NCI-MICCAI Challenge on

Multimodal Brain Tumor Segmentation

Proceedings of
NCI-MICCAI BRATS 2013
September 22nd, Nagoya, Japan

.

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 16th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2013) on September 22nd, 2013 in Nagoya, Japan. This event succeeds the MICCAI-BRATS 2012 challenge that was held in conjunction with MICCAI 2012.

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, adding another 20 multimodal image volume from high and low grade glioma patients to the BRATS 2012 data set. All images – in both the publicly distributed training data set, and the blinded test data set- are annotated through clinical experts who annotated four different types of tumor substructurs (edema, enhancing core, non-enhancing core, necrotic core).

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 10 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.

Bjoern Menze, ETH Zurich
Mauricio Reyes, University of Bern
Andras Jakab, Debrecen University
Elizabeth Gerstner, Harvard Medical School
Justin Kirby, SAIC-Frederick
Jayashree Kalpathy-Cramer, Harvard Medical School
Keyvan Farahani, National Cancer Institute, NIH

August 2013

Challenge Program

Organizers

Bjoern Menze, ETH Zurich
Mauricio Reyes, University of Bern
Andras Jakab, Debrecen University
Elizabeth Gerstner, Harvard Medical School
Justin Kirby, SAIC-Frederick
Jayashree Kalpathy-Cramer, Harvard Medical School
Keyvan Farahani, National Cancer Institute, NIH

Contributions

A Grouping Artificial Immune Network for Segmentation of Tumor Images

Patricia Buendia[1], Thomas Taylor[1], Michael Ryan[1], Nigel John[2]

[1]INFOTECH Soft, Inc., 1201 Brickell Avenue, Suite 220, Miami, FL 33131, USA
[2]University of Miami, Coral Gables, FL 33124, USA

{paty,thomas,mryan}@infotechsoft.com

Abstract. GAIN$^+$ is an enhanced version of the original Grouping Artificial Immune Network that was developed for fully automated MRI brain segmentation. The model captures the main concepts by which the immune system recognizes pathogens and models the process in a numerical form. GAIN$^+$ was adapted to support a variable number of input patterns for training and segmentation of tumors in MRI brain images and adapted to train on multiple images. The model was demonstrated to operate with multi-spectral MR data with an increase in accuracy compared to the single spectrum case. Using the BRATS High Grade 2013 dataset with the 2012 tissue labels for Edema and Tumor, the model's Dice scores were compared to published results and proved to be as accurate as the best methods. Using the 4 labels from the BRATS 2013 data sets, a Dice overlap of 73% for the complete tumor region and 64% for the enhancing tumor region were obtained for the high grade BRATS images when applying pre- and post-processing. This was attained with speed optimizations allowing segmentation at 21s per case with post-processing of all 4 tissues.

1 Introduction

GAIN is a biologically inspired artificial immune model developed for the segmentation of MRI brain images. The model captures the main concepts by which the immune system recognizes pathogens and models the process in a numerical form. A performance enhancement using bit grouping was added to improve training speed and segmentation accuracy.

The biological immune system is able to differentiate between pathogens and normal body cells by classifying the proteins presented on the surface of cells as being self proteins, which belong to the body or as non-self proteins. The brain segmentation GAIN approach is inspired by this biological process of recognition. The GAIN process leads to a generalized brain segmentation and is not tailored towards finding only one targeted tissue or disease. The model presented is best described as a segmentation and labeling model that results in segments to which tissue classes are assigned.

Artificial Immune Systems as described in [1] were good primary candidates for their binary nature; however the number of generated detectors that guarantee high

adfa, p. 1, 2011.

accuracy were sometimes larger than the training set. Modeling recognition through energies and reactions as proposed in [2] provided inspiration for parts of the described model. It had the disadvantage of computing those energies from a random process that did not relate directly to the training set, and placed a limit on the molecule size. The GAIN model solves those two specific problems by using a single comprehensively defined detector and relating the different model parameters directly to the training set using an information theory approach.

2 Method

GAIN$^+$ is an enhanced version of the original Grouping Artificial Immune Network, GAIN program [3] supporting a variable number of input patterns for training and segmentation of MRI brain images. The new input patterns include voxel intensities from 2D or 3D blocks or shapes of varying sizes customized to each MRI sequence (T1, T2, FLAIR, etc) individually, and include feature and textural patterns such as mean and variance of selected block sizes, slice or radial distance, co-occurrence matrices, among others. GAIN$^+$ has been stripped down of platform dependant libraries and now runs on different platforms and offers a selection of new features, such as different types of initializations, training on multiple images, and different post-processing options.

Due to the representation of the tissues as multi-bit values, it can be shown that not every bit carries the same entropy. That is, each bit does not contribute equally to the final interpretation of the data. The GAIN algorithm makes use of this fact to increase the speed of its operation. Bits are grouped into groups of size s bits. The current GAIN$^+$ implementation uses groups of size s=2 bits. By forming groups of size s, G groups are developed. Each tissue is represented by a T-Cell that recognizes the tissue in an input pattern. Thus a general form for the GAIN output per T-Cell is formulated as follows:

$$R_x(P) = \sum_{g=1}^{G} \chi\big(x, g, P(g)\big)$$

In the above equation, P(g) represents the value of group g. The parameter x denotes the weights of function X, and are specific to the T-Cell x. R_x represents the affinity of the T-Cell for the input pattern P. The pattern will be recognized by the T-Cell showing the highest affinity. The bit groups are not necessarily of neighboring bits, rather bit groups are defined by finding the pairs of bits with the highest mutual information with the output. Mutual information is defined as:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x)p(y)}{p(x,y)}$$

In GAIN$^+$, bit pairs x and y are grouped together and ordered in descending order based on their mutual information value I(x,y). During training the weights of function X for higher information bit groups will be defined before groups with lower information values.

Given X different classes, it is necessary to train X T-Cells of affinities R_x where x=0...X-1 towards the classes C_x. Training GAIN entails reducing the false negatives and false positives during the classification of each class. In the context of automated 3D MRI segmentation the Dice [4] and Tanimoto's similarity metrics are regularly used to compare the segmentation performance against the manual segmentation and are both implemented in GAIN$^+$. The goal of the training algorithm is to maximize the average of the chosen similarity metric computed for each class and is termed the average overlap D.

Training is carried by using iterated conditional modes implemented numerically. The modes that are searched for their maxima are the weights X. Since the equation to compute R_x is implemented as a sum of weights, the training algorithm utilizes that fact and pre-calculates the responses of each cell over the training set. When the training algorithm varies the weights instead of recalculating the whole again, it is only updated using the difference between the old weight and the new weight as follows:

$$\Delta R_x\left(\chi(x,g,v)\right) = \chi_n(x,g,v) - \chi_0(x,g,v)$$

where $X_n(x,g,v)$ is the new value of the weight and $X_0(x,g,v)$ is the old value of the weight. Thus the computation time is minimized during training. For each update iteration, the average overlap D is recalculated and finally the weight corresponding to the highest value of the average overlap is kept. It is well known that iterated conditional modes is a local maximization/minimization algorithm. However tests on real and simulated data sets and over different range of parameters, have shown it did not get trapped in local maximas.

2.1 GAIN$^+$ New Features

Expanded Feature Vector

GAIN$^+$ supports the use of a varying number of intensity values from each MRI spectrum to be included in the training and segmentation input vector. For example, knowing that FLAIR images return good results for Edema classification and T1C for Tumor, the user may choose to read 7 voxels of FLAIR, 7 of T1C, but only 1 each of T1, and T2. Each voxel intensity value is stored in a byte. In addition to the intensity values, a list of input features may be added to the input vector as described in [5], [6].

New Multi-Image Grouping Algorithm

The mutual information (MI) grouping approach was shown in [3] to produce good results with an increase in training speed. Yet, it was tested on only small input vectors and requires a different MI grouping for each training case. When including intensity values from all 4 spectra, T1, T2, T1C and FLAIR and using a larger number of labels, the MI grouping approach can be quite slow, taking an average of 70 minutes for 28 voxels and 2 labels. Thus, a new grouping approach was implemented based on the location of each voxel, and the significance of the input features giving higher priority to higher order bits and overall to voxels at closer distance to the center voxel. This grouping approach runs in just a few seconds and the same grouping file can be used for all cases.

Training on multiple images.
The original GAIN method was designed to train on a single case. As the mutual information grouping is designed to be used with 1 training case only, a different grouping algorithm was used to train on multiple cases (see New Multi-Image Grouping Algorithm above). GAIN$^+$ implemented two new options to train on multiple images:

1. A sequential training option that adds new cases to an already trained network and trains on the new cases by updating the existing weights.
2. A cross-segmentation training option, which segments a list of cases during training and computes the average of the average overlap of the selected cases, compares it to the previous score and either retains the new set of weights if the score improves or revert to the previous one otherwise.

Pre- and Post-processing options
The preprocessing pipeline was designed to remove noise and inhomogeneities due to MR scanner bias fields, and match each spectrum's intensity histogram to the volumes used for training. Several Post-processing options were added to the program, such as finding and extracting connected components, and performing dilation and erosion on those components.

3 Evaluation

We performed 20-fold cross validation on the real high-grade BRATS 2012 training set using the two-label ground truth (Edema and Tumor) provided by the challenge. Several of the input features described earlier were evaluated, yet they consistently improved Edema but not Tumor and we left them out of the final evaluations. Using 4x7=28 voxels, 7 from each spectrum, we achieved a mean accuracy (Dice score) of 62.3% for edema and 62.6% for tumor in the real cases when evaluated with the Dice score. While not directly comparable, the scores of BRATS 2012 competitors on the 2012 high grade test data were reported to be below the GAIN$^+$ Dice scores, with Zikic et al [7] obtaining 59.8% for edema and 47.6% for tumor, and Bauer et al. [8] 53.6% for edema and 51.2% for tumor.
GAIN$^+$ performance was evaluated with the four BRATS 2013 labels: (1) Necrosis, (2) Edema, (3) Non-Enhancing tumor, and (4) Enhancing Tumor. In this case, GAIN$^+$ was run with an input pattern of 7 FLAIR+ 7 T1C + 1 T1 + 1 T2 voxels and the new faster grouping algorithm achieving a speed of 32 minutes for training and 21 seconds for segmentation of a single case. The segmented images were uploaded to the BRATS 2013 Virtual Skeleton web site. The evaluation is done for 3 different tumor sub-compartments:

- Region 1: Complete tumor (labels 1+2+3+4 for patient data)
- Region 2: Tumor core (labels 1+3+4 for patient data)
- Region 3: Enhancing tumor (label 4 for patient data)

Table 1 below shows the GAIN$^+$ per region results reported by the Virtual Skeleton web site for the high grade BRATS 2013 images.

Table 1. Evaluation Results for BRATS 2013 Training Data

	Dice	**Jaccard**	**Specificity**	**Sensitivity**	**Kappa**
Mean Region 1	0.73412	0.598006	0.857127	0.659063	0.992449
Mean Region 2	0.607888	0.485163	0.677739	0.597957	0.995753
Mean Region 3	0.635123	0.507829	0.668212	0.65881	0.997692

4 Acknowledgement

This work was partially supported by National Cancer Institute grant #HHSN261201200032C

5 References

[1] S. A. Hofmeyr, An Immunological Model of Distributed Detection and Its Application to Computer Security. 1999.

[2] M. Oprea and S. Forrest, "How the immune system generates diversity: Pathogen space coverage with random and evolved antibody libraries," in In GECCO 99, Real-world Applications Track, 1999, pp. 1651–1656.

[3] Younis, A.A., Soliman, A.T., Kabuka, M.R., and John, N.M., "MS Lesions Detection in MRI Using Grouping Artificial Immune Networks," in Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, 2007. BIBE 2007, Boston, MA, 2007, pp. 1139 – 1146.

[4] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany, M. R. Kaus, S. J. Haker, W. M. Wells, F. A. Jolesz, and R. Kikinis, "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index," Acad. Radiol., vol. 11, no. 2, pp. 178–189, Feb. 2004.

[5] B. A. Abdullah, A. A. Younis, and N. M. John, "Multi-Sectional Views Textural Based SVM for MS Lesion Segmentation in Multi-Channels MRIs," Open Biomed. Eng. J., vol. 6, pp. 56–72, 2012.

[6] B. A. Abdullah, "Textural Based SVM for MS Lesion Segmentation in FLAIR MRIs," Open J. Med. Imaging, vol. 01, no. 02, pp. 26–42, 2011.

[7] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," Med. Image Comput. Comput.-Assist. Interv. Miccai Int. Conf. Med. Image Comput. Comput.-Assist. Interv., vol. 15, no. Pt 3, pp. 369–376, 2012.

[8] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," Med. Image Comput. Comput.-Assist. Interv. Miccai Int. Conf. Med. Image Comput. Comput.-Assist. Interv., vol. 14, no. Pt 3, pp. 354–361, 2011.

Patch-based Segmentation of Brain Tissues

Nicolas Cordier[1], Bjoern Menze[1,2], Hervé Delingette[1], and Nicholas Ayache[1]

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

Abstract. We describe our submission to the Brain Tumor Segmentation Challenge (BraTS) at MICCAI 2013. This segmentation approach is based on similarities between multi-channel patches. After patches are extracted from several MR channels for a test case, similar patches are found in training images for which label maps are known. These labels maps are then combined to result in a segmentation map for the test case. The labelling is performed, in a leave-one-out scheme, for each case of a publicly available training set, which consists of 30 real cases (20 high-grade gliomas, 10 low-grade gliomas) and 50 synthetic cases (25 high-grade gliomas, 25 low-grade gliomas). Promising results are shown on the training set, and we believe this algorithm would perform favourably well in comparison to the state of the art on a testing set.

1 Introduction

We describe our submission to the Brain Tumor Segmentation Challenge (BraTS) at MICCAI 2013. This segmentation approach is based on ideas similar to these developed for human brain labelling in [1]. A database of multi-channel patches is first built from a set of training cases, for which label maps are known. Then, given a test case, patches are extracted from several Magnetic Resonance (MR) channels, and similar multi-channel patches are retrieved in the patch database. Since each multi-channel intensity patch from the database is associated with a label patch, a combination of the labels can result in a segmentation map for the test case. First, we detail the pre-processing steps, mainly a global intensity alignment and tumour localization. Second, we recall the currently naive procedure to build the database of patches and retrieve similar patches. Third, we describe the label fusion step. Then, the labelling is performed, in a leave-one-out scheme, for each case of a publicly available training set, which consists of 30 real cases (20 high-grade gliomas, 10 low-grade gliomas) and 50 synthetic cases (25 high-grade gliomas, 25 low-grade gliomas). Finally, results and promising improvements to the current algorithm are discussed.

2 Method: Patch-based Segmentation

In this section, our patch-based segmentation algorithm is described in its most recent state. Current limitations could be bypassed with several promising improvements, which are still work-in-progress at the time of the submission of the article.

2.1 Pre-processing

In order to decrease the computation time, we first sub-sample , with a nearest-neighbour interpolation scheme, all the images to 2-mm isotropic resolution, as in [2]. In the end, to compare the resulting segmentation to the ground truth, we would perform an up-sampling to 1-mm isotropic resolution.

Moreover, we define a bounding box surrounding the tumour, and we crop the images. There are three main reasons to back this pre-processing step:

- computations are faster since there are less patches,
- patches are more relevant since we are only interested in segmenting the tumor and we focus on image parts containing a tumor,
- we avoid any problem related to image parts missing for some channels, since these happen outside these bounding boxes.

While these bounding boxes are known for the training cases, we do not know their localization for a test case. Given more time, we would intend to roughly localize the bounding box by applying a threshold to T_2 FLAIR images. In the meantime, we currently assume a bounding box around the tumor is known.

2.2 Global intensity alignment

Let D be the number of training cases, each of which consists of four MR channels and one label map. Before we build a multi-channel patch database, we would like to normalize each MR image in order to make all the D cases appear more similar to each other. This is a very important step since the similarity criterion, which drives the patch retrieval, is not invariant to affine intensity changes.

Since different Magnetic Resonance Imaging (MRI) scanners could be used for the acquisition of each case, a pre-processing step is mandatory to normalize the image intensities. Similarly to [2], we "align the mean intensities within each channel by a global multiplicative factor" K.

Let c be a channel, among T_1-weighted, T_2-weighted, FLAIR T_2-weighted, and contrast enhanced T_1-weighted images. Let x be a voxel position in the MR image I_c. Let $< I_c >$ be the mean intensity for channel c.

The global intensity alignment is such that:

$$\forall c,\ \forall x,\ I_c(x) \leftarrow I_c(x) \times \frac{K}{< I_c >}$$

where K is arbitrarily set to 700, which is a value close to the mean intensities. Currently, this step is performed with *cropped* images.

2.3 Building a database of patches and retrieving similar patches

For each case I of the D training cases, for each voxel x of I, we extract a 3x3x3 patch $P_x^{I_c}$ per channel c, and we concatenate these patches to get a multi-channel patch P_x^I.

Given a new case I^{test}, which consists of the four MR channels previously mentionned, we compute a multi-channel patch $P_x^{I^{\text{test}}}$ for each voxel x.

For each patch in I^{test}, for each of the D reference cases, we find the k-nearest-neighbours with respect to our similarity measure, which is here the sum of squared differences $d(\cdot,\ \cdot)$ between multi-channel patches. In the following, $k = 5$.

For each voxel x of the test case I^{test}, we have retrieved the k most similar patches for each reference case, which makes $D \times k$ similar multi-channel patches in total, for which we know the corresponding 3D label patches.

2.4 Label fusion

In order to define the label map S for the test case I, a label fusion method is defined.

Vote weight for each retrieved patch For each voxel x in I^{test}, each retrieved similar patch $(P_{y_{n,m}}^{I_{n,m}})_{n\in[|1,D|],m\in[|1,k|]}$ contributes to a voting scheme, with a vote weight $w(x,\ \cdot)$ which depends on the similarity measure $d(\cdot,\ \cdot)$:

$$w(x,\ P_y^I) = exp(-\frac{d(P_y^I,\ P_x^{I^{\text{test}}})}{\sigma^2})$$

where σ^2 is the maximum (over every voxel $x \in I^{\text{test}}$ and every training case $n \in [|1, D|]$) of the L_2 error made with first-neighbour patches only ($m = 1$).

Aggregation of the votes Given vote weights for each retrieved patch, there are two possibilities to aggregate the votes. First, if we consider a voxel x, the simplest method increments the votes for the label found at the center of each retrieved patch. Second, the method we use in this paper makes use of the labels found in the whole patch, i.e. at the center x and in the 3x3x3 neighbourhood of x of each label patch. Thus, we increment the votes for 27 voxels for each retrieved patch. This results in regularized vote maps.

Vote maps For each label, the aggregation of votes results in a *vote map*, which consists in a map of the aggregated vote weights. These vote maps can be seen as probability maps: they can be scaled by the total of vote increments so that the total of votes for each foreground voxel is 1, in which each vote is a non-negative value between 0 and 1. This rescaling does not affect the segmentation procedure though, as opposed to the rescaling described in the following.

Label selection In the Brain Tumor Segmentation Challenge, we are interested in the segmentation of three regions (complete tumor, tumor core, enhancing tumor) which show an interesting property: these regions are interlocked. Therefore, during label selection, we proceed with three consecutive steps:

1. distinguish between the background and the complete tumor,
2. inside the complete tumor, distinguish between the tumor core and what remains of the complete tumor (the edema),
3. inside the tumor core, distinguish between the enhancing tumor and what remains of the tumor core (the non-enhancing tumor and the necrotic core).

For each of these steps, we want to give more weights to labels which are less represented in the reference dataset, typically tumor classes. Indeed, less represented labels are less likely to be picked at the patch retrieval step, and in the case of tumor classes, the number of false negatives could be increased. Given a pool of two region labels to select from, we proceed as follows in order to achieve a penalization of the biggest region: for each region separately, we divide voxel-wise region votes by the number of voxels for which the region votes are non-zero. Therefore, we slightly favour labels for which only a few voxels have non-zero votes. Finally, for each voxel x, the label with the highest scaled vote is picked.

3 Evaluation

Our segmentation procedure is evaluated on 30 real cases (20 high-grade gliomas, 10 low-grade gliomas) and 50 synthetic cases (25 high-grade gliomas, 25 low-grade gliomas). The dataset is publicly available through the MICCAI 2013 Brain Tumor Segmentation challenge. We perform the labelling of each image of the training set, in a leave-one-out framework. The Dice score, Jaccard index, specificity and sensitivity are computed by the online evaluation tool provided by the organizers of the Brain Tumor Segmentation Challenge. Average scores are given in tables 1 to 4, per-patient scores are given in figures 1 and 2. The best segmentation maps for each of the four training datasets are shown in figures 3 to 10.

Dice score	Complete tumor	Tumor core	Enhancing tumor
mean	0.79	0.60	0.59
std-dev	0.17	0.26	0.25
median	0.87	0.72	0.69

Table 1. Evaluation summary for 20 real high-grade cases.

Dice score	Complete tumor	Tumor core	Enhancing tumor
mean	0.76	0.64	0.44
std-dev	0.18	0.21	0.40
median	0.80	0.71	0.45

Table 2. Evaluation summary for 10 real low-grade cases. Regarding the right-most column, 3 cases show enhancing tumor, and the Dice score is only computed for these.

Dice score	Complete tumor	Tumor core
mean	0.84	0.78
std-dev	0.08	0.12
median	0.85	0.83

Table 3. Evaluation summary for 25 synthetic high-grade cases.

Dice score	Complete tumor	Tumor core
mean	0.83	0.67
std-dev	0.04	0.19
median	0.83	0.74

Table 4. Evaluation summary for 25 synthetic low-grade cases.

4 Discussion

We have described a generic framework for brain tumor segmentation, which relies on similarities between multi-channel patches. The algorithm performs well on the training cases provided for the Brain Tumor Segmentation Challenge. We believe this algorithm would perform favourably well in comparison to the state of the art on a testing set. Promising improvements to the current algorithm are still in progress.

Regarding the pre-processing step, we do not expect the results to differ significantly with slightly different bounding boxes. Indeed, bounding boxes restrict the images in coronal and saggital views, but the axial view is not cropped at all in order to keep more examples of healthy tissues, as is shown in figures 3 to 10. Moreover, let us assume we make an error during the localization of the bounding box:

- if we find a bounding box containing the true bounding box, then no false negative would be added, and we would expect the number of additional false positives to be low since the specificity of the classifier is close to 1 in practice,
- if we find a bounding box included in the true bounding box, then this may result in less false positives and this would necessarily result in more false negatives since parts of the tumour are necessarily wrongly omitted by the classifier. This is the case for synthetic high-grade cases, due to a human error.

Regarding the global intensity alignment step, we believe results would not be fundamentally different if the full brain mask were used instead of a bounding box. In fact, a global intensity alignment applied to the full images would be the first pre-processing step in order to localize the tumor.

Regarding the retrieval of similar patches in the patch database, a cluster is used since the search is naive and exhaustive. This paper in its current state is mostly a proof-of-concept, and we expect in the foreseeable future to tremendously decrease the computation time by relying on approximate hash-based searches.

References

1. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. Medical Imaging, IEEE Transactions on **30**(10) (2011) 1852–1862
2. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., Price, S.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel mr. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012. Springer (2012) 369–376

Fig. 1. Per-patient evaluation for the two real BraTS data sets (Real-HG, Real-LG). We show the results for three regions: complete tumor, tumor core, and, for all high-grade cases and a few low-grade cases only, enhancing tumor. We report the following measures: Dice score, Specificity, and Sensitivity.

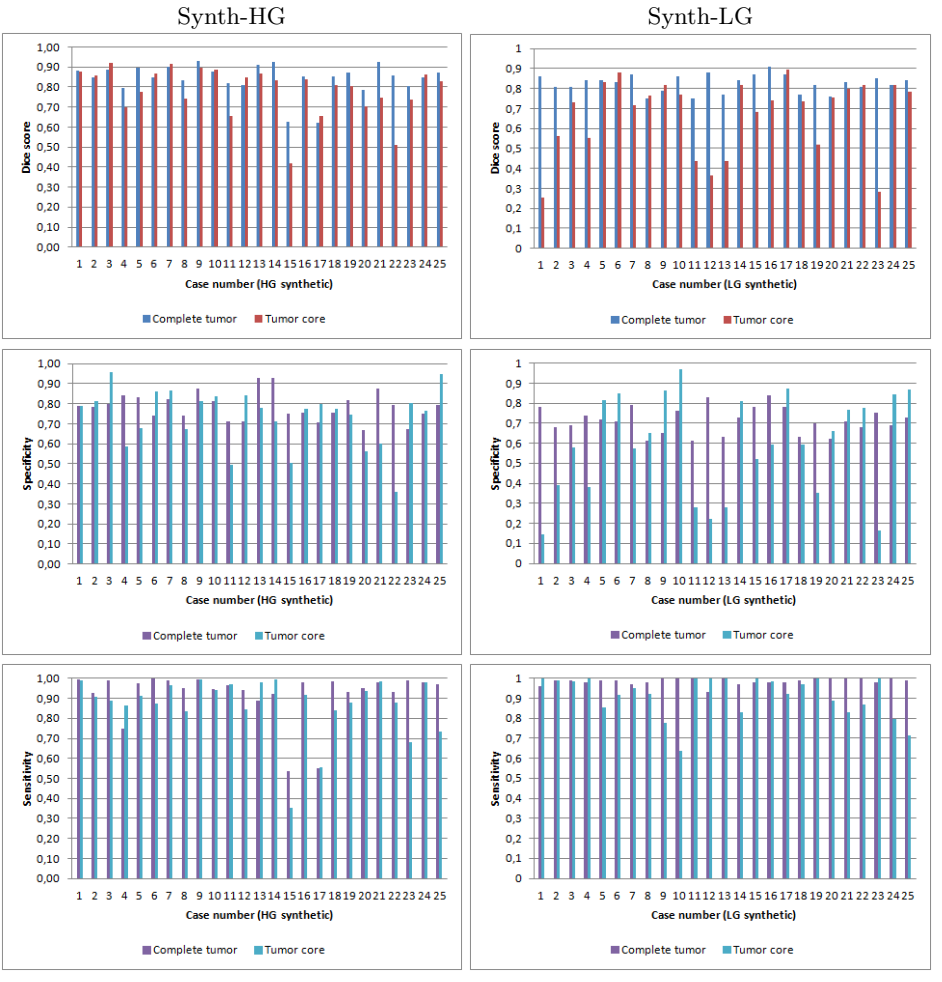


Fig. 2. Per-patient evaluation for the two synthetic BraTS data sets (Synth-HG, Synth-LG). We show the results for two regions: complete tumor, and tumor core. We report the following measures: Dice score, Specificity, and Sensitivity.

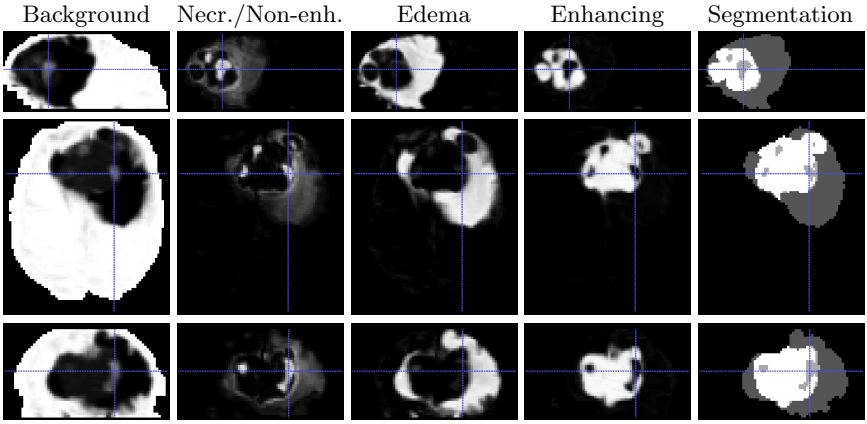


Fig. 3. Real high-grade case 15. From left to right: Vote maps for background, necrosis and non-enhancing tumor (merged), edema, enhancing tumor; Segmentation map. From top to bottom: saggital, axial, and coronal views. Images are cropped.

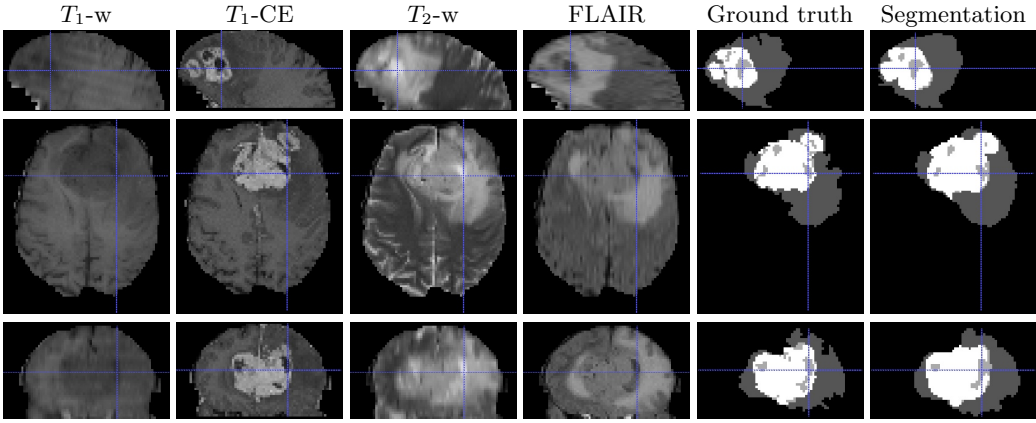


Fig. 4. Real high-grade case 15. From left to right: MR images, ground truth, and segmentation. From top to bottom: saggital, axial, and coronal views. Images are cropped.

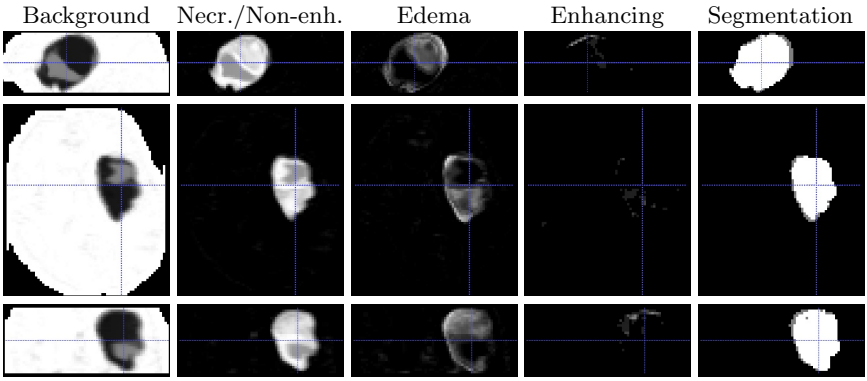


Fig. 5. Real low-grade case 06. From left to right: Vote maps for background, necrosis and non-enhancing tumor (merged), edema, enhancing tumor; Segmentation map. From top to bottom: saggital, axial, and coronal views. Images are cropped.

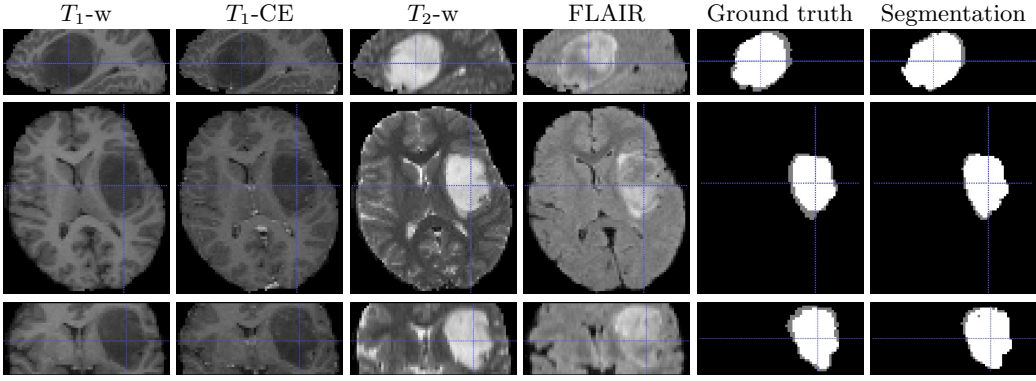


Fig. 6. Real low-grade case 06. From left to right: MR images, ground truth, and segmentation. From top to bottom: saggital, axial, and coronal views. Images are cropped.

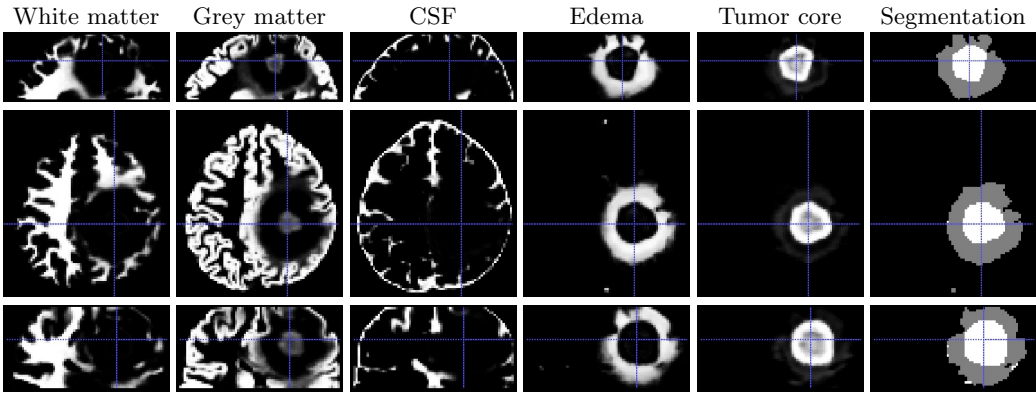


Fig. 7. Synthetic high-grade case 07. From left to right: Vote maps for white matter, grey matter, cerebro-spinal fluid (CSF), edema, and tumor core; Segmentation map. From top to bottom: saggital, axial, and coronal views. Images are cropped.

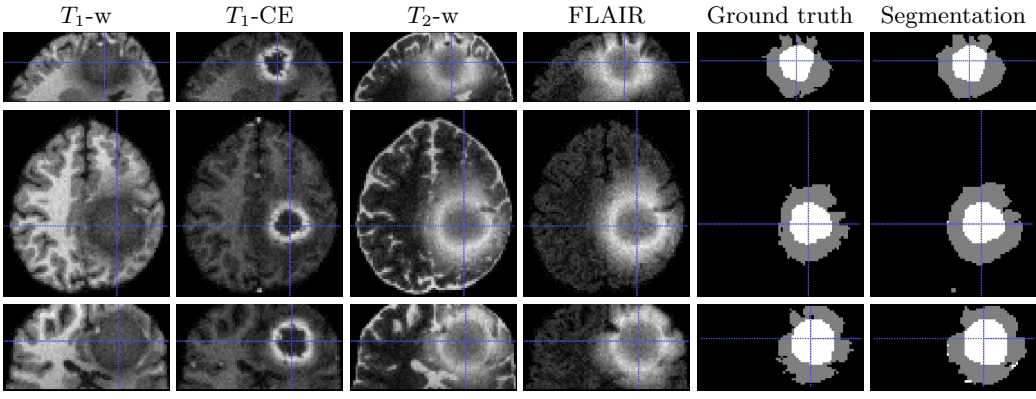


Fig. 8. Synthetic high-grade case 07. From left to right: MR images, ground truth, and segmentation. From top to bottom: saggital, axial, and coronal views. Images are cropped.

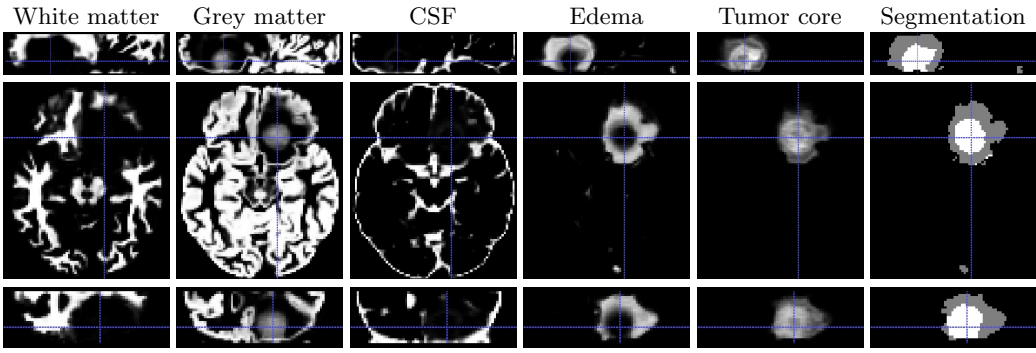


Fig. 9. Synthetic low-grade case 17. From left to right: Vote maps for white matter, grey matter, cerebro-spinal fluid (CSF), edema, and tumor core; Segmentation map. From top to bottom: saggital, axial, and coronal views. Images are cropped.

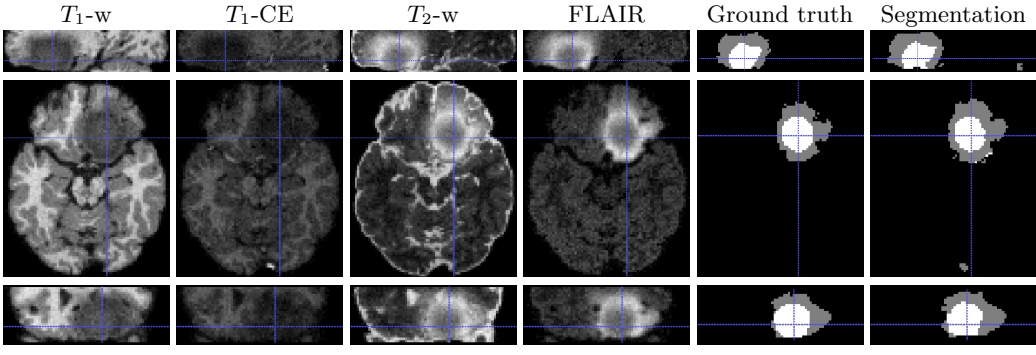


Fig. 10. Synthetic low-grade case 17. From left to right: MR images, ground truth, and segmentation. From top to bottom: saggital, axial, and coronal views. Images are cropped.

Fully Automatic Brain Tumor Segmentation from Multiple MR Sequences using Hidden Markov Fields and Variational EM

S. Doyle[1], F. Vasseur[1], M. Dojat[2], and F. Forbes[1]

[1] INRIA Rhône-Alpes, Grenoble, France, senan.doyle@inria.fr,
[2] INSERM, U836, Grenoble, France

Abstract. A fully automatic algorithm is proposed to segment glioma MR sequences, by availing of the complimentary information provided by multiple Magnetic Resonance (MR) sequences, and prior information incorporated in a Bayesian framework. The image data is described using a hidden Markov random field model, which accounts for spatial dependencies between neighbouring voxels, and allows for the inclusion of *a priori* spatial information. The algorithm is designed to perform without the need for a preliminary training phase by availing of this prior knowledge, and by imposing constraints on the estimation of MRF interaction parameters. The method is evaluated on high-grade and low-grade gliomas from the real-patient images of the BRATS2013 dataset. The computation time is 30 minutes per patient, giving an average Dice coefficient for high-grade and low-grade complete tumor volume of 0.84 and 0.81 respectively.

1 Introduction

The segmentation of healthy and diseased voxels in multi-sequence tumor MR images is addressed as a classification problem. For multiple MRI volumes, intensity distributions are commonly modelled as multi-dimensional Gaussian distributions. This provides a way to combine the multiple sequences in a single segmentation task replacing the continuous intensity field by a partition model. Such a model involves the introduction of a finite number of tissue classes and allocation variables to assign voxels to one of these classes. Spatial dependencies are then taken into account by modeling the allocation variables as a discrete state-space Markov field, namely a spatial Potts model. This results in a discrete HMRF model.

We build on the standard hidden Markov field model by considering a more general formulation that is able to encode more complex interactions than the standard Potts model (Section 2). In particular, we encode the fact that certain tissue combinations in the neighborhood are penalized more than others, whereas the standard Potts model penalizes dissimilar neighboring classes equally, regardless of the tissues they represent.

A solution to the model is found using the Expectation Maximization (EM) framework [1] combined with variational approximation for tractability in the presence of Markov dependencies. In particular, we consider the so-called mean field principle that provides a deterministic way to deal with intractable Markov Random Field (MRF) models [2] and has proven to perform well in a number of applications. Furthermore, in the absence of explicit expert knowledge, we show how a statistical atlas can be used to guide the model toward lesion identification. Experiments are reported to illustrate the performance of our approach.

2 Tumor Model Specification

We consider a finite set V of N voxels on a regular 3D grid. The intensity values observed at each voxel are denote by $\mathbf{y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$. Each $\mathbf{y}_i = \{y_{i1}, \ldots, y_{iM}\}$ is itself a vector of M intensity values corresponding to the M different MR sequences. The classification task is to assign each voxel i to one of K classes. This assignment is considered latent data that better describes the observed data in the context of the statistical model, and is denoted by $\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$. Typically, the $\mathbf{z}_i$'s corresponding to class memberships, take their values in $\{e_1, \ldots, e_K\}$ where e_k is a K-dimensional binary vector whose k^{th} component is 1, all other components being 0. The set in which $\mathbf{z}$ takes its values is represented by $\mathcal{Z} = \{e_1, \ldots, e_K\}^N$. The set of voxels V is associated to a neighborhood system. Spatial dependencies between voxels are modeled by assuming a Markov Random Field (MRF) prior. Denoting $\psi = \{\beta, \phi\}$ additional parameters, we assume that the joint distribution $p(\mathbf{y}, \mathbf{z}; \psi)$ is a MRF with the following energy function:

$$H(\mathbf{y}, \mathbf{z}; \psi) = H_{\mathbf{Z}}(\mathbf{z}; \beta) + \sum_{i \in V} \log g(\mathbf{y}_i | \mathbf{z}_i; \phi) \ , \tag{1}$$

where the $g(\mathbf{y}_i|\mathbf{z}_i;\phi)$'s are probability density functions of $\mathbf{y}_i$.

For brain data, the data term $\sum_{i \in V} \log g(\mathbf{y}_i|\mathbf{z}_i;\phi)$ in (1) corresponds to the modelling of tissue dependent intensity distributions. For our multi-dimensional observations, we consider M-dimensional Gaussian distributions with diagonal covariance matrices. For each class k, $(\mu_{k1}, \ldots, \mu_{kM})$ is the mean vector and $\{s_{k1}, \ldots, s_{kM}\}$ the covariance matrix components. We will use the notation $\mu_m = {}^t(\mu_{km}, k = 1 \ldots K)$ and $s_m = {}^t(s_{km}, k = 1 \ldots K)$. When $\mathbf{z}_i = e_k$ then $\mathcal{G}(y_{im}; \langle \mathbf{z}_i, \phi_m \rangle)$ and $\mathcal{G}(y_{im}; \langle \mathbf{z}_i, \mu_m \rangle, \langle \mathbf{z}_i, s_m \rangle)$ both represent the Gaussian distribution with mean μ_{km} and variance s_{km}. The whole set of Gaussian parameters is denoted by $\phi = \{\phi_{km}, k = 1, \ldots K, m = 1, \ldots, M\}$. Our data term is then defined by setting $g(\mathbf{y}_i|\mathbf{z}_i;\phi) \propto \prod_{m=1}^{M} \mathcal{G}(y_{im}; \langle \mathbf{z}_i, \phi_m \rangle)$ ..

The missing data term $H_{\mathbf{Z}}(\mathbf{z}; \beta)$ in (1) describes the dependencies between neighboring Z_i's, and is specified by further assuming that the joint distribution of $\{Z_1, \ldots, Z_N\}$ is a discrete MRF on the voxels grid :

$$P(\mathbf{z}; \beta) = W(\beta)^{-1} \ \exp\left(-H_{\mathbf{Z}}(\mathbf{z}; \beta)\right) \tag{2}$$

where $W(\beta)$ is a normalizing constant and $H_{\mathbf{Z}}$ is a function restricted to pairwise interactions,

$$H_{\mathbf{Z}}(\mathbf{z};\beta) = -\sum_{i\in S} z_i^t \alpha_i - \sum_{\substack{i,j \\ i\sim j}} z_i^t \mathbb{B} z_j,$$

where we write z_i^t for the transpose of vector z_i and $i \sim j$ when areas i and j are neighbors. The set of parameters β consists of two sets $\beta = (\alpha, \mathbb{B})$. Parameter α_i is a $K-$dimensional vector which acts as weights for the different values of z_i. The statistical atlas and tumor ROI are used to define these weights. $\mathbb{B}$ is a $K \times K$ matrix that encodes the interactions between different classes. If in addition to a null α, $\mathbb{B} = b \times I_K$ where b is a real scalar and I_K is the $K \times K$ identity matrix, parameters β reduce to a single scalar interaction parameter b and we get the Potts model traditionally used for image segmentation.

Note that the standard Potts model is often appropriate for classification since it tends to favor neighbors that are in the same class. However, this model penalizes pairs that have different classes with the same penalty, regardless of the tissues they represent. In practice, it may be more appropriate, to allow $\mathbb{B}$ to express higher penalties when the tissues are unlikely neighbors. For example, the penalty for a white matter and extraventricular CSF pair is expected to be greater than that of a grey matter and extraventricular CSF pair, as these two classes are more likely to form neighborhoods.

We adopt a data model comprising of five *normal* tissue classes; *white matter*, *grey matter*, *ventricular CSF*, *extraventricular CSF*, and *other*. The glioma is modeled by a further four classes representing the diseased tissue state; *edema*, *non-enhancing*, *enhancing* and *necrotic*. In the absence of sufficient data to robustly and accurately estimate a full free $\mathbb{B}$ with $K = 9$, further constraints are imposed on the MRF interaction matrix. The four glioma classes are considered a single *structure*, whose interaction with the normal tissue classes is not dependant on the specific glioma tissue state. Letting τ be the set of classes comprising the glioma structure, $\mathbb{B}$ is a matrix defined by:

$$\begin{aligned} \mathbb{B}(k,k') &= b_t \qquad \forall\ k,k' \in \tau \\ \mathbb{B}(k,k') &= b_{\{k,k'\}} \quad \text{otherwise.} \end{aligned} \tag{3}$$

For the distribution of the observed variables $\mathbf{y}$ given the classification $\mathbf{z}$, the usual conditional independence assumption is made. It follows that the conditional probability of the hidden field $\mathbf{z}$ given the observed field $\mathbf{y}$ is

$$P(\mathbf{z}|\mathbf{y};\psi,\beta) = W(\beta)^{-1} \exp\left(-H_{\mathbf{Z}}(\mathbf{z};\beta) + \sum_{i\in S} \log g(y_i|z_i,\phi)\right).$$

Parameters are estimated using the variational EM algorithm, which provides a tractable solution for non trivial Markov models.

3 Experiments

The deformable transform that describes the mapping between the International Consortium for Brain Mapping (ICBM) template and the data space is found using tools provided by the Insight Segmentation and Registration Toolkit (ITK). The transform is used to register the probabilistic tissue atlases to the MR sequences. An initial 5-class segmentation is performed, and the tumor region of interest (ROI) is detected by a simple morphological method comparing the segmentation result and the 5 tissue atlases. The prior probabilistic tissue atlas and the tumor ROI are incorporated *a priori* in the final segmentation algorithm via the α parameter in the MRF.

The algorithm was tested on real-patient data from the BRATS 2013 dataset. No training was performed; the initial labeling z_{init} was random, and all model parameters were estimated iteratively. Figure 1 & 2 show the correspondence between the ground truth (GT) and segmentation result for different slices of HG004. Table 1 details the average Dice coefficient for the high-grade and low-grade patients.

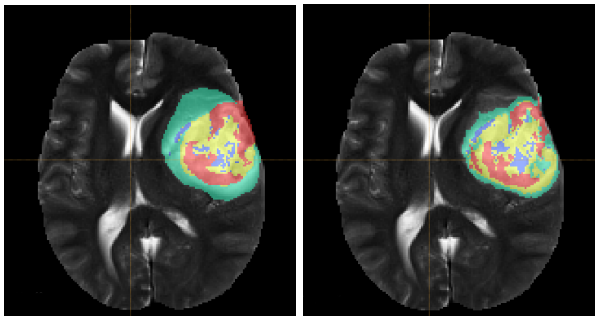


Fig. 1. Slice 98; GT & Result

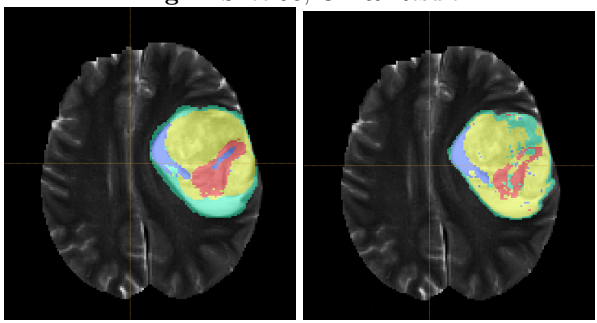


Fig. 2. Slice 116; GT & Result

	Complete Tumor	Tumor Core	Enhancing Tumor
High Grade	0.84	0.54	0.67
Low Grade	0.81	0.54	0.11

Table 1.

4 Discussion

We proposed an adaptive scheme of multiple MR sequences for brain tumor segmentation. Our approach is fully automatic and requires no training. The model parameters are instead estimated using a variational EM algorithm with MRF constraints and the inclusion of *a priori* probabilistic maps to provide a stable parameter trajectory during optimization. As our method requires no training, it could be used to perform the tedious task of creating training databases for other algorithms.

References

1. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B **39** (1977) 1–38
2. Celeux, G., Forbes, F., Peyrard, N.: EM procedures using mean field-like approximations for Markov model-based image segmentation. Pat. Rec. **36**(1) (2003) 131–144
3. McLachlan, G., Peel, D.: Finite Mixture Models. Wiley (2000)
4. Scherrer, B., Forbes, F., Garbay, C., Dojat, M.: Fully Bayesian Joint Model for MR Brain Scan Tissue and Structure Segmentation. In: MICCAI. (2008) 1066–1074

Automatic Brain Tumor Segmentation of Multi-sequence MR images using Random Decision Forests

MICCAI Grand Challenge: BRATS 2013

Joana Festa[1], Sérgio Pereira[1], José António Mariz[2,3], Nuno Sousa[2,3], Carlos A. Silva[1]

[1] Department of Electronics, University Minho, Portugal - csilva@dei.uminho.pt
[2] Life and Health Science Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
[3]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal

Abstract. This work is integrated in the MICCAI Grand Challenge: Multiparametric Brain Tumor Segmentation (BRATS). The main goal is to segment brain tumors, differentiating among several tumor labels. The proposed algorithm applies a trained Random Decision Forest to classify the voxels, based on meaningful features. The method is able to segment different tumor constituents, performing better for high grade tumors than for low grade.

Keywords: Random decision forests, brain tumor segmentation, multi-sequence, MRI

1 Introduction

We propose an algorithm for the segmentation of both low grade and high grade brain tumors, using a supervised classification method, namely a Random Decision Forest. The entire pipeline is fully automated. It starts with a two steps pre-processing stage, then the features are extracted and each voxel is classified, finally a post-processing step is applied to the classified volumes.

The provided dataset included T1, T2, FLAIR and T1 with contrast (T1C), all were co-registered to T1C and interpolated to 1 mm isotropic resolution. All the sequences are utilized, making our approach a multi-sequence one. The dataset consisted of 30 real glioma patients (20 high grade (HG) and 10 low grade (LG)), and 50 simulated glioma patients (25 of each grade). A manual segmentation was also provided distinguishing necrosis, edema, non-enhanced tumor and enhanced tumor.

2 Method

Proposed algorithm. Since the MR Datasets were not bias corrected, the first step of the pre-processing stage consisted in applying the N4ITK method [1]. Next, the intensity scale of each sequence was normalized to a chosen reference, by a histogram matching algorithm, implemented in ITK [2], [3].

A Random Decision Forest is used to classify the voxels in the brain, based on several features extracted from the training data. This classifier is a supervised one, which means that it must be trained before the classification. A Random Forest is a group of several Random Trees, which makes it easy to parallelize the process. Besides, there is some randomness during the training, making each tree unique and improving the generalization for unseen data. Other two advantages are the capability to deal with many features, even if they are redundant, and the possibility to be used in multilabel classification [4]. Finally, a post-processing was applied to the classified volumes to remove isolated points.

The features set, as well as the hyperparameters for the Decision Forest, were found using a leave-one-out cross-validation. The final and submitted classification of the real dataset was also performed using a leave-one-out cross validation. The training step uses all 4 sequences from all patients, except the patient being tested, from both low and high grade simultaneously, but separating the real dataset from the synthetic one.

Random Decision Forests. The main parameters in a Decision Forest are the number of trees and its depth. For this work, the forest is comprised of 50 trees with a depth of 25. Due to computational limitations, a total of 120 000 point per training subject was sampled from a set including all tumor and edema points and the double of their number in background points.

For each voxel, a set of meaningful features was extracted.

- MR sequences intensities: these features include both the intensity of the brain images in all sequences provided and the difference between each two sequences. The total number of this type of feature was 10.

- Neighborhood information: these features comprise the mean, sum, median and the intensity range of the 3D neighborhoods centered at the voxel that is being labeled. It was assumed that the neighborhoods would be cubes with edges of length 3, 9, 15 and 19 mm. Similar to the intensity features, these ones were calculated for each of the 4 sequences and for the differences between sequences. The total number of this kind of feature was 160.

- Context information: These features were calculated as the difference between the voxel being labeled and the mean of a 3x3x3 mm cube, which center was at a distance of 3 mm from the voxel. A total of 6 cubes per voxel were considered (2 per axis). Since this feature was extracted from all sequences, the total number of context features was 24.

- Texture: These features include edge density and local binary partition (signal and magnitude) extracted from a 3x3x3 mm neighborhood and the Laws texture features [5], extracted from a 2D neighborhood with 3x3 mm, in each of the 3 dimensions. All sequences were used, so the total number of texture features was 120.

Post processing. It was assumed that very small isolated regions of one label type should not exist. So, if a 3D connected component of one label type has 7 voxels or less, it is substituted by the mode of the surrounding voxels.

3 Results

Results were obtained using an online evaluation tool provided by the challenge organization. Table 1 summarizes the quantitative results for the segmentation of the data for high grade and low grade tumors in real cases. Figure 1 shows 3 examples of segmentations with the automatic algorithm.

Table 1. Results for the real data with high (HG) and low (LG) grade tumors, in terms of mean value (Mean) and standard deviation (Std) for 5 different metrics.

		Dice		**Jaccard**		**Specificity**		**Sensitivity**		**Kappa**	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
HG	1	0.83	0.08	0.72	0.12	0.87	0.07	0.80	0.13	0.99	0.00
	2	0.70	0.24	0.58	0.25	0.85	0.13	0.64	0.28	1.00	0.00
	3	0.75	0.16	0.63	0.18	0.80	0.14	0.74	0.20	1.00	0.00
LG	1	0.72	0.17	0.58	0.19	0.72	0.14	0.78	0.22	1.00	0.00
	2	0.47	0.23	0.33	0.19	0.70	0.30	0.42	0.26	1.00	0.00
	3	0.21	0.35	0.17	0.28	0.20	0.35	0.24	0.38	1.00	0.00
All	1	0.79	0.13	0.67	0.15	0.82	0.12	0.80	0.16	1.00	0.00
	2	0.62	0.26	0.50	0.26	0.80	0.21	0.57	0.29	1.00	0.00
	3	0.57	0.35	0.47	0.31	0.60	0.37	0.57	0.36	1.00	0.00

1. Complete tumor (tumor core and edema);
2. Tumor core (necrosis, non-enhancing tumor and enhancing tumor);
3. Enhancing tumor.

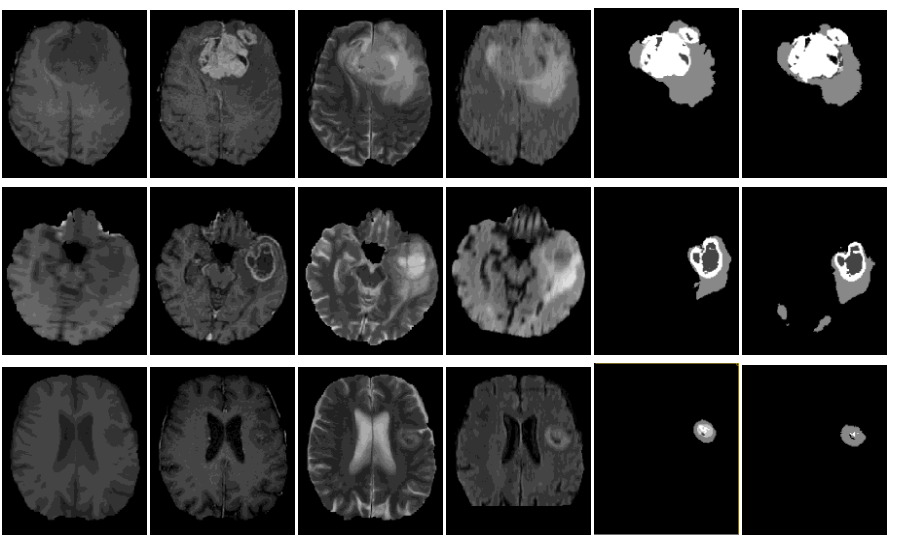


Fig. 1. Three examples of tumor segmentations. From left to right, in the columns are the database MRI sequences: T1, T1-contrast, T2 and Flair, the provided manual segmentation and the automatic segmentation, respectively. First row shows the best result achieved (HG0015), second row shows an image with evaluation close to the mean (HG0022), third row corresponds to the worst segmentation (HG0012).

The quantitative results indicate that by looking at the complete tumor (which includes edema) it is possible to achieve a good result, correcting some misclassification between tumor classes. Even so, the enhancing tumor class is well segmented by the algorithm, especially in high grade tumors. Many of the low grade images do not have this class or the necrotic, making the results much worst for this tumor grade.

4 Conclusion

In this paper, a fully automatic multi-sequence tumor segmentation in 4 different classes is reported. The algorithm is capable of achieving state-of-the-art results.

The total execution time is about 20 to 25 minutes for each test subject, using the programming language Python on a computer with an Intel processor (i7-3930k, 3.2 GHz) and 24 GB of RAM.

Our method is highly dependent on the sequences' intensities. This problem is maximized using a supervised algorithm because all images need to have a similar histogram distribution. Given this, for future work, we aim to study different image normalization methods and their impact on the segmentation results.

5 Acknowledgements

This work is supported by FEDER through Operational Program for Competitiveness Factors – COMPETE and by national funds through FCT – Fundação para a Ciência e Tecnologia in the scope of the project: FCOM-01-0124-FEDER-022674.

WiseRF™ is a product of wise.io, Inc. We thank wise.io for making an academic license freely available.

6 References

[1] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich and J. C. Gee, "N4ITK: Improved N3 Bias Correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310-1320, 2010.

[2] L. G. Nyul, J. K. Udupa and X. Zhang, "New Variants of a Method of MRI Scale Standardization," *IEEE Transactions on Medical Imaging*, vol. 19, no. 2, pp. 143-150, 2000.

[3] T. S. Yoo, M. J. Ackerman, W. E. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas and R. Whitaker, "Engineering and Algorithm Design for an Image Processing API: A Technical Report on ITK - The Insight Toolkit," *roc. of Medicine Meets Virtual Reality, J. Westwood, ed., IOS Press Amsterdam*, 2002.

[4] A. Criminisi, Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Foundations and Trends® in Computer Graphics and Vision, 2011.

[5] K. Laws, "Rapid Texture Identification," in *Proc. SPIE 0238, Image Processing for Missile Guidance*, San Diego, 1980.

Semi-automatic Segmentation of Multimodal Brain Tumor Using Active Contours

Xiaotao Guo, Lawrence Schwartz, Binsheng Zhao

Department of Radiology, Columbia University, New York, NY, USA
xg2145@columbia.edu

Abstract. In this paper, we present a semi-automatic segmentation method for multimodal brain tumors. It requires only that a user manually draw a region of interest (ROI) roughly surrounding the tumor on a single image. The algorithm combines the image analysis techniques of region and edge-based active contours and level set approach, and has the advantages of easy initialization, quick segmentation, and efficient modification. The typical run-time for each case in the training dataset can be within 1 minute.

1 Introduction

Fast and accurate segmentation of brain tumor images is an important but challenging task in many clinical applications. In recent years, a number of automatic and semi-automatic approaches have been proposed [1]. Fully automatic methods, in which the tumor contours are obtained without any human interaction, are very attractive in theory. Unfortunately, when the type of tumor doesn't fit the segmentation model as learned from the training dataset, the segmentation result may be erroneous. We present a fast semi-automatic brain tumor segmentation method, which only requires drawing a region of interest that roughly surrounds the tumor on a single image.

2 Methods

The general idea is based on a computer-aided tumor segmentation approach disclosed in our invention report [2]. The approach can be applied to segment images of highly heterogeneous tumors, helping to monitor response to therapy. The technique is modified and tailored here in order to segment multimodal brain tumors. The principle procedure of the segmentation is as follows. First, a user manually draws a region of interest on a single image (called reference image hereafter), followed by automatically refining the tumor contour on the reference image and propagating the tumor contour to a neighboring image using region- and edge-based active contour models [3][4]. For cases of high-grade gliomas, the ROI can be drawn on a post-

Gadolinium T1 image. The algorithm takes this ROI as an initialization and segments the total tumor regions, including enhancing tumor core and necrotic regions. Once the segmentation result is obtained, an adaptive threshold is determined based on the statistic analysis of the intensity distribution inside the tumor, and is used to separate the necrotic region from the enhancing tumor core. For edema segmentation, the ROI is first mapped to the FLAIR images as an initialization of the segmentation algorithm. A combination of global and local region-based active contour model is employed to segment the hyperintense region. The edema part is obtained by subtracting the total tumor region obtained in the previous step from the hyperintense region. The active contour is implemented using level set method [5], which allows topological changes, such as merging and splitting, which is a desirable property for the edema segmentation. Notice that the enhancing tumor and hyperintense region can be segmented separately, this means that registration of both type of images is not required for enhancing tumor core segmentation.

The system offers a user-friendly interface for quick modification of the segmented results. While reviewing the resulting contour overlaid on the image, an operator can modify the result either on one slice by using the "refine" function, or on multiple slices by using the "propagation" function, which allows a user to quickly and easily refine the result. With the help of the present automated refinement tool, the contour can be easily modified if the current segmentation result is not optimal.

3 Result

The method is implemented on Matlab environment, using mex files for core algorithms. For each case, user interaction and typical run-time for each case can be within 1 minute for a HP Z800 workstation. Table 1 shows the segmentation result obtained on the HG training dataset. For LG training dataset, we will present the result soon.

4 Discussion and Conclusions

We have presented a method for the segmentation of brain tumors, which achieves encouraging results with fast speed on clinical multimodal MR images. It also provides a mechanism for efficient modification of the segmented result. For a few very challenging cases, in which certain automatic segmentation models face difficulties, the present interactive system is a good alternative for obtaining the segmentation result in a fast and efficient way. Combining this segmentation method with the modification tool makes it a highly viable brain tumor segmentation system in practice.

Table 1: Segmentation results obtained on the HG training dataset

TYPE	CASE	Result			Dice			Jaccard			Positive Predictive Value			Sensitivity			Kappa
		Complete	Tumor	Enhancing													
HG	001	tumor	core	tumor	0.89	0.96	0.87	0.8	0.92	0.76	0.97	0.96	0.82	0.82	0.96	0.91	1
		Complete	Tumor	Enhancing													
HG	002	tumor	core	tumor	0.72	0.73	0.73	0.56	0.58	0.58	1	1	1	0.56	0.58	0.58	1
		Complete	Tumor	Enhancing													
HG	003	tumor	core	tumor	0.78	0.86	0.85	0.63	0.75	0.74	0.99	1	0.99	0.64	0.75	0.75	0.99
		Complete	Tumor	Enhancing													
HG	004	tumor	core	tumor	0.93	0.88	0.37	0.87	0.79	0.23	0.96	0.87	0.24	0.91	0.89	0.84	1
		Complete	Tumor	Enhancing													
HG	005	tumor	core	tumor	0.74	0.85	0.54	0.58	0.74	0.37	0.94	0.87	0.44	0.6	0.84	0.7	0.99
		Complete	Tumor	Enhancing													
HG	006	tumor	core	tumor	0.54	0.53	0.76	0.37	0.36	0.61	1	1	0.91	0.37	0.36	0.65	0.98
		Complete	Tumor	Enhancing													
HG	007	tumor	core	tumor	0.82	0.87	0.79	0.69	0.77	0.65	0.77	0.89	0.83	0.88	0.86	0.74	1
		Complete	Tumor	Enhancing													
HG	008	tumor	core	tumor	0.9	0.93	0.86	0.81	0.88	0.75	0.9	0.96	0.85	0.9	0.91	0.86	0.99
		Complete	Tumor	Enhancing													
HG	009	tumor	core	tumor	0.91	0.73	0.86	0.84	0.58	0.75	0.91	0.97	0.93	0.91	0.59	0.79	0.99
		Complete	Tumor	Enhancing													
HG	010	tumor	core	tumor	0.74	0.87	0.81	0.59	0.77	0.68	0.88	0.99	0.98	0.64	0.77	0.68	1
		Complete	Tumor	Enhancing													
HG	011	tumor	core	tumor	0.92	0.89	0.82	0.85	0.8	0.69	0.99	1	0.88	0.85	0.8	0.76	1
		Complete	Tumor	Enhancing													
HG	012	tumor	core	tumor	0.9	0.58	0.16	0.81	0.41	0.08	0.97	0.63	0.09	0.83	0.54	0.72	1
		Complete	Tumor	Enhancing													
HG	013	tumor	core	tumor	0.85	0.89	0.76	0.74	0.8	0.62	0.99	0.96	0.94	0.75	0.83	0.64	1
		Complete	Tumor	Enhancing													
HG	014	tumor	core	tumor	0.94	0.94	0.84	0.88	0.89	0.72	0.96	0.96	0.83	0.91	0.92	0.84	1
		Complete	Tumor	Enhancing													
HG	015	tumor	core	tumor	0.9	0.9	0.89	0.82	0.81	0.8	0.85	0.99	0.97	0.95	0.82	0.81	0.99
		Complete	Tumor	Enhancing													
HG	022	tumor	core	tumor	0.9	0.9	0.8	0.83	0.83	0.66	0.9	0.89	0.76	0.91	0.92	0.85	1
		Complete	Tumor	Enhancing													
HG	024	tumor	core	tumor	0.67	0.88	0.77	0.51	0.79	0.62	0.69	0.98	0.89	0.66	0.8	0.67	0.99
		Complete	Tumor	Enhancing													
HG	025	tumor	core	tumor	0.71	0.68	0.32	0.55	0.51	0.19	0.79	0.64	0.22	0.64	0.72	0.61	0.99
		Complete	Tumor	Enhancing													
HG	026	tumor	core	tumor	0.81	0.7	0.6	0.68	0.53	0.43	0.88	0.78	0.57	0.75	0.63	0.64	0.99
		Complete	Tumor	Enhancing													
HG	027	tumor	core	tumor	0.86	0.78	0.72	0.76	0.64	0.56	0.86	0.8	0.65	0.86	0.76	0.79	0.99

Reference

1. N. Gordillo, E. Montseny, P.Sobrevilla, “State of the art survey on MRI brain tumor segmentation,” *Magnetic Resonance Imaging*, Article in Press , 2013.
2. X. Guo, B. Zhao, L. Schwartz, “Computer Aided Tumor Segmentation” (Reference: 2906) http://innovation.columbia.edu/technologies/2906/cumc-radiology-imaging-lab-available-technologies
3. V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," *International Journal of Computer Vision*, vol. 22, pp. 61-79, 1997.
4. D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Appl. Math.*, vol. 42, pp. 577-685, 1989.
5. S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, pp. 12-49, 1988.

A Hybrid Model for Multimodal Brain Tumor Segmentation

Raphael Meier[1], Stefan Bauer[1,2], Johannes Slotboom[2], Roland Wiest[2], and Mauricio Reyes[1]

[1] Institute for Surgical Technologies and Biomechanics, University of Bern
[2] Inselspital, Bern University Hospital, Switzerland
raphael.meier@istb.unibe.ch

Abstract. We present a fully automatic segmentation method for multimodal brain tumor segmentation. The proposed generative-discriminative hybrid model generates initial tissue probabilities, which are used subsequently for enhancing the classification and spatial regularization. The model has been evaluated on the BRATS2013 training set, which includes multimodal MRI images from patients with high- and low-grade gliomas. Our method is capable of segmenting the image into healthy (GM, WM, CSF) and pathological tissue (necrotic, enhancing and non-enhancing tumor, edema). We achieved state-of-the-art performance (Dice mean values of 0.69 and 0.8 for tumor subcompartments and complete tumor respectively) within a reasonable timeframe (4 to 15 minutes).

1 Introduction

Brain tumor segmentation in multimodal image data plays a major role in different clinical areas such as radiology, radiotherapy planning and longitudinal studies [3]. Fully automatic segmentation methods greatly facilitate the otherwise cumbersome and time-consuming manual segmentation process. Moreover, such methods do not suffer from bias by the observer.

This work is a continuation of the model previously published by Bauer et al. [1][2]. We extend the discriminative model to a generative-discriminative hybrid model. The usage of a hybrid model is motivated by recent work on brain tumor segmentation by Geremia et al. and Zikic et al. [6][10], concepts from auto-context [9] and well-known results from machine learning [8]. Compared to [6] and [10] we also employ tissue probabilities from a generative model. However, our method does not induce a spatial regularization via features but as a postprocessing step. By using tissue probabilities to model the relationship between neighboring voxels during regularization, we improve final classification results.

2 Methods

In this section we consider a supervised learning setting in which we are given a fully labeled training set $\mathcal{S} = \left\{ \left(\mathbf{x}^{(i)}, y^{(i)} \right) : i = 1, ..., |\mathcal{S}| \right\}$, where $\mathbf{x}$ denotes the

feature vector and y the corresponding class label. Furthermore, the superscripts index the sample data and $|\mathcal{S}|$ represents the cardinality of our training set. Our method can be subdivided into three main parts: an estimation of initial tissue probabilities based on a generative model, a supervised classification based on a discriminative model and a spatial regularization. In contrast to our previous approach, this model will subdivide the image into four pathological subcompartments (necrotic component, enhancing and non-enhancing tumor, edema). The pipeline is visualized in figure 1. The magnetic resonance imaging (MRI) modalities used are: T_1, T_{1c}, T_2, $FLAIR$.

2.1 Density Forest

In this paper we explore a recently published generative version of the well-known random or decision forest [4] called *density forest* [5]. We utilize density forests to model the class-dependent likelihood $p(\mathbf{x}|y)$ which we will use to estimate initial tissue probabilites. More specifically, for each of the seven tissue classes $c \in C$ we employ a density forest to model the respective likelihood $p(\mathbf{x}|y = c)$. Notice that only the feature vectors of the respective class are used to train such a density forest, which means that $\mathcal{S} = \left\{\mathbf{x}^{(i)} : \forall i.(y^{(i)} = c)\right\}$. Our feature vector contains the four voxel-wise intensities: $\mathbf{x}_I = (I_{T1}, I_{T1c}, I_{T2}, I_{FLAIR})^T$.

Analogously to a standard forest model we consider split nodes where a binary split is performed via axis-aligned hyperplanes. During training, the parameters of the weak learners are chosen such that they maximize the information gain. The main difference between a density forest and a classification forest arises in the node model for both split and leaf nodes. Every node in a decision tree of a density forest represents a component of the density modeled by a multivariate normal distribution. Discrete node statistics in classification forests are replaced by continuous statistics. Thus, the information gain IG for a node j is expressed in terms of differential entropy

$$IG = \log(\det \Sigma(\mathcal{S}_j)) - \sum_{i \in \{L,R\}} \frac{\left|\mathcal{S}_j^i\right|}{|\mathcal{S}_j|} \log(\det \Sigma(\mathcal{S}_j^i)) \tag{1}$$

where $\mathcal{S}_j$, $\mathcal{S}_j^i$ denote the training data before and after the split and Σ is a $n \times n$ covariance matrix.

A density forest can be seen as a hierarchical Gaussian mixture model with hard label assignments. Beginning at the root we have a single multivariate normal distribution modelling the whole density. The children of the root node subdivide the density into two components each represented by a multivariate normal distribution. This recursive partitioning of the probability density continues with increasing depth of the trees. For a more detailed explanation of the model we refer to the book of Criminisi et al. [5].

During testing, a previously unseen feature vector $\mathbf{x}_I$ is pushed down the tree and ends up in a leaf $l(\mathbf{x}_I)$. The estimate of the class-dependent likelihood

$p_t(\mathbf{x}_I|y=c)$ of a single decision tree t is then given by

$$p_t(\mathbf{x}_I|y=c) = \frac{\left|\mathcal{S}_{l(\mathbf{x}_I)}\right|}{|\mathcal{S}|}\frac{1}{Z_t}\mathcal{N}(\mathbf{x}_I;\mu_{l(\mathbf{x}_I)},\Sigma_{l(\mathbf{x}_I)}) \tag{2}$$

where $|\mathcal{S}_{l(\mathbf{x}_I)}|$ indicates the amount of training data contained in leaf $l(\mathbf{x}_I)$. As suggested in [5], in case of axis-aligned hyperplanes the partition function Z_t corresponds to the cumulative multivariate normal distribution function. We assumed Σ to be diagonal which greatly simplifies calculations and allows us to obtain Z_t via the error function. Parameters of node-based multivariate normal distributions are estimated using maximum likelihood. The final likelihood is given as the average over the whole forest $p(\mathbf{x}_I|y=c) = 1/T\sum_{t=1}^{T} p_t(\mathbf{x}_I|y=c)$. The prior distribution $p(y)$ is estimated from the empirical class histogram over the training set. In summary, each of the seven tissue classes has been modeled by a forest-based likelihood term and the corresponding prior. The tissue probabilities for class c are then computed as posterior probabilities:

$$p(y=c|\mathbf{x}_I) = \frac{p(\mathbf{x}_I|y=c)p(y=c)}{\sum_{\bar{c}} p(\mathbf{x}_I|y=\bar{c})p(y=\bar{c})}. \tag{3}$$

The rationale to use those tissue probabilities is twofold. First, they will enhance the feature vector of our discriminative model. Second, they will be integrated into the spatial regularization. In the following we will refer to the tissue probabilities as $p_G(y|\mathbf{x}_I)$ where the subscript G should indicate that they stem from a generative model.

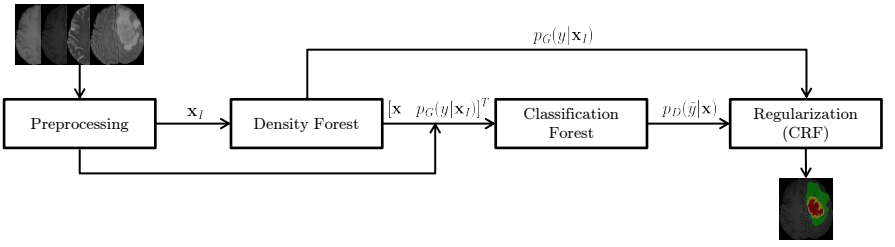


Fig. 1: Segmentation pipeline. After the multimodal image has been preprocessed, initial tissue probabilities are estimated, followed by classification and subsequent spatial regularization.

2.2 Classification Forest

The discriminative part of our model performs the supervised classification of a voxel. For every voxel we extract 44 features including first-order texture measures, gradient information and symmetry features. Additionally, we add the 7

voxel-wise tissue probabilities estimated by the density forest, which yields a 51-dimensional feature vector $\mathbf{x}$. As a classifier we use a classification forest which is trained on the fully labeled training set $\mathcal{S}$. The predicted class label is defined according to the MAP-rule: $\tilde{y} = \arg\max_y p(y|\mathbf{x})$. The probabilistic output $p(\tilde{y}|\mathbf{x})$ of the classification forest will be used later on in the regularization step. In the following we will refer to the posterior probability of the classification forest as $p_D(y|\mathbf{x})$ where the subscript D indicates that it stems from a discriminative model.

2.3 Regularization

The regularization is based on our hierarchical approach from [1], where it is formulated as an energy minimization problem of a conditional random field (CRF). The image is represented as a grid graph $G = (V, E)$ where the energy

$$E = \sum_{i \in V} \psi_i(\mathbf{x}^{(i)}, y^{(i)}) + \sum_{(i,j) \in E} \psi_{i,j}(\mathbf{x}^{(i)}, y^{(i)}, \mathbf{x}^{(j)}, y^{(j)}) \tag{4}$$

of the CRF corresponds to a sum of unary and pairwise potentials. The unary potential describes the affinity of a voxel to a possible class c, whereas the pairwise potential resembles the coherence between neighboring voxels. By using the posterior probability $p_D(\tilde{y}^{(i)}|\mathbf{x}^{(i)})$ given by the classification forest the unary potentials are defined as

$$\psi_i(\mathbf{x}^{(i)}, y^{(i)}) = p_D(\tilde{y}^{(i)}|\mathbf{x}^{(i)}) \cdot (1 - \delta(\tilde{y}^{(i)}, y^{(i)})) \tag{5}$$

where $\delta(\cdot)$ is the Kronecker delta.

The pairwise potential is defined differently from our previous approach to incorporate initial tissue probabilities. This term is motivated by the idea that voxels belonging to the same segment show similar discrete tissue probability distributions estimated by the generative model. Similarity between the probability distributions of neighboring voxels i and j is measured in terms of the Bhattacharyya-distance:

$$D_B = -\log(BC(i,j)) \tag{6}$$

where

$$BC(i,j) = \sum_{c \in C} \sqrt{p_G(y^{(i)} = c|\mathbf{x}_I^{(i)}) p_G(y^{(j)} = c|\mathbf{x}_I^{(j)})}. \tag{7}$$

If we replace the Pseudo-Chebyshev distance measure used for the discontinuity penalty of our previous approach by (6), the pairwise potential reduces to

$$\psi_{i,j}(\mathbf{x}^{(i)}, y^{(i)}, \mathbf{x}^{(j)}, y^{(j)}) = w_s(i,j)(1 - \delta(y^{(i)}, y^{(j)}))BC(i,j)D_t(y^{(i)}, y^{(j)}) \tag{8}$$

where $w_s(i,j)$ denotes a weighting function, which depends on the voxel spacing, $(1 - \delta(y^{(i)}, y^{(j)}))$ enforces a Potts model and D_t allows to incorporate prior knowledge about tissue dependencies. For the optimization of (4) we employed the Fast-PD algorithm [7].

3 Results

We evaluated the performance of our proposed model on the training data of the BRATS2013 dataset. This data set comprises 20 high-grade cases and 10 low-grade cases including manual expert annotations. We performed a 5-fold cross validation for the high-grade cases and leave-one-out cross validation in case of the low-grade tumors. For implementing density and classification forests we used the Sherwood library [5]. The hyperparameters of the decision forests were chosen according to a gridsearch. The evaluation of the results has been conducted online on the Virtual Skeleton Database[3] (VSD). An exemplary segmentation of a high-grade case is shown in figure 2. Quantitative results are shown in table 1. The testing time of our (unoptimized) algorithm ranges from 4 to 15 minutes depending on the size of the image volume. We performed a nonparametric statistical test (Wilcoxon signed-rank test, $\alpha = 0.05$) for the high-grade cases to compare the estimated Dice-coefficients of our hybrid model against the ones from our previous model. The test yielded a statistical significant difference ($p = 0.03, W = 47$) regarding the tumor core (necrotic component + enhancing and non-enhancing tumor) suggesting an increased accuracy.

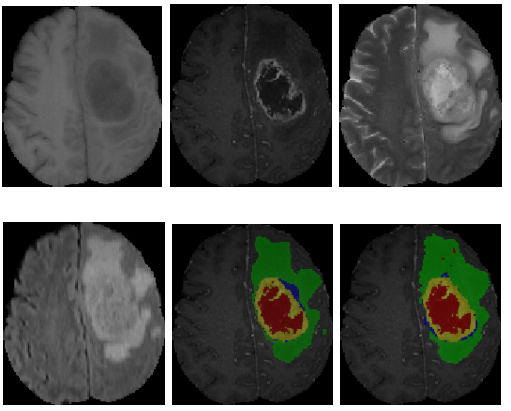


Fig. 2: Segmentation result for case HG0011. First row (left to right): T_1, T_{1c}, T_2. Second row: $FLAIR$, overlayed ground truth on T_{1c} image (necrotic = red, enhancing tumor = yellow, non-enhancing tumor = blue, edema = green), overlayed segmentation result for the four tumor subclasses .

[3] https://vsd.unibe.ch/WebSite/BRATS/Start2013

Region	Dice	Jaccard	PPV	Sensitivity
Complete tumor (HG)	0.802 ± 0.124	0.684 ± 0.153	0.773 ± 0.151	0.859 ± 0.126
Tumor core (HG)	0.691 ± 0.220	0.561 ± 0.212	0.712 ± 0.143	0.719 ± 0.262
Enhancing tumor (HG)	0.698 ± 0.247	0.578 ± 0.232	0.754 ± 0.213	0.680 ± 0.270
Complete tumor (LG)	0.764 ± 0.105	0.628 ± 0.131	0.738 ± 0.171	0.836 ± 0.136
Tumor core (LG)	0.585 ± 0.229	0.446 ± 0.228	0.728 ± 0.319	0.570 ± 0.200
Enhancing tumor (LG)	0.201 ± 0.326	0.152 ± 0.250	0.241 ± 0.388	0.178 ± 0.296

Table 1: Results of online evaluation for high- (HG) and low-grade (LG) cases. Performance measures are given as mean values ± standard deviation.

4 Discussion and Conclusion

We presented a fully automatic segmentation algorithm capable of segmenting multimodal images into four tumor subcompartments similar to the 'VASARI' guidelines[4]. We utilized the novel generative model proposed in [5] to derive initial tissue probabilities, which are subsequently employed in a voxel-wise classification and spatial regularization. The used modalities are standard in clinical acquisition protocols (T_1, T_{1c}, T_2, $FLAIR$) and the algorithm completes processing within a reasonable timeframe. The described algorithm is being integrated into the BraTumIA software suite, which offers an easy-to-use interface for radiologists to perform brain tumor image analysis.

Acknowledgments. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°600841. The research was also partially funded by the Bernese Cancer League.

References

1. S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L.-P. Nolte, and M. Reyes. Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization. In *Proceedings of MICCAI-BRATS 2012*, Jan. 2012.
2. S. Bauer, L.-P. Nolte, and M. Reyes. Fully Automatic Segmentation of Brain Tumor Images using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 14(3), Jan. 2011.
3. S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes. A survey of MRI-based medical image analysis for brain tumor studies. *Physics in Medicine and Biology*, 58(13), July 2013.
4. L. Breiman. Random Forests. *Machine Learning*, 2001.

[4] https://wiki.cancerimagingarchive.net/display/Public/VASARI+Research+Project

5. A. Criminisi and J. Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013.
6. E. Geremia, D. Zikic, O. Clatz, B. H. Menze, Glocker, E. Konukoglu, J. Shotton, O. M. Thomas, S. J. Price, T. Das, R. Jena, N. Ayache, and A. Criminisi. Classification Forests for Semantic Segmentation of Brain Lesions in Multi-channel MRI. In *Decision Forests for Computer Vision and Medical Image Analysis*, pages 245–260. 2013.
7. N. Komodakis, G. Tziritas, and N. Paragios. 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), Oct. 2008.
8. A. Y. Ng and M. I. Jordan. On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes. In *NIPS 2001*, 2001.
9. Z. Tu and X. Bai. Auto-context and Its Application to High-level Vision Tasks and 3D Brain Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10), 2009.
10. D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 15(3), Jan. 2012.

Multi-class Abnormal Brain Tissue Segmentation Using Texture Features

S. Reza and K. M. Iftekharuddin

Vision Lab, Department of Electrical and Computer Engineering,
Old Dominion University, Norfolk, VA 23529, USA.

Abstract

In this work, we propose fully automated multi-class abnormal brain tissue segmentation in multimodality brain MRI. Different brain tissues are characterized using novel texture features such as piece-wise triangular prism surface area (PTPSA), and textons, along with intensity difference and regular intensity in multimodal MRI scans. Classical Random Forest (RF) classifier is used for segmentation and classification of these features in multi-modal MRI (T_1, T_2, Flair, $T_{1contrast}$). Efficacy of abnormal brain tissue segmentation is evaluated using the publicly available BRATS2013 training dataset. Quantitative Dice and Jaccard scores are obtained using the online evaluation tool from Virtual Skeleton database website and reported for verification.

Methods

This basic idea of this work is based on our novel multiresolution-fractal based brain tumor segmentation works [2] [3]. In our prior works, we exploit SVM as weak classifier with a novel modified Adaboost algorithm for tumor and other tissue segmentation. In contrast, this work uses Random Forests [4] for multiclass abnormal brain tissue classification we obtain 2D MRI slices from 3D volume MRI for subsequent processing. Since BRATS 2013 dataset is already skull stripped and co-registered, we restrict our preprocessing steps to bias and inhomogeneity correction only. The overall flow diagram of our approach is shown in Fig. 1.

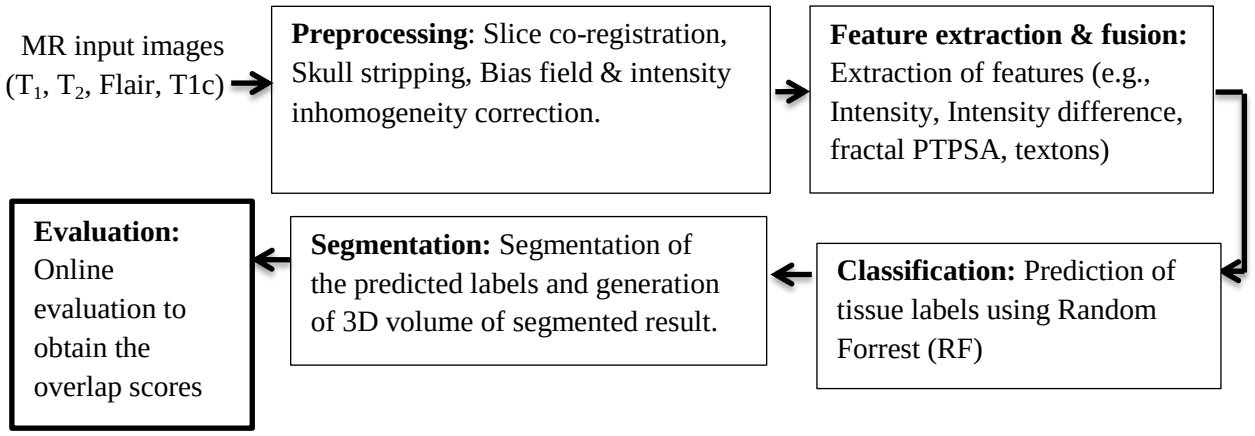


Figure 1: Generic flow diagram of the proposed method

Preprocessing: To minimize the intensity bias of the MR image, intensity normalization is used as pre–processing step. N4ITK [5] MRI bias correction tool of slicer3D is used for bias correction. We

implement a two- step normalization method [6], where the image histograms are modified such that the histograms match a mean histogram obtained using the training data. .

Feature Extraction & Fusion: For each slice of the input images we extract both spatial and non-local features as follows.

Non-local feature: We extract intensities of four MRI modalities (I_{T1}, I_{T2}, I_{Fl}, I_{T1c}) and the corresponding difference of pixel intensities among the modalities ($d_1=I_{T1}- I_{T2}$, $d_2=I_{T1}- I_{Fl}$, $d_3=I_{T1}- I_{T1c}$). These difference features (d_1, d_2, d_3) captures the amount of intensity variation at each point among the MR modalities. Similar type of difference features are also used in [7].

Spatial/texture feature: We employ our novel texture features such as fractal PTPSA [2] [3] [8] and texton [2] [9] to characterize the tumor surface variation which is expected to be different from the non-tumor region. The efficacy of texture based tumor detection, segmentation and classification has been discussed [3] [8]. More details on our texture features can be found in ref [10]. After extraction, we perform feature domain fusion with all our features.

Abnormal Brain Tissue Classification

We use RF [4] [11] for classification of different abnormal tissue types in the brain. The RF is an ensemble learning algorithm that generates many classifiers and aggregates their results in order to make decisions. The RF is based on classification trees and adds an additional layer of randomness to bagging. Each branch of a tree is divided by the best among a randomly selected subset of predictors [12] [13]. This added randomness helps RF to perform better compared to other classifiers such as, support vector machines and neural networks. We use Classification Forests (CFs) that are ensembles of (binary) classification trees [7]. At each node n, classification tree randomly takes a subset of training samples X_n and predicts a class $p_t^n(\omega|x)$, where p_t is the probability of the sample x in class ω [7]. Based on the features; classification forests continually splits the training samples at every node, and assign the partitions $X_L \; and \; X_R$ to the left/right nodes. This splitting is done with a random dimension. Tree growing is continues up to a certain tree depth, D_T. In testing phase, data points to be classified are pushed through each tree t, with the learned split functions. The leaf node probability is directly used as the tree probability i.e.

$$p_t(\omega|x) = p_t^l(\omega|x) \quad (1)$$

where, p_t^l denotes the probability of at leaf node, l of sample x in class ω.

The overall probability is calculated by the following equation,

$$p(\omega|x) = \frac{1}{T}\sum_{t=1}^{T} p_t(\omega|x) \quad (2)$$

where, $p(\omega|x)$ is the average probability of sample x in class ω and T is the total number of trees.

Finally the class with highest probability is estimated $\hat{\omega}$ as the actual class, i.e.

$$\hat{\omega} = \arg max_{\omega} \; p(\omega|x) \quad (3)$$

More details about CFs can be found in [14].

We perform 3-fold cross-validation for each training patient data.

Results and Discussions

We obtain 2D segmented tissues using the predicted pixel labels from RF. These 2D abnormal tissue segments are then stacked to generate volume image. Example tissue segments using two slices from a patient are shown in Fig. 2.

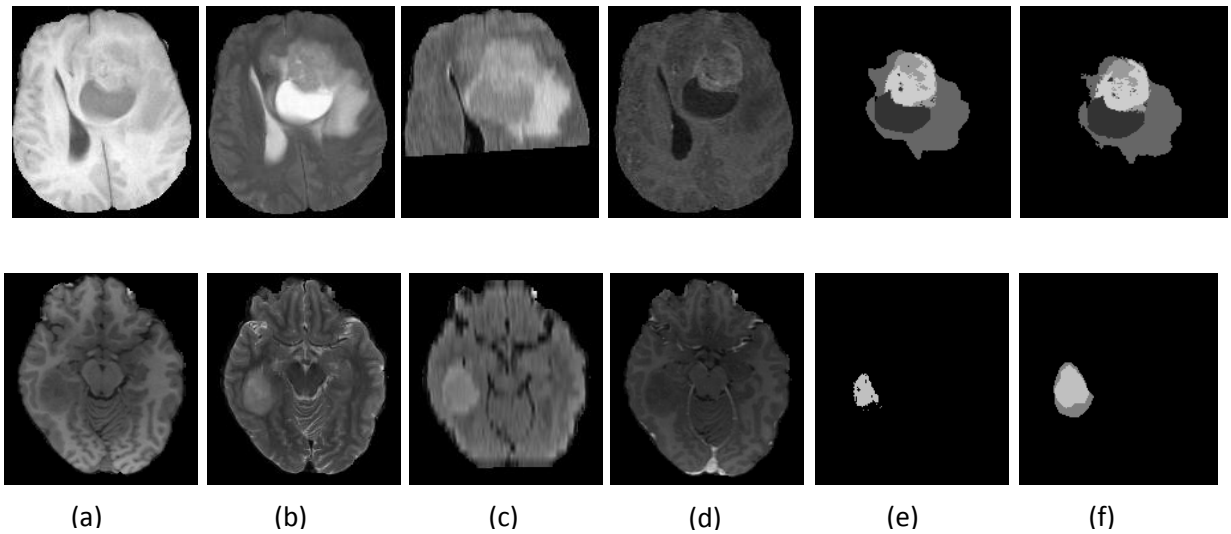


Figure 2: Segmented tissues with corresponding input and ground-truth images. Each row represents two example set of multimodality MRI slices. ; Input: (a) T1, (b) T2, (c) Flair (d) T1contrast. (e) Segmented image (f) ground-truth. Labels in the ground-truth: 1-necrosis, 2- edema, 3-non-enhancing tumor, 4- enhancing tumor, 0-everything else.

Quantitative evaluation: We evaluate our preliminary abnormal tissue segmentation results using the BRATS-2013 online evaluation tool from the Virtual Skeleton database. Overall the results suggest comparatively higher score for complete tumors than enhancing tumor parts. Three different categories such as complete tumor, tumor core and enhanced tumor are considered for the evaluation. The details on these three categories are as follows: Complete Tumor: (1-necrosis, 2-Edema, 3-non-enhancing tumor, 4-enhance tumor); Tumor Core: (3-non-enhance tumor, 4-enhance tumor); and Enhance tumor: (4-enhacne tumor). Summary of quantitative overlap scores using Dice and Jaccard metrics for Low Grade (LG) and High Grade (HG) are presented in Table 1 and 2 respectively. The results in these Tables show that our proposed method is effective in segmenting different grades of glioma tumors. Our segmentation rate varies between 88% to 92% using Dice overlap metric for enhanced tumor, tumor core, complete tumor respectively. Low standard deviation, 4% - 7% indicates that the proposed method offers consistent results for both LG and HG segmentation examples in this study.

Table 1: 3-fold cross validation results of BRATS-2013 training cases. Dice and Jaccard score for complete tumor, tumor core and enhancing tumor for Low Grade (LG) tumor. High Grade

Pat ID	Dice	Jaccard

	Complete Tumor	Tumor Core	Enhancing tumor*	Complete Tumor	Tumor Core	Enhancing tumor*
L015	0.93	0.93	--	0.87	0.86	--
L014	0.77	0.85	--	0.62	0.73	--
L013	0.86	0.87	0.91	0.75	0.77	0.83
L012	0.94	0.93	--	0.89	0.88	--
L011	0.96	0.97	--	0.92	0.95	--
L008	0.96	0.95	--	0.92	0.90	--
L006	0.94	0.91	--	0.89	0.83	--
L004	0.93	0.87	--	0.88	0.77	--
L002	0.97	0.89	0.83	0.95	0.80	0.70
L001	0.90	0.90	0.91	0.83	0.83	0.84
Mean	0.92	0.91	0.88	0.85	0.83	0.79
Std.	0.06	0.04	0.46	0.01	0.07	0.08

(*Note: -- indicates the absence of the corresponding tissue [1] and reported -1 by the online evaluation tool.

Table 2: 3-fold cross validation results of BRATS-2013 training cases. Dice and Jaccard score for complete tumor, tumor core and enhancing tumor for High Grade (HG) tumor.

Pat ID	Dice			Jaccard		
	Complete Tumor	Tumor Core	Enhancing tumor*	Complete Tumor	Tumor Core	Enhancing tumor*
H027	0.95	0.92	0.9	0.90	0.86	0.81
H026	0.82	0.74	0.76	0.69	0.59	0.61
H025	0.92	0.88	0.76	0.85	0.79	0.61
H024	0.91	0.94	0.90	0.84	0.88	0.81
H022	0.93	0.93	0.88	0.87	0.88	0.78
H015	0.95	0.95	0.97	0.90	0.91	0.93
H014	0.95	0.96	0.91	0.91	0.92	0.84
H013	0.93	0.96	0.95	0.87	0.93	0.91
H012	0.93	0.88	0.67	0.86	0.78	0.50
H011	0.97	0.96	0.94	0.94	0.93	0.88
H010	0.84	0.90	0.88	0.73	0.81	0.79
H009	0.96	0.93	0.94	0.91	0.86	0.88
H008	0.94	0.91	0.90	0.88	0.84	0.82
H007	0.90	0.90	0.88	0.81	0.82	0.78
H006	0.97	0.95	0.91	0.93	0.90	0.84
H005	0.86	0.90	0.86	0.76	0.82	0.75
H004	0.97	0.95	0.89	0.93	0.91	0.80
H003	0.94	0.93	0.94	0.89	0.87	0.88
H002	0.93	0.87	0.90	0.87	0.77	0.81
H001	0.95	0.97	0.93	0.90	0.94	0.86
Mean	0.92	0.91	0.88	0.86	0.85	0.79
Std.	0.04	0.05	0.07	0.07	0.08	0.11

Conclusion

In this work we propose novel abnormal brain tissue segmentation method and investigate the efficacy of our technique. Experimental results with 30 clinical LG and HG patient data confirm the efficacy of our method for multi-class abnormal brain tissue segmentation. Our training results show comparable performance when compared to other state-of-the art works posted on the VSD website [1]. Our future works include study of feature selection and fusion of other effective features such as level set [3], and multi-fractal texture (mBm) [2].

Acknowledgements

This work is partially supported through a grant from NCI/NIH (R15CA115464).

References

[1] https://vsd.unibe.ch/WebSite/BRATS/Start2013/

[2] A. Islam, S. Reza, and K. M. Iftekharuddin, "Multi-fractal texture estimation for detection and segmentation of brain tumors," *IEEE Transactions on Biomedical Engineering, To Appear*, 2013.

[3] S. Ahmed, K. Iftekharuddin, and A. Vossough, "Efficacy of texture, shape, and intensity feature fusion for posterior-fossa tumor segmentation in MRI," *IEEE Transactions on Information Technology in Biomedicine*, pp. 206-213, 2011.

[4] A. Cutler and L. Breiman, "Random forests-classification description," *Technical report, University of California, Berkeley*, 2004.

[5] N. Tustison and J. Gee. N4ITK: Nick's N3 ITK implementation for MRI bias field correction. *The Insight Journal*, 2010.

[6] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Transaction on Medical Imaging*, vol. 19, no. 2, pp. 143-150, 2000.

[7] D. Zikic, B. Glocker, E. Konkoglu, J. Shotton, A. Criminisi, D. H. Ye, C. Demiralp, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Context-sensitive classification forests for segmentation of brain tumor tissues.," in *Proceedings MICCAI-BRATS*, 2012, pp. 1-9.

[8] K. M. Iftekharuddin, S. Ahmed, R. J. Ogg, and F. H. Laningham, "Efficacy of texture , shape and intensity features for robust posterior-fossa tumor segmentation in MRI." *SPIE Med. Imag.*, vol. 7260, Feb. 2009.

[9] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29 – 44, 2001.

[10] K. M. Iftekharuddin,W. Jia, and R. March, "Fractal analysis of tumor in brain MR images," *Machine Vision and Applications*, vol. 13, pp. 352-362, 2003.

[11] L. Breiman, "Random forest," Statistics department, University of California, Berkeley, CA 94720, January, 2001.

[12] M. Wiener and A. Liaw, "Classification and regression by random forests," *R News*, 2002.

[13] Wei-Yin Loh, "Classication and regression tree methods, " *Encyclopedia of Statistics in Quality and Reliability*, 2008.

[14] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests: A unified frame-work for classification, regression, density estimation, manifold learning and semi-supervised learning," *Foundations and Trends in Computer Graphics and Vision*, 7(2-3), 2012.

Map-Reduce Enabled Hidden Markov Models for High Throughput Multimodal Brain Tumor Segmentation

Thomas Taylor[1], Nigel John[2], Patricia Buendia[1], Michael Ryan[1]

[1]INFOTECH Soft, Inc.
[2]Department of Electrical and Computer Engineering, University of Miami

Abstract
We have developed a novel extension to Hidden Markov Models (HMMs) to enable high-throughput training and segmentation of tumors and edema in multimodal magnetic resonance images of the brain. Our method has been evaluated on the two-label BRATS2013 training dataset for both simulated and real patient high-grade glioma cases. We achieve an mean accuracy (Dice score) of [66.7]% for edema and [89.2]% for tumor in the simulated cases and [59.5]%for edema and [65.6]% for tumor in the real cases. The Map-Reduce enabled HMM is able to train on all cases simultaneously, performing 220% faster on an 8-node cluster than on a single node. Segmentation of a single patient case takes less than one minute.

Introduction
Brain tumor segmentation in Magnetic Resonance Imaging (MRI) is an important task for neurosurgeons, oncologists and radiologists to assess disease burden and measure tumor response to treatment. An estimated 69,720 new cases of primary malignant and non–malignant brain and CNS tumors are expected to be diagnosed in the United States in 2013. In 2012, approximately 13,700 deaths were attributed to primary malignant brain and central nervous system tumors in the United States [1]. There is a large variation in five-year survival estimates depending upon tumor histologies, ranging from 94% for pilocytic astrocytomas to less than 5% for glioblastomas and the survival period of patients with gliomas after initial diagnosis is generally limited to 12-14 months [15]. The glioblastoma multiforme brain tumor is the most common primary tumor of the CNS, accounting for approximately 40% of brain tumors across patients of all ages [14].

Manual segmentation of brain-tumor images for volume measurement has been a common practice in clinics, but it is time-consuming, labor intensive, and subject to considerable variation in intra- and inter-operator performance [5]. In the current clinic, the tumor volume is often approximated by the area of maximal cross-section, which is often further approximated to an ellipse [14][6]. Rough approximation is used because the time cost to compute a more accurate manual volume estimate is too high. A study by Vaidyanathan [7] measured manual segmentation of 3D tumor volumes from MR images to average between 30 minutes and 3.5 hours per patient. In a study by Kaus [8], the time required for segmentation of the complete 3D volume was on the order of 3-5 hours per patient. In addition to time cost, manual segmentation is subject to inter-operator variability and human error [8][9]. Variability in serial volume measurements and inconsistency between observers has been noted in studies involving tumor volume measurement [10][11][12]. In a study by Kaus [8] intra-observer variability of manual segmentation for four observers over 20 patient cases ranged between 0.24% and 4.11%, while inter-observer variability ranged between 2.62% and 14.42% for meningioma and low-grade glioma. The process of manual diagnosis using MRI is error prone primarily because of large numbers of image slices and variation between intensities of different images [13].

The Map-Reduce enabled Hidden Markov Model brain tumor segmentation approach has been developed to specifically address fast and accurate delineation of tumor boundaries in multimodal MRI data of the brain. In this paper, we focus on describing our BRATS submission.

Methods
Preprocessing prepares the input MR spectra, T1, T1 with gadolinium contrast (T1C), T2, and FLAIR, for segmentation. The preprocessing pipeline has been designed to remove spatial inhomogeneities due to patient movement, remove image artifacts (skull, eyes) not related to the segmentation problem, remove inhomogeneities due to MR scanner bias fields, and match each spectrum's intensity histogram to the volumes used for training.

Training the HMM involves extracting a feature vector for each voxel in the

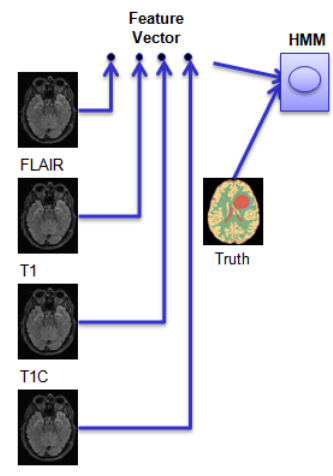


Figure 1. Training the HMM Model

source case. We extract voxels from FLAIR, T1, T1C, and T2 MR spectra. Neighboring voxels are added to the feature vector. The corresponding truth labels for the voxel neighborhood in the feature vector is utilized for supervised training of the HMM.

Extending the HMM model to Map-Reduce involved adapting the HMM supervised learning algorithm to incrementally update based on individual feature vectors and coding a Mapper to perform feature extraction. In our current case, a single Mapper handles a single training case, extracting all of the feature vectors for the case and providing the vectors to the Reducer. The Reducer collects the feature vectors from all of the Mappers and incrementally updates the HMM model as new feature vectors are produced. A final Controller normalizes the probabilities in the HMM (initial, transition, emission) and stores the HMM to a file.

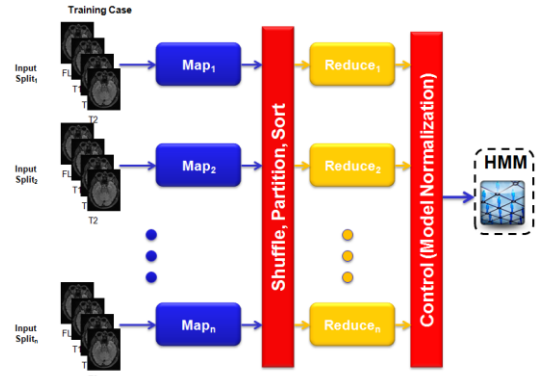


Figure 2. MapReduce Model for HMM-based Brain Tumor Segmentation

Segmenting with the HMM involves extracting the feature vector for each voxel in the target case in the same manner as HMM training. Voxels from FLAIR, T1, T1C, and T2 in a neighborhood around the voxel of interest are organized into the feature vector and provided to the trained HMM model. The HMM model produces a predicted label for the feature vector.

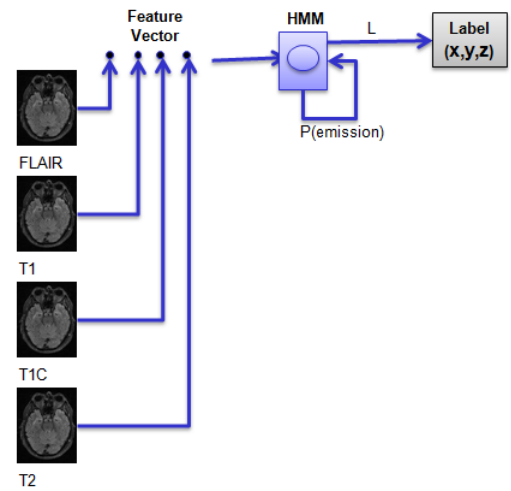
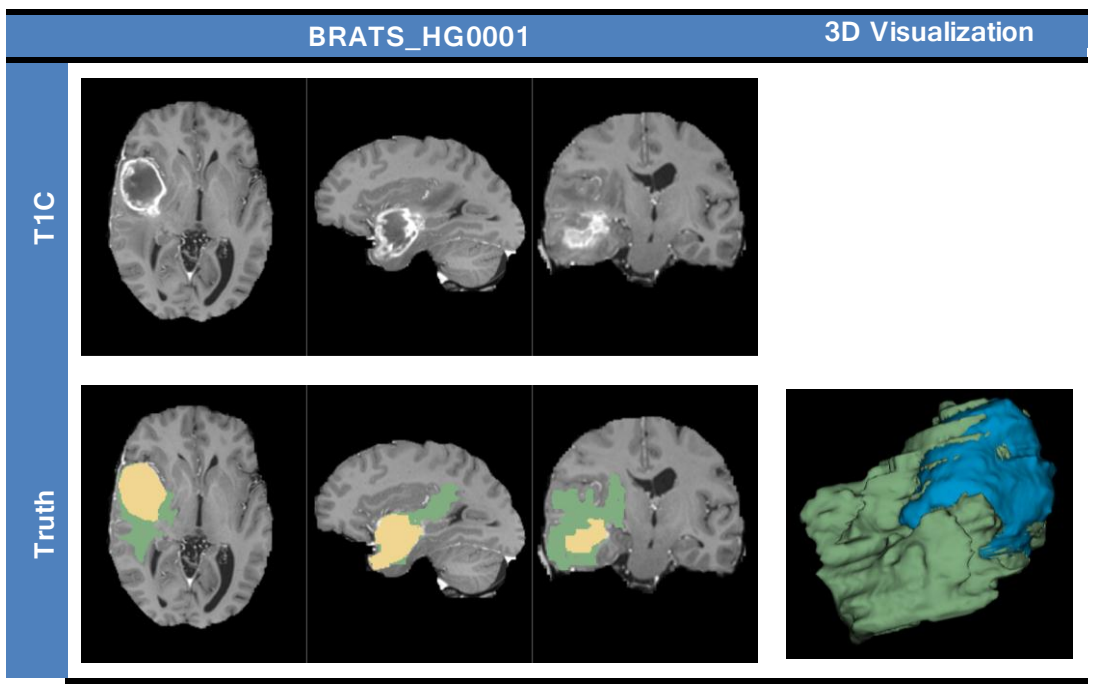


Figure 3. Applying the HMM Model for Segmentation

Postprocessing involved filtering out small objects and applying dilation and erosion operations on each segmented class.

Results

We performed 20-fold cross validation on the real high-grade BRATS2013 training set and 25-fold cross validation on the simulated high-grade training set using the two-label ground truth provided by the challenge. We achieve an mean accuracy (Dice score) of [66.7]% for edema and [89.2]% for tumor in the simulated cases and [59.5]%for edema and [65.6]% for tumor in the real cases.



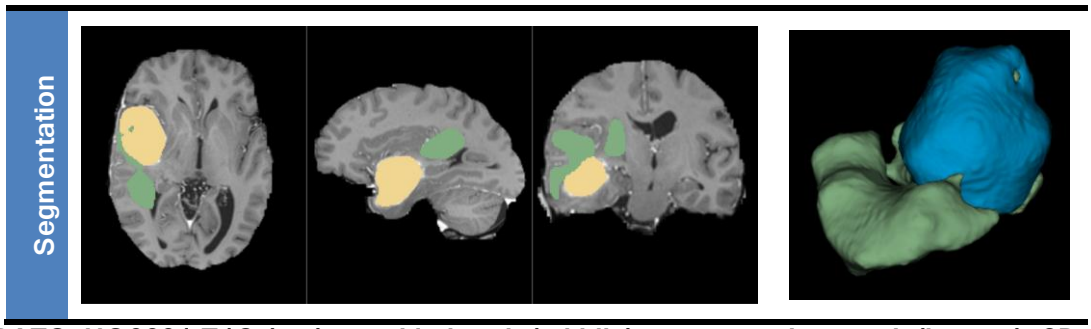


Figure 4. BRATS_HG0001 T1C (top), provided truth (middle), segmentation result (bottom). 3D visualization (right) of tumor (blue) and edema (green).

High-grade (synthetic)		High-grade (real)	
Edema	Tumor	Edema	Tumor
0.6626	0.8928	0.5962	0.6263

Table 1. Average cross-segmentation scores for the BRATS2013 training dataset.

We compared the runtime required to train a single HMM on all 25 cases in the SimBRATS_HG dataset. The dataset was loaded into Hadoop distributed file system and the tests measured the runtime of training the HMM given a single computational node, 2, 4, 5, 6, 7, 8, 10, 15, 20, and 25 nodes. A single Reducer task was assigned with collecting the Mapper results and adjusting the HMM accordingly. As the number of Mappers increase, the total runtime decreases until 7 Mappers are used as computational nodes begin to be assigned multiple mapping tasks. The total reduce time increases slightly as the number of Mappers increase as we only have a single Reducer collecting the Mapper's feature vectors to produce the HMM.

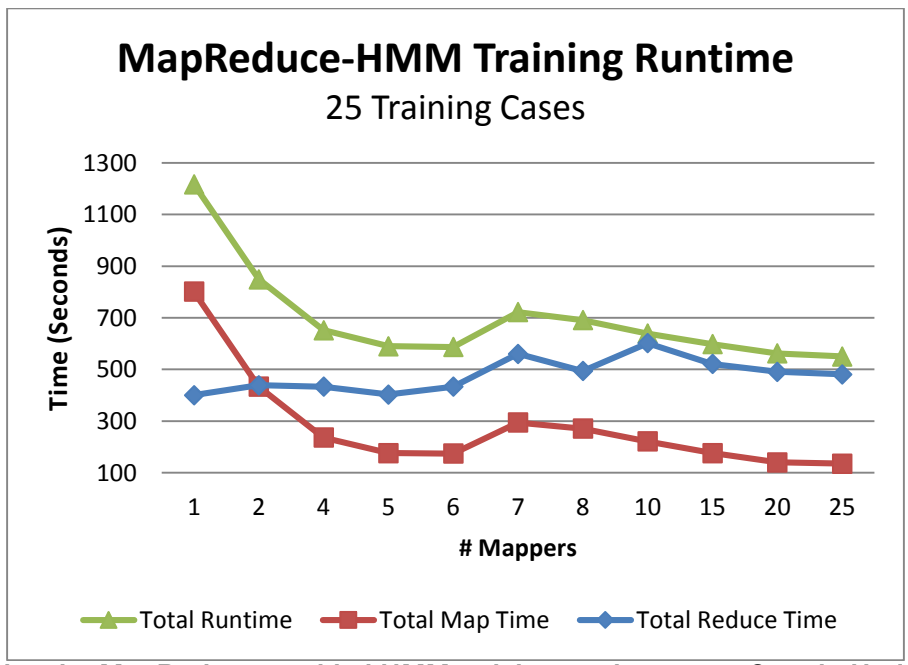


Figure 5. Comparing the MapReduce-enabled HMM training runtime on an 8-node Hadoop cluster as the number of Mappers increase from 1 to 25.

Acknowledgement

This work was partially supported by National Cancer Institute grant #HHSN261201200032C.

References

[1] Central Brain Tumor Registry of the United States (CBTRUS), Fact Sheet, http://www.cbtrus.org/factsheet/factsheet.html

[2] Surveillance, Epidemiology, and End Results (SEER), SEER Stat Database, http://www.seer.cancer.gov, accessed October 2011

[3] Kumar M, et al. Novel Membrane-Permeable Contrast Agent for Brain Tumor Detection by MRI, Magnetic Resonance in Medicine 63:617-624 (2010)

[4] Corso JJ, Sharon E, Dube S, El-Saden S, Sinha U, Yuille A: Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification. Medical Imaging 2008, 28(5):629-640.
[5] Alfano B, A. Brunetti, M. Larobina M. Quarantelli, E. Tedeschi, A. Ciarmiello, E. M. Covelli, and M. Salvatore, "Automated Segmentation and Measurement of Global White Matter Lesion Volume in Patients with Multiple Sclerosis," *Journal of Magnetic Resonance Imaging* 12:799–807, 2000.
[6] Sorensen AG, et al. Response criteria for gioma, Nature Reviews Clinical Oncology 5:634-644 (November 2008)
[7] Vaidyanathan M, et al. Monitoring Brain Tumor Response to Therapy Using MRI Segmentation, Magnetic Rsonance Imaging, Vol. 15 (3) 323-334, 1997
[8] Kaus MR, Warfield S, Nabavi A, Black PM, Jolesz FA, Kikinis R. Automated segmentation of MR images of brain tumors. Radiology, 218(2):586-591, 2001.
[9] Vos MJ, Interobserver variability in the radiological assessment of response to chemotherapy in glioma, Neurology 60:826-830 (2003)
[10] Clarke LP, et al., MRI segmentation: methods and applications. Magnetic Resonance Imaging 13:323-334 (1995)
[11] Joe BN, et al. Brain Tumor Volume Measurement: Comparison of Manual and Semiautomated Methods, Radiology 212:811-816 (September 1999)
[12] Zijlstra EJ, et al. Radiotherapy response of cerebral metastases quantified by serial MR imaging, Journal of Neurooncology 21:171-176 (1994)
[13] Dubey RB, et al. Evaluation of Three Method for MRI Brain Tumor Segmentation, Eighth International Conference on Information Technology: New Generations, 2011
[14]Corso JJ, Sharon E, Dube S, El-Saden S, Sinha U, Yuille A: Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification. Medical Imaging 2008, 28(5):629-640.
[15]Kumar M, et al. Novel Membrane-Permeable Contrast Agent for Brain Tumor Detection by MRI, Magnetic Resonance in Medicine 63:617-624 (2010)

ANTs and Árboles

Nick Tustison[1], Max Wintermark[1], Chris Durst[1], and Brian Avants[2]

[1] University of Virginia, Charlottesville VA 22903, USA
[2] University of Pennsylvania, Philadelphia PA 18104,USA

Abstract. Given the success of random forest approaches for segmentation, particularly for the BRATS 2012 tumor segmentation challenge, we implemented a variant framework for our own research. The innovation of our methodology and implementation is characterized by the following four-fold contribution: 1) generation of novel feature images in addition to what has been previously reported which significantly enhances classification, 2) concatenated application of random forest models for improved performance, 3) the use of ANTsR (a packaging of the ANTs library plus additional analysis tools for the R statistical project) for direct access to robust random forest functionality with parallelization, and 4) public availability of all scripts to recreate the leave-one-out evaluation study performed with the provided training data.

Keywords: ANTsR, Atropos, N4, R, random forests, segmentation

1 Introduction

The success of random forest (RF)-based approaches in the BRATS 2012 challenge, our own clinical research needs, and the lack of publicly available tools for such processing motivated the implementation that we describe herein. Although we borrowed many ideas from related previous work, our approach expands on this previous work in algorithmic and implementation terms both of which rely heavily on our Advanced Normalization Tools (ANTs)[3] including its R packaging known as ANTsR. This includes concatenated application of RF models (one based on Gaussian mixture modeling (GMM), similar to previous efforts, used as input to the succeeding one based on maximum a priori estimation and Markov random fields (MAP-MRF)). We then refine the resulting labeling using binary morphological processing.

2 Methods

The proposed workflow for estimating tumor-based labeling from multi-modal images involves the following steps:

1. Symmetric template construction [3][4] using the data described in [4].

[3] http://stnava.github.io/ANTs
[4] Implementation in the script `antsMultivariateTemplateConstruction.sh`.

2. Image preprocessing.
   - Windowing image intensities (quantiles $[0.01, 0.99]$).
   - N4 bias correction [5].
   - Rescaling intensity range to $[0, 1]$.
3. Stage 1 processing:
   - generation of feature images,
   - construction of the Stage 1 RF model and probability images.
4. Stage 2 processing:
   - generation of single-modality MAP-MRF images using the Stage 1 RF probability images as spatial priors,
   - construction of the Stage 2 RF model and labelings.
5. Refinement of Stage 2 labelings using binary morphological processing.

2.1 Multi-Modal Feature Image Generation

Based on previous work and our own experience, we selected the following feature images for our supervised segmentation framework:

- Per modality (FLAIR, T1, T1C, T2)
  - First-order neighborhood statistical images: mean, variance, skewness, and entropy. Neighborhood radius $\in \{1, 3\}$.
  - GMM (stage 1) and MAP-MRF (stage 2) posteriors: CSF, gray matter, white matter, necrosis, edema, non-enhancing tumor and enhancing tumor (or just CSF, gray matter, white matter, edema, and tumor for the simulated data).
  - GMM (stage 1) and MAP-MRF (stage 2) connected component geometry features: distance to tumor core label, volume, volume to surface area ratio, eccentricity, and elongation
  - Template-based: symmetric template difference and contralateral difference with Gaussian smoothing ($\sigma = 4$mm).
- Miscellaneous: normalized Euclidean distance based on cerebral mask, log Jacobian image, and (T1C - T1) difference image.

Prior cluster centers for specific tissue types learned from training data are used in the first stage to construct multiple GMM-based feature images [2]. The resulting spatial priors derived from application of the random forest model for the first stage were used as input to an iterative n-tissue N4 $\rightleftharpoons$ Atropos MAP-MRF segmentation protocol (encapsulated in the ANTs script `antsAtroposN4.sh`) [2]. These are used to create modified feature images for the second stage.

ANTs registration capabilities [1] are also used to determine the transform from each subject to the symmetric template (using the tool `antsRegistration`). This transform provides three sets of feature images: the log Jacobian determinant image (assuming that the presence of tumor causes abnormal displacements), voxelwise intensity differences between each modality of each subject and the corresponding symmetric template component, and voxelwise contralateral differences.

2.2 Implementation

As mentioned previously, motivating this work were the limited public resources for performing multi-modal tumor segmentation. Github is used to make available all scripts for the evaluation study as well as the source for this document[5] in addition to ANTs,[6] ANTsR,[7] and some additional utilities.[8] The available scripts include:

- `createTruthLabels.pl` – performs a 3-tissue segmentation of the tumor training data. These three labels are then combined with the given labels.
- `createNormalizedImagesForCohort.pl` – windows and rescales the images (commented out are previous attempts at N4 bias correction and intensity normalization).
- `createFeatureImagesForCohort.pl` – calculates the features images by calling `createFeatureImages.sh` for each subject.
- `runLeaveOneOutCrossValidation.pl` – calls `createRandomForestModel.pl` for each subject using the training data from the other subjects.
- `createRandomForestModel.pl` – calls the R script `createModel.R`.
- `applyTumorSegmentationModelForCohort.pl` – creates the random forest probability maps for each label using `applyTumorSegmentationModel.sh`.
- `refineTumorSegmentationResultsForCohort.pl` – refines the final labels from the random forest model using STAPLE.
- `createFeatureImages.sh` – creates features images for a specific subject.
- `applyTumorSegmentationModel.sh` – given a random forest model (.RData), produces the subject-specific random forest probability images.
- `createModel.R` – R script interface to the `randomForest` R package. Provides optional parallelization with the `snowfall` package.
- `applyModel.R` – R script interface to the `randomForest` R package. Provides optional parallelization with the `snowfall` package.
- `createCSVFileFromModel.R` – produces a csv file containing a data frame of feature images.
- `plotVariableImportance.R` – plots feature importance given a random forest model.

3 Evaluation

A leave-one-out evaluation strategy was adopted for each of the four cohorts (high vs. low grade and real vs. simulated data). The metrics for performance assessment were given by the organizers and include combining labels to assess overlap of complete tumor (labels 1–4), tumor core (labels 1,3, and 4), and enhancing tumor (label 4). The resulting assessment metrics are provided in Table 1. All processing was performed using the computational cluster at the

[5] https://github.com/ntustison/BRATS2013
[6] https://github.com/stnava/ANTs
[7] https://github.com/stnava/ANTsR
[8] https://github.com/ntustison/Utilities

University of Virginia.[9] Given the number of subjects, all processing was single-threaded although multi-threading can easily be employed. The timing for the various stages were approximately as follows: normalization (0.5 hours), Stage 1 feature image creation (8-10 hours due to image registration), Stage 2 feature image creation (1-1.5 hours), Stage 1 RF model construction (1-2 hours), and Stage 2 RF model construction (1-2 hours), and all remaining steps (0.5 hours).

Table 1. Scores from the MICCAI 2013 BRATs Evaluation Data

	Dice			Pos. predictive value			Sensitivity		
	complete	core	enhancing	complete	core	enhancing	complete	core	enhancing
Real	0.88	0.76	0.55	0.88	0.80	0.65	0.89	0.79	0.53
Sim.	0.89	0.90	0.00	0.82	0.88	0.00	0.99	0.93	0.00

4 Discussion and Conclusions

Our experience matches with previous research demonstrating good segmentation performance using random forests. We found that a two-stage model construction incorporating a second MAP-MRF step while incorporating additional feature images to what has been proposed previously can significantly improve classification. This includes symmetric template-based features, normalized Euclidean distance and log Jacobian images. Additionally, we provide our implementation as open source.

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54(3), 2033–44 (Feb 2011)
2. Avants, B.B., Tustison, N.J., Wu, J., Cook, P.A., Gee, J.C.: An open source multivariate framework for n-tissue segmentation with evaluation on public data. Neuroinformatics 9(4), 381–400 (Dec 2011)
3. Avants, B.B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., Gee, J.C.: The optimal template effect in hippocampus studies of diseased populations. Neuroimage 49(3), 2457–66 (Feb 2010)
4. Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A.D., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., Smith, S.A., Joel, S., Mori, S., Pekar, J.J., Barker, P.B., Prince, J.L., van Zijl, P.C.M.: Multi-parametric neuroimaging reproducibility: a 3-T resource study. Neuroimage 54(4), 2854–66 (Feb 2011)
5. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29(6), 1310–20 (Jun 2010)

[9] http://www.uvacse.virginia.edu

Automatic Brain Tumor Segmentation with MRF on Supervoxels

Liang Zhao, Duygu Sarikaya, and Jason J. Corso

Computer Science and Engineering, SUNY at Buffalo

Abstract. Segmenting brain tumors from multi-modal imaging remains to be a challenging task despite the growing interest in the area. Brain tumors have a highly variable shape, appearance and spatial location. In this paper, we propose an algorithm that automatically segments brain tumors in magnetic resonance human brain images. Our method uses a Markov random field model based on supervoxels and terms that capture the intensity probabilities and edge cues. The evaluation results on the BRATS2013 training data shows the efficiency and robustness.

1 Introduction

Despite the growing interest and many approaches [3] proposed in recent years, the segmentation of tumor in brain MRI remains a challenging task. There is still a great need for an automated segmentation method that works accurately and efficiently. Working with MRI to segment tumors is challenging because the tumors have a high variance in shape, appearance and spatial location, leading to limited applicability of prior models and a difficulty in modeling. In particular, labeling a single voxel or a voxel in a small neighborhood can lead to errors because of this variability. Markov random field on superpixels is widely used on natural image segmentation [4]. In this work, we rely on a supervoxel over-segmentation of the MRI image first and then label within each supervoxel, constraining all constituent voxels to be the same label. Our likelihood model on the intensities is based on histogram matching and forms the unary potential in a Markov random field defined over the supervoxel graph. Inference is ultimately performed with graph cuts.

2 Method

We normalize the data and estimate the likelihood of pixels by the registration of a 3D joint histogram. Then for each MRI case, we first perform over-segmentation, which results in a set of supervoxels. We then solve the voxel labeling problem directly on the supervoxels.

2.1 Pre-processing

For each channel of each MRI case, we first denoise with SUSAN [6]; then we compute the standardized z-scores (zero mean and unit covariance) to put the data in the same scale.

2.2 Oversegmentation of the Image with Supervoxels

Supervoxels provide a perceptually consistent unit that we can use to compute local image features. Rather than working on voxels itself, working with meaningful entities obtained from a low-level grouping process is more efficient. Using supervoxels highly reduces the cost of subsequent image processing steps. In order to obtain supervoxels of MRI scan images, we use SLIC 3D [1] which generates supervoxels by clustering voxels based on their color similarity and proximity in the image volume.

2.3 Segmentation with Graph Cuts on a Markov Random Field

For the actual segmentation, we define a Markov random field using the probability distributions over the pixels as the unary term and the edge cues of shared boundaries of two superpixels to define the binary term. We group the pixel features into supervoxels at each step and construct a model based on these supervoxels.

Consider a Markov random field defined over the supervoxels $\mathcal{S}$. A labeling f assigns a label $f_P \in \text{Ł}$ to each supervoxel $P \in \mathcal{S}$, where $\text{Ł} = \{N,\ E,\ nonET,\ ET,\ C,\ B\}$, necrosis, edema, non-enhancing tumor, enhancing tumor, cerebrospinal fluid and background (white matter and gray matter), respectively. The energy function,

$$E(f) = \sum_{Q \in \mathcal{S}} D_Q(f_Q) + \sum_{(P,Q) \in N_{\mathcal{S}}} V_{PQ}(f_P, f_Q)\ , \tag{1}$$

where $\mathcal{S}$ is the set of supervoxels and $N_{\mathcal{S}}$ is the set of adjacent supervoxels, captures the cost of a certain labeling f. We define the data term as $D_Q(f_Q) = \sum_{q \in Q} -log(P(I(q)|f_Q))$, where $P(I(q)|f_Q)$ is the node class likelihood estimated by histogram based method (Sec. 2.4) and $I(q)$ denotes the feature of voxel q. We define the smoothness term to capture the edge presence along the common boundary of the two supervoxels:

$$V_{PQ}(f_P, f_Q) = \delta(f_P \neq f_Q) \cdot \left[\left(\alpha + \beta \frac{1}{\sqrt[3]{|P||Q|}} \right) \sum_{p \in P, q \in Q \cap N_p} (1 - \max(Edge(p), Edge(q))) \right] \tag{2}$$

where α,β are two nonnegative parameter, and N_p is the neighborhood of p. $Edge(p)$ is defined as

$$\begin{aligned} Edge(p) &= \max_{q \in N} P\left(f_{r_{q,p}} \neq f_q | I(q), I(r_{q,p})\right) \\ &= \max_{q \in N_p} \frac{Pr\left(I(q), I(r_{q,p}) | f_{r_{q,p}} \neq f_q\right)}{Pr\left(I(q), I(r_{q,p})\right)}\ , \end{aligned} \tag{3}$$

where $r_{q,p}$ is a voxel, such that q and $r_{q,p}$ are symmetric about p.

Finally, we solve the supervoxel labeling energy minimization problem using graph cuts [2].

2.4 Histogram Based Likelihood Estimation

Histogram Construction We describe the unary term for measuring supervoxel likelihood. For each case, we first quantize each of the three channels, $T1c$, $Flair$, $T2$, to λ different values, resulting in I_i=($I_{Tic,i}$, $I_{Flair,i}$, $I_{T2,i}$) for image Img_i. In our experiments, $\lambda = 40$ based on empirical analysis on the training data. In this way, we can construct a 3D joint histogram with λ^3 distinct bins.

We define $H_{t,i}$ as the 3D histogram of image Img_i given the label of the pixel is t. In other words, $H_{t,i}(a,b,c) = Pr(I_i(p) = (a,b,c)|f_p = t)$. H_i is defined as the histogram of the whole imge Img_i.

Histogram Matching Following the nonrigid local image registration method proposed in [5], we register a subject 3D joint histogram H_i to the reference joint histogram H_j treating them as 3D images, resulting in a deformed histogram H'_i, and compute the transformation fields T^i_j. We explicitly register a testing image with each labeled training image separately and then integrate the set of resulting likelihood terms, described below.

Likelihood Estimation Given a testing image Img_x and a labeled training image Img_i, we estimate the likelihood $Pr_i(I(p)|f_p)$ for each voxel $p \in Img_x$ with Algorithm 1.

Algorithm 1 Likelihood Estimation

Input: Img_x, labelled image Img_i

1: Compute I_i, I_x with quantization
2: Compute H_i, H_x, $H_{t,i}$, $t \in$ Ł
3: With H_i, H_x, compute T^i_x
4: With T^i_x and $H_{t,i}$, compute the deformed Histogram, $H'_{t,i}$, $t \in$ Ł
5: $Pr_i(I(p)|f_p)$=$H'_{f_p,i}(I_x(p))$

We then integrate information from each training image as follows. Because different cases may have different kinds of tumor, but must have the same kind of gray matter, white matter and CSF, we define $Pr(I(p)|f_p)$ as the following

$$Pr(I(p)|f_p) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} Pr_i(I(p)|f_p) & \text{if } f_p \in \{B, C\} \\ \max_{i=1}^{n} Pr_i(I(p)|f_p) & \text{if } f_p \in \text{Ł} \setminus \{B, C\} \end{cases} . \quad (4)$$

3 Experiments

We have evaluated our method on BRATS 2013 Training data with a leave-one-out cross validation. In this data set, each voxel is labeled as necrosis, edema, non-enhancing tumor, enhancing tumor or everything else (normal brain and

non-data voxels). For each case, we use a small threshold to remove the non-data voxels. To improve the performance, for each case in the training data, we make a two class segmentation on the normal brain with k-means, and manually label them as C (cerebrospinal fluid) or B (white matter and gray matter). The runtime is about $0.6n + 4$ minutes for each case, where n is the number of labeled training cases. The evaluation of our method on the whole tumor (tumor+edema), tumor, and enhancing tumor is shown in Table 1.

	Dice (%)		Specificity(%)		Sensitivity(%)	
Structure	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Whole Tumor(Low-grade)	83.44	9.32	86.32	11.29	82.33	13.19
Tumor(Low-grade)	58.74	22.60	74.02	22.02	54.88	24,74
Enhancing Tumor(Low-grade)	51.79	13.01	61.29	9.37	49.17	15.70
Whole Tumor(High-grade)	83.39	9.92	84.84	12.76	83.87	12.61
Tumor(High-grade)	74.42	23.27	85.79	16.90	70.65	26.68
Enhancing Tumor(High-grade)	68.25	22.13	85.68	18.28	64.17	24.16

Table 1. Evaluation results of our method for low-grade and high-grade patients are shown in the table.

4 Summary

In summary, we have described and evaluated a method for tumor labeling in MRI that is based on a Markov random field defined over a supervoxel over-segmentation of the image volume. Voxel likelihood is computed by histogram matching over the full training data set. Our evaluation demonstrates the quality of our proposed method on the BRATS challenge data.

References

1. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE TPAMI*, 34(11):2274–2282, 2012.
2. Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE TPAMI*, 20(12):1222–1239, 2001.
3. J. J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha, and A. L. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE TMI*, 27(5):629–640, 2008.
4. S. Gould. Multiclass pixel labeling with non-local matching constraints. pages 2783–2790, 2012.
5. D.-J. Kroon and K. Slump. Mri modalitiy transformation in demon registration. pages 963–966, 2009.
6. S. M. Smith and J. M. Brady. Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.