3D Graph cut with new edge weights for cerebral white matter segmentation

Ashish K. Rudra, Mainak Sen, Ananda S. Chowdhury, Ahmed Elnakib, Ayman El-Baz

Abstract

Accurate and efficient automatic or semi-automatic brain image segmentation methods are of great interest to both scientific and clinical researchers of the human central neural system. Cerebral white matter segmentation in brain Magnetic Resonance Imaging (MRI) data becomes a challenging problem due to a combination of several factors like low contrast, presence of noise and imaging artifacts, partial volume effects, intrinsic tissue variation due to neurodevelopment and neuropathologies, and the highly convoluted geometry of the cortex. In this paper, we propose a new set of edge weights for the traditional graph cut algorithm (Boykov and Jolly, 2001) to correctly segment the cerebral white matter from T1-weighted MRI sequence. In this algorithm, the edge weights of Boykov and Jolly (2001) are modified by comparing the probabilities of an individual voxel and its neighboring voxels to belong to different segmentation classes. A shape prior in form of a series of ellipses is used next to model the contours of the human skull in various 2D slices in the sequence. This shape constraint is imposed to prune the original graph constructed from the input to form a subgraph consisting of voxels within the skull contours. Our graph cut algorithm with new set of edge weights is applied to the above subgraph, thereby increasing the segmentation accuracy as well as decreasing the computation time. Average segmentation errors for the proposed algorithm, the graph cut algorithm (Boykov and Jolly, 2001), and the Expectation Maximization Segmentation (EMS) algorithm Van Leemput et al., 2001 in terms of Dice coefficients are found to be $(3.72 \pm 1.12\%)$, $(14.88 \pm 1.69\%)$, and $(11.95 \pm 5.2\%)$, respectively.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Segmentation of the cerebral white matter and other brain structures has always been one of the most challenging and essential tasks in any computer assisted diagnosis system. In this work, we focus on segmenting the cerebral white matter from the input 3D T1-weighted Magnetic Resonance Imaging (MRI) data. The segmentation problem becomes quite difficult due to a combination of several factors like low contrast, presence of noise and imaging artifacts, partial volume effects, intrinsic tissue variation due to neurodevelopment and neuropathologies, and the highly convoluted geometry of the cortex. So, this work is of general interest to the computer vision community. In addition to the technical challenges, the present problem has a lot of clinical significance, e.g., it helps us in the understanding of brain disorders like autism, Alzheimer's disease (Warfield et al., 1999), and cognitive defects in elderly subjects (Yang et al., 2004). Application of graph cut for solving image segmentation problems, including those appearing in medical domain, is a well-studied topic. For one of the earliest works, please see the paper by Boykov et al. (2001). In another work, Boykov and Jolly (2001) presented experimental results of graph cut-based segmentation on abdominal CT images in 2D and 3D. In another seminal work, Ishikawa showed (Ishikawa, 2003) how graph cut can find globally optimum solution for convex energy functions in the context of image segmentation. In this paper, we propose a new set of edge weights for the traditional graph cut algorithm (Boykov and Jolly, 2001). In particular, the edge weights of Boykov and Jolly (2001) are modified by comparing the probabilities of the individual voxels and their neighboring voxels to belong to different segmentation classes. A preliminary version of the paper appears in (Chowdhury et al., 2010). A shape prior in form of a series of ellipses is applied to model the contours of the human skull in the 2D slices of an input sequence. This shape constraint is used to prune the original graph constructed from the input. A subgraph consisting of voxels within the skull contours is being generated. Our graph cut algorithm with new set of edge weights is applied to the above subgraph, thereby increasing the segmentation accuracy as well as decreasing the computation time.
The rest of the paper is organized as follows: in Section 2, we do the literature review and highlight our contribution. In Section 3, we describe the proposed graph cut algorithm with new set of edge weights, and the details of the shape prior. In Section 4, we discuss the experimental results. Finally, in Section 5, we conclude the paper with an outline of future research directions.

2. Literature review

The most popular computer vision techniques for extracting a brain structure from the MRI data can be grouped into three broad categories, namely, (a) statistical models-based segmentation, (b) atlas-based segmentation, and (c) deformable models-based segmentation. The EMS algorithm proposed by Leemput et al. is one such statistical method, which has been extensively used as a tissue classifier for brain MRI. The statistical approaches (Van Leemput et al., 2001; Atkins and Mackiewich, 1998; Farag et al., 2006; Collins et al., 1994) depend heavily on the high accuracy of registration between the atlas brain and the unsegmented brain, which may be hard to obtain in many cases. The main drawbacks of the deformable models (Kass et al., 1987; Caselles et al., 1997; Zeng et al., 1998) include slow speed, strong sensitivity to initialization and incomplete segmentation in case of complicated object boundary with concavities. Graph cuts (Boykov and Jolly, 2001; Boykov et al., 2001; Ishikawa, 2003; Kim and Zabih, 2003; Boykov and Funka-Lea, 2006), have been applied to medical image segmentation problems including structure detections in the human brain (Wolz et al., 2009). The main advantage of graph cuts is that they yield a globally optimal segmentation, which can potentially overcome the problems of oversegmentation and undersegmentation. However, the success of the graph cut algorithm depends on the variation in the capacities of different edges in the graph, which in turn depends on existing contrast in the image. The input MRI images in the present problem exhibit extreme low contrast. Therefore, suitable modifications of the existing graph cut algorithm become necessary to handle the current image segmentation problem.

Our principal contribution in this work is the proposition of an improved graph cut algorithm with a new set of edge weights. This new set of edge weights effectively handles the low contrast between the cerebral white matter and the surrounding anatomical structures in the MRI data. We now state another potential benefit of the proposed method. It has been observed that the choice of the weighting parameter for the data and the smoothness terms in the energy function can considerably affect the performance of the graph cut algorithm (Peng and Veksler, 2008). The performance of our improved graph cut algorithm, in sharp contrast, is much less sensitive to the variation in the weighting parameter. We further employ a simple yet effective shape prior in form of a series of ellipses to model the human skull in individual 2D slices of the input sequence. The third contribution lies in the manner in which the shape prior is used in the improved graph cut algorithm. Unlike the methods described in (Freedman and Zhang, 2005; Song et al., 2006), where the shape constraints are added in form of additional terms in the energy function, we use the shape constraint in a different manner to generate a subgraph by pruning a significant part of the 3D graph (constructed from the input). The proposed algorithm is applied on the newly formed subgraph resulting in a considerable reduction of the execution time and significant improvement of the segmentation accuracy.

3. Methods

The graph cut algorithm proposed by Boykov-Jolly has been used quite effectively in medical image segmentation problems (Wolz et al., 2009). For the sake of completeness, we present their central idea: the user marks some pixels as being part of the object of interest, and some as lying outside the object, i.e., within the background. The segmentation is governed by the following criteria (Boykov and Jolly, 2001):

(1) Each pixel inside the object is given a value according to whether its intensity matches the object’s appearance model; low values represent better matches.
(2) Each pixel in the background is given a value according to whether its intensity matches the background’s appearance model; low values represent better matches.
(3) A pair of adjacent pixels, where one is inside the object and the other is outside, is given a value according to whether the two pixels have similar intensities; low values correspond to contrasting intensities (i.e., to an edge).

Given these criteria for scoring a segmentation, the goal is to devise an algorithm that can find an optimal segmentation. The edge weights in this graph cut can be formulated such that a binary partition derived from the solution of the max-flow min-cut algorithm minimizes certain energy function. As mentioned earlier, success of the graph cut algorithm depends on the correct identification of the inexpensive edges. Thus, coining of a proper energy function to capture the variations in the capacities of the edges becomes the most essential step. In the present problem, the existing intensity distribution is not particularly amenable to a correct minimum cut, which thus leads to an improper segmentation. Therefore, some probabilistic adaptation of the standard graph cut is proposed.

We first discuss the construction of the graph from the input sequence. The input MRI sequence is modeled as a weighted undirected 3D graph $G = G(V, E)$. Let $X$ denote the set of all voxels. Each voxel $x \in X$ is a node/vertex in $G$. In addition, two special terminal nodes, called the ‘source’ $s$ and the ‘sink’ $t$, are considered (Song et al., 2006). There are two types of edges/links in the present network, namely, the $t$-links ($T$) and the $s$-links ($N$). The $t$-links connect all the individual voxels $x$ to the source node $s$ and the sink node $t$. A neighborhood set $Ne(x)$ is assumed for each node $x$. In this paper, we have used the 26-neighborhood, i.e., $|Ne(x)| = 26$. The $s$-links are constructed between each voxel $x$ and its neighboring voxels $y$, where, $y \in Ne(x)$. Therefore, we have $V = X \subset s \subset t$ and $E = T \subset N$.

3.1. New set of edge weights for graph cut

Here, we describe our improved graph cut algorithm with new set of edge weights. Let $A$ define a segmentation, i.e., a classification of all voxels into either the “object” class or the “background” class. Then, following (Boykov and Jolly, 2001), we can write the energy function to be minimized via graph cut as:

$$E(A) = B(A) + \lambda R(A)$$  \hspace{1cm} (1)

where $B(A)$ and $R(A)$ respectively denote the smoothness term/ boundary properties of segmentation $A$ and the data term/region properties of segmentation $A$. Mathematical expressions for these two terms are given below:

$$B(A) = \sum_{x \in X} B_{(x,y)}$$  \hspace{1cm} (2)

$$R(A) = \sum_{x \in X} R_{d}(Ax)$$  \hspace{1cm} (3)
To overcome the problem of low contrast between the cerebral white matter and its surrounding structures, we modify the smoothness term and the data term to obtain significant variations in their values across any two regions in the image with little difference in their intensities. The modification is based on comparison of the probabilities of any voxel and its neighboring voxels to belong to the “object” class (white matter) and the “background” class (rest of the image). Note the following two important observations about the t-link(s) and the n-link(s), respectively in this connection:

(A) For t-links: The higher the weight of a link between the source (sink) and a node, the higher is the probability of that particular node to belong to the background (object) class.

(B) For n-links: The higher the weight of a link between two neighboring nodes, the higher is the probability of both the nodes to be in the same segmentation class.

The improved smoothness term is given by:

\[ B_{(x,y)} = K_{(x,y)} \exp\left(-\left(I_x - I_y\right)^2/2\sigma^2\right) \times \frac{1}{d(x,y)} \]  

(4)

where the term \(d(x,y)\) denotes distance between two voxels \(x\) and \(y\) and the term \(K_{(x,y)}\) depends on the probabilities of these two voxels to belong to the same segmentation class.

Next, let us discuss another problem of the traditional graph cut method. The parameter \(\lambda\) determines the relative importance of the smoothness term and the data term in a graph cut-based segmentation framework (see Eq. (1)). Proper choice of this parameter largely influences the quality of the segmented output (Boykov and Jolly, 2001; Peng and Veksler, 2008). In particular, large values of \(\lambda\) result in a non-smooth boundary thereby leading to oversegmentation. On the other hand, small values of \(\lambda\) make the boundary too smooth causing undersegmentation. Thus selecting a correct \(\lambda\), which would cause neither oversegmentation nor undersegmentation, remains a fundamental problem. Significant user interaction may be necessary at many cases to obtain optimal segmentation. In order to circumvent the above problem, (Peng and Veksler, 2008) developed an automatic parameter selection strategy with a combination of different image features using AdaBoost. However, the above method involves many steps like feature extraction, feature normalization and training the AdaBoost classifier and hence potentially requires significantly more time for completion compared to that of the standard graph cut. In this work, we define the data term in the graph cut algorithm in a judicious manner, which not only solves the low-contrast problem, but also makes the segmentation output quite insensitive to the exact choice of the weighting parameter \(\lambda\). The formulation of the data term is guided by the following careful observations:

(A) If a voxel lies well within the foreground or the background region, the smoothness term should get more importance. In such cases, a voxel is typically surrounded by voxels whose probabilities to belong to any of the segmentation class are considerably higher than the other class.

(B) If a voxel is close to the boundary, then we should give more importance to the data term.

Therefore, the value of \(\lambda\) for any voxel should depend on its position within the image. Let \(O\), \(B\), and \(I\) be the histogram of the object seeds, histogram of the background seeds, and intensity of any voxel \(x\) respectively. Further, let \(Pr(I,O)\) and \(Pr(I,B)\) denote the probability of a certain voxel \(x\) to belong to the class “object” and to the class “background” respectively. In addition, let \(A_x\) be the set of neighbors of any voxel \(x\) for which \(Pr(I,O) > Pr(I,B)\) and \(B_{x}\) be the set of neighbors of any voxel \(x\) for which \(Pr(I,O) < Pr(I,B)\). The parameter \(\lambda\) determines the relative importance to the data term. Consequently, the smoothness term gets more importance when a voxel is well within the object region, we have \(S_A \gg S_B\). Similarly, when a voxel is well within the background region, we have \(S_A \gg S_B\). In both these cases, the exponential term of Eq. (5) and (6) takes a small value, which in turn decreases the value of the data term compared to the smoothness term in the Eq. (4). Consequently, the smoothness term gets more priority. When a voxel is on or near the edge of the foreground and the background, we have \((SA - SB)^2 \rightarrow 0\). Then, following Eqs. (5) and (6), we get larger capacity in the data link. Consequently, the data term gets more importance compared to smoothness term. In this manner, our probabilistic strategy provides appropriate weights to the data term and the smoothness term to the individual voxels, based on their position of within the image.

Algorithm.

Step 1: Evaluation of \(K_{(x,y)}\) – Compare the probabilities of any two voxels \(x \in X\) and \(y \in Ne(x)\) to belong to the same segmentation class using the terms \(Pr(I,O)\), \(Pr(I,B)\), \(Pr(I,O)\) and \(Pr(I,B)\). If both \(x\) and \(y\) have higher probabilities to belong to the same class, assign \(K_1\) to \(K_{(x,y)}\). Otherwise, assign \(K_2\) to \(K_{(x,y)}\).

Step 2: Evaluation of \(K^b_1\) and \(K^b_2\) – Compare the probability of a single voxel \(x\) to belong to the segmentation class “object” or “background” using the terms \(Pr(I,O)\) and \(Pr(I,B)\). If \(x\) has a higher probability to belong to the “object” class, assign \(K_1\) to \(K^b_1\) and \(K_2\) to \(K^b_2\). Otherwise, assign \(K_2\) to \(K^b_1\) and \(K_1\) to \(K^b_2\).

Step 3: Evaluation of exponential term in the data link functions – Partition the neighboring voxels \(y\) of a single voxel \(x\) into two groups called \(A_x\) and \(B_x\), depending on their probabilities to belong to either segmentation class. If a neighbor voxel \(y\) has a higher probability to belong to the “object” class, then it will be grouped into the set \(A_x\), otherwise, in the set \(B_x\). Then, compute the sum of the distance-weighted probabilities of the elements of set \(A_x\) of belong to the “object” class which is termed as \(S_A\) and the sum of the distance-weighted probabilities of the
elements of set $B_0$ of belong to the “background” class which is termed as $S_B$.

Step 4: Apply graph cut with new set of edge weights using the Eqs. (2)–(6).

Note that the nature of the energy function, as defined in Eq. (1) still remains convex, in spite of the modifications. Thus, we can clearly apply the max-flow min-cut algorithm to find a global minimum of this energy function, which corresponds to the optimal segmentation (Kolomogorv and Zabih, 2004).

### 3.2. Imposition of shape prior

A simple geometric shape prior (Slabaugh and Unal, 2005) is applied to extract a subgraph $G_1 = G_1(V_1, E_1)$ by pruning the original graph $G = G(V, E)$. Generation of shape constraint-induced subgraph is in a way similar to the banded graph cut idea (Lombaert et al., 2005). As an alternative to more complex methods for shape modeling like Gradient Vector Flow (GVF) snakes Xu and Prince, 1998, the current approach does not need any further user inputs (other than seeds) and has negligible computational overhead. The contour of the skull in the input axial data is modeled as an ellipse with center $(x, y)$, semi-major axis $a$, and semi-minor axis $b$. The equation for the ellipse is:

$$\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1 \quad (10)$$

The contours of the skull appearing in various 2D slices in the input can be appropriately modeled by a series of ellipses. The center of such a typical ellipse in any slice is obtained from the seeds inputted by the user in that particular slice. The other two parameters of the ellipse, namely, the semi-major axis $(a)$ and the semi-minor axis $(b)$ are extracted from the image using simple geometry. In this way, we can get the ellipse in only a few slices (where the user has provided seeds). For the rest of the slices, bilinear interpolation is employed to obtain the corresponding ellipses. Let $p(x, y)$ and $p_m(x, y)$ respectively denote the coordinates of a pixel $p$ lying on the ellipse in the slice $k$ and that of in the slice $m$, respectively. Then, the coordinates of the pixel $p$ lying on the ellipse in the slice $l$, where $k < l < m$, are given by:

$$p_i(l) = \frac{(l-k)}{(m-k)}p_i(k) + \frac{(m-l)}{(m-k)}p_i(m) \quad (11a)$$

$$p_j(l) = \frac{(l-k)}{(m-k)}p_j(k) + \frac{(m-l)}{(m-k)}p_j(m) \quad (11b)$$

The potential target voxels lie inside the contour of the skull. Once these target voxels, i.e., the vertices of the newly formed subgraph $(P_i \subset P)$ are obtained; the $t$-links $(T_1)$ and the $n$-links $(N_1)$ are constructed. It is obvious that $T_1 \subset T$ and $N_1 \subset N$. So, we can write $V_1 = P_i \cup T \cup E_1$. The subgraph construction has two distinct advantages. Firstly, the time-complexity of the Ford–Fulkerson algorithm, used to find a minimum cut, is reduced considerably. Note that the augmenting paths are detected using the Edmonds–Karp algorithm (Cormen et al., 2001). The time-complexity of the Edmonds–Karp method, on the original graph $G = G(V, E)$ is $O(|V||E|^2)$ and that on the subgraph $G_1 = G_1(V_1, E_1)$ is $O(|V_1||E_1|^2)$. Let us assume that: $V_1 = V \setminus E_1$ and $E_1 = E \setminus E_1$. We decided to fix the value of parameter $\lambda$ is kept fixed at 70. The value of the parameter $\sigma$, which can be treated as “camera noise,” is estimated as 3. For a given low contrast image, there is a lower limit of the parameter $K$ (the ratio of $K_1$ and $K_2$) above which the accuracy of segmentation is independent of the value of $K$. If a suitable value of $K$ (above the lower limit) is chosen, then desired segmentation results can be achieved for the images of same or higher contrast. A plot of Dice coefficient of the proposed improved graph cut algorithm with variation of $K$ for an image with 5% contrast is shown in Fig. 1.

From the graph in Fig. 1, we find that after the parameter $K$ has reached a value of 8.7, the Dice coefficient remains almost constant. We decided to fix the value of $K$ at 10 throughout because the experimental datasets have contrasts greater than or equal to 5%. We next show in a single graph, the variation of Dice coefficient with variation of $\lambda$ for (i) our method and (ii) traditional graph cut method (Boykov and Jolly, 2001). Fig. 2 shows the impact of variations of lambda on the Dice coefficient, used as performance measures of the proposed method. The graph clearly demonstrates the following facts:
(a) Dice coefficient is significantly higher for any $k$ in our proposed method compared to the traditional graph cut. This describes the necessity of the probabilistic modifications in the boundary and region terms.

(b) Variation of Dice coefficient with $k$ is much lower in our proposed method compared to the variation of Dice coefficient with $k$ in the case of traditional graph cut. This emphasizes that the proposed graph cut method with improved

---

Table 1
Performance comparison of various segmentation algorithms in term of Dice coefficients (D.C.).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Min. D.C. (%)</th>
<th>Max. D.C. (%)</th>
<th>Mean D.C. (%)</th>
<th>Std. Dev. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph cut (Boykov and Jolly, 2001)</td>
<td>12.46</td>
<td>15.41</td>
<td>14.88</td>
<td>1.69</td>
</tr>
<tr>
<td>EMS Van Leemput et al., 2001</td>
<td>7.9</td>
<td>18.5</td>
<td>11.95</td>
<td>5.2</td>
</tr>
<tr>
<td>Graph cut with new edge weights</td>
<td>3.16</td>
<td>4.81</td>
<td>3.72</td>
<td>1.12</td>
</tr>
</tbody>
</table>

---

Fig. 2. Variation of Dice coefficient with variation of $\lambda$ in the cases of proposed method and traditional graph cut method.

Fig. 3. From left (row-wise): input, graph cut output, output of our method, and ground truth. Row 1: axial view; Row 2: coronal view; and Row 3: sagittal view. The red circle indicates oversegmentation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
edge weights is more robust compared to traditional graph cut.

The above facts also justify the choice of the functional form of data terms and smoothness terms in Eqs. (4)–(6).

Now, we present a comparative study of the proposed algorithm with two other well-known segmentation algorithms. As we introduce a probabilistic variation of the traditional graph cut (Boykov and Jolly, 2001), it is but natural to compare the proposed improved graph cut algorithm with that of the traditional one. In addition, we choose a non-graph based statistical method, namely the Expectation Maximization Segmentation (EMS) algorithm Van Leemput et al., 2001 for performance comparison.

Table 1 clearly demonstrates the superiority of the proposed method. Our proposed graph cut algorithm, with new set of edge weights, results in most accurate segmentation by generating the best minimum, the best average and the best maximum errors among the three competing methods. The above table also indicates that the graph cut algorithm generates worst minimum error (min. D.C. = 12.46%) and worst average error (mean D.C. = 14.88%) while the EMS algorithm generates the worst maximum error (mean D.C. = 18.5%). We next show the segmented outputs of one typical MRI sequence for these three algorithms. The first row of Fig. 3 shows from left the axial view of the input, the output resulting from Boykov and Jolly (2001), the output resulting from our method and the ground-truth, respectively. The second and third rows, respectively, show the coronal and sagittal views of the same. The graph cut output clearly shows some oversegmentation, as indicated by a red circle in the last row.

In Fig. 4, we show from left the 3D segmented surfaces from the graph cut algorithm, from our algorithm, from the EMS algorithm and the ground truth. These figures indicate that the proposed graph cut algorithm with new
set of edge weights is closest in appearance to the ground truth. In addition, we show the result of our proposed algorithm on a second MRI sequence. Fig. 5 is similar to Fig. 3 where we show the three different views of the input data, 3D segmented output from our improved graph cut algorithm with new set of edge weights and the ground-truth. Fig. 6 is similar to Fig. 4 where we show the 3D segmented surfaces from our proposed algorithm and the ground-truth. The proposed scheme is not only insensitive to the variations of \( \lambda \) (as discussed before), it is found to be quite insensitive to the user inputs as well. For example, even if the centers of the ellipses are off by few pixels, only a negligible change (<0.5%) is noticed in the value of the Dice coefficient. The only basic requirement was object seeds had to be placed in the object region and the background seeds had to be placed in the background region. The average execution time of the proposed algorithm, when run on a MRI dataset, is 5 s on an Intel\textsuperscript{®} Core\textsuperscript{TM} 2 processor with a speed of 2.8 GHz.

5. Conclusion and future work

In this paper, we have addressed the problem of segmentation of cerebral white matter from T1-weighted MRI data. The problem is of great interest to both the computer vision as well as the medical community. The segmentation problem becomes very challenging due to the complicated shape of the white matter and extremely poor contrast between the object and the surrounding structures in the input data. From the medical point of view, accurate segmentation of cerebral white matter will enable us to better understand different neurological disorders. We propose an improved graph cut algorithm with a new set of edge weights based on some probabilistic modifications of the traditional graph cut. The proposed method effectively handles the low contrast problem. It also makes the segmentation performance less sensitive to the variation of the parameter. A shape prior in form of a series of ellipses is introduced to model the human skull in the 2D slices of the input sequence. This shape constraint is used to prune the graph constructed from the input sequence to form a subgraph consisting of voxels lying within the skull. Application of the improved graph cut algorithm on the newly formed subgraph increases the speed of execution as well as the accuracy of the segmentation. The method proposed in this paper is extremely fast and has been shown to outperform both the traditional graph cut and the EMS algorithms.

As a part of the future work, we plan to use more accurate techniques like GVF snakes for modeling the human skulls, which can further improve the segmentation accuracy. The proposed graph cut algorithm with the new set of edge weights has successfully solved a low contrast image segmentation problem. So, one direction of future research will be to employ this algorithm to segment various low contrast images in different domains.

References