

INTERACTIVE SEGMENTATION OF MR IMAGES FROM BRAIN TUMOR PATIENTS

Stefan Bauer^{1,2}, *Nicole Porz*^{2,3}, *Raphael Meier*¹,
*Alessia Pica*⁴, *Johannes Slotboom*², *Roland Wiest*², *Mauricio Reyes*¹

¹ Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland

²Department of Neuroradiology, Inselspital, Bern University Hospital, Switzerland

³Department of Neurosurgery, Inselspital, Bern University Hospital, Switzerland

⁴Department of Radiation Oncology, Inselspital, Bern University Hospital, Switzerland

ABSTRACT

Medical doctors often do not trust the result of fully automatic segmentations because they have no possibility to make corrections if necessary. On the other hand, manual corrections can introduce a user bias. In this work, we propose to integrate the possibility for quick manual corrections into a fully automatic segmentation method for brain tumor images. This allows for necessary corrections while maintaining a high objectiveness. The underlying idea is similar to the well-known Grab-Cut algorithm, but here we combine decision forest classification with conditional random field regularization for interactive segmentation of 3D medical images. The approach has been evaluated by two different users on the BraTS2012 dataset. Accuracy and robustness improved compared to a fully automatic method and our interactive approach was ranked among the top performing methods. Time for computation including manual interaction was less than 10 minutes per patient, which makes it attractive for clinical use.

1. INTRODUCTION

Segmentation of tumor-bearing brain images is an important pre-requisite for diagnosis as well as for planning and monitoring the treatment of brain tumor patients in clinical practice. Manual segmentation is tedious and time-consuming and therefore not generally used. Additionally, it suffers from intra- and inter-expert variability. On the other hand, automatic segmentation of brain tumor images has made significant improvements over the last few years [1], but due to the highly complex scenario of tumor appearance the results are still not accurate and robust enough in some cases. This motivates the idea to combine automatic segmentation with manual interaction. Deeley et al. [2] showed that manual post-editing of automatically generated segmentations reduces not only operator time, but also inter-expert variations compared

to completely manual segmentations. Recently, also Heckel et al. [3] argued in favor of the need for more efficient segmentation editing tools in clinics.

In the context of semi-automatic brain tumor segmentation Birkbeck et al. [4] suggested to employ user inputs for a graph-cut segmentation of the tumor area, however they considered only binary tumor / non-tumor segmentation. Hamamci et al. [5] combined cellular automata with graph-cut methods to finally evolve a level-set towards the tumor boundaries for enforcing smoothness. The user had to provide input information by brush strokes to separate background from enhancing tumor and the necrotic region. Recently, Zhao et al. [6] introduced another semi-automatic approach to segment tumor core and edema. They required user input by drawing rectangles or curves on one single slice. The information collected from this input was propagated to the complete image volume based on structural trajectories in combination with constrained Markov random fields.

In this contribution, we propose to integrate smart interactive post-editing into an existing fully automatic segmentation method for tumor-bearing brain images. In contrast to other methods, the interactive part comes only after a first automatic segmentation step in order to decrease observer variability. The tumor is segmented into 4 different sub-compartments (necrosis, non-enhancing tumor, enhancing tumor, edema).

2. METHODOLOGY

We start with a fully automatic segmentation of the multi-modal volumetric images using [7], which can be corrected by the user where necessary. The fully automatic segmentation is based on integrating a decision forests classifier for voxel-wise tissue classification using a 40-dimensional feature vector \mathbf{x}_i (multi-modal intensities, textures and gradient statistics) with conditional random field (CRF) regularization for enforcing spatial smoothness in an energy minimization scheme (more details can be found in [7]). This framework can be extended, allowing us to incorporate additional information from user corrections in an elegant way similar to the

This work was partially supported by the Swiss Cancer League, the Bernese Cancer League and the Swiss National Science Foundation.
Corresponding author: stefan.bauer@istb.unibe.ch

“Grab-Cut” algorithm [8]. Figure 1 shows a schematic of the pipeline of the proposed interactive segmentation method.



Fig. 1. Pipeline of the interactive segmentation algorithm. Manual corrections are performed after a first fully automatic segmentation stage and incorporated into the final result in a second segmentation update stage.

The automatic segmentation, i.e. the first stage of the presented method, produces a segmentation of the tumor into 4 sub-compartments as mentioned earlier. This result is shown to the user as a color overlay, together with the original $T_{1\text{contrast}}$ image and the FLAIR image. The decision forest classifier generates posterior probabilities for each voxel, indicating how reliable the assigned tissue class is. This information is used to create a confidence map, which can guide the user in the manual correction after the first segmentation stage. Manual corrections by brush strokes will be most beneficial in regions of low confidence of the classifier. The user can apply manual corrections using 5 different labels: 4 labels for the different tumor sub-compartments and one label for healthy tissues. As we are mostly interested in a correct segmentation of the tumor and not so much in a very accurate segmentation of the healthy tissues into cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM), we employ only one healthy tissue class for the manual corrections in order to make the interactive part easier and faster for the user. It is important to mention that approximate quick brush strokes by the user are sufficient to indicate errors of the automatic segmentation. It is not necessary to perform a very accurate manual segmentation at this point.

In the next stage of the algorithm, an energy minimization method according to equation (1) is used to perform the final segmentation, integrating the manual user corrections. This corresponds to applying the random field regularization part of the fully automatic segmentation stage once again, but with some modifications, which are described in the following.

$$E = \sum_i V(y_i, \mathbf{x}_i) + \sum_{ij} W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) \quad (1)$$

The first term in equation (1) corresponds to the voxel-wise singleton potentials and the second term to the pairwise potentials, enforcing a regularization in the neighborhood of each voxel.

$$V(y_i, \mathbf{x}_i) = \alpha \cdot p(\tilde{y}_i | \mathbf{x}_i) \cdot (1 - \delta(\tilde{y}_i, y_i)) \quad (2)$$

In equation (2), $p(\tilde{y}_i | \mathbf{x}_i)$ is the probabilistic output from the initial fully automatic decision forest classifier, \tilde{y}_i is the label before, and y_i the label after the final regularization stage, \mathbf{x}_i is the feature vector. δ is the Kronecker- δ function. If

a label has been manually edited, this new manual label is selected for \tilde{y}_i , otherwise the output label from the fully automatic segmentation stage is kept for \tilde{y}_i . α is a coefficient to ensure that manual corrections are preserved during the random field regularization, which is set to a high value if the voxel has been manually edited, otherwise it is set to 1.

$$W(y_i, y_j, \mathbf{x}_i, \mathbf{x}_j) = w_s(i, j) \cdot (1 - \delta(y_i, y_j)) \cdot \exp\left(\frac{-\text{PCD}(\mathbf{x}_i, \mathbf{x}_j)}{2 \cdot \bar{x}}\right) \cdot D(y_i, y_j) \quad (3)$$

Equation (3) ensures that not only the voxels, which have been actually manually edited, are changed, but that these manual edits also have an effect on the surrounding neighborhood (a 6-neighborhood is used in 3D). This is enforced by the pairwise terms. $w_s(i, j)$ is a weighting factor which depends on the voxel spacing in x,y,z-direction, $(1 - \delta(y_i, y_j))$ penalizes different labels of adjacent voxels, $\exp\left(\frac{-\text{PCD}(\mathbf{x}_i, \mathbf{x}_j)}{2 \cdot \bar{x}}\right)$ ensures that regularization is stronger in regions of similar intensity and $D(y_i, y_j)$ integrates prior knowledge about the likelihood of tissue adjacencies (more details about the individual terms can be found in [7]).

The case at hand is slightly more complicated than the manual post-editing discussed in [8] for binary foreground / background segmentation in natural 2D images. Here we have 7 different labels instead of 2, therefore ruling out one label does not necessarily directly yield the label to be selected. It can happen that the manual correction produces only a “background” label (in this case e.g. “healthy”), but it is unknown if it should be CSF, GM or WM. For these few undefined voxel labels after manual correction, a quick reclassification using the decision forest classifier is performed under the condition that only healthy tissue classes are allowed.

Finally, after manual editing, the energy minimization algorithm in (1) is solved using [9] to yield an updated segmentation result with improved accuracy. This saves a substantial amount of time compared to running the complete classification and regularization algorithm again after manual editing and offers the possibility to provide the user with the final result very quickly.

3. RESULTS

The supervised classifier was trained on the training dataset from the MICCAI BraTS 2012 challenge¹ and the method was evaluated on the corresponding testing dataset. The testing dataset contains multi-modal MR images (T_1 , $T_{1\text{contrast}}$, T_2 , FLAIR) from 11 high-grade glioma patients and 4 low-grade glioma patients. The images were already skull-stripped and registered, therefore no additional pre-processing was necessary. We compared the results of a fully automatic

¹<http://www2.imm.dtu.dk/projects/BRA TS2012/>

algorithm (this corresponds to the first stage of the presented method, which is basically the same approach as described in [7]) with the proposed interactive method, where the manual editing was performed once by a neurosurgeon and once by an experienced researcher in brain tumor imaging. The users were presented with an overlay of the automatic segmentation on the $T_{1\text{contrast}}$ and FLAIR image along with the confidence image (see figure 2 and 3). Corrections could be performed on a separate $T_{1\text{contrast}}$ image. The users were asked to make approximate corrections only and to spend no more than 5 minutes for the manual editing of one patient dataset in order to guarantee a realistic clinical scenario.

Figure 2 shows results on an axial slice of patient HG-0130. It can be seen that the false tumor positives in the frontal region are effectively suppressed after manual corrections (user brush strokes in cyan and red). The part of the tumor core, which is initially wrongly classified as edema, is later corrected to be necrotic tissue. Especially for the necrotic region it can also be seen that it is sufficient to do rough corrections. The neighborhood of the corrected voxels will be re-classified and corrected by the random field segmentation, which is run after the manual editing.

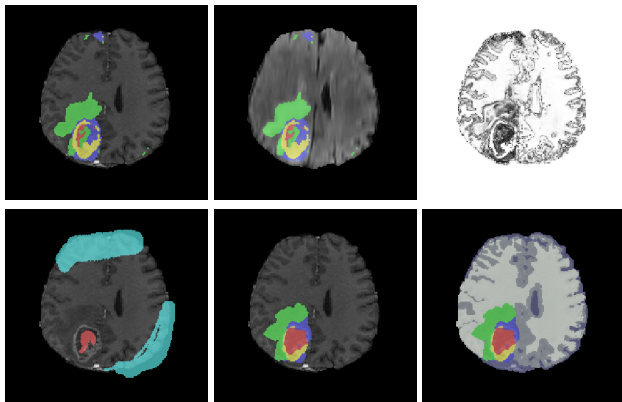


Fig. 2. Example patient BRATS-HG-0130. Upper row from left to right: overlay of the fully automatic segmentation result on the $T_{1\text{contrast}}$ image and the FLAIR image, and the confidence map image (brighter means more confidence, darker less confidence). Lower row from left to right: brush strokes with manual corrections, final result of the interactive segmentation algorithm (only tumor tissues shown), final result of the interactive segmentation algorithm (all tissues shown). Color code: red: necrosis, green: edema, blue: non-enhancing tumor, yellow: enhancing tumor, cyan: brush stroke for healthy tissues.

Figure 3 shows results on an axial slice of patient LG-0103. A subdural hematoma cavity that is wrongly classified as being tumor is corrected after user intervention. In patients with cavities it is difficult to obtain a correct segmentation result without manual interaction.

Quantitative results in terms of Dice overlap coefficient

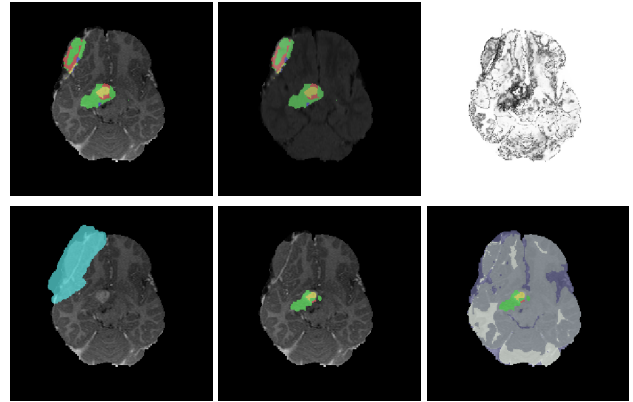


Fig. 3. Example patient BRATS-LG-0103. Upper row from left to right: overlay of the fully automatic segmentation result on the $T_{1\text{contrast}}$ image and the FLAIR image, and the confidence map image (brighter means more confidence, darker less confidence). Lower row from left to right: brush strokes with manual corrections, final result of the interactive segmentation algorithm (only tumor tissues shown), final result of the interactive segmentation algorithm (all tissues shown). Color code: red: necrosis, green: edema, blue: non-enhancing tumor, yellow: enhancing tumor, cyan: brush stroke for healthy tissues.

with the ground-truth segmentation were obtained from the BraTS2012 online evaluation tool, details are listed in table 1. For both users overlap improved for most compartments after corrections. Interestingly, the enhancing compartment showed a decrease in Dice overlap after manual corrections for both users. Most likely this is due to the fact that enhancing structures are usually very thin and the users did not spend sufficient time to capture these thin structures with their manual corrections very accurately. Another reason might be that the conditional random field segmentation, which is run after manual corrections, is based on a graph-cut method and graph-cut methods are known to have a “shrinking bias” that is most prominent for thin structures.

The difference in overlap with the ground-truth between the semi-automatic segmentations of the two different users was not statistically significant ($p > 0.1$ for every sub-compartment, Wilcoxon signed-rank test). Furthermore, we found that the mutual overlap between the semi-automatic segmentations of both users, also expressed as Dice coefficient, was comparatively high (except for the low grade gliomas where results were dominated by one ambiguous case with a very small tumor, see also table 1). This can be seen as an indication that post-editing of an automatic segmentation can reduce inter-observer variability compared to completely manual segmentations. As a comparison, Mazzarra et al. [10] found much higher values for inter-observer variability in completely manual segmentations (different image data used).

Table 1. Quantitative results expressed as Dice similarity coefficient. HG stands for high-grade, LG for low-grade gliomas. We report results for the fully automatic segmentation versus ground-truth (auto-GT), the interactive segmentation by a neurosurgeon versus ground-truth (NS-GT) and the interactive segmentation by an experienced researcher versus ground-truth (ER-GT) for 3 different tumor compartments: complete tumor, tumor core and enhancing tumor. Additionally, we also report the Dice coefficient of the neurosurgeon versus the experienced researcher after interactive segmentation (NS-ER), in this case the neurosurgeon segmentation is considered as the ground-truth.

		Complete	Core	Enhancing
auto-GT	HG	0.72±0.15	0.56±0.20	0.53±0.25
	LG	0.35±0.39	0.27±0.41	0±0
NS-GT	HG	0.77±0.14	0.58±0.31	0.48±0.21
	LG	0.51±0.38	0.31±0.40	0.25±0.50
ER-GT	HG	0.78±0.11	0.61±0.27	0.51±0.22
	LG	0.69±0.16	0.34±0.42	0.25±0.50
NS-ER	HG	0.92±0.06	0.86±0.12	0.71±0.30
	LG	0.64±0.44	0.43±0.50	0.19±0.38

On the BraTS2012 testing dataset the algorithm was ranked fourth among all submitted segmentation methods according to the online table (the ranking includes both fully and semi-automatic methods and is updated continuously). As this is a method which is based on user editing, better results could be achieved at the cost of allowing more time for user corrections.

When analyzing the manual corrections, it could be shown that they were mostly performed in regions where the automatic classifier had less confidence (posterior probability of 0.84 in regions with manual corrections versus posterior probability of 0.89 in other regions). This shows that manual intervention is especially useful in regions where it is difficult for the automatic classification method to distinguish individual tissue compartments.

Computation time after manual editing was only slightly more than half of the original computation time for automatic segmentation (on average 102 seconds after user editing versus 176 seconds for the initial fully automatic segmentation).

4. DISCUSSION & CONCLUSIONS

We proposed an interactive method for semi-automatic segmentation of brain tumor images. An initial fully automatic segmentation, together with a confidence map, is shown to the user who can make corrections where necessary. These corrections are integrated into the segmentation in a second stage using a fast conditional random field segmentation approach.

The users experienced the manual editing as being quick, easy and useful for having the possibility to make corrections of an automatic segmentation interactively.

Total computation time, including automatic segmentation, manual intervention and final segmentation step after manual intervention was less than 10 minutes per patient. Overlap in terms of Dice similarity coefficient improved after manual intervention and there was no statistically significant difference between two observers. This shows that an interactive method for brain tumor segmentation has the potential to achieve improved accuracy and robustness over automatic methods within a reasonably short time, while maintaining a low inter-observer variability. The approach can be useful for performing quick tumor segmentation and volumetry of clinical patients. Yet it has to be acknowledged that while an interactive method might be more accurate and robust, the complete objectiveness of a fully automatic method is lost. It depends on the use-case what is more important.

We are planning to integrate the proposed interactive segmentation method as an option into our BraTumIA software suite for brain tumor image analysis.

References

- [1] 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*, vol. 58, no. 13, pp. R97–R129, 2013.
- [2] M. A. Deeley, A. Chen, J. Noble, A. Cmelak, E. Donnelly, A. Malcolm, L. Moretti, J. Jaboin, K. Niermann, E. S. Yang, D. S. Yu, and B. M. Dawant, "Segmentation editing improves efficiency while reducing inter-expert variation and maintaining accuracy for normal brain tissues in the presence of space-occupying lesions." *Physics in medicine and biology*, vol. 58, no. 12, pp. 4071–4097, May 2013.
- [3] F. Heckel, J. H. Moltz, C. Tietjen, and H. K. Hahn, "Sketch-Based Editing Tools for Tumour Segmentation in 3D Medical Images," *Computer Graphics Forum*, 2013.
- [4] N. Birkbeck, D. Cobzas, M. Jagersand, A. Murtha, and T. Kesztyues, "An interactive graph cut method for brain tumor segmentation," *2009 Workshop on Applications of Computer Vision (WACV)*, 2009.
- [5] A. Hamamci, N. Kucuk, K. Karaman, K. Engin, and G. Unal, "Tumor-cut: Segmentation of Brain Tumors on Contrast Enhanced MR Images for Radiosurgery Applications." *IEEE transactions on medical imaging*, 2011.
- [6] L. Zhao, W. Wu, and J. J. Corso, "Semi-Automatic Brain Tumor Segmentation By Constrained MRFs using Structural Trajectories," in *MICCAI*. Nagoya: Springer LNCS, 2013, p. preprint.
- [7] 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 *Miccai Brats Workshop*, 2012.
- [8] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut," in *ACM SIGGRAPH 2004 Papers on - SIGGRAPH '04*, vol. 23, no. 3. New York, New York, USA: ACM Press, 2004.
- [9] N. Komodakis, G. Tziritis, 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*, vol. 112, no. 1, pp. 14–29, 2008.
- [10] G. P. Mazzara, R. P. Velthuisen, J. L. Pearlman, H. M. Greenberg, and H. Wagner, "Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation," *International journal of radiation oncology, biology, physics*, 2004.