Published in final edited form as:

Magn Reson Med. 2010 September ; 64(3): 893–901. doi:10.1002/mrm.22416.

Automatic Segmentation of Rodent Spinal Cord Diffusion MR Images

Vanessa K. Tidwell¹, Joong H. Kim², Sheng-Kwei Song², and Arye Nehorai¹

¹Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

²Department of Radiology, Washington University School of Medicine, St. Louis, MO 63110, USA

Abstract

Magnetic resonance imaging (MRI) is a key tool for non-invasive spinal cord lesion analysis; however, accurate, quantitative methods for this analysis are lacking. A new, multi-step, multidimensional approach, utilizing the Classification Expectation Maximization (CEM) algorithm, is proposed for MRI segmentation of spinal cord tissues. Diffusion tensor imaging is used to generate multiple images of each spinal slice with different diffusion direction weightings. The maximum likelihood tissue classifications are then jointly estimated to produce a binary classification image, corresponding to voxels containing either spinal cord or background. Edge detection is employed to find a non-parametric curve encapsulating the entire spinal cord. The algorithm is evaluated using data from *in vivo* DTI of control and injured mouse spinal cords. The algorithm is shown to remain accurate for whole spinal cord, white matter, and hemorrhage segmentation in the presence of significant injury. The results of the method are shown to be at least on par with expert manual segmentation.

Keywords

spinal cord injury; diffusion tensor imaging (DTI); tissue classification; classification expectation maximization (CEM) algorithm

In spinal cord injury (SCI), the amount of total parenchyma or surviving white matter is known to be strongly related to post-injury neurological function (1–4). Objective quantification of these regions of interest is critical in both fundamental pathophysiological study and the development of effective treatment. The most universally accepted method for accurate segmentation analysis is histology (5–7), but its use is limited to postmortem study due to its invasive nature. In contrast, MRI is well suited for noninvasive diagnosis of living tissue. MRI-based spinal cord lesion reporting, both *in vivo* and *ex vivo*, shows good agreement with conventional histology validation and reflects clinical disabilities (8–11). However, there is a lack of objective and precise quantitative methods for segmentation of total parenchyma or white matter in MR images.

Existing methods for *in vivo* transaxial spinal cord segmentation in MRI can generally be categorized into two broad classes. The first, most common class requires significant human intervention, and ranges from entirely manual segmentation to computer-aided manual edge selection. These segmentation methods are subject to human bias and are therefore unreliable

and generally not reproducible. They are also slow and therefore impractical for analyzing large data sets. The second class seeks to define a contour around the cord automatically, based on image gradients and pixels intensities within and outside the contour, with minimal human intervention. These approaches generally use contour methods such as snakes (12) or level-sets (13), and they vary in speed but are in general much faster than manual segmentation. A recent example of this second class is presented in (14), which uses a B-spline snake approach to find the spinal cord contour from *in vivo* MR images of healthy and mildly-injured rat spines. This method relies on human intervention to select the midpoint of the spine as the seed point for the snake algorithm, and it generates segmentations in an average of 1.6 seconds per slice (14).

The automatic methods outlined above all make assumptions about the cord shape that do not necessarily hold in the case of an injured spinal cord. Injured cords shrink as the tissue atrophies and can assume very irregular shapes. Contour-based methods such as snakes or manual edge-tracing assume relatively smooth edges. Additionally, severely injured cords may have voids or hemorrhage within the spinal cord, which violates the basic assumption of contour-based methods—that a single continuous boundary can be found to separate the tissue of interest from the rest of the image. Robust methods that eliminate these faulty assumptions are needed.

We propose a multi-dimensional, multi-step Classification Expectation Maximization (CEM)-based algorithm for spinal cord segmentation. Our algorithm is multi-dimensional in that it segments the cord based on a set of MR images collected for Diffusion Tensor Imaging (DTI) analysis. This joint segmentation incorporates significantly more data than is available in a single MRI image, and thus is more robust to noise in individual images.

We also extend our algorithm to automatic *in vivo* segmentation of spared white matter and regions of hemorrhage in injured spines. Previous studies of automatic MRI segmentation of spinal white/gray matter have only attempted to validate their algorithms for uninjured cords (15,16), or excised cords, imaged *ex vivo* (17). To our knowledge, no previous work has been done on automatically segmenting areas of spinal hemorrhage.

Our method is novel in medical image segmentation due to its multi-step approach, which allows improved segmentation accuracy by incorporating both prior knowledge of cord geometry and the distinct information contained in the different images in successive steps. For example, in the proposed algorithm, the initial step for rough spinal cord segmentation is based on the $b=0$ image, which is acquired without diffusion-sensitizing gradient pairs and so in general is a T2-weighted (T2W) image. The next steps provide further, more exact, segmentation based jointly on the diffusion weighted images (DWIs). To our knowledge, no previous work on automatic *in vivo* transaxial MRI spinal cord segmentation has incorporated the information in T2W images in addition to the DWIs (14,15,17).

Additionally, to our knowledge the CEM algorithm has not been used for spine segmentation. Our algorithm differs greatly from previous, contour-based cord segmentation approaches in that our algorithm is a voxel classification algorithm—voxels are classified individually rather than grouped according to a single contour. In contrast to existing contour-based algorithms, the proposed algorithm defines contours only as a means to generate localization constraints on which voxels may be classified as particular tissue types.

BACKGROUND

Spinal cord histology allows high resolution tissue segmentation with clear separation between the tissue types. The goal for MRI spinal cord segmentation is to use *in vivo* imaging to approximate, as closely as possible, the tissue types that one could generate via histology, thus aiding the treatment and evaluation of spinal cord injuries. MRI is, however, much lower in

resolution, and therefore generating these accurate segmentations is non-trivial. In this section, we discuss the relevant attributes of spinal cord diffusion MR images, and the challenges faced when automatically segmenting these images, particularly in the case of injured cords.

Characteristics of DTI

DTI provides micro-structural information with greater sensitivity to tissue integrity than conventional MRI. Many studies have reported the potential of 6-direction DTI derived parameters to reveal the morphological integrity and pathophysiological changes of living tissue in rodent spinal cord studies (8,18–20).

In the T2W image, the brightest area is located within the spinal canal, containing both cord and cerebrospinal fluid (CSF), with minimal differentiation in voxel intensity level between the white and gray matter tissue types. In severely injured cords, the intensity level of the CSF can on occasion differ slightly from that of the cord; however, even when there is variation between the cord and CSF, the difference in intensity level between the background and the spinal cavity is always much greater.

DWIs are brightest only in the region corresponding to spinal cord, not cerebrospinal fluid. The voxel intensity level of the CSF in the DWIs is equivalent to that of the background tissues. This is because the signal from CSF attenuates significantly in diffusion weighted imaging in any direction. Voxel intensities in DWIs differentiate strongly between white and gray matter; in some diffusion gradient directions, the white matter is brighter than the gray matter, while in other directions, the gray matter is brightest.

From the DWIs, DTI maps can be calculated. These DTI maps—in particular, the relative anisotropy, the axial diffusivity, and the radial diffusivity—do not have the same useful properties as the T2W image or the DWIs to allow for simple segmentation of the spine from the background tissue. However, they have clearer intensity separation between the white and gray matter than the DWIs, particularly in the case of injured cords.

Challenges for Automatic Segmentation

There are many challenges for accurate voxel-by-voxel classification of the spinal MR images. For instance, there are frequently scattered bright spots in both the T2W image and in the DWIs that lie outside the spinal cord. Because the two types of images are sensitive to different tissue properties (spin-spin relaxation and diffusion), these extraneous bright spots are in general not co-located in the T2W image and the DWIs. Thus, by using both types of images, we can achieve more accurate voxel-by-voxel classification of the tissues.

Additionally, there are occasionally co-located bright areas in both the T2W image and the DWIs, e.g., at points where nerves branch off from the spinal cord. Because such nerve tissue is similar to the cord tissue, it cannot be classified as background by voxel intensity alone. Therefore, accurate segmentation requires localization constraints in addition to the pure voxel classification of the CEM algorithm.

Finally, in assessing chronic posttraumatic changes in injured cords, we observe two major effects of injury—atrophy and hemorrhage. As the cord atrophies, the size of the spinal cord decreases, making the shape of the injured cord, as well as white and gray matter areas, unpredictable. The additional space within the spinal cavity is filled by CSF. In addition, the tissue contrast of the surviving white and gray matter in injured cords is much less clear than in the control cords.

In injured cords, the presence of hemorrhage appears as dark areas within the bright spinal cord. Hemorrhage has a similar voxel intensity distribution to background, non-cord voxels in

the MR images because its relaxation and diffusion properties differ from those of the spinal cord, for example, the T2* value is shorter in hemorrhage. It is useful to automatically produce two spinal cord segmentations, one including hemorrhage and one excluding it. This allows quantitative analysis of the size of the hemorrhage relative to the whole cord size, which is useful in SCI evaluation.

By using a multi-step approach to the automatic spine segmentation problem, we are able to take advantage of the unique tissue differentiation abilities of the MR image types. Our multidimensional approach allows us to generate segmentations that are more robust to noise in individual images than existing approaches that rely on a single MR image.

METHODS

Theory: Classification Using the CEM Algorithm

We propose using intensity levels for automatic segmentation rather than the existing contour-fitting approaches, to allow more accurate segmentation of injured spinal cords. We assume the intensity values in the T2W image come from a sum of two distinct Gaussian distributions—that of the background and that of the cord and CSF. Similarly, we assume the intensity values in each DWI come from a sum of three distinct Gaussian distributions—those of the background (all non-spinal cord tissues, including bone), the white matter, and the gray matter (Fig. 1). Although we make the assumption that the above distributions are Gaussian for the purposes of our classification algorithm, these distributions are known to be Rician. We also compute results under a Rician assumption for purposes of comparison.

To find an optimal, unbiased separation of voxels into these classes, we employ the CEM algorithm in a series of stages. Each stage of our algorithm separates the voxels into one of two classes, e.g., initially, the pixels are classified as either background or spinal cord, and then in a later stage, the cord voxels are classified as either spared white or gray/injured white matter.

We observe the vector of image intensity values $\mathbf{x}_i \in \mathbb{R}^D$ from voxels $i = 1, \dots, n$ in D input images. These intensity values come from K possible tissue classifications $k = 1, \dots, K$, where K is known. Let y_i denote the classification corresponding to $\mathbf{x}_i$ ($i = 1, \dots, n$), taking a value from 1 to K . The intensity distribution, $\mathbf{f}_k(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$, for each class k is assumed to be a multi-dimensional Gaussian with mean $\boldsymbol{\mu}_k$ and covariance matrix Σ_k both unknown. Each classification has prior probability π_k , which is also assumed to be unknown. We assume noise independence within an individual voxel across the T2W image and the six DWIs, and across the three DTI maps used; this reduces Σ_k to a diagonal matrix. Therefore, the estimated distribution parameters are reduced to the priors π_k , means $\boldsymbol{\mu}_k$, and variances σ_k of multi-dimensional Gaussian intensity distributions.

The CEM algorithm (21) is a variation of the EM algorithm (22), with a classification step (C-step) added between the expectation step (E-step) and the maximization step (M-step). It finds a classification maximum likelihood estimate of y_i , maximizing the classification likelihood, C_1 :

$$C_1 = \sum_{k=1}^K \sum_{i|y_i=k} \log \mathbf{f}(\mathbf{x}_i|\boldsymbol{\mu}_k, \Sigma_k). \quad [1]$$

Starting from initial cluster parameter values, the CEM algorithm iteratively converges to a maximum *a posteriori* estimate of y_i , which is known to be in general initialization-sensitive. For example, if the initial cluster parameter values are far from the optimal parameters, the

estimate for y_i may be locally, not globally, optimal. To avoid such 'bad' initialization, for each CEM initialization in our method, we apply the commonly used K-means algorithm with random sample seeding to find a preliminary clustering of the data. For our application, 30 random K-means seedings produced exactly identical final spine segmentations for each MRI data set.

Given our assumptions, the steps of the CEM algorithm for iteration m are as follows (21), assuming Gaussian distributions:

- E-step: Compute the current posterior probabilities $p_{i,k}$ for all data points $i = 1, \dots, n$ and all clusters $k = 1, \dots, K$

$$p_{i,k} = \frac{\pi_k \mathbf{f}(\mathbf{x}_i | \mu_k^{(m)}, \sum_k^{(m)})}{\sum_{k'=1}^K \pi_{k'} \mathbf{f}(\mathbf{x}_i | \mu_{k'}^{(m)}, \sum_{k'}^{(m)})}. \quad [2]$$

- C-step: Assign each data point $\mathbf{x}_i$ to the cluster with the largest posterior probability.

$$y_i^{(m+1)} = \arg \max_k p_{i,k}. \quad [3]$$

- M-step: Compute the maximum likelihood estimates for the parameters, $\pi_k^{(m+1)}$, $\mu_k^{(m+1)}$, and $\sum_k^{(m+1)}$:

$$\pi_k^{(m+1)} = \frac{\sum_{i=1}^n \widehat{p}_{i,k}}{n}, \quad [4]$$

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^n \widehat{p}_{i,k} \mathbf{x}_i}{\sum_{i=1}^n \widehat{p}_{i,k}}, \quad [5]$$

$$\sigma_{k,d}^{(m+1)} = \frac{\sum_{i=1}^n \widehat{p}_{i,k} (x_{i,d} - \mu_{k,d}^{(m+1)})^2}{\sum_{i=1}^n \widehat{p}_{i,k}}, \quad [6]$$

$$\sum_k^{(m+1)} = \begin{bmatrix} \sigma_{k,1}^{(m+1)} & 0 & \dots & 0 \\ 0 & \sigma_{k,2}^{(m+1)} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{k,D}^{(m+1)} \end{bmatrix} \quad [7]$$

$$\text{where } \widehat{p}_{i,k} = \begin{cases} 1 & \text{if } y_i^{(m+1)} = k \\ 0 & \text{else} \end{cases}.$$

These steps repeat until the algorithm converges, i.e., when no voxel changes classification from one iteration to the next.

Method for Automatic Segmentation

The MRI output in this study consists of seven images of a specific slice of the spinal column—one T2W image and six DWIs from independent diffusion gradient directions. Figures 2 and 3 show examples of these images for a control animal and an injured animal, respectively. We choose to use the DWIs for all cord/background tissue segmentation because of their useful properties, as described in the Background section.

Find initial spinal cord segmentation—In the T2W image of each spinal slice (Figs. 2(a) and 3(a)), we use the CEM algorithm to classify voxels into two sets, background and spinal cavity. We refer to this set of spinal cavity voxels as S_{T2W} . This step provides a loose constraint on the spinal cord's location, since the T2W image does not generally differentiate between cerebrospinal fluid and spinal cord, so the bright region will be larger than the cord but should completely encapsulate it.

Next, we apply the multi-dimensional CEM algorithm to the set of DWIs (Figs. 2(b)–(g) and 3(b)–(g)) to jointly classify voxels as background or as spinal cord according to all six DWIs. We refer to this set of spinal cavity voxels as S_{DWI} . This step more accurately separates the spinal cord from the background than the T2W image step. By accepting voxels as spinal cord, S_0 , only when they are so classified according to both the T2W image and the DWIs, i.e.,

$$S_0 = S_{T2W} \cap S_{DWI}, \quad [8]$$

most outlying bright spots are correctly classified as background, without loss of correctly classified cord voxels.

Constrain by location—We automatically classify outlying voxels as background by adding a localization constraint. In most cases, the previous two steps will cleanly differentiate between background and spinal cord. This step accounts for potential co-located bright areas in both image types, such as nerve tissues, that are not actually within the cord.

To correctly classify bright, non-cord areas as background, we reclassify all small bright areas in S_0 as background, a process demonstrated in Fig. 4. First, we find the set of boundary curves B_0 that separate the voxels currently classified as spinal cord from the rest of the image (Fig. 4(c,i,o)). Finding these boundary curves is very simple, unlike the usual contour methods, because we find the boundaries using edge detection on the binary image of voxel classifications, in which voxels have an intensity value of 1 if they are members of S_0 , and 0 intensity otherwise. Then, we refine the set of spinal cord voxels by including in the new set S_1 only the voxels in S_0 that lie within the largest such curve ($B_{0,max}$) (Fig. 4(d,j,p)).

Account for regions of hemorrhage—If a region of hemorrhage exists that is completely contained within non-hemorrhaging spinal cord tissue, as in Fig. 4(g–l), generating two segmentations, one excluding the hemorrhage (S_{excl}) and one including it (S_{incl}), does not require additional work. The segmentation excluding hemorrhage is simply

$$\mathbf{S}_{\text{excl}} = \mathbf{S}_1, \quad [9]$$

while in this case, the segmentation including hemorrhage is the set of all voxels enclosed by $\mathbf{B}_{0,\text{max}}$.

Unfortunately, regions of hemorrhage can lie along the cord boundary, as in Fig. 4(m-r), and in such cases, producing the second segmentation is more challenging. However, because the whole cord including hemorrhage is surrounded by CSF, it is possible to identify regions of hemorrhage by locating holes in the set $\mathbf{S}_{\text{T2W}}$, which contains spinal cord and all of its surrounding CSF, even if the hemorrhage is not encapsulated by surviving spinal cord. To identify these holes, first, we find the set of boundary curves $\mathbf{B}_{\text{T2W}}$ around the voxels in $\mathbf{S}_{\text{T2W}}$. Next, we exclude all outliers, retaining only the largest boundary curve $\mathbf{B}_{\text{T2W,max}}$. $\mathbf{B}_{\text{T2W,max}}$ is therefore comprised of spinal cord, hemorrhage, and CSF. We then classify as hemorrhage, $\mathbf{H}$, all points encapsulated by $\mathbf{B}_{\text{T2W,max}}$ that are not members of the set $\mathbf{S}_{\text{T2W}}$, thus removing the voxels containing CSF and surviving spinal cord.

By combining the two sets, that of hemorrhage $\mathbf{H}$ and that of cord excluding hemorrhage $\mathbf{S}_{\text{excl}}$, we are able to find a segmentation for the entire spinal cord,

$$\mathbf{S}_{\text{incl}} = \mathbf{S}_{\text{excl}} \cup \mathbf{H}. \quad [10]$$

The final boundary curve around the entire cord is then simply the curve separating the voxels in $\mathbf{S}_{\text{incl}}$ from the rest of the image. This step is applied to all cords. In the case of healthy spine, this step will not change the segmentation. In the case where there are regions of hemorrhage, this step allows us to identify the actual boundary of the spinal cord along with precisely where hemorrhage is, and therefore allows us to calculate, for instance, the amount of hemorrhage relative to the area of the total cord.

Apply a bounding shape—The final step in segmenting the spinal cord is the application of a bounding shape. We fit a modified superellipse to the set of voxels in $\mathbf{S}_{\text{incl}}$ to remove possible protrusions from the cord, such as nerve tissues branching away from the spine. This modified superellipse is described as follows:

$$\begin{aligned} \left| \frac{(x-x_0)}{a} \right|^2 + \left| \frac{(y-y_0)}{b} \right|^2 &\leq 1 \quad \text{if } y > 0, \\ \left| \frac{(x-x_0)}{a} \right|^3 + \left| \frac{(y-y_0)}{b} \right|^3 &\leq 1 \quad \text{if } y < 0, \end{aligned} \quad [11]$$

where (x_0, y_0) is the center of the superellipse, a is the horizontal semi-diameter, and b is the vertical semi-diameter. This pair of equations generates a shape that can somewhat tightly encapsulate the spinal cord over the range of shapes it takes in varying states of injury and at varying points along its length. No tight bounding shape can be chosen because of the wide array of injured spinal cord shapes.

Segment spared white matter—The final step is the automatic classification of the spared white matter within the segmented cord tissue (Fig. 5). We use the multi-dimensional CEM algorithm to create an initial spared white matter segmentation. Next, we apply a localization constraint by first finding boundary curves around each disjoint group of voxels that the CEM step classified as spared white matter, and then reclassifying the smallest such groups into the gray/injured white matter classification. For this step, only voxel groups with a boundary length of at least 10 pixels retain their classification as spared white matter.

Rodent Spinal Cord MRI Experiments^{1,2}

We evaluated the segmentation performance of our algorithm using *in vivo* MR images of uninjured and injured rodent spinal cords. Ten twelve-week-old female C57BL/6 mice weighing 18 ~ 20 g (Harlan, Indianapolis, IN) were anesthetized with an isoflurane and oxygen mixture (7% for knock out and 1.5% for maintenance). After dorsal laminectomy at the T8 and T9 vertebral levels, the mice received contusive spinal cord injury utilizing a modified Ohio State University device (23). The injury group underwent contusion injury at 0.2 m/s with 0.6 mm impact displacement. After impact, the site was closed in layers with 4–0 silk sutures. Enrofloxacin (2.5 mg/kg) and lactated ringers (1 ml) were administered subcutaneously. The control group received sham operations including laminectomy and zero point contact of impactor tip on the surface of the spinal cord to establish a reference position, but no impact.

Animal preparation for *in vivo* diffusion tensor imaging—All mice were delivered to the MR facility and anesthetized with an isoflurane and oxygen mixture (1.0 – 1.5% for maintenance) at 14 days post injury. The body temperatures were maintained at 37°C with a circulating warm water pad. An inductively-coupled surface coil covering T8 - T10 vertebral segments (15 mm × 8 mm) was used as the RF receiver. A 9-cm i.d. Helmholtz coil was employed as the RF transmitter. The entire preparation was placed in an Oxford Instruments 200/330 magnet (4.7 T, 33-cm clear bore) equipped with a 15-cm inner diameter, actively shielded Oxford gradient coil (18 G/cm, 200-μs rise time). The magnet, gradient coil, and Techon gradient power supply were interfaced with a Varian UNITY-INOVA console (PaloAlto, CA) controlled by a Sun Microsystems Blade 1500 workstation.

***In vivo* diffusion tensor imaging**—A conventional spin-echo imaging sequence was modified by adding Stejskal-Tanner diffusion weighting gradients (24). The repetition time (TR, ~1.2 s) was varied according to the period of the respiratory cycle (~270 ms). The spin echo time (TE) = 38 ms, time between application of gradient pulses (Δ) = 20 ms, and diffusion gradient on time (δ) = 7 ms, were fixed throughout the experiment. For each animal, three consecutive slices were collected to cover the epicenter of the contusion-injured cord, with a total scan time of 2 hours. Diffusion weighted images were obtained with diffusion sensitizing gradients applied in six orientations, $(G_x, G_y, G_z) = (1, 1, 0), (1, 0, 1), (0, 1, 1), (-1, 1, 0), (0, -1, 1),$ and $(1, 0, -1)$, using diffusion sensitizing factors (b values) of 1.0 ms/μm². One image (the b=0 or T2W image) was collected without diffusion sensitizing gradient to serve as a reference. Six scans were averaged per k-space line. The field of view was 10×10 mm² with 1.0 mm slice thickness and the image data matrix for each slice was 128 (phase encoding)× 256 (read out) (zero filled to 256×256). Of the total image area for control animals, on average 1100 voxels contain spinal cord.

A weighted linear least-squares method was used to estimate diffusion tensors for each voxel from the diffusion-weighted images (25). The eigenvalue decomposition was then applied to each tensor, yielding a set of eigenvalues ($\lambda_1 \geq \lambda_2 \geq \lambda_3$) and eigenvectors for each voxel. Maps of diffusion indices including relative anisotropy (RA) and axial and radial diffusivities ($\lambda_{||}$ and $\lambda_{\perp}$) were generated by applying the following equations for each voxel:

$$\lambda_{||} = \lambda_1, \quad [12]$$

¹All surgical interventions and both pre- and post-surgical care were performed in accordance with the Public Health Service Policy on Humane Care and Use of Laboratory Animals, *Guide for the Care and Use of Laboratory Animals* (Institute of Laboratory Animal Resources, National Research Council, 1996), and with the approval of the Washington University Institutional Care and Use Committee.

²Standard postoperative care including bladder expression was provided in accordance to the manual of Spinal Cord Injury Research Training Program held at Spinal Trauma and Repair Laboratories in Ohio State University (<http://medicine.osu.edu/sci/>).

$$\lambda_{\perp} = \frac{\lambda_2 + \lambda_3}{2}, \quad [13]$$

$$\langle D \rangle = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}, \quad [14]$$

$$RA = \frac{\sqrt{\sum_{i=1}^3 (\lambda_i - \langle D \rangle)^2}}{\sqrt{3} \langle D \rangle}. \quad [15]$$

Quantification of Segmentation Accuracy

We automatically segmented the spinal cords and the white matter from the MRI images of all ten mice at each of the spinal slice locations. Because the MRI data we use to test our algorithm has a very large field of view relative to the size of the spine (the area of the spine is roughly 2% of the entire field of view), the data is manually cropped prior to application of the algorithm to a rectangle around the spine. This cropping yields an image for segmentation in which the spine comprises roughly 20% of the field of view.

The spinal cord contains an area of white matter called the dorsal column, which is disconnected from the rest of the white matter and is not included in the white matter segmentations of our experts. Because of this, to compare our automatic segmentations to the manual segmentation, the dorsal column needs to be excluded from the automatic white matter segmentations. To eliminate this area, we automatically exclude the pixels in a small trapezoidal section of the spinal cord, defined relative to the superellipse of Eq. [11], with four vertices at $(x_0 - a/4, y_0 + b/2)$, $(x_0 + a/4, y_0 + b/2)$, $(x_0 + a/2, y_0 + b)$, and $(x_0 - a/2, y_0 + b)$. We then apply a localization constraint to remove any small sections of dorsal white matter that may not have been fully contained in the trapezoid.

Our algorithm ran in an average of 0.709 seconds (using DWIs for all segmentation) per spinal slice using MATLAB (Mathworks, Natick, MA, USA) on an Intel Core 2 Quad CPU, 2.4 GHz PC. We compared the performance of the algorithm when using only DTI maps, using only DWIs, or using both for the spared white matter segmentation. The algorithm ran in an average of 0.448 seconds when using DWIs for the cord/hemorrhage tissue segmentations and DTI maps for the spared white matter segmentation, and 0.515 seconds when using DWIs for the cord/hemorrhage tissue segmentations and all data (DWIs and DTI maps) for the spared white matter segmentation.

For verification of our algorithm's performance, three types of manual segmentations were created. The entire spinal cord (gray and white matter), the cord excluding hemorrhage, and the white matter were manually segmented by five experts for both control and injured cords utilizing diffusion weighted images and calculated diffusion maps. This was done for one spinal slice from the MRI images of each animal, at the location of the sham operation for control animals and at the epicenter of surgically-induced SCI for injury group animals. From the five expert segmentations for each tissue type, we are able to find a measure of the variation in manual segmentations across experts, to which we can compare the variation between manual and automatic segmentations.

Because our study uses *in vivo* imaging, there is no ground truth available, such as histology. Lacking a ground truth by which to calculate percent error, we instead evaluate our algorithm using the overlap of our automatic segmentation results with manual segmentations of the same data. This overlap is calculated using the binary classification images, i.e., images that take a value of 1 only the in region of cord, of cord excluding hemorrhage, or of white matter, and a value of 0 outside that region. We calculate overlap as

$$\text{Overlap}(\mathbf{A}, \mathbf{B}) = \frac{\mathbf{A} \cap \mathbf{B}}{\mathbf{A} \cup \mathbf{B}}, \quad [16]$$

where $\mathbf{A}$ and $\mathbf{B}$ are the two binary classification images to be compared.

RESULTS

Table 1 shows the average overlap between each pair of independently drawn manual segmentations. This establishes a baseline for variation in manual segmentations to which the variation between the manual and automatic segmentations (Table 2) can be compared. It is clear from the poor correspondence of the injured cord segmentations of hemorrhage and particularly white matter that manual segmentation cannot be treated as the ground truth, as it is in some works.

Table 2 shows the total average overlap of our segmentation results with the manual segmentations. For segmentation of the tissue regions of control spinal cords, the performance of our algorithm is on average equivalent to that of the expert manual segmentations. Similarly, for all segmentations of the injured cords—the whole cords, the spinal cord excluding hemorrhage, and the white matter—our automatic segmentations are on average at least as good as the expert manual segmentations. The extreme lack of consistency between expert segmentations of injured white matter makes it impossible for our algorithm to have a high overlap with all experts, but considering the overlap between expert segmentations, the relative performance is strong.

DISCUSSION

In comparison of our algorithm to the individual manual segmentations, we note that the maximum overlap between our algorithm and any manual segmentation is always greater than the minimum overlap between any two manual segmentations, for all tissue-type segmentations. That is to say, our algorithm performs at least as well as the worst human expert, given that we do not know which is the most accurate segmentation in the group. It is important to note that, as we do not have a ground truth segmentation, it is not possible to say if a particular manual segmentation is superior or inferior to our automatic segmentation. We can only say with certainty that our algorithm has higher consistency with manual segmentations than the consistency between manual segmentations for injured spinal cords.

We found that, although the properties of DWIs are very useful for segmentation of spinal cord and hemorrhage, our results for white matter segmentation using the DTI maps were both qualitatively and quantitatively more accurate for moderately to severely injured cords and also more robust with respect to image quality. In fact, the relative noise level in the DWIs as compared to the DTI maps is such that consideration of all the data, both DWIs and DTI maps, reduces the segmentation performance.

As a point of comparison, in (15,16) and in (17), intensity-based, fuzzy classification methods are used for spine and white/gray matter segmentation. However, the method in (17) relies on the fact that the spinal cord is excised for finding the spinal mask, and so cannot be applied to

in vivo images. Additionally, they do not measure the agreement between manual segmentations and their segmentation, but compare only the intensity statistics of the two regions, so the segmentation accuracy of their method is not validated even for excised spines. The authors of (15) attempt to validate their method on five uninjured spinal cords, taking a fixed template as the ground truth. The percent overlap between their automatic results and the fixed template ranged from 84.4% to 89.2% for the spinal cord segmentation, and their percent correct classifications for white and gray matter were 67.1% and 86.5%, respectively. The authors then in (15) validate only the manual template-alignment step for estimation of the anisotropy statistics of white matter, gray matter, and CSF from test images, not the final segmentation performance in the case of injured cords.

As a final note, if, instead of assuming Gaussian distributions, we treat the intensity distributions as Rician, our algorithm is slower and performs equivalently with respect to overlap with the manual segmentations. The algorithm is slower because there is no analytical expression for the maximum likelihood estimate for Rician parameters given samples from the distribution, so an iterative estimation method is necessary. This iterative method must be applied at each iteration of the CEM algorithm, which causes slowing of the total algorithm run-time. The average runtimes are 0.644 seconds for using DWIs for all segmentations, 0.420 seconds using DWIs for the cord/hemorrhage tissue segmentations and DTI maps for the spared white matter segmentation, and 0.488 seconds using DWIs for the cord/hemorrhage tissue segmentations and all data (DWIs and DTI maps) for the spared white matter segmentation. Table 3 shows the total average overlap of our segmentation results with the manual segmentations.

These results are roughly equivalent to the results when the Gaussian assumption is used because the distributions of the white and gray matter are nearly Gaussian. In addition, although a Gaussian is not a good fit for the background distribution, it can be seen in Fig. 1 that the decision threshold for spinal cord and background will not be affected significantly by assuming a Gaussian rather than a Rician distribution.

CONCLUSION

We have proposed a new multi-step, CEM-based approach to spinal cord and white matter segmentation from *in vivo* MR images; and we have validated that its performance is on par with that of expert manual segmentation. We have demonstrated that our algorithm, unlike previous approaches, remains reliable for spinal cord segmentation in the presence of moderate and severe cord injury, not just extremely mild injury. In addition, we have demonstrated that our algorithm is as reliable as the average human expert for hemorrhage and white matter segmentation for injured rodent spinal cords. Future work includes incorporation of a more detailed physical model for the spinal tissues, as well as comparison of manual and automatic MRI segmentation results to spinal histology results. We expect that our algorithm will yield closer agreement with histology than an average manual expert segmentation does. Future work also includes adaptation of the algorithm to MR images of human spinal cords, which have lower resolution than MR images of rodent cords.

Acknowledgments

This work was supported in part by the NIH under Grant NS047592, in part by a National Science Foundation (NSF) Graduate Research Fellowship, and also in part by a Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

1. Scheff SW, Rabchevsky AG, Fugaccia I, Main JA, Lump JJ. Experimental modeling of spinal cord injury: Characterization of a force-defined injury device. *J Neurotrauma* 2003;20(2):179–193. [PubMed: 12675971]
2. Ma M, Basso DM, Walters P, Stokes BT, Jakeman LB. Behavioral and histological outcomes following graded spinal cord contusion injury in the C57BL/6 mouse. *Exp Neurol* 2001;169(2):239–254. [PubMed: 11358439]
3. Basso DM, Beattie MS, Bresnahan JC. Descending systems contributing to locomotor recovery after mild or moderate spinal cord injury in rats: Experimental evidence and a review of literature. *Restor Neurol and Neurosci* 2002;20(5):189–218.
4. Basso DM. Neuroanatomical substrates of functional recovery after experimental spinal cord injury: Implications of basic science research for human spinal cord injury. *Phys Ther* 2000;80(8):808–817. [PubMed: 10911417]
5. McEwen ML, Springer JE. Quantification of locomotor recovery following spinal cord contusion in adult rats. *J Neurotrauma* 2006;23(11):1632–1653. [PubMed: 17115910]
6. Magnuson DS, Lovett R, Coffee C, Gray R, Han Y, Zhang YP, Burke DA. Functional consequences of lumbar spinal cord contusion injuries in the adult rat. *J Neurotrauma* 2005;22(5):529–543. [PubMed: 15892599]
7. Basso DM, Beattie MS, Bresnahan JC. Graded histological and locomotor outcomes after spinal cord contusion using the NYU weight-drop device versus transection. *Exp Neurol* 1996;139(2):244–256. [PubMed: 8654527]
8. Kim JH, Loy DN, Liang HF, Trinkaus K, Schmidt RE, Song SK. Noninvasive diffusion tensor imaging of evolving white matter pathology in a mouse model of acute spinal cord injury. *Magn Reson Med* 2007;58(2):253–260. [PubMed: 17654597]
9. Loy DN, Kim JH, Xie M, Schmidt RE, Trinkaus K, Song SK. Diffusion tensor imaging predicts hyperacute spinal cord injury severity. *J Neurotrauma* 2007;24(6):979–990. [PubMed: 17600514]
10. Cohen-Adad J, Benali H, Hoge RD, Rossignol S. In vivo DTI of the healthy and injured cat spinal cord at high spatial and angular resolution. *NeuroImage* 2008;40(2):685–697. [PubMed: 18201909]
11. Schwartz ED, Duda J, Shumsky JS, Cooper ET, Gee J. Spinal cord diffusion tensor imaging and fiber tracking can identify white matter tract disruption and glial scar orientation following lateral funiculotomy. *J Neurotrauma* 2005;22(12):1388–1398. [PubMed: 16379577]
12. Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *Int'l J of Computer Vision* 1987;1(4):259–268.
13. Osher S, Sethian JA. Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J of Comput Phys* 1988;79:12–49.
14. Deng X, Ramu J, Narayana PA. Spinal cord atrophy in injured rodents: High-resolution MRI. *Magn Reson Med* 2007;57(3):620–624. [PubMed: 17326166]
15. Ellingson BM, Ulmer JL, Schmit BD. Gray and white matter delineation in the human spinal cord using diffusion tensor imaging and fuzzy logic. *Acad Radiol* 2007;14(7):847–858. [PubMed: 17574135]
16. Ellingson BM, Ulmer JL, Schmit BD. Morphology and morphometry of human chronic spinal cord injury using diffusion tensor imaging and fuzzy logic. *Ann Biomed Eng* 2008;36(2):224–236. [PubMed: 18066663]
17. Younis AA, Ramirez N, Pattany PM, Burns RJ, Sharawy MI. Automated segmentation of spinal diffusion tensor MR imaging. *SoutheastCon, 2005 IEEE Proc* 2005:187–192.
18. Budde MD, Kim JH, Liang HF, Russell JH, Cross AH, Song SK. Axonal injury detected by *in vivo* diffusion tensor imaging correlates with neurological disability in a mouse model of multiple sclerosis. *NMR Biomed* 2007;21(6):589–597. [PubMed: 18041806]
19. DeBoy CA, Zhang J, Dike S, Shats I, Jones M, Reich DS, Mori S, Nguyen T, Rothstein B, Miller RH, Griffin JT, Kerr DA, Calabresi PA. High resolution diffusion tensor imaging of axonal damage in focal inflammatory and demyelinating lesions in rat spinal cord. *Brain* 2007;130(8):2199–2210. [PubMed: 17557778]

20. Gullapalli J, Krejza J, Schwartz ED. *In vivo* DTI evaluation of white matter tracts in rat spinal cord. *J Magn Reson Imaging* 2006;24(1):231–234. [PubMed: 16767707]
21. Celeux G, Govaert G. A classification EM algorithm for clustering and two stochastic versions. *Comput Stat Data Anal* 1992;14(3):315–332.
22. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statist Soc Ser B* 1977;39(1):1–38.
23. Jakeman LB, Guan Z, Wei P, Ponnappan R, Dzwonczyk R, Popovich PG, Stokes B. Traumatic spinal cord injury produced by controlled contusion in mouse. *J Neurotrauma* 2000;17(4):299–319. [PubMed: 10776914]
24. Stejskal EO, Tanner JE. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J Chem Phys* 1965;42(1):288–292.
25. Koay CG, Chang LC, Carew JD, Pierpaoli C, Basser PJ. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J Magn Reson* 2006;182:115–125. [PubMed: 16828568]

Figure 1. Histogram of a representative cropped and normalized DWI for a control spine (see image in Fig. 2(b)). The distributions of the automatically segmented tissue types (background, white matter, and gray matter) are superimposed assuming data comes from (a) Gaussian and (b) Rician distributions.

Figure 2. Representative example of cropped and normalized MRI data for a control spine slice: (a) T2W image; (b)–(g) diffusion weighted images (DWIs); (h)–(j) DTI maps (relative anisotropy (RA), axial diffusivity ($\lambda_{\parallel}$), and radial diffusivity ($\lambda_{\perp}$), respectively); (k)–(o) manual segmentations of spinal cord (red/light gray curves) and spared white matter (blue/dark gray curves), superimposed over the DWI in (b), for the five separate manual segmentations.

Figure 3.

Representative example of cropped and normalized MRI data for an injured spine slice: (a) T2W image; (b)–(g) diffusion weighted images (DWIs); (h)–(j) DTI maps (relative anisotropy (RA), axial diffusivity ($\lambda_{\parallel}$), and radial diffusivity ($\lambda_{\perp}$), respectively); (k)–(o) manual segmentations of spinal cord and hemorrhage (red/light gray curves) and spared white matter (blue/dark gray curves), superimposed over the DWI in (b), for the five separate manual segmentations.

Figure 4.

Representative examples of spine and hemorrhage segmentation for a control spine slice (a–f) and for injured spine slices with hemorrhage within (g–l) and along (m–r) the spinal cord boundary: (a,g,m) sample cropped, normalized DWI; (b,h,n) after application of the CEM algorithm to the DWIs; (c,i,o) with boundary curves outlined; (d,j,p) after removal of small boundary curves; (e,k,q) after inclusion of holes from T2W image; (f,l,r) after application of composite superellipse bounding shape, with hemorrhage also outlined.

Figure 5.

Representative examples of spared white matter segmentation for control (a–d) and injured (e–h) spine slices: (a,e) sample cropped, normalized relative anisotropy (RA) map after spinal cord segmentation; (b,f) after application of the CEM algorithm to the three DTI maps; (c,g) with boundary curves outlined; (d,h) after removal of small boundary curves.

Table 1

Average Percent Overlap between the 5 Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.47 $\pm$ 1.78	90.43 $\pm$ 2.45
Spinal Cord Excluding Hemorrhage	91.47 $\pm$ 1.78	68.10 $\pm$ 12.53
Spared White Matter	80.21 $\pm$ 3.28	51.41 $\pm$ 9.14

Values are given as the mean $\pm$ standard deviation

Table 2

Average Percent Overlap between the Automatic Segmentations and the 5 Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.32 $\pm$ 2.39	90.20 $\pm$ 2.13
Spinal Cord Excluding Hemorrhage	91.28 $\pm$ 2.27	72.18 $\pm$ 9.39
Spared White Matter Using DWIs	73.24 $\pm$ 6.80	11.87 $\pm$ 8.32
Spared White Matter Using DTI Maps	79.65 $\pm$ 5.48	54.82 $\pm$ 9.88
Spared White Matter Using All Data	77.76 $\pm$ 4.93	37.26 $\pm$ 16.13

Values are given as the mean $\pm$ standard deviation

Table 3

Average Overlap between the Automatic Segmentations Assuming Rician Distributions and the 5 Manual MRI Tissue Segmentations

	Control Group	Injury Group
Entire Spinal Cord	91.51 $\pm$ 2.33	90.22 $\pm$ 2.05
Spinal Cord Excluding Hemorrhage	91.48 $\pm$ 2.29	72.11 $\pm$ 9.41
Spared White Matter Using DWIs	73.47 $\pm$ 6.96	11.85 $\pm$ 8.32
Spared White Matter Using DTI Maps	79.42 $\pm$ 5.13	54.99 $\pm$ 9.63
Spared White Matter Using All Data	78.15 $\pm$ 4.96	38.11 $\pm$ 15.12

Values are given as the mean $\pm$ standard deviation