Manual labeling strategy for Ground Truth estimation in MRI Glial Tumor Segmentation.

Valentina Pedoia*
Dipartimento di Scienze Teoriche e Applicate, Universitá degli Studi dell'Insubria
Varese, Italy

Giuseppe Renis
Dipartimento di Scienze Teoriche e Applicate, Universitá degli Studi dell'Insubria
Varese, Italy

Sergio Balbi
Dipartimento di Biotecnologie e Scienze della Vita, Universitá degli Studi dell'Insubria
Varese, Italy

Alessandro De Benedictis
Dipartimento di Biotecnologie e Scienze della Vita, Universitá degli Studi dell'Insubria
Varese, Italy

Emanuele Monti
Dipartimento di Biotecnologie e Scienze della Vita, Universitá degli Studi dell'Insubria
Varese, Italy

Elisabetta Binaghi†
Dipartimento di Scienze Teoriche e Applicate, Universitá degli Studi dell'Insubria
Varese, Italy

ABSTRACT

In this paper we focused our attention on the problem of determining reliable ground truth for validating unsupervised, fully automatic MRI brain tumor segmentation procedures in the clinical context of Glial Tumor treatment . The goal was achieved by proposing an integrated "visual knowledge elicitation strategy" centered on the use of *GliMAn*(**Gli**al Tumor **Ma**nual **An**notator), a 3D MRI navigator that allows to view and manually labeling MRI volumes. As seen in our experimental context, the manual labeling process benefits from the insertion of a software tool taylored on the experts visual and usability requirements.

Categories and Subject Descriptors

D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Design,Verification

1. INTRODUCTION

*Correspond to: valentina.pedoia@uninisubria.it
†Correspond to: elisabetta.binaghi@uninisubria.it

VIGTA '12 May 21 2012, Capri, Italy

The use of automated procedures for MRI segmentation is still limited by several crucial aspects such as optimal selection of features in multispectral MRI, uncertainty management and minimize of the computational complexity [3, 2, 1]. Novel solutions are investigated in an attempt to fulfil stringent accuracy and efficiency requirements imposed by new fields of application. In this critical context the availability of reliable quantitative measures of accuracy and reproducibility for the proposed segmentation method should plays a key role. Unfortunately, the application of validation method to automated segmentation poses significant problems. Several methods have been proposed including the use of MRI contrast studies, phantom validation, MRI simulation studies, correlation with pathologic findings and reproducibility studies [3]. In recent works the validation problem has been addressed using experts to manually trace the boundaries of the different tissue regions [7, 11]. Manual labeling has the great potential of mimicking the radiologist's decision attitude which realistically is the only truth available [3]. However, some important drawbacks arise in terms of the labor intensive and, over all, in terms of the intra-inter- observer variation which strongly limits the ground truth determination process [8]. The considerable variation is related to limitations in the visualization process, MRI imaging conditions and to the intrinsic subjective character of the interpretation process with which an expert decides whether to assign a region under a given category. Several solutions have been proposed to address the problem. Warfield et.al, for example, proposed an automated algorithms based on Expectation-Maximization in in an attempt to remove the variability introduced by experts [13]. In other studies a solution based on the development of software tools for computer-assisted manual labeling was investigated [9]. Proceeding from these considerations, we focused our attention on the problem of determining reliable ground truth for validating unsupervised, fully automatic MRI brain tumor

segmentation procedures in the clinical context of Glial Tumor treatment [8]. The goal was achieved by proposing an integrated "visual knowledge elicitation strategy" centered on the use of *GliMAn*(**Gli**al Tumor **Ma**nual **An**notator), a 3D MRI navigator that allows to view and manually label MRI volumes. A contextual analysis has been developed with the aim of describing the clinical domain, the clinical practice in use and assessing how and how much the physicians perceive the problem. A quantitative analysis has been developed providing an objective measure of the intra- and inter- operator variability in manual labeling procedures accomplished with standard tools in use in the specific domain of interest. The results of this analysis created the basis for deriving solutions and validation procedures.

2. CONTEXTUAL ANALYSIS

A precise volumetric computation of the pathological MRI signal has several fundamental implications in clinical practice. In fact, the accurate definition of both the topographical features and the growing pattern of the tumor is crucial in order to select the most appropriate treatment, to plan the best surgical approach and, postoperatively, to correctly evaluate the extent of resection and monitoring the evolution over time of any possible residue [5]. However, it is worth noting that gliomas are characterized by a constant local growth (4 mm/year) within the brain parenchyma, migration along white matter pathways, both in ipsilateral and even contralateral hemisphere and unavoidable anaplastic transformation [4]. Because of the their infiltrative nature, the pathological signal revealed in MRI does not correspond to the exact boundaries of gliomas. On the contrary, especially in the case of slow-growing lesions, it was demonstrated, by taking multiple biopsy samples, that tumor cells are present in a consistent number, but not sufficient to give an hyperintense signal, at a distance of at least 20 mm from the tumor landmarks shown by MR imaging [10]. For these reasons, the main problem in radiological detection and segmentation for gliomas depends from their histopathological features, especially at the periphery of the hyperintensity detected by MRI. As a consequence, since it is not easy or even impossible to objectively establish the limits between the tumor and the normal brain tissue, a large intra- interpersonal variability is usually revealed during the manual segmentation of MRI sequences.

2.1 Assessment of Inter- and Intra- Personal Variability In Glial Tumor Segmentation

A team of medical experts has segmented axial, sagittal and coronal slices of several cases of Low Grade Glioma in MRI. The manual segmentation was performed using a simple image manual annotator, already in use, that includes the standard tools of an image viewer. MRI segmentation is performed with the purpose of determining the volume of tissues and their 3D spatial distribution. We proceed by measuring the reliability of the ground truth determination with respect to both the above purposes.

Volume Estimation Error:

Let be $V_1^i\ V_2^i\ V_3^i$ the volume estimation from the axial, sagittal and coronal plane segmentation respectively, performed by the the $i-th$ expert. The Intra- and Inter- volume estimation errors for the plane p with $p \in [1-3]$ and the $i-th$ expert are computed as follows:

$$intraVolErr_p^i = \frac{V_p^i - \frac{1}{N_{seg}} \sum_{j=1}^{N_{seg}} V_j^i}{\frac{1}{N_{seg}} \sum_{j=1}^{N_{seg}} V_j^i};$$
$$interVolErr_p^i = \frac{V_p^i - \frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} V_p^j}{\frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} V_p^j} \quad (1)$$

where N_{seg} is the number of segmentations performed by the same expert on the same volume and N_{exp} is the total number of experts.

Spatial Distribution Variability:

Let be $M_1^i\ M_2^i\ M_3^i$ the volumetric masks obtained from the axial sagittal and coronal plane segmentation respectively, performed by the $i-th$ expert. The Intra- and Inter- spatial distribution variability, based on the Jaccard Index [6], are computed as follows:

$$J_i^{p,t} = \frac{M^p i \cap M^t i}{M^p i \cup M^t i}; \quad J_{i,j}^p = \frac{M^p i \cap M^p j}{M^p i \cup M^p j}; \quad (2)$$

where i and j are indexes related to the experts and p and t to the segmentation planes.

A strong Disagreement was found, with maximum volume estimation error equal to 10% and 20% for the intra- and inter- personal variability respectively. For that concerns the spatial distribution, the minimum percentage of overlap for the intra- and inter- variability are equal to 73% and 71% respectively.

3. MANUAL LABELING STRATEGY FOR GROUND THROUGH ESTIMATION IN MRI GLIAL TUMOR SEGMENTATION

The proposed strategy contemplates, as preliminary phase, the organization of tuning sessions aimed at establishing a consensus among experts trough discussion of the most controversial cases. In addition to this preliminary phase, experts were invited to monitor the design and development phases of the manual labeling tool *GliMAn* for giving intermediate feedbacks and suggestions.

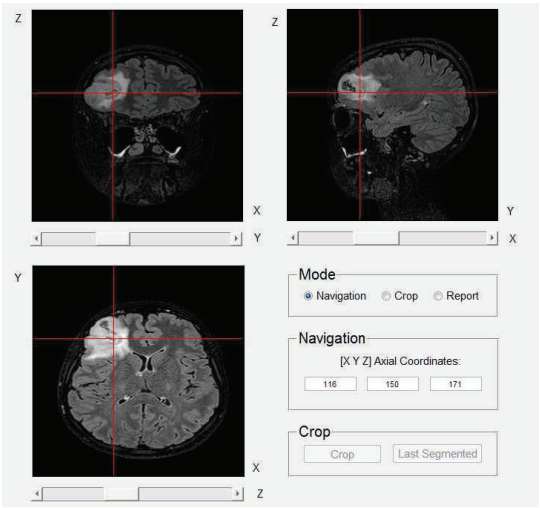

Figure 1: *GliMAn* central zone

3.1 GliMAn Design

GliMAn is a MATLAB application that allows to view and manipulate MRI volumes with the aim of supporting reliable collection of ground truth in Glial brain tumor segmentation process. Input data of the application are DICOM images. *GliMAn* has been designed according to Interaction Design framework that structures the phases design within an iterative process [12] in which partial evaluations supported by experts and refinements take place. The design of *GliMAn* started with the collection and analysis of requirements in which the user model and operational conditions are outlined. Cognitive and/or perceptual processes, attitudes and limitations involved in visual inspection and manual segmentation tasks have been assessed. Operation requirements concerning hardware facilities, input and output devices are inherited form protocols in use in biomedical radiology domain. Conceptual design phase was focused on the analysis of factors involved in the operator variability phenomenon and in the formulation of solutions conceived as *GliMAn* main functional requirement. The analysis was conducted through a close dialogue between physicians and computer scientists and joint meetings in which working sessions were held. The main cause of variability has been attributed to the loss of the spatial continuity constraint in the orthogonal direction with respect to the segmentation plane. It was concluded that the simultaneous viewing from different points is needed to resolve uncertainties and for making reliable decisions on detecting boundaries and labeling regions of interest. The physician should explore a resonance volume for subsequent axial coronal and sagittal slices. The decision on the single slice must be contextually related to the inspection of previous and subsequent slices. Proceeding from these considerations, we conceived, as main feature of the *GliMAn* tool, the preservation of the volumetric nature of the data through the simultaneous display of the three orthogonal planes (axial,sagittal and coronal) and the synchronized visualization of the manual annotation activity. Human-computer interaction principles and usability guidelines have been strictly observed in the *GliMAn* physical design, in order to limit in the GUI interaction, eyestrain and ambiguities the would interfere in the effectiveness of conceptual solutions. The GUI is composed of 3 principal areas: upper, central and lateral. The first includes standard image viewer I/O and management tools, the central zone shows the orthogonal planes and the lateral zone allows to change the execution mode. Plan layout has been designed in accordance with solutions adopted in standard image processing and viewer environments for medical applications (Figure 1). Moreover the method of the orthogonal projections is universally used to represent in a simple, objective and dimensionally accurate way the volumetric object. The essential feature of this visualization method is to preserve the correct proportions between the elements of the volume. The visualization in three planes is synchronized: choosing a point of coordinates (x', y', z'), the three images represented are the intersection of the MRI volume with the sagittal coronal and axial planes respectively passing through the point. The manual labeling is obtained through the identification of a series of points on one of the three planes. The remaining planes are used for control purposes. The boundary detection task is accomplished element by element and is organized as follows: 1. the user points and selects a candidate point in a given plane; 2. the same point is highlighted in the other two planes and analyzed; 3. the expert confirms the decision by re-selecting the same point or decides to examine another point. Figure 2 shows a crop of a brain MRI axial section of a low grade glioma, the high degree of infiltration makes the identification of the pathology's boundary a very complex task. The analysis of the axial section alone is not sufficient to make a reliable decision about the point identified with the red circle to the edge of the tumor. As illustrated in Figure 3, the visualisation of the orthogonal control planes in *GliMAn* interface reduces the uncertainty in the assignment of the point to the boundary. The selected points are then joined by a broken line Clicking on the first point the broken line becomes a polygon that encloses the area of interest and the segmentation performed is shown superimposed on the original MRI. During the segmentation of the $N-th$ slice, the segmentation performed on the slice $N-(th-1)$ can be viewed as a guide. Report mode allows to obtain information concerning the number of slices segmented during the work session, the volume of the pathological area segmented and the surface of the last slice segmented. Moreover, the trend of the areas of tumor's sections along the slices can be viewed; the objective is to control the consistency, in terms of smoothness, of the detected area, in the direction perpendicular to the segmentation plane.

Figure 2: Axial slice of a critical case example

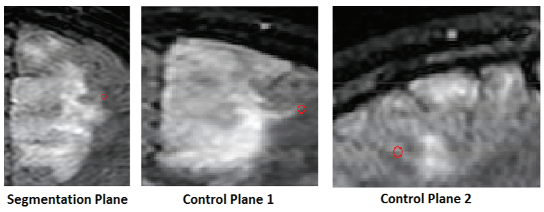

Figure 3: Example of critical case solved using the control orthogonal planes

4. EXPERIMENTAL RESULTS

GliMAn was evaluated in terms of inter- personal variability. Moreover, the GUI usability has been evaluated through an evaluation test of the experience of use. The dataset used for the evaluation process is composed of 2 brain MRI, gray scale, 12 bit depth, volumes of patients with Low Grade Glioma. All the MRI in the dataset are volumetric acquisition (contiguous slices are acquired, there are no jumps) with isotropic voxel (0.57 mm). 4 axial sections, identified by the experts as critical cases, were manually segmented, performing the segmentation slice by slice independently and using *GliMAn*. This evaluation phase reflects the method used in the Section 2 However we had available 4 slices processed by

Table 1: Comparison between Slice by Slice and *GliMAn* manual segmentation method a) Mean of the surface estimation errors $interSurErr_p^i$ for each segmented plane p in the first and second case b) Mean of the spatial distribution agreements $J_{i,j}^p$ for each segmented plane p and for each couple of experts i, j

(a)

Expert	Case 1		Case 2	
	Slice by Slice	GliMaAn	Slice by Slice	GliMaAn
1	14,31%	8,05%	15,87%	8,25%
2	10,90%	11,89%	16,76%	6,02%
3	20,90%	6,62%	9,69%	9,76%
4	26,14%	20,09%	15,73%	9,64%
5	17,85%	8,48%	19,34%	6,17%

(b)

Expert	Case 1		Case 2	
	Slice by Slice	GliMaAn	Slice by Slice	GliMaAn
1	78,26%	84,57%	75,79%	78,41%
2	77,48%	86,60%	74,92%	80,05%
3	75,76%	85,55%	74,33%	80,89%
4	75,76%	85,55%	75,31%	79,53%
5	77,42%	87,68%	77,00%	80,93%

experts and then we assessed the error in the area estimation instead of in the volume estimation; for what concern the spatial distribution similarity the evaluation process is the same.
As regards the interpersonal variability in the surface estimation, the error made for each expert is computed for both the slice by slice and *GliMAn* segmentation as follows:

$$interSurErr_p^i = \frac{S^i - \frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} S_p^j}{\frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} S_p^j} \quad (3)$$

where S^i is the surface estimation from segmentation performed by the the $i-th$ expert. In Table 1(a) the average for each segmented plane p of surface estimation errors computed using the slice by slice and *GliMAn* tools are collected. The segmentation with *GliMAn* prevails in average in both the MRI cases. For the first case the average surface error is 11.02% and 18.02% for slice by slice and *GliMAn* respectively. For the second case the same indexes are 7.97% and 15.48%. The comparison in terms of spatial distribution is highly in favor of *GliMAn* with maximum increasing of agreements percentages 26.79%, passing from 62% to 89%. In two cases the level of agreement slightly decreases (-1.93% and -0.30%). In Table 1(a) the mean of the spatial distribution agreements $J_{i,j}^p$ for each segmented plane p and for each couple of experts i, j are reported.

5. CONCLUSION

In this paper we have addressed the problem of defining a reliable validation of MRI brain tumor segmentation. In the literature there are several works that emphasize the unreliability of the current ground truth collection methods, however relatively few works have proposed operative solutions. We conducted a study with the purpose of analyzing and assessing factors involved in the observer variation during manual labeling process and proposing a computer assisted labeling strategy. As seen in our experimental context, the manual labeling process benefits from the insertion of a software tool taylored on the experts visual and usability requirements. All the phases contemplated in the elicitation strategy have made a significant contribution; preliminary discussions, structured and unstructured interviews created the premise for a successful subsequent design phase. As side effects, experts improved their knowledge through discussions and comparison of their decision attitudes.

6. REFERENCES

[1] M. S. Atkins, B. Mackiewich, and K. Whittall. Fully automatic segmentation of the brain in MRI. 1998.

[2] M. Balafar, A. Ramli, M. Saripan, and S. Mashohor. Review of brain MRI image segmentation methods. *Artificial Intelligence Review*, 33:261–274, 2010.

[3] L. Clarke, R. Velthuizen, M. Camacho, J. Heine, M. Vaidyanathan, L. Hall, R. Thatcher, and M. Silbiger. MRI segmentation: Methods and applications. *Magnetic Resonance Imaging*, 13(3):343 – 368, 1995.

[4] H. Duffau. Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity. *Lancet Neurol*, 4(8):467–487, 2005.

[5] H. Duffau. Surgery of low-grade gliomas: towards a 'functional neurooncology'. *Current Opinion In Oncology*, 21:543–549, 2009.

[6] P. Jaccard. The distribution of the flora in the alpine zone. *New Phytologist*, 11(2):37–50, 1912.

[7] C. Jason J., S. Eitan, D. Shishir, E.-S. Suzie, S. Usha, and A. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Transactions on Medical Imaging*, 27(5), 2008.

[8] M. Kaus, S. Warfield, A. Nabavi, P. M. Black, F. A. Jolesz, , and R. Kikinis. Automated segmentation of MRI of brain tumors. *Radiology*, 218:586–591, 2001.

[9] R. Kikinis, P. L. Gleason, T. M. Moriarty, M. R. Moore, E. I. Alexander, P. E. Stieg, M. Matsumae, W. E. Lorensen, H. E. Cline, and P. M. Black. Computer-assisted interactive three-dimensional planning for neurosurgical procedures. *Neurosurgery*, 38(4), 1996.

[10] J. Pallud, P. Varlet, B. Devaux, S. Geha, M. Badoual, and C. Deroulers. Diffuse low-grade oligodendrogliomas extend beyond MRI-defined abnormalities. *4. Neurology*, 74:1724–1731, 2010.

[11] M. Sabuncu, B. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *Medical Imaging, IEEE Transactions on*, 29(10):1714 –1729, oct. 2010.

[12] H. Sharp, Y. Rogers, and J. Preece. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, 2 edition, Mar. 2007.

[13] S. K. Warfield, K. H. Zou, and W. M. Wells. Validation of image segmentation and expert quality with an expectation-maximization algorithm. In *in Proceedings of Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part I*, pages 298–306. Springer-Verlag, 2002.