2013 IEEE 10th International Symposium on Biomedical Imaging:
From Nano to Macro
San Francisco, CA, USA, April 7-11, 2013

BRAIN TUMOR SEGMENTATION WITH SYMMETRIC TEXTURE AND SYMMETRIC INTENSITY-BASED DECISION FORESTS

Anthony Bianchi[1,2], *James V. Miller*[1], *Ek Tsoon Tan*[1], *Albert Montillo*[1*]
[1]GE Global Research, Niskayuna, NY, USA, [2]University of California Riverside, Riverside, CA USA

ABSTRACT

Accurate automated segmentation of brain tumors in MR images is challenging due to overlapping tissue intensity distributions and amorphous tumor shape. However, a clinically viable solution providing precise quantification of tumor and edema volume would enable better pre-operative planning, treatment monitoring and drug development. Our contributions are threefold. First, we design efficient gradient and LBPTOP based texture features which improve classification accuracy over standard intensity features. Second, we extend our texture and intensity features to symmetric texture and symmetric intensity which further improve the accuracy for all tissue classes. Third, we demonstrate further accuracy enhancement by extending our long range features from 100mm to a full 200mm. We assess our brain segmentation technique on 20 patients in the BraTS 2012 dataset. Impact from each contribution is measured and the combination of all the features is shown to yield state-of-the-art accuracy and speed.

Index Terms— Lesion segmentation, brain tumor, MRI, decision forest, texture, symmetry.

1. INTRODUCTION

There are many diseases that cause lesions in the brain, a primary example being brain tumors. Being able to automatically and accurately segment tumors allows for better treatment planning by reducing inter and intra operator error. Brain tumors have many properties that make them difficult to segment using MRI. Tumors appear with: low contrast, intersubject intensity variability, shape variability, and appear at variable locations. In any single MR modality the intensity values of the tumor will overlap substantially with the intensity values of the healthy brain tissue. Therefore to discriminate between healthy tissue and tumor, multiple MR modalities such as T_1, T_2, FLAIR and post-Gadolinium T_1 are acquired and a segmentation algorithm must integrate information from all channels.

Previously, approaches have been proposed to automatically segment brain tumors. Both generative and discriminative approaches have been proposed. In this paper a discriminative decision forest classifier is constructed. In such an approach, the posterior class probability $p(L(\boldsymbol{x}) = l|\{f_i(\boldsymbol{x})\}_1^N)$ of the label, *L,* for voxel $\boldsymbol{x}$ given *N* observed features $\{f_i\}_1^N$ is estimated directly. Our contributions are threefold: (1) We design fast texture features (both gradient based and 3D Local Binary Pattern texture called LBPTOP) and show how texture improves classification accuracy over intensity only features. (2) We present a method for quickly and automatically extracting a symmetry map and use it to design both symmetric texture and symmetric intensity features. We show these features further improve the results over non-symmetric features. Symmetric features take advantage of the fact that healthy brains are largely symmetric, and symmetry can be used with both appearance and texture features. (3) We implement long range features (*up to* 200mm) relative to typical brain dimensions (avg length 167mm) and demonstrate that using very long range features spanning the whole brain gives higher accuracy than local features in the immediate vicinity of the voxel to classify or medium range features (up to 100mm).

Recently, several approaches employing the decision forest have been proposed. Zikic et al. [9] use context aware features along with a generative tissue model. However, they do not take advantage of symmetry or texture. Menze et al. [5] and Geremia et al. [3] use context rich features along with symmetry features and spatial tissue priors for segmentation. While there are some commonalities with our method, there are some major differences. We implement texture features and our implementation of symmetry allows for variable neighborhood sizes and shapes; they use a point compared to a fixed region. This gives our approach more flexibility for estimating symmetry. Some texture features were used by Bauer et al. [1], but they do not take advantage of symmetry nor long range context. The top ranked methods in the BraTS challenge, [5] and [9], combine the decision forest with a generative model. This paper improves the decision forest, which lends itself to improving the overall performance of these methods and enabling higher lesion segmentation accuracy.

2. TECHNICAL APPROACH

2.1. Decision Forest Voxel Classifier

The decision forest has favorable properties including: it can produce maximum-margin boundaries, it is resistant to overfitting, and it performs intrinsic feature selection. This

*Corresponding author: Tel:(518)-387-4791, Email: montillo@ge.com

has made it a leading classification approach. It is an ensemble approach using the outputs of T separately trained decision trees. Each tree differs due to randomness injected during the training of each node. We inject randomness through the selection of a random subset of features and thresholds to try at each node, because this combination was shown [4] to give superior performance over bagging. The feature pool we use to randomly create a candidate set includes: modality, feature type and feature parameters (size and long range offset). During *training* at each node, the feature parameters that give the maximum information gain is found and stored with the node. During *testing*, these winning features are used to guide the input voxels to a leaf where the T class posteriors from the trees are averaged together to give a final, maximum a posteriori voxel label.

2.2. Long Range Contextual Features

The appearance of neighboring voxels gives information on how the observed voxel should be classified. We hypothesize that to correctly distinguish pathology from healthy tissue at a voxel, both short range (neighboring voxels) and very long range information will be useful. Many features, such as Haar- like features, use only the immediate neighborhood, therefore we implement many Haar-like features but rather than keeping the sub-rectangles in each box adjacent, we allow them to be separated by 1 to 200mm. The appearance features we use include the difference in the mean intensity of two cuboidal probe regions, as illustrated in Fig. 1A. This feature is the difference between the mean value from the two regions (R_1 and R_2) at vector offsets (Δ_1 and Δ_2) from the voxel P, and with an averaging box size B (mm) which can vary from 1 to 200mm in each dimension. Each of these features is a parameter which is explored by the forest by testing parameterized features at each node. With these intensity features described, we now turn our attention to our first contribution, the design of texture features.

2.3. Texture Features

Texture is a compact representation of a local neighborhood. In brain MRI, texture gives a description of the underlying tissue in a region. We use two types of texture features: gradient magnitude and Local Binary Patterns (LBP). We use the *gradient magnitude* in the axial, coronal, and sagittal directions (assuming basic intersubject alignment via affine registration) because this fundamental description of the local neighborhood of a voxel is fast to compute and, we hypothesize, discriminative for tumor segmentation. LBP features describe the texture at an observed voxel by using its intensity to threshold the voxel's local neighborhood. The thresholded neighborhood is then encoded as a binary number. This number is mapped into a uniform pattern space making it rotationally invariant and yielding 59 possible values. LBP is extended to an approximate 3D feature, by computing the 2D LBP in each of *t*hree *o*rthogonal *p*lanes (TOP) (axial, coronal, and sagittal in our case) hence the feature is called LBPTOP [8].

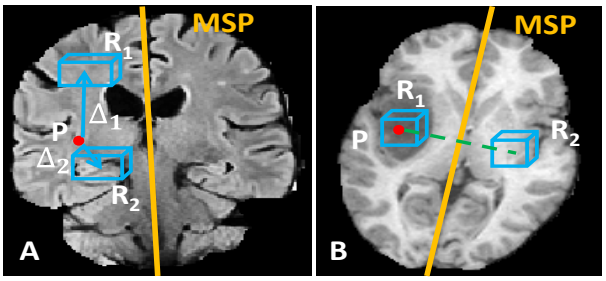

Figure 1: (A) Long range contextual features use information from around the brain. (B) Symmetric feature measuring the difference from the observed region and contralateral region.

Other true 3D extensions of LBP are possible, however LBPTOP is computationally attractive and affords the ability to select the plane for feature computation when data is acquired non-isometrically, as is the case in BraTS data.

To illustrate consider Fig. 2. Upper right shows the input intensity image with bright tumor and gray edema. The corresponding LBP images computed in the axial, coronal and sagittal directions are shown in the remaining quadrants. We observe LBP patterns are visibly correlated with the tumor and edema regions. We combine these texture maps with the long range context concept described previously to efficiently produce long range texture context.

2.4. Symmetry Features

A healthy brain exhibits bilateral symmetry causing a region reflected across the mid-sagittal-plane (MSP) to look similar, while areas affected by pathology appear different. Before symmetry features can be calculated, the MSP must be estimated. We achieve this by locating symmetric interest points which vote for a MSP as in [6]. Once the MSP is known, the contralateral reflection of every point can be estimated. The difference between the mean of the observed region and the mean of the reflective region is used, as illustrated in Fig. 1B. Averaging over the region helps account for mass effect and MSP estimation error. The size

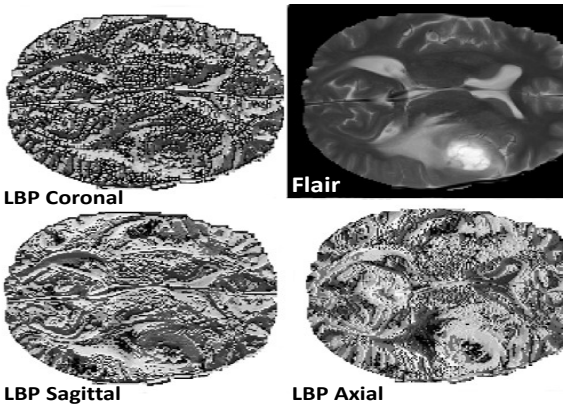

Figure 2: Example of the LBPTOP feature in the Flair modality. Note the edges of the edema and tumor are easier to distinguish in the LBP images.

of the averaging box is a random parameter in the training process. We further combine our texture measures with symmetry by computing the texture difference with the symmetric region. This symmetric texture contribution plays a large role in regions with significant texture (e.g. sulci).

3. EXPERIMENTAL RESULTS

3.1. Dataset

The dataset we use for experimentation is from the MICCAI 2012 Mutlimodal brain tumor segmentation (BraTS) challenge dataset. It is comprised of 20 real high grade (HG) glioma patients with the following MR modalities: T_1, T_2, FLAIR and post-Gadolinium T_1. We also use the 50 simulated HG and low grade (LG) BraTS cases. The data channels are co-registered, skull-stripped and resampled to 1mm isometric volume voxels. To make the intensities across the subjects more consistent, we mode-align (shift only) the intensity histograms across subjects per modality.

3.2. Effect of Feature Parameters

3.2.1. Maximum Box Size

In most decision forest implementations, only individual samples (e.g. voxels) are used as the features [7]. For brain lesion segmentation, however, individual voxels can be affected by noise or imaging artifact, therefore, to make features more robust we construct features from cuboidal regions of voxels. Fig. 3A shows the effect of increasing the *maximum* side length of the cuboidal regions from 3mm to 20mm with a minimum of 1mm on a side. The light blue curve shows the Dice overlap of the tumor+edema label, when using separate *real* HG glioma cases as training and testing through fivefold cross-validation. The red curve shows the Dice overlap for the 50 *simulated* HG and LG glioma cases. We observe that the simulated cases are easier to segment than the real HG cases. This is somewhat expected; tumors in the simulated data have slightly higher contrast to healthy tissue than in the real cases. More importantly, we observe that when varying just this parameter, i.e. allowing the forest to try larger and larger features at each node during training, the accuracy of the forest monotonically increases until an asymptote of roughly 0.62 Dice for HG glioma and 0.92 for simulated HG and LG tumors. This suggests that large 3D cuboidal features (up to 20mm on a side) are useful for discriminating healthy and pathological tissues.

3.2.2. Effect of Long Range Feature Distance

Traditionally, voxel classifiers are trained with local features which sample the target brain in the immediate vicinity (within several mm) of the voxel to be classified. However, we use very *long range* features which extend up to 200mm from the voxel to be classified, because even in the relatively constrained space of the intra cranial cavity, such long range features, which may span the whole brain, yield

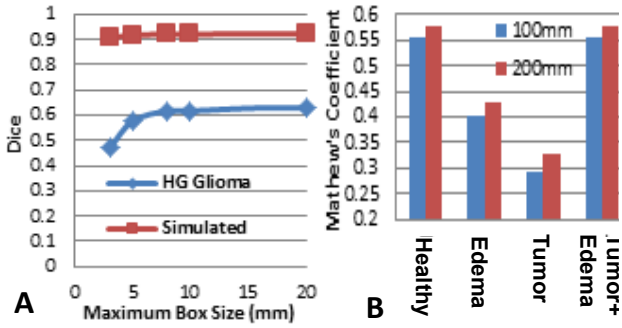

Figure 3: (A) Allowing the box size to range from a small (1mm) to large (20mm), rather than 1mm to 3mm, yields the optimum accuracy. (B) Allowing the long range context features to span 200mm yields the highest accuracy.

higher accuracy than local features. Fig. 3B shows the segmentation improvement. Using the real 20 HG datasets, the Mathew's coefficient (similar to dice) is of higher accuracy for 200mm than 100mm both when considering tumor+edema together as one class and when they are considered separately. While in [2] it was shown that a range of 200mm is beneficial to localize organs in *full body CT*, we show the nonobvious result that such long range features are also beneficial to segment *brain MRI*.

3.2.3. Qualitative Effect of Symmetry

Symmetry uses the patient as their own control. Fig. 4A, shows the input FLAIR image for subject 1 in column 1 and the true tumor+edema segmentation in light blue in column 2. Healthy labels are transparent. The tumor+edema segmentation using traditional, non-symmetric, intensity only features without texture is shown in column 3. We observe the result is substantially improved (column 4) by adding symmetric intensity features, which corrects false negatives and false positives (arrows, column 3).

3.2.4. Qualitative Effect of Texture

Texture features compactly represent the pattern of intensities in the local neighborhood around a voxel. Fig. 4B shows the input FLAIR image for subject 2 in column 1, while the true tumor+edema segmentation is in column 2. The tumor+edema segmentation using traditional, non-symmetric intensity features without texture is shown in column 3. The result is notably improved (column 4) when texture features are added. Texture corrects missed tumor voxels (false negatives) and false positives (arrows, column 3).

3.2.5. Qualitative Effect of Texture + Symmetry

Fig. 4C shows a slice from the same subject 3. We observe: (a) intensity only features yield many false negative voxels (red, column 3), (b) adding texture or symmetry corrects many of them (columns 4,5) and (c) *adding both* texture and symmetry (column 6) yields the best segmentation: few false positives (blue) and false negatives (red).

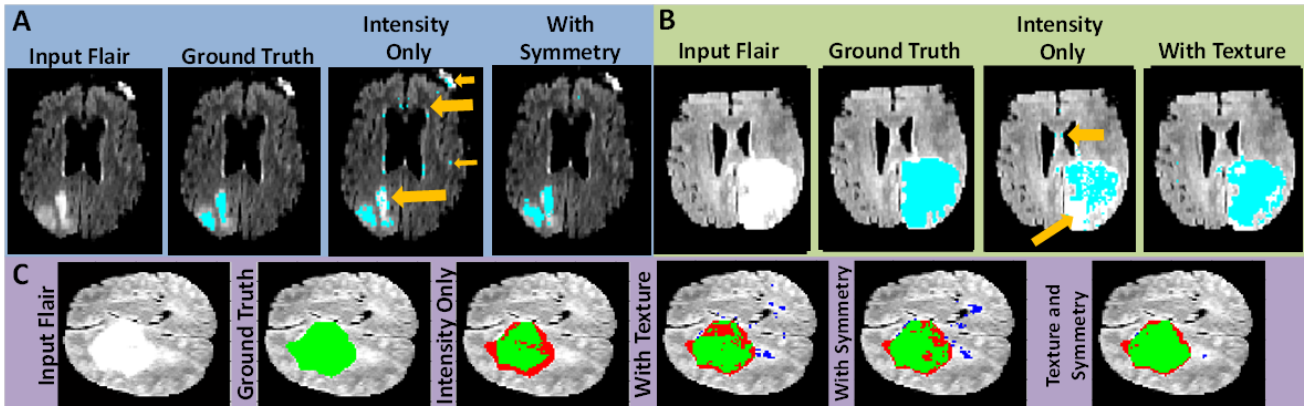

Figure 4: Qualitative results from texture and symmetry tests. A) When symmetry is introduced many false positives and negatives are corrected. B) Adding texture also corrects many false negatives. C) Adding symmetry *and* texture yields best segmentation (green-true positive, blue-false positive, red-false negative, clear-true negative).

3.2.6. Quantitative Feature Impact

In Fig. 5 we quantitatively assess the impact of texture, and combined texture and symmetry. These tests use just 8 of the 20 HG subjects to train the forest to rapidly detect relative changes in performance. As a baseline, single voxel (side=1mm) intensity features are used (dark blue bars). For tumor+edema segmentation a dice of 0.49 is attained. Allowing larger cuboidal features (side≤10mm) dice increases to 0.6 (light blue). Adding texture features increases dice to 0.625 (pink); while adding symmetric texture and symmetric intensity features increases dice to 0.655 (orange).

In subsequent experiments we doubled the number of training subjects from 8 to 16. This boosts intensity only dice for tumor+edema to 0.64 while the intensity+texture result increases to 0.696. All other parameters are held constant. We also performed these tests on the 25 synthetic BraTS subjects obtaining dice of 0.92 (red curve in Fig. 3A). Our dice scores for tumor+edema (0.696 real HG, 0.92 synthetic HG) compare to the best reported during the BraTS challenge: 0.7 to 0.8 for real HG and 0.9 for synthetic HG. Segmentations were performed rapidly requiring just 1.5 seconds per test brain volume on an Intel Core 2 Quad core CPU (2.4GHz) with 16GB RAM.

4. CONCLUSIONS

To the best of our knowledge, this paper is the first to show that in brain analysis an extremely long range of 200mm yields higher segmentation accuracy than medium range features (100mm). We also show that gradient and LBPTOP based texture features improve segmentation accuracy. We present a method for extending the texture and intensity features to symmetric texture and symmetric intensity features and show these features further improve classifier accuracy. The winning methods of the BraTS 2012 challenge used the decision forest as a central component. We suggest that combing our improvements to the decision forest with these hybrid generative-discriminative methods will enable new state of the art performance for brain lesion segmentation.

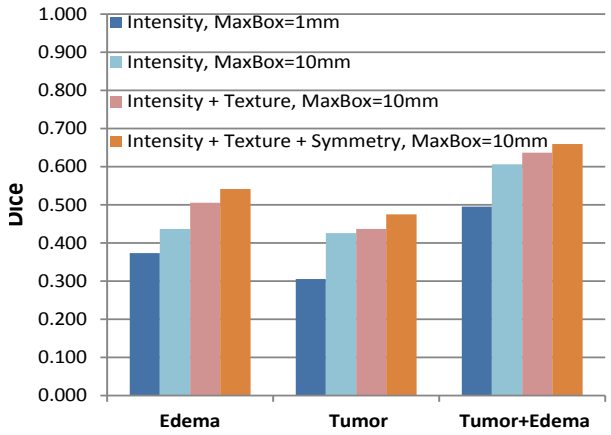

Figure 5: Performance monotonically increases with increased box size, adding texture, and adding symmetry + texture.

5. REFERENCES

[1] Bauer S., et al., "Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization." MICCAI BraTS Challenge, Nice, France 2012.

[2] Criminisi, A., et al. "Regression forests for efficient anatomy detection and localization in CT." MICCAI, Berlin, Germany, 2010, pp. 106–117.

[3] Geremia E., et al., "Spatial Decision Forest for Glioma Segmentation in Multi-Channel MRI." MICCAI BraTS Challenge, Nice, France 2012.

[4] Geurts P., et al. "Extremely randomized trees." Mach. Learn., 63(1):3–42,2006.

[5] Menze, B., et al., "Segmenting Glioma in Multi-Modal Images using a generative-Discriminative Model for Brain Lesion Segmentation.", MICCAI BraTS Challenge, Nice, France 2012

[6] Sun Y. and Bhanu B., "Reflection Symmetry-Integrated Image Segmentation." PAMI, IEEE Transactions on, 34(9):1827–1841, 2012

[7] Shotton J., et al. "TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context." Int. J. Comput. Vision, 81(1):2–23, 2009.

[8] Zhao G. and Pietikainen M. "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions." PAMI, IEEE Trans, 29(6): 915–928, 2007.

[9] Zikic D., et al., "Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR." MICCAI, Nice, France, 2012.