Automatic metastatic brain tumor segmentation for stereotactic radiosurgery applications

To cite this article: Yan Liu et al 2016 Phys. Med. Biol. 61 8440

View the article online for updates and enhancements.

Related content
- A survey of MRI-based medical image analysis for brain tumor studies
  Stefan Bauer, Roland Wiest, Lutz-P Nolte et al.
- Automatic delineation of tumor volumes by co-segmentation of combined PET/MR data
  S Läibfarth, F Eckert, S Welz et al.
- A discriminative model-constrained EM approach to 3D MRI brain tissue classification
  Michael Wels, Yeleng Zheng, Martin Huber et al.

Recent citations
- A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery
  Yan Liu et al.
Automatic metastatic brain tumor segmentation for stereotactic radiosurgery applications

Yan Liu¹,², Strahinja Stojadinovic², Brian Hrycushko², Zabi Wardak², Weiguo Lu², Yulong Yan², Steve B Jiang², Robert Timmerman², Ramzi Abdulrahman², Lucien Nedzi² and Xuejun Gu²

¹ College of Electrical Engineering and Information Technology, Sichuan University, Chengdu 610065, People’s Republic of China
² Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

E-mail: xuejun.gu@utsouthwestern.edu

Received 20 May 2016, revised 15 August 2016
Accepted for publication 29 September 2016
Published 15 November 2016

Abstract
The objective of this study is to develop an automatic segmentation strategy for efficient and accurate metastatic brain tumor delineation on contrast-enhanced T1-weighted (T1c) magnetic resonance images (MRI) for stereotactic radiosurgery (SRS) applications. The proposed four-step automatic brain metastases segmentation strategy is comprised of pre-processing, initial contouring, contour evolution, and contour triage. First, T1c brain images are preprocessed to remove the skull. Second, an initial tumor contour is created using a multi-scaled adaptive threshold-based bounding box and a super-voxel clustering technique. Third, the initial contours are evolved to the tumor boundary using a regional active contour technique. Fourth, all detected false-positive contours are removed with geometric characterization. The segmentation process was validated on a realistic virtual phantom containing Gaussian or Rician noise. For each type of noise distribution, five different noise levels were tested. Twenty-one cases from the multimodal brain tumor image segmentation (BRATS) challenge dataset and fifteen clinical metastases cases were also included in validation. Segmentation performance was quantified by the Dice coefficient (DC), normalized mutual information (NMI), structural similarity (SSIM), Hausdorff distance (HD), mean value of surface-to-surface distance (MSSD) and standard deviation of surface-to-surface distance (SDSSD). In the numerical phantom study, the evaluation yielded a DC of 0.98 ± 0.01, an NMI of 0.97 ± 0.01, an SSIM of...
0.999 ± 0.001, an HD of 2.2 ± 0.8 mm, an MSSD of 0.1 ± 0.1 mm, and an SDSSD of 0.3 ± 0.1 mm. The validation on the BRATS data resulted in a DC of 0.89 ± 0.08, which outperform the BRATS challenge algorithms. Evaluation on clinical datasets gave a DC of 0.86 ± 0.09, an NMI of 0.80 ± 0.11, an SSIM of 0.999 ± 0.001, an HD of 8.8 ± 12.6 mm, an MSSD of 1.5 ± 3.2 mm, and an SDSSD of 1.8 ± 3.4 mm when comparing to the physician drawn ground truth. The result indicated that the developed automatic segmentation strategy yielded accurate brain tumor delineation and presented as a useful clinical tool for SRS applications.

Keywords: automatic segmentation, brain metastases, stereotactic radiosurgery, contrast enhanced T1-weighted MRI

(Some figures may appear in colour only in the online journal)

1. Introduction

Stereotactic radiosurgery (SRS) is commonly used to treat metastatic brain tumors and delivers a potent dose of highly conformal radiotherapy to the target in a single fraction (Leksell 1983, Iorio-Morin et al. 2014, Nieder et al. 2014, Bae et al. 2015, Wu et al. 2016a). Brain tumor delineation/segmentation is an active research topic and has been investigated for decades (Bauer et al. 2013, Gordillo et al. 2013, Menze et al. 2015). Current delineation methods are either (1) single- or multi-modality segmentation based on the source of imaging modality; or (2) manual, semi-automatic, and automatic based on the degree of required human interaction. The multi-modality segmentation methods take advantage of synergistic information from different imaging modalities and delineate treatment regions in a comprehensive manner. For example, T1-weighted MRI provides distinguishable visualization of tissues, T2-weighted MRI is sensitive to edema, while diffusion weighted (DW)-MRI contributes to biological characterization of tissue (van der Heide et al. 2012). Researchers have developed various multi-modality co-segmentation methods to utilize these geometric and physiological information (Bauer et al. 2011, Geremia et al. 2012, Bagci et al. 2013, Buendia et al. 2013, Song et al. 2013, Leibfarth et al. 2015, Wu et al. 2016b). However, multi-modality segmentation is not always favorable or feasible for metastatic brain tumors delineation in SRS applications. As for Gamma Knife SRS, T1-weighted MRI with Gadolinium contrast (T1c) is used for planning on the treatment day. For proliferative metastatic brain tumors, T1c images show a clear tumor boundary due to accumulation of the contrast agent in regions of blood–brain barrier disruption, hence this is a clinically accepted mono-modality image used for radiotherapy planning. This is contrary to the T2-weighted MRI or DW-MRI, which are often acquired on routine diagnostic scans. Moreover, the planning T1c images are acquired with image slice thickness ≤1.5 mm to provide the accuracy required for SRS (Klein et al. 2009), while other modality images are often acquired with larger slice thickness (~3–5 mm). Integrating pre-treatment low-resolution images for tumor delineation requires additional image processing, such as resampling and/or deformable registration, which adds additional uncertainty in segmentation.

Brain tumor delineation in current clinical practice is a manual task where clinicians review planning images slice by slice to identify and delineate treatment regions. For Gamma Knife SRS, both target delineation and treatment planning are performed on the same day of treatment delivery. Semi-automatic brain tumor segmentation may improve the efficiency of the radiosurgery process. The main components of semi-automatic brain tumor segmentation
include user interaction and automatic software computation. Users are required to analyze the visual information and provide initial delineation inputs and feedback for the software to perform segmentation (Raviv et al. 2009, Riklin-Raviv et al. 2010, Fitton 2011, Hamamci et al. 2012, Guo et al. 2013, Havaei 2014, Cui et al. 2016). The feedback is subjective and prone to intra- and inters-observer variation. Semi-automatic segmentation methods can produce different results whether the method is guided by different experts or even the same expert at different times. Fully automatic brain tumor segmentation is an attractive and popular research topic, especially with the recent development of machine learning algorithms. In automatic brain segmentation, the computer determines the segmentation of brain tumors using artificial intelligence and prior knowledge without human interaction. Conventional automatic segmentation mainly uses standard image processing methods such as thresholding and region-growing (Hsieh et al. 2011, Huber et al. 2015). These methods are often applied to 2D images. Machine learning based auto-segmentation (Bauer et al. 2011, 2012, Liu 2011, Geremia et al. 2012, Buendia et al. 2013, Cordier et al. 2013, Altman et al. 2015) has been actively developed and widely applied to image segmentation in diagnostic radiology and segmentation quality assurance in radiotherapy. These algorithms can learn complex relationships or patterns from empirical data and make accurate decisions. Machine learning algorithms mainly belong to either classification or clustering families. Classification methods deduce a relationship from the training dataset and apply the learned relationships to auto-segment new images. The performance of classification methods relies on the quality of the learned models, which heavily depend on the quality of training datasets. The complexity of brain lesions, variations in size, location, and number, make generation of an effective training dataset challenging. It has been shown that some developed classification models (Huang et al. 2014, Moeskops et al. 2015) work for normal brain structure segmentation but fail at brain tumor delineation. Clustering methods are simple but always have high storage requirements and time issues. The performance of clustering methods depends on the extracted features and similarity metrics. Active contouring methods are attractive in automatic segmentation thanks to their capability of segmenting images of anatomic structures by exploiting constraints derived from the image data. Contours obtained from these methods are connected and smooth, but require a meaningful initialization to avoid a local minimum in optimization procedures.

In this work, we report an automatic metastatic brain tumor delineation strategy based on a single T1c image modality, which is a commonly available image in GammaKnife-based radiosurgery treatment planning (Sze et al. 1990, Barajas and Cha 2012). The entire segmentation strategy takes advantage of both conventional auto-segmentation methods to localize the tumor(s) and modern clustering and active contour techniques to extract the contours. The developed single-modality multi-method segmentation strategy later can be easily extended to multi-modality MRI images. In this study, the workflow of this segmentation strategy is presented and theoretical details of each step are described. Validation results are based on numerical phantoms from the BRATS dataset (Menze et al. 2015) and clinical data.

2. Methods and materials

2.1. Auto-segmentation workflow

Figure 1 shows the workflow of the proposed automatic brain metastatic tumor delineation strategy. The workflow is comprised of four steps: image preprocessing, initial contouring, contour evolution, and false-positive contour removal. The process starts with skull stripping to remove non-cerebral tissues such as skull and scalp and enhance the intensity difference between contrast-enhanced tumor and normal brain tissue. The initial contouring step
includes tumor localization and super-voxel clustering. The tumor localization finds a suspicious rectangle region with high-intensity tissue appearance on each 2D slice. Here, we utilize the lateral symmetry of normal brain to identify possible tumor locations through unbalanced histograms of the two cerebral hemispheres. The super-voxel clustering step uses simpler linear iterative clustering (SLIC) (Achanta et al. 2012). Next, a localizing region-based active contour method (Lankton and Tannenbaum 2008) is used to evolve initial contours to match tumor boundaries. Lastly, false-positive contour removal is accomplished using tumor geometrical characteristic information.

2.1.1. Skull stripping. In T1c images, solid tumor regions tend to have an enhanced intensity from contrast agents. The skull has a similar high intensity on T1c images and needs to be removed. We adopted an existing robust learning-based MRI brain extraction system (ROBEX) (Iglesias et al. 2011) to conduct skull stripping. ROBEX combines a discriminative and a generative model to achieve the final result. When a new image is presented to ROBEX, the tool uses a Random Forest classifier to detect the brain boundary. Then the generative model is explored to find a highest likelihood contour. The brain contour is refined by free deformation and used for skull stripping.

2.1.2. Initial contouring.
2.1.2.1. Tumor localization. The tumor localization aims to find a series of axis-parallel bounding boxes around the tumor to facilitate initial tumor contouring. Since a healthy human brain is nearly symmetrical in an axial plane (Saha et al. 2012), the bounding boxes circumscribing tumors can be found by asymmetry detection. Descriptive information hidden in the intensity distribution of the image and histogram are ideal to present asymmetry. Researchers have used the Bhattacharya distance between histograms of brain hemispheres to localize tumors (Saha et al. 2012). However, this hemisphere-based method intrinsically lacks sensitivity to detect small and multiple brain tumors common to SRS cases. In addition, some brain tumors are accompanied with edema and necrosis which manifest in different intensity from normal brain as well. Thus, a general histogram unbalance can be disturbed and cause a false positive detection.

In this study, we developed a multi-scaled, adaptive-threshold bounding box method to accurately localize tumors (figure 2). Starting from the coarsest scale axial MR slice, we find the axis of symmetry (blue line of figure 2(a)), dividing the brain into left and right hemispheres. A horizontal line (red) is placed to separate the stripped pair into four regions, left top (LT), right top (RT), left bottom (LB), and right bottom (RB). For each line position, a cost function \( E(l) = BC(H_{LT}, H_{RT}) - BC(H_{LB}, H_{RB}) \) is calculated based on the Bhattacharya distance. Here, \( H_{LT}, H_{RT}, H_{LB}, H_{RB} \) are the histograms of LT, RT, LB, and RB region, respectively and \( BC(H_1, H_2) = \sum \sqrt{H_1(i)H_2(i)} \) is the Bhattacharya distance between...
two normalized histograms $H_1(i) \in [0, 1]$ and $H_2(i) \in [0, 1]$ with $i$ indicating a histogram bin. As the line sweeps from the top to bottom, we obtain a plot of the cost function with respect to the position $l$, as shown in figure 2(a). The upper and the lower edge of a tumor region are identified at the peak and valley in the cost function plot. Similarly, the left and the right edge of a tumor region are detected by sweeping line through the transposed left and right strips. These four edges (upper, lower, left, and right) form a boundary box circumscribing the tumor. To make this method sensitive to small tumors, we further divide both hemispheres in strips. As shown in the figure 2(b) at the $k$th scale, each hemisphere is divided into $2^k$ equal strips. The boundary box detection procedure is independently conducted on each symmetrical strip pair. In the end, all detected bounding boxes which overlap in a 3D space, are combined to get several 3D bounding boxes.

To eliminate interference from edema and necrosis, image thresholding is performed. Edema and necrosis often exhibit low intensity relative to the healthy brain on T1c images, hence, the threshold is chosen at a crossing point between normal tissue section and tumor section on the normalized error histogram curve. Pixels in T1c image with intensity less than the threshold are replaced by black pixels. Figure 3 shows a sample axial T1c MRI of left (without tumor) and right (with tumor) hemispheres and their normalized histograms before (a) and after thresholding (b). The differences between left and right hemispheres are enhanced, which improves the detectability using the bounding box method.

2.1.2.2. Super-voxel clustering. Super-voxel clustering aims to group artefactual rigid image voxels into perceptually meaningful anatomical regions. Image redundancy is captured and provides for a convenient primitive to compute image features. In this work simple linear iterative
clustering (SLIC) is adapted for MRI voxel clustering. A voxel in an MR image is represented by its intensity and spatial location. We adopt an intensity-distance space based Euclidean distance as a similarity metric to group rigid image voxels. The Euclidean distance \( D_{ij} \) is defined as:

\[
D_{ij} = \sqrt{(d_{\text{intensity}})^2 + m^2 \left( \frac{d_{\text{spatial}}}{S} \right)^2}
\]

with \( d_{\text{intensity}} = |I_i - I_j| \)

\[
d_{\text{spatial}} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}
\]

(1)

where \( d_{\text{intensity}} \) and \( d_{\text{spatial}} \) are the intensity difference and spatial distance between super voxel \( i \) and image voxel \( j \) respectively. \( I_i \) and \( [x, y, z]^T \) are the intensity and the coordinates of a voxel. \( S \) is the size of super voxel grid interval and \( m \) is a weighting factor designed to balance the effect of intensity similarity and distance similarity in \( D_{ij} \). The SLIC algorithm evenly splits voxels into cubes and uses them as initial values. During an iterative step, a super voxel is represented by its centroid voxel and the distance between super voxels are calculated. A rigid voxel belonging to several neighborhood super voxels is assigned to the one with the shortest distance. The centroids are re-calculated following each iteration. The iterations stop when the Euclidean distance between centroids in two consecutive iterations is smaller than a preset threshold. Figure 4(a) shows a cross-section view of a super-voxel clustering result on a sample brain MRI.

Figure 3. An adaptive-threshold scheme: (a) sample axial T1c MRI’s left (without tumor) and right (with tumor) hemisphere and their normalized histograms before thresholding, (b) sample axial T1c MRI’s left and right hemisphere and their normalized histograms after thresholding.
2.1.2. Initial contouring. Initial contours are obtained by combining bounding boxes from tumor localization and super-voxels from voxel clustering (figure 4). The bounding boxes from tumor localization are used to group involved super-voxels and the boundaries of grouped super-voxels are connected. The proportion of a super-voxel covered by the bounding box reveals its relationship with the potential tumor area; therefore, the grouping is performed with reference to the overlapping proportion threshold. For example, in the enlarged image of figures 4(b) and (c), the three super voxels (a)–(c) highlighted by yellow arrows are also covered by the red bounding box. However, the area ratio between the overlapping region and the super voxel is smaller than pre-defined threshold (0.1 in this study) and these three super voxels are excluded from the initial contour.

2.1.3. Contour evolving. Tumors and surrounding tissue heterogeneity (Chan and Vese 1999), as well as inherent noise of MR images, cause the initial contours generated in the above section to often over- or under-segment tumors. In this step, we adopt a localized region-based active contour algorithm (Lankton and Tannenbaum 2008) to refine initial contours. The premise of the active contour is to allow an initial contour to deform so as to minimize an energy function to produce the desired segmentation. The details and numerical solution of the model have been described (Lankton and Tannenbaum 2008) and here we just give a brief overview.

Let \( \Omega \) be an image domain and \( x, y \) be independent spatial coordinates representing a single point in \( \Omega \) respectively. An energy function can be defined as:

\[
E(\phi) = \int_{\Omega_1} \delta \phi(\lambda) \int_{\Omega_1} B(x,y) F(I(y), \phi(y)) dy dx + \lambda \int_{\Omega_1} \delta \phi(\lambda) \| \nabla \phi(x) \|
\]

With \( \delta \phi(\lambda) = \begin{cases} 
0, & \text{if } \phi(x) = 0 \\
1, & \text{if } |\phi(x)| < \varepsilon \\
\frac{1}{2\varepsilon} \left[ 1 + \cos \frac{\pi \phi(x)}{\varepsilon} \right], & \text{otherwise}
\end{cases} \)

\( B(x,y) = \begin{cases} 
1, & \text{if } \|x - y\| < r \\
0, & \text{otherwise}
\end{cases} \)

and \( F(\phi) = H(\phi(y))(I(y) - u_y)^2 + (1 - H(\phi(y)))(I(y) - v_y)^2. \)  

(2)

Here \( \phi \) is a signed distance function representing the distance of a given point from the boundary of \( \Omega \) and \( \delta \phi(\lambda) \) is a smooth Dirac delta function. \( B(x,y) \) is a region delimitation and can be
a unit ball or cubic. $r$ is the radius of the $B(x, y)$. $\varepsilon$ is a small constant approaching 0. $F(\phi)$ is a measurement of the adherence between a point and a contour boundary. The complete energy function $E(\phi)$ consists of an image energy term to model the foreground and background in the localized region piecewise constant and one regularization term to denote the arc length of contours. $\lambda$ is a regularization parameter balancing the two terms. In function $F(\phi)$, $H$ is a Heaviside function. $u_x$ and $v_x$, defined as followed, are the local means of the interior and exterior regions defined by $B(x, y)$.

$$u_x = \frac{\int_{\Omega} B(x, y) H(\phi(y)) I(y) dy}{\int_{\Omega} B(x, y) I(y) dy}, \quad v_x = \frac{\int_{\Omega} B(x, y) (1 - H(\phi(y))) I(y) dy}{\int_{\Omega} B(x, y) (1 - H(\phi(y))) dy}. \quad (3)$$

An evolution equation is obtained from the first variation of equation (2) with respect to $\phi$:

$$\frac{\partial \phi}{\partial t}(x) = \delta \phi(x) \int_{\Omega} B(x, y) \nabla \phi(y) F(I(y), \phi(y)) dy + \lambda \delta \phi(x) \sum \left\{ \frac{\nabla \phi(x)}{\sqrt{\nabla \phi(x)^2}} \right\}. \quad (4)$$

A finite differences implicit scheme is used to discretize equation (4) and get the numerical solution (Chan and Vese 1999). Following the iterative process produced by equation (4), the initial surface gradually converges to the boundary of a tumor. We calculate the Euclidean distance between the two evolved contours with a 10 steps interval as a measure of progress. The iterative process ends when the Euclidean distance is smaller than a preset threshold.

2.1.4. False-positive contours removal. The intensity-based delineation algorithm proposed in this paper is sensitive to intensity variations such as image noise. Brain structures, for example the superior sagittal sinus and a confluence of sinuses, often present with a high intensity in T1c images. The algorithm may converge to these undesired structures and lead to a false-positive as shown in figure 5. Theoretically, the high intensity tissue interference cannot be avoided since our segmentation strategy is based purely on image intensity. However, we can utilize metastatic brain tumors and sinuses geometric shape characteristics to reduce the false-positive rate. SRS brain metastases are often small and without large branching into surrounding normal tissues, while the superior sagittal sinus is often characterized by geometrical twigs. We adopt a geometry based metric, sphericity, to quantify the structure shape. Sphericity $\psi$ is defined as $\frac{4}{3} \pi V_p^3 / A_p$, where $V_p$ is the volume of convergence and $A_p$ is a measure of the surface area of the volume (Hakon 1935). Due to the irregularity of the segmented volume, it is difficult to find an analytical solution of the segmented volume surface. We implemented a marching cubes algorithm (Cline 1987) to generate a triangular surface mesh and to approximate the surface area as a summation of areas of triangles. The segmented volume having its sphericity value lower than the threshold is considered a false detection and removed from the brain tumor contour sets.

2.2. Evaluation dataset

2.2.1. Numerical phantom. To evaluate the proposed auto segmentation strategy, we generated a set of numerical synthetic brain MRIs with eight wide-ranging tumors as described below:

a. A healthy adult brain MRI was chosen. The original dataset is comprised of 27 transverse slices of an MRI scan of a human cranium and an image size of 128 $\times$ 128. The resolu-
tion on the transverse plane was estimated as $dx = dy = 2.0 \text{ mm}$ and the slice thickness $dz = 5 \text{ mm}$. We interpolated images along the $z$ direction to $2.0 \text{ mm}$ to mimic the SRS MRI slice thickness, which increases the data size to $128 \times 128 \times 68$.

b. Artificial tumors were added to simulate diverse clinical situations. The details of the synthetic brain tumors are listed in Table 1. The added tumors were spherical with the radius varied from 0.8 to 3.5 cm and the contrast to background ratio ranged from 1.0 to 2.0, which covers the majority of clinical cases. Tumors #1–4 had a fixed contrast of 2.0, but were varied in size. Tumors #5 and #6 had the same size as tumor #2, but had lower contrast. Tumor #7 was placed at the same axial slide as tumor #2, which was designed to examine whether the auto-segmentation strategy can detect tumors symmetrically located on both hemispheres. Tumor #8 was placed on the cerebellum across the symmetry line of the left and right brain to examine whether the developed strategy can segment tumors which cross hemispheres. The intensity in the simulated lesions has a uniformly distributed random number to reflect the tumor heterogeneity. Figure 6 shows a 3D tumor distribution of eight simulated tumors.

c. Gaussian noise and Rician noise (Gudbjartsson and Patz 1995) were added to the image to mimic realistic image quality for the purpose of testing the auto-segmentation strategy in noise polluted MRIs. The intensity of clear MRI image varied from 0 to 88. Different noise levels were added to validate the robustness of our method for different image qualities. The mean of Gaussian noise was 0 and the variance was varied from 1 to 25. The Rician noise was generated by combining two orthogonal Gaussian noise distributions with a mean of 0 and the variance was varied from 1 to 25. Sample transverse slices of the clear synthetic data and noisy data are shown in Figure 7.
2.2.2. BRATS data. Multimodal brain tumor image segmentation (BRATS) (Menze et al 2015) are published as benchmark data for evaluating different brain tumor MRI auto-segmentation algorithms. The BRATS dataset comprises patient data and synthetic data. Since the BRATS are not specifically designed for metastases and its patient datasets include various kinds of tumors, we chose the synthetic dataset with tumor cores to validate our method and compare our method to other existing auto-segmentation methods. BRATS datasets in resolution of 1.0 mm × 1.0 mm × 1.0 mm have all relevant structures manually delineated,
including necrosis, edema, non-enhancing tumor, enhancing tumor core, and etc. In this study, we randomly selected 21 synthetic T1c images with enhancing tumor cores to validate our auto-segmentation strategy. Seventeen cases have one tumor while the rest have two tumors overlapping along an axial direction.

2.2.3. Clinical data. T1c images of eight patients who received Gamma knife treatment at the University of Texas Southwestern Medical Center (UTSW) were used for auto-segmentation strategy performance evaluation. All T1c images were acquired on a SIEMENS 3T MRI scanner. The image size was $256 \times 256 \times 176$ with a resolution of $1.0\,\text{mm} \times 1.0\,\text{mm} \times 1.0\,\text{mm}$. The brain tumors on the acquired images were manually drawn by radiation oncologists following UTSW standard clinical protocol and confirmed by neurosurgeons. In this study, we used physician drawn contours as the ground truth to validate the accuracy of auto-segmented contours. The details of the patient data are illustrated in table 2. The size of the metastases varied from 0.78 to 4.00 cm, covering majority of SRS clinical cases.

2.3. Parameter selection

Selecting proper parameters is critical to the performance of the developed auto-segmentation workflow. There are four steps requiring parameters selection in the entire auto-segmentation workflow, including (1) designing a proper multi-scale level $k$ in the bounding box step, (2) choosing a reasonable super voxel grid interval $S$ and cluster compactness $m$ in the super-voxel clustering step, (3) selecting a regularization parameter $\lambda$ and a local region radius parameter $r$ in the contour evolving step; and (4) defining a sphericity threshold $\delta$ in false-positive contours removal step. Due to the large variation of patient images quality in routine clinical practice, image-specific parameters may exist to achieve an optimal auto-segmentation performance. In this study, instead of fine-tuning parameters for each individual case, we selected a fixed set of parameters empirically, which allows us to ensure the robustness of the developed method in clinical applications.

2.3.1. Multiscale level $k$. The multi-scale level decides the width of strip pair used to locate the tumor. The larger the $k$, the narrower the strip, and consequently the smaller tumor could be detected. However, increasing $k$ increases scales and raises the computational cost, which deteriorates the segmentation workflow efficiency and practicability. To balance the algorithm’s small tumor detectable and the computational cost, we choose $k$ equal to 3. This
An empirical value is validated in the numerical phantom, BRATS synthetic data, and SRS patient T1c images in our study.

### 2.3.2. Super-voxel grid interval $S$ and cluster compactness $m$.

The super-voxel grid interval $S$ refers to an initial distance between the centers of two super-voxels. In our case, the to-be-delineated tumor is as small as 5 mm in one dimension, thus, we assign $S$ equal to 6 mm by adding 1 mm margin to enclose the tumor. The parameter $m$ is introduced in equation (1) allowing the adjustment of compactness of the super-voxel cluster. The greater value of $m$, the more spatial proximity is emphasized and the more compact the cluster. A range of values varied from 1 to 40 was recommended by researchers who studied nature images (Achanta et al 2012). In our study, the SLIC achieve the best perform at $m = 60$.

### Table 2. Details of clinical data.

<table>
<thead>
<tr>
<th>#</th>
<th>Contrast</th>
<th>Size (cm)</th>
<th>Necrosis</th>
<th>Multiple in one slice</th>
<th>Crossing symmetrical line</th>
<th>Tumor position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.74</td>
<td>0.78</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Cerebellum</td>
</tr>
<tr>
<td>2</td>
<td>1.84</td>
<td>1.16</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Anterior lobe of cerebellum</td>
</tr>
<tr>
<td>3</td>
<td>1.58</td>
<td>2.37</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Frontal lobe</td>
</tr>
<tr>
<td>4</td>
<td>1.33</td>
<td>2.44</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Frontal lobe</td>
</tr>
<tr>
<td>5</td>
<td>1.50</td>
<td>2.11</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Precentral gyrus</td>
</tr>
<tr>
<td>6</td>
<td>2.34</td>
<td>3.93</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Cerebellum</td>
</tr>
<tr>
<td>7</td>
<td>2.36</td>
<td>4.00</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Temporal lobe</td>
</tr>
<tr>
<td>8</td>
<td>2.67</td>
<td>2.06</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Pons</td>
</tr>
<tr>
<td>9</td>
<td>1.71</td>
<td>0.94</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Frontal lobe</td>
</tr>
<tr>
<td>10</td>
<td>1.27</td>
<td>0.65</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Precentral gyrus</td>
</tr>
<tr>
<td>11</td>
<td>1.08</td>
<td>2.93</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Precentral gyrus</td>
</tr>
<tr>
<td>12</td>
<td>2.19</td>
<td>2.94</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Temporal lobe</td>
</tr>
<tr>
<td>13</td>
<td>1.17</td>
<td>1.78</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Precentral gyrus</td>
</tr>
<tr>
<td>14</td>
<td>1.85</td>
<td>2.48</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Frontal lobe</td>
</tr>
<tr>
<td>15</td>
<td>1.23</td>
<td>2.70</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Precentral gyrus</td>
</tr>
</tbody>
</table>

### Table 3. Quantitative metrics used for auto-segmentation accuracy evaluation.

<table>
<thead>
<tr>
<th>Metric name</th>
<th>Metric definition</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume-based metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>DC = 2(A ∩ B)/(A + B)</td>
<td>Normalized mutual information</td>
</tr>
<tr>
<td>NMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIM</td>
<td></td>
<td>Structural similarity index</td>
</tr>
<tr>
<td>Surface-based metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HD</td>
<td>HD = max(h(C, D), h(D, C))</td>
<td>Structural similarity index</td>
</tr>
<tr>
<td>MSSID</td>
<td>h(C, D) = max_{c∈C, d∈D}</td>
<td></td>
</tr>
<tr>
<td>SDSSD</td>
<td>MSSID = mean_{c∈C}(min_{d∈D}</td>
<td></td>
</tr>
<tr>
<td>SDSSD</td>
<td>SDSSD = std_{c∈C}(min_{d∈D}</td>
<td></td>
</tr>
</tbody>
</table>

$a$ and $B$ are auto- and manual- segmented volumes, $C$ and $D$ are the auto- and manual segmented surfaces, $c$ and $d$ are the points located on the surfaces.
2.3.3. Regularization parameter \( \lambda \) and a local regional radius parameter \( r \). The regularization parameter \( \lambda \) is used to penalize the arc length of contour curves. We set \( \lambda = 0.5 \) to make sure the smooth of final tumor contour. The radius \( r \) determines the effective \( B(x, y) \) region. \( B(x, y) \) will be 1 only when the point \( y \) is within a ball of radius \( r \) centered at point \( x \), 0 otherwise. Theoretically, the parameter \( r \) can be treated as a function of the to-be-delineated tumor size. In this study, we simplify parameter \( r \) selection by using an empirical value \( r = 20 \). This valued was selected based on metastatic brain tumors size treated with radiosurgery in our clinical center and validated on varied clinical cases.

2.3.4. Sphericity threshold \( \delta \). The sphericity threshold \( \delta \) is a critical parameter for removing the false positive contours. Although brain tumors have a large variance in geometrical shape, brain metastasis tumors usually have the shape close to spheres. In this study, physician delineation tumors are available as our study benchmark data. The sphericities of these delineated tumors were calculated and their values varied from 0.58 to 0.82. Meanwhile, we also calculated sphericity of cranial sinuses, which has low values around 0.1–0.3. Based on these calculation results, we choose \( \delta = 0.5 \) as the sphericity threshold.

2.4. Evaluation metrics for segmentation accuracy

Volume and surface based metrics were selected to compare the auto-delineated tumor contour with the ground truth. The volume based metrics calculate the intensity of voxels inside
the tumor. The surface based metrics measure the position of voxels on the surface only. Table 3 lists all metrics used in this study. Among volume based metrics, the DICE coefficient (DC) is most commonly used and is a standard metric to judge the segmentation accuracy of different algorithms in the current BRATS report. The normalized mutual information (NMI) is a similarity metric derived from information theory and measures the correlation of the intensity of two volumes defined by the contours. The structural similarity (SSIM) is used

Table 4. Average DC achieved by six algorithms participating BRATS 2012 segmentation challenge and our method.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average DC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bauer et al (2012)</td>
<td>0.81</td>
</tr>
<tr>
<td>Geremia et al (2013)</td>
<td>0.62</td>
</tr>
<tr>
<td>Hamamci et al (2012)</td>
<td>0.69</td>
</tr>
<tr>
<td>Shin (2012)</td>
<td>0.03</td>
</tr>
<tr>
<td>Subbanna et al (2013)</td>
<td>0.41</td>
</tr>
<tr>
<td>Zikic et al (2012)</td>
<td>0.86</td>
</tr>
<tr>
<td>Our segmentation strategy</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Figure 9. Average metric values of 8 tumors with various noise levels: (a) DICE coefficient, (b) normalized mutual information, (c) structural similarity, (d) Hausdorff distance, (e) mean value of surface-to-surface distance, (f) standard deviation of surface to surface distance.
to measure the perceivable differences of structural information, which is hidden behind the relationship of nearby voxels. DC, NMI and SSIM, all range from 0 to 1 with 1 indicating the best. For surface metric evaluation, the marching cubes algorithm (Cline 1987) is utilized to generate the surface mesh based on each tumor volume and the similarity metrics are calculated between the vertices located on the iso-surface. Hausdorff distance (HD), a bi-direction metric, is used to measure the resemblance of the two contours. Surface-to-surface distance (SSD) is calculated as the shortest Euclidean distance from the automatic delineation result to the ground truth. Further, we used the mean value of the SSD (MSSD) to quantify the match of two surfaces and the standard deviation of the SSD (SDSSD) to quantify the scatter.

3. Results

3.1. Numerical phantom

Figure 8 shows a sample segmentation result from a numerical phantom. The green lines represent automatic delineation results and the red lines are the ground truth. The automatic delineated contours are shown to match well with the ground truth even on noisy MRIs. Figure 9 shows the average metrics values of eight tumors under different noise levels. The x-axis is the noise level added to the numerical phantom, where label 'clear image' indicates a numerical phantom without any noise, g1, g2, g3, g4 and g5 indicate that the numerical phantom with Gaussian noise with variance of 1, 4, 9, 16 and 25 and r1 to r5 indicate that the numerical phantom with Rician noise with variance of 1, 4, 9, 16 and 25 respectively. Quantitative metrics plotted in figure 9 were in an acceptable range even with a high noise level. In general, as the noise level goes up, the accuracy of auto-segmentation is decreased. However, the trend is not obvious in the evaluated cases. This is because we didn’t fine tune the regularization parameter and radius parameter in the contour evolving step to achieve best performance.
Instead, we use a fixed set of parameters as described in section 2.3 to verify the algorithm robustness. The numerical validation results show that our auto-segmentation strategy has consistent performance on noisy MRIs and indicate it is a reliable method for clinical data.

3.2. BRATS data

The purpose of the study was to compare the developed brain metastases auto-segmentation strategy with other existing competitive algorithms. In the BRATS 2012 segmentation challenge (Menze et al. 2015), seven algorithms participated in automatic segmentation of synthetic tumor data. Six of the seven algorithms achieved reasonable results with an average DC values ranging from 0.30 to 0.86 in segmenting tumor cores. Table 4 lists the average DC values achieved by these six algorithms as well as ours. We applied our segmentation strategy on the same dataset. Our results have a DC varying from 0.71 to 0.97 with an average value of 0.89 ± 0.08. Specific to single tumor cases, DC varies from 0.81 to 0.97, with average value

Figure 11. Auto-segmentation sub-step results on a sample patient: (a) a tumor region identified in the tumor localization step, (b) initial contour obtained after incorporating super voxel clustering and tumor localization, (c) final contours after active contour evolving.
of 0.93 ± 0.05. For multiple tumors cases, DC varies from 0.71 to 0.87, with average value of 0.80 ± 0.05. Based on these results, the developed algorithm performs better than other BRATS challenge competing algorithms.

Figure 10 shows the evaluation results for all six metrics among all test cases (T), cases with a single tumor (S), and cases with multiple tumors (M), respectively. We can see that the proposed algorithm performs better in the cases with a single tumor scenario compared to multiple tumors.

3.3. Clinical data

Figure 11 plots auto-segmented sub-step results from our algorithms in coronal, sagittal, transverse, and 3D views. In the 3D view, the ground truth segmentation delineated by a
physician is plotted in red while the auto-segmentation results are in green. In the tumor localization step, the bound tumor regions were obtained by scanning the image slice by slice and a 3D contour was generated by combining bounding boxes as highlighted by the green lines in figure 11(a). After overlapping the bounding boxes with the super voxels resulted from the super voxel clustering, an initial contour was created as shown by green lines in figure 11(b). Then, the initial contour was evolved to final segmentation contours as shown in figure 11(c). This was a case which had the necrosis inside the tumor and the tumor crossed the midline of the brain. If we use the 3D contour given in figure 11(a) as an initial contour, it is highly possible to converge to only the active tumor tissue which is indicated by high intensities. With our initial contour generated with super voxel clustering assistance, the final segmentation result is reasonable.

Quantitative metrics of the 15 cases are listed in table 5. Most values of evaluation metrics are acceptable with several exceptional cases. The reason for these exceptional cases will be detailed in the discussion section.

4. Discussion and conclusions

In this study, we successfully designed and implemented an automatic brain metastases segmentation strategy for radiosurgery applications. The entire automatic segmentation workflow was systematically evaluated on numerical phantoms, the BRATS dataset, and patient images. The results of numerical phantom study show that our segmentation strategy is robust to image noise. Qualitative and quantitative evaluation on BRATS dataset showed that the developed algorithm performed better than existing competing algorithms. Results from patient dataset showed good match between auto and manual results, which indicates that our developed segmentation strategy is promising for clinical applications.

Our method has advantages for clinical applications. Firstly, our method is very simple to use. It only requires a T1c MR image, which matches the current SRS clinical workflow. Secondly, the method doesn’t require parameter tuning for each patient individually since a set of parameters was selected to make the method robust enough across cases. Commonly, active contour based methods are sensitive to input parameters. Due to the similarity among metastatic brain tumor cases and a well-determined initial value, the dependence on parameters is seldom an issue. All evaluated cases presented in previous sections used the same parameters.
Thirdly, our method has a high accuracy. The proposed method achieved a DC of $0.98 \pm 0.01$ in a numerical phantom with different noise levels, a DC of $0.89 \pm 0.08$ in the BRATS data, and a DC of $0.86 \pm 0.09$ in patient data. Fourthly, the algorithm is not sensitive to the number of the brain tumors and was proven to be robust in both single and multiple tumor cases. Last but not least, the method can simply be extended to other brain abnormality segmentation. Following the assumption of brain symmetry, the main idea of our segmentation strategy is to detect a target by capturing any asymmetry in intensity. The algorithm may be used to find any target as long as there is intensity asymmetry. For example, in T2 weighted MRI images edema could be recognized as highlighted areas. Furthermore, asymmetric intensity changes captured from multi-modality images could be combined to provide additional information not seen from a single modality.

The developed auto-segmentation strategy was shown to perform better on a numerical phantom and BRATS data than patient data. This is due to the complexity of real patient cases. The first issue is the existence of cerebral veins. Even though we have used geometrical characteristics to exclude incorrect convergence to cerebral veins, there were still some cases which our algorithm could not properly handle. As shown in Table 5, all evaluation metrics of tumor 7 were worse than the group average. We further investigated this case and found that the metastatic tumor and cerebral vein were spatially overlapped. Figure 12 shows a typical case that the venous overlaps with the tumor. The algorithm treats both vein (in green) and tumor as one target because of their similar intensity and regions are connected. Obviously, we cannot simply exclude the vein by setting a sphericity threshold. In such cases, it is hard to isolate the tumor using our current algorithm. To solve this problem, we plan to take more distinguished features, such as texture, into account in future work (Islam et al. 2013, Brynolfsson et al. 2014).

The tumor detectability in our auto-segmentation strategy depends on the image intensity contrast between tumor and background. In certain cases, the contrast may be low. As shown in Figure 13(a), the contrast between tumor and brain is only 1.197, which requires evaluation by the physician in multiple image planes and potentially utilizing alternative imaging sequences. The red contour shown in Figure 13(b) is the result of T1c image assisted with other image modalities. Currently, these low contrast tumors cannot be detected with our segmentation strategy because the low contrast abnormality is not reflected in asymmetrical intensity changes. In future work, we will consider cooperating contrast-enhancement methods (Celik and Tjahjadi 2012) and multimodality images (Menze et al. 2015) to improve tumor detectability.

Currently the developed strategy can effectively localize a tumor if it partially crosses the middle line of the brain. Theoretically speaking, if the tumor is absolutely symmetrical with respect to the brain middle line, it wouldn’t be detected by the developed strategy. However, in reality, a tumor mass often is asymmetrical. Thanks to our multi-scale strategy, the developed bounding box method can locate the tumor partially. Later, the active contour step can evolve the partial contours to a complete contour set to enclose the entire tumor.

Computational efficiency is another key issue for the practicability of using the proposed strategy in clinical environment. In this study, all algorithms in the workflow were written in C++. For a typical $256 \times 256 \times 176$ T1c image set, segmentation took ~5–7 minutes depending on number of tumors, when running on a laptop with E3-1505M CPU and 32 GB memory. This computational speed is comparable or faster than several algorithms reported in BRATS multiple brain tumor segmentation challenge. Zikic et al. (2012) evaluated their decision forest method and concluded that training takes tens of minutes, while testing takes a few minutes. Bauer et al. (2012) reported that computation time of segmentation on BRAST 2012 dataset with their hierarchical regularization integrated classification method ranged from 4 to 12 minutes.
depending on the size of the dataset. Hamamci et al (2012) claimed their tumor-cut algorithm can finish brain tumor segmentation ~1 s to 16 min depending on number of tumors. Shin (2012) reported that it took 5–10 min to segment a tumor volume depending on the whole head size.

To improve the efficiency, we are planning to implement all algorithms involved in the workflow using graphical processing units (GPU) based on our previous experience (Gu et al 2010, 2011a, 2011b). The tumor localization is conducted slice by slice and will be easily implemented in parallel. The super voxel clustering, SLIC, has a GPU parallel version with NVIDIA CUDA (Ren and Reid 2011), which may speed up 10–20 times compared to the original CPU implementation. The localized level set evolution algorithm can be accelerated by adopting a sparse field method (Lankton 2009). These programming tricks can considerably raise the computational speed to meet clinical needs.

Acknowledgments
This work is supported by China Scholarship Council, Grant No. 201406245043.

References


Bagci U, Udupa J K, Mendhiratta N, Foster B, Xu Z, Yao J, Chen X and Mollura D J 2013 Joint segmentation of anatomical and functional images: applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images Med. Image Anal. 17 929–45


Bauer S, Fejes T, Slotboom J, West R, Nolte L-P and Reyes M 2012 Segmentation of brain tumor images based on integrated hierarchical classification and regularization MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) (Nice, October 2012) p 77


Cordier N, Menze B, Delingette H and Ayache N 2013 Patch-based segmentation of brain tissues MICCAI Challenge on Multimodal Brain Tumor Segmentation (BRATS) (Nagoya, September 2013) p 57


Geremia E, Menze B and Ayache N 2013 Spatially adaptive random forest 2013 IEEE Int. Symp. on Biomedical Imaging: from Nano to Macro pp 1332–5

Geremia E, Menze B H and Ayache N 2012 Spatial decision forests for glioma segmentation in multi-channel MR images MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) (Nice, October 2012) p 77


Guo X, Schwartz L and Zhao B 2013 Semi-automatic segmentation of multimodal brain tumor using active contours MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) (Nagoya, September 2013) p 57

Hakon W A 1935 Volume, shape and roundness of quartz particles J. Geol. 43 250–80


Nieder C, Grosu A L and Gaspar L E 2014 Stereotactic radiosurgery (SRS) for brain metastases: a systematic review Radiat. Oncol. 9 155
Ren C and Reid I 2011 gSLIC: a real-time implementation of SLIC superpixel segmentation Technical Report (Department of Engineering, University of Oxford)
Shin H C 2012 Hybrid clustering and logistic regression for multi-modal brain tumor segmentation MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS), (Nice, October 2012) p 77
Wu J, Aguilera T, Shultz D, Gudur M, Rubin D L, Loo B W Jr, Diehn M and Li R 2016a Early-stage non-small cell lung cancer: quantitative imaging characteristics of 18F fluorodeoxyglucose PET/CT allow prediction of distant metastasis Radiology 281 270–8