Segmentation of Skull-infiltrated Tumors Using ITK: Methods and Validation

Aleksandra Popovic[1], Martin Engelhardt[2], and Klaus Radermacher[1]

[1] Helmholtz Institute for Biomedical Engineering, RWTH Aachen University, Germany, popovic@hia.rwth-aachen.de,
[2] Neurosurgical Hospital, Ruhr University Bochum, Germany,

Abstract. Methods for segmentation of skull infiltrated tumors in Computed Tomography (CT) images using Insight Segmentation and Registration Toolkit ITK (www.itk.org) are presented. Pipelines of filters and algorithms from ITK are validated on the basis of different criteria: sensitivity, specificity, dice similarity coefficient, Chi-squared, and Hausdorff distance measure. The method to rate segmentation results in relation to validation metrics is presented together with analysis of importance of different goodness measures. Results for one simulated dataset and three patient are presented.

1 Introduction

In the framework of CRANIO[1] project for Computed and Robot Aided Craniotomy [1] a module for automatic and semiautomatic detection and delineation of tumors invading skull, i.e. primary skull tumors, skull metastases, primary intradural meningiomas, and primary extradural meningiomas with skull extension, as defined in [2], is being developed. Those lesion typically involve both soft and calcified tissue, occupying wide range of Hounsfield Units (HU), therefore are difficult to segment. In this phase of our research we focus on CT imaging modality, since CT shows better geometrical precision needed for intraoperative navigation [3], but is also a golden standard for imaging of bone changes [4].

Recent developments in the ITK open source library made numerous image processing algorithms available for research and development. WE develop CT head image analysis pipelines to address the problem of skull tumor segmentation. As the building blocks of each pipeline, ITK functions and classes connected by the means of the same software paradigm and framework used in ITK itself, are used. Therefore, the compatibility with ITK, as well as with new versions of the library is preserved.

Measuring the performance of medical image processing algorithms requires the knowledge of the true borders of the object under investigation and a precisely defined validation metrics. However in medical image processing society there is no established method to define the *goodness* of a segmentation result.

[1] This project is supported by German Research Foundation (DFG RA548/2-1)

Software tools recently introduced for segmentation validation (e.g. [5]) are helpful but are not available as source code. We have used ITK functionalities in order to implement ITK-like functions/classes for computation of different validation metrics. Here, metrics are desribed and a proposal how to analyze the results is given.

2 Materials and Methods

One simulated CT dataset and three patients datasets, with manual segmentation done by a neurosurgical expect, two diagnosed with meningioma and one with the interosseous brain metastasis, are used. Patients are scanned with Siemens Somatom Plus CT scanner (0.43mm x 0.43mm x 2mm). The simulated dataset image is generated on the basis of two superimposed images: one of the healthy patient and one of the patient with diagnosed and postoperatively validated large meningioma. The simulated dataset was used to train the segmentation methods and optimize algorithmic parameters, since the true borders of the tumor are known.

2.1 Segmentation methods

Prior to evaluation, we have tested and visually inspected various results obtained from the three main groups of the segmentation algorithms:

- Image intensity filters
- Region-based methods
- Model-based methods.

Image intensity filters failed in nearly all cases, since the HU range of the tumor that infiltrated cranial bone is wide, given that image intensities in soft and calcified tissue are significantly different. The segmented region is either too large, comprising healthy parts as well, or too small comprising either bone or soft tissue only but no the entire tumor affecting calcified and soft tissue. For the region based segmentation methods we have used mathematical morphology operators, erosion and dilation, in order to remove significant over-segmentation, present in all the cases. Preprocessing of images included smoothing and edge enhancement using a Sigmoid filter around the tumor HU values. Figure 1 shows typical pipelines of two main processes. For Region Growing (RG) were used: Neighborhood Connected (NC), Confidence Connected (CC), Connected Threshold (CT) algorithm. Following Level Sets (LS) are used: Geodesic Active Contours (GAC), Laplacian (LLS), Threshold Level Sets (TLS) and Narrow Band Curves (NB).

More details on the algorithms could be found in [6] and [7].

2.2 Validation

As the golden standard for the validation a manual segmentation done by a neurosurgical expert is used. Although it is indicated that a larger number of

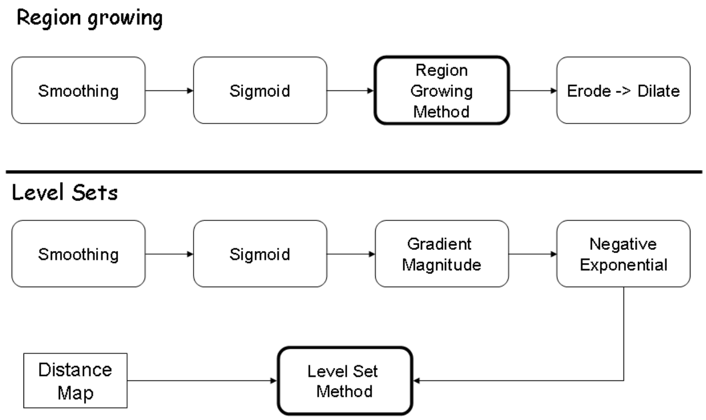


Fig. 1. Pipelines of two segmentation approaches

human raters is needed for a validation (e.g. in [8]), for the development process one segmentation done by an expert is sufficient. Validation criteria are often classified into two groups, statistical metrics and geometrical metrics, in respect which region/edge properties are being measured. The most common statistical criteria are sensitivity and specificity, defined as follows:

$$sensitivity = \frac{TP}{TP + FN} \qquad specificity = \frac{TN}{TN + FP}, \tag{1}$$

where TP, TN, FP, and FN are the numbers of voxels classified true positive, true negative, false positive, and false negative, respectively. Although they are widely used in the medical image processing domain, they are not sufficient since it is often ambiguous which criteria is more relevant for the given problem. The major pitfall of sensitivity/specificity is their dependability of the relation between sizes of image and object under the investigation. Statistical metric not dependable on the size of samples, Dice Similarity Coefficient (DSC) [9], is defined as follows:

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}. \tag{2}$$

Another size independent statistical metric is χ^2 [10]:

$$\chi^2 = \frac{sensitivity - (TP + FP)}{1 - (TP + FP)} \cdot \frac{specificity - (1 - (TP + FP))}{TP + FP}. \tag{3}$$

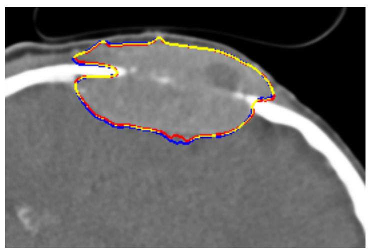
For geometrical validation, Hausdorff distance is used, available in ITK.

3 Results

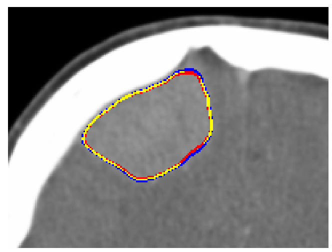
Simulated dataset was used to test the performance and to narrow the number of algorithms and parameter space used for the real patients. As mentioned in

Patient	Method	Sensitivity	Specificity	*DSC*	χ^2	Hausdorff
Phantom	LS	0.81±0.28	0.99±0.01	0.73±0.25	0.72±0.25	5.64±4.33
	RG	0.77±0.24	0.97±0.05	0.58±0.29	0.56±0.29	29.02±24.05
A	LS	0.69±0.12	0.89±0.24	0.46±0.14	0.39±0.15	28.96±21.25
	RG	0.69±0.08	0.93±0.05	0.39±0.07	0.29±0.08	74.2±8.23
B	LS	0.87±0.10	0.99±0.01	0.79±0.09	0.78±0.09	17.10±4.66
	RG	0.82±0.13	0.97±0.02	0.56±0.09	0.51±0.01	100.65±19.54
C	LS	0.79±0.10	0.97±0.001	0.70±0.09	0.68±0.09	19.86±6.38
	RG	0.88±0.07	0.97±0.03	0.66±0.12	0.63±0.13	50.65±25.73

Table 1. Mean and standard deviation of validation metrics 3 patients and the simulated dataset (Phantom); LS = Level Sets; RG = Region Growing



(a) Patient B



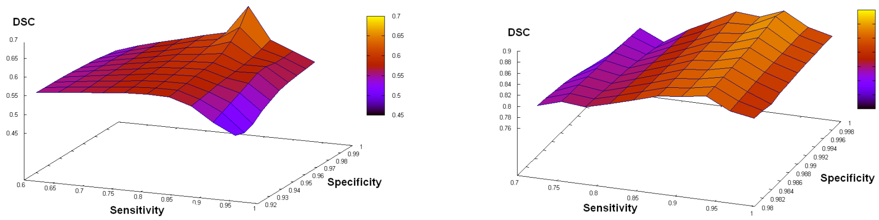
(b) Patient C

Fig. 2. Example CT slices with the best segmentation result (in both cases: Geodesic Active Contours Level Set algorithm). Red line is a manual segmentation, blue line is the best segmentation result according to the analysis of validation metrics; yellow line represents a perfect match between manual and automatic segmentation

the Section 2.1 image intensity filter failed and were not further considered. For further three patients segmentation methods were executed with varying parameters. Results of validation procedure for all segmentation outcomes is presented in the Table 1.

Analysis showed that there is no significant difference in segmentation result rating between DSC and χ^2, thus only one of them (DSC) is used for the validation purposes. In order to assess the quality of segmentation results, 3D plots of statistical parameters as in Figure 3 are used. For the segmentation results having all three statistical parameters similar, in the second stage of the validation, a geometrical measure (Hausdorff) is employed.

Figure 3 shows the obvious advantage of LS algorithms in comparison to RG approaches. As discussed in the Section 2.1 RG methods tend to over-segment, due to the wide range of HU values in which the tumor is situated. Erosion and dilation, combined, can suppress it to some amount but the problem persists. Among the LS methods, Geodesic Active Contours LS showed the best performance.



(a) RG (b) LS

Fig. 3. 3D plot of sensitivity, specificity, and DSC measures of two major groups of segmentation methods (RG and LS).

4 Discussion and Conclusion

Statistical methods commonly used for validation in medical image processing can often fail to detect errors in a segmentation result. For example, RG algorithms for Patient B, Table 1, show both sensitivity and specificity slightly better in comparison with the LS. However, after analysis of DSC, χ^2, and Hausdorff measures, it is clear that this conclusion is false since all later three are significantly in favor of Level Sets. Visual inspection showed that RG algorithms had significant over-segmentation. However, for a large amount of segmentation results, visual inspection is too time consuming, and should be avoided. Therefore, we have used a stepwise validation: 3D analysis in statistical metrics space (as in Fig.3), followed by geometric measure (Hausdorff) for segmentation results with similar statistical parameters. This validation pipeline takes into consideration multiple goodness measures and is able to appropriately classify segmentation results.

Advantage of Level Set algorithms to other moving surface approaches is in their capability to advance in two or more fronts and finally merge them together into one bounding surface. This behavior was observed here as well. The fronts were propagating separately in parts of images with soft and calcified tissue, building a common edge in the final phases of spreading. However in parts of images with a high texture, typically in calcified tissue, fronts were separating around high HU values.

Our ongoing work is focused on image intensity and mathematical morphology analysis of tumors in order to improve existing ITK algorithms to incorporate *a priori* knowledge in the segmentation pipeline.

References

1. Popovic, A., Engelhardt, M., Wu, T., Portheine, F., Schmieder, K., Radermacher, K.: CRANIO - Computer Assisted Planning for Navigation and Robot-assisted Surgery on the Skull. In Lemke, H., Vannier, M., Inamura, K., Farman, A., Doi,

K., Reiber, J., eds.: Proceedings of the 17th International Congress and Exhibition (CARS). Volume 1256 of International Congress Series., Elsevier (2003) 1269–1276
2. Lang, F.F., Macdonald, O., Fuller, G., DeMonte, F.: Primary Extradural Meningiomas: A Report of Nine Cases and Review of the Literature from the Era of Computerized Tomography Scanning. J Neurosurg **93** (2000) 940–950
3. Grover, S., Aggarwal, A., Uppal, P.S., Tandon, R.: The CT Triad of Malignacy in Meningioma - Redefinition, with a Report of Three New Cases. Neuroradiology **45** (2003) 799–803
4. Grunert, P., anf J. Espinosa, K.D., Filippi, R.: Computer-aided Navigation in Neurosurgery. Neurosurg Rev **26** (2003) 73–99
5. Gerig, G., Jomier, M., Chakos, M.: Valmet: A New Validation Tool for Assessing and Improving 3d Object Segmentation. In: MICCAI 2001: Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention. Volume 2208 of Lecture Notes in Computer Science. (2001) 516–528
6. Yoo, T.S., ed.: Insight into Images. Principles and Practice for Segmentation, Registration, and Image Analysis. A K Peters, Ltd. (2004)
7. Ibanez, L., Schroeder, W., Cates, L.N.J.: The ITK Software Guide. 2 edn. ITK Software Consortium (2005)
8. Warfield, K.H.Z., Wells, W.M.: Simultaneous Truth and Performance Level Estimation (staple): An Algorithm for the Validation of Image Segmentation. IEEE Trans. on Medical Imaging **23** (2004) 903–921
9. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Ron, K.: Statistical Validation of Image Segmntation Quality Based on a Spatial Overlap Index. Acad Radiol **11** (2004) 178–189
10. Yitzhaky, Y., Peli, E.: A Method for Objective Edge Detection Evaluation and Detector Parameter Selection. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003) 1–7