koen Van Leemput[3], Danial Lashkari[4]
Marc-André Weber[5], Nicholas Ayache[2], and Polina Golland[4]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Asclepios Research Project, INRIA Sophia-Antipolis, France
[3] Radiology, Massachusetts General Hospital, Harvard Medical School, USA
[4] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA
[5] Diagnostic Radiology, Heidelberg University Hospital, Germany

1 Introduction

In this paper, we evaluate a fully automated method for channel-specific tumor segmentation in multi-dimensional images proposed by us in [1]. The method represents a tumor appearance model for multi-dimensional sequences that provides channel-specific segmentation of the tumor. Its generative model shares information about the spatial location of the lesion among channels while making full use of the highly specific multi-modal signal of the healthy tissue classes for segmenting normal tissues in the brain. In addition to tissue types, the model includes a latent variable for each voxel encoding the probability of observing tumor at that voxel, based on the ideas from [2, 3]. This extends the general "EM segmention" algorithm for situations when specific spatial structures cannot be described sufficiently through population priors. Different from [1], we now use a simplified EM algorithm for estimating the tissue state that also allows us to enforce additional constraints for segmenting lesions that are either hyper- or hypo-intense with respect to other tissues visible in the same image.

2 Generative Tumor Model

We use a generative modeling approach, in which we first build an explicit statistical model of image formation and subsequently use this model to derive a fully automatic segmentation algorithm. We follow the description of the model from [1].

Normal state We model the *normal state* of the healthy brain using a spatially varying probabilistic prior $\boldsymbol{\pi}_k$ for each of the K tissue classes. This population prior (atlas) is estimated from prior examples and is assumed to be known. At each voxel i, the atlas defines a multinomial distribution for the tissue label k_i:

$$p(k_i = k) = \pi_{ki}. \tag{1}$$

The normal state k_i is shared among all C channels at voxel i. In our experiments we assume $K = 3$, representing gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF).

Tumor state We model the *tumor state* using a spatially varying "latent" probabilistic atlas $\boldsymbol{\alpha}$, similar to [2], that is specific to the given image data set or patient. At each voxel i, this atlas provides a scalar parameter α_i that defines the probability of observing tumor at that voxel. Parameter α_i is unknown and is estimated as part of the segmentation process. We define a latent tumor state $t_i^c \in \{0, 1\}$ that indicates the presence of tumor in channel c at voxel i and model it as a Bernoulli random variable with parameter α_i. We form a binary tumor state vector $\boldsymbol{t}_i = [t_i^1, \ldots, t_i^C]^T$ indicating the tumor presence for all c observations at voxel i, with probability

$$p(\boldsymbol{t}_i; \alpha_i) = \prod_c p(t_i^c; \alpha_i) = \prod_c \alpha_i^{t_i^c} \cdot (1 - \alpha_i)^{1-t_i^c}. \tag{2}$$

Observation model The *image observations* y_i^c are generated by Gaussian intensity distributions for each of the K tissue classes and the C channels, with mean μ_k^c and variance v_k^c, respectively. In tumor tissue (i.e., if $t_i^c = 1$) the normal observations are replaced by intensities from another set of channel-specific Gaussian distributions with mean μ_T^c and variance v_T^c, representing the tumor class. Letting $\boldsymbol{\theta}$ denote the set of all mean and variance parameters, and $\boldsymbol{y}_i = [y_i^1, \ldots, y_i^C]^T$ denote the vector of the intensity observations at voxel i, we define the data likelihood:

$$\begin{aligned} p(\boldsymbol{y}_i | \boldsymbol{t}_i, k_i; \boldsymbol{\theta}) &= \prod_c p(y_i^c | t_i^c, k_i; \boldsymbol{\theta}) \\ &= \prod_c \left[\mathcal{N}(y_i^c \ ; \ \mu_{k_i}^c, v_{k_i}^c)^{1-t_i^c} \cdot \mathcal{N}(y_i^c \ ; \ \mu_T^c, v_T^c)^{t_i^c} \right], \end{aligned} \tag{3}$$

where $\mathcal{N}(\cdot \ ; \ \mu, v)$ is the Gaussian distribution with mean μ and variance v.

Joint model Finally, the *joint probability* of the atlas, the latent tissue class and the observed variables

$$p(\boldsymbol{y}_i, \boldsymbol{t}_i, k_i; \boldsymbol{\theta}, \alpha_i) = p(\boldsymbol{y}_i | \boldsymbol{t}_i, k_i; \boldsymbol{\theta}) \cdot p(\boldsymbol{t}_i; \alpha_i) \cdot p(k_i) \tag{4}$$

is the product of the components defined in Eqs. (1-3).

3 Inference

Maximum Likelihood Parameter Estimation We seek Maximum Likelihood estimates of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$:

$$\langle \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}} \rangle = \arg\max_{\langle \boldsymbol{\theta}, \boldsymbol{\alpha} \rangle} \ p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \arg\max_{\langle \boldsymbol{\theta}, \boldsymbol{\alpha} \rangle} \ \prod_{i=1}^N p(\boldsymbol{y}_i; \boldsymbol{\theta}, \boldsymbol{\alpha}),$$

where N is the number of voxels in the volume and

$$p(\boldsymbol{y}_i;\boldsymbol{\theta},\boldsymbol{\alpha}) = \sum_{\boldsymbol{t}_i}\sum_{k_i} p(\boldsymbol{y}_i,\boldsymbol{t}_i,k_i;\boldsymbol{\theta},\boldsymbol{\alpha}) = \sum_{\boldsymbol{s}_i} p(\boldsymbol{y}_i,\boldsymbol{s}_i;\boldsymbol{\theta},\boldsymbol{\alpha}).$$

For summing over values of $\boldsymbol{t}_i$ and k_i in Eq. (4), we follow the same approach as in [1], but – rather than summing over the two parameters individually – we now sum over tissue state vector $\boldsymbol{s}_i$ that is obtained by expanding $\boldsymbol{t}_i$ and k_i into one state vector. This state vector $\boldsymbol{s}$ indicates tumor $s_i^c = T$ in all channels with $t_i^c = 1$, and normal tissue $s_i^c = k_i$ for all other channels. As an example, with $\boldsymbol{t}_i = [0,0,1,1]$ and $k_i = WM$ indicating tumor in channels 3 and 4 and a white matter image intensity in all healthy channels, we obtain a tissue state vector $\boldsymbol{s}_i = [WM, WM, T, T]$. Letting $\{\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\alpha}}\}$ denote the current parameter estimates, we can compute the posterior probability of any of the resulting $K * 2^C$ tissue state vectors $\boldsymbol{s}_i$ that may characterize the multimodal image intensity pattern observed in voxel i. Writing out the components of Eq. (4) we obtain for $\boldsymbol{s}_i$ (using the corresponding $\boldsymbol{t}_i(\boldsymbol{s}_i)$ and $k_i(\boldsymbol{s}_i)$ for simplicity of the notation):

$$p(\boldsymbol{s}_i|\boldsymbol{y}_i;\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\alpha}}) \;\propto\; \pi_{ki}\prod_c \left[\mathcal{N}(y_i^c\;;\;\tilde{\mu}_T^c,\tilde{v}_T^c)^{t_i^c}\alpha_i^{t_i^c}\cdot\mathcal{N}(y_i^c\;;\;\tilde{\mu}_k^c,\tilde{v}_k^c)^{1-t_i^c}(1-\alpha_i)^{1-t_i^c}\right] \tag{5}$$

As an additional constraint we only consider state vectors $\boldsymbol{s}_i$ that are biologically reasonable. We rule out, for example, state vectors that indicate at the same time CSF and tumor, or that correspond to observing tumor-specific changes in the T1gad channel (that is characteristic for the tumor core), while T2 and FLAIR do not show tumor specific changes in the same location. Choosing appropriate constraints reduces the total number of states $|\boldsymbol{S}|$ to be summed over in Eq. 5 significantly. Similar to the double EM-type minorization from [1] – that updated $\boldsymbol{t}_i$ and k_i iteratively – we arrive at closed-form update expressions that guarantee increasingly better estimates of the model parameters. The updates are intuitive: the latent tumor prior $\tilde{\alpha}_i$ is an average of the corresponding posterior estimated and the intensity parameters $\tilde{\mu}_k^c$ and $\tilde{v}_k^c$ are updated with the weighted statistics of the data for the healthy tissues and for the tumor class. We iterate the estimation of the parameters $\{\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\alpha}}\}$ and the computation of the posterior probabilities $p(\boldsymbol{s}_i|k_i,\boldsymbol{y}_i;\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\alpha}})$ until convergence that is typically reached after 10-15 updates. During the iterations we enforced that tumor voxels are hyper- or hypo-intense with respect to the current average μ_k^c of the white matter tissue (hypo-intense for T1, hyper-intense for T1gad, T2, FLAIR) by reducing the class weight for observations that do not comply with this constraint, similar to [4].

Spatial regularization Little spatial context is used in the basic model, as we assume the tumor state $\boldsymbol{t}_i$ in each voxel to be independent from the state of other voxels Eq. 3). It is only the atlas $\boldsymbol{\pi_k}$ that encourages smooth classification for the healthy tissue classes by imposing similar priors in neighboring voxels. To encourage a similar smoothness of the tumor labels, we extend the latent atlas α to include a Markov Random Field (MRF) prior, relaxing the MRF to

a mean-field approximation with an efficient approximate algorithm. Different from [1], we now use channel-specific regularization parameters β^c that are all in the range of .3 to 1.

Channel-specific tumor probabilities, and semantic interpretation Once we have an estimate of the model parameters $\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\}$, we can evaluate the probability that tumor is visible in channel c of voxel i by summing over all the configurations $\boldsymbol{t}_i$ for which $s_i^c = T$ or $t_i^c = 1$, respectively:

$$p(t_i^c = 1|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = p(s_i^c = T|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{s}_i} \delta(s_i^c, T)\ p(\boldsymbol{s}_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}), \qquad (6)$$

where δ is the Kroneker delta that is 1 for $s_i^c = T$ and 0 otherwise.

We then assign channel c of voxel i to tumor if $p(t_i^c = 1|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) > 0.5$. For a semantic interpretation that is in line with the class definitions of the segmentation challenge, we label voxels that show tumor specific changes in the T2 channel as *edema*, and voxels that show hyper-intense tumor specific changes as *tumor core*. All other image voxels are considered to be normal. Moreover, we remove any isolated regions that is smaller than .5 cm^3 in size.

4 Experiments

We evaluate our model on a the BRATS challenge data set of 25 patients with glioma. The data set comprises T_1, T_2, FLAIR-, and post-Gadolinium T_1 MR images, all images are skull stripped and co-registered using an affine registration. We segment the volume into the three healthy and an outlier class using a freely available implementation of the EM segmentation with bias correction [5, 4]. Outliers are defined as being more than three standard deviations away from the centroid of any of the three normal tissue classes. We apply our algorithm to the bias field corrected volumes and initialize intensity parameters with values estimated in the initial segmentation. We initialize the latent atlas $\boldsymbol{\alpha}$ to 0.7 time the local prior for the presence of gray or white matter.

Channels-specific segmentations returned by our algorithm are transformed to *Edema* and *Core* classes as detailed above. Exemplary segmentations are shown in Figure 1 and quantitative results from a leave-one-out cross-validation are shown in Table 1. Note that the definition of "core" labels differs between ground truth (where it also includes the T1 hypo-intense center of the tumor) and the algorithm tested (where it is only the T1gad hyper-intense area of the tumor) leading to misleading evaluation scores for low-grade cases and, to some degree, for high-grade core labels. Please note that another submission to the BRATS challenge [6] deals with further processing the probability maps presented here.

Acknowledgements. This work was supported by the German Academy of Sciences Leopoldina (Fellowship Programme LPDS 2009-10), the Academy of Finland (133611), INRIA CompuTumor, NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NINDS R01-NS051826, NIH R01-NS052585, NIH R01-EB006758, NIH R01-EB009051, NIH P41-RR014075 and the NSF CAREER Award 0642971.

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_HG0027	0.633	0.649	0.995	0.728	0.619	0.999
BRATS_HG0026	0.681	0.616	0.998	0.443	0.369	0.999
BRATS_HG0025	0.643	0.704	0.996	0.154	0.087	1
BRATS_HG0024	0.652	0.685	0.998	0.71	0.639	0.999
BRATS_HG0022	0.689	0.683	0.998	0.463	0.311	1
BRATS_HG0015	0.762	0.699	0.998	0.691	0.534	1
BRATS_HG0014	0.217	0.453	0.989	0.457	0.303	1
BRATS_HG0013	0.429	0.647	0.999	0.74	0.983	1
BRATS_HG0012	0.373	0.58	0.997	0.068	0.043	1
BRATS_HG0011	0.606	0.464	0.999	0.57	0.54	0.998
BRATS_HG0010	0.381	0.792	0.996	0.724	0.77	1
BRATS_HG0009	0.697	0.594	0.997	0.486	0.38	0.997
BRATS_HG0008	0.652	0.56	0.996	0.697	0.556	1
BRATS_HG0007	0.542	0.492	0.997	0.775	0.727	0.999
BRATS_HG0006	0.649	0.621	0.997	0.65	0.505	1
mean	0.574	0.616	0.997	0.557	0.491	0.999
median	0.643	0.621	0.997	0.65	0.534	1

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_LG0015	0.373	0.523	0.998	0	0	1
BRATS_LG0014	0.182	0.335	0.998	0	0	1
BRATS_LG0013	0.185	0.324	0.995	0.17	0.099	1
BRATS_LG0012	0.42	0.79	0.997	0.005	0.002	1
BRATS_LG0011	0.344	0.777	0.996	0.001	0	1
BRATS_LG0008	0.471	0.386	1	0.675	0.547	1
BRATS_LG0006	0.625	0.809	0.998	0.591	0.507	1
BRATS_LG0004	0.75	0.764	0.998	0.011	0.006	1
BRATS_LG0002	0.584	0.622	0.991	0.109	0.059	1
BRATS_LG0001	0.3	0.495	0.997	0.838	0.777	1
mean	0.423	0.582	0.997	0.24	0.2	1
median	0.396	0.572	0.998	0.06	0.0325	1

Table 1. Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`), indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyper-intense regions as class 2, irrespectively of the grading (low/high) of the disease.

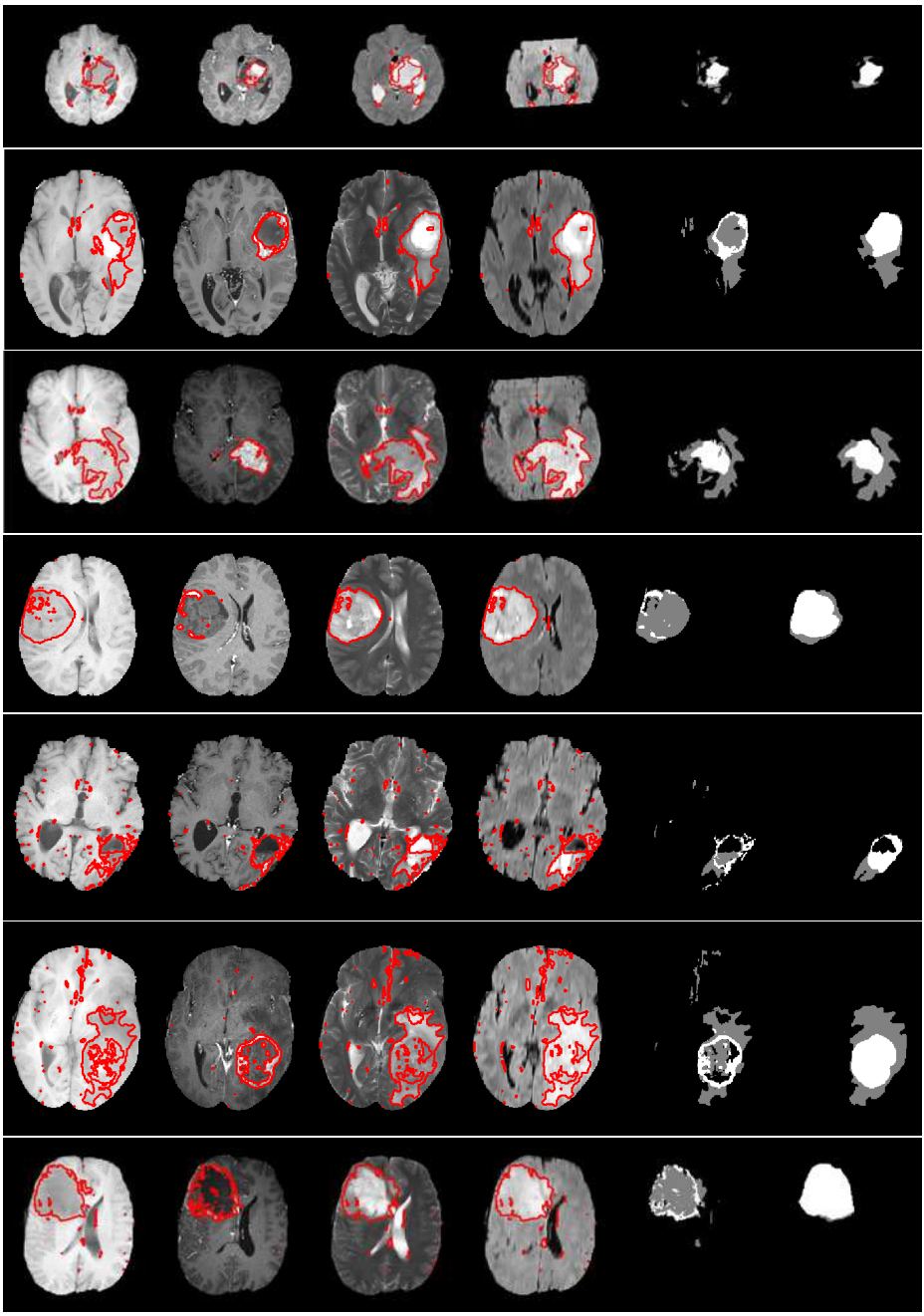


Fig. 1. Representative results of the channel-wise tumor segmentation. Shown are the MR images together with the most likely tumor areas (outlined red). The first four columns show T1, T1gad, T2 and FLAIR MRI, lesions are hyper-intense with respect to the gray matter for T1gad, T2 and FLAIR, they are hypo-intense in T1. The last two columns show the labels inferred from the channel-specific tumor segmentation (column 5), and the ground truth (column 6). The examples show that expert annotation may be disputable in some cases (e.g., rows 4, 5, 6).

References

1. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Proc MICCAI. (2010) 151–159
2. Riklin-Raviv, T., Menze, B.H., Van Leemput, K., Stieltjes, B., Weber, M.A., Ayache, N., Wells, W.M., Golland, P.: Joint segmentation via patient-specific latent anatomy model. In: Proc MICCAI-PMMIA (Workshop on Probabilistic Models for Medical Image Analysis). (2009) 244–255
3. Riklin-Raviv, T., Van Leemput, K., Menze, B.H., Wells, 3rd, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Med Image Anal **14** (2010) 654–665
4. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE T Med Imaging **20** (2001) 677–688
5. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE T Med Imaging **18** (1999) 885–896
6. Menze, B.H., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation. In: Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge). (2012) 8p

Segmenting Glioma in Multi-Modal Images using a Generative-Discriminative Model for Brain Lesion Segmentation

Bjoern H. Menze[1,2], Ezequiel Geremia[2], Nicholas Ayache[2], and Gabor Szekely[1]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Asclepios Research Project, INRIA Sophia-Antipolis, France

1 Introduction

In this paper, we evaluate a generative-discriminative approach for multi-modal tumor segmentation that builds – in its generative part – on a generative statistical model for tumor appearance in multi-dimensional images [1] by using a "latent" tumor class [2, 3], and – in its discriminative part – on a machine learning approach based on a random forest using long-range features that is capable of learning the local appearance of brain lesions in multi-dimensional images [4, 5]. The approach combines advantageous properties from both types of learning algorithms: First, it extracts tumor related image features in a robust fashion that is invariant to relative intensity changes by relying on a generative model encoding prior knowledge on expected physiology and pathophysiological changes. Second, it transforms image features extracted from the generative model – representing tumor probabilities in the different image channels – to an arbitrary image representation desired by the human interpreter through an efficient classification method that is capable of dealing with high-dimensional input data and that returns the desired class probabilities. In the following, we shortly describe the generative model from [1], and input features and additional regularization methods used similar to our earlier discriminative model from [4].

2 Generative Tumor Model

We use a generative modeling approach, in which we first build an explicit statistical model of image formation and subsequently use this model to derive a fully automatic segmentation algorithm. We follow closely our description of the method from [1]. The structure of the generative probabilistic model provides strong priors on the expected image patterns. For segmenting magnetic resonance (MR) images, it has the advantage that model parameters describing the observed image intensities serve as nuisance parameters. This makes it robust against tissue specific changes of the image intensity, and the algorithm does not depend on intensity calibration methods – often required for learning approaches that use image intensities as input – that may be prone to errors in the presence of lesions that vary in size and image intensity. Moreover, generative image appearance model can be combined with other parametric models

used, for example, for registration [6] or bias field correction [7], and even more complex image modification that can be "regressed out" in the same way.

Tumor appearance model We model the *normal state* of the healthy brain using a spatially varying probabilistic prior $\boldsymbol{\pi}_k$, a standard population atlas, for each of the $K = 3$ tissue classes that are visible from the given images (gray matter, white matter, and cerebrospinal fluid). The normal state k_i is shared among all C channels at voxel i. We model the *tumor state* using a spatially varying "latent" probabilistic atlas $\boldsymbol{\alpha}$, similar to [2]. At each voxel i, this atlas provides a scalar parameter α_i that defines the probability of observing a tumor transition at that voxel. Parameter α_i is unknown and is estimated as part of the segmentation process. We further assume that *image observations* y_i^c are generated by Gaussian intensity distributions for each of the K tissue classes and the C channels, with mean μ_k^c and variance v_k^c, respectively. If the image in channel c shows a transition from normal tissue to tumor in voxel i (i.e., if tissue state $s_i^c = T$), the normal observations are replaced by intensities from another set of channel-specific Gaussian distributions with mean μ_T^c and variance v_T^c, representing the tumor class.

Biological constraints on the estimated parameters We seek Maximum Likelihood estimates of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$ by estimating the tissue state vector $\boldsymbol{s}_i$ of every voxel i that indicates the type of tissue visible in the different image modalities. The vector has $\boldsymbol{s}_i^c = T$ in all channels that show tumor, and has value $s_i^c = k_i$ in all channels that appear normal. With $K = 3$ tissues types and $C = 4$ channels (for the given data), the cardinality of the state vector is $|\boldsymbol{s}| = K * 2^C = 3 * 2^4 = 48$. However, plausibility constraints on the expected tumor appearance in the different channels apply, for example ruling out tumor-induced intensity changes in T1gad unless the same location also shows tumor-induced changes in both T2 and FLAIR, and only gray and white matter being able to show tumor transitions, the number of possible tissue states reduces to $|\boldsymbol{s}| = 7$. We estimate the most likely state vector $\boldsymbol{s}_i$ in a standard expectation maximization procedure, similar to the "EM segmention" algorithm, with iterative updates of the parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}\}$ and the posterior probabilities $p(\boldsymbol{s}_i | k_i, \boldsymbol{y}_i; \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}})$. Updates can be performed using intuitive closed-form expressions: the latent tumor prior $\tilde{\alpha}_i$ is an average of the corresponding posterior estimated, and the intensity parameters $\tilde{\mu}^c$ and $\tilde{v}^c$ are updated with the weighted statistics of the data for the healthy tissues and for the tumor class. During the iteration we enforced that tumor voxels are hyper- or hypo-intense with respect to the current average gray value of the white matter tissue (hypo-intense for T1, hyper-intens for T1gad, T2, FLAIR) similar to [8]. Also we encourage smoothness of the tumor labels by extending the latent atlas α to include a Markov Random Field (MRF) prior, relaxing the MRF to a mean-field approximation with an efficient approximate algorithm. Different from [1], we now use channel-specific regularization parameters β that are all in the range of .3 to 1. Typically convergence is reached after 10-15 updates.

Channel-specific tumor and tissue probabilities Once we have an estimate of the model parameters $\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\}$, we can evaluate the probability that tumor is visible in channel c of voxel i by summing over all the configurations $\boldsymbol{S}_i$ for which $S_i^c = T$:

$$p(s_i^c = T|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{t}_i} \delta(s_i^c, T)\ p(\boldsymbol{t}_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}), \tag{1}$$

where δ is the Kroneker delta that is 1 for $s_i^c = T$ and 0 otherwise. The generative model returns C tumor appearance map $p(s_i^c = T|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$, one for each channel of the input volume. It also returns the probability maps for the K healthy tissues $p(k_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}})$, with global estimates for each voxel i that are valid for all C images.

3 Discriminative Lesion Model

The present generative model returns probability maps for the healthy tissues, and probability maps for the presences of characteristic hypo- or hyper-intens changes in each of the image volumes. While this provides highly specific information about different pathophysiological processes induced by the tumor, the analysis of the multimodal image sequence may still require to highlight specific structures of the lesion – such as edema, the location of the active or necrotic core of the tumor, "hot spots" of modified angiogenesis or metabolism – that cannot directly be associated with any of these basic parameter maps returned. As a consequence, we propose to use the probabilistic output of the generative model, together with few structural features that are derived from the same probabilistic maps, as input to a classifier modeling the posterior of the desired pixel classes. In this we follow the approach proposed by [4] that prove useful for identifying white matter lesion in multiple input volumes. The building blocks of this discriminative approach are the input features, the parametrization of the random forest classifier used, and the final post-processing routines.

Image features As input feature describing the image in voxel i we use the probabilities $p(k_i)$ for the $K = 3$ tissue classes ($\boldsymbol{x}_i^k$). We also use the tumor probability $p(s_i^c = T)$ for each channel $C = 4$ ($\boldsymbol{x}_i^c$), and the $C = 4$ image intensities after calibrating them with a global factor that has been estimated from gray and white matter tissue ($\boldsymbol{x}_i^{im}$). From these data we derive two types of features: the "long range features" that calculate differences of local image intensities for all three types of input features ($\boldsymbol{x}_i^k$, $\boldsymbol{x}_i^c$,$\boldsymbol{x}_i^{im}$), and a distance feature that calculates the geodesic distance of each voxel i to characteristic tumor areas.

The first type of features calculate the difference between the image intensity, or scalar of any other map, at voxel j that is located at $\boldsymbol{v}$ and the image intensity at another voxel k that is located at $\boldsymbol{v} + \boldsymbol{w}$ (with $\boldsymbol{v}$ here being 3D spatial coordinates). For every voxel j in our volume we calculate these differences $\boldsymbol{x}_j^{diff} = \boldsymbol{x}_j - \boldsymbol{x}_k$ for 20 different directions $\boldsymbol{w}$ around $\boldsymbol{v}$ with spatial offsets

between 3mm to 3cm. To reduce noise the subtracted value at $\boldsymbol{v}+\boldsymbol{w}$ is extracted after smoothing the image intensities locally around voxel k (using a Gaussian kernel with 3mm standard deviation).

The second type of features calculates the geodesic distance between the location $\boldsymbol{v}$ of voxel j to specific image feature that are of particular interest in the analysis. The path is constrained to areas that are most likely gray matter, white matter or tumor as predicted by the generative model. More specifically, we use the distance of $\boldsymbol{x}_j^{\delta tissue}$ of voxel j to the boundary of the the brain tissue (the interface of white and gray matter with CSF), and the distance $\boldsymbol{x}_j^{\delta edema}$ to the boundary of the T2 lesion representing the approximate location of the edema. The second distance $\boldsymbol{x}_j^{\delta edema}$ is calculated independently for voxels outside $\boldsymbol{x}_j^{\delta edema+}$ and inside $\boldsymbol{x}_j^{\delta edema-}$ the edema. In total we have 269 image features $\boldsymbol{x}$ for each voxel when concatenating the vectors of $\boldsymbol{x}^k$, $\boldsymbol{x}^c$, $\boldsymbol{x}^{im}$, $\boldsymbol{x}^{diff}$, and $\boldsymbol{x}^{\delta}$.

Classifier and spatial regularization We use the image features $\boldsymbol{x}$ defined above to model the probabilities $p(L;\boldsymbol{x})$ of class labels L in the BRATS data set, and the labels of the K normal tissue. For the normal classes (that are not available from the manual annotation of the challenge data set) we infer the maximum a posterior estimates of the generative model and use them as label during training. We choose random forests as our discriminative model as it uses labeled samples as input and returns class probabilities. Random forests learn many decision trees from bootstrapped samples of the training data, and at each split in the tree they only evaluate a random subspaces to find the best split. The split that separates samples of one class best against the others (with respect to Gini impurity) is chosen. Trees are grown until all nodes contain sample of a single class only. In prediction new samples are pushed down all trees of the ensemble and assigned, for each tree, to the class of the terminal node they end up in. Votes are averaged over all trees of the ensemble. The resulting normalized votes approximate the posterior for training samples that are independent of each other [9]. To minimize correlation in the training data, and also to speed up training, we draw no more 2000 samples from each of the $\approx 10^6$ voxels in each of the 25 data set. We train an ensemble with 300 randomized decision trees, and choose a subspace dimensionality of 10. We use the random forest implementation from Breiman and Cutler. To improve segmentation, we use a Markov Random Field (MRF) imposing a smoothness constraint on the class labels. We optimize the function imposing costs when assigning different labels in a 6 neighbourhood on the cross-validated predictions on the training data.

4 Experiments

We evaluate our model on the BRATS challenge data set of 25 patients with glioma. The data set comprises T_1, T_2, FLAIR-, and post-Gadolinium T_1 MR images, all images are skull stripped and co-registered. We segment the volume

into the three healthy and an outlier class using a freely available implementation of the EM segmentation with bias correction [7, 8]. Outliers are defined as being more than three standard deviations away from the centroid of any of the three normal tissue classes. We apply the generative model to the bias field corrected volumes and initialize intensity parameters with values estimated in the initial segmentation; we also use the the bias field and intensity corrected images as input for the discriminative model. More details about these data is given in another submission to the BRATS challenge that focuses on evaluating the generative model [10].

Exemplary segmentations that are returned from the present approach are shown in Figure 1 and quantitative results from a leave-one-out cross-validation are shown in Table 1. Note that the definition of "core" labels differs between ground truth (where it also includes the T1 hypo-intense center of the tumor) and the algorithm tested (where it is only the T1gad hyper-intense area of the tumor) which results in misleading evaluation scores for the "core" class in low-grade cases.

Acknowledgements. This work was supported by NCCR Co-Me of the Swiss National Science Foundation, and INRIA CompuTumor.

References

1. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: A generative model for brain tumor segmentation in multi-modal images. In: Proc MICCAI. (2010) 151–159
2. Riklin-Raviv, T., Menze, B.H., Van Leemput, K., Stieltjes, B., Weber, M.A., Ayache, N., Wells, W.M., Golland, P.: Joint segmentation via patient-specific latent anatomy model. In: Proc MICCAI-PMMIA (Workshop on Probabilistic Models for Medical Image Analysis). (2009) 244–255
3. Riklin-Raviv, T., Van Leemput, K., Menze, B.H., Wells, 3rd, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Med Image Anal **14** (2010) 654–665
4. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel MR images. In: Proc MICCAI. (2010)
5. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage **57** (2011) 378–90
6. Pohl, K.M., Warfield, S.K., Kikinis, R., Grimson, W.E.L., Wells, W.M.: Coupling statistical segmentation and -pca- shape modeling. In: Medical Image Computing and Computer-Assisted Intervention. Volume 3216/2004 of Lecture Notes in Computer Science., Rennes / St-Malo, France, Springer, Heidelberg (2004) 151 – 159
7. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE T Med Imaging **18** (1999) 885–896

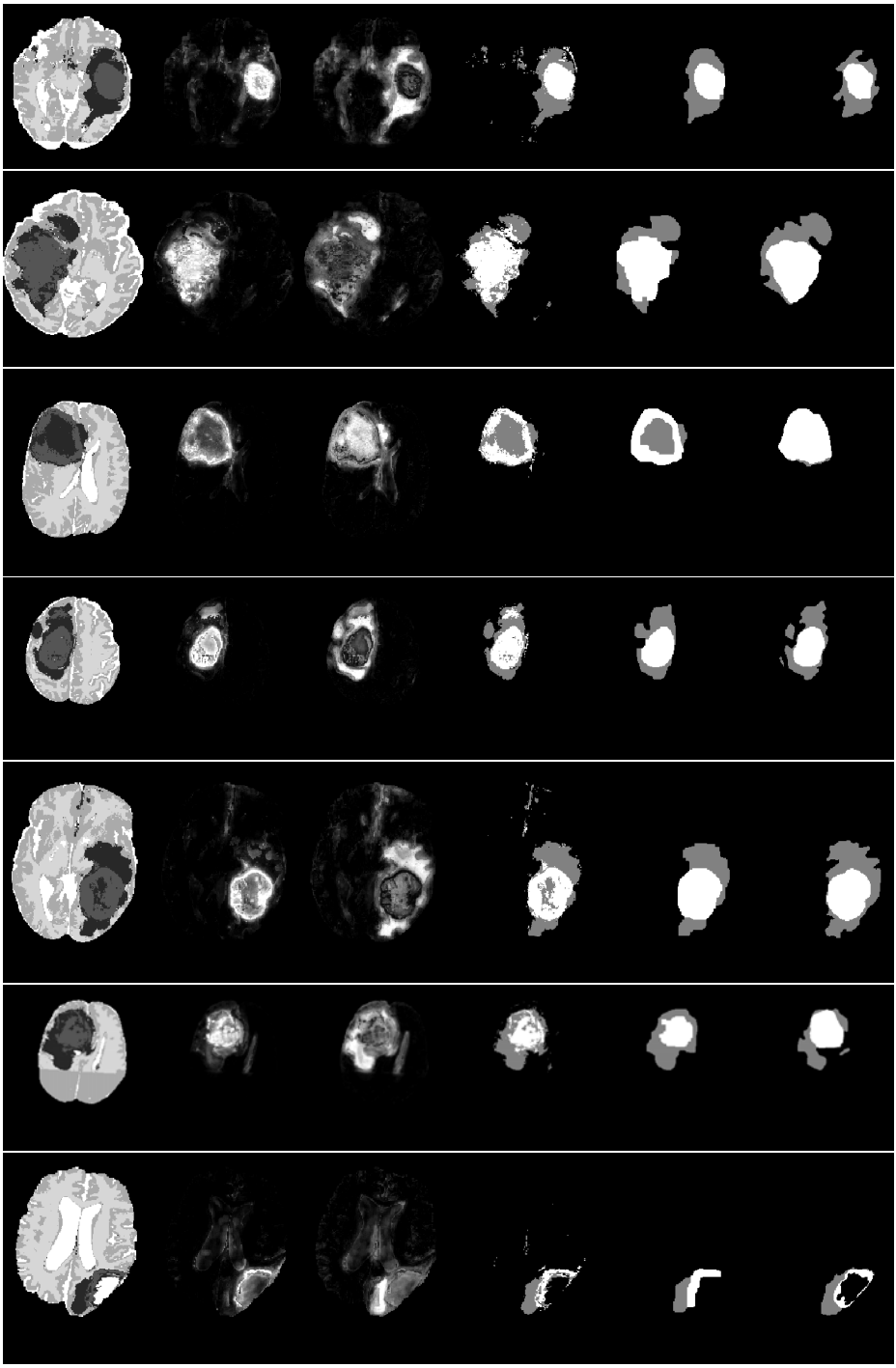


Fig. 1. Representative results of the tumor segmentation. Shown are the maximum-a-posteriori (MAP) estimates as obtained from the random forest for normal and tumor classe), the probabilities for core and edema (column 2,3), the MAP estimates of the two tumor classes before and after spatial smoothing (column 4,5), and the ground truth (column 6). The examples show that expert annotation may be disputable in some cases.

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_HG0027	0.735	0.823	0.995	0.822	0.898	0.997
BRATS_HG0026	0.738	0.758	0.997	0.412	0.401	0.998
BRATS_HG0025	0.641	0.934	0.992	0.06	0.031	1
BRATS_HG0024	0.7	0.834	0.997	0.896	0.982	0.999
BRATS_HG0022	0.779	0.806	0.998	0.821	0.729	1
BRATS_HG0015	0.8	0.82	0.996	0.894	0.879	0.999
BRATS_HG0014	0.327	0.476	0.994	0.761	0.696	0.998
BRATS_HG0013	0.7	0.661	1	0.887	0.985	1
BRATS_HG0012	0.629	0.704	0.999	0	0	1
BRATS_HG0011	0.808	0.763	0.998	0.9	0.889	0.999
BRATS_HG0010	0.664	0.788	0.999	0.836	0.879	1
BRATS_HG0009	0.833	0.822	0.997	0.749	0.604	1
BRATS_HG0008	0.784	0.679	0.999	0.917	0.979	0.998
BRATS_HG0007	0.644	0.508	0.999	0.838	0.942	0.999
BRATS_HG0006	0.7	0.795	0.994	0.793	0.731	0.999
mean	0.699	0.745	0.997	0.706	0.708	0.999
median	0.7	0.788	0.997	0.822	0.879	0.999

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_LG0015	0.402	0.751	0.997	0	0	1
BRATS_LG0014	0.405	0.605	0.999	0	0	1
BRATS_LG0013	0.29	0.492	0.996	0.164	0.089	1
BRATS_LG0012	0.424	0.94	0.996	0	0	1
BRATS_LG0011	0.3	0.908	0.994	0	0	1
BRATS_LG0008	0.419	0.53	0.999	0.521	0.397	1
BRATS_LG0006	0.767	0.992	0.998	0.788	0.74	1
BRATS_LG0004	0.813	0.898	0.998	0	0	1
BRATS_LG0002	0.652	0.767	0.989	0.017	0.009	1
BRATS_LG0001	0.454	0.552	0.999	0.843	0.915	0.999
mean	0.493	0.744	0.996	0.233	0.215	1
median	0.422	0.759	0.998	0.009	0.005	1

Table 1. Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`)indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyper-intense regions as class 2, irrespectively of the grading (low/high) of the disease.

8. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE T Med Imaging **20** (2001) 677–688
9. Breiman, L.: Random forests. Mach Learn J **45** (2001) 5–32
10. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.A., Ayache, N., Golland, P.: Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation. In: Proc MICCAI-BRATS (Multimodal Brain Tumor Segmentation Challenge). (2012) 7p

Multi-modal Brain Tumor Segmentation via Latent Atlases

Tammy Riklin Raviv[1,2,3], Koen Van Leemput[4,5], and Bjoern H. Menze[6,7]

[1] Psychiatry and NeuroImaging Laboratory, Harvard Medical School, Brigham and Women's Hospital, Boston, USA
[2] Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, USA
[3] The Broad Institute of MIT and Harvard, Cambridge, USA
[4] Harvard Medical School, Mass General Hospital, Boston, USA
[5] Technical University of Denmark, Lyngby, Denmark
[6] Computer Vision Laboratory, ETH Zurich, Switzerland
[7] Asclepios Research Project, INRIA Sophia-Antipolis, France

Abstract. In this work, a generative approach for patient-specific segmentation of brain tumors across different MR modalities is presented. It is based on the latent atlas approach presented in [7, 8]. The individual segmentation of each scan supports the segmentation of the ensemble by sharing common information. This common information, in the form of a spatial probability map of the tumor location is inferred concurrently with the evolution of the segmentations. The joint segmentation problem is solved via a statistically driven level-set framework. We illustrate the method on an example application of multimodal and longitudinal brain tumor segmentation, reporting promising segmentation results.

1 Introduction

Modeling patient-specific anatomy is essential in longitudinal studies and pathology detection. We present a generative approach for joint segmentation of MR scans of a specific subject, where the latent anatomy, in the form of spatial parameters is inferred concurrently with the segmentation. The work is based on the latent atlas approach presented in [7, 8]. While the methodology can be applied to a variety of applications, here we demonstrate our algorithm on a problem of multimodal segmentation of brain tumors. Patient-specific datasets acquired through different modalities at a particular time point are segmented simultaneously, yet individually, based on the specific parameters of their intensity distributions. The spatial parameters that are shared among the scans facilitate the segmentation of the group.

The method we propose is almost fully automatic. No prior knowledge or external information is required but a couple of mouse clicks at approximately the center and the boundary of a single tumor slice used to generate a sphere that initializes the segmentations. All model parameters, spatial and intensity, are inferred from the patient scans alone. The output of the algorithm consist of

individual segmentations for each modality. This is in contrast to many discriminative methods, e.g., [9], that use multimodal datasets for multivariate feature extraction, assuming spatial coherence of the tumor outlines in different image modalities. Here we relax this assumption and search for systematic, structural differences of the visible tumor volume acquired by different imaging protocols.

2 Problem definition and probabilistic model

This section summarizes the formulation of [6–8] for the joint segmentation of N aligned MR images . The input consists of N scans of a specific patient acquired via different imaging protocols. Our objective is to extract a brain tumor that may appear slightly differently across the images. Let $I_n\colon \Omega \to \mathbb{R}^+$, be a gray level image with V voxels, defined on $\Omega \subset \mathbb{R}^3$ and let $\Gamma_n\colon \Omega \to \{0,1\}$ be the unknown segmentation of the image I_n, $n = 1,\ldots,N$. We assume that each segmentation Γ_n is generated iid from a probability distribution $p(\Gamma;\ \theta_\Gamma)$ where θ_Γ is the set of the unknown spatial parameters. We also assume that Γ_n generates the observed image I_n, independently of all other image-segmentation pairs, with probability $p(I_n|\Gamma_n;\theta_{I,n})$ where $\theta_{I,n}$ are the parameters corresponding to image I_n. Since the images are acquired by different imaging protocols we assign a different set of intensity parameters to each of them. Our goal is to estimate the segmentations Γ. This, however, cannot be accomplished in a straightforward manner since the model parameters are also unknown. We therefore jointly optimize Γ and Θ:

$$\{\hat{\Theta},\hat{\Gamma}\} = \arg\max_{\{\Theta,\Gamma\}} \log p(I_1\ldots I_N,\Gamma_1\ldots\Gamma_N;\Theta) \tag{1}$$

$$= \arg\max_{\{\Theta,\Gamma\}} \sum_{n=1}^{N} \left[\log p(I_n|\ \Gamma_n;\theta_{I,n}) + \log p(\Gamma_n;\theta_\Gamma)\right]. \tag{2}$$

We alternate between estimating the maximum a posteriori (MAP) segmentations and updating the model parameters. For a given setting of the model parameters $\hat{\Theta}$, Eq. (2) implies that the segmentations can be estimated by solving N separate MAP problems:

$$\hat{\Gamma}_n = \arg\max_{\Gamma_n} \left[\log p(I_n|\ \Gamma_n;\theta_{I,n}) + \log p(\Gamma_n;\ \theta_\Gamma)\right]. \tag{3}$$

We then fix $\hat{\Gamma}$ and estimate the model parameters $\Theta = \{\theta_\Gamma,\theta_{I,1},\ldots\theta_{I,N}\}$ by solving two ML problems:

$$\hat{\theta}_{I,n} = \arg\max_{\theta_{I,n}} \log p(I_n;\ \Gamma_n,\theta_{I,n}), \tag{4}$$

$$\hat{\theta}_\Gamma = \arg\max_{\theta_\Gamma} \sum_{n=1}^{N} \log p(\Gamma_n;\ \theta_\Gamma). \tag{5}$$

3 Level-set framework

Now we draw the connection between the probabilistic model presented above and a level-set framework for segmentation. Let $\phi_n \colon \Omega \to \mathbb{R}$ be the level-set function associated with image I_n. The zero level $C_n = \{\mathbf{x} \in \Omega |\ \phi_n(\mathbf{x}) = 0\}$ defines the interface that partitions the image space of I_n into two disjoint regions ω and $\Omega \backslash \omega$. Similar to [4, 5] we define the level-set function ϕ_n using the log-odds formulation instead of the conventional signed distance function:

$$\phi_n(\mathbf{x}) \triangleq \epsilon \ \mathrm{logit}(p) = \epsilon \log \frac{p(\mathbf{x} \in w)}{1 - p(\mathbf{x} \in \omega)} = \epsilon \ \log \frac{p(\mathbf{x} \in \omega)}{p(\mathbf{x} \in \Omega \setminus \omega)}, \tag{6}$$

where $p(\mathbf{x} \in \omega)$ can be viewed as the probability that the voxel in location $\mathbf{x}$ belongs to the foreground region. The constant ϵ determines the scaling of the level-set function ϕ_n with respect to the ratio of the probabilities. The inverse of the logit function for $\epsilon = 1$ is the logistic function:

$$H_\epsilon(\phi_n) = \frac{1}{2}\left(1 + \tanh\left(\frac{\phi_n}{2\epsilon}\right)\right) = \frac{1}{1 + e^{-\phi_n/\epsilon}}. \tag{7}$$

Note, that $H_\epsilon(\phi_n)$ is similar, though not identical, to the regularized Heaviside function introduced by Chan and Vese [1]. We use this form of Heaviside function and its derivative with respect to ϕ in the proposed level-set formulation. To simplify the notation, we omit the subscript ϵ in the rest of the paper.

Cost functional for segmentation The joint estimation problem of the hidden variables Γ and the unknown model parameters $\{\theta_\Gamma, \theta_I^n\}$ can be solved as an energy minimization problem. As in [6–8], we establish the correspondence between the log probability and the level-set energy terms. We also look for the fuzzy labeling functions $H(\phi_n)$ rather than the hard segmentations Γ_n.

Let us consider first the prior probability $p(\Gamma_n;\ \theta_\Gamma)$ in Eq. (2) and its corresponding energy terms. Specifically, we construct an MRF prior for segmentations:

$$\begin{aligned} \log p(\Gamma_n; \theta_\Gamma) = &\sum_{v=1}^{V} \left[\Gamma_n^v \log(\theta_\Gamma^v) + (1 - \Gamma_n^v) \log(1 - \theta_\Gamma^v)\right] \\ &- \sum_{v=1}^{V} f(\Gamma_n^v, \Gamma_n^{\mathcal{N}(v)}) - \log Z(\theta_\Gamma), \end{aligned} \tag{8}$$

where $Z(\theta_\Gamma)$ is the partition function and $\mathcal{N}(v)$ is the set of the closest neighbors of voxel v. We define the spatial energy term E_S based on the singleton term in Eq. (8). Using the level-set formulation we obtain:

$$E_S(\phi_n, \Theta) = -\int_\Omega \left[\log \theta_\Gamma(\mathbf{x}) H(\phi_n(\mathbf{x})) + \log(1 - \theta_\Gamma(\mathbf{x}))\,(1 - H(\phi_n(\mathbf{x})))\right] d\mathbf{x}. \tag{9}$$

The dynamically evolving latent atlas θ_Γ is obtained by optimizing the sum of the energy terms that depend on θ_Γ :

$$\hat{\theta}_\Gamma(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N} \tilde{H}(\phi_n(\mathbf{x})). \tag{10}$$

The standard smoothness term used in level-set framework:

$$E_{\mathrm{LEN}}(\phi_n) = \int_\Omega |\nabla H(\phi_n(\mathbf{x}))| d\mathbf{x}, \tag{11}$$

can be obtained as an approximation of the pairwise term in Eq. (8).

The energy term $E_I(\phi_n, \theta_I^n)$ corresponds to the image likelihood term in Eq. (3):

$$\begin{aligned} E_I(\phi_n, \Theta) = -\int_\Omega \big[&\log p_{\mathrm{in}}(I_n; \theta_{I,n}^{\mathrm{in}}) H(\phi_n(\mathbf{x})) \\ &+ \log p_{\mathrm{out}}(I_n; \theta_{I,n}^{\mathrm{out}})\,(1 - H(\phi_n(\mathbf{x})))\big]\, d\mathbf{x}. \end{aligned} \tag{12}$$

We assume that the intensities of the structure of interest are drawn from a normal distribution such that the pair of scalars $\theta_{I,n}^{\mathrm{in}} = \{\mu_n^{\mathrm{in}}, \sigma_n^{\mathrm{in}}\}$ are the mean and standard deviation of the foreground intensities. We use a local-intensity model for the background intensity distributions in the spirit of [3], where $\theta_{I,n}^{\mathrm{out}}(\mathbf{x}) = \{\mu_n^{\mathrm{out}}(\mathbf{x}), \sigma_n^{\mathrm{out}}(\mathbf{x})\}$ are the local mean and standard deviation of a small neighbourhood of $\mathbf{x}$ that exclusively belongs to the background.

We construct the cost functional for $\phi_1 \dots \phi_N$ and the mode parameters by combing Eq. (12), (11) and (9):

$$E(\phi_1 \dots \phi_N, \Theta) = \gamma E_{\mathrm{LEN}} + \beta E_I + \alpha E_S \tag{13}$$

where α, β and γ are positive scalars.

Gradient descent and parameter estimation. We optimize Eq. (13) by a set of alternating steps. For fixed model parameters the update of each level-set function ϕ_n in each iteration is determined by the following gradient descent equation:

$$\begin{aligned} \frac{\partial \phi_n}{\partial t} = \delta(\phi_n) \Big\{ &\gamma \operatorname{div}\left(\frac{\nabla \phi_n}{|\nabla \phi_n|}\right) + \beta \left[\log p_{\mathrm{in}}(I_n(\mathbf{x}); \theta_{I,n}) - \log p_{\mathrm{out}}(I_n(\mathbf{x}); \theta_{I,n})\right] \\ &+ \alpha \left[\log \theta_\Gamma - \log(1 - \theta_\Gamma)\right] \Big\}, \end{aligned} \tag{14}$$

where $\delta(\phi_n)$ is the derivative of $H(\phi_n)$ with respect to ϕ_n. For fixed segmentations ϕ_n, the model parameters are recovered by differentiating the cost functional in Eq. (13) with respect to each parameter.

4 Experiments

We evaluate our model on the BRATS challenge data set of 25 patients with glioma. The data set comprises T_1, T_2, FLAIR-, and post-Gadolinium T_1 MR images, all images are skull stripped and co-registered. The tumor is initialized through a sphere of 1-3 cm diameter, that is placed in the center of the tumor.

Exemplary segmentations that are returned from the present approach are shown in Figure 1 and quantitative results from a leave-one-out cross-validation are shown in Table 3. Note that the definition of "core" labels differs between ground truth (where it also includes the T1 hypo-intense center of the tumor) and the algorithm tested (where it is only the T1gad hyper-intense area of the tumor) which results in misleading evaluation scores for the "core" class in low-grade cases.

5 Discussion and future directions

We presented a statistically driven level-set approach for joint segmentation of subject-specific MR scans. The latent patient anatomy, which is represented by a set of spatial parameters is inferred from the data simultaneously with the segmentation through an alternating minimization procedure. Segmentation of each of the channels or modalities is therefore supported by the common information shared by the group. Promising segmentation results on scans of 25 patients with Glioma were demonstrated.

Acknowledgements. This work was supported by NCCR Co-Me of the Swiss National Science Foundation, and INRIA CompuTumor.

References

1. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE TIP*, 10(2):266–277, 2001.
2. L. Dice. Measure of the amount of ecological association between species. *Ecology*, 26(3):297–302, 1945.
3. S. Lankton and A. Tannenbaum. Localizing Region-Based Active Contours. *IEEE Transactions on Image Processing*,17(11):2029–2039, 2008.
4. K.M. Pohl and R. Kikinis and W.M. Wells. Active Mean Fields: Solving the Mean Field Approximation in the Level Set Framework. *IPMI*, 4584:26–37, 2007.
5. Pohl *et al.* Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(6):465–477, 2007.
6. T. Riklin Raviv, K. Van Leemput, W.M. Wells, and P. Golland. Joint segmentation of image ensembles via latent atlases. In *Medical Image Computing and Computer-Assisted Intervention*, pages 272–280, 2009.
7. T. Riklin Raviv, B.H. Menze, K. Van Leemput, B. Stieltjes, N. Weber, M.A. Ayache, W.M. Wells, and P. Golland. Joint segmentation via patient-specific latent anatomy model. In *Proc. of MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, 2009.

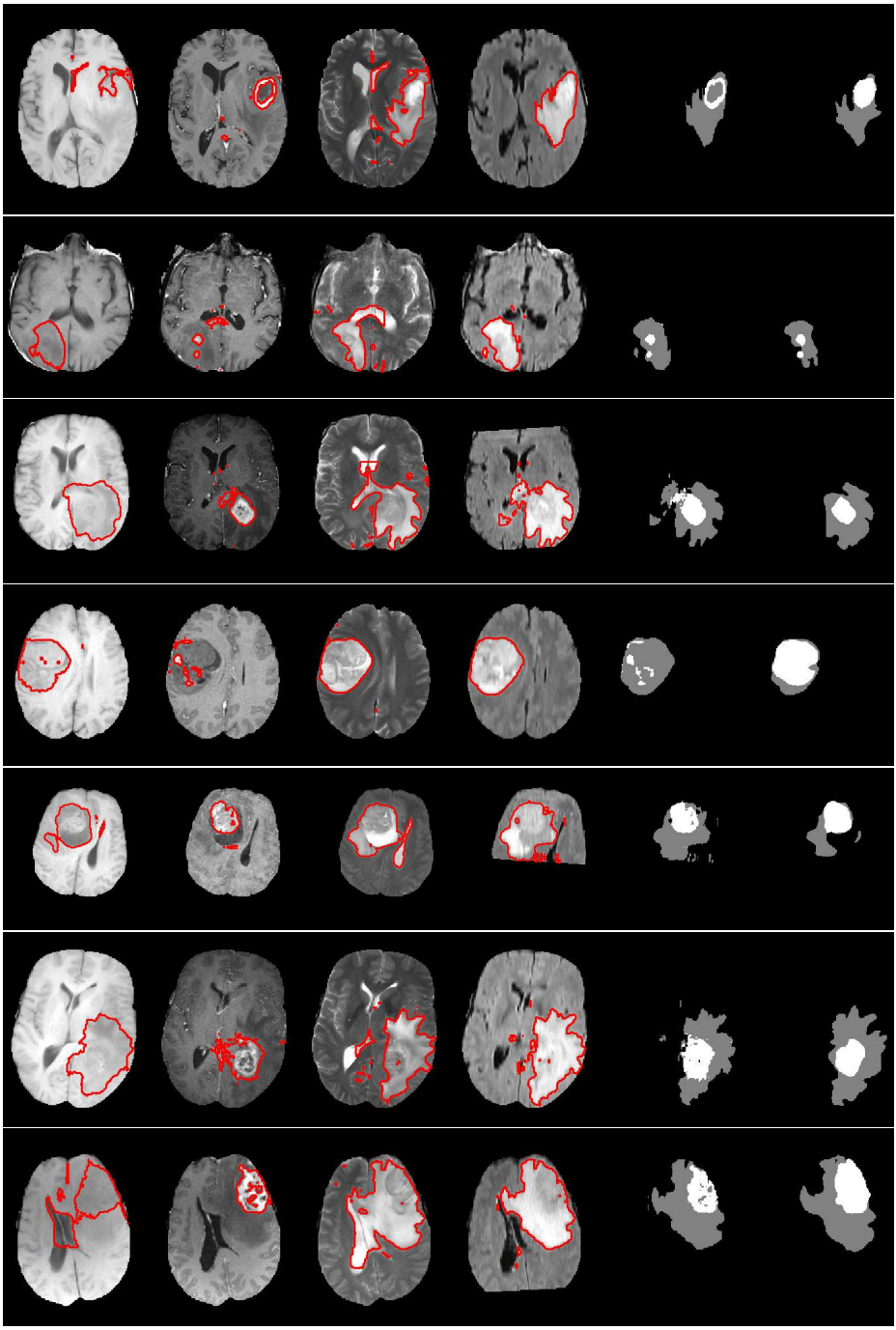


Fig. 1. Representative results of the tumor segmentation. Shown are the segmentations for the four different modalities (columns 1-4), the labels inferred from the channel-wise segmentation (column 5), and the ground truth (column 6). The examples show that expert annotation may be disputable in some cases.

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_HG0027	0.65	0.752	0.993	0.813	0.812	0.998
BRATS_HG0026	0.678	0.602	0.998	0.413	0.293	0.999
BRATS_HG0025	0.59	0.933	0.991	0.083	0.043	1
BRATS_HG0024	0.659	0.873	0.996	0.825	0.779	1
BRATS_HG0022	0.699	0.82	0.997	0.608	0.498	0.999
BRATS_HG0015	0.756	0.908	0.991	0.831	0.731	1
BRATS_HG0014	0.27	0.665	0.987	0.59	0.45	0.999
BRATS_HG0013	0.684	0.713	1	0.894	0.996	1
BRATS_HG0012	0.637	0.709	0.999	0.098	0.077	1
BRATS_HG0011	0.798	0.742	0.998	0.882	0.971	0.998
BRATS_HG0010	0.097	0.145	0.997	0.276	0.945	0.996
BRATS_HG0009	0.795	0.8	0.995	0.548	0.377	1
BRATS_HG0008	0.734	0.771	0.992	0.841	0.885	0.998
BRATS_HG0007	0.407	0.361	0.996	0.278	0.298	0.996
BRATS_HG0006	0.648	0.843	0.991	0.817	0.716	1
mean	0.607	0.709	0.995	0.586	0.591	0.999
median	0.659	0.752	0.996	0.608	0.716	0.999

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
BRATS_LG0015	0.37	0.712	0.997	0.116	0.066	1
BRATS_LG0014	0	0	1	0	0	1
BRATS_LG0013	0.326	0.631	0.995	0.452	0.293	1
BRATS_LG0012	0.563	0.721	0.999	0.822	0.762	1
BRATS_LG0011	0.262	0.958	0.993	0	0	1
BRATS_LG0008	0	0	1	0	0	1
BRATS_LG0006	0.556	0.985	0.996	0.73	0.832	1
BRATS_LG0004	0.513	0.492	0.997	0.022	0.072	0.997
BRATS_LG0002	0.636	0.734	0.989	0.242	0.178	0.997
BRATS_LG0001	0.345	0.648	0.997	0.843	0.774	1
mean	0.357	0.588	0.996	0.323	0.298	0.999
median	0.358	0.68	0.997	0.179	0.125	1

Table 1. Real data. Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`)indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyper-intense regions as class 2, irrespectively of the grading (low/high) of the disease.

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
SimBRATS_HG0025	0.023	0.105	0.996	0.631	0.593	0.998
SimBRATS_HG0024	0.217	0.993	0.999	0.968	0.948	1
SimBRATS_HG0023	0.007	0.078	0.997	0.454	0.419	0.997
SimBRATS_HG0022	0.689	0.682	0.996	0.002	0.001	0.999
SimBRATS_HG0021	0.312	0.214	0.998	0.02	0.011	1
SimBRATS_HG0020	0.138	0.127	0.996	0.315	0.249	0.998
SimBRATS_HG0019	0.45	0.349	0.998	0	0	1
SimBRATS_HG0018	0.01	0.047	0.996	0.579	0.552	0.997
SimBRATS_HG0017	0.147	0.179	0.998	0.499	0.348	1
SimBRATS_HG0016	0.033	0.091	0.995	0.681	0.667	0.997
SimBRATS_HG0015	0.36	0.289	0.998	0.234	0.186	0.998
SimBRATS_HG0014	0.362	0.3	0.998	0.451	0.406	0.998
SimBRATS_HG0013	0.623	0.564	0.996	0.004	0.002	0.999
SimBRATS_HG0012	0.44	0.36	0.999	0.035	0.022	1
SimBRATS_HG0011	0.453	0.518	0.997	0.366	0.235	1
SimBRATS_HG0010	0.528	0.867	0.999	0.974	0.978	1
SimBRATS_HG0009	0.762	0.788	1	0.958	0.977	1
SimBRATS_HG0008	0.381	0.352	0.996	0.454	0.386	0.999
SimBRATS_HG0007	0.635	0.689	0.995	0.559	0.75	0.997
SimBRATS_HG0006	0.011	0.037	0.998	0.373	0.274	0.999
SimBRATS_HG0005	0.63	0.615	0.996	0.019	0.015	0.999
SimBRATS_HG0004	0.33	0.311	0.996	0.485	0.475	0.998
SimBRATS_HG0003	0.63	0.593	0.998	0.317	0.314	0.999
SimBRATS_HG0002	0.405	0.819	0.999	0.924	0.875	1
SimBRATS_HG0001	0.592	0.856	0.999	0.971	0.982	1
mean	0.367	0.433	0.997	0.451	0.427	0.999
median	0.381	0.352	0.998	0.454	0.386	0.999

Table 2. Simulated data (high grade). Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`)indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyperintense regions as class 2, irrespectively of the grading (low/high) of the disease.

ID	Dice1	Sens1	Spec1	Dice2	Sens2	Spec2
SimBRATS_LG0025	0.042	0.528	0.993	0.05	0.026	1
SimBRATS_LG0024	0.404	0.997	0.993	0.137	0.074	1
SimBRATS_LG0023	0.662	0.74	0.997	0.008	0.004	1
SimBRATS_LG0022	0.404	0.551	0.997	0.007	0.004	1
SimBRATS_LG0021	0	0	1	0	0	1
SimBRATS_LG0020	0.367	0.702	0.997	0.023	0.012	1
SimBRATS_LG0019	0.378	0.367	0.998	0.015	0.008	1
SimBRATS_LG0018	0	0	1	0	0	1
SimBRATS_LG0017	0.632	0.678	0.998	0	0	1
SimBRATS_LG0016	0.699	0.858	0.993	0.08	0.045	1
SimBRATS_LG0015	0.46	0.578	0.997	0.01	0.005	1
SimBRATS_LG0014	0.02	0.416	0.996	0.025	0.013	1
SimBRATS_LG0013	0.402	0.447	0.997	0.014	0.007	1
SimBRATS_LG0012	0	0	1	0	0	1
SimBRATS_LG0011	0	0	1	0	0	1
SimBRATS_LG0010	0.078	0.694	0.991	0.21	0.117	1
SimBRATS_LG0009	0.394	0.507	0.997	0.035	0.018	1
SimBRATS_LG0008	0.051	0.998	0.994	0.13	0.07	1
SimBRATS_LG0007	0	0	1	0	0	1
SimBRATS_LG0006	0.395	0.857	0.995	0.054	0.028	1
SimBRATS_LG0005	0.483	0.994	0.993	0.089	0.047	1
SimBRATS_LG0004	0.317	0.316	0.998	0.002	0.001	1
SimBRATS_LG0003	0.359	0.546	0.997	0.007	0.004	1
SimBRATS_LG0002	0	0	1	0	0	1
SimBRATS_LG0001	0.489	0.39	0.998	0.001	0.001	1
mean	0.281	0.487	0.997	0.0359	0.0194	1
median	0.367	0.528	0.997	0.01	0.005	1

Table 3. Simulated data (low grade). Performance measures as returned by the online challenge tool (`challenge.kitware.com/midas/`)indicating Dice score, sensitivity and specificity (top: high-grade cases; bottom: low-grade cases). Class "1", with results shown in the left column, refers to the "edema" labels. Class "2", with results shown in the right column, refers to the "tumor core" labels (for both low and high grade cases). Note that this definition differs somewhat from the labels returned by the algorithm that only indicates T1gad hyperintense regions as class 2, irrespectively of the grading (low/high) of the disease.

8. T. Riklin Raviv, K. Van Leemput, B.H. Menze, W.M. Wells, and P. Golland. Segmentation of image ensembles via latent atlases. *Medical Image Analysis*, 14(5):654–665, 2010.
9. Van Leemput *et al.* Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE TMI*, 20:677–688, 2001.