OUT-OF-ATLAS LABELING: A MULTI-ATLAS APPROACH TO CANCER SEGMENTATION

Andrew J. Asman[1], Bennett A. Landman[1,2]

[1]Electrical Engineering, Vanderbilt University, Nashville, TN, 37235
[2]Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21218

ABSTRACT

Conventional automated segmentation techniques for magnetic resonance imaging (MRI) fail to perform in a robust and consistent manner when brain anatomy differs wildly from expectations – as is often the case in brain cancers. We propose a novel out-of-atlas technique to estimate the spatial extent of abnormal brain regions by combining multi-atlas based segmentation with semi-local non-parametric intensity analysis. In a study with 30 clinically-acquired MRI scans of patients with malignant gliomas and 29 atlases of normal anatomy from research acquisitions, we demonstrate that this technique robustly identifies cancerous regions. The resulting segmentations could be used to study cancer morphometrics or guide selection/application/refinement of tumor analysis models or regional image quantification approaches.

Index Terms— Cancer Segmentation, Tumors, Multi-Atlas Segmentation, Out-of-Atlas Labeling

1. INTRODUCTION

Segmentation of brain structures from magnetic resonance imaging has become central to research in the structural underpinnings of the mind and how these activities evolve in development, aging and disease. Extensive image processing efforts have been directed to optimally exploit variations in intensity, shape, and inter-object relationships to precisely capture anatomical structures – such efforts have largely hinged on codifying expert knowledge and efficient intensity modeling based on healthy tissues. In contrast to the case of healthy anatomy, brain tumor location, morphometry, relative pixels intensity, and local texture are exceptionally variable among different patients, but global histograms and local intensity patterns on MRI are highly similar to those of healthy patients. Although recent efforts with learning classifiers and extensive training data have had some success in isolating tumors, brain cancers and other substantial disruptions of normal anatomy remain persistent problems when segmenting clinical populations given the difficulty in defining a comprehensive training set and overlap with normal populations [1].

Herein, we approach the challenge of segmenting tumors from the dual perspective – we seek to identify all tissues that appear normal with high probability and indicate the remainder as suspect. Previously, a single atlas of normal anatomy has been used for cancer segmentation by detecting outliers and abnormalities on patients of interest [2, 3]. Such approaches have been dependent upon accurate non-rigid registration of the atlas to the target [2, 3] and manual intervention [2]. Both of which are substantial impediments for clinical imaging. We build on the success of multi-atlas based segmentation methods [4, 5] to increase the robustness of atlas-based approaches.

We propose a novel "out-of-atlas labeling" approach which combines multi-atlas based segmentation [6] and semi-local intensity analysis [7, 8] to detect abnormalities in the human brain. This approach utilizes only affine registrations between the atlases and the target and is not dependent upon utilizing multiple modalities or for atlases to be acquired using a representative acquisition sequence.

2. THEORY

To construct the out-of-atlas estimate, a non-parametric estimate of the *expected* semi-local intensity distribution is constructed using the registered atlases and their associated labels and compared with an estimate of the *observed* semi-local intensity distribution constructed from the target. The label fusion result is refined by removing "normal labels" and performing a spatial regularization (Figure 1).

Problem definition: Consider an image of N voxels with unknown target labels $\boldsymbol{T}$, $T_i \in \{0, 1\}$ (i.e. 0: normal and 1: abnormal). R registered atlases (or "raters" in common fusion terminology) each provide an observed delineation of all N voxels exactly once. The set of labels on these atlases, **L**, represents the set of possible values that an atlas can assign to all N voxels. Let $\boldsymbol{D}$ be an N × R matrix that indicates the label decisions of the R registered at all N voxels where each element $D_{ij} \in \{0, 1, \dots, L-1\}$. Let $\tilde{\boldsymbol{I}}$ be another N × R matrix that indicates the associated post-registration atlas intensities for all R atlases and N voxels where $\tilde{I}_{ij} \in \mathbb{R}$. Let $\boldsymbol{I}$ be the N-vector target intensity where $I_i \in \mathbb{R}$. Note that to approximately match absolute values, atlas intensities are linearly normalized to the 5th and 95th percentiles of the target intensities.

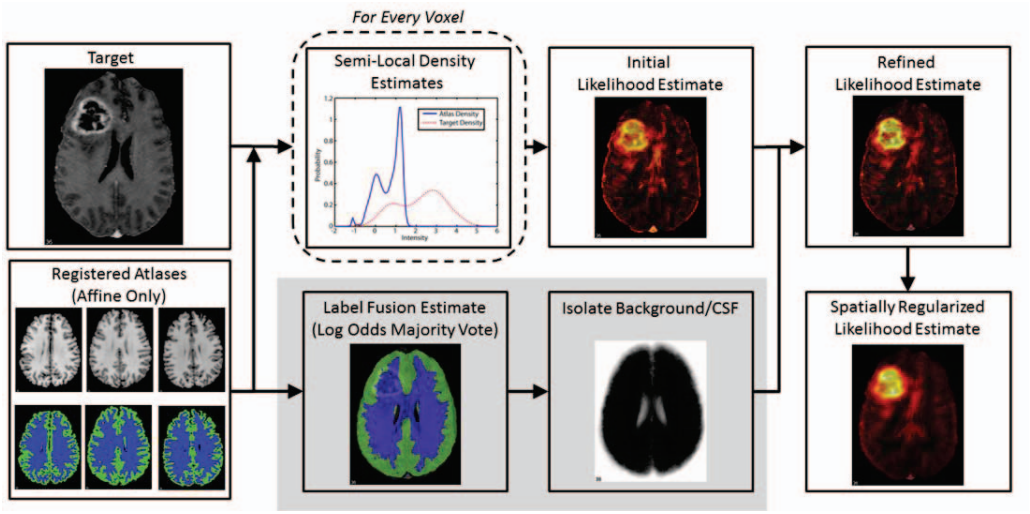


Fig 1. Flowchart demonstrating the multi-atlas based cancer segmentation technique presented in this paper. Non-parametric intensity distributions are constructed from the atlases and the target in order to construct an initial cancer likelihood estimate. The label fusion estimate is then used to refine the likelihood estimate. Finally, the refined estimate is then spatially regularized using a mean filter.

2.1 Semi-local Intensity Based Segmentation

The first step in our technique is to define the *expected* intensity distribution in the semi-local neighborhood surrounding each voxel on the target. To limit the influence of noise and bias in this estimate, this distribution is constructed by summing the non-parametric conditional probabilities for each atlas of observing all possible intensity values given the observed atlas labels. In other words, let

$$p(\gamma|D_{ij}=l)=\frac{\sum_{i:D_{ij}=l}K\left(\frac{\gamma-\tilde{I}_{ij}}{h}\right)}{h\sum_i\delta(D_{ij},l)} \quad (1)$$

be the conditional probability density function (PDF) of intensity for rater j and label l, where γ is all possible intensity values, K is a standard Gaussian kernel and h is the bandwidth associated with the kernel density approach. The bandwidth value was selected using the standard approach for an assumed underlying Gaussian distribution. Given (1), the *expected* PDF of the semi-local neighborhood surrounding each voxel is then given by

$$p_i^E(\gamma)=\frac{1}{Z_E}\sum_{i\in\mathcal{N}_i}\sum_j p(\gamma|D_{ij}) \quad (2)$$

where $\mathcal{N}_i$ is the semi-local neighborhood surrounding voxel i and Z_E is the partition function. Here, a $10\times10\times3$ neighborhood centered at the voxel of interest was utilized.

Next, we estimate the *observed* semi-local intensity distribution using the target intensities with a kernel density estimation technique similar to (1).

$$p_i^O(\gamma)=\frac{\sum_{i\in\mathcal{N}_i}K\left(\frac{\gamma-I_i}{h}\right)}{h|\mathcal{N}_i|} \quad (3)$$

where $|\mathcal{N}_i|$ is the cardinality of the semi-local neighborhood. The parameters associated with the kernel density approach are the same as (1).

To capture the difference between the *expected* and *observed* distributions we integrate over the difference between p_i^E and p_i^O where p_i^O is greater than p_i^E. Thus, the initial value for the cancer likelihood at voxel i is

$$\mathcal{L}_i^0=\frac{1}{Z_0}\int_{-\infty}^{+\infty}I\left(p_i^O(\gamma)>p_i^E(\gamma)\right)\left[p_i^O(\gamma)-p_i^E(\gamma)\right]d\gamma \quad (4)$$

where Z_0 is the partition function and I is the indicator function.

2.2 Multi-Atlas Based Label Fusion

In order to construct an estimate of the underlying segmentation, we use a LogOdds majority vote technique [6]. In this technique, the probability that the underlying segmentation given the observed labels is

$$p(T_i=l|\boldsymbol{D})=\frac{1}{Z_T}\sum_j e^{\rho\tilde{D}_{ij}^l} \quad (5)$$

where ρ is the decay coefficient (herein unity) and $\tilde{D}_{ij}^l$ is the value of the signed distance transform of the observed labels at voxel i for rater j and label l. Note that this probability is then normalized (via Z_T) such that $\sum_l p(T_i=l|\boldsymbol{D})=1$.

2.3 Combining Intensity and Multi-Atlas Based Approaches to Estimate Cancer Likelihood

The final estimate of the cancer likelihood is obtained by combining equations (4) and (5). Let $\boldsymbol{L}'$ be the set of labels for which observing cancer is possible. Herein, this is simply the set of labels excluding the CSF and background

labels. The final estimate for the cancer likelihood, $\mathcal{L}$, is then given by

$$\mathcal{L} = \frac{1}{Z_F}(K_m * [\mathcal{L}^0 \cdot p(T \in \boldsymbol{L}'|\boldsymbol{D})]) \quad (6)$$

where Z_F is the partition function, K_m is the kernel for a mean filter and $*$ is the convolution operator.

3. METHODS AND RESULTS

3.1 Experimental Setup

Thirty pre-operative T1-weighted brain MRI scans based on varied (but standard of care) imaging protocols with malignant gliomas were obtained in anonymous form under IRB approval. The resolution of each of the patient images was $1 \times 1 \times 3\,\mathrm{mm}^3$. The corresponding "ground truth" labels associated with each of the tumor regions were obtained from an experienced neuroanatomist. The normal atlases used in this experiment are a collection of 29 whole-brain segmentations with 5 labels per brain (in this case, gray matter, white matter, CSF, cerebellum and brain stem). Each of the brains (which are not gadolinium enhanced) is part of the OASIS [9] data set and the labels were acquired using a simplified version of the brainCOLOR protocol (http://www.braincolor.org/). All targets and atlases were skull-stripped using BET (FMRIB, Oxford, UK) and all registrations were performed using an affine registration with FLIRT (FMRIB, Oxford, UK). The technique presented in this paper was implemented in MATLAB (Natick, MA). In order to quantitatively and qualitatively assess the efficacy of this approach, we applied the proposed technique to all 30 cancerous brains using all 29 of the normal atlases.

3.1 Qualitative Results

Figure 2 presents three representative examples. As shown, for the first two examples (the top two rows) the cancer likelihood is able to detect the cancerous regions with high likelihood. To juxtapose these best-case scenarios, the third example presents a scenario where the cancer is not as accurately captured. As seen by the cancer likelihood estimate in the third row of Figure 3E, while the cancer is detected, additional regions are considered anomalous. This is likely due to several factors, including: i) the intensity profile of the cancerous region is not significantly different from the standard white matter regions, ii) the estimate of the CSF regions is not particularly accurate for this example, and iii) the cancerous region is relatively small and not easily captured by the proposed semi-local neighborhoods..

3.1 Quantitative Results

Second, quantitative analysis of the technique presented in this paper is considered. Using a simple thresholding technique on the cancer likelihood, the accuracy of our ability to segment the cancerous regions is considered. The results from this experiment can be seen in Figure 3. The receiver operating characteristic (ROC) curve for the presented technique can be seen in Figure 3A. The results indicate that, even using a simple thresholding technique, that the classifier is significantly better than chance. Additionally, the positive predictive value (PPV) and negative predictive value (NPV) associated with the thresholding-based segmentation can be seen in Figures 3B and 3C. The PPV results indicate that for high thresholds (e.g. above 0.8) that over 70% of voxels declared to be abnormal, represent regions that were labeled cancer by the

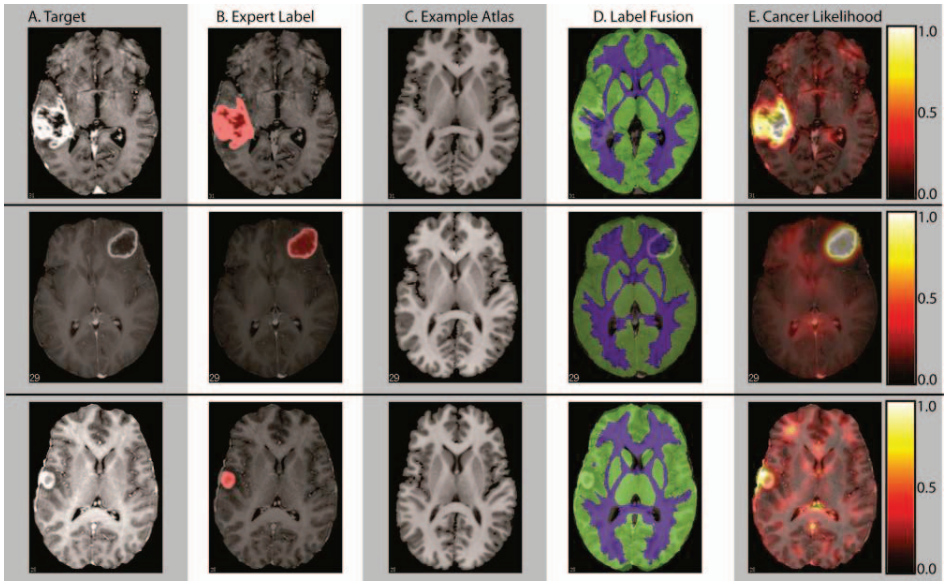


Fig 2. Example qualitative results. Three representative slices from three different brains are presented for comparison. For each example, the target, truth labels, an example atlas, the label fusion estimate and the cancer likelihood estimate are presented. Note the high quality likelihood estimate in the first two examples. On the third example, the cancer is detected, but it is not as well localized as in the first two examples.

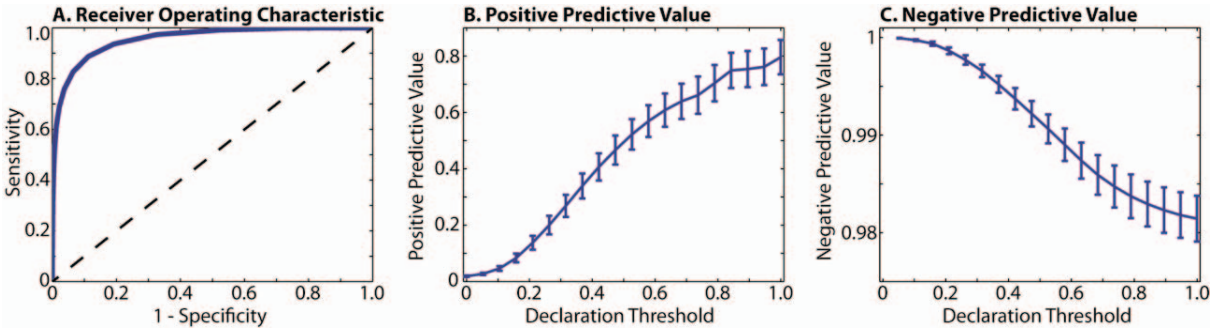


Fig 3. Quantitative results for the technique presented in this paper. The results in (A) show the ROC curve using a simple thresholding technique on the cancer likelihood estimate. The positive and negative predictive values are presented in (B) and (C). Note that for high thresholds, the positive predictive value is over 0.7, while the negative predictive value remains over 0.98.

experienced neuroanatomist. The results in Figure 3C show that with increasing thresholds the NPV is not dramatically affected, which indicates that the cancer likelihood estimate is accurately characterizing the cancerous voxels.

4. DISCUSSION

The ability to accurately and robustly estimate segmentations of brain cancer is a critical problem. Due to anatomical variability (e.g. size, shape, location, and intensity) of cancerous regions in the brain, commonly used segmentation techniques often fail to perform consistently. The proposed approach offers a high degree of sensitivity (>0.5 fraction by tumor volume) with nearly negligible false positive rate (Figure 3A). For example, a simple declaration threshold (e.g., abnormal > 0.8) provides a robust classifier, whose PPV is over 0.7 and NPV is over 0.98.We note that positive predictive values falls off rapidly if increasingly lenient thresholds are used, so a direct single-threshold for the estimated probability maps is unlikely to lead to precise tumor boundary definition.

Rather, this technique localizes abnormalities in brain tissues with a computationally simple algorithm with only weak dependencies on tumor morphometry or appearance. We envision that this approach would allow for application-specific algorithms to focus on regions of high tumor likelihood, which would (1) reduce the need to human intervention, (2) reduce the propensity for false positives. Alternatively, this approach would allow image quantification methods to focus on regions of relatively normal anatomy to (3) ascertain image quality or (4) model/adapt to image appearance characteristics.

In conclusion, out-of-atlas labeling shows great promise for robust and rapid identification of brain abnormalities. Continued innovation in local modeling (e.g., ventricular structure), optimality of the neighborhood size, and application of probabilistic abnormality classifiers promises to be a fascinating area of continued research. Lastly improved performance would be expected if the normal atlases were acquired in the same manner as the abnormal brains (e.g. gadolinium enhanced).

ACKNOWLEDGEMENTS: The authors are grateful to Dr. Reid Thompson for the cancer imaging data and Dr. Andy Worth of providing the labeled datasets. This work was supported in part by NIH/NINDS 1R01NS056307 and NIH/NINDS 1R21NS064534.

5. REFERENCES

[1] S. L. Cichosz, S. Vangsgaard, A. S. Jørgensen, K. E. Kannik, E. Steffensen, S. F. Eskildsen, Y. Xiao, and T. Amon, "Brain tumor segmentation from MRI: a comparative study," *CGVCVIP,* 2010.

[2] M. R. Kaus, S. K. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, and R. Kikinis, "Automated Segmentation of MR Images of Brain Tumors," *Radiology,* vol. 218, p. 586, 2001.

[3] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Medical Image Analysis,* vol. 8, pp. 275-283, 2004.

[4] T. Rohlfing, D. B. Russakoff, and C. R. Maurer, "Performance-Based Classifier Combination in Atlas-Based Image Segmentation Using Expectation-Maximization Parameter Estimation," *IEEE Transactions on Medical Imaging,* vol. 23, pp. 983-994, 2004.

[5] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *Neuroimage,* vol. 33, pp. 115-26, 2006.

[6] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A Generative Model for Image Segmentation Based on Label Fusion," *IEEE Transactions on Medical Imaging,* vol. 29, pp. 1714-1729, 2010.

[7] J. M. Lotjonen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *Neuroimage,* vol. 49, pp. 2352-65, Feb 1 2010.

[8] N. Weisenfeld and S. Warfield, "Learning Likelihoods for Labeling (L3): A General Multi-Classifier Segmentation Algorithm," *MICCAI 2011,* pp. 322-329, 2011.

[9] D. Marcus, T. Wang, J. Parker, J. Csernansky, J. Morris, and R. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience,* vol. 19, pp. 1498-1507, 2007.