

Segmentation of Intervertebral Discs in 3D MRI Data Using Multi-atlas Based Registration

Chunliang Wang^{1,2} and Daniel Forsberg^{1,2}(✉)

¹ Sectra, Linköping, Sweden

`daniel.forsberg@sectra.com`

² Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

`chunwan@kth.se`

Abstract. This paper presents one of the participating methods to the intervertebral disc segmentation challenge organized in conjunction with the 3rd MICCAI Workshop & Challenge on Computational Methods and Clinical Applications for Spine Imaging - MICCAI–CSI2015. The presented method consist of three steps. In the first step, vertebral bodies are detected and labeled using integral channel features and a graphical parts model. The second step consists of image registration, where a set of image volumes with corresponding intervertebral disc atlases are registered to the target volume using the output from the first step as initialization. In the final step, the registered atlases are combined using label fusion to derive the final segmentation. The pipeline was evaluated using a set of 15 + 10 T2-weighted image volumes provided as training and test data respectively for the segmentation challenge. For the training data, a mean disc centroid distance of 0.86 mm and an average DICE score of 91 % was achieved, and for the test data the corresponding results were 0.90 mm and 90 %.

1 Introduction

Lower back pain is considered as one of the most common neurological ailments in the United States and as such costs associated to lower back pain form a significant portion of the total annual spending on healthcare. Degeneration of intervertebral discs (IVDs), as caused by aging, trauma, mechanical load, nutritional or genetic factors, is a common underlying cause of lower back pain. The degree of degeneration is typically assessed by means of magnetic resonance imaging (MRI), given the superior ability of MRI to distinguish between soft tissues and its absence of ionizing radiation. Because of significant inter-observer variation in grading IVD degeneration and constantly increasing workloads for radiologists, more and more research is devoted to develop computer assisted diagnosis systems to support the radiologists in their work and thereby improving the diagnostic process [1]. A crucial initial step in such a system is an accurate segmentation of the IVDs.

Over the last years, a number of different methods have been proposed for segmentation of IVDs [1–6]. The performance ranges from a mean DICE score

of a mediocre 74% to an impressive 92%. However, a drawback of some referenced methods is the limitation to two-dimensional (2D) image data, as typically given by the mid-sagittal slice of an image volume covering the lumbar spine. In addition, thus far no comparison between methods has been possible since all have been evaluated on different data sets. To this end, that is promoting three-dimensional (3D) segmentation of IVDs along with enabling a valid comparison of different methods, an IVD segmentation challenge was set up and organized in conjunction with the 3rd MICCAI Workshop & Challenge on Computational Methods and Clinical Applications for Spine Imaging - MICCAI-CIS2015.

In this paper, we present one of the challenge participants of said challenge. As such, the presented method is capable of both localizing and segmenting IVDs in MRI data. The method builds upon earlier work as presented by Lootus et al. [7] for vertebral body detection and labeling in MRI data and by Forsberg [8] for multi-atlas based segmentation of vertebrae in computed tomography data. The two approaches are combined and adapted to the task of localization and segmentation of IVDs in MRI data. Results are presented and discussed as pertaining to the training and test data provided for the challenge.

2 Materials and Methods

2.1 Image Data

The image data provided for training/testing and initial/final evaluation consisted of MRI data from 15 respectively 10 subjects, where each subject had been scanned with a 1.5 T scanner (Siemens Magnetom Sonata, Siemens Healthcare, Erlangen, Germany) and form a subset of the data used in the work of Chen et al. [2]. The image data consisted of sagittal T2-weighted turbo spin echo image volumes with a spatial resolution of $2.00 \times 1.25 \times 1.25 \text{ mm}^3$ and a size of $39 \times 305 \times 305$ or $48 \times 304 \times 304$. The IVDs have been manually segmented using the original sagittal images. Examples of the image data and ground truth segmentations are given in Fig. 1.

2.2 Segmentation Pipeline

The proposed segmentation pipeline consists of the following three components, detection and labeling of vertebral bodies, multi-atlas based segmentation per IVD and finally label fusion. Note that the presented method is a 3D-based method providing an accurate segmentation in 3D, even though the initial detection and labeling step is performed on individual 2D images.

2.3 Detection and Labeling

Detection and labeling of vertebrae is a challenge of its own and a number of methods have been presented in recent years. In our pipeline, it was decided to mimic the approach presented by Lootus et al. [7]. However, we employed aggregated channel features [9] coupled with an AdaBoost classifier for detection of the

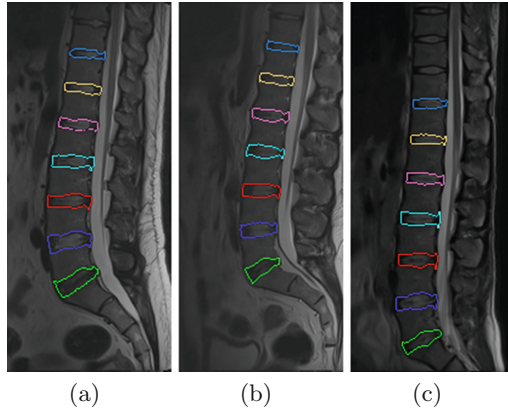


Fig. 1. Example of data used for the evaluation along with ground truth segmentations.

individual vertebrae. The reasons for not choosing the deformable parts model based on histogram of oriented gradients, as employed by Lootus et al. [7], were twofold. Firstly, it did not perform as well as the chosen approach, and secondly, it was more computationally demanding. Similar to the work of Lootus et al. [7], two different detectors were trained, one general vertebra detector and one for the fused S1 and S2 segments of the sacrum. To remove a significant portion of the false positive detections, a greedy non-maxima suppression algorithm was employed. In order to improve the performance of the detector, by increasing the number of detections, a set of sagittal slices (the three mid-sagittal slices) were fed as input to the vertebra detector. The individually detected objects are then combined using a pictorial structures model [10], further removing false positive detections along with labeling the detected vertebra. The object vertebra detectors along with the graphical parts model had previously been trained on a separate data set. Figure 2 provides a few examples depicting the output from the detection and labeling step.

The output from the detection and labeling consisted of a pair of y and z coordinates along with a label for each detected and labeled vertebra (assuming a coordinate system where x goes from right to left, y anterior to posterior and z inferior to superior). Corresponding x coordinates for each detection and labeling were simply set to the x coordinate of the mid-sagittal slice in the data set. Note that this works well as long as the orientation of the image volume is parallel with the spinal column and the subject has limited spinal deformities in the coronal plane.

2.4 Multi-atlas Based Segmentation

The output from the previous step provided a set of landmarks denoting the centerpoints of vertebra T11 to L5 along with the centerpoint of sacrum S1-S2. These centerpoints were used to provide rough estimates of the centerpoints of the corresponding IVDs, simply by using the midpoint between the centerpoints

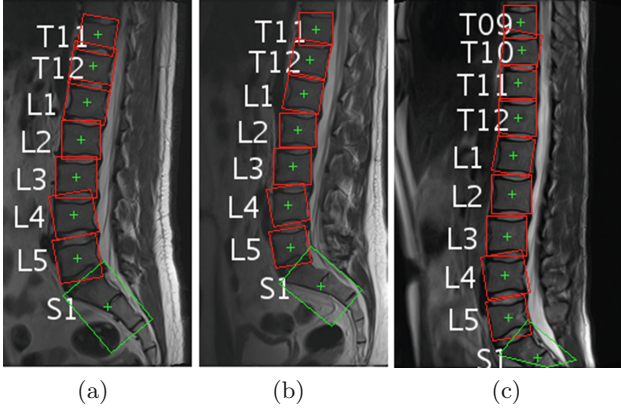


Fig. 2. Example results from the detection and labeling of vertebra. The detection and labeling works well even in cases where the sacrum is not fully depicted (c).

of two consecutive vertebrae. This served as input to the image registration in which a registration per disc was performed by extracting a sub-block (sized $40 \times 96 \times 96$ voxels, determined empirically to ensure a good coverage of each disc) of the data from each volume centered around the respective centerpoints of each disc. Note that for a general segmentation pipeline, the size of the sub-blocks should preferably be set in millimeters and with the possible extension to scale the size depending on sex, age and length of the patient. Each disc of the image data to segment is registered with multiple atlases. The registration was executed in two steps, where an initial affine registration was performed to account for differences in size and pose, and where a subsequent deformable registration was applied to account for local differences in shape. In both cases, local phase-based image registration approaches were applied.

Affine Registration. The affine phase-based registration was defined as an L_2 norm:

$$\epsilon^2 = \frac{1}{2} \sum_k \sum_{\mathbf{x} \in \Omega} c_k(\mathbf{x}) \left[\nabla \varphi_k(\mathbf{x})^T B(\mathbf{x}) \mathbf{p} - \Delta \varphi_k(\mathbf{x}) \right]^2, \quad (1)$$

where φ_k refers to the local phase-difference in orientation $\hat{\mathbf{n}}_k$ between the two images to be registered, c_k is a measure of certainty related to φ_k , and $B(\mathbf{x})\mathbf{p}$ corresponds to a linear parameterization of the local displacement $\mathbf{d}(\mathbf{x})$. A more detailed description is found in the works of Hemmendorff et al. [11] and Eklund et al. [12], and provides details on the employed graphics processing unit (GPU) implementation.

Deformable Registration. Similarly, as for the affine registration, a voxel-wise L_2 norm based upon local phase-differences was defined for the deformable counterpart:

$$\epsilon^2(\mathbf{x}) = \sum_k [c_k(\mathbf{x}) \mathbf{T}(\mathbf{x}) (\varphi_k(\mathbf{x}) \hat{\mathbf{n}}_k - \mathbf{u}(\mathbf{x}))]^2. \quad (2)$$

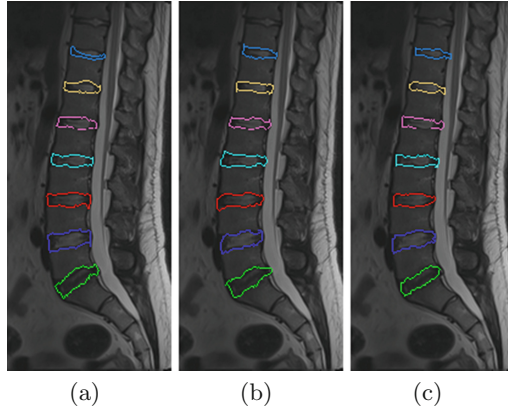


Fig. 3. Example results from the atlas-based registration. The individual segmentations as provided by the atlas-based segmentation are highly irregular and far from perfect, for example note the top disc in (a).

In this case, φ_k and c_k are as before, and \mathbf{T} refers to local structure tensor.

Solving for \mathbf{u} provides a voxel-wise update field $\mathbf{u}(\mathbf{x})$, which can be iteratively regularized and added to the final displacement field \mathbf{d} . Details on the registration algorithm can be found in the works of Knutsson and Andersson [13], and Forsberg et al. [14,15] for the employed GPU implementation.

Examples of output from the whole atlas-based segmentation step are shown in Fig. 3.

2.5 Label Fusion

The final step is to merge the labels of the multiple deformed atlases into a single label volume. In this case, a modified majority voting has been employed for label fusion, where instead of a standard majority vote only a minimum number of votes were required to render a valid segmentation. The reason for this approach was that since the discs are well-separated, it is only the background that provides a competing label. The minimum number of votes required for a segmentation was set to five, a number which was empirically determined. Example visualizations of the final segmentations are given in Fig. 4.

2.6 Evaluation

Given that 15 data sets were available for training, including ground truth data, a leave-one-out evaluation was performed for the training data in which one data set is segmented using the 14 others as atlases. This is then repeated for each available data set. Both IVD localization and segmentation results were included in the evaluation. For the evaluation on the test data, all 15 data sets were used as atlases.

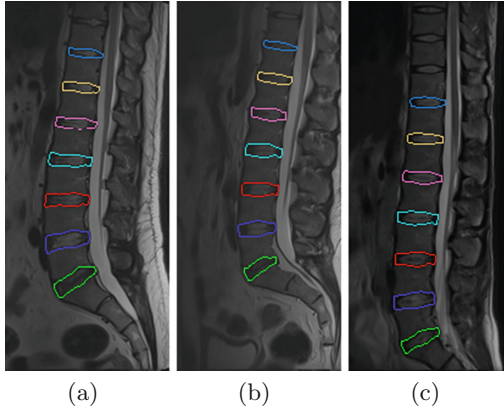


Fig. 4. Example results from the final step of label fusion. Previous irregularities are now gone and no apparent errors in the final segmentations are visible.

The evaluation of the test data was performed in a two-step process, with five data sets released prior to the challenge for an off-site evaluation and with remaining five data sets released on the day of the challenge for an on-site evaluation. The challenge organizers only provided results in terms of mean disc centroid distance and DICE score.

Localization. For each segmented disc, the disc centroid distance was computed as the Euclidean distance between the centroid of the ground truth IVD and the centroid of the segmented IVD obtained from the presented method. Based upon the disc centroid distance a disc localization was considered as successful if the distance was less than 2 mm.

Segmentation. The ground truth data was compared with the segmentations obtained from the multi-atlas based segmentation using the DICE score. The DICE score is defined as:

$$DICE = \frac{2 * |GT \cap S|}{|GT| + |S|}, \quad (3)$$

where GT and S refer to the ground truth and the computed segmentations respectively, and $|\dots|$ denotes the number of voxels, i.e. no respect was given to the anisotropic resolution.

To complement the agreement measure of the DICE score, false negative (FN) and false positive (FP) ratios were also computed as:

$$FN = \frac{|GT \setminus S|}{|GT|} \quad (4)$$

respectively

$$FP = \frac{|S \setminus GT|}{|S|}. \quad (5)$$

Table 1. Average DICE score, and false negative (FN) and false positive (FP) ratios per disc for the training data.

| Disc | DICE | FN | FP |
|---------|-----------------|-----------------|-----------------|
| T11/T12 | 0.89 ± 0.03 | 0.10 ± 0.04 | 0.11 ± 0.05 |
| T12/L1 | 0.91 ± 0.02 | 0.09 ± 0.02 | 0.09 ± 0.04 |
| L1/L2 | 0.91 ± 0.03 | 0.09 ± 0.04 | 0.10 ± 0.04 |
| L2/L3 | 0.92 ± 0.01 | 0.08 ± 0.03 | 0.07 ± 0.02 |
| L3/L4 | 0.92 ± 0.02 | 0.07 ± 0.03 | 0.08 ± 0.03 |
| L4/L5 | 0.92 ± 0.02 | 0.08 ± 0.03 | 0.08 ± 0.02 |
| L5/S1 | 0.89 ± 0.03 | 0.09 ± 0.04 | 0.12 ± 0.05 |

3 Results

3.1 Training Data

The mean disc centroid distance was 0.86 ± 0.45 mm. Only three disc centroid distances were larger than 2 mm (2.05, 2.91 and 2.06 respectively) providing an IVD detection rate of 97%.

The average achieved DICE score was 0.91 ± 0.01 along with an average FN and FP of 0.08 ± 0.02 and 0.09 ± 0.02 respectively. Detailed results per disc and subject are given in Tables 1 and 2 respectively.

Table 2. Average DICE score, and false negative (FN) and false positive (FP) ratios per subject for the training data.

| Subject | DICE | FN | FP |
|---------|-----------------|-----------------|-----------------|
| 1 | 0.91 ± 0.02 | 0.09 ± 0.03 | 0.10 ± 0.01 |
| 2 | 0.91 ± 0.02 | 0.07 ± 0.02 | 0.10 ± 0.04 |
| 3 | 0.91 ± 0.01 | 0.09 ± 0.03 | 0.08 ± 0.02 |
| 4 | 0.89 ± 0.05 | 0.09 ± 0.06 | 0.14 ± 0.05 |
| 5 | 0.92 ± 0.02 | 0.09 ± 0.03 | 0.06 ± 0.04 |
| 6 | 0.92 ± 0.03 | 0.09 ± 0.05 | 0.06 ± 0.02 |
| 7 | 0.91 ± 0.03 | 0.08 ± 0.01 | 0.11 ± 0.05 |
| 8 | 0.92 ± 0.02 | 0.10 ± 0.03 | 0.06 ± 0.02 |
| 9 | 0.93 ± 0.01 | 0.09 ± 0.01 | 0.06 ± 0.01 |
| 10 | 0.93 ± 0.01 | 0.07 ± 0.02 | 0.06 ± 0.02 |
| 11 | 0.90 ± 0.03 | 0.08 ± 0.04 | 0.12 ± 0.03 |
| 12 | 0.92 ± 0.01 | 0.04 ± 0.02 | 0.12 ± 0.02 |
| 13 | 0.88 ± 0.02 | 0.11 ± 0.02 | 0.13 ± 0.05 |
| 14 | 0.91 ± 0.01 | 0.10 ± 0.04 | 0.07 ± 0.02 |
| 15 | 0.91 ± 0.02 | 0.08 ± 0.02 | 0.10 ± 0.04 |

3.2 Test Data

For the off-site evaluation the mean disc centroid distance was 0.81 ± 0.42 mm and a detection rate of 97%. Corresponding results for the on-site evaluation was 0.99 ± 0.78 mm and a detection rate of 80%. The achieved DICE score was 0.90 ± 0.03 for both the off-site and on-site evaluation.

4 Discussion

In this paper, we have presented one of the methods participating in the IVD segmentation challenge, organized in conjunction with MICCAI–CSI2015, a method relevant for both robust IVD detection and accurate IVD segmentation. The method has been evaluated using training and testing data provided by the challenge organizers. Performance was assessed using disc centroid distance, the DICE score along with computing the ratios of false negatives and false positives (the latter two only computed for the training data).

The presented method achieved a mean disc centroid distance of 0.86 ± 0.45 mm, with a success rate of 97% given a threshold of 2 mm for the training data. This can be compared with a mean disc centroid distance of 2.08 mm as reported by Ghosh and Chaudhary [3] for the 2D case, 1.23 mm by Law et al. [4] and 1.6 – 2.0 mm by Chen et al. [2] (both the latter for the 3D case). As such, the presented results are superior to earlier results. Note that the results obtained for the on-site evaluation was somewhat lower than for the training data and for off-site evaluation.

In terms of segmentation accuracy, the presented method performs on par with current state-of-the-art methods for IVD segmentation. For example, Michopoulou et al. [5] achieved an impressive mean DICE score of 0.92, however, only for segmentation of 2D image data and using manual interaction for performing the initial atlas-based registration. Similar DICE scores were presented in the work of Law et al. [4], but again only for 2D image data. Neubert et al. [16] presented an extension of their initial work [6] using multi-level statistical model and achieved a mean DICE score of 0.91 on 3D data.

In Tables 1 and 2 it can be noted that the segmentation accuracy is stable over both discs and subjects, i.e. there exists no failed cases and any future improvements are, thus, rather related to fine-tuning of parameters than making major changes in the presented pipeline. The ratios of false negatives and false positives show that there appear to be an equal distribution of under- and over-segmentation between discs and subjects.

Limitations of the presented results are given by the small size of the data set employed for the evaluation along with its homogeneity. Further, the data set lacks in cases of degenerated IVDs, hence, it is difficult to foresee the performance of the presented segmentation pipeline on more clinically relevant data, including a variety of degenerated IVDs. Another limitation is given by the dependence of the registration step on the detection and labeling step. A missed vertebra in-between other vertebrae can be handled with some additional heuristics to account for long distance between detected vertebra. In the case the detection

and labeling is off by one or two labels, i.e. in the case when S1-S2 is missed and instead L5-S1 is labeled as S1-S2, the segmentation is still expected to perform well but neglects to segment the most inferior IVD.

In all, the presented method along with the evaluation results, a mean disc centroid distance of 0.86 ± 0.45 mm and an average DICE score of 0.91 for the training data, show that robust localization and accurate segmentation of IVDs are achievable.

Acknowledgements. The work of D.F. was partially funded by the Swedish Innovation Agency (VINNOVA), grant 2014-01422.

References

1. Oktay, A., Albayrak, N., Akgul, Y.: Computer aided diagnosis of degenerative intervertebral disc diseases from lumbar MR images. *Comput. Med. Imaging Graph.* **38**(7), 613–619 (2014)
2. Chen, C., Belavy, D., Yu, W., Chu, C., Armbrecht, G., Bansmann, M., Felsenberg, D., Zheng, G.: Localization and segmentation of 3D intervertebral discs in MR images by data driven estimation. *IEEE Trans. Med. Imaging* **34**(8), 1719–1729 (2015)
3. Ghosh, S., Chaudhary, V.: Supervised methods for detection and segmentation of tissues in clinical lumbar MRI. *Comput. Med. Imaging Graph.* **38**(7), 639–649 (2014)
4. Law, M., Tay, K., Leung, A., Garvin, G., Li, S.: Intervertebral disc segmentation in MR images using anisotropic oriented flux. *Med. Image Anal.* **17**(1), 43–61 (2013)
5. Michopoulou, S., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A.: Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. *IEEE Trans. Biomed. Eng.* **56**(9), 2225–2231 (2009)
6. Neubert, A., Fripp, J., Engstrom, C., Schwarz, R., Lauer, L., Salvado, O., Crozier, S.: Automated detection, 3D segmentation and analysis of high resolution spine MR images using statistical shape models. *Phys. Med. Biol.* **57**(24), 8357–8376 (2012)
7. Lootus, M., Kadir, T., Zisserman, A.: Vertebrae detection and labelling in lumbar MR images. In: Yao, J., et al. (eds.) *CSI 2013. LNCV*, vol. 17, pp. 219–230. Springer, Switzerland (2014)
8. Forsberg, D.: Atlas-based registration for accurate segmentation of thoracic and lumbar vertebrae in CT data. In: Yao, J., et al. (eds.) *CSI 2014. LNCVB*, vol. 20, pp. 49–59. Springer, Switzerland (2015)
9. Dollár, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: *Proceedings of 2009 British Machine Vision Conference, BMVC 2009*, pp. 91.1–91.11. BMVA Press (2009)
10. Felzenszwalb, P., Huttenlocher, D.: Pictorial structures for object recognition. *Int. J. Comput. Vis.* **61**(1), 55–79 (2005)
11. Hemmendorff, M., Andersson, M., Kronander, T., Knutsson, H.: Phase-based multidimensional volume registration. *IEEE Trans. Med. Imaging* **21**(12), 1536–1543 (2002)

12. Eklund, A., Andersson, M., Knutsson, H.: Phase based volume registration using CUDA. In: Proceedings of 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2010, pp. 658–661. IEEE (2010)
13. Knutsson, H., Andersson, M.: Morphons: segmentation using elastic canvas and paint on priors. In: Proceedings of 12th IEEE International Conference on Image Processing, ICIP 2005, vol. 2, pp. 1226–1229. IEEE (2005)
14. Forsberg, D.: Robust image registration for improved clinical efficiency: using local structure analysis and model-based processing. Ph.D. thesis, Linköping University, Sweden (2013)
15. Forsberg, D., Eklund, A., Andersson, M., Knutsson, H.: Phase-based non-rigid 3D image registration - from minutes to seconds using CUDA. In: Proceedings of Joint MICCAI Workshop on High Performance and Distributed Computing for Medical Imaging - HP/DCI 2011 (2011)
16. Neubert, A., Fripp, J., Engstrom, C., Crozier, S.: Segmentation of lumbar intervertebral discs from high-resolution 3D MR images using multi-level statistical shape models. In: Yao, J., et al. (eds.) CSI 2013. LNCVB, vol. 17, pp. 173–183. Springer, Switzerland (2014)