

Combining Outlier Detection with Random Walker for Automatic Brain Tumor Segmentation

Vasileios G. Kanas, Evangelia I. Zacharaki, Evangelos Dermatas, Anastasios Bezerianos, Kyriakos Sgarbas, Christos Davatzikos

▶ **To cite this version:**

Vasileios G. Kanas, Evangelia I. Zacharaki, Evangelos Dermatas, Anastasios Bezerianos, Kyriakos Sgarbas, et al.. Combining Outlier Detection with Random Walker for Automatic Brain Tumor Segmentation. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.26-35, 10.1007/978-3-642-33412-2_3 . hal-01523040

HAL Id: hal-01523040
https://hal.science/hal-01523040

Submitted on 16 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Combining outlier detection with random walker for automatic brain tumor segmentation

Vasileios G. Kanas[1, 4*], Evangelia I. Zacharaki[2, 3], Evangelos Dermatas[1], Anastasios Bezerianos[2], Kyriakos Sgarbas[1], Christos Davatzikos[3]

[1]Department of Electrical and Computer Engineering, University of Patras, Patras, Greece
[2]Department of Medical Physics, School of Medicine, University of Patras, Patras, Greece
[3]Department of Radiology, University of Pennsylvania, Philadelphia PA, USA
*(vaskanas@upatras.gr)

Abstract. The diagnosis of brain neoplasms has been facilitated by the emerging of high-quality imaging techniques, such as Magnetic Resonance Imaging (MRI), while the combination of several sequences from conventional and advanced protocols has increased the diagnostic information. Treatment planning and therapy follow-up require the detection of neoplastic and edematous tissue boundaries, a very time consuming task when manually performed by medical experts based on the 3D MRI data. Automating the detection process is challenging, due to the high diversity in appearance of neoplastic tissue among different patients and, in many cases, similarity between neoplastic and normal tissue. In this paper, we propose an automatic brain tumor segmentation method based on a multilabel multiparametric random walks approach initialized by an outlier detection scheme. Segmentation assessment is performed by measuring spatial overlap between automatic segmentation and manual segmentation performed by medical experts. Good agreement is observed in most of the 26 cases for both neoplastic and edematous tissue. The highest achieved overlapping values were 0.74 and 0.79 for neoplastic and edematous tissue, respectively.

Keywords: brain neoplasms, segmentation, random walks, outlier detection, k- means, magnetic resonance imaging

1 Introduction

Diagnosis and characterization of brain neoplasms appears of greatest importance for remedial management. The procedure of diagnosis consists of two main steps, the extraction of the pathological region in the brain (tumor segmentation), and the assessment of malignancy of brain neoplasms (tumor classification). These two steps are essential for treatment planning and future assessment.

Despite various efforts and promising results in automatic brain tumor segmentation using MRI data, accurate segmentation and characterization of abnormalities is still a demanding and difficult task due to the large variability in shape, location and image intensity of brain neoplasms. Moreover, some neoplasms also distort the surrounding healthy structures or appear together with edema that changes intensity properties of

the nearby region making the extraction of neoplastic area a very complex assignment. There have been considerable efforts to develop automated computer algorithms for locating neoplasms in brain using MRI data. A review of pattern recognition methods for automatic brain tumor segmentation is presented in [1]. Other methods are based on learning a statistical distribution intensity model of the abnormal tissue types using manually labeled samples [2]. Such supervised classification methods are subject to inter- but also intra-rater variability[1]. In other relevant study [3], an atlas of the tumoral region has been extracted in order to augment a probabilistic map of healthy tissue priors aiming at identifying tumor boundaries. They incorporate prior knowledge about the normal brain tissue, which may not be accessible for all situations. In [4] it is propose a supervised variational method that incorporates appearance priors to better disambiguate the tumor from the surrounding deformed brain tissue. Yet, these appearance priors are not suitable for all patients. Nan Zhang et. al [5] integrate Support Vector Machine classification with a selection of the features in a kernel space to learn the tumor profile from the first MRI examination. Then, they proceed to refine the tumor boundaries using a region growing technique to follow up the brain tumor evolution. Though, they used a supervised learning process at the beginning of the patients' treatment.

Neoplastic tissue classification generally requires information of several MR protocols and contrasts, as T1, T1 contrast-enhanced T2, FLAIR (FL) or diffusion tensor imaging (DTI), yet not all sequences are acquired in clinical routine, especially during regular follow-up. The focus of this work is the development of an automatic method for brain tumor segmentation combining multiple MR sequences acquired in general clinical practice. The MR sequences used in this study were the following: axial 3D T1-weighted (T1), Fluid Attenuated Inversion Recovery (FLAIR), axial 3D contrast enhanced T1 contrast-enhanced (T1CE) images. We have chosen these modalities due to the complementary information they incorporate. In T1CE, the vital neoplastic tissue appears hyperintense and in FL datasets, the perifocal edema is hyperintense.

The proposed framework applies a hybrid methodology incorporating unsupervised clustering in the first step and the random walker algorithm [6] in the second step. The main contribution of this work is the automatic generation of seed points using a succession of procedures, which exploit the information provided by multi- modal MRI datasets. Such an approach is applicable to enhancing masses with different malignancy grades. It aims to segment normal and abnormal tissue including the non-necrotic enhancing neoplastic tissue (ET) and the peritumoral edema with possible inclusion of neoplastic infiltration (EDM). Its application to several datasets with various tumor shapes, intensities and locations shows that it can automatically detect and segment enhancing brain neoplasms and edematous regions with a good quality.

The rest of a paper is organized as follows. In the next section, the methods incorporated on our proposed framework are explained. In section 3, the data are described and experimental results are shown. Finally the last section is devoted to some discussion and concluding remarks.

[1] The same rater selecting different samples in two different time points

2 Materials and Methods

In this paper, we focus on the segmentation process of the different abnormal tissue types (neoplastic and edematous) and healthy brain tissue, i.e. white matter (WM), gray matter (GM), cerebrospinal fluid (CSF). The main purpose is to develop an automated tool to avoid the manual tracing of all tissue types, which is extremely time consuming when performed in 3D data and also highly dependent on the expertise of operators. Specifically, we develop a method based on random walker algorithm. This algorithm treats images as a graph, where user- defined seeds are required. Our contribution focuses on automatically finding seed points and segmenting intracranial neoplastic regions by combining several MRI sequences. This unsupervised approach incorporates imaging data which are acquired in a routine clinical protocol, such as multi-parametric conventional MRI. During our experiments we concluded that T1CE, FL and T1 images provide us with adequate information appropriate to our analysis, as inputs to random walker. Our approach could be separated in the following core steps (**Fig. 1**), which will be discussed in detail during this paper:

- Data acquisition and preprocessing.
- Automatically calculating seeds based on unsupervised clustering and outlier detection.
- Segmentation of all issue types using the random walker algorithm.

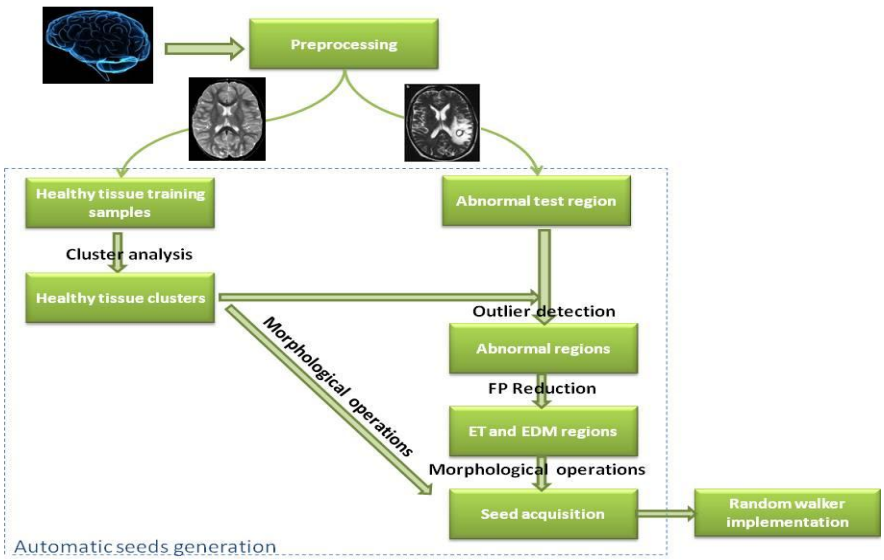


Fig. 1. Proposed framework for brain enhancing tumor (ET) and edema (EDM) segmentation

Particularly, based on the hypothesis that neoplastic and edematous tissue profile deviates from the healthy tissue profile, we first applied a non-parametric distance-based outlier detection methodology to approximately segment tumor and edema. We verified our hypothesis by plotting the T1CE and FL image intensities for abnormal and normal tissue based on the expert defined ROIs, as illustrated in **Fig. 2**. Generally it can be noticed that healthy tissues and abnormalities overlap with each other. This makes thresholding techniques inaccurate for the segmentation process. However, it can be observed that some of the pathological tissue (for both neoplastic and edematous regions) deviates significantly from the healthy tissue. The aim is to automatically detect these voxels and use them later as seeds to the random walker algorithm in order to better determine the separation boundary between normal and abnormal tissue based on local image properties.

More details on each methodological component are given next.

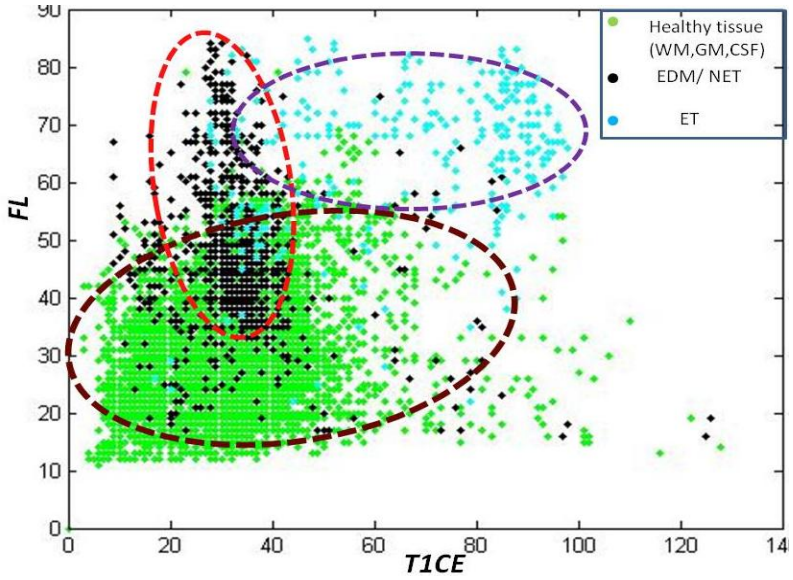


Fig. 2. T1CE and FL intensity distribution for healthy tissue (WM, GM, CSF), enhancing tumor (ET) and non-enhancing (NET) or edematous tissue (EDM)

2.1 Data description

The data and ROIs are acquired from a previous study [9] and described also here for completeness. Twenty-six patients (age 45-83 years) with a diagnosis of brain neoplasm were selected out of 97 with the criterion of the presence of tumor enhancement. All patients underwent biopsy or surgical resection of the neoplasm with histopathological diagnosis. The total of 26 brain masses were graded based on WHO criteria as metastasis (14), meningiomas grade I (3), gliomas grade II (2), gliomas grade III (2), and glioblastomas grade IV (5).

The images were preprocessed following a number of steps including noise reduction, bias-field correction, co-registration of all sequences (T1, T2, T1CE, FL), skull stripping and histogram matching to make the intensities comparable across patients. For further information about the preprocessing step see [9].

2.2 Extraction of healthy tissue clusters

The k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. The healthy tissue priors are calculated by segmenting the patient's T1 image into WM, GM and CSF/necrosis using the k-means algorithm. Several distance metrics were tested, however we used the Euclidean distance metric because it provided more accurate segmentation results for healthy tissue. During our analysis we combined tumor necrosis with the CSF cluster since the intensity profiles are similar and because necrosis is a non-vital tissue and consequently cannot be considered as malignancy. As a note, fuzzy c-means algorithm [7] was also tested without producing more accurate results.

2.3 Outlier detection

After calculating the cluster centers of healthy tissue, abnormalities are detected as samples that are distant to all three cluster centers (for WM, GM, CSF). The Mahalanobis distance (*MD*) is used as distance metric, and is calculated for voxel *j* as follows:

$$MD_i^2(j) = \sum_{i=1}^{3} (\mathbf{h}_j - \mathbf{c}_i)' \cdot \sigma_i^{-2} \cdot (\mathbf{h}_j - \mathbf{c}_i) \quad (1)$$

where $\mathbf{h}_j \in R^{\mathrm{p}}$ (p is the number of sequences used in this step) is the feature representation of voxel *j*, $\mathbf{c}_i$ is the center location of cluster *i* and σ_i^{-2} indicates the inverse intensity variance of the data in class *i* and it expresses the probability of each voxel to be an outlier. We implemented univariate analysis (p=1), e.g. T1CE and FL as two distinct feature vectors for ET and EDM were used, respectively. By calculating *MD* for every voxel and then thresholding the obtained *MD* maps for ET and EDM, we detect the neoplastic tissue and edematous tissue, as well as intensity artifacts.

The selected threshold value is not selected the same for all subjects because images might differ in contrast and brightness levels even if preprocessing is applied to factor out this variance. The threshold is automatically determined by examining the histogram of the MD maps for each subject and by selecting the 4% largest MD values for ET and EDM.

2.4 False Positive reduction

At this point, we have noticed that the thresholding technique introduces some false positives that must be eliminated. For example, the thresholding technique when applied on T1CE image detected pixels belonging to cerebellum. In order to reduce false

positives we constraint the number of possible tumor clusters to equal a predefined parameter (N_f) describing the number of tumor foci (e.g. 1 or 2). The retained tumor clusters were the N_f largest (regarding total volume) spatially connected components. Connected-component analysis was performed by application of the union-find algorithm [8].

2.5 Final seeds acquisition

Once the seeds regions have been segmented, we selected as seeds voxels with higher likelihood to be correct detection. Those are voxels in the interior of the regions since the uncertainty is higher towards the boundaries due to the diffusive nature of many high grade tumors and due to partial volume effects. Particularly, we applied skeletonization operation on the labeled image, because the skeleton retains spatial and topological properties of the object while containing all the information necessary to reconstruct the object. Assuming the boundaries of the region are unclear, the skeletonization technique retains seeds belonging to the specific region with higher certainty. Furthermore, we implemented the same morphological operations to the segmented healthy tissue (**section 2.2**) in order to obtain seeds for the healthy tissues (WM, GM and CSF).

2.6 Random Walker algorithm (Theoretical background)

We propose a multi-parametric framework for automated brain tumor segmentation by extracting intensity attributes and defining a precise notion for a graph. For a given image, we can use the random walk model or circuit model to solve the segmentation problem [6]. The random walk method consists of computing the probability that a "random walker", starting at an arbitrary node, first reaches a pre-assigned node with a particular label. The starting node is then given the label with the highest probability. Assuming that the image is a given graph, each pixel in the graph represents a vertex of the circuit, which connects with 4 neighborhood pixels (if the algorithm is applied on a 2D slice) with weights. The total weight between pixel *i* and pixel *j* is then defined as

$$r_{ij} = e^{-\lambda ||h_i - h_j||^2} \quad (2)$$

where $\mathbf{h}_i$ and $\mathbf{h}_j$ are the intensities of pixels *i, j* respectively and λ is a constant penalizing dissimilarity. Low edge weights indicate high probabilities of region margin evidence between two neighboring pixels and avoid a random walker crossing these boundaries. We used 3- dimensional vectors for $\mathbf{h}_{i,j}$ corresponding to T1CE, T1 and FL intensities.
Equation (2) is a typical Gaussian weighting function and during our experiments the free parameter λ was assigned to its default value (λ=90). Results showed that the segmentation was not sensitive to the choice of λ.
The steps of random walker algorithm are the following:

(i) Construct an undirected weighted graph representing the image. Each pixel corresponds to vertices of the graph and image intensities are mapped to edge weights (equation (2)).

(ii) Obtain automatically a set of marked nodes with K labels. In general, seeds are produced interactively by the user. However, we generate the seeds automatically, as described in the previous section. Moreover, in our case the number of labels are K=6 (ET, EDM, GM, WM, CSF, background of the image).

(iii) Solve a linear system of equations (see [6] for more details) to obtain probabilities for the unmarked nodes.

(iv) Assign to each node the label corresponding to maximum probability obtaining a final segmentation of the image.

2.7 Method evaluation

The results of our automated method are compared with manually segmented ROIs. Manual segmentation was performed by experts by tracing hyperintensities in FL images and labeling them as EDM (including possible tumor infiltration). Hyperintensities in T1CE were marked as ET. The automatic segmentation of edematous and neoplastic regions is assessed using the Dice coefficient as measure of spatial overlap. The Dice coefficient (DSC) over two segmentations, L1 and L2, is the volume of intersection of the two segmentations divided by the mean of the two segmentation volumes:

$$DSC = 2\,|L_1 \cap L_2|/(|L_1| + |L_2|) \tag{3}$$

3 Experimental Results

3.1 Segmentation of enhancing tumor, edematous region and healthy tissue

Fig. 3 and **Fig.4** illustrate some examples of enhancing tumor and edema segmentation by the proposed method and **Fig. 5** demonstrates the quantitative segmentation assessment of our method for both ET and EDM tissues. Overall good agreement with manual segmentation is observed. In general, the results shown in **Fig. 3** and **Fig.4** are encouraging and prove the efficiency of the method to segment particular tumor tissues. It's noteworthy that this method can utilize the characteristics of the enhancing neoplastic regions and take more general and complete pathologic information into account if different weighting functions or features are chosen. A limitation of random walker algorithm is observed in situations, where the number of unmarked regions exceeds the number of seed/label groups. In such cases ambiguous segmentation results could occur. The probability inside an unseeded region could be defined as the average of its neighboring regions, weighted by the contrast between the re-

gions and the level of mutual surface area. Therefore, an unmarked region that shares an equal surface area with two labeled regions will be assigned to the region for which it has a lower contrast. On the other hand, if an unmarked region has the same contrast with two labeled regions, it will be assigned to the seeded region with which it shares a greater boundary. However, if the contrast and the overlapping surface area of an ambiguous region with two seeded regions are equal, the ambiguous region would be divided in half with respect to the two labels [6]. The above hypothesis becomes clear, if we examine patient 10, where necrotic tissue is surrounded by the ET region. Such a case is plausible since we have not taken into account necrosis as a separate class; thus the calculation of DSC is affected.

In addition, observing either T1CE or FL sequences, we notice that in most cases the ROIs are only approximately drawn by the experts and do not follow the exact boundaries visible in the MR images. This might be due to either insufficiency of the incorporated modalities to accurately reflect boundaries or due to incorporation of additional knowledge by the expert during manual segmentation. For example experts might better appreciate the biophysical tissue changes. Recent approaches try to predict these changes by modeling tumor growth and infiltration instead of only looking at image intensity characteristics [10].

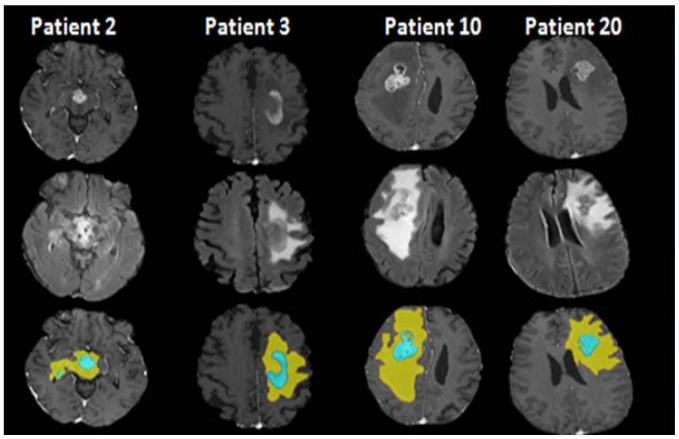


Fig. 3. Four patients diagnosed with brain neoplasms. From top to bottom: T1CE sequence, FL sequence, ROIs traced by expert (yellow: EDM, light blue: ET),

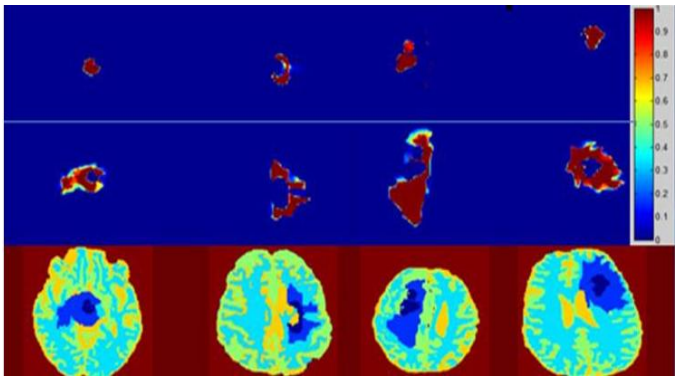


Fig. 4. Segmentation of enhancing tumor by the proposed method. From top to bottom: Probability maps of ET, Probability maps of FL, final segmented image (dark blue: ET, light blue: EDM, cyan: WM, green: GM, yellow: CSF, red: background)

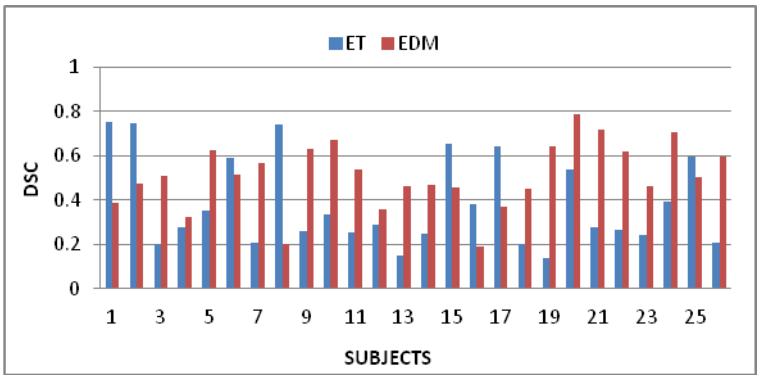


Fig. 5. Segmentation assessment for edematous and neoplastic tissues for 26 subjects

4 Conclusion

The segmentation of brain neoplasms is a complex task. Our system attempted to segment the enhancing tumor area, the region that appears bright on T1CE sequences, and edematous area, the region that appears bright on FL sequences. Note that it is not sufficient to simply threshold T1CE images by brightness because other tissues can have the same range of intensities. Here, we present an unsupervised hybrid method, combining outlier detection with random walker. The proposed method includes two steps: 1) The automatic generation of seeds and 2) implementation of multiparametric random walker algorithm. Firstly, abnormal regions were extracted using an outlier detection process. Secondly, reduction of false positives was implemented using con-

nected components analysis resulting in the construction of ET and EDM components. The seeds for abnormal and normal tissue were obtained by retaining voxels with high certainty level. Finally, multilabel multiparametric random walker was applied and the segmentation results were validated against manual segmentation. The results showed that our method can produce very satisfactory segmentation outcomes while being fully automated. In this work, we used the typical Gaussian weighting function and investigated only intensity attributes (equation (2)). The analysis in this paper concerns only enhancing tumors with possible edema; Consequently, in future, we intend to extract more subtle attributes using more advanced methods, such as wavelet transform or shape features to exploit the information provided by MRI sequences.

Acknowledgments

This research was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme and an NIH grant R01 NS042645.

References

1. Jason J. Corso, Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha, and Alan Yuille, "Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification", IEEE Transactions on medical imaging, Vol. 27, No. 5, 2008
2. Verma, R., Zacharaki, E.I., Ou, Y., Cai, H., Chawla, S., Lee, S.-K., Melhem, E.R., Wolf, R., Davatzikos, C. Acad Radiol, "Multiparametric Tissue Characterization of Brain Neoplasms and Their Recurrence Using Pattern Classification of MR Images", Academic Radiology, Vol 15, No 8, pp. 966–977, 2008
3. Menze, B.H., Van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P., "A generative model for brain tumor segmentation in multi-modal images". Med Image Comput Comput Assist Interv, Vol. 13, Part 2, pp. 151–159, 2010
4. Popuri K, Cobzas D, Murtha A, Jagersand M, "3D variational brain tumor segmentation using Dirichlet priors on a clustered feature set". Int J Comput Assist Radiol Surg doi: 10.1007/s11548- 011-0649-2, 2011
5. Zhang N., Ruan, S., Lebonvallet, S., Liao, Q., Zhu, Y., "Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation", Computer Vision and Image Understanding Vol. 115, Issue 2, pp. 256-269, 2011
6. Leo Grady, "Random Walks for Image Segmentation", IEEE transactions on pattern analysis and machine intelligence, vol. 28, No. 11, 2006
7. J.-C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, 1981
8. Sedgewick, Robert, "Algorithms in C", 3rd Ed., Addison-Wesley, pp. 11-20, 1998
9. Zacharaki EI, Wang S, Chawla S, Yoo DS, Wolf R, Melhem ER, Davatzikos C, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme", Magnetic resonance in Medicine, Vol.62, pp.1609-1618, 2009
10. Ali Gooya, George Biro, and Christos Davatzikos, "Deformable Registration of Tumor Invaded Images Using EM Algorithm and Glioma Growth Modeling", IEEE transactions on medical imaging, 2010