Labeling of Lumbar Discs Using Both Pixel- and Object-Level Features With a Two-Level Probabilistic Model

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

***Abstract*—Backbone anatomical structure detection and labeling is a necessary step for various analysis tasks of the vertebral column. Appearance, shape and geometry measurements are necessary for abnormality detection locally at each disc and vertebrae (such as herniation) as well as globally for the whole spine (such as spinal scoliosis). We propose a two-level probabilistic model for the localization of discs from clinical magnetic resonance imaging (MRI) data that captures both pixel- and object-level features. Using a Gibbs distribution, we model appearance and spatial information at the pixel level, and at the object level, we model the spatial distribution of the discs and the relative distances between them. We use generalized expectation-maximization for optimization, which achieves efficient convergence of disc labels. Our two-level model allows the assumption of conditional independence at the pixel-level to enhance efficiency while maintaining robustness. We use a dataset that contains 105 MRI clinical normal and abnormal cases for the lumbar area. We thoroughly test our model and achieve encouraging results on normal and abnormal cases.**

***Index Terms*—Gibbs distribution, hierarchical model, lumbar disc detection, magnetic resonance imaging (MRI), spine.**

I. Introduction

ACCORDING to the National Institute of Neurological Disorders and Stroke (NINDS), back pain is the second most common neurological ailment in the United States after headache. Over 12 million Americans have some sort of intervertebral disc disease (IDD) [1]. Localization and labeling of the vertebral column anatomical structures has thus been a focus of recent studies [2], [3] due to the high demand of analysis of the vertebral column structures such as disc size, disc shape, and water content percentage in discs. This analysis is a core requirement for the diagnosis of the vertebral column as a whole and for anatomical structures such as discs, vertebrae, and soft tissues.

Accurate labeling of the backbone structures is a necessary step for performing any type of analysis, diagnosis, or surgical planning. One key use of labeling is the design of a computer aided diagnosis system for lumbar area. In the clinical practice, the neuroradiologist reports the diagnosis at each disc level as shown in Fig. 2. Thus, the localization and labeling step is necessary to be able to perform any subsequent diagnosis. We have been using this labeling model in our subsequent diagnosis work in designing a computer aided diagnosis system for the lumbar area [4]–[7].

However, although accurate labeling of the backbone structures is necessary, the backbone has wide variabilities including degree of bending of the vertebral column, sizes, shapes, count (number), and appearances of discs and vertebrae. In addition, existing abnormality conditions such as vertebral fusion, degenerative disc diseases, spinal infections, and spinal scoliosis add more variability [8].

In this paper, we extend our work [9], where we proposed a two level probabilistic model for labeling disc structures from a clinical dataset. In our model, we incorporate two levels of information: low- and high-level. On the low-level, we model the local pixel properties of discs, such as appearance. On the high-level, we capture the object-level geometrical and contextual relationships between discs. Our two-level model allows the assumption of conditional independence at the pixel-level to enhance efficiency while maintaining robustness. We estimate the model parameters from manually labeled cases (supervised learning). We test our model using a dataset of 105 clinical magnetic resonance (MR) normal and abnormal cases. This dataset has wide variability in the types of abnormalities, patient ages (17–81 years old), and patient heights which affects the size and appearance of the discs. Fig. 1 shows a normal T2-weighted MRI sagittal view with labels of all discs connected to all lumbar vertebrae.

Whereas some radiologists rely on computed tomography (CT) for diagnosis, the clinical standard among radiologists is to use MRI for diagnosis because MRI provides better resolution images for most spinal abnormalities related to internal structures of the spinal cord such as degenerative disc diseases. Furthermore, MRI is preferred over CT due to patient safety in terms of ionizing radiation. CT is usually preferred when the abnormality includes bone fracture such as vertebral lesions which is out of scope of this research paper [10], [11]. Fig. 2

Manuscript received December 31, 2009; revised March 18, 2010; accepted March 24, 2010. Date of publication April 08, 2010; date of current version December 30, 2010. The work of R. S. Alomari was supported by the University of Jordan, Amman, Jordan. The work of J. J. Corso was supported by NSF CAREER grant IIS 0845282. The work of V. Chaudhary was supported by a grant from the New York State Foundation for Science, Technology and Innovation (NYSTAR). *Asterisk indicates corresponding author.*

*R. S. Alomari is with the Department of Computer Science and Engineering, State University of New York–Buffalo, Buffalo, NY 14260 USA (e-mail: ralomari@buffalo.edu).

J. J. Corso and V. Chaudhary are with the Department of Computer Science and Engineering, State University of New York–Buffalo, Buffalo, NY 14260 USA (e-mail: jcorso@buffalo.edu; vipin@buffalo.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2010.2047403

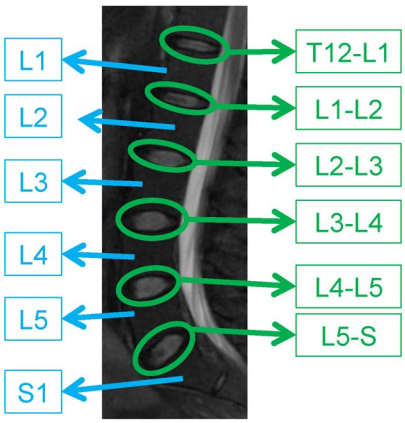

Fig. 1. T2-Weighted MR sagittal view of lumbar region showing lumbar vertebrae (L1-L5) and the six discs connected to them (T12-L1 down to L5-S). The disc L5-S is sometimes called L5-S1.

Type of exam:	MR-Lumbar spine w/o contrast.
Date of birth:	DELETED FOR ANONYMITY.
History:	Backpain.
Findings:	At the L5-S1 level, there is a 6mm right paracentral disc herniation indenting the anterior aspect of the thecal sac and impinging on the right S1 nerve root in the recess. The neural foramina are patent. The L4-5, L3-4, L2-3, L1-2, and T12-L1 levels show no evidence of disc herniation or spinal stenosis. The neural foramina are patent. The conus and cauda equina are normal. The paravertebral soft tissues are normal. No fracture or dislocation is identified.

Fig. 2. Sample portion of the clinical diagnosis report showing the diagnosis at each disc level of lumbar area. Labeling is a necessary step.

shows a sample from a typical clinical report from our dataset. Fig. 3 shows sagittal normal and abnormal MRI cases from our clinical dataset.

The remainder of this paper is organized as follows. The related work is introduced in Section II. Then we present our model in Section III. Section IV presents description of our clinical data and Section V shows our experimental settings and results. Section VI concludes our paper and describes our future directions.

II. Related Work

Many researchers have investigated the problem of labeling the anatomic structures of the vertebral column. However, these studies are limited by their dataset size and thus limit the significance of the validation of the proposed method beside the extensibility of these methods on real clinical datasets.

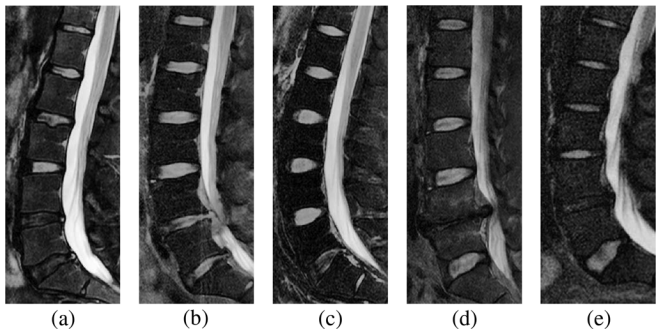

Fig. 3. Variabilities in disc appearances, shapes, locations, and sizes in different abnormal cases. (a) Variability in appearance of discs. The lower two discs ($L4-L5$ and $L5-S$) have less intensity levels due to abnormalities (herniation, stenosis, and desiccation). (b) Variabilities in shape of discs with close intensity levels due to abnormalities in the lower two discs (herniation and stenosis). (c) Clear difference at the lowest disc level ($L5-S$) as well as the difference in bending of the lumbar vertebral column which results in variability in location. (d) Variability in location of discs from other figures, sizes of discs, and the missing disc at $L4-L5$ disc. (e) Variabilities in disc sizes between the upper four discs and the lowest disc $L5-S$. Ages of these patients are 35, 36, 29, 47, and 27, respectively, from (a) to (e). All images have been edited by cropping and contrast enhancement for better visualization.

We point out that researchers have been working on image segmentation using multilayer Markov models such as [12]–[14], and more recently [15], with various applications on grayscale and color images. However, we do not perform segmentation and rather, we localize the discs and label them. In the literature, there is neither a similar model that mimics our model idea for object detection and labeling nor one that has been used for lumbar disc labeling or a similar problem.

In mid-1980s, Jenkins *et al.* [16] concluded that quantitative MR analysis may assist in the diagnosis of intervertebral disc degeneration. They performed a valuable analysis study on 107 normal and 18 abnormal cases. They analyzed the relation between proton density and age in normal discs. Next, we discuss the related work stratified by the driving anatomical structure: vertebrae, soft tissues, and discs.

A. First Trend: Vertebrae

Chwialkowski *et al.* [17] study the localization of discs, vertebrae, and spinal cord in only one MRI case using intensity profiles and edge detectors. Long *et al.* [18] investigate segmentation of the C2 and C3 vertebrae from the cervical area using active shape models [19] as a first step for building image based retrieval system for a dataset consisting of 7000 lumbar X-rays and 10 000 cervical spine X-rays. They have built the Web-based Medical Information Retrieval System (WebMIRS) based on the National Health and Nutrition Examination Surveys (NHANES). Another X-ray-oriented work is presented by de Bruijne *et al.* [20] who use Shape Particle Filter [21] and k-nearest neighbor pixel-level classification for segmenting lumbar vertebrae. However, for this segmentation technique, manual intervention cannot be avoided.

Having a manual step in labeling can be time consuming and will depend on the experience of the user. In some cases, if

manual intervention is restricted to picking a set of points, errors are unlikely to occur and it does not take much time. However, it is desirable to automate the whole process and avoid the manual intervention especially if the size of the workload is large and batches of cases are delivered for labeling.

In MRI modalities, Carballido–Gamio [2] discuss the segmentation of vertebral bodies from sagittal T1-weighted MR images using normalized cuts [22] with the Nyström approximation method [23]. T1-weighted MR images are first preprocessed by Anisotropic Diffusion algorithm [24] that smooths the image without distorting the edges. However, they test their work on only six subjects for lumbar area. Zheng *et al.* [25] propose a method for segmenting the lumbar vertebrae from digital video fluoroscopic images, based on shape descriptors and a Hough transform (relatively high dimension). The method is validated on synthetic data and a single *vivo* sequence. Videofluoroscopic images are noisier than the standard MR radiograph images but have a time component.

B. Second Trend: Soft Tissues

Ghebreab *et al.* [26] present a combination of Strings [27] and Necklaces [28] to model the spine in the lumbar area. The Strings model focuses on learning the most relevant biological variation in the visual appearance of the spine as a whole, and Necklaces aims at exploiting inhomogeneities in multiple continuous shape and gray-level features of vertebrae. Thus they are able to use both *a priori* knowledge about natural variation and anatomical saliency in the visual appearance of the spine. However, they test their method on only six CT images and with minimal spinal and vertebral deformations. Furthermore, manual intervention is used for initialization of their model.

Valdes *et al.* [29] use Kass snakes [30] to segment the trachea[1] from eight CT images. However, they use a Canny filter for initializing snake contours which needs constant data-dependent threshold values. Booth *et al.* [31] also use active contours (snakes) to segment the spinal cord from X-rays, MRI and CT scans of the lumbar area as well as edge detection of the vertebrae bodies. Nyl *et al.* [32] propose a semi-automatic method using 2-D snakes for segmenting the spinal cord in a slice-by-slice manner in 27 CT images for the thorax region. The 3-D volume is then generated by interpolation.

Karangelis *et al.* [33] propose a semi-automatic 3-D segmentation method of the spinal canal and test that on 14 CT volumes. Their method is based on 2-D boundary tracking [34] that starts from an initial seed point and moves horizontally and vertically until facing an edge determined by a specific threshold. Archip *et al.* [35] use knowledge-based techniques and proposed the framework "the plan solver" to detect structures such as spine using other neighboring anatomy structures. They test their work on 23 full volumes of the thorax area.

Recently, Seifert *et al.* [3] have used the Hough transform and knowledge about spine curvature to find initial seed points for discs which are then refined using clustering by considering the center of gravity of the cluster as the disc center. Disc centers are then used to segment the soft tissues from nine T1-weighted MR images of the cervical area.

[1]Trachea is a common term for an airway through which respiratory air passes in organisms.

C. Third Trend: Intervertebral Discs

Peng *et al.* [36] extract intensity profiles of discs and use a convolution operation to match a template of the disc. They perform labeling for discs and vertebrae on only five cases. Another study by Weiss *et al.* [37] proposes a semi-automatic technique for disc labeling. The upper and lower halves of the spine are separately labeled after histogram processing, filters and the use of threshold values. They test their algorithm on 50 MRI cases. Masaki *et al.* [38] propose a method for automated geometry planning based on intensity and a Hough transform to localize the spine and the discs. They use ten MRI normal cases for validation.

A more recent study has been done by Pekar *et al.* [39] for automation of MRI scan planning from the survey (scout) scans for the cervical and lumbar areas. Initially, a set of disc candidates are located by a filter using eigenvalues analysis of the Hessian matrix. Then using prior structural knowledge of the spine, they select the disc centers from the candidates. After that labeling takes place starting from the first spine point and moving upward/downward. They also use a distance constraint for locating the next disc, otherwise a new point is introduced and that disc is considered missing due to abnormality. They use 15 different subjects for validation producing 60 image volumes for lumbar and cervical areas with two poses for each subject.

The Schmidt *et al.* [40] model, which shares mathematical basics with our proposed model [9], proposed a probabilistic model that finds the most probable configuration of part (intervertebral discs) locations $\hat{x}$ among all possible configurations $x = \{x_1, x_2, x_3, \ldots x_{|V|}\}$ given the image $\mathtt{I}$ and a set of parameters Θ where V is the vertices in the graph. Thus, they maximize

$$\hat{x} = \arg\max_{x \epsilon \Omega} P(x|\mathtt{I}, \Theta). \tag{1}$$

Their model incorporates appearance (unary potential) and relative geometry of pairs of parts (which has two versions: 1-D Gaussian for distance and 2-D Gaussian for location) and then uses the A* algorithm, which does a best-first greedy coordinate search for the solution. They use T1-weighted MRI while we use T2-weighed MRI because our empirical analysis indicates that the difference in intensity between discs and vertebrae is clearer in T2-weighted than T1-weighted, especially in degenerated discs [41], [42].

III. Approach

We develop a two level generative probabilistic model that incorporates low- (pixel-) level information such as disc appearance and high- (object-) level information such as location and context. We do full and direct inference using expectation-maximization for localization of the discs of interest.

A. Conventional Labeling Model

One standard probabilistic approach is to formulate the labeling problem with one label per object and do inference by assigning the most probable label to each pixel. The brain structure

labeling work of Fischl *et al.* [43] 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 intensities,[2] $\mathtt{I} : \Lambda \mapsto \mathbb{R}$. Define the set of labels $\mathcal{A} = \{1, \ldots, 7\}$ [there are six discs connected to lumbar vertebrae that take the labels ($\{1, \ldots, 6\}$) besides the last label that means "not-disc"], and a set $\mathcal{T}$ of label variables t_s with one for each pixel $s \in \Lambda$. Then, given an image $\mathtt{I}$, one seeks the maximum a posteriori estimate of the labels

$$\mathcal{T}^* = \arg\max_{\mathcal{T}} P(\mathcal{T}|\mathtt{I}). \tag{2}$$

However, this inference problem is nontrivial; 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.* [40] (but we do not intend to imply they use this conventional labeling approach, as they do not). 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., desiccation). Furthermore, it is difficult to incorporate high- or object-level information (such as disc shape or the relative context of discs) into this formulation since all of the variables are represented at the same scale, the pixel level.

B. Two-Level Model

We instead propose a two-level probabilistic model that only requires conditional independence and adequately separates the localization variables from the pixel intensities while at the same time modeling the exact disc geometry rather than solely pixel-level labels. Let $\mathcal{D} = \{d_i : i \in \{1, 2, \ldots, 6\}\}$ be the set of disc variables with each $d_i = (x_i, y_i)^\mathsf{T}$ representing the disc center i (it 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 by $\mathcal{L} = \{l_s, \forall s \in \Lambda\}$. Each disc-label variable can take a value of $\{-1, +1\}$ for non-disc or disc, respectively. 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 the full set of discs. Note the simpler situation than above where we had a particular label for each disc. Fig. 4 presents and compares the two modeling situations.

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

$$\begin{aligned}\mathcal{D}^* &= \arg\max_{\mathcal{D}} \sum_{\mathcal{L}} P(\mathcal{L}, \mathcal{D}|\mathtt{I}) \\ &= \arg\max_{\mathcal{D}} \sum_{\mathcal{L}} \frac{P(\mathtt{I}|\mathcal{D}, \mathcal{L})P(\mathcal{D}, \mathcal{L})}{P(\mathtt{I})} \\ &= \arg\max_{\mathcal{D}} \sum_{\mathcal{L}} P(\mathtt{I}|\mathcal{L})P(\mathcal{L}|\mathcal{D})P(\mathcal{D}) \end{aligned} \tag{3}$$

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

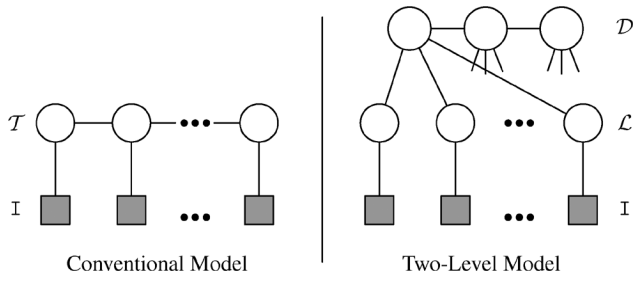

Fig. 4. Graphical models depicting the conventional probabilistic model and our proposed two-level probabilistic model. Our model (right) separates the low- or pixel-level information from the high- or object-level information adding disc localization robustness to the common intensity and structure 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 2-D or 3-D image lattice Λ, but the high level is a single 1-D chain with one node per disc.

where the second equality follows from the multilevel nature of the model (the disc variables are assumed independent of the intensities) and the $P(\mathtt{I})$ in the denominator is independent of the maximization over $\mathcal{D}$ and can be ignored. Note the summation is over a very large set of possible assignments ($2^{|\Lambda|}$). We model it as a Gibbs distribution

$$P(\mathtt{I}|\mathcal{L}) = \frac{1}{Z[\mathtt{I}|\mathcal{L}]} \times \exp\left[-\beta_1 \sum_{s\in\Lambda} U_{\mathrm{I}}(l_s, \mathtt{I}(s))\right] \quad \leftarrow \text{intensity} \tag{4}$$

$$P(\mathcal{L}|\mathcal{D}) = \frac{1}{Z[\mathcal{L}|\mathcal{D}]} \times \exp\left[-\beta_2 \sum_{s\in\Lambda} U_{\mathrm{D}}(l_s, \mathcal{D})\right] \quad \leftarrow \text{spatial} \tag{5}$$

$$P(\mathcal{D}) = \frac{1}{Z[\mathcal{D}]} \exp\left[-\beta_3 \sum_{d_i\in\mathcal{D}} U_{\mathrm{L}}(d_i) \quad \leftarrow \text{location} \right. \\ \left. -\beta_4 \sum_{(i\sim j)} V_{\mathrm{D}}(d_i, d_j)\right] \quad \leftarrow \text{context} \tag{6}$$

where s is a pixel on the lattice Λ ($s \in \Lambda$), $\mathtt{I}(s)$ is the intensity level of the pixel s, $\beta_k \geq 0$, $k = \{1, \ldots, 4\}$ are tunable parameters, the $(\cdot \sim \cdot)$ notation denotes the set of neighboring elements on the disc chain. $Z[\mathtt{I}|\mathcal{L}]$, $Z[\mathcal{L}|\mathcal{D}]$, and $Z[\mathcal{D}]$ are the partition functions that make the normalizing constant for the Gibbs distribution for each model, respectively. The potentials $U_{\mathrm{I}}(l_s, \mathtt{I}(s))$ and $U_{\mathrm{D}}(l_s, \mathcal{D})$ describe the low-level disc intensity and the spatial relationship at the disc level, respectively, while the potentials $U_{\mathrm{L}}(d_i)$ and $V_{\mathrm{D}}(d_i, d_j)$ describe the high-level object location and context, respectively. Both i and j are indexes for the six disc center *xy*-coordinates. All four potentials are discussed in detail in the next two subsections.

The first level, $P(\mathtt{I}|\mathcal{L})P(\mathcal{L}|\mathcal{D})$, 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

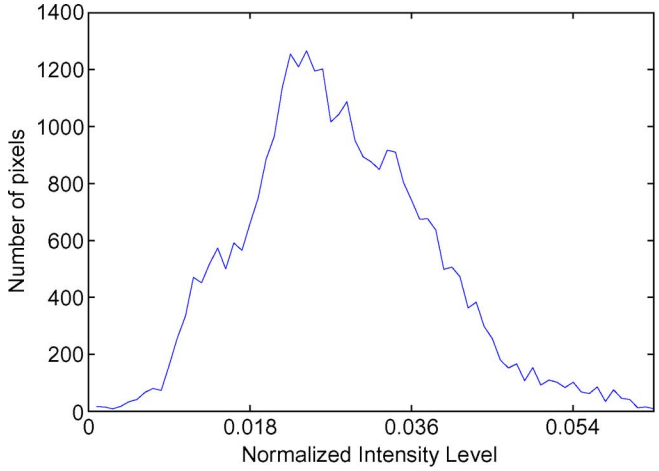

Fig. 5. Empirical distribution at the pixel-level of intensities for normal and abnormal discs in 50 cases.

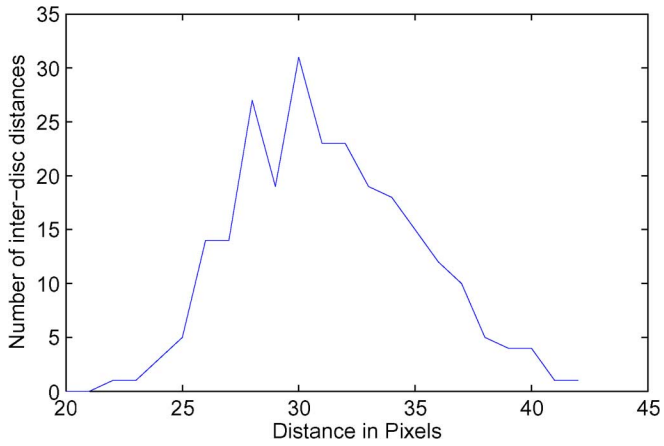

Fig. 6. Empirical distribution of disc distances in pixels for 50 cases.

$P(\mathcal{D})$ models the high-level information about the disc locations and context.

C. Low Level Terms

Intensity: The potential, $U_{\mathrm{I}}(l_s, \mathtt{I}(s))$, models the pixel appearance $\mathtt{I}(s)$ based on its current label l_s. Fig. 5 shows a histogram for all pixels that are labeled as disc pixels. The histogram distribution suggests modeling disc pixels' intensity levels by a Gaussian distribution (discarding the slight heavy-tail). We take the negative-log of the Gaussian for the potential. The parameters of the Gaussian, μ_I and σ_I^2, are learned from labeled training data (using the subscript I does not mean that the Gaussian distribution is image-specific. Rather it is based on intensities of the pixels in the discs. Specific details in experiments Section V). We drop the normalizing term since it does not depend on the specific intensity value

$$U_{\mathrm{I}}\left(l_s, \sigma_{\mathtt{I}(s)}\right) = \begin{cases} \frac{(\mathtt{I}(s)-\mu_{\mathrm{I}})^2}{2\sigma_{\mathrm{I}}^2}, & \text{if } l_s = +1 \\ -\log\left(1 - \exp\left[-\frac{(\mathtt{I}(s)-\mu_{\mathrm{I}})^2}{2\sigma_{\mathrm{I}}^2}\right]\right), & \text{if } l_s = -1 \end{cases} \tag{7}$$

where $s \in \Lambda$. Since there are a (small) finite number of intensities, the second case (with the log) can be precomputed and cached without incurring a large computational burden.

Spatial: The potential $U_{\mathrm{D}}(l_s, \mathcal{D})$ 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 disc variables, i.e., disc centers. 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

$$U_{\mathrm{D}}(l_s, \mathcal{D}) = l_s \cdot \min_{d_i \in \mathcal{D}} \left[(s - d_i)^{\mathsf{T}} \Sigma_i^{-1} (s - d_i)\right] \tag{8}$$

where i is the disc center *xy*-coordinates index and $s \in \Lambda$. The potential assigns energy proportional to the distance of the closest disc, and the label variable acts as a switch, i.e., when $l_s = -1$, being far from the closest disc is lower energy than being closer to it and vice versa. We incorporate shape information by modeling the location of the discs. Generally, discs have elliptical shape which is the motivation behind selecting a 2-D Gaussian for modeling the disc spatial location. We are able to use a simple Gaussian to capture the shape of the discs because of their simple elliptical nature. We find that this model is sufficient for our datasets.

D. High Level Terms

Location: The potential $U_{\mathrm{L}}(d_i)$ measures the distance of disc d_i to its expected location. Because our data are clinical, patients are aligned in a certain way, by the MRI technician, with respect to the MRI machine. Thus, it is a valid assumption to assume that images are approximately aligned with respect to scaling, rotation, and translation. Similar to the spatial low level term in (8), 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

$$U_{\mathrm{L}}(d_i) = (d_i - \mu_i)^{\mathsf{T}} \Sigma_i^{-1} (d_i - \mu_i) \tag{9}$$

where i is an index for discs centers *xy*-coordinates.

Disc-Context: The potential $V_{\mathrm{D}}(d_i, d_j)$ captures the high-level contextual relationship between two *neighboring* discs. Considering the discs as 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 (Fig. 6) suggests a Gaussian parameterized by μ_D, σ_D^2, and its negative log for the energy. Let

$$e_{ij} = \|d_i - d_j\|_2. \tag{10}$$

Then

$$V_{\mathrm{D}}(d_i, d_j) = \frac{(e_{ij} - \mu_D)^2}{\sigma_D^2}. \tag{11}$$

E. Inference Using gEM

The exact inference is infeasible for our model because of the dependencies of $\mathcal{D}$ on all $\mathcal{L}$ despite that $\mathcal{D}$ is a Markov chain.

1. Initialize each $d_i = \mu_i$ for all $\mathcal{D}$.
2. for (Iter1 = 0; Iter1 < globalInterations; Iter1++){
3. - for each disc $d_i \in \mathcal{D}$
4. a. Evaluate Expectation (Eq.12)
5. b. for (Iter2=0; Iter2 < localIterations; Iter2++){
6. - Evaluate (13) for current d_i.
7. - Evaluate (13) for each direction.
8. - Update d_i to max (13).}
9. c. If d_i did not change, stop.}
10. d_i are the disc locations.

Fig. 7. Pseudo Code for the general expectation maximization implementation for the Inference [(12) and (13)].

We use the generalized expectation maximization (gEM) algorithm [44] to optimize (3). Whereas an EM algorithm requires maximization in the M step, a generalized EM algorithm simply requires an improvement over the current state. However, both are guaranteed to converge [44].

We initialize $\mathcal{D}^0$ by setting each disc d_i in its mean location μ_i, which we have estimated from the 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)

$$\mathbf{E-step} \rightarrow F^t(\mathcal{L}) = P(\mathcal{L}|\mathcal{D}^t, \mathtt{I}) \tag{12}$$

$$\mathbf{M-step} \rightarrow \mathcal{D}^{t+1} = \arg\max_{\mathcal{D}} \left[\log P(\mathcal{D}) + \sum_{\mathcal{L}} F^t(\mathcal{L}) \log P(\mathcal{L}|\mathcal{D}^t, \mathtt{I}) \right] \tag{13}$$

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 $\mathcal{L}$ as we assume that disc labels $\mathcal{L}$ are conditionally independent given the discs $\mathcal{D}$, we factor them into independent local terms

$$P(\mathcal{L}|\mathcal{D}, \mathtt{I}) = \prod_{s \in \Lambda} \frac{1}{Z[l_s]} \exp\left[-\beta_1 U_{\mathrm{I}}(l_s, I(s)) - \beta_2 U_{\mathrm{D}}(l_s, \mathcal{D})\right]. \tag{14}$$

Thus, we directly evaluate the full posterior (12) and log terms (13) during optimization. Since the partials are not analytically available, we execute a finite differences-based gradient ascent algorithm to iteratively maximize the ELL.

Our implementation of the inference (Fig. 7) starts by initializing the disc variables d_i with the learned mean locations μ_i (Line 1). We then optimize $d_i \in \mathcal{D}$ simultaneously where the coordinates chain moves reserving context constraints. Then, iteratively, for each disc variable d_i, we evaluate the local gradient of the ELL by perturbing d_i by a set of changes $\{\delta\}$ in a set of directions (four directions as shown in Line 7). For each perturbation, we fully evaluate (13) 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 (Line 10). We fixate δ to two pixels steps.

IV. Clinical Data Description

Our available dataset is clinical MRI that we obtained from our collaborative radiologist group. Currently, it contains 25 normal cases and 80 abnormal cases. The images have been acquired on a 3T Philips Medical Systems Intera Scanner. Each sagittal view includes at least the six discs connected to the lumbar vertebrae. The total number of discs connected to lumbar vertebrae is $105 \times 6 = 630$ discs. Among them 371 are normal discs and the rest (259 discs) are abnormal. Abnormalities include herniation, desiccation, spondylolistheses, and many others.

Each case contains five, six, or seven acquisition protocols. Most of the cases contain: 1) T1-weighted sagittal, 2) T2-weighted sagittal, 3) T2-SPIR weighted sagittal, 4) T2-weighted axial for a set of selected discs, and 5) Myelo MR images (usually six slices). Sagittal views have 4.5 mm thickness and 5.0 mm separation. Axial views are acquired for each selected disc (not all discs have axial views). The radiologist decides which discs should be captured with full axial scans depending on his judgment for abnormality. All images are 512×512 pixels.

V. Experimental Results

We build our model on T2-weighted SPIR modality as disc intensities have better discrimination from other structures in the image [41], [42]. However, the intensity signal is affected by the MRI protocol as well as other reasons such as magnetic field inhomogeneities. Thus radiologists measure the disc signal with respect to an adjacent intra-body [45], [46]. Cerebrospinal fluid (CSF) is usually the standard reference for this purpose in the lumbar spine [42], [47], [48]. Recently, the spine cord signal is suggested for this purpose as well [49]. We use the spinal cord signal for this purpose.

Both T1-weighted and T2-weighted are useful for performing pre-processing steps to enhance performance of our method. For example, the spinal cord can be suppressed in most cases just by using a subtraction operation between the normalized T1-weighted and T2-weighted images. However, in this work, we directly use our model on the T2-SPIR images because we do not assume the availability of various registered protocols in other datasets [9].

We perform ground truth annotation for our dataset by specifying three things: 1) selecting a contour for every disc, 2) selecting a point inside every disc that roughly represents the center for that disc (nonuniform shape), and 3) determining whether the disc is normal or abnormal.

To train our model, we use the disc contours and the disc centers to infer our model parameters. Using disc contours, we extract intensity values for the disc pixels (these pixels inside the disc). Fig. 5 shows a histogram for all disc pixel intensities (both normal and abnormal). We created histograms for both normal and abnormal discs to study the necessity of building a separate model for each. However, we found that the distribution difference is insignificant and we justify that by the type of abnormalities we have in our dataset. Desiccation and herniation, for example, cause noticeable reduction in disc intensity levels. However, most other abnormalities are not related to drying out of the disc water contents which means that the disc intensity

TABLE I
SUMMARY OF EXPERIMENTS

Experiment	Training size	Testing size	Description
1	50	55	Normal and Abnormal
2	30	50	Abnormal

TABLE II
NUMBER OF CORRECTLY LOCALIZED AND LABELED DISCS FOR BOTH EXPERIMENTS. RESULTS OF LOCALIZATION ON THE TRAINING SET (SECOND COLUMN) AND THE TESTING SET (THIRD COLUMN) OF THE FIRST EXPERIMENT. FOURTH COLUMN SHOWS THE RESULTS OF LOCALIZATION ON THE TESTING SET IN THE SECOND EXPERIMENT

Subset	Experiment 1		Experiment 2
	Training set (50)	Testing set (55)	Testing set (50)
1	29	29	28
2	29	25	27
3	28	25	26
4	29	26	25
5	26	28	26
6	26	28	24
7	27	24	25
8	27	27	27
9	24	29	28
10	27	25	27
11	-	28	-
Accuracy	**90.7 %**	**89.1 %**	**87.7 %**

levels are not affected. Furthermore, we use disc centers to compute both spatial model distribution and disc context distribution parameters. Fig. 6 shows a histogram for disc center distances for both normal and abnormal cases in our dataset.

We validate our model on one dataset that contains 105 cases: 25 normal cases and 80 abnormal cases. In Table I, we show a summary of the experiments we perform and the amount of cases we use in each.

In the first experiment (second and third columns in Table II), we train our model on 50 cases. We report a quantitative measure for accuracy of testing on the training set itself (second column in Table II) and on the rest of the cases (the remaining 55 cases) in the third column in Table II. This quantitative measure counts the number of correctly localized discs. In Table II, each row represents the results of five cases by summing up the number of correctly localized discs (5 cases $\times$ 6 discs per case $=$ 30 discs for each row). We consider that the disc is correctly localized if the resulting points lie inside the disc by visually inspecting all the resulting images. We report both training and testing results to show that our model did not over fit on the training data. The accuracy is comparable between the training and the testing set which is around 90% (i.e., we are not overfitting). Because the disc contours are manually drawn for the nonuniform disc shapes, especially in abnormal cases, we do not use it for measuring our localization accuracy by the idea of automatically checking if the converged point is inside this contour. We find that if we do so, errors in accuracy reporting appear due to the nonuniformity of the manually-drawn disc contour. Thus we do visual inspection of each converged point inside the disc and decided whether it is accurately localized or not.

In Table II, each row represents the sum of accuracies in randomly selected five cases. Each case can only be selected once. This brings the summation of all cases to 50 in first and third columns and to 55 in the second column. The selection of the cases in each subset does not change the average accuracy because we randomly do so. The reported accuracy for all subset lies within a small interval of 24/30 and 29/30. Thus, our reported average accuracy of 91%, 89.1%, and 87.7% represents the expected localization accuracy of a randomly selected case.

We also report the results with a second accuracy measure that is based on the Euclidean distances E_i between the ground truth disc center g_i and the detected disc center from our model d_i where E_6 represents the Euclidean distance at the L5-S1 level (Fig. 1), E_5 represents L4-L5, E_4 represents L3-L4, E_3 represents L2-L3, E_2 represents L1-L2, and E_1 represents T12-L1

$$E_i = \frac{1}{5}\sum_{c=1}^{5}\left\| g_i^{(c)} - d_i^{(c)} \right\|_2 \tag{15}$$

where $i \in \{1,\ldots,6\}$ is an index for the discs' centers xy-coordinates, c is a counter for the randomly selected five cases, with each denoted $g_i^{(c)}$ and $d_i^{(c)}$, that we use for the average calculation of the Euclidean distances at each disc level i, $d_i = (x_i, y_i)^{\mathsf{T}}$ represents the disc center coordinates resulting from the automatic convergence while $g_i = (x_i, y_i)^{\mathsf{T}}$ represents the gold standard for the disc center.

We report distances in millimeters by converting from the resulting distances in pixels. In our dataset, a pixel is 0.5273×0.5273 mm^2. The automatic localization point might go in any direction with respect to the gold standard point. We, however, do not keep track of the direction (anterior/posterior) as far as the point is inside the disc boundary.

Fig. 8 shows the summary of statistics of the Euclidean distances over the testing cases of the first experiment. The line in each box is the median, the top and bottom are the 75th and 25th percentiles, and pluses are statistical outliers. We define any values larger than 1.5 times the inter-quartile range from both ends of the box as statistical outliers. We note that the lower discs (L5-S1, L4-L5) are more difficult than the upper discs which we believe due to the larger variability in the lower part of the lumbar vertebral column in terms of location and appearance. Furthermore, the lower lumbar area has more possibility for abnormality than the upper lumbar area. Over 80% of the abnormal cases have the abnormality at the lower one or two levels in our dataset.

We count the number of normal and abnormal discs that are incorrectly labeled to see if there is any bias in accuracy of labeling normal or abnormal discs. In the first experiment, incorrectly labeled discs are around 45% normal and 55% abnormal for training and testing sets. Thus, we believe that the abnormality condition of discs has insignificant effect on labeling accuracy. Table III shows the summary of these numbers for the first experiment.

To test the sensitivity of our method to the selection of the training set, we repeated experiment 1 again with a different random subset of 50 cases. In this cross-validation-like experiment, we again achieve 90.4% and 89.7% localization accuracy on the training and testing sets, respectively. Our localization accuracy demonstrates the robustness of our method to the selection of the training set.

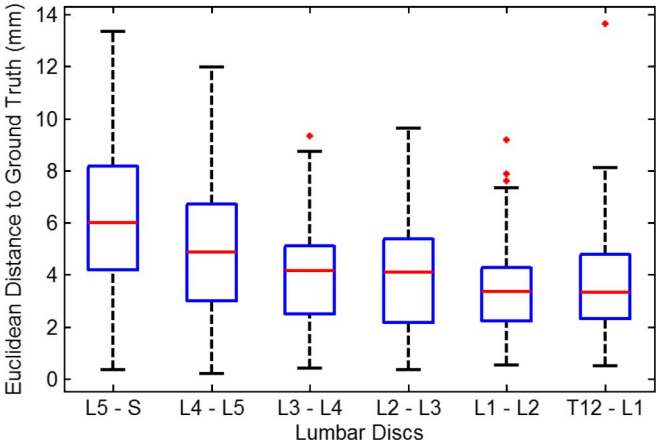

Fig. 8. Box-plot showing the summary statistics of Euclidean distances over the testing set of the first experiment. The line in each box is the median, the top and bottom lines are the 75th and 25th percentiles, and the pluses are statistical outliers.

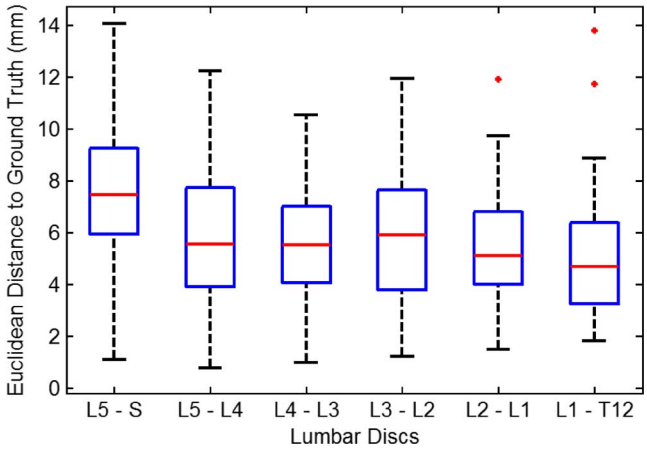

Fig. 9. Box-plot showing the summary statistics of Euclidean distances over the testing set of the second experiment. The line in each box is the median, the top and bottom lines are the 75th and 25th percentiles, and the pluses are statistical outliers.

TABLE III
STUDY OF INCORRECTLY LABELED DISCS FROM THE FIRST EXPERIMENT ON BOTH TRAINING AND TESTING. THIRD COLUMN IS THE COUNT OF INCORRECTLY LABELED DISCS

Experiment 1 data	Size	Count	Percent Normal	Percent Abnormal
Training	50	28	46 %	54%
Testing	55	36	44 %	56 %

In the second experiment, we use 80 abnormal cases: 30 cases for training and the remaining 50 for testing. We perform localization for the testing set only which shows an accuracy over 87% as shown in the last column of Table II. Fig. 9 shows the summary of statistics of Euclidean distances. Similar to the box plot of the first experiment, whiskers extend from each end of the box to the most extreme values within 1.5 times the inter-quartile range from the ends of the box. Thus, all Euclidean distances larger than 1.5 times the inter-quartile range are considered statistical outliers. By examining both box plots, we notice strong consistency in our labeling results. Both box plots show similar ranges of Euclidean distances. All discs show small change in Euclidean distances in the box plot of the second experiment compared to the first one. However, the second experiment contains only abnormal cases while the first one has both normal and abnormal cases. Furthermore, having similar Euclidean distance ranges between both experiments demonstrate good degree of robustness of our model in labeling and its extensibility to wide range of abnormal cases.

Fig. 10 shows the labeling result in four different cases: The first two cases (top two cases) show a full success in localization for a normal case (top-right) and an abnormal case (top-left) at levels L5-S1 and L4-L5. The third case (bottom-left) has an abnormality at the L4-L5 level. However, our model is able to localize this disc because of the contextual model. The fourth case (bottom-right) has all the six discs as abnormal. This reduces the effect of disc intensity model and thus having unsuccessful localization at three (L1-L2, L2-L3, and L3-L4) of the six discs. Thus we are working on adding more global aspects to our model to capture these cases where abnormality exist in most discs because the contextual model that works in the bottom-left example could not suffice for convergence in the bottom-right figure.

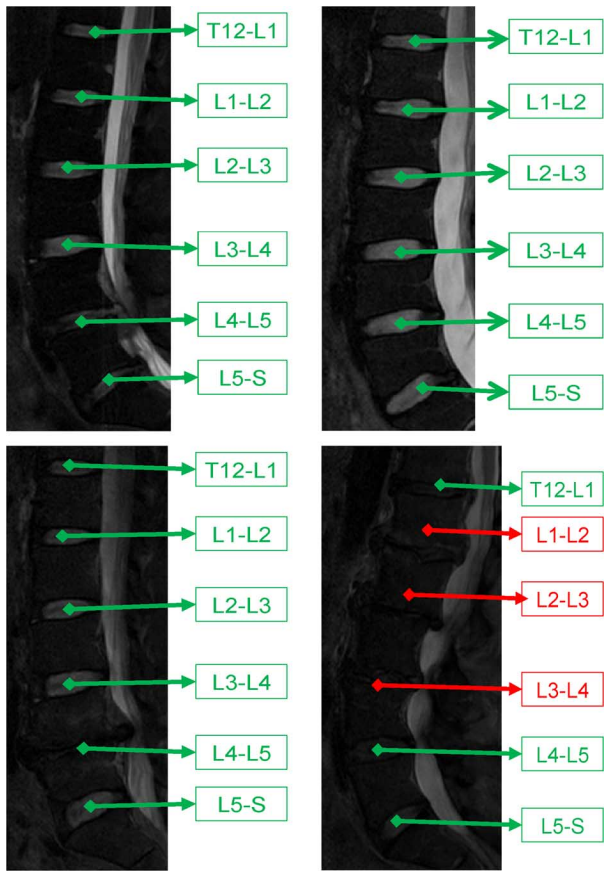

Fig. 10. Labeling results: The top two cases show normal (right) and abnormal (left) cases where our model successfully localize the discs. The bottom left figure shows an abnormal case where the context model succeeds in localizing disc L4-L5 despite its high level of abnormality. The bottom right case shows unsuccessful localization of the discs L1-L2, L2-L3, and L3-L4 due to abnormalities in all discs as well as the missing discs.

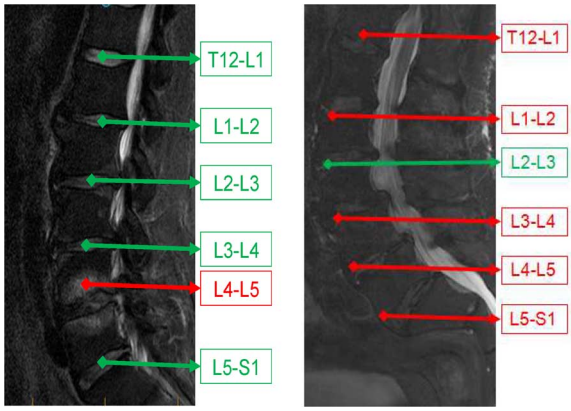

Fig. 11. Unsuccessful localization examples: (Left) vertebrae and disc abnormality causing high intensity level at both L4 and L5 vertebrae. (Right) discs intensity are indistinguishable from vertebrae.

A. *Failure Modes*

We show two examples where our labeling model does not perform accurate labeling. Most cases that failed the localization have either many severely abnormal discs as shown in Fig. 11 (right) or unusual intensity profile for the vertebrae due to specific abnormality conditions such as Fig. 11 (left). In Fig. 11 (right), the discs have indistinguishable intensity levels with the vertebrae which lead to failure of the intensity-based model. Moreover, the locations of the discs have major effect due to the high abnormality level of these cases. In Fig. 11 (left), both L4 and L5 vertebrae have high signal intensity surrounding the L4-L5 disc due to abnormality in the bone and the disc. This caused the intensity model to trap into the high intensity and consider that as the disc. However, these cases are very rare in the clinical settings. Usually these patients have special circumstances including immobility and high levels of back bending. These cases require special attention by the radiologist. We believe that patient history information is the key for accurate labeling of these rare cases.

VI. Conclusion and Future Work

We proposed a robust and accurate generative two-level probabilistic model for localization of intervertebral discs from T2-Weighted sagittal MRI. We showed that our model robustly localizes intervertebral discs from clinical data. The uniqueness of our model is that we include two levels of information in a generative transparent matter. The two levels are: 1) the pixel- (low-) level that includes appearance and shape, and 2) the object- (high-) level that includes location and context. Our stratification lets us assume only conditional independence at the pixel level to increase efficiency and maintain robustness rather than the more common full independence.

Appearance (Intensity levels) of discs are fairly distinguished from surrounding vertebrae in T2-SPIR images. We performed an empirical study for manually specified disc pixels and concluded that they have a Gaussian distribution (Fig. 6). We capture shape information of the discs by assuming the elliptical shape of the disc. The Mahalanobis distance in our Gibbs model performs this task. Normal and most abnormal cases in our dataset have elliptical shape. However, some abnormalities do not obey this rule and thus this is not a sufficient information for determining the shape of the disc which motivates why we needed other information (appearance) at low level in our model.

At the high level, we express our discs as objects represented by coordinates. The location information of the discs measures the distance to the expected location which is learned offline in our model. The context information represents the relative distances between discs. Discs have a general relation in a virtually connected spine which is motivated by our empirical study of the distances (Fig. 6).

We validated our model on 105 cases that include normal and abnormal cases. We tested our model in two different experimental settings: using the whole dataset and using only abnormal cases. When using the whole dataset, we perform model training on 50 cases and then perform localization for both the training and the testing set (test set is the rest 55 cases). We obtained over 91% as a general accuracy for our quantitative accuracy measure on the training set and a similar accuracy (over 89%) for the testing set. When using only the abnormal cases, we trained our model on 30 cases and performed localization on 50 abnormal cases and we obtained over 87% acuracy which shows the robustness of our model for accurate localization.

We are working on performing quantitative analysis and abnormality detection for intervertebral discs after this localization step. We are working on two levels: a local level where abnormalities are local at every disc such as desiccation, and a global level where abnormalities are global for the whole spine such as spinal scoliosis.

Acknowledgment

The authors would like to thank Dr. G. Dhillon, MD, from ProScan Imaging Inc., Buffalo, NY, for his interest and support of the anonymized clinical dataset.

References

[1] National Institute of Neurological Disorders and Stroke (NINDS), "Low back pain fact sheet," NIND Brochure 2008 [Online]. Available: http://www.ninds.nih.gov/disorders/backpain

[2] J. Carballido-Gamio, S. Belongie, and S. Majumdar, "Normalized cuts in 3-D for spinal MRI segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 1, pp. 36–44, Jan. 2004.

[3] S. Seifert, I. Waechter, G. Schmelzle, and R. Dillmann, "A knowledge-based approach to soft tissue reconstruction of the cervical spine," *IEEE Trans. Med. Imag.*, vol. 28, no. 4, pp. 494–507, Apr. 2009.

[4] R. S. Alomari, J. J. Corso, and V. Chaudhary, "Abnormality detection in lumbar discs from clinical MR images with a probabilistic model," presented at the Comput. Assist. Radiol. Surg. Conf., Berlin, Germany, Jun. 2009.

[5] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Desiccation diagnosis in lumbar discs from clinical MRI with a probabilistic model," in *Proc. 6th IEEE Int. Conf. Symp. Biomed. Imag.: From Nano to Macro (ISBI)*, Boston, MA, Jun. 2009, pp. 546–549.

[6] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 5, no. 3, pp. 287–293, May 2010.

[7] R. S. Alomari, J. J. Corso, V. Chaudhary, and G. Dhillon, "Automatic diagnosis of disc herniation with shape and appearance features from MRI," in *Proc. SPIE*, Feb. 2010, vol. 7624, pp. 76241A–76241A-9.

[8] A. F. Dalley and A. M. R. Agur, *Atlas of Anatomy*. New York: Lippincott Williams and Wilkins, 2004.

[9] J. J. Corso, R. S. Alomari, V. Chaudhary, and G. Dhillon, "Lumbar disc localization and labeling with a probabilistic model on both pixel and object features," in *Proc. Med. Image Computing Computer Assist. Intervent. (MICCAI)*, 2008, vol. 5241, LNCS Part 1, pp. 202–210.

[10] I. Beggs and J. Addison, "Posterior vertebral rim fractures," *Br. J. Radiol.*, vol. 71, no. 845, pp. 567–572, 1998.

[11] E. Ne, "Lumbar surgery for 56 limbus fractures emphasizing noncalcified type iii lesions," *Spine*, vol. 17, no. 12, pp. 1489–1496, 1992.

[12] R. Hu and M. M. Fahmy, "Texture segmentation based on a hierarchical Markov random field model," in *Proc. IEEE Symp. Circuits Syst.*, 1991, vol. 1, pp. 512–515.

[13] C. D'Elia, G. Poggi, and G. Scarpa, "A tree-structured markov random field model for bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1259–1273, Oct. 2003.

[14] J. Sun and D. Gu, "Bayesian image segmentation based on an inhomogeneous hidden Markov random field," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR'04)*, 2004, vol. 1, pp. 596–599.

[15] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical multiple Markov chain model for unsupervised texture segmentation," *IEEE Trans. Image Process.*, vol. 18, pp. 1830–1843, Aug. 2009.

[16] J. P. Jenkins, D. S. Hickey, X. P. Zhu, M. Machin, and I. Isherwood, "MR imaging of the intervertebral disc: A quantitative study," *Br. J. Radiol.*, vol. 58, no. 692, pp. 705–709, 1985.

[17] M. P. Chwialkowski, P. E. Shile, R. M. Peshock, D. Pfeifer, and R. W. Parkey, "Automated detection and evaluation of lumbar discs in MR images," in *Proc. IEEE EMBS*, 1989, vol. 2, pp. 571–572.

[18] L. R. Long and G. R. Thoma, "Use of shape models to search digitized spine X-rays," in *Proc. 13th IEEE Symp. Computer-Based Med. Syst, (CBMS)*, Jun. 2000, pp. 255–260.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[20] M. de Bruijne and M. Nielsen, "Image segmentation by shape particle filtering," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, 2004, pp. 722–725.

[21] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[22] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[23] C. Fowlkes, Q. Shan, S. Belongie, and J. Malik, "Extracting global structure from gene expression profiles," *Methods Microarray Data Anal. II*, pp. 888–905, 2002.

[24] P. Perona, T. Shiota, and J. Malik, *Geometry-Driven Diffusion in Computer Vision*. New York: Springer, 1994.

[25] Y. Zheng, M. S. Nixon, and R. Allen, "Automatic segmentation of lumbar vertebrae in digital videofluoroscopic imaging," *IEEE Trans. Med. Imag.*, vol. 23, no. 1, pp. 45–52, Jan. 2004.

[26] S. Ghebreab and A. W. M. Smeulders, "Combining strings and necklaces for interactive three-dimensional segmentation of spinal images using an integral deformable spine model," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 10, pp. 1821–1829, Oct. 2004.

[27] S. Ghebreab and A. W. Smeulders, "Strings: Variational deformable models of multivariate ordered features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1399–1410, Nov. 2003.

[28] S. Ghebreab, P. R. Pfluger, and A. W. M. Smeulders, "Necklaces: Inhomogeneous and point-enhanced deformable models," *Comput. Vis. Image Understand.*, pp. 96–117, 2002.

[29] R. Valdes, O. Yanez-Suarez, and V. Medina, "Trachea segmentation in CT images using active contours," in *Proc. 22nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2000, pp. 3184–3187.

[30] M. Kass, A. Wittkin, and D. Terzopoulos, "Snakes, active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1987.

[31] S. Booth and D. A. Clausi, "Image segmentation using MRI vertebral cross-sections," in *Proc. Canadian Conf. Electrical Comput. Eng.*, May 2001, vol. 2, pp. 1303–1307.

[32] L. G. Nyl, J. Kany, E. Mt, G. Makay, E. Balogh, M. Fidrich, and A. Kuba, "Method for automatically segmenting the spinal cord and canal from 3-D CT images," in *Computer Analysis of Images and Patterns*. New York: Springer, 2005, vol. 3691, pp. 456–463.

[33] G. Karangelis and S. Zimeras, "An accurate 3-D segmentation method of the spinal canal applied to CT data," Bildverarbeitung für die Medizin 2002, pp. 370–373.

[34] H. R. M. and S. L. G. , "Image segmentation techniques," *Comput. Vis., Graphics, Image Process.*, pp. 100–132, 1985.

[35] N. 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 Trans. Med. Imag.*, vol. 21, no. 9, pp. 1504–1516, Sep. 2002.

[36] Z. Peng, J. Zhong, W. Wee, and J. Lee, "Automated vertebra detection and segmentation from the whole spine MR images," presented at the IEEE EMBS Conf., Sep. 1–4, 2005.

[37] K. L. Weiss, J. M. Storrs, and R. B. Banto, "Automated spine survey iterative scan technique," *Radiology*, vol. 239, no. 1, pp. 255–262, 2006.

[38] T. Masaki, Y. Lee, D. Y. Tsai, M. Sekiya, and K. Kazama, "Automatic determination of the imaging plane in lumbar MRI," in *Proc. SPIE Med. Imag.*, 2006, pp. 1252–1259.

[39] V. Pekar, D. Bystrov, H. S. Heese, S. P. M. Dries, S. Schmidt, R. Grewer, C. Harder, R. C. Bergmans, A. W. Simonetti, and A. Muisinkel, "Automated planning of scan geometries in spine MRI scans," in *Proc. MICCAI*. New York: Springer, 2007, vol. 4791, Lecture Notes Computer Science, pp. 601–608.

[40] S. Schmidt, J. Kappes, M. Bergtholdt, V. Pekar, S. P. Dries, D. Bystrov, and C. Schnoerr, "Spine detection and labeling using a parts-based graphical model," in *Proceedings of the IPMI*. : Springer, 2007, vol. 4584, Lecture Notes Computer Science, pp. 122–133.

[41] H. S. An, P. A. Anderson, V. M. Haughton, J. C. Iatridis, J. D. Kang, J. C. Lotz, R. N. Natarajan, T. R. Oegema, P. Roughley, L. A. Setton, J. P. Urban, T. Videman, G. B. Andersson, and J. N. Weinstein, "Disc degeneration: Summary," *Spine*, vol. 29, pp. 2677–2678, Dec. 2004.

[42] T. Videman, P. Nummi, M. Battie, and K. Gill, "Digital assessment of MRI for lumbar disc desiccation: A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings," *Spine*, vol. 19, pp. 192–198, 1994.

[43] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Deiterich, C. Haselgrove, A. Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Monttillo, N. Makris, B. Rosen, and A. M. Dale, "Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, pp. 341–355, 2002.

[44] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[45] K. Luoma, R. Raininko, P. Nummi, and R. Luukkonen, "Is the signal intensity of cerebrospinal fluid constant? Intensity measurements with high and low field magnetic resonance imagers," *Magn. Reson. Imag.*, vol. 11, pp. 549–555, 1993.

[46] E. K. Luoma, R. Raininko, P. J. Nummi, R. Luukkonen, H. I. Manninen, and H. A. Riihimäki, "Suitability of cerebrospinal fluid as a signal-intensity reference on MRI: Evaluation of signal-intensity variations in the lumbosacral dural sac," *Neuroradiology*, vol. 39, pp. 728–732, Oct. 1997.

[47] T. Videman, M. Battié, L. E. Gibbons, K. Maravilla, and H. Manninen, "Associations between back pain history and lumbar MRI findings," *Spine*, vol. 28, pp. 582–588, 2003.

[48] T. Videman, M. Battié, S. Ripatti, K. Gill, H. Manninen, and J. Kaprio, "Determinants of the progression in lumbar degeneration: A 5-year follow-up study of adult male monozygotic twins," *Spine*, vol. 31, pp. 671–678, 2006.

[49] R. Niemeläinen, T. Videman, S. Dhillon, and M. Battié, "Quantitative measurement of intervertebral disc signal using MRI," *Clin. Radiol.*, vol. 63, no. 3, pp. 252–255, 2008.