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

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

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

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

## 1 Materials and methods

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

### 1.1 Dataset

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

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

### 1.2 Context-rich decision forest

Our detection and segmentation problem can be formalized as a multi-class classification of voxel samples into either background, edema or tumor core. This classification problem is addressed by a supervised method: discriminative random decision forest, an ensemble learner using decision trees as base learners. Decision trees are discriminative classifiers which are known to suffer from overfitting. A random decision forest [4] achieves better generalization by growing an ensemble of many independent decision trees on a random subset of the training data and by randomizing the features made available to each node during training [5].

**Forest training.** The forest has $T$ components with $t$ indexing each tree. The training data consists in a set of labeled voxels $\mathcal{T} = \{\mathbf{v}_k, Y(\mathbf{v}_k)\}$ where the label $Y(\mathbf{v}_k)$ is given by an expert. When asked to classify a new image, the classifier aims to assign every voxel $\mathbf{v}$ in the volume a label $y(\mathbf{v})$. In our case, $y(\mathbf{v}) \in \{0, 1, 2\}$, 2 for the tumor core, 1 for edema and 0 for background.

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

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

where $\theta$ is a function identifying the visual feature extracted at position $\mathbf{x}_k$. There are several ways of defining $\theta$, either as a local intensity-based average, local spatial prior or context-rich cue. These are investigated in more detail in the next section. The value of the extracted visual feature is thresholded by $\tau_{low}$ and $\tau_{up}$. The voxel $\mathbf{v}_k$ is then sent to one of the two child nodes based on the outcome of this test. Training the classifier means selecting the most discriminative binary test for each node by optimizing over $(\tau_{low}, \tau_{up}, \theta)$ in order to maximize the information gain on the input data partition [10], noted $\mathcal{T}_p$, defined as follows: $IG_{\tau_{low},\tau_{up},\theta}(\mathcal{T}_p) = H(\mathcal{T}_p) - H(\mathcal{T}_p|\{t^{\tau_{low},\tau_{up},\theta}(\mathbf{v}_k)\})$ where $\mathcal{T}_p \subset \mathcal{T}$, $H$ stands for the entropy.

Only a randomly sampled subset $\Theta$ of the feature space is available for internal node optimization, while the threshold space is uniformly discretized. The optimal $(\tau^*_{low}, \tau^*_{up}, \theta^*)$ is selected by exhaustive search jointly over the feature and threshold space. Random sampling of the features leads to increased internode and inter-tree variability which improves generalization. Nodes are grown to a maximum depth $D$. Another stopping criterion is to stop growing a node when too few training points reach it, i.e. when the information gain is below a minimal value $IG_{min}$.

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

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

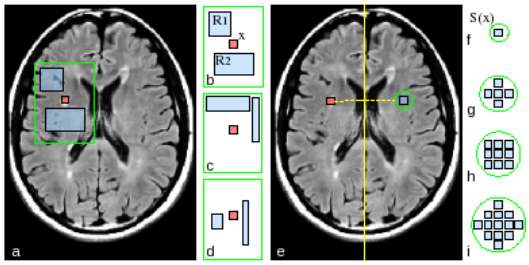


**Fig. 1. 2D view of context-rich features.** (a) A context-rich feature depicting two regions $R_1$ and $R_2$ with constant offset relatively to $\mathbf{x}$. (b-d) Three examples of randomly sampled features in an extended neighborhood. (e) The symmetric feature with respect to the mid-sagittal plane.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

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

### 1.3 Visual features

In this section, two kinds of visual features are computed: 1) local features: $\theta_c^{loc}(\mathbf{v}) = C_c(\mathbf{x})$ where $c$ indexes an intensity or a prior channel; 2) context-rich features comparing the voxel of interest with distant regions . The first context-rich feature looks for relevant 3D regions $R_1$ and $R_2$ to compare within an extended neighborhood: $\theta^{cont}_{c_1,c_2,R_1,R_2}(\mathbf{v}) = C_{c_1}(\mathbf{x}) - \frac{1}{vol(R_1 \cup R_2)} \sum_{\mathbf{x}' \in R_1 \cup R_2} C_{c_2}(\mathbf{x}')$ where $c_1$ and $c_2$ are two signal channels. The regions $R_1$ and $R_2$ are sampled randomly in a large neighborhood of the voxel $\mathbf{v}$ (cf. Fig. 1). The sum over these regions is efficiently computed using integral volume processing [6]. The second context-rich feature compares the voxel of interest at $\mathbf{x}$ with its symmetric counterpart with respect to the mid-sagittal plane, noted $S(\mathbf{x})$: $\theta_c^{sym}(\mathbf{v}) = C_c(\mathbf{x}) - C_c \circ S(\mathbf{x})$ where $c$ is an intensity channel. Instead of comparing with the exact symmetric $S(\mathbf{x})$ of the voxel, we consider, respectively, its 6, 26 and 32 neighbors in a sphere $\mathcal{S}$ (cf. Fig. 1), centered on $S(\mathbf{x})$. We obtain a softer version of the symmetric feature which reads: $\theta^{sym}_{c,\mathcal{S}}(\mathbf{v}) = \min_{\mathbf{x}' \in \mathcal{S}}\{C_c(\mathbf{x}) - C_c(\mathbf{x}')\}$.

## 2 Results

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

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

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

For quantitative evaluation, a three-fold cross-validation is carried out on this dataset: the forest is trained on $\frac{2}{3}$ of the cases and tested on the other $\frac{1}{3}$, this operation is repeated three times in order to collect test errors for each case. Note that the random forest is trained on the preprocessed data. Prediction on a single image lasts for approximately 10 minutes.

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

## Aknowledgments

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

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

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

## References


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