Locating Brain Tumors from MR Imagery Using Symmetry

Nilanjan Ray, Baidya Nath Saha, and Matthew Robert Graham Brown
{nray1, baidya, mbrown}@cs.ualberta.ca

Department of Computing Science
University of Alberta
Canada

***Abstract*— Tumor/abnormality segmentation from magnetic resonance imagery (MRI) can play a significant role in cancer research and clinical practice. Although accurate tumor segmentation by radiologists is ideal, it is extremely tedious. Experience shows that for MRI database indexing purposes approximate segmentations can be adequate. In this paper, we propose a straightforward, real-time technique to find a bounding box around the brain abnormality in an MR image. Our algorithm exploits left-to-right symmetry of the brain structure. The proposed detection algorithm can play a useful role in indexing and storage of bulk MRI data, as well as provide an initial step or seed to assist algorithms designed to find accurate tumor boundaries.**

I. Introduction

Currently, the large databases of brain tumor magnetic resonance (MR) images maintained by most clinics are not indexed, and cannot be searched based on clinically relevant tumor characteristics such as location and size. On the one hand, manual tumor segmentation is extremely laborious and tedious given the sheer volume of MR data. On the other hand, reliable and fast automated off-the-shelf tumor segmentation algorithms are equally hard to obtain. Presently available automatic brain tumor segmentation algorithms that attempt to segment the tumor exactly do not perform as reliably as it is desired even from the database indexing purpose. Instead, we propose in this paper an algorithm that finds a less exact segmentation but does so reliably and in real-time.

Exact, automatic segmentation of tumors/edema from brain MRI is a difficult and unsolved problem. For a nice account of this topic from an image analysis and machine learning perspective see [7]. The potential road blocks seem to come from incorporating domain specific knowledge into the algorithms. Voxel-wise classification algorithms, such as those via support vector machine, are typically dependent on the locally computed features (computed within a window around a voxel). On the other hand, incorporation of global region-based features is non-trivial and computationally intensive ([2], [5]). These algorithms typically require registration of MR images, standardization of image intensities, and noise removal [7]. Moreover, many advanced algorithms can be slow and unsuitable for database indexing purposes.

Keeping these views in consideration, we propose here a fast method for locating brain tumors in magnetic resonance (MR) images. Specifically, the algorithm defines bounding boxes around abnormal regions caused by primary brain tumors in typical MR scanning modalities, including T1, T1 with gadolinium, T2, and FLAIR [7]. We believe that our algorithm has at least the following two direct applications in the treatment of brain cancers. Clinical centers currently maintain large amounts of archived brain tumor MR data that is not indexed for easy retrieval according to image properties. Database indexing based on automated tumor location would allow a clinician to retrieve historical cases relevant to the diagnosis and treatment of new patients' cancers. In addition, our proposed automated tumor location algorithm could be used to seed or constrain an automated brain tumor segmentation system.

The key observation underlying our approach is that normal brain structure is roughly symmetric: the left part and the right part can be divided by an axis of symmetry. Tumors/edemas typically disturb this symmetry. We utilize this property to design a real-time algorithm to locate bounding boxes. The proposed algorithm uses only a single MR image and avoids the non-trivial issues of image registration and intensity standardization. Our bounding box finding algorithm is unsupervised and requires no prior training. It uses only two user-tuned parameters. Empirical studies with our algorithm show that these two parameters can be conveniently set by a user to run on a large amount of data.

II. Proposed Technique

We cast the problem of finding a bounding box around brain abnormalities from MRI as a change detection problem [6]. We assume that the region of abnormality is located in one of the two halves of the brain. Thus, in the MRI one half of the brain acts as a reference image, and the other half, as a test image. We need to compare the test image with the reference image to find out the region of abnormality.

To formulate this change detection problem let us consider the setting shown in Fig. 1(a), where we show two images, I and R. Here, I is the test image and R is the reference image. The task is to find the anomalous region D on I that is not present in R. Often, straightforward methods such as taking the point-wise subtraction image $|I-R|$ fail to identify the

978-1-4244-2110-7/08/$25.00 ©2007 IEEE

region differences correctly. The reasons are various; however, the most prominent one is that the left and right halves of a brain do not precisely match on a point-to-point basis, even when we consider a normal brain without tumor/edema. Noise is also present in the difference image.

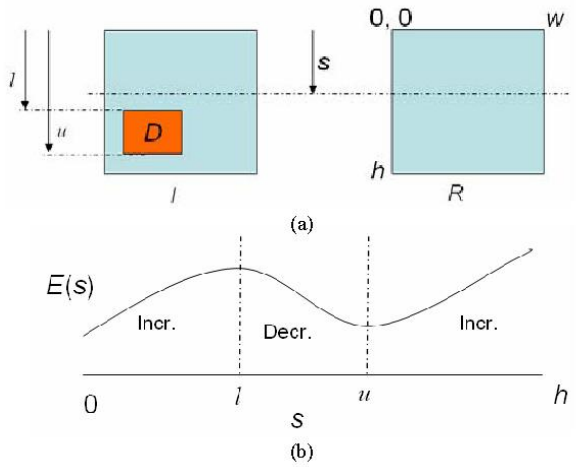


Fig 1: (a) Finding D from image I, using a reference image R, (b) A typical score function plot.

In this paper we use a region-based approach rather than a point-to-point comparison that most of the change detection algorithms use [6]. To detect the region of abnormality D on I, we consider a score function defined by the Bhattacharya coefficient (BC) [4]. We define our proposed score metric as follows. Consider a horizontal dotted line drawn across the images I and R in Fig. 1(a) at a distance s from the tops of the images. We define two rectangles on the image domain, $A(s) = [0,w] \times [0,s]$ and $B(s) = [0,w] \times [s,h]$, where w and h are respectively the width and the height of the images I and R. Note that $A(s)$ and $B(s)$ respectively denote the image domain above and below the horizontal line. We define the proposed score function as:

$$E(s) = \langle \sqrt{P_I^{A(s)}}, \sqrt{P_R^{A(s)}} \rangle - \langle \sqrt{P_I^{B(s)}}, \sqrt{P_R^{B(s)}} \rangle,$$

where P's are normalized gray level intensity histograms (probability mass functions). The subscript on P denotes whether P is constructed on the test image I or on the reference image R. The superscript of P denotes on which portion of the image domain P is constructed. For example, $P_I^{A(s)}$ denotes the normalized intensity histogram of image I within $A(s)$. Here $\langle X, Y \rangle$ is the inner product of two vectors X and Y. Thus the score function is the difference of BC's between the image domains above and below the reference line at s.

BC is a number between 0 and 1. BC between two probability mass functions (p.m.f.) is 1 when they are exactly equal. If two p.m.f.'s are very different, *e.g.*, with disjoint supports, then their BC value will be 0. From this perspective our score metric essentially measures: (a) how similar are the two upper histograms and (b) how dissimilar are the two lower histograms. Thus when our score function has a high value, we expect the two upper histograms to achieve a good match, while the two lower histograms are likely to have a poor match. On the other hand, a low value of $E(s)$ indicates a mismatch between the upper histograms and a good match between the lower histograms.

We claim that the score function helps determine the location of D in a very fast computation. Our claim is expressed in Fig. 1(b), which shows that the score function should first increase, then decrease and then increase again as s increases from 0 to h. The increasing and decreasing segments meet at $s=l$ and $s=u$, at the lower and upper bounds of D, respectively. So, from a score plot we can locate the upper and lower bounds for D quickly. To find the left and right bounds of D, we simply rotate I and R by 90 degrees and follow the same procedure. The following two propositions establish this claim along with the necessary assumptions about the data, *i.e.*, I and R.

Proposition 1: $M(s) + L(s) \geq E(s) \geq L(s)$, where

$$L(s) = \sqrt{\frac{|A(s) \setminus D|}{|A(s)|}} \left\langle \sqrt{P_I^{A(s) \setminus D}}, \sqrt{P_R^{A(s)}} \right\rangle - \sqrt{\frac{|B(s) \setminus D|}{|B(s)|}} \left\langle \sqrt{P_I^{B(s) \setminus D}}, \sqrt{P_R^{B(s)}} \right\rangle,$$

and

$$M(s) = \sqrt{\frac{|A(s) \cap D|}{|A(s)|}} \left\langle \sqrt{P_I^{A(s) \cap D}}, \sqrt{P_R^{A(s)}} \right\rangle + \sqrt{\frac{|B(s) \cap D|}{|B(s)|}} \left\langle \sqrt{P_I^{B(s) \cap D}}, \sqrt{P_R^{B(s)}} \right\rangle.$$

Proof: Note that $P_I^{A(s)}$ can be written as:

$$P_I^{A(s)} = \frac{|A(s) \cap D|}{|A(s)|} P_I^{A(s) \cap D} + \frac{|A(s) \setminus D|}{|A(s)|} P_I^{A(s) \setminus D}.$$

Using the above decomposition for $P_I^{A(s)}$, one can show:

$$\left\langle \sqrt{P_I^{A(s)}}, \sqrt{P_R^{A(s)}} \right\rangle \geq \sqrt{\frac{|A(s) \setminus D|}{|A(s)|}} \left\langle \sqrt{P_I^{A(s) \setminus D}}, \sqrt{P_R^{A(s)}} \right\rangle,$$

and

$$\sqrt{\frac{|A(s) \setminus D|}{|A(s)|}} \left\langle \sqrt{P_I^{A(s) \setminus D}}, \sqrt{P_R^{A(s)}} \right\rangle + \sqrt{\frac{|A(s) \cap D|}{|A(s)|}} \left\langle \sqrt{P_I^{A(s) \cap D}}, \sqrt{P_R^{A(s)}} \right\rangle \geq \left\langle \sqrt{P_I^{A(s)}}, \sqrt{P_R^{A(s)}} \right\rangle.$$

A similar set of two inequalities holds for $\langle \sqrt{P_I^{B(s)}}, \sqrt{P_R^{B(s)}} \rangle$. Combining these four inequalities yields the result. Q.E.D.

Proposition 2: If the following four conditions, *viz.*,

$$(i)\quad \langle \sqrt{P_I^{A(s)\cap D}}, \sqrt{P_R^{A(s)}} \rangle << \langle \sqrt{P_I^{A(s)\setminus D}}, \sqrt{P_R^{A(s)}} \rangle,$$

$$(ii)\quad \langle \sqrt{P_I^{B(s)\cap D}}, \sqrt{P_R^{B(s)}} \rangle << \langle \sqrt{P_I^{B(s)\setminus D}}, \sqrt{P_R^{B(s)}} \rangle,$$

$$(iii)\langle \sqrt{P_I^{A(s)\setminus D}}, \sqrt{P_R^{A(s)}} \rangle \approx c_1, \text{ and}$$

$$(iv)\langle \sqrt{P_I^{B(s)\setminus D}}, \sqrt{P_R^{B(s)}} \rangle \approx c_2,$$

hold where c_1 and c_2 are two constants, then $E(s)$ is (a) increasing when $0 \le s \le l$, (b) decreasing when $l \le s \le u$, and (c) increasing when $u \le s \le h$.

Proof: Assumptions (*i*) and (*ii*) together with Proposition 1 imply: $E(s) = L(s)$. Next applying (*iii*) and (*iv*) in the expression for $L(s)$ yields:

$$L(s) = c_1\sqrt{\frac{|A(s)\setminus D|}{|A(s)|}} - c_2\sqrt{\frac{|B(s)\setminus D|}{|B(s)|}}.$$

Now it is straightforward to verify that $L(s)$ is increasing when $0 \le s \le l$, decreasing when $l \le s \le u$, and again increasing when $u \le s \le h$. Q.E.D.

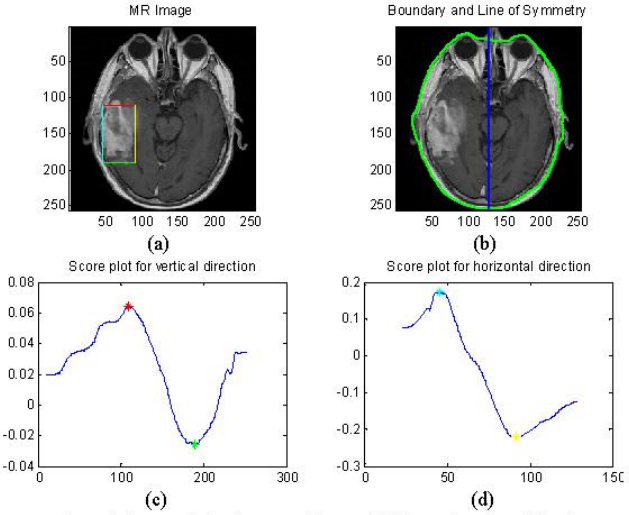


Fig 2: (a) T1 weighted MRI with overlaid bounding found by the proposed method, (b) skull boundary and line of symmetry, (c) $E(s)$ plot for vertical direction, (d) $E(s)$ plot for horizontal direction.

To establish the nature of $E(s)$ as illustrated in Fig. 1(b), the assumptions (*iii*) and (*iv*) can be relaxed. One only needs that the rate of area change with respect to s occurs faster than that of the BC's. On the other hand, the assumptions (*i*) and (*ii*) mean that the abnormality portion of I is different from the rest, when compared with respect to a reference histogram.

The use of our computation is illustrated in Fig 2. In Fig. 2(a), we show a brain MRI (T1 weighted with gadolinium, henceforth referred to as T1C) in which the abnormality is observed on the left side of the image (corresponding to the patient's right side). First, we compute a gradient vector flow snake [8] to find the skull boundary as shown in Fig. 2(b). Next, a vertical line is drawn through the centroid of the snake. This vertical line serves as a line of symmetry (LOS) (also shown on Fig. 2(b)). For an image having considerable rotation, one can fit an ellipse to the snake and take the major axis of the ellipse as the LOS. Now the portion of MRI to the left of the LOS serves as the test image I and the portion of MRI to the right of LOS, after taking a reflection, serves as the reference image R. Note that we do not need to know a priori if the tumor is located on the left or the right side. After finding the bounding box on one side, say on the left side, we test if the average intensity within the bounding on the left side is greater than that on the right side, assuming that tumor/edema will produce stronger signal on T1C images. A score function plot for the vertical direction and another plot for the horizontal direction are shown, respectively, in Fig. 2(c) and Fig. 2(d).

Finally, to locate the bounding box, we detect the extrema of the score plots as shown in Fig. 2(c) and 2(d). The bounding box found is overlaid on Fig. 2(a). Note that the score function has a number of local extremum points. Also note that our propositions cannot guarantee that the extremum points we are seeking are the global ones. In practice, we find these extremum points by the following algorithm.

Algorithm 1

Step 1: Locate all maxima and minima from a score plot. These extrema are found within a neighborhood of size N pixels (a user defined parameter, we take N=41 for all our experiments)

Step 2: Consider all the pairs of consecutive extremum points: (maxp, minp) where maxp is a maximum and minp is a minimum point. From among all such pairs, find the pair (l, u) for which the difference $(E(l) - E(u))$ is the maximum.

Note that in Algorithm 1, following the claim of Proposition 2, we are essentially looking for a pair of points comprised of a consecutive maximum point and minimum point that, respectively, corresponds to the upper and the lower bound of the bounding box. The difference $(E(l) - E(u))$ is the amount of decrease in score function between two consecutive extrema. The neighborhood size N limits the size of the abnormal region we can find. For example, Algorithm 1 will be able to find bounding boxes with heights larger than N.

III. Results

This section illustrates the results of applying our bounding box finding algorithm to brain MRI data. In Fig. 3, we show bounding boxes found by our algorithm on four T1C images. In these examples, the bounding boxes include the regions of abnormality as intended. The two parameters of the proposed bounding box finding algorithm is the number of histogram bins, which we take 64 for all our experiments, and the neighborhood size N that we take 41 pixels for all the experiments.

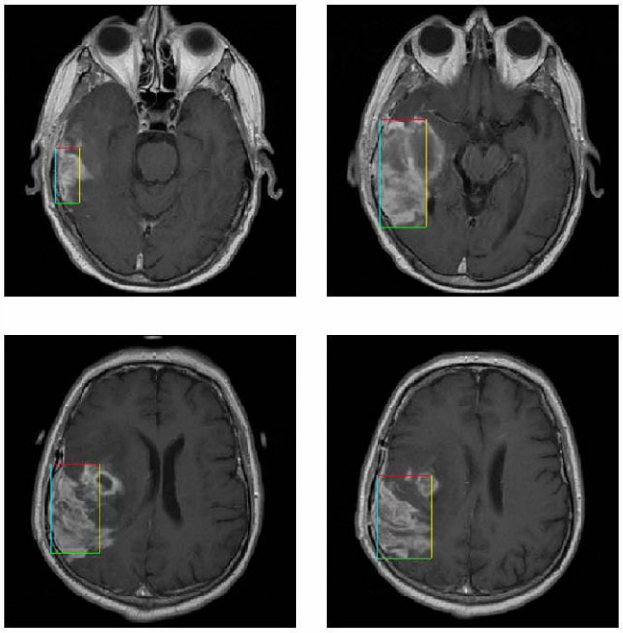


Fig 3: MR images and bounding boxes around abnormal regions.

For performance evaluation of our algorithm we consider the Dice coefficient (DC) [3]: $DC = \frac{2|G \cap S|}{|G| + |S|}$, where S is the set of pixels within a bounding found by our algorithm, and G is the set of pixels belonging to a bounding box computed around the tumor/edema boundary as drawn by a human expert radiologist. The modulus sign in DC denotes the number of pixels belonging to a set. The ideal value of DC is 1, in which case $S = G$. A DC value closer to 1 implies a better segmentation. Fig. 4 shows encouraging Dice coefficient values for two sets of brain MRI data taken from two patients. In Fig. 4 slice numbers refer to the axial slices taken at different heights through the patients' brains.

As already mentioned, other than providing a fast means for database indexing, our algorithm can also provide an initialization for other segmentation algorithms. We elaborate this view in Fig. 5. In Fig. 5(a), a bounding box is first computed by our proposed method. In Fig. 5(b), we show segmentation by the Chan-Vese method [1] starting with an initial contour from the bounding box of Fig. 5(a). Fig. 5(b) shows the final boundary computed by the Chan-Vese algorithm. In Fig. 5(c), we show the Chan-Vese segmentation from a different initial curve (a shrunk skull boundary in this case). We observe that spurious segmentation boundaries are generated in Fig. 5(c). This example illustrates that our proposed bounding box algorithm can aid other algorithms to delineate the region of abnormality.

IV. Future Work And Conclusions

We propose a technique for computing bounding boxes around brain abnormality in standard MR images based on symmetry. The technique uses a scoring function that provides a measure of the similarity or difference between two regions in terms of the Bhattacharya coefficient computed on those region intensity histograms. We provide a mathematical basis of the behavior of this scoring function that essentially locates the bounding box. This region-based and image feature histogram-based approach can open new avenues of brain tumor boundary delineation.

Our approach has several advantages: (a) It exploits approximate left-right symmetry of the brain. (b) No pre-processing, such as intensity standardization or noise removal, is required by our algorithm. (c) It requires no labeled image data, nor any training. (d) It does not require image registration. (e) Only two user defined parameters are used. (f) It can be implemented in real-time. One limitation of our algorithm is that it assumes the tumor/abnormality is confined to the left or right side of the brain and does not cross the LOS. Also, when the tumor is fragmented into multiple parts, our algorithm tends to detect only the most prominent region of abnormality. In the future, we would like to relax these two limitations.

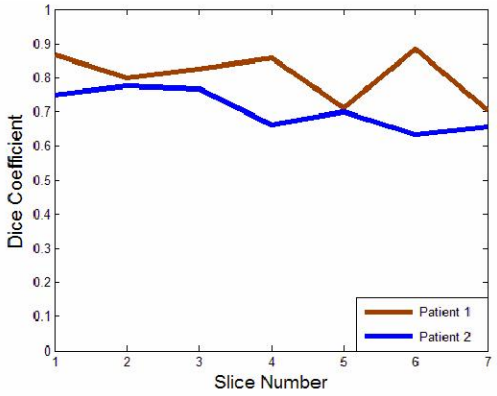
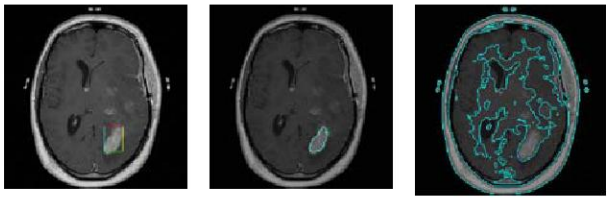


Fig 4: Dice coefficients for MR images for two studies.



(a) (b) (c)

Fig 5: (a) Bounding box, (b) segmentation within bounding box, (c) segmentation on the entire image.

Initial, encouraging results from a few patient studies have prompted us to conduct extensive testing on the patient image database maintained at the Cross Cancer Institute on the University of Alberta campus. We also plan to couple this bounding box finding algorithm with other in-house segmentation algorithms (visit: http://www.cs.ualberta.ca/~btgp/).

Extensions of the proposed algorithm to 3D are straightforward. We are currently making this effort. Last but

not the least, being a change detection algorithm, we plan to extend the application areas of our algorithm to other areas such as video surveillance.

REFERENCES

[1] T.F. Chan and L.A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol.10, no.2, pp.266-277, 2001.

[2] D. Cobzas, N. Birkbeck, M. Schmidt, M. Jägersand, A. Murtha, "3D variational brain tumor segmentation using a high dimensional feature set," In *Mathematical Methods in Biomedical Image Analysis*, a workshop in conjunction with International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brazil, October 2007.

[3] L.R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, pp.297-302, 1945.

[4] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, 2nd ed., 1990.

[5] C-H. Lee, S. Wang, F. Jiao, R. Greiner, D. Schuurmans, "Learning to model spatial dependency: semi-supervised discriminative random fields," *Neural Information Processing Systems*. Vancouver, BC. December 2006.

[6] R.J. Radke, S. Andra, O. Al-Kofahi, B. Roysam, "Image Change Detection Algorithms: A Systematic Survey," *IEEE Trans. Image Processing*, vol.14, no.3, pp.294-307, March 2005.

[7] M. Schmidt, *Automatic brain tumor segmentation*, M.Sc. Thesis, University of Alberta, 2005.

[8] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol.7, no.3, pp.359-369, 1998.