Automatic Brain Tumor Segmentation from MR Images via a Multimodal Sparse Coding Based Probabilistic Model

Yuhong Li*, Qi Dou†, Jinze Yu†§, Fucang Jia*, Jing Qin‡ and Pheng-Ann Heng†
*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
†Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong
‡National-Regional Key Technology Engineering Laboratory for Medical Ultrasound
School of Medicine, Shenzhen University, Shenzhen, China
§Corresponding author, e-mail: jzyu@cse.cuhk.edu.hk

***Abstract*—Accurate segmentation of brain tumor from MR image is crucial for the diagnosis and treatment of brain cancer. We propose a novel automated brain tumor segmentation method based on a probabilistic model combining sparse coding and Markov random field (MRF). We formulate the brain tumor segmentation task as a pixel-wise labeling problem with regard to three classes: tumor, edema and healthy tissue. For each class, dictionary learning is performed independently on multi-modality gray scale patches. Sparse representation is then extracted based on a joint dictionary which is constructed by combing the three independent dictionaries. Finally, we build the probabilistic model aiming to estimate maximum a posterior (MAP) probability by introducing the sparse representation into likelihood probability and prior probability using the Markov random field (MRF) assumption. Compared with traditional methods, which employed hand-crafted low level features to construct the probabilistic model, our model can better represent the characteristics of a pixel and its relation with neighbors based on the sparse coefficients obtained from the learned dictionary. We validated our method on the MICAAI 2012 BRATS challenge brain MRI dataset and achieved comparable or better results compared with state-of-the-art methods.**

***Keywords*-brain tumor segmentation, multi-modality, sparse coding, probabilistic model, maximum a posterior**

I. Introduction

Brain tumor is the abnormal tissue proliferation which causes the increase of intra-cranial pressure, resulting in the damage of central nervous system, and endangers the lives of patients. Reliable segmentation of brain tumors from Magnetic Resonance (MR) images can aid surgical planning and therapy assessment during clinical treatment [1]. Currently standard manual segmentation is time consuming and suffers from inter- and intra- rater variability with limited reproduction. Therefore, automatic and semi-automatic brain tumor segmentation methods play an increasingly important role in modern medical image analysis. However, this remains a challenging problem due to the unpredictable appearance, location, shape, size of tumor and the overlapping intensity range of the tumor and healthy tissues.

In the past decades, a variety of brain tumor segmentation algorithms have been proposed. Early investigations employing active contours and geometric deformable models suffer from the inhomogeneity, unpredictable location and complex structural appearance of brain tumor [2]. Recently, probability-based [3] and graph-based [4] methods have been proposed to meet these challenges. Unfortunately, the hand-crafted low level features employed in these methods are still insufficient to deal with the unpredictable appearance of brain tumor. Machine learning-based methods [5] are then proposed, attempting to exploit high level features generated from training samples. However, these methods tend to neglect the spatial constraints, as they usually directly adopt the output of their classifier as the segmentation mask, which degrades the segmentation performance.

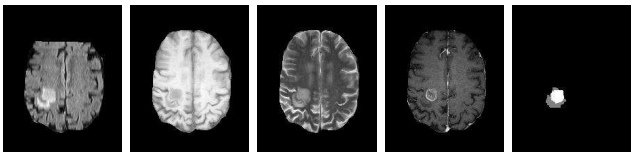

Figure 1. Examples of various modalities of brain tumor images. Form left to right are: Flair, T1, T2, T1C and our segmentation result (white and gray parts represent the tumor and edema, respectively).

In this paper, we build machine learning strategies into a graph model so that the spatial constraints can be sufficiently considered. We formulate the brain tumor segmentation task as a pixel-wise labeling problem with regard to three classes: tumor, edema and healthy tissue. Our automatic segmentation model includes three steps: multi-modality dictionary learning, sparse coding from a joint dictionary and maximum a posterior (MAP) probabilistic model construction. First, we learn a collection of overcomplete atoms by online dictionary learning based on multi-modality brain MR images. Second, given a target image, sparse representation is acquired from the learned joint dictionary. Finally, the sparse representation is adopted for building the likelihood probability density and Markov Random Field (MRF) is used to construct a probabilistic graph. The MAP prediction is taken as the output label. Compared with traditional methods, which employed hand-crafted low level features to construct the probabilistic model, our model can better

978-1-4673-7145-2/15 $31.00 © 2015 IEEE
DOI 10.1109/PRNI.2015.18

represent the characteristics of a pixel and its relation with neighbors based on the sparse coefficients obtained from the learned dictionary. In addition, the sparse representation calculated from multiple modalities should be more powerful than representations from single modality in capturing the complex structure of brain tumor. Experimental results show that our method is comparable to or better than state-of-the-art approaches.

II. Method

A. Multi-modality Dictionary Learning

We first independently trained three dictionaries for the three classes of tissue respectively. For each dictionary, we get training samples from four modalities: FLAIR, T1-weighted (T1), T2-weighted (T2) and contrast enhanced T1-weighed (T1C). In this way, the learned dictionary can benefit from the complementary information of multiple modalities. For example, the T1 weighted modality edema and non enhancing tumor areas share the range of intensities with the gray matter whereas the FLAIR modality can separate the edema with a hype-intense signal around the tumor, as shown in Fig 1. Specifically, we sampled small patches from the four modalities and construct a sample vector $\mathbf{x}$ by concatenating the four smaller vectors which are reshaped from the four patches.

When training a dictionary, with a given set of N training samples $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N], \mathbf{x}_n \in \mathbb{R}^M$, we aim to solve the following optimization problem, where the ℓ_1-norm constraint helps to yield a sparse solution $\boldsymbol{\alpha}$:

$$\min_{\boldsymbol{D},\, \boldsymbol{\alpha}\in\mathbb{R}^{K\times N}} \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{2}\|\mathbf{x}_n - \mathbf{D}\boldsymbol{\alpha}_n\|_2^2 + \lambda\|\boldsymbol{\alpha}_n\|_1\right) \quad \text{s.t.} \quad \|\mathbf{d}_k\|_2^2 \le 1,\ \forall\ k = 1, ..., K. \tag{1}$$

where $\boldsymbol{D} \in \mathbb{R}^{M\times K}$ is the learned dictionary whose column vector $\mathbf{d}_k$ is constrained to prevent $\boldsymbol{D}$ from being arbitrarily large or equivalently to prevent $\boldsymbol{\alpha}$ from being arbitrarily small, and $\boldsymbol{\alpha}_n$, corresponding to the n-th column of matrix $\boldsymbol{\alpha}$, is the sparse coefficients vector of the n-th input $\mathbf{x}_n$.

When training the dictionary, we need to jointly minimize the objective function (1) over $\boldsymbol{D}$ and $\boldsymbol{\alpha}$. This non-convex optimization problem can be converted into two convex problems with respect to $\boldsymbol{D}$ or $\boldsymbol{\alpha}$ by fixing the other. In our implementation, we adopt the online dictionary learning algorithm to train the dictionary [6].

We obtain three dictionaries $\boldsymbol{D}_l \in \mathbb{R}^{M\times K}$ $(l = T, E, B)$ which are corresponding to the tumor, edema and background healthy tissue respectively. Then, the three dictionaries are combined to construct the joint dictionary $\boldsymbol{D}_J = [\boldsymbol{D}_T, \boldsymbol{D}_E, \boldsymbol{D}_B] \in \mathbb{R}^{M\times 3K}$ which is used for seeking sparse representation (or equivalently sparse coding) towards a new test sample.

B. Sparse Representation of Joint Dictionary

Given the joint dictionary $\boldsymbol{D}_J$, we can represent a given new test patch with sparse linear combination of the atoms. We denote the test sample as $\mathbf{x}_i \in \mathbb{R}^M$ and the sparse coefficients vector as $\boldsymbol{\beta}_i \in \mathbb{R}^{3K}$. The sparse representation of a test sample is obtained by minimizing the following objective function over $\boldsymbol{\beta}_i$

$$\min_{\boldsymbol{\beta}_i\in\mathbb{R}^{3K}} \frac{1}{2}\|\mathbf{x}_i - \boldsymbol{D}_J\boldsymbol{\beta}_i\|_2^2 + \lambda\|\boldsymbol{\beta}_i\|_1. \tag{2}$$

where the λ is used to balance the sparsity of the solution and fidelity of the approximation to $\mathbf{x}_i$. Different from the dictionary learning stage, here we only minimize the objective over $\boldsymbol{\beta}_i$ with the joint dictionary $\boldsymbol{D}_J$ fixed. We employ the LARS-Lasso algorithm [7] to solve this ℓ_1-norm regularization problem.

C. MAP Model Based on Sparse Representation

Given the test sample $\mathbf{x}_i$ $(i = 1, ..., I)$, we can get its sparse representation $\boldsymbol{\beta}_i$ based on the joint dictionary $\boldsymbol{D}_J$. We then extract the coefficients with respect to each class for classification. More specifically, for each potential class $l_i \in \{T, E, B\}$, we denote by $\delta_{l_i} : \mathbb{R}^{3K} \to \mathbb{R}^{3K}$ the characteristic function which has value 1 on class l_i and 0 otherwise. Then the given test sample $\mathbf{x}_i$ can be approximated as

$$\hat{\mathbf{x}}_i = \boldsymbol{D}_J\delta_{l_i}(\boldsymbol{\beta}_i), \tag{3}$$

where $\hat{\mathbf{x}}_i$ can be used for modeling estimation of classification. We propose to wrap the sparse representation into the maximum a posterior probabilistic (MAP) model by combing a Markov random field (MRF). The MAP prediction is taken as the output label:

$$\hat{l}_i = \arg\max_{l_i}\ p\,(l_i \mid \mathbf{x}_i), \tag{4}$$

where $\hat{l}_i \in \{T, E, B\}$ is the output label of the sample $\mathbf{x}_i$, from the Bayes rule, we can obtain:

$$p\,(l_i \mid \mathbf{x}_i) = \frac{p\,(\mathbf{x}_i \mid l_i)\,p\,(\,l_i\,)}{p(\mathbf{x}_i)} \propto p\,(\mathbf{x}_i \mid l_i)\,p\,(\,l_i\,), \tag{5}$$

where $p\,(\mathbf{x}_i \,|\, l_i)$ is the likelihood function, $p\,(\,l_i\,)$ is the prior probability for the class label l_i and $p\,(\mathbf{x}_i)$ is the density of $\mathbf{x}_i$ which is a constant given the sample $\mathbf{x}_i$.

We model the likelihood probability by evaluating the distance between the reconstructed data and the original data as follows:

$$p\,(\mathbf{x}_i \,|\, l_i) \propto \exp(-\mathcal{R}(\mathbf{x}_i, l_i)),\ \mathcal{R}(\mathbf{x}_i, l_i) = \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2, \tag{6}$$

Meanwhile, we use the MRF to model the prior probability $p\,(\,l_i\,)$ in Eq. 5 of the sample $\mathbf{x}_i$ as follows:

$$p\,(\,l_i\,) \propto \exp\,(-\sum_{j\in\mathcal{N}(i)} V_{i,j}(l_i, l_j)\,),$$

$$V_{i,j}(l_i, l_j) = c \cdot \exp\,(\frac{-\|\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j\|^2}{\sigma^2}) \cdot (1 - \boldsymbol{\delta}(l_i, l_j)), \tag{7}$$

where $\mathcal{N}(i)$ is the spatial 4-neighborhood of pixel i, l_i and l_j are output labels of pixel i and pixel j, respectively, and $1-\boldsymbol{\delta}(l_i, l_j)$ penalizes different class labels of adjacent pixels. $V_{i,j}(l_i, l_j)$ is the spatial smoothness term which regulates the degree of smoothness of the local intensity variation.

From Eq. 5, 6 and 7, we get the posterior probability as:

$$p\left(l_i \mid \mathbf{x}_i\right) \propto \exp\left(-\mathcal{R}(\mathbf{x}_i, l_i) - \sum_{j \in \mathcal{N}(i)} V_{i,j}(l_i, l_j)\right), \quad (8)$$

which is converted to the following by taking logarithm:

$$\log p\left(l_i \mid \mathbf{x}_i\right) \propto -(\mathcal{R}(\mathbf{x}_i, l_i) + \sum_{j \in \mathcal{N}(i)} V_{i,j}(l_i, l_j)), \quad (9)$$

Denoting by $L = [l_1, ..., l_i, ..., l_I]$, the MAP problem can be written into an energy minimizing problem:

$$\min_{L} \sum_{i=1}^{I} \mathcal{R}(\mathbf{x}_i, l_i) + \sum_{i=1}^{I} \sum_{j \in \mathcal{N}(i)} V_{i,j}(l_i, l_j). \quad (10)$$

We adopt the graph cuts to approximately solve the minimizing energy optimization problem by finding the minimal cuts in a graph model with reference to the multi-label optimization via α-expansion move in [8].

III. Experiments

A. Dataset

To validate our method, we employed the BRATS 2012 high-grade brain tumor dataset of MICCAI Challenge, which is consist of both synthetic dataset (25 training and 10 test cases) and real dataset (20 training and 11 test cases). For each subject, four modalities of MR volumes are available: FLAIR, T1, T2 and T1C. In the pre-processing stage, the volumes were linearly co-registered to the T1 contrast images with skull stripped and intensity normalized to the interval $[0, 1]$. The parameters were set as patch size of 7×7, the dictionary size of $K = 784$ and the sparse factor of $\lambda = 0.15$. To further boost the performance, we applied largest connected component labeling for post-processing.

B. Evaluation on Synthetic Dataset

As the ground truth mask for test data is unavailable for public downloading, we validated our model on the training set with the leave-one-out strategy for visual evaluation. As shown in Fig. 2, our method can accurately segment out the brain tumor core and the surrounding edema tissue when comparing with the ground truth.

Meanwhile, we evaluated our segmentation results on the synthetic test dataset with the public online evaluation tool provided by the challenge which is accessible to the ground truth mask. The segmentation results were evaluated with three metrics including dice similarity coefficient, specificity and sensitivity[1]. The evaluation results shown in Table I proved the effectiveness of our algorithm for segmenting out the tumor and edema tissues.

[1]http://www2.imm.dtu.dk/projects/BRATS2012/evaluation.html

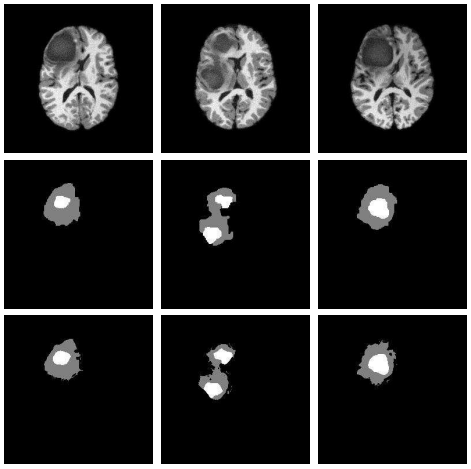

Figure 2. Examples of segmentation results on synthetic subjects. Each column corresponds to one subject, from top to bottom are: T1, our results and ground truth images.

Table I
SEGMENTATION RESULTS ON SYNTHETIC DATASET.

Metrics	Dice		Specificy		Sensitivity	
	edema	tumor	edema	tumor	edema	tumor
mean	0.852	0.861	0.998	0.999	0.871	0.927
std	0.080	0.097	0.001	0.001	0.083	0.099

C. Evaluation on Real Dataset

Typical segmentation results on the real training dataset are shown in Fig. 3. Visually, our results are very close to the ground truth. The evaluation results in Table II further demonstrate the validity of our algorithm. Despite the intensity normalization during the pre-processing stage, the inter-patients intensity variation of the real clinical data still degraded the effectiveness of the learned dictionary, and our algorithm failed on the cases of HG131 and HG137.

Table II
SEGMENTATION RESULTS ON REAL DATASET.

Metrics	Dice		Specificy		Sensitivity	
	edema	tumor	edema	tumor	edema	tumor
mean	0.551	0.515	0.995	0.999	0.630	0.494
std	0.157	0.294	0.005	0.001	0.200	0.348

D. Comparison with State-of-the-art Approaches

We compared our method with the top four ranking algorithms in MICCAI BRATS 2012 challenge [9]. These algorithms are context-sensitive classification forests (CCF) [10], hierarchical classification and regularization (HCR) [11], hierarchical random walker (HRW) [12], Gabor and markov random fields (GMRF) [13]. To make the comparison more convincing, we employed the online published challenge results[2] of BRATS2012 which utilized the same dataset and evaluation tool.

We compared three performance metrics on the high-grade synthetic and real dataset in Table III and IV, respectively. For synthetic data, it is observed that our method is comparable in all the metrics to CCF which obtained the best

[2]http://www2.imm.dtu.dk/projects/BRATS2012/results.html

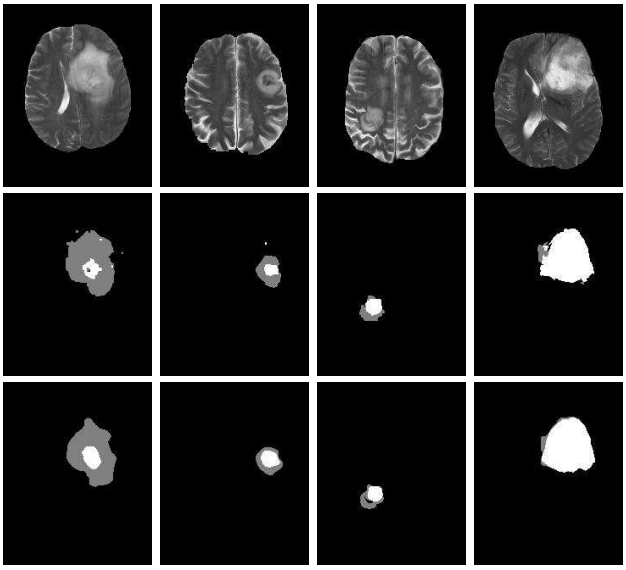

Figure 3. Examples of segmentation results on real data. Each column corresponds to one subject, from top to bottom are: T2, our results and ground truth images.

Table III
COMPARISON OF DIFFERENT METHODS ON SYNTHETIC DATA.

Methods	Dice		Specificy		Sensitivity	
	edema	tumor	edema	tumor	edema	tumor
CCF	0.850	**0.869**	**0.999**	**0.999**	0.851	**0.970**
HCR	0.785	0.779	**0.999**	**0.999**	0.789	0.809
GMRF	0.696	0.398	**0.999**	0.992	0.598	0.733
HRW	0.343	0.414	0.997	0.989	0.324	0.881
Ours	**0.852**	0.861	0.998	**0.999**	**0.871**	0.927

evaluation scores in challenge and our method outperforms the other three methods in most metrics. For the real dataset, our dice overlap is lower than CCF for the edema, but higher than CCF for the tumor. Compared with the synthetic data, the segmentation accuracy decreased upon the real data, due to the inter-slices and and inter-patients intensity variation. Nevertheless, our performance still achieved good results compared to state-of-the-art algorithms.

IV. CONCLUSION

In this paper, we formulate the multimodal MR brain tumor segmentation task as a three class (tumor, edema and normal tissue) classification problem by a sparse-coding based probabilistic model. Experimental results on both synthetic and real patient datasets validated the efficacy of our model. In the future, we will extend our model into 3D by training the joint dictionary with volume patches and segment the target image into more classes.

ACKNOWLEDGMENT

The work described in this paper was supported by the National Basic Research Program of China, 973 Program (Project no. 2015CB351706), Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 412412) and Natural Science Foundation of Guangdong (Project No. S2013010014973).

Table IV
COMPARISON OF DIFFERENT METHODS ON REAL DATA.

Methods	Dice		Specificy		Sensitivity	
	edema	tumor	edema	tumor	edema	tumor
CCF	**0.598**	0.476	**0.997**	**1**	**0.67**	0.397
HCR	0.536	0.512	0.994	0.999	0.627	0.471
HRW	0.539	0.337	0.983	0.998	0.63	0.359
GMRF	0.166	0.248	-	-	0.056	**0.527**
Ours	0.551	**0.515**	0.995	0.999	0.63	0.494

REFERENCES

[1] N. Gordillo, E. Montseny, and P. Sobrevilla, "State of the art survey on MRI brain tumor segmentation," *Magnetic resonance imaging*, vol. 31, no. 8, pp. 1426–1438, 2013.

[2] T. Wang, I. Cheng, and A. Basu, "Fluid vector flow and applications in brain tumor segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 3, pp. 781–789, 2009.

[3] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *MICCAI*. Springer, 2011, pp. 354–361.

[4] M. Wels, G. Carneiro, A. Aplas, M. Huber, J. Hornegger, and D. Comaniciu, "A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-d MRI," in *MICCAI 2008*. Springer, 2008, pp. 67–75.

[5] A. Davy, M. Havaei, D. Warde-Farley, A. Briard, L. Tran, P.-M. Jodoin, A. Courville, H. Larochelle, C. Pal, and Y. Bengio, "Brain tumor segmentation with deep neural networks," in *MICCAI - BRATS*, 2014.

[6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *ICML*. ACM, 2009, pp. 689–696.

[7] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 11, pp. 1222–1239, 2001.

[9] B. Menze, M. Reyes, and K. Van Leemput, "The multimodal brain tumorimage segmentation benchmark (brats)," *Medical Imaging, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[10] D. Zikic, B. Glocker, E. Konukoglu, J. Shotton, A. Criminisi, D. Ye, C. Demiralp, O. Thomas, T. Das, R. Jena *et al.*, "Context-sensitive classification forests for segmentation of brain tumor tissues," *Proc MICCAI-BRATS*, 2012.

[11] S. Bauer, T. Fejes, J. Slotboom, R. Wiest, L. P. Nolte, and M. Reys, "Segmentation of brain tumor images based on integrated hierarchical classification and regularization," in *Proc. MICCAI-BRATS Challenge 2012 on Multimodal Brain Tumor Segmentation*, 2012, pp. 10–13.

[12] Y. Xiao and J. Hu, "Hierarchical random walker for multimodal brain tumor segmentation," in *Proc. MICCAI-BRATS Challenge 2012 on Multimodal Brain Tumor Segmentation*, 2012, pp. 36–40.

[13] N. K. Subbanna and T.Arbel, "Probabilistic gabor and markov random fields segmentation of brain tumor in MRI volumes," in *Proc. MICCAI-BRATS Challenge 2012 on Multimodal Brain Tumor Segmentation*, 2012, pp. 36–40.