

# Simultaneous Localization of Lumbar Vertebrae and Intervertebral Discs with SVM based MRF

Ayşe Betül Oktay and Yusuf Sinan Akgül

**Abstract**—This paper presents a method for localizing and labeling the lumbar vertebrae and intervertebral discs in mid-sagittal MR image slices. The approach is based on a Markov-chain-like graphical model of the ordered discs and vertebrae in the lumbar spine. The graphical model is formulated by combining local image features and semi-global geometrical information. The local image features are extracted from the image by employing Pyramidal Histogram of Oriented Gradients (PHOG) and a novel descriptor that we call image projection descriptor (IPD). These features are trained with Support Vector Machines (SVM) and each pixel in the target image is locally assigned a score. These local scores are combined with the semi-global geometrical information like the distance ratio and angle between the neighboring structures under the Markov Random Field (MRF) framework. An exact localization of discs and vertebrae is inferred from the MRF by finding a maximum a posteriori solution efficiently by using dynamic programming. As a result of the novel features introduced, our system can scale-invariantly localize discs and vertebra at the same time even in the existence of missing structures. The proposed system is tested and validated on a clinical lumbar spine MR image dataset containing 80 subjects of which 64 have disc and vertebra related diseases and abnormalities. The experiments show that our system is successful even in abnormal cases and our results are comparable to the state of the art.

**Index Terms**—Lumbar vertebrae, intervertebral disc, labeling, Markov Random Field, Pyramidal Histogram of Oriented Gradients, Support Vector Machines

## I. INTRODUCTION

The vertebral column, or the spine, is the main axis of the human body and it is crucial for supporting the weight of the torso and protecting the spinal cord. The spine is usually divided into four curves: cervical curve, thoracic curve, lumbar curve, and sacral curve. Lower back pain, which is one of the most common pain types, is usually caused by the lumbar region [1]. Lumbar region is commonly viewed by X-Rays, Magnetic Resonance (MR) imaging, and computed tomography (CT) modalities. In MR images soft tissues are shown better than X-Rays and CT [2].

In clinical practice, a crucial step in the analysis, diagnosis, and application tasks of the vertebral column is the localization and labeling of vertebrae and discs. The radiologists report the diagnosis after labeling the vertebrae and the discs [3].

Ayşe Betül Oktay is with the Department of Computer Engineering, Istanbul Medeniyet University, Istanbul, Turkey. (e-mail: abetul.oktay@medeniyet.edu.tr)

Yusuf Sinan Akgül is with the Vision Lab, Department of Computer Engineering, Gebze Institute of Technology, Gebze, Kocaeli 41400 Turkey. (e-mail: akgul@bilmuh.gyte.edu.tr.)

For vision lab see <http://vision.gyte.edu.tr/>.

The dataset used in this study can be downloaded from the multimedia section of the journal and <http://vision.gyte.edu.tr/lumbarmri/>

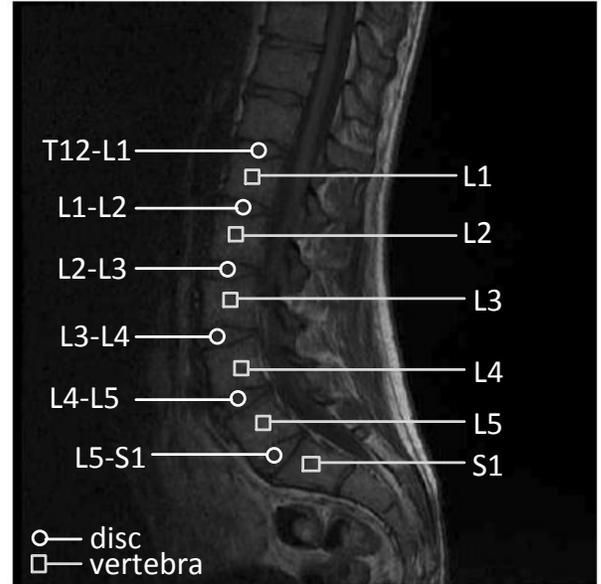


Fig. 1. A mid-sagittal view T1-weighted MR image of the lumbar vertebrae and intervertebral discs. L1, L2, L3, L4, and L5 are the lumbar vertebrae. S1 is the first vertebra of the sacral curve and T12 is the last vertebra of the thoracic curve. The intervertebral disc between L1 and L2 is labeled by L1-L2.

Similarly, the manually labeled vertebral structures are used in some orthopedic and neurological applications and surgeries [4]. A system for automatically performing various tasks about vertebral column needs the accurate localization and labeling of the vertebral structures. For example, localization and labeling are necessary for automatic positioning of volume of interest for the MR spectroscopy of the lumbar structures. Therefore, the core requirement of building a computer based system for spine is automatically labeling of vertebrae and discs.

In this paper, we propose a novel method for automatic localization and labeling of the lumbar vertebrae and discs from the 2D sagittal MR images (Figure 1) as an extension of our previous work [5]. Our method has two main steps (Figure 2). First, the image features are extracted by using Pyramidal Histogram of Oriented Gradients (PHOG) and a novel method that we call *Image Projection Descriptor* (IPD). These features are employed by Support Vector Machines (SVM) to produce candidate disc and vertebra positions with their scores. The first step basically uses the local image gradient information of each vertebra and disc and it locally searches for the candidate structure positions. The second step

takes the advantage of the Markov-chain-like structure of the spine by assuming latent variables for the disc and vertebra positions. An exact inference from the chain-like graphical model based on dynamic programming is employed to find the final lumbar disc and vertebra positions and labels. The second step uses semi-global geometric information of the spine like the angles and distance ratios between the discs and vertebrae.

We call the first and second steps image inference and positional inference steps, respectively because the first step locally infers the candidate structure positions from the image and the second step infers the final localizations from the candidate structure positions. The positional inference is performed at a more abstract level and it is isolated from direct image influence which brings many advantages for our system. First, the separation of the image inference from the positional inference makes our system flexible and modular. A modification in one step does not affect the other step and our method can be conveniently applied to other medical applications like localization of the teeth from panoramic X-ray images [6] with small modifications. Second, at the positional inference step, it is more convenient to capture the one-dimensional nature of the spine with a Markov-chain-like model which makes it possible to infer a globally optimal solution by an efficient polynomial time algorithm. By defining factors on triples within the chain-like structure, it is feasible to make the inference system scale-invariant and image position independent. The other advantages of the positional inference includes handling of the missing discs and vertebrae and simultaneous localization of these structures.

The rest of this paper is organized as follows. In Section II, the related work is presented. The image inference step is described in Section III. Section IV includes the positional inference step with the graphical model and exact inference. The validation of the method is presented in Section V. Finally, we provide concluding remarks in Section VI.

## II. RELATED WORK

There are many image analysis studies about spinal column including vertebra and disc detection, segmentation, and labeling [7], [8]. Peng et al. [9] use intensity profiles to localize the 24 articulated vertebrae. Huang et al. [10] detect and segment vertebra with wavelet transform based Adaboost and iterative normalized cut. An automatic approach for the segmentation and reconstruction of intervertebral discs from peripheral quantitative CT images is introduced in [11]. Donner et al. [12] propose an MRF method that uses symmetry based interest points as the nodes of a graphical model. The methods explained above are tested and validated on limited sizes of datasets.

In [13], a knowledge based approach that combines segmentation algorithms, object recognition, and anatomical knowledge is presented for the spine reconstruction from MR images. However, their work is developed for only the cervical spine. Zheng et al. [14] use an Hough Transform (HT) based method to localize discs from videofluoroscopic CT images. Due to the similarity of the discs and local nature of the HT algorithm, their system needs to use human vision to suggest approximate disc positions.

Zhan et al. [15] present a hierarchical strategy and local articulated model to detect (anchor and bundle) vertebrae and discs from 3D MR images. They use a Haar filter based Adaboost classifier and employ a local articulated model for calculating the spatial relations between vertebrae and discs. Glocker et al. [16] propose a method for localization and identification of vertebra in arbitrary field-of-view CT scans. Their method employs regression forests and Hidden Markov Models. It works even on cropped scans and partially visible spines.

Schmidt et al. [17] introduce a probabilistic inference method that measures the possible locations of the spinal discs in 3D MR images. Their approach uses a part-based model that describes the disc appearances by employing a tree classifier. The relationship between the discs is enforced under a graphical model framework. The inference algorithm uses a heuristic based A\* search to prune the exponential search space for efficiency.

Alomari et al. [3], [18] use a graphical model for the lumbar disc localization. Their model assumes local and global levels with latent variables. The global level latent variables make it possible to enforce restrictions on the dependencies of disc positions. The local latent variables are used as an abstraction level that separates the disc variables from the image intensity values. The inference on the resulting graphical model is performed by the generalized expectation maximization method, which is an approximate and iterative inference technique. In order to make the inference routines converge to final positions, the method of [3] needs to include high level terms such as expected disc locations on the image which is the main drawback of the system.

Similar to our method, localization vertebra/discs from MR images methods of [15], [3], and [17] are based on detecting the candidate positions with different ways and employing graphical models for the labeling. Our main contribution is using a polynomial time and an exact inference algorithm. In addition, our method is insensitive to scale and orientation changes of the 2D MR images of lumbar structures between patients. Furthermore, another important contribution of this study is the presentation of IPD which is a novel feature extraction method. There is no common large dataset on which the methods can be evaluated and the lack of a benchmark dataset causes absence of the fair comparisons between the localization systems in the literature. Therefore, we publish our dataset containing 80 MR images<sup>1</sup>.

Our current work extends our previous work [5] in many respects. First, our current system labels the lumbar vertebrae and discs together simultaneously, while [5] deals with discs only. Another important difference of our current method from [5] is being scale and patient-orientation invariant because we employ a second order MRF chain in our current work. In addition, [5] cannot handle the missing structures in the MR images. Finally, [5] is tested on a smaller dataset containing 40 images.

<sup>1</sup>The dataset can be downloaded from the multimedia section of the journal and <http://vision.gyte.edu.tr/lumbarmri/>

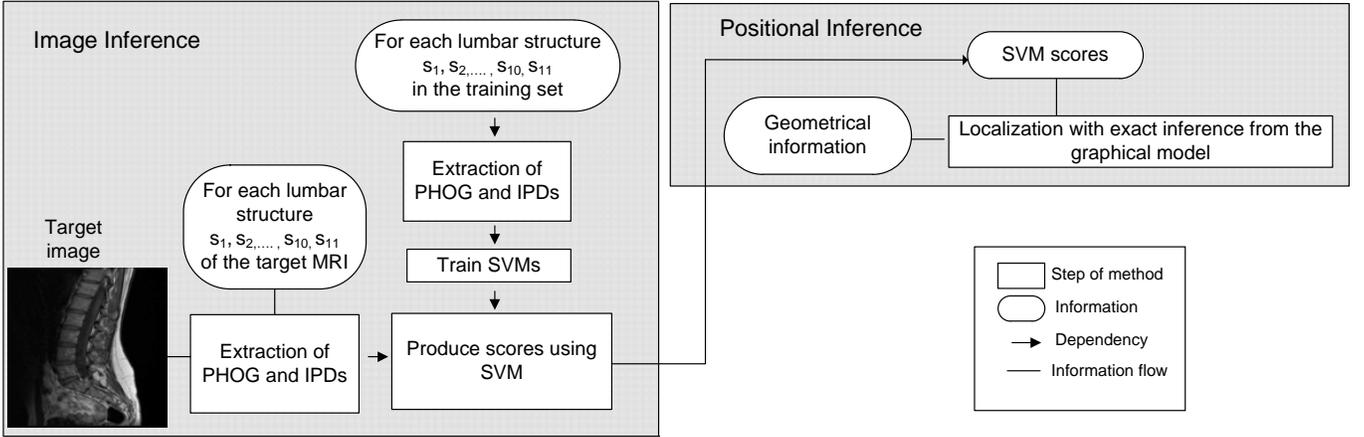


Fig. 2. The architecture of the proposed system. In the image inference step, disc scoring is performed with SVM after extracting PHOG and IPD. In the positional inference step, final localization is performed with inference from the chain-like graphical model.

### III. IMAGE INFERENCE STEP: THE SCORING PROCESS

In the image inference step, the PHOG and IPD features are extracted and candidate lumbar structure positions are scored with SVM (Figure 2).

#### A. Extraction of PHOG and IPD

The local intervertebral disc and vertebra detection methods in the literature use image gradient magnitude information [17], spatial location and intensity of discs [18], [3], [9], and image edges [13]. However, these local methods may have difficulties in lumbar disc and vertebra detection because the lumbar vertebrae and discs vary in the size, location, shape, and appearance due to pathologies and individual variations (See Figure 9). For example, disc degeneration<sup>2</sup> and disc herniation<sup>3</sup> change the intensity and shape of the discs [19]. Similarly, the scoliosis, the lateral curvature in the vertical line of the spine, affects the shape and location of the discs and vertebrae.

Histogram of Oriented Gradients (HOG) descriptor [20] represents a shape by a histogram of gradient orientations which are quantized into a number of bins. Each bin shows the number of pixels that have gradient magnitude orientations within a specific angular range in a given window. The PHOG [21] method combines the local image shape and the spatial pyramid kernel [22]. The pyramidal spatial information is incorporated by dividing the window into a sequence of smaller spatial grids like a quadtree where each division process forms a new level in the pyramid. For each grid cell, a HOG vector is calculated and the combination of these HOG vectors forms the PHOG descriptor. Although they are not often used for medical imaging applications, the insensitivity of the PHOG descriptors against intensity variations and shape deformations make them very useful for a wide range of medical applications.

For extracting the PHOG features, we employ a sliding window approach with different sized windows. We use the

integral histogram technique [23] to speed up the PHOG feature extraction process because direct PHOG descriptor extraction is computationally expensive.

One problem with the sliding windows based PHOG is that it can produce very similar descriptors for the overlapping windows which cause multiple detections for the same target. In order to make our feature extraction process more robust against multiple detections, we introduce a novel descriptor IPD to be used with the PHOG descriptors. For a given window, IPD horizontally projects the intensity values into a feature vector. This vector is then normalized and resized. IPD helps us to localize the object in a window more accurately. By combining PHOG with IPD, we obtain a very robust feature descriptor that is free from multiple detections for the same target.

Let  $W$  be an  $f \times d$  sized two dimensional window and  $W(i, j)$  be a pixel in the window  $W$  where  $1 \leq i \leq f$  and  $1 \leq j \leq d$ . An integral intensity vector  $V$  of size  $f$  is formed by

$$V(i) = \sum_{k=1}^d W(i, k), \quad (1)$$

where  $1 \leq i \leq f$ . The vector  $V$  is normalized and resized to a new vector of size  $r$  with linear interpolation to form the final IPD. The calculation of horizontal IPD is shown in Figure 3. Note that by rotating the window  $W$ , we could get other IPDs for different orientations. However, since the orientations of the lumbar structures do not vary greatly, we only use the horizontal IPD.

For the feature extraction process, instead of searching the whole image for the candidate disc and vertebra positions, the area around the spinal cord is searched. We extract the spinal cord with a method similar to [24]. First, we find the difference between T2-weighted and T1-weighted registered mid-sagittal slices of a subject. Then, we use a single threshold at a pixel intensity value of 100 in order to eliminate the unrelated parts. The connected part with greatest size is labeled as the spinal cord. This process produces the soft segmentation of the spinal

<sup>2</sup>Deterioration of the disk [19].

<sup>3</sup>Leak of the nucleus pulposus through a tear in the disc wall [19].

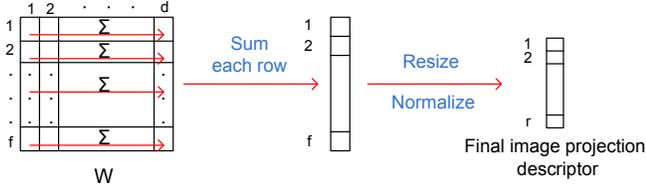


Fig. 3. The calculation of the horizontal IPD. The intensity values of each row is summed and then they are resized.

cord since our aim is detecting the search region around the spinal cord. The search region is decreased nearly to 1/10 of the original image size and this preprocessing step decreases the feature extraction time without losing any valid search regions.

### B. Scoring with SVM

The extracted PHOG and IPD features are first normalized and then linearly combined before SVM training. We use the Sequential Minimal Optimization [25] for training.

In the testing phase, the pixel in the center of each window  $W$  is assigned a score that indicates if the window contains a lumbar structure (vertebra or disc). These scores are generated by fitting a logistic regression model to the outputs of the SVM [26].

Let  $s_1, s_2, \dots, s_{11}$  denote the labels for the lumbar structures (T12-L1, L1, L1-L2, L2, L2-L3, L3, L3-L4, L4, L4-L5, L5, and L5-S1). Each structure  $s_i$  is trained and scored separately, i.e., we employ a separate SVM classifier for each  $s_i$ . Therefore for a given image  $I$ , each pixel in the search region is assigned 11 different scores that indicate being a lumbar structure or not. A number of candidates,  $m$ , with the best scores are chosen and used in the positional inference step. These score values are used in the formulation of the chain-like graphical model which is described in the next section.

## IV. POSITIONAL INFERENCE STEP: STRUCTURE LABELING WITH GRAPHICAL MODEL

The scoring process in the image inference step is achieved by using only local information for each lumbar structure. However, in order to ultimately localize and label the structures, more geometrical information is needed. In the positional inference step, we propose a second-order chain-like graphical model that combines the scores from the image inference step with the geometrical neighborhood information between the lumbar structures. This graphical model allows us to use dynamic programming based exact inference for the localization of the lumbar structures.

### A. The Graphical Model

We built our graphical model on triples containing 3 nodes and 2 edges. Each node represents a lumbar structure and edges in triples represent the relation between the nodes. The graphical model is a second-order Markov-chain-like

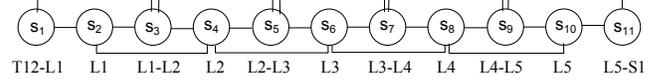


Fig. 4. The second-order chain-like graphical model of our system.

model. The correspondence between the labels and the lumbar structures is shown in Figure 4.

Consider  $x_k \in \mathbb{R}^2$  as a random variable that assigns a scored candidate structure  $s_k$  to its image location. Let  $p_k(x_k, I)$  be a function that gives the score value of a pixel at location  $x_k$  in the image  $I$  for the structure  $s_k$  where  $1 \leq k \leq 11$ .  $x = \{x_1, x_2, \dots, x_{11}\}$  denotes a configuration of candidate structures. The optimal configuration  $x' = \{x'_1, x'_2, \dots, x'_{11}\}$  assigns the centers of all lumbar structures  $s = \{s_1, s_2, \dots, s_{11}\}$  to their exact locations.

We find the optimal configuration  $x'$  with the maximum a posteriori (MAP) estimate

$$x' = \arg \max_x P(x|I, \alpha), \quad (2)$$

where  $\alpha$  represents the geometrical parameters learned from the training set and  $I$  is the image.  $P(x|I, \alpha)$  captures the local information about being a lumbar structure and its relation with the neighboring structures. It is defined in terms of a Gibbs distribution as

$$P(x|I, \alpha) = \frac{1}{Z} \exp^{-E(x, I, \alpha)}, \quad (3)$$

where the energy function  $E(x, I, \alpha)$  includes two potentials  $E_{local}$  and  $E_{geo}$

$$E(x, I, \alpha) = \left[ \sum_{k=1, \dots, 11} E_{local}(x_k, I) + \lambda \sum_{k=1, \dots, 11} E_{geo}(x_{k-1}, x_k, x_{k+1}, \alpha) \right], \quad (4)$$

where  $\lambda$  is a weighting parameter and it is selected as 0.5 in this study. The potential function  $E_{local}(x_k, I)$  is for the local information about the target structures for which we directly use the score values generated by fitting a logistic model to outputs of SVM.

$$E_{local}(x_k, I) = -p_k(x_k, I). \quad (5)$$

The potential function  $E_{geo}(x_{k-1}, x_k, x_{k+1}, \alpha)$  is for the geometrical information. It is defined as

$$E_{geo}(x_{k-1}, x_k, x_{k+1}, \alpha) = \left[ D(x_{k-1}, x_k, x_{k+1}, \alpha) + O(x_{k-1}, x_k, x_{k+1}, \alpha) \right], \quad (6)$$

where the functions  $D$  and  $O$  capture the ratio of the positional and orientational relations between the neighboring structures  $x_{k-1}$ ,  $x_k$ , and  $x_{k+1}$ .  $\alpha$  includes the geometric parameters between the neighboring structures learned from the training set.

The distance function  $D(x_{k-1}, x_k, x_{k+1}, \alpha)$  measures the distance ratio between the positions of 3 neighboring structures

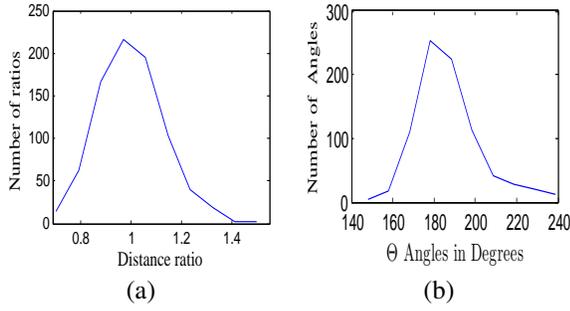


Fig. 5. The empirical distribution of (a) the distance ratios and (b) the  $\Theta$  of the lumbar structures in the dataset. They show the feasibility of using a Gaussian distribution for modeling the structures.

$x_{k-1}$ ,  $x_k$ , and  $x_{k+1}$  which is defined as  $\frac{\|x_{k-1}-x_k\|}{\|x_k-x_{k+1}\|}$ . The histogram of the distance ratios between all of the structures  $x_{k-1}$ ,  $x_k$ , and  $x_{k+1}$  in the training set is shown in Figure 5(a). This empirical distribution suggests the feasibility of using a Gaussian distribution for modeling the structure distance ratios. The mean distance ratio  $\mu_k^d \in \alpha$  and standard deviation  $\sigma_k^d \in \alpha$  are learned from the training set for the lumbar structure  $s_k$ . The distance ratio function  $D(x_{k-1}, x_k, x_{k+1}, \mu_k^d, \sigma_k^d)$  is

$$D = \begin{cases} \frac{(\frac{\|x_{k-1}-x_k\|}{\|x_k-x_{k+1}\|} - \mu_k^d)^2}{2(\sigma_k^d)^2}, & \text{if } \frac{\|x_{k-1}-x_k\|}{\|x_k-x_{k+1}\|} \text{ in } r-d \pm \tau_d \\ \infty, & \text{otherwise,} \end{cases} \quad (7)$$

where  $\tau_d$  is a threshold value and  $r-d$  is the range of minimum and maximum distance ratio values in the training set. The function  $D$  produces small values when the distance ratio between the structures is closer to the mean of the ratios in the training set. If the ratio  $\frac{\|x_{k-1}-x_k\|}{\|x_k-x_{k+1}\|}$  is outside of the range of the ratios in the training set, the function  $D$  is  $\infty$ . This means that we do not allow the neighboring structures to be further or nearer than the structures in the training set.

In Equation 7, the distance ratio is defined on a triple containing 3 nodes and 2 edges. The ratio does not change in images at different scales and at patients that have different vertebrae size. As a result, using distance ratios instead of distances makes our system scale-independent. The other lumbar structure detection methods such as [5], [17], and [3] use the direct distance values in their formulations which makes them scale sensitive. The employment of the distance ratios instead of the direct distances makes our system scale independent.

The function  $O(x_{k-1}, x_k, x_{k+1}, \mu_k^o, \sigma_k^o)$  measures the orientational relations between the neighboring structures. Let  $\Theta_k$  be the angle between the structures  $x_{k-1}$ ,  $x_k$ , and  $x_{k+1}$ . The histogram of all angles  $\Theta_k$  in the training set is shown in Figure 5(b). This empirical distribution again suggests the feasibility of using a Gaussian distribution for modeling the orientational relations. The parameters  $\mu_k^o \in \alpha$  and  $\sigma_k^o \in \alpha$  are learned from this training set. The orientation function  $O(x_{k-1}, x_k, x_{k+1}, \mu_k^o, \sigma_k^o)$  is defined as

$$O = \begin{cases} \frac{(\Theta_k - \mu_k^o)^2}{2(\sigma_k^o)^2}, & \text{if } \Theta_k \text{ in } r-o \pm \tau_o \\ \infty, & \text{otherwise,} \end{cases} \quad (8)$$

where  $\tau_o$  is a threshold value and  $r-o$  defines the range of angles in the training set. Note that similar to distance ratio function  $D$ , function  $O$  makes our system image and patient orientation independent since  $\Theta$  does not change with orientation of the image.

### B. Exact Inference with Dynamic Programming

Given a target image  $I$ , our objective is to infer the optimal configuration  $x'$  by maximizing the Equation 2. There are  $m$  best scored candidates for each structure  $s_k \in \{s_1, s_2, \dots, s_n\}$ . Searching for the optimal configuration  $x'$  of  $n$  structures that minimizes the energy term  $E(x, I, \alpha)$  in Equation 4 has running time of  $O(m^n)$  with the brute force technique which is intractable. However, note that in our chain-like graphical model, a node in a triple is conditionally dependent only on its two neighboring nodes which makes it possible to find  $x'$  in polynomial time by using the principle of optimality and dynamic programming [27], [28], [29], [30], instead of brute-force. It is observed that the energy term  $E$  of Equation 4 can be written in terms of separate energy terms  $E_1, E_2, \dots, E_{n-2}$ .

$$E(x_1, x_2, \dots, x_n) = \left[ E_1(x_1, x_2, x_3) + E_2(x_2, x_3, x_4) + \dots + E_{n-2}(x_{n-2}, x_{n-1}, x_n) \right], \quad (9)$$

where

$$E_{i-1}(x_{i-1}, x_i, x_{i+1}) = E_{local}(x_i, I) + \lambda E_{geo}(x_{i-1}, x_i, x_{i+1}, \alpha). \quad (10)$$

Each energy term  $E_{i-1}$  is written only in terms of the neighboring structures  $x_{i-1}$ ,  $x_i$ , and  $x_{i+1}$ . We define optimal value functions  $c_{j-1}(x_j, x_{j+1})$  that involve the best energy configuration up to the  $j^{th}$  structure starting from the first structure in the chain-like model.

$$\begin{aligned} c_0(x_1, x_2) &= E_{local}(x_1, I), \\ c_1(x_2, x_3) &= \min_{x_2} E_1(x_1, x_2, x_3), \\ c_2(x_3, x_4) &= \min_{x_3} E_2(x_2, x_3, x_4) + c_1(x_2, x_3), \\ &\dots \\ c_{n-2}(x_{n-1}, x_n) &= \min_{x_{n-1}} E_{n-2}(x_{n-2}, x_{n-1}, x_n) \\ &\quad + c_{n-3}(x_{n-2}, x_{n-1}) + E_{local}(x_n, I). \end{aligned} \quad (11)$$

The minimum element  $c_{n-2}(x_{n-1}, x_n)$  gives us the minimum energy  $E$ .

$$\min E = \min_{x_{n-1}, x_n} c_{n-2}(x_{n-1}, x_n). \quad (12)$$

By using the standard memoization and re-tracking the  $x$  with minimum energies, we find the optimal configuration  $x'$  of Equation 2. In Equation 11, each optimal value function is calculated by searching only three neighboring structure candidates and it has  $O(m^3)$  cost. Since there are totally  $(n-1)$  optimal value functions, the running time of the dynamic programming based algorithm is  $O(nm^3)$ . As a result, our method can benefit from the advantages of the chain-like structure such as scale and orientation independence while performing the exact optimization in polynomial time.

In some medical cases, discs might be missing because of high levels of disc degeneration. There also might be cases where the lumbar structure might not be found in the image inference due to image noise. We use a technique as in [17] for handling missing structures. For each missing structure  $s_i$ , we consider an additional dummy  $(m + 1)^{th}$  candidate. The potential function  $E_{local}$  and the global potential function  $E_{geo}$  for these dummy candidate positions are assigned high values. The optimal inference mechanism chooses the dummy candidate positions with the high energy if there are no suitable alternatives, thus handling the missing structure problem without affecting the localization of the other structures.

## V. EXPERIMENTAL RESULTS

The system was tested and validated on a clinical MR image dataset containing MR volumes of lumbar spinal column of 80 subjects. For each subject, there are 3 acquisition protocols: T1-weighted sagittal, T2-weighted sagittal, and T2-weighted axial. Each image is 512x512 pixels in size. The pixel spacing is 0.625 mm and thickness is 4 mm between the sagittal slices. Only mid-sagittal slices of T1-weighted and T2-weighted sagittal MR images were used in the system. The T2-weighted images were only used for the extraction of the spinal cord. For the feature extraction, T1-weighted images were used because the image gradient information is clearer in the T1-weighted images.

The sagittal view contains at least 5 lumbar vertebrae and 6 lumbar inter-vertebral discs. These views also include a number of sacrum and thoracic vertebrae and intervertebral discs. The factors on triples cannot be defined for the beginning node  $s_1$  (T12-L1) and end node  $s_{11}$  (L5-S1) of the chain without using their anterior and posterior nodes. In order to define factors on triples for the beginning and end nodes of the chain and in order to improve the localization results, we had extended our graphical model with the T12 and S1 vertebrae. More information is incorporated by using the anterior and posterior nodes of the chain.

In the dataset, 16 of the subjects are pathology-free and the remaining 64 subjects have pathologies. There are totally  $80 \times 5 = 400$  lumbar vertebrae and  $80 \times 6 = 480$  lumbar intervertebral discs. 158 of the lumbar intervertebral structures have pathologies and problems like herniation, degeneration, scoliosis, etc. The data set includes 31 vertebra abnormalities such as lumbarization, nodular lesion, etc.

For each image, the approximate centers of the lumbar structures were marked manually by an expert. The contour of each structure was also delineated by the same expert. The delineations and markings of the expert were used for training the features and the evaluation of the experimental results.

We evaluated the results of the image inference step and the positional inference step separately. For both of these two steps, we performed a subset of leave-20-out cross-validation. We randomly divided the dataset into 4 subsets, each containing 20 subjects. In each sub-experiment, 60 MR image slices from 3 subsets were trained and 20 slices from the 4<sup>th</sup> subset were tested. We performed 4 sub-experiments in total, so each of the 80 subjects was tested by our system.

### A. Image Inference Results

For the PHOG descriptor, we used 8 bins each containing 45 degrees angle range and two pyramidal levels. For the IPD, we extracted 10 features for each window, i.e., the number of bins  $r$  was selected as 10. The sizes of the minimum and maximum windows enclosing the structures in the training set delineated by the expert were found. Several values between the minimum and maximum window sizes are used in the test phase by extracting PHOG and IPDs. The features extracted at all scales are tested simultaneously.

Each structure  $s_k$  was trained and scored separately. 180 positive samples and 1800 negative samples were used for each structure in the SVM training. We run each of the 11 SVM classifiers separately for each possible candidate window size and position in testing. If the score of a window is in the best  $m = 500$  scores, then the window is considered as possibly containing a lumbar structure and it is used by the positional inference step.

It is not possible to evaluate the given scores directly, so we consider the binary classification results of SVM. In order to evaluate our image inference results, we use the windows around the spinal cord in the testing phase. Let  $L$  be the smallest window that contains the boundaries of the target expert delineated structure. Consider a pair  $(W, c)$  for a window  $W$  and label  $c \in \{-1, 1\}$  as the classification label of  $W$ . The windows  $(W, 1)$  are labeled as containing a lumbar structure and the windows  $(W, -1)$  are labeled as not containing a lumbar structure. A window  $(W, 1)$  is considered as correctly classified if it contains all of the boundaries of the target expert delineated structure, i.e.  $L \in (W, 1)$ . A window  $(W, -1)$  is considered as correctly classified if it does not include all of the boundaries of the target expert delineated structure ( $L \ni (W, -1)$ ). The classification rate (CR) is calculated by dividing the number of correctly detected windows by the number of all tested windows. The CR of our system is 89.41%.

In order to evaluate the effect of IPD, we make a classification with only PHOG descriptors for the same dataset. The CR of our system is 82.78% with only PHOG descriptor, which shows that using IPD with PHOG increases the classification accuracy.

Figure 6-a shows the disc centers detected using the T12-L1 disk classifier. The scores of candidate discs are shown with different colors where dark red colored pixels show low scored disc centers and light yellow colored pixels show high scored disc centers. Figure 6-b to 6-f show the image inference results for the same subject with other SVM disc classifiers (L1-L2, L2-L3, etc.). As seen in Figure 6, not only the target structure, but also the neighboring structures are detected. The confusion between the neighboring structures, whose borders have similar orientations, is expected because image inference is a local process. The analysis of Figure 6 indicates that although image inference produces valuable information about the lumbar structures, the final structure labeling and positions cannot be reliably produced from this local procedure, hence a more abstract level inference such as our positional inference step is required.

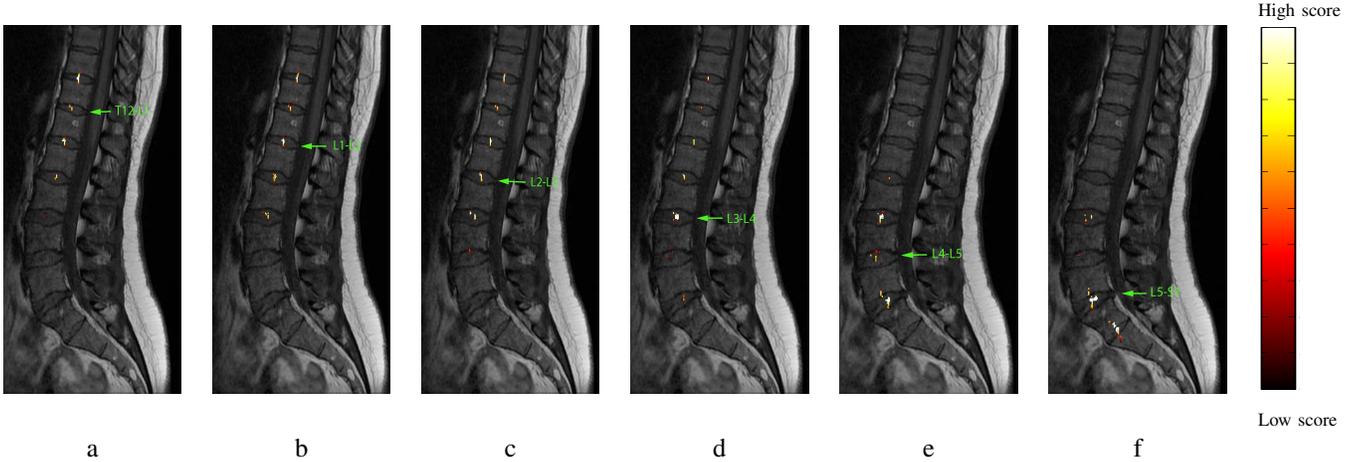


Fig. 6. The scores of the candidate disc centers detected by different structure classifiers are shown with different colors. Top 500 scores for the target structure is shown in each figure. Best viewed electronically.

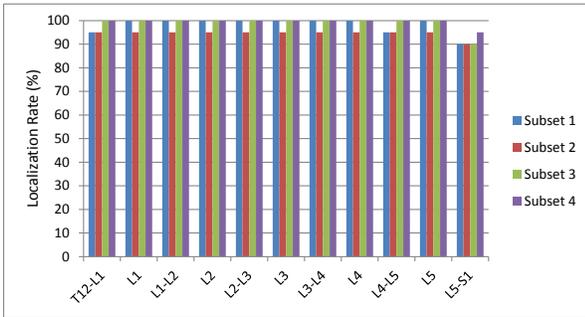


Fig. 7. The average disc labeling rates (%) of inference from our chain-like graphical model.

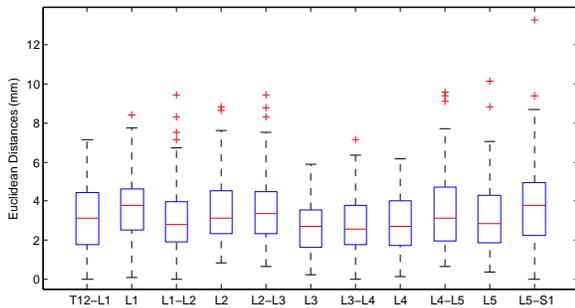


Fig. 8. The boxplot shows the Euclidean distances (in mm) between structure center labeled by our system and corresponding expert-delineated structure for 80 MR images in the dataset.

### B. Positional Inference Results

We also tested the final results of our system in the positional inference step. We estimated the parameters  $\mu_k^d, \sigma_k^d, \mu_k^o, \sigma_k^o \in \alpha$  from the same training data as in the image inference step. The weighting parameter  $\lambda$  in Equation 4 was empirically selected as 0.5,  $\tau_d$  was empirically selected as 0.2 and  $\tau_o$  was empirically selected as 15.

In order to evaluate the labeling performance of the system, we used two different methods. In the first method, the

structure is considered as correctly localized if the center of the structure detected by our system is inside the expert delineated contour. The points on the expert delineated contour are also considered as correctly localized. The labeling results according to the first metric are shown in Figure 7. The average of the labeling results in Figure 7 is 97.82%. The middle discs and vertebrae have higher accuracies than beginning (T12-L1) and end discs (L5-S1) because of the available information flow from the neighboring structures towards the middle structures. Note that although our system can handle most of the pathological cases, there might be some abnormalities that confuse our system such as the lumbarization<sup>4</sup> case in the subset 2. Figure 9-k shows the lumbarization example in subset 2 where all the lumbar structure labels detected by our system are shifted down to the next structure. Although our system can handle three other lumbarization cases in the dataset, we do not expect our system to localize all structures in the presence of abnormalities like lumbarization, metal artifacts, or large tumors because in medical practice extra image features from other structures are used for the diagnosis of these abnormalities.

The second evaluation method for the structure labeling was the Euclidean distances from the labeled structure center to the structure center labeled by the expert. The Euclidean distance is defined as  $\|gr_c - l_c\|$  where  $gr_c$  is the coordinate of the structure center labeled by the expert and  $l_c$  is the coordinate of the structure center labeled by our system. The Euclidean distances are reported in millimeters by converting from the pixels where each pixel is  $0.625 \times 0.625 \text{ mm}^2$ . The boxplot of the Euclidean distances for the lumbar structures for 80 images in the dataset is shown in Figure 8. The centerline of the box is the median, the top and bottom lines of the box are 25th and 75th percentiles and the pluses are the statistical outliers. The Euclidean distances of the unsuccessful lumbarization case are not included in the boxplot.

Figure 9 shows a few visual results of the final localizations

<sup>4</sup>Lumbarization is the nonfusion of the S1 and S2 vertebrae where the lumbar spine subsequently appears to have six vertebrae. It is an anatomic anomaly in the human spine [31].

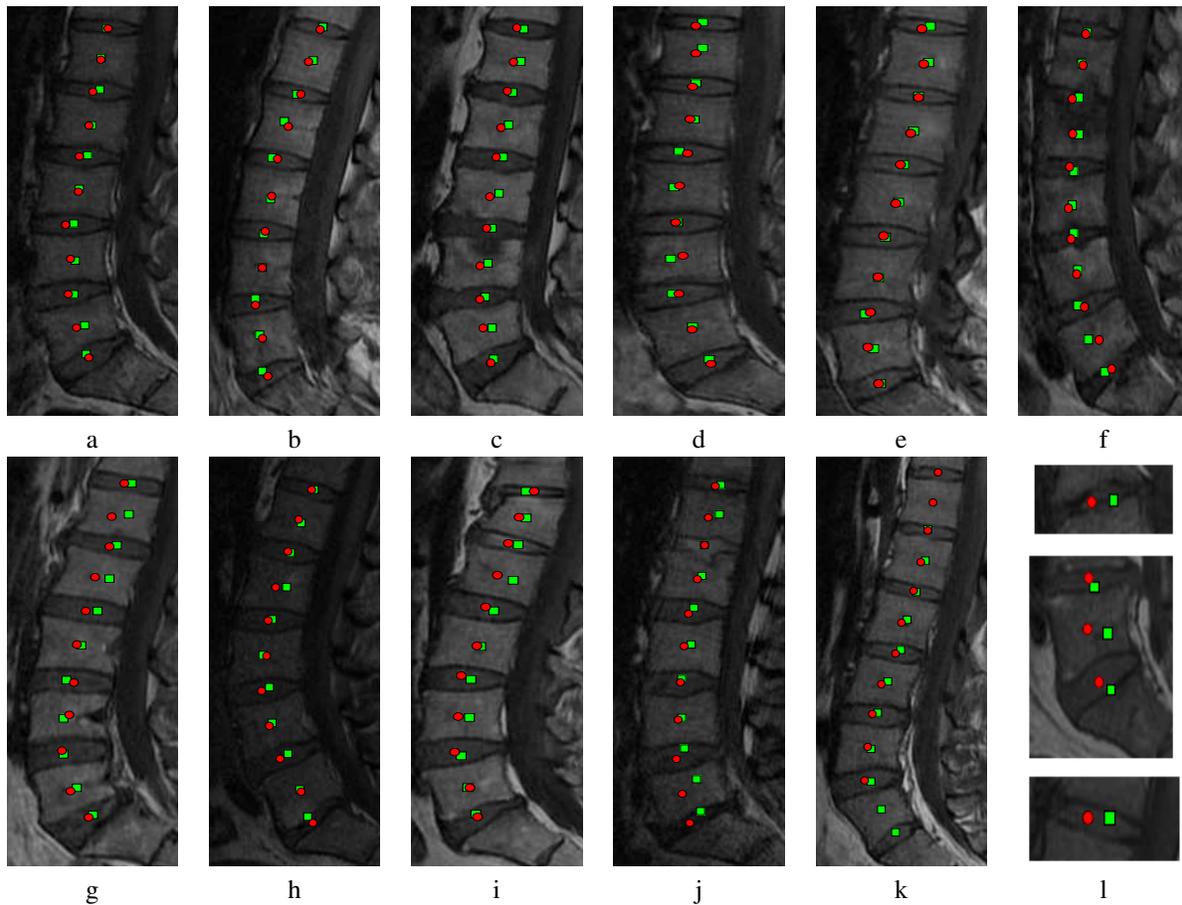


Fig. 9. (Best viewed electronically.) The results of the disc and vertebra labeling. Red dots are the ground truth and green squares are the structure centers localized by our system. The localization results of the normal subjects are shown in a-d and the localization results of the abnormal subjects are shown in e-k. In (k), our system does not localize the structures successfully where the lumbarization abnormality exists in the patient. In (l), the centers which have the largest Euclidean distances between our system and expert localizations. All of the images are cropped for better visualization.

of our system. Figure 9-a to 9-d show the labeling of lumbar structures of normal cases. Figure 9-e to 9-k show the labeling of abnormal cases where the subjects have diseases like herniation and disc degeneration. As mentioned before, Figure 9-k shows the lumbarization case where all of the structures are mislabeled. Figure 9-l shows 3 cases where the difference between our localization and expert-delineation are high. The experiments show that our system can be conveniently used for the medical applications for disc and vertebra localization.

### C. Scale and Orientation Invariance Evaluation

We performed another experiment in order to evaluate the scale and orientation invariance gained by defining triples on a second order Markov-chain-like structure. We randomly select 10 subjects from the dataset. The mid-sagittal slices of these subjects are randomly rotated between  $-40$  and  $40$  degrees and randomly scaled between  $0.5$  and  $1.5$  times. Other 70 subjects are used for training. In this experiment, the rotated versions of positive samples (between  $-40$  and  $40$  degrees) are also used for training. Other settings are same with the previous experiment. The boxplot of the Euclidean distances (in mm) between structure center labeled by our system and corresponding expert-delineated structure is shown in Figure

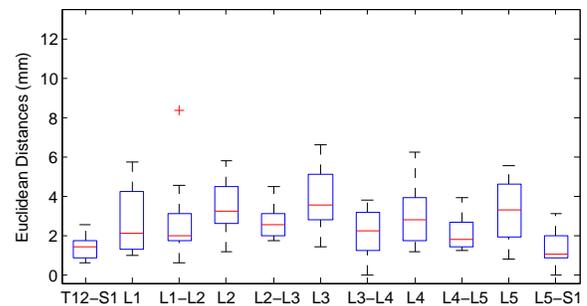


Fig. 10. The boxplot shows the Euclidean distances (in mm) between structure center labeled by our system and corresponding expert-delineated structure for the rotated and scaled images.

10. For the rotated and scaled images, the mean Euclidean distance to the expert-delineated center is  $2.07$  mm for the discs and  $3.25$  mm for the vertebra. The mean distance is  $3.1$  mm for the discs and  $2.95$  mm for the vertebra when there is no rotation and scaling. The distances are close to each other in original and rotated/scaled images. This experiment shows that rotation and scaling do not negatively affect our labeling results.

## VI. CONCLUSIONS

We presented a novel method for the simultaneous localization of the lumbar structures using machine learning and MRF. The machine learning handles the local image gradient orientation features of each structure and assigns scores to the candidate structures. The semi-global geometrical information of the lumbar spine is combined with the local score information under the MRF framework.

The separate handling of the local and geometrical information makes the system more modular and flexible. It is validated by experiments that the PHOG and IPDs produce very successful scoring results with SVM. More local features, such as 3D features and combined T1 and T2-weighted MR features, can be easily added to the system. Our local scoring process eliminates some structure positions and only a small number of candidate structures are used in the final inference process. This makes our system very efficient. In addition, the chain-like graphical model keeps the MRF model at manageable levels so that the dynamic programming based inference mechanism finds an optimal solution in polynomial time. At the same time, our system scale and image position invariant. The method can trivially be extended to localize all structures of the whole spine.

The clinical MRI dataset that we used to validate our system includes 80 cases where 64 of them have at least one abnormal lumbar structure. The localization results show that our system detects the structure centers even in abnormal cases successfully. In addition, the experiment performed with rotated and scaled images show that our system is scale and orientation invariant.

The presented method is the first and crucial step of the computer aided diagnosis and image-guided surgeries about the spine. It can be conveniently used for the disease and abnormality detection and orthopedic or neurological applications. Our system is very robust and our results are comparable with the state of the art.

## REFERENCES

- [1] L. Giles and K. Singer, *Clinical Anatomy and Management of Low Back Pain*. Butterworth-Heinemann, 2007.
- [2] L. Remonda, A. Lukes, and G. Schroth, "Spinal stenosis: current aspects of imaging diagnosis and therapy," *Schweiz Med Wochenschr*, vol. 126, no. 6, pp. 220–9, 1996.
- [3] R. S. Alomari, J. J. Corso, and V. Chaudhary, "Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model," *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1–10, 2011.
- [4] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz, "Automated model-based vertebra detection, identification, and segmentation in CT images," *Medical Image Analysis*, vol. 13, no. 3, pp. 471–482, 2009.
- [5] A. B. Oktay and Y. S. Akgul, "Localization of the lumbar discs using machine learning and exact probabilistic inference," in *Proc. of MICCAI*, vol. LNCS 6893, 2011, pp. 158–165.
- [6] Y. H. Lai and P. L. Lin, "Effective segmentation for dental x-ray images using texture-based fuzzy inference system," in *Proc. 10th Int. Conf. ACIVS*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 936–947.
- [7] L. R. Long and G. R. Thoma, "Use of shape models to search digitized spine x-rays," in *Proc. 13th IEEE Symposium on CBMS*, Washington, DC, USA, 2000, pp. 255–260.
- [8] 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 Transactions on Biomedical Engineering*, vol. 51, no. 10, pp. 1821–1829, 2004.
- [9] Z. Peng, J. Zhong, W. Wee, and J. Lee, "Automated vertebra detection and segmentation from the whole spine MR images," in *Conf. Proc. IEEE EMBS*, vol. 3, 2005, pp. 2527–2530.
- [10] S.-H. Huang, Y.-H. Chu, S.-H. Lai, and C. L. Novak, "Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI," *IEEE Transactions on Medical Imaging*, vol. 28, no. 10, pp. 1595–1605, 2009.
- [11] A. Wong, A. K. Mishra, J. Yates, P. W. Fieguth, D. A. Clausi, and J. P. Callaghan, "Intervertebral disc segmentation and volumetric reconstruction from peripheral quantitative computed tomography imaging," *IEEE Trans. Biomed. Engineering*, vol. 56, no. 11, pp. 2748–2751, 2009.
- [12] R. Donner, G. Langs, B. Mičušik, and H. Bischof, "Generalized sparse MRF appearance models," *Image Vision and Computing*, vol. 28, pp. 1031–1038, June 2010.
- [13] S. Seifert, I. Wachter, G. Schmelzle, and R. Dillmann, "A knowledge-based approach to soft tissue reconstruction of the cervical spine," *IEEE Transactions on Medical Imaging*, vol. 28, no. 4, pp. 494–507, 2009.
- [14] Y. Zheng, M. Nixon, and R. Allen, "Automated segmentation of lumbar vertebrae in digital videofluoroscopic images," *IEEE Transactions on Medical Imaging*, vol. 23, no. 1, pp. 45–52, 2004.
- [15] Y. Zhan, D. Maneesh, M. Harder, and X. S. Zhou, "Robust mr spine detection using hierarchical learning and local articulated model," in *Proceedings of the 15th international conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I*, ser. MICCAI'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 141–148.
- [16] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view ct scans," in *MICCAI (3)*, 2012, pp. 590–598.
- [17] S. Schmidt, J. Kappes, M. Bergholdt, V. Pekar, S. Dries, D. Bystrov, and C. Schnörr, "Spine detection and labeling using a parts-based graphical model," in *Proc. of the 20th Int. Conf. IPMI*, vol. LNCS 4584, 2007, pp. 122–133.
- [18] J. J. Corso, R. S. Alomari, and V. Chaudhary, "Lumbar disc localization and labeling with a probabilistic model on both pixel and object features," in *Proc. of MICCAI*, vol. LNCS 5241, 2008, pp. 202–210.
- [19] D. F. Fardon and P. C. Milette, "Nomenclature and classification of lumbar disc pathology," *Spine*, vol. 26, no. 5, pp. E93–E113, 2001.
- [20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of IEEE Int. Conf. CVPR*, vol. 2, June 2005, pp. 886–893.
- [21] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proc. of the Int. CIVR*, 2007, pp. 401–408.
- [22] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. of IEEE Int. Conf. CVPR*, 2006, pp. 2169–2178.
- [23] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," in *Proc. of IEEE Int. Conf. CVPR*, 2005, pp. 829–836.
- [24] C. Bhole, S. Kompalli, and V. Chaudhary, "Context-sensitive labeling of spinal structures in MRI images," in *In Proc. of SPIE Medical Imaging*, 2009, pp. 803–806.
- [25] J. C. Platt, *Advances in Kernel Methods*. Cambridge, MA, USA: MIT Press, 1999, pp. 185–208.
- [26] —, *Advances in Large Margin Classifiers*, 1999, pp. 61–74.
- [27] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2007.
- [28] Y. S. Akgul and C. Kambhamettu, "A coarse-to-fine deformable contour optimization framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 174–186, February 2003.
- [29] A. A. Amini, T. E. Weymouth, and R. C. Jain, "Using dynamic programming for solving variational problems in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 855–867, September 1990.
- [30] Y. Amit and A. Kong, "Graphical templates for model registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 225–236, 1996.
- [31] F. Wang and L. Zhang, "Forensic identification for 16 cases with lumbar sacralization and lumbarization," *Fa Yi Xue Za Zhi*, vol. 22, no. 1, pp. 67–9, 2006.