

# ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images

Paul A. Yushkevich\*, Yang Gao<sup>†</sup>, Guido Gerig<sup>‡</sup>

\*Department of Radiology, University of Pennsylvania, Philadelphia, USA

<sup>†</sup>Department of Computer Science, University of Utah, Salt Lake City, USA

<sup>‡</sup>Department of Computer Science and Engineering, New York University, New York, USA

**Abstract**—Obtaining quantitative measures from biomedical images often requires *segmentation*, i.e., finding and outlining the structures of interest. Multi-modality imaging datasets, in which multiple imaging measures are available at each spatial location, are increasingly common, particularly in MRI. In applications where fully automatic segmentation algorithms are unavailable or fail to perform at desired levels of accuracy, semi-automatic segmentation can be a time-saving alternative to manual segmentation, allowing the human expert to guide segmentation, while minimizing the effort expended by the expert on repetitive tasks that can be automated. However, few existing 3D image analysis tools support semi-automatic segmentation of multi-modality imaging data. This paper describes new extensions to the ITK-SNAP interactive image visualization and segmentation tool that support semi-automatic segmentation of multi-modality imaging datasets in a way that utilizes information from all available modalities simultaneously. The approach combines Random Forest classifiers, trained by the user by placing several brushstrokes in the image, with the active contour segmentation algorithm. The new multi-modality semi-automatic segmentation approach is evaluated in the context of high-grade glioblastoma segmentation.

## I. INTRODUCTION

Segmentation is one of the most studied problems in the field of biomedical image analysis. Most research has focused on fully automatic approaches. Among the most successful and broadly applicable classes of fully-automatic approaches are statistical clustering techniques based on Gaussian mixture models or more complex models, including graph-theoretic approaches; atlas-based and multi-atlas label fusion approaches; and deformable modeling techniques with shape priors. The latter two classes of techniques can handle challenging segmentation problems where structures of interest are separated from adjacent tissues by subtle or unseen boundaries; but at the cost of needing substantial training data. However, in many problems, fully automatic segmentation is not available, or is not as accurate as manual segmentation by an expert with deep knowledge of underlying anatomy and pathology. Even in problems where fully automatic segmentation works well, manual segmentation is still often needed to train the automatic algorithms. However, with the increasing size and complexity of imaging datasets, manual segmentation is becoming more and more time consuming. For instance, in the past a typical MRI dataset for one subject would have consisted of one or two scans, often with thick slices. Segmentation of anatomical structures or lesions would have required tracing on a few slices. And

most studies were limited to a few tens of subjects. Today, MRI studies often recruit hundreds of subjects, contain four to eight different tissue contrasts (T1-weighted, T2-weighted, FLAIR, diffusion-weighted, etc.), and have sub-millimeter resolution images. For such studies, manual segmentation becomes increasingly impractical, due to the sheer volume of data to be processed manually. Furthermore, the task of fusing information from multiple imaging modalities during segmentation is challenging, and in contexts where multiple modalities are available, it is uncertain whether human experts fully leverage this rich information when tracing anatomical structures and lesions.

Semi-automatic segmentation approaches that allow the human expert to make “big picture” decisions, while relying on the computer to perform time-consuming repetitive tasks, may be the ideal strategy for complex problems where automatic segmentation is unreliable or unavailable and manual segmentation is impractical. Designing effective semi-automatic approaches is part science and part art, as powerful algorithms must be paired with efficient and intuitive interactive tools for user guidance. In the context of multi-modality imaging data, this is particularly challenging, as ways must be found to present the user with a wealth of data in a way that allows her to make executive guidance decisions without being overwhelmed by the complexity of the dataset.

ITK-SNAP (<http://itksnap.org>) is a general-purpose interactive tool for image visualization, manual segmentation, and semi-automatic segmentation. It was created in the early 2000s with the vision of an uncluttered tool that would be easy for computer non-experts to learn and would be focused almost entirely on the problem of segmentation, with the creep of additional non-segmentation and domain-specific features kept to a minimum. Since 2006, ITK-SNAP has been cited by over 1700 articles (source: Google Scholar), suggesting its wide adoption by the biomedical imaging community. However, until recently, ITK-SNAP only supported single-modality image segmentation, and its semi-automatic mode relied on the structures of interest being distinguished from surrounding tissues by different image intensity or strong image edges. The main contribution described in this paper is the development of a multi-modality segmentation capability in the new ITK-SNAP version 3.4 that uses machine learning to differentiate tissue classes based on a more complex range of features, including texture, location, and intensity.

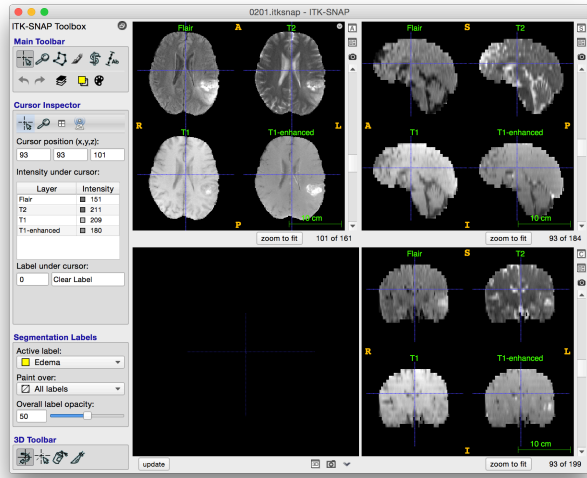


Fig. 1. ITK-SNAP tool (version 3.4) with four MRI modalities from a BRATS tumor dataset loaded.

This paper is organized as follows. Section II introduces the ITK-SNAP tool and the new multi-modality capabilities. Section III describes a validation study, in which ITK-SNAP was used to **segment brain tumors from multiple MRI contrasts simultaneously**. Section IV provides discussion and concludes the paper.

## II. ITK-SNAP TOOL OVERVIEW

### A. Image Viewing and Navigation

ITK-SNAP (Fig. 1) allows the user to load multiple image volumes in a variety of common file formats, including DICOM and NIFTI. The main program window shows three orthogonal views through the images, in axial, coronal and sagittal planes. Each view displays the lines along which it intersects with the other views as orthogonal lines, or a crosshair. Moving this crosshair in one slice view adjusts the slices visualized in other views. This “linked crosshair” concept provides a seamless way to navigate through 3D volumes, with **all three views focused on the same location in the 3D image**. Multiple image modalities can be visualized in three ways: a *tiled* layout, where the coronal, axial and sagittal slice views each display the same slice through all loaded modalities; a *thumbnail* layout, where one modality occupies most of each slice view, while others are shown as small thumbnails, clicking on which switches to that modality; and with selected modalities shown as semiopaque overlays shown on top of other modalities. Multiple color maps can be used to view images, and linear and non-linear contrast adjustment tools are provided.

### B. Manual Segmentation and Editing

ITK-SNAP represents segmentations by assigning a distinct integer value to each voxel in the image volume. Non-zero values correspond to different anatomical labels (e.g., tumor, edema). The segmentation is visualized as a semiopaque layer overlaid on the anatomical images, with

each label rendered using a different color, configurable by the user. An additional view is used to show the 3D surface rendering of each anatomical label in the segmentation.

Manual segmentation is performed using “**polygon**” and “**paintbrush**” tools. With the polygon tool, the user outlines polygonal region in any of the slice views. Polygons can be edited by moving vertices in the slice plane. Once accepted, the polygon is assigned the current label and integrated into the 3D segmentation volume. The paintbrush tool allows quick drawing and touch-up editing using the mouse, with masks of different shape and size. An adaptive paintbrush mask is also provided, wherein only the neighboring voxels similar in intensity to the voxel clicked on by the user are assigned the foreground label. Lastly, tools for manipulating segmentations in the 3D render window, such as partitioning into two labels using a cut plane, are provided.

### C. Semi-Automatic Segmentation

Semi-automatic segmentation in ITK-SNAP uses a two-stage pipeline. In Stage 1, the multiple image modalities used for segmentation are combined to produce a scalar “speed” image  $g$  with intensities in range  $[-1, 1]$ , defined as  $g(x) = P(x \in S) - P(x \in \Omega \setminus S)$ , where  $S$  is the structure of interest (foreground) and  $\Omega$  is the image domain. Multiple ways of obtaining these foreground/background probabilities are available, including thresholding of a given single modality; **Gaussian mixture model-based clustering**; and **supervised classification using Random Forests [1], [2], which is the focus of this paper**. Using the paintbrush or polygon tools, the user marks examples of different tissue classes present in the image. For each voxel labeled by the user, a vector of features is generated. **Features include the intensity of each image modality at the voxel**. Optionally, **intensities of the neighboring voxels** are also included as features, allowing implicit incorporation of texture information into the classifier. **The feature vector may also optionally include the coordinates of the voxels**, allowing spatial constraints to be incorporated into the classifier. The classifier is trained using these features, and then applied to each voxel in the image domain  $\Omega$ , yielding a posterior probability  $P^k(x)$  for each voxel  $x$  and each class  $k$ . The user then specifies which class, or which combination of classes, form the foreground  $S$ , and which classes form the background. **The sum of the posterior probabilities for the foreground/background is used to derive  $P(x \in S)$  and  $P(x \in \Omega \setminus S)$ , respectively**. Fig. 2 illustrates ITK-SNAP during the computation of the speed image using Random Forest classification.

Stages 2 of the pipeline involves active contour segmentation guided by the speed function  $g(x)$  and **user-placed initialization seeds**. The active contour algorithm [3], [4] is an iterative approach in which a parametric contour  $C$  representing the boundary of the segmented region undergoes evolution over time  $t$  according to the equation

$$\frac{\partial C}{\partial t} = [g(C) + \alpha \kappa_C] \mathcal{N}_C \quad (1)$$

where  $\kappa_C$  is the mean curvature of  $C$ ,  $\mathcal{N}_C$  is the unit outward normal of  $C$ , and  $\alpha$  is a scalar parameter. For

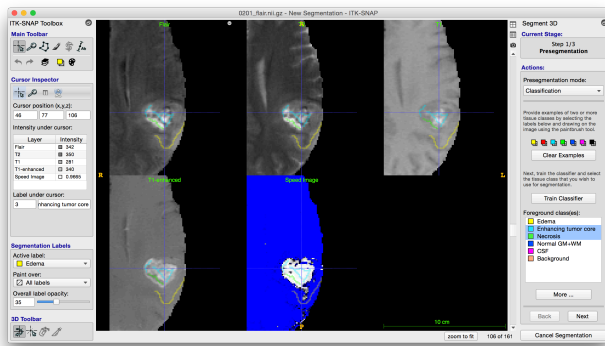


Fig. 2. ITK-SNAP tool during speed image computation in the semi-automatic segmentation mode. Examples of three labels are seen on the selected MRI slices, and the speed image is visualized using the blue (negative) to white (positive) colormap. The speed image is computed for the ET and necrosis as foreground labels, and other labels mapped to background.

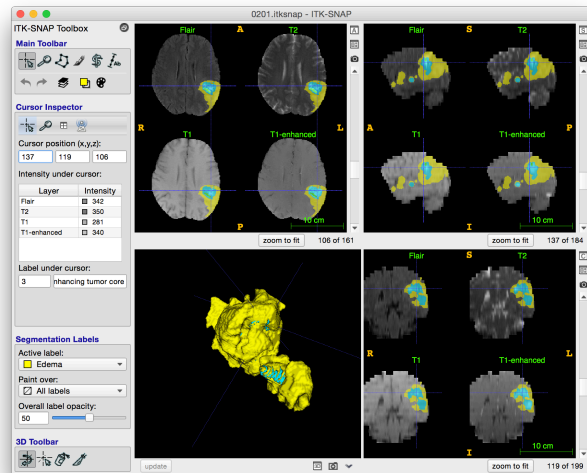


Fig. 3. ITK-SNAP tool after completion of the segmentation of the edema (yellow), ET (blue) and necrosis (green).

numerical stability and computational efficiency,  $C$  is represented implicitly as the zero level set of a function  $\phi$  defined on the image volume, and the evolution equation (1) is approximated as an evolution of  $\phi$  in the narrow region around the zero level set [5], [6]. In ITK-SNAP,  $C$  is initialized by the user placing one or more spherical seeds in the structure of interest. During evolution, the contour expands into the regions where  $g(x)$  is positive, i.e., where  $P(x \in S) > P(x \in \Omega \setminus S)$ , and contracts in regions where  $g(x)$  is negative. The curvature term, modulated by user-controllable parameter  $\alpha$ , imposes smoothness on the contour  $C$ . As the contour evolves, a visualization in 2D slices and 3D is provided in real time, allowing the user to stop evolution and reinitialize if necessary.

Stages 1 and 2 are repeated for different structures in the image. Often, segmenting different structures in the same image does not require re-training of the Random Forest

classifier, but simply the assignment of different labels or combination of labels to the foreground and background. Fig. 3 shows ITK-SNAP with a completed segmentation.

#### D. Software Architecture

ITK-SNAP contains multiple image processing pipelines written using the C++ Insight Toolkit library [7]. Pipelines are optimized for speed and interactivity. For instance, during Stage 1 of semi-automatic segmentation, Random Forest classifiers are applied only to parts of the 3D volume currently visible to the user, and are reapplied when the user moves the 3D crosshair. This provides almost real-time feedback when the user trains and re-trains the classifier using additional brushstrokes. Classification is applied to the entire 3D volume only when the active contour stage is entered. To reduce memory use, ITK-SNAP stores segmentation volumes, which are almost always largely contiguous, in compressed format using run-length encoding.

### III. EVALUATION

We evaluated ITK-SNAP segmentation using data from the 2013 MICCAI Brain Tumor Segmentation (BRATS) challenge. BRATS data consists of multi-modality brain MRI scans of patients with glioblastomas. For each patient, four co-registered scans are available: T1-weighted, T1-weighted contrast enhanced (T1CE) with gadolinium contrast agent; T2-weighted; and FLAIR. Evaluation focused on high-grade glioblastomas (HGG). Typically, 3-4 types of pathology are present in HGGs: edema (bright on T2 and FLAIR), enhancing tumor (ET; bright on T1CE), non-enhancing tumor (NET; abnormal in T2 but appears as normal gray/white matter in T1CE), and necrosis (dark in T1CE, T1). A semi-automatic protocol was developed for ITK-SNAP. It involves training Random Forest classifiers using 7 labels (edema, enhancing tumor, non-enhancing tumor, necrosis, normal brain tissue, cerebrospinal fluid, air/background). Active contour is performed in four sequential stages, by taking different combinations of labels to form the foreground. First, the combined (edema+NET+ET+necrosis) region is segmented; then (NET+ET+necrosis); then (ET+necrosis); and finally necrosis only. This sequence takes advantage of the fact that in most tumors necrosis lies within the ET, which is within the NET, which in turn is within the edema, and minimized the need to label structures with holes.

Users trained by performing segmentation using this protocol and comparing with the reference segmentations provided by the BRATS organizers. Once trained, they performed segmentation on BRATS “leaderboard” datasets for which the reference segmentation is not publicly available. One user performed repeat segmentations after a one week delay. The average segmentation time for one dataset was  $11.7 \pm 4.2$  minutes for rater 1 attempt 1, and  $10.4 \pm 5.5$  minutes for rater 1 attempt 2, and  $14.8 \pm 9.7$  minutes for rater 2.

Fig.4 reports the Dice similarity coefficient, a measure of relative overlap, between different segmentations of each dataset. Following the convention for reporting results adopted by the BRATS challenge, the four anatomical

Case	Intra-Rater (R1-1 vs R1-2)			Inter-Rater (R1-1 vs R2)			R1-1 vs. BRATS Reference		
	complete	core	enhancing	complete	core	enhancing	complete	core	enhancing
117	0.930	0.243	0.809	0.845	0.022	0.030	0.861	0.059	0.512
119	0.962	0.900	0.866	0.963	0.916	0.747	0.949	0.850	0.808
131	0.877	0.792	0.761	0.835	0.690	0.568	0.876	0.853	0.763
133	0.861	0.716	0.627	0.398	0.000	0.000	0.785	0.726	0.697
135	0.945	0.810	0.851	0.882	0.790	0.694	0.899	0.838	0.794
137	0.410	0.687	0.576	0.000	0.000	0.000	0.519	0.249	0.749
201	0.913	0.665	0.817	0.481	0.489	0.814	0.881	0.737	0.769
203	0.957	0.915	0.873	0.960	0.899	0.867	0.875	0.856	0.841
205	0.929	0.968	0.940	0.916	0.886	0.812	0.941	0.945	0.895
207	0.905	0.830	0.766	0.950	0.856	0.722	0.957	0.836	0.702
209	0.973	0.949	0.900	0.939	0.906	0.888	0.906	0.906	0.853
Average	0.878	0.770	0.799	0.743	0.587	0.558	0.859	0.714	0.762

Fig. 4. Quantitative evaluation of tumor segmentation accuracy by two raters using ITK-SNAP. Intra-rater reliability (repeat segmentation of same images by Rater 1), inter-rater reliability (segmentation of same images by Rater 1 and 2), and agreement between Rater 1 and the BRATS reference segmentation are reported for three labels used in the BRATS framework.

labels are combined into composite labels “complete” = (edema+NET+ET+necrosis), “core” = (NET+ET+necrosis), and “enhancing” = (ET). To put these numbers in the context, the top-ranked fully automatic method in the BRATS leaderboard dataset (as of March 2016) has average Dice overlap 0.84 for the complete label, 0.72 for the tumor core, and 0.62 for the enhancing label.

Note, however, that these numbers may not be compared directly, since BRATS reports average Dice for all leaderboard datasets (high-grade and low-grade glioblastomas combined), while our evaluation only examined a subset of 11 high-grade tumors. Overall, intra-rater reliability is generally high, with few exceptions where repeat segmentation grossly disagreed with the initial segmentation – interestingly this occurs in the same cases/labels when the agreement with the overall BRATS reference is low. Agreement between two raters was lower, indicating that there were some cases where the raters disagreed on what constitutes the tumor. In one case (137) rater 2 did not find a tumor.

#### IV. DISCUSSION

The evaluation demonstrates the feasibility of quickly segmenting anatomical structures using multiple imaging modalities simultaneously using ITK-SNAP. While the results reported here cannot be directly compared with the results of automatic segmentation reported by the BRATS challenge [8], the overall magnitude of the Dice scores suggests that the semi-automatic segmentation is competitive with the best of the fully automatic methods. While fully automatic methods that performed well in the BRATS challenge must be specifically retrained for each new anatomical domain and imaging modality, semi-automatic segmentation is easily adaptable to new domains, and the algorithms used in ITK-SNAP employ no domain-specific knowledge. This makes ITK-SNAP an excellent tool for accelerating segmentation in complex problems for which reliable automatic segmentation is not yet available.

The main contribution in this paper is the extension of single modality segmentation in the previously reported version of ITK-SNAP [9] to a pipeline that combines user-guided multimodality preprocessing and level set object

segmentation where information from multiple channels is used jointly. While the Random Forest [1], [2] and level set [3], [4], [10] algorithms are widely known and broadly used in image segmentation, the concept of combining them in the context of semi-automatic multi-modality image segmentation tool is innovative. To our knowledge, similar capability is not available in other semi-automatic software tools available today. For instance, the 3D Slicer software includes various modules for semi-automatic 3D segmentation, including the GrowCut module [11], which has been evaluated for glioblastoma segmentation. However, [11] use a single contrast-enhanced T1-MRI modality, rather than multiple modalities, and focus on the enhancing tumor and necrosis, but not edema.

#### ACKNOWLEDGEMENTS

ITK-SNAP development is supported by NIH grants R01 EB014346 and R01 EB017255. We thank the tool’s many contributors ([itksnap.org/credits.php](http://itksnap.org/credits.php)).

#### REFERENCES

- [1] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] A. Criminisi, J. Shotton, and E. Konukoglu, “Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 7, no. 2–3, pp. 81–227, 2012.
- [3] V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” *Int J Comput Vision*, vol. 22, pp. 61–79, 1997.
- [4] S. C. Zhu and A. Yuille, “Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 9, pp. 884–900, 1996.
- [5] J. A. Sethian, *Level set methods and fast marching methods*. Cambridge University Press, 1999.
- [6] A. E. Lefohn, J. M. Kniss, C. D. Hansen, and R. T. Whitaker, “A streaming narrow-band algorithm: interactive computation and visualization of level sets,” *IEEE T Vis Comput Gr*, vol. 10, no. 4, pp. 422–433, 2004.
- [7] T. S. Yoo and M. J. Ackerman, “Open source software for medical image processing and visualization,” *Commun. ACM*, vol. 48, no. 2, pp. 55–59, 2005.
- [8] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, Ç. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, “The multimodal brain tumor image segmentation benchmark (brats),” *IEEE Trans Med Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct 2015.
- [9] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, Jul 2006.
- [10] J. Sethian, “A fast marching level set method for monotonically advancing fronts,” in *Proc. Nat. Acad. Sci.*, vol. 93, 1996, pp. 1591–1595. [Online]. Available: [citeseer.ist.psu.edu/sethian95fast.html](http://citeseer.ist.psu.edu/sethian95fast.html)
- [11] J. Egger, T. Kapur, A. Fedorov, S. Pieper, J. V. Miller, H. Veeraraghavan, B. Freisleben, A. J. Golby, C. Nimsy, and R. Kikinis, “Gbm volumetry using the 3d slicer medical image computing platform,” *Sci Rep*, vol. 3, p. 1364, 2013.