

Lumbar Disc Localization and Labeling with a Probabilistic Model on Both Pixel and Object Features

Jason J. Corso, Raja' S. Alomari, and Vipin Chaudhary*

Department of Computer Science and Engineering
University at Buffalo, State University of New York
jcorso@cse.buffalo.edu

Abstract. Repeatable, quantitative assessment of intervertebral disc pathology requires accurate localization and labeling of the lumbar region discs. To that end, we propose a two-level probabilistic model for such disc localization and labeling. Our model integrates both pixel-level information, such as appearance, and object-level information, such as relative location. Utilizing both levels of information adds robustness to the ambiguous disc intensity signature and high structure variation. Yet, we are able to do efficient (and convergent) localization and labeling with generalized expectation-maximization. We present accurate results on 20 normal cases (96%) and a promising extension to a pathology case.

1 Introduction

Past [1] and recent [2] studies have suggested a need for repeatable quantitative intervertebral disc degeneration (IDD) assessment methods. IDD is prevalent, especially in modern times—about 12 million Americans have some type of IDD—and can cause very high pain levels. One example of an IDD is herniation, where the disc, a structure that acts like a shock-absorber between vertebrae, bulges out of place causing fragments of disc material to press on the nerve roots in the spinal column. In the clinic, a radiologist diagnoses an IDD by first localizing and labeling each intervertebral disc and then identifying any pathology in a given disc based on its local geometry and appearance.

Accurate localizing and labeling is necessary for diagnosis of an IDD pathology. However, the underlying image signal is ambiguous: the intensity of a disc greatly overlaps with that of the spinal nerve fibers. Even the structure can change from case to case, with possible bending of the spinal column (scoliosis) and missing or additional vertebrae [3]. In Figure 1, we show a normal example of the lumbar region of the spinal column with the discs labeled according to the standard scheme.

In this paper, we propose a method to automate the detection and labeling of the intervertebral discs from single MR radiographs. Although we study lumbar discs, our method is directly extensible to the full spinal column. Our method is based on a two-level probabilistic model. On the

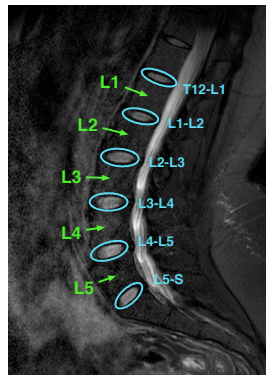


Fig. 1. Lumbar region

* This work was supported in part by a grant from NYSTAR.

first level, we model the local pixel-level properties of the discs, such as appearance. On the second level, we capture the object-level geometrical and contextual relationships between discs. The model parameters are estimated from previously labeled cases (supervised learning). The two-level model insulates the localization of the discs from the complex appearance variation in the underlying radiograph leading to a robust and efficient algorithm.

2 Related Work

Automatic localization and labeling of inter-vertebral discs has been a focus of the medical imaging community for over two decades, and it has largely evaded researchers due to the ambiguous disc appearance and high structural variation. Early methods, e.g., Chwialkowski et al. [4], are primarily model- or heuristic-based such as analyzing a horizontal cross-section of intensities to search for the posterior and anterior disc edge. These early models have been influential (e.g., [5]), but were generally tested on few cases and it is not known if they scale.

The recent work by Peng et al. [5] uses such a model-based approach to localize the disc position based on template convolution and an intensity profile. Output from this process is used to help automatically pick the sagittal slice from the *volume* to work with; the slice with minimum variance for peaks of the intensity profile are used. Pekar et al. [6] also develop an approach for labeling the vertebral column as part of scan geometry planning (a step beyond selecting a particular slice from an existing scan). They search for possible disc locations and then do 3D connected components to find disc centers. Masaki et al. [7] also propose a method for automated geometry planning based on intensity and a Hough transform. Five radiologists agree in 9 out of 10 cases that the automatic image plane is as good as or better than manual.

Weiss et al. [8] propose a semi-automatic technique for labeling discs. The user manually marks one disc and then the algorithm proceeds by an iterative intensity-analysis based method. The upper and lower halves of the spine are independently processed using threshold values, filters and noise suppression operators. The technique is highly dependent on imaging quality and data dependent thresholding values. Zheng et al. [9] do segmentation of lumbar vertebrae using digital videofluoroscopic images, which are more noisy than the standard MR radiographs but have a time component. They propose a method based on shape descriptors and a Hough transform (relatively high dimension); the method is validated on synthetic data and a single *in vivo* sequence.

Our proposed method is most similar to the recent work by Schmidt et al. [10]. These two works take a step away from the heuristic driven approaches and construct probabilistic models. [10] define a probabilistic graphical model based on the vertebrae. It incorporates appearance and shape information and uses the A* algorithm, which does a best-first greedy coordinate search for the solution. Their method assumes a full 3D volume, which is not a current clinical standard. In contrast, our method uses a single T2 radiograph (current clinics use multi-spectral radiographs). Their method uses comparatively opaque discriminative models, while we develop a generative probabilistic model that also incorporates appearance and object information but in a more direct,

transparent manner. We structure our model in two levels to increase robustness to variation and make it plausible to do full inference using expectation-maximization.

3 Approach

3.1 Conventional Labeling Approach

Our goal is to localize each disc connected to the lumbar vertebrae. One standard probabilistic approach is to formulate a labeling problem with one label per disc and do inference by assigning the most probable label at each pixel. The brain structure labeling work of Fischl et al. [11] is an example of this standard problem formulation. Concretely, let $\Lambda = \{s = (x, y) : 0 \leq x < n, 0 \leq y < m\}$ be the image lattice and consider the image as a map from the lattice to intensities¹, $I: \Lambda \mapsto \mathbb{R}$. Define the set of disc labels $\mathcal{A} = \{1, \dots, 6\}$ (there are six discs connected to lumbar vertebrae), and a set $\mathcal{T}$ of label variables t_i with one for each pixel $i \in \Lambda$. Then, given an image I , one seeks the maximum a posteriori estimate of the labels:

$$\mathcal{T}^* = \arg \max_{\mathcal{T}} P(\mathcal{T}|I) . \quad (1)$$

However, this inference problem is non-trivial; one typically must make an assumption of independence or rely on (opaque) discriminative models such as the randomized tree classifier as done by Schmidt et al. [10]. Both assumptions may break down due to the high degree of intensity similarity to neighboring anatomy and across discs, the large spatial variability of the discs, and pathological discs (e.g., herniation). Furthermore, it is difficult to incorporate high- or object-level information (such as the relative disc ordering) into this formulation since all of the variables are represented at the pixel level. This difficulty not only affects the robustness of modeling, but also makes it necessary to add post-processing constraints to enforce additional problem specific constraints, such as each disc being a closed elliptical region.

3.2 Two-Level Probabilistic Model

We instead propose a two-level probabilistic model that only requires conditional independence, is generative (more transparent), and adequately insulates the localization variables from the pixel intensities while at the same time modeling the exact disc geometry rather than solely pixel-level labels. Let $\mathcal{D} = \{d_1, d_2, \dots, d_6\}$ be the set of disc variables with each $d_i = (x_i, y_i)^T$ representing the disc center (they could also include disc angle, boundary, etc.). Inferring these from an image is our ultimate goal, but we avoid doing it directly due to the difficulties mentioned above. Rather, we introduce a set of auxiliary variables, called disc-label variables and denoted $\mathcal{L} = \{l_i, \forall i \in \Lambda\}$. The disc-labels make it plausible to separate the disc variables from the image intensities; i.e., the disc-label variables will capture the local pixel-level intensity models while the disc variables will capture the high-level geometric and contextual models of

¹ These are MR images, but we consider the pixel values as simple intensities without incorporating any special MR related model.

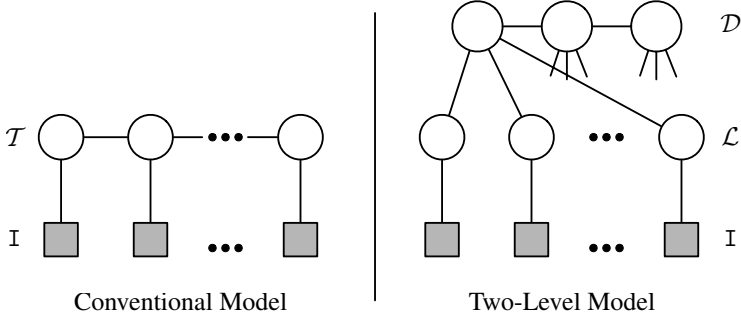


Fig. 2. Graphical models depicting the conventional probabilistic model and our proposed two-level probabilistic model. Our model (right) separates the low- or pixel-level information from the high- or object-level information adding disc localization robustness to the common intensity and structure variation. In the figure, filled squares are observed data (the image) and hollow circles are latent variables. The image and first level ($\mathcal{T}$ and $\mathcal{L}$) sit on the 2D or 3D image lattice Λ , but the high level is a single 1D chain with one node per disc.

the full set of discs. Each disc-label variable can take a value of $\{-1, +1\}$ for non-disc or disc, respectively. Note the simpler situation than above where we had a particular label for each disc. Figure 2 presents and compares the two modeling situations.

We marginalize over the possible disc-labelings since these are auxiliary variables giving the following optimization function:

$$\mathcal{D}^* = \arg \max_{\mathcal{D}} \sum_{\mathcal{L}} P(\mathcal{L}, \mathcal{D} | \mathcal{I}) = \arg \max_{\mathcal{D}} \sum_{\mathcal{L}} P(\mathcal{L} | \mathcal{D}, \mathcal{I}) P(\mathcal{D}) , \quad (2)$$

where the second equality follows from the multi-level nature of the model (the disc variables are assumed independent of the intensities). Note the summation is over a very large set of possible assignments ($2^{|\Lambda|}$). We model it as a Gibbs distribution:

$$\begin{aligned}
 P(\mathcal{L} | \mathcal{D}, \mathcal{I}) &= \frac{1}{Z[\mathcal{L}]} \exp \left[-\beta_1 \sum_{s \in \Lambda} U_I(l_s, I(s)) \right] && \leftarrow \text{intensity (3)} \\
 &\quad -\beta_2 \sum_{s \in \Lambda} U_D(l_s, \mathcal{D}) && \leftarrow \text{spatial} \\
 P(\mathcal{D}) &= \frac{1}{Z[\mathcal{D}]} \exp \left[-\beta_3 \sum_{d_i \in \mathcal{D}} U_L(d_i) \right] && \leftarrow \text{location} \\
 &\quad -\beta_4 \sum_{(i \sim j)} V_D(d_i, d_j) && \leftarrow \text{disc-context}
 \end{aligned}$$

where $\beta_k \geq 0, k = \{1, \dots, 4\}$ are tunable parameters and $Z[\cdot]$ are the partition functions. The $(\cdot \sim \cdot)$ notation denotes the set of neighboring elements on the disc chain. Each term will be explicitly defined in the following sections.

The first level, $P(\mathcal{L} | \mathcal{D}, \mathcal{I})$, captures the probability of a particular labeling given both the underlying image and the overlying disc variables. Each potential function models

a different aspect of the local pixel-level information (the aspect is mentioned on the right of each equation-line). The second level $P(\mathcal{D})$ models the high-level information about the disc locations and context.

3.3 Low Level Terms

Intensity. The potential, $U_I(l_s, \mathbf{I}(s))$, models the pixel appearance $\mathbf{I}(s)$ based on its current label l_s . We use a Gaussian for the disc pixels, motivated by empirical study of the distribution of disc pixels (figure 3), and take the negative-log of it for the potential. The parameters of the Gaussian, μ_I and σ_I^2 , are learned from labeled training data (specific details in experiments section §4). We drop the normalizing term since it does not depend on the specific intensity value:

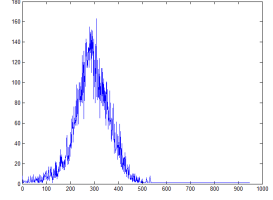


Fig. 3. Disc intensity empirical distribution

$$U_I(l_s, \mathbf{I}(s)) = \begin{cases} \frac{(I(s) - \mu_I)^2}{2\sigma_I^2} & \text{if } l_s = +1 \\ -\log \left(1 - \exp \left[-\frac{(I(s) - \mu_I)^2}{2\sigma_I^2} \right] \right) & \text{if } l_s = -1 \end{cases} \quad (4)$$

Since there are a (small) finite number of intensities, the second case (with the log) can be precomputed and cached without incurring great computational burden.

Spatial. The potential $U_D(l_s, \mathcal{D})$ models the spatial relationship between a disc-label and the set of discs. Intuitively, a disc-label is more likely to take value $+1$ the closer it is to the location specified by one of the discs. We compute the covariance matrix Σ_i (i.e., shape) of each disc d_i during training (roughly, the discs are elliptically shaped) and then base the spatial potential on the squared Mahalanobis distance:

$$U_D(l_s, \mathcal{D}) = l_s \cdot \min_{d_i \in \mathcal{D}} \left[(s - d_i)^\top \Sigma_i^{-1} (s - d_i) \right] . \quad (5)$$

The potential assigns energy proportional to the distance of the closest disc, and the label variable acts as a switch: when $l_s = -1$, being far from the closest disc is lower energy than being closer to it and vice versa.

3.4 High Level Terms

Location. The potential $U_L(d_i)$ measures the distance of disc d_i to its expected location. Similar to the spatial low level term in (5), we estimate the covariance Σ_i for each disc (same as above) and, in this case, the mean location μ_i . We do the estimation offline from training data. The squared Mahalanobis distance defines the potential:

$$U_L(d_i) = (d_i - \mu_i)^\top \Sigma_i^{-1} (d_i - \mu_i) . \quad (6)$$

Disc-Context. The potential $V_D(d_i, d_j)$ captures the high-level contextual relationship between two *neighboring* discs. Consider the discs form a chain; then the neighboring

pairs are nearest neighbors on the chain. Since we include only the spatial location for each disc variable, the distance between neighboring discs is a natural measure for this potential. Inspecting the empirical distribution of these distances (figure 4) suggests a Gaussian parameterized by μ_D, σ_D^2 , and its negative log for the energy. Let $e_{ij} = |d_i - d_j|_2$, then

$$V_D(d_i, d_j) = \frac{(e_{ij} - \mu_D)^2}{\sigma_D^2} . \quad (7)$$

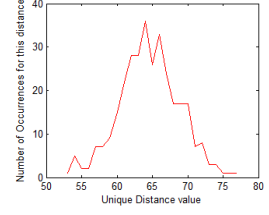


Fig. 4. Empirical distribution of disc distances

3.5 Inference Using Generalized Expectation-Maximization

We use the generalized EM (gEM) algorithm [12] to solve (2). We initialize $\mathcal{D}^0$ by setting each disc d_i in its mean location μ_i , which we've learned offline from training data. Then, we iteratively estimate the posterior over the disc-label variables, $\mathcal{L}$, and refine the disc variables by maximizing the expected log likelihood (ELL):

$$\textbf{E-step} \rightarrow F^t(\mathcal{L}) = P(\mathcal{L} | \mathcal{D}^t, \mathbf{I}) \quad (8)$$

$$\textbf{M-step} \rightarrow \mathcal{D}^{t+1} = \arg \max_{\mathcal{D}} \left[\log P(\mathcal{D}) + \sum_{\mathcal{L}} F^t(\mathcal{L}) \log P(\mathcal{L}, \mathbf{I} | \mathcal{D}) \right] \quad (9)$$

Inference with gEM is tractable without resorting to monte carlo methods because of the underlying structure of our two-level model. Since no dependencies are defined among the disc-label variables $\mathcal{L}$, we factor them into independent local terms:

$$P(\mathcal{L} | \mathcal{D}, \mathbf{I}) = \prod_{s \in \Lambda} \frac{1}{Z[l_s]} \exp \left[-\beta_1 U_I(l_s, \mathbf{I}(s)) - \beta_2 U_D(l_s, \mathcal{D}) \right] . \quad (10)$$

Thus, we directly evaluate the full posterior (8) and log terms (9) during optimization. Since the partials are not analytically available, we execute a finite differences-based gradient ascent algorithm to iteratively maximize the ELL. It is beyond the scope of this paper to go into full detail; basically, iteratively for each disc variable d_i , we evaluate the local gradient of the ELL by perturbing d_i by a set of changes $\{\delta\}$. For each perturbation, we fully evaluate (9) and change d_i by the δ yielding the maximum (if no δ increases the function, we do not change d_i). We stop when no d_i has changed.

4 Experimental Results

We use 20 pathology-free cases of clinical T2-weighted MR data for the lumbar spinal column (e.g., figure 1). The images have been acquired on a 3T Philips Medical Systems Intera Scanner with a voxel resolution of $0.5 \times 0.5 \times 4.5 \text{ mm}^3$. From each case, we extract the four middle slices giving a total of 80 radiographs. We work in 2D because there is 5mm inter-slice space and it is standard clinical practice. We have manually annotated the images to specify each disc center as well as to label a subset

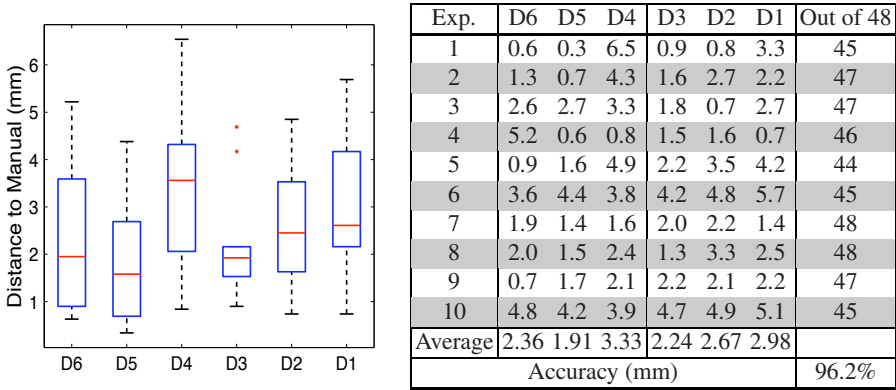


Fig. 5. Results of cross validation (CV). Every time two cases (8 images) are pulled out and training is performed on the other 72 images. (Left) A box-plot showing summary statistics of Euclidean distances over all testing cases in CV. The line in each box is the median, the top and bottom are the 75th and 25th percentiles, and plusses are statistical outliers. (Right) The mean Euclidean distances for each of the 10 CV experiments and the labeling results (right-most column) marking how many discs were localized inside of the correct disc. Accuracy is 96.2%.

of the disc pixels so that we could perform training of the necessary parameters for our model (intensity, spatial, etc.). We set the parameters of the model β_1 through β_4 by hand; these could be learned by various supervised parameter estimation algorithms.

We have performed a leave-two-out cross validation (CV) experiment. In each CV sub-experiment, we separate all four images from two cases to be used for testing and leave the rest (72 images in 18 cases) for training; this prevents bias since multiple slices are taken from the same case. Figure 5 summarizes the results of this experiment using two quantitative mechanisms: the Euclidean distance to the known disc center and a boolean score that is yes if the inferred label lies anywhere inside of the disc (this is scored manually). A distance of less than 3mm is within the margin of error in human labeling of the disc center. The discs are labeled from bottom to top, i.e., D1 is S-L5. Our scores are comparable or superior to [10] (depending on which variant of their algorithm is analyzed), but they use a full 3D volume and opaque discriminative models while we use a 2D image and transparent generative models. We have focused exclusively on the lumbar region while they do the full spine (our model is directly extensible).

Figure 6 shows the labeling in four test images; the first three are from our dataset and the fourth is a pathology case. The images show a green line that traces the path during disc optimization and a plus for the converged point. We can see (e.g., left-most D4) that even when initialization is far from the true disc, the two-level model is able to drive the variable to an accurate position. Although the model is not designed to handle the abnormalities presented in the fourth case (disc degeneration, bottom two), we are able to successfully localize each disc in this image too, suggesting our model could extend to the more complicated cases of abnormalities (with some enhancements). One failure we see is D3 in the left-most image; here, the disc is labeled but the model fails to accurately localize the disc center.

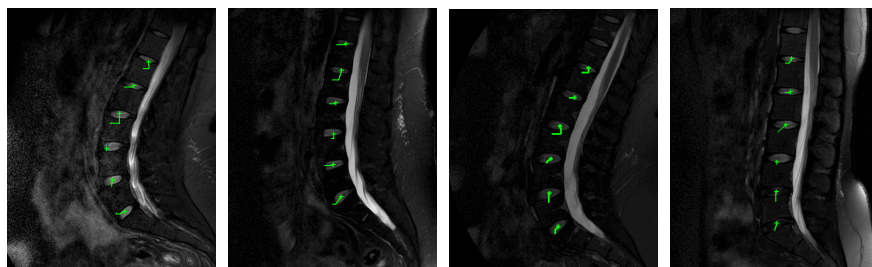


Fig. 6. Labeling results. Each image shows disc optimization path and a plus for the converged point. The right-most image is an abnormal case not in our data-set.

5 Summary and Discussion

We have proposed a two-level probabilistic model for robustly localizing and labeling intervertebral discs. The unifying model incorporates both pixel-level information (e.g., appearance) and high-level information (e.g., spatial relationship between discs) in a single coherent generative model. This two-level approach insulates the high-level inferences (about disc localization) from the pixel-level variations.

Our quantitative results show good localizing and labeling performance: about 96% labeling accuracy in a cross validation experiment. We plan to extend the size of the dataset and adapt the method to pathological cases, which is our ultimate goal. As the example in figure 6 shows, the generalization to pathological cases is possible with the straightforward potentials proposed in this paper, but will require extensions to the model. To that end, we plan to enhance the model by incorporating local texture into the low-level terms and disc orientation and shape into the high-level terms.

References

1. Jenkins, J.P., Hickey, D.S., Zhu, X.P., Machin, M., Isherwood, I.: Mr imaging of the intervertebral disc: A quantitative study. *British Journal of Radiology* 58(692), 705–709 (1985)
2. Antoniou, J., Mwale, F., Demers, C.N., Beaudoin, G., Goswami, T., Aebi, M., Alini, M.: Quantitative magnetic resonance imaging of enzymatically induced degraded of the nucleus pulposus of intervertebral discs. *Spine* 31(14), 1547–1554 (2006)
3. Dalley, A.F., Agur, A.M.R.: *Atlas of Anatomy*. Lippincott Williams and Wilkins (2004)
4. Chwialkowski, M.P., Shile, P.E., Peshock, R.M., Pfeifer, D., Parkey, R.W.: Automated detection and evaluation of lumbar discs in mr images. In: *Proc. of IEEE EMBS* (1989)
5. Peng, Z., Zhong, J., Wee, W., Lee, J.: Automated vertebra detection and segmentation from the whole spine MR images. In: *Proc. of IEEE EMBS*, vol. 3 (2005)
6. Pekar, V., Bystrov, D., Heese, H.S., Dries, S.P.M., Schmidt, S., Grever, R., Harder, C., Bergmans, R.C., Simonetti, A.W., Muisinkel, A.: Automated planning of scan geometries in spine MRI scans. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 601–608. Springer, Heidelberg (2007)
7. Masaki, T., Lee, Y., Tsai, D.Y., Sekiya, M., Kazama, K.: Automatic determination of the imaging plane in lumbar mri. In: *Proc. of SPIE Med. Img.*, pp. 1252–1259 (2006)

8. Weiss, K.L., Storrs, J.M., Banto, R.B.: Automated spine survey iterative scan technique. *Radiology* 239(1), 255–262 (2006)
9. Zheng, Y., Nixon, M.S., Allen, R.: Automatic segmentation of lumbar vertebrae in digital videofluoroscopic imaging. *IEEE Trans. on Medical Imaging* 23(1), 45–52 (2004)
10. Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S.P., Bystrov, D., Schnorr, C.: Spine Detection and Labeling Using a Parts-Based Graphical Model. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007. LNCS*, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
11. Fischl, B., Salat, D.H., Busa, E., Albert, M., Deiterich, M., Haselgrove, C., Kouwe, A.v.d., Killiany, R., Kennedy, D., Klaveness, S., Monttilo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33, 341–355 (2002)
12. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood From Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society – Series B* 39(1), 1–38 (1977)