Published in final edited form as:

Neuroimage. 2016 February 15; 127: 435–444. doi:10.1016/j.neuroimage.2015.09.032.

Automated Cerebellar Lobule Segmentation with Application to Cerebellar Structural Analysis in Cerebellar Disease

Zhen Yang^{a,*}, Chuyang Ye^b, John A. Bogovic^c, Aaron Carass^{a,d}, Bruno M. Jodynak^e, Sarah H. Ying^f, and Jerry L. Prince^{a,d,e,f}

^aDepartment of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA

^bBrainnetome Center and National Laboratory of Pattern Recognition Institute of Automation, The Chinese Academy of Sciences, Beijing, 100190 China

^cJanelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147 USA

^dDepartment of Computer Science, The Johns Hopkins University, Baltimore, MD 21218 USA

^eDepartment of Applied Math and Statistics, The Johns Hopkins University, Baltimore, MD 21218 USA

^fDepartment of Radiology, The Johns Hopkins School of Medicine, Baltimore, MD 21287 USA

Abstract

The cerebellum plays an important role in both motor control and cognitive function. Cerebellar function is topographically organized and diseases that affect specific parts of the cerebellum are associated with specific patterns of symptoms. Accordingly, delineation and quantification of cerebellar sub-regions from magnetic resonance images are important in the study of cerebellar atrophy and associated functional losses. This paper describes an automated cerebellar lobule segmentation method based on a graph cut segmentation framework. Results from multi-atlas labeling and tissue classification contribute to the region terms in the graph cut energy function and boundary classification contributes to the boundary term in the energy function. A cerebellar parcellation is achieved by minimizing the energy function using the α -expansion technique. The proposed method was evaluated using a leave-one-out cross-validation on 15 subjects including both healthy controls and patients with cerebellar diseases. Based on reported Dice coefficients, the proposed method outperforms two state-of-the-art methods. The proposed method was then applied to 77 subjects to study the region-specific cerebellar structural differences in three spinocerebellar ataxia (SCA) genetic subtypes. Quantitative analysis of the lobule volumes show distinct patterns of volume changes associated with different SCA subtypes consistent with known patterns of atrophy in these genetic subtypes.

*Please address correspondence to: Zhen Yang, Dept. of Electrical and Computer Engineering, Johns Hopkins University, 105 Barton Hall, 3400 N. Charles St., Baltimore, MD 21218. zyang11@jhu.edu, Fax: +14105165566.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Cerebellum; cerebellar lobule segmentation; graph cuts; magnetic resonance imaging; multi-atlas segmentation; random forest classifier; spinocerebellar ataxia

1. Introduction

The cerebellum plays an important role in motor control and sensory integration and is also involved in cognitive functions such as attention, language, and the regulation of fear and pleasure responses. As in the cerebrum, cerebellar function is specialized by location [37, 33, 38], and therefore diseases that affect specific parts of the cerebellum are associated with specific patterns of symptoms [15, 43]. Pronounced regional cerebellar atrophy is known to occur in the spinocerebellar ataxias (SCAs) [15, 43, 24], while more subtle regional cerebellar atrophy has been observed in several other neurological diseases such as Parkinson's disease [8, 42], Huntington's disease [30, 32], Alzheimer's disease [36], and multiple sclerosis [26]. Accurate delineation of the cerebellum as a whole as well as further parcellation into lobes and lobules can be used to better understand cerebellar structural change and to diagnose and monitor disease.

Although the cerebellum has a fairly regular structure [34]—known, in particular, to be less variable than the cerebrum [27]—it nevertheless remains challenging for both manual raters and automatic algorithms to consistently and reliably label the parts of the cerebellum from magnetic resonance (MR) images. For example, experts (requiring thousands of hours of training), take 50–60 hours to label the whole cerebellum, its lobes, and its lobules [10]. Although time and cost can both be reduced by using multiple inexpert raters [10], the time and effort still remains considerable, and consistency of the results across different sets of raters may be poor.

The cerebellar cortex consists of a thin sheet of convoluted gray matter (GM) wrapped around white matter (WM) branches emanating from a central mass of white matter called the corpus medullare (CM). Parts of the cerebellum including and surrounding specific white matter branches are called lobules and are referred to using roman numerals from I to X [34]. Some of these lobules are themselves subdivided and referred to by appending an uppercase letter of the alphabet; for example, lobule VII is divided into VIIA and VIIB. Specific groups of lobules are referred to as lobes; for example, lobules I through V make up the anterior lobe. The boundary between lobes is called a fissure. Figure 1 shows manually labeled regions with their names overlaid on MR images from different orientations. All manual labels are provided by a trained expert according to the manual delineation protocol in Bogovic et al. [10]. See Table 1 for the list of lobes and its lobules used in this paper. Figure 1 also shows manually labeled regions overlaid on the MR images of a healthy subject and an ataxia patient. Note the significant gray matter atrophy in the ataxia patient's cerebellum.

Various automated methods such as FreeSurfer [20], TOADS [35], and others [16] provide a cerebellum segmentation result as part of their whole-brain segmentation method. These methods typically provide a whole cerebellum segmentation and also segment its gray

matter (GM) and white matter (WM). Several specialized methods for segmenting just the cerebellum itself have also been developed; one method uses atlas registration and local image descriptors [29] and another uses an active contour model with a shape prior [22].

For a parcellation of the cerebellum into lobes and lobules, atlas based methods are typically used. In these methods one or more atlas brains are registered to the subject to be segmented and the atlas labels are transferred to the subject brain. Diedrichsen [18] developed a spatially unbiased atlas template of the human cerebellum (SUIT) and later revised it with a probabilistic atlas [19]. Multi-atlas methods reduce the bias towards a specific atlas, and have been shown to be effective combined with various label fusion strategies [40, 21, 2]. However, the accuracy of multi-atlas methods are limited by the quality of the registration result, which is in turn affected by the difference in cerebellar geometry of the atlas and subject. Cerebellar parcellation of subjects with extreme atrophy can therefore be expected to be poor when using these approaches, which is undesirable for quantitative analysis of group structural difference or studying structural-functional correlations. Various label fusion strategies have been proposed to solve these problems in different segmentation tasks [1, 3, 5, 39], including segmenting cerebellar sub-structure [41]. There has been limited work on studies containing both healthy and patients with cerebellar atrophy. Bogovic et al. proposed ACCLAIM (Automatic Classification of Cerebellar Lobules Algorithm using Implicit Multi-boundary evolution) [9], which used a multi-object geometry deformable model driven by random forest boundary classification. The method demonstrated superior performance in comparison to the SUIT atlas-based labeling method and a multi-atlas label fusion approach [40], and was shown to perform well on both healthy controls and cerebellar ataxia patients.

The method described in this paper combines multi-atlas label fusion result and tissue/boundary classification in a graph cut segmentation framework. In particular, the multi-atlas and tissue classification results together determine the region terms in the graph cut energy function while the boundary classifier determines a boundary term in the energy function. The multi-atlas component ensures the correct localization of different lobules, and tissue/boundary classification promotes accurate delineation of lobule boundaries. The proposed method is assessed on a cohort of 15 subjects, including both healthy controls and patients with various degree of cerebellar atrophy, for which expert manual labels are also available. The proposed method outperformed both a state-of-the-art cerebellar lobule segmentation method, ACCLAIM, and a state-of-the-art multi-atlas label fusion method, non-local STAPLE [5], especially on patients with large cerebellar atrophy. To further validate the proposed method and demonstrate its effectiveness in group study, it was applied to a cohort of 77 subjects containing both normal controls and patients experiencing cerebellar ataxia. Quantitative analysis of the lobule volumes show distinct patterns of volume changes associated with different SCA subtypes consistent with known patterns of atrophy in these genetic subtypes.

2. Materials and methods

2.1. Materials and image preprocessing

T1-weighted magnetization prepared rapid gradient echo (MPRAGE) images of 92 subjects were acquired on a 3.0T MR scanner (Intera, Phillips Medical Systems, Netherlands). The parameters of the MPRAGE are: 132 slices, axial orientation, 1.1mm slice thickness, 8° flip angle, TE = 3.9ms, TR = 8.43ms, FOV 21.2 × 21.2cm, matrix 256 × 256 (resolution: 0.828125 × 0.828125 × 1.1mm). 15 of the 92 subjects (containing both HC and patients) were manually delineated by an expert with over 5000 hours of training, using the protocol described in Bogovic et al. [10]. Table 1 lists the lobules that were delineated, grouped by lobes. Table 2 summarizes the demographic information of each diagnosis group, in the expert delineated subject set and in the set with no expert delineation.

The MR images were preprocessed using FreeSurfer version 5.3.0 [20]. As part of this procedure, each MR scan was transformed into MNI space and a skull-stripped and intensity normalized image I_s was generated along with an initial tissue classification of the cerebellum into gray matter (GM) and white matter (WM). We then cropped I_s to contain only the cerebellum, and the cropped image I was used in all subsequent processing. For the 15 subjects having manually delineated labels, the labels were transformed from the native image space into the same space as the preprocessed images so that the images and their corresponding manual labels are aligned.

2.2. graph cut segmentation framework

Let $x \in \Omega$ be a voxel location in a set of discrete voxel locations Ω in the image domain. An image segmentation is a label assignment $A(x)$ that maps each voxel location x to an integer label $k \in \{0, 1, 2, \dots, K\}$. In our formulation, 0 represents background and each non-zero integer represents a finest-grain anatomical structure (see Table 1). The number of structures to be segmented is 26 (i.e., $K = 26$).

Graph cut methods [12, 11] are widely used in various image segmentation tasks for their robustness and accuracy. It casts the energy-based image segmentation problem in a graph structure and finds the optimal solution by efficient min-cut algorithms. A typical graph cut energy function $E(A)$ includes a region term (unary potential) and a boundary term (pairwise potential), evaluated for the label assignment $A(x)$. In our approach, the graph cut energy function is formulated as

$$E(A) = \lambda_l \underbrace{\sum_{x \in \Omega} R_l(x, A(x))}_{\text{Lobule region term}} + \lambda_c \underbrace{\sum_{x \in \Omega} R_c(x, A(x))}_{\text{Cerebellum region term}} + \underbrace{\sum_{\substack{(x, y) \in \Gamma \\ A(x) \neq A(y)}} B(x, y)}_{\text{Boundary term}}, \quad (1)$$

where the constants λ_l and λ_c weight the two region terms relative to the boundary term. Γ is the set of all unordered 26-connected neighborhood voxel pairs. The lobule region term $R_l(x, A(x))$ specifies the cost of assigning a lobule label $A(x)$ to a voxel x . It is computed based on the multi-atlas labeling result, and controls the overall shape of the final segmentation. The cerebellum region term $R_c(x, A(x))$ specifies the cost of assigning cerebellar ($A(x) = 0$) or

non-cerebellar label ($A(x) = 0$), i.e., background, to a voxel x . It is computed based on a cerebellar tissue classification result and aims to refine the lobule-background boundary. The boundary term $B(x, y)$ specifies the cost of assigning a pair of neighboring voxels (x, y) with different lobule labels. It is computed using a boundary voxel classification result and it refines both the lobule-lobule and the lobule-background boundaries. The definition of these energy terms is the key to a good segmentation, which we will describe in detail in the following sections.

For this multiple label assignment problem, we use the α -expansion optimization strategy [12, 11] which successively segments all α and non- α voxels (where α represents a particular label) using the max-flow/min-cut algorithm and then iterates over each label (assigned as the α label) until the label assignments converge.

Figure 2 shows a diagram of the proposed method. First the MR image is preprocessed as described in Section 2.1. Then, multi-atlas registration and label fusion is carried out to obtain both an initial segmentation of the cerebellum and the lobule region term. Tissue classification is performed to obtain the cerebellum region term and boundary classification is performed to obtain the boundary term. Finally, the graph cut energy function is minimized to produce the final segmentation result. In the following subsections, we describe in detail the specification and computation of each energy term in our graph cut formulation.

2.3. Lobule region term

In this section, we define an energy term $R_l(x, A(x))$, which ideally yields a smaller cost when voxel x is assigned the correct label $A(x)$. We derive the necessary information to define this energy term from a multi-atlas labeling result that is carried out as another preprocessing step. Multi-atlas labeling methods register a set of labeled images to the test subject, transfer their labels, and then decide on a single label at each voxel given the set of transferred labels. Various approaches have been proposed for label fusion—e.g., majority voting [31], similarity weighted voting [2, 23, 17], statistical fusion [40, 4, 28]. Statistical methods like STAPLE [40] and its variants [4, 28] integrate a model of rater behavior, and are among the very best performers in the label fusion task. In this work, we use the non-local STAPLE (NL-STAPLE) algorithm proposed by Asman and Landman [5]. NL-STAPLE models the registered atlases as collections of volumetric patches containing both intensity and label information. It uses the non-local criteria [17, 14] to improve results that are otherwise flawed by imperfect registration results.

Each atlas is made up of a preprocessed image I together with a registered expert delineation. To carry out multi-atlas labeling, each of the atlas images is deformably registered to the subject image using the symmetric image normalization method (SyN) [6] implemented in the Advanced Normalization Tools (ANTs) [7]. We use mean square intensity difference as the similarity term in registration. The optimization in registration is performed over two resolutions with a maximum of 10 iterations at the coarse level and 10 at the full resolution. The atlas label is transformed to the subject space using the corresponding deformation field. The transformed labels from all atlases are then fused using NL-STAPLE to produce $A_m(x)$, a multi-atlas subject label.

This multi-atlas labeling result represents a state-of-the-art cerebellar segmentation result by itself, but here we use it as a preprocessing step to help specify the lobule region term in our graph cut energy. In particular, given the multi-atlas label result $A_m(x)$ we define the lobule region term as

$$R_l(x, A(x)) = \min_{y \in \Phi_{A(x)}} \|x - y\|, \quad (2)$$

where $\Phi_l = \{y \in \Omega \mid A_m(y) = l\}$ is the region labeled l in the multi-atlas labeling result. Evaluated at each voxel x , the term $R_l(x, A(x))$ yields the distance to region $\Phi_{A(x)}$, the region with the same label as that of x in the multi-atlas labeling result. If the label at x actually agrees with the multi-atlas label at x , then the cost $R_l(x, A(x))$ is zero. The cost grows larger as x grows farther away from $\Phi_{A(x)}$. Consequently, this lobule region term encourages the final segmentation result to agree with the multi-atlas labeling result.

2.4. Cerebellum region term

Neither FreeSurfer nor multi-atlas labeling provides boundary between the cerebellar and non-cerebellar region. FreeSurfer consistently under-estimates the cerebellum region, i.e., it labels cerebellar lobule voxels as background. Multi-atlas labeling, on the other hand, tends to over-estimate the cerebellum region, i.e., it labels background voxels as cerebellar lobule voxel, especially when the cerebellum has experienced atrophy, such as that which occurs in spinocerebellar ataxia. This occurs even when atrophied cerebellum are included among the atlases used to provide the multi-atlas segmentation result. A better result can be achieved by directly carrying out a cerebellum segmentation step. Several approaches for whole cerebellum segmentation have been proposed [16, 22]. To differentiate cerebellum from non-cerebellum, these methods typically use intensity and texture information together with prior information about the position and shape of the cerebellum. In our work, both image and spatial information are used to train a voxelwise cerebellar tissue classifier. This result is then used to define the cerebellum region term in our graph cut energy function.

Two types of features are used to distinguish cerebellar voxels from non-cerebellar voxels in a preprocessed image I . The first type of features are image features including the intensity, the magnitude of the image gradient, and the trace and determinant of the Hessian matrix of the intensity. The second type of features are spatial features including the spatial coordinates relative to the centroid of the estimated corpus medullare (CM) and the signed distance of the voxel to the boundary of the estimated CM. The estimated CM is obtained by applying a morphological opening operation (using a circular structuring element of diameter 5 pixels) to the cerebellar WM mask obtained from the FreeSurfer segmentation result. This opening operation removes WM branches inside the lobules, yielding a better estimate of the CM. A feature vector $u(x)$ comprising the values of these eight features is constructed for each voxel x .

We use random forests [13] to perform the classification of cerebellar and non-cerebellar voxels given the feature vector $u(x)$. Random forests have been shown to achieve robust and accurate classification while avoiding over-fitting [13]. To form the training samples, voxels were sampled from the two classes, cerebellar and non-cerebellar tissue, among the images

of the training subjects (which are the same set of labeled subjects used in the multi-atlas segmentation step described in Section 2.3). To generate the training samples, 1000 voxels were randomly sampled from each class of each subject. For the random forest configuration, an ensemble of 500 decision trees were constructed, and each decision node within a decision tree used a random subset of two of the eight input features.

Let $A_c(x)$ be the class output by the random forest classifier at voxel x where $A_c(x) = 1$ represents a cerebellar voxel and $A_c(x) = 0$ represents a non-cerebellar (background) voxel. We use this to define the cerebellum region term used in our graph cut energy as follows

$$R_c(x, A(x)) = \begin{cases} \min_{y \in \Psi_0} \|x - y\|, & \text{if } A(x) = 0 \\ \min_{y \in \Psi_1} \|x - y\|, & \text{if } A(x) \neq 0 \end{cases}, \quad (3)$$

where $\Psi_0 = \{y \in \Omega \mid A_c(y) = 0\}$ is the region classified as non-cerebellum by the random forest classifier and $\Psi_1 = \{y \in \Omega \mid A_c(y) = 1\}$ is the region classified as cerebellum. The cerebellum region term $R_c(x, A(x))$ acts in a fashion very similar to the lobule region term $R_l(x, A(x))$ in that it penalizes labels that disagree with the preliminarily estimated labels (in this case by the random forest classifier) by the distance from the region with agreeing label. For example, if the label $A(x)$ is a cerebellar lobule label (i.e., $A(x) = 0$) and the random forest classifies x as cerebellar voxel (i.e., $A_c(x) = 1$) then the cost for x is zero. However, if $A(x)$ is a cerebellar lobule label but the random forest classifies x as a non-cerebellar voxel (i.e., $A_c(x) = 0$), the cost is non-zero and increases with its distance to Ψ_1 , the region classified as cerebellum by the random forest classifier. In this way, when this cost is included in the graph cut energy, as in Equation (1), the final labeling is encouraged to agree with the random forest classification of cerebellar and non-cerebellar tissue.

2.5. Boundary term

A boundary term is used to further refine the lobule boundaries. Lobule boundaries are typically characterized by high image gradients (at boundaries between lobule and non-cerebellar tissue) or high second order directional derivatives (at boundaries between different lobules). In order to encourage our graph cut segmentation algorithm to make cuts that favor such boundaries between labels, we trained another random forest classifier to identify boundary voxels, including both lobule-background and lobule-lobule boundaries.

The same input feature vector $u(x)$ as in Section 2.4 was used, and the same set of atlas images as the multi-atlas segmentation algorithm was used in training. A multiclass classifier was trained to output the class label $k \in \{1, 2, 3\}$. $k = 1$ indicates a boundary voxel, defined as any voxel having any neighboring voxel (within a 26-voxel neighborhood) with a different label. So a boundary voxel can lie between any pair of lobules or between a lobule and the background. $k = 2$ indicates a lobule voxel, which is a cerebellar (non-background) voxel that is not a boundary voxel, and $k = 3$ indicates a non-cerebellar (background) voxel that is not a boundary voxel. A random forest classifier with the same configuration as in Section 2.4 was trained. The training samples were formed by sampling 1000 voxels from each class on each training subject. As before, 500 decision trees were used and each decision node considered a random subset of two of the eight input features.

Rather than combining all decision trees in the random forest to yield a hard classification (of $k = 1, 2$, or 3), we used the outputs of the decision trees to generate a probability of class assignment. Let $h_i(u(x))$ be the class prediction made by the i -th decision tree given the feature vector $u(x)$. Then the probability that the voxel with feature $u(x)$ belongs to class k can be computed as

$$p(k|u(x)) = \frac{1}{500} \sum_{i=1}^{500} \delta[h_i(u(x)) - k], \quad k \in \{1, 2, 3\}, \quad (4)$$

where $\delta[k]$ is the delta function. Figure 3 shows a training image and its 3-class labeling (derived from a manual labeling result) and a test image with the probability of its boundary class output from the trained random forest.

We then define the boundary term as

$$B(x, y) = 1 - \max\{p(1|u(x)), p(1|u(y))\}, \quad (5)$$

which yields a small value when either one of the two neighboring voxels has high boundary class probability and a large value when neither voxel is likely to be a boundary voxel. When this boundary term is used in Equation (1), the optimal cut is encouraged to be positioned along highly probable boundary voxels.

2.6. Energy Minimization

After the energy terms described in Section 2.3, 2.4, and 2.5 are computed, the graph cut energy function in Equation (1) is minimized using the α -expansion algorithm [12]. We used the multi-label energy optimization library available online at <http://vision.csd.uwo.ca/code/>. We chose $\lambda_l = \lambda_c = 1.0$ in Equation (1) empirically according to experimental observations. The algorithm produces highly similar results when λ_l and λ_c are selected in the range $[0.5, 5]$.

3. Experiments

3.1. Accuracy evaluation

In this experiment, 15 subjects with expert delineated labels were used to perform a leave-one-out validation study. In this study, each subject was used as a test subject while the remaining 14 subjects were used in the multi-atlas labeling phase and also as training data for both the cerebellum tissue and boundary classifiers. In this way, results were computed for each of the 15 subjects and statistical performance measures were computed across these 15 subjects. See Table 2 for a summary of the diagnosis and demographic information on the 15 subjects. Three segmentation approaches were compared: ACCLAIM [9], NL-STAPLE (as described in Section 2.3), and the proposed method. To quantitatively evaluate the segmentation results, we first examined the overlap between the true and automatically obtained labels for each lobule using the Dice similarity coefficient (DSC). The DSC of a segmented region R against its manual delineation T is computed as

$$DSC = \frac{2|R \cap T|}{|R| + |T|},$$

where $|\cdot|$ represents the volume of a region.

Figure 4 shows the statistics of the DSC between the manual and automatic labels. For the proposed method, most of the lobules had mean DSC above 0.70, and many were above 0.80 (CM, lobule I-IV, lobule VI, sub-structures of lobule VII, VIII vermis, and lobule IX). The worst mean DSC values were still above 0.60, and these occurred in lobules with relative small volumes such as lobules V and X. These structures also had a larger variation in DSC. Comparing the three segmentation algorithms, NL-STAPLE and the proposed method outperformed ACCLAIM in most of the lobules, with higher mean DSC for all the lobules and smaller DSC variance for almost all the lobules. The proposed method outperformed NL-STAPLE in most of the lobules, especially the lobules in the middle and caudal portions of the cerebellum (lobules XII, IX, and X).

In order to evaluate the overall performance of segmentation on a subject, we computed the average volume weighted DSC (ADSC) over all lobules for each subject. The ADSC is computed as

$$ADSC = \frac{\sum_{i=1}^N |T_i| DSC_i}{\sum_{i=1}^N |T_i|}$$

where DSC_i is the DSC of the i -th lobule, and $|T_i|$ is the volume of the i -th lobule computed from the manual delineation. N is the total number of lobules.

Table 3 lists the ADSC of each subject in the leave-one-out experiment, grouped into three groups: 1) HCs, which have no cerebellar atrophy; 2) SCA6 patients, with atrophy varying from mild to severe; 3) CBs, which are patients with symptoms of cerebellar dysfunction but no genetic diagnosis or other diagnosis. In this particular 15 subject cohort, all the CBs happened to have severe cerebellar atrophy. The Wilcoxon two-sided signed rank test was carried out to compare the ADSCs of the 15 subjects produced by the proposed method and the other two, at a 5% confidence interval. The test results indicated that the proposed method improved the segmentation results over the other two methods significantly in terms of ADSC, with a p-value of 1.2e-4 when compared with ACCLAIM and 3.3e-3 when compared with NL-STAPLE. The mean and standard deviation (std) of ADSC for each group were also computed, as listed in Table 3. The proposed method has the highest mean ADSC among the three algorithms in each group. The ADSCs of the proposed method and NL-STAPLE were similar in the HC group. The ADSCs of the proposed method were higher than NL-STAPLE for all the subjects in both SCA6 and CB groups. The improvements were most prominent in the CB group where subjects have large cerebellar atrophy. Figure 5 shows example segmentation results generated by the three algorithms. We can see that NL-STAPLE and the proposed method outperformed ACCLAIM for both

lobule location and boundary accuracy. The proposed method showed improvements over NL-STAPLE at the lobule boundaries, being able to capture the deep sulci and fissures.

3.2. Volumetric analysis in SCA

In this experiment, we applied the proposed segmentation method to 77 subjects (with no expert delineation) to study the region-specific cerebellar atrophy patterns in ataxia patients and to further validate the proposed method. The 15 expert labeled subjects were used for training. We studied three sub-types of SCA based on the volumetric measurements of the parcellated regions of interest (ROIs) from our method. The demographic information of the different groups is summarized in Table 2. In order to compare the degree of atrophy across different ROIs, we evaluated a relative ROI volume computed as follows. Let $v_{r,s}$ be the volume of ROI r in subject s . Let $\bar{v}_{r,HC}$ be the average volume of ROI r over all healthy subjects. The relative ROI volume of ROI r for subject s is computed as

$$\tilde{v}_{r,s} = \frac{v_{r,s}}{\bar{v}_{r,HC}}.$$

Figure 6 shows the statistics of the relative ROI volumes of gross anatomical divisions, while Figure 7 shows the statistics of relative ROI volumes of cerebellar lobules. A two-sided Wilcoxon rank sum test was carried out between the relative ROI volumes of each ataxia type and the HCs for each ROI, at a 5% confidence interval. The Wilcoxon rank sum test had a null hypothesis that the ROI volumes of the two groups have equal medians, against the alternative that they do not. The test results are marked at the bottom of the box-plots in Figure 6 and 7, where one star (p-value between 0.01 and 0.05) indicates the volume difference between the patient and healthy group is statistically significant, and two stars (p-value less than 0.01) indicates the difference has strong statistical significance.

As shown in Figures 6 and 7, SCA6 exhibits a global atrophy in comparison to the HCs, with significant atrophy across all lobules. The average relative ROI volumes against HCs were between 0.6 and 0.8 for CM, lobules in anterior and middle lobe, and between 0.7 and 0.9 for lobules in posterior lobes. SCA2 showed significant atrophy of the CM, middle lobe, and superior vermis with relative sparing of the posterior-inferior regions of the cerebellum. The above results were consistent with the observations reported in [24, 25]. For SCA3, significant atrophy was found in the CM region and near significant atrophy was found in the anterior lobes. The fact that no significant volume decrease was found in other ROIs partially agreed with [15], which reported significant volume decrease in SCA3 against HCs in cerebellar vermis, and no significant volume change in cerebellar hemisphere.

4. Conclusion and discussion

This paper proposed a fully automated cerebellar lobule segmentation method that combines a multi-atlas result and tissue/boundary classification in a graph cut framework. The multi-atlas component ensures the correct localization of different lobules, and tissue/boundary classification promotes accurate delineation of lobule boundaries. The method was validated using a leave-one-out study on 15 expert delineated subjects including both healthy controls

and patients with various degree of cerebellar atrophy. The proposed method yielded accurate results in three groups of subjects with mild to severe cerebellar atrophy, with mean ADSC 82.0% for control group, 80.5% for SCA6 group, and 78.6% for CB group. The proposed method outperformed two state-of-the-art cerebellum segmentation methods, especially on subjects with moderate to large cerebellar atrophy.

Although the proposed method outperformed the other two methods in identifying cerebellar atrophy, the segmentation results were still biased towards the majority of the training data, in our case, subjects with none to moderate atrophy. The segmentation tended to over-estimate cerebellum tissue in subjects with large atrophy, resulting in a lower dice in the CB group.

Experiments on a larger cohort demonstrated the effectiveness of the proposed method for ROI volumetric analysis of the cerebellum. The proposed method was applied to label the cerebellar lobules of a cohort consisting of healthy controls and patients with different SCA sub-types. Quantitative analysis of the ROI volumes showed significant regional volume decreases in all SCA sub-types and distinct patterns of atrophy for each subtypes. The observed atrophy patterns for different SCA sub-types are consistent with previous findings [25, 24], in which cerebellar lobules were manually labeled by experts.

The run time of the algorithm for one subject is ~7 hours on a computer with 48GB Ram and 12 cores each with a 2.7Ghz Intel processor. It includes ~6 hours of FreeSurfer processing, ~40 mins for multi-atlas registration and label fusion, and ~20 mins for computing the graph cut energy terms and optimization. One way to reduce the algorithm runtime is to replace the FreeSurfer processing part by other brain MR image analysis software that provides cerebellar tissue segmentation, e.g. TOADS [35], which take ~3 hours to produce a whole brain segmentation. Since it can take 50–60 hours for an expert rater to label the cerebellar lobules, the proposed method has potential for MRI based volumetric studies involving large numbers of subjects. Due to the data-driven nature of multi-atlas methods and learned tissue/boundary classifiers, the proposed method can be applied to any cerebellar parcellation protocol and any image modality, providing manually labeled training data with the same protocol and image modality.

Acknowledgments

This work is supported by NIH/NINDS 5R01NS056307-08.

References

1. Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*. 2009; 46:726–738. [PubMed: 19245840]
2. Artaechevarria X, Munoz-Barrutia A, Ortiz-de Solorzano C. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Trans. Med. Imag.* 2009; 28:1266–1277.
3. Asman AJ, Landman B, et al. Formulating spatially varying performance in the statistical fusion framework. *IEEE Trans. Med. Imag.* 2012; 31:1326–1336.
4. Asman AJ, Landman BA. Robust statistical label fusion through consensus level, labeler accuracy, and truth estimation (COLLATE). *IEEE Trans. Med. Imag.* 2011; 30:1779–1794.

5. Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis*. 2013; 17:194–208. [PubMed: 23265798]
6. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. 2008; 12:26–41. [PubMed: 17659998]
7. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ants similarity metric performance in brain image registration. *NeuroImage*. 2011; 54:2033–2044. [PubMed: 20851191]
8. Bhattacharya K, Saadia D, Eisenkraft B, Yahr M, Olanow W, Drayer B, Kaufmann H. Brain magnetic resonance imaging in multiple-system atrophy and parkinson disease: a diagnostic algorithm. *Archives of Neurology*. 2002; 59:835–842. [PubMed: 12020268]
9. Bogovic, JA.; Bazin, PL.; Ying, SH.; Prince, JL. Automated segmentation of the cerebellar lobules using boundary specific classification and evolution. In: Gee, JC.; Joshi, S.; Pohl, KM.; Wells, WM.; Zöllei, L., editors. *Information Processing in Medical Imaging*. Springer; 2013a. p. 62-73.
10. Bogovic JA, Jedynak B, Rigg R, Du A, Landman BA, Prince JL, Ying SH. Approaching expert results using a hierarchical cerebellum parcellation protocol for multiple inexperienced human raters. *NeuroImage*. 2013b; 64:616–629. [PubMed: 22975160]
11. Boykov Y, Funka-Lea G. Graph cuts and efficient ND image segmentation. *International Journal of Computer Vision*. 2006; 70:109–131.
12. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell*. 2001; 23:1222–1239.
13. Breiman L. Random forests. *Machine Learning*. 2001; 45:5–32.
14. Buades, A.; Coll, B.; Morel, JM. A non-local algorithm for image denoising, in: *Computer Vision and Pattern Recognition, 2005.. CVPR 2005. IEEE Computer Society Conference on; IEEE; 2005.* p. 60-65.
15. Bürk K, Abele M, Fetter M, Dichgans J, Skalej M, Laccone F, Didier-jean O, Brice A, Klockgether T. Autosomal dominant cerebellar ataxia type i clinical features and MRI in families with sca1, sca2 and sca3. *Brain*. 1996; 119:1497–1505. [PubMed: 8931575]
16. Ciofalo C, Barillot C. Atlas-based segmentation of 3D cerebral structures with competitive level sets and fuzzy control. *Medical Image Analysis*. 2009; 13:456. [PubMed: 19362876]
17. Coupé P, Manjón JV, Fonov V, Pruessner J, Robles M, Collins DL. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*. 2011; 54:940–954. [PubMed: 20851199]
18. Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. *NeuroImage*. 2006; 33:127–138. [PubMed: 16904911]
19. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. *NeuroImage*. 2009; 46:39–46. [PubMed: 19457380]
20. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002; 33:341–355. [PubMed: 11832223]
21. Heckemann RA, Hajnal JV, Aljabar P, Rueckert D, Hammers A. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*. 2006; 33:115–126. [PubMed: 16860573]
22. Hwang J, Kim J, Han Y, Park H. An automatic cerebellum extraction method in T1-weighted brain MR images using an active contour model with a shape prior. *Magnetic Resonance Imaging*. 2011; 29:1014–1022. [PubMed: 21616622]
23. Isgum I, Staring M, Rutten A, Prokop M, Viergever MA, van Ginneken B. Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans. *IEEE Trans. Med. Imag*. 2009; 28:1000–1010.
24. Jung BC, Choi SI, Du AX, Cuzzocreo JL, Geng ZZ, Ying HS, Perlman SL, Toga AW, Prince JL, Ying SH. Principal component analysis of cerebellar shape on MRI separates SCA types 2 and 6 into two archetypal modes of degeneration. *The Cerebellum*. 2012a; 11:887–895. [PubMed: 22258915]

25. Jung BC, Choi SI, Du AX, Cuzzocreo JL, Ying HS, Landman BA, Perlman SL, Baloh RW, Zee DS, Toga AW, et al. MRI shows a region-specific pattern of atrophy in spinocerebellar ataxia type 2. *The Cerebellum*. 2012b; 11:272–279. [PubMed: 21850525]
26. Kutzelnigg A, Faber-Rod JC, Bauer J, Lucchinetti CF, Sorensen PS, Laursen H, Stadelmann C, Brück W, Rauschka H, Schmidbauer M, et al. Widespread demyelination in the cerebellar cortex in multiple sclerosis. *Brain Pathology*. 2007; 17:38–44. [PubMed: 17493036]
27. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. Automated talairach atlas labels for functional brain mapping. *Human Brain Mapping*. 2000; 10:120–131. [PubMed: 10912591]
28. Landman BA, Asman AJ, Scoggins AG, Bogovic JA, Xing F, Prince JL. Robust statistical fusion of image labels. *IEEE Trans. Med. Imag.* 2012; 31:512–522.
29. van der Lijn, F.; de Bruijne, M.; Hoogendam, YY.; Klein, S.; Hameeteman, R.; Breteler, MM.; Niessen, WJ. Cerebellum segmentation in MRI using atlas registration and local multi-scale image descriptors. *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on; IEEE; 2009.* p. 221-224.
30. Rodda RA. Cerebellar atrophy in Huntington's disease. *J. Neurol. Sci.* 1981; 50:147–157. [PubMed: 6453209]
31. Rohlfing, T.; Brandt, R.; Maurer, CR., Jr; Menzel, R. Bee brains, b-splines and computational democracy: Generating an average shape atlas. *Mathematical Methods in Biomedical Image Analysis, 2001. MMBIA 2001. IEEE Workshop on; IEEE; 2001.* p. 187-194.
32. Rub U, Hoche F, Brunt ER, Heinsen H, Seidel K, Del Turco D, Paulson HL, Bohl J, von Gall C, Vonsattel JP, Korf HW, den Dunnen WF. Degeneration of the cerebellum in Huntington's disease (HD): possible relevance for the clinical picture and potential gateway to pathological mechanisms of the disease process. *Brain Pathol.* 2013; 23:165–177. [PubMed: 22925167]
33. Schlerf JE, Verstynen TD, Ivry RB, Spencer RM. Evidence of a novel somatopic map in the human neocerebellum during complex actions. *Journal of Neurophysiology*. 2010; 103:3330–3336. [PubMed: 20393055]
34. Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *NeuroImage*. 1999; 10:233–260. [PubMed: 10458940]
35. Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL. A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage*. 2010; 49:1524–1535. [PubMed: 19766196]
36. Sjoberg M, Englund E. Alzheimer's disease and the cerebellum: a morphologic study on neuronal and glial changes. *Dement Geriatr Cogn Disord*. 2001; 12:211–218. [PubMed: 11244215]
37. Stoodley CJ, Schmahmann JD. Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *NeuroImage*. 2009; 44:489–501. [PubMed: 18835452]
38. Strick PL, Dum RP, Fiez JA. Cerebellum and nonmotor function. *Annual Review of Neuroscience*. 2009; 32:413–434.
39. Tang X, Oishi K, Faria AV, Hillis AE, Albert MS, Mori S, Miller MI. Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. *PloS ONE*. 2013; 8:e65591. [PubMed: 23824159]
40. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.* 2004; 23:903–921.
41. Weier K, Fonov V, Lavoie K, Doyon J, Collins DL. Rapid automatic segmentation of the human cerebellum and its lobules (RASCAL)–implementation and application of the patch-based label-fusion technique with a template library to segment the human cerebellum. *Human Brain Mapping*. 2014
42. Wu T, Hallett M. Reply: The cerebellum in Parkinson's disease and parkinsonism in cerebellar disorders. *Brain*. 2013; 136:e249. [PubMed: 23739172]
43. Ying S, Choi S, Perlman S, Baloh R, Zee D, Toga A. Pontine and cerebellar atrophy correlate with clinical disability in SCA2. *Neurology*. 2006; 66:424–426. [PubMed: 16476946]

Highlights

- Proposed an automated cerebellar lobule segmentation method
- Validated on expert delineated cohort containing both healthy and patients
- Performed better than two state-of-the-art segmentation methods
- Applied to ataxia dataset and showed group differences consistent with literature

Figure 1.

Annotated examples of the cerebellar lobules in three different views. The top and bottom rows are MR images overlaid with expert delineated lobules from a healthy control and an ataxia patient respectively.

Figure 2.
Diagram of the proposed algorithm.

Figure 3.

Example training and prediction result of the boundary classifier. (a) Preprocessed MR image of a subject used for training. (b) Voxel class overlaid with the image in (a), where yellow indicates boundary voxels, blue indicates lobule voxels and otherwise non-cerebellar voxels. (c) Preprocessed MR image of a test subject. (d) Boundary probability output from the random forest classifier overlaid with the image in (c).

Figure 4. Box plots of Dice similarity coefficient comparing ACCLAIM, NL-STAPLE, and the proposed method.

Figure 5.

Example lobule segmentation results. Each column contains the preprocessed MR image (coronal slice), the segmentation results and the expert delineation of a subject. From the top to the bottom row are in turn the preprocessed MR images, the segmentation results obtained by ACCLAIM, NL-STAPLE, the proposed method, and the expert delineations.

Figure 6.
Box plots of relative lobe volumes for healthy controls and three SCA subtypes.

Figure 7.
Box plots of relative lobule volumes for healthy controls and three SCA subtypes.

Table 1

Cerebellar lobe and lobules.

Lobe	Lobule
Left anterior lobe	Left lobule I-IV, Left lobule V
Right anterior lobe	Right lobule I-IV, Right lobule V
Left middle lobe	Left lobule VI, Left lobule VIIA (CRUS I), Left lobule VIIA (CRUS II)
Right middle lobe	Right lobule VI, Right lobule VIIA (CRUS I), Right lobule VIIA (CRUS II)
Left caudal lobe	Left lobule VIIB, Left lobule VIIIA, Left lobule VIIIB, Left lobule IX, Left lobule X
Right caudal lobe	Right lobule VIIB, Right lobule VIIIA, Right lobule VIIIB, Right lobule IX, Right lobule X
Superior vermis	Vermis VI, Vermis VII
Inferior vermis	Vermis VIII, Vermis IX, Vermis X

Left lobule I-IV indicates the union of lobules I through IV.

Table 2

Demographic information on both the set of 15 subjects with expert delineation and the set of 77 subjects without expert delineation.

	Expert delineated				No expert delineation			
	<i>N</i>	(m/f)	Age	(SD)	<i>N</i>	(m/f)	Age	(SD)
HC	6	(2/4)	54.3	(14.7)	44	(19/25)	56.5	(12.9)
CB	3	(1/2)	54.3	(8.0)	-	(-/)	-	(-)
SCA2	-	(-/)	-	(-)	4	(3/1)	48.8	(8.8)
SCA3	-	(-/)	-	(-)	7	(2/5)	51.7	(9.3)
SCA6	6	(2/4)	55.3	(12.6)	22	(6/16)	58.8	(8.4)

Key: *N* is the number of subjects; m/f is the male/female ratio; Age is the mean age; SD is the standard deviation of the age; healthy controls (HC); people who have symptoms of cerebellar dysfunction but no genetic diagnosis (CB); spinocerebellar ataxia type 2 (SCA2), type 3 (SCA3), and type 6 (SCA6).

Table 3

The average volume weighted dice similarity coefficient (ADSC) for each subject in the leave-one-out experiment, grouped by diagnosis.

Method	Healthy control						mean $\pm$ std.
	1	2	3	4	5	6	
ACCLAIM	70.9	77.0	66.0	73.9	79.1	77.2	74.2 $\pm$ 4.9
NL-STAPLE	82.4	82.6	83.2	76.0	84.3	82.5	81.8 $\pm$ 2.9
Proposed	82.1	82.5	83.8	76.2	84.2	83.0	82.0 $\pm$ 2.9

Method	SCA6						mean $\pm$ std.
	1	2	3	4	5	6	
ACCLAIM	77.2	79.0	73.9	73.2	76.5	73.6	75.6 $\pm$ 2.3
NL-STAPLE	82.2	82.6	80.4	80.1	74.4	79.0	79.8 $\pm$ 3.0
Proposed	83.1	82.7	80.5	81.9	75.5	79.4	80.5 $\pm$ 2.8

Method	CB			mean $\pm$ std.
	1	2	3	
ACCLAIM	71.5	67.9	72.5	70.6 $\pm$ 2.4
NL-STAPLE	72.9	73.1	80.3	75.4 $\pm$ 4.2
Proposed	77.2	77.5	81.2	78.6 $\pm$ 2.2

See Table 2 for a key of the diagnoses.