Insubria University
Major Ph.D. Thesis

Functional and Structural MRI Image Analysis for Brain Glial Tumors Treatment

Advisor: Prof. Elisabetta Binaghi
Co-Advisor: Dott. Sergio Balbi

Major Ph.D. Thesis
Valentina Pedoia

Abstract

This Ph.D Thesis is the outcome of a close collaboration between the *Center for Research in Image Analysis and Medical Informatics (CRAIIM)* of the Insubria University and the *Operative Unit of Neurosurgery*, *Neuroradiology* and *Health Physics* of the University Hospital "Circolo Fondazione Macchi", Varese.
The project aim is to investigate new methodologies by means of whose, develop an integrated framework able to enhance, and in some cases make possible, the use of the information contained in the Magnetic Resonance Images, in order to support clinical experts in the treatment of patients with brain Glial tumor.
Both the most common uses of MRI technology for non-invasive brain inspection were analyzed. From the *Functional* point of view, the goal has been to provide tools for an objective reliable and non-presumptive assessment of the brain's areas locations, to preserve them as much as possible at surgery.
From the *Structural* point of view, methodologies for fully automatic brain segmentation and recognition of the tumoral areas, for evaluating the tumor volume, the spatial distribution and to be able to infer correlation with other clinical data or trace growth trend, have been studied.
Each of the proposed methods has been thoroughly assessed both qualitatively and quantitatively; this second aspect is particularly critical in the context of medical imaging where it is very difficult to derive a unique ground truth.
All the Medical Imaging and Pattern Recognition algorithmic solutions studied for this Ph.D. Thesis have been integrated in *GliCInE: Glioma Computerized Inspection Environment*, which is a MATLAB prototype of an integrated analysis environment that offers, in addition to all the functionality specifically described in this Thesis, a set of tools needed to manage Functional and Structural Magnetic Resonance Volumes and ancillary data related to the acquisition and the patient.

Sommario

Questa tesi é il frutto di una stretta collaborazione tra il *Centro di Ricerca in Analisi di Immagini e Informatica Medica (CRAIIM)* dell'Universitá degli Studi dell'Insubria e le *Unitá Operative di Neurochirurgia*, *Neuroradiologia* e *Fisica Sanitari*a dell'Ospedale del Circolo Fondazione Macchi di Varese.
L'obbiettivo del progetto é stato quello di studiare metodologie innovative attraverso le quali sviluppare un framework integrato di strumenti software in grado di migliorare, e in alcuni casi rendere possibile l'utilizzo delle informazioni contenute nelle Immagini di Risonanza Magnetica, nell'intento di supportare gli esperti clinici nella trattazione di pazienti affetti da tumore Gliale cerebrale.
Entrambi i maggiori impieghi della tecnologia MRI per l'ispezione cerebrale non invasiva sono stati analizzati: il punto di vista *Funzionale* con l'obbiettivo di fornire strumenti per una valutazione oggettiva ripetibile e non presuntiva della dislocazione delle aree cerebrali al fine di preservarle quanto piú é possibile in sede chirurgica e il punto di vista *Strutturale/Morforlogico* studiando metodologie completamente automatiche per la segmentazione del parenchema cerebrale e l' identificazione, segmentazione e riconoscimento delle aree tumorali in modo da poterne valutare il volume la distribuzione spaziale ed essere in grado di inferire correlazione con altri dati clinici o tracciarne un trand di accrescimento. Ognuna delle metodologie proposte é stata accuratamente valutata sia dal punto di vista qualitativo che quantitativo; aspetto particolarmente critico nell'ambito del Medical Imaging in cui risulta molto difficile ricavare una veritá univoca dall'analisi svolta da un singolo esperto.
Tutte le soluzioni algoritmiche di Medical Imaging e Pattern Recognition studiate per questo progetto di tesi sono state integrate in *GliCInE: Glioma Computerized Inspection Environment* prototipo MATLAB di un ambiente di analisi che offre, oltre alle funzionalitá specificatamente descritte in questa Tesi, anche una serie di tools accessori necessari alla gestione dei volumi di Risonanza Magnetica Funzionale e Strutturale e i dati ancillari relativi all'aquisizione e al paziente.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Glial Brain Tumors Overview

Brain tumors are abnormal and uncontrolled proliferations of cells. Some originate in the brain itself, in which case they are termed primary. Others spread to this location from somewhere else in the body through. Metastasis, and are termed secondary. Primary brain tumors do not spread to other body sites, and can be malignant or benign. Secondary brain tumors are always malignant. Both types are potentially disabling and life threatening. Because the space inside the skull is limited, their growth increases intracranial pressure, and may cause edema, reduced blood flow, and displacement, with consequent degeneration, of healthy tissue that controls vital functions.
Brain tumors are, in fact, the second leading cause of cancer-related deaths in children and young adults. According to the Central Brain Tumor Registry of the United States (CBTRUS), there will be 64,530 new cases of primary brain and central nervous system tumors diagnosed by the end of 2011. Overall, more than 600,000 people currently live with the disease. Although the causes of brain tumors are unknown, a few risk factors have been proposed. These include head injuries, hereditary syndromes, immunosuppression, prolonged exposure to ionizing radiation, electromagnetic fields, cell phones, or chemicals like formaldehyde and vinyl chloride. None of these, however, is proven to actually cause the disease. Symptoms of brain tumors include persistent headache, nausea and vomiting, eyesight, hearing and/or speech problems, walking and/or balance difficulties, personality changes, memory lapses, problems with cognition and concentration, and seizures.

Gliomas are the most common primary malignant brain tumors. They originate from non-neuronal (glial) brain cells called astrocytes. The World Health Organization (WHO) classifies them, from the least to the most aggressive, into four grades:

- Grade I pilocytic astrocytoma
- Grade II diffused astrocytomas
- Grade III anaplastic astrocytomas
- Grade IV glioblastomas

Gliomas are typically associated with low survival. This is due to a combination of factors, including high relapse rates, which hinder successful treatment. Also, gliomas are highly invasive. They can infiltrate to surrounding and remote brain areas as single cells. So, there is often no clear boundary between tumor and healthy tissue, which makes surgical resection difficult. Infiltration may involve eloquent areas; that is, those with specific functions, such as language, vision, and motor control. In this case, gliomas are referred to as infiltrating. IV-grade gliomas include glioblastoma multiforme, the most aggressive type of brain tumor. Its median survival rate is just 12 months.
Diffuse World Health Organization grade II gliomas (GIIG) represent approximately 15-25% of all gliomas with an average incidence around 1 per 100.000 population per year [125]. These slow-growing primary brain tumors usually affect young adults aged between 30 and 40, fully enjoying their socio-professional life [184]. The clinical picture frequently remains silent for a long time and the neurological onset, typically paucisymptomatic, comprehends seizures or other mild disorders concerning language, cognitive aspects or sensory-motor functions [14]. Nevertheless, during the last decade, more exhaustive insights on the epidemiological and biological features of these tumors allowed to understand that the natural history of GIIG expects, in all cases, three consecutive evolutionary tendencies [59]: a constant local growing (4 mm/year); the tendency to migrate along white matter pathways, both homolaterally to the lesion and even contralaterally via the interhemispheric connecting fibers of the corpus callosum; an unavoidable anaplastic transformation after a median delay of around 7-8 years, conditioning a fatal progression (median overall survival around 10 years) [110, 111]. Therefore, given the inescapable biological progression, there is today a general consensus in considering GIIG not longer as benign tumor, but rather as precancerous lesions [157]. As a consequence, the better understanding on the natural course of GIIG allowed to improve their management, moving from a traditional *wait-and-see* attitude to a multistage and multidisciplinar therapeutic strategy, comprehending surgery, chemotherapy and radiotherapy [157, 60]. In particular, total or subtotal surgical resection, when possible, is now indicated as the first therapeutic option to adopt in symptomatic subjects, because it was demonstrated effective in dramatically changing the natural history of slowly

evolving gliomas with a positive impact on overall survival by delaying the anaplastic transformation [83]. Moreover, the even more safe and feasible clinical use of intraoperative neurofunctional methods for mapping and monitoring the eloquent brain structures at the cortico-subcortical level, enables the surgeon to reach the maximal quality of resection while preserving the functional outcome of the patient at the long term [18].

1.2 Structural MRI role in Glial Tumors Treatment

Structural Magnetic Resonance Imaging (MRI) fills a central place in GIIG neuroradiological assessment for several reasons. First, it represents the first adopted exam to forward a diagnostic hypothesis, topically based on the finding of an iperintensity on T2/FLAIR-weighted sequences with no or slight enhancement after gadolinium injection [59]. An accurate characterization, both qualitative and quantitative, of the pathological MRI signal is crucial for defining the topographical features and evolutionary trend of the tumor. This evaluation is critical in order to select the most appropriate treatment, to plan the best surgical approach and, when the resection appears not at once indicated, for istance in the case of more diffusive lesions, to envisage possible alternative solutions supporting surgery. In fact, some preliminary experiences reported the positive contribution of a neo-adjuvant chemotherapy in inducing a shrinkage of tumor hiperintensity on the post-treatment neuroimaging estimation [24]. Postoperatively, MRI is necessary for evaluating the extent of resection which constitutes, in addition to histopathology, the most important prognostic factor [83]. In addition, a precise volumetric computation is of basilar relevance during the follow-up, especially for monitoring the evolution over time of any possible residue which was necessarily not resected because localized within territories critical for neurological integrity of the patient, and it is essential for establishing any potential alteration suggestive for undesirable biological modifications [111]. Finally, thanks to a more widespread availability of MRI equipment, an earlier identification of lesions still in preclinical stages is actually possible [150]. Even in these cases, since the natural history for asymptomatic GIIG does not differ from clinically manifested forms, it was recently suggested to propose a preventive surgery to these patients with the aim, once again, to optimize the prognostical impact before a tumor diffusion surgically inaccessible. However, it should be remarked that the pathological image reveled on MRI does not define the exact boundaries of the tumor. On the contrary, as for high-grade gliomas [87], also for GIIG it was demonstrated, through multiple biopsy samples, that tumor cells are present in a consistent number, but not sufficient to generate an hyperintense signal, at a distance of at least 20 mm from the tumor landmaks shown by MR imaging. For these reasons, it was lately advocated to push surgery beyond the radiological limits so achieving a supra-total resection, according to the functional cortico-subcortical landmarks specific for the patient, with the purpose to

get the best possible oncological control over time [61]. In summary, MRI is actually the first fundamental neuroradiological tool for evaluating and tracing the morphological, topographical and evolutionary characteristics of a GIIG. The development of methods able to improve the accuracy of the segmentation of the infiltrated cerebral tissue in of primary interest to enable the neuroncological team to reach the best therapeutic planning at the individual level.

1.3 Functional MRI role in Glial Tumors Treatment

Functional Magnetic Resonance Imaging (fMRI) is a technique for measuring brain activity. It works by detecting the changes in blood oxygenation and flow that occur in response to neural activity. When the brain area is more active it consumes more oxygen and to meet this increased demand blood flow increases to the active area. fMRI can be used to produce activation maps showing which parts of the brain are involved in a particular mental process [84, 98, 69]. fMRI now has a small but growing role in clinical neuroimaging. It is used in pre-surgical planning to localize brain function [119]. Presurgical fMRI is carried out to facilitate function preserving and safe treatment in patients with brain tumors and epilepsies by non invasively localizing and lateralizing specific brain functions or epileptic activity [84]. This diagnostic information is not available from morphological brain imaging or invasive measures prior to treatment. Consequently presurgical fMRI measurements are always performed in individual patients to achieve a 'functional' diagnosis. They thereby differ fundamentally from the fMRI experiments in basic neuroscience research where usually normal or altered brain function is investigated in group studies. The experimental setup as well as data processing and evaluation need to be adapted to the clinical environment so that patients who may present with neurological or cognitive deficits can be examined successfully in a standardized way. It is important to note that fMRI can not be considered a definitively established method routinely used in clinical diagnostic neuroimaging. Clinical fMRI should be carried out in the framework of scientific studies until the hardware and software components for medical applications have been certified and official recommendations or guidelines have been issued by relevant medical associations. Until then, individual routines and standards need to be established for measuring techniques and examination protocols, data processing and evaluation, as well as for medical interpretation and documentation of clinical fMRI findings. As a basic principle, clinical fMRI examinations should be performed, evaluated and interpreted by trained and experienced examiners with particular expertise in this area, since careless use of this very promising technique could endanger patients [119, 84].
With the help of functional landmarks provided by fMRI, the spatial relationship between brain tumor or epileptogenic zone and functional areas can be evaluated preoper-

atively, which facilitates selection of the most cautious therapy. fMRI permits planning and implementation of the most radical and at the same time function-preserving operation, particularly if these functional landmarks are available for functional neuronavigation [118]. This goal is rarely achieved with morphological information alone, especially in the presence of brain malformation, anatomical variants, disturbed or destroyed anatomy due to tumor growth or pathology related neuroplastic changes of brain function. fMRI yields relevant diagnostic information on anatomy, pathology and function prior to surgery in one single examination. The indication for presurgical fMRI results from clinical signs and symptoms. There is a large body of evidence from validation studies with established reference procedures and multimodal investigations comparing fMRI to other methods of functional neuroimaging suggesting reliability of fMRI in (presurgical) localization of functional areas in primary motor and somatosensory cortices and in determination of Broca and Wernicke language areas, if indications and limits of the method are considered . Additional attention needs to be paid to technique related inaccuracies due to the superposition of functional and morphological images, the referencing of fMRI data in neuronavigation systems or to surgery induced localization errors following removal of brain tissue or loss of cerebrospinal fluid (brain shift). There is also potential for clinical fMRI in applications including pre symptomatic diagnosis, drug development, individualization of therapies and understanding functional brain disorders. Early studies also suggest that fMRI has the potential to be used as bio-feedback for conditions such as chronic pain [187, 62].

Part I

Structural Magnetic Resonance Imaging

2
MRI Segmentation

2.1 Overview

The problem of segmenting all or part of the brain in MRI imagery continues to be investigated giving rise to a variety of approaches attempting to satisfy the high accuracy demand in diversified clinical and neuroimaging application [42, 10, 6, 88]. The rationale behind this effort is to improve upon existing early methods suffering from incorrect detection due to noise, difficulties in assigning threshold in the edge image and over and under estimation. The proposed solutions make use of a single, gray scale 2D or 3D image or multiple MR images with different gray scale contrast [42].
Focusing on MRI segmentation from gray scale images, the early most intuitive approach proposed is automatic thresholding. The application of this approach to MR brain image segmentation includes works based on iterative, knowledge guided thresholding [167, 36] and histogram analysis [2] complemented in some cases with image processing techniques such as morphological filtering [102]. The solutions proposed suffer in general from the difficulty in setting threshold values and present strong limitations in the presence of inhomogeneity, image artefact and anatomical variability.
The Region Growing approach extends thresholding by combining it with connectivity. Methods within this approach need an operator for manual setting of seed for each region and their results depend on the heuristic assessment of thresholding for homogeneity [44, 127]. As a consequence, suitable accuracy is achieved only detecting the contour in slices where the brain is one homogeneous region without pathological alterations and inhomogeneity.

Supervised statistical pattern recognition approaches are proposed in MR image segmentation to circumvent the problem of explicitly and analytically describing the specific segmentation procedure and related parameters, lying to a learning stage the charge of inducing the classifier from supervised data available. Parametric and non parametric methods are proposed showing a competitive behavior in reducing the effect of Radio Frequency(RF) inhomogeneity but requiring some interaction to provide tissue training pixels and to perform post processing adjustments [80, 43]. The parametric probabilistic classification is sensitive to accurate estimation of the probability density function (PDF) [158]. Nonparametric approaches obtain the parameters of the PDF from data without any assumptions, so these approaches are accurate but expensive [115, 39]. Computational complexity and low generalization of trained classifiers limit their applicability. Unsupervised methods employed for MR brain image segmentation include K-Means and its fuzzy equivalent Fuzzy c-Means(FCM) [74]. From a viewpoint of reproducibility, unsupervised methods are clearly desirable. Many algorithms are introduced to make FCM robust against noise and inhomogeneity but most of them still present low accuracy [193, 75]. Other pattern recognition methods such as neural networks and template matching have been also applied to MR brain image segmentation [75, 171].
Edge-based segmentation approaches have been widely investigated to segment MR images of the head [42, 25, 53]. Base methods apply local operations to every point of the input picture, in order to extract short edge elements. Some of these operations are very simple, like the gradient or the Laplacian, whereas other operations are more complex and allow the elimination of most of the local noise [13]. When small edge elements have been obtained, edge following strategies are applied to connect them to form borders. Snell et al. use an active surface template to find the intracranial boundary in MRI volumes of the head [156]. The method is based on the active contour model algorithm Snakes [85]. However the resulting procedure requires user interaction to provide a good initial contour for subsequent snake contour refinement. The Bayesian dynamic contour (BDC) model [100] shows advantages over the snakes active contour as the energy functions used to find the boundary can be more generally based on information about the whole region, rather than just the local boundary characteristics. The results are promising for the central slices of multispectral images presented, but have yet to be developed for isolating the whole brain. Contour-based deformable models are applied successfully to MRI brain segmentation although the approach suffer from the difficulty of determining the initial contour and tuning parameters [88, 2]. Khotanlou et al. [88] explicitly propose a segmentation method based on deformable models and asymmetry analysis. Proceeding from the assumption that in normal brain the symmetry plane of the head is approximately equal to the symmetry plane of the segmented brain, a MRI brain segmentation algorithm robust in presence of tumor has been developed. Results obtained are satisfactorily but they heavily depend from the validity of the symmetry

assumption.
A boundary tracing method using dynamic programming for noisy brain sections with vague boundaries, is proposed but successful application for global segmentation of MR images remains to be demonstrated since a good initial guess for the boundaries is required [183]. Of the many techniques available for automated border detection, those based on graph searching principles [112] have been successful in several applications. The value of graph search method lies in robustness which derives from the property typical of the optimization methods of embedding global information about edges and the structure of object in a figure of merit allowing accurate border detection from noisy pictures [116]. Graph searching has become one of the best investigated segmentation tool for medical image data with applications to coronary angiography, ultrasound and cardiac MRI imaging [170, 162, 160, 65].

2.2 Experimental evaluation of expert Inter- Intra- Variability in fully manual segmentation

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 [60]. 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 [59]. 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 [126, 87]. 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- inter- personal variability is usually revealed during the manual segmentation of MRI sequences.
Before detailing the proposed segmentation strategy, an analysis aimed to observe and measure the difficulties met from the physicians in the manual segmentations is shown. The aim of this variability analysis is twofold: to assess the agreements of segmentations performed by different experts (Inter-Variability) and to assess the reproducibility of the

manual segmentation process of the same expert (Intra-Variability).
The dataset used is composed of four FLAIR MRI gray scale volumes with the following acquisition parameters:

- gray scale
- 12 bit depth
- Volume Size $[432 \times 432 \times 300]$
- Slice Thickness $0.6mm$
- Spacing Between Slice $0.6mm$
- Pixel Spacing $(0.57, 0.57)mm$
- Repetition Time 8000
- ECHO Time 282.89

All dataset volumes are altered by the presence of glial tumors, heterogeneous in terms of position, dimension, intensity and shape.
A team of five medical experts was asked to segment axial, sagittal and coronal slices of these volume data by employing a simple image annotator normally in use in the clinical practice and offering standard image viewing facilities. Figure 2.1 shows an example in which slice-by-slice manual segmentations provided by experts are superimposed on a MRI slice.
MRI segmentation is performed with the purpose of determining the size of pathological tissues and their spatial distribution in two or three dimensions depending on the nature of the data. The metrics adopted in size estimation error and spatial distribution error are described below:

Size Estimation Error: Let be S_1^i S_2^i S_3^i the size estimation of one region extracted from the axial, sagittal and coronal plane segmentation respectively, performed by the the $i-th$ expert. The Intra- and Inter- size estimation errors along the plane p with $p \in [1-3]$ and the $i-th$ expert are computed as follows:

$$intraSizefErr_p^i = \frac{S_p^i - \frac{1}{N_{seg}} \sum_{j=1}^{N_{seg}} S_j^i}{\frac{1}{N_{seg}} \sum_{j=1}^{N_{seg}} S_j^i};$$
$$intraSizefErr_p^i = \frac{S_p^i - \frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} S_p^j}{\frac{1}{N_{exp}} \sum_{j=1}^{N_{exp}} S_p^j} \tag{2.1}$$

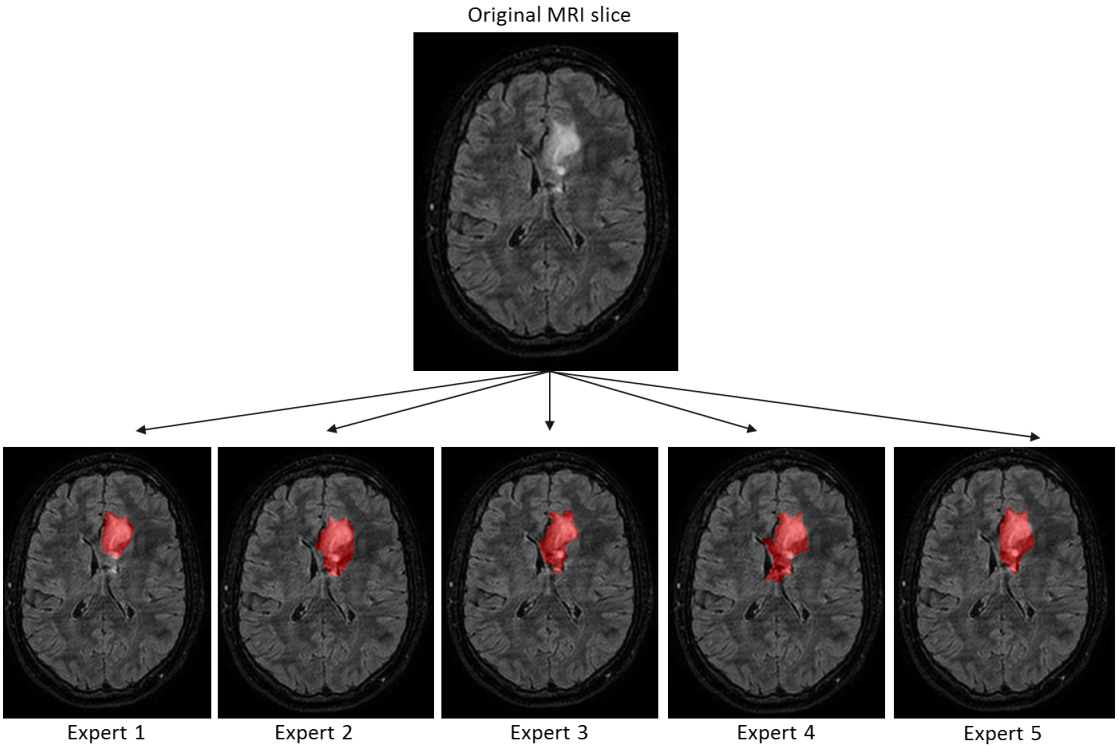


Figure 2.1: Slice-by-slice manual segmentations of Low Grade Glioma Brain Tumor performed by 5 medical experts

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 Error: Let be M_1^i M_2^i M_3^i the 2D or 3D masks obtained from the segmentation along axial sagittal and coronal plane respectively, performed by the $i-th$ expert. The Intra- and Inter- spatial distribution error is evaluated in terms of Jaccard Distance [81], are computed as follows:

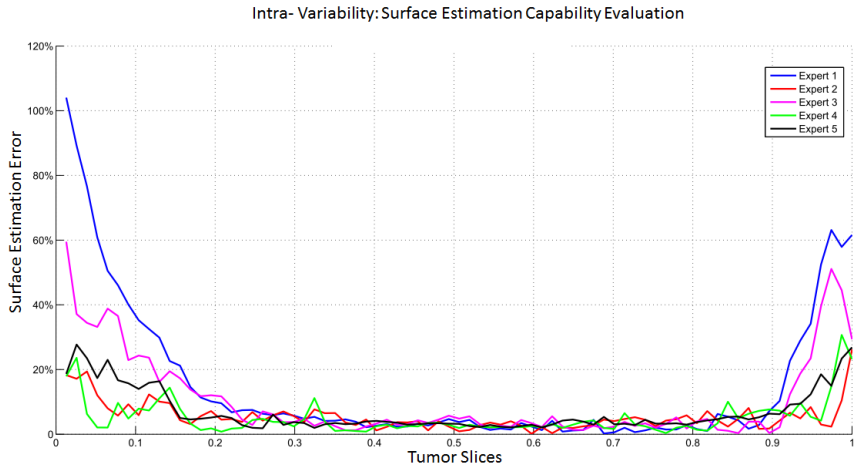
$$J_{p,t}^i = 1 - \frac{M^i p \cap M^i t}{M^i p \cup M^t i}; \quad J_p^{i,j} = 1 - \frac{M^i p \cap M^j p}{M^i p \cup M^j p}; \tag{2.2}$$

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

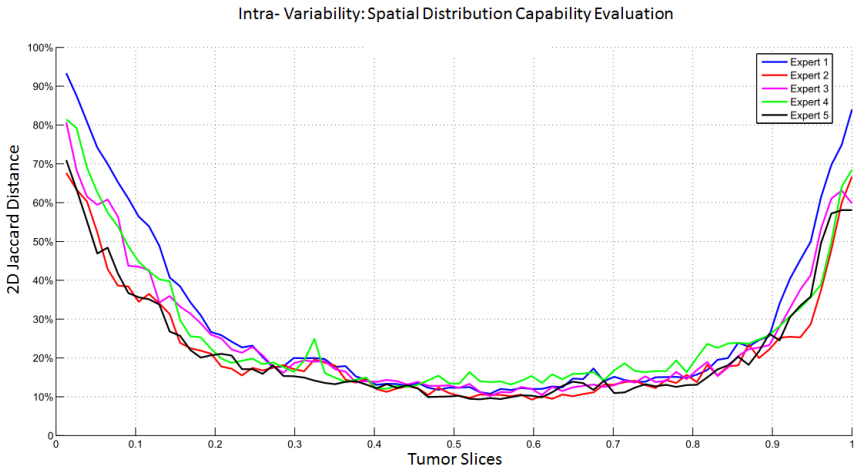
2.2.1 2D Inter- Intra- Variability Evaluation

Figure 2.2(a) shows the mean of the the intra- size estimation error for each expert and for each tumor slice of one MRI volume in the data set computed in terms of mean of $intraSizeErr_p^i$ varying segmentation plane p.
Figure 2.2(b) shows the mean of the spatial distribution error for each expert and for each experts for the same volume computed in terms of mean of the $J_{p,t}^i$ varying al the possible pair of plane p, t.
Both the intra- variability measures adopted confirm consistently an acceptable level of reproducibility for what concerns the central slices of the tumor, with values lower then 15% and 20% for the surface estimation error and Jaccard distance respectively. The intra-variability increases considerably in the slices adjacent to the tumor boundary with peaks of 103% in surface estimation error and 92% in spatial distribution error. This results can be interpreted in the light of two main reasons; boundary masks are smaller and an error computed on few pixels results in a large percentage error; secondly boundary slices are difficult to segment showing a high level of infiltration in the healthy tissue.

Figure 2.3(a) shows the mean of the the inter- size estimation error for each volume in the data set and for each segmentation along the axial plane, computed in terms of mean of $interSurfErr_p^i$ varying the expert i.
Figure 2.3(b) shows the mean of the spatial distribution error or each volume in the data set and for each segmentation along the axial plan computed in terms of mean of the $J_p^{i,j}$ varying al the possible pair of experts i, j.
both the inter- variability measures adopted confirm a high level of variability when segmenting both central and boundary slices with picks exceeding 50% and definitely

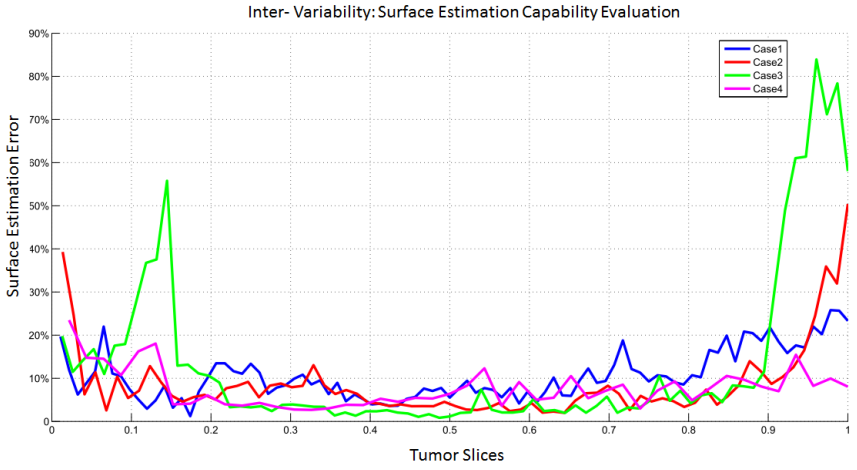


(a)

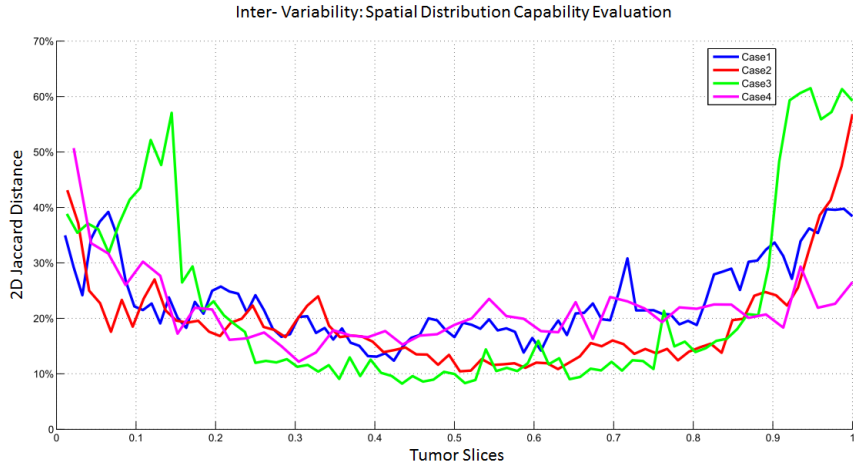


(b)

Figure 2.2: 2D Intra-Variability analysis for each expert on one MRI volume: a) Surface Estimation Error b) 2D Spatial Distribution (Jaccard Distance)



(a)



(b)

Figure 2.3: 2D Inter-Variability analysis for 4 MRI volumes: a)Surface Estimation Average Error b) 2D Spatial Distribution (Jaccard Distance)

unacceptable results in the boundary slices.

2.2.2 3D Inter- Intra- Variability Evaluation

Table 2.1 reports the results of intra-variability analysis both in terms of volume estimation error and 3D spatial distribution for 2 cases of the dataset. Volume estimation results shows an acceptable level of inter- variability. Otherwise the Jaccard Distances indicate a high level of variability in spatial distribution. The inconsistency of the two metrics revile the compensation of omission and commission errors.

Table 2.1: 3D Intra-Variability analysis for each expert on two MRI volumes

	Case 1		**Cace 2**	
	Volume Estimation Error	**3D Jaccard Distance**	**Volume Estimation Error**	**3D Jaccard Distance**
Expert 1	0.32%	23.33%	1.20%	15.03%
Expert 2	4.58%	24.67%	2.46%	24.33%
Expert 3	6.46%	24.33%	6.09%	24.33%
Expert 4	1.68%	22.00%	4.83%	16.33%
Expert 5	7.79%	25.00%	9.10%	20.00%

Table 2.2 reports the results of the inter-variability analysis for all the 4 cases for all the 5 experts of the dataset. The results obtained lead to the same conclusion made in the previous case. The low variance values equal to 0.14% and 0.10% computed on the volume estimation and the 3D spatial distribution errors respectively indicates that dissimilarities are equally distributed on the experts. Anomalous behaviors of individuals or sub-groups of experts (i.e. neuroradiologists and neurosurgeons) can not be identified.

Table 2.2: 3D Inter-Variability analysis for 4 MRI volumes

	Volume Estimation Error	**3D Jaccard Distance**
Case 1	7.20%	21.52%
Case 2	5.72%	15.56%
Case 3	3.00%	16.08%
Case 4	1.80%	20.78%

2.3 Discussion of the Uncertainty Components In Manual MRI Segmentation

Results obtained were discussed and interpreted by means of a close dialogue between physicians and computer scientists and joint meetings in which working sessions were

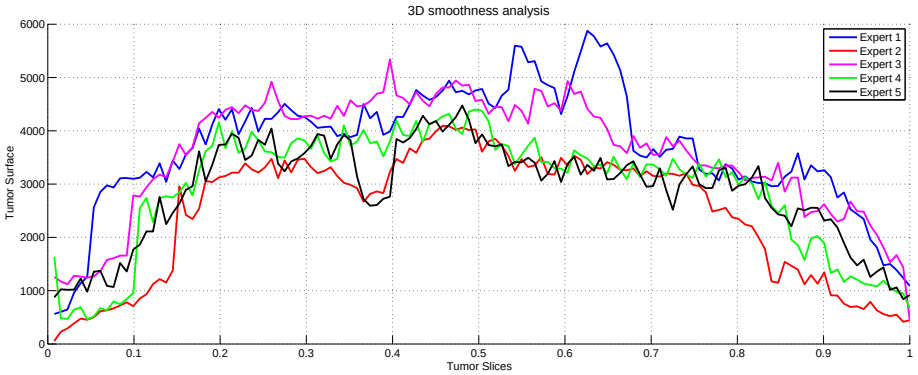


Figure 2.4: Trends of the areas of tumor sections manually segmented along the axial direction by a group of experts

held. Our analysis substantiate the general claim that the inter- observed variation strongly limits the reliability of the truth labels produced by individual segmentation and create difficulties in combining them for reference estimation [186, 141, 26].

Dissimilarity among experts have been traced back to two main sources. A first source of uncertainty is identified in the lack of information during the visual inspection phase. Considering the trend of the areas of tumor sections, manually annotated by each expert and reported in Figure 2.4, we may notice large transitions between consecutive slices, indicating non-compliance with the constraint of continuity. We concluded that physicians should explore a resonance volume for subsequent axial coronal and sagittal slices and the inspection on the given slice must be contextually related on the inspection of previous and subsequent slices. The second independent source of uncertainty has been identified within the decision process with which an expert decides whether to assign a region under a given category basing on complex and/or intrinsically vague clinical signs. The assignment of crisp labels is accomplished by experts, arbitrarily reducing uncertainty and forcing a boolean decision. We concluded that label assignment is a matter of degrees and we insert manual labeling procedure and reference label estimation within the fuzzy set framework.

Proceeding from this experimental experience in the quantitative evaluation of expert inter- intra-variability, we focus our attention on a new common agreement estimation method for reference standard definition in brain MRI segmentation.

A great relevance has been given in this Ph.D Thesis to the collection of a reliable Ground Truth, essential step for the validation of automated methods proposed in the following. It is our conviction that the solution proposed must take motivation from the different sources of uncertainty that characterize the labeling process and that cause dissimilarity in individual manual segmentation results. As mentioned above, the considerable

inter- observer variation which strongly limits the reference determination process, is in general composed of two interdependent components. A first component is related to technical, sensory gaps such as MRI imaging conditions, limitations in the visualization process during the acquisition process. In this case the solutions must be sought in the context of tools that allows to view and efficiently manipulate MRI volumes and to collect reliably expert data. The second uncertainty component is related to the intrinsic subjective character of the visual inspection and interpretation process with which an expert decides whether to assign a region under a given category. The first source of uncertainty is addressed by proposing tools that allows to view and efficiently manipulate MRI volumes and to collect reliably expert data.
The uncertainty arising in the visual interpretation and decision process is managed within the fuzzy set framework.
Images are, by nature, fuzzy, and object intensities come from different factors such as the material heterogeneity of the object and from degradation introduced by the imaging device. Under these critical conditions, the labeling process have to be conceived vague in nature and this kind of uncertainty have to be properly modeled, e not reduced in order to completely represent the expert decision attitudes in connecting heterogeneous image elements forming objects and rationally attribute decision labels. Fuzzy sets techniques have had an enormous impact on medical diagnosis in which uncertainty plays a key role [164, 136, 20]. The Fuzzy-Connectedness theory and algorithms [176]. In particular attempt to capture the intrinsically vague processes underlying visual inspection and interpretation and provide a computational framework usually proposed for semi-automatic three-dimensional (3D) image segmentation tasks. The method is minimally interactive and operation interaction consists in selecting representative seed points that is, seeding the segmentation within the object of interest. Results obtained in previous studies are very satisfactory and promising but the method suffer from the limitations common to semi-automatic approaches. We make a different and, at best of our knowledge, novel use of fuzzy connectedness which is is conceived here as a method of multiple label fusion within a common agreement estimation strategy. Labels are provided by the experts who are asked to manually trace few representative points within the objects of interest. Starting from these multiple seeds the fuzzy connectedness algorithm computes the segmentation. The proposed strategy inherits the advantages of fuzzy connectedness segmentation method which is naturally oriented to integrate uncertain information and then it is expected to manage dissimilarity among input labels; interaction is drastically limited with respect to a complete manual tracing and the formal fuzzy framework supports in the overall process of estimation without arbitrary solutions. In the follow we present *GliMAn*(**Gli**al Tumor **M**anual **An**notator). *GliMAn* is a MATLAB® application that allows to view and manipulate MRI volumes and integrates the interdependent technical and methodological solutions conceived for validating glial brain

tumor segmentations. Other than the above mentioned experimental investigation of the operation variability in manual labeling, a set of experiment have been also conceived and conducted with the purpose of quantitatively assess the reliability of our solutions and their contribution in the process of truth label collection and estimation.

2.4 Fuzzy connectedness-based common agreement strategy for reference segmented data estimation

In the next subsection we summarize the formal ingredients of the Fuzzy Connectedness framework. Subsequently, we illustrate the segmentation strategy based on these principles, and in the last subsection we present a specialization procedure for adapting the original framework to the context of reference estimation by common agreement.

2.4.1 Fuzzy connectedness principles

In the task of image segmentation, as presented by Udupa et al. in [176], the Fuzzy Connectedness method works to define as realistically as possible the global relation of *hanging togetherness*, between each pair of image elements.
In this framework, an n-dimensional image is represented as a *scene* $\mathcal{C} = (C, \mathbf{f})$ over a digital space, where C is an n-dimensional matrix of space elements, aka *spels* (n=2 pixels, n=3 voxels) and $\mathbf{f}$ is a function called *scene intensity*, which assigns to every spel an intensity value, chosen from a set of integers.
In this space, any pair of spels $(c,d) \in C$, has a degree of adjacency defined by a *fuzzy adjacency relation* α and defined by a membership function $\mu_\alpha(c, d)$. This relation assigns to every pair of spels $(c,d) \in C$, a value between zero and one inversely related to their distance. An example of *fuzzy adjacency relation* for a 3D digital space, is given below:

$$\mu_\alpha(c, d) = \begin{cases} 1, & \text{if } c = d, \text{ or } c \text{ and } d \text{ are 6-adjacents,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

The Fuzzy Connectedness is a *global* fuzzy relation, based on a *local Affinity* fuzzy relation k defined by the membership function $\mu_k(c, d)$. Affinity is defined between each pair of spels $(c,d) \in C$ and $\mu_k(c, d)$ which express how they are similar (*local hanging togetherness*), because of their spatial adjacency ($\mu_\alpha(c, d)$), and intensity values *f(c)*, *f(d)*.
As presented in [147], given $(c,d) \in C$, there are several ways to define $\mu_k(c, d)$ in such away that the features of the object of interest should be at best represented. The Affinity relation constitutes the link between Fuzzy Connectedness and image segmentation; its definition is a critical aspect in the overall segmentation process strongly affecting its

performance. More precisely, other then the spatial adjacency, k should considers the degree of homogeneity between the pair of spels and their degree of membership with respect to known intensity, for the object of interest and for the background.
With these components, given a pair of spels $(c,d) \in C$, $\mu_k(c,d)$ can be formalized as follows:

$$\mu_k(c,d) = \mu_\alpha(c,d)\, g(\mu_\psi(c,d), \mu_\phi(c,d)) \tag{2.4}$$

Where $\mu_\psi(c,d)$ and $\mu_\phi(c,d)$ represent the *homogeneity based* and *object-feature-based* components of affinity between the spels c and d respectively. Both *homogeneity based* and *object-feature-based* can be considered as fuzzy relations, $\mu_\psi(c,d)$ indicates how the spels c and d belong to the same object on the basis of their similarities of intensities, whereas $\mu_\phi(c,d)$ expresses how c and d are similar on the basis of their similarities to some known object intensity.
The function g combines these two components together; any functional form ensuring non decreasing behavior can be used for this purpose; using the well known functional form g [147], we obtain the following affinity relation :

$$\mu_k(c,d) = \mu_\alpha(c,d) \sqrt{\mu_\psi(c,d), \mu_\phi(c,d)} \tag{2.5}$$

Focusing on the homogeneity-based component: given a pair of spel $(c,d) \in C$, $\mu_\psi(c,d)$ is

$$\mu_\psi(c,d) = H_\psi(\mid f(c) - f(d) \mid) \tag{2.6}$$

Where, $H_\psi(\mid f(c) - f(d) \mid)$ is monotonically non increasing function defined in $[0,1]$, with $H_\psi(0) = 1$ and H_ψ.
H_ψ must be also able to assign, to the same value of intensity difference, a proper value from 0 to 1 which takes into account the degree of heterogeneity of the region which belong its arguments.
In [147] are proposed three functional forms for H_ψ, all with a monotonically descendent trend together with the increasing of intensity difference.
as regards object-feature-based component, $\mu_\phi(c,d)$, it considers both object-feature and background feature, by means of two *membership functions*: I_o and I_b, which express the notion of *be closed to an expected value* for the object and background respectively. Given two spels c and d, they will have an high value for the object-feature base component $\mu_\phi(c,d)$, if and only if both c and d have a high value for I_o and a low value for I_b, that is if and only if the pair of spels has high object membership together with low background membership.

A functional form for $\mu_\phi(c,d)$ which express this latter condition may be the following:

$$\mu_\phi(c,d) = \begin{cases} 1, & \text{if } c = d, \\ \frac{min(I_o(c),I_o(d))}{min(I_o(c),I_o(d))+max(I_b(c),I_b(d))}, & \text{if } min(I_o(c),I_o(d)) \neq 0, \\ 0, & \text{othewise.} \end{cases} \tag{2.7}$$

To make the affinity more robust with respect to noise, the definition of this relation can involves the concept of *scale*.
The *scale*, is referred at any spel $c \in C$, and is defined as the radius *r(c)* of the largest ball centered at c, which lies entirely inside an homogeneous region. To employ a *scale-based affinity*, *r(c)* must be computed first for any $c \in C$ and subsequently, the definition for the components $\mu_\psi(c,d)$ and $\mu_\phi(c,d)$ of $\mu_k(c,d)$ will considers the pair of spels *(c,d)*, together with their neighbours, as defined by *r(c)* and *r(d)*.
The Fuzzy Connectedness K, is a global fuzzy relation which for any pair of spels $(c,d) \in C$, associates a value $\mu_K(c,d)$, between zero and one expression of the *strength of the strongest path connecting c to d*.
Given a pair of spels $(c,d) \in C$ their *strength of connectedness* $\mu_K(c,d)$ is calculated considering all possible paths between them. A path from c to d, denoted as p_{cd}, is a sequence $(s_1, s_2, ..., s_{n-1}, s_n)$ of $n \geq 2$ spels, where $s_1 = c$ and $s_n = d$ and for any pair of successive spels $(u,v) \in p_{cd}$: $\mu_\alpha(c,d) > 0$.
Each path $p_{cd} = (s_1, s_2, ..., s_n)$ connecting c to d , has a strength of connectedness, indicated with $\mathcal{F}(p_{cd})$ which is the smallest affinity of pairwise spels along that path, that is $\mathcal{F}(p_{cd}) = min((\mu_k(s_1,s_2), \mu_k(s_2,s_3), ..., \mu_k(s_{n-2},s_{n-1}), \mu_k(s_{n-1},s_n))$.
Given a pair of spels $(c,d) \in C$ and denoting as $\mathcal{P}_{cd}$ the set of all possible paths between them, employing the concept of $\mathcal{F}(p_{cd})$, the *strength of connectedness* $\mu_K(c,d)$*, between* c *and* d, can be formulated as follows:

$$\mu_K(c,d) = max_{p_{cd} \in \mathcal{P}_{cd}}\{\mathcal{F}(p_{cd})\} \tag{2.8}$$

2.4.2 Fuzzy Connectedness in Image Segmentation

The application of Fuzzy Connectedness in image segmentation, involves the concept of *Fuzzy Object*: defined as subset of spels in the scene domain having significant strength of connectedness with respect to one textitseed, which is spel representative for the object of interest.
A Fuzzy Object is represented by one seed spel, and can be *absolute*, *relative* or *iterative-relative*.
An *Absolute Fuzzy Object* indicated as $\mathcal{O}_\theta(s)$, contains all the spels in the scene $\mathcal{C}$ of the image having a strength of connectedness with respect to the seed s greater than θ, with $\theta \in [0,1]$.

The process of *Fuzzy Objects extraction*, is based on the production of the so called *connectivity scene* $\mathcal{C}_{K_s} = (C, f_{K_s})$, where for any spel $c \in C$, $f_{K_s}(c) = \mu_K(s, c)$. Algorithms which output $\mathcal{C}_{K_s} = (C, f_{K_s})$ receive as input the following parameters:

- an image represented by a scene $\mathcal{C} = (C, \mathbf{f})$,
- one or more seed spels,
- an affinity relation k.

The production of a connectivity scene, means the computation of the strength of connectedness between seeds spels and the rest of the scene.
This process can be viewed as a connectivity propagation process, which starts from seed's neighbourhood and proceed to finally reach the more farthest spels taking all the available spatial directions.
This is a onerous computational task, which involves numerous Max-Min time consuming operations and so powerful algorithms based either on Dijkstra algorithm or dynamic programming have been devised to solve this problem [38, 122].
An *absolute-fuzzy object*, is defined on the basis of a specific strength of connectedness on its own, while a *relative-fuzzy object* is based on a competition between all the co-objects that lie in the scene, to have spels as their members.
Following this strategy, each spel in the scene will have a strength of connectedness with respect to each object under competition and will belong to the object for which this value will be the highest.
In the definition of *relative-fuzzy object*, as presented by Udupa et. al. [175], the input scene is considered as composed of an object of interest over a background, that is a binary case which can however extended to a multiclass context.
First step in the definition of a *fuzzy object $\mathcal{O}$ containing a seed o relative to a background containing a seed b*, is the construction of the so called *support of $\mathcal{O}$*, indicate as P_{ob_k} and defined as follow

$$P_{ob_k} = \{c \mid c \in C \; and \; \mu_K(o, c) > \mu_K(b, c)\} \tag{2.9}$$

The support, P_{ob_k} contains all the spels having a strength of connectedness with respect to the object, represented by the seed o, greater than to the background represented by the seed b.
Once P_{ob_k} has been delineated, a *fuzzy object $\mathcal{O}$ containing a seed spel o, relative to a background containing a seed spel b* is defined as the fuzzy subset of C defined by the following membership function:

$$\mu_{\mathcal{O}}(c) = \begin{cases} \eta(f(c)), & if \; c \in P_{ob_k}, \\ 0, & otherwise. \end{cases} \tag{2.10}$$

where η, as indicated in [175], is an objecteness function with range [0,1], which express the degree of membership of each support spel to the fuzzy relative object $\mathcal{O}$, on the basis of their intensity value (f).
An *iterative-relative fuzzy object*, can be considered as an extension for the concept of relative fuzzy object, where the competition rule for the membership between object and background is properly refined within an iterative process which produces as output the so called *iterative-relative fuzzy object*. This iterative strategy has the purpose to capture more precisely as fuzzy subset of the scene, the object of interest, starting from its core part and then proceed to identify its subtle and fine parts, which usually pose segmentation challenges.

2.4.3 Seeding the segmentation by means of common seeds provided by experts

As pointed out in the previous subsection, the process of fuzzy object extraction which computes the connectivity scene, represents the core of the segmentation process and receives as inputs, other than the scene $\mathcal{C}$ the set of seeds spels, and the affinity relation k. The key point in the use of Fuzzy Connectedness for reference data estimation, is the design of a proper strategy for setting the above parameters. This strategy has to be able to represent and to embed within the fuzzy connectedness framework the different decision attitudes of the experts in manual labeling.
The affinity relation k should be in general defined on the basis of representative regions for the object of interest. In our case the affinity relation definition must proceed by collecting the regions identified by each expert as representative region and performing a suitable integration of them.
Each expert i is then called to identify by visual inspection the following set of information:

1. object and background region of interest, R_i^o and R_i^b, respectively;
2. sets of object and background seeds, O_i and B_i, respectively.

Recalling the functional forms of (2.5),(2.6) and (2.7), respectively for k and its components μ_ψ and μ_ϕ, given in the above section, the affinity is defined in this context, starting with the homogeneity-component μ_ψ.
Being $\mu_\psi = H_\psi(\mid f(c) - f(d) \mid)$ and $\delta = \mid f(c) - f(d) \mid$, for any pair of spels $(c, d) \in C$, $H_\psi(\delta)$ is defined as follow:

$$H_\psi(\delta) = \sum_{i=1}^{N} h(\delta, x_{\psi i}) \tag{2.11}$$

where N is the number of the experts and

$$h(\delta, x_{\psi i}) = e^{-\frac{\delta^2}{2x_{\psi i}^2}} \tag{2.12}$$

with $x_{\psi i} = \mu_{\delta i} + v\sigma_{\delta i}$. $\mu_{\delta i}$ and $\sigma_{\delta i}$ are mean and standard deviation of the intensity differences, calculated between adjacent spels within each expert region of interest R_i^o. v, is a number between 2 and 3, in this context with a value of 2.5, expression of the concept that in a normal distribution, 2-3 standard deviations on the two sides of the mean cover 95.4-99.7% of the population.
Depending from the homogeneity component $h(\delta, x_{\psi i})$ defined by each expert i, $H_\psi(\delta)$ then represents the aggregation of all the heterogeneity descriptions made by every expert.
For what is concerned with the feature-based-component μ_ϕ, as defined by (2.7), its *object-membership function* I_o is computed with the following formula: (the process for I_b is analogous)

$$I_o = \sum_{i=1}^{N} w_i G(\mu_i, \sigma_i) \tag{2.13}$$

Where: μ_i and σ_i, parameters of a gaussian membership function G, are the mean and standard deviation of the intensity values within each expert region of interest R_i^o, N is the number of experts and the weight w_i is estimated as follow:

$$\bar{w_i} = \sum_{j=1}^{M} \frac{1}{d_{ij}} \tag{2.14}$$

where M is the number of seeds, fixed for any expert and d_{ij} is the distance between intensity values of the seed $o_j \in O_j$ and the mean μ_i. All the weights $\bar{w_i}$ are then normalized to define w_i
This formulation allows to take into account the performance level of each expert measured in terms of consistency between seeds and region indicated as object of interest. Consistent information will play a major role in the definition of the feature based component of affinity.
With these components so defined, the affinity k is employed in the process of producing a *fuzzy collective truth*, through the definition of a connectivity scene for the object of interest. This process employs as set of seeds $S_o = \bigcup_{i=1}^{N} O_i$, namely the sets resulting from the union of the seeds selected by every expert for object of interest. The connectivity scene so computed, is then used to produce a *fuzzy truth estimation*.

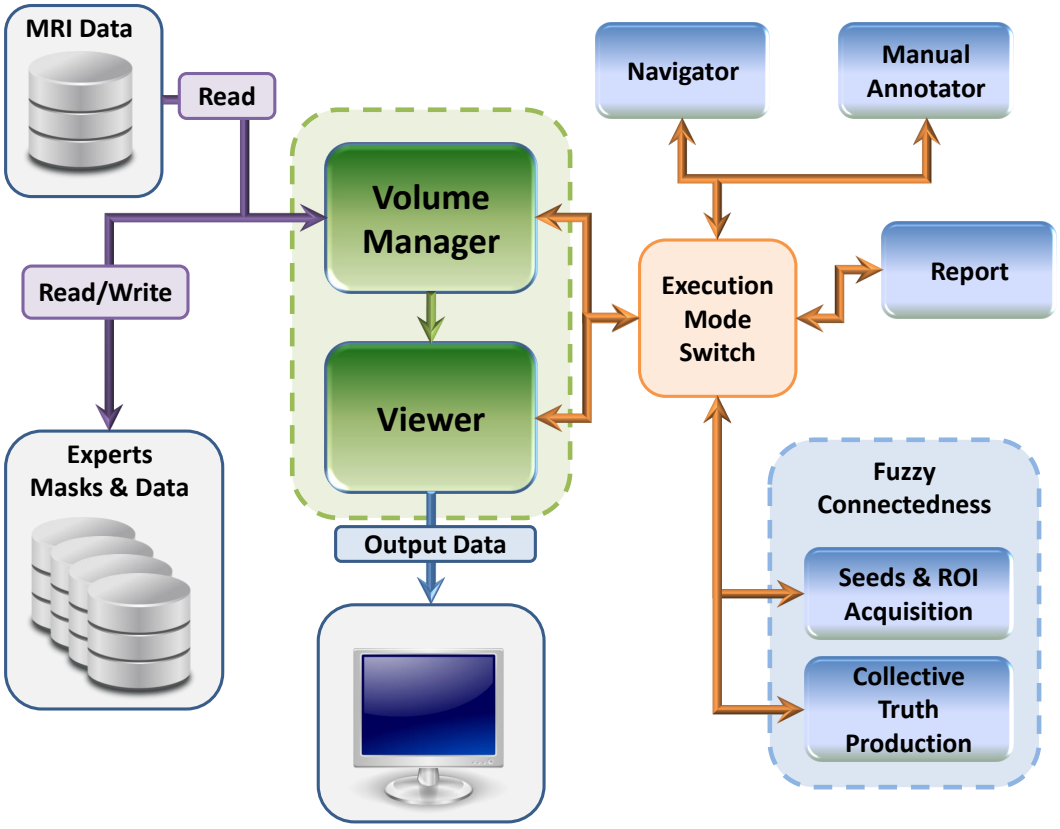


Figure 2.5: *GliMAn* Architecture diagram.

2.5 GliMAn Design and Architecture

GliMAn is a MATLAB application for Ground Truth labeling and estimation, conceived and developed on the basis of conclusions reached in the analysis of observer variations in MRI Glial Tumor segmentation. As detailed in Figure 2.5, The *GliMAn* architecture includes modules for viewing, manage and process MRI volumes supporting the following main tasks:

1. fully manual labeling of glial tumors by individual experts
2. semi-automatic generation of reference segmented data through the use of fuzzy connectedness-based common agreement assessment

Conceptual design phase of *GliMAn* proceeds from the conclusion that experts should explore a resonance volume for subsequent axial coronal and sagittal slices and the decisions on the given slice must be contextually related on the inspection of previous and subsequent slices.

The main feature of *GliMAn* is then 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 input labels. 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 that could undermine 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 2.5). Moreover the method of the orthogonal projections

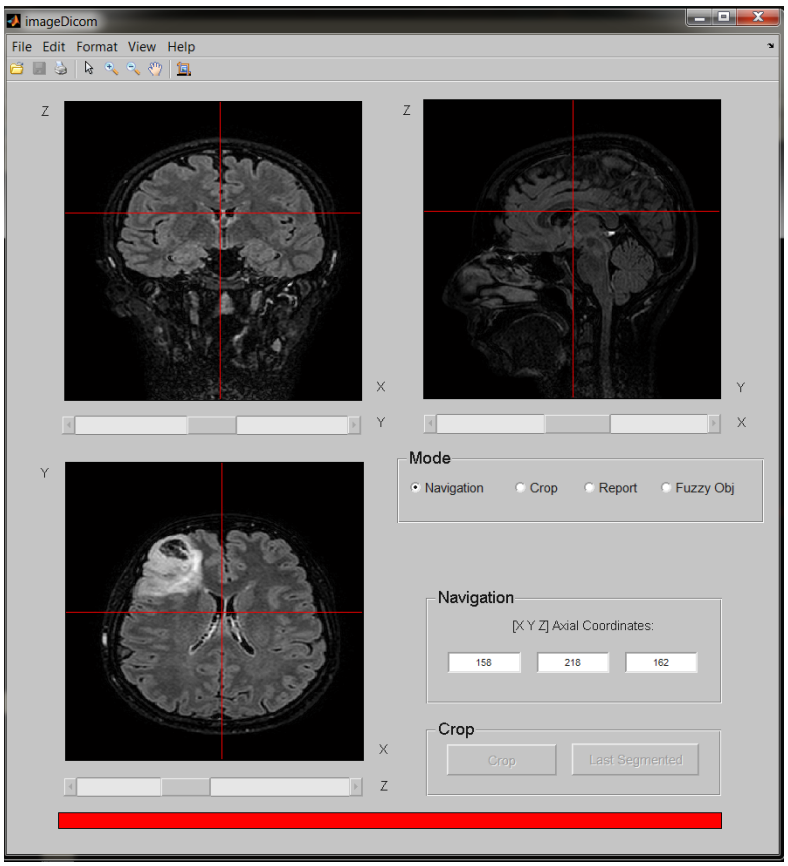


Figure 2.6: *GliMAn* Graphical User Interface(GUI): Central Zone in Navigation Mode.

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 $(x_0; y_0; z : 0)$, the three images

represented are the intersection of the MRI volume with the sagittal coronal and axial planes respectively passing through the point.

2.5.1 Fully manual labeling of glial tumors by individual experts

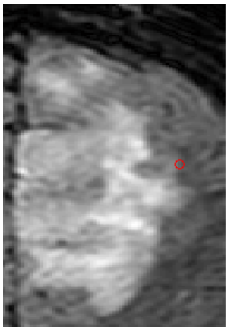


Figure 2.7: Axial slice of a critical case example.

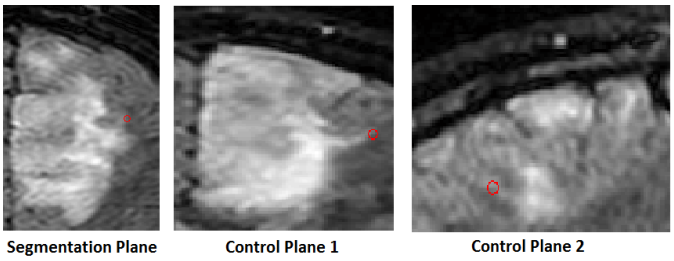
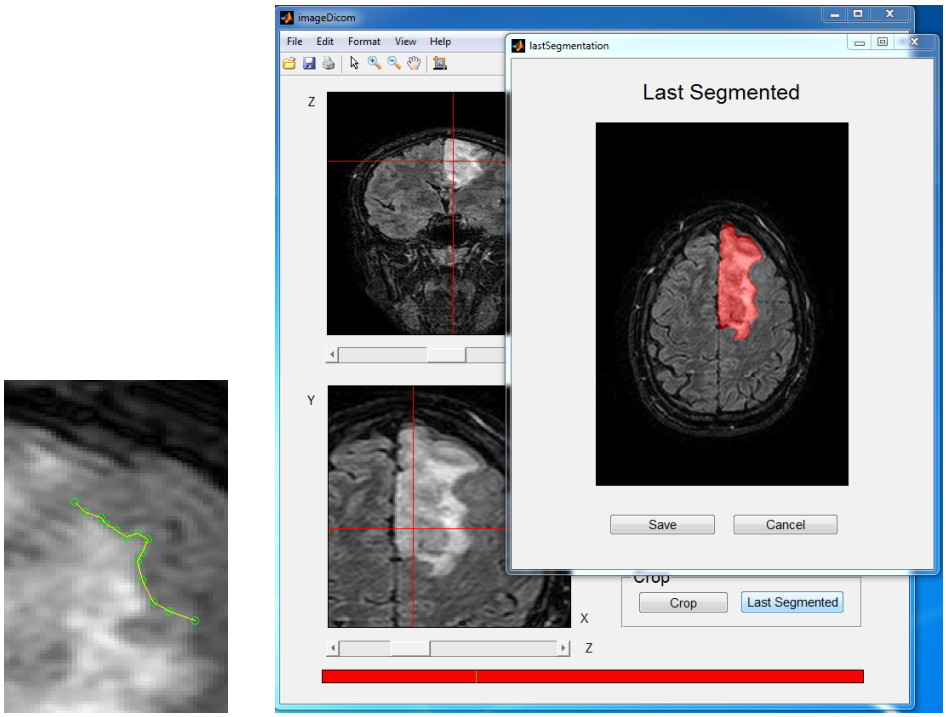


Figure 2.8: Example of critical case solved using the control orthogonal planes.

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 it is organized as follows:

- the user points and selects a candidate point in a given plane;
- the same point is highlighted in the other two planes and analyzed;
- the expert confirms the decision by re-selecting the same point or decides to examine another point.

Figure 2.7 shows a crop of a brain MRI axial section of a Low Grade Glioma Tumor, 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 enough to make a reliable decision about the point identified with the red circle to the edge of the tumor. As



(a) (b)

Figure 2.9: *Gliman* Manual Segmentation interactive procedure: a) broken line manual tracing b)Segmentation superimposing on the original MRI.

illustrated in Figure 2.8, the visualization 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 (Figure 2.9(a)). 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 (Figure 2.9(b)). During the segmentation of the $N-th$ slice, the segmentation performed on the slice $N-(th-1)$ can be viewed as a guide.

2.5.2 Semi-automatic generation of reference segmented data through the use of fuzzy connectedness-based common agreement assessment

The semi-automatic generation of reference segmented data based common agreement assessment is composed of two phases:

- the collection of initialization by every expert,

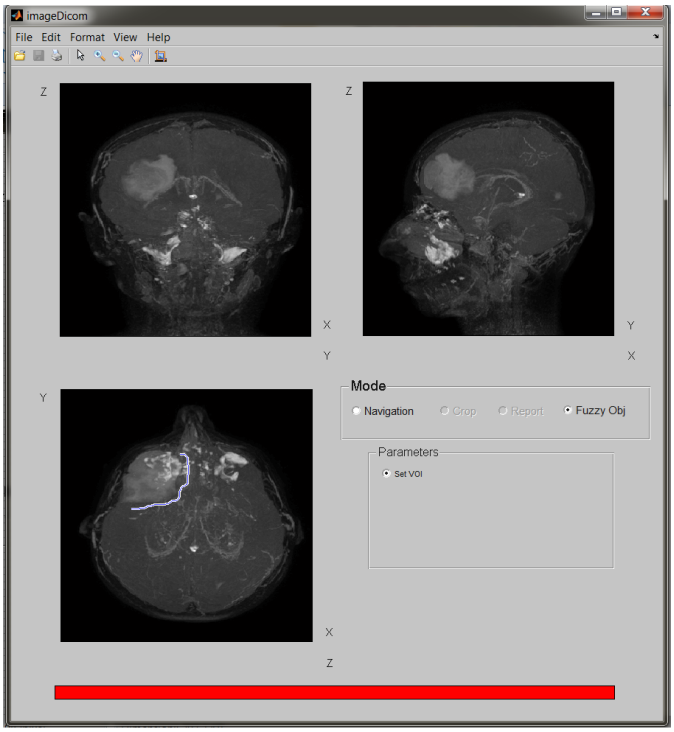


Figure 2.10: Axial sagittal and coronal Maximun Intensity Projection(MIP) Images shown by *GliMAn* for the Volume Of Interest(VOI) identification.

- the fusion of initialization and execution of the fuzzy connectedness.

To accomplish the first stage, *GliMAm* provides a specific execution mode *Fuzzy Obj*. The transition in this mode imposes a visualization change. The Maximum Projection Intensity Images (MIP) computed in axial, sagittal and coronal direction are shown on the 3 plans (Figure 2.10). The MIP is a volume rendering method for 3D data that projects in the visualization plane the voxels with maximum intensity that fall in the way of parallel rays traced from the viewpoint to the plane of projection [185].
Each user draw a Region of Interest on each plane in whose the tumor is included. Through the intersections of the 3 projection a *Volume of Interest* is identify and the display of orthogonal planes moves in the VOI extracted from the original MRI. Each user must now identify, selecting them from a mouse click, a set of *object and background seed*. Also in this phase the points position display is synchronous on the three planes ensuring a more robust choice of the points of interest; in addition to the seeds will be also identified *two regions* characterizing the object and the background. All initialization parameters are then stored using a unique identifier of the user.
The second phase involves the loading of all the initialization from the data base and the common agreement initialization building using the procedure detailed in section 3.3.

and the execution of fuzzy connectedness for the generation of reference segmentation. Two display modes are available: object and background *Absolute Fuzzy Object*, with an interactive tool for apply threshold values for creating a crisp segmentations; and the *Relative Fuzzy Object*, that displays only the points for which the membership to the object is stronger then the membership to the background. Each user then has the opportunity to display the superimposing of the manual segmentation produced by him and the fuzzy object that says the common opinion of all experts.
The use of the*GliMAn* system and the proposed strategy for the fuzzy estimation of a common agreement ground truth allow us to estimate a reliable reference data with which to evaluate all the segmentation strategy included in the *Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation* detailed in the following section. It was decided to treat very deeply this first part regarding the collection of the truth considering the the particular field of application of the proposed strategies.

2.6 How to use the Fuzzy Connectedness truth estimation

The intimate aim of the present study was the definition of a simultaneous truth estimation strategy based on Fuzzy connectedness principle and valuable for MRI brain tumor segmentation studies. We would like to emphasize, however, that the proposed fuzzy framework actually may play the role of a semiautomatic segmentation tool in which a set of experts provided multiple initialisation information.
Focusing however on to the use for which it was conceived, we may outline potential employments within the MRI brain glial tumor studies. The richness of information that characterizes the fuzzy estimation is particularly valuable for a direct feedback to individual experts in performing the complex task of MRI tumor segmentation. As exemplified in Figure 2.11 the superimposition of the manual traced boundary on the fuzzy segmentation output allows to synthetically compare the individual decision with the consensus decision.
Moreover, the simultaneous visualization of fuzzy grades and crisp labels allows to easily identify critical regions in which the labeling was performed under a high level of uncertainty. Basing on this evaluation, the experts may operatively update labels and/or acquire new insights.

Consider now the case in which the proposed fuzzy connectedness method is adopted as reference standard to quantitatively evaluate segmentation results, manually or automatically produced. The use of conventional metrics for agreement measures, involves a hardening process to obtain final crisp partitions, causing inevitably loss of information. An alternative is offered by able keep track of the uncertainty underlying the segmentation task. As regards our context, we propose the use of fuzzy set based metrics for comparing results obtained by fuzzy simultaneous estimations and crisp segmentations.

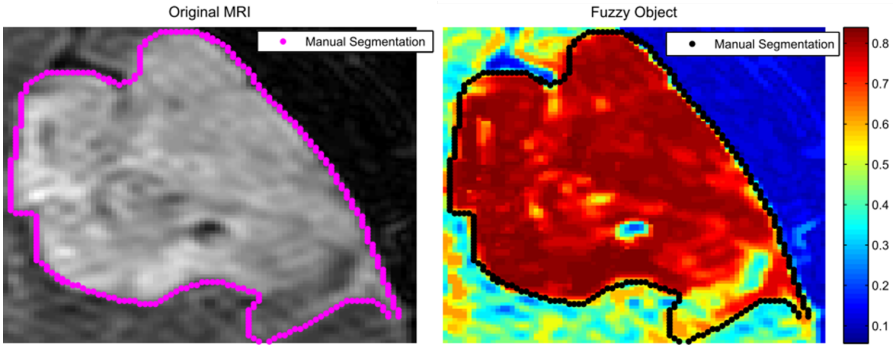


Figure 2.11: Fully manual segmentation boundary traced on brain MRI and superimposed on the fuzzy segmentation output

Considering first the comparison of hard segmentations, the minimal problem of assessing the agreement between two binary maps B_1 and B_2, representing reference and segmented data respectively is traditionally done by measuring the number of voxels at which both B_1 and B_2 score "1" (True Positive T_p) or "0" (True Negative T_n), the number of voxels at which B_1 scores "0" and B_2 scores "1" (False Positive F_p) and viceversa (False Negative F_n). These measures can be accommodated in 2×2 confusion matrix, as follows:

	1	0
1	T_p	F_n
0	F_p	T_n

Let R represents the set of locations of voxels labeled "1" in B_1 and S the set of locations of voxels labeled "1" in B_2. Measurements allocated in the matrix can be interpreted in the light of Set Theory obtaining the following expressions: $T_p = |R \cap S|$, $T_n = |\bar{R} \cap \bar{S}|$, $F_n = |R \cap \bar{S}|$ and $F_p = |\bar{R} \cap S|$.
The applicability of the confusion matrix can be easily extended to the case of fuzzy agreement assessment by expressing their coefficients with Fuzzy Set Theory ingredients, according to the formal framework presented in this work [21].

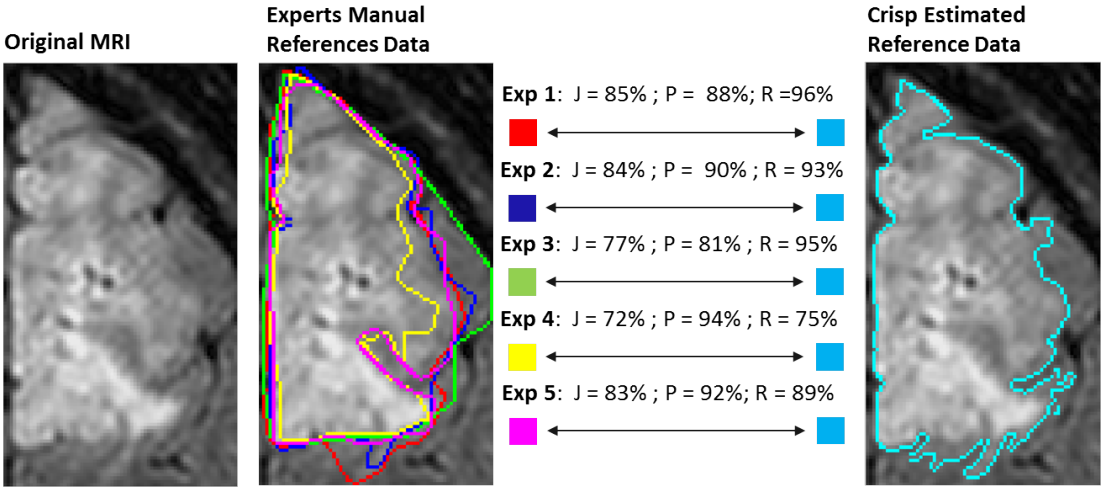


Figure 2.12: Example of an evaluation procedure performed using the crisp metrics Jaccard (J) Precision (P) Recall (R).

Let $v \in V$ be the generic voxel belonging to the total volume V:

$$\begin{aligned}
\mathcal{T}_p &= |\mathcal{R} \cap \mathcal{S}| = \sum_{v \in V} min(\mu_{\mathcal{R}}(v), \mu_{\mathcal{S}}(v)); \\
\mathcal{T}_n &= |\bar{\mathcal{R}} \cap \bar{\mathcal{S}}| = \sum_{v \in V} min(\mu_{\bar{\mathcal{R}}}(v), \mu_{\bar{\mathcal{S}}}(v)); \\
\mathcal{F}_n &= |\mathcal{R} \cap \bar{\mathcal{S}}| = \sum_{v \in V} min(\mu_{\mathcal{R}}(v), \mu_{\bar{\mathcal{S}}}(v)); \\
\mathcal{F}_p &= |\bar{\mathcal{R}} \cap \mathcal{S}| = \sum_{v \in V} min(\mu_{\bar{\mathcal{R}}}(v), \mu_{\mathcal{S}}(v))
\end{aligned} \tag{2.15}$$

where:

- $\mu_{\mathcal{R}}$ and $\mu_{\mathcal{S}}$ are the membership function associated to the fuzzy objects $\mathcal{R}$ and $\mathcal{S}$ defined in reference and under evaluation segmentations respectively
- $\mu_{\bar{\mathcal{R}}}$ and $\mu_{\bar{\mathcal{S}}}$ are the membership function associated to the fuzzy background $\bar{\mathcal{R}}$ and $\bar{\mathcal{S}}$ defined in reference and under evaluation segmentations respectively

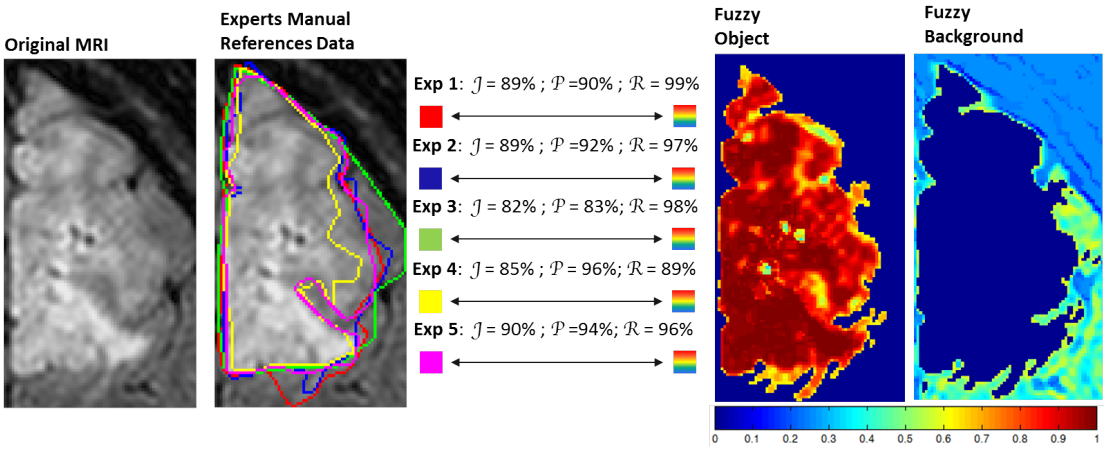


Figure 2.13: Example of an evaluation procedure performed using the fuzzy metrics $\mathcal{J}accard(J)$ $\mathcal{P}recision(P)$ $\mathcal{R}ecall(R)$.

The fuzzy variant of Jaccard, Precision and Recall indexes is then formulated as follows:

$$\begin{aligned} \mathcal{J} &= \frac{\mathcal{T}_p}{\mathcal{T}_p + \mathcal{F}_n + \mathcal{F}_p} \\ \mathcal{P} &= \frac{\mathcal{T}_p}{\mathcal{T}_p + \mathcal{F}_p}; \\ \mathcal{R} &= \frac{\mathcal{T}_p}{\mathcal{T}_p + \mathcal{F}_n}; \end{aligned} \tag{2.16}$$

Figures 2.12 and 2.13 present an example of crisp and fuzzy evaluation of segmentation results. In particular, Figure 2.12 compares manual segmentations provided by experts with the "support" of the fuzzy object computed by the fuzzy connectedness simultaneous estimation strategy. The agreement is computed in terms of conventional crisp agreement indexes, Jaccard, Precision and Recall. In Figure 2.13 manual segmentations are instead compared with the fuzzy partition produced by the fuzzy connectedness strategy using the fuzzy variant of the above crisp indexes.
Focusing on the agreements reported in Figure 2.12, the crisp indexes highlight low performance for the manual segmentation performed by the expert 3 for which the combined analysis of Jaccard (77%) and Precision (81%) values denotes a overestimation and by the expert 4 and 5 for which the combined analysis of Jaccard and Recall values ($J = 72\%$ and $R = 75\%$ for expert 4, $J = 83\%$ and $R = 89\%$ for expert 5) denotes a underestimation.
Focusing on the agreements reported in Figure 16, the fuzzy evaluation differs significantly from the crisp evaluation when computing Jaccard and Recall indexes for the manual segmentations performed by the experts 4 and 5 with an increment of the 12,88%

and 12,96% for the Jaccard and Recall indexes related to expert 4 and 6,95% and 6,46% for the Jaccard and Recall indexes related to expert 5. The less conservative behavior of fuzzy measurements can be interpreted in the light of the fact that misclassification errors occur in regions having a low degree of membership to the fuzzy object. These degrees concur to the definition of the fuzzy quantities in the fuzzy Jaccard and Recall indexes with resulting measures quite different from the crisp case. Symmetrically to these cases, the crisp and fuzzy evaluation of the manual segmentation of the expert 3 are quite similar as the overestimation error are located in regions having a high degree of membership to the fuzzy background.
Considering globally the results obtained with the fuzzy and crisp agreement measurements, we may conclude that they agree when the errors few and occur in regions of low uncertainty; they disagree in the cases in which errors occur in regions of high uncertainty. Proceeding from these observations, we propose for a complete evaluation a combined use of the crisp and fuzzy indexes that allows not only to quantify the errors but also to evaluate their severity.

2.7 Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation

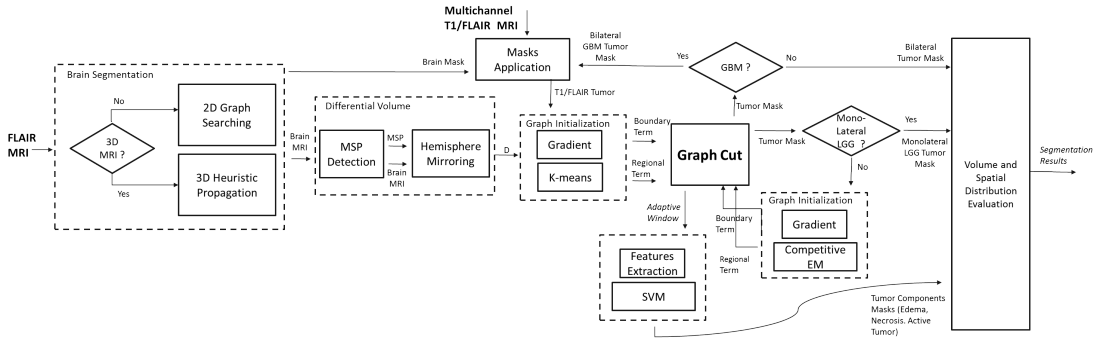


Figure 2.14: Overall schema of the Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation

After having extensively dealt with the problem of how to derive and use a reliable Ground Truth, the proposed Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation is presented. The main purpose of our work is to develop an integrated framework able to analyze volumetric MRI for detecting, segmenting and evaluating volume and spatial distribution of different class of Glial tumors. The attempt is to execute the various stages in a fully automatic way, with the goal to make the procedures as much as possible objective and reliable.

In Figure 2.14 the block diagram of the overall procedure is shown. In this diagram, only the main features of the framework are mentioned but several ancillary tools are implemented in the framework required for the management visualization and processing of the MRI volumetric data.
The Main steps of the execution pipeline are here summarized and detailed in the follows:

Brain Segmentation The first step of our execution pipeline is the brain segmentation. Two segmentation modalities are provided depending on whether the volume that is being analyzed is a "dense" or "sparse" MRI. We identify with "dense" the volumes for which the *spacing between slice* is equal to the *slice thickness*, then the discrete volume is completely defined; for this type of MRI a 3D approach is studied. The sparse MRI volumes are those for which between the slices acquired there is a layer of air. For these volumes the topological constraints in the direction orthogonal to the acquisition planes must be weak or non-existent. Both the brain segmentation strategy are based on the explicit description of the brain boundary and to graph searching principles. The first is a 2D technique whose input data will be sparse MRI volumes and the second is the volumetric extension of the first solved with an heuristic propagation technique to be applied to the dense volumes [129, 128].

Differential Volume Estimation A key point of our tumor segmentation strategy is the analysis of asymmetric components; therefore an important step of the overall procedure is the automatic identification of the Mid Sagittal Plane(MSP) and the computation of a differential volume able to highlight the asymmetric components.

Tumor Segmentation The well know Graph Cut technique is applied to the volume computed in the previous step in a non conventional way, obtaining a fully automatic tumor segmentation procedure. As shown in the schema in Figure 2.14 the Graph cut is the core of all the segmentation procedures applied in the framework. The novelty of our work lies in the automatization of the initialization procedure, usually supplied by the user in the classical Graph Cut.
Some of the solutions found will be closely linked to the problem addressed and other are more general purpose [130]. The technique applied for solving the tumor segmentation operates under the assumption that the tumor is located in only one hemisphere. This constraint is not met by a particular class of tumors. For overcoming this problem a new hybrid formulation of the Competitive Expectation Maximization and Graph Cut is studied.

Recognition of the heterogeneous parts of the tumor The last step of our strategy concerns the recognition of the heterogeneity that compose the tumor (*Active Tumor, Edema and Necrose*). This phase is accomplished using a Multichannel

supervised technique. the Machine Learning tool used is the Support Vector Machine (SVMs). Our effort and the novelty of this application lies in the study of a graph based technique used for extract a features vector able to describe in an adaptive way the contextual information of a single voxel.

2.7.1 Automatic Whole Brain Segmentation by Graph Searching

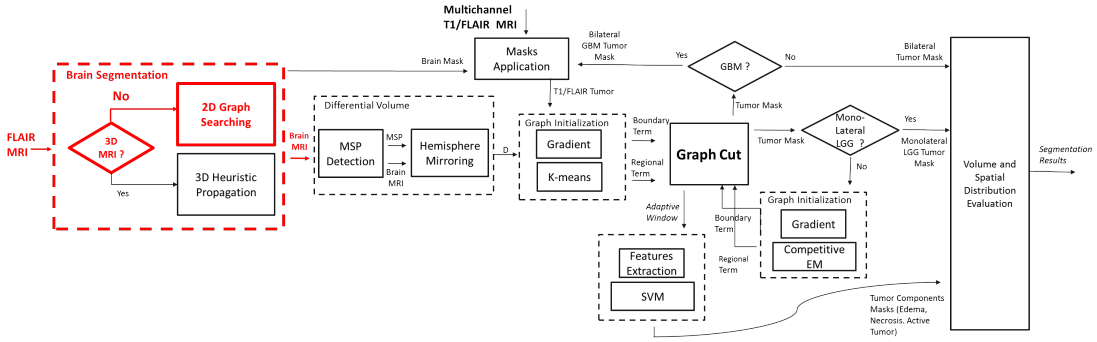


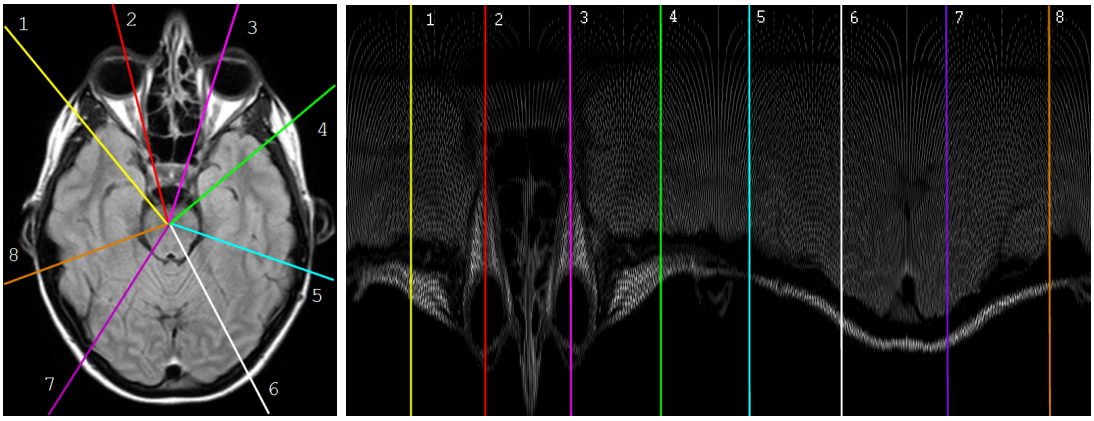
Figure 2.15: Schema of the Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation with highlight the part detailed in Section (2.7.1)

Accurate and efficient segmentation of the whole brain in magnetic resonance (MR) images is a key task in many neuroscience and medical studies either because the whole brain is the final anatomical structure of interest or because the automatic extraction facilitates further analysis [6, 88]. Whole brain segmentation is a task available in software such as Brain Visa [190], FSL [155] and Brainsuite [151], however in most cases they give poor results especially in case of tumors located on the border of the brain [88].

We focus our attention on fully automatic whole brain segmentation from MRI volumes, specifically addressing those critical situations in which the presence of pathologies such tumors of great dimension and/or with location close to the border causes significant morphological deformations and consistent alteration of homogeneity. We derive our method from the edge-based approach adopting Graph Searching technique for 2D boundary detection proposed by Martelli [112]. The solutions investigated in this paper are an extension of those adopted in a previous work [116] from which we inherit the graph searching model. In our strategy the MRI volume is processed slice by slice, the overall segmentation procedure is made fully automatic and easily reproducible by the automatic detection of frames including eyes and by computing the internal main parameters directly from the image data.

Graph Searching Technique for Radial Contour Detection

The salient aspect of our approach to whole brain segmentation is the use of graph searching for identifying optimal brain borders. Brain contour has a radial shape that can be conveniently treated by using polar coordinates (Figure 2.16). We proceed then by



(a) (b)

Figure 2.16: Polar conversion of a brain MRI image, a) original image b) polar converted image

formulating graph searching conceptual ingredients for contour detection in polar space [112, 159]. Let be $I_{cart}(x, y)$ an image described on a square cartesian discrete grid Figure 2.16(a). The image can be 'unwrapped' converting it in polar coordinates, obtaining $I_{pol}(\rho, \theta)$ where $1 \leq \rho \leq M$, $1 \leq \theta \leq N$, $\rho = \sqrt{(x - x_o)^2 + (y - y_o)^2}$, $\theta = arctan(\frac{y-y_o}{x-x_o})$ and (x_o, y_o) coordinates of the centroid of the object Figure 2.16(b). Working in polar space, the radial boundary of a given object can be represented by a transformation $\rho = f(\theta)$ characterized by the following feasibility constraints:

Boundary as a Function - $f(\theta)$ is single valued and the value ρ exists for each θ with $1 \leq \theta \leq N$

Connectivity Constraint - $|f(\theta + 1) - f(\theta)| \leq 1$ for $1 \leq \theta \leq N - 1$

Closing Constraint - $|f(1) - f(N)| \leq 1$ imposing that the first and last pixels satisfy the connectivity constraint.

Each feasible function $f(\theta)$ is a candidate object boundary. The goal is then to find the minimum cost boundary subject to the feasibility constraints. The boundary detection task within the graph searching framework is modeled by embedding the properties of the boundary in a cost function and formulating the boundary extraction as the problem of minimizing this function subject to the feasibility constraints. The boundary cost is

defined as follows:

$$B_{cost} = \sum_{\theta=1}^{N} C(\theta, f(\theta)) \tag{2.17}$$

where $C(\theta, f(\theta))$ is a cost image. The value of each pixel in the cost image must be inversely related to the likelihood that an edge is present at that point. The likelihood is usually determined by the application of a low-level local edge operator [170]. In general, the definition of this cost function depends on a priori knowledge on the object to be segmented. The boundary cost allows to express both local and global information that can be incorporated within the constrained minimization strategy for optimal boundary detection. Representing the constrained minimization problem within the graph search framework, a graph is build where nodes are cost image pixels whose values are node weights. The connectivity of the graph is derived from the feasibility constraints. Minimal cost feasibility boundary is the minimum cost path in the weighted node graph that can be identified by using A* Algorithm [159].
The global solution is implemented by a bottom-up scheme: the solution of the whole problem is built from the solutions of the subproblems. A "path cost image" collecting the intermediate information useful for defining the global optimal solution is computed with the Algorithm 1. In the path cost image the minimum value of the last column is

Algorithm 1 algorithm for computing the Path Cost Image P basing on the cost image C

```
for j = 0 to M-1 do
    P(j,1) ← C(j,1)
end for
for i = 1 to N-1 do
    for j = 0 to M-1 do
        if i = 0 then
            P(j,i) ← C(j,i) + min(P(j,i−1), P(j+1,i−1))
        else
            if i = N − 1 then
                P(j,i) ← C(j,i) + min(P(j,i−1), P(j−1,i−1))
            else
                P(j,i) ← C(j,i) + min(P(j−1,i−1), P(j,i−1), P(j+1,i−1))
            end if
        end if
    end for
end for
```

the cost of the minimal path. Starting from this point and passing through the graph up to the first column, the minimal path is found by choosing for each visited point the adjacent nodes having the minimum value. In this way the feasible function that minimizes the cost is detected. Figure 2.17 shows the steps with which the 2D segmentation method based on graph searching is organized. Figure 2.17(b) shows the polar representation of the original radial object represented in cartesian space (Figure 2.17(a)). In Figure 2.17(c) the superimposing of the minimum path on the cost image is shown.

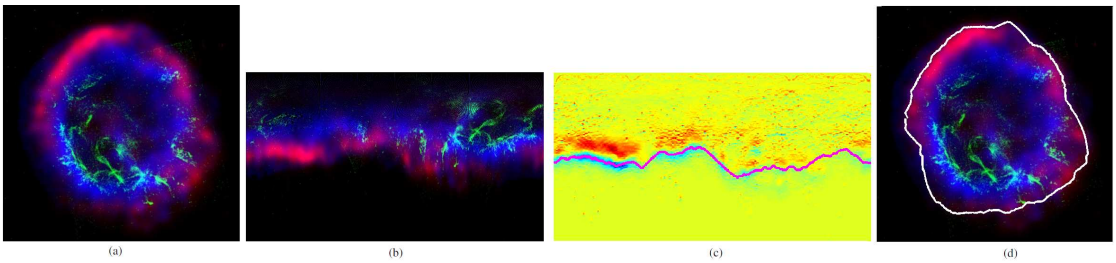


Figure 2.17: Sequence of the steps which where the 2D segmentation method based on graph searching is organized: a)cartesian image b)polar image c) superimposing of the minimum path on the cost image d) segmentation result

Figure 2.17(d) represents the solution in the original cartesian space.

Graph Search Technique for Border Detection in Brain MRI

In this section the application of the above described general framework to the solution of whole brain segmentation in MRI imagery is detailed, addressing specific imaging related aspects.
We consider that between the brain and the skull there is a layer of liquor and in the MRI imagery used in our study, this anatomical part is acquired with voxel values close to zero. Secondly, we take into account that the bone is more intense than the brain tissue and the transitions from the brain and the 'liquor cushion' and from the skull and the external air are similar and can be confused as shows in Figure 2.18. Proceeding from these considerations, a first cost image with which to identify the internal skull boundary is firstly defined; the actual cost image for the brain boundary detection is subsequently defined after a masking procedure. The overall detection procedure includes a preprocessing phase aimed to emphasize the brain boundary before starting with the boundary graph-based detection.

The more the object is similar to a circle in cartesian coordinates or, in other words, an horizontal line in the θ, ρ plan, the more the description of the cost through the Sobel filter application is effective. In axial images of the brain this feature is very strong and could be easily emphasized by changing the aspect ratio of the image. After this preprocessing phase the 2D brain segmentation is accomplished in the three following phases.

Phase 1 Polar Conversion- The image is converted in polar coordinates. The polar sampling is a topic widely discussed in literature presenting a high level of complexity if optimally solved [57]. We adopt a suboptimal solution considering a square grid for the description of the polar space and using an average filter applied on the polar image.

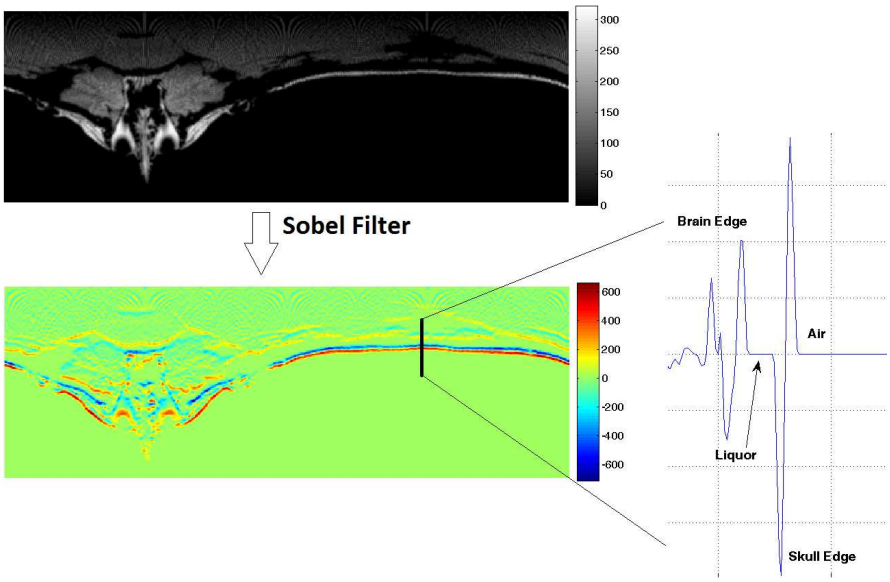


Figure 2.18: Polar representation of brain axial slice and result of the vertical Sobel filter application, with emphasis on a vertical line of profile

In order to fulfill the fully automation requirement, both the angular sampling and the center of the polar space are automatically computed from the data. Let be A_{max} the major axis of the ellipse that circumscribes the image points whose intensity is higher than average $[x_+, y_+]$, the number of angle N_{ang} is computed as follows:

$$N_{ang} = \pi A_{max} \tag{2.18}$$

and the coordinates of the center of the polar space (x_o, y_o) are computed as the centroid coordinates of the points $[x_+, y_+]$; formally:

$$x_o = \frac{1}{N_{x_+}} \sum_{i=1}^{N_{x_+}} x_{+,i}; \quad y_o = \frac{1}{N_{y_+}} \sum_{i=1}^{N_{y_+}} y_{+,i}.$$

The resulting centroid coordinates approximate the actual brain centroid with an estimation error related to the accuracy of the segmentation by trivial thresholding. We have experimentally verified that the approximation does not affect the graph segmentation results.

Phase 2 Skull Boundary Detection- The cost image is computed by applying the vertical Sobel filter to the polar image. Every pixel of this cost image are weights of graph nodes and using the feasibility constraints the graph connections are identified. The minimal path in the graph is computed using Algorithm A* and

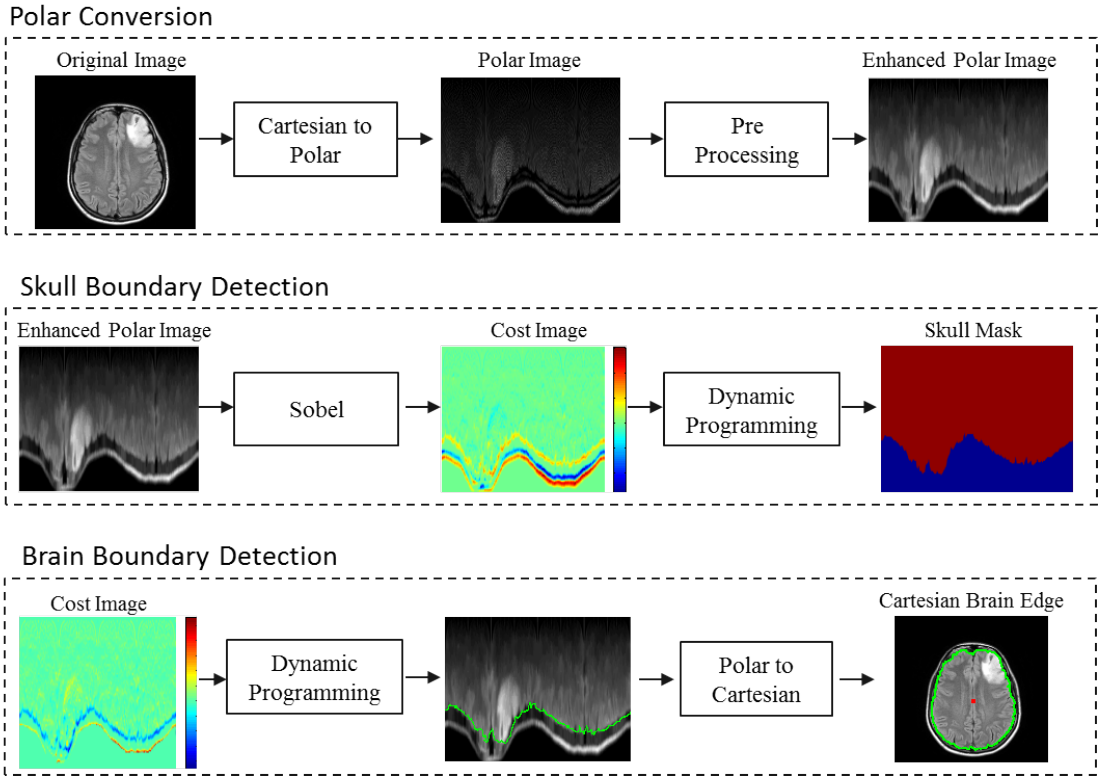


Figure 2.19: 2D brain segmentation strategy

skull boundary is found. A binary mask is computed distinguishing between pixels with ρ less and greater than the edge. This mask is applied to cost image and it is inverted for finding a new cost image for the second step: the actual brain segmentation.

Phase 3 Brain Boundary Detection- The Algorithm A* for the extraction of the minimal path is applied on this new cost image. The minimal path in the graph is the brain boundary in the polar space. The last step is the conversion in cartesian coordinates of the detected boundary.

The detection of the 3D boundary of the brain is computed iterating the 2D segmentation process on every slice composing the whole brain. The schema of the overall 2D brain segmentation procedure is shown in Figure 2.19.
Often, the brain MRI volume starts from the shoulders and includes a portion of air above the head, especially if we consider axial slices of a volume acquired trough subsequent sagittal sections. It would not be appropriate to compute the initial and the final image slice to be processed in the volume as a percentage of the number of slices, because parts of the body included in the the volume varies greatly depending to the acquisition

mode and to the intersubject variability. To make our technique completely independent from the acquisition mode and fully automatic, both the starting and stopping slices are computed directly from the data. The final slice is identified by segmenting the central sagittal slice with the algorithm detailed above. Let be r and c two vectors containing the row and column values of the pixels belonging to the segmented portion, the final slice is $Stop = min(r)$.
Instead, the initial slice is related to the first slice containing the eyes S_e. The eyes detection is a widely studied topic especially in biometric application [52, 70]. Considering that the main feature of the eyes is the circular shape, the Slice of Interest(SoI) detection is performed in the Hough Transform domain [11]. This approach is a well know feature extraction tool, applied on a thresholded edge image. The Hough Transform allows to obtain an accumulation matrix in which the intensity peaks correspond to the position of the center of the circular objects with a specific radius in the image. The eyes radius, very stable in our data is set to $1cm$ value. The SoI detection procedures is detailed below. The edge extraction process is composed of 3 phases iterated for each slice:

- the image is thresholded looking at values greater than zero;
- a close morphological operation is applied on the binary image, with a circular structural element with radius less than $1cm$.
- edges are extracted using Sobel filter applied on the image.

The Hough Transform is applied on the binary edge image and the accumulation matrix A is computed. In Figure 2.20 three examples of the above detailed process are shown. Figure 2.20(a) and Figure 2.20(b) show the processing of a slice under the SoI, at the dental arch level and a slice over the SoI respectively. The third case (Figure 2.20(c)) is related to the Slice of Interest.
For each slice in the volume a Eyes Presence Index (EPI) is computed. Let A^i be the accumulation matrix related to the i^th slice with dimension $[J \times K]$ the EPI is obtained as follows:

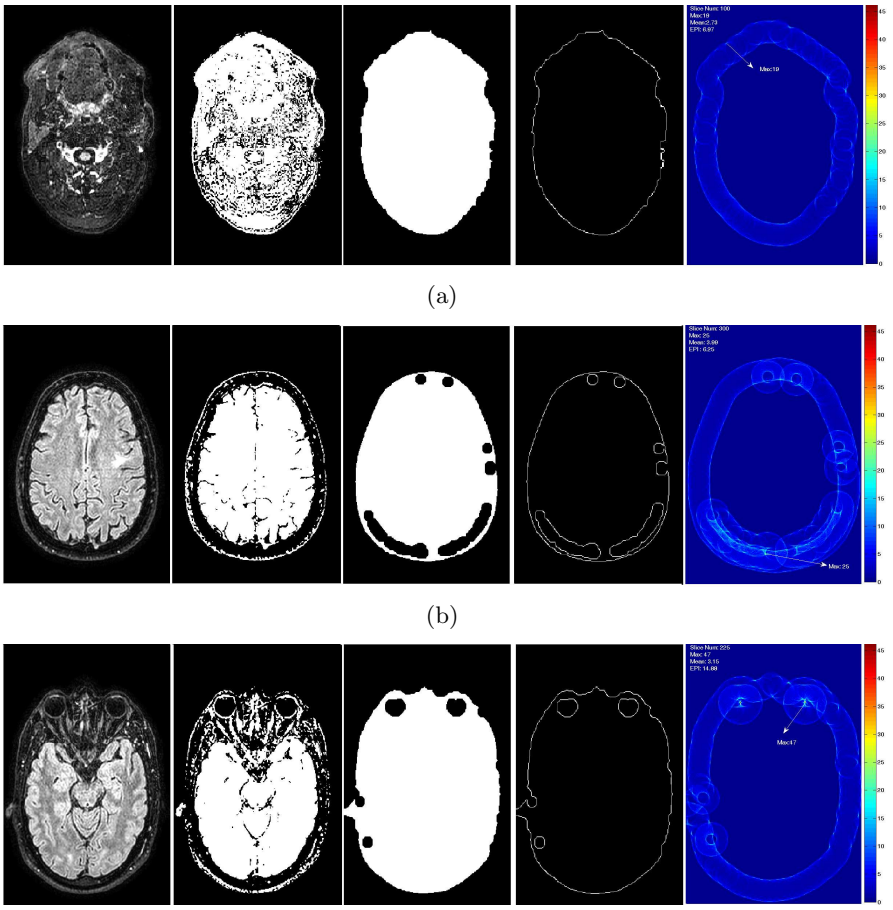
$$\tilde{A^i} = \left\{ \tilde{a^i} \mid \tilde{a^i} = a^i_{j,k} \quad \wedge \quad a^i_{j,k} > 0 \quad for \quad 1 \leq j \leq J \quad and \quad 1 \leq k \leq K \right\} \tag{2.19}$$

$$EPI(i) = \frac{max(\tilde{a^i})}{\frac{1}{W}\sum_{w=1}^{W} \tilde{a^i_w}} \tag{2.20}$$

where W is the cardinality of $\tilde{A^i}$, than the Slice Of Interest is computed as follows:

$$SoI = \underset{i}{argmax}\, EPI(i) \tag{2.21}$$

The $EPI(i)$ values in the three examples given above are respectively: $6.25, 6.97, 14.88$,



(a)

(b)

(c)

Figure 2.20: Phases of the eyes slice detection algorithm a) slice at the dental arc level under the Slice of Interest b) slice over the Slice of interest c) Slice of Interest

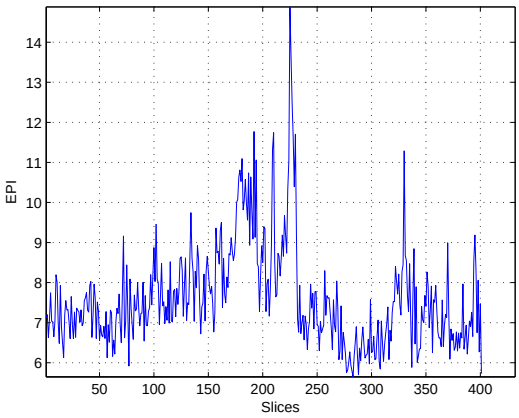


Figure 2.21: EPI signal

the eyes slice has an $EPI(i)$ value greater than twice the other two. In Figure 2.21 an example highlighting the significance of the the EPI index is shown in function of the slice number. The max peak corresponds to The SoI that identifies the central slice of the eyes. The starting slice is computed by shifting the SoI of a number of slices equal to the radius of the eyes.

2.7.2 2D Vs 3D Graph Based Segmentation

The value of the graph search methods detailed in the previous section lies both in robustness and improvements in computing time. Robustness derives from the fact that it inherits the property of optimization methods of embedding global information about edges and the structure of object in a figure of merit allowing accurate border detection from noisy pictures [116]. The formulation of the optimization problem as the search of the shortest path allows to use heuristic information to speed up the search, leading to improvements in computing time with respect to other optimization representation [159].
Unluckily, when dealing with volumetric segmentation, the application, slice-by-slice, of 2D graph searching techniques for the identification of 3D object boundary suffers of some limitations that essentially can be reconduct to lack of contextual spatial information.
In recent years we assist to a wide spread utilization of three dimensional models, often conceived as an extension of previously defined 2D solutions [79]. All path-based methods are limited to 2D applications because object boundary in 3D volumes can not be represented by a path ; therefore, the standard strategy of graph searching for optimal path cannot be directly applied to perform on optimal surface search. The topic

of extending the graph-searching techniques to a direct volumetric medical image segmentation has been investigated in literature [170, 169, 161, 64] giving rise to two main approaches, one acting on graph representation in order to take advantage from the existing search techniques, and the other acting on search algorithm that directly performs an optimal surface determination. Both of them attempt to create a balance between two opposite criteria: to limit the computational complexity and to reach a global optimality. In particular Thedens et al.[170] proposes a method of graph searching belonging to the first category. To make the method efficient, the proposed transformation of the lattice representing the graph imposes the restriction that the number of points to which a single point within the lattice is linked, should be very small and secondly, one dimension of the lattice should be small compared to the other two. The first requirement is satisfied by imposing a smoothness constraint, the second, unfortunately, prevents the applicability of the method to several imaging such as volumetric brain MRI Imaging in which voxels are isotropic and the volume dimensions in the three axis are comparable. The method, applied to MRI cardiac imagery, resulted computational expensive [6].
A graph cut segmentation method for optimal hypersurface detection has been developed by Li et al. [97]. The solution proposed is based on the representation of the problem in terms of a geometric 3D graph and on the computation of a minimum *s-t cut* in a derived directed graph of reduced complexity [31]. Despite the low-order polynomial time complexity and general efficiency, the method maintains the drawback of a high memory demand.
We have investigated a novel solution alternative to both the above mentioned approaches circumventing the problem of defining both a renewed lattice structure and a novel optimization strategy. Contextual information related to the sequence of 2D images has been embedded directly in the cost function defined in the 2D graph searching stage. Acting directly on the node weights we avoid the construction of a 3D graph topology which implies high resource demand, preserving however the explicit description of the boundary features. Basing on this main idea, we propose a 3D graph-based boundary detection strategy oriented to radial object segmentation.

2.7.3 Heuristic 3D Graph Propagation Technique for Volumetric MRI Segmentation

In this section we formally describe a method for 3D radial surface detection conceived as an heuristic extension of the 2D graph-based boundary detection.
A 3D radial object can be easily treated using the spherical coordinates. Let be $V_{cart}(x, y, z)$ a volume described on a square cartesian discrete grid. The volume can be 'unwrapped'

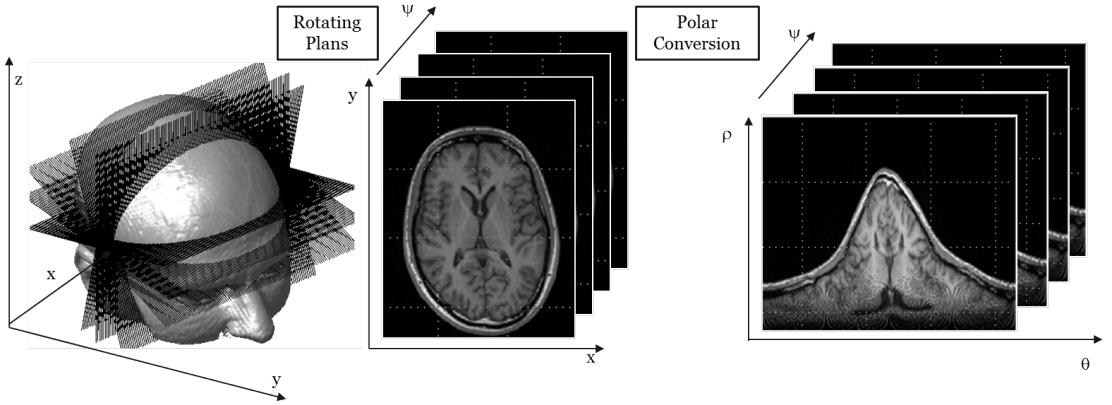


Figure 2.22: Spherical transformation of brain MRI volume

converting it in spherical coordinates, obtaining $V_{sph}(\theta, \rho\, \phi)$:

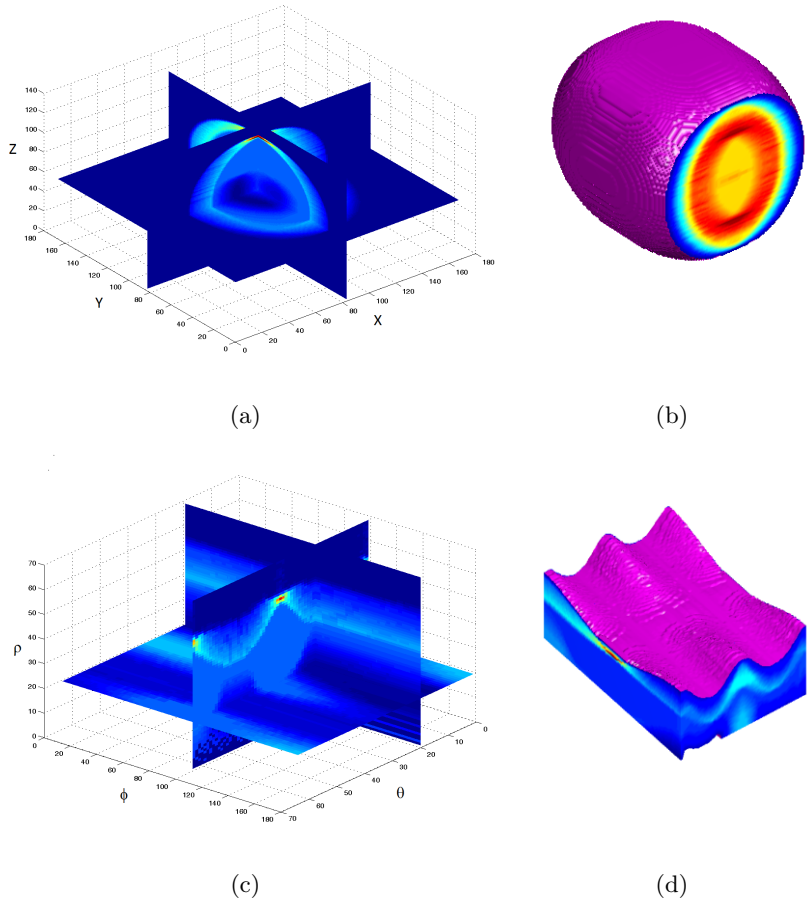
$$\begin{aligned} \rho &= \sqrt{(x-x_o)^2 + (y-y_o)^2 + (z-z_o)^2} \\ \phi &= arccos(\frac{z-z_o}{\rho}) \\ \theta &= \begin{cases} arcsin(\frac{y-y_o}{S}), & x > 0 \\ \pi - arcsin(\frac{y-y_o}{S}), & x <= 0 \end{cases} \end{aligned} \tag{2.22}$$

where (x_o, y_o, z_o) are the coordinates of the 3D object centroid and

$$S = \sqrt{(x-x_o)^2 + (y-y_o)^2} \tag{2.23}$$

The spherical sampling is a topic widely discussed in literature presenting a high level of complexity if optimally solved [63]. In order to circumvent the problem of drawing up a regular sampling grid of independent variables $V(\theta, \phi)$ on a sphere without singularity at the poles [63], we introduce a non-conventional spherical transformation whose application on brain MRI volume is illustrated in Figure 2.22. Considering the sheaf of planes Π passing through the straight line lying on the axis x, each plan $\Pi_\phi \in \Pi$ may be identified by the angle ϕ between the normal vector of the plane and the y axis. The image formed by all points of the volume $V(x, y, z)$ that lie on each plane Π_ϕ, is unwrapped in polar coordinates (θ, ρ). A proper spherical description $V(\theta, \rho, \phi)$ of the cartesian volume $V(x, y, z)$ is then obtained at the expense of the independence of variables which is not however an essential requirement in the present domain.

Figure 2.23 shows a volumetric radial phantom both in cartesian and spherical space. The representation by slices and the boundary surface of the original volume are shown in Figures 2.23(a) and 2.23(b) respectively. The same contents in spherical space are



(a) (b)

(c) (d)

Figure 2.23: Spherical Conversion a) Volumetric Phantom in Cartesian Space b) Phantom segmentation in Cartesian Space c) Volumetric Phantom in Spherical Space d) Phantom segmentation in Spherical Space

represented in Figures 2.23(c) and 2.23(d) respectively.
As highlighted in Figure 2.23(d), the boundary of a 3D radial object can be represented by the function $\rho = f(\theta, \phi)$ described by the following set of feasibility constraints, which are an extension of those listed for the 2D case (Section 2):

Boundary as a Function - in each column and slice the 3D boundary exists and it has a single value so that it can be represented by a 2D function $\rho = f(\theta, \phi)$.

Connectivity Constraints -

1. θ)- $f(\theta + 1, \phi) - f(\theta, \phi)| \leq 1$ for $1 \leq \theta \leq N - 1$ where N is the number of columns of the volume, that is the number of sampled θ angles.

2. ϕ)- $|f(\theta, \phi + 1) - f(\theta, \phi)| \leq 1$ for $1 \leq \phi \leq K - 1$ where K is the number of sampled ϕ angles.

Closing Constraints -

1. θ)- $|f(1, \phi) - f(N, \phi)| \leq 1$ that is the closing constraints in θ direction.
2. ϕ)- $|f(\theta, 1) - f(\theta, K)| \leq 1$ that is the closing constraints in ϕ direction.

Each feasible 2D function $f(\theta, \phi)$ is a candidate object boundary. The goal is then to find the minimum cost 3D boundary subject to the feasibility constraints. The 3D boundary cost is defined as follows:

$$B_c = \sum_{\phi=1}^{K} \sum_{\theta=1}^{N} C(\theta, \phi, f(\theta, \phi)) \tag{2.24}$$

where $C(\theta, \phi, f(\theta, \phi))$ is a volumetric cost. Similarly to the 2D case, the value of each voxel in the cost volume must be inversely related to the likelihood that an edge is present at that point. The likelihood is usually determined by the application of a low-level local edge detection operator [170]. As in the 2D case, the definition of this cost function depends on a priori knowledge about the object to be segmented. The boundary cost allows to express both local and global information that can be incorporated within the constrained minimization strategy for optimal boundary detection. Figure 2.24 shows cost volume of the radial phantom object in Figure 2.23 extracted using a vertical gradient local operator.

Conceiving the 3D boundary detection as an extension of the graph based solutions

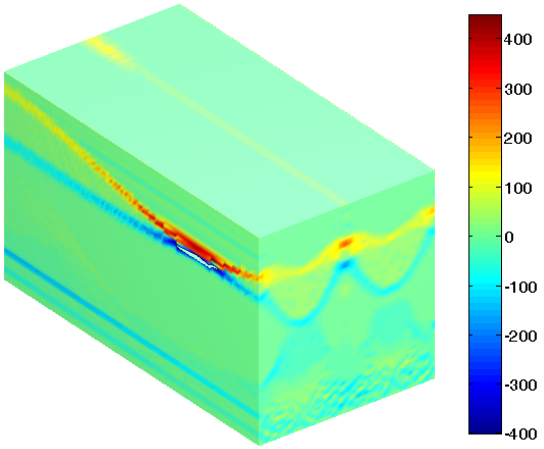


Figure 2.24: Cost volume of the radial phantom object in Figure 2.23 extracted using Sobel filter

adopted for the 2D contour detection, the underling minimization problem has to cope

with a 3D graph. As discussed in the introduction, the impossibility of representing a 3D boundary as a path, has led to new optimal and/or heuristic solutions acting on the minimization schema and/or graph topology respectively [97, 170]. In the present work we want to explore a third heuristic way ascribing to the cost function the role of exploiting the contextual slice-to-slice information.
Ignoring either one of the connectivity constraints (we choose here that related to the ϕ variable), we consider the 3D graph as a stack of 2D graphs without connection between subsequent ϕ slices. The continuity in ϕ direction is represented by changing the weights of the nodes of the $k-th$ graph according to the minimal path found in the $k-th-1$ graph. Let be $B_{k-1}(\theta)$ the boundary of the $(k-1)-th$ slice obtained by applying the 2D graph searching method (Section 2). We introduce a weight function $f_w(x)$ that produces a weight surface $Ws_k(\theta,\rho)$ combined with the resulting boundary $B_{k-1}(\theta)$. Any formulation is accepted as long as $f_w(x)$ is monotonically increasing with $f_w(1)\rightarrow 1$ and $f_w(0)\rightarrow 0$. A possible choice is:

$$f_w(x) = \exp\frac{-1}{(\beta\cdot x)} \tag{2.25}$$

where the β value makes the continuity constraint more or less strict: if $\beta >> 1$ the constraint is hard otherwise if $\beta \rightarrow 1$ the constraint is more relaxed. Another formulation for the weight function, could be a sigmoid

$$f_w(x) = \frac{1}{1+exp{-\beta(x-0.5)\cdot 10}} \tag{2.26}$$

The analytical formulation of the weight surface $Ws_k(\theta,\rho)$ is piecewise defined as follows:

$$Ws_\phi(\theta,\rho) = \begin{cases} 0, & if \quad (B_{k-1}(\theta)-1) < \rho < B_{k-1}(\theta)-1) \wedge \forall\theta \\ f_w(\frac{|\rho-\bar{\rho}|}{\bar{\rho}}), & if \quad \rho < (B_{k-1}(\theta)-1) \wedge \forall\theta \\ f_w(\frac{|\rho-\bar{\rho}|}{M-\bar{\rho}}), & if \quad \rho > (B_{k-1}(\theta)+1) \wedge \forall\theta \end{cases} \tag{2.27}$$

Each $Ws_\phi(\theta,\rho)$ can assume values between 0 and 1.
In order to process the $k-th$ slice, the normalized cost image $C_k^n(\theta,\rho)$ is combined with the weight function $Ws_k(\theta,\rho)$ obtaining a new weighted cost image $C_k^w(\theta,\rho)$ which represents the balance between the likelihood of being an edge and the likelihood of belonging to a continuous surface. An analytical formulation may be the following:

$$C_k^w(\theta,\rho) = C_k^n(\theta,\rho) + \alpha Ws_k(\theta,\rho) \tag{2.28}$$

were α gives more or less emphasis on the continuity term with respect to the original values of the cost function. In Figure 2.25(a) and 2.25(b) two weight surfaces derived

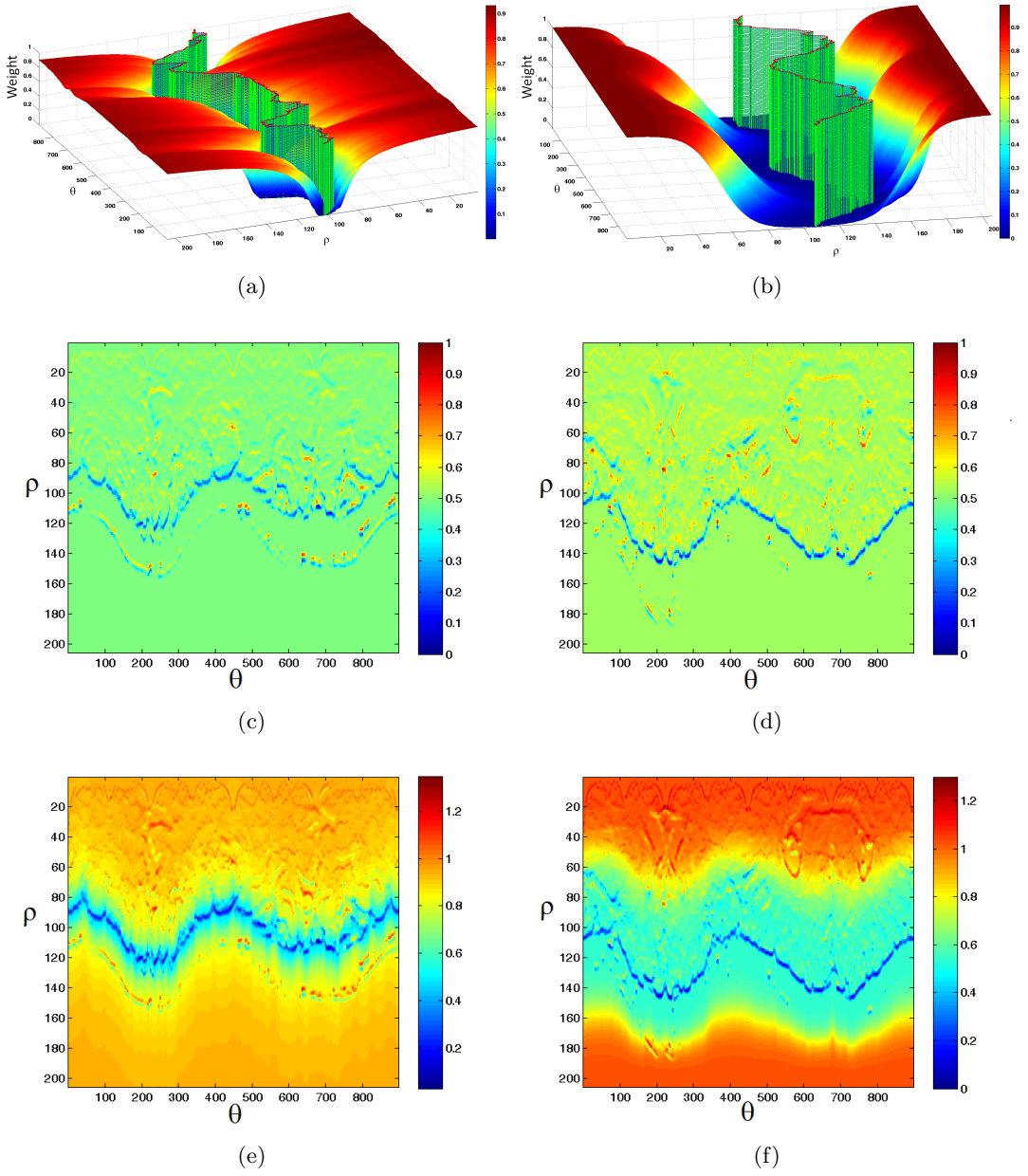


Figure 2.25: Examples of cost function weighing process: a) Exponential weight surface b)Sigmoid weight surface c)d) Original cost functions e)f) Weighted cost functions

from exponential and sigmoid weight function are shown, the first more restrictive than the second. Figure 2.25(c) and 2.25(d) and Figure 2.25(e) 2.25(f) show the normalized cost image $C_k^n(\theta, \rho)$ and the weighted cost images $C_k^w(\theta, \rho)$ for the exponential and sigmoid weight function respectively. When applying the 2D graph searching strategy to the current $k-th$ slice, the weight function $Ws_k(\theta, \rho)$ determines a kind of inhibition in the search of the minimal path.
In the overall strategy the solution presented for the $k-th$ slice is iterated to process the entire volume starting from a slice segmented without connectivity constraints. The overall procedure is formalized in Algorithm 1.

In order to exploit the great potential of graph searching technique of embedding a priori knowledge about the object to be segmented within the overall minimization strategy, we may further specialize the weight function considering, for example, how the size of the object varies across the slices. The piecewise definition of $Ws_k(\theta, \rho)$ may include a more relaxed $f_w(x)$ in the direction of growth, inversely a more conservative $f_w(x)$ could be used in the opposite direction.
A limitation of the overall procedure could be the propagation of the errors made on individual slices. This critical aspect can be however compensated by two facts. Firstly, the connectivity component, responsible of the propagation of the errors, is regularized by the original 2D cost function. Secondly, errors can be reduced by processing at first less problematic slices.
Both time and memory complexity are drastically reduced by embedding the topological information in the cost function, thus avoiding the search of the optimal surface on a 3D fully connected graph. The lattice structure stored for each iteration is only that relating to a single slice of the 2D graph stack. The time complexity of the 2D graph searching algorithm used for finding the minimal path is polynomial: $O(n^m)$ where n is the number of pixel in a single cost image and m the polynomial order. This execution is simply iterated for each plan k, the total computational complexity is $K \cdot O(n^m)$.

Application of the Heuristic 3D Graph Propagation Technique for Automatic MRI Brain Segmentation

In this section we describe the application of the above presented general framework to the problem of segmenting whole brain in MRI imagery. We have to proceed from preliminary considerations regarding the MRI imagery and the anatomical district considered.
As regard the radial shape of brain volume the more the object is similar to a sphere in cartesian coordinates or, in other words, an horizontal plane in the θ, ϕ, ρ space, the more the description of the cost by applying a volumetric local edge detector is effective. We introduce a preprocessing phase that change the aspect ratio of the MRI volume in and attempt to inscribe the brain volume as much as possible in a sphere.

Algorithm 2 Heuristic 3D Graph Propagation Algorithm

$V \leftarrow InputVolume$
$\alpha \leftarrow 0.5$
$\beta \leftarrow 10$
$fw = \exp \frac{-1}{(\beta \cdot [0:1])}$
$V_{sph} \leftarrow car2sph(V)$
$StartPol \leftarrow V_{sph}(:,:,0)$
$C \leftarrow VerticalGradient(StartPol)$
$[Mask_{pol}, MinPath] \leftarrow AlgA * (C)$
$Mask_{sph}(:,:,0) \leftarrow Mask_{pol}$
for $i = 0$ **to** N **do**
  for $j = 0$ **to** M **do**
    if $j < MinPath(i) - 1$ **then**
      $Ws_1(i,j) \leftarrow f_w(\frac{|j - MinPath(i)|}{MinPath(i)})$
    else
      if $j > MinPath(i) + 1$ **then**
        $Ws_1(i,j) \leftarrow f_w(\frac{|j - MinPath(i)|}{M - MinPath(i)})$
      else
        $Ws_1(i,j) = 0$
      end if
    end if
  end for
end for
for $k = 1:1$ **to** K **do**
  $Pol \leftarrow V_{sph}(:,:,k)$
  $C \leftarrow VerticalGradient(Pol)$
  $C \leftarrow C + |min(C)|$
  $C^n \leftarrow \frac{C}{max(C)}$
  $C^w \leftarrow C^n + \alpha * Ws_k$
  $[Mask_{pol}, MinPath] \leftarrow AlgA * (C_w)$
  $MaskV_{sph}(:,:,\phi) \leftarrow Mask_{pol}$
  for $i = 0$ **to** N **do**
    for $j = 0$ **to** M **do**
      if $j < MinPath(i) - 1$ **then**
        $Ws_{k+1}(i,j) \leftarrow f_w(\frac{|j - MinPath(i)|}{MinPath(i)})$
      else
        if $j > MinPath(i) + 1$ **then**
          $Ws_{k+1}(i,j) \leftarrow f_w(\frac{|j - MinPath(i)|}{M - MinPath(i)})$
        else
          $Ws_{k+1}(i,j) = 0$
        end if
      end if
    end for
  end for
end for
$MaskV_{car} = sph2car(MaskV_{sph})$

To make the overall procedure fully automated, both the angular sampling and the center of the spherical space have to be automatically computed from the data. Let A_{max} be the major axis of the 3D ellipsoid that circumscribes the set of volume points $E = (x_+, y_+, z_+)$ whose intensities are higher than average intensity computed overall the volume points. The number of angles, N_{ang}, is computed as follows:

$$N_{ang} = \pi A_{max} \tag{2.29}$$

The inscription of the object to be segmented within a sphere, accomplished in the preprocessing phase, suggests the use of the same sampling rate for the angles θ and ϕ. The coordinates of the center of the spherical space (x_o, y_o, z_o) are computed as the coordinates of the centroid of the points (x_+, y_+, z_+):

$$x_o = \frac{1}{N_+}\sum_{i=1}^{N_{x_+}} x_{+,i}; \quad y_o = \frac{1}{N_+}\sum_{i=1}^{N_{y_+}} y_{+,i}.z_o = \frac{1}{N_+}\sum_{i=1}^{N_{z_+}} z_{+,i}.$$

where N_+ is the cardinality of the set E.
To properly design the skull boundary detection within the overall segmentation strategy, we have to consider that there is a layer of liquor between the brain and the skull and in the MRI imagery used in our study, this anatomical part is acquired with voxel values close to zero. Moreover, we have to take into account that the bone is more intense than the brain tissue and the transitions from the brain and the 'liquor cushion' and from the skull and the external air are similar and can be confused.

Brain segmentation is achieved trough an iterative process. Figure 2.26 provides the overall schema of the segmentation strategy highlighting in particular the novelties introduced with respect to the previous 2D segmentation strategy. New tasks are inserted between the results obtained by the *skull boundary detection* task and the minimal path computation for the *brain boundary detection*. In the *cost image weighting* task, the cost image is refined according to the results of the *weight surface definition* task computed in the previous iteration step. The optimization strategy is applied on this renewed cost image and the computed boundary, combined with the weight function, contributes to the definition of the weight surface for the next step according to the Equation 2.22.

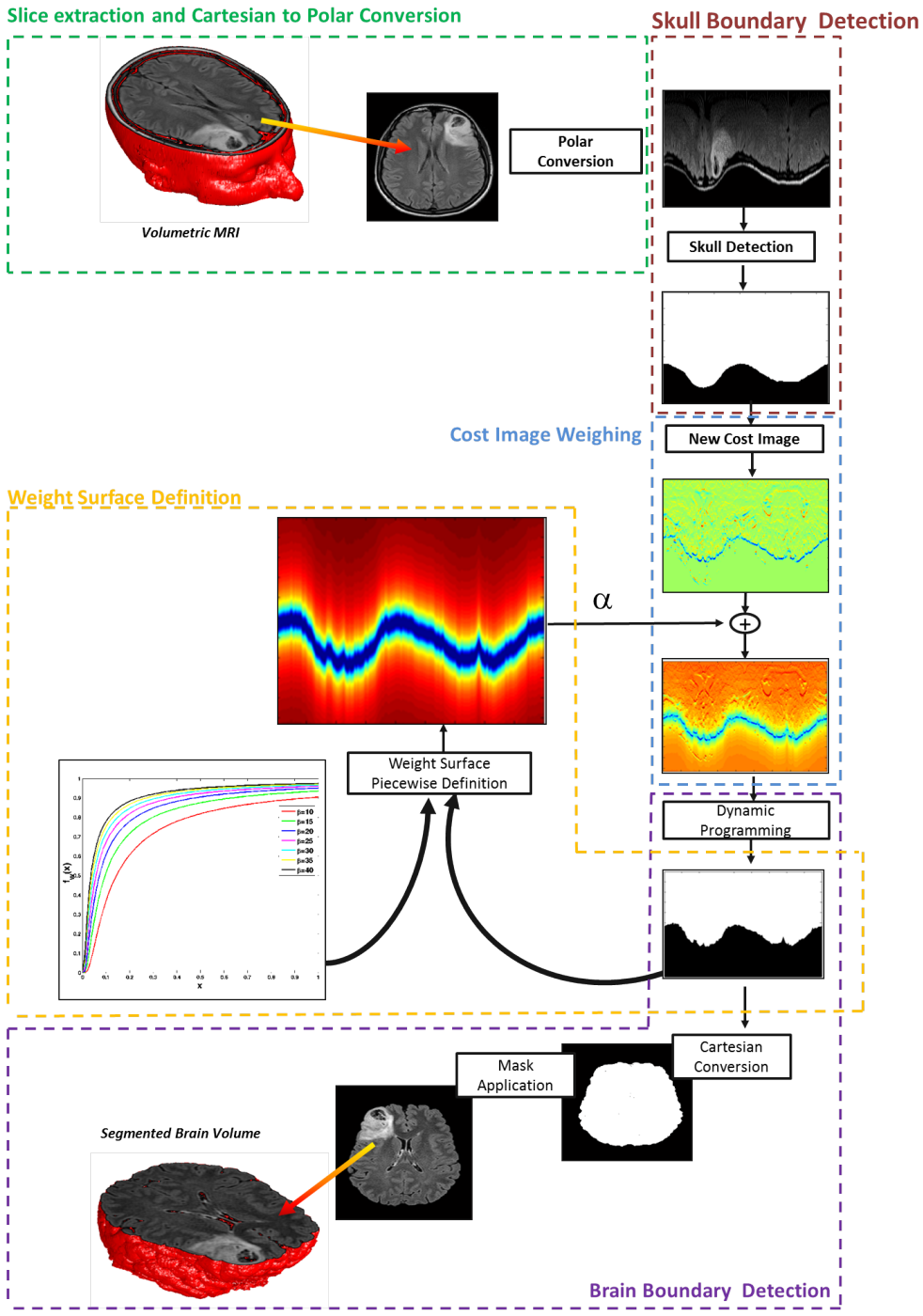


Figure 2.26: Schema of Heuristic 3D Graph Propagation Segmentation strategy

2.7.4 Graph Cut in Symmetry Space for Unilateral Glial Tumor Segmentation

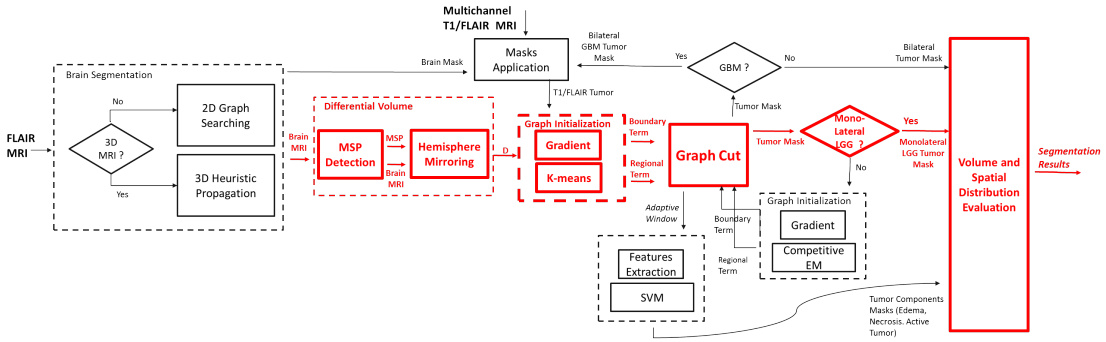


Figure 2.27: Schema of the Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation with highlight the part detailed in Section (2.7.4)

The framework for fully automatic MRI Glial tumor segmentation, presented in this Ph.D. Thesis, after the brain segmentation step, provides the challenging task of the brain tumor segmentation. (Figure 2.27)
The quantitative analysis of brain tumor allows to obtain useful key indicators (volume, shape...) of tumor progression. Moreover, it is helpful in post-treatment radiological follow-up (evaluation of tumor remnants) and in the planning of other therapies such as radiotherapy. The automatic segmentation has the great potential in clinical medicine of freeing physicians from the burden of manual labeling; only a quantitative measurements allows to track and modeling precisely the disease.
Despite the undisputed usefulness of automatic tumor segmentation, this is not yet a widespread clinical practice, therefore the automatic brain tumor segmentation is still a widely studied research topic. The main difficulties are related to the fact that the brain tumors are very heterogeneous in terms of shape, color, texture and position and they often deform other nearby anatomical structures.
Many techniques are developed and experimentally evaluated in the last two decades. The vast amount of works can be categorized with respect to the type of brain tumor treated and the segmentation techniques applied. A possible taxonomy is provided below summarized between a great choice.
Kaus et.al. developed an algorithm for segmenting the Low Grade Glioma and Meningioma using a statistical classification with atlas prior [103, 146]. Instead, the methods described in [99, 182] work on the segmentation of Glioblastoma both using the fuzzy clustering. Liu et. al.[99], proposes a segmentation method besed on Fuzzy Connectedness principles, reaching a high level of accuracy; however, this system is interactive and the results are strongly related with the parametrization. Prastawa and Clarck [104, 101]

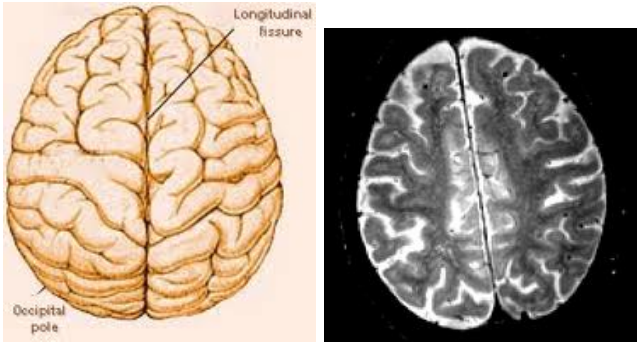
have worked on Glioblastoma segmentation using a Knowledge-based algorithm. Many other approaches use the deformable model like level set [145]. The Machine Learning is acquiring in the last few years an important role in medical imaging also in field of brain tumor segmentation: for example Zhang et al [194] use a Support Vector Machine to perform the brain tumor segmentation.
A promising technique particulary used in recent years is the segmentation based on the global minimization of an energy function by using a Graph [82, 22, 196]. The Graph Cut framework for segmentation is based on an implicit representation of object boundaries. Thus, it allows the same flexible topological properties of segments as in techniques where the edge properties are explicit, level-sets for example [145]. Since the introduction of the approximation energy minimization via Graph Cut proposed by Boykov et al. [31] and the subsequent technique formalization for solving segmentation problem on N-D Images [32, 27], many methods have been proposed. The Graph Cut has undergone evolutions in different directions: some with the aim of improving the interactivity [144], other incorporating other techniques as watershed or gaussian mixture model [23, 135] and other integrating high-level contextual information [92]. Brain tumor segmentation strategy based on these technique apply Graph Cut using information directly readable in the image as the intensity values, without considering a features extraction phase. The aim of our work is to develop a technique for the brain tumor segmentation using Graph Cut in a feature space.
Proceeding from the idea of describing the volume with a feature invariant with respect to the internal inhomogeneity of the tumor (edema, necrose, infiltration...) we are inspired by other authors [140, 88] that use the feature symmetry in their analysis, considering the tumor as an anomaly in the symmetry of the brain with respect to the Mid Sagittal Plane.
In the visual analysis of the MRI, the first step of the experts cognitive process, is the detection of an anomaly with respect to the normal tissue, whatever its nature. Identifying an anomaly requires the knowledge of an *expected*. There are two main approaches to identify abnormalities in the brain, which differ from the procedure by which the *expected* is found.

Inter-subject It makes use of an external template to detect the abnormalities. Many segmentation algorithms use probabilistic atlas to describe the expert knowledge about brain structure [138, 103, 146]. These algorithms suffer from two problems: the difference in brain structure caused from the normal interpersonal variability, and the need to apply co-registration algorithms to align the patient brain and the atlas. Moreover, these algorithms are often time consuming and with a low level of robustness.

Intra-subject It makes use of information obtained from the image itself for defining an *expected*, avoiding the drawbacks mentioned above for the Inter-subject techniques.

An important brain feature is the symmetry computed with respect to the Mid Sagittal Plane; indeed, many tumor segmentation algorithms use it for the anomaly detection, considering that tumors are generally not symmetrically placed in both hemispheres [88].

The use of the feature symmetry requires the automatic identification of the Mid Sagittal Plane(MSP), the plane that best separates brain hemispheres (Figure 2.28(a)).
In a real MRI acquisition there is no guarantee that the MSP lies on the sagittal acquisition plane (Figure 2.28(b)). It is therefore necessary a technique for an automatic identification.



(a) (b)

Figure 2.28: Brain Symmetry Plane description a) Explanatory image of the brain symmetry b) Example of MSP not coincident with the sagittal plane of the acquisition

For this purpose we use a method that expresses this problem as a registration task and computes a degree of similarity between the image and its reflection with respect to a plane. The best plane is then obtained by maximizing the similarity measure [173]. For completeness we report only the main elements of this strategy.
We deal with 3D gray-level images that can be represents es a 3D function $f : F \in \mathbb{R}^3 \rightarrow V$ let be $\boldsymbol{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ a given point and $f_{\boldsymbol{b}}$ the translation of the 3D image f at point $\boldsymbol{b}$. Let be $\boldsymbol{\Pi}_{\boldsymbol{u},d}$ a plane orthogonal to a vector $\boldsymbol{u}$ and at the distance d to the point $\boldsymbol{b}$. We define $e_{\boldsymbol{u},d}(f)$ the reflation of the image f with respect to the plane $\boldsymbol{\Pi}_{\boldsymbol{u},d}$:

$$e_{\boldsymbol{u},d}(f)(x, y, z) = f(e_{\boldsymbol{u},d}(x, y, z)) \tag{2.30}$$

An image f is called plane symmetrical if exists a plane $\boldsymbol{\Pi}_{\boldsymbol{u},d}$ such that $e_{\boldsymbol{u},d}(f) = f$. We say in this case that $\boldsymbol{\Pi}_{\boldsymbol{u},d}$ is a plane of symmetry for f. Of course the actual MRI 3D images of a brain doesn't meet exactly this constraint but closely approximates it with respect the Mid Sagittal Plane; thus the goal is to find the plane $\boldsymbol{\Pi}_{\boldsymbol{u},d}$ that minimize $|e_{\boldsymbol{u},d}(f) - f|$. The computation of symmetry measure given by Eq. (3.12) is a time consuming non-convex optimization problem. Therefore it seems useful to check

first the most appropriate positions of symmetry planes. The initial positions can be interpreted, for example, in terms of ellipsoid of inertia In this case we consider the image f as a 3D body F with a mass distribution $f(x, y, z)$ at every point $(x, y, z) \in F$. The axes of the ellipsoid of inertia for a 3D body are defined by the eigenvectors of the corresponding covariance matrix:

$$\begin{pmatrix} m_{2,0,0} & m_{1,1,0} & m_{1,0,1} \\ m_{1,1,0} & m_{0,2,0} & m_{0,1,1} \\ m_{1,0,1} & m_{0,1,1} & m_{0,0,2} \end{pmatrix} \tag{2.31}$$

where $m_{p,q,r}$ is the central moment of the order $p + q + r$ and:

$$m_{p,q,r}(f) = \int\int\int f(x, y, z)(x - x_c)^p (y - y_c)^q (z - z_c)^r dxdydz \tag{2.32}$$

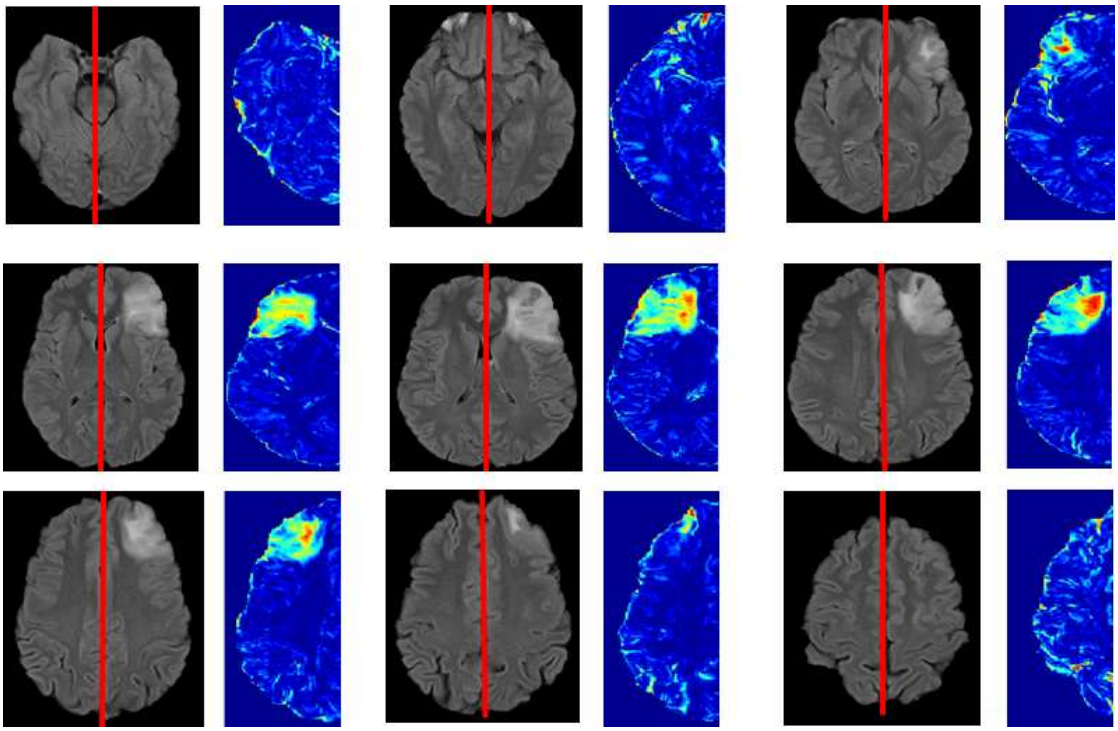
where (x_c, y_c, z_c) are the coordinates of the center of mass of the body. It is known that if a body possesses a plane symmetry then the symmetry plane passes through its center of mass orthogonal to some ellipsoid axis. Therefore, for the computation of a symmetry measure in this case it is sufficient to check three planes only those defined taking in pairs the eigenvectors. Among these three plane what better meet the equation Eq. (3.12) is the estimated symmetry plane. Reflecting the right hemisphere across this plane and computing the normalized difference, voxel by voxel, from the left hemisphere and this mirrored hemisphere, it derives a new volume that highlights the asymmetric volume components as shown in Figure 2.29. Formally the new volume D consists of a set of voxels d defined as

$$d(i, j) = \frac{|V(i, j) - f_{Mir}(V(i, j))|}{V(i, j) + f_{Mir}(V(i, j))} \tag{2.33}$$

where $f_{Mir}(\cdot)$ is the function which mirrors a point with respect to the plane of symmetry.
In this differential volume D, it is easy to isolate portions with greater intensity with respect to the background, that is the anomalies computed in function of the feature of symmetry. We address this phase by developing a two class clustering solved with 3D graph. In the next section all the formal element are provided.

Graph Cut Strategy for Volume Segmentation

The Graph Cut framework offers a globally optimal object extraction method for N-dimensional images. This strategy belongs to the category of the energy minimization methods that, strictly proceeding from the definition just stated, formalize the segmen-



(a)

Figure 2.29: Examples of Differential Volumes

tation problem in terms of optimization, defining an energy function whose the minimization is the solution of the segmentation. A fairly general cost function that can include both region and boundary properties of segments and certain types of topological constraints are described. These components naturally fit into our global optimization framework. The main technical novelty of the object extraction approach of the Graph Cut is that segmentation energy is formalized over binary variables whose values only indicate whether the pixel is inside or outside the object of interest. In contrast to the earlier path-based combinatorial methods, this can be seen as a region-based approach to encode image segments. In fact, the difference between path-based and region-based representations of segments on a discrete graph is analogous to the difference between explicit contour representation, e.g. snakes, and the implicit level-sets approach.
Moreover, Boykov and Kolmogorov [29] have studied the practical efficiency of combinatorial min-cut/max- flow algorithms on applications in computer vision. It was shown that some max-flow techniques could solve 2D and 3D segmentation problems in close to real-time using regular PCs. The Graph Cut framework for segmentation is based on an implicit representation of object boundaries. Thus, it allows the same flexible topological properties of segments as in level-sets techniques. The segmentation results may

have isolated objects which may also contain holes. However, it may be useful to impose some topological constraints reflecting certain high-level contextual information about the object of interest. Graph cut can incorporate some types of topological constraints. For example, the hard constraints can indicate some image pixels a priori known to be a part of the object or background.
In the follow we describe the *object/background* segmentation framework in detail. The aim of a volumetric segmentation problem is to assign to each voxel $v \in V$ a label that says the membership of the voxel to a specific region L_i in the case of binary segmentation objects/background the goal is to find the optimal labeling $L = (L_{Bac}, L_{Obj})$.
We assume that each voxel v of the volume V has two cost values concerning the association of the specific voxel to the background $R_v(O_{Bkg})$ and object $R_v(O_{Obj})$. Moreover, we assume each pair of adjacent voxels $v, w \in N$ has a cost $B_{v,w}$ associated with the fact that the pair of voxel have different labeling, where N is a set of pairs of 3D neighboring voxel. Our goal is to find the optimal labeling $L = (L_{Bkg}, L_{Obj})$ in assigning each voxel v to the object or to the background by the minimization of the following cost function:

$$E(L) = \lambda R(L) + B(L) \tag{2.34}$$

where

$$R(L) = \sum_{v \in V} R_v(O_v) \tag{2.35}$$

is called *Regional Term* and

$$B(L) = \sum_{v,w \in N} B_{v,w} \delta_{L_i, L_j}; \tag{2.36}$$

where

$$\delta_{L_i, L_j} = \begin{cases} 0 & \text{if } L_i = L_j \\ 1 & \text{if } L_i \neq L_j \end{cases} \tag{2.37}$$

is called *Boundary Term*.
The segmentation problem can be formulated as finding the optimal set of voxel belonging to a desired 3D object; considering the translation of the segmentation problem in an optimization problem solvable with 3D graph $\mathcal{G}\{\mathcal{V}, \mathcal{E}\}$ we have to look for the partition of the graph that minimizes the cost of the cut.
Each $v \in V$ becomes a graph node $\mathcal{V}$ that is connected to the 26 3D neighbors with a set of links called *n-link*. There are also two specially designated terminal nodes, $S(source)$ and $T(sink)$ that represent "object" and "background". Each node is also connected to both the terminal node with a link called *t-link*. To every link both *t-link* and *n-link* is given a weight w_e. The weights assigned to *n-link* represent the distance between two neighbor nodes, thus the *Boundary Term* $B_{v,w}$. Costs $B_{v,w}$ may be based on local in-

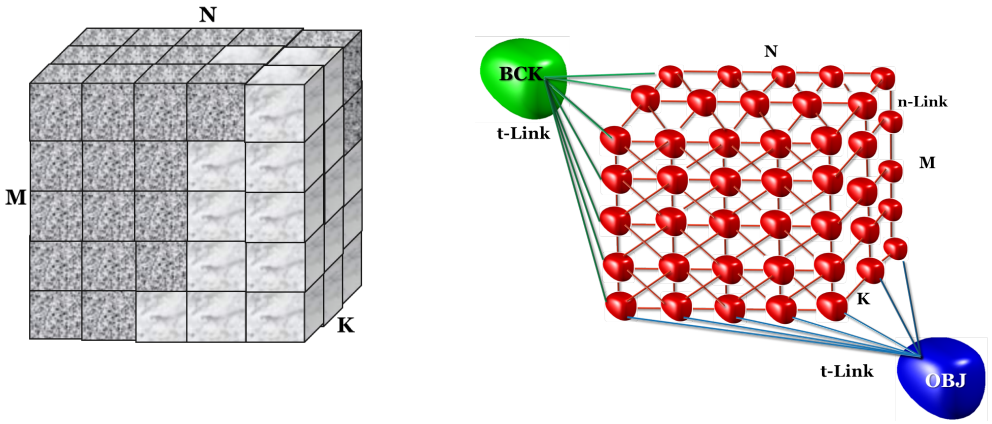
tensity gradient, Laplacian zero-crossing, gradient direction, geometric or other criteria [28, 30]. Often the following function is used:

$$B_{v,w} \propto \exp(-\frac{(I_v - I_w)^2}{2\sigma}) \tag{2.38}$$

The weights assigned to *t-link* represent the *Regional Terms* $R_v(O_{Bkg})$ and $R_v(O_{Obj})$. For example, $R_v(\cdot)$ may reflect on how the intensity of the voxel v fits into given intensity models of the object and background

$$\begin{aligned} R_v(O_{Obj}) &= -\ln Pr(I_v|Obj) \\ R_v(O_{Bkg}) &= -\ln Pr(I_v|Bkg) \end{aligned} \tag{2.39}$$

An s-t cut is a subset of edges $\mathcal{C} \in \mathcal{E}$ such that the terminals S and T become com-



(a) (b)

Figure 2.30: Formalization of a volumetric segmentation using Graph Cut framework

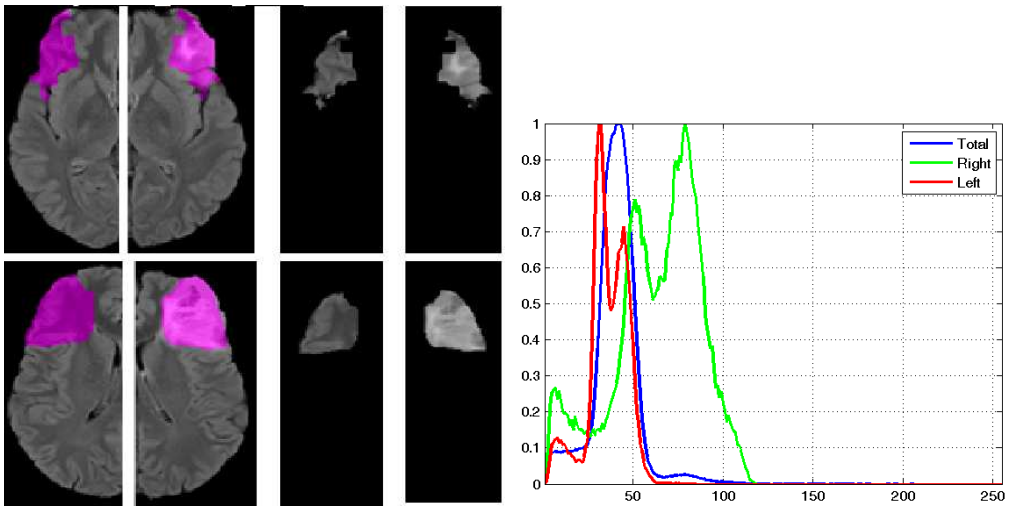
pletely separated on the graph $\mathcal{G}(\mathcal{C}) = \{\mathcal{V}, \mathcal{E}|\mathcal{C}\}$. Any cut corresponds to some binary partitioning of an underlying image into 'object' and 'background' segments. Note that, in general, cuts can generate binary segmentation with arbitrary topological properties. The cost of a cut is sum of all the weight crossed by the cut.

$$C_{cost} = \sum_{e \in C} w_e; \tag{2.40}$$

The segmentation goal is to identify the cut that minimize the cost using the max-flow/min-cut algorithm [31, 29, 90, 91, 8]. Figure 2.30 shows a volumetric graph definition example.

On one hand severed *n-links* are located at the segmentation boundary. Thus, their total cost represents the cost of the segmentation boundary. On the other hand, severed *t-links* can represent the regional properties of segments. Thus, a minimum cost cut may correspond to a segmentation with a desirable balance of boundary and regional properties. Numerically, graph cut is based on a well-known combinatorial optimization for the fact that a globally minimum s-t cut can be computed efficiently in low-order polynomial time.
The Graph Cut approach is the first global optimization object extraction technique that extends to N-dimensional images. The underlying numerical optimization scheme is straightforward and robust. The strength of this method, as for all the graph based methods, lies into the fact that the embedding of both the *local* and *global* information in the graph search segmentation allows to identify the complete object 3D boundary even in a complex scene. On one hand, the graph construction is based on *local* information of the volume: the homogeneity of the region and the strength of the transition; on the other hand the minimization process on the graph is *global*.
A critical aspect in the application of Graph Cut is the objects and background prototypes identifications. Usually this task is accomplished through an interactive session in which users select seeds on the image to initialize the process. Interaction reduce the reproducibility of the method making preferable a fully automatic procedure [41]

Graph Cut in Symmetry Space



(a) (b)

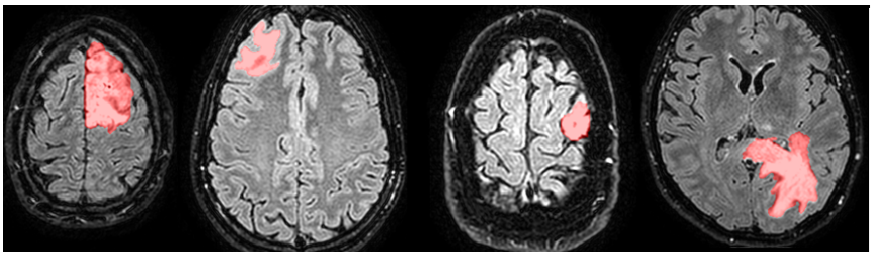
Figure 2.31: a)Differential Volume Clustering b) Left and Right Cluster Histograms

In this Ph.D Thesis novel solutions have been investigated to make the Graph Cut

segmentation method fully automatic, in an attempt to improve the reproducibility and operativity of the overall framework.
The conventional approach builds the graph starting from the voxels $v \in V$ that make up the scene we want to label. In our approach we have defined a volume D computed reflecting the right hemisphere across the brain plane of symmetry and computing the normalized difference, voxel by voxel, from the left hemisphere and this mirrored hemisphere. In this volume the components that are not symmetric with respect to the Mid Sagittal Plane in the original volume will have greater average values. From this observation it descends that prototypes for the object and the background can be easily identified by applying a trivial clustering on the features space. The whole segmentation process according to the Graph Cut framework can be applied to the volume D thus obtaining a fully automatic strategy.
The segmentation of the anomaly in the differential volume involves the loss of information on what is the ill hemisphere. Thus, a histogram analysis to identify the hemisphere where the tumor is localized is necessary. The mask of the object resulting from the clustering is mirrored across the mid sagittal plane and only the voxels selected by applying this mask on the original volume are taken into account to compute the left and right normalized histograms H_l and H_r. Let be H_{all} the normalized histogram of whole the MRI volume after the brain segmentation the ill hemisphere is the one whose histogram differs, by using the Euclidean distance, most from the histogram of the total volume as shown in Figure 2.31.

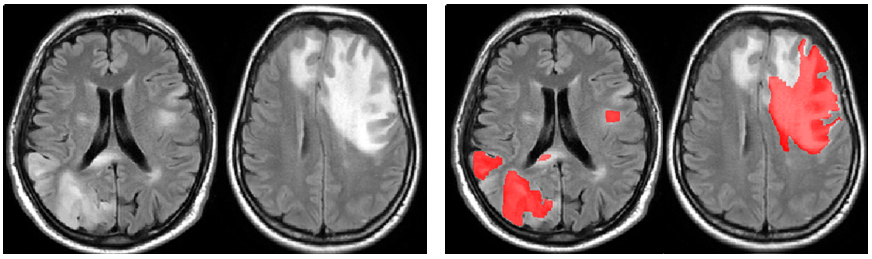
Competitive EM Graph Cut for Multicentric and Butterfly Bilateral Glial Tumor Segmentation

The use of the Graph Cut in Symmetry Space technique leads to very promising results and the masks obtained allow to compute with high accuracy, both volume and spatial distribution of the tumor (all the detail in Chapter 3). As highlighted by the comparison between Fig 2.32 e 2.33, high accuracy results can be obtained under the condition that pathological areas are localized in a single hemisphere.
This is a feature of many brain expansive mass especially in the class of Low Grade Glioma, but there are some cases in which the tumor can be bilateral or multicentric. To overcome this limitation, a new procedure has been designed and built on the top of the first segmentation that acts as initialization in the overall segmentation process. It is based on the combined use of Competitive Expectation Maximization [192] and Graph Cut. The combined use of the two techniques allows to integrate statistical and topological information reinforcing the robustness of the method which continue to be fully automated. In the next session the formal background related to the new general segmentation strategy is provided.



(a)

Figure 2.32: Examples of segmentation results obtained applying the proposed Graph Cut in Symmetry Space on Homolateral Gliomas, satisfactory results are observed



(a) (b)

Figure 2.33: Examples of segmentation results obtained applying the proposed Graph Cut in Symmetry Space on Bilateral Gliomas, poor results are observed

Competitive EM for Gaussian Mixture Model description

Gaussian mixtures have been extensively used in the field of statistical pattern recognition. Expectation Maximization algorithm (EM) is one of the most used tool for solving the Gaussian Mixture Model(GMM) [19, 54]. Unfortunately, this method often gets trapped at local maxima and sometimes converges to the boundary of the parameter space. Proceeding from this consideration in literature a lot of work was done for covering this gap [55, 178, 177]. The Competitive EM (CEM) acts in this direction. This algorithm, proposed by Zhang at al. in [192], is capable of automatically choosing the clustering number and selecting the split or merge operations efficiently based on competitive mechanisms avoiding the falling in the local minimums of the classical EM strategy. It is insensitive to the initial configuration of the mixture component number and model parameters. CEM is a very robust clustering method with a slightly heavier computation cost, especially in very complicated cases. In the follow we briefly summarize the CEM's formal elements needed to explain our strategy in the next section.
Let be $\boldsymbol{x} = [x_1, x_2, ..., x_d]^T$ a d-dimensional random variable whose probability distribution can be written as the composition of k components. Assuming that they are

Gaussian the formulation is the follows:

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{m=1}^{k} \alpha_m p(\boldsymbol{x}|\boldsymbol{\theta}_m) \tag{2.41}$$

where α_m is the prior probability of the m^{th} component and

$$p(\boldsymbol{x}|\boldsymbol{\theta}_m) = \frac{1}{2\pi^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T|\boldsymbol{\Sigma}|^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \tag{2.42}$$

The distribution is completely defined when the parameters $\boldsymbol{\theta_m} = [\boldsymbol{\mu}, \boldsymbol{\Sigma}]$, and the weight α_m of each gaussian component and the number of gaussian k are known. The goal of the CEM algorithm is to find the best set of these parameters. The basic step of the CEM is the classical EM. A brief overview is reported.
Let be considered a process in which you have some observed data X and some unknown data Z and parameters θ. The maximum likelihood estimate (MLE) of the unknown parameters is determined by the marginal likelihood of the observed data. The aim of the EM algorithm is to find MLE estimate of the marginal likelihood by iteratively applying two steps: *Expectation* that calculate the expected value of the log likelihood function, with respect to the conditional distribution of Z given X under the current estimate Θ and *Maximization* in which the parameters that maximize the marginal likelihood are found. The EM algorithm performs the E- and M-steps iteratively, and the convergence is theoretically guaranteed. When EM encounters local maxima, the components usually overpopulate in some regions, i.e. the model over-fits the data, but underpopulate in other regions. To overcome this problem, CEM provides the split or merge operation when EM has converged to a local maximum. The Kullback divergence is used for choosing if the split and marge operations are needed. To estimate the appropriate number of components the CEM use the same criterion presented by Figueiredo et. al. [67] that use a deterministic criteria for evaluating the models' state. If the S/M operations increase the value of model evaluation function the new parameters are accepted. However, if the operation decreases figure of merit of the models, they will not be rejected directly, but accepted via certain probability. This mechanism avoids the errors propagation. Moreover, to prevent the algorithm from converging to the bounds of parameter space, the mechanism of component annihilation is introduced.
Contrary to what was seen for the Graph Cut this method is fully statistical therefore ignore all the structural and topological information includes in the image data. Therefore, even though it is a very strong statistical clustering method it is inadequate for the image segmentation in which the use of contextual information is essential. Moreover, the final aim of the tumor segmentation is to find the optimal labeling $L = (L_{Bac}, L_{Obj})$ CEM provides only a clustering not a labeling.

Competitive-EM for Graph Cut Initialization(CEM Graph Cut)

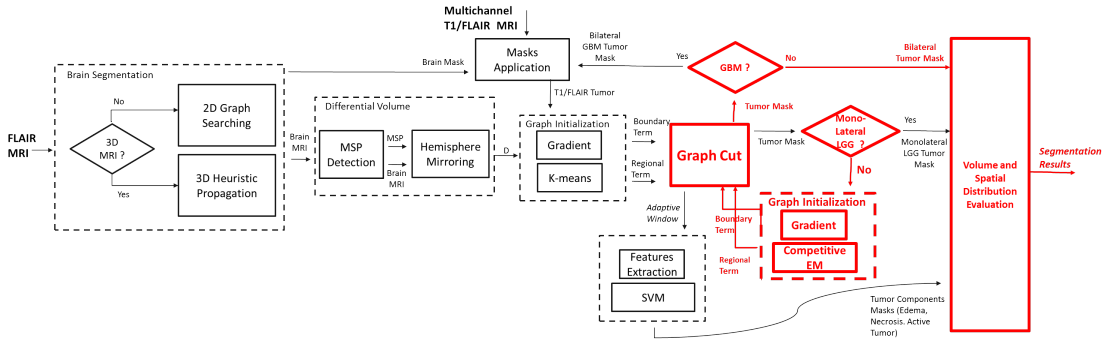


Figure 2.34: Schema of the Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation with highlight the part detailed in Section (2.7.4)

In the previous sections we have seen strengths and weaknesses of both the Graph Cut and Competitive EM, two promising and widely used techniques applicable to the segmentation problem.
By this analysis we realize that, if applied alone, both this techniques are inadequate for solving the complex task of MRI tumor segmentation. The Graph Cut is a very promising method, the combined management of local and global information makes it very powerful but the seeding phase is critical and hardly automatized with trivial techniques. The Competitive EM, because of its split and marge strategies, works in fully automatic way for solving the GMM, but its application in segmentation of the tumor is bounded to the inability of assigning the object and background labels. It is only capable to cluster the image regions. Moreover this is a statistical method that works in the features space, unless to give contextual information directly in the features vector the image topological constraints are completely neglected.
The salient aspect of the proposed generalized segmentation strategy is the complementary use of CEM and Graph Cut, allowing the integrated use of statistical and topological information managed by the two methods respectively. A hybrid strategy has been then defined in which the role of CEM is that of performing an automated initialization procedure(schema in Figure 2.34).
Consider to achieve a set of observation X_o X_b in a d-dimensional space of the unknown object and background distribution. Applying the Competitive EM strategy on these data we obtain two sets of parameters $\boldsymbol{\theta}^o$ and $\boldsymbol{\theta}^b$ that fully describes the statistics of the object and of the background:

$$\boldsymbol{\theta}_o = [k_o, \boldsymbol{\alpha}_o, \boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o]; \quad \boldsymbol{\theta}_b = [k_b, \boldsymbol{\alpha}_b, \boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b]; \tag{2.43}$$

where

$$\begin{aligned} \boldsymbol{\alpha}_o &= \{\alpha_{o,1}, ..., \alpha_{o,k}\} \quad \boldsymbol{\alpha}_b = \{\alpha_{b,1}, ..., \alpha_{b,k}\} \\ \boldsymbol{\mu}_o &= \{\boldsymbol{\mu}_{o,1}, ..., \boldsymbol{\mu}_{o,k}\} \quad \boldsymbol{\mu}_b = \{\boldsymbol{\mu}_{b,1}, ..., \boldsymbol{\mu}_{b,k}\} \\ \boldsymbol{\Sigma}_o &= \{\boldsymbol{\Sigma}_{o,1}, ..., \boldsymbol{\Sigma}_{o,k}\} \quad \boldsymbol{\Sigma}_b = \{\boldsymbol{\Sigma}_{b,1}, ..., \boldsymbol{\Sigma}_{b,k}\} \end{aligned} \tag{2.44}$$

The sets $\boldsymbol{\theta}^o$ and $\boldsymbol{\theta}^b$ are used for describe the terminal nodes of the graph. k_o $S(source)$ and k_b $T(sink)$ terminal nodes are identified. By this way the segmentation is no longer binary but both the object and background are the composition of a set of components each described by a gaussian component of the mixture model. Let be $\mathcal{O}$ the set of the voxels v labeled object and $\mathcal{B}$ the set of the voxels v labeled background they are identified as:

$$\mathcal{O} = \bigcup_{i=1}^{k_o} \mathcal{O}_i \quad and \quad \mathcal{B} = \bigcup_{i=1}^{k_b} \mathcal{B}_i \tag{2.45}$$

The Regional Term $R(L) = \sum_{v \in V} R_v(O_v)$ of the Graph Cut is defined using the follows:

$$R_v(O_{Obj}) = \sum_{i}^{k_o} R_{v,i}(O_{Obj}); \tag{2.46}$$

$$R_v(B_{Bkg}) = \sum_{i}^{k_b} R_{v,i}(O_{Bkg}); \tag{2.47}$$

where

$$R_{v,i}(O_{Obj}) = \frac{1}{\alpha_{o,i}} (\boldsymbol{I}_v - \boldsymbol{\mu}_{o,i})^T \boldsymbol{\Sigma}_{o,i}^{-1} (\boldsymbol{I}_v - \boldsymbol{\mu}_{o,i}) \tag{2.48}$$

$$R_{v,i}(B_{Bkg}) = \frac{1}{\alpha_{b,i}} (\boldsymbol{I}_v - \boldsymbol{\mu}_{b,i})^T \Sigma_{b,i}^{-1} (\boldsymbol{I}_v - \boldsymbol{\mu}_{b,i}) \tag{2.49}$$

where $\boldsymbol{I}_v$ is the descriptio of the volxel v in the d-dimensional space. The Boundary term $B(L) = \sum_{v,w \in N} B_{v,w} \delta_{L_i, L_j}$ can be computed using a formulation:

$$B_{v,w} \propto \exp(-\frac{1}{2} (\boldsymbol{I}_v - \boldsymbol{I}_w)^T |\boldsymbol{\Sigma}_{v,w}|^{-1} (\boldsymbol{I}_v - \boldsymbol{I}_w)) \tag{2.50}$$

where $\boldsymbol{\Sigma}_{v,w}$ is the sum of the covariance matrix $\boldsymbol{\Sigma}_{o,i}$ and $\boldsymbol{\Sigma}_{b,j}$ relative to the object component and background component to which the voxels v and w are assign. This function penalizes a lot discontinuities between elements of similar value in the d-dimensional space. The segmentation solution is the min-cut of the graph built using the $t-links$ and $n-links$ described by the Regional and Boundary term defined. Using this hybrid strategy we emphasize the strengths of both the Competitive EM and Graph Cut com-

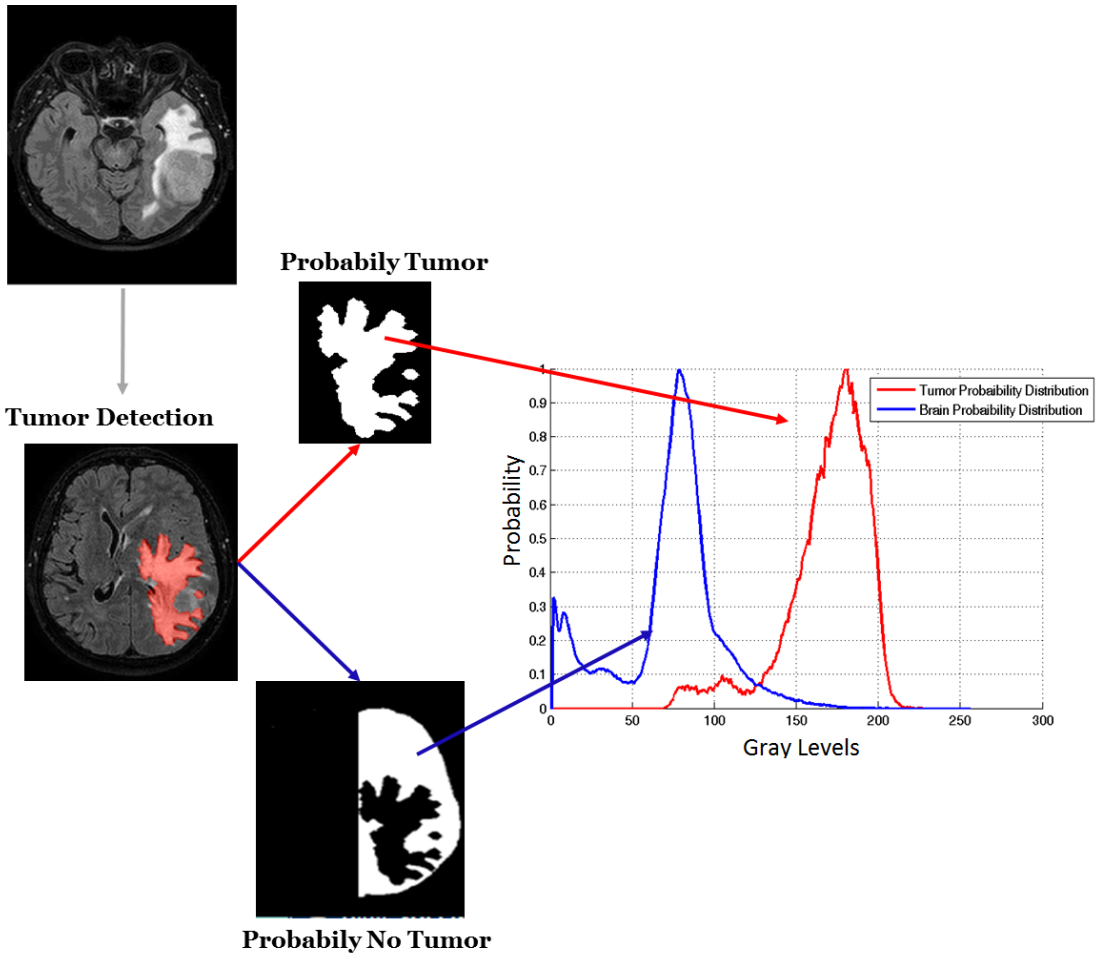


Figure 2.35: Brain and Tumor probability distributions extracted from the detected voxel

pensating for the weaknesses.

It is worth to note that the observed data X_o X_b may be directly used as a seed of the graph with the result that these points would have strong constraint to belong respectively to the object and to background would not be allowed errors in the collection of these observations. Our strategy instead provides a statistical description of these observations using the parameters obtained by the CEM application for the Graph initialization.

2.7.5 Competitive Expectation Maximization and Graph Cut for Brain Tumor Segmentation

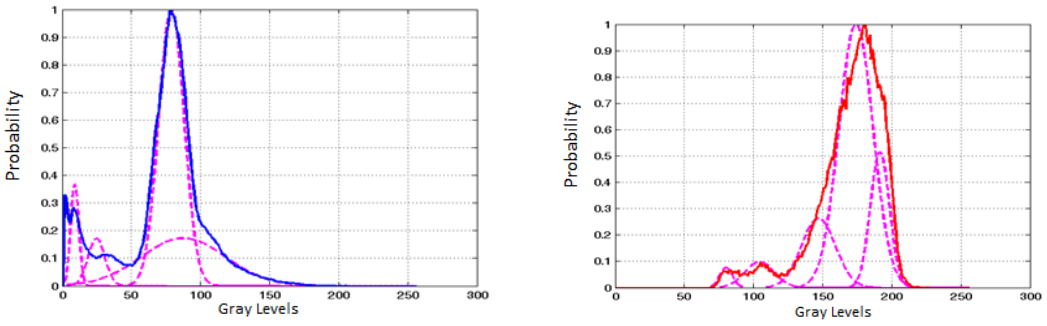
This section describes how the CEM Graph Cut can be applied for multicentric and butterfly tumor segmentation. Mimicking the human visual inspection process the first step of the tumor analysis process, is the detection of an anomaly with respect to the normal tissue, whatever its nature. As show in the previous section the use of the our Graph Cut in the symmetry space technique leads to very promising results for the homolateral tumors but in the cases of bilateral tumors reveals this technique inadequate. However, the segmented voxels may be consider a set of observation X_t X_b of the unknown tumor and brain gray scale level distribution. As done until now, we consider these models as a weighted sum of gaussian:

$$p_t(x|\boldsymbol{\theta}_t) = \sum_{m=1}^{k_t} \alpha_{t,m} \frac{1}{\sigma_{t,m}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_{t,m}}{\sigma_{t,m}})} \tag{2.51}$$

$$p_b(x|\boldsymbol{\theta}_b) = \sum_{m=1}^{k_b} \alpha_{b,m} \frac{1}{\sigma_{b,m}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_{b,m}}{\sigma_{b,m}})} \tag{2.52}$$

In this case the distributions are one-dimensional because we use the intensity level of the voxels as feature this procedure is detailed in Figure 2.35. It is worth noting that the proposed method is, however, applicable in d-dimensional space this is an essential feature especially in the field of tumors segmentation that often requires multichannel description of the MRI volume.

Applying the Competitive EM strategy on X_t X_b data we obtain the sets of parameters



(a) (b)

Figure 2.36: Brain and Tumor Gaussian Mixture Model

$\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_b$ required for describe the $p_t(x|\boldsymbol{\theta})$ and $p_b(x|\boldsymbol{\theta})$

$$\boldsymbol{\theta}_t = [k_t, \alpha_t, \boldsymbol{\mu}_t, \boldsymbol{\sigma}_t]; \quad \boldsymbol{\theta}_b = [k_b, \alpha_b, \boldsymbol{\mu}_b, \boldsymbol{\sigma}_b]; \tag{2.53}$$

The components of Regional Term, therefore the strength of the $t-links$, is defined using the follows:

$$R_{v,i}(T_{Tum}) = \frac{(I_v - \mu_{t,i})^2}{\alpha_{t,i}\sigma_{t,i}} \tag{2.54}$$

$$R_{v,i}(B_{Bre}) = \frac{(I_v - \mu_{b,i})^2}{\alpha_{b,i}\sigma_{b,i}} \tag{2.55}$$

where I_v is the intensity value of the voxel v. The strength of the $n-link$ that connect to the pair of neighboring voxels $\{v, w\}$ is:

$$B_{v,w} \propto \exp(-\frac{(I_v - I_w)^2}{2\sigma}) \tag{2.56}$$

where I_v and I_w are the intensity value of the voxels v and w. The min-cut of the graph as built allows to identify the subsets of voxels that compose of the tumor and the brain. The final binary segmentation is found with the union of these subsets for the tumor and brain respectively.

$$\mathcal{T} = \bigcup_{i=1}^{k_t} \mathcal{T}_i; \quad and \quad \mathcal{B} = \bigcup_{i=1}^{k_b} \mathcal{B}_i \tag{2.57}$$

In Figure 2.37 a graphical representation of the graph transformation is shown.

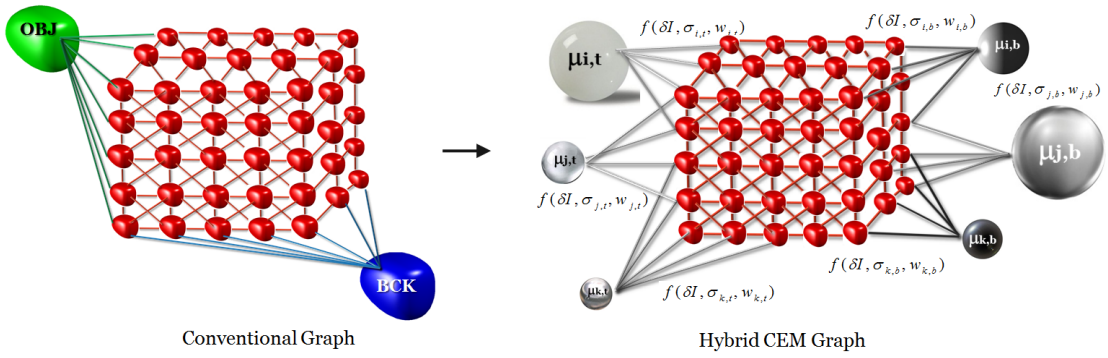


Figure 2.37: Transformation of the traditional graph in the graph defined by the hybrid CEM Graph Cut strategy

2.7.6 Multichannel Tumor Heterogeneity Recognition

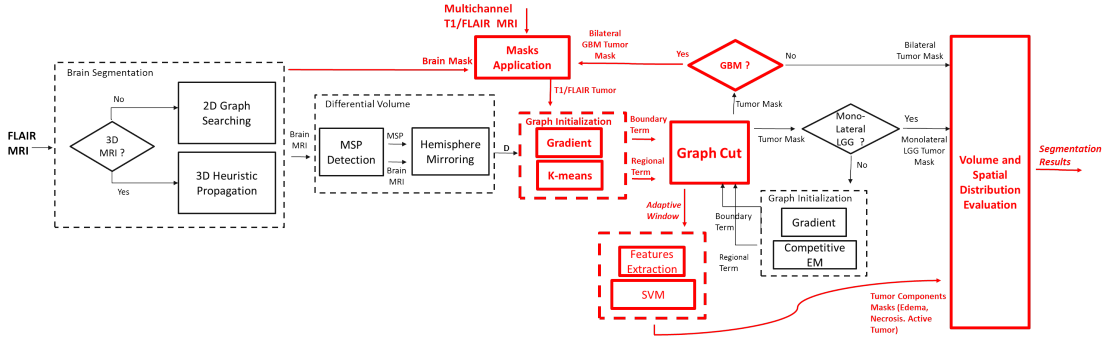


Figure 2.38: Schema of the Integrated Framework for Fully Automatic MRI Glial Tumor Segmentation with highlight the part detailed in Section (2.7.6)

The last part of the proposed integrated framework conceived the use of a supervised technique for recognizing the heterogenous parts of the tumor: *Necrose Active Part and Edema*(Figure 2.38). The Machine Learning tool used is the Support Vector Machine(SVM). Founded on the Statistical Learning Theory [180], SVMs are currently the state-of-the-art algorithm for solving binary classification problems [51, 77, 89] and have shown excellent results for various pattern recognition tasks also in medical imaging and more particular in brain tumor segmentation [7, 71, 195]. Support Vector Machines are a very powerful tool for classification tasks due to their appealing generalization properties. However, SVMs assume that data (in the simplest case of application, individual voxels) is *independently and identically distributed(iid)*, which is not appropriate for tasks such as segmenting medical images. In particular, SVMs can not consider dependencies in the labels of adjacent pixels/voxels.
Some authors have worked for overcoming this gap using together with the SVMs Random Field and Graph Theory [16, 96]. In both these approaches the learning procedure of the SVM is modified to develop a new hybrid strategy able on one hand to menage the topological information through the use of Graph or Random Filed and on the other hand to exploit the good classification performances of the SVMs.

For solving our recognition problem we decided to adopt a quite different and more simple procedure, privileging the method's transferability, although with encouraging performance. It is worth to note that the task we have to accomplish is easier than the aim of the cited study. The goal of the last step of our execution pipeline is the recognition of the Edema Necrose and Active part after a preliminary tumor segmentation (Figure 2.39). In the cited work the SVMs technology is used for solving the whole segmentation problem; but, this so pretentious goal often requires the use of post-processing heuristic techniques for the errors rejection.

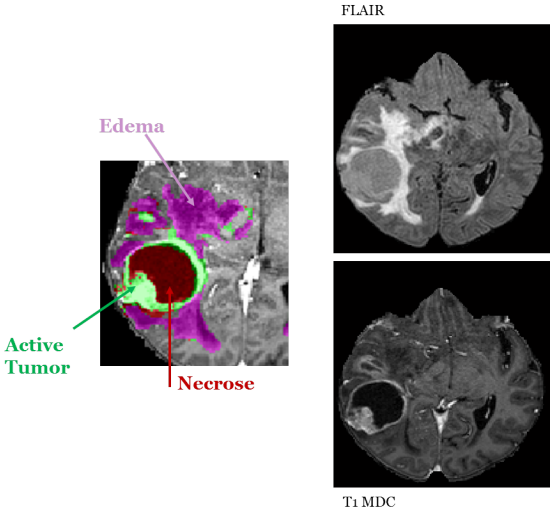


Figure 2.39: Example of Multichannel Tumor Heterogeneity Recognition

The MRI images depicted in Figure 2.39 represent the same axial slice of two different volumetric MRI acquired from the same patient with two different acquisition modality; the first is a FLAIR MRI, and the second is a post gadolinium T1 MRI. Observing the expected results in Figure 2.39 is easy to note as the alone use of each modality would prove insufficient for accomplishing the recognition task. Proceeding from this consideration we adopt for this last phase a Multichannel approach. In the recent years this has been the common trend also for tumor segmentations purpose [49, 121]. In our proposed framework we preferred to investigate segmentation solutions that can good perform, also if it is applied on a single MRI channel, imposing the need of a multichannel data only at this last phase.
In the work presented by Kwang et.al. [89] the SVM potentialities in texture classification are investigated. In this paper is detailed as the both feature extraction and classification are performed within the SVM architecture. Indeed, in the experiment performed by Kwang et.al. gray-level values of the raw pixels are directly fed to the SVM. The only preprocessing of the input, before feeding it to the SVM, is the selection of certain pixels following the configuration of autoregressive (AR) features. SVM incorporates feature extractors, therefore, nonlinear mapped input patterns can be used as feature vectors. Actually, in an SVM, feature extraction is implicitly performed by a kernel, which is defined as the dot product of two mapped patterns. Furthermore, SVMs with such a kernel can provide solutions to the problems inherent in conventional feature extraction methods. The main advantage of the proposed method is that there is no need for a carefully designed feature extraction mechanism because the feature extraction task is reduced to the problem of training the SVMs.

Proceedings form these considerations seems to be inadequate to implement a features extraction procedure, both problem based and statistical, for accomplishing the recognition task. The solutions that seems the simplest and theoretical the best is to build the SVM training pattern with the couple of voxel coming from FLAIR and post gadolinium T1 MRI; but as mentioned above SVMs assume that data is *independently and identically distributed(iid)*, then the contextual information should be included in the training pattern. The most immediate solution would be to use a sliding volumetric window building the pattern which represents the central voxel of the window with all the voxels in both the FLAIR and post gadolinium T1 MRI. The problem become however how to chose the size of the window. Must also be considered that working on volumetric data is difficult to apply pyramidal or multiresolution approaches having regard that the size of the pattern quickly become unmanageable. Our goal is to keeping down the size of the training patterns resulting in an improved generalization performance and classification speed.
A last premise should be made, SVMs are originally designed for binary classification and the optimal design of multiclass SVM classifiers is still an area of active research [50, 181, 4]. The conventional way is to decompose the $M-class$ problem into a series of $2-class$ problems and construct several binary classifiers. One frequently used method is *one-against-others* decomposition [40], which constructs M SVM classifiers with the i_{th} one separating class i from all the remaining classes. In our strategy we use this method.

Graph Features Extraction

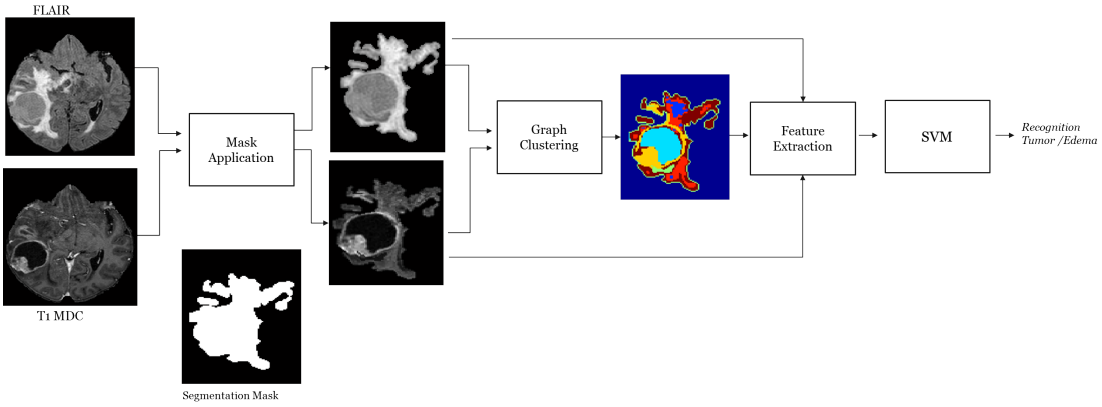


Figure 2.40: Schema of the SVM Multichannel recognition strategy

Our procedure is based on the definition of an adaptive windows using a Graph clustering strategy. The Schema of the features extraction and the binary SVM classification is shown in Figure 2.40. The recognition procedure is applied only on the voxels that

have been labeled as tumor in the previous stage.
TO do (adaptive dilatation)
The vectorial volume consisting of FLAIR and post gadolinium T1 tumor voxel is clusterized using the same Graph cut technique applied on the differential volume (Section 2.7.4) the number of cluster is an external parameter of the method.
Each region of connected pixels i belonging to a cluster j is characterized by a simple statistical moments the means the variance and the co-variance:

$$\boldsymbol{\mu_k} = [\mu_{k_{T1}}, \mu_{k_F}]; \tag{2.58}$$

$$\boldsymbol{\sigma_k} = [\sigma_{k_{T1}}, \sigma_{k_F}, \sigma_{k_{T1,F}}]; \tag{2.59}$$

where the index k is $j * i$. The features vector used for recognize each voxel $v(x, y, z)$ belonging to the k_{th} region is made up by different components:

- The couple of voxel coming from FLAIR and post gadolinium T1 MRI : $[v(x, y, z)_{T1}, v(x, y, z)_F]$.
- The six adjacent for both the volumes $[v_(x-1, y, z)_{T1}, v(x, y-1, z)_{T1}, ..., v(x, y, z+1)_{T1}, v_(x-1, y, z)_F, v(x, y-1, z)_F, ..., v(x, y, z+1)_F]$.
- The statistical description of the membership cluster $[\boldsymbol{\mu_k}, \boldsymbol{\sigma_k}]$.

This vector with 19 components is feeding to the SVM. The use of the adaptive windows allows to give a dynamic description of the contextual information of the voxel both in terms of shape and dimension. With this strategy are not necessary padding operations for the border elements and also the voxels belonging to an edge between two classes are well described.

3

MRI Analysis: Experimental Results

3.1 MRI Segmentation Evaluation

The aim of this chapter is to evaluate the performance of the methods proposed for brain and tumor segmentation and for the recognition. In this critical context the availability of reliable quantitative measures of accuracy and precision for the proposed segmentation method plays a key role [42, 186].
According to Warfield et al. [186] the accuracy of an algorithm in creating a segmentation is the degree to which the segmentation corresponds to the true segmentation, against which it may be compared. Precision is determined by the stability of the results obtained repeatedly from the same imaging conditions. Differently from precision that may be assessed without comparison, the assessment of accuracy of a segmentation requires a reference standard, representing the true segmentation. Since a direct measure of ground truth in this context, is not logistically feasible, or even possible with pathologic correlation, several methods have been proposed to circumvent the problem, including the use of MRI contrast studies, correlation with pathologic findings, phantom MRI simulation studies. These methods show potentialities but also strong limitations due to invasiveness, low degree of confidence, unrealistic assessment, partial evaluation, dependence from method and scene considered. Physical or digital phantoms in particular, can provide a reliable level of known "ground truth" [42, 186], however they have so far been unable to reproduce the full range of imaging characteristics such as partial volume effects, intensity inhomogeneity, noise and anatomical variability observed in clinical data.

Due to the difficulty of establishing a standard for this type of data, recent works focus the attention on methods which do not require a ground truth, but instead rely on behavioral comparison [186, 141, 26].
In this approach, accuracy assessment is addressed using a group of experts to manually trace the boundaries of the regions of interest. Automated results are then compared with those produced by manual labelling to decide if the automated segmentation algorithm can be considered an acceptable substitute for the experts. The motivation of the behavioral comparison method is that direct labeling has the undisputed value of mimicking the radiologist decision attitude which realistically is the only truth available; human raters hold some prior knowledge of the ground truth that is reflected in their manual tracings [186, 141, 26].
Basing on a behavioral comparison strategy, the evaluation involves two important components: first, the choice of appropriate similarity metrics and second, the design of a reliable common agreement strategy able to define a suitable reference standard by taking a combination of expert segmentations.
The most appropriate way to carry out both these aspects is still under investigation [26].
A number of metrics have been proposed to compare segmentations. Simply measuring the volume of segmented structures has often been used. However, volume estimates may be quite similar when the segmented structures are located differently, have different shapes or have different boundaries and so alternative measures are introduced [95]. Other measures used in practice include measures of spatial overlap, such as the Dice and Jaccard Similarity Coefficients [56, 81].
As regard the second aspect concerning the definition of a reference standard, strategies for combination of multiple segmentations, also mentioned as label fusion, have recently been introduced as effective methods to obtain segmentations that are more accurate than any of the individual input segmentations. The intimate aim of the methods is to produce a single, discrete element or label from a combination of several independent candidate segmentations results. The underlying assumption is that the merged result is potentially more accurate and representative of a common agreement, with respect to the exact, sought segmentation, than each individual input. Majority Voting is the generally used rule to fuse the segmentations [78], but more sophisticated methods have also been proposed. A well-known technique called Simultaneous Truth and Performance Level Estimation (STAPLE), initially proposed by Warfield et al. [186, 142], is used in a variety of studies [5, 142]. A third approach, referred to as Shape-Based Averaging (SBA), incorporates spatial information [143].
Robitaille and Duchesne [141] developed an empirical studies in which the three above mentioned label fusion techniques were compared. The study comes to the important conclusion that none of the above techniques could be considered superior depending on

the dissimilarity between the input elements.
In the follow the experimental results obtained using the segmentation strategies proposed in this thesis are presented. All the methods, both for the brain and tumor segmentation and recognition are qualitative and quantitative evaluated. By way of introduction to all the experimental phase some results concerning the use of the proposed strategies for the manual annotation are presented.

3.2 Ground Truth Estimation Evaluation

We carried out a set of experiments addressing the following main issues:

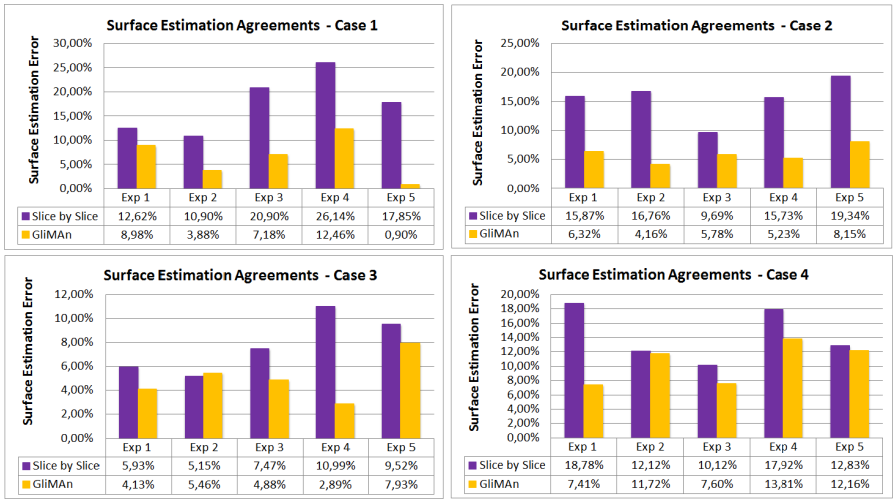
- to evaluate how and at which extent *GliMAn* contributes to the reduction of operation variability in fully manual segmentation
- to isolate the contribution of the heuristic solutions adopted in the initialization procedure for the fuzzy connectedness segmentation algorithm
- to compare the fuzzy connectedness based strategy with other simultaneous truth estimation approaches.

3.2.1 Experiments in fully manual labeling using *GliMAn*

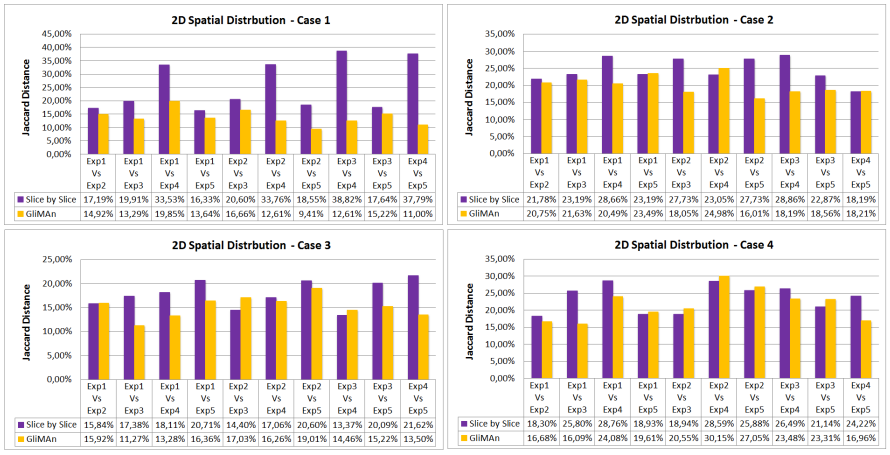
The same group of experts, who worked in the operation variability analysis, has been involved again to segment with the support of *GliMAn* 4 slices for each of the 4 MRI volumes (case 1-4) of our dataset. We measured the 2D inter- variability using the metrics described in section 2 and compare the results with those obtained using the conventional annotator (see Figure 3.1(a) and Figure 3.1(b)). Results are expressed in terms of mean of the surface estimation error and mean of the Jaccard distance respectively, varying the 4 slices segmented by each experts. The use of *GliMAn* has determined a significant reduction of the surface estimation error, with a maximum value equal to 16.95% for case 1 expert 5 and minimum equal to -0.30% for case 3 expert 2. The average reduction of the Jaccard distance is equal to 5.14%, with a maximum value equal to 26.79% for case 1 between experts 4 and 5; and minimum equal to -2.63%. for case 3 between experts 2 and 3.

3.2.2 Experiments of novel solutions proposed in fuzzy connectedness segmentation

A second set of experiments is developed with the aim of investigating how the performances of the fuzzy segmentation method are influenced by the specific initialization procedure conceived in our strategy simultaneous estimation. In the first experiment we evaluate the performances of the modified fuzzy connectedness algorithm when solving



(a)



(b)

Figure 3.1: Mean of the Surface Errors(a) and 2D Jaccard distances computed for each expert, varying the 4 slices segmented using conventional and *GliMAn* tools.

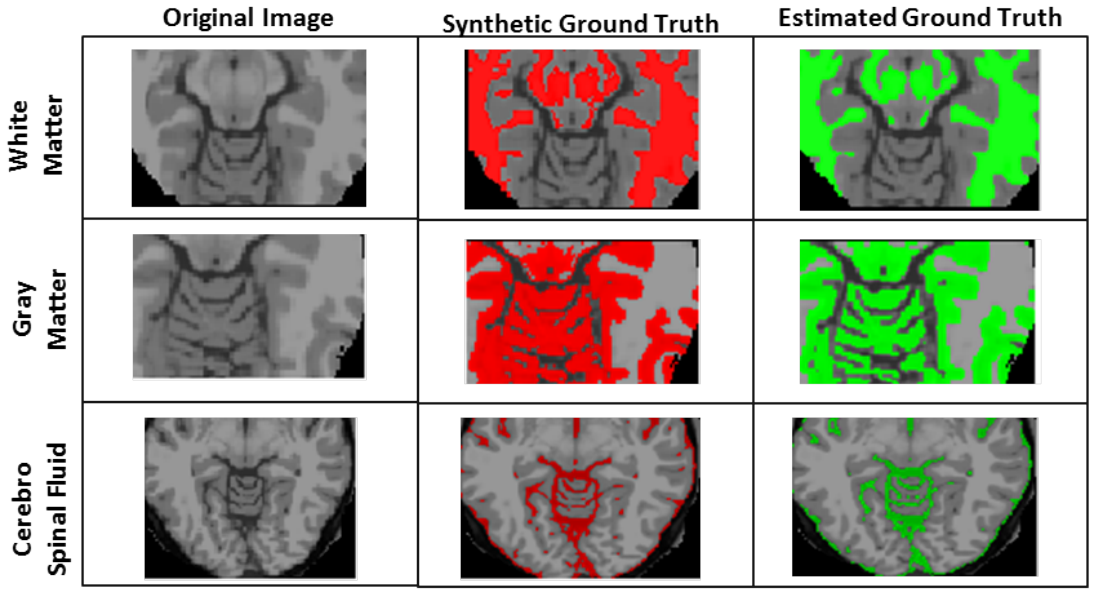


Figure 3.2: Comparison between fuzzy based consensus segmentation and syntectic ground truth in White Matter(WM) Grey Matter(GM) and Cerebral Spinal Fluid(CSF) Segmentation

Table 3.1: Accuracy results of Fuzzy Connectedness consensus segmentation in White Matter(WM) Grey Matter(GM) and Cerebral Spinal Fluid(CSF) Segmentation.

	Syntetic True Volume	Fuzzy Estimation Volume	Volume Estimation Error	Jaccard Distance
WM	426805	421231	1.30%	1.99%
GM	232758	234264	0.64%	1.88%
CSF	218427	213270	2.36%	9.50%

a segmentation problem in brain MRI domain.
We use data from the Simulated Brain Database (BrainWeb) including simulated human brain MR images and ground truth for spatial distribution of White Matter (WM) Gray Matter(GM) and Cerebral Spinal Fluid (CSF). [45]. A simulated T1-weighted MRI (isotropic voxel 1 mm) of a healthy brain was used in our experiments. The fuzzy truth estimation algorithm has been initialized with information provided by 5 experts for the WM, GM and CSF segmentation. The obtained fuzzy objects are hardened through the application of a suitable threshold easy set by visual inspection. Figure 3.2 shows an example of the results in complex areas. Table 3.1 reports accuracy results in terms of Volume Estimation Error and 3D Jaccard Distance. Results obtained are aligned with those obtained in experimental studies presented in [147] and [174], in which the standard formulation of the fuzzy connectedness method has been applied to solve the same segmentation problem with the same benchmark dataset. From this comparison

Table 3.2: Willian Indexes obtained by manual labeling of five experts and the two configuration Algorithm 1 (a) and Algorithm 2 (b) of fuzzy connectedness simultaneous truth estimation configured

(a)

Cases	**Slices**	**Exp1**	**Exp2**	**Exp3**	**Exp4**	**Exp5**	**Algorithm 1**
Case 1	Slice 1	1.01	1.01	0.96	1.01	1.01	**1,01**
	Slice 2	1.01	0.99	0.99	1	1.02	**0,99**
	Slice 3	0.97	1.03	1	1	1.03	**0,97**
	Slice 4	0.96	1.06	0.95	1.02	1.08	**0,94**
Case 2	Slice 1	0.99	1.01	1.02	0.95	1.02	**1**
	Slice 2	1.03	1.04	0.99	0.92	1.02	**1,01**
	Slice 3	1.01	1.01	0.99	0.98	1.02	**1**
	Slice 4	0.85	1.02	1.03	1.02	1.07	**1,02**
Case 3	Slice 1	1.04	0.99	1.07	1.09	1.08	**0,99**
	Slice 2	1.01	0.98	1.01	1	1	**1**
	Slice 3	1	1.01	1	1.01	0.98	**1**
	Slice 4	1.03	1.01	1.01	0.98	0.98	**1**
Case 4	Slice 1	1.04	0.93	1.01	0.96	1.04	**1,02**
	Slice 2	1.02	1	1.01	0.97	0.99	**1,01**
	Slice 3	1.02	1.01	1.01	0.97	0.98	**1,01**
	Slice 4	1.05	1.02	1.02	0.98	0.97	**0,98**

(b)

Cases	**Slices**	**Exp1**	**Exp2**	**Exp3**	**Exp4**	**Exp5**	**Algorithm 1**
Case 1	Slice 1	1.01	1.01	0.96	1.01	1.01	**1.01**
	Slice 2	1.01	0.99	0.99	1	1.02	**0.99**
	Slice 3	0.97	1.03	1	1	1.03	**0.97**
	Slice 4	0.96	1.06	0.95	1.02	1.08	**0.94**
Case 2	Slice 1	0.99	1.01	1.02	0.95	1.02	**1**
	Slice 2	1.02	1.03	1	0.92	1.02	**1.01**
	Slice 3	1	1.01	0.99	0.97	1.02	**1.01**
	Slice 4	0.86	1.02	1.03	1.01	1.07	**1.01**
Case 3	Slice 1	1.08	1.02	1.1	1.12	1.11	**0.56**
	Slice 2	1.01	0.99	1.01	1	1.01	**0.98**
	Slice 3	1	1.01	1	1.01	0.99	**0.98**
	Slice 4	1.04	1.02	1.01	1	1	**0.92**
Case 4	Slice 1	1.04	0.93	1.01	0.96	1.03	**1.03**
	Slice 2	1.02	1	1.01	0.97	0.99	**1**
	Slice 3	1.02	1.01	1.01	0.97	0.98	**1**
	Slice 4	1.04	1.02	1.02	0.98	0.98	**0.96**

we can have a first positive evaluation of the solutions adopted to integrate a common agreement in the fuzzy framework to the extent that they do not worsen the performance of the standard version when coping with segmentation tasks.
As a second step, we proposed a direct use of the modified fuzzy connectedness framework for simultaneous truth estimation, and evaluate its goodness within a behavioral comparison strategy. We make use of Williams' index [188], which is used in neuroimaging studies [26] and formulated in the following.
Consider a set of r raters labeling the finite grid L of n voxels with labels $1, 0$. Let X_j be the set of voxels labeled 1 by rater j and $s(X_j, X'_j)$ the similarity between rater j and j' over all n voxels. Williams' index for rater j is defined as:

$$WI_j = \frac{(r-2)\sum_{j' \neq j}^{r} s(X_j, X'_j)}{(2)\sum_{j' \neq j}^{r} \sum_{j' \neq j}^{j'-1} s(X_j, X'_j)} \tag{3.1}$$

Several similarity measures can be used, in our experiments we express $s(X_i, X_j)$ in terms of Jaccard similarity index.
Initially we consider as realters the five experts and the fuzzy connectedness algorithm modified according the strategy presented in Section 4; secondly the William Index analysis considers as realters again the five experts the fuzzy connectedness algorithm in which multiple seeds and regions provided by experts are interpreted as provided by a unique experts and then trivially fused. Segmentations are performed on the same dateset used in Section 6.1. Results obtained are shown in Table 3.2. The performance levels obtained by the two configurations considered for the fuzzy connectedness algorithm are comparable with the exception of case 3 which includes highly heterogenous scenes difficult to interpret and for which Algorithm 1 strongly prevails in estimating a simultaneous truth We may conclude that the fuzzy connectedness framework benefits from the insertion of the information fusion strategy presented in section 4.

3.2.3 Comparison Analysis

A set of experiments are conducted to compare Fuzzy Connectedness-based truth estimation with Majority Voting rule and STAPLE. As a first step we intend to visually compare the results obtained by manual labeling produced with the support of *GliMAn* and the estimation strategy considered.
In Figure 3.3 results obtained by manual labeling, Majority Voting, STAPLE and Fuzzy Connectedness, when processing slice 3 of case 2 are shown. To correctly interpret the results obtained in the comparison, we have to take into account that Staple works within a statistical framework and directly processes the total sets of labels produced by the experts in the fully manual segmentation. Majority Voting manages the same information as Staple in a simple determinist manner. The fuzzy connectedness frame-

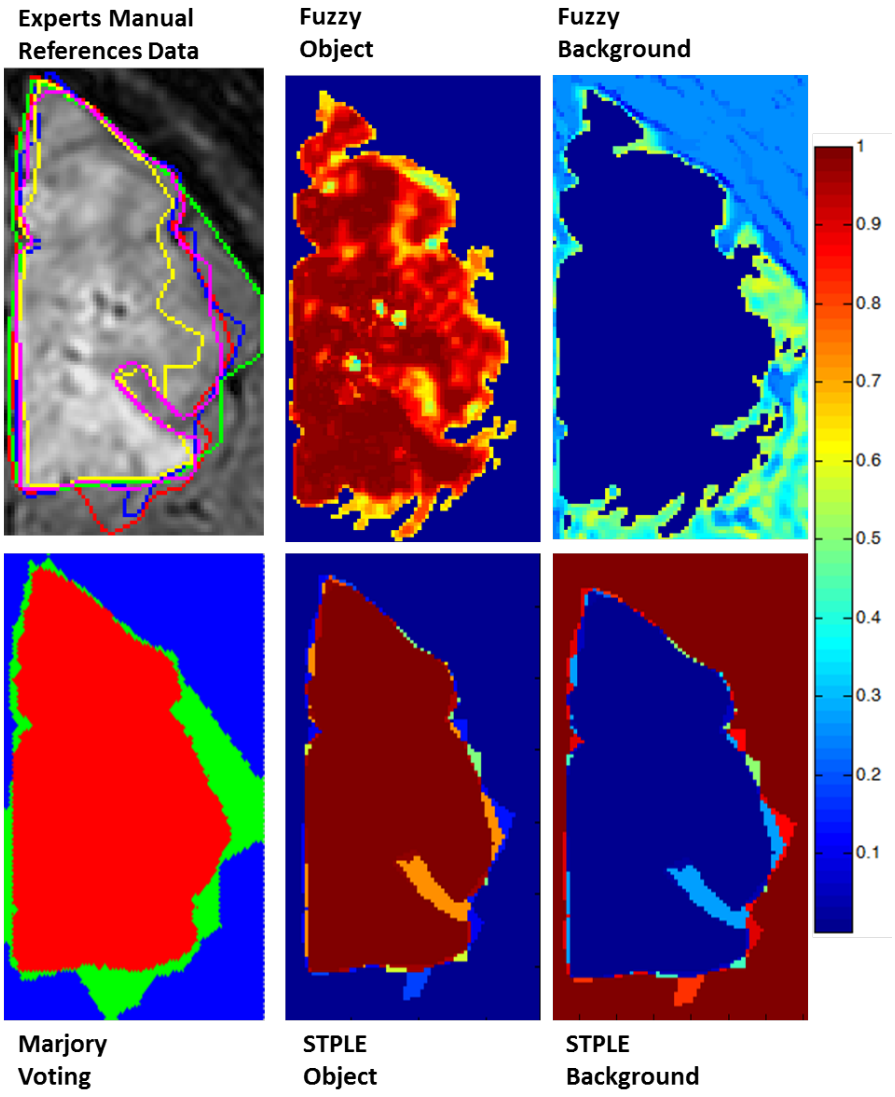


Figure 3.3: Comparison among segmentation manually traced by five medical experts and the results obtained by Fuzzy Connectedness (fuzzy object and fuzzy background), Majority Voting and STAPLE (soft STAPLE object and soft STAPLE background). The regions in which Majority Voting decides for tumor label are red and regions in which frequency for tumor label is non zero but Majority Voting decides for background label are green. Both the STAPLE and Fuzzy Connectedness output are depicted in MATLAB jet scale forom 0 to 1)

Table 3.3: Means of fuzzy output values computed in the regions in which Majority Voting and STAPLE decide for tumor label (Agreement) and regions in which frequency for tumor label is non zero but Majority Voting and STAPLE decide for background label (Disagreement)

Cases	Slices	Majory Voting		STAPLE	
		Agree Tumor	Disagree	Agree Tumor	Disagree
Case 1	1	0.761	0.443	0.759	0.445
	2	0.760	0.460	0.758	0.453
	3	0.749	0.485	0.745	0.487
	4	0.692	0.426	0.692	0.426
Case 2	1	0.858	0.549	0.856	0.546
	2	0.853	0.583	0.852	0.581
	3	0.886	0.591	0.885	0.590
	4	0.803	0.572	0.800	0.569
Case 3	1	0.756	0.660	0.755	0.657
	2	0.811	0.666	0.810	0.666
	3	0.825	0.666	0.825	0.666
	4	0.797	0.678	0.797	0.678
Case 4	1	0.826	0.593	0.830	0.594
	2	0.810	0.592	0.802	0.602
	3	0.808	0.546	0.807	0.547
	4	0.761	0.499	0.760	0.482

work prevents the exploiting of the total set of labels in manual labeling, and proceed from of high reliable but incomplete information (seeds and regions). Looking into the details of the Figure, we note that manual segmentations, reported in the first image of the first row in Figure 3.3, presents strong dissimilarities. These differences, that we may consider intrinsic in the labeling process and related to uncertainty in decision, are conservatively reduced by Majority Voting that inevitably cause a loss of information (see the first image of the second row in Figure 3.3). As expected the estimation performed by STAPLE is highly consisted with the results obtained by Majority Voting but in the soft output for both object and background (second and third image of the second row in Figure 3.3) we may recognize regions at different level of probability in a number equal to the number of experts. The Fuzzy Connectedness estimation accommodates grade independently from the dissimilarities of the manual segmentation. Focusing on the second and third images of the first row in Figure 3.3 corresponding with the fuzzy relative object and background respectively, we may note that regions with a low level of agreement among experts judgments presents a high level of fuzziness.

To better investigate the consistency among the estimation methods considered, we computed the average of fuzzy output values in the regions in which Majority Voting and STAPLE decide for tumor label and those for which frequency for tumor label is non zero but they decide for background label. Results obtained are reported in Table 3.3 showing a high level of consistency between the estimation procedures.

In support to this quantitative analysis visual results are shown in Figure in terms of manual segmentations, Majority Voting STAPLE and Fuzzy Connectedness estimations

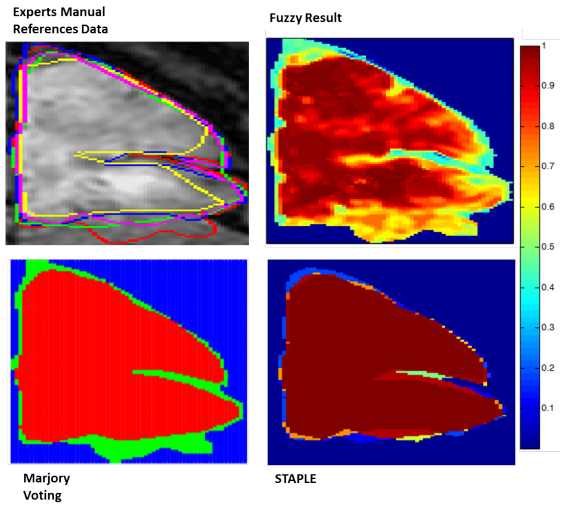


Figure 3.4: Qualitative evaluation of the fuzzy output values located in the regions in which Majority Voting and STAPLE decide for tumor label (Red in Majority Voting) and regions in which frequency for tumor label is non zero but Majority Voting and STAPLE decide for background label (Green in Majority Voting)

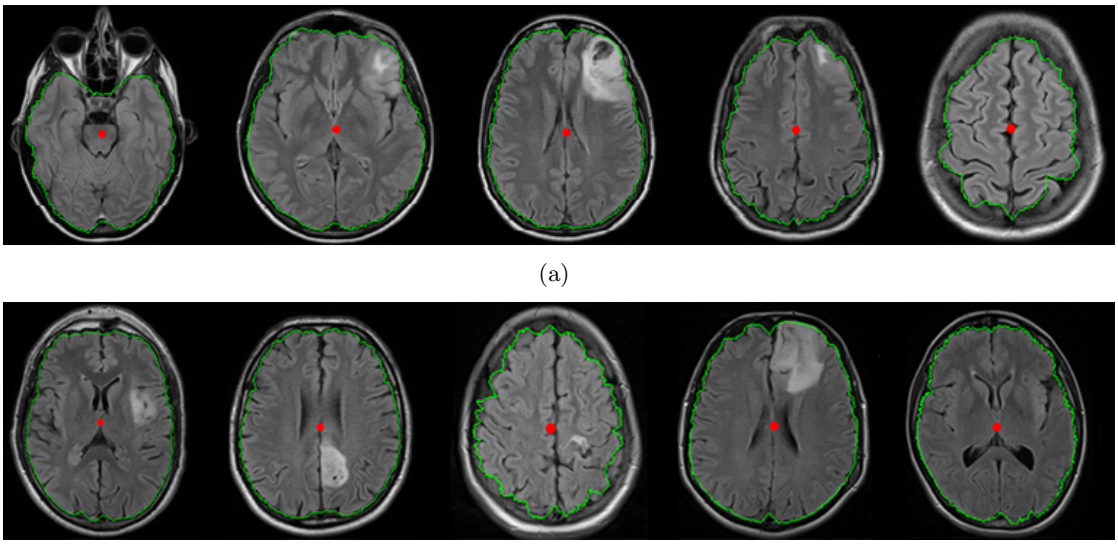
outputs, when processing slice 1 of case 2. The Fuzzy Connectedness output, (Figure second image of the first row, is obtained by accommodating corresponding grades in non zeros frequency tumor area in STAPLE and Majority Voting output. This qualitative analysis is aligned with the quantitative results.

3.3 2D Brain Segmentation Evaluation

Table 3.4: Data set for the brain segmentation evaluation Description

Case	Acquisition Mode	Volume Size (voxel)	Slice Thickness (mm)	Spacing Between Slice(mm)	Pixel Spacing (mm)	Repetition Time (ms)	ECHO Time (ms)
1	Assial FLAIR-T2	$[228 \times 228 \times 22]$	5	6	(0.80,0.80)	11000	140
2	Assial FLAIR-T2	$[320 \times 320 \times 23]$	5	6	(0.75,0.75)	11000	140
3	Assial FLAIR-T2	$[228 \times 228 \times 23]$	5	6	(0.80,0.80)	11000	140
4	Assial FLAIR-T2	$[320 \times 320 \times 23]$	5	6	(0.75,0.75)	11000	95
5	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.94,0.94)	7800	95
6	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.90,0.90)	7800	95
7	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.90,0.90)	7800	95
8	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(1.02,1.02)	7800	95
9	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.98,0.98)	7800	95
10	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.98,0.98)	7800	95
11	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	282.89
12	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	281.93
13	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	281.38
14	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	282.43

The dataset used for the 2D brain segmentation evaluation process is composed of 14 sparse FLAIR MRI gray scale 12 bit depth volumes (Table 3.4). All dataset volumes are altered by the presence of glial tumors heterogeneous in terms of position, dimension, intensity and shape. The dataset is composed of two parts. The first 10 MRI images are obtained with a FLAIR sequence; the slice thickness is 5 mm, the sequence is characterized by a long repetition times and by a reversal of the spin pulse at 180 degrees; the measurement is performed when the value of the liquid is close to 0 allowing the removal of the signal of liquid normally hyperintense in sequences with long repetition time. The Case 9 and 10 are post surgical MRI playing the role of assessing the performances of our segmentation strategy also in cases in which the most commonplace assumptions about the shape of the brain such as symmetry are are not valid.
The second part of the dataset is composed of volumetric dense acquisition with isotropic voxel (0.57 mm). This high resolution allows to detect fine anatomical detail, to reconstruct on different orthogonal planes (axial, sagittal, coronal) and to make accurate measurements. These advantages however are obtained at the expense of signal to noise ratio. The results obtained from the varied set of experiments have been evaluated qualitatively by a group of experts by comparing the results produced by the automatic segmentation with those obtained by manual segmentation. In order to limit the effects of the inter- and intra-observer variation, our strategy contemplates the organization of tuning sessions aimed at establishing a consensus among experts through discussion of the most controversial segmentation cases.
An analysis of the stability of the results in function of the variations of the center coordinates in the polar reference system was performed. The eyes slice detection algorithm is tested on all the dataset volumes and the statistical significance of the EPI index is



(a)

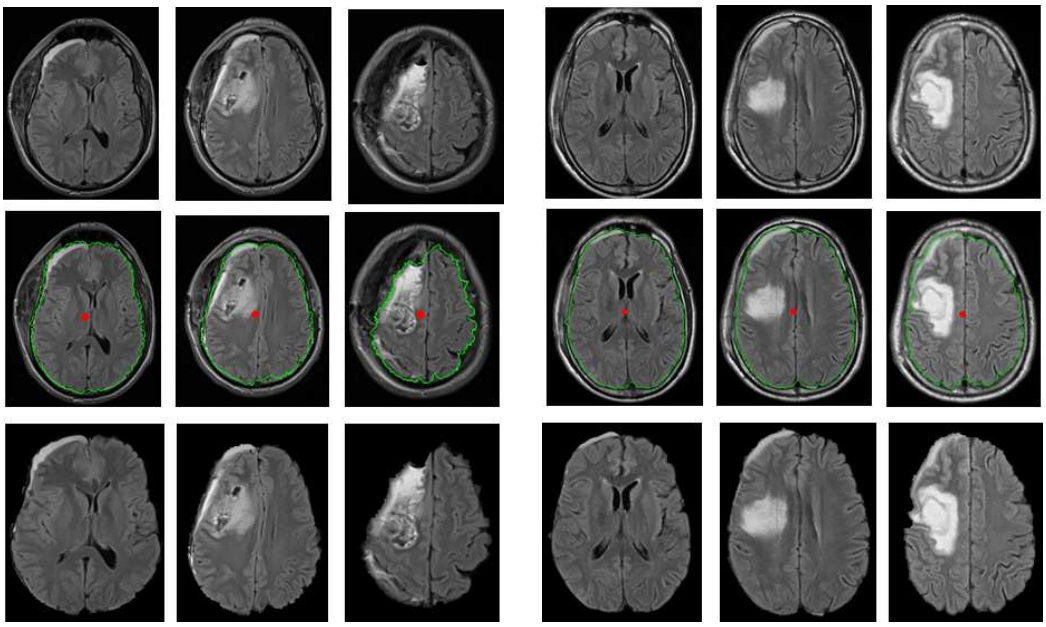
(b)

Figure 3.5: Example of 2D Brain Segmentation: a) different slices of the same patient b) different patients

proved.
The last section of the experimental phase regards the comparative analysis between our algorithm and a well known and more used segmentation algorithm: FEAT FSL Brain Extraction Tool (BET) [155] that uses a deformable model which evolves to fit the brains surface.

3.3.1 Qualitative Evaluation

Segmentation results shown in Figure 3.5 are obtained by processing five slices at different level from the same MRI volume (Figure 3.5(a)) and five slices from different cases (Figure 3.5(b)).
The team of expert performed a visual inspection of the resulting segmented images judging satisfactory the results obtained also in cases in witch tumors are very invasive. The segmentation procedure has a good behavior also when processing critical slices such as those including eyes or slices near to the top skull characterized by high texture and complex morphological patterns. In Figure 3.6(a) and 3.6(b) two examples of post surgery MRI are shown. In the first row of both Figures 3.6(a)(b) the original slices are displayed, in the second the boundary detected by the algorithm is superimposed on the original image and in the last row the brain extraction is shown. One of the goal of the post-operative MRI inspection is to identify any tumor residual. In this context, the whole brain segmentation, is used as a preprocessing phase for subsequent tumor



(a) (b)

Figure 3.6: Post surgery examples

detection, segmentation and analysis.
In the first case (Figure 3.6(a)), an uneven area can be seen at the site of intervention due to the surgical cavity in which cerebrospinal fluid is observed, together with traces of blood and air bubbles. It is visible a rim of perilesional edema (hyperintense) and into the deepest parts, in the vicinity of the lateral ventricle, the signal hyperintensity is consistent with a small tumor residual. Among the brain and the opercolo, which has been repositioned at the end of the intervention, there is a small flap hyperintense, consistent with a subdural flap, located in the meningeal subdural space between the brain and the skull, with a maximum thickness of 5 mm. Outside of the skull there is an important thickening of the subcutaneous tissues overlying the operculum, clearly uneven, due to edematous imbibition. In the outer part there are the stitches with a typical signal distortion. Again in Figure 3.6(b) an MRI after surgical removal of right frontal expansive mass is shown. Contrary to the previous case, here the operculum was removed to allow decompression of the brain around the lesion, hyperintense edema-gliotico halo is visible. Small right frontal subdural flaps with a maximum thickness of 3 mm. All these anomalies compared with the structure of a standards brain MRI, make the automatic brain segmentation an hard task. Moreover, the high variability between the cases highlights the need for a robust algorithm working without a-priori conditions to solve the segmentation problem. In both cases the segmentation algorithm results have been well evaluated by the experts. The presence or absence of the operculum, the

asymmetries and large deformations do not affect the performance.

3.3.2 Quantitative Evaluation

The quantitative analysis is conducted on the volumetric MRI. For each volume, the segmentation results, related to 10 slices evenly sampled, are evaluated in terms of both surface estimation and segmentation similarity. For each selected slice a ground truth mask has been manually identified and used in the evaluation. As regards the evaluation of the algorithm capability to estimate the surface, the error made is computed by $E = (S_{true} - S_{alg})/S_{true}$ where S_{true} is the mean of the surface obtained by the manual labeled masks produced by individual experts and S_{alg} is the surface of the region detected by the segmentation algorithm. As regards the segmentation similarity three performance indexes are used: the Jaccard index $J = \frac{T_p}{T_p+F_p+F_n}$ [81], the Precision $P = \frac{T_p}{T_p+F_p}$ and The Recall $R = \frac{T_p}{T_p+F_n}$ [124] where Tp are True Positive, Tf True Negative, Fp False Positive and Fn the False Negative. Table 5.1 reports results for both surface and similarity evaluation. Performances are highly satisfactory with mean value of Jaccard, Precision and Recall indexes equal to $\bar{J} = 91,43$,$\bar{P} = 93.36$,$\bar{R} = 97.78$ respectively. It is worth to note that there is a common trend through the slices in all four cases. Indeed, best results are obtained in central slices (35-110) with mean values $\bar{J} = 96.55$,$\bar{P} = 98.314$,$\bar{R} = 98.35$; results slightly decrease overestimating at eye level (Slices 5,20) with mean values $\bar{J} = 86.15$,$\bar{P} = 87.72$,$\bar{R} = 98.89$. The worst performances are found in top slices with mean values $\bar{J} = 80.98$,$\bar{P} = 84.15$,$\bar{R} = 95.53$. In the last column of Table 5.1 the expert variability, measured with Jaccard index, is reported making transparent the level of inconsistency under which the quantitative evaluation process has taken place.

3.3.3 Stability Evaluation

As mentioned above the identification of the center of the polar reference system is a critical aspect in the overall procedure. A sensitivity analysis is conducted with the aim of measuring how much the segmentation accuracy depend on this parameter. Varying the center of polar image different optimal feasible functions are obtained. Differently from the number of polar angle parameters whose setting simply affects the sampling rate, the variation of the center of polar reference system involves significant variation of the directional components of the polar image and then vertical gradient values that are in correspondence with the likelihood that an edge is present at that point. We base our analysis on Cases 1-8. For each volume three slices are selected: one from the eyes level, the other two from central and top slices. For each selected slice with an independent procedure we identify an ideal center. Under this condition we proceed in the brain segmentation identifying the boundary $B_{cent}(\theta)$ whose accuracy is qualitatively verified.

Table 3.5: 2D Brain Surface computed using the segmentation manually annotated by a set of medical experts (S_t), Brain Surface computed using the segmentation automatically obtained (S_e), Surface Estimation Error (E), Jaccard Index (J) Precision (P) and Recall (R) obtained when processing FLAIR MRI volumes of brain altered by glial tumor. Jaccard index obtained by compare manual segmentation performed by the expert ($IterJ$)

(a)

Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	11428.20	12111	5.97%	82.57%	88.44%	99.02%	97.58%
20	13721.18	13732	0.08%	96.67%	98.27%	98.35%	97.12%
35	14845.66	14920	0.50%	97.00%	98.23%	98.73%	97.23%
50	15917.66	15942	0.15%	97.81%	98.82%	98.97%	97.88%
65	15452.57	15548	0.62%	97.42%	98.39%	99.00%	97.32%
80	14809.10	14746	0.43%	96.44%	98.40%	97.98%	97.40%
95	13076.74	12910	1.28%	96.88%	99.05%	97.79%	96.88%
110	10979.51	10216	6.96%	91.74%	99.25%	92.39%	94.58%
125	7674.46	11632	51.56%	60.80%	62.76%	95.12%	91.39%
140	3797.59	6995	74.20%	54.51%	54.52%	99.95%	93.12%

(b)

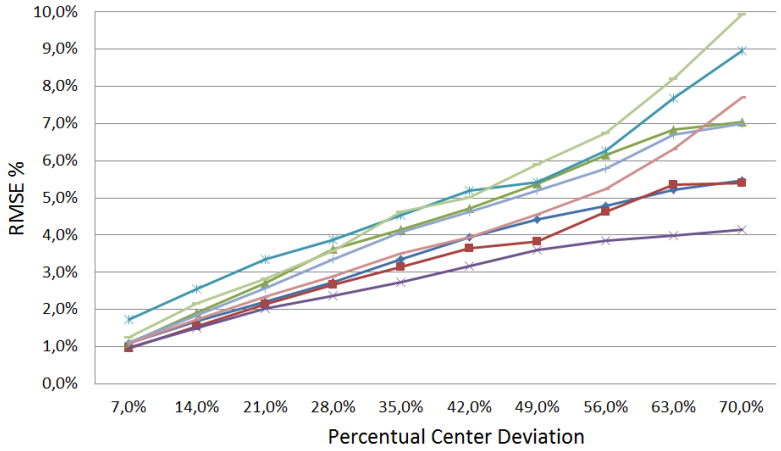
Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	10884.96	12897	18.48%	81.98%	83.07%	98.43%	95.44%
20	13653.11	13697	0.32%	97.94%	98.80%	99.12%	97.86%
35	15359.81	15357	0.02%	95.67%	97.80%	97.78%	97.77%
50	15342.59	15431	0.58%	97.76%	98.59%	99.15%	97.72%
65	15041.08	15150	0.72%	97.36%	98.31%	99.02%	97.86%
80	14107.81	14272	1.16%	97.40%	98.12%	99.26%	97.93%
95	12695.79	12769	0.57%	96.12%	97.74%	98.30%	97.23%
110	10580.53	10412	1.59%	95.69%	98.59%	97.02%	97.51%
125	7682.91	7760	1.00%	94.36%	96.61%	97.58%	96.09%
140	4808.36	5179	7.71%	68.66%	78.48%	84.59%	91.83%

(c)

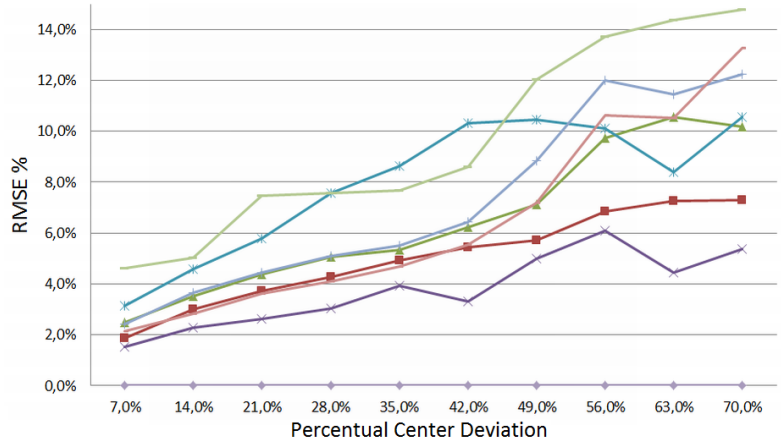
Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	6713.41	9860	46.87%	58.29%	58.34%	99.79%	86.37%
20	10634.79	11168	5.02%	91.50%	93.28%	97.96%	96.45%
35	13445.50	13221	1.67%	95.35%	98.45%	96.81%	96.76%
50	15323.42	15512	1.23%	97.85%	98.31%	99.52%	97.84%
65	15371.99	15517	0.94%	97.73%	98.39%	99.32%	97.83%
80	14901.21	14990	0.60%	97.92%	98.66%	99.25%	97.83%
95	13802.73	13949	1.06%	97.54%	98.24%	99.28%	98.00%
110	12233.95	12453	1.79%	96.54%	97.37%	99.12%	97.52%
125	10173.11	10079	0.92%	96.47%	98.65%	97.76%	95.84%
140	7307.00	7429	1.67%	94.20%	96.22%	97.82%	95.22%

(d)

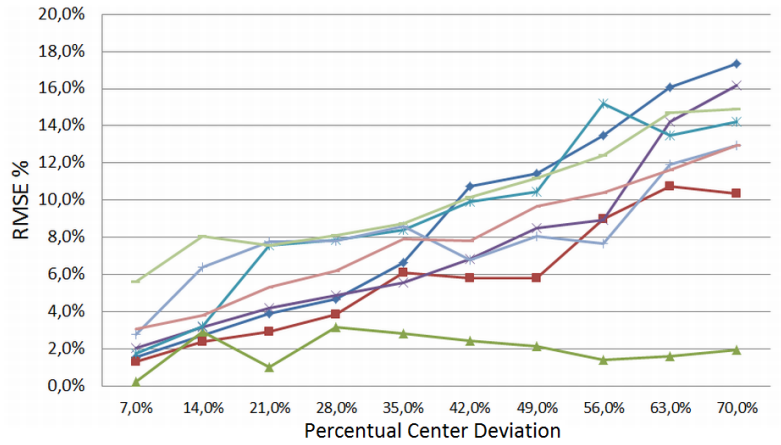
Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	10800.49	11847	9.68%	89.23%	90.15%	98.88%	96.88%
20	13682.35	14897	8.88%	91.03%	91.42%	99.54%	97.06%
35	15006.81	15213	1.38%	96.95%	97.78%	99.13%	97.60%
50	15705.18	15921	1.37%	97.54%	98.08%	99.43%	97.76%
65	15680.32	15848	1.07%	97.54%	98.23%	99.29%	99.76%
80	15141.48	15288	0.97%	97.69%	98.36%	99.31%	97.87%
95	14003.68	14014	0.07%	96.14%	98.00%	98.07%	96.57%
110	12281.54	11968	2.56%	93.99%	98.17%	95.67%	96.09%
125	9230.08	9211	0.21%	95.40%	97.75%	97.54%	95.50%
140	6046.55	6431	6.35%	83.41%	88.24%	93.86%	93.08%



(a)

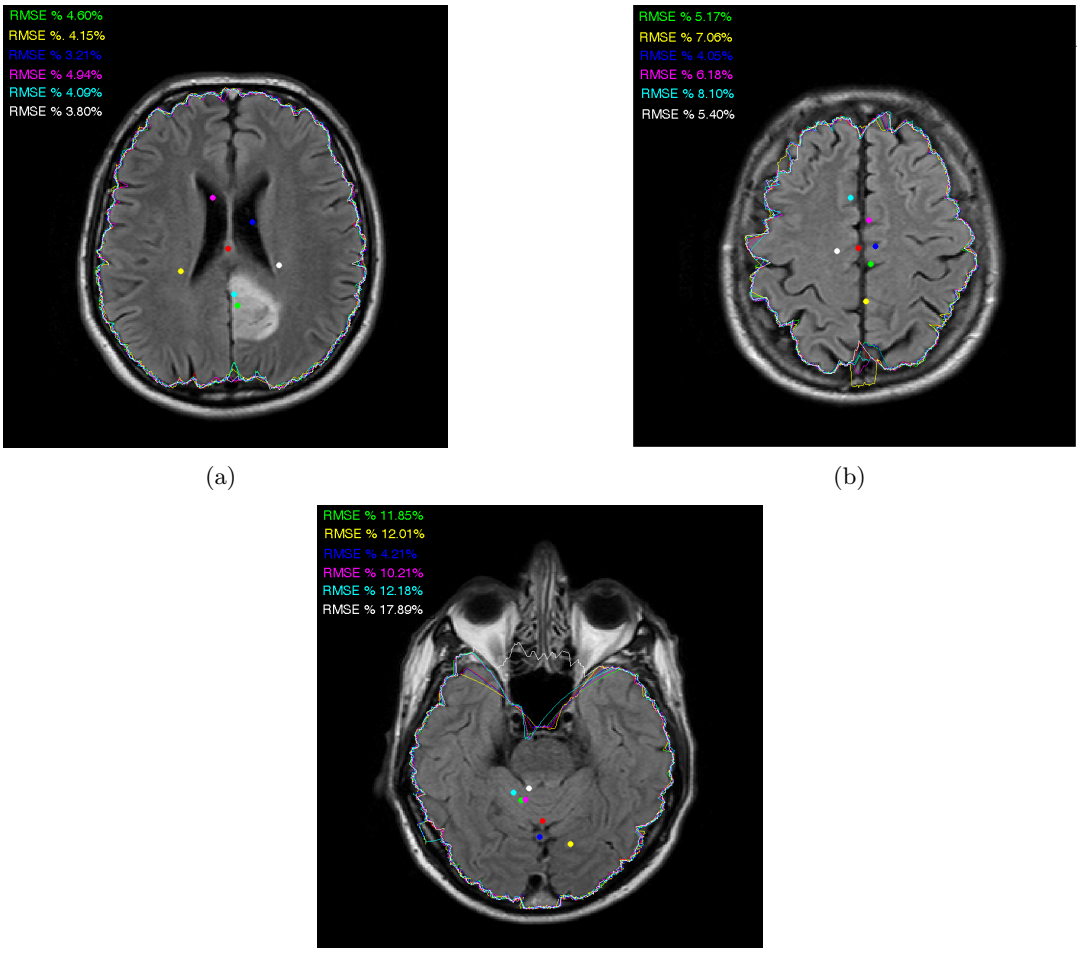


(b)



(c)

Figure 3.7: Stability analysis: trend of errors as a function of the center deviation a) central slice b) top slice c) eyes slice



(a) (b)

(c)

Figure 3.8: Stability analysis: qualitative examples a) central slice b) top slice c) eyes slice

We generate a set of new center coordinates moving randomly the centroid of N_p pixels. For each new generated center the boundary $B_{move,Np}(\theta)$ is identified and detection is evaluated as follows:

$$E_{N_p} = \frac{1}{N} \sum_{\theta=1}^{N} \frac{|B_{cent}(\theta) - B_{move,Np}(\theta)|}{B_{cent}(\theta)} \tag{3.2}$$

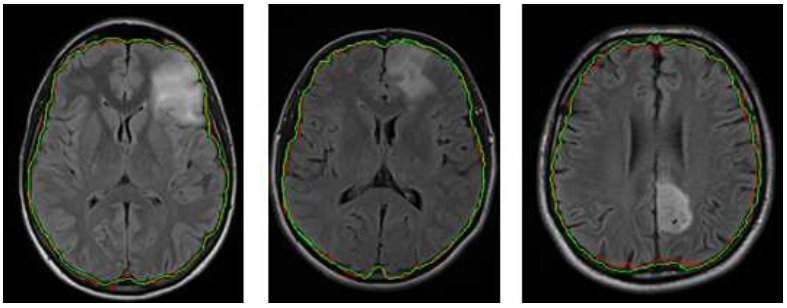
where N is the number of sampled angles. For each N_p the experiment is iterated for fifty trials and the results are averaged. The results obtained form central top and eyes level slices are shown in Figure 3.7(a)(b)(c) respectively. The values N_p are expressed as a percentage compared to the maximin radius of the ellipse where the brain is inscribed, to make the results invariant to image resolution. The relation between errors and distances between the center of the polar system approximates a liner law. The interpretation of these results is twofold: the behavior of the system is predictable and controllable and the variations of the parameter under study are not invasive. Moreover, the mean ratio between the percentage error and the percentage centroid deviation is equal to 0.11 for central slices, 0.17 for top slices and 0.21 for the eyes slices that show the higher instability.
To conduct a qualitative analysis in Figure 3.8 three visual examples are shown. The centers and the corresponding boundaries detected are superimposed on the MRI original slices. The visual and the above quantitative results analyzed tallied in general.

3.3.4 Eyes Detection Process Evaluation

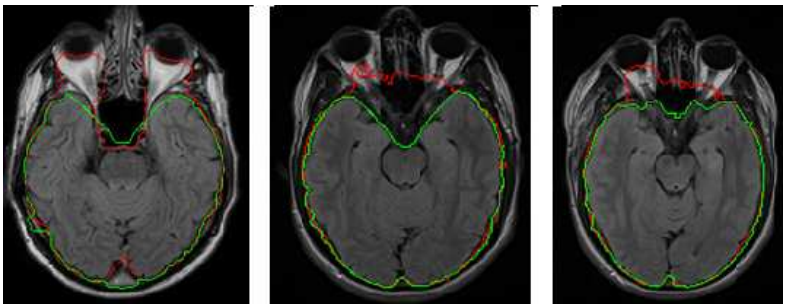
The eyes detection procedure identifying the initial slice SoI of the volume segmentation is evaluated by comparison with a manual detection using all the 14 Cases of the Dataset. The manual procedure is based on the identification in each volume of the first and last axial slices containing eyes; the slice lying in the middle between these two is considered as ground truth SoI_{gt} in the evaluation procedure. For each volume we compute the detection error $E = |SoI_{gt} - SoI|$ which is expressed in terms of number of slices; using the *Slice Thickness*, which is accessible in the DICOM header, it is converted into millimeters. Table 3.5 details the obtained results. The first 12 Cases show errors always smaller than 5 mm, under the vertical resolution of the voxel, except for case 1 in which the error is equal to the vertical voxel resolution.
As regards the volumetric images 11-14, in two cases (11,14) the error is less than the Slice Thickness (0.6 mm), for the case 12 and 13 the error is 1.5 and 3.9 mm respectively. These results can be reconducted to a slight variation of eyes diameter. In the last two column of the Table there are the EPI value of the detected slice and the mean value of the EPI signal computed for every slices respectively. Considering the results obtained we may conclude that the EPI index is robust considering the separability of the mean

Table 3.6: Eyes Detection Results

Case	Start Eyes	Stop Eyes	*SOI*	*E*	EPI	EPI Mean
1	6	10	9	5	13.84	6.79
2	5	10	7	2.5	14.40	7.55
3	7	10	9	2.5	10.82	7.63
4	4	9	7	2.5	10.05	6.75
5	3	8	5	2.5	8.58	5.37
6	5	9	7	0	11.09	6.78
7	6	9	8	2.5	10.98	6.95
8	6	9	7	2.5	12.24	6.77
9	5	10	8	2.5	11.89	6.66
10	5	8	7	2.5	11.14	6.97
11	209	242	225	0.3	14.88	7.01
12	190	227	215	3.9	15.12	7.30
13	175	208	189	1.5	15.32	7.81
14	200	235	218	0.3	15.69	7.31



(a)



(b)

Figure 3.9: Performance Comparison of our Algorithm Vs FSL BET: a) central slice, good overlapp b) slices at the eyeball level, bad overlapp

of EPI signal and the EPI value of the detected slice. The distribution separability was evaluated with a T-Test obtaining a confidence value equal to 99.96%

Table 3.7: Quantitative Comparison of our Algorithm Vs FSL BET: a) Jaccard b) Precision c) Recall

(a)

Slice	Jaccard 2D Graph	Jaccard FSL BET
5	78.02%	88.75%
20	94.29%	92.31%
35	96.24%	93.14%
50	97.74%	96.09%
65	97.51%	94.71%
80	97.36%	95.10%
95	96.67%	95.01%
110	94.49%	93.69%
125	86.76%	90.18%
140	75.20%	83.03%

(b)

Slice	Precision 2D Graph	Precision FSL BET
5	80.00%	88.89%
20	95.44%	92.42%
35	98.06%	93.26%
50	98.45%	96.30%
65	98.33%	94.84%
80	98.38%	95.80%
95	98.26%	95.22%
110	98.35%	95.39%
125	88.94%	94.32%
140	79.37%	90.66%

(c)

Slice	Recall 2D Graph	Recall FSL BET
5	99.03%	99.82%
20	98.74%	99.87%
35	98.11%	99.86%
50	99.27%	99.78%
65	99.15%	99.86%
80	98.95%	99.24%
95	98.36%	99.77%
110	96.05%	98.16%
125	97.00%	95.37%
140	94.06%	90.81%

3.3.5 Comparison Analysis

Our strategy has been compared with the widely used FEAT FSL Brain Extraction Tool(BET) [155]. The BET algorithm describes the brain surface starting from a sphere composed of vertices and triangles obtained by dividing iteratively the triangles of an icosahedron, each surface vertex is iteratively moved by a set of forces, who consider the homogeneity of the region inside the surface and the intensity of the transitions. The BET algorithm is here summarized as you can read in [155]. Firstly, the intensity histogram is processed to find 'robust' lower and upper intensity values for the image, and a rough brain/non-brain threshold. The centre-of-gravity of the head image is found, along with the rough size of the head in the image. Next a triangular tessellation of a sphere's surface is initialised inside the brain, and allowed to slowly deform, one vertex at a time, following forces that keep the surface well-spaced and smooth, whilst attempting to move towards the brain's edge. If a suitably clean solution is not arrived at then this process is re-run with a higher smoothness constraint. Finally, if required, the outer surface of the skull is estimated.

A qualitative comparison has been conducted using cases 1-8; cases 11-14 have been compared quantitatively. In Figure 3.9 some examples of the comparison are shown.

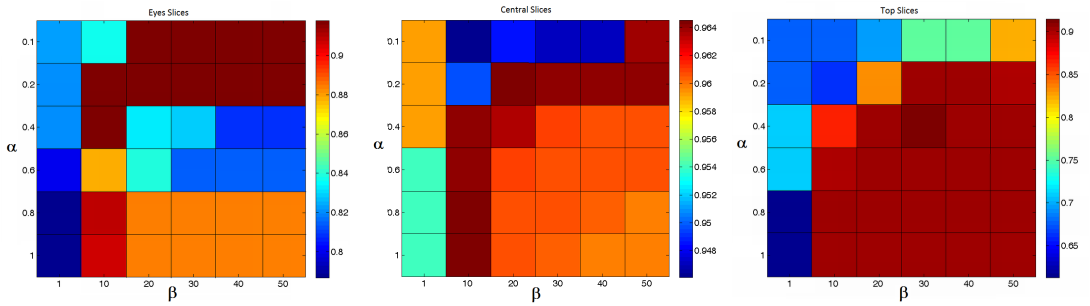
The brain boundaries are superimposed on six MRI slices. The red boundary is found by using the Brain Extraction Tool and the green one with our method. When processing the central slices, the two techniques have comparable behavior (Figure 3.9(a)); when processing eyes level slices (Figure 3.9(b)) our strategy definitely prevails. The results can be interpreted in the light of the fact that the embedding of both the local and global information in the graph search segmentation allows to identify the complete brain boundary even in a complex scene. Instead, the BET deformable model strategy is deceived by hight transition zone related to different anatomical parts. The ground truth masks and the evaluation metrics described above are used to develop the quantitative analysis. The analysis is focused on cases 11-14 that record scenes in which the transition between the tissue are shaded and the noise level is high. Results obtained for all cases considered (11-14) highlight a comparable behavior with slightly better performance of the BET when processing eyes and top level slices. As mentioned above the edges are shaded, thus the BET is not deceived by hight transition zone related to different anatomical structure. Moreover, it is worth to note that BET is a 3D algorithm that considers connectivity information in all directions. For this reasons it shoes advantages in cases in which the segmentation tasks are more difficult as at the eyes and top level. Table 3.7 shows the comparison analysis for an illustrative example related to one volumetric case

3.4 3D Brain Segmentation Evaluation

The dataset used for the 3D brain segmentation evaluation process is composed of volumetric brain MRI gray scale 12 bit depth volumes. All the MRI in the dataset are volumetric acquisition with isotropic voxel (from 0.57 to 1 mm). This high resolution makes the problem of the memory complexity, a critical aspect that makes significant the application of the strategy on this dataset. The dataset MRI are FLAIR, T1 and T2 weighted. Some dataset volumes are altered by the presence of glial tumors heterogeneous in terms of position, dimension, intensity and shape, others concern patients with cerebral atrophy.

3.4.1 Sensitivity Analysis

Preliminarily, a sensitivity analysis was performed on the MRI volumes, in order to evaluate how did the results depend on the strategy's main parameters (α, β). The results obtained from the varied set of experiments have been evaluated qualitatively by a group of experts. The quantitative evaluation process was performed by comparing the results produced by the automatic segmentation with those obtained by manual segmentation. In order to limit the effects of the well known phenomenon of inter-and



(a) (b) (c)

Figure 3.10: Jaccard indexes obtained by applying 36 configurations of the segmentation strategy distinguished by different α and β values when processing Eyes level Slices (a) Central Slices (b) Top Slices (c)

intra-observer variation, our strategy contemplates the organization of tuning sessions aimed at establishing a consensus among experts through discussion of the most controversial segmentation cases. The results obtained with the 3D heuristic graph based strategy are compared with the results obtained by applying slice-by-slice the 2D strategy. The last section of the experimental phase regards the comparative analysis between our algorithm and a well known and more used brain segmentation algorithm: Brain Extraction Tool (BET) [155] available in FMRIB Software Library (FSL).
The quantitative evaluation procedure is the same adopted for the 2D brain segmentation evaluation. For each volume, the segmentation results, related to 10 axial slices, evenly sampled, have been evaluated. For each selected slice 2 ground truth masks manually identified by experts are used in the evaluation. As regards the evaluation of the algorithm capability to estimate the surface, the error made is computed by $E = (S_t - S_e)/S_t$ where S_t is the mean of the surface obtained by the manual labeled masks produced by individual experts and S_e is the surface of the region detected by the segmentation algorithm. As regards the evaluation of the spatial distribution three performance indexes are used: the Jaccard index $J = \frac{T_p}{T_p+F_p+F_n}$ [81], the Precision $P = \frac{T_p}{T_p+F_p}$ and The Recall $R = \frac{T_p}{T_p+F_n}$ [124] where Tp are True Positive, Tf True Negative, Fp False Positive and Fn the False Negative. Experiments were performed by adopting the exponential weight function (Eq. 4) for the weight surface definition (Eq. 6) in the segmentation algorithm. Thirty-six configurations of the segmentation procedure were considered distinguished by different α and β values in ranges $[0-1]$ and $[1-50]$ respectively. The values considered for α are (0.1,0.2,0.4,0.6,0.8,1) and for β are (1,10,20,30,40,50). When α is equal to 0 the overall segmentation strategy performs precisely as the iteration of the 2D graph searching technique; otherwise when α is equal to 1 the likelihood that an edge is present at each point and the value of the weight surface at that point equally contribute.

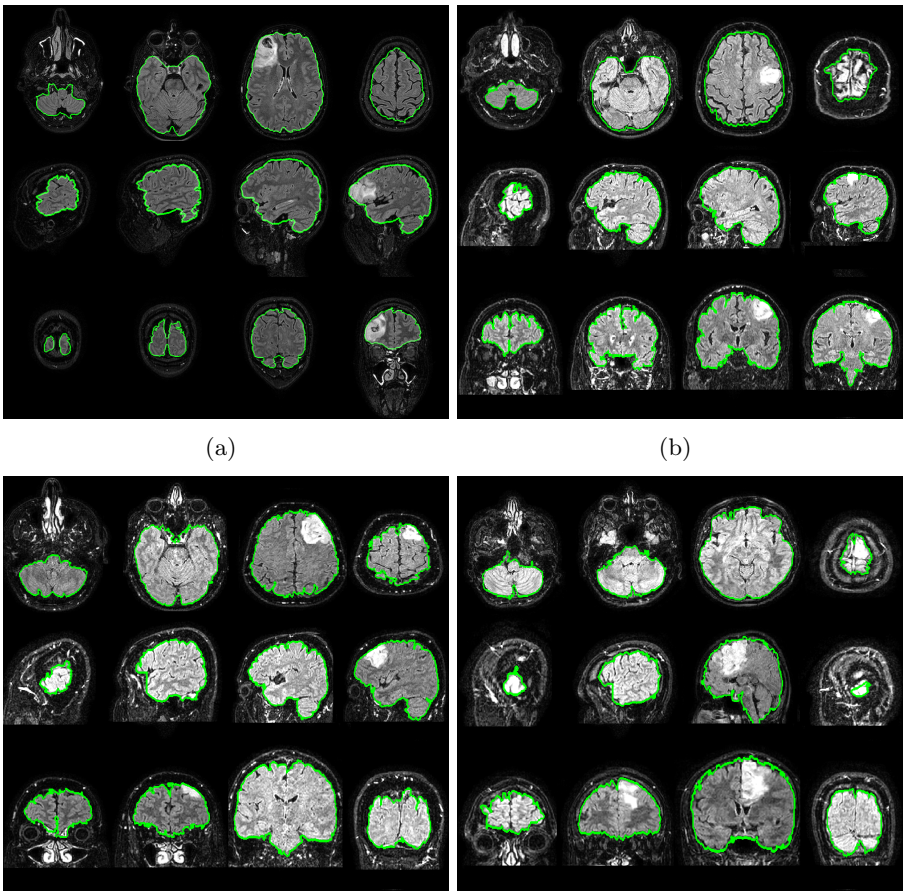
The results obtained in terms of Jaccard indexes are shown in Figures 3.10(a), 3.10(b), 3.10(c) when processing central top and eyes level slices respectively. The maximum of the performance at the eyes level slices is 92% at the point ($\alpha = 0.4, \beta = 10$) , the minimum is 79% at the point ($\alpha = 1, \beta = 1$) the average deviation is 4.5%. The maximum of the performance for central slices is 96% at the point ($\alpha = 0.8, \beta = 10$) , the minimum is 94% at the point ($\alpha = 0.1, \beta = 10$) the average deviation is 0.5%. The maximum of the performance for the top slices is 92% at the point ($\alpha = 0.4, \beta = 30$) , the minimum is 62% at the point ($\alpha = 1, \beta = 1$) the average deviation is 10.3%.
This analysis allows to conclude that the method has a good stability in function of the parameter variation while having considered ranges including unlikely values. A good setting can be achieved creating a balance between the effects of the two parameters whose values can be easily identified within reasonable more restricted ranges. In our experiments we assign the value 0.5 to the connectivity component α and value 10 to the exponential wight function β.

3.4.2 Qualitative Evaluation

Figure 3.11 shows results obtained when processing four FLAIR MRI volumes acquired with sagittal scans. The sequence is characterized by a long repetition times and by a reversal of the spin pulse at 180 degrees; the measurement is performed when the value of the liquid is close to 0 allowing the removal of the signal of liquid normally hyperintense in sequences with long repetition time. The way in which results are visualized reflects the habit of the medical expert in MRI inspection. In particular, four axial, sagittal and coronal sections, with superimposition of the brain boundary intended as the 2D section of the volume surface detected, are considered. The slices in Figure 3.11 shown the presence of Low grade glial tumor that alters the brain shape and intensity at varying degrees. The experts judgment confirmed that the results tallied in general even in situations in which the pathology significantly alters the brain anatomy. This is evident in some slices, for example first row third column and second row fourth column in Figure 3.11(a). Good results have been also obtained in situations characterized by strong concavities and irregular boundaries.
The visual inspection of the segmentation results obtained when processing T1 weighted volumes related to patients with diffuse axonal injury confirmed the validity of the proposed strategy which is able to accurately delineate atrophic brain (Figure 3.12).

3.4.3 Quantitative Evaluation

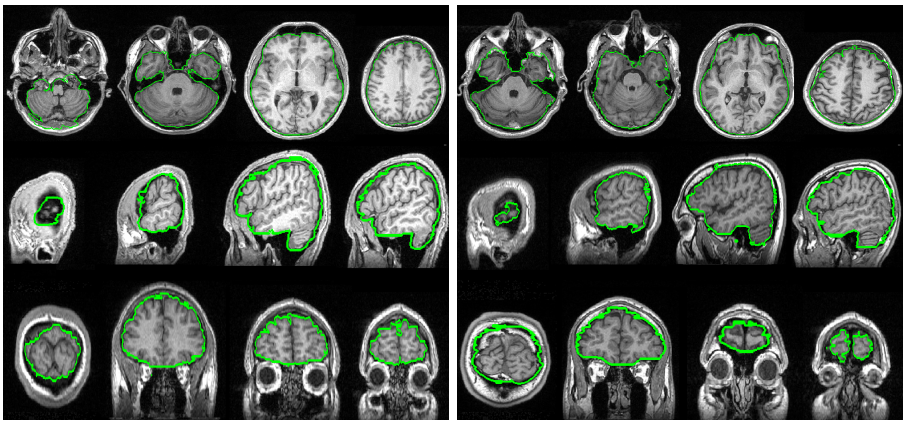
The quantitative analysis is conducted on four FLAIR MRI volumes. Results are evaluated in terms of both surface estimation and spatial distribution using the evaluation strategy detailed in section 4.1. Table 3.8 reports indexes obtained by averaging the in-



(a) (b)

(c) (d)

Figure 3.11: Segmentation results obtained by 3D graph heuristic propagation when processing FLAIR MRI volumes of brain altered by glial tumor



(a) (b)

Figure 3.12: Segmentation results obtained by 3D graph heuristic propagation when processing T1 MRI volumes of atrophic brain

dividual score of the two experts involved in the evaluation. In the last column of Table 3.8 the experts variability, measured with Jaccard index, is reported making transparent the level of inconsistency under which the quantitative evaluation process has taken place.

Comparing E and J indexes values we may conclude that the two evaluation criteria, surface estimation and spatial distribution are highly consistent. Performances are very satisfactory with mean value of Jaccard, Precision and Recall indexes equal to $\bar{J} = 95.31\%$, $\bar{P} = 96.94\%$, $\bar{R} = 98.24\%$ respectively. Moreover, the segmentation errors are uniformly distributed along the slices without trends of overestimation and underestimation.

Table 3.8: 3D Brain Surface computed using the segmentation manually annotated by a set of medical experts (S_t), Brain Surface computed using the segmentation automatically obtained (S_e), Surface Estimation Error (E), Jaccard Index (J) Precision (P) and Recall (R) obtained when processing FLAIR MRI volumes of brain altered by glial tumor. Jaccard index obtained by compare manual segmentation performed by the expert ($IterJ$)

(a)

Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	11428	11984	4.86%	93%	94%	99%	98%
20	13721	13562	1.16%	96%	99%	98%	97%
35	14846	14825	0.14%	97%	98%	98%	97%
50	15918	15898	0.13%	98%	99%	99%	98%
65	15453	15632	1.16%	97%	98%	99%	97%
80	14809	14917	0.73%	97%	98%	99%	97%
95	13077	13028	0.37%	97%	99%	98%	97%
110	10980	10681	2.72%	95%	99%	96%	95%
125	7674	7747	0.94%	93%	95%	96%	91%
140	3798	4020	5.86%	92%	93%	98%	93%

(b)

Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	10885	11255	3.40%	92%	94%	97%	95%
20	13653	13698	0.33%	98%	99%	99%	98%
35	15360	15151	1.36%	95%	98%	97%	98%
50	15343	15394	0.34%	98%	99%	99%	98%
65	15041	15154	0.75%	97%	98%	99%	98%
80	14108	14312	1.45%	97%	98%	99%	98%
95	12696	12743	0.37%	97%	98%	98%	97%
110	10581	10547	0.32%	96%	98%	98%	98%
125	7683	7914	3.01%	95%	96%	99%	96%
140	4808	5056	5.15%	92%	93%	98%	92%

(c)

Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	6713	7534	12.23%	85%	87%	97%	86%
20	10635	10311	3.04%	92%	97%	94%	97%
35	13445	13197	1.85%	95%	98%	97%	97%
50	15323	15478	1.01%	98%	98%	99%	98%
65	15372	15517	0.94%	98%	98%	99%	98%
80	14901	14993	0.61%	98%	99%	99%	98%
95	13803	13972	1.23%	97%	98%	99%	98%
110	12234	12380	1.19%	97%	98%	99%	98%
125	10173	10241	0.66%	96%	98%	99%	98%
140	7307	7629	4.41%	94%	95%	99%	95%

(d)

Slice.	S_t	S_e	E	J	P	R	$IterJ$
5	10800	11124	2.99%	94%	95%	98%	97%
20	13682	13541	1.04%	96%	98%	97%	97%
35	15007	14866	0.94%	95%	98%	97%	98%
50	15705	15753	0.31%	97%	98%	99%	98%
65	15680	15856	1.12%	98%	98%	99%	99%
80	15141	15301	1.06%	98%	98%	99%	98%
95	14004	14132	0.92%	97%	98%	99%	97%
110	12282	12064	1.78%	95%	98%	96%	96%
125	9230	9528	3.22%	95%	96%	99%	95%
140	6047	6762	11.83%	88%	89%	99%	93%

3.4.4 Comparison Analysis

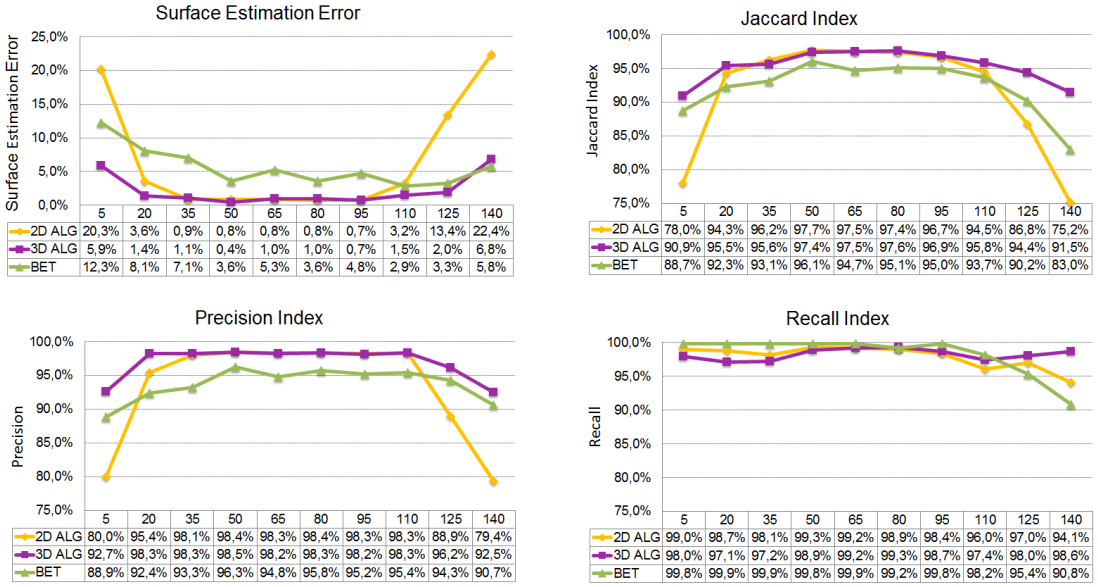
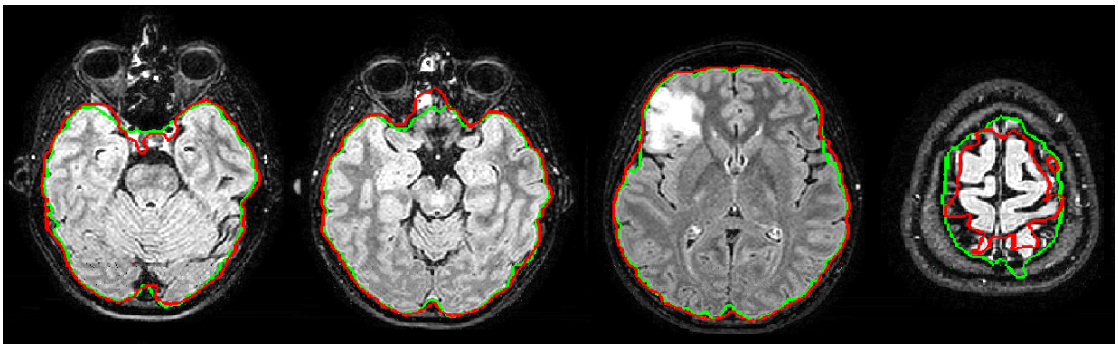


Figure 3.13: Comparison between Heuristic 3D Graph Propagation, 2D Graph Searching segmentation methods and Brain Extraction Tool(BET) computed in terms of a) Surface Estimation Error b) Jaccard Index c) Precision d) Recall for the slices from 5 to 140 evenly sampled every 15 slices.

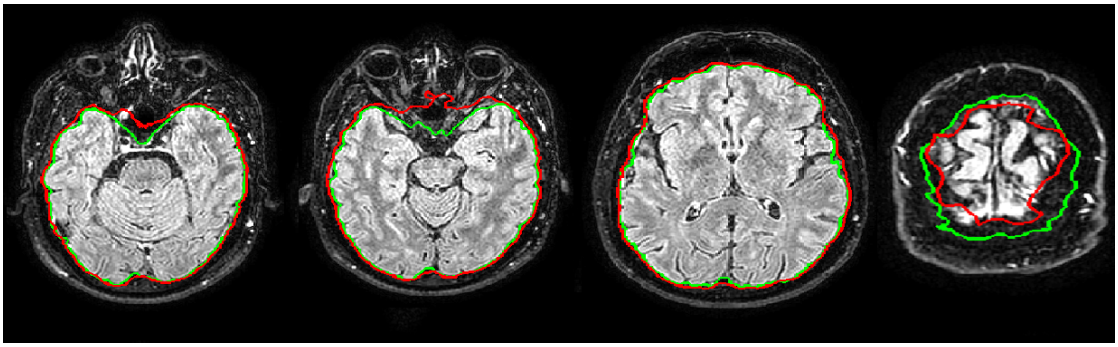
The aim of the comparison analysis is twofold:

- To quantify the contribution of the novel solutions with respect to the previous 2D graph searching technique.
- To see whether the 3D heuristic propagation strategy can be considered competitive with FEAT FSL Brain Extraction Tool(BET) [155], which include a deformable model strategy and it is a widely used tool in biomedical and clinical research.

Four MRI volumes were segmented by applying the slice-by-slice 2D graph searching method, the FSL BET and the heuristic 3D graph propagation strategy. Results in Figure 3.13 show that the 3D heuristic propagation strategy prevails for both surface and spatial distribution evaluation. Averaging for all the slices considered, we obtain $\bar{J} = 91.4\%$, $\bar{P} = 93.4\%$, $\bar{R} = 97.9\%$ for the 2D strategy, $\bar{J} = 92,2\%$,$\bar{P} = 93.7\%$,$\bar{R} = 98.3\%$ for the BET, $\bar{J} = 95.3\%$, $\bar{P} = 96.9\%$, $\bar{R} = 98.2\%$ for the propose 3D strategy. In more detail, comparing 2D and 3D strategies, it is worth to note that the performances are comparable in the central slices (35-110) where the average differences of the corresponding evaluation indexes J, P, R are equal to $J_{dif} = 0.1\%$,$P_{dif} = 0.1\%$,$R_{dif} = 0.1\%$ respectively. The 3D strategy works much better at the top and eyes level slices (5,20)



(a)



(b)

Figure 3.14: Segmentation results obtained by Heuristic 3D Graph Propagation (green) and FSL BET superimposed on axial slice (red) extracted by FLAIR MRI volumes

and (125,140) where the average differences of the evaluation indexes between the 2D and 3D strategy are equal to $J_{dif} = 7.1\%$, $P_{dif} = 7.8\%$, $R_{dif} = -1.4\%$ and $J_{dif} = 11.9\%$, $P_{dif} = 10.2\%$, $R_{dif} = 2.8\%$ respectively. The reason of this behavior lies in the fact that the local information embedded in the cost function for 2D boundary characterization in the current slice is ambiguous and a reliable detection can be accomplished only by exploiting the integration information of the spatial sequence of images.
Focusing on the comparison between BET and 3D strategy, the analysis highlights that performances are always slightly worst but degrade significantly in correspondence with the top slices where it shows a tendency to underestimate. This behavior is confirmed by the visual results shown in Figure 3.14 where it is possible to localize the errors and interpret their seriousness. The images in Figure 3.14 show the superimposition of the segmentation contour obtained by the 3D graph strategy and FSL BET. Looking into the details, the underestimation errors performed by FSL BET are highly visible in the fourth image of Figure 3.14(a) and Figure 3.14(b). The slight overestimation performed by the 3D graph-based method was judged preferable by the experts. The second images in Figure 3.14(a) and 3.14(b) refer to cases in which high performances have been

evaluated for both the segmentation procedures. However the few misclassified voxels are located in critical zones making the better results obtained by our proposed strategy highly valuable.

3.5 Brain Tumor Segmentation Evaluation

The segmentation performances are assessed both qualitatively, by a visual inspection of an expert group and quantitatively by comparing the automatic segmentation masks with the masks obtained by manual labeling.
Some of the cases in the dataset will be assessed through the proposed Fuzzy based reference estimation process.
To make our results comparable with those obtained by other authors, the last set of experiments will be conducted using a dataset published for the *MICCAI 2012 Multimodal Brain Tumor Segmentation Challenge (BraTS)*.

3.5.1 MRI Brain Tumor Dataset Collection

Table 3.9: Description of the acquisition parameters of the Data set for Brain Tumor Segmentation evaluation

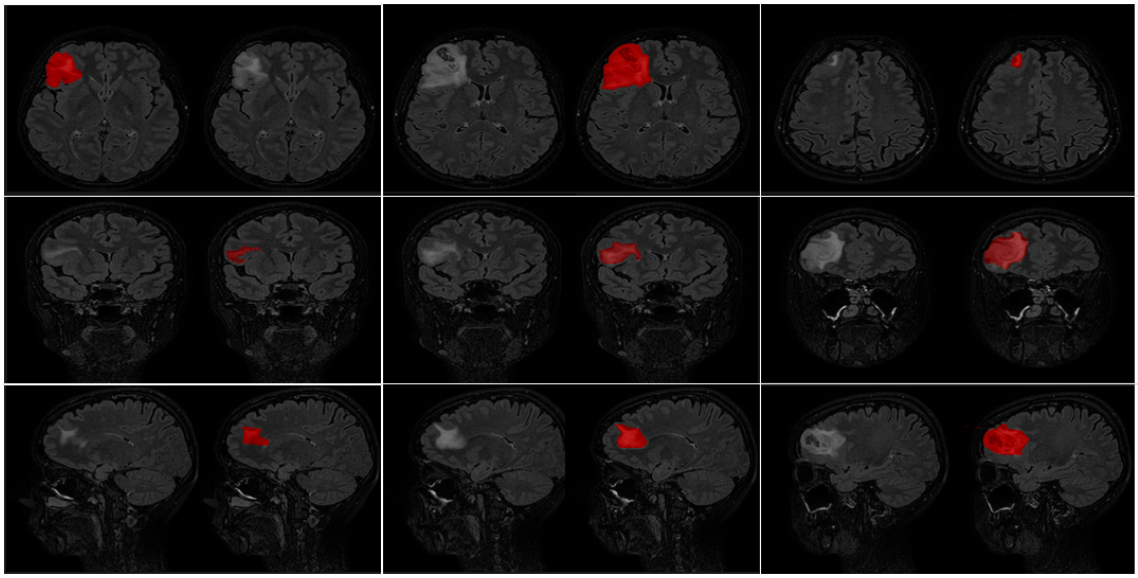
Case	Acquisition Mode	Volume Size (voxel)	Slice Thickness (mm)	Spacing Between Slice(mm)	Pixel Spacing (mm)	Repetition Time (ms)	ECHO Time (ms)
1	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	282.89
2	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	281.93
3	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	281.38
4	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	282.43
5	Sagittal FLAIR-T2	$[432 \times 432 \times 271]$	0.6	0.6	(0.57,0.57)	8000	280.93
6	Sagittal FLAIR-T2	$[432 \times 432 \times 300]$	0.6	0.6	(0.57,0.57)	8000	282.89
7	Assial FLAIR-T2	$[320 \times 320 \times 22]$	0.6	0.6	(0.75,0.75)	11000	140
8	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(0.94,0.94)	7800	95
9	Assial FLAIR-T2	$[256 \times 204 \times 20]$	5	6	(1.02,1.02)	7800	95
10	Assial FLAIR-T2	$[288 \times 288 \times 22]$	5	6	(0.79,0.79)	11000	140
11	Sagittal FLAIR-T2	$[432 \times 432 \times 280]$	0.6	0.6	(0.57,0.57)	8000	282.89
12	Sagittal FLAIR-T2	$[432 \times 432 \times 280]$	0.6	0.6	(0.57,0.57)	8000	281.33
13	Assial FLAIR-T2	$[288 \times 288 \times 22]$	5	6	(0.79,0.79)	11000	140
14	Assial FLAIR-T2	$[288 \times 288 \times 22]$	5	6	(0.79,0.79)	11000	140
15	Assial FLAIR-T2	$[288 \times 288 \times 22]$	5	6	(0.79,0.79)	11000	140
16	Assial FLAIR-T2	$[256 \times 204 \times 23]$	5	6	(0.94,0.94)	7800	95
17	Assial FLAIR-T2	$[320 \times 320 \times 24]$	5	6	(0.75,0.75)	11000	140
18	Assial FLAIR-T2	$[256 \times 204 \times 24]$	5	6	(0.94,0.94)	7800	95
19	Assial FLAIR-T2	$[288 \times 288 \times 22]$	5	6	(0.79,0.79)	11000	140
20	Assial FLAIR-T2	$[256 \times 204 \times 22]$	5	6	(1.02,1.02)	7800	95

The dataset used for the brain tumor segmentation evaluation process is composed of twenty FLAIR MRI, gray scale, twelve bit depth, volumes. In Table 3.9 all the acquisition parameter are detailed.
The dataset includes six "dense" and fourteen "sparse" MRI volume; the first ten cases

involve Low Grade Glioma tumors and the last ten High Grade Glioma, Butterfly and Multicentric. The tumor cases that are considered are heterogeneous in terms of shape position and intensity level, making this volumes set a good test set for the brain tumor segmentation evaluation. The heterogeneity of the cases makes our dataset also a good set for the training test and validation of supervised strategies.
This dataset should not be considered just a collection of MRI images but all the cases are widely discussed by a group of six medical experts and a detailed report was written for each case. In the following section, that concern the qualitative evaluation, all the information collected by the expert for every cases are detailed. Moreover, all the report and a visualization of the results are shown.
The task of describing the tumor highlighting the criticality in determining the tumor boundary has been evaluated very interesting by the expert team. They were led to reflect on the tasks they perform automatically based the judgement of their experience but not "knowing" the features that they use. This job was used by physicians to rationalize their choices and for the computer scientist developing team for eliciting the expert knowledge. Thanks to this close collaboration the algorithmic choices have been always substantiated by experts opinion by creating a system based on features actually used by physicians in the cognitive process of tumor diagnosis.
Four cases of the dataset (Case 1-4) have been manually fully annotated by the experts. Moreover the Case 3 and 4 are segmented 3 times changing the reference plane (Axial, Sagittal and Coronal) in order to relate the results of the automatic segmentation with both the inter- and intra- personal variability. To have an idea of the amount of work done by the experts we have to consider that every of these volumes consists of 432 axial images; moreover only the 30% of the slices includes the tumor and the mean annotation time of one slice is two minutes. Each expert worked for more than 4 hours on each case. These considerations should lead to emphasize the absolute necessity of automatic tumor brain segmentation methods; it is unthinkable to invest so much time to manage a single case in clinical practice. For assessing the adequacy of the number of cases included in our dataset, consider that the dataset that has had far greater use in filed of brain tumor segmentation is composed of ten cases related to patients with Meningiomas and Low Grade Gliomas. Only one central slice is manually segmented by four medical experts for each MRI volume. This dataset is distributed by the Surgical Planning Laboratory of the Harvard University [86]. In the last few years, the scientific community that deals with the brain tumor segmentation is working widely towards creating a standard benchmark dataset; recently, a dataset was published for the *MICCAI 2012 Multimodal Brain Tumor Segmentation Challenge (BraTS)*. Proceeding from this consideration, we believe that our work in collecting describing and fully annotating clinical cases is of great scientific value, independent of the algorithmic solutions proposed for the brain tumor segmentation.

3.5.2 Qualitative Evaluation

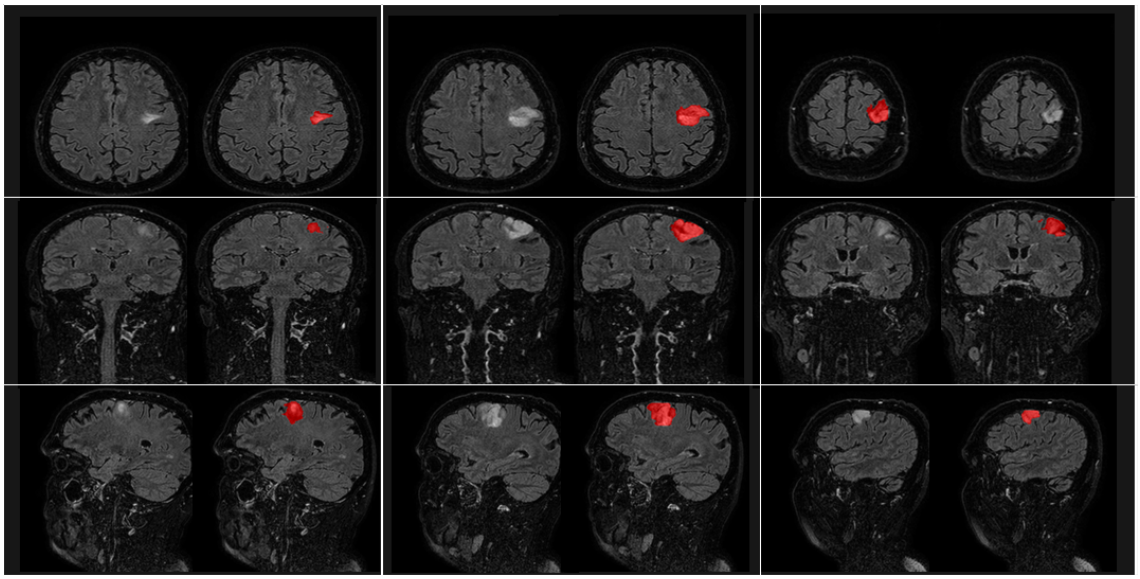
In this section the qualitative evaluation of the proposed tumor segmentation strategy is provided.
In an attempt to standardize the qualitative feeling of the experts we have decided to submit to them a rigid questionnaire. The structure of this module has been agreed with an interdisciplinary discussion, trying to highlight both the clinical and technical aspects. For each data set volumes the acquisition mode are specified; moreover, the tumor description in terms of type dimension position and all the specific characteristics which could affect the both the manual and automatic contours delineations capability are detailed. In particular, for each case, the specific problems that would face in the manual execution of the segmentation task that make the procedure arbitrary and not repeatable will be listed. The automatic segmentation performed with the proposed procedure will be evaluated qualitatively both in term of volume estimation and spatial distribution capabilities. For dense MRI volumes a comparative evaluation will be carried out on 3 planes Axial, Coronal and Sagittal. Specific comments are contemplated on the ability of the method to segment the perilesional edema.
In the follows all the qualitative analysis performed with the medical expert is summarized. For each MRI volumes (Cases 1-20) a Table that summarizes the qualitative judgment expressed by a medical expert is shown. For the cases involving dense volumetric MRI (Case 1-6,10,11) three original MRI slices for each plane (Axial,Sagittal and Coronal) and their superimposition with the segmentation masks are shown. For the remaining cases, involving sparse MRI (Case 7-9,12-20), only three Axial slices are presented.
In average the ratings are positive and there were no trends prevail in error (overestimation or underestimation). What emerges from this analysis is a valuable segmentation tool but difficult to assess because experts have great difficulty in defining with certainty the contours of the tumor. In the next section we try to quantify this feeling.



Case 1 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 282.89.
Image Quality Assessment:	Fair, too dark a window level used.
Tumor Description:	Presented as left frontal expansive lesion.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred peripheric edges in the white matter.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Generally good despite some punctual under- and overestimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	**Axial views** n 1 good, n 2 good, n 3 slightly overestimated, **Coronal views** n 1: good, n 2 good, n 3 superior lesion edge overestimated, underestimated the inferior one. **Sagittal views** n 1: extremely bad, jagged line too widely squared off, not following lesion edges; n 2: fair, underestimated, wrong evaluation of the superior side, n 3 good, minimal underestimation along the borders
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	axial plane: good coronal plane: fair sagittal plane: poor, low-quality segmentation
Comment on the perilesional edema segmentation:	good inclusion.
Need for multimodal confirmation	

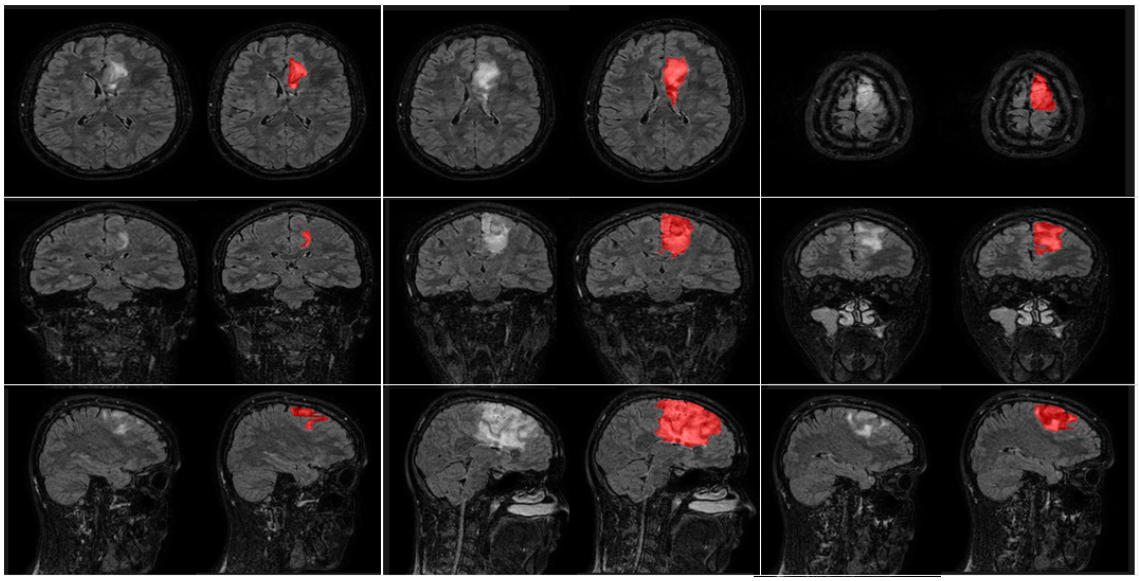
Case 1 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 2 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 281.93.
Image Quality Assessment:	Good quality images.
Tumor Description:	Left fronto-parietal neoplastic lesion close to the central sulcus.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred margins in the white cerebral matter, especially in the antero-lateral border of the lesion, toward the frontal lobe.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Overestimated in coronal plane number three good in the other planes.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	**Axial views** n 1 good, n 2 very good, n 3 good, slightly underestimated laterally next to the cerebral cortex; the problem may be caused by a slight underestimation in brain segmentation. **Coronal views** n 1 fair, underestimated superiorly next to the cortex, n 2 very good, n 3 very good with the exception of the inclusion of an unlinked median point in the lesion. **Sagittal views** n 1 good, underestimated supero-posteriorly to the side of the cortex, n2: very good, n3: very good.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	Sagittal plane may be described as the best, fair the coronal planes, axial views generally good.
Comment on the perilesional edema segmentation:	Good as it's included in most of the cases.
Need for multimodal confirmation	

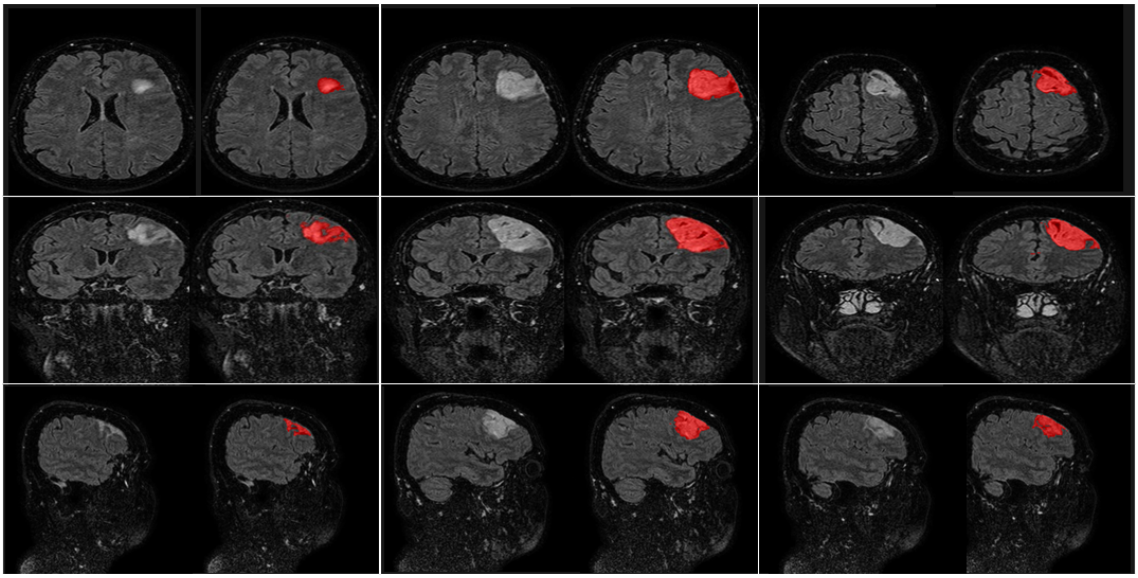
Case 2 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 3 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 281.38.
Image Quality Assessment:	Good quality images.
Tumor Description:	Left fronto-parietal expansive lesion adjacent and infiltrating the lateral wall of the left lateral ventricle.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred margins towards the lateral edges of the left omolateral ventricle.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Highly underestimated on sagittal view number one, quite precise on the other planes.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	**Axial views** n 1 good, slightly underestimated on the posterior side of the lesion; n 2 good, the same as n 1; n 3: very good. **Coronal views** n 1 fair, impossible to state if underestimated on the superior medial side because of the blurred typical characteristics, n2: very good, n 3: good, slightly underestimated medially and inferiorly. **Sagittal views** n 1 bad as it's highly underestimated, n 2: very good, n 3: very good.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	Axial and sagittal may be considered as the best ones, being equally distributed on estimation.
Comment on the perilesional edema segmentation:	Excluded on sagittal plane number one, as previously stated.
Need for multimodal confirmation	

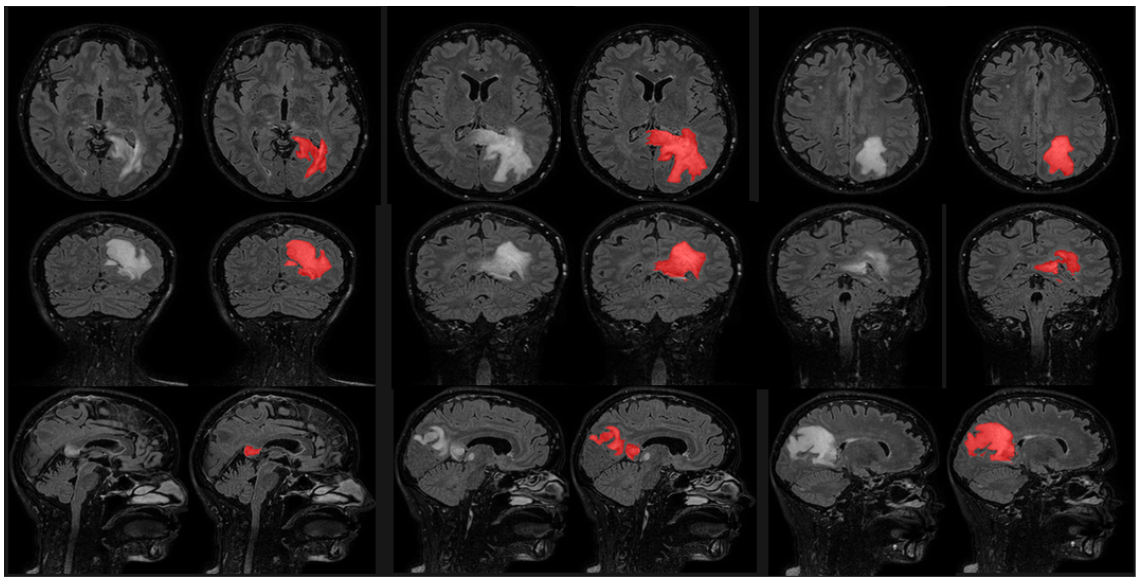
Case 3 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 4 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 282.43
Image Quality Assessment:	Substantially good, good contrast.
Tumor Description:	Right anterior frontal involving centrum semiovale.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Critical issues here are due to blurred edges; critical points are therefore located close to cortical sulci and gyri, that represents natural constraints to glioma extension, and the cortico-subcortical junction.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Generally correct estimation here. The visual inspection reveals no particular unclear area, since tumor boundaries in images seems to be well enhanced. Sometimes it's not easy to decide if an area that seems to be a sulcus has to be included into the segmented image or if it would be more correct cut that out by a line that passes around it.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	On the first axial there's a clear overestimation of the depicted lesion on its right side; a similar error happened on the third axial image (the tiny portion extending medially). The results in the coronal plane appear satisfactory in general except for some very little dots around the lesion in 1st and 3rd pair of slices. No critical aspects referring to the sagittal plane.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	Sagittal segmentation is the best.
Comment on the perilesional edema segmentation:	not much edema
Need for multimodal confirmation	

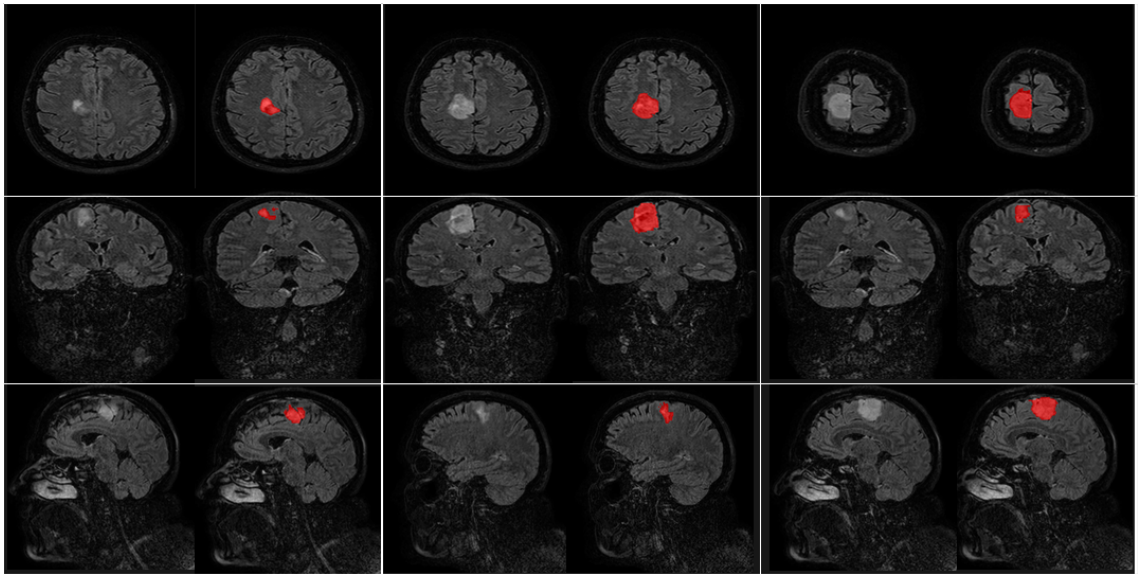
Case 4 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 5 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,271), Repetition Time 8000 ms, ECHO time 280.93.
Image Quality Assessment:	Good.
Tumor Description:	Tumor with homogeneous aspect in left occipito-parietal region, with perilesional edema and infiltration of corpus callosum.
Highlight the critical issues in manual segmentation. *The critical points are located close to: cortex, white matter, bone, mid sagittal plane*	The critical points are periventricular regions and the edges between white and gray matter regions.
General evaluation of the automatic segmentation: *capability of volume estimation (overestimates underestimates)*	The volume estimation is good; perhaps a little underestimation.
General evaluation of the automatic segmentation: *capability of spatial distribution estimation (detailed and punctual comments)*	Lack of detection of a pathological spot in the thalamus in the sagittal plane.
Assessment on the Axial Sagittal and Coronal Planes: *In which plane the tumor appear better/worse segmented?*	Best in the axial plane, worst in the coronal plane.
Comment on the perilesional edema segmentation:	Description of edema is good, little underestimated in the contralateral infiltration along corpus callosum.
Need for multimodal confirmation	No, regarding the general segmentation of the pathology (both tumor and edema) and because of the low grade nature of the tumor.

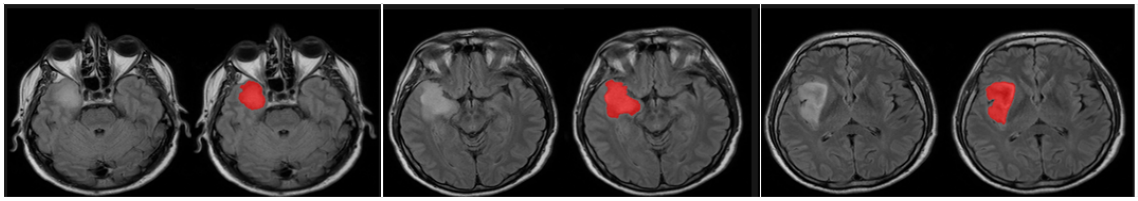
Case 5 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 6 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 282.89.
Image Quality Assessment:	Generally good, good contrast and texture, field inhomogeneity on middle sagittal plane.
Tumor Description:	left mesial fronto-parietal lesion extending downside towards the corpus callosum and delimitated by the falx cerebri medially.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurring of edges make difficult, particularly in some slice to trace a clear cut to delineate the lesion from the normal tissue. Critical points are located close to the gray-white matter junction.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Globally the segmentation task has been solved correctly, with no clear tendency to under- or overestimation, even if there are some cases (1st and 3rd coronal pair of slice) where respectively an underestimation and an overestimation occurred, in such a way we cannot accept it.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	In first axial image, due to the softness of lesion edges there's a clear overestimation; in 3rd axial there are two little flaps that represent a clear error, while the medial side is precise for the presence of a structural limit to glioma extension. First coronal: great underestimation error; third coronal :overestimation error that could be accepted looking at the other planes for the same points that should be excluded here by simply inspecting.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	Best: sagittal, Worst: coronal.
Comment on the perilesional edema segmentation:	Few edema.
Need for multimodal confirmation	

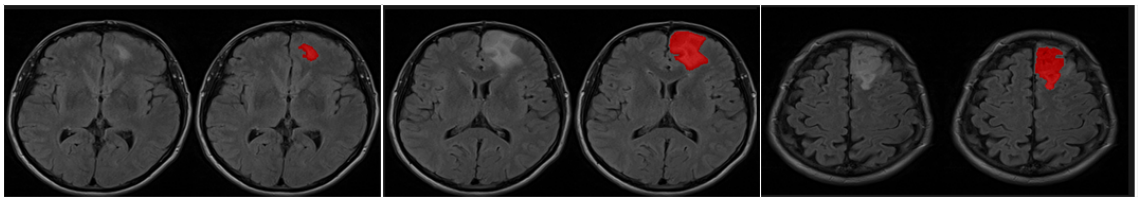
Case 6 - qualitative evaluation of the segmentation results provided by a medical expert



Case 7 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0,75, 0.75, 5) mm, volumetric matrix size (320,320,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Good quality images.
Tumor Description:	Fronto-temporal deep cerebral lesion extended to insula and partially to limbic lobe.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Quite blurred borders in many places all around the lesion between the neoplasm and white matter.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Good even if slightly underestimated.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	Axial plane n 1 good, n 2 good but medially underestimated next to the basal cisterns, n 3 good even if the posterior part of the lesion was not detected.
Comment on the perilesional edema segmentation:	Fair; evaluation loss of some tumor parts

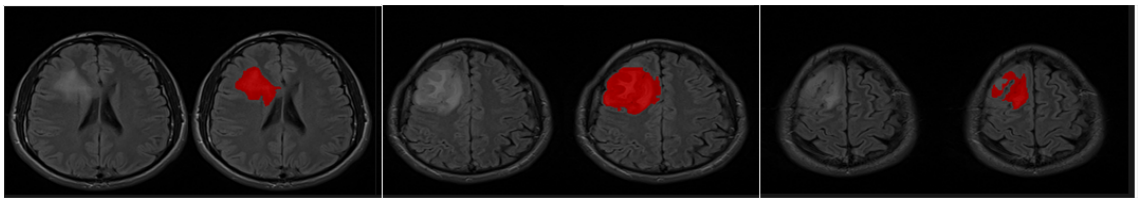
Case 7 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 8 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,300), Repetition Time 8000 ms, ECHO time 282.89.
Image Quality Assessment:	Good
Tumor Description:	Left mesial anterior frontal
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred edges. Another critical aspect for manual segmentation is the absence of coronal and sagittal plane sequences.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Globally accurate.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	For the first axial I agree with software results but here is really critical to cut the image without referring to other planes. The third axial appear underestimated in the portion close to the cortex.
Comment on the perilesional edema segmentation:	Correct inclusion

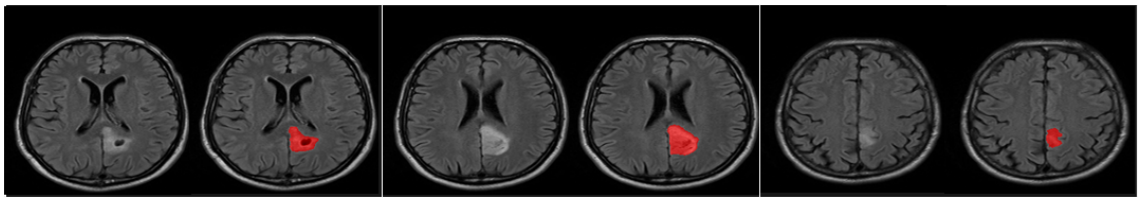
Case 8 - qualitative evaluation of the segmentation results provided by a medical expert



Case 9 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (1.02,1.02,5) mm, volumetric matrix size (256,204,20), Repetition Time 7800 ms, ECHO time 95.
Image Quality Assessment:	Fair.
Tumor Description:	Left frontal Low Grade Glioma.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Inhomogeneity of the lesion, edema fading lesion boundaries.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	1st and 2nd are similar to what expert would do by manually segmenting. The 3rd segmentation represent a critical problem.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	1 Image is ok, In 2 Image overestimation on the medial part; 3 concave aspect, but a convex shape lesion in this case, would be more adherent to real.
Comment on the perilesional edema segmentation:	Correct inclusion

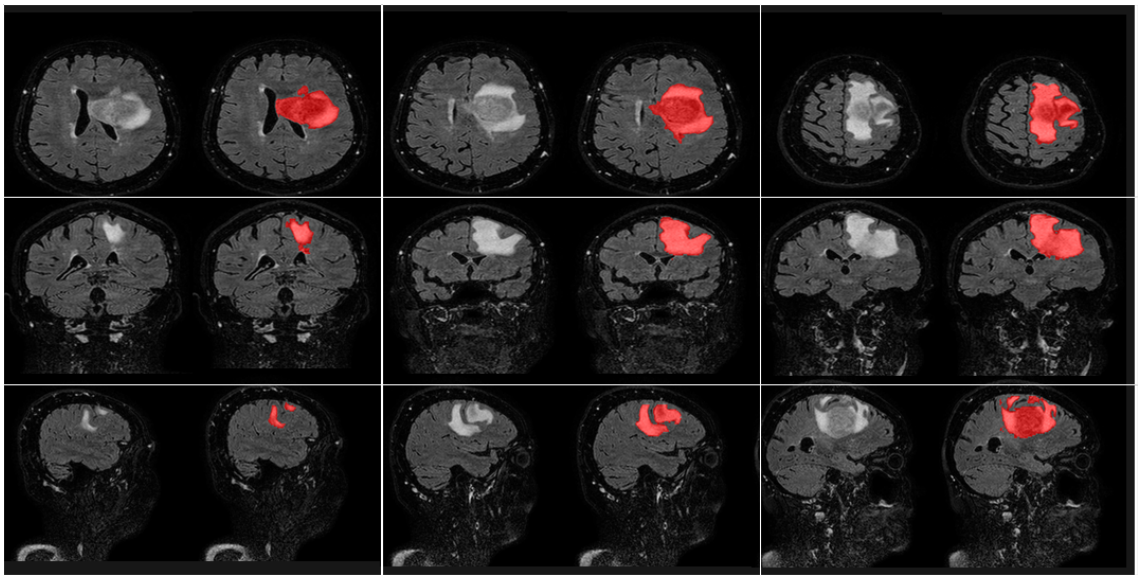
Case 9 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 10 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.79,0.79,5) mm, volumetric matrix size (288,288,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Good.
Tumor Description:	Left Occipital parasagittal tumor with cystic component and mild perilesional edema.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	No critical issue in this case because of the well delineated margins of the tumor.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Minimal underestimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	In the third image the underestimation is due to the presence of little sulci and gray matter.
Comment on the perilesional edema segmentation:	The edema is not enough to give an evaluation

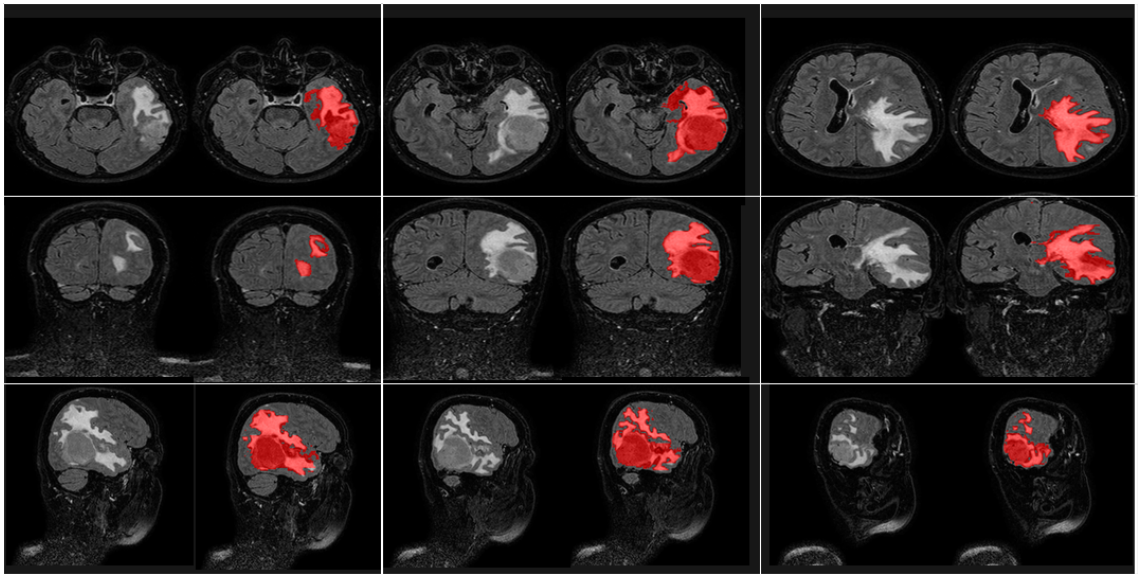
Case 10 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 11 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0,6 0.57 0,57) mm, volumetric matrix size (432,432,280), Repetition Time 8000 ms, ECHO time 282.89.
Image Quality Assessment:	Good quality images.
Tumor Description:	Left frontal expansive cerebral lesion involving the corpus callosum and initial crossing of the median line, clear mass effect on homolateral cerebral ventricle.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Doubtful determination of cortex involvement.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	very good.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	very good result with the exception of only one point in coronal plane n1, where the evaluation was overestimated: some of leukoaraiosis abnormalities are considered neoplasm.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	axial plane: very good; sagittal plane: very good; coronal plane: good.
Comment on the perilesional edema segmentation:	good
Need for multimodal confirmation	

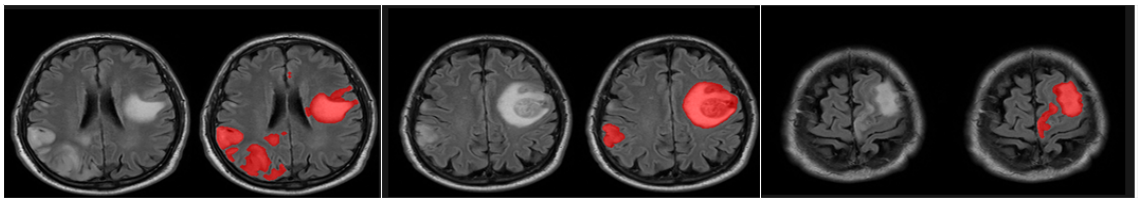
Case 11 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 12 - Superimposion of the tumor mask automatically computed on three Axial, Coronal and Sagittal slices.

Acquisition Modality:	Dense Volumetric MRI, Sagittal FALIR T2, voxel dimension (0.6 0.57 0,57) mm, volumetric matrix size (432,432,280), Repetition Time 8000 ms, ECHO time 282.89.
Image Quality Assessment:	Excellent.
Tumor Description:	Tumor with cistico necrotic component and severe perilesional edema in the left posterior temporal region.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	In the coronal plane manual segmentation is difficult in the peripheral digitation of brain edema (first image of coronal plane).
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Excellent mainly for the extreme complexity of the edema infiltration; in general a little overestimation (see below).
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	In the axial plane, in the medial areas of temporal lobe (uncus) there is an overestimation of pathology.
Assessment on the Axial Sagittal and Coronal Planes: ***In which plane the tumor appear better/worse segmented?***	Very good in the sagittal plane; worst in the axial plane (see above).
Comment on the perilesional edema segmentation:	Very good.
Need for multimodal confirmation	No regarding the segmentation of the whole pathology, T1+ Gadolinium regarding tumor detection and necrosis.

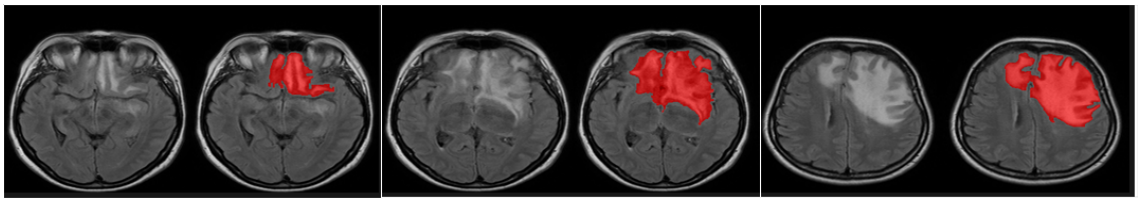
Case 12 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 13 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.79,0.79,5) mm, volumetric matrix size (288,288,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Very good.
Tumor Description:	Multicentric tumor with a solid and well-defined component in the left frontal region (coronaric region) and a contralateral part (cortical and subcortical) in right parietal region.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	The multicentric nature of the lesion and the proximity with cortical sulci at vertex are the main difficulties of the manual segmentation.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Moderate underestimation because some parts of edema are not included in the segmentation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	In the image on the right segmentation of pathological sulci is underestimated and bad.
Comment on the perilesional edema segmentation:	Edema segmentation is not good because the lack of segmentation at vertex (see above)

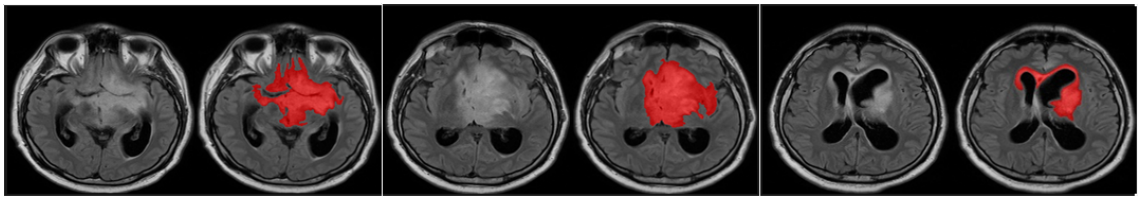
Case 13 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 14 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.79,0.79,5) mm, volumetric matrix size (288,288,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Good quality images.
Tumor Description:	Large infiltrating bilateral fronto-temporal lesion, whose edges are adjacent expecially to left thalamus, causing displacement of the left lateral ventricle.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred edges with the white matter.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Quite precise evaluation in all of the choosen images. bias due to the absence of sagittal and coronal planes.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	Axial n 1: good, impossible to determine if underestimated medially to the right orbit axial n 2: very good axial n 3: very good.
Comment on the perilesional edema segmentation:	Generally good

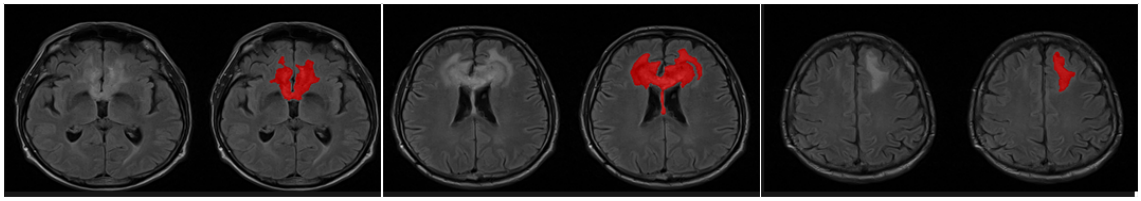
Case 14 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 15 - Superimposion of the tumor mask automatically computed on three slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.79,0.79,5) mm, volumetric matrix size (288,288,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Good.
Tumor Description:	Bilateral fronto-basal infiltrative expansive lesion developing towards the basal nuclei on the left side and towards the brainstem inferiorly.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Determination of lesion edges and its extension in brain cortex.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Modest overestimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	Some of the probably uninvolved brain cortex has been detected as part of the lesion. difficult to discern between neoplasm and periventricular tissue alterations in the third image.
Comment on the perilesional edema segmentation:	Fair.

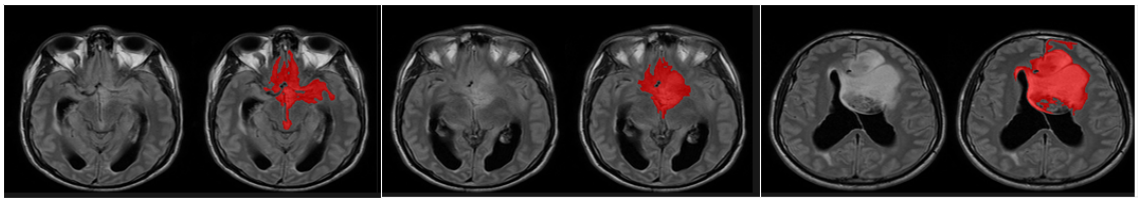
Case 15 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 16 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.94,0.94,5) mm, volumetric matrix size (256,204,23), Repetition Time 7800 ms, ECHO time 95.
Image Quality Assessment:	Good.
Tumor Description:	Infiltrative lesion with infiltration of the anterior part of the corpus callosum (rostrum): Bilateral infiltration of the frontal lobes and perilesional edema.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	The critical points are corpus callosum near the ventricles.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	A mild underestimation because of the not precise segmentation of brain edema.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	In the second and third image segmentation of the lateral part of brain edema is not precise.
Comment on the perilesional edema segmentation:	See above

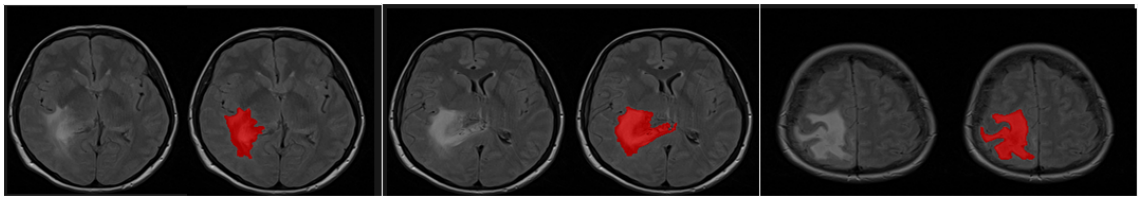
Case 16 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 17 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0,75,0.75,5) mm, volumetric matrix size (320,320,24), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Very good quality images.
Tumor Description:	Bilateral fronto-temporal expansive lesion whose medial side determines infiltration and controlateral dislocation of the left lateral ventricle.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred edges especially on the median line.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Generally very good estimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	Axial image n1: very good axial image n2: very good axial image n3: fair, as it includes a portion of frontal anterior cortex, apparently not infiltrated by the tumor. in addition it excludes some of the posterior medial part of the lesion.
Comment on the perilesional edema segmentation:	Good segmentation.

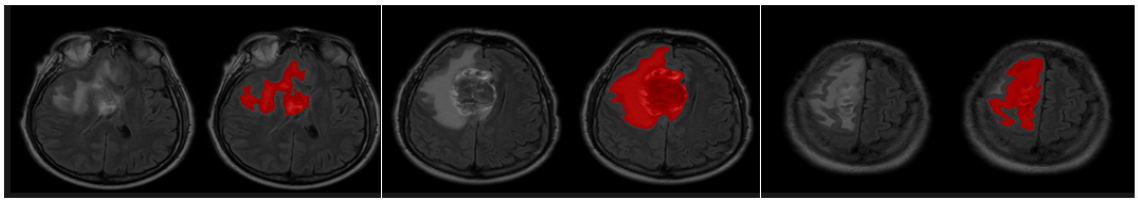
Case 17 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 18 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.94,0.94,5) mm, volumetric matrix size (256,204,24), Repetition Time 7800 ms, ECHO time 95.
Image Quality Assessment:	Fair, good contrast.
Tumor Description:	Deep parietal lesion involving basal ganglia.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Blurred edges, critical points are located near the junction between deep gray matter and deep white matter; another critical point is the presence of edema.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Tendency to a global overestimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	1 little overestimation, 2 overestimation (difficult to idealize ground truth), 3 correct (edema is not separable from the lesion).
Comment on the perilesional edema segmentation:	Correctly included.

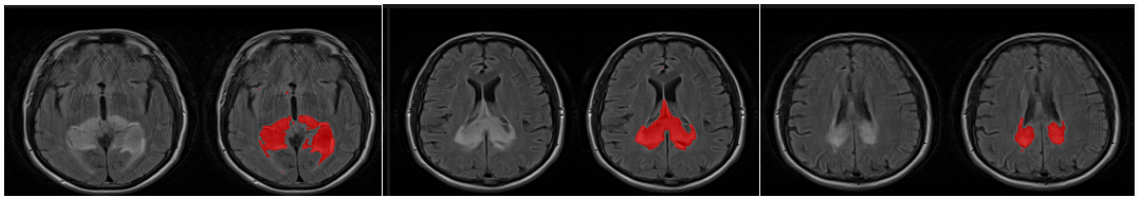
Case 18 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 19 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (0.79,0.79,5) mm, volumetric matrix size (288,288,22), Repetition Time 11000 ms, ECHO time 140.
Image Quality Assessment:	Good Images Quality.
Tumor Description:	frontal expansive lesion; strong mass effect and contralateral overrun by corpus callosum and anterior commissure.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Expansive lesion detection in lateral cerebral ventricles proximity. Only axial images, other plane views would be considered useful for segmentation
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	underestimated in the first image correct in the others.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	Good recognition in particular of perilesional edema; scarce identification of isointense (compared to parenchima) neoplastic tissue in the first image.
Comment on the perilesional edema segmentation:	Very good.

Case 19 - Qualitative evaluation of the segmentation results provided by a medical expert



Case 20 - Superimposion of the tumor mask automatically computed on three Axial slices.

Acquisition Modality:	Sparse MRI, Axial FALIR T2, voxel dimension (1.02,1.02,5) mm, volumetric matrix size (256,204,22), Repetition Time 7800 ms, ECHO time 95.
Image Quality Assessment:	bad quality images (blurred because of patients movements during MRI data acquisition).
Tumor Description:	Bilateral parieto-occipital expansive lesion, whose anterior edges surround the postero-lateral walls of the cerebral ventricles.
Highlight the critical issues in manual segmentation. ***The critical points are located close to: cortex, white matter, bone, mid sagittal plane***	Critical points are located close to lateral ventricles and midline with inclusion of connection interemispheric structures.
General evaluation of the automatic segmentation: ***capability of volume estimation (overestimates underestimates)***	Fair, due to general underestimation.
General evaluation of the automatic segmentation: ***capability of spatial distribution estimation (detailed and punctual comments)***	1: underestimation of posterior edema branches in both the occipital lobes, 2: good estimation, axial n3: insufficient estimation of the portions surrounding the lateral ventricle walls.
Comment on the perilesional edema segmentation:	Fair, as it was underestimated in one image

Case 20 - Qualitative evaluation of the segmentation results provided by a medical expert

Table 3.10: Assessment of the Volume Estimation Error(VEE) of the automatic brain tumor segmentation

VEE	Case 1	Case 2	Case 3			Case 4		
			Axial	Coronal	Sagittal	Axial	Coronal	Sagittal
EXP 1	7.84%	9.43%	0.04%	7.09%	6.29%	9.97%	4.55%	2.75%
EXP 2	5.30%	0.99%	8.47%	21.94%	3.50%	1.79%	14.77%	1.96%
EXP 3	2.86%	0.14%	14.82%	18.92%	15.54%	3.27%	5.85%	5.92%
EXP 4	6.41%	0.99%	8.77%	17.74%	21.82%	2.82%	17.58%	19.67%
EXP 5	1.05%	0.27%	3.49%	4.04%	4.27%	7.57%	5.82%	2.92%
EXP 6	3.39%	5.79%	0.46%	4.93%	3.64%	10.52%	11.93%	3.42%
AUTOMATIC	**0.56%**	**3.22%**	**0.65%**	**4.69%**	**4.13%**	**0.24%**	**4.19%**	**2.71%**
Mean	*3.92%*	*2.97%*	*5.24%*	*11.34%*	*8.45%*	*5.17%*	*9.24%*	*5.62%*
Var	0.07%	0,12%	0,31%	0,61%	0,53%	0,17%	0,30%	0,40%
Max	*7.84%*	*9.43%*	*14.82%*	*21.94%*	*21.82%*	*10.52%*	*17.58%*	*19.67%*
Min	*0.56%*	*0.14%*	*0.04%*	*4.04%*	*3.50%*	*0.24%*	*4.19%*	*1.96%*

3.5.3 Quantitative Evaluation

The aim of this phase is to quantitatively assess the capability of our segmentation strategy to estimate tumor volume and spatial distribution. Three different sets of experiments were conducted. The first has the aim of comparing manual segmentations provided by medical experts and those obtained automatically, and investigating how the agreement and disagreement observed lie in the inter- experts variability statistics, taking into account that a unique ground truth is not available. The second set of experiments uses the fuzzy estimated reference data. In this regard we also introduce some metrics that allow to compare fuzzy and crisp masks. The last part of the this experimental phase was conducted using a benchmark dataset (*BraTS*) composed of 30 real dense MRI volume (20 Low Grade Glioma and 10 High grade Glioma) and 50 simulated MRI volume (25 Low Grade Glioma and 25 High grade Glioma).

Segmentation Evaluation using manually labeled reference data

Table 3.10 shows the percentage volume estimation errors considering, as a reference volume, the mean of the volumes estimated by every experts. The Table 3.10 also shows the errors in estimating the volume committed by each expert and same statistical values for summary.

For the Cases 1, 3 and 4 the volume estimation error of the automatic segmentation is lower than the average of the errors committed by the experts; only in the second case the error is slightly greater but still contained within the Gaussian described by the mean and variance of the errors distribution. Moreover, in Cases 1,3 and 4 even the automatic segmentation volume estimation errors are out of the statistics: they are outlier for their goodness. It is worth to note that our segmentation is fully volumetric, then spatial continuity constraints are respected in all directions which can not ever be guaranteed by the manual segmentation produced by experts.

In Figure 3.15 the trends of the tumor areas computed slice by slice by the 2D masks

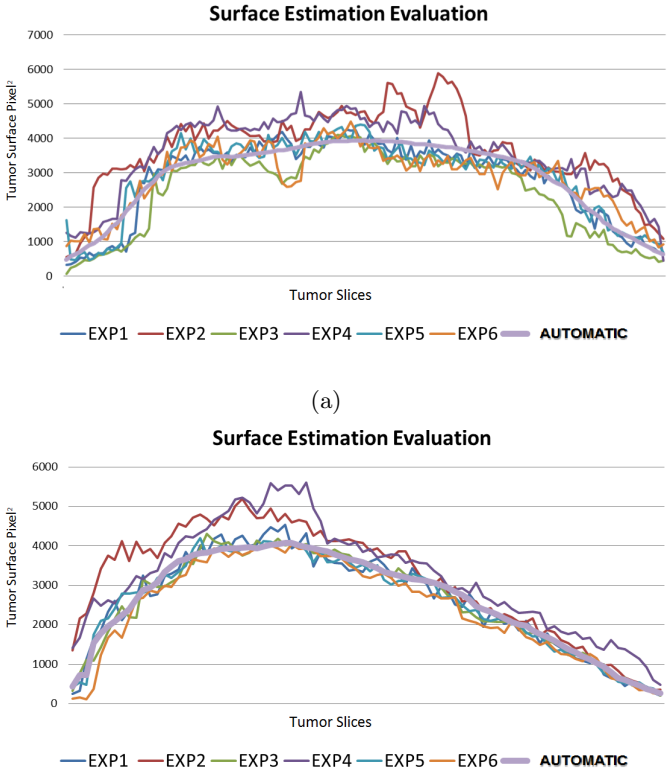


Figure 3.15: Analysis of the Surface Estimation Errors trends

segmented manually by the experts on the axial plane and the automatic segmented masks for the Cases 1 and 4 are shown. This representation highlights the observation made earlier about the loss of spatial continuity in the manual segmentation. Instead, the 3D surface of the tumor automatically computed is more smoothed and for this reason the volume estimation errors are lower than those committed by the experts. The automatic segmentation seems to approximate the average behavior of the experts, and a problem of compensation of underestimation and overestimation errors may occur. Unfortunately, this could result in a poor resolution in the ability of precisely identify tumor contours. To investigate this aspect and quantify the effects we consider an another set of experiments that assess the masks agreement in terms of spatial distribution are preformed. Differently from what was done previously it is not easy to derive a common reference segmentation starting from the annotation of several experts. In the introductory part of this chapter we have seen that several techniques for label fusion are present in literature [186, 142, 143, 141], but at this stage we have preferred to consider

Table 3.11: Assessment of the spatial distribution agreement between the automatic segmented volumetric masks and the manually annotated masks (Jaccard Index)

Jaccard Index	Case 1	Case 2	Case 3			Case 4		
			Axial	Coronal	Sagittal	Axial	Coronal	Sagittal
EXP 1 vs EXP2	82,79%	80.16%	82.62%	75.40%	81.31%	83.59%	69.49%	83.59%
EXP 1 vs EXP3	79,18%	77.15%	83.04%	82.26%	83.04%	81.56%	77.67%	81.56%
EXP 1 vs EXP4	79.98%	77.43%	82.23%	73.30%	75.22%	74.67%	73.33%	74.67%
EXP 1 vs EXP5	82.02%	78.28%	78.50%	82.60%	81.84%	84.29%	78.90%	84.29%
EXP 1 vs EXP6	82.50%	77.65%	78.05%	82.46%	78.05%	86.07%	76.72%	86.07%
EXP 2 vs EXP3	82.97%	77.85%	85.14%	75.18%	82.85%	77.18%	63.36%	77.18%
EXP 2 vs EXP4	83.66%	80.53%	78.42%	75.90%	75.85%	79.09%	81.10%	79.09%
EXP 2 vs EXP5	84.79%	79.26%	81.76%	74.95%	81.43%	85.05%	72.65%	85.05%
EXP 2 vs EXP6	82.96%	81.34%	81.80%	73.73%	73.73%	83.57%	72.39%	83.57%
EXP 3 vs EXP4	84.37%	77.44%	84.64%	73.23%	75.94%	67.69%	66.85%	67.69%
EXP 3 vs EXP5	84.45%	78.22%	83.31%	83.03%	84.61%	76.68%	74.97%	76.68%
EXP 3 vs EXP6	80.65%	77.70%	82.43%	83.60%	73.23%	80.79%	74.47%	80.79%
EXP 4 vs EXP5	86.59%	81.44%	81.35%	72.76%	75.22%	79.95%	76.29%	79.95%
EXP 4 vs EXP6	84.08%	78.80%	81.41%	71.83%	75.85%	76.91%	75.48%	76.91%
EXP 5 vs EXP6	84.47%	81.40%	84.17%	82.89%	75.94%	83.96%	78.66%	83.96%
AUTO vs EXP1	**81.20%**	**74,96%**	**77.85%**	**77.90%**	**77.53%**	**84.20%**	**79.92%**	**84.20%**
AUTO vs EXP2	**82.94%**	**78.90%**	**75.59%**	**70.34%**	**76.66%**	**81.52%**	**71.31%**	**81.52%**
AUTO vs EXP3	**80.03%**	**74.55%**	**78.93%**	**77.87%**	**77.93%**	**80.62%**	**75.82%**	**80.62%**
AUTO vs EXP4	**81.45%**	**77.94%**	**75.63%**	**68.29%**	**71.95%**	**73.66%**	**74.30%**	**73.66%**
AUTO vs EXP5	**80.79%**	**77.92%**	**77.34%**	**76.37%**	**75.84%**	**81.39%**	**79.51%**	**81.39%**
AUTO vs EXP6	**80.43%**	**77.48%**	**77.08%**	**76.67%**	**77.53%**	**84.79%**	**76.58%**	**84.79%**
Mean	*82.49%*	*78.40%*	*80.54%*	*76.69%*	*77.69%*	*80.34%*	*74.75%*	*80.34%*
Var	0.04%	0.03%	0.09%	0.21%	0.12%	0.21%	0.19%	0.21%
Max	*86.59%*	*81.44%*	*85.14%*	*83.60%*	*84.61%*	*86.07%*	*81.10%*	*86.07%*
Min	*79.18%*	*74.55%*	*75.59%*	*68.29%*	*71.95%*	*67.69%*	*63.36%*	*67.69%*

separately the different reference masks considering all combinations of pairs of experts and all the pairs formed by expert and automatic system; in an attempt to understand, also in this case, if the segmentation performances may be considered observations of the interpersonal variability statistical distribution. Table 3.11 summarizes this analysis.
We define the *Expert to Expert Mean Agreement(EEMA)* and the *Expert to Expert Standard Deviation Agreement(EESA)* as the mean and the standard deviation of the agreements between the various experts respectively, and the *Expert to Automatic Mean Agreement(EAMA)* and the *Expert to Automatic Standard Deviation Agreement(EASA)* as the mean and standard deviation of the agreements between all the experts and the automatic segmentation. Assuming a Gaussian distribution for the agreements, by means of these parameters we can study the overlap between the *Expert to Expert* and *Expert to Automatic* distributions. Figure 3.16 shows the results of this analysis. In all the cases a high degree of overlap is observed and even for case 4, the mean value *EEMA* is lower then the *EAMA* (see Figure 3.16). This behavior shows that the automatic system performs as one of the experts.

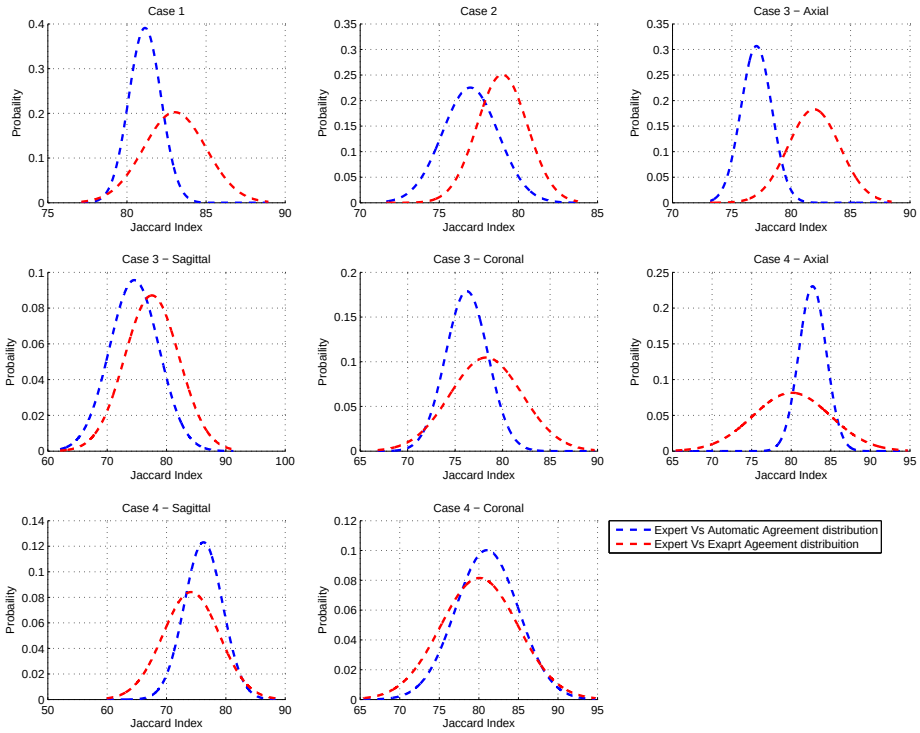


Figure 3.16: Evaluation of the overlap between the Expert Vs Expert agreement and Expert Vs Automatic agreement distribution

Fuzzy Evaluation of the Tumor Segmentation Strategy

Another set of experiments are performed to assess the agreement between the masks obtained with the automatic procedure and the fuzzy reference masks obtained with the proposed Fuzzy Connectedness based common agreement strategy for reference data estimation. The output of this procedure is a graduated mask computed staring from a set of seeds listed by several experts. The labeling process have to be conceived vague in nature and this kind of uncertainty have to be properly modeled, e not reduced in order to completely represent the expert decision attitudes in connecting heterogeneous image elements forming objects and rationally attribute decision labels. Unfortunately the segmentation result is hard mask therefore a non standard method for comparing the crisp segmented mask and the fuzzy reference mask is proposed. We consider to adopt the relative Fuzzy connectedness procedure for estimating the reference data, by this way we have a rigid support in which to evaluate the agreement: all the points for which the membership of the object is greater than that belonging to the background.

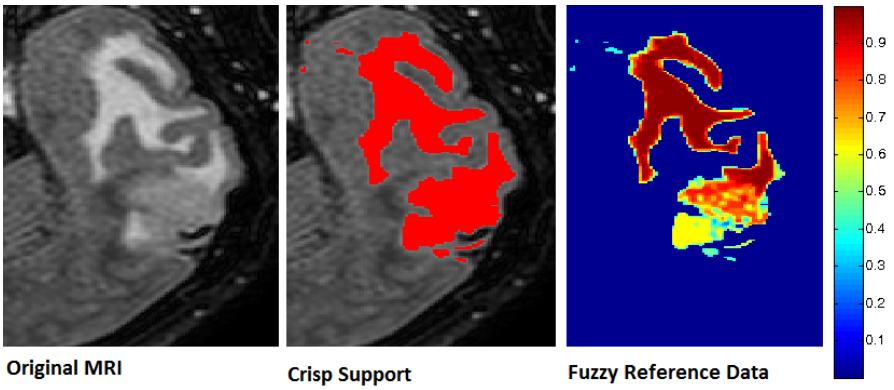


Figure 3.17: Example of the computation of a fuzzy ground trough

Within this rigid mask every the voxels assume the value of the membership to the object. Figure 3.17 shows an example of the fuzzy reference data extracted from an axial slice of the Case 12 of our dataset. The lighter values (MATLAB jet color scale) indicate less plausibility that those points are belonging to the object then actually belonging to the tumor. This areas are located where two different experts may have different opinions areas where errors of commission and omission should be considered in a weighted way. Our evaluation procedure tries to formalize this heuristic idea. If we apply the formulation of Jaccard Index, Precision and Recall by comparing fuzzy and crisp masks they must also take into account the degrees of the soft mask. Following the intuitive principle that errors of omissions in points where the value of membership to the object is high are to be considered more serious than errors made in points where the membership to the object is low both the True Positive and False Negative have to be weighted directly by the fuzzy value; this means that for computing True Positive index, for example, instead of considering the number of points in agreement, we have to consider the cardinality of the corresponding fuzzy set. It is the same for False Negatives. On the contrary, the seriousness of the commission errors is inversely proportional to the value of belonging to the object of the points erroneously included in the mask; therefore the value of the False Positive is the fuzzy cardinality of the background set. All the dense MRI volume of our datesets are evaluated with this strategy: (Case 1-6,10,12). What is said here in an intuitive way is formally discussed in section .
In Table 3.12 the results are summarized.

MICCAI BraTS Dataset

To make our work comparable with that of other authors, we also evaluated our results with respect to a public dataset created on the occasion of Multimodal Brain Tumor Segmentation Challenge, a satellite event of the MICCAI(Medical Image Computing

Table 3.12: Fuzzy Evaluation of the brain tumor segmentation strategy

Case	Jaccard	Precision	Recall
1	78.25%	81.77%	94.80%
2	84.91%	91.80%	91.88%
3	69.52%	97.13%	70.97%
4	82.69%	89.52%	91.55%
5	86.50%	87.61%	98.56%
6	89.72%	90.74%	98.77%
11	81.34%	98.37%	82.45%
12	91.12%	98.53%	92.38%
Mean	83.01%	91.93%	90.17%
Std	6.91%	5.87%	9.29%
Max	91.12%	98.53%	98.77%
Min	69.52%	81.77%	70.97%

and Computer Assisted Intervention Society) Conference. The dataset consists of multi-contrast MR scans of 30 glioma patients (both low-grade and high-grade, and both with and without resection) along with expert annotations for active tumor and edema. For each patient, T1, T2, FLAIR, and post-Gadolinium T1 MR images are available. All volumes were linearly co-registered to the T1 contrast image, skull stripped, and interpolated to 1mm isotropic resolution. No attempt was made to put the individual patients in a common reference space. The MR scans, as well as the corresponding reference segmentations, are distributed in the ITK- and VTK-compatible MetaIO file format. Patients with high- and low-grade gliomas have file names BRATS HG and BRATS LG, respectively. All images are stored as signed 16-bit integers, but only positive values are used. The dataset also contains simulated images for 25 high-grade and 25 low-grade glioma subjects. These simulated images closely follow the conventions used for the real data, except that their file names start with "SimBRATS"; they are all in BrainWeb space; and their MR scans and ground truth segmentations are stored using unsigned 16 bit and unsigned 8 bit integers, respectively. The manual segmentations were performed in the Axial plane. a specific guidelines were used for the manual delineations of "active tumor" and "edema" from the real MR data, you can find all the detail here (http://www2.imm.dtu.dk/projects/BRATS2012/data.html). In this phase of the evaluation we do not consider the separation between edema and an active tumor but the ability of our method to segment whole the pathological area. Figures 3.18 and 3.19 show 2 examples: the first regards an High Grade Glioma case BRATS HG0024 and the second a Low Grade Glioma case BRATS LG0004. These Figures are shown with two purposes: the first is to give a qualitative feeling of the capability of the proposed method to work on this dataset and the second is to comment the difficulty in accepting a mask manually segmented by a single expert as an absolute Ground Truth. Both the Figures are organized as follows: in the first Figure the original MRI image is shown, in the second Figure the superimposion of the ground truth mask in blue is shown, in the third figure the superimposion of the automatic segmented mask is shown and the

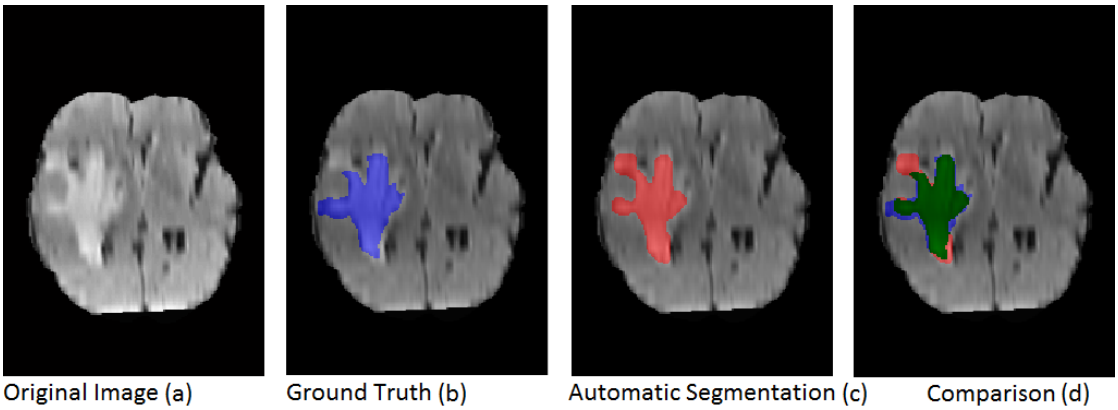


Figure 3.18: Brain Tumor Segmentation Strategy Evaluation on the High Grade Glioma of the BraTS dataset: Qualitative Example Case BRATS HG0024 a) Original Image b) Overlap between original image and mask computed with the automatic strategy c) Overlap between original image and the Ground Truth mask d) Qualitative comparison between the Ground Truth mask and the mask computed with the automatic strategy. Green points are True Positive, Red points are False Positive, Blue points are False Negative

last figure represents the comparison between the two masks, the True positive points are colored in Green the False Positive in Red and the False false negative in Blue. In both cases the qualitative evaluation of the automatic brain tumor segmentation strategy is very good even if there are areas with commission errors in the automatic segmentation. Nevertheless, by means of a purely visual inspection it would seem that it is actually the ground truth mask to underestimate the spread of the perisional edema. The quantitative evaluation is performed in terms of volume estimation error and spatial distribution using the same indexes adopted for the previous evaluation phases. In the following Tables all the numerical results are reported.

Table 3.13: Brain Tumor Segmentation Strategy Evaluation on the High Grade Glioma of the BraTS dataset

id	Data Set file Name	Volume			Spatial Distribution		
		Estimated Vol	True Vol	Vol Est Error	Jaccard	Precision	Recall
1	BRATS HG0001	108169	108819	0.60%	78.26%	88.07%	87.54%
2	BRATS HG0002	46801	48701	3.90%	79.37%	90.30%	86.77%
3	BRATS HG0003	144875	153589	5.67%	80.64%	91.97%	86.75%
4	BRATS HG0004	98872	106776	7.40%	80.68%	92.88%	86.00%
5	BRATS HG0005	68301	60293	13.28%	60.20%	70.75%	80.14%
6	BRATS HG0006	188428	179794	4.80%	69.53%	80.15%	84.00%
7	BRATS HG0007	87092	88337	1.41%	58.75%	74.54%	73.49%
8	BRATS HG0008	201809	200151	0.83%	80.79%	89.01%	89.74%
9	BRATS HG0009	223976	219584	2.00%	79.16%	87.50%	89.25%
10	BRATS HG0010	14371	15957	9.94%	60.00%	67.78%	69.98%
11	BRATS HG0011	113974	144416	21.08%	76.37%	98.17%	77.47%
12	BRATS HG0012	19129	20621	7.24%	64.01%	68.75%	73.53%
13	BRATS HG0013	11469	10887	5.35%	75.81%	84.05%	88.55%
14	BRATS HG0014	146994	141372	3.98%	79.31%	86.77%	90.22%
15	BRATS HG0015	174576	181016	3.56%	76.90%	88.54%	85.39%
16	BRATS HG0022	77743	84636	8.14%	67.04%	83.82%	77.00%
17	BRATS HG0024	80817	83648	3.38%	68.32%	82.60%	79.80%
18	BRATS HG0025	121814	111589	9.16%	68.69%	70.31%	73.49%
19	BRATS HG0026	94208	121670	22.57%	61.52%	63.26%	38.59%
20	BRATS HG0027	213573	194287	9.93%	66.78%	71.98%	90.24%
	Mean			*7.21%*	*71.61%*	*81.56%*	*80.40%*
	Std			*6.03%*	*7.93%*	*9.96%*	*11.80%*
	Max			*22.57%*	*80.79%*	*98.17%*	*90.24%*
	Min			*0.60%*	*58.75%*	*63.26%*	*38.59%*

Table 3.14: Brain Tumor Segmentation Strategy Evaluation on the Low Grade Glioma of the BraTS dataset

id	Data Set file Name	Volume					
		Estimated Vol	True Vol	Vol Est Error	Jaccard	Precision	Recall
1	BRATS LG0001	27907	30879	10%	63%	73%	66%
2	BRATS LG0002	216409	241873	11%	82%	95%	85%
3	BRATS LG0004	59740	59084	1%	70%	82%	83%
4	BRATS LG0006	24664	22551	9%	60%	71%	78%
5	BRATS LG0008	21782	24748	12%	81%	95%	84%
6	BRATS LG0011	59538	53241	12%	86%	88%	98%
7	BRATS LG0012	43027	42727	1%	77%	87%	87%
8	BRATS LG0013	79798	77531	3%	62%	76%	78%
9	BRATS LG0014	17319	16407	6%	68%	79%	83%
10	BRATS LG0015	39916	34886	14.42%	60%	68%	67%
	Mean			*7.80%*	*70.90%*	*81.40%*	*80.96%*
	Std			*4.88%*	*9.76%*	*9.65%*	*9.41%*
	Median			*9.50%*	*68.98%*	*80.35%*	*83.00%*
	Max			*14.42%*	*86.11%*	*95.36%*	*98.01%*
	Min			*0.70%*	*59.92%*	*68.00%*	*66.18%*

Table 3.15: Brain Tumor Segmentation Strategy Evaluation on the High Grade Glioma of the BraTS dataset

id	Data Set file Name	Volume			Spatial Distribution		
		Estimated Vol	True Vol	Vol Est Error	Jaccard	Precision	Recall
1	SimBRATS HG0001	49368	49698	0.66%	89.78%	94.93%	94.30%
2	SimBRATS HG0002	35493	35779	0.80%	86.28%	93.01%	92.27%
3	SimBRATS HG0003	123373	123296	0.06%	84.50%	91.57%	91.63%
4	SimBRATS HG0004	157248	124438	26.37%	74.85%	76.68%	96.90%
5	SimBRATS HG0005	132602	138098	3.98%	87.48%	95.25%	91.46%
6	SimBRATS HG0006	119864	94597	26.71%	75.35%	76.89%	97.42%
7	SimBRATS HG0007	157526	148998	5.72%	86.18%	90.07%	95.23%
8	SimBRATS HG0008	131379	121557	8.08%	83.88%	87.82%	94.92%
9	SimBRATS HG0009	53566	57674	7.12%	92.55%	99.82%	92.71%
10	SimBRATS HG0010	48619	50066	2.89%	91.03%	96.72%	93.93%
11	SimBRATS HG0011	79559	63235	41.63%	68.21%	69.18%	97.98%
12	SimBRATS HG0012	85097	80934	5.14%	83.50%	88.78%	93.35%
13	SimBRATS HG0013	170356	201802	15.58%	79.97%	97.07%	81.94%
14	SimBRATS HG0014	114988	130398	11.82%	60.37%	84.89%	65.33%
15	SimBRATS HG0015	137835	119764	15.09%	68.90%	67.45%	98.10%
16	SimBRATS HG0016	166651	131468	26.76%	77.63%	78.18%	99.10%
17	SimBRATS HG0017	68988	71803	3.92%	71.61%	72.45%	69.43%
18	SimBRATS HG0018	124190	91387	35.89%	71.16%	72.17%	98.07%
19	SimBRATS HG0019	141681	148148	4.37%	84.20%	97.14%	86.34%
20	SimBRATS HG0020	148391	120826	22.81%	64.17%	65.56%	96.80%
21	SimBRATS HG0021	155312	176936	12.22%	85.54%	98.63%	86.57%
22	SimBRATS HG0022	180764	166466	8.59%	81.64%	86.34%	93.75%
23	SimBRATS HG0023	85219	76429	11.50%	65.14%	65.61%	98.91%
24	SimBRATS HG0024	26837	29750	9.79%	86.50%	97.80%	88.22%
25	SimBRATS HG0025	115621	104426	10.72%	70.29%	70.87%	98.83%
	Mean			*12.73%*	*78.83%*	*84.60%*	*91.74%*
	Std			*1.26%*	*0.85%*	*1.40%*	*0.73%*
	Max			*41.63%*	*92.55%*	*99.82%*	*99.10%*
	Min			*0.06%*	*60.37%*	*65.56%*	*65.33%*

Table 3.16: Brain Tumor Segmentation Strategy Evaluation on the Simulated Low Grade Glioma of the BraTS dataset

id	Data Set file Name	Volume			Spatial Distribution		
		Estimated Vol	**True Vol**	**Vol Est Error**	**Jaccard**	**Precision**	**Recall**
1	SimBRATS LG0001	148036	146533	1.03%	82,36%	89,87%	90.79%
2	SimBRATS LG0002	48614	39765	22.25%	71,77%	75,96%	92.86%
3	SimBRATS LG0003	53291	39300	35.60%	61.95%	66.47%	90.13%
4	SimBRATS LG0004	68969	60305	14.37%	76.63%	81.32%	93.00%
5	SimBRATS LG0005	91336	71358	28.00%	72.09%	74.62%	95.51%
6	SimBRATS LG0006	61719	53150	16.12%	63.09%	65.49%	94.52%
7	SimBRATS LG0007	40761	40606	0.38%	73.60%	84.63%	84.95%
8	SimBRATS LG0008	45558	40171	13.41%	73.20%	79.53%	90.19%
9	SimBRATS LG0009	65569	66081	0.77%	83.24%	91.21%	90.50%
10	SimBRATS LG0010	102186	91993	11.08%	83.41%	86.42%	95.99%
11	SimBRATS LG0011	74735	65412	14.25%	77.15%	81.67%	93.31%
12	SimBRATS LG0012	75857	89427	15.17%	76.96%	94.76%	80.38%
13	SimBRATS LG0013	74812	71481	4.66%	79.00%	86.31%	90.33%
14	SimBRATS LG0014	45906	42853	7.12%	69.36%	79.18%	84.82%
15	SimBRATS LG0015	68550	61626	11.24%	78.20%	83.33%	92.70%
16	SimBRATS LG0016	152072	156728	2.97%	83.22%	92.23%	89.49%
17	SimBRATS LG0017	56059	59043	5.05%	77.01%	89.33%	84.81%
18	SimBRATS LG0018	61511	51633	19.13%	68.35%	74.68%	88.97%
19	SimBRATS LG0019	62233	72888	14.62%	80.32%	96.71%	82.57%
20	SimBRATS LG0020	45004	37568	19.79%	73.27%	77.58%	92.94%
21	SimBRATS LG0021	64417	70127	8.14%	79.63%	92.59%	85.05%
22	SimBRATS LG0022	62701	63277	0.91%	78.79%	88.54%	87.73%
23	SimBRATS LG0023	97299	95042	2.37%	75.39%	84.97%	86.99%
24	SimBRATS LG0024	75191	65713	14.42%	80.50%	83.58%	95.63%
25	SimBRATS LG0025	60555	58847	2.90%	82.90%	89.37%	91.97%
	Mean			*11.43%*	*76.06%*	*83.61%*	*89.85%*
	Std			*0.82%*	*0.35%*	*0.66%*	*0.18%*
	Max			*35.60%*	*83.41%*	*96.71%*	*95.99%*
	Min			*0.38%*	*61.95%*	*65.49%*	*80.38%*

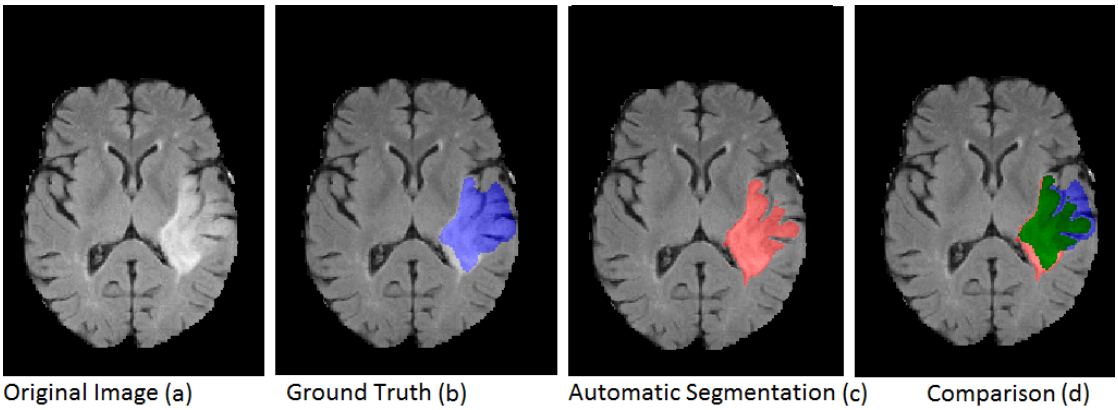
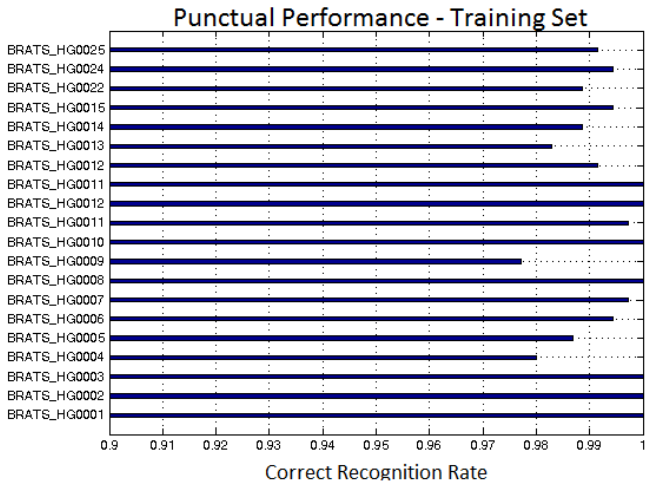


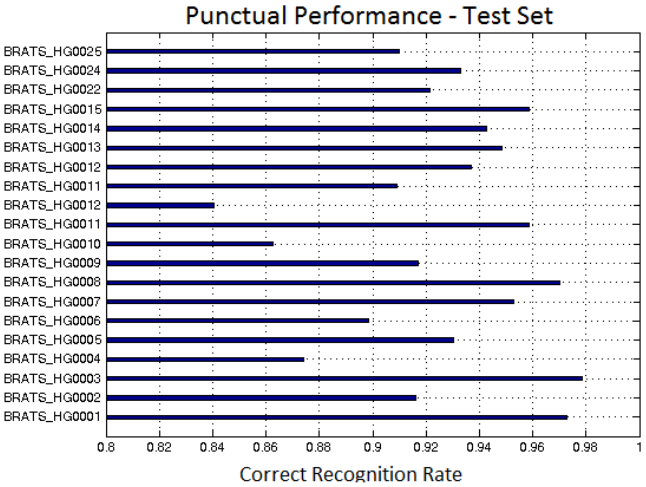
Figure 3.19: Brain Tumor Segmentation Strategy Evaluation on the Low Grade Glioma of the BraTS dataset: Qualitative Example Case BRATS LG0004 a) Original Image b) Overlap between original image and mask computed with the automatic strategy c) Overlap between original image and the Ground Truth mask d) Qualitative comparison between the Ground Truth mask and the mask computed with the automatic strategy. Green points are True Positive, Red points are False Positive, Blue points are False Negative

3.6 Evaluation of Multichannel Tumor Heterogeneity Recognition

In this section the last phase of our MRI segmentation strategy is evaluated. The aim of this phase is to recognize the heterogenous part that compose the tumor(Edema, Necrose and Active Part) using the Machine Learning. In this experimental phase the



(a)



(b)

Figure 3.20: Tumor heterogeneity Recognition Evaluation: Punctual Performance. a) Training Set b) Test Set

performances are assessed reducing the recognition problem to a binary classification that consider uniquely Edema and Active Part. The rationale behind this choice lies in

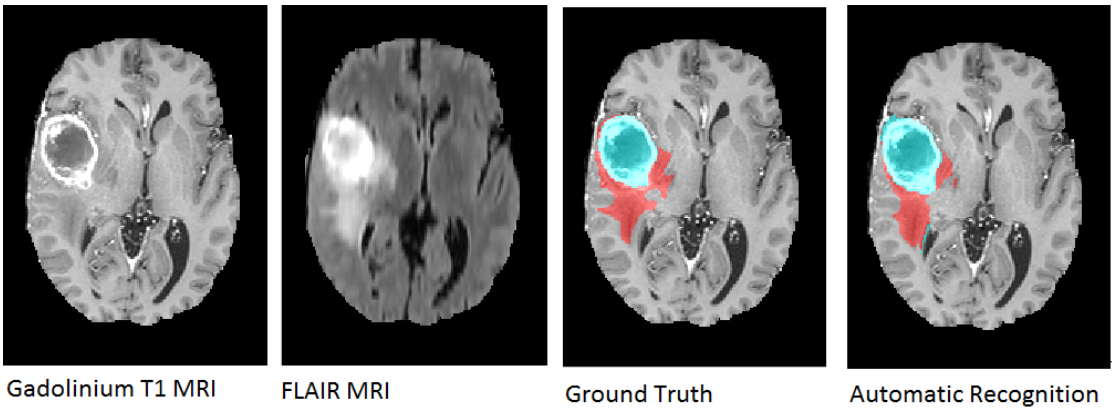


Figure 3.21: Tumor heterogeneity Recognition Evaluation on the High Grade Glioma of the BraTS dataset: Qualitative Example Case BRATS HG0001 a) Gadolinium T1 MRI Original Image b) FLAIR MRI Original Image c) Overlap between Gadolinium T1 MRI original image and the Ground Truth masks d) Overlap between Gadolinium T1 MRI original image and the automatic recognized masks.(Red points are Edema and Cyan point are the Active Tumor)

the fact that the MICCAI dataset used as reference data set in the experiments includes images annotated in terms of Edema and Active Tumor classes. The use of these data has the aim to make our results comparable with the state of the art tumors segmentation algorithms presented in this Challenge. All the real High Grade Glioma of the dataset are used for this evaluation phase (BRATS HG, 20 cases).
The presented results are obtained using the cross validation technique choosing for each repetition of the experiment 750 random points for each class equally distributed in 10 of the 20 cases, also randomly selected. The recognition performances are evaluated both on the training and test sets. Figure 3.20 shows the mean of the punctual recognition accuracies obtained at each cross validation experiment (performed 100 times) both for the training and test sets. The performances are very promising with a mean value of the percentage of correct recognition equal to 99% and 92% for the training and test respectively. The cross validation experiment then showed a good stability in the obtained performances, showing very low variances in the percentage of correct recognition for the various trials: 0.72% in training data and 3.7% in test data.
In addition to this punctual analysis the ability of the method to generalize over the whole MRI volume also in points never used in previous experiments was evaluated. In Figure 3.6 and 2 cases are presented for the qualitative analysis of the results (BRATS HG0001, BRATS HG0002). In these Figures one slice for both the MRI modalities (Gadolinium T1 and FLAIR) used in this phase is showed together with the superimposion between the Edema and Active Tumor mask manually labeled by the expert

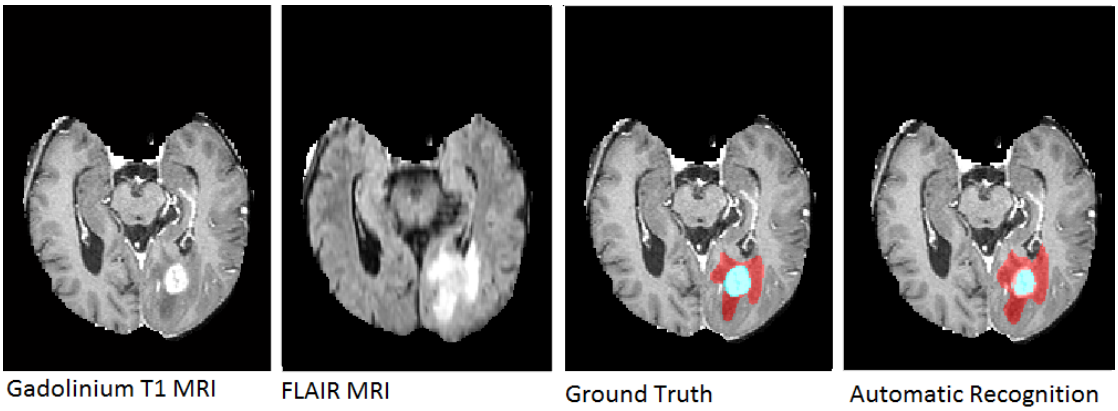


Figure 3.22: Tumor heterogeneity Recognition Evaluation on the High Grade Glioma of the BraTS dataset: Qualitative Example Case BRATS HG0001 a) Gadolinium T1 MRI Original Image b) FLAIR MRI Original Image c) Overlap between Gadolinium T1 MRI original image and the Ground Truth masks d) Overlap between Gadolinium T1 MRI original image and the automatic recognized masks.(Red points are Edema and Cyan point are the Active Tumor)

and automatically recognized by the presented strategy. In both of these examples, the performance can be considered good in average but with some inaccuracy. In the first case Figure 3.6 the active tumor is slightly overestimated in the lower right and it has an underestimation of the edema points. In this regard it is noteworthy that the recognition si done downstream the segmentation of pathological area, then only the points considered tumor during previous phase are analyzed. By this way any omission errors cannot be corrected. It would be good that the process of segmentation rather have a tendency to overestimate and that at this stage of recognition the class healthy tissue is contemplated. Also the second case (Figure 3.6) can be judged fairly good with some errors in the boundary between active tumor and edema.

The quantitative analysis is performed in comparison with the paper collected in the "*Proceedings of MICCAI-BRATS 2012 October 1 st , Nice, France*". All these works were evaluated on the same dataset used here. Considering the prestige of the organization, but also the fact that the event took place very recently, these few works can well represent a trend of the state of the art in this topic. The evaluation metric used is that provided by online performances assessment system of the competition: the Dice coefficient

$$D = \frac{2A \cap B}{|A| + |B|} \tag{3.3}$$

Algorithm	Mean of the Dice Coefficients	
	Edema	Tumor
Context-sensitive Classification Forests for Segmentation of Brain Tumor Tissues [197]	70%	71%
Segmentation of Brain Tumor Images Based on Integrated Hierarchical Classification and Regularization [15]	61%	62%
Spatial Decision Forests for Glioma Segmentation in Multi-Channel MR Images [73]	56%	58%
Multimodal Brain Tumor Segmentation Using The Tumor-cut Method on The BraTS Dataset [76]	56%	73%
Probabilistic Gabor and Markov Random Fields Segmentation of Brain Tumours in MRI Volumes [166]	56%	67%
Automatic Brain Tumor Segmentation a Coupled Global-Local Intensity Bayesianbased on Model [172]	43%	55%
Segmenting Glioma in Multi-Modal Images using a Generative Model for Brain Lesion Segmentation [114]	57%	56%
Segmenting Glioma in Multi-Modal Images using a Generative-Discriminative Model for Brain Lesion Segmentation [113]	69%	70%
Multi-modal Brain Tumor Segmentation Latent Atlases [139]	60%	58%
Mean	**58,68%**	**63,28%**
Std	**8,00%**	**7,03%**
Our Method	**59,45%**	**70,42%**

Table 3.17: Comparison analysis between the performance obtained proposed algorithm and the methods collected in: "*Proceedings of MICCAI-BRATS 2012 October 1 st , Nice, France*"; metric used Dice Coefficient.

where A and B are two binary mask; with respect to the Jaccard coefficient can be easily computed with the following formula:

$$D = \frac{2J}{1+J} \tag{3.4}$$

In Table 3.17 the mean of the dice coefficient obtained on all the 20 High Grade Glioma cases of the BRATS datasets are reported for the most of the presented algorithm and also for the proposed method for the Edema and Active Tumor recognition. In both the classes the performances obtained with the presented strategy are better than those obtained on average with the other methods. In particular the Dice's index is higher than the average of only 0.77% for the edema recognition but nicely of 7.14% for the active tumor class. I believe that this result can be further improved by solving the problem of underestimation of edema discussed by analyzing the case in Figure 3.6. Only three of the papers presented at the workshop are not considered in this comparative analysis because performance are reported in a way not directly comparable with our results.

Part II

Functional Magnetic Resonance Imaging

4
fMRI Data Analysis

4.1 Overview

The fMRI is used to understand physiological changes that accompany brain activation. Currently there exist a number of different imaging modalities that allow researchers to study brain functionality. They differ in what they attempt to measure, as well as in the temporal and spatial resolution.

Electro/Magneto-encephalography EEG/MEG: this methods study the electrical and magnetic activity in the brain. They provide temporal resolution on the order of milliseconds but uncertain spatial localization.

fMRI and positron emission tomography (PET:) this methods provide information on blood flow changes that accompany neuronal activity with relatively high spatial resolution, but with a temporal resolution limited by the much slower rate of brain hemodynamics.

fMRI in the past few years has taken a dominant position in the field of neuroimaging [98]. During the course of an fMRI experiment, a series of brain images is acquired while the subject performs a set of tasks the changes in the measured signal between individual images are used to make inferences regarding task-related activations in the brain. Data consist of a sequence of Magnetic Resonance Images (MRI) each consisting of a number of uniformly spaced volume elements, or voxels, that partition the brain into equally sized boxes. The image intensity from each voxel represents the spatial

distribution of the nuclear spin density in that area. Changes in brain hemodynamics, in reaction to neuronal activity, impact the local intensity of the MR signal Changes in voxel intensity across time can be used to infer when and where activity is taking place. A good number of these voxels consist solely of background noise, and can be excluded from further analysis, the total amount of data that needs to be analyzed is staggering. The data exhibit a complicated temporal and spatial noise structure with a relatively weak signal.
To construct an image, the subject is placed into the field of a large electromagnet. The magnet has a very strong magnetic field, typically between $1.5 - 7.0$ Tesla, functional Magnetic Resonance Imaging is most commonly performed using blood oxygenation level dependent (BOLD) contrast to study local changes in deoxyhemoglobin concentration in the brain. BOLD imaging takes advantage of inherent differences between oxygenated and deoxygenated hemoglobin. Each of these states has different magnetic properties, diamagnetic and paramagnetic respectively, and produces different local magnetic fields. The cerebral blood flow refreshes areas of the brain that are active during the execution of a mental task with oxygenated blood. During the acquisition the patient perform cognitive or motor tasks with a given repetition time.
The acquisition paradigm is often block designed: the patient execute the task by alternating periods of rest. The stimulus is then described by a square wave. When designing an experiment it is important to balance the need for adequate spatial resolution with that of adequate temporal resolution. The manner in which fMRI data is collected makes it impossible to simultaneously increase both, as increases in temporal resolution limit the number of space measurements that can be made in the allocated sampling window and thereby directly influence the spatial resolution of the image.
One of the benefits of MRI is its ability to provide detailed anatomical scans of gray and white matter with a spatial resolution well below $1mm$. However, the time needed to acquire such scans is prohibitively high and currently not feasible for use in functional studies.
The fMRI spatial resolution is typically on the order of $[3 \times 3 \times 5]mm^3$ corresponding to image dimensions on the order of $[128 \times 128 \times 30]$, which can readily be sampled in approximately 2 seconds. fMRI provides relatively high spatial resolution compared with many other functional imaging techniques The temporal resolution of an fMRI study depends on the time between acquisition of each individual image: repetition time (TR). In most fMRI studies the TR ranges from $0.5 - 4.0$ seconds. These values indicate a fundamental disconnect between the underlying neuronal activity. The main component of the fMRI signal is the BOLD signal that is a nonlinear function of the results of neuronal and vascular changes. The shape of the response depends both on the applied stimulus and the hemodynamic response to neuronal events. A big number of methods for modeling the BOLD response exist in the literature. A major difference between methods lies

in how the relationship between the stimulus and BOLD response is modeled: linear and non-linear. While the non linear models are more biophysically plausible than their linear counterparts, they have a number of drawbacks. They require the estimation of a large number of model parameters, they do not always provide reliable estimates with noisy data and they do not provide a direct framework for performing inference. So despite the flexibility of nonlinear models is attractive, linear models provide robust and interpretable characterizations in noisy systems. It is common to assume a linear relationship between neuronal activity and BOLD response. Studies have shown that under certain conditions the BOLD response can be considered linear with respect to the stimulus particularly if events are spaced at least 5 seconds. The ability to assume linearity is important, as it allows the relationship between stimuli and the BOLD response to be modeled using a linear time invariant system LTI Figure 4.1. The measured fMRI

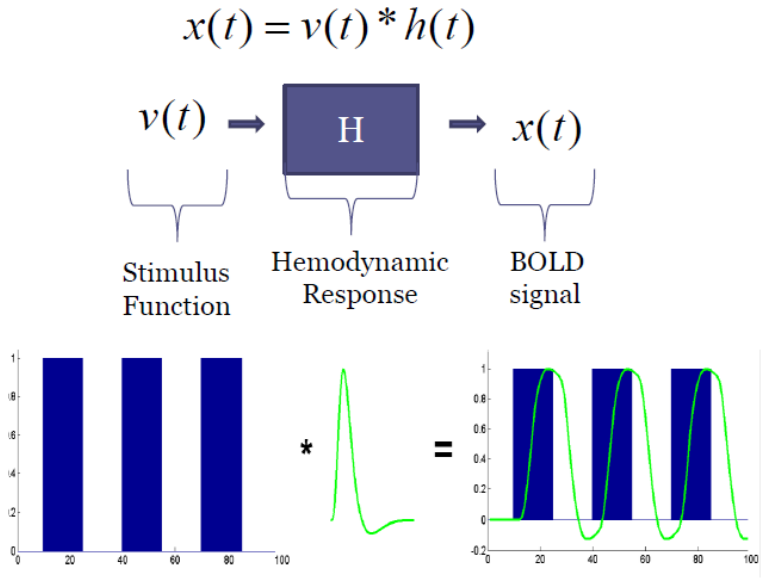


Figure 4.1: Description of linear time invariant system that modeled the bold response to a block design stimulus

signal is corrupted by random noise and various nuisance components that arise due both to hardware reasons and the subjects themselves. Fluctuations in the MR signal intensity caused by thermal motion of electrons within the subject and the scanner gives rise to noise that tends to be highly random and independent of the experimental task. Scanner drift, caused by scanner instabilities. Physiological noise due to patient motion, respiration and heartbeat cause fluctuations in signal across both space and time. It can often be modeled and the worst of its effects removed.

There are several common objectives in the analysis of fMRI data. All of these objec-

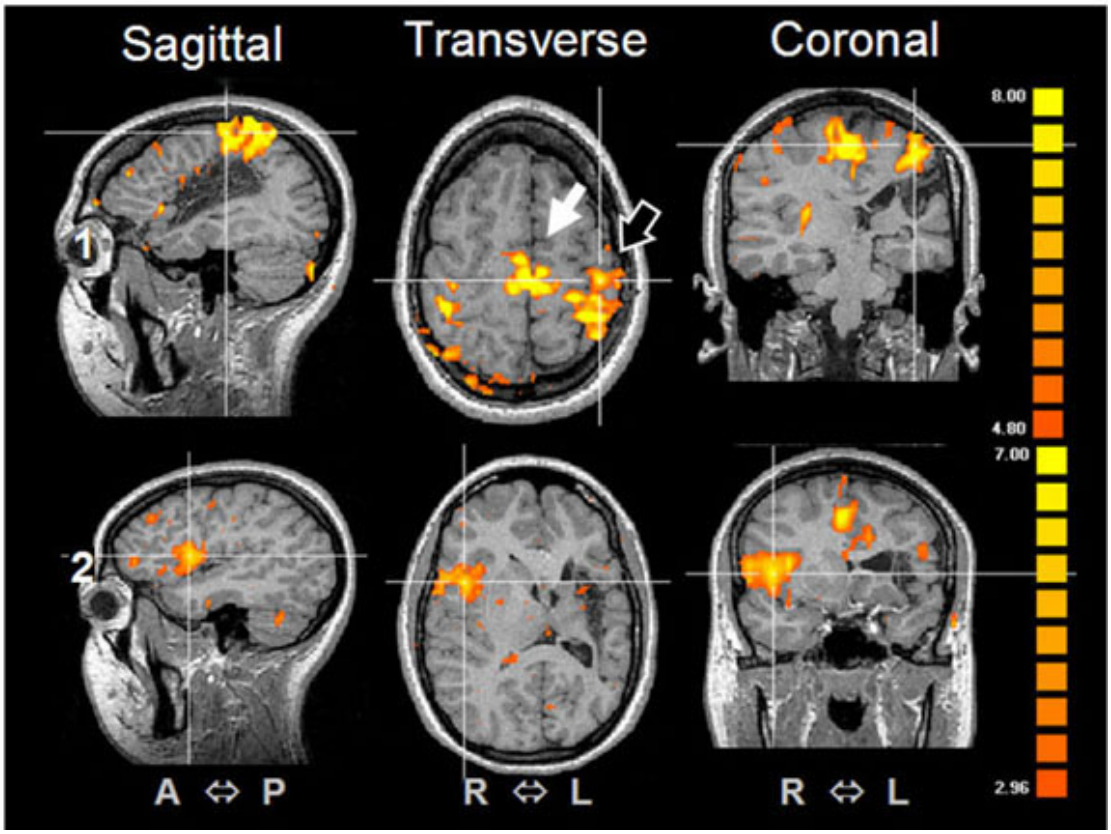


Figure 4.2: Example of functional MRI Statistical Parametric Map (SPM)

tives are related to understand how the application of certain stimuli leads to changes in neuronal activity the common fMRI aims are: to localize regions of the brain activated by a certain task, to describe distributed networks that correspond to brain function and to make predictions about psychological or disease states; in this study we focus on the first. The state of the art method used for fMRI time series analysis aimed to localize the functional activation is the General Linear Model(GLM). It is mathematically identical to a multiple regression analysis but it stresses its suitability for both multiple qualitative and multiple quantitative variables. The GLM is suited to implement any parametric statistical test with one dependent variable. Because of its flexibility to incorporate multiple quantitative and qualitative independent variables, the GLM has become the core tool for fMRI data analysis after its introduction into the neuroimaging community by Friston and colleagues [17]. This method evaluates for each time series associated with each voxel of the fMRI sequences the statistical significance of the activation hypothesis. Then the result is a volumetric Statistic Parametric Map(SPM) that says, for each zone, the plausibility that it is involved in the execution of the task that

the patient has performed during the acquisition.
The physician's task is to analyze this map to locate the area in which the activations are concentrated. The fMRI has an essential role in the clinical life of a brain tumor because the functional analysis has a direct impact on the surgical planning as detailed in the introduction chapter.

4.2 Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM)

The qualitative evaluation of the SPM is often a challenging and time consuming task for the physicians. Our goal is to provide an integrated framework able to analyze the fMRI Statistical Parametric Map to assist the radiologist in its decisions concerning the location of the functional zones of the patient's brain. The Schema of the the Computer Assisted Analysis is shown in Figure 4.3. In this diagram only the main features are mentioned but several ancillary tools are implemented in the framework for the management visualization and processing of the fMRI volumetric data. The mains

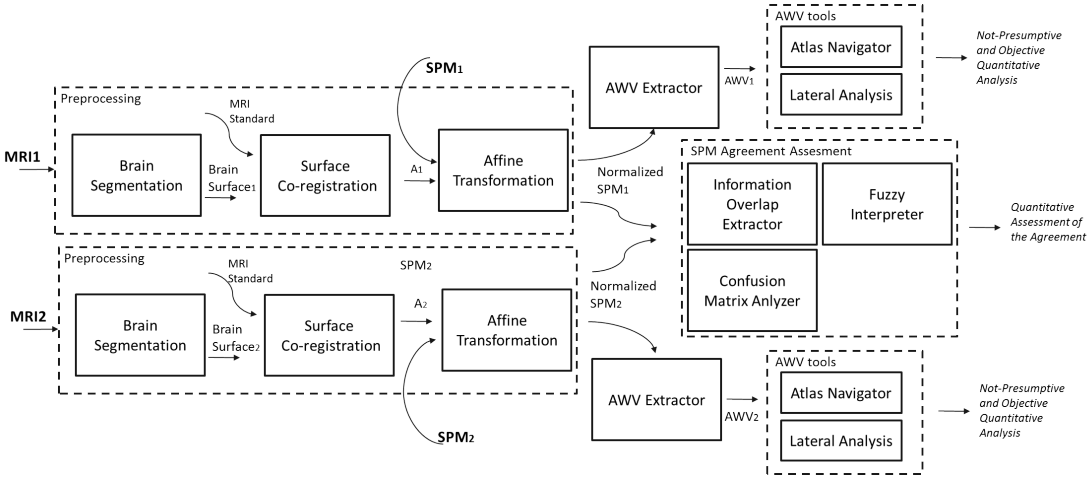


Figure 4.3: Schema of the Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM)

steps of the execution pipeline are here summarized and detailed in the follows:

Brain Segmentation Some software tools for fMRI time series analysis do not include brain segmentation in the processing pipeline, therefore all MRI voxels are analyzed and may be active, with the result that active zones can be found outside the brain. Brain Segmentation allows to restrict the quantitative analysis in the brain. Moreover, the segmentation allows to extract brain surface that will be used in the next registration phase.

Brain Surface Co-Registration During fMRI exam, a set of brain volumes is acquired and SPM is the result of the time series analysis. The maps are computed in relation with one of the acquired volumes, usually the first or the last. The aim of this step is to align this volume to a standard one to which an anatomical atlas is linked. By this way we compute the geometric transformation to be applied to the SPM for bringing it in the atlas space. Our strategy makes use of a surface registration method based on spherical harmonic decomposition **SPH**erical **HARM**onic (SPHARM) [132].

Comparison Between Activation Maps This stage concern the comparison of two different Maps to assess the statistical agreement [131, 133]. nt time. The study of the statistical agreement will be carried out following two ways. The first way considering the SPM as two classifiers and extracting the confusion matrix and the Cohen's Kappa index to assess the agreement [46]. Some important considerations will be made on the statistical dependence of classes and a new formulation of the Kappa index will be proposed to overcame this problem. The second way considers the SPM as two 3D images, and computes the similarity of the SPM images with a fuzzy formulation of the Jaccard Index [81].

Activation Weighted Vector Extraction The aim of this phase is to briefly describe the volumetric Map for giving to the expert a synoptic view of the SPM in a simply feature vector. For this purpose an index that describe the information contained in the map *Activation Weighted Vector(AWV)* was proposed in [131]. *AWV* is brief but comprehensive and can describe quickly and easily the SPM. This vector can be used as a guide to the expert for the exploration of the maps. Moreover, for its brevity and specificity, it can be used in a pattern recognition process for the identification of a particular situation. Many application of the AWI vector can be developed for example, the hemisphere dominance is an aspect that can be studied through the AWI vector [134].

In the follows all the mentioned steps are detailed.

4.2.1 Preprocessing: Segmentation and Surface Co-Registration

The preprocessing phase is composed of two phase the brain segmentation and the surface co-registration. (Figure 4.4) The brain segmentation strategy used in this work is a fully automatic and unsupervised method [129]. This is a 2D boundary-based strategy that usees the graph searching principles [112, 159, 170]. The brain boundary is described using a set of feasibility constraints and the segmentation problem is described in terms of constrained optimization obtaining solvable as the search of a minimal path in a weighted graph. More details on the use of this technique are provides in Chapter 2.

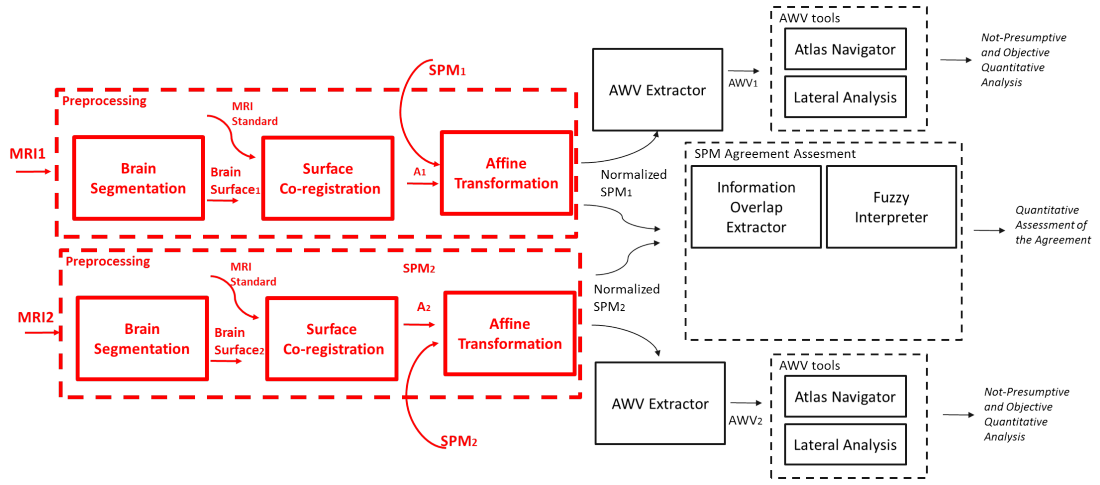


Figure 4.4: Schema of the Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM) with highlight the part detailed in Section (4.2.1)

The output of the brain segmentation is a 3D boundary that represent the brain surface. The aim of the second preprocessing phase is to find the geometric transformation that bring the MRI of the patient in a standard space.

AffineSPHARM Registration

The 3D surface registration was dealt with extensively in machine vision and computer graphics literature. In the last few years a lot of techniques were proposed [189]. Particular attention was given on a parametric surface modeling. The most widely used technique employs a description of a radial or stellar surfaces $v(\theta, \varphi)$ with the spherical harmonic decomposition [12]. An extension of this work allows to describe more general 3D simply connected surface using three radial functions $v(\theta, \varphi) = (x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi))^T$ called SPHARM (SPHerical HARMonic modeling) [33].
The aim of the SPHARM registration technique is the use of the spherical parametrization of a 3D closed surface for the description of the moving shape and static template. Consider a 3D radial object represented by a set of vertices in the cartesian space $v = (x, y, z)$. The mapping of these vertices in the spherical domain $v(\theta, \varphi) = \rho$ where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$ is performed with surface parametrization [68]. The spherical homogeneous sampling of the space is obtained starting with an icosahedron and iteratively subdividing each triangle into four smaller triangles. A spherical surface can be decomposed in a set of orthogonal bases through an integral transformation. The

synthesis functions is the following:

$$v(\theta,\varphi)=\sum_{l=0}^{L}\sum_{m=-l}^{l}c_l^m Y_l^m(\theta,\varphi) \tag{4.1}$$

The spherical surface modeling is specially appropriate to the brain description for the big dimensionality reduction of the spherical harmonic representation. In Figure 4.5 is shown a reconstruction of a brain surface obtained by increasing number of harmonics, the original surface is described by 22511 vertices. Synthesizing the surface with only 10 harmonics (100 coefficients which only 50 significants for the symmetry properties of the transformation) the 99.95% of the signal energy is preserved with a compression rate 0.0024. SPHARM surface modeling of a radial object benefits of the rotation property.

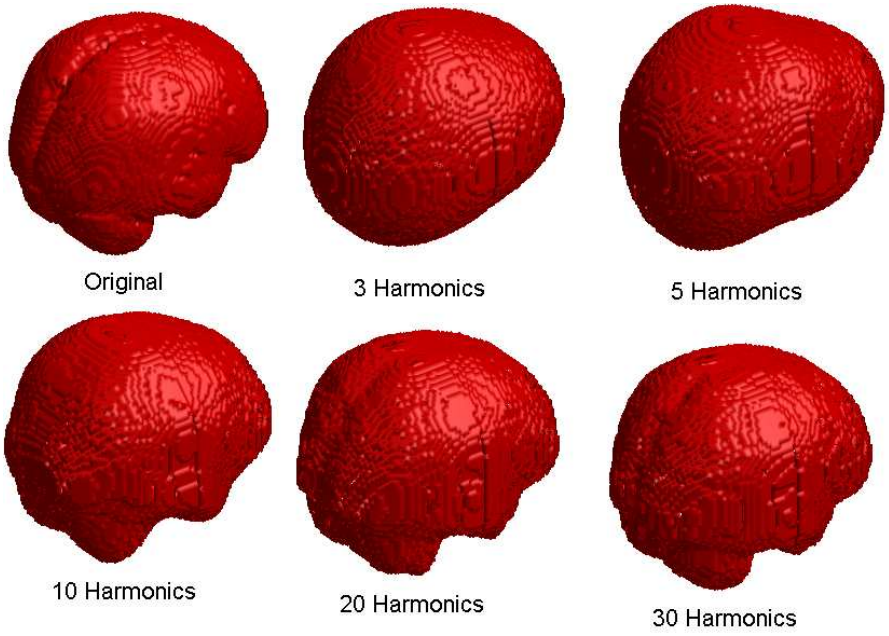


Figure 4.5: Synthesis of brain surface with an increasing number of harmonics

The rotation of a surface, defined trough the three Euler angles (α,β,γ) can be compute directly in the spherical domain. If the spherical function represents a radial object, the rotation coefficients rotates, the parametrization and also the object. The possibility to rotate a surface only by rotating the harmonic expansion coefficients makes the SPHARM alignment algorithms very efficient but restricted only to the rigid transformations. Our goal is to generalize this approach to the Affine transformation making it applicable to the alignment of brain surfaces of different patients or the registration with a Standard Atlas. To exploit the good features of SPHARM modeling is necessary to perform the affine registration in the spherical domain. To this purpose, a transformation of

the spherical coefficients that guarantees an affine transformation in a space domain is necessary.
Instead of finding an analytical solution, we attempt to solve the problem trough a Radial Basis Function (RBF) Neural Network. The affinity is a class of linear transformations that maps variables in new variables, it consists of a linear transformation followed by a translation.

$$x' = Ax + t \tag{4.2}$$

To find the affine transformation in the SPHARM domain we start by considering, at first, only the rotation: as shown by Li Shan, all the coefficients $c_l^m(\alpha, \beta, \gamma)$ of the rotated surface are a linear combination of all the coefficients of the same order and lower degree.

$$c_l^m(\alpha, \beta, \gamma) = \sum_{n=-l}^{l} D_{mn}^l(\alpha, \beta, \gamma) c_l^n \tag{4.3}$$

Observing that the affinity is a linear transformation but don't preserve the orthogonality of the basis we can suppose that all the coefficients $c_l^m(\mathbf{a})$ of the surface after affine transformation are a linear combination of all the other coefficients.

$$c_l^m(\mathbf{a}) = \sum_{k=0}^{L'} \sum_{n=-k}^{k} T_{mn}^{lk}(\mathbf{a}) c_k^n \tag{4.4}$$

The analytical definition of the function $T_{nm}^{kl}(\mathbf{a})$ is a critical aspect and is not guaranteed a closed-form expression. To asses this problem the RBF networks were introduced to regress this function. One of the easiest and effective way to model regression is that of using a finite dimensional space of function spanned by a given basis. The RBF neural network solves the regression problem by this way with a very simple structure and, differently from other types of neural network, like Multy Layer Perceptron (MLP), with a faster training [35]. Moreover, the RBF works well if it is trained with many examples in this specific application, the ground truth set can be arbitrarily large. For each $c_l^m(\mathbf{a})$ one RBF network is involved. The generation of the training set is easy: let be $v'(\theta, \varphi)$ the surface $v(\theta, \varphi)$ after the affine transformation $A(\mathbf{a})$ we can extract the c_l^m of the original surface and the $c_l^m(\mathbf{a})$ of the transformed surface. The input pattern of the RBF network approximates the $c_l^m(\mathbf{a})$ is $[a_1, ...a_{12}, c_0^0, c\_1^1, c_0^1, c_1^1 .... c_l^l, c_l^l]$. The training set is composed of a series of affine transformations of the same object, the network can generalize only in the domain of the training shape and the affine transformations of this. Depending to the training set the network can be specialized to a particular affine deformations (scaling shearing,reflection...). The spherical description of a surface is intrinsically a metric of the shapes similarity. The surfaces alignment is obtained by aligning the SPHARM models minimizing the root mean squared distance (RMSD)

between the harmonic coefficients.

$$RMSD = \sqrt{\frac{1}{4\pi} \sum_{l=0}^{L_{max}} \sum_{m=-l}^{l} ||\mathbf{c}_{1,l}^{m} - \mathbf{c}_{2,l}^{m}||^2} \tag{4.5}$$

In Figure 4.6 some registration results between brain surface of fMRI patients and a standard brain.

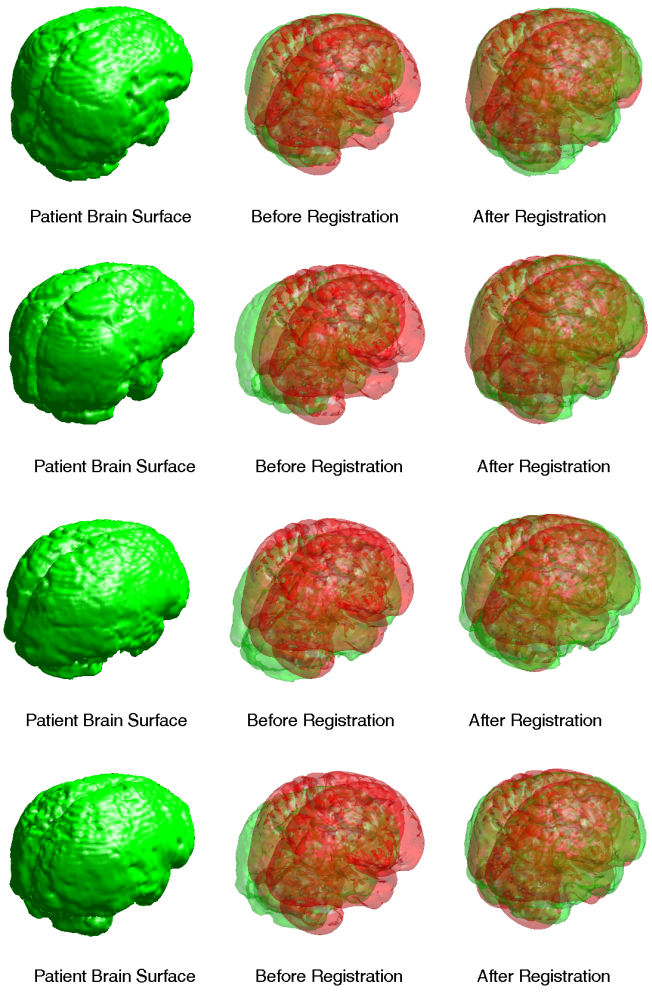


Figure 4.6: Examples of registration results on Real MRI patients

After the execution of the detailed preprocessing phase the core processing phases aimed to assess the agreement of SPMs and to provide a concise but comprehensive description of the SPM can be applied.

4.2.2 Comparison Between Activation Maps Computation Pipelines in a Clinical Context

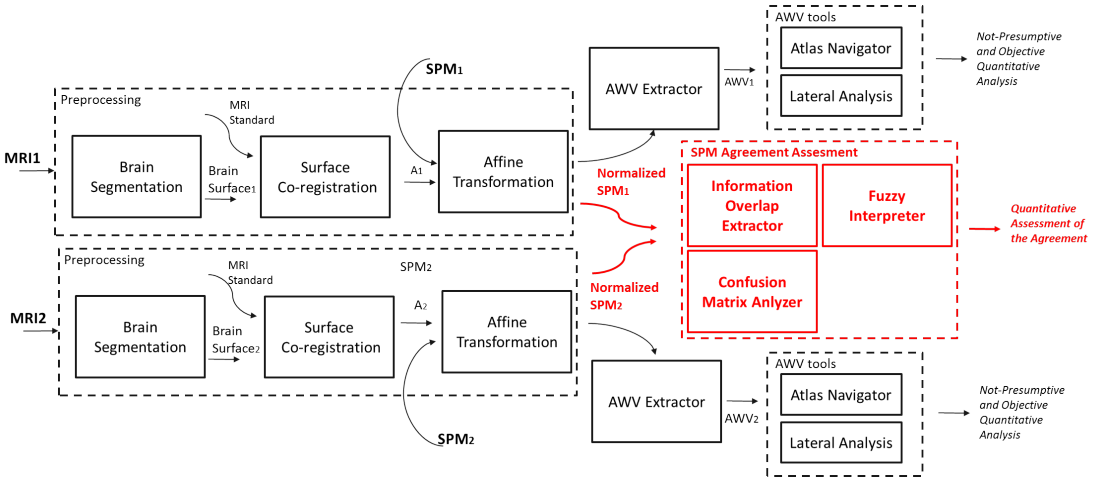


Figure 4.7: Schema of the Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM) with highlight the part detailed in Section (4.2.2)

A focal part of the integrated framework for computer assisted analysis of fMRI SPM presented in this Ph.D. Thesis concerns the comparison of different brain activation maps(Figure 4.7). This goal also has a large clinical return, wherein in the today practice the comparison between different maps occurs qualitatively. Especially in the glial tumors treatment, to have a procedures for the SPM comparing allows for objective measurement the potential of the nervous system to reshape itself during ontogeny, learning or following injuries: brain plasticity. Moreover, in a last few year a large number of software tools for time series analysis was developed [179, 1]. The user is then faced with the question of how to choose among them. Of course, many considerations are important at this stage, including costs, operating systems and software compatibility, and more. In our opinion an important question should also be whether all the pipelines considered supply the same results when they have the same information as an input. We have especially investigated if it is possible to measure in a quantitative way such differences.

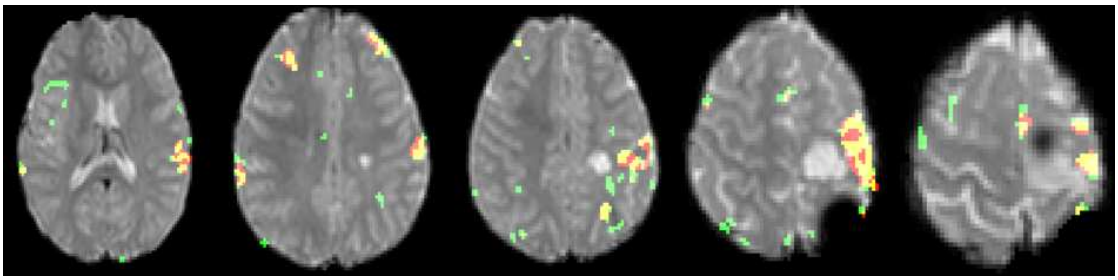
A large amount of work has been done in comparing SPMs both in medical and computer science fields [193, 191, 123]. The state of the art in this field can be analyzed from both the aim of the evaluation strategy and the metrics used in the evaluation process.

As regards the first aspect, some authors used synthetic phantoms to evaluate the performances of commonly used fMRI software packages with respect to motion correction [117, 137]. The definition of a phantom that can simulate real data well is a critical part of the research in the field of fMRI tools evaluation. Piknes et. al. published a study

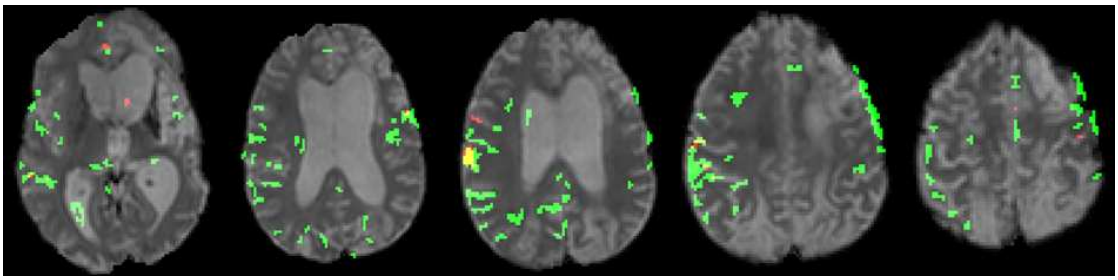
focused on the development of a synthetic phantom based on real subject data [137]. Often, the strategies based on simulated data used Receiver Operating Characteristic (ROC) curves for assessing the methods sensitivity and specificity [163, 153], with all the drawbacks arising from the use of synthetic data.
Other approaches are focused on the repeatability of the exam pipeline. Strother et al. and Tegeler at el.[165, 168] have conducted longitudinal studies repeating the same pipeline on couples of sequential fMRI acquisitions. The results obtained show that the SPMs for the finger tapping tasks obtained by high-field experiments are well reproducible. On the other hand, Machielsen et. al. [105] measured the reliability of a visual encoding task highlighting inconsistencies in the activation in repeated examinations. These findings indicate that, whereas consistent patterns of activation exist, more insight is needed into what determines the volume of activation, especially to assess cognitive alterations in patients over time.
Other authors have employed their effort to study the repeatability in terms of statistical methods used for fMRI volumes analysis, observing that the relative performance of the methods varies considerably across subjects and classification tasks [154, 148].
From the point of view of the metrics used for the evaluation two categories of approaches can be considered: *threshold* and *non-threshold*. SPM is a 3D map that describes brain activation where each voxel value identifies the statistical significance of the hypothesis of activation, thus SPM is a volume of discrete values between 0 and the maximum value of statistical significance obtained by the analysis. The assessment of the agreement between two such maps has been considered a complex task for a long time. Therefore, many approaches have defined a threshold value by means to which create binary activation maps more manageable. Genovese et al. and Maitra et al. have proposed a metric for evaluating binary maps based on pseudo-ROC analysis [72, 107]; Strother et al. and Tegeler et al. performed the repeatability analysis mentioned above on binary maps, using a metric designed ad hoc for the study. Moreover, classical statistical agreement tools as the Cohen's kappa index [46], Jaccard index [81] and Dice coefficient [56] were applied to assess accuracy and reliability in fMRI considering binary maps [94, 48]. In recent years the trend is to study evaluation strategy and metrics that can be applied on *non-threshold* maps. Maitra reformulated the pseudo ROC approach to eliminate the thresholding requirement [106]. The Intraclass Correlation Coefficient(ICC) [152] is a widely used *non-threshold* metric [37]. A new approach based on measuring the approximate Mutual Information between the fMRI time-series of a validation dataset and a calculated activation map is proposed in [3]. This method can be applied on thresholded label maps or continuous maps.
From the cited literature one can draw two important interrelated issue which are still being studied: on one hand, the management and quantitative assessment of inconsistencies of the results obtained from different processing methods and on the other hand,

the choice of appropriated evaluation metrics.
Proceeding from these considerations, a first objective of our work is to make a contribution to the ongoing debate on the agreement between the different processing methods, breaking down the problem in the clinical context in which fMRI is a widely accepted tool in glial tumor pre-surgical planning procedures. For this purpose we developed a detailed comparative analysis aimed to show how and at what extent two of the most relevant and widely used processing methods differ in producing SPMs.
The first method considered is IViewBOLD, a widely used tool in clinical practice, and the second method considered is FMRI Expert Analysis Tool (FEAT) included in FMRIB Software Library (FSL) [179], which is one of the state-of-the-art methods within the scientific community.
In Figure 4.8 two SPMs computed with the processing pipeline studied are overlapped. Figure 4.8(a) shows an example of good overlapping and in Figure 4.8(b) shows an example of bad overlapping; our goal is to make this qualitative feeling a quantitative agreement measure.

Dealing with the problem of choosing an adequate evaluation metric, we proceeded



(a)



(b)

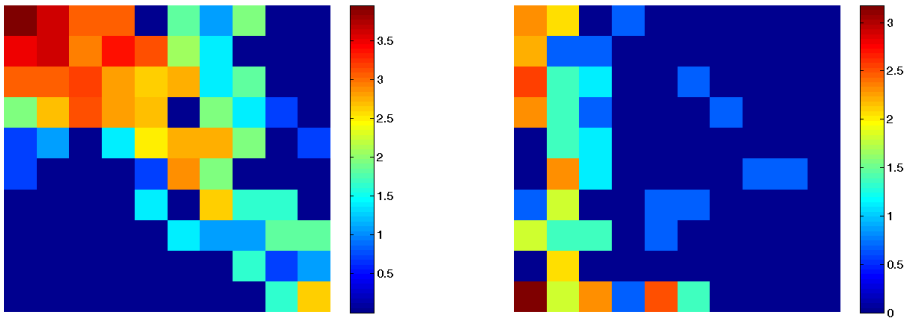
Figure 4.8: IViewBOLD and FEAT FSL SMPs superimposed on MRI acquisition (Red Map FEAT FSL, Green Map IViewBOLD, Yellow intersection) a) Good Agreement b) Bad

from the assumption that the evaluation conducted on graduate *non-threshold* maps is preferable in our context. A number of metrics work on *non-threshold* maps, some-

one conceived ad hoc [152], others derived from the standard indexes [3]. Consistently with what was done in literature we conceived a new evaluation strategy to compare *non-threshold* SPMs based on a continuous formulation of kappa and Jaccard statistical indexes. All the experiments were carried out on real fMRI data. The dataset represents well the typical clinical data: the subjects were real neurosurgery patients awaiting surgery for removal of brain tumors and they underwent widely used clinical tasks, including visual, cognitive and motor aspects.

Statistical Agreement: Confusion Matrix

The first method for assessing the agreement of activation maps proposed considers the statistical nature of the data. It analyzes the software packages as two classifiers using a confusion matrix that contains information about actual and predicted classifications made by the systems. Performance is commonly evaluated using data in the matrix with a set of specific indices, the most widely used is Kappa index [46]. Each SPM was normalized in relation to its own maximum value, hence each voxel assumes a value between 0 and 1. Let SPM_1 and SPM_2 be two normalized and Co-registered maps obtained from different analysis of the same fMRI exam. A set of quantization levels are defined from 0 to 1 and in each i, j position of the confusion matrix there is the number of overlapped voxels that belong to the i^{th} quantization level for SPM_1 and j^{th} quantization level for SPM_2. In this way, as in a classical confusion matrix, on the main diagonal there are the agreements and outside the diagonal the disagreements. Figure 4.11(a)(b) shows confusion matrixes, on 10 quantization levels, computed for examples in Figure 4.8. Following the classical way to assess statistical agreement, Kappa index



(a) (b)

Figure 4.9: Examples of Confusion Matrices a) Good Agreement b) Bad Agreement

is computed to evaluate the confusion matrix. For the first example K is 0.17 and for the second K is -0.01. Kappa Index takes values in the interval $[-1, 1]$. Considering

both Landis and Fleiss Kappa evaluation tables [93, 66] a poor agreement is found in both cases. In the second case the agreement is underestimated due to the statistical dependence of classes In general Kappa index can be computed if the following conditions are true:

- Analyzed instances are independent.
- Classifiers judge independently.
- Classes are mutually exclusive, exhaustive and independent.

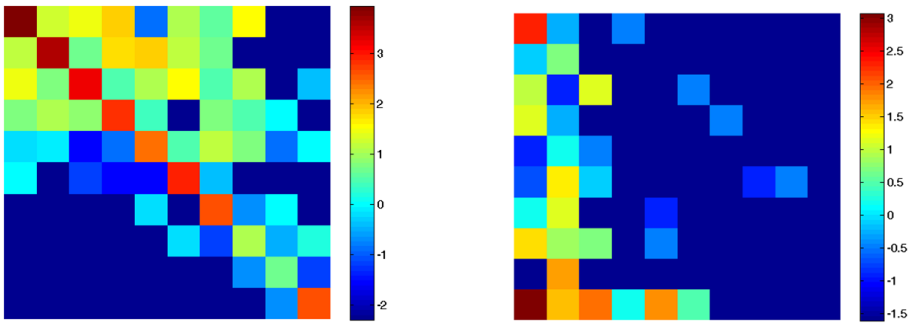
In this context, considering the way in which the confusion matrix is defined, the first two hypotheses are verified. Otherwise, the last condition is not true. This third condition allows the agreement to be computed if all the errors are equal. Indeed, Kappa index does not distinguish between major and minor errors. But, in this case, the misclassification between adjacent classes should weigh less in agreement evaluation than misclassification between far classes. The problem to weigh disagreement differently in the evaluation of the confusion matrix has been extensively studied in literature proposing new indexes such as Weighted Kappa [9, 47]. The method applied in this work takes into account statistical dependence between classes and computes for each class, i.e. in our context each level of quantization, a Mean Error based on the distance (ϵ) which takes into account the distance between the classes:

$$\epsilon_{Colj} = \frac{1}{C(j,j)} \sum_{i=1}^{N} \frac{|i-j|}{N-1} C(i,j)$$
$$\epsilon_{Rowi} = \frac{1}{C(i,i)} \sum_{j=1}^{N} \frac{|i-j|}{N-1} C(i,j) \tag{4.6}$$

where C is the confusion matrix built as detailed above and N is the number of the quantization levels. From the equations it is clear that the calculation of ϵ involve the application of a set of weights to the confusion matrix. The indexes ϵ_{Colj} and ϵ_{Rowi} can be interpreted as the Producer Accuracy and User Accuracy of a new confusion matrix C_w to which has been applied a matrix of weights through the element by element product: $C_w(i,j) = C(i,j) * W(i,j)$. The weight matrix is defined as follows:

$$W(i,j) = \begin{cases} 1, & if \quad i = j \\ \frac{|i-j|}{N-1}, & if \quad i \neq j \end{cases} \tag{4.7}$$

In this way the method takes into account classes dependence not defining new indices but weighing the confusion matrix and calculating traditional metrics of evaluation. In Figure 4.10(a)(b) weighted confusion matrixes computed for the examples above are



(a) (b)

Figure 4.10: Examples of Weighted Confusion Matrix a) Good Agreement b) Bad Agreement

shown. New Kappa values computed respectively are 0.61 and -0.08, that are consistent with visual inspection. It is worth noting that the application of the weight matrix does not imply an unconditioned amplification of the level of agreement. Indeed, if the agreement is under estimated because of the classes dependence the application of the weight matrix leads to an increase in agreement. Otherwise, where the two classifiers, are completely in disagreement, the application of the weight matrix does not change the comparison response.

Information Overlap

The SPMs Comparison method based on weighted confusion matrix, although very good because it honors statistical nature of the data, lacks in that it moves away from how the map is qualitatively evaluated: as a volumetric image. Moreover, the weighted confusion matrix puts little emphasis on what classes are in agreement. One way could be to observe the trend of the Producer and User Accuracy, but this loses the overall vision of the problem. For these reasons, the evaluation framework presented here provides a second method to evaluate SPMs agreement. As before we consider SPM_1 and SPM_2 both normalized and co-registered maps obtained from different analysis of the same fMRI exam. The similarity measure here presented is based on the Jaccard index [81].Computed as the ratio of the intersection cardinality the union cardinality of two binary mask. In Our approach intersection and union are computed using fuzzy logic definitions and $InformationOverlapIndex(IO)$ is defined as:

$$IO = \frac{SPM_1 \cap SPM_2}{SPM_1 \cup SPM_2} = \frac{1}{N}\sum_{i=1}^{N}\frac{min(SPM_{1,i}, SPM_{2,i})}{max(SPM_{1,i}, SPM_{2,i})} \tag{4.8}$$

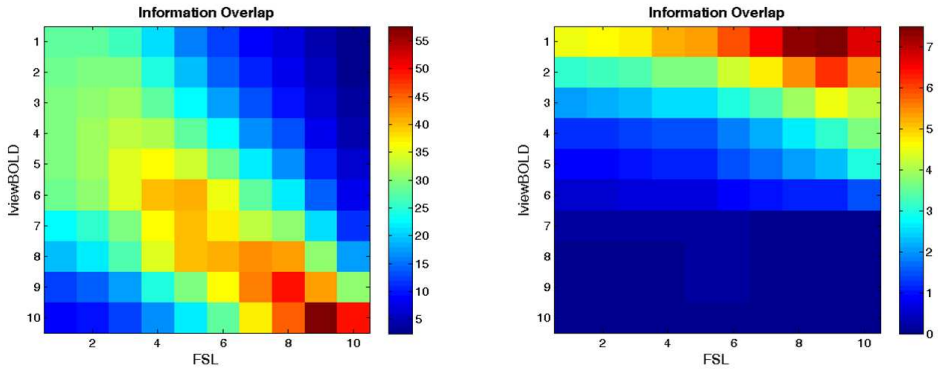
where N is the total number of voxels where $max(SPM_{1,i}, SPM_{2,i}) > 0$. This index gives a quantitative evaluation of behavior similarity of different tools. However, if SPMs are very noisy, IO can be "polarized" by the overlap of noise (voxels with low activation). For this reason, IO is computed as a function of different activation thresholds. Consider to obtain a binary mask M_{th} applying the threshold Th on the activation map SPM. The activation map SPM_{th} is computed applying the mask M_{th} to the original SPM:

$$SPM^{Th}(i,j) = \begin{cases} SPM(i,j), & if \quad SPM(i,j) \geq Th \\ 0, & if \quad SPM(i,j) < Th \end{cases} \tag{4.9}$$

Information Overlap matrix is computed as follows:

$$IO(j,k) = \frac{1}{N} \sum_{i=1}^{N} \frac{min(SPM_{1,i}^{Th_j}, SPM_{2,i}^{Th_k})}{max(SPM_{1,i}^{Th_j}, SPM_{2,i}^{Th_k})} \tag{4.10}$$

The IO Matrix values are high if the SPMs are in agreement. A joint variation of the two thresholds describe IO matrix from the high left corner to the low right one. IO matrix of two identical SPMs would be a perfect symmetrical matrix with all the values on the diagonal line equal to one. In real cases we can have IO matrixes approximating this one (Excellent Agreement). So peaks along the main diagonal line or parallel to it mean that activation areas are in the same position in the two SPMs and have similar statistical significance. Along the diagonal line, if a peak is on the right side it corresponds to the overlapping of more significant areas in the original SPMs. Otherwise, if it is on the left it corresponds to the overlapping of areas with less statistical significance. In all cases, if a peak is high SPMs overlap well (Good Agreement), if it is quite high SPMs have similar significance (Acceptable Agreement). If a peak dose not have a sym-



(a) (b)

Figure 4.11: Examples of Information Overlap(IO) Matrix a) Good Agreement b) Bad Agreement

metry in relation to the diagonal line it means that statistical relevance of the activation areas is different in the two SPMs (Poor Agreement). IO matrixes which exhibit no peaks describe SPMs that, even if partially overlapping have different clinical meanings (Disagreement). Maps obtained from tasks that involve language necessarily involve also the visual cortex, this side effect is caused by as the stimulus is produced showing images to the patient. We consider that the overlap of these areas is not significant in evaluating maps clinical agreement. For this reason IO matrixes are also evaluated eliminating data from voxels known to belong to visual functions (from a registration with Juelich Functional Atlas [179]). So we could also evaluate the concordance of the two pipelines for small functional areas, as the visual area is always a big and bright one, when stimulated by our paradigms. It is worth noting that the registration process allows removal of activation of functional areas that are noise for the evaluation process (like the visual cortex in our application); or to assess agreement between the maps separately for different functional areas of interest. Figure 4.11 shows the IO matrixes for the above examples with 10 threshold values.
If overlap is really related to information and not to noise, IO increases on the main diagonal jointly with the threshold. This surface gives a quantitative description, therefore allowing a comparison of different experiments. If the tools lead to the same conclusions about brain activation, the surface has an unmistakable pattern similar to a comet as is shown in Figure 4.11(a). Otherwise, IO matrix shown in Figure 4.11(b) looks completely different and has much smaller values of agreement. This representation, differing from that made by the confusion matrix, highlights whether the agreement is on the high or low values of activation. To summarize IO matrix characteristics in an only figure of merit, a grid with 5 scores from -2 to 2 was defined based on the previous observations:

Score 2: Excellent Agreement Plot symmetric in relation to the diagonal line or to a line parallel to diagonal one

Score 1: Good Agreement Well defined and high absolute maximum

Score 0: Acceptable Agreement Well defined and quite high absolute maximum

Score -1: Poor Agreement Not well defined absolute maximum, monotone decreasing plot

Score -1: Disagreement Quite uniformly flat plot

Despite the judge being subjective, it is driven by clear rules. Moreover, this figure is founded on bidimensional image analysis typical of human vision. As a result, IO matrix assessment is easier and less error prone than direct SPMs comparison with the same meaning.
Considering the fact that the analysis of IO matrix characteristics is accomplished

through an approximate evaluation of qualitative visual features, it was decided to design a fuzzy production system that could formalize and then automatically reproduce the expert decision attitudes [109]. We consider the following Visual Features computed on the IO matrixes normalized to 100 (SPM_{100})

M: Maximum max value of IO matrix:
$Max = max(SPM_{100})$.

I: Integral sum of all values in the SPM_{100}:
$I = \sum_{i=1}^{N_q} \sum_{j=1}^{N_q} SPM_{100}(i,j)$ where N_q is the number of the quantization levels.

DFD: Distance From the Diagonal is the distance between the point that contains the maximum value and the matrix main diagonal. Let i_{max} and j_{max} be the row and column of the maximum value the Distance From the Diagonal is $DFD = |i_{max} - j_{max}| \cos(\frac{\pi}{4})$

DL: Diagonal Length is the distance between the point that contains the maximum value and the left top corner $DL = \sqrt{i_{max}^2 + j_{max}^2}$.

The overall IO matrix evaluation process was then modeled in terms of fuzzy relationships between visual features and levels of agreement and is organized in three conceptual steps:

- Linguistic labeling of Visual Features and Classes/Levels of Agreement.
- Diagnostic rules definition.
- Inference for class assignment.

We formalize these concepts using fuzzy declarative propositions of the form X is A to represent the linguistic description of visual features (referenza). X is a linguistic variable denoting a given feature such as, for example, *Distance from the diagonal*(DFD); A is a term, such as *Low* or *Medium* or *High*; it belongs to a given term set and represents a fuzzy set in a given universe of discourse U characterized by a membership function $\mu_{Low,Med,High} : U \to [0,1]$. U contains all the possible numerical values assumed by X. Letting u belonging to U, a value associated with the visual feature DFD $\mu_{Low}(u)$ is interpreted as the degree of possibility of the expert describing the feature concerned with the term *Low*, or in other words, the degree of compatibility between the numerical value u and the linguistic term *Low*. The 5 Classes of Agreement listed above are modeled as crisp classes. All the linguistic variable (M, I, DFD and DL) are normalized in a unit range $[0,1]$, the associated term sets are *Low, Medium and High* and the the membership functions of the corresponding fuzzy sets are three Gaussian $G(\mu, \sigma)$ with mean equal to 0,0.5 and 1 respectively and standard deviation equal to 0.2.

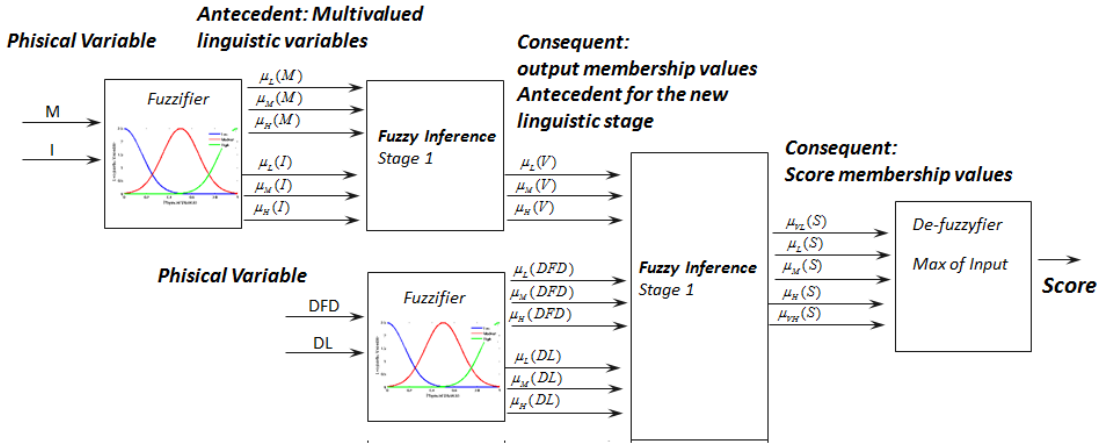


Figure 4.12: Scheme of the Fuzzy Recognition System

Fuzzy production rules describing how combinations of visual features, expressed in terms of multiple fuzzy declarative propositions, relate to Classes of Agreement, have the following general form:
The theoretical operator *min* proposed in the original formulation of fuzzy set theory is chosen for implementing the aggregation connective AND in the antecedent of rules. [58]

In the fuzzy framework IO matrix evaluation is modeled as a classification process and consists in the deduction of a conclusion regarding a given matrix, the premises of which are the values of visual features and the classification rules in the knowledge-base. As in any other logic, a rule of inference governs this deduction. When specific values are provided in input the fuzzy logic inference mechanism interprets the set of fuzzy production rules and deduces the final diagnostic results. A well-known deductive paradigm is the MAX-MIN method provided by Mamdani and Assilian [108]. The MAX-MIN method tests the magnitudes of each rule and selects the highest one.
Table 4.1 lists the complete set of rules elicited from experts indicating for each rule the antecedent and consequent part. A first set of rules Table 4.1(a) is used to aggregate visual features M and I and to infer intermediate variable IO_{Val} that combined with the other features DL and DFD allow to infer the Class of Agreement *Score* using the set of rules in Table 4.1(b). The classification process is performed choosing the class with the maximum value of plausibility. In Figure 4.12 the scheme of the overall procedure is shown.

Table 4.1: Rules elicited from the experts a First Stage: $(I, M) \rightarrow IO_{val}$, (b) Second Stage: $(IO_{val}, DFD, DL) \rightarrow Score$

(a)

Maximum	Integral	IO_{Val}
Low	Low	**Low**
Low	Medium	**Low**
Low	High	**Medium**
Medium	Low	**Low**
Medium	Medium	**Medium**
Medium	High	**High**
High	Low	**Medium**
High	Medium	**High**
High	High	**High**

(b)

IO_{Val}	DFD	DL	Score
Low	Low	Low	**Very Low**
Low	Low	Medium	**Very Low**
Low	Low	High	**Low**
Low	Medium	Low	**Very Low**
Low	Medium	Medium	**Very Low**
Low	Medium	High	**Low**
Low	High	Low	**Very Low**
Low	High	Medium	**Very Low**
Low	High	High	**Low**
Medium	Low	Low	**Low**
Medium	Low	Medium	**High**
Medium	Low	High	**Very High**
Medium	Medium	Low	**Medium**
Medium	Medium	Medium	**Medium**
Medium	Medium	High	**High**
Medium	High	Low	**Medium**
Medium	High	Medium	**Medium**
Medium	High	High	**Medium**
High	Low	Low	**High**
High	Low	Medium	**Very High**
High	Low	High	**Very High**
High	Medium	Low	**Medium**
High	Medium	Medium	**High**
High	Medium	High	**Very High**
High	High	Low	**Medium**
High	High	Medium	**High**
High	High	High	**High**

4.2.3 Activation Weighted Vector for SPM Synoptic Description

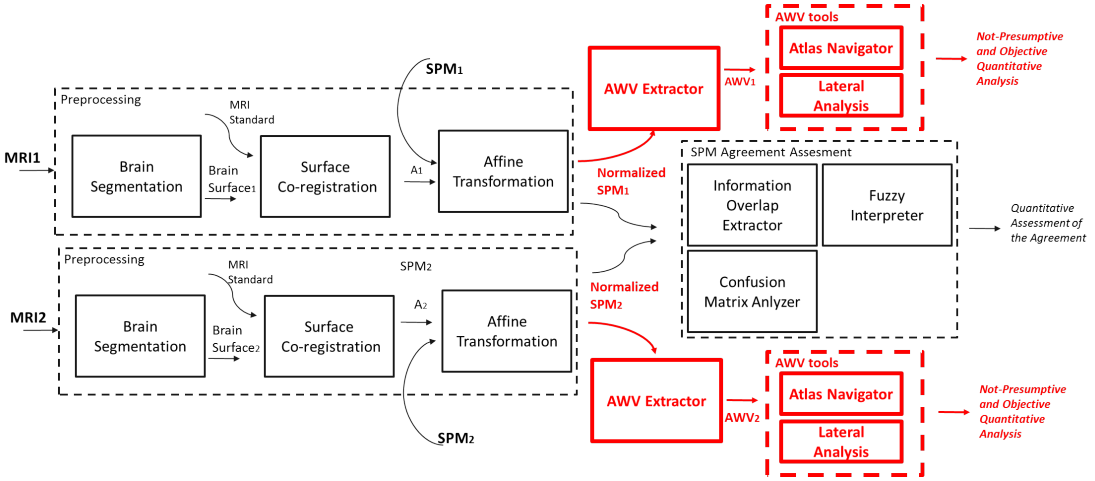


Figure 4.13: Schema of the Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM) with highlight the part detailed in Section (4.2.3)

The objective of this step is to briefly describe the volumetric map in an attempt to give to the radiologist a synoptic view of the SPM in a simply "features vector". For this purpose an index that describe the information contained in the map was studied (Figure 4.13). After brain segmentation and registration of patient brain on a gold standard, for which the labeling of each voxel with respect to an anatomical atlas is known, it is easy to compute the *Activation Weighted Index (AWI)* that summarizes the activation of the specific brain portion. The SPM is normalized with respect to its own maximum and each active voxel is assigned to the relevant atlas anatomical area. The *AWI* is computed as follows:

$$AWI_J = \frac{1}{N_J} \sum_{i=1}^{N_J} w_i \tag{4.11}$$

where N_J is voxel total number in the J^{th} area and w_i is the normalized voxel value in the J^{th} area. *AWI* value can vary between zero and one and has greater value in areas where activations are clustered. Collecting these indexes for every zones a feature vector can be defined. *AWV (Activation Weighted Vector)* is brief but comprehensive and can describe quickly and easily the SPM.

Figure 4.14 shows an example of SPM extracted form the execution of a Word Generation task, the map includes widespread activations in both hemispheres with high levels of noise, that create difficulties in determining which activities are significant and which are merely acquisition artifacts. Moreover, the volumetric nature of the data makes it

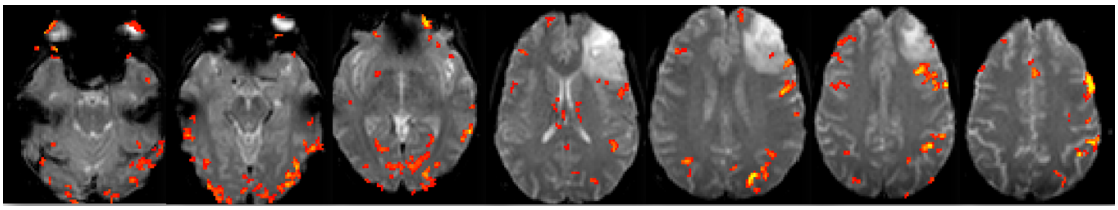


Figure 4.14: Statistical Parametric Map of a Word Generation task

difficult to assess the general activation distribution and the radiologist must proceed through each SPM axial plan. In Figure 5.7 the *Activation Weighted Vector* computed

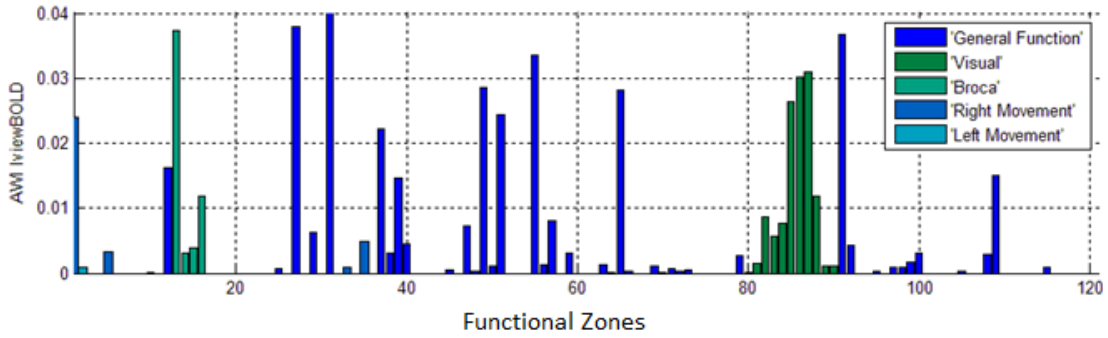


Figure 4.15: *Activation Weighted Vector* of a Word Generation task

for the above example is shown. Each bar in the histogram represents an anatomical zone in the atlas chosen as reference. This brief representation gives a synoptic view of all the Map, giving to the expert a complete overview of the SPM. The *Activation Weighted Vector* can be used to recognize particular patterns, extract indexes, as the hemisphere dominance, or as a guide to the full exploration of the map. The proposed Integrated Framework for Computer Assisted Analysis of SPMs includes several AWV tools; one of these, called Atlas Navigator, allows to link the original SPM with the AWV value as shown in Figure 4.16. By this way the physicians can use the AWV as a guide for explore the volumetric SPM with the guarantee of focus on the functional areas that are actually related with the performed task and not to be influenced by what one expects to happen. The same analysis carried out on the raw data would be challenging time consuming. Analyzing the activation map by means of *AWV*, makes the analysis more objective, avoiding a presumptive approach in which the expert only verifies a priori hypothesis of activation. Qualitative analysis can not be exhaustive on the map; thus the expert will intend only to verify that the activation occurred in areas classically related to the task performed. But it is now known that the brain is a plastic organ particularly in presence of diseases. Moreover, functional areas locations are very individual-dependent. In these terms comprehensive and quantitative analysis of the

map is essential.

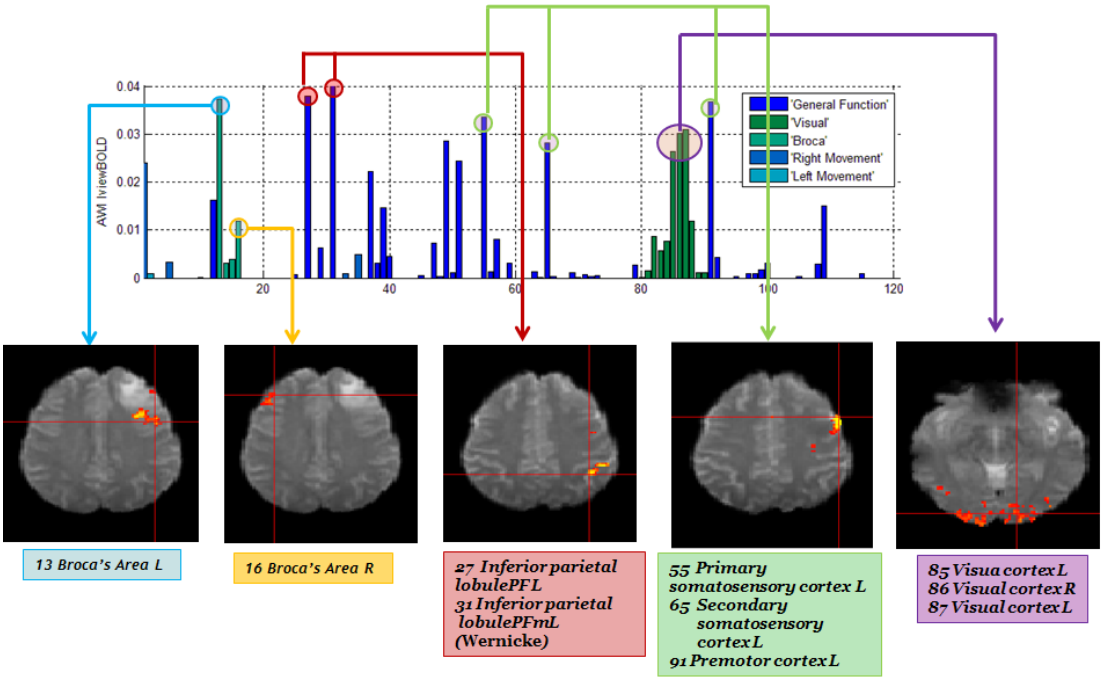


Figure 4.16: Example of Activation Weighted Vector Use for the Statistical Parametric Map exploration

4.2.4 Hemisphere Dominance Evaluation by using fMRI Activation Weighted Vector

One of the challenging goal of the fMRI is to predict which of the two Cerebral Hemispheres is dominant with respect to a specific task. the proposed framework includes a specific tool for the hemispheric evaluation starting from the AVW description of the SPM. Before the use of fMRI the hemispheric dominance evaluation for the presurgical planning involves the use of invasive techniques as the Wada test to identify the hemisphere of the areas of language [34]. fMRI allows to obtain similar results in a non-invasive manner and an index of hemispheric dominance has been introduced (Lateralization Index)[149]:

$$LI = 100 * \frac{Q_l - Q_r}{Q_l + Q_r} \tag{4.12}$$

where Q_l and Q_r are representative quantities measured by fMRI for the left and right contributions, respectively. The variability range of the index is [-100,100], it is positive and close to 100 if the right hemisphere is dominant and it is negative and close to -100 if the right hemisphere is dominant, values close to zero indicate bilateral brain.

Recently the use of this index has highlighted some limitations essentially due to its low reliability causing artifacts and to a new definition of hemispheric dominance that makes the test inadequate. The concept of hemispheric dominance stems from the hypothesis of functional asymmetry identified by Broca in 1861. Systematic studies on the subject begun in the sixties. Initially the left hemisphere was considered dominant, indeed it is the sites of language and reasoning in right-handed. fMRI studies allow to exploit new complex aspects. Lateralization is less clear (the areas of language are mainly localized in the left hemisphere for about 95% of the right-handed and for about 75% of the left-handed), and both hemispheres contribute to brain functions with interpersonal variability. Moreover, for different functions there may be a different degree of lateralization. Currently, hemispheric specialization concept, (functional difference detected in the hemispheres with certain tasks and the subsequent specialization in certain skills) is preferred.
The brain is considered *'asymmetric but integrated'*. Hemispheric specialization is mostly recognized and accepted because supported by several studies, but at the same time the importance of cooperation between the hemispheres in a healthy brain it is emphasized. In the presurgical studies is essential to identify topologically and quantitatively, the location of eloquent areas in order to preserve them as far as possible. A complete analysis requires the use of DTI in an attempt to identify the white matter bundles and the corresponding networks.
Our goal in this phase is to assess hemispheric dominance for different functional areas, emphasizing the importance of the activation clustering with respect to the singular activation value. The method makes use of the concise description of the SPM obtained using the AWV. The AWV is splitted in two vectors AWV_l and AWV_r considering the activation of only the left and right hemisphere respectively as shown in Figure 4.17 The hemispheric specialization can be simply evaluated as follows:

$$LI_{AWV} = 100 * \frac{Q_l^{AWV} - Q_r^{AWV}}{Q_l^{AVW} + Q_r^{AWV}} \tag{4.13}$$

$$Q_l^{AWV} = \sum_{i \in AWV_l} AWI_i; \quad Q_r^{AWV} = \sum_{i \in AWV_r} AWI_i$$

For the example in Figure 4.14 the LI_{AWV} value is 60.4 indicating a clear overall dominance of the left hemisphere; this quantitative assessment agrees with that made in qualitative way by looking at the map. It is worth to note that the calculation of hemispheric dominance as here proposed is very different from the classical one in which the left and right hemisphere activations are directly compared. LI_{AVW} allows to attribute less importance to isolated active voxels which are probably artifacts and gives signif-

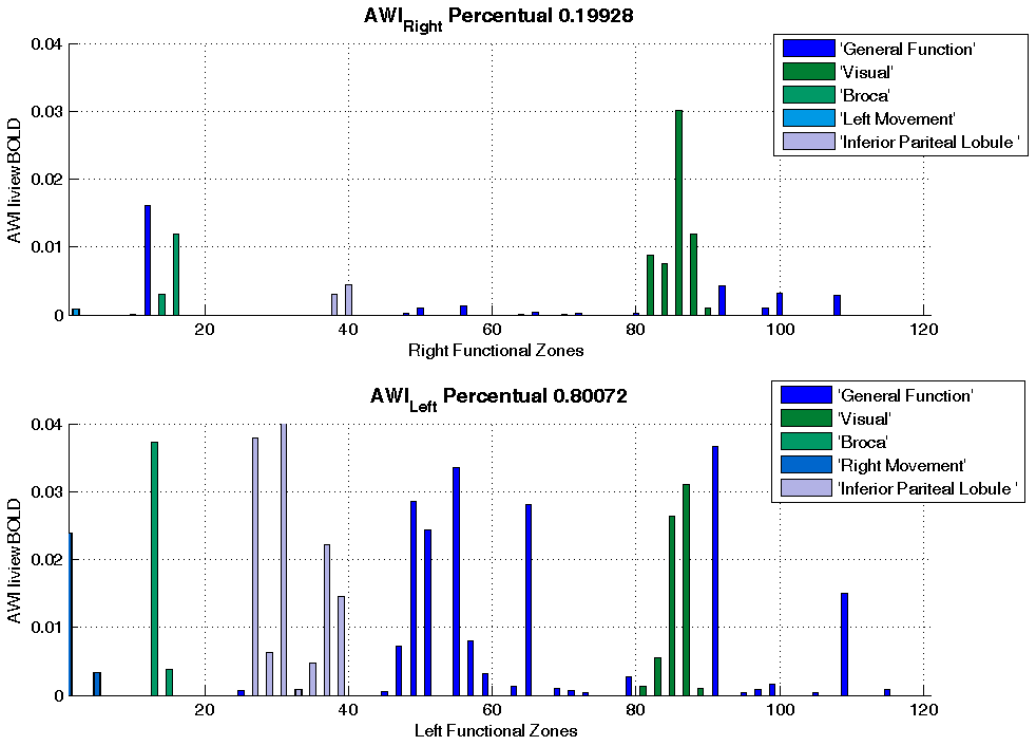


Figure 4.17: AWV Left and Right splitting

icance to voxels belonging to clusters. Indeed, in the example above, the classical LI doesn't detects a predominant hemisphere (LI = 0.86) due to the accumulation of noise voxels in the right hemisphere. Moreover, LI_{AVW} can be computed also to evaluate a quantitative hemispheric specialization considering separately the brain areas classically linked to some of the more relevant functions.

Figure 4.18 shows an example of the hemispheric dominance evaluation performed using the AWV. The patient is left handed and has an expansive mass in the right parietal lobe. The example in Figure 4.18 is related to the Word generation task. The SPMs obtained are very noisy and difficult to be interpreted. The qualitative analysis of the SPM performed by a radiologist didn't highlight a clear hemispheric dominance in this language task; thus, the patient is judged with bilateral dominance. The AWV analysis allows to quantify the observation and to study the SPM comprehensively, faithfully and quickly. As regards the Broca's area the right hemisphere is dominant ($LI_{AWV}^{Broca} = -50.2$). Instead, as regards the Wernicke's area, located in the inferior parietal lobe, the left hemisphere is dominant ($LI_{AWV}^{Wernicke} = 80.97$). The AVW automatic analysis agrees with the expert comments, with the additional merit of giving

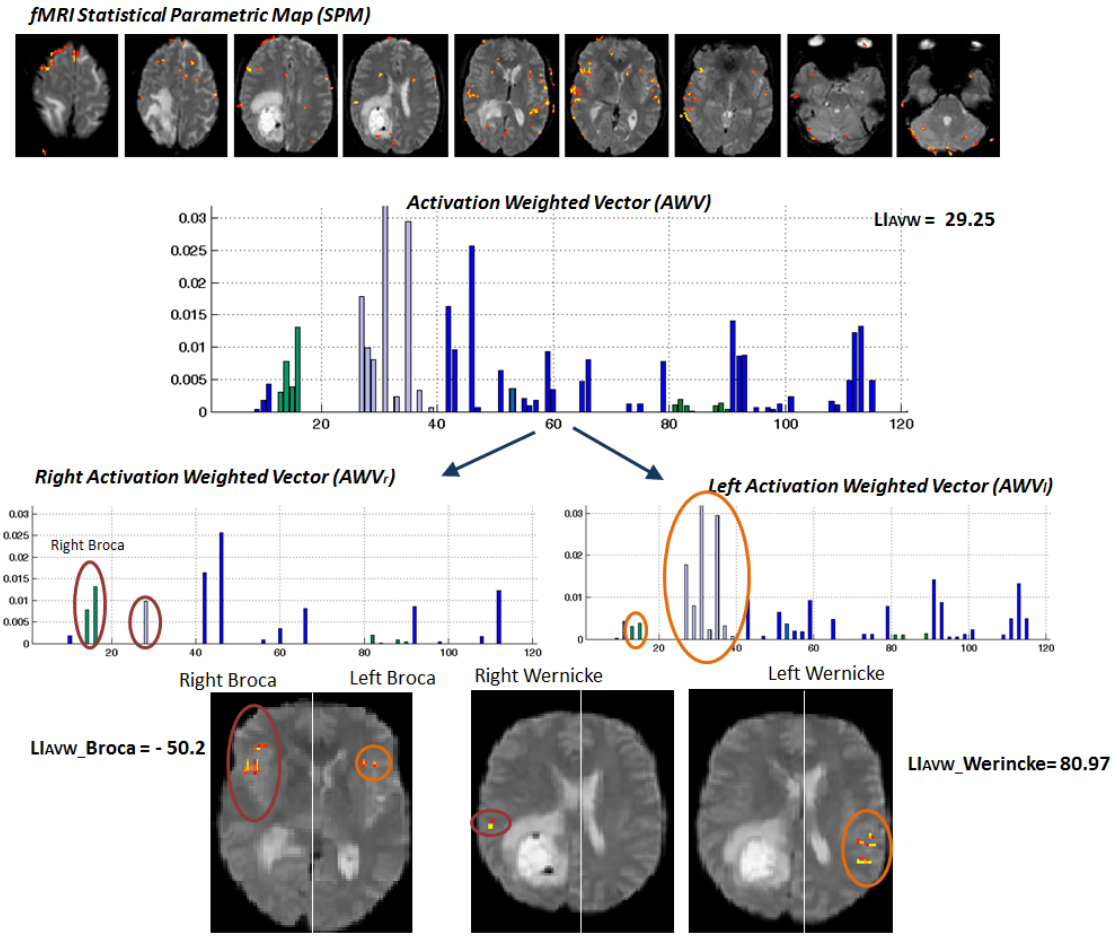


Figure 4.18: Example of the Hemispheric Dominance evaluation using AVW: Verb Generation task

a quantitative evaluation thus, comparable with other cases or usable for the clinical follow-up.

5

fMRI Analysis: Experimental Results

5.1 Evaluation of Integrated Framework for Computer Assisted Analysis of fMRI SPM

The aim of this chapter is to evaluate the proposed Integrated Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM). Where it was possible we tried to make objective the assessment, in other points we only report some examples and use cases.
In the following sections the datasets used in this experimental phase and the acquisition modality are detailed.

5.1.1 Stimulation paradigms

Stimulation paradigms are those most often used in neuroradiology: motor (finger tapping) and language (verb generation, word generation, object naming) tasks. All tasks are block designs.
The *Finger tapping* task is a monolateral thumb- forefinger movement of 9 total blocks: 4 times OFF-ON plus a last OFF block. Each block lasts 30 seconds, where 10 volumes are acquired. The patient is vocally instructed by the operator about when to perform the activity.
Object naming is performed showing simple color images (ON block) alternating a white cross on a black background (OFF block). During the viewing of the image (an image every 3 seconds) the patient has to think of the word representing the object shown.

During the OFF block the patient has to think nothing. Each block lasts 30 seconds, where 10 volumes are acquired. The block sequence is: 4 times OFF-ON plus a last OFF block.
Verb generation is performed showing simple black and white drawings representing an action (ON block) alternating a white cross on a black background (OFF block). During the viewing of the drawings (an image every 3 seconds) the patient has to think of the verb representing the action shown. During the OFF block the patient has to think nothing. Each block lasts 30 seconds, where 10 volumes are acquired. The block sequence is: 6 times OFF-ON plus a last OFF block.
Word generation is performed showing single white letters on a black background (ON block) alternating a white cross on a black background (OFF block). During the viewing of the letters (a letter every 6 seconds) the patient has to think of all the words beginning with that letter that he/she can. During the OFF block the patient has to think nothing. Each block lasts 30 seconds, where 10 volumes are acquired. The block sequence is: 6 times OFF-ON plus a last OFF block.

5.1.2 MRI images acquisition

During the tasks, the brain volume was acquired using standard EPI BOLD acquisition sequence (TR 3000ms, TE 50 ms, FA 90, matrix 128 x 128, FOV 230 mm, axial orientation, slice thickness 4 mm, spacing 4 mm, number of slices 30). For anatomical image reference, a T1-weighted volumetric sequence was chosen (TR 7113 ms, TE 3208 ms, FA 8, matrix 256 x 256, FOV 256 mm, sagittal orientation, slice thickness 1 mm, spacing 1 mm, number of slices 180).

5.2 SPM Comparison Tools Evaluation

5.2.1 Compared SW pipelines

From the acquired volumes (90 for finger tapping and object naming, 130 for verb generation and word generation), activation maps were calculated as statistical parametric maps (SPMs) through two different pipelines.
The first pipeline used IViewBOLD software supplied by Philips®. Statistical analysis was carried out applying t-test to each voxel time-course, after retrospective image registration for motion correction. SPMs are function of t, minimum t value 3, maximum t value as results from analysis. Contiguous voxels = 4 and default mask were set.
The second pipeline used FMRI Expert Analysis Tool (FEAT) included in FMRIB Software Library (FSL) [179], which uses univariate General Linear Model (GLM) for voxel time-course analysis [120]. Firstly brain was extracted from an anatomical reference sequence through Brain Extraction Tool [155]. Then FEAT was launched with parameters

set as detailed in Appendix I. SPMs are function of z, minimum and maximum z values as result from analysis.

5.2.2 Patients

We used images of 15 consecutive patients, who underwent different fMRI procedures, for a total of 49 tasks acquired (13 finger tapping, 13 object naming, 12 word generation, 11 verb generation). The patients underwent fMRI for clinical reasons, mainly pre-surgical for oncologic illness.

5.2.3 Experimental Results of the SPM Comparison Tools

This section presents our results with two aims. The first aim is to evaluate our method to assess the agreement of $SPMs$. The second aim is to evaluate agreement between the results obtained by the fMRI time series analysis software tools considered. Concerning the first purpose, five examples of agreement evaluation, for both confusion matrix and Information Overlap analysis, are detailed. The shown examples are presented in order of increasing agreement, with a score from -2 to 2. The quantization step for both IO and confusion matrix is 0.1. For each example a representation of SPMs overlap is shown. Figures 5.1,5.2,5.3,5.4,5.5 show in order: SPMs overlap, confusion matrixes before and after the correction using a logarithmic color scale and IO matrix using the same color scale for each example.

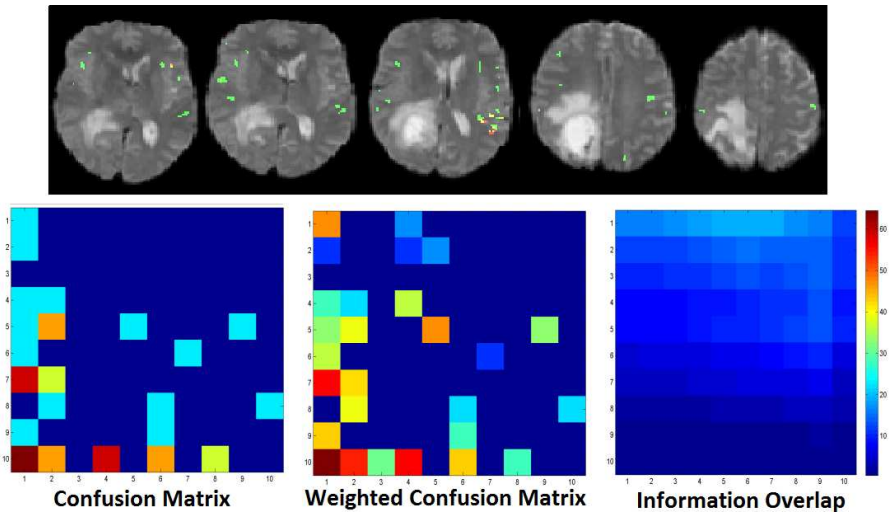


Figure 5.1: Agreement Analysis Example, IO Score -2

Figure 5.1 shows the agreement analysis of a Verb generation task. Both softwares tools have identified a poor activation with few overlap points focused in the inferior

parietal lobule zone. The first software tool detects activations only in the left hemisphere. Otherwise the second processing detects some contralateral activations. Kappa indexes computed on confusion matrix and on weighted confusion matrix are 0 and 0.08 respectively, which indicate a complete disagreement. It is worth that these values are computed without considering agreement in class 0, which would lead to an overestimation of agreement. Score -2 has been attributed to the IO matrix. It is quite uniformly flat with a small maximum value that is located far from the main diagonal line. Quantities that describe the IO matrix have the following values: $M = 19.79$, $I = 824.49$ thus with mean value 8.24, $DFD = 5$, $DL = 6.08$.

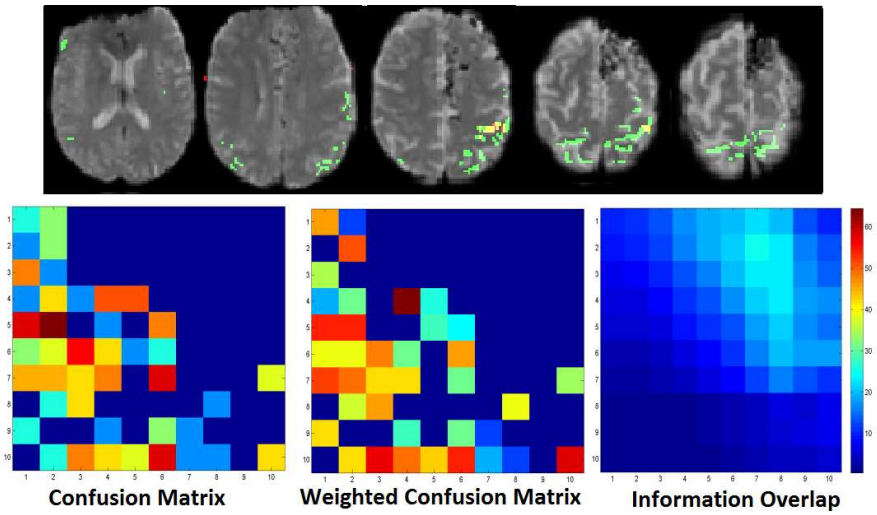


Figure 5.2: Agreement Analysis Example, IO Score -1

Figure 5.2 shows the agreement analysis of a Right Finger Tapping task. Activations are mainly in the left hemisphere for both processing. First the software detects activations focused in a small and specific area, otherwise there is a greater dispersion as regards the second processing. Kappa indexes computed on the confusion matrix and on weighted confusion matrix are 0.04 and 0.22 respectively. Although they are higher than in the previous case, they indicate little agreement. Qualitative analysis of the maps leads us to believe that one of the software is more conservative, but through the analysis of confusion matrices we can not verify or refute this thesis. Otherwise, analysis of IO matrix clarifies this doubt. Indeed, if the thesis is verified, IO matrix increases on the main diagonal line jointly with the threshold. The location of maximum far from the diagonal line tells us that the most activated voxels do not overlap. Score -1 has been attributed to the IO matrix. Absolute maximum is not well defined, and the IO matrix is monotone decreasing. Quantities that describe the IO matrix have the following values: $M = 25.02$, $I = 1046.30$ thus with mean value 10.46, $DFD = 5$, $DL = 7.28$.

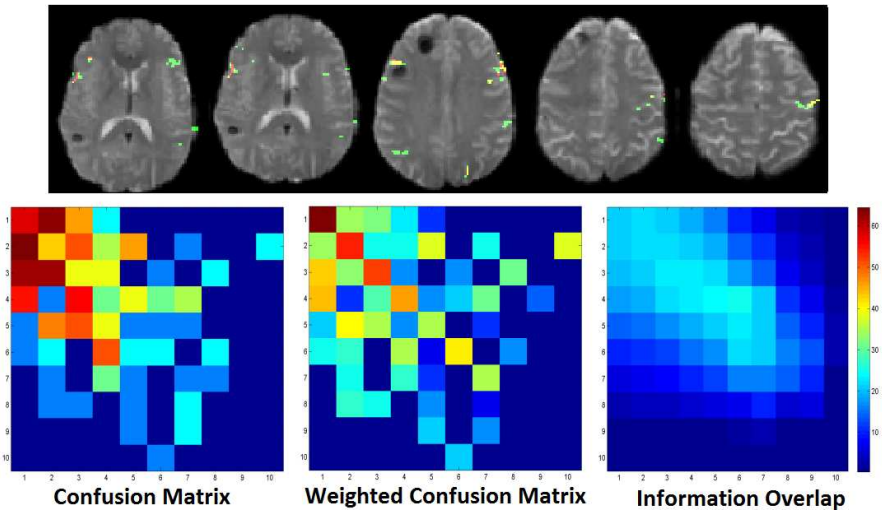


Figure 5.3: Agreement Analysis Example, IO Score 0

Figure 5.3 shows the agreement analysis of an Object Naming task. For both the processing, activations are focused in Broca areas and a contralateral activation is noticed. Activation area in the left hemisphere detected by the second software is widest. Qualitative analysis leads to the conclusion that the maps are quite overlapping. Kappa indexes computed on confusion matrix and on weighted confusion matrix are -0.02 and 0.37 respectively. The IO matrix looks not very different from the previous case because maximum and integral are similar. The main difference is the lower distance between maximum and main diagonal line, that indicates a better overlap. Score 0 has been attributed to the IO matrix. Quantities that describe the IO matrix have the following values: $M = 25.02$, $I = 1083.84$ thus with mean value 10.82, $DFD = 2$, $DL = 7.21$.

Figure 5.4 shows the agreement analysis of a Verb Generation task. For both the processing activations are focused in Broca and Wernicke areas in the left hemisphere; both software tools detected a small contralateral activation. Qualitative analysis of the maps leads to judge the agreement good; despite the fact the maps are not identical the conclusions drawn from the analysis would probably be the same. Kappa indexes computed on confusion matrix and on weighted confusion matrix are respectively, -0.06 and 0.47. It is worth noting that, in the cases where the agreement is higher, to weight the confusion matrix becomes essential for the correct analysis. Score 1 has been attributed to the IO matrix. The high absolute maximum is well defined and is located on the main diagonal line, IO matrix is not evaluated with score 2 because maximum value is located at the level of quantization 0.6. An excellent IO matrix has a maximum located in the high levels of quantization. Quantities that describe the IO matrix have

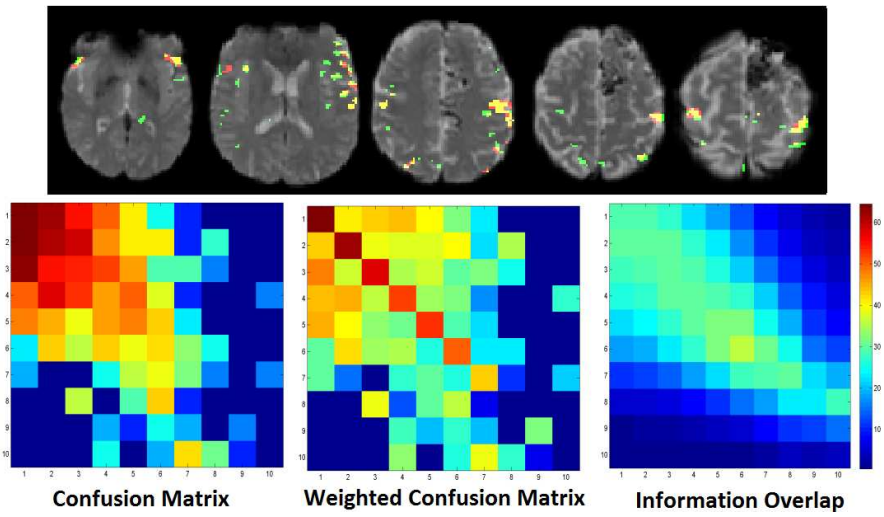


Figure 5.4: Agreement Analysis Example, IO Score 1

the following values: $M = 36.73$, $I = 1701.14$ thus with mean value 17.01, $DFD = 0$, $DL = 8.49$.

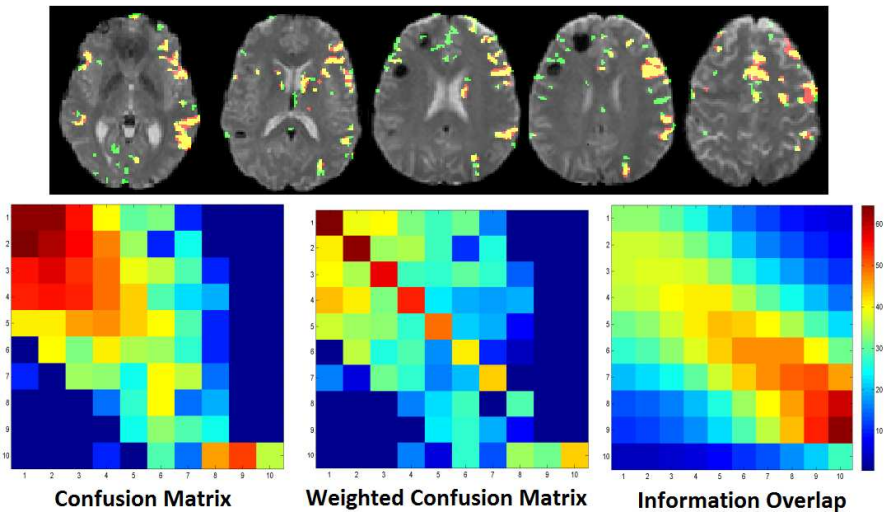


Figure 5.5: Agreement Analysis Example, IO Score 2

Figure 5.5 shows the agreement analysis of a Word Generation task. There are many active areas. Indeed, it is difficult to identify an activation core. SPMs are overlapped on every zone of accumulation of active voxels. Disagreement points are isolated and appear to be noise. Qualitative analysis of the maps leads to judge the agreement excellent. Kappa indexes computed on confusion matrix and on weighted confusion matrix

are respectively, 0.08 and 0.52. Considering Landis Kappa evaluation tables agreement is only Moderate. It seems that the Kappa evaluation method tends to underestimate the qualitative agreement impression. Otherwise, IO matrix is evaluated with the highest score 2. The maximum value is high, well defined and located near the main diagonal line at the quantization level 0.9. Quantities that describe the IO matrix have the following values: $M = 63.39$, $I = 2920.37$ thus with mean value 29.37, $DFD = 1$, $DL = 13.45$.
In the second phase of our experiments the 49 fMRI exams mentioned above were considered. Both the Kappa and IO Matrix evaluation were performed. The SPMs agreements were assessed using the confusion matrix and an observer attributed scores to all the IO matrixes twice (to verify the reproducibility of the attribution). All the observed IO matrix were displayed with the same fixed color scale. The same IO matrixes are evaluated using the fuzzy system.
In Table 5.1 all the results are reported. For each task the quantities describing IO matrix used as linguistic variable and confusion matrix are detailed.
For the IO matrix:

- maximum value (**M**)
- total surface integral (**I**)
- horizontal distance of maximum from diagonal line (**DFD**)
- diagonal length (**DL**)
- mean of the scores assigned in both subjective evaluation of the IO the matrix (**Score**).
- mean scores automatically assigned to the IO matrix by the fuzzy system (**Fuzzy Score**).

For the confusion matrix:

- classical Kappa, evaluated with threshold 0.1 (**K**)
- Kappa including dependent classes correction, evaluated with threshold 0.1 (**WK**)

IO matrix description quantities were not known to the observer attributing subjective scores during his assessing process. Linear correlation coefficient between "classical" Kappa and subjective score is $R = 0.55$. Linear correlation coefficient between corrected Kappa and subjective score is $R = 0.75$.
So we can say that IO matrix comparison method correlates well with well established statistical evaluation methods as Cohen's Kappa one. Moreover, the correction made to account for partial dependence of classes is valid, and causes a better correlation between this statistical method and a visual one. Moreover, our results show that the

Table 5.1: IViewBOLD and FEAT FSL Activation Map Agreement Study

Case	Task	M	I	DFD	DL	Human Score	Fuzzy Score	K	WK
1	Object Naming	36.50	1494.37	0	11.31	0.5	2	0.04	0.47
2	Word Generation	30.98	1579.36	2	3.16	-0.50	-1	0.03	0.23
3	Verb Generation	26.61	1325.93	0	2.83	0.00	-1	0.04	0.43
4	Object Naming	37.17	1928.36	2	7.21	1.00	0	0.03	0.36
5	Word Generation	30.55	1429.11	7	8.06	-1.00	0	0.00	0.10
6	Verb Generation	37.60	1743.06	2	10.00	1.00	1	0.05	0.44
7	Verb Generation	33.58	1775.96	1	5.00	0.00	0	0.09	0.52
8	Object Naming	33.86	1690.39	1	3.61	0.5	0	0.06	0.46
9	Finger Tapping	42.24	2310.11	0	4.24	1.50	1	0.06	0.40
10	Finger Tapping	43.32	2390.59	1	6.40	1.50	1	0.04	0.36
11	Word Generation	39.22	1866.45	0	2.83	1.00	1	0.07	0.44
12	Object Naming	16.43	627.86	2	4.47	-2.00	-2	0.07	0.41
13	Finger Tapping	37.03	2221.04	1	12.04	0.50	2	0.04	0.25
14	Word Generation	18.19	600.72	6	10.77	-1.50	-1	-0.01	0.09
15	Verb Generation	19.79	824.49	5	6.08	-1.50	-2	0.00	0.08
16	Object Naming	27.72	1303.59	1	6.40	0.5	0	0.07	0,43
17	Word Generation	14.59	391.51	9	10.05	-2.00	-2	-0.05	-0.02
18	Verb Generation	35.36	2287.47	2	8.60	1.00	0	0.05	0.26
19	Verb Generation	26.98	729.81	9	10.05	-1.00	-2	0.02	0.04
20	Object Naming	35.15	1742.91	3	10.82	1.00	1	0.02	0.27
21	Word Generation	22.55	1043.57	3	4.12	-0.50	0	0.02	0.20
22	Finger Tapping	29.97	1702.37	1	10.63	0.00	2	0.08	0.47
23	Finger Tapping	47.92	1925.96	1	12.04	2.00	2	0.08	0.47
24	Verb Generation	8.05	311.92	2	11.40	-2.00	-1	-0.01	0.37
25	Object Naming	12.33	249.44	4	5.10	-2.00	-2	0.11	0.35
26	Word Generation	47.74	2879.23	2	11.40	2.00	2	0.07	0.40
27	Verb Generation	32.17	940.52	0	12.73	1.00	-1	0.10	0.50
28	Object Naming	25.02	1083.84	2	7.21	0.00	0	-0.02	0.33
29	Finger Tapping	6.42	201.07	4	5.10	-2.00	-2	-0.03	0.13
30	Finger Tapping	34.53	1861.89	3	4.12	0.00	0	0.03	0.21
31	Word Generation	63.39	2920.37	1	13.45	2.00	2	0.08	0.52
32	Object Naming	11.25	406.62	3	4.12	-2.00	-2	-0.01	0.25
33	Word Generation	26.96	1184.04	5	11.18	0.00	1	0.04	0.22
34	Object Naming	27.96	1239.38	0	9.90	1.00	0	0.10	0.53
35	Finger Tapping	25.02	1046.30	5	7.28	-1.00	0	0.04	0.22
36	Finger Tapping	13.23	638.51	5	7.28	-2.00	-2	-0.06	0.04
37	Word Generation	42.52	2332.12	3	12.21	1.00	1	0.11	0.46
38	Verb Generation	36.73	1701.14	0	8.49	1.00	0	0.06	0.47
39	Verb Generation	23.31	995.97	4	11.66	-1.00	1	-0.04	0.01
40	Object Naming	58.48	2656.55	1	13.45	2.00	2	0.27	0.49
41	Word Generation	38.26	2078.56	1	13.45	1.00	2	0.08	0.52
42	Object Naming	0.31	2.36	1	2.24	-2.00	-2	0.00	0.00
43	Finger Tapping	9.33	228.00	7	8.06	-2.00	-2	-0.01	0.08
44	Finger Tapping	6.49	118.10	9	10.05	-2.00	-2	0.07	0.14
45	Verb Generation	0.01	0.01	0	1.41	-2.00	-2	0.00	0.00
46	Finger Tapping	57.61	2455.89	1	13.45	2.00	2	0.17	0.60
47	Finger Tapping	73.92	3863.13	0	9.90	2.00	2	0.12	0.60
48	Object Naming	49.61	1722.27	3	10.82	2.00	1	0.04	0.36
49	Word Generation	41.27	2092.11	4	10.30	1.00	1	0.02	0.28

IO matrix evaluation performed using the fuzzy system has a high degree of agreement with that performed by the subjective analysis $R = 0.85$. In this way the analysis is objective despite the linguistic nature of the problem being respected. Thus, our method is repeatable and fully transferable to other contexts.
The mean scores attributed to *IO* matrixes of all the tasks analyzed, divided by task itself for the analysis with and without the visual cortex are reported in Table 5.2.
When the comparison of the two pipelines is performed including visual cortex, Table

Table 5.2: Mean scores by task

Mode	Finger Tap.	Object Nam.	Word Gen.	Verb Gen.
with visual cortex	0.2	1.3	0.5	0.5
without visual cortex	0.0	-0.2	0.1	0.4

5.2 shows that there is concordance between the results of the two pipelines. It seems that concordance is better for language tasks than for motor task. If visual cortex is excluded from the analysis, results are completely different. For all tasks concordance is much smaller, with scores almost always zero or negative. This shows that the previously observed concordance is almost always based on the concordance for the big and bright visual area. This is particularly true for language tasks, and this is logical: our language tasks are stimulated with visual stimuli. The difference is truly big for object naming (stimulated by color photos of objects) and verb generation (stimulated by hand-drawing of actions). For word generation, the stimulus is a letter on a black background, that is a simpler visual stimulus.
So if one wants to evaluate a coarse concordance between fMRI pipelines, he can include visual cortex. In view of a coarse evaluation, the two pipelines considered are in agreement. However, in our opinion, in a clinical context a fine concordance is necessary. In fact, positions of small language or motor areas are used in brain surgical planning. So we consider that the concordance of the two pipelines, to be real and useful, should last also without visual cortex. In this case, the concordance between the two considered fMRI pipelines vanishes. Finger tapping, known as a robust task, simple to perform for patients, is the only task obtaining a concordance evaluated as mild ("0" mean score).
Summarizing, in this experimental phase we presented a comparison analysis of two fMRI software tools: Philips IViewBOLD and FMRI Expert Analysis Tool (FEAT) included in FMRIB Software Library (FSL), obtaining a quantitative assessment of the inconsistencies between the SPM generated. The experimental analysis adopted two *non threshold* evaluation strategies based on standard indexes adapted to graduated "soft" data.
Results obtained in our clinical context substantiate the general claim that often same sequence, processed in different ways, gives different results. This effect has been observed and quantitatively analyzed in the specific relevant context of clinical surgery in

which variation in software require attention.

Our conclusions about the inconsistency of the software pipeline are not generalizable to other time series analysis methods, but the proposed comparison method can be applied to them. The strength of our work lies into the fact that the SPMs analyzed concern real patients and are extracted using the software tool embedded in acquisition machines widely used in clinical practice.

5.3 Activation Weighted Vector Evaluation

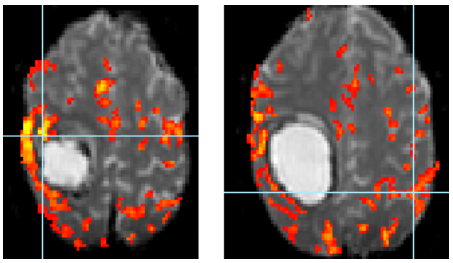


Figure 5.6: Example: SPM of 2 slices in hot scale (red to yellow) superimposed on MR anatomical images

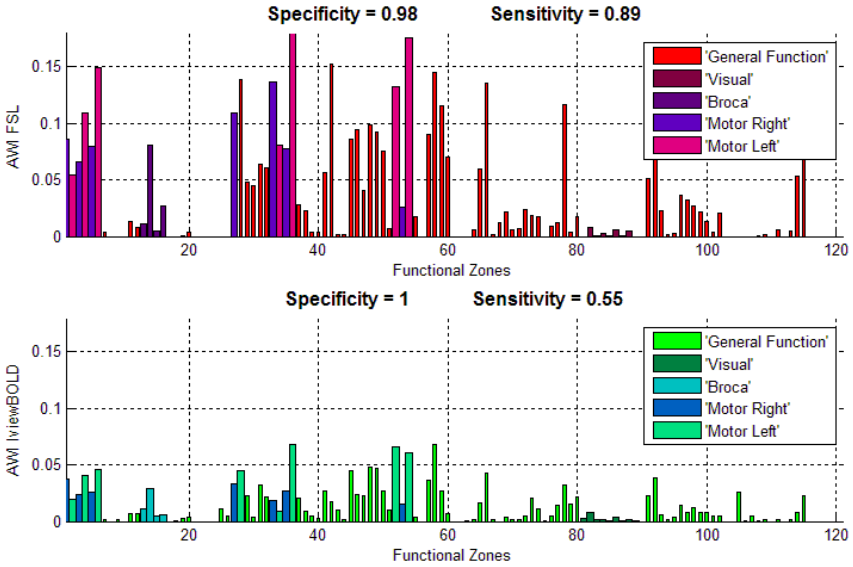


Figure 5.7: Example: SPM of 2 slices in hot scale (red to yellow) superimposed on MR anatomical images

In Figures 5.6 and 5.7 an example of AWI analysis is shown (left finger tapping). The patient has a tumor in the right hemisphere, precisely near the area that regards the movement of body left side. In Figure 5.6 two slices of SPM are depicted in hot scale (max: yellow, min: red). The first slice shows the biggest and the most intense activation core. A qualitative analysis of this SPM shows that the tumor has compressed the functional zone near the skull: this effect can be quantified with AWI analysis. In the second slice the contro-lateral activation is shown. Also in this case, a similar effect of the tumor on the allocation of the functional skill can be easily observed, but hardly measured. In Figure 5.7, AWI vector of FSL FEAT SPM and iViewBOLD SPM are shown. The briefness of AWI vector is its best feature. It can show all the information in a one dimensional vector. The comparison of the vectors with ground truth allows to evaluate software performance. Moreover, deductions made by observing the SPM can be quantified. Some observations can be made comparing the left finger tapping ground truth and the FSL FEAT AWI vector. Activation mean is similar and the maximum is in the same functional zone (36 GM Inferior parietal lobule PFT Right) respectively with value 0.19 and 0.18. On the other hand, the average activity of iViewBOLD is lower, even if the same pattern is recognizable. The maximum is in the same zone 36 with value 0.08.

Reference ground truth for each considered functional task was built considering AWIs

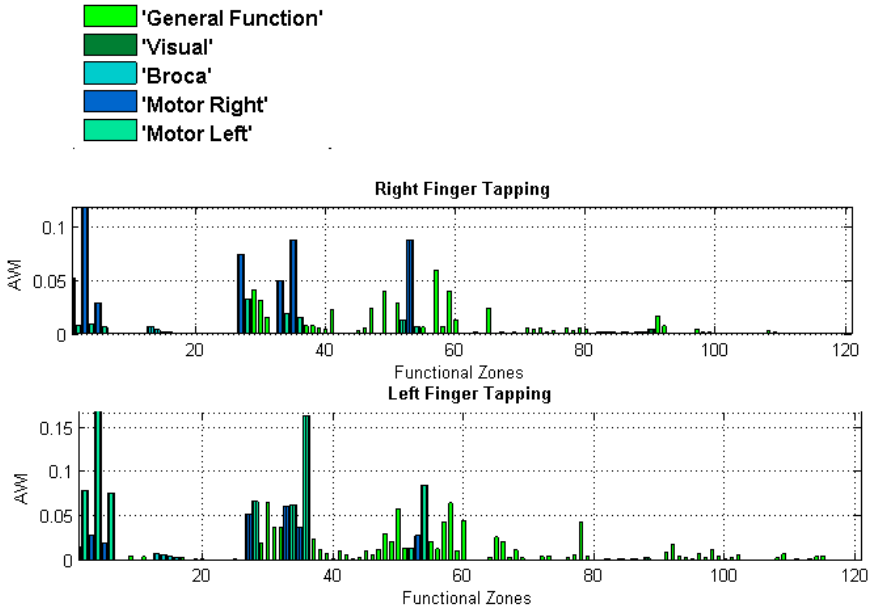


Figure 5.8: Ground truth Activation weighted vector for 2 analyzed tasks: right finger tapping and left finger tapping

resulting from fMRI exams of patients under study. All exams were analyzed with FEAT FLS and iViewBOLD; at first results were considered separately. For a particular area, AWIs of all patients were considered together and this area was considered actually related to the task under inspection only if it is active in at least 50% of patients. This was repeated for all functional areas, obtaining a vector in which only areas related to the task under inspection are included. Finally, vectors coming from FSL FEAT and iViewBOLD analysis were averaged, obtaining the definitive reference ground truth considered for subsequent evaluations. It is evident that these vectors depend on the number of patients better and better when this number increases. However, this procedure remains valid also for a single patient. Moreover, reference ground truth obtained with this procedure is independent of atlas choice, in fact Juelich atlas was used only in order to split brain into a certain number of well defined areas. For each functional task, let be $TRUE$ the ground truth vector, then true positives T_p, false positives F_n, true negatives T_n, and false negatives F_n are computed using fuzzy logic definitions:

$$\begin{aligned} T_p &= \sum_i min(AWI_j, TRUE_j) \\ F_p &= \sum_i max(0, AWI_j - TRUE_j) \\ T_n &= \sum_i min(1 - AWI_j, 1 - TRUE_j) \\ T_n &= \sum_i max(0, TRUE_j - AWI_j) \end{aligned} \tag{5.1}$$

sensitivity and specificity indexes were defined as follows:

$$\begin{aligned} sensitivityindex &= \frac{T_p}{T_p + F_n} \\ specificityindex &= \frac{T_p}{T_p + F_p} \end{aligned} \tag{5.2}$$

Figure 5.8 shows the ground truth used for this experiment. The sensitivity index is 0.98 for FSL FEAT and 0.55 for iViewBOLD. This index represents the software package capability of correctly recognize functional areas actually related to the corresponding functional task. False negative lowers index value. Specificity index represents the capability of correctly show as not activated areas not related to the functional task under inspection. In this example specificity index is 0.98 for FSL FEAT and 1 for iViewBOLD, so slightly higher for the latter. Areas 52 Primary somatosensory Cortex BA1 Right and 54 Primary somatosensory Cortex BA1 Right, although active in the ground truth vector with values 0.04 and 0.1, shown a strange behavior with values 0.13 and 0.18. These are the zones related to left movement nearest to the skull where

the tumor has displaced the functional areas. Comparing AWI values a quantitative evaluation is possible. In the same way the contro-lateral activation can be analyzed: quantifying the 'hyperactivity' compared to ground truth observed in SPM. The AWI analysis may be useful to study brain plasticity and redundancy. This example shows the usefulness and effectiveness of AWI quantitative analysis of fMRI SPMs, not only for software evaluation but also for clinical use.

5.3.1 Hemisphere Dominance Evaluation

Table 5.3: Hemispheric Dominance Evaluation

idx	Task	LI	LI_{AWV}	$LI^{B.}_{AWV}$	$LI^{W.}_{AWV}$	Report
1	L_1	-84.6	-2.1	-54.4	-7.5	Bilateral
	L_2	-2.5	20.1	-27.0	72.5	
	L_3	-33.6	64.2	74.8	88.4	
2	L_1	-47.0	-39.4	18.8	-29,0	Bilateral
	L_2	40.2	-14,2	-8.5	-22.7	
	L_3	-9.1	35.9	36.2	59.9	
3	L_1	-28.8	-18.0	-27.0	-60.1	Bilateral
	L_2	13.2	29.6	-50.2	81.0	
	L_3	1.8	55.8	71.9	37.6	
4	L_1	-44.4	0.3	-35.8	57.9	Bilateral with tendency to the Left
	L_2	-63.7	25.5	19.6	71.9	
	L_3	-36.4	25.9	39.9	49.7	
5	L_1	-85.9	19.2	21.4	35.3	Left Dominance
	L_2	-78.1	45.9	71.0	85.4	
	L_3	-26.4	25.2	26.6	15.9	
6	L_1	-16.7	25.7	50.3	20.7	Bilateral
	L_2	-26.9	15.4	38.2	16.0	
	L_3	-14.4	2.3	21.0	3.6	
7	L_1	-57.1	15.8	12.0	69.8	Left Dominance
	L_2	-50.4	46.2	41.6	100.0	
	L_3	36.3	64.9	59.8	77.7	
8	L_1	-63.4	1.5	-3.3	52.9	Left Dominance with Controlateral Attivation
	L_2	-87.1	21.1	-25.0	68.1	
	L_3	-24.5	18.5	-11.9	55.1	
9	L_1	-76.3	-9.4	-36.7	46.7	Left Dominance
	L_2	-62.8	25.9	77.3	76.4	
	L_3	0.8	60.1	46.7	88.8	
10	L_1	-70.6	40.5	19.2	75.1	Left Dominance
	L_2	-52.9	33.1	17.3	45.3	
	L_3	20.9	43.2	-7.4	100.0	

A set of experiments have been conducted to assess the capability of the AWV analysis to describe the hemispheric specialization. The hemispheric dominance is evaluated on all the 10 patient in our dataset. SPMs were qualitatively assessed by an expert, the classic LI is computed and the AWV hemispheric dominance is performed considering the global LI_{AWV} and the Broca and Wernicke language zones LI^{Borca}_{AWV} and $LI^{Wernicke}_{AWV}$ respectively. The fMRI exam pipeline includes three language tasks: Object Naming (L_1), Verb Generation (L_2) and Word Generation (L_3). the dataset used is composed by 10 patients, 6 males and 4 females, mean age 50.5 years and age range between 22

to 70 years 9 right-handed and 1 left-handed. Functional tests were performed preoperatively in patients awaiting surgery for removal of brain tumors. Table 5.3 shows the results obtained for all the cases analyzed. The AVW analysis is always in agreement with the radiologist opinion. Otherwise the classical LI is often in disagreement. It is worth to note that the analysis for specific areas is essential for a correct evaluation. The activation distribution can be very complex, therefore appears to be inadequate to describe the hemispheric specialization with a single index.

6
Conclusion and Future Work

In this Ph.D. Thesis some Madical Imaging and Pattern Recognition strategies are presented to analyze both the *Functional* and *Structural* Magnetic Resonance Imagines in order to support clinical experts in the treatment of patients with brain Glial tumor. All the algorithmic solutions studied have been integrated in *GliCInE: Glioma Computerized Inspection Environment* which is a MATLAB prototype of an integrated analysis environment that offers, in addition to functionality specifically described in this thesis, a set of tools needed to manage Functional and Structural Magnetic Resonance Volumes and ancillary data related to the acquisition and the patient. Figure 6.1 shows an example of the *GliCInE* GUI. The aim of this thesis was twofold: on the one hand to study innovative approaches in the context of applied computer science with great attention to the algorithmic and strictly methodological point of view. An evidence of this attitude can be appreciated in the sections of the thesis dedicated strictly to the methods (Chapters 2 and 4) in which an attempt was made to describe the methods in a formal way and independently of the specific application. On the other hand this Ph.D. Thesis is a multidisciplinary work whose main driving force is the medical problem of treating patients suffering from glial tumor. The medical objective is the shared aspect to all strategies studied that are instead heterogeneous in terms of type of data needed or of algorithmic methodology used. It is our conviction that only a complex analysis that use mutually Functional and Structural information can achieve an efficient use of the potential of such resonance imaging tools.
From the Functional point of view two challenging topics have been studied. The first

Figure 6.1: *GliCInE: Glioma Computerized Inspection Environment* Graphical User Interface

topic regards the comparison between two different Maps for assessing the statistical agreement. This comparison may be very interesting, whether it is applied on SPMs derived from the processing of two software package, or from two patients, or from the same patient at different time. The main contribution concerns the use of evaluation metrics. Both the indexes studied (weighted kappa index and information overlap) have been found useful in the clinical domain. The use of the first metric proposed, based on the weighted kappa, is facilitated by the similarity with the well known kappa index. The second, more complex IO matrix strategy, better reflects the typical visual method used by clinicians to evaluate SPMs.

The second aim is the quantitative analysis of the SPM. The objective of this phase is to briefly describe the volumetric Map for giving to the expert a synoptic view of the SPM in a simply feature vector. For this purpose an index that describe the information contained in the map it was studied the *Activation Weighted Index (AWI)*. This vector can be used as a guide to the expert for the exploration of the maps. Moreover, for its

brevity and specificity, it can be used in a pattern recognition process for the identification of a particular situation. Many application of the AWI vector can be developed; as for example the study of hemispheric dominance.
A comprehensive analysis requires that these functional information are integrated by structural data such as the position, dimension and spatial distribution of the tumor. Therefore, was mandatory, in this Ph.D. Thesis, deal with the automatic segmentation. The complexity of the brain tumor segmentation task has been addressed in a modular way taking care before brain segmentation in following tumor segmentation and then recognition.
The brain segmentation strategy proposed exploits a 2D graph-based technique to reliably detect boundary in complex scenes characterized by heterogeneous morphological patterns. A full level of automation is achieved through the automatic detection of frames including eyes and the computation of values for internal main parameters directly from the image data. As seen in our experimental context, the overall segmentation procedure is less sensitive to parameters variations, it shows a robust behavior while segmenting images with increasing levels of pathological deformation and has proven competitive compared with standard state-of-the-art solutions. It is then eligible as an operational tool specifically oriented to actively support surgical planning and post-surgical assessment activities. The encouraging results we have obtained in this work have prompted us to continue with further work with the aim to improve the quality of segmentation results. In particular our purpose was to extend the 2D solutions to a 3D taking into account the information available in the spatial sequence of MRI images
The rationales behind the proposed solutions are twofold: to exploit the potentialities of graph searching in volumetric segmentation adding properly spatial information and, at the same time, to limit the complexity in managing graph in 3D domain. In the experimental evaluation we cope with the critical task of automatically segmenting brain showing invasive pathology in MRI imagery. Results obtained demonstrate that the segmentation benefits from the insertion of contextual information within the 2D graph based strategy. the sensitivity analysis substantiate the conclusion that the heuristics in setting parameter values is reliably compensated by the well theoretical funded graph searching framework giving rise to an easy configurable method that shows a stable behavior. Comparison results show that significant improvements have been obtained with respect to slice-by-slice 2D segmentation and to deformable models approaches.
Two tumor segmentation strategies are proposed that can be used for the Unilateral tumor segmentation and for the butterfly and multicentric tumor segmentation. For the first task, the feature used is the symmetry, considering the tumor as an anomaly in the symmetry of the brain with respect to the mid sagittal plane. A clustering method based on energy minimization through Graph-Cut is applied on the volume computed as a difference between the left hemisphere and the right hemisphere mirrored across the

symmetry plane. The use the Graph Cut in Symmetry Space leads to very promising results and the masks obtained allow to compute both volume and spatial distribution of tumors with high accuracy under the condition that pathological areas are localized in a single hemisphere. To deal with bilateral and/or multicentric tumors, a new segmentation procedure has been designed and built on the top of the first method. It is based on the combined use of Competitive Expectation Maximization and Graph Cut. The combined use of the two techniques allows to integrate statistical and topological information reinforcing the robustness of the method which continue to be fully automated.

An essential aspect of this Ph.D. Thesis is the experimental effort. All the segmentation proposed solutions are evaluated qualitatively and quantitatively and moreover, the problem of defining a reliable validation of MRI brain tumor segmentation was addressed. 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. In this activity multidisciplinary discussion with medical experts was fundamental.

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 tailored on the experts visual and usability requirements. The solutions have been proposed on the basis of the results obtained in a study aimed to analyze and assess the uncertainty underlying the manual labeling process and have been implemented in the system GliMAn. 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. The contribution of this Thesis in this specific topic is centered of the reference data estimation method conceived within a common agreement strategy. The proposed Fuzzy Connectedness based method allows to estimate consensus segmentation, avoiding a complete manual segmentation. Experts are asked to provide essential reliable information, leaving to the automated method the estimation of the complete segmentation. Preliminary accuracy results obtained are encouraging. Future plans contemplate the use in the clinical practice of GliMAn and then the development of a robust evaluation, basing on which to refine the implementation of the reference estimation facilities and to eventually improve the overall usability.

The final task of the proposed framework concerns the classification of different heterogeneous tissues that compose the tumor area (Active Tumor, Edema and Necrose) from multiple MR images. The recognition task has been conceived as a supervised contextual pattern recognition problem. The evaluation of this last phase has the big plus of being

carried out in a comparative way with the state of the art algorithms in this context showing a very satisfactory behavior.
Several have been the contribution of this Ph.D Thesis, but the main result of this research has been a framework eligible as an operational tool to actively support surgical planning and post-surgical activities in brain glial tumor treatment. Future plans contemplate the application of the solutions proposed in an attempt to solve open clinical problems for deeply experiment the *GliCInE* potentialities in neuroimaging but also in other biomedical imaging domains.

Summary of the Innovative Contributions

In this section the innovative contributions detailed in the presented Ph.D Thesis are summarized and the related publications are listed.
The activities carried out during the Ph.D. project presented here can be categorized into four classes:

Collection of Medical Datasets: during the course of this Ph.D. a long time was spent in the collection of clinical cases both from the Structural and Functional point of view. We first consider the Structural aspect. Twenty real MRI Volumes was collected, all cases concern glial tumors both homolateral low-grade gliomas, butterfly and multicentric Glioblastoma. The volumes are heterogeneous in terms of resolution and Image quality. The built datasets should not be considered just a collection of MRI images but all the cases are discussed by a group of 5 medical experts and a detailed report was written for each case. In this report the acquisition mode, the tumor description in terms of type dimension position and all the specific characteristics which could affect the both the manual and automatic contours delineations capability are reported. Moreover, four cases of the dataset have been fully annotated: both the brain and the tumor are manually contoured by the experts. The segmentation of the same case by more than one expert has allowed us to observe and measure the inter-expert variability phenomenon. Two of these volumes were then segmented two more times by the same experts changing the reference plane for the manual segmentation allowing us to observe and measure the Intra-Expert variability phenomenon. To have multiple segmentations of the same data has also allowed us to evaluate the performance of our strategy with respect the expert variabilities measurement.
Form the Functional point of view we have collected images of 15 patients, who underwent different fMRI procedures, for a total of 49 tasks acquired (13 finger tapping, 13 object naming, 12 word generation, 11 verb generation). The patients underwent fMRI for clinical reasons, mainly pre-surgical for oncologic illness. A team of neuroradiology ana-

lyzed the data with and without the help of the proposed *Framework for Computer Assisted Analysis of fMRI Statistical Parametric Maps(SPM)* highlighting the contribution of the automatic support.

Algorithmic Innovation: focusing only on the algorithmic aspect, some solutions that exceed the limitations of the state of art have been proposed.
In Brain Segmentation, 2D and 3D approaches are studied finding in both cases novel solutions that which have proved competitive with state of the art methods. The problem of segmenting brain MRI images has been extensively addressed by many researchers. Despite the relevant achievements obtained, automated segmentation of brain MRI imagery is still a challenging problem whose solution has to cope with critical aspects such as anatomical variability and pathological deformation. Methods for segmenting brain from MRI images basing on 2D graph searching principles for border detection are studied. The 2D procedure is made fully automatic and easily reproducible by the automatic by computing the internal main parameters directly from the image data.

- Valentina Pedoia, Elisabetta Binaghi, Sergio Balbi, Alessanbro De Benedictis, Emanuele Monti, Renzo Minotto **2D MRI Brain Segmentation by using feasibility constraints**, *In Proceedings Vision And Medical Image Processing*, VipIMAGE 2011, 251-256.
- Valentina Pedoia, Elisabetta Binaghi, **Automatic MRI 2D Brain Segmentation using Graph Searching Technique** *International Journal Numerical Method in Biomedical Engineer*, John Wiley & Sons. 2012

Also a 3D strategy has been studied attempting to circumvent the problem of defining both a renewed 3D lattice structure and a novel optimization strategy. Contextual information related to the sequence of 2D images has been embedded directly in the cost function defined in the 2D graph searching stage. Acting directly on the node weights we avoid the construction of a 3D graph topology which implies high resource demand, preserving however the explicit description of the boundary.

- Valentina Pedoia, Elisabetta Binaghi, **Heuristic 3D Graph Propagation Technique for Automatic MRI Brain Segmentation** *International Journal in Pattern Recognition and Artificial Intelligence*, Under Review.

In Tumor Segmentation, a clustering method based on energy minimization through Graph-Cut is applied on the volume computed as a difference between the left hemisphere and the right hemisphere mirrored across the symmetry plane. The feature used is the symmetry, considering that the tumor as an anomaly in the symmetry of the brain with respect to the mid sagittal plane. This unconventional use of graph cut in the space of symmetries allows the fully automatic solution of the problem. The proposed strategy leads to very promising results and the masks obtained allow to compute both volume and spatial distribution of tumors with high accuracy under the condition that pathological areas are localized in a single hemisphere. To deal with bilateral and/or multicentric tumors, a new segmentation procedure has been designed and built on the top of the first method. It is based on the combined use of Competitive Expectation Maximization and Graph Cut. The combined use of the two techniques allows to integrate statistical and topological information reinforcing the robustness of the method which continue to be fully automated.

- Valentina Pedoia, Elisabetta Binaghi, Sergio Balbi, Alessanbro De Benedictis, Emanuele Monti, Renzo Minotto **Glial Brain Tumor Detection by using symmetry analysis**, *In Proceedings of SPIE Medical Imaging: Image Processing 8314*, pages 831445-831445-8, Feb 2012.

In this Ph.D. Thesis was also presented a new method for the common agreement estimation of truth labels based on Fuzzy-Connectedness. The proposed strategy is naturally oriented to integrate uncertain information and then it is expected to manage dissimilarity among input labels; interaction is drastically limited with respect to a complete manual tracing and the formal fuzzy framework supports in the overall process of estimation without arbitrary solutions.

- Elisabetta Binaghi, Valentina Pedoia, Desiree Lattazzi, Emanuele Monti, Sergio Balbi, Renzo Minotto **Collection and Fuzzy Estimation of Truth Labels in Fully Automatic MRI Segmentation**, *Computerized Medical Imaging and Graphics.* Under Review

Considering the Functional use of the Magnetic Resonance Imaging a new method for describing in a brief but comprehensive manner the brain activation maps was studied. The method requires registration

of the brain of a patient with an atlas brain; for solving this task a non conventional use of the spherical harmonic decomposition is proposed. Moreover, the feature description studied was used for a novel Hemispherical Dominance evaluation strategy, which has been found very useful by the group of medical experts who evaluated the proposed tools.

- Valentina Pedoia, Vittoria Colli, Sabina Strocchi, Cristina Vite, Elisabetta Binaghi, Leopoldo Conte. **fMRI analysis software tools: an evaluation framework**, *Proceedings of SPIE Medical Imaging: Image Processing: Biomedical Applications in Molecular, Structural, and Functional Imaging 7965* , pages, 796528-796528-9, Feb 2011.
- Valentina Pedoia, Ignazio Gallo, Elisabetta Binaghi, **Affine Spharm Registration: Neural Estimation of Affine Transformation in Spherical Domain**, In Leonid Mestetskiy and Jose Braz editors, VISAPP, pages 197-200. SciTePress, 2011.
- Valentina Pedoia, Sabina Strocchi, Renzo Minotto, Elisabetta Bianghi, **Hemisphere Dominance Evaluation by using fMRI Activation Weighted Vector**, *In Proceedings Of Computational Modeling of Object Presented In Images: Fundamentals, Methods and Application*, CompIMAGE 2012, Pages 303-306 Set 2012 ISBN: 978-0-415-62134-2 DOI: 10.1201/b12753-56.

Another important part of the work concern the comparison of two different Maps to assess the statistical agreement. We conceived new methods to compare *non-threshold* SPMs based on a continuous formulation of Kappa and Jaccard statistical indexes.

- Valentina Pedoia, Sabina Strocchi, Vittoria Colli, Elisabetta Binaghi, Leopoldo Conte, **Functional Magnetic Resonance Imaging: comparison between activation maps and computation pipelines in a clinical context**. *Magnetic Resonance Imaging*, Elsevier, 12 Dec- 2012 (Article in Press DOI: 10.1016/j.mri.2012.10.013).

Technical Contributions *(interfaces, data structures, user support)* In this Ph.D. project an important role has been given to the clinical operational aspects. Therefore, some operational prototypes have been developed. This applications are actually usable by the physicians allowing the use of the proposed algorithms for medical research studies.
GLIMAN: GLIoma Manual ANnotator is a MATLAB application that allows to view and manipulate MRI volumes with the aim of reducing

the intra- and inter- personal variability in the manual Glial brain tumor segmentation process. 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.

- Valentina Pedoia, Giuseppe Renis, Sergio Balbi, Alessandro De Benedictis, Emanuele Monti, Elisabetta Binaghi **Manual labeling strategy for Ground Truth estimation in MRI Glial Tumor Segmentation**, *In Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA 2012, pages 8:1-8:4, New York, NY, USA, 2012. ACM.

GLICINE: GLIoma Computerized INspection Environment is a MATLAB application that collect all the tools for analyzing Functional and Structural MRI presented in this Ph.D. Thesis. The main results of the outlined research has been this framework eligible as an operational tool to actively support surgical planning and post-surgical activities in brain glial tumor treatment.

Medical Results The use of the proposed tools has contributed in carrying out some studies in the medical field:

- Emanuele Monti,Valentina Pedoia, Elisabetta Bianaghi, Alessandro De Benedictis, Sergio Balbi, **Graph based MRI analysis for evaluating the prognostic relevance of longitudinal brain atrophy estimation in post-traumatic diffuse axonal injury**, *In Proceedings Of Computational Modeling of Object Presented In Images: Fundamentals, Methods and Application*, CompIMAGE 2012, Pages 300-305 Set 2012.
- A. De Benedictis, Valentina Pedoia, E.Binaghi, S. Sangiorgi, D. Lattazzi, S. Strocchi, R. Minotto, A. Giorgianni, G. Tomei, S. Balbi, **Tawards a Non-Presumptive Interpretation of fMRI Data Quantitative Analysis of Statistic parametric Maps**, *ISMRM Workshop on Mapping Functional Networks for Brain Sourgery.* Milan, Sep 2011.
- Renzo Minotto, Andrea Giorgianni, Sergio Balbi, Alessandro De Benedictis, Vittoria Colli, Sabina Strocchi, Elisabetta Binaghi, Valentina Pedoia, **Ruolo degli studi di fMRI perioperatoria e loro validazione mediante dati ottenuti da stimolazione cerebrale diretta in pazienti sottoposti ad asportazione di gliomi celebrali**, *I Congresso Nazionale di Neuroradiologia Funzionale*, Feb 2011.
- Sabina Strocchi, Vittoria Colli, Valentina Pedoia, Elisabetta Binaghi, Leopoldo Conte, **Mappe Statistiche Parametriche in fMRI: in-**

postazione di una metodologia di Confronto e di una visualizzazione Ottimizzata, *VII Congresso Nazionale Associazione Italiana di Fisica Medica*, Squillace (CZ), Set 2011.

Bibliography

[1] http://www.fil.ion.ucl.ac.uk/spm/.

[2] G. B. Aboutanos and B. M. Dawant. Automatic brain segmentation and validation: Image-based versus atlas-based deformable modelsn. In *in Proc. SPIE-Medical Imaging 1997*, volume 3034, pages 299–310, 1997.

[3] Babak Afshin-Pour, Hamid Soltanian-Zadeh, Gholam-Ali Hossein-Zadeh, Cheryl L. Grady, and Stephen C. Strother. A mutual information-based metric for evaluation of fMRI data-processing approaches. *Hum. Brain Mapp.*, 32(5):699–715, 2011.

[4] Mohamed Aly. Survey on multiclass classification methods, 2005.

[5] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain mr data. *Medical Imaging, IEEE Transactions on*, 28(8):1266 –1277, aug. 2009.

[6] M. Stella Atkins, Blair Mackiewich, and Ken Whittall. Fully automatic segmentation of the brain in MRI. 1998.

[7] Raouia Ayachi and Nahla Ben Amor. Brain tumor segmentation using support vector machines. In Claudio Sossai and Gaetano Chemello, editors, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 5590 of *Lecture Notes in Computer Science*, pages 736–747. Springer Berlin Heidelberg, 2009.

[8] Shai Bagon. Matlab wrapper for robust higher order potentials, January 2009.

[9] R. Bakeman, Gottman, and J.M. *Observing interaction: An introduction to sequential analysis*. Cambridge University Press, Cambridge, UK, 1997.

[10] M. Balafar, A. Ramli, M. Saripan, and S. Mashohor. Review of brain MRI image segmentation methods. *Artificial Intelligence Review*, 33:261–274, 2010. 10.1007/s10462-010-9155-0.

[11] D. H. Ballard. Readings in computer vision: issues, problems, principles, and paradigms. chapter Generalizing the hough transform to detect arbitrary shapes, pages 714–725. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[12] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, N.J., 1982.

[13] Issac N. Bankman. *Handbook of medical image processing and analysis*. Academic press, 2008.

[14] Luc Bauchet, Valérie Rigau, Hélène Mathieu-Daudé, Dominique Figarella-Branger, Delphine Hugues, Loreleï Palusseau, Fabienne Bauchet, Michel Fabbro, Chantal Campello, Laurent Capelle, Anne Durand, Brigitte Trétarre, Didier Frappaz, Dominique Henin, Philippe Menei, Jérome Honnorat, and François Segnarbieux. French brain tumor data bank: Methodology and first results on 10,000 cases. *Journal of Neuro-Oncology*, 84:189–199, 2007.

[15] 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 Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[16] Stefan Bauer, Lutz-P. Nolte, and Mauricio Reyes. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *Proceedings of the 14th international conference on Medical image computing and computer-assisted intervention - Volume Part III*, MICCAI'11, pages 354–361, Berlin, Heidelberg, 2011. Springer-Verlag.

[17] Christian F. Beckmann, Mark Jenkinson, and Stephen M. Smith. General multilevel linear modeling for group analysis in fmri. *NeuroImage*, 20(2):1052 – 1063, 2003.

[18] G. Bertani, E. Fava, G. Casaceli, G. Carrabba, A. Casarotti, C. Papagno, A. Castellano, A. Falini, S. M. Gaini, and L. Bello. Intraoperative mapping and monitoring of brain functions for the resection of low-grade gliomas: technical considerations. 27(4), 10 2009.

[19] Jeff Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, 1998.

[20] Elisabetta Binaghi. A fuzzy logic inference model for a rule-based system in medical diagnosis. *Expert Systems*, 7(3):134–141, 1990.

[21] Elisabetta Binaghi, Pietro A. Brivio, Paolo Ghezzi, and Anna Rampini. A fuzzy set-based accuracy assessment of soft classification. *Pattern Recognition Letters*, 20(9):935 – 948, 1999.

[22] N. Birkbeck, D. Cobzas, M. Jagersand, A. Murtha, and T. Kesztyues. An interactive graph cut method for brain tumor segmentation. In *Applications of Computer Vision (WACV), 2009 Workshop on*, pages 1 –7, dec. 2009.

[23] Andrew Blake, Carsten Rother, M. Brown, Patrick Perez, and Philip H. S. Torr. Interactive image segmentation using an adaptive gmmrf model. In *ECCV (1)'04*, pages 428–441, 2004.

[24] Marie Blonski, Luc Taillandier, Guillaume Herbet, IgorLima Maldonado, Patrick Beauchesne, Michel Fabbro, Chantal Campello, Catherine Gozé, Valérie Rigau, Sylvie Moritz-Gasser, Christine Kerr, Roberta RudA, Riccardo Soffietti, Luc Bauchet, and Hugues Duffau. Combination of neoadjuvant chemotherapy followed by surgical resection as a new strategy for who grade ii gliomas: a study of cognitive status and quality of life. *Journal of Neuro-Oncology*, 106:353–366, 2012.

[25] M. Bomans, K.-H. Hohne, U. Tiede, and M. Riemer. 3-D segmentation of MR images of the head for 3-D display. *Medical Imaging, IEEE Transactions on*, 9(2):177 –183, jun 1990.

[26] S. Bouix, M. Martin-Fernandez, L. Ungar, M. Nakamura M.-S. Koo, R. W. McCarley, and M. E. Shenton. On evaluating brain tissue classifiers without a ground truth. *NeuroImage*, 36:1207–1224, 2007.

[27] Boykov, Yuri, Funka-Lea, and Gareth. Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision*, 70:109–131, November 2006.

[28] Yuri Boykov and Vladimir Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 26–, Washington, DC, USA, 2003. IEEE Computer Society.

[29] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 26:1124 – 1137, September 2004.

[30] Yuri Boykov, Vladimir Kolmogorov, Daniel Cremers, and Andrew Delong. An integral solution to surface evolution pdes via geo-cuts. In *In ECCV*, pages 409–422, 2006.

[31] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:1222–1239, November 2001.

[32] Yuri Y. Boykov and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. 2001.

[33] Ch. Brechbühler, G. Gerig, and O. Kübler. Parametrization of closed surfaces for 3-d shape description. *Comput. Vis. Image Underst.*, 61(2):154–170, 1995.

[34] Hermann Bruce. Wada test failure and cognitive outcome. *Epilepsy Currents*, 5(2):61–62, 2005.

[35] Martin D. Buhmann and M. D. Buhmann. *Radial Basis Functions*. Cambridge University Press, New York, NY, USA, 2003.

[36] D. B. Godlgof C. Li and L. O. Hall. Knowledge-based classification and tissue labeling of mr images of human brain. *IEEE Transaction On Medical Imaging*, 12:740–750, 1993.

[37] Alejandro Caceres, Deanna L. Hall, Fernando O. Zelaya, Steven C.R. Williams, and Mitul A. Mehta. Measuring fmri reliability with the intra-class correlation coefficient. *NeuroImage*, 45(3):758 – 768, 2009.

[38] B.M. Carvalho, C.J. Gau, G.T. Herman, and T.Y. Kong. Algorithms for fuzzy segmentation. *Pattern Anal. Appl.*, 2.

[39] Ping-Lin Chang and Wei-Guang Teng. Exploiting the self-organizing map for medical image segmentation. *Computer-Based Medical Systems, IEEE Symposium on*, 0:281–288, 2007.

[40] Jia-Lin Chen and A. Kundu. Unsupervised texture segmentation using multichannel decomposition and hidden markov models. *Trans. Img. Proc.*, 4(5):603–619, May 1995.

[41] Krzysztof C. Ciesielski, Jayaram K. Udupa, A. X. Falcão, and P. A. V. Miranda. Comparison of fuzzy connectedness and graph cut segmentation algorithms. pages 796203–796203–12, 2011.

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

[43] L.P. Clarke, R.P. Velthuizen, S. Phuphanich, J.D. Schellenberg, J.A. Arrington, and M. Silbiger. Mri: Stability of three supervised segmentation techniques. *Magnetic Resonance Imaging*, 11(1):95 – 106, 1993.

[44] H.E. Cline, W.E. Lorensen, R. Kikinis, and F.A. Jolesz. Three-dimensional segmentation of mr images of the head using probability and connectivity. *J Comput Assist Tomogr*, 14(6):1037–45, 12 1990.

[45] Chris A. Cocosco, Vasken Kollokian, Remi K.-S. Kwan, G. Bruce Pike, and Alan C. Evans. BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage*, 5:425, 1997.

[46] Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960.

[47] Jacob Cohen. Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220, 1968.

[48] R. K. Colwell and J. A. Coddington. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond*, 345, 1994.

[49] Jason J. Corso, Eitan Sharon, Shishir Dube, Suzie El-saden, Usha Sinha, and Alan L. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Transactions on Medical Imaging*, 27:629–640, 2008.

[50] Koby Crammer and Yoram Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Mach. Learn. Res.*, 2:265–292, March 2002.

[51] N. Cristianini and J. Shawe-Taylor. *Support Vector Machines*. Cambridge University Press, 2000.

[52] Kok Houi Lim David B.L. Bong. Application of fixed-radius hough transform in eye detection. *International Journal of Intelligent Information Technology Application*, 2:121–127, 2009.

[53] S. Dellepiane. Image segmentation: Errors, sensitivity, and uncertainty. In *Engineering in Medicine and Biology Society, 1991. Vol.13: 1991., Proceedings of the Annual International Conference of the IEEE*, pages 253 –254, oct-3 nov 1991.

[54] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[55] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[56] L. R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, July 1945.

[57] M. Dincic, Z. Peric, and A. Jovanovic. Optimal polar image sampling. *Opto-Electronics Review*, 19:249–255, 2011. 10.2478/s11772-011-0013-7.

[58] D. Dubois and H. Prade. A review of fuzzy set aggregation connectives. *Information Sciences*, 36(1–2):85 – 121, 1985.

[59] 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.

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

[61] Hugues Duffau. Awake surgery for incidental who grade ii gliomas involving eloquent areas. *Acta Neurochirurgica*, 154:575–584, 2012.

[62] Chang Edward F, Aaron Clark, Justin S Smith, Mei-Yin Polley, Susan M Chang, Nicholas M Barbaro, Andrew T Parsa, Michael W McDermott, and Mitchel S Berger.

[63] Murray Eisenberg and Robert Guy. A proof of the hairy ball theorem. *The American Mathematical Monthly*, 86(7):pp. 571–574, 1979.

[64] Alexandre X. Falcao and Jayaram K. Udupa. A 3d generalization of user-steered live-wire segmentation. *Medical Image Analysis*, 4(4):389 – 402, 2000.

[65] A.X. Falcao, J.K. Udupa, and F.K. Miyazawa. An ultra-fast user-steered image segmentation paradigm: live wire on the fly. *Medical Imaging, IEEE Transactions on*, 19(1):55 –62, jan. 2000.

[66] L. Feliss. *Statistical Method for Rates and Proportions.* Johon Wiley and Sons, 3 edition, 1981.

[67] Mario A. T. Figueiredo and Anil K. Jain. Unsupervised learning of finite mixture models. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 24:381–396, 2000.

[68] Michael S. Floater and Kai Hormann. Surface parameterization: a tutorial and survey, 2005.

[69] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4):189–210, 1994.

[70] Jingge Gao, Shuqiang Zhang, and Wei Lu. Application of hough transform in eye tracking and targeting. In *Electronic Measurement Instruments, 2009. ICEMI '09. 9th International Conference on*, pages 3–751 –3–754, 2009.

[71] Cristina Garcia and Jose Ali Moreno. Kernel based method for segmentation and modeling of magnetic resonance images. In Christian Lemaitre, Carlos A. Reyes, and Jesus A. Gonzalez, editors, *Advances in Artificial Intelligence - IBERAMIA 2004, 9th Ibero-American Conference on AI, Puebla, Mexico, November 22-26, 2004, Proceedings*, volume 3315 of *Lecture Notes in Computer Science*, pages 636–645. Springer, 2004.

[72] Christopher R. Genovese, Douglas C. Noll, and William F. Eddy. Estimating test-retest reliability in functional MR imaging I: Statistical methodology. *Magnetic Resonance in Medicine*, 38:497–507, 1997.

[73] Ezequiel Geremia, Bjoern H. Menze, and Nicholas Ayache. Spatial decision forests for glioma segmentation in multi-channel MR images. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[74] Guido Gerig, John Martin, Ron Kikinis, Olaf Kubler, Martha Shenton, and Ferenc A. Jolesz. Unsupervised tissue type segmentation of 3d dual-echo mr head data. *Image Vision Comput.*, 10:349–360, July 1992.

[75] L.O. Hall, A.M. Bensaid, L.P. Clarke, R.P. Velthuizen, M.S. Silbiger, and J.C. Bezdek. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *Neural Networks, IEEE Transactions on*, 3(5):672 –682, sep 1992.

[76] A. Hamamci and G. Unal. Multimodal brain tumor segmentation using the "tumor-cut" method on the brats dataset. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[77] Marti A. Hearst. Support vector machines. *IEEE Intelligent Systems*, 13(4):18–28, July 1998.

[78] Rolf A. Heckemann, Joseph V. Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain mri segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115 – 126, 2006.

[79] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, August 2009.

[80] L. O. Hall J. C. Bezdek and L. P. Clarke. Review of mr image segmentation techniques using pattern recognition. *Medical Physic*, 20(4):1033–1048, 1993.

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

[82] Shishir Dube Suzie El-Saden Usha Sinha Jason J. Corso, Eitan Sharon and Alan Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Transactions on Medical Imaging*, 27(5), 2008.

[83] Chang. Kathleen R Lamborn. Susan M Chang. Michael D Prados. Soonmee Cha. Tarik Tihan. Scott Vandenberg. Michael W McDermott. Mitchel S Berger. Justin S, Smith. Edward F. *Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas.* 2008.

[84] Christoph Stippich K. Sartor. *Clinical Functional MRI: Presurgical Functional Neuroimaging (Medical Radiology / Diagnostic Imaging)*. Springer Berlin Heidelberg, 2 edition, 2010.

[85] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1988. 10.1007/BF00133570.

[86] M.R. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz, and R. Kikinis. Automated segmentation of brain tumors. 12 2007.

[87] P J Kelly, C Daumas-Duport, D B Kispert, B A Kall, B W Scheithauer, and J J Illig. Imaging-based stereotaxic serial biopsies in untreated intracranial glial neoplasms. *Journal Of Neurosurgery*, 66(6):865–874, 1987.

[88] Hassan Khotanlou, Olivier Colliot, Jamal Atif, and Isabelle Bloch. 3d brain tumor segmentation in mri using fuzzy classification, symmetry analysis and spatially constrained deformable models. *Fuzzy Sets Syst.*, 160:1457–1473, May 2009.

[89] Kwang In Kim, Keechul Jung, Se Hyun Park, and Hang Joon Kim. Support vector machines for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1542–1550, 2002.

[90] Pushmeet Kohli, Lubor Ladicky, and Phillip Torr. Graph cuts for minimizing robust higher order potentials. Technical report, Oxford Brookes University, UK., 2008.

[91] Pushmeet Kohli, Lubor Ladicky, and Phillip Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.

[92] Pawan M. Kumar, Philip H. S. Torr, and A. Zisserman. Obj cut. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, pages 18–25, Washington, DC, USA, 2005. IEEE Computer Society.

[93] J. Richard Landis and Gary G. Koch. The mesurament of observer agreement for the categorial data. *Biometrics*, 33:159–174, 1977.

[94] Tuong H. Le and Xiaoping Hu. Methods for assessing accuracy and reliability in functional MRI. *NMR in Biomedicine*, 10(4-5):160–164, 1997.

[95] D N Levin, X P Hu, K K Tan, and S Galhotra. Surface of the brain: three-dimensional MR images created with volume rendering. *Radiology*, 171(1):277–280, 05 1989.

[96] Hongming Li, Ming Song, and Yong Fan. Segmentation of brain tumors in multi-parametric mr images via robust statistic information propagation. In *Proceedings of the 10th Asian conference on Computer vision - Volume Part IV*, ACCV'10, pages 606–617, Berlin, Heidelberg, 2011. Springer-Verlag.

[97] Kang Li, Xiaodong Wu, Danny Z. Chen, and Milan Sonka. Optimal surface segmentation in volumetric images-a graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28:119–134, January 2006.

[98] Martin A. Lindquist. The statistical analysis of fmri data. *Statistice Science*, 23(4):439–464, 2008.

[99] J. Liu, J. K. Udupa, D. Odhner, D. Hackney, and G. Moonis. A system for brain tumor volume estimation via mr imaging and fuzzy connectedness. *Comput. Med. Imag. Graphics*, 29(1):21–34, 2005.

[100] Arvid Lundervold and Geir Storvik. Segmentation of brain parenchyma and cerebrospinal fluid in multispectral magnetic resonance images. *IEEE Transactions on Medical Imaging*, 14:339–349, 1995.

[101] D. B. Goldgof R. Velthuizen R. Murtagh M. C. Clark, L. O. Hall and M. S. Silbiger. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17(2):187–201, 1998.

[102] R. L. Eisner M. E. Brummer, R. M. Mersereau and R. R. J. Lewine. Automatic detection of brain contours in mri data sets. *IEEE Transaction On Medical Imaging*, 12:153–166, 1993.

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

[104] S. Ho M. Prastawa, E. Bullitt and G. Gerig. A brain tumor segmentation framework based on outlier detection. *Med. Image Anal. J.*, 8(3):275–283, 2004.

[105] Willem C. M. Machielsen, Serge A. R. B. Rombouts, Frederik Barkhof, Philip Scheltens, and Menno P. Witter. fMRI of visual encoding: Reproducibility of activation. *Human Brain Mapping*, 9(3):156–164, 2000.

[106] Ranjan Maitra. Assessing certainty of activation or inactivation in test retest fMRI studies. *NeuroImage*, 47(1):88 – 97, 2009.

[107] Ranjan Maitra, Steven R. Roys, and Rao P. Gullapalli. Test-retest reliability estimation of functional MRI data. *Magnetic Resonance in Medicine*, 48(1):62–70, 2002.

[108] E.H. Mamdani. Advances in the linguistic synthesis of fuzzy controllers. *International Journal of Man-Machine Studies*, 8(6):669 – 678, 1976.

[109] E.H. Mamdani and B.R. Gaines. *Fuzzy Reasoning and its Applications,*. Academic Press, London, 1981.

[110] Emmanuel Mandonnet, Laurent Capelle, and Hugues Duffau. Extension of paralimbic low grade gliomas: toward an anatomical classification based on white matter invasion patterns. *Journal of Neuro-Oncology*, 78:179–185, 2006.

[111] Emmanuel Mandonnet, Johan Pallud, Denys Fontaine, Luc Taillandier, Luc Bauchet, Philippe Peruzzi, Jacques Guyotat, Valerie Bernier, Marie-Helene Baron, Hugues Duffau, and Laurent Capelle. Inter- and intrapatients comparison of WHO grade II glioma kinetics before and after surgical resection. *Neurosurgical Review*, 33:91–6, 2009.

[112] Alberto Martelli. Edge detection using heuristic search methods. *Computer Graphics and Image Processing*, 1(2):169 – 182, 1972.

[113] B. H. Menze, N. Ayache E. Geremia, and G. Szekely. Segmenting glioma in multi-modal images using a generative-discriminative model for brain lesion segmentation. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[114] B. H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland. Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[115] J. Ross Mitchell, Stephen J. Karlik, Donald H. Lee, and Aaron Fenster. Computer-assisted identification and quantification of multiple sclerosis lesions in mr imaging volumes in the brain. *Journal of Magnetic Resonance Imaging*, 4(2):197–208, 1994.

[116] Ugo Montanari. On the optimal detection of curves in noisy pictures. *Commun. ACM*, 14:335–345, May 1971.

[117] V. L. Morgan, B. M. Dawant, Y. Li, and D. R. Pickens. Comparison of fMRI statistical software packages and strategies for analysis of images containing random and stimulus-correlated motion. *Computerized Medical Imaging and Graphics*, 31(6):436 – 446, 2007.

[118] J. Morioka, T. Nishizaki, T. Tokumaru, S. Uesugi, K. Yamashita, H. Ito, and M. Suzuki. Functional magnetic resonance imaging-controlled neuronavigator-guided brain surgery: a case report. *J Clin Neurosci*, 8(3):283–5, 2001.

[119] C. Moritz and V. Haughton. Functional mr imaging: paradigms for clinical preoperative mapping. *Magn Reson Imaging Clin N Am*, 11(4):529–42, v, 2003.

[120] J Nelder and R Wedderburn. Generalized Linear Models. *Jurnal of the Royal Statistical Society*, 135(3):370 – 384, 1972.

[121] Jingxin Nie, Zhong Xue, Tianming Liu, Geoffrey Young, Kian Setayesh, Lei Guo, and Stephen T. C. Wong. Automated brain tumor segmentation using spatial accuracy-weighted hidden markov random field. *Comp. Med. Imag. and Graph.*, 33(6):431–441, 2009.

[122] L.G. Nyul, A.X. Falcao, and J.K. Udupa. Fuzzy-connected 3d image segmentation at interactive speeds. *Graph. Models Image Processing*, 64.

[123] T.R. Oakes, T. Johnstone, K.S. Ores Walsh, L.L. Greischar, A.L. Alexander, A.S. Fox, and R.J. Davidson. Comparison of fMRI motion correction software tools. *NeuroImage*, 28(3):529 – 543, 2005.

[124] Delen Dursun Olson, David L. *Advanced Data Mining Techniques*. Springer, 1 edition, 2008.

[125] Kleihues P and Cavenee W. *Pathology and genetics of tumours of the nervous system*. Lyon, 2000.

[126] 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.

[127] F. Pannizzo, M. J. B. Stallmeyer, J. Friedman, R. J. Jennis, J. Zabriskie, C. Plank, R. Zimmerman, J. P. Whalen, and P. T. Cahill. Quantitative mri studies for assessment of multiple sclerosis. *Magnetic Resonance in Medicine*, 24(1):90–99, 1992.

[128] Valentina Pedoia and Elisabetta Binaghi. Automatic mri 2d brain segmentation using graph searching technique. *International Journal for Numerical Methods in Biomedical Engineering*, pages n/a–n/a, 2012.

[129] Valentina Pedoia, Elisabetta Binaghi, Sergio Balbi, Alessandro De Benedictis, Emanuele Monti, and Renzo Minotto. 2d MRI brain segmentation by using feasibility constraints. In *Proceedings Vision And Medical Image Processing, VipIMAGE 2011*, pages 251–256, 2011.

[130] Valentina Pedoia, Elisabetta Binaghi, Sergio Balbi, Alessandro De Benedictis, Emanuele Monti, and Renzo Minotto. Glial brain tumor detection by using symmetry analysis. In *Proceedings of SPIE Medical Imaging: Image Processing 8314*, pages 831445–831445–8, 2012.

[131] Valentina Pedoia, Vittoria Colli, Sabina Strocchi, Cristina Vite, Elisabetta Binaghi, and Leopoldo Conte. fMRI analysis software tools: an evaluation framework. In *Proc. SPIE 7965*, 2011.

[132] Valentina Pedoia, Ignazio Gallo, and Elisabetta Binaghi. Affine spharm registration - neural estimation of affine transformation in spherical domain. In Leonid Mestetskiy and Jose Braz, editors, *VISAPP*, pages 197–200. SciTePress, 2011.

[133] Valentina Pedoia, Sabina Strocchi, Vittoria Colli, Elisabetta Binaghi, and Leopoldo Conte. Functional magnetic resonance imaging: comparison between activation maps and computation pipelines in a clinical context. *Magnetic Resounce Imaging*, pages n/a–n/a, 2012.

[134] Valentina Pedoia, Sabina Strocchi, Renzo Minotto, and Elisabetta Binaghi. Hemispheric dominance evaluation by using fmri activation weighted vector. *Computational Modelling of Objects Represented in Images III: Fundamentals, Methods and Applications*, page 303, 2012.

[135] Yanmin Peng and Rong Liu. Object segmentation based on watershed and graph cut. In *Image and Signal Processing (CISP), 2010 3rd International Congress on*, volume 3, pages 1431 –1435, oct. 2010.

[136] Nguyen Hoang Phuong and Vladik Kreinovich. Fuzzy logic and its applications in medicine. *International Journal of Medical Informatics*, 62:2–3, 2001.

[137] D. R. Pickens, Y. Li, V. L. Morgan, and B. M. Dawant. Development of computer-generated phantoms for fMRI software evaluation. *Magnetic Resonance Imaging*, 23(5):653 – 663, 2005.

[138] M. Prastawa, E. Bullit, N. Moon, K.V. Leemput, and G. Gerig. Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology*, 10(12):1341–1348, 2003.

[139] T. Riklin Raviv, K. Van Leemput, and B. H. Menze. Multi-modal brain tumor segmentation via latent atlases. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[140] N. Ray, B.N. Saha, and M.R. Graham Brown. Locating brain tumors from mr imagery using symmetry. In *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pages 224 –228, nov. 2007.

[141] Nicolas Robitaille and Simon Duchesne. Label Fusion Strategy Selection. *International Journal of Biomedical Imaging*, 2012(7), 2012.

[142] T. Rohlfing, D. B. Russakoff, and C. R. Maurer. Expectation maximization strategies for multi-atlas multi-label segmentation. *Inf Process Med Imaging*, 18:210–221, July 2003.

[143] Torsten Rohlfing and Calvin R. Maurer Jr. Shape-based averaging. *IEEE Transactions on Image Processing*, pages 153–161, 2007.

[144] Rother, Carsten, Kolmogorov, Vladimir, Blake, and Andrew. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[145] E. Bullitt S. Ho and G. Gerig. Level set evolution with region competition: Automatic 3-d segmentation of brain tumors. In *in Proc. Int. Conf. Pattern Recognit*, volume 1, pages 523–535, 2002.

[146] F. A. Jolesz S. K. Warfield, M. R. Kaus and R. Kikinis. Adaptive template moderated spatially varying statistical classification. In *in Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 431–438. Springer-Verlag, 1995.

[147] P.K. Saha, J.K. Udupa, and D.Odhner. Scale-based fuzzy connectedness image segmentation: Theory, algorithms, and validation. *Comput. Vision Image Understanding*, 77.

[148] Tanya Schmah, Grigori Yourganov, Richard S. Zemel, Geoffrey E. Hinton, Steven L. Small, and Stephen C. Strother. Comparing classification methods for longitudinal fmri studies. *Neural Comput.*, 22(11):2729–2762, 2010.

[149] Mohamed L Seghier. Laterality index in functional mri: methodological issues. *Magnetic Resonance Imaging*, 26(5):594–601, 2008.

[150] A.H. Shah, K. Madhavan, D. Heros, D.M.S. Raper, J.B. Iorgulescu, B.E. Lally, and R.J. Komotar. The management of incidental low-grade gliomas using magnetic resonance imaging: systematic review and optimal treatment paradigm. *Neurosurg Focus*, 31(6):E12, 2011.

[151] David W Shattuck and Richard M Leahy. Brainsuite: An automated cortical surface identification tool. *Medical Image Analysis*, 6(2):129 – 142, 2002.

[152] P. Shrout and J. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*, January 1979.

[153] Pawel Skudlarski, R.Todd Constable, and John C. Gore. Roc analysis of statistical methods used in functional mri: Individual subjects. *NeuroImage*, 9(3):311 – 329, 1999.

[154] Martha Skup. Longitudinal fmri analysis: A review of methods. *Statistics and its interface*, 3(2):235–252, 2010.

[155] Stephen M. Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155, November 2002.

[156] J. W. Snell, M. B. Merickel, J. M. Ortega, J. C. Goble, J. R. Brookeman, and N. F. Kassell. Segmentation of the brain from 3d mri using a hierarchical active surface template. In *In Proceedings of the SPIE Conference on Medical Imaging. SPIE*, pages 2–9, 1994.

[157] R. Soffietti, B.G. Baumert, L. Bello, A. Von Deimling, H. Duffau, M. Frenay, W. Grisold, R. Grant, F. Graus, K. Hoang-Xuan, M. Klein, B. Melin, J. Rees, T. Siegal, A. Smits, R. Stupp, and W. Wick. Guidelines on management of low-grade gliomas: report of an EFNS EANO Task Force. *European Journal of Neurology*, 17(9):1124–1133, 2010.

[158] Tao Song, M.M. Jamshidi, R.R. Lee, and Mingxiong Huang. A modified probabilistic neural network for partial volume segmentation in brain mr image. *Neural Networks, IEEE Transactions on*, 18(5):1424 –1432, sept. 2007.

[159] M Sonka, V Hlavac, and R Boyle. *Image Processing, Analysis and Machine Vision*. London: Chapman and Hall, 3 edition, 1993.

[160] M. Sonka, M.D. Winniford, and S.M. Collins. Robust simultaneous detection of coronary borders in complex images. *Medical Imaging, IEEE Transactions on*, 14(1):151 –161, mar 1995.

[161] M. Sonka, M.D. Winniford, and S.M. Collins. Robust simultaneous detection of coronary borders in complex images. *Medical Imaging, IEEE Transactions on*, 14(1):151 –161, mar 1995.

[162] M. Sonka, Xiangmin Zhang, M. Siebes, M.S. Bissing, S.C. Dejong, S.M. Collins, and C.R. McKay. Segmentation of intravascular ultrasound images: a knowledge-based approach. *Medical Imaging, IEEE Transactions on*, 14(4):719 –732, dec 1995.

[163] James A. Sorenson and X. Wang. ROC methods for evaluation of fMRI techniques. *Magnetic Resonance in Medicine*, 36(5):737–744, 1996.

[164] Friedrich Steimann. Fuzzy set theory in medicine, 1997.

[165] S. C. Strother, N. Lange, J. R. Anderson, K. A. Schaper, K. Rehm, L. K. Hansen, and D. A. Rottenberg. Activation pattern reproducibility: Measuring the effects of group size and data analysis models. *Human Brain Mapping*, 5:312–316, 1997.

[166] Subbanna and T. Arbel. Probabilistic gabor and markov random fields segmentation of brain tumours in mri volumes. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[167] Hidetomo Suzuki and Jun ichiro Toriwaki. Automatic segmentation of head mri images by knowledge guided thresholding. *Computerized Medical Imaging and Graphics*, 15(4):233 – 240, 1991. ¡ce:title¿NMR Image Processing and Pattern Recognition¡/ce:title¿.

[168] C. Tegeler, S. C. Strother, J. R. Anderson, and S. G. Kim. Reproducibility of BOLD-based functional MRI obtained at 4 T. 7(4):267–283, 1999.

[169] D.R. Thedens, D.J. Skorton, and S.R. Fleagle. A three-dimensional graph searching technique for cardiac border detection in sequential images and its application to magnetic resonance image data. In *Computers in Cardiology 1990, Proceedings.*, pages 57 –60, sep 1990.

[170] D.R. Thedens, D.J. Skorton, and S.R. Fleagle. Methods of graph searching for border detection in image sequences with applications to cardiac magnetic resonance imaging. *IEEE Transaction On Medical Imaging*, 14:42–55, 1995.

[171] D. Tian and L. Fan. A brain mr images segmentation method based on som neural network. In *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, pages 686 –689, july 2007.

[172] X. Tomas-Fernandez and S. K. Warfield. Automatic brain tumor segmentation based on a coupled global-local intensity bayesian mode. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.

[173] Alexander V. Tuzikov, Olivier Colliot, and Isabelle Bloch. Evaluation of the symmetry plane in 3d mr brain images. *Pattern Recognition Letters*, 24(14):2219 – 2233, 2003.

[174] Jayaram K. Udupa, Punam K. Saha, and Roberto A. Lotufo. Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(11):1485–1500, November 2002.

[175] J.K. Udupa and A. Lotufo. Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graph. Models Image Processing.*

[176] J.K. Udupa and S. Samarasekera. Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graph. Models Image Processing*, 58.

[177] Naonori Ueda and Ryohei Nakano. Deterministic annealing em algorithm. *Neural Networks*, 11(2):271 – 282, 1998.

[178] Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. Smem algorithm for mixture models. *NEURAL COMPUTATION*, 12(12):200–0, 1999.

[179] Analysis Group FMRIB Oxford UK. FMRIB software library. http://www.fmrib.ox.ac.uk/fsl/.

[180] Vladimir N. Vapnik. *Statistical learning theory.* 1998.

[181] Volkan Vural and Jennifer G. Dy. A hierarchical method for multi-class support vector machines. In *Proceedings of the twenty-first international conference on Machine learning*, ICML '04, pages 105–, New York, NY, USA, 2004. ACM.

[182] S. Phupanich L. O. Hall L. P. Clarke W. E. Phillips, R. P. Velthuizen and M. L. Silbiger. Application of fuzzy c-means segmentation technique for tissue differentiation in mr images of a hemorrhagic glioblastoma multiforme. *Magn. Reson. Imag*, 13(2):277–290, 1995.

[183] Amir Waks and Oleh J. Tretiak. Recognition of regions in brain sections. *Computerized Medical Imaging and Graphics*, 14(5):341 – 352, 1990. ¡ce:title¿Progress in Imaging in the Neurosciences Using Microcomputers and Workstations¡/ce:title¿.

[184] DC Walker and AH Kaye. Low grade glial neoplasms. *Jurnal of Clinical Neurosciece*, 10(1), Jan 2003.

[185] J.W. Wallis, T.R. Miller, C.A. Lerner, and E.C. Kleerup. Three-dimensional display in nuclear medicine. *Medical Imaging, IEEE Transactions on*, 8(4):297 –230, dec 1989.

[186] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*, 23(7):903–921, July 2004.

[187] Nikolaus Weiskopf. Real-time fMRI and its application to neurofeedback. *NeuroImage*, October 2011.

[188] G W Williams. Comparing the joint agreement of several raters with another rater. *Biometrics*, 32(3):619–27, 1976.

[189] G. Xiao, S. H. Ong, and K. W. C. Foong. Efficient partial-surface registration for 3d objects. *Comput. Vis. Image Underst.*, 98(2):271–294, 2005.

[190] L. Garnero J.-B. Poline H. Benali Y. Cointepas, J.-F. Mangin. Brain visa: software platform for visualization and analysis of multi-modality brain data. *Neuroimage*, 6:339–349, 2001.

[191] J. Zang, L. Liang, J. R. Anderson, L. Gatewood, D. A. Rottenberg, and S. C. Strother. A Java-based processing pipeline evaluation system for assessment of univariate general linear model and multivariate canonical variate analysis-based pipelines. *Neuroinformatics*, 6(2):123 – 134, 2008.

[192] Baibo Zhang, Changshui Zhang, and Xing Yi. Competitive em algorithm for finite mixture models. *Pattern Recognition*, 37(1):131 – 144, 2004.

[193] Dao-Qiang Zhang and Song-Can Chen. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 32(1):37 – 50, 2004.

[194] J. Zhang, K. Ma, M. H. Er, and V. Chong. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. In *in Proc. 2004 Int. Conf. Intell. Mechatronics Automation*, pages 207–211, 2004.

[195] Jianguo Zhang, Kai kuang Ma, Meng Hwa Er, and Vincent Chong. Tumor segmentation from magnetic resonance imaging by learning via one-class support vector machine. In *INTERNATIONAL WORKSHOP ON ADVANCED IMAGE TECHNOLOGY*, pages 207–211, 2004.

[196] Zhenzhen Zhang, Victor Chen, and Su Ruan. Graph cut based segmentation of brain tumor from mri images. In *9th international conference on Sciences and Techniques of Automatic control and computer engineering*, Sousse, Tunisie, December 2008.

[197] D. Zikic, B. Glocker, E. Konukoglu, J. Shotton, A. Criminisi, C. Demiralp D. H. Ye, O. M. Thomas, R. Jena T. Das, and S. J. Price. Context-sensitive classification forests for segmentation of brain tumor tissues. In *MICCAI Challenge on Multimodal Brain Tumor Segmentation*, October 2012.