Semiautomatic tumor segmentation with multimodal images in a conditional random field framework

Yu-chi Hu
Michael Grossberg
Gikas Mageras

Semiautomatic tumor segmentation with multimodal images in a conditional random field framework

Yu-chi Hu,^{a,b,*} Michael Grossberg,^b and Gikas Mageras^a

^aMemorial Sloan Kettering Cancer Center, Department of Medical Physics, 1275 York Avenue, New York, New York 10065, United States

^bCity College of New York, Department of Computer Science, 160 Convent Avenue, New York, New York 10031, United States

Abstract. Volumetric medical images of a single subject can be acquired using different imaging modalities, such as computed tomography, magnetic resonance imaging (MRI), and positron emission tomography. In this work, we present a semiautomatic segmentation algorithm that can leverage the synergies between different image modalities while integrating interactive human guidance. The algorithm provides a statistical segmentation framework partly automating the segmentation task while still maintaining critical human oversight. The statistical models presented are trained interactively using simple brush strokes to indicate tumor and nontumor tissues and using intermediate results within a patient's image study. To accomplish the segmentation, we construct the energy function in the conditional random field (CRF) framework. For each slice, the energy function is set using the estimated probabilities from both user brush stroke data and prior approved segmented slices within a patient study. The progressive segmentation is obtained using a graph-cut-based minimization. Although no similar semiautomated algorithm is currently available, we evaluated our method with an MRI data set from Medical Image Computing and Computer Assisted Intervention Society multimodal brain segmentation challenge (BRATS 2012 and 2013) against a similar fully automatic method based on CRF and a semiautomatic method based on grow-cut, and our method shows superior performance. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.3.2.024503](https://doi.org/10.1117/1.JMI.3.2.024503)]

Keywords: semiautomatic segmentation; tumor; multimodality imaging; conditional random field; logistic regression.

Paper 15218RR received Nov. 11, 2015; accepted for publication Jun. 2, 2016; published online Jun. 28, 2016.

1 Introduction

Accurate spatial delineation of the tumor and the critical normal organs is an essential factor of the treatment outcome in radiotherapy. The treatment plan is designed to deliver a high dose of radiation to the tumor while restricting a tolerable dose to adjacent normal tissues. For this purpose, multiple image modalities are often used for better visualization of the tumor and its extent. Computed tomography (CT) provides high resolution of anatomical structures and remains the clinical standard for volume definition and dose calculation. One limitation of CT is the lack of contrast resolution for soft-tissue structures. Magnetic resonance imaging (MRI), on the other hand, improves soft tissue contrast in many anatomical sites. Positron emission tomography (PET) reveals functional information that is invisible in CT and MRI. However, challenges arise in delineation by human experts when multiple image modalities are used simultaneously. Studies have shown that tumor volumes delineated in a patient with different image modalities do not agree well^{1–3} and are not interchangeable,² leading to variations in dose conformation, which is critical in modern high-precision radiotherapy techniques such as intensity modulated radiotherapy and proton beam therapy.

We have developed a semiautomatic approach that addresses the need within computational methods to facilitate the human interpretation of quantitative information from multiple image modalities, with a goal to obtain a more consistent segmentation across modalities. Despite increasing progress in fully automatic segmentation, methods have not yet reached reliable quality.^{4–6}

Human interventions are, therefore, still needed and required in clinical practice to proofread the results, resulting in little or no significant time saving compared to semiautomatic approaches.⁴ Therefore, clinicians have expressed a need for semiautomatic tools to delineate anatomic volume in radiotherapy,⁷ which is the focus of our work. Our method makes use of adaptive machine learning, guided by human experts, in the conditional random fields (CRFs) framework.⁸ We define purely probabilistic regional and boundary terms in the energy function of a CRF, which can be minimized with a max-flow/min-cut algorithm to obtain a segmentation. The terms are statistically estimated from logistic regression models of which the parameters are interactively learned from the expert user using brush strokes to indicate tumor and nontumor tissue samples. The same tool allows for correction of the segmentation as well as refining of the model parameters. An overview of our segmentation method is shown in Fig. 1.

To our knowledge, a semiautomatic approach that uses patient-specific statistical boundary model of the tumor from multimodal medical images for the pairwise potentials in CRF has not been previously reported. In addition, using the contour delineation tool familiar to human experts for training both regional and boundary models is highly desirable in clinical practice. We evaluated our method with 20 high-grade and 10 low-grade tumor cases from four modalities of MRI images, which were obtained from Medical Image Computing and Computer Assisted Intervention Society (MICCAI) brain tumor segmentation competitions in 2012 and 2013.

*Address all correspondence to: Yu-chi Hu, E-mail: huj@mskcc.org

Fig. 1 The overview of the proposed segmentation work flow. In the initial training phase, brush strokes indicating tumor and nontumor samples are drawn on a few training multimodal image slices. Parameters of the regional logistic regression model (P1) are calculated. In CRF, the regional term of the energy function is estimated from the regional model along with a contrast-based boundary term to obtain MAP estimation of the segmentation of the training slices via graph cut minimization (S1). Additional brush strokes are optional for correction to re-estimate the regional model parameters for CRF graph cut until the segmentation on the training slices is accepted. The parameters of the boundary logistic regression model (P2) are estimated from the accepted segmentation. Once both regional and boundary regression models are trained for our CRF energy function, MAP estimation segmentation can be obtained automatically from the other image slices via minimization of the CRF energy function (S2).

2 Related Works

Most of the semiautomatic or interactive methods for MRI brain tumor segmentation rely on an active contour or a level set. The user initializes a contour inside or outside the target structure to be segmented. The contour then evolves toward the boundary of the target iteratively via optimization of an energy function formulated with image gradient (edge) and curve regulation. Wang et al.⁹ proposed the fluid vector flow contour model to overcome the limitation of traditional active contour models that cannot capture acute concave shapes of a tumor. To integrate multiple MRI modalities into these algorithms, Cobzas et al.¹⁰ incorporated tissue classification with high-dimensional features into the level-set method such that the level-set energy minimization is equivalent to maximizing the *a posteriori* probability of the partitioning (inside/outside the contour). Ho et al.¹¹ fit a histogram of the difference between pre- and postcontrast T1 with a mixture of Gaussian and Poisson models to derive tumor and normal tissue probability maps for region competition in level-set evolution. Hamamci et al.¹² introduced a tumor-cut algorithm that constructs a tumor probability map from a seeded cellular automata segmentation,¹³ followed by a level set evolving on the tumor probability map to impose spatial smoothness.

Fully automatic methods usually formulate the MRI brain tumor segmentation task as a supervised voxel-wise classification

problem, requiring learning a model that uses the MR image intensities or textures to discriminate between normal and tumor voxels. Many approaches are based on machine-learning techniques. Cai et al.¹⁴ trained a support vector machine (SVM) with interpatient samples to obtain a more generalizable tissue classification, while Ruan et al.¹⁵ used inpatient samples. Zikic et al.¹⁶ combined additional probabilities from a Gaussian mixture model of image features for tissue appearance as the input to the decision forest classification. Voxel-wise classification approaches, however, cannot achieve smooth segmentation results without using neighborhood relationships for spatial regularization after initial classification. One of the options is to use Markov random fields (MRFs) or CRF to penalize discontinuity. MRFs were first applied to binary image denoising by Greig et al.¹⁷ to model the *a priori* distribution of the class assignments in the Bayesian formulation of a *a posteriori* probability of a clean image, given a noisy image. The prior reflects the belief that a cleaner image consists of large homogeneous patches. Therefore, a locally dependent MRF that models pairwise interaction is a natural choice for the prior that penalizes discontinuity in class assignments of a pair of neighboring pixels. Additionally, the authors showed that the maximum *a posteriori* (MAP) estimation of the class assignments can be obtained by a graph max-flow/min-cut algorithm. CRFs introduced by Lafferty

et al.,⁸ in which unknown labels are globally conditioned by the observations, relax the local dependency assumption that is needed in MRFs to model observations for factorization, thus allowing the use of arbitrary attributes of the observations without explicitly modeling them. Nie et al.¹⁸ introduced spatially weighted MRF for the prior in the same Bayesian formulation of the MAP problem for brain tumor segmentation. The discontinuity penalty is weighted according to the spatial resolutions of different MRI modalities. Instead of processing on individual voxels, Zhao et al.¹⁹ applied an MRF on oversegmented super-voxels, a technique commonly seen in natural scene segmentations. Lee et al.²⁰ computed posterior probability distributions from SVMs for the CRF unary potential and presented a local-consistency measure for pairwise potential that encourages spatial continuity. Bauer et al.²¹ used a multistage approach that obtains a coarse segmentation from SVM and then renders smoother results by using three-dimensional (3-D) and two-dimensional (2-D) CRF regularization. Motivated by Zikic et al., Meier et al.²² extended this approach by incorporating a probability map from a density forest as additional input to a classification forest and using the Bhattacharyya distance for pairwise potential in CRF as the discontinuity penalty.

Previously, we have proposed a CRF framework with statistical pairwise potential for liver and kidney segmentation in CT images.^{23,24} In this work, we extend our framework to multimodal images. Taking advantage of direct modeling of *a posteriori* distribution in CRF, we introduce a purely probabilistic pairwise boundary term for multimodal images while maintaining the MAP inference. In our framework, we formulate the segmentation problem as an energy-minimization problem of a CRF. Such an energy-minimization scheme has also been used in Bauer et al.²¹ and Meier et al.²² In their works, the energy-minimization functions as a spatial regularization on the classified labels in a postprocessing step after the initial voxelwise classification is obtained. In contrast, the energy minimization in our work is classification rather than just regularization since we directly derive both a regional (voxelwise) discriminative classification model and a boundary (pairwise) discriminative classification model from the energy function. Both models are case-specific and are trained from the manual inputs.

Another difference between our method and those of Bauer et al. and Meier et al. is the degree of automation in the segmentation process. The goal of Bauer et al. and Meier et al. is to create a fully automated technique, with separate training and application steps. In clinical practice, this implicitly means that the required expert oversight and correction of possible errors must be handled as a separate, fully manual postprocessing step. Automatic methods are potentially faster in cases in which the study images are well represented by the training set since the amount of subsequent correction of segmentation errors may be small. In this work, we relax the goal of full automation using the CRF to merge the expert manual input more in the spirit of online or reinforcement learning. The required manual interaction is balanced by the ability to adapt to differences in instrumentation, differences between training and study patient groups, or changes in protocols. Studies²⁵ have reported that—even within the same protocol, body region, scanner, and patient—MRI images acquired at different times may appear different due to a variety of scanner-dependent variations. Therefore, models implemented with interpatient training for segmentation might not be robust.^{26,27}

3 Methods

3.1 Conditional Random Fields

Let $G = (V, E)$ be the graph representing an image with N voxels or nodes indexed by i , where $V = \{i | 1 \leq i \leq N\}$, $E = \{\{i, j\} | i \in V, j \in N_i\}$, and N_i are neighbors of voxel i in a neighborhood system such as four-connected or eight-connected systems in 2-D. We treat the segmentation problem as a classification problem in a stochastic process, i.e., assigning a class label for each node in the graph based on the observed image features. Let $X = (X_i)_{i \in V}$ be the multivariate random variable of such assignments. x is an assignment instance and x_i is the class assignment for node i . Let $Y = (Y_i)_{i \in V}$ be the multivariate random variable of images and y_i be the extracted image feature vector at node i . For example, $y_i = (\text{FLAIR}_i, T1_i, T2_i)$ is the image intensity of MRI Flair, T1 and T2 sequences at voxel i . The segmentation problem can be simply described as finding an assignment x such that the conditional probability $P(X = x | Y = y)$ or $P(x|y)$ is maximum, i.e., obtaining an MAP estimate of x .

A CRF obeys the Markov property as does MRF. By definition, the factorization of the conditional probability characterized by a Gibbs distribution has the form

$$P(x|y) \propto \exp - \sum_i \left[r_i(x_i, y) + \lambda \sum_{j \in N_i} u_{ij}(x_i, x_j, y) \right]. \quad (1)$$

For image segmentation, r_i represents a regional term that describes the relationship between the multimodal images y and label assignment x_i . The boundary term u_{ij} describes the relationship between the multimodal images y and label assignments for the pair of neighboring x_i and x_j . The constant λ is a weight to penalize discontinuity. It should be pointed out that the global image data y are contained in both r_i and u_{ij} . That is, CRF allows the use of not only the image features limited at voxel i but also, without scarifying the MAP inference of random fields, more remote image features such as contextual local pattern²⁸ and image patches.²⁹ In this work, we use additional first-order statistical features in a region centered by a voxel. More details are described in Sec. 4.2.1. The extracted image features simply become a new image modality in our framework.

3.2 Logistic Regression Models

To keep the CRF framework purely statistical without restrictive assumptions, a probabilistic function is a sensible choice for both regional and boundary terms. We define

$$r_i(x_i, y) = -\ln p(x_i | y_i) \quad (2)$$

and

$$u_{ij}(x_i, x_j, y) = \begin{cases} 0, & \text{if } x_i = x_j \\ -\ln p(x_i \neq x_j | y_i, y_j), & \text{if } x_i \neq x_j \end{cases}. \quad (3)$$

For biclass segmentation, i.e., $x_i \in \{0, 1\}$, the posterior probabilities can be modeled with logistic regression models. Let $\beta = (\beta_0, \beta_1, \dots, \beta_K)$, where K is number of modalities, be the model parameters for the regional term and $y'_i = (1, y_i)$. Thus, the probability that pixel i belongs to target class, $x_i = 0$, is

$$p_i = p(x_i = 0|y_i) = \frac{1}{1 + \exp(-\beta \cdot y'_i)}, \quad (4)$$

and the probability for nontarget class $p(x_i = 1|y_i)$ is simply

$$p(x_i = 1|y_i) = 1 - p_i. \quad (5)$$

As for the boundary term, we define a new feature vector $z_{ij} = (z_{ij_1}, z_{ij_2}, \dots, z_{ij_K})$, $j \in N_i$, where $z_{ij_k} = |y_{i_k} - y_{j_k}|$. Let the model parameters be $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_K)$ and $z'_{ij} = (1, z_{ij})$. Then, the probability that there is a boundary between the neighboring pixel i and j is

$$p_{ij} = p(x_i \neq x_j|y_i, y_j) = \frac{1}{1 + \exp(-\gamma \cdot z'_{ij})}. \quad (6)$$

In practice, we fit the model parameters using a stochastic gradient descent.

3.3 Energy Minimization

Following the minimization scheme proposed by Greig et al.¹⁷ and Boykov et al.,³⁰ we show that our energy function in Eq. (1) can be minimized by a graph max-flow/min-cut algorithm. With the definition of Eq. (3), we have

$$u_{ij}(0,0,y) = u_{ij}(1,1,y) = 0, \quad \text{for } x_i = x_j, \quad (7)$$

therefore

$$u_{ij}(0,0,y) + u_{ij}(1,1,y) \leq u_{ij}(1,0,y) + u_{ij}(0,1,y). \quad (8)$$

According to Kolmogorov and Zabih,³¹ with Eq. (8), our energy function is graph representable and can be minimized by a graph cut.

4 Experiments

4.1 Data Set

To evaluate our method, we use the training data set from the MICCAI Brain Multi-Modal Tumor Segmentation Challenge [BRATS 2012 and 2013 (Private communication with the organizer has confirmed that BRATS 2012 and BRATS 2013 have the same training cases)]. Twenty clinical high-grade (HG) cases and 10 low-grade (LG) cases, each with four MRI modalities—flair, T1, T1 contrast (T1c), and T2 images—were evaluated. In the context of radiation treatment planning, we focus on gross tumor volume (GTV) segmentation. The GTV is defined as the abnormal region that can be seen in the image, which is the entire visible tumor. A GTV in the data set consists of up to four different tumor tissue types presented: (1) edema, (2) enhancing core, (3) nonenhancing core, and (4) necrotic. We, however, applied our method to biclass segmentation, in particular, GTV or non-GTV. For each case, the union of ground truths of the four possible tumor tissue types serves as the single GTV ground truth for our evaluation. For active tumor and edema segmentations in BRATS 2012, we first obtained a GTV segmentation and then further partitioned the GTV into active tumor and edema by performing an additional biclass (active tumor and edema) segmentation.

4.2 Image Feature Extraction and Training

4.2.1 Image feature selection

We first used three cases for probing important features in multimodal MRI images for GTV segmentation. In addition to the original image intensities, seven additional candidate features were extracted: minimum, maximum, gradient, entropy, spot, ripple, and edge, where minimum is the local minimum of a neighborhood around each pixel, maximum is the local maximum of a neighborhood around each pixel, and entropy is the minimum number of bits needed to encode the local gray-level

Table 1 Feature ranking reported by a random forest in three probing HG cases. Four features are ranked among the 10 most important in all three cases across different modalities: intensity (int), maximum (max), minimum (min), and entropy (ent).

Rank	Case 1			Case 2			Case 3		
	Modality	Feature	Score	Modality	Feature	Score	Modality	Feature	Score
1	Flair	max	0.137	Flair	max	0.156	Flair	max	0.177
2	Flair	int	0.107	Flair	min	0.104	Flair	int	0.108
3	T2	int	0.105	Flair	int	0.102	T2	int	0.105
4	Flair	min	0.078	T2	int	0.088	Flair	min	0.096
5	T2	max	0.076	T2	min	0.059	T2	max	0.062
6	T2	ent	0.063	T1	min	0.050	T2	min	0.055
7	T1	min	0.042	T2	max	0.046	Flair	ent	0.050
8	T2	min	0.042	T1c	min	0.045	T2	ent	0.042
9	T1c	max	0.038	Flair	ent	0.036	T1	min	0.041
10	Flair	ent	0.036	t1	ent	0.036	T1c	ent	0.033

distribution of a neighborhood around each pixel. The last three features (spot, ripple, and edge) are Laws' energy features. This results in a 32-D feature vector (4 modalities $\times$ 8 features per modality) for each voxel. Those features are extracted using the Scikit-Image toolkit.³² We generated the features with small five-pixel radius patches (a common window size in medical image analysis^{33,34}), chosen to capture the heterogeneous characteristics presented in the GTV. For selecting important features, a random forest with 100 decision trees was trained using 32-D features from the three probing cases with ground truth provided in the MICCAI data set. The feature selection random forest ranked intensity, minimum, maximum, and entropy as the most important features across all four MRI modalities. Laws' image features were not as important and thus were excluded from our models. Table 1 shows the rankings of features in the three probing cases. As a result, feature images were generated with the selected features for each of the MRI

Fig. 2 Image features. The first row shows the original MRI images of three modalities: flair in column (a), T1 contrast in column (b), and T2 in column (c). The second to fourth rows show the corresponding three texture feature images: minimum, maximum, and entropy, respectively.

modalities, resulting in a 16-D feature vector per voxel. Figure 2 shows the example of the resulting feature images from one HG case.

4.2.2 Training

An experienced specialist drew the brush strokes on three image slices for each case to obtain samples for training the case-specific models. One training slice is in the middle section of the tumor and the other two training slices are 10 to 20 slices apart superiorly and inferiorly from the middle slice. GTVs are visible across 30 to 100 slices depending on cases in the data set. Figure 3 shows examples of training slices with brush strokes from one case for GTV segmentation [Fig. 3(a)] and active tumor segmentation [Fig. 3(b)]. Case-specific models are trained as follows: sample voxels under the red and blue brush strokes, indicating tumor tissue class and nontumor tissue class, respectively, were used to train the regional logistic regression model in Eq. (4); neighboring voxel pairs straddling the boundaries of the ground truth, treated as the accepted segmentation in our proposed work flow, and randomly selected neighboring voxel pairs not on the boundaries on those training slices were collected to train the boundary logistic regression model in Eq. (6). The patient specific models obtained from the training slices were used in the CRF framework to automatically segment all other slices in the same case. The segmentation for each slice is done in 2-D. The graphical framework can be applied for volumetric (3-D) segmentation, however, as manual segmentation is performed slice-wise in the clinical setting, we believe that the tool might be more intuitive to use in the same fashion. We did not perform any manual postediting on the results in this evaluation, although such editing using brush strokes for retaining the models to improve the segmentation can be easily done in our framework as we have demonstrated in our previous work²⁴ for single-modality segmentation.

Fig. 3 Example brush strokes on three training slices in a case: (a) the superior slice, (b) the middle slice, and (c) the inferior slice. The first row shows the brush strokes superimposed on flair images for training models used for GTV segmentation. The second row shows another set of brush strokes superimposed on T1c images for training models used for active tumor segmentation. For each case, only three slices were used for training models to segment all other slices in the same case.

Table 2 GTV Dice coefficients compared to Meier et al., showing means and standard deviations of 20 HG cases and 10 LG cases.

	Our method (4-D)	Our method (16-D)	Meier et al. (44-D)
HG	0.79 ± 0.13	0.84 ± 0.09	0.80 ± 0.12
LG	0.81 ± 0.10	0.84 ± 0.08^a	0.76 ± 0.11

^a $P < 0.05$ for one-tail independent two-sample *t*-test.

Table 3 Dice coefficients of active tumor and edema compared to Hamamci et al., showing means and standard deviations of 20 HG cases and 10 LG cases.

	Our method (16-D)		Hamamci et al.	
	HG	LG	HG	LG
Active tumor	0.80 ± 0.14^a	0.77 ± 0.09^a	0.73 ± 0.21	0.71 ± 0.13
Edema	0.68 ± 0.16^a	0.46 ± 0.26	0.56 ± 0.20	0.38 ± 0.16

^a $P < 0.05$ for one-tail paired two-sample *t*-test.

4.3 Results

We compare the similarity of the segmented results and the ground truth using Dice coefficient (Dice). Dice from the models with intensity-only feature [four-dimensional (4-D)] and texture feature (16-D) are compared against the Dice from Meier et al.²² reported in BRATS 2013, in which 44-D texture feature vectors were used. The BRATS 2013 onsite challenge winner was Tustison et al.,³⁵ while Meier et al. placed second in the challenge. Tustison et al. reported 0.88 in Dice of complete tumor for training set. The score includes both high-grade and low-grade training cases. However, they did not clearly specify the number of cases that they used to obtain the scores, so it is difficult to make a direct comparison. Further their method requires tremendous computing resources that are not practical in clinical settings. Meier et al. performed well in both onsite challenge cases and offsite training cases and also used CRF. This is the main reason we choose Meier et al. for comparison. Dices of active tumor and edema segmentations are compared against the Dices from Hamamci et al.¹² reported in BRATS 2012. Himamci et al. is the only semiautomatic method presented in BRATS 2012. They also used brush strokes for user guidance to initiate a grow-cut segmentation with a tumor probability map, followed by a level-set evolution.

Fig. 5 GTV segmentation results from two cases: the first to third rows show three example slices from one case and the fourth to sixth rows are from another case. Segmentations are overlaid on flair image slices: column (a) ground truth, (b) our method with 4-D feature, (c) our method with 16-D feature, and (d) the overlay of our 16-D results with the ground truth.

Table 2 shows the GTV Dice coefficients of the comparison, and Table 3 shows the active tumor and edema Dice coefficients of the comparison. Our method with models of 16-D features for GTV segmentation, although using a smaller number of texture features, performs better than Meier et al. (0.84 ± 0.09 versus 0.80 ± 0.12 in HG cases and 0.84 ± 0.08 versus 0.76 ± 0.11 ,

Fig. 4 Case-by-case Dice coefficients of our segmentations. Results with 4-D features are in light gray and 16-D features are in dark gray. In 25 of 30 cases, using the selected 16-D features improves the performance.

Fig. 6 Comparison of the result from a slice that Meier et al. reported in the proceeding: from left to right: (a) the flair image, (b) the T1c image with GTV ground truth overlay, (c) the reported segmented image with all four labels from Meier et al., and (d) the segmented GTV from our method. The arrows point to parts of the segmentation where our method more closely resembles the ground truth.

$p < 0.05$, in LG cases.) Even with models of intensity-only (4-D) features, GTV Dice coefficient of our method is close to and comparable to Meier et al. (0.79 ± 0.16 versus 0.80 ± 0.12) in HG cases and better in LG cases (0.81 ± 0.10 versus 0.76 ± 0.11). Case-by-case GTV Dice coefficients are shown in Fig. 4. Figure 5 shows examples of segmented GTV from our method with 4-D feature vectors and 16-D feature vectors. In general, compared with intensity-only models, models with 16-D feature vectors in our framework, provide less noisy tumor segmentation. Figure 6 shows the visual comparison of the result from a slice that Meier et al. reported in the proceeding. Our method performs better than Hamamci et al. in HG cases for both active tumor (0.80 ± 0.14 versus 0.73 ± 0.21 , $p < 0.05$) and edema (0.68 ± 0.16 versus 0.56 ± 0.20 , $p < 0.05$), as

well as in LG cases for both active tumor (0.77 ± 0.09 versus 0.71 ± 0.13 , $p < 0.05$) and edema (0.46 ± 0.26 versus 0.38 ± 0.16). Figures 7 and 8 show our segmentation results of active tumor and edema from one HG case and one LG case of which the Dice coefficients are close to the mean Dice coefficients.

Figures 9 and 10 demonstrate two challenging examples for HG and LG cases, respectively. In the HG example, the tumor indicated by the ground truth has very indefinite boundary appearance (pointed by the arrow in the figure) in all MRI modalities. The optimal cut found by the algorithm may not be decisive. In the LG example, the first row shows a slice close to a training slice. Classification models generated from the training slice produced the result (the overlay in red color) in column (d) in the figure. The second row shows a slice at

Fig. 7 Examples of active tumor and edema segmentation from an HG case that has a Dice coefficient close to the mean of all HG cases: (a) flair image, (b) T2 image, (c) T1c image, (d) ground truth superimposed on the T1 image, and (e) our result superimposed on the T1 image.

Fig. 8 Examples of active tumor and edema segmentation from an LG case that has a Dice coefficient close to the mean of all LG cases: (a) flair image, (b) T2 image, (c) T1c image, (d) ground truth superimposed on the T1 image, and (e) our result superimposed on the T1 image.

Fig. 9 An example in an HG case where lower Dice coefficient is received. Images from left to right: (a) flair, (b) T2, (c) T1c, and (d) overlay of our result in red, with the ground truth in cyan. The arrow indicates an area where the tumor boundary is difficult to discern.

the inferior part of the tumor. Our algorithm failed to segment the tumor indicated by the ground truth (the overlay in cyan color). This may be because (1) the classification models learned with image features from the training slices are not applicable due to the substantially different image intensity distribution in the tumor on the slice to be segmented and (2) the tumor is difficult to distinguish from the normal tissue, and, in such case, it becomes a difficult task for any classifier.

In this study, we did not examine the effect of user interaction on segmentation quality. We have examined this aspect in a previous study²⁴ of kidney segmentation in CT using the same framework. In that study, agreement of interobserver segmentations using our framework was 99% compared to 98% in manual segmentation.

5 Conclusions and Future work

We have presented a semiautomatic approach for multimodality medical image segmentation using CRF framework. In the energy function, we have introduced purely probabilistic regional and boundary terms that are estimated from logistic

regression models. The case-specific models are directly trained from expert user inputs with brush strokes and accepted segmentations on the training slices, allowing for the method to adaptively learn from human guidance.

In our evaluation, with just a few training slices (three per case), we showed that tumor segmentations from our method on multimodality images are more accurate than a similar automatic method using CRF and a semiautomatic method using grow-cut and active contour in terms of Dice coefficients when compared to ground truth. Since our regional term in CRF is common among methods using a Bayesian approach, the results suggest the major advantage of using a discriminative model introduced in our method for the pairwise term in CRF. It directly models target-specific boundary probability distribution via training, instead of using some dissimilarity metric between two tissue classes, which may not represent the true boundary of the target. Further, our semiautomatic approach with a tool familiar to users increases its usability in the clinical practice.

Using the same purposed framework, future works will include addressing multiclass segmentation and experimenting

Fig. 10 An example in an LG case where lower Dice coefficient is received. The first row shows a slice close to a training slice and the second row shows a slice close to the inferior part of the tumor. Images from left to right: (a) flair, (b) T2, (c) T1c, and (d) overlay of our result, in red with the ground truth, in cyan. The classification models learned from image features on the training slice are successful on the top slice but not the bottom slice. The tumor is difficult to distinguish from the normal tissue on the bottom slice.

alternative statistical models for the regional and boundary. We also are considering attending future BRATS onsite challenges in the near future once we have extended our method to multi-class segmentation.

Acknowledgments

This research was funded in part through the NIH/NCI Cancer Center Support under Grant P30 CA008748.

References

1. R. J. Steenbakkers et al., "Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis," *Int. J. Radiat. Oncol. Biol. Phys.* **64**(2), 435–448 (2006).
2. B. Stall et al., "Comparison of t2 and flair imaging for target delineation in high grade gliomas," *Radiat. Oncol.* **5**(1), 5 (2010).
3. M. Braendengen et al., "Delineation of gross tumor volume (GTV) for radiation treatment planning of locally advanced rectal cancer using information from {MRI} or fdg-pet/ct: a prospective study," *Int. J. Radiat. Oncol. Biol. Phys.* **81**(4), e439–e445 (2011).
4. T. Heimann et al., "Comparison and evaluation of methods for liver segmentation from CT datasets," *IEEE Trans. Med. Imaging* **28**, 1251–1265 (2009).
5. D. P. Huyskens et al., "A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer," *Radiother. Oncol.* **90**, 337–345 (2009).
6. N. Sheth et al., "Comparison of model-based segmentation systems for contouring of male pelvic structures," *Med. Phys.* **37**, 3125 (2010).
7. G. A. Whitfield et al., "Automated delineation of radiotherapy volumes: are we going in the right direction?," *Br. J. Radiol.* **86**(1021), 20110718 (2013). PMID: 23239689
8. J. Lafferty, A. McCallum, and A. F. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. on Machine Learning*, pp. 282–289 (2001).
9. T. Wang, I. Cheng, and A. Basu, "Fluid vector flow and applications in brain tumor segmentation," *IEEE Trans. Biomed. Eng.* **56**, 781–789 (2009).
10. D. Cobzas et al., "3D variational brain tumor segmentation using a high dimensional feature set," in *IEEE 11th Int. Conf. on Computer Vision (ICCV 2007)*, pp. 1–8 (2007).
11. S. Ho, E. Bullitt, and G. Gerig, "Level-set evolution with region competition: automatic 3-D segmentation of brain tumors," in *Proc. 16th Int. Conf. on Pattern Recognition 2002*, Vol. 1, pp. 532–535 (2002).
12. A. Hamamci et al., "Tumor-cut: segmentation of brain tumors on contrast enhanced mr images for radiosurgery applications," *IEEE Trans. Med. Imaging* **31**, 790–804 (2012).
13. V. Vezhnevets and V. Konouchine, "grow-cut" - interactive multi-label n-d image segmentation," in *Proc. Graphicon*, pp. 150–156 (2005).
14. H. Cai et al., "Probabilistic segmentation of brain tumors based on multi-modality magnetic resonance images," in *4th IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI 2007)*, pp. 600–603 (2007).
15. S. Ruan et al., "Tumor segmentation from a multispectral MRI images by using support vector machine classification," in *4th IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro (ISBI 2007)*, pp. 1236–1239 (2007).
16. D. Zikic et al., "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," *Lect. Notes Comput. Sci.* **7512**, 369–376 (2012).
17. D. Greig, B. Porteous, and A. Scheult, "Exact maximum a posteriori estimation for binary images," *J. R. Stat. Soc.* **51**, 271–279 (1989).
18. J. Nie et al., "Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov random field," *Comput. Med. Imaging Graphics* **33**(6), 431–441 (2009).
19. L. Zhao, W. Wu, and J. J. Corso, "Semi-automatic brain tumor segmentation by constrained MRFs using structural trajectories," in *Proc. of Medical Image Computing and Computer Aided Intervention* (2013).
20. C.-H. Lee et al., "Segmenting brain tumors with conditional random fields and support vector machines," *Lect. Notes Comput. Sci.* **3765**, 469–478 (2005).
21. 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," *Lect. Notes Comput. Sci.* **6893**, 354–361 (2011).
22. R. Meier et al., "A hybrid model for multimodal brain tumor segmentation," in *MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) MICCAI*, pp. 31–37, Nagoya, Japan (2013).

23. Y.-C. Hu, M. Grossberg, and G. Mageras, "Semi-automatic medical image segmentation with adaptive local statistics in conditional random fields framework," in *30th Annual Int. Conf. of the IEEE EMBS 2008*, pp. 3099–3102 (2008).
24. Y.-C. Hu et al., "Interactive semiautomatic contour delineation using statistical conditional," *Med. Phys.* **39**, 4547–4558 (2012).
25. L. Nyul, J. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging* **19**, 143–150 (2000).
26. L. Clarke et al., "MRI: stability of three supervised segmentation techniques," *Magn. Reson. Imaging* **11**(1), 95–106 (1993).
27. J. Bezdek, L. Hall, and L. Clarke, "Review of MR image segmentation techniques using pattern recognition," *Med. Phys.* **20**(4), 1033–1048 (1993).
28. J. Cui et al., "Transductive object cutout," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2008)*, pp. 1–8 (2008).
29. Y. Schnitman et al., "Inducing semantic segmentation from an example," *Lect. Notes Comput. Sci.* **3852**, 373–384 (2006).
30. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *Proc. Int. Conf. of Computer Vision*, pp. 105–112 (2001).
31. V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(2), 147–159 (2004).
32. S. van der Walt et al., "scikit-image: image processing in Python," *PeerJ* **2**, e453 (2014).
33. M. Rachidi et al., "Laws' masks descriptors applied to bone texture analysis: an innovative and discriminant tool in osteoporosis," *Skeletal Radiol.* **37**(6), 541–548 (2008).
34. H. A. Elnemr, "Statistical analysis of law's mask texture features for cancer and water lung detection," *Int. J. Comput. Sci. Issues* **10**(6), 196–202 (2013).
35. N. Tustison et al., "Ants and arboles," in *MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS)*, pp. 47–50, Nagoya, Japan (2013).

Yu-chi Hu is a senior programmer analyst in the Department of Medical Physics at Memorial Sloan Kettering Cancer Center and is a PhD candidate of the Computer Science at Graduate Center, City University of New York. His research interests are medical image registration and segmentation algorithms utilizing machine learning techniques in the applications of radiotherapy.

Michael Grossberg is an assistant professor of CS at City College, City University of New York. He has held teaching and research positions at Columbia University, MPI for Math in Bonn, and Hebrew University. He received his PhD in math from MIT. His current interests are in applied machine learning for remote sensing, computer vision, and climate science as well as data visualization.

Gikas Mageras is an attending medical physicist and chief of the Medical Physics Computer Service at Memorial Sloan Kettering Cancer Center. His research interests are in modeling and management of patient motion in radiation treatment plans and in image-guided radiotherapy. He leads a team in the development of software systems for treatment planning and verification, image registration and segmentation. He became a fellow of the American Association of Physicists in Medicine in 2006.