3D brain tumor segmentation in multimodal MR images based on learning population- and patient-specific feature sets

Jun Jiang 1, Yao Wu 2, Meiyan Huang 1, Wei Yang 1, Wufan Chen 3, Qianjin Feng *

School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

ABSTRACT

Brain tumor segmentation is a clinical requirement for brain tumor diagnosis and radiotherapy planning. Automating this process is a challenging task due to the high diversity in appearance of tumor tissue among different patients and the ambiguous boundaries of lesions. In this paper, we propose a method to construct a graph by learning the population- and patient-specific feature sets of multimodal magnetic resonance (MR) images and by utilizing the graph-cut to achieve a final segmentation. The probabilities of each pixel that belongs to the foreground (tumor) and the background are estimated by global and custom classifiers that are trained through learning population- and patient-specific feature sets, respectively. The proposed method is evaluated using 23 glioma image sequences, and the segmentation results are compared with other approaches. The encouraging evaluation results obtained, i.e., DSC (84.5%), Jaccard (74.1%), sensitivity (87.2%), and specificity (83.1%), show that the proposed method can effectively make use of both population- and patient-specific information.

Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Brain tumor contouring is an important and challenging task in cancer radiotherapy. However, manual segmentation is time-consuming, and the intra- and inter-observer variability potentially leads to substantial inconsistency in the segmentation. Multimodal magnetic resonance imaging (MRI) images are extensively used in brain disease diagnosis and radiotherapy because of their ability to provide complementary information for the diagnosis. Different modal MRI images can enhance specific brain tissues. For example, TIC (T1, with a gadolinium contrast agent) highlights the abnormal regions while FLAIR (fluid attenuated inversion recovery) restrains the gray level of the cerebrospinal fluid. Information from multimodal MRI images can be fully used in the delineation of brain tumor because they provide the essential distinction between lesions and normal tissues.

Even with multimodal images, brain tumor segmentation is quite a challenging task because the pathological process responsible for the creation and growth of brain tumors is inherently unpredictable. Consequently, the geometric properties of the tumor do not conform to a particular shape/size distribution [1], which makes the tumor boundaries universally and irregularly distorted. Furthermore, the tumor is heterogeneous and the border is difficult to localize. Another aspect that complicates segmentation is that artifacts and noise can be easily interfused into images during data acquisition. These obstacles make it impossible to use any kind of shape prior on these normal structures to aid in tumor segmentation, and make it difficult to design a fully automatic segmentation method. Subsequently, there have been numerous efforts to develop semi-automatic and fully automatic segmentation algorithms to delineate tumors in MRI images [1-4].

Clearly, it is difficult to segment a brain tumor through simple unsupervised thresholding [5]. Many other methods have been proposed by previous investigators. These methods were divided into four groups based on their characteristics: (1) atlas-based, (2) contour/surface evolution, (3) graph-based, and (4) learning-based.

1.1. Atlas-based methods

The Atlas-based method is widely used in medical image segmentation [6–8]. One major advantage of segmenting medical images as opposed to natural scenes is that the structural and intensity characteristics of a natural biological variability or the presence of pathology are well known. Therefore, a geometric prior can be used by atlas-based segmentation in which a fully labeled template MR volume is registered to an unknown dataset. Population atlases provide an important prior to improve segmentation by measuring the deviation from the normal brain. However, the deformable
registration of brain images with tumor to the population atlas is an extremely challenging problem and is still an active research area because of the intensity variations around the tumor mainly caused by edema/infiltration and the tumor mass effect, which also deforms the healthy tissue morphology [6]. Therefore, the atlas-based method is limited by the fact that the segmentation results depend highly on the quality of affine registration. A slight misalignment of the issues usually leads to a dramatic decrease in segmentation accuracy [7,8].

1.2. Contour/surface evolution methods

Contour/surface evolution methods are successfully used in tumor contouring for 2D and 3D datasets [9–11]. The Active Contour Model/Snake (ACM) and Level sets are very influential approaches proposed in past decades. The common ground of these methods is their method in evolving a curve/surface under some constraints to extract the desired object. Early active contour models such as ACM are formulated in terms of a dynamic parametric contour. Substantially, the curve evolution of ACM can be converted to a level set formulation by embedding the dynamic contour as the zero level set of a time-dependent level set function (LSF) [10]. One advantage of level set methods is their representation of the contours of complex topology and their ability to handle topological changes such as splitting and merging in a natural and efficient way, which is not allowed in parametric active contour model. Although level set methods are used to solve a wide range of scientific and engineering problems, their applications are plagued with the irregularities of the LSF developed during the level set evolution [10]. When level sets expand to 3D space, they can be relatively slow to compute [11] and their formulation usually entails several free parameters that can be very difficult to tune correctly for specific applications.

1.3. Learning-based methods

Machine learning classification techniques, including supervised [12–14] and unsupervised (clustering or fuzzy clustering) [2,15–17], are also introduced into brain tumor segmentation. Trained classifiers estimate the probability for each voxel in the testing volume, judging whether the voxel belongs to the target or the background. The threshold of the probability map is calculated to obtain the segmentation result [2] or provide for post-processing [4]. These techniques make it possible for high-dimensional features to be utilized in order to achieve a better discriminatory power for tumors compared with sole dependence on intensity information [5]. Moreover, the approaches applied in the field of pattern analysis can be transplanted into medical image segmentation, such as distance metric learning algorithm [18,19], to make the intra-class samples closer while keeping extra-class samples as far away from each other as possible. Unfortunately, these classification-based segmentation approaches consider the voxels in the image to be independent of each other, with no spatial correlation both in the training and testing phases.

1.4. Graph-based methods

Many graph-based segmentation algorithms were introduced into image segmentation [21–25]. All these algorithms share the same idea that images can be modeled as an undirected graph by formulating a cost function to define the link strength between each node. Most of these algorithms construct the graph based only on the intensity information in the individual testing image. However, the observed intensity distribution in the tumor region spans a wide range, and the voxel intensities in the tumor regions often overlap with the surrounding normal tissues, which severely affects the performance of the segmentation methods. Our previous work [26] showed that it is helpful to introduce high-dimensional features into the graph-cut framework to improve segmentation performance. Unfortunately, due to the increase in dimensionality of the feature, it becomes difficult to directly estimate the probability distributions of the target and the background in the feature space when the sample size is limited.

It is actually a trend to adopt more than one of the methods mentioned above to complement the drawbacks in a single one. Hamamci et al. [4] adopted the CA (Cellular Automata) algorithm to set two strength maps that are combined to obtain the tumor probability map. Then, a level set surface is initialized, which converges to the final segmentation. Cobzas et al. [3] presented a MRI tumor segmentation method that incorporates a learned statistical model and atlas-based features into the variational framework.

In the current paper, we present a graph-cut based method for the segmentation of brain tumors in multimodal MR images. A Gabor filter bank is used to capture the texture properties. An optimal distance metric is learned, aimed at improving the discrimination in the feature space by employing Closed-Form Metric Learning [18]. Furthermore, a Real-AdaBoost classifier is introduced to estimate the probabilities of voxels belonging to the target and the background. Generally, the training samples taken only from the interaction seed points are not sufficient for classifier learning. Thus, the population feature set is exploited to make the classifier more accurate and robust. Based on this idea, a new cost function is constructed and optimized via graph-cut. Our main contributions are listed as follows:

1) Instead of using only gray-level information, high-dimensional features are adopted to identify the tumors from the surrounding tissues. Meanwhile, to enlarge the discrimination between lesion and tumor, Closed-Form Metric Learning is employed to learn a task-specific distance metric in the feature space using the interactive inputs.

2) As opposed to the learning method constantly used in the task of medical image segmentation, such as KNN and SVM, we introduce a Real-AdaBoost technique with the learned distance metric to estimate the probability density of the tumor and the background through learning high-dimensional feature set.

3) In addition to the use of patient-specific feature set to train the classifier, a population feature set is also exploited to make the classifier more accurate and robust.

4) Since the classification method considers voxels in the image to be independent of each other and a probability is assigned to each voxel independently, this may lead to the deficiency of spatial information for neighboring voxels. We adopt a graph-cut framework, which combines the region term and boundary term, in order to add spatial constrains and yield the final segmentation.

2. Method

2.1. Background of graph-cut

The conventional graph-cut has been very popular in studies of image segmentation in recent years [27,28]. This algorithm models images as an undirected graph. Each graph node corresponds to a voxel in the image volume, and the link strength between nodes can be quite different. By introducing both a region term and a boundary term into the graph-cut energy function, a minimum cost s/t cut is
computed on the appropriately constructed graph [20,21]. The cost function is usually defined as follows:

\[ l^* = \arg \min_l \left( \beta \sum_{p \in P} D_p(l_p) + \sum_{p,q \in N} V_{p,q}(l_p, l_q) \right) \tag{1} \]

where \( l = \{l_p | l_p \in L\} \) denotes a labeling, \( L = \{0, 1\} \) is the label set, 0 corresponds to the background, 1 corresponds to the foreground, coefficient \( \beta \) specifies a relative importance of the region term versus the boundary term, \( P \) is a set of sites, and \( N \) is a set of neighboring pixels.

On the right hand side of Eq. (1), the first term is the region term. It measures the sum of penalty for assigning each individual pixel \( p \) to the label \( l_p \). The region item is usually defined as:

\[
\begin{align*}
D_p(l_p = l_{obj}) &= -\ln P(l_p = l_{obj}) \\
D_p(l_p = l_{bgk}) &= -\ln P(l_p = l_{bgk})
\end{align*}
\tag{2}
\]

where \( l_{obj} \) is the label of the object and \( l_{bgk} \) is the label of the background, \( P(l_p = l_{obj}) \) denotes the probability of assigning the voxel \( p \) to \( l_{obj} \), and \( P(l_p = l_{bgk}) \) denotes the probability of assigning the voxel \( p \) to \( l_{bgk} \). Traditionally, \( P(l_p) \) is estimated by utilizing the gray histogram. However, the overlap of voxel intensities in the tumor regions and normal tissues severely affects the performance of segmentation.

The second term is the boundary term. It is interpreted as a penalty for discontinuity between pixels \( p \) and \( q \). In the conventional cost function of the graph-cut [20,21], the boundary item is usually defined as follows:

\[
V_{p,q} = \exp \left[ -\frac{G(l_p, l_q)}{2\sigma^2} \right] \frac{1}{\text{dist}(p, q)} \quad \text{if } l_p \neq l_q
\]

\[ 0 \quad \text{if } l_p = l_q \tag{3} \]

where \( G(l_p, l_q) \) denotes the difference between the intensities of pixel \( p \) and \( q \), \( \text{dist}(p,q) \) is the Euclidean distance between the locations of \( p \) and \( q \), and parameter \( \sigma \) is a constant relating to the gray level that measures noise standard deviation [20]. This penalty term is introduced when pixels \( p \) and \( q \) have different labels.

2.2. Overview of the proposed method

The proposed method is performed on multimodal brain MR images, including T1, T2, post-Gadolinium T1, and FLAIR sequences. The main idea of the proposed method is to exploit two classifiers to compute the probability of a voxel belonging to the tumor or the background. The first classifier, called the global classifier, is trained by using samples from the population feature set. The second one, called custom classifier, is trained by using samples from seed points in the testing image. The outputs of these two classifiers are weighted and then used to construct the region term of the cost function, as detailed in Section 2.4. The training of the global classifier is notably an off-line procedure, that is, training it once and then using if for all the testing images. The outputs of the global classifier reflect the distribution of features in the population feature set. The custom classifier needs to be training on-line. It changes according to the testing image and corresponds to the characteristic of features from the testing image.

Figs. 1 and 2 illustrate the procedure of the proposed method, which mainly contains four components. (1) Feature extraction: the responds of Gabor filter banks and the image modality features are extracted from the multimodal MR images. The feature volumes are generated by stacking the features extracted on the 2D image slices, as detailed in Section 2.4. (2) Distance metric learning: A project matrix is obtained through learning training set in the feature space, by adopting Closed-Form Metric Learning [18]. The feature set of testing samples is projected to another space before classification to improve the discrimination between lesions and normal tissues, as detailed in Section 2.5. (3) Classifier training: To construct the population training set, the positive and negative samples are gathered according to the manual segmentation. For the current testing image, training samples are taken from manually selected seed points. (4) Optimization by graph-cut: A new cost function are constructed and optimized by using the max-flow/min-cut algorithm. If the results are not satisfied, the additional interactions can be done by the users.

2.3. Data acquisition

Brain tumor image data used in this work were obtained from the MICCAI 2012 Challenge on Multimodal Brain Tumor Segmentation (http://www.imm.dtu.dk/projects/BRATS2012) organized by B. Menze, A. Jakab, S. Bauer, M. Reyes, M. Prastawa, and K. Van Leemput. The challenge database contains fully anonymized images from

Fig. 1. Flow diagram of the training phase of the proposed method.

Fig. 2. Flow diagram of the testing phase of the proposed method.
the following institutions: ETH Zurich, University of Bern, University of Debrecen, and University of Utah.

According to the description of the Challenge database, there are 45 high-grade and 35 low-grade glioma clinical data, which contain realistically generated synthetic brain tumor datasets (simulated images) with 25 high-grade and 25 low-grade glioma subjects, and 30 cases of clinical image data (Table 1). In the clinical images, the tumors and edema regions are manually delineated. For each patient, T1, T2, FLAIR, and post-Gadolinium T1 MR images are available. All volumes are linearly co-registered to the T1 contrast image, skull stripped, and interpolated to 1 mm isotropic resolution.

We normalize the intensity of all modalities of MR images before any procedural analysis or transformation because the intensities of the MRI images are sensitive to acquisition conditions such as scanner type, magnetic field, and so forth. For each image volume, intensity values at the 5% and 95% quantiles are computed for the brain region. Then, the two values are used to normalize intensity to [0, 1] through the min-max method [29]. This correction strategy helps to improve tolerance to the variation of image acquisition condition and makes our implementation more reliable.

2.4. Feature extraction

2.4.1. Image modality features

Excluding ordinary T1 and T2 images, T1C and FLAIR modalities are also in the MR image bank. All the modalities are directly included into the feature set. At the same time, we notice that the T1C images highlight the abnormal “enhancement” regions, which indicate the presence of a tumor. The FLAIR images restrain the magnetic resonance signals of hydrogen atom and lower the contrast of the cerebrospinal fluid with surrounding tissues. The different modalities of the images enhance unique information in the same region of human body. This motivates the use of the image difference between the two modalities as a sub-feature [1]. As such, there are $C_2^d = 6$ kinds of sub-features from the image subtraction, including T1-T2, T1-T1C, T2-FLAIR, and so on.

2.4.2. Gabor features

In MR images, lesions usually appear quite different in texture from normal tissues. Texture features provide an important cue in the perception and discrimination of a tumor. As a bio-computational model of object-recognition [30–32], Gabor features are widely used in the field of image analysis and computer vision [13]. In the current paper, we construct a Gabor filter group with several scales and directions to extract the texture features of the image. The following formulators are utilized to generate the Gabor filter kernel.

$$G(x, y) = W(x, y) \times S(x, y)$$

In Eq. (4), $G(x,y)$ is the Gabor wavelet kernel. $W(x,y)$ and $S(x,y)$ are defined as follows:

$$W(x, y) = K \exp \left[ -\pi \left( \frac{x^2}{A^2} + \frac{y^2}{B^2} \right) \right]$$

$$S(x, y) = \exp(i2\pi R(x \cos \theta + y \sin \theta))$$

where $W(x,y)$ is a Gaussian kernel, the vertical and horizontal sizes are $A$ and $B$, respectively (usually we let $A = B$), the amplitude is $K$; and $S(x,y)$ is a sine-wave surface function, wherein the direction is defined by $\theta$ and the frequency is defined by $R$ (usually we let $R = 1$). In our approach, we use four scales with eight directions Gabor filters to form a filter group.

By convolving the MR image $I$ with the above filter bank $f$ for one pixel located at $(x,y)$, a feature vector $F_C(x,y)$ can be obtained as follows:

$$F_C(x, y) = |I^T F_{a,0} |(x, y)$$

wherein $A = 2, 4, 6, 8$ and $\theta = N \times \pi / 8, N = 0, 1, 2, \ldots, 7$. $F_C(x,y)$ represents the output at location $(x,y)$ by convolving the image with the Gabor filters. In the current implementation, four scales and eight directions are considered. Thus, $F_C(x,y)$ contains 32 elements. With the 10 image modality features mentioned in the upper subsection, the image features are 42 dimensional vectors in the proposed method.

2.5. Distance metric learning

In our image segmentation system, all the features prepared for training are projected to another feature space before boosting learning, which aims to enlarge the difference/distance between voxels with different labels while reducing that of voxels with the same labels. In this section, CFML (Closed-Form Metric Learning), a distance metric learning algorithm for automatically learning a distance metric with labeled data is presented as proposed by Ali-panahi et al. [18].

Let $x_i$ and $x_j$ be the D-dimensional feature vectors of two voxels with different labels in the images. The squared Mahalanobis distance between feature vectors $x_i$ and $x_j$ is

$$d_M(x_i, x_j) = ||L(x_i - x_j)||^2 = (x_i - x_j)^T L(x_i - x_j)$$

where $L$ is a transformation matrix with $d \times D$, $M = L^T L$, and $M$ is a positive semi-definite matrix. Distance metric learning aims to find a linear transformation matrix $L$ to project the image features to a new feature space. This makes the squared Mahalanobis distance between data with the same labels closer while making the distance of data with different labels farther.

Numerous researchers have proposed distance metric learning algorithms to find an optimal projection $L$ or a metric $M$ for minimizing an objective function. Among the various kinds of distance metric learning algorithms, CFML is a simple and effective algorithm that can achieve a closed-form solution. The feature vectors of voxels with the same label are defined as similar, whereas those with different labels are defined as dissimilar. Let $S$ and $D$ be the set of similar pairs and the set of dissimilar pairs, respectively, defined by

$$S : (x_i, x_j) \in S \quad \text{if} \quad x_i, x_j \quad \text{are similar},$$

$$D : (x_i, x_j) \in D \quad \text{if} \quad x_i, x_j \quad \text{are dissimilar}.$$
where \( \text{tr}(\cdot) \) denotes the matrix trace and
\[
M_s = \frac{1}{|S|} \sum_{(x_i, x_j) \in S} (x_i - x_j)(x_i - x_j)^T, \tag{10}
\]
\[
M_D = \frac{1}{|D|} \sum_{(x_i, x_j) \in D} (x_i - x_j)(x_i - x_j)^T. \tag{11}
\]

The solution of the optimal transformation matrix \( L^* \) is given by the matrix of eigenvectors associated with the largest eigenvalues of the matrix \( M_s^{-1} M_D \). With the optimal transformation matrix, the squared Mahalanobis distance between two feature vectors in two different images can be computed. Here, CFML tries to minimize the squared Mahalanobis distance between similar pairs while maximizing the squared Mahalanobis distance between dissimilar pairs.

In our implementation, the optimal transformation matrices are learned from global and custom training sets, and are symbolized as \( DT_C \) and \( DTC \), respectively, as illustrated in the flow diagram (Figs. 1 and 2). The projection of global and custom training sets can be formulated as:
\[
\begin{align*}
F_C &= DT_C^T \cdot f_C \\
F_C &= DTC^T \cdot f_C
\end{align*}
\]

where \( F_C \) and \( F_C \) are the transformed global and custom training sets, respectively; \( DT_C \) and \( DTC \) are the transposed optimal transformation matrices learned from global and custom training sets, respectively; and \( f_C \) and \( f_C \) are the original global and custom training sets, respectively.

2.6. Voxel classification

The Adaboost algorithm [33,34] proposed by Yoav Freund and Robert Schapire is one of the most important ensemble methods because of its solid theoretical foundation and very accurate prediction. There is evidence that Adaboost performance is more accurate than many other algorithms [35–37] in the classification task. Many variants of Adaboost have been developed during the past decades. To estimate the probability of a voxel belonging to the tumor or the background, Real-Adaboost [33,34,38] is introduced to solve our problem. Since the feature extraction procedure has been detailed, this section focuses on the negative/positive label sampling.

2.6.1. Training phase

As mentioned in Section 2.2, we want to train two classifiers to formulate the region term of the cost function. The global classifier is trained by using samples from the population feature set, and the custom classifier is trained by using samples from seed points in the testing image (Fig. 1). For the population training set, the positive samples are the ground truth posted with image sequences, and the negative samples are the outer circles of the ground truth. They are generated following the formulation:
\[
S_N = (T \oplus P) \cap T^T
\]

where \( T \) is the binary truth of the lesion, \( S_N \) is the surrounding negative sample set, and \( P \) is a morphological structuring element, wherein \( P \) is a ball, with a radius equal to 15. The binary dilation of \( T \) by \( P \) is denoted as \( T \oplus P \). This strategy allows the population training set to contain much more information near the tumor edges.

Notably, the global training sets are projected to another space by multiplying the transform matrix \( DT_C \) before the boosting procedure in the training phase.

2.6.2. Testing phase

The training of the global classifier is an off-line procedure while the custom classifier needs to be trained on-line. Therefore, the training of the custom classifier is detailed in the testing phase (Fig. 2). In this phase, a graphic user interface is developed to pick the seeds from the testing images, where the sagittal, transverse, and coronal views are displayed synchronously and the region of interest (ROIs) is drawn on them (Fig. 3). This strategy makes the interaction more convenient and efficient. Orthogonal planes are also visualized on each projection, allowing for a better sense of where you are in the 3D space.

In the testing phase, the manually labeled samples are used to learn an optimal transformation matrix \( DTC \). The boosting procedure that follows yields a custom classifier. \( DT_C \) and \( DTC \) are utilized to project all the testing samples in ROI to a new feature space. According to the transformed samples, the custom and global classifiers estimate the probability of whether a sample belongs to the target or the background, respectively. The two probability maps are weighted by the coefficient \( w \). The weighted probability maps compose the base of region term in the cost function.

2.7. Cost function design

In this section, the probability, which measures the likelihood of a voxel belonging to the target or the background, is assigned to each voxel in the ROI. Our optimization procedure is based on the graph-cut framework. Thus, the cost function is quite similar to that of the conventional graph-cut and contains two parts: region term and boundary term.

2.7.1. Region term construction

To estimate these two probabilities, two classifiers are introduced. Thus, \( P(l_p = l_{bg}) \) and \( P(l_p = l_{t}) \) are defined as:
\[
\begin{align*}
P(l_p = l_{bg}) &= wD_C(f_p) + (1 - w)D_C(f_p) \\
P(l_p = l_{t}) &= 1 - P(l_p = l_{bg})
\end{align*}
\]

where \( D_C(f_p) \) is the output of the custom classifier when its input is \( f_p \) and \( D_C(f_p) \) is the output of the global classifier. Notably, \( D_C(f_p) \) and \( D_C(f_p) \) are normalized to \([0,1]\) because they represent the probabilities here. The parameter \( w \) is a weighting factor to leverage both population- and patient-specific statistics. By weighting the predictions from the custom and global classifiers previously trained, the algorithm obtains the best performance when \( w \) is set to 0.7 in our practice. The selection of \( w \) is detailed in Section 3.2.

2.7.2. Boundary term construction

Since graph-cut is a discrete approach, the image is modeled as an undirected graph and the graph nodes represent the isolated voxels in the image. It is important to incorporate contextual information in the form of spatial dependencies between voxels. The integration of spatial information usually yields smoother and more reliable results. In the proposed method, the boundary term is formulated as follows:
\[
V_{p, q} = \exp \left[ -\frac{E(f_p - f_q)}{2\sigma^2} \right] \cdot \frac{1}{\text{dist}(p, q)}
\]

where \( \text{dist}(p, q) \) is the physical distance between voxel \( p \) and \( q \), and \( E(f_p - f_q) \) is the similarity between two features \( f_p \) and \( f_q \). Here, the Euclidian distance in the feature space is adapted to measure feature similarity. Typically, we set the parameter \( \sigma \) equal to 10.
3. Experiments and results

3.1. Assessment of segmentation performance

The proposed method was applied to 23 sets of brain tumor MR images. Qualitative results are shown in Fig. 4, in which (a) and (b) are the synthetic brain tumor image cases, and (c)–(e) are the clinical cases. The goal was to accurately delineate the boundaries of the brain tumor for each image in the first row of Fig. 4. The lesion boundaries of cases (c)–(e) are quiet blurry due to the existence of edema, but our proposed method yields encouraging results. The global and custom predictions are shown in the second and third rows, respectively. The fourth row shows the ground truth posted with the image data. The last row shows the segmentation results of the proposed method. We must declare that the cysts inside a tumor, filled with blood/csf/liquid are excluded in the ground truth, and that circumstance inevitably lowers the segmentation assessment quantitatively. Fig. 5 gives an intuitionistic visualization of the segmentation results for case (a) and (c).

To quantitatively compare the performance of our method with those of others, several metrics are introduced. For each segmentation, we compute the false positive (FP), false negative (FN), true positive (TP), and true negative (TN) counts. These four metrics are defined as follows:

\[ TP = R \cap T; \quad TN = R \cup T; \quad FP = R \cap \bar{T}; \quad FN = \bar{R} \cap T \]

where \( T \) is the true set and \( R \) is the result set. From these, we can then compute the following additional error metrics (Table 2), namely, including Dice similarity coefficient (DSC), Jaccard similarity, Sensitivity and Specificity.

\[ Jaccard = \frac{TP}{FP + TP + FN} \]

\[ \text{Dice coefficient:} \]

\[ DSC = \frac{2 \times TP}{(FP + TP) + (TP + FN)} \]

\[ \text{Sensitivity: (fraction of positives that are correctly detected)} \]

\[ \text{Sens.} = \frac{TP}{TP + FN} \]

\[ \text{Specificity: (fraction of negatives that are correctly detected)} \]

\[ \text{Spec.} = \frac{TN}{TN + FP}. \]

We compare the segmentation result with that yielded by conventional Graph-cut [20] and Random walks [39], wherein we extend the algorithm into 3D. As detailed in Section 2.1, conventional Graph-cut solely adopt pixel intensity rather than sufficient features for region/boundary term construction. That situation limits its application to complex target segmentation. Somewhat alike to graph-cut, random walk is a graph-based segmentation method. Each node of the graph corresponds to a voxel in the volume. During algorithm optimization, random walk decisions are made on the graph, where at each step the location jumps to another site according to some probability distribution. Since both methods need human interaction, we initial the algorithms with the same seeds. Because of high computation complexity, random walks algorithm takes far more time than graph-cut in our experiments. They both exhibit a lower score due to the segmentation leakage of some cases. Moreover, the limited amount of initial seeds makes the segmentation prone to failure.

Fig. 3. Interaction interface for manually picking the seeds from the 3D images. Positive (target) samples are labeled as green scatter points, while negative (background) samples are labeled as red scatter points.
Fig. 4. Qualitative segmentation results for brain tumor. First row: original T2 images. Second row: global prediction. Third row: custom prediction. Fourth row: ground truth posted by the MICCAI workshop. Fifth row: segmentation yielded by our method. Last row: contours of our segmentation results (in red) and true contours (in green) are shown. Apparent cysts inside a tumor, filled with blood/csf/liquid, are excluded in the ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
Fig. 5. Solid tumor rendering with mesh on the surface. They are reconstructed for case (a) and (c) in Fig. 4, each case represents both segmentation result (red) and ground truth (blue). The Dice coefficient of the segmentation result for (a) and (c) are 96.6% and 83.3% respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 6. Segmentation performance of different algorithms. The bar in the figure shows the means and standard deviations of evaluation metrics.

We also compare our method with a method that does not use population samples, which shows higher sensitivity but lower specificity. This means that the method that does not use population training samples is too sensitive to accurately predict the lesion in the image, making numerous misjudgments. On the other hand, our method exhibits a relatively higher score but a lower coefficient of variation (std/mean) on the average compared with the other methods used in validation (Fig. 6).

3.2. Parameter optimization

The region term of our graph-cut framework plays an important role in the segmentation; thus, a series of experiments was performed to evaluate the effectiveness of the combination of the custom and global prediction.

We let weighting factor \( w \) vary in \([0, 1]\) with the step of 0.1. As \( w \) increases, the weighting of global prediction decreases and the custom prediction increases. For each value of \( w \), we calculate the dice similarity coefficient to evaluate the segmentation result (Fig. 7). Our method performs best when \( w \) equals 0.7.

4. Discussion and conclusions

Results show that our newly proposed method significantly outperforms both the traditional graph-cut and optimized methods without using population information. Even in some challenging cases, the segmentation result is comparable to that of a clinician expert.

4.1. Advantages of the reported method

In our proposed method, high-dimensional features and Real-Adaboost technique are used for voxel classification. Information from all modalities is effectively used for distinguishing lesions and normal tissues. Additionally, the training samples from the global training set that provides probability constrains are inherently integrated in the graph-based segmentation method, which further improves the segmentation accuracy. All these strategies make our approach more efficient and stable.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph-cut [3]</td>
<td>69.01 ± 12.2</td>
<td>53.85 ± 15.9</td>
<td>87.48 ± 9.0</td>
<td>59.67 ± 20.8</td>
</tr>
<tr>
<td>Random walk [8]</td>
<td>73.65 ± 16.2</td>
<td>60.24 ± 18.8</td>
<td>75.79 ± 26.9</td>
<td>75.86 ± 7.5</td>
</tr>
<tr>
<td>Without population</td>
<td>75.99 ± 6.9</td>
<td>61.67 ± 8.9</td>
<td>92.72 ± 6.9</td>
<td>64.57 ± 7.8</td>
</tr>
<tr>
<td>Our method</td>
<td>84.5 ± 9.4</td>
<td>74.10 ± 14.5</td>
<td>87.22 ± 7.9</td>
<td>83.11 ± 14.8</td>
</tr>
</tbody>
</table>

Please cite this article in press as: Jiang J, et al. 3D brain tumor segmentation in multimodal MR images based on learning population- and patient-specific feature sets. Comput Med Imaging Graph (2013). http://dx.doi.org/10.1016/j.compmedimag.2013.05.007
4.2. Limitations of the reported method

There are some limitations in our approach. The main limitation is that the method is a semi-automatic method. Although an interaction interface for picking the initial seeds has been developed, a fully automatic segmentation method is still our aspiration. Another limitation is instantaneity. Due to the high dimension of the features and the slow convergence of the boosting algorithm, the most time-consuming stage is the training of the global classifier. It takes nearly 2 h to train the global classifier, including 57 cases with 4 modalities (that is 228 image volumes). On the other hand, the time consumed in the training of the custom classifier is much shorter, upon which its ROI size depended, and is usually less than five minutes. Since the training of the global classifier is an off-line procedure, it is trained once and used for all the testing images. This limitation can be overcome by optimizing our source code from Matlab into C.

In summary, we present an effective semi-automatic method for brain tumor contouring that performs very well in challenging cases. Our framework uses a transformation matrix to enlarge the distance between two samples with different labels in the feature space. It also employs the Real-AdaBoost classification technique to estimate the probability of whether a voxel belongs to the foreground or the background. Through weighting predictions made by summarizing localized images and population images, a new cost function is constructed and optimized via graph cuts. This strategy makes our segmentation result more reliable and robust. Furthermore, we intend to investigate the performance of our method for the more general problem of lesion contouring.

Conflict of interest statement

Hereby the authors confirm that they do not have any conflict of interest.

Acknowledgments

This work was supported by the grants from the National Basic Research Program of China (973 Program) (No. 2010CB732505), National Science & Technology Pillar Program of China (No. 2012BAI14B02), and National Natural Science Funds of China (NSFC, No. 81101109).

References


