AN AUTOMATIC SEGMENTATION METHOD OF THE SPINAL CANAL FROM CLINICAL MR IMAGES BASED ON AN ATTENTION MODEL AND AN ACTIVE CONTOUR MODEL

Jaehan Koh, Peter D. Scott, Vipin Chaudhary*

University at Buffalo, SUNY
Department of Computer Science and Engineering
Buffalo, NY 14260, USA
{jkoh,peter,vipin}@buffalo.edu

Gurmeet Dhillon, MD

Proscan Imaging of Buffalo
Williamsville, NY 14221, USA
gdhillon@proscan.com

ABSTRACT

The spinal cord is a vital organ that serves as the only communication link between the brain and the various parts of the body. It is vulnerable to traumatic spinal cord injury and various diseases such as tumors, infections, inflammatory diseases and degenerative diseases. The exact segmentation and localization of the spinal cord are essential to effective clinical management of such conditions. In recent years, due to the advances in imaging technology, the structure of internal organs and tissues can be captured accurately, and various abnormalities are diagnosed based on scanned images. In this paper, we present an unsupervised segmentation method that automatically extracts the spinal canal in the sagittal plane of magnetic resonance (MR) images. This segmentation method based on a novel saliency-driven attention model and a standard active contour model requires no human intervention and no training. Experiments based on 60 patients' data show that this procedure performs segmentation robustly, achieving the Dice's similarity index of 0.71 between the segmentation by our model and reference segmentation, as compared to the Dice's similarity index of 0.90 between two observers.

Index Terms— spinal canal, segmentation, saliency map, active contour, level set

1. INTRODUCTION

The spinal cord is a vital structure of the central nervous system that relays messages back and forth between the brain and the body. Thus, damage to the spinal cord can result in loss of sensation and loss of the ability to voluntarily control muscles, sometimes leading to paralysis. It is also of great clinical importance since it can develop traumatic spinal cord injury and various diseases such as tumors, infections, inflammatory diseases and degenerative diseases. As a result, the exact segmentation and localization of the spinal cord are considered as a prerequisite to understanding the conditions of it. Moreover, the quantification based on the segmenta-

tion of the spinal cord helps analyze and diagnose many cord-related diseases. Recently, due to the advances in imaging technology that can capture the structure of internal organs and tissues with high precision, a lot of techniques have been proposed and clinically used in the diagnosis of spine-related diseases based on computed tomography (CT) and magnetic resonance imaging (MRI). In the sagittal view of the lumbar MRI, the spinal cord terminates around the level $L2$. Below the spinal cord, the thecal sac, the protective insulating membrane that wraps around the spinal cord and cauda equina, is housed. These two structures are relatively bright compared to vertebrae and discs in MR images and constitute the central spinal canal. Clinically, T2-weighted sagittal images are more frequently examined in diagnosing conditions related to the spinal cord since they give better contrast than T1-weighted ones, allowing to capture boundaries of soft tissues and bones more accurately as in Figure 1.

The spinal cord segmentation is a hard problem since its morphology is usually diverse in different regions: cervical, thoracic, lumbar and sacral. Inhomogeneity in intensity due to varying amount of white matter and gray matter makes the segmentation task
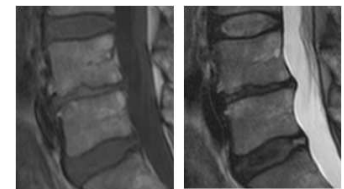

Fig. 1. Examples of T1-weighted (left) and T2-weighted (right) sagittal images of the same patient

much harder. In spite of those challenges, semi-automatic or fully-automatic segmentation is known to be less time-consuming, more accurate and less sensitive to intra- and inter-observer variability than manual segmentation [1], [2], [3]. Accordingly there have been many attempts made in the semi-automatic segmentation of the spinal cord.

1.1. RELATED WORK

Nieniewski and Seneels [4] proposed a morphological method for segmenting the spinal cord from MR images. The method

was tested on images from cross-sections of the spinal cord. Archip *et al.* [5] proposed a top-down knowledge-based technique that identified the spinal cord in the CT images. The technique was flexible enough to handle a large amount of inter-patient variation but required a high computational cost. Schmit and Cole [6] employed three-dimensional seeded region growing technique for the segmentation of the spinal cord. However, the authors did not give any validation analysis on the segmented results. Uitert *et al.* [2] presented a semi-automatic process for segmenting the spinal cord from MRI. However, they did not provide any validation on the computed quantification. Nyúl *et al.* [3] proposed an semi-automatic segmentation method that extracted the spinal cord/canal from CT images. With an initial seed point set by a human, this method was validated by comparing with gold standards, *i.e.*, manual segmentation. McIntosh and Hamarneh [7] proposed a semi-automatic spinal cord segmentation and analysis technique from MRI with high-level control mechanisms. The model was quantitatively validated and compared against manual, and the Itk-SNAP method. However, it had very high computational cost and user intervention was required. Horsfield *et al.* [8] proposed a semi-automatic method for the segmentation of the spinal cord from MRI. The method utilizing active surface model was applied to assessing multiple sclerosis. Validation was done by evaluating the intra-observer reproducibilities.

The above-mentioned methods are either semi-automatic or not validated in terms of performance metrics. In this paper, we present a novel fully-automatic segmentation method of the spinal cord in T2-weighted sagittal images of lumbar spine that combines an attention model with the standard active contour model. Since T2-weighted sagittal images are frequently evaluated in the preliminary stage of diagnosing lumbar spinal disorders in the clinical settings, we tested our segmentation method on images in that protocol. As [3] mentioned that the spinal cord can be only detected with some prior knowledge, we use the saliency map as prior knowledge and compare the segmentation results with reference segmentation in terms of a similarity index, Dice's coefficient.

1.2. Active Contour Model

Deformable image segmentation has been widely investigated since it allows for variability of object structures indexed by time or by cross-section location. It may also serve as a preprocessing step to image analysis and understanding. Among a variety of deformable contour models, active contours (or snakes) are widely adopted and applied to many problems such as edge detection, object classification, and motion tracking since they can reduce human intervention in the task of segmentation [9]. A snake approximates the boundary of an object through the evolution of a curve by minimizing a selected energy functional. However, it has some disadvantages. One of them is that the initial contour must be specified before the curve evolution can begin, and the resulting bound-

ary is dependent on this selection. In addition, the evolving snake contour is sensitive to noise.

The first active contour model was proposed by Kass *et al.* [9] and is defined as $\mathcal{C}(s) := (x(s), y(s))^\top : \mathbb{R}^2 \to \Omega$ where $x(s), y(s)$ are x, y-coordinates along the contour and $s \in [0, 1]$. Plus Ω represent the entire domain of an image $I(x, y)$. Here the energy functional $F(\mathcal{C})$ to be minimized can be defined on the contour $\mathcal{C}$ as follows: $F(\mathcal{C}) := \int_0^1 \left(E_{int} + E_{ext} \right) ds$, where E_{int} is the internal spline energy function term, and E_{ext} is external energy function term, respectively. Given the set of parameters, a solution to the snake problem finds the contour $\mathcal{C}$ that minimizes the total energy functional. The internal energy function determines the smoothness of the contour and is usually defined as $E_{int} := \alpha \left| \frac{d\mathcal{C}}{ds} \right|^2 + \beta \left| \frac{d^2\mathcal{C}}{ds^2} \right|^2$, where α and β specify the elasticity and stiffness of the contour, respectively. The external energy term E_{ext} determines the constraints of contour evolution depending on the image $I(x, y)$ and is defined as $E_{ext} := \int_0^1 E_{img}(\mathcal{C}(s)) ds$, where $E_{img}(x, y)$ denotes a scalar function defined on the image plane. Thus, the local minima of E_{img} attract the snakes to boundaries. Given the set of weights α, β, and λ, a solution to the snake problem finds the contour $\mathcal{C}$ that minimizes the total energy functional.

The rest of the paper is organized as follows. In Section 2, our attention model and the level set method are introduced. Experimental results and discussion are given in Section 3 and Section 4, respectively. Section 5 concludes this paper and states future work.

2. PROPOSED METHOD

In this section, we study our models for spinal cord segmentation. Our method is composed of the linking of two models, the bottom-up attention model and the active contour model, followed by a morphological process that removes noise-like small blobs. The attention model is based on linear filtering and the superposition of filtered components. The active contour model is based on level sets. In the postprocessing stage, a morphological analysis of blobs is conducted based on the size of each blob.

Since intensities of the pixels that constitute the spinal canal tend to be high compared with those of nearby bony structures, we attend to bright regions using the bottom-up approach. The attention model is then fused into the active contour model, providing an initial contour that is required for contour evolution.

2.1. Attention Model

The attention model is inspired by the structure and the functionality of the primary visual cortex. Many computational models of visual attention have been published in terms of bottom-up, top-down or hybrid approaches. The top-down

mechanism usually simulates our long-term cognitive process while the bottom-up mechanism attends to specific visual cues.

Since the simple cells in the primary visual cortex are sensitive to some localized signal orientation and frequencies, we model them using a Gabor filter bank based on the following formula as discussed in [10].

$$g(x, y, f_0, \theta_i, \sigma_x, \sigma_y) = \frac{1}{\sigma_x \sigma_y \pi} \cdot e^{-\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right)} \cdot e^{(2\pi \cdot i \cdot f_0 \cdot x')}, \tag{1}$$

where x, y denote the center coordinates of the filter, f_0 is the central frequency, θ_i determines the orientation of the filter, and $x' = x \cdot \cos\theta_i + y \cdot \sin\theta_i$ and $y' = -x \cdot \sin\theta_i + y \cdot \cos\theta_i$, respectively. In addition, σ_x and σ_y specify the width of the Gabor filter bank in terms of horizontal and vertical directions. We set both σ_x and σ_y to be 1 for computational simplicity. For the numerical studies to follow, we select 3 central frequency bands $f_0 = \{0.0625, 0.125, 0.25\}$ and 8 orientations $\theta_i = \{\frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}, \pi\}$. Gabor filter banks of different orientations with a central frequency of 0.25 are shown in Figure 2. The filtered responses $S_1, \ldots, S_8, M_1, \ldots, M_8, L_1, \ldots, L_8$ are computed by convolving the image with a Gabor filter of different frequencies as $g(\cdot) * \mathtt{I}$ where $g(\cdot)$ is a Gabor filter, $\mathtt{I}$ is an image, and $*$ is a convolution operator.

Then the interactions between simple cells are modeled by a weighted bipolar linear combination of simple cells. These process represents excitatory and inhibitory connections between simple cells. For our model, the filtered responses with $f_0 = 0.25$ are labeled L_1 through L_8 in the order of increasing angles. Similarly, the responses with $f_0 = 0.125$ are labeled from M_1 through M_8 and the filter outputs with $f_0 = 0.0625$ from S_1 through S_8. The interaction, R_j, is then formulated as follows:

$$R_j = M_{j+1} - \frac{1}{2}(M_j + M_{j+2}) + \frac{1}{2}(S_{j+1} + L_{j+1}). \tag{2}$$

The filtered responses become excitatory or inhibitory depending on the responses of neighboring filters. That is, if the neighboring responses are in the same direction to the current response, they are fortified. On the contrary, if the neighboring responses are orthogonal to the current response, they are suppressed as seen in Figure 3. Finally, the saliency map is generated by the normalized sum of the interaction maps: $S = \frac{\sum |R_j|}{\max\{\sum |R_j|\}}$ where $j \in \{1, 2, 3, 4, 5, 6\}$. In this way, the

salient locations are chosen and those regions are used as the contextual information for the segmentation model.
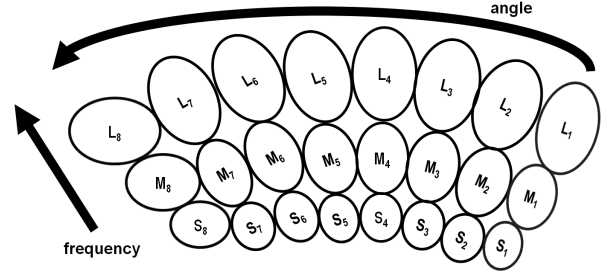

Fig. 3. Interactions between filtered responses

2.2. Level Set Method

Unlike classic snakes that only detect one boundary at one time, the level set methods allow us to find multiple boundaries simultaneously [11]. The Chan-Vese active contour algorithm solves a special case of the segmentation problem formulated by Mumford and Shah [12]. Specifically, the problem is described as follows: given an image u_0, find a partition Ω_i of Ω and an optimal piecewise smooth approximation u of u_0 such that u smoothly evolves within each Ω_i and across the boundaries of Ω_i.

To solve this problem, Mumford and Shah [12] proposed the following minimization problem:

$$\inf_{\mathcal{C}} \Big\{ F^{MS}(u, \mathcal{C}) = \int_\Omega (u - u_0)^2 dxdy + \mu \int_{\Omega \setminus \mathcal{C}} |\nabla u|^2 dxdy + \nu|\mathcal{C}| \Big\}, \tag{3}$$

where μ, ν are nonnegative parameters and $|\mathcal{C}|$ is the magnitude of boundary curve. If the segmented image u is restricted to piecewise constant function inside each connected component Ω_i, then the problem becomes the minimal partitioning problem and its functional is given by

$$F^{MS}(u, \mathcal{C}) = \sum_i \int_\Omega (u - c_i)^2 dxdy + \nu|\mathcal{C}|. \tag{4}$$

According to Chan and Vese [11], given the curve $\mathcal{C} = \partial\omega$ where $\omega \in \Omega$, an open subset and two unknown constants c_1 and c_2 as well as $\Omega_1 = \omega$ and $\Omega_2 = \Omega - \omega$, the minimum partitioning problem becomes the problem of minimizing the energy functional with respect to c_1, c_2, and $\mathcal{C}$ in accordance with:

$$F(c_1, c_2, \mathcal{C}) = \int_{\Omega_1 = \omega} (u_0(x, y) - c_1)^2 dxdy + \int_{\Omega_2 = \Omega - \omega} (u_0(x, y) - c_2)^2 dxdy + \nu|\mathcal{C}|. \tag{5}$$

In level set formulation, $\mathcal{C}$ becomes $\{(x, y)|\phi(x, y) = 0\}$. Thus, the energy functional becomes

$$F(c_1, c_2, \mathcal{C}) = \int_\Omega (u_0(x, y) - c_1)^2 H(\phi)dxdy + \int_\Omega (u_0(x, y) - c_2)^2 (1 - H(\phi)) dxdy + \nu \int_\Omega |\nabla H(\phi)|dxdy, \tag{6}$$

Fig. 2. An example of Gabor filters with a large frequency and different orientations

where $u_0(x,y)$ is the given image, c_1 and c_2 represent the mean intensities of the inside and outside of the segmented objects, respectively. Here, $H(z)$ is the unit step function. To get the minimum of F, we need to take the partial derivatives of F and set them to 0.

$$c_1(\phi) = \frac{\int_\Omega u_0(x,y)H(\phi(t,x,y))dxdy}{\int_\Omega H(\phi(t,x,y))dxdy}, \quad (7)$$

$$c_2(\phi) = \frac{\int_\Omega u_0(x,y)(1 - H(\phi(t,x,y)))dxdy}{\int_\Omega (1 - H(\phi(t,x,y)))dxdy}, \quad (8)$$

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\left[\nu \mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right) - (u_0 - c_1)^2 - (u_0 - c_2)^2\right], \quad (9)$$

where $\delta(z)$ is the Dirac function defined by $\delta(z) = \frac{d}{dz}H(z)$.

2.3. Postprocessing

After the segmentation is performed based on the active contour model, small blobs containing foreground pixels are removed in a way similar to the procedure used in [13]. To be specific, all holes within each blob are filled by morphological operations and small blobs are removed based on the number of pixels in each filled blob. Empirically, a small blob is determined by the following criterion, that is, one smaller than the half of the largest blob: $0.5 \times \max_{b \in B}|b|$ where b is an individual blob, B is the set of blobs and $|\cdot|$ the number of pixels in the blob.

3. EXPERIMENTS

Experiments were done based on linking the bottom-up attention model and the level-set active contour model. In the preprocessing step, the middle slice in a set of sagittal images per patient is chosen since it gives relatively clear snapshot of the lumbar spine. Then, the region of interest is extracted as in Figure 4(a). Empirically, pixels between 150 from left corner and 180 from the right corner of each slice contain the spinal cord across all patient data. Only these regions of interest (ROIs) are considered for subsequent processing. After a saliency map is generated by the attention model as in Figure 4(b), it is binarized with a threshold of 0.9 and the contour of the thresholded response forms the initial curve for the active contour model as in Figure 4(c). The maximum number of iterations for contour evolution is set to 1200. The threshold and the maximum iteration values are set empirically. The performance is compared with the manually marked boundaries by two trained medical specialists. Since the contextual information is obtained automatically throughout the process, no human intervention is required. Also our method does not require any training stages in segmentation.

3.1. Image Dataset

For MRI segmentation, 60 T2-weighted mid-sagittal MR images from 60 patients were used. Images were taken by a 3-T Philips scanner. The set consists of slices on the T2-weighted
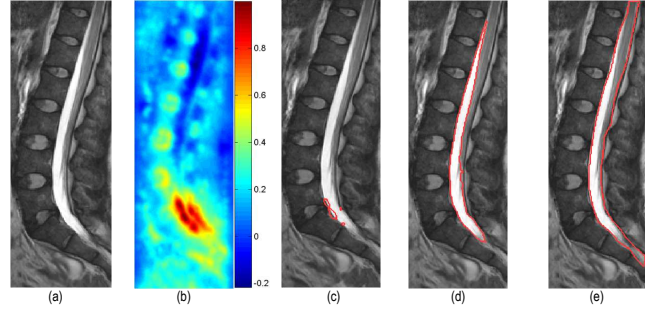

Fig. 4. (a) ROI image; (b) Its saliency map; (c) Initial contour; (d) Segmentation output; and (e) Reference segmentation

sagittal plane with flip angle $90°$, echo time of 100 *ms*, repetition time of 2622 *ms*, slice thickness of 4.5 *mm*, and the dimension of 512×512.

3.2. Automatic Segmentation Based on Our Model

To compare the segmentation performance between the "reference segmentation" marked by two specialists and an outline from our model, we employed the Dice's coefficient as a similarity index (SI). The Dice's coefficient $D(G,M)$ is defined as the ratio of twice the intersection over the sum of the two segmented results, the gold standard G and the our model-generated result M, *i.e.*, $D(G,M) = 2 \cdot \frac{n\{G \cap M\}}{n\{G\} + n\{M\}}$ where $n\{G\}$ is the number of elements in set G [14]. This measure is derived from a reliability measure known as the kappa (κ) statistic to evaluate the inter-observer agreement in regard to categorical data. According to this $D > 0.8$ indicates near-perfect agreement, $0.6 < D \leq 0.8$ represents substantial agreement, and $0.4 < D \leq 0.6$ moderate agreement [15].

4. DISCUSSION

As we expect, the saliency map correctly extracted meaningful information about the spinal canal in a given image and provided that to the segmentation model as shown in Figures 4 and 5. Since the intensities of the spinal canal in the sagittal scans of MRI are relatively brighter than the vertebra or disc regions, the saliency map successfully finds that area.

We classified the initial contour into three classes depending on how much of the initial contour is contained in the spinal cord as in Figure 5. If more than half of the contour is contained in the spinal cord, it is classified as *fully contained*. If the initial contour is partly contained in the spinal cord, we called it *semi-contained*. If the initial contour is not contained in the spinal cord, it is classified as *not contained*. Table 1 also shows the hit rate of the initial contour landing on the spinal canal. Overall, 95% of initial contours fall within the spinal canal. Figures 4(d) and 4(e) compare the segmentation results from our model with a reference segmentation. Table 2 shows that the means and the standard deviations of SI for two different reference segmentation by two medical specialists.
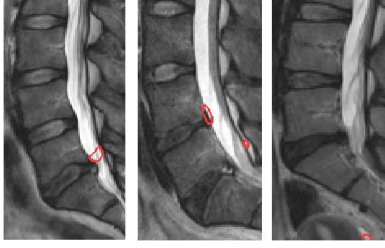
Fig. 5. Types of initial contours: Fully contained (left), Semi-contained (middle), Not contained (right) in the spinal cord

Hit criterion	Full hit	Semi-hit
Rate	67%	28%

Table 1. Hit rates of initial contours based on saliency map

Overall, SI remains relatively consistent across all gold standards (*i.e.*, reference segmentation). Specifically, the value of similarity index D calculated from our model is 0.71 on average, which indicates substantial agreement. Along with the comparison between results from our model and reference segmentation, we compared two reference segmentations in terms of the same similarity index D. The mean index value D between two observers is 0.90, implying that 90 percent of the reference segmentation by two observers is overlapping. The difference of the similarity index between two observers accounts for inter-observer variability.

The saliency map generated from the model is a global data structure. Thus, in order to extract the local features such as vertebrae or intervertebral discs, other elements might be added to this model.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present a fully automatic segmentation method using contextual information derived from MR images. In order to extract contextual information, a bottom-up Gabor filter attention model is used. The thresholded saliency map from the attention model is then integrated into an level-set active segmentation model for extracting out the spinal canal that houses the spinal cord and the thecal sac, followed by a morphological blob removal process. Experiments show that this procedure performs segmentation robustly achieving the Dice's similarity index of 0.71, a score considered to reflect substantial agreement. Segmentation results on medical MR images with comparison against hand annotation of trained medical observers also show a high degree of consistency between the automated and human-intervention methods. In the future, we plan to study the performance of our model in terms of speedup and over a larger variety of test environments.

Metric	Observer 1	Observer 2	Inter-observer
Mean of SI	0.70	0.71	0.90
Standard deviation of SI	0.069	0.063	0.024

Table 2. Intra-observer variability in terms of the mean and standard deviation of similarity index D for two different reference segmentation

6. REFERENCES

[1] O. Coulon, S.J. Hickman, G.J. Parker, G.J. Barker, D.H. Miller, and S.R. Arridge, "Quantification of spinal cord atrophy from magnetic resonance images via a b-spline active surface model," *Mag. Res. in Med.*, vol. 47, pp. 1176–1185, 2002.

[2] R.V. Uitert, I. Bitter, and J.A. Butman, "Semi-automatic spinal cord segmentation and quantification," *CARS*, pp. 224–229, 2005.

[3] L.G. Nyúl, J. Kanyó, E. Máté, G. Makay, E. Balogh, M. Fidrich, and A. Kuba, "Method for automatically segmenting the spinal cord and canal form 3d ct images," *CAIP 2005*, vol. 3691, pp. 456–463, 2005.

[4] M. Nieniewski and R. Serneels, "Segmentation of spinal cord images by means of watershed and region merging together with inhomogeneity correction," *IJMGV*, vol. 11, pp. 101–121, 2002.

[5] N.A. Archip, P-J. Erard, M. Egmont-Petersen, J-M. Haefliger, and J-F. Germond, "A knowledge-based approach to automatic detection of the spinal cord in ct images," *IEEE TMI*, vol. 21, pp. 1504–1516, 2002.

[6] B.D. Schmit and M. K. Cole, "Quantification of morphological changes in the spinal cord in chronic human spinal cord injury using magnetic resonance imaging," *IEEE EMBS*, pp. 4425–4428, 2004.

[7] C. McIntosh and G. Hamarneh, "Spinal crawlers: Deformable organisms for spinal cord segmentation and analysis," *MICCAI 2006*, vol. 4190, pp. 808–815, 2006.

[8] M.A. Horsfield, S. Sala, M. Neema, M. Absinta, A. Bakshi, M.P. Sormani, M.A. Rocca, R. Bakshi, and M. Filippi, "Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis," *NeuroImage*, vol. 50, pp. 446–455, 2010.

[9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *ICCV*, pp. 259–268, 1987.

[10] M. Guironnet, N. Guyader, D. Pellerin, and P. Ladret, "Static and dynamic feature-based visual attention model: Comparison to human judgement," *EUSIPCO*, 2005.

[11] T.F. Chan, J. Shen, and L. Vese, "Variational pde models in image processing," *Notices-AMS*, vol. 20, pp. 14–26, 2003.

[12] D. Mumford and J. Shah, "Optimal approximation by piecewise smooth functions and associated variational problems," *CPAM*, vol. 42, pp. 577–685, 1989.

[13] J. Koh, T. Kim, V. Chaudhary, and G. Dhillon, "Segmentation of the spinal cord and the dural sac in lumbar mr images using gradient vector flow field," *EMBC*, pp. 3117–3120, 2010.

[14] S.K. Michopoulou, L. Costaridou, E. Panagiotopoulos, R. Speller, G. Panayiotakis, and A. Todd-Pokropek, "Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine," *TBME*, vol. 56, pp. 2225–2231, 2009.

[15] H.L. Kundel, "Measurement of observer agreement," *RSNA*, pp. 303–308, 2003.