Spine detection in CT and MR using iterated marginal space learning

B. Michael Kelm [a,*], Michael Wels [a], S. Kevin Zhou [b], Sascha Seifert [a], Michael Suehling [a], Yefeng Zheng [b], Dorin Comaniciu [b]

[a] Imaging and Computer Vision, Siemens Corporate Technology, Erlangen, Germany
[b] Imaging and Computer Vision, Siemens Corporate Research and Technology, Princeton, NJ, USA

ABSTRACT

Examinations of the spinal column with both, Magnetic Resonance (MR) imaging and Computed Tomography (CT), often require a precise three-dimensional positioning, angulation and labeling of the spinal disks and the vertebrae. A fully automatic and robust approach is a prerequisite for an automated scan alignment as well as for the segmentation and analysis of spinal disks and vertebral bodies in Computer Aided Diagnosis (CAD) applications. In this article, we present a novel method that combines Marginal Space Learning (MSL), a recently introduced concept for efficient discriminative object detection, with a generative anatomical network that incorporates relative pose information for the detection of multiple objects. It is used to simultaneously detect and label the spinal disks. While a novel iterative version of MSL is used to quickly generate candidate detections comprising position, orientation, and scale of the disks with high sensitivity, the anatomical network selects the most likely candidates using a learned prior on the individual nine dimensional transformation spaces. Finally, we propose an optional case-adaptive segmentation approach that allows to segment the spinal disks and vertebrae in MR and CT respectively. Since the proposed approaches are learning-based, they can be trained for MR or CT alike. Experimental results based on 42 MR and 30 CT volumes show that our system not only achieves superior accuracy but also is among the fastest systems of its kind in the literature. On the MR data set the spinal disks of a whole spine are detected in 11.5 s on average with 98.6% sensitivity and 0.073 false positive detections per volume. On the CT data a comparable sensitivity of 98.0% with 0.267 false positives is achieved. Detected disks are localized with an average position error of 2.4 mm/3.2 mm and angular error of 3.9°/4.5° in MR/CT, which is close to the employed hypothesis resolution of 2.1 mm and 3.3°.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Examinations of the spine in both Magnetic Resonance (MR) imaging and Computed Tomography (CT) require a standardized alignment with the vertebral column. While some physicians may prefer to align an oblique multi-planar reconstruction in the acquired 3D volume, improved image quality is achieved when incorporating the alignment into image generation. In CT, for example, specialized reconstruction algorithms can compute oblique slices from the spiral raw data. In MR, oblique acquisition matrices as well as saturation bands for suppressing lipid artifacts have to be aligned with the spine based on a scout scan. As manual alignment is both time-consuming and operator-dependent, it is desirable to have a robust, fully automatic, and thus reproducible approach.

For reporting, physicians are also interested in an automatic labeling of the vertebrae (C1, C2, . . . , T1, . . . , L5, S1) that avoids error-prone manual counting. Such a labeling may further support systems for semantic body parsing (Seifert et al., 2009; Seifert et al., 2010; Criminisi et al., 2011) and semantic description of anatomical locations through vertebra names as frequently done by physicians (Klinder et al., 2009; Seifert et al., 2010). Furthermore, an automatic system for localizing the spinal disks can also be part of a computer-aided diagnosis system for analyzing pathologies of the spine, the spinal disks or the vertebrae (Wels et al., 2012). Approaches modeling biomechanical behavior of the spinal column (Adams and Dolan, 2005) could also benefit from such a system by supporting patient-specific analyses to guide therapeutic decisions (Tschirhart et al., 2007; Seifert and Dillmann, 2007; Adams and Dolan, 2011).

An automatic procedure for detecting and segmenting the spinal column faces various challenges, however. Varying contrasts, image artifacts and pathologies can compromise the detection of spinal disks based on local image features. Thus, a global spine model is required to robustly identify individual disks from their context. Such a model must also cope with missed detections and pathologic deformations of the spine such as various types of scoliosis, kyphosis and lordosis. Finally, the overall approach

* Corresponding author.
E-mail address: michael.kelm@siemens.com (B. M. Kelm).

should run reasonably fast (within seconds) to allow clinical application.

In this article we propose a novel approach that combines efficient local object detection based on an iterative version of Marginal Space Learning (MSL) (Zheng et al., 2008) with a global probabilistic prior model for the vertebral column. Furthermore, we propose an adaptive segmentation approach that, initialized with an oriented bounding box around a detected spinal disk or a vertebra, may be used to obtain detailed boundaries. Since our whole system is learning-based, it can easily be trained for use with CT as well as various MR sequences (Georgescu et al., 2005). Compared to a previous version of this work published in (Kelm et al., 2010), here, we elaborate on some details of the proposed algorithm and present additional results on CT data, structure segmentation and an analysis of hyperparameters.

1.1. Related work

Recently, spine detection and spinal shape analysis have regained interest. Boisvert et al. (2008) present a model that describes the statistical variations of the spine in terms of sequential rigid transformations of the local vertebra coordinate systems. Using principal component analysis on the Riemannian manifold of rigid transformations, clinically meaningful eigenmodes can be extracted. Although relying on the same metrics, we extend upon that by formulating a probabilistic spine model that is applied for detection rather than statistical analysis.

The detection of spinal disks in 3D MR scout scans has recently been addressed by Pekar et al. (2007). The authors propose a three-step approach using a special-purpose 2D image filter for disk candidate detection, followed by a customized spine tracking method and a final labeling step based on counting. Since their approach is designed to work on MR data only, it might not be easily adapted to CT image volumes.

Schmidt et al. (2007) propose a trainable approach based on extremely randomized trees in combination with a complete graphical model. They employ an A*-search based inference algorithm for exact maximum a posteriori (MAP) estimation. The approach only considers the position of the spinal disks, while we also determine their orientations and scales. However, their parts-based 3D approach is similar to ours.

Corso et al. (2008) and Alomari et al. (2011) argue that a two-level probabilistic model is required to separate pixel-level properties from object-level geometric and contextual properties. They propose a generative graphical model with latent disk variables which they solve by generalized expectation maximization (EM). Although the approach only provides position estimates and has only been evaluated for lumbar disks in 2D T_2-weighted MR data, it could in principle be extended to full 3D estimation. But since EM only finds a local optimum of the expected log likelihood, which can render such an approach very sensitive to initialization, it is not clear how the approach would scale to higher-dimensional estimation including 3D position, orientation, and scale.

Huang et al. (2008), Huang et al. (2009) propose a three-step approach comprising a vertebra detector based on AdaBoost, spinal column fitting using the random sample consensus (RANSAC) algorithm and a final segmentation step using an iterative normalized-cut algorithm. They apply their method to sagittal slices from 22 spinal MR datasets, thus only present results based on 2D spinal MR images.

Various approaches for automated spinal column extraction and partitioning have also been proposed for CT data. One such method is presented by Yao et al. (2006). It consists of the following steps: thresholding for initial spine segmentation, spinal canal extraction, four-part model-based vertebral region segmentation, and intensity profile analysis for spinal column partitioning. Vali-

dation has been carried out by visual examination. By lacking a quantitative analysis of the method's accuracy it is difficult to assess its practical applicability. Several design decision are made paying respect to particular intensity and contrast characteristics of CT data. For this reason adaptation of the method to MR data may not be straight-forward.

Jäger et al. (2009) rely on an iterative spinal cord segmentation method based on Markov random fields for assessing spinal geometry in terms of computed planes through the vertebral bodies, which is a less specific representation of the final detection results than chosen by us. In order to pay respect to, in particular, scoliotically deformed spinal columns, their method does not depend on any prior information about the shape of the spine nor on the shape of the vertebral bodies. It aligns with other methods for the segmentation of the spinal cord (McIntosh and Hamarneh, 2006; Coulon et al., 2002).

A comprehensive system for spine segmentation was proposed by Klinder et al. (2009). The approach starts with the tracking of the spinal canal which is subsequently used for a curved planar reformation (CPR) of the CT volume. The detection of vertebra candidates employs the generalized Hough transform on that CPR volume. Vertebrae are then identified and labeled using an appearance model learned from annotated data. Finally, the vertebrae are segmented using a vertebra-specific point distribution model. The algorithm, in particular the vertebra detection step, is tuned towards CT and is not directly applicable to MR data. It was evaluated on 64 CT volumes including cases with various pathologies. Further methods for the localization of the spine in tomographic images substantially rely on the segmentation of individual vertebrae like, for example, the method of Peng et al. (2006).

While a motivation for our probabilistic model is found in the great success of conditional random fields (Lafferty et al., 2001), i.e. discriminative models, their training with nonlinear potential functions is difficult (Dietterich et al., 2008) and has been avoided. Instead we followed the common approach of combining a probabilistic generative model and a discriminative model with independent training like, for example, Wels et al. (2008) and Donner et al. (2011). Alternative approaches for combining the strengths of generative and discriminative probabilistic models have been proposed by Lasserre et al. (2006), Kelm et al. (2006) and Tu et al., 2008.

2. Methods

We propose a learning based approach that can be applied to MR and CT alike. Our approach can be subdivided into four major steps (cf. Fig. 1). To constrain the search range for the disks, the spine is roughly located within the given volume first. To this end we use the MSL approach (Zheng et al., 2008) for detecting oriented bounding boxes around the cervical, the thoracic and the lumbar spine parts. Second, disk candidates are generated with high sensitivity using a novel iterative extension of the MSL approach. Third, a global probabilistic spine model is used to select the most likely disk candidates based on their appearance probabilities and their relative poses. Simultaneously, this model also assigns an appropriate label to each disk. In the final, optional step, the oriented bounding boxes around the detected disks are used for a case-adaptive segmentation (detailed boundary delineation) of the spinal disks in MR and the vertebral bodies in CT respectively.

In the following, we first elaborate on the global probabilistic spine model used for detection. We then review standard MSL (Zheng et al., 2008) followed by the introduction of its iterative extension. Finally, we describe a case-adaptive segmentation approach based on graph cuts that exploits the bounding boxes of the detected spinal disks or the vertebrae.
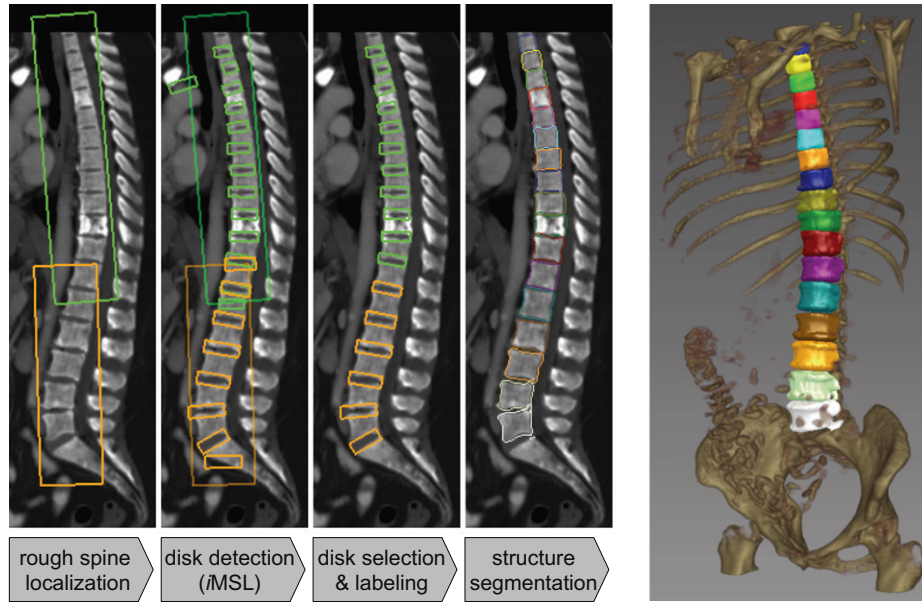
Fig. 1. **Left**: Four-step approach to the detection and segmentation of the spinal column. **Right**: Volume rendering showing the final 3D segmentations of the vertebral bodies in CT.

2.1. Global probabilistic spine model

The typical spatial structure of the spine gives rise to a prior on the relative poses of the spinal disks. This has been modeled by the factor graph (Kschischang et al., 2001) depicted in Fig. 2. We have chosen a chain model with potentials considering position, orientation and scale of the spinal disks. Each of the (vector-valued) random variables $\boldsymbol{b}_1$ to $\boldsymbol{b}_N$ represents the pose of a certain spinal disk, thus $\boldsymbol{b}_s$ holds a 3D position $\boldsymbol{p}_s = [x_s, y_s, z_s]^T$, a unit quaternion $\boldsymbol{q}_s$ representing the orientation and an anisotropic scale $\boldsymbol{s}_s = [s_s^x, s_s^y, s_s^z]^T$ for every disk $s \in \{1, \dots, N\}$. Thus, a distribution over disk poses is defined by the log probability

$$\log \Pr(\boldsymbol{b}_1, \dots, \boldsymbol{b}_N | \Theta, \boldsymbol{X}) = \sum_s V_s(\boldsymbol{b}_s | \theta_s, \boldsymbol{X}) + \sum_{s \sim t} V_{st}(\boldsymbol{b}_s, \boldsymbol{b}_t | \theta_{st})$$
$$- A(\Theta) \qquad (1)$$

where $A(\Theta)$ is the log partition function, $\boldsymbol{X} = (x_i)_{i=1, \dots, M}$ represents the image data and $\Theta = \{\theta_s, \theta_{st}\}$ subsumes all model parameters which are detailed in the following.

The pair potential between two neighboring disk $\boldsymbol{b}_s$ and $\boldsymbol{b}_t$ combines relative position, relative scale and relative orientation terms

$$V_{st}(\boldsymbol{b}_s, \boldsymbol{b}_t | \theta_{st}) = V_{pos,st} + V_{sca,st} + V_{rot,st}. \qquad (2)$$

Relative position is modeled by a Gaussian pair potential

$$V_{pos,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{1}{2} \boldsymbol{d}_{pos}^T(\boldsymbol{b}_s, \boldsymbol{b}_t) \Sigma_{pos,st}^{-1} \boldsymbol{d}_{pos}(\boldsymbol{b}_s, \boldsymbol{b}_t), \qquad (3)$$

where the displacement vector $\boldsymbol{d}_{pos}(\boldsymbol{b}_s, \boldsymbol{b}_t) = \mathbf{R}_s^{-1}(\boldsymbol{p}_t - \boldsymbol{p}_s) - \boldsymbol{\mu}_{pos,st}$ is computed relative to the orientation of disk candidate $\boldsymbol{b}_s$. Thus the position term is invariant with respect to translations and rotations applied to both disk candidates $\boldsymbol{b}_s$ and $\boldsymbol{b}_t$. The parameters of the position term are the mean displacement vector $\boldsymbol{\mu}_{pos,st}$ and the covariance matrix $\Sigma_{pos,st}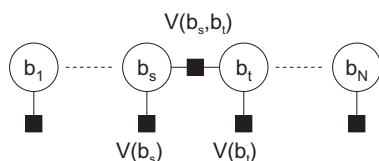$ which is assumed to be diagonal here. Mean and covariance of the displacement vector are estimated from the annotated training data for each pair potential of the spinal chain model independently.

The scale term is modeled by a Gaussian potential as well:

$$V_{sca,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{1}{2} \boldsymbol{d}_{sca}^T(\boldsymbol{b}_s, \boldsymbol{b}_t) \Sigma_{sca,st}^{-1} \boldsymbol{d}_{sca}(\boldsymbol{b}_s, \boldsymbol{b}_t), \qquad (4)$$

where the scale difference is $\boldsymbol{d}_{sca}(\boldsymbol{b}_s, \boldsymbol{b}_t) = \boldsymbol{s}_t - \boldsymbol{s}_s - \boldsymbol{\mu}_{sca,st}$. Again, mean scale difference $\boldsymbol{\mu}_{sca,st}$ and diagonal covariance $\Sigma_{sca,st}$ are estimated from the training data.

For the orientation term, we follow an approach from Pennec (2006) which suggests to employ the intrinsic metric of the underlying Riemannian manifold $\mathcal{SO}^3$. For the spine model this intrinsic metric is the rotation angle between the orientations of the disk candidates $\boldsymbol{b}_s$ and $\boldsymbol{b}_t$. Thus, the orientation term is modeled as the univariate Gaussian potential

$$V_{rot,st}(\boldsymbol{b}_s, \boldsymbol{b}_t) = -\frac{\alpha\left(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{\mu}_{rot,st}^{-1}\right)^2}{2\sigma_{rot,st}^2} \qquad (5)$$

where $\boldsymbol{q}_s$ ($\boldsymbol{q}_t$) is the quaternion associated with the rotation matrix $\mathbf{R}_s$ ($\mathbf{R}_t$) and the rotation angle $\alpha(\boldsymbol{q}) = \alpha([q_0 \ q_1 \ q_2 \ q_3])$ is computed from the quaternion as $2 \arccos(q_0)$. The orientation term defines the variance parameter $\sigma_{rot,st}$ and the quaternion mean parameter $\boldsymbol{\mu}_{rot,st}$, which is determined as the Fréchet mean (Pennec, 2006). Collecting all instances of a certain disk pair $(\boldsymbol{b}_s, \boldsymbol{b}_t)$ into the training sample $\mathcal{P}_{st}$, the Fréchet mean for the corresponding orientation term is determined as

$$\boldsymbol{\mu}_{rot,st} = \operatorname{argmin}_{|\boldsymbol{q}|=1} \sum_{(\boldsymbol{b}_s, \boldsymbol{b}_t) \in \mathcal{P}_{st}} \alpha\left(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{q}^{-1}\right)^2. \qquad (6)$$

Here, the mean was computed using the eigen-decomposition proposed by Karney (2007) and the variance was estimated with

$$\sigma_{rot,st}^2 = \frac{1}{|\mathcal{P}_{st}| - 1} \sum_{(\boldsymbol{b}_s, \boldsymbol{b}_t) \in \mathcal{P}_{st}} \alpha\left(\boldsymbol{q}_t \boldsymbol{q}_s^{-1} \boldsymbol{\mu}_{rot,st}^{-1}\right)^2. \qquad (7)$$

Finally, the single site potentials, which are determined by iterated marginal space learning as described in Section 2.2.2, encode image-based likelihood, i.e.,

$$V_s(\boldsymbol{b}_s | \theta_s, \boldsymbol{X}) = \log(\Pr(\boldsymbol{b}_s | \theta_s, \boldsymbol{X})). \qquad (8)$$

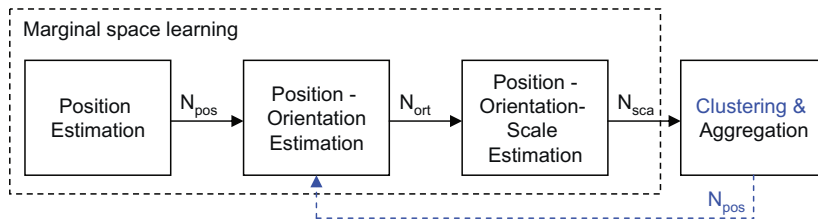**Fig. 2.** Factor graph modeling the relation between the spinal disks.

Fig. 3. Graphical overview of marginal space learning and its iterative extension (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Note that while the pair potentials, by their definition, are invariant under global rigid transformations (translation and rotation), the single site potentials that capture the appearance of disks are not necessarily. However, since a common disk detector is trained for different spine parts, the learned appearance models capture many disk poses and thus exhibit some invariance towards rigid transformations. Hence, the resulting global probabilistic spine model is insensitive towards slight changes of the spine pose. Furthermore, models capturing only parts of the complete spine can be easily constructed by just omitting certain disk variables. Since all potential parameters are determined independently (i.e., the likelihood decouples), no retraining is required and a probabilistic model appropriate for the current acquisition protocol, e.g., a lumbar spine protocol, can be assembled at runtime.

2.2. Local probabilistic detection model

The single site potentials defined in Eq. (8) require a probability estimate for the presence of a spinal disk at every possible position, orientation and scale in the image volume. However, an exhaustive evaluation of the single site as well as the pair potentials on the uniformly discretized nine dimensional parameter space (3 position, 3 orientation and 3 scale parameters) would quickly become computationally infeasible. Hence, we adopt the MSL paradigm (Zheng et al., 2008), a novel approach to object detection that has recently proven successful in numerous applications (e.g. Feng et al., 2009; Ionasec et al., 2009; Wels et al., 2009; Chen et al., 2009).

2.2.1. Object detection using marginal space learning

Instead of searching the whole nine dimensional parameter space for possible objects, the MSL heuristic proposes to employ three estimation steps (cf. Fig. 3) and to greedily follow the most promising partial candidate solutions in a fashion similar to beam search (Koller and Friedman, 2009, p. 1156).

For the first step, a machine learning classifier is trained on Haar-like features using integral volumes similar to the 2D approach proposed by Viola and Jones (2001). Like Zheng et al. (2008), we used the probabilistic boosting tree (PBT) classifier but other classifiers such as random forests (Breiman, 2001) work as well. This position classifier is used to evaluate a number of H_{pos} positions (e.g. some volume on a rectangular grid) and to collect the N_{pos} most likely candidate positions for the sought object. For the second step, each candidate position is augmented by a discrete number of H_{ort} orientation hypotheses yielding $N_{pos}H_{ort}$ partial candidate solutions. These are evaluated by a second classifier using steerable features (Zheng et al., 2008) and the N_{ort} most likely candidates are collected. Similarly, the third steps augments the position-orientation candidates with H_{sca} scale hypotheses and determines the N_{sca} most likely detections among the $N_{ort}H_{sca}$ candidate solutions. In the final step, an aggregate estimate for position, orientation and scale is constructed from the N_{sca} estimates obtained in the third step, for which a simple mean (Fréchet mean for the orientations) is usually sufficient and may yield errors below the discretization resolution.

Compared with exhaustive search, which would require the evaluation of $N = H_{pos}H_{ort}H_{sca}$ candidate solutions, MSL only requires the evaluation of $N = H_{pos} + N_{pos}H_{ort} + N_{ort}H_{sca}$ (partial) candidate solutions, which is considerably smaller for reasonably chosen parameters N_{pos} and N_{ort}. Beyond this obvious speed-up obtained through the MSL heuristic, it also leads to more robust and accurate detection results than an exhaustive full space search (Zheng et al., 2009b; Zheng et al., 2009a). This effect can be attributed to the application of three classifiers, the first two of which are tuned towards high sensitivity. Starting with a relatively low specificity ('first discard the obvious cases'), each stage is designed for rejecting false positives without loosing the true positives, which resembles the idea of the cascade approach proposed by Viola and Jones (2001). A crucial difference to the cascade approach is, however, that the two first classifiers employed in the MSL approach only model marginal distributions of the object parameters, which simplifies the classification problem and leads to better generalization performance (Zheng et al., 2009b). The approach of discarding candidates with low probabilities at early stages also resembles the hierarchical approach of Zhou et al. (2007) and can be generalized into applying whole networks of detectors as described by Sofka et al. (2011).

2.2.2. Iterated marginal space learning

MSL has been designed to detect a single, specific object such as a particular organ or landmark. If multiple objects of the same type, such as spinal disks, are to be detected, the final aggregation step has to be adapted to include some type of clustering before the aggregation. But even then, this MSL approach may end up with detections for the most salient disks only, i.e., many disks may be missed. Although the sensitivity could be improved by drastically increasing the number of considered candidates in each step, MSL would then loose its computational efficiency and render such an approach impracticable.

Algorithm 1. Iterated marginal space learning (iMSL).

Input: R, N_0, N_{pos}, N_{ort}, N_{sca}
Output: Set $\mathscr{D}$ of detected disk candidates
$\mathscr{D} := \{\}$
$\mathscr{P}_0 :=$ top N_0 position candidates
repeat
  $\mathscr{P}_0 := \{p \in \mathscr{P}_0 : \text{dist}(p, q) > R \; \forall q \in \mathscr{D}\}$
  $\mathscr{D}_{pos} :=$ top N_{pos} position candidates from $\mathscr{P}_0$
  $\mathscr{D}_{ort} :=$ top N_{ort} position-orientation candidates from $\mathscr{D}_{pos}$
  $\mathscr{D}_{sca} :=$ top N_{sca} position-orientation-scale candidates from $\mathscr{D}_{ort}$
  Perform hierarchical agglomerative clustering on $\mathscr{D}_{sca} \cup \mathscr{D}$
  foreach cluster $\mathscr{C}$ **do**
    if $|\mathscr{C}| \geqslant N_A$ **then**
      Aggregate the top N_A candidates
      Add the resulting box to $\mathscr{D}$
until $|\mathscr{P}_0| = 0$ or $|\mathscr{D}|$ remains constant

---

We therefore propose a novel extension to MSL, iterative MSL (*i*MSL), to cope with multiple objects of the same type (cf. Algorithm 1 and Fig. 3 blue parts). It is designed to achieve a higher sensitivity than usual MSL at moderate computational costs. Like for standard MSL, the position detector is evaluated in each voxel of the given image volume region first. The N_0 most likely candidates are collected in the set of initial position candidates $\mathcal{P}_0$. Then, the best N_{pos} ($N_{pos} \ll N_0$) candidates from $\mathcal{P}_0$ are evaluated using the orientation detector whose top candidates are evaluated using the scale detector. The resulting set $\mathcal{D}_{sca}$ contains disk candidate detections with all estimated parameters. Using pairwise average-linkage clustering (Hastie et al., 2001) with Euclidean distance, clusters of candidate disks are obtained. The number of clusters is determined by the cut in the dendrogram for which the distance of the merged clusters first exceeds twice a specified radius R. The most likely N_A box candidates of each resulting cluster are averaged and added to the set of detected disk candidates $\mathcal{D}$. After removing all position candidates from $\mathcal{P}_0$ that are closer than the cluster radius R to any of the detections in $\mathcal{D}$, orientation and scale detection are repeated on the remaining position candidates until no position candidates are left or no new disk candidates are detected.

After the first iteration, *i*MSL produces the same results as standard MSL with equivalent parameters. Formally, the number of candidate evaluations that are necessary for *i*MSL is $N = H_{pos} + N_{iter}(N_{pos}H_{ort} + N_{ort}H_{sca})$ where N_{iter} is the number of executed iterations (which is case-specific and depends on N_0). Since the number of positions examined with the position classifier H_{pos} is the same for both, *i*MSL and MSL, and since it dominates the number of evaluations N for sufficiently small N_{pos} and N_{ort}, *i*MSL does not require significantly more computation time than MSL with the same candidate numbers N_{pos} and N_{ort} ($\Delta N = (N_{iter} - 1)(N_{pos}H_{ort} + N_{ort}H_{sca})$). Starting with the second iteration, however, *i*MSL will examine less likely partial solution candidates obtained with the position detector thus increase the detection rate, i.e., the sensitivity of the detector.

2.2.3. Spinal disk detection

The probabilistic spine model described in the previous section is discretized using the disk candidates detected with *i*MSL. Each random variable $\boldsymbol{b}_s$ is transformed into a discrete random variable where each state represents one of the detected disk candidates. In order to allow for missed detections, an extra "missing" state is introduced. Note, that *i*MSL detects disk candidates with high sensitivity which usually results in more disk candidates than actual disks (cf. Fig. 1). The MAP estimate, i.e., the maximum of Eq. (1), provides the optimum assignment of a disk candidate to one of the disk variables according to the probabilistic spine model. Thus, only those disk candidates that form a valid spine are selected and are implicitly assigned a suitable label. The MAP is efficiently computed by belief propagation where, due to the tree structure of the factor graph (cf. Fig. 2), a single forward–backward pass yields the exact solution (Kschischang et al., 2001).

2.3. Patient-adaptive structure segmentation

Once the position, orientation and scale of the spinal disks $\boldsymbol{b}_s = (\boldsymbol{p}_s, \boldsymbol{q}_s, \boldsymbol{s}_s), s \in \{1, \ldots, N\}$ are found, they can be exploited for initializing and guiding a detailed segmentation of disks and vertebrae in MR and CT respectively. Since the vertebral disks and bodies in one patient data set exhibit similar contrasts and thus share the same basic appearance, we propose the following case-adaptive segmentation method based on the graph cuts framework (Boykov and Funka-Lea, 2006) as an optional post-processing step.

In this formulation, graph cuts are used to efficiently determine the MAP estimate of a Markov random field by minimizing an en-

ergy function defined by submodular potential functions (Kolmogorov and Zabin, 2004) that weighs between a local data-based likelihood term and a global smoothness term:

$$\boldsymbol{Y}^* = \arg \min_{\boldsymbol{Y}} \left(\sum_{i=1}^M V(x_i | y_i) + \sum_{i \sim j} V_{ij}(y_i, y_j) \right) \tag{9}$$

where i and $j \in \{1, \ldots, M\}$ are indices to an image voxel, x_i is the observed intensity at voxel i and $i \sim j$ denotes neighboring voxels according to the employed 26-neighborhood system on the regular 3D lattice. The vector of fore-/background labels y_i for all voxels in the examined subvolume is denoted by $\boldsymbol{Y} \in \{+1, -1\}^M$.

Here we use the spatially varying pair potential proposed by Boykov and Funka-Lea (2006),

$$V_{ij}(y_i, y_j) \propto \exp \left(-\frac{(x_i - x_j)^2}{2\sigma} \right) \cdot \frac{\delta_{y_i \neq y_j}}{\mathrm{dist}(i, j)}, \tag{10}$$

where x_i and x_j denote the observed intensities at voxels i and j, respectively, and the Kronecker delta $\delta_{y_i \neq y_j}$ is one for $y_i \neq y_j$ and zero otherwise. The function dist (i, j) denotes the physical distance between voxels i and j, which varies when working on image volumes with anisotropic voxel spacing. The model emphasizes homogeneous classifications among neighboring voxels but weighs penalties for heterogeneity according to intensity similarities of the voxels involved. It assumes the noise among neighboring voxels of an input volume to be distributed in a Gaussian manner. Discontinuities between voxels of similar intensities are penalized if the intensity difference $|x_i - x_j|$ is smaller than σ. However, if the voxel intensities are very different, that is to say, the difference $|x_i - x_j|$ is larger than σ the penalty is small.

The single site potentials $V(x_i | y_i)$ are derived from a probabilistic model that is estimated from the detected spinal disks and vertebrae for each individual patient prior to the structure segmentation. Let c_s be the index closest to the detected disk position $\boldsymbol{p}_s$, $\mathcal{F}_s = \{j \in \{1, \ldots, M\} | \mathrm{dist}(j, c_s) \leqslant r\}$ the set of voxel indices within a certain radius r around c_s, and $\mathcal{B}_s = \{j \in \{1, \ldots, M\} | j \in \mathcal{C}_s\}$ the set of voxel indices covered by the surface $\mathcal{C}_s$ of $\boldsymbol{b}_s$. Certainly, the radius r should be chosen small enough to not exceed the size of the object to be segmented. The combined random samples $\mathcal{F} = \bigcup_{t=1}^N \mathcal{F}_t$ for the foreground and $\mathcal{B} = \bigcup_{t=1}^N \mathcal{B}_t$ for the background over all detected objects, i.e., vertebral disks or bodies, are used for histogram estimates $\widehat{\Pr}(x | +1)$ and $\widehat{\Pr}(x | -1)$ of the associated distributions of fore- and background intensities, respectively. Also taking into account fore- and background seed points we have,

$$V(x_i | y_i = +1) = \begin{cases} 0 & \text{if } i \in \mathcal{F}, \\ \infty & \text{if } i \in \mathcal{B}, \\ -\log \widehat{\Pr}(x_i | +1) & \text{otherwise,} \end{cases} \tag{11}$$

and

$$V(x_i | y_i = -1) = \begin{cases} \infty & \text{if } i \in \mathcal{F}, \\ 0 & \text{if } i \in \mathcal{B}, \\ -\log \widehat{\Pr}(x_i | -1) & \text{otherwise.} \end{cases} \tag{12}$$

By computing the observation models based on detection results obtained on-the-fly during processing, it is ensured that the fore- and background models accurately characterize the intensity distributions that are present in the current patient data set. This way, a case-adaptive segmentation model is generated that is more precise than models generated off line from a collection of patient data sets. Due to degenerative diseases or slightly changed acquisition parameters, fore- and background models may vary a lot between different patient data sets. Imperfections of individual detection results

are compensated as the models are computed from multiple structures.

3. Experimental

3.1. Data

Experiments have been conducted based on 42 MR and 30 CT volumes. T_1-weighted MR volumes (FL3D-VIBE sequence) were obtained from 42 healthy volunteers. About one half of the volumes has been acquired on two 1.5T scanner models (MAGNETOM Avanto and MAGNETOM Espree, Siemens AG, Erlangen) with TR = 5/4 ms, TE = 2 ms and a flip angle of 10°. The other half has been obtained from two 3T scanner models (MAGNETOM Trio, MAGNETOM Verio, Siemens AG, Erlangen) with TR = 4/3 ms, TE = 1 ms and again a flip angle of 10°. Each of the volumes was recorded in a two station scan and subsequently combined to a volume covering the whole spine (approximately 860 mm × 350 mm × 190 mm). Susceptibility artifacts and intensity variations due to magnetic field inhomogeneities were present in the data. No bias field correction was performed.

CT data covering the thoracic and lumbar spine was acquired from 30 patients with bone lesions. It was reconstructed in sagittal slices, most at about 3 mm distance (three volumes with 2 mm and one with 1.2 mm), using various medium to hard convolution kernels (B50f, B60f, B70f). The in-plane resolution was between 0.48 mm and 1.70 mm. Various hyperdense disease patterns were present in the data stemming from osteoblastic bone lesions and diffuse sclerotic areas within the vertebral bodies. In some patients, metallic spinal fixators yielded severe streaking artifacts that contributed to the challenge of this data set.

Both, CT and MR volumes were resampled to an isotropic resolution of 2.1 mm for spinal disk detection and to an isotropic resolution of 6 mm for spine part localization. The segmentation experiments were performed in the resolution of the original reconstruction.

3.2. Ground truth and hyperparameters

Each spinal disk has been manually annotated with four defined landmarks. From these, ground truth boxes have been derived for the spinal disks as well as for the lumbar, thoracic and cervical spine regions.

For spine part detection, standard MSL was run using $N_{pos} = 500$ position, $N_{ort} = 50$ orientation and $N_{sca} = N_A = 20$ scale candidates. For hypothesis generation we used position hypotheses at the volume resolution of (6 mm), orientation hypotheses at a resolution of 3.3° and scale hypotheses at 6 mm resolution.

For disk detection, *i*MSL was employed with a cluster radius of $R = 6$ mm, $N_0 = 3000$ initial position candidates and *500* candidates for the remaining detection steps ($N_{pos} = 500$, $N_{ort} = 500$, $N_{sca} = 500$). In each cluster, the top $N_A = 20$ candidates were then averaged to obtain a disk candidate. Hypotheses were again generated at volume resolution (2.1 mm), orientation hypotheses again at 3.3° and scale hypotheses at 4.0 mm resolution.

Note that the choice of hyperparameters is not critical for the performance of *i*MSL as long as they are within a reasonable range. The cluster radius R, for example, determines the minimum distance that can be obtained between any two detections. For the cervical disks we measured a distance of 18 ± 2 mm in our training data which means that the cluster radius R should certainly be smaller than 9 mm to avoid merging candidates from neighboring disks. The number of initial position candidates N_0, for example, should exceed a certain minimum number. It can be estimated

from a few examples by examining the minimum number of candidates required to cover all disks after position detection which we did on some MR volumes. Detailed results from an experimental analysis of the influence of the hyperparameters are provided in the following.

4. Results

All evaluation results have been obtained using 10-fold cross validation, ensuring that training and testing data never originated from the same subject. Every ground truth annotation for which no disk within a distance of 10 mm was detected, was counted as a missed detection.

In MR, spinal disks were detected with a sensitivity of 98.64% and only 0.0731 false positives per volume, yielding a positive predictive value of 99.68%. The overall processing time on a 2.2 GHz dual core laptop computer was between 9.9 s and 13.0 s, and 11.5 s on average where most of the time was spent on disk candidate detection. In CT, a sensitivity of 98.04% was achieved while 0.267 false positives per volume were obtained, yielding a positive predictive value of 98.43%.

The accuracy of the detected spinal disks has been evaluated by the position distance and the angle between the disk plane normals of the detected spinal disks and the ground truth annotation (cf. Table 1). On average, a position error of 2.42 mm (about 1 voxel) and an angular error of 3.85° was obtained for the MR data. For the CT data, the position error of 3.22 mm and the angular error of 4.47° was slightly worse which can be attributed to both, the smaller data set available for training and the presence of pathologies and artifacts (cf. Fig. 5).

Four examples from the MR data set are shown in Fig. 4. The right-most example shows a case where the volunteer has been instructed to lie down twisted in order to obtain a spine recording with unusual pose. Still the proposed approach could locate and label all spinal disks reliably.

Fig. 5 shows four examples from the CT data set. The examples exhibit various spinal pathologies and age-related degenerations. Nevertheless, our method could locate and label the spinal disks with high reliability. The right-most example shows that even in the presence of severe metal artifact, our spinal disk detector is sensitive and robust enough to detect the spinal disks that are hardly visible. However, our method fails to detect the last spinal disk between L5 and the sacrum, resulting in a shifted labeling of the detected disks in this example. A detailed analysis reveals that the L5/S1 disk is detected with *i*MSL but not selected with the global prior model.

The influence of different hyperparameters on the performance of the proposed *i*MSL algorithm is visualized in Fig. 6. Each graph displays the behavior of position error, angular error and sensitivity when varying one of the hyperparameters while keeping all others fixed to their default values ($N_A = 20$, $R = 6$ mm, $N_0 = 3000$,

Table 1
Disk detection results using 10-fold cross validation based on 42 T_1-weighted MR volumes and 30 CT volumes. Left: position error. Right: angular error between normals.

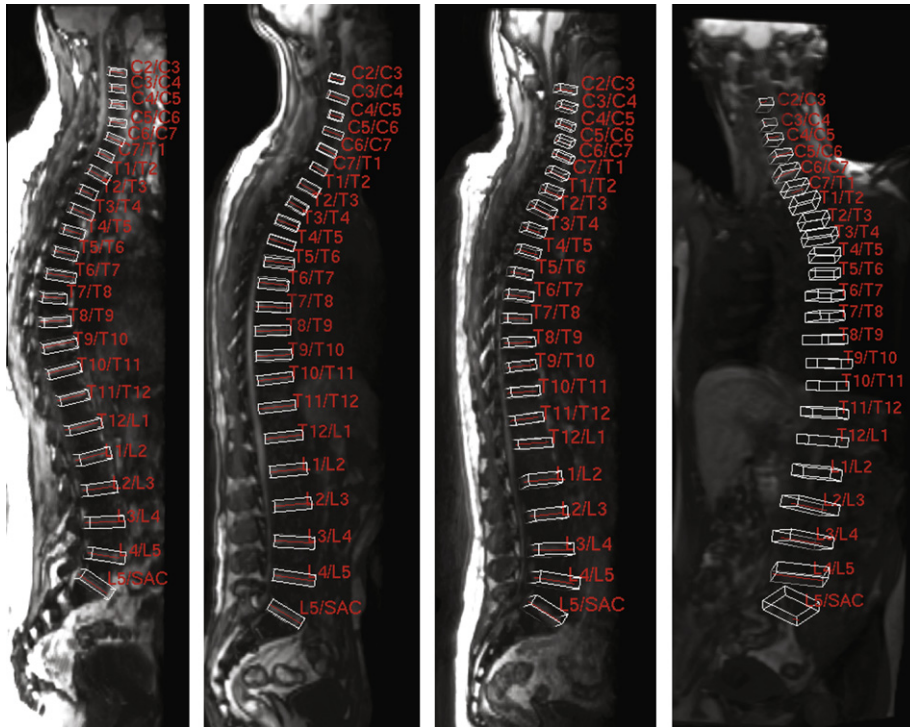|		Position error (mm)				Angular error (degree)			
		Cerv.	Thor.	Lumb.	Avg.	Cerv.	Thor.	Lumb.	Avg.
MR	Mean	2.09	2.41	2.86	2.42	4.86	3.38	3.80	3.85
	Stdev	1.06	1.33	1.30	1.28	3.34	2.11	2.40	2.62
	Median	1.84	2.18	2.68	2.19	3.89	2.90	3.37	3.17
CT	Mean	–	3.40	2.80	3.22	–	4.77	3.72	4.47
	Stdev	–	1.78	1.59	1.75	–	3.07	2.57	2.97
	Median	–	3.11	2.54	2.94	–	4.32	3.21	3.97

Fig. 4. Four examples from the MR data with detection results. Although the volunteer in the rightmost example lay down in an unusual pose, all spinal disks were detected and labeled correctly.
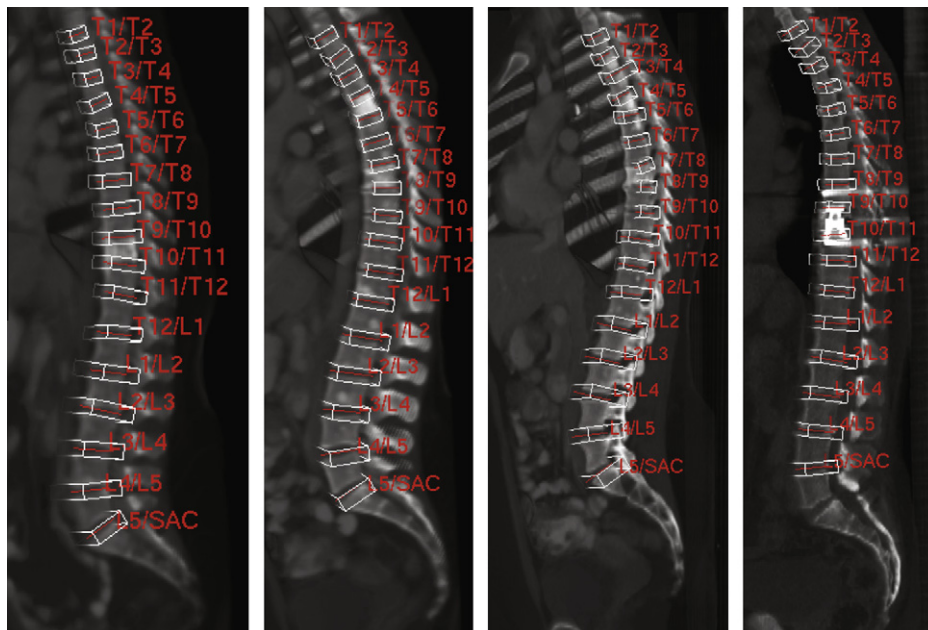

Fig. 5. Four examples from the CT data set with detection results. The hyperdense areas within some of the vertebral bodies are either well-defined osteoblastic bone lesions or diffuse sclerotic areas. The right-most example shows that our method can also cope with severe artifacts, here stemming from a spinal fixation. Note however, that in this example the last spinal disk was missed and the labeling was erroneously shifted by one.

$N_{pos} = N_{ort} = N_{sca} = 500$). All three performance measures are provided relative to their values with default parameters which are marked with vertical, red and dotted lines. The top left graph in Fig. 6 shows that, as expected, sensitivity starts dropping and the position error deteriorates for an aggregation radius R above 10 mm (the distance between cervical disks is about 18 mm). For very small values (<2 mm) the position error also increases while the angular error remains nearly constant across all tested values.

The evaluation of the number of aggregation candidates N_A reveals that our default choice was not optimal and that all performance measures could be improved by just increasing that number. For the number of initial position candidates N_0 our chosen default value is very close to the minimum required number since any smaller choice would quickly decrease sensitivity and increase both position and angular error. Any bigger choice, however, would be safe and leave the performance measures unaltered. A similar
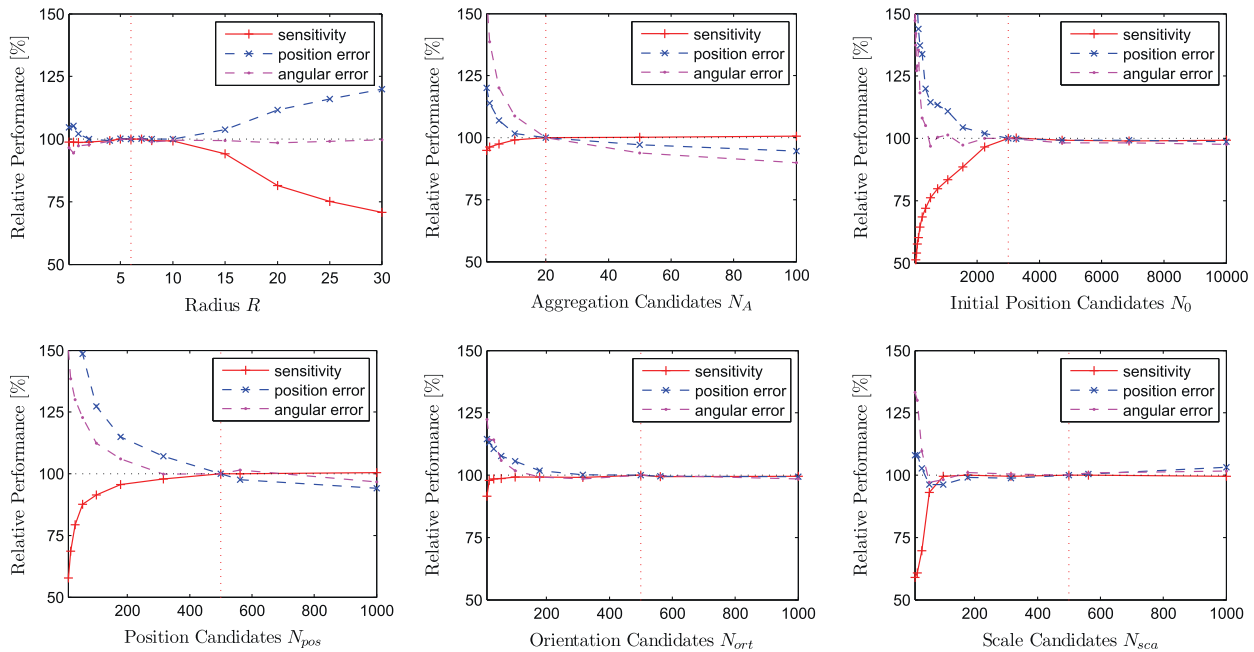
Fig. 6. Analysis of the influence of individual hyperparameters on the position error, angular error and sensitivity of the proposed algorithm on the CT data. All performance measures are reported in percent relative to the performance obtained with the default parameters employed for all other evaluations, which are indicated by the vertical, red and dotted lines. The top row shows from left to right results for the cluster radius R, the number of aggregation candidates N_A and the number of initial position candidates N_0. The bottom row shows results for varying the number of position candidates N_{pos}, orientation candidates N_{ort} and scale candidates N_{sca} used in the iteration of the iMSL algorithm.

behavior is observed for the number of position (N_{pos}), orientation (N_{ort}) and scale candidates (N_{sca}) considered in each iMSL iteration. While their default choice appears not to be optimal, e.g. N_{ort} and N_{sca} could be chosen smaller, a broad range of values beyond a certain minimum number does not impair the accuracy of the iMSL algorithm.

Exemplary spinal disk segmentation results for the vertebrae in CT and the spinal disks in MR are presented in Fig. 7. Due to the renunciation of an overly restrictive shape prior, the obtained segmentations accurately delineate spinal disks and vertebrae also in the presence of pathologic shape deformations as seen in the CT examples. The patient-adaptive appearance model helps to cope
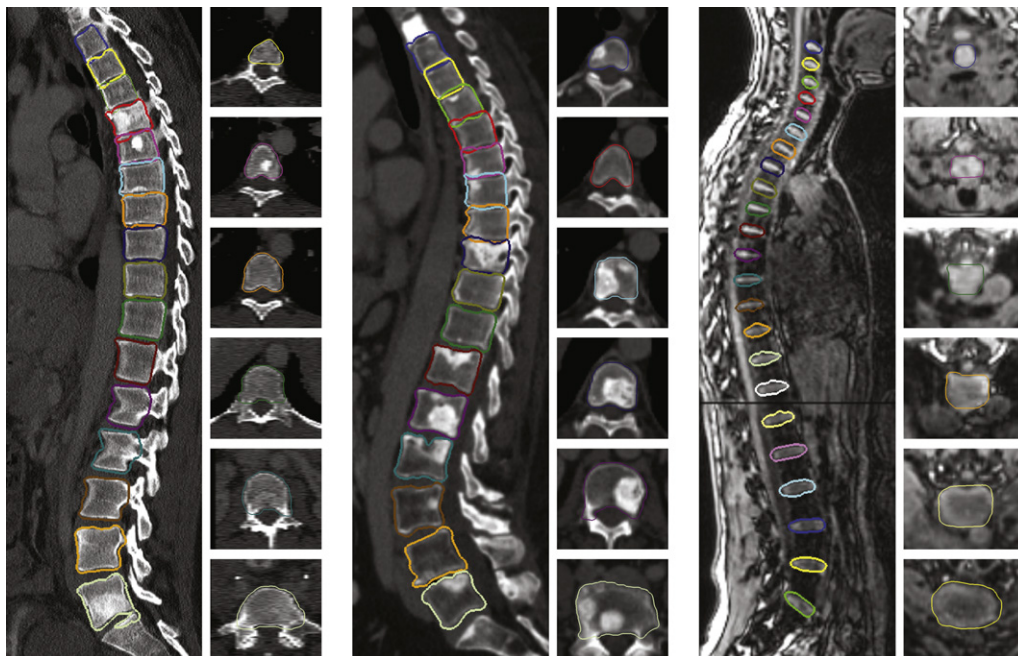

Fig. 7. Segmentation results. The two left examples show segmentations of the vertebral bodies obtained for two of the CT bone lesion patients. The right example shows segmentations of the spinal disks as obtained on one of the MR data sets. Even in the presence of severe pathologic degenerations like spinal disk herniations the vertebral bodies are well delineated.

with various patient-specific intensity characteristics that differ from the standard due to pathology or degenerations caused by old age (cf. Fig. 7).

5. Discussion

The results of our proposed method compare favorably with results presented in previous works. While with only 6s processing time the approach by Pekar et al. (2007) runs faster than ours, it has lower sensitivity (95.6% before candidate selection) and does not provide orientation estimates.

Compared with the best cross validation results on MR data by Schmidt et al. (2007), the results obtained with our approach are significantly better. While a competitive but still inferior sensitivity of 97% is reported, they only achieve a position error of 5.1 mm. Furthermore, no orientation estimates are provided and the approach takes several minutes to run. In contrast to Schmidt et al. (2007), we did not perform any posterior search at the positions of missing disks which could further increase our sensitivity.

Although Klinder et al. (2009) do not provide position errors for their vertebra detection step, the example shown in Figure 9 of their article shows a very blurred Hough score and considerable variance in the detected local coordinate systems which suggests that their localization error is considerably bigger than the 3.2 mm achieved by our method. On average, their method requires 27.9 s for candidate generation and another 36.5 min for vertebra identification. Further considering the required processing steps including two Hough transformations and the generation of a curved planar reformation, their method appears to be computationally much more demanding than ours. Unlike our approach, the method proposed by Klinder et al. (2009) works on volumes with varying scan ranges and can reliably identify vertebrae if more than 16 of them are within the scanned range. To this end, our current approach still requires external information, e.g. from user input, from the employed scan protocol or from an automatic scan range detector (Emrich et al., 2010). A combination of our method with a vertebra identification model similar to Klinder et al. (2009) is also conceivable and might be a promising approach if its computational requirements can be reduced.

Our results provide evidence that the presented approach also works in the presence of pathological deformations of the spinal column (such as various types of scoliosis, kyphosis, and lordosis) and pathologies that result in unusual appearance of individual disks (e.g. degeneration, herniation, desiccation, etc.) and vertebral bodies (e.g. compression fractures, hemivertebrae, diffuse sclerosis, etc.). In this as well as other applications we have observed, that MSL is very robust to imaging artifacts and unusual appearances of the sought object. Using iMSL increases sensitivity and helps detect disks with very unusual appearance. The CT examples in Fig. 5 show cases in which disks between diseased vertebrae and disks in the presence of severe artifacts can reliably be detected. Furthermore, since the global spine model is restricted to candidates provided by the disk detector, abnormalities such as scoliosis, kyphosis, lordosis, compression fractures and hemivertebrae can be robustly handled. The volunteer with the unusual pose in Fig. 4 provides evidence towards this. Simple retraining of our system with more abnormal cases added, enables the detectors as well as the prior model to handle pathologies and degenerations even more reliably.

6. Conclusion

In this article, we have presented a novel approach to the fully automatic detection of 3D spinal geometry and labeling of the spinal disks. The approach uses an iterative extension of MSL for disk candidate detection along with an anatomical network that incorporates spatial context in form of a prior on the nine dimensional disk poses. Since the entire approach is learning-based, it can be trained for CT and MR alike.

Using 42 MR and 30 CT volume data sets, superior sensitivity and localization accuracy was obtained as compared to previous works. With an overall processing time of only 11.5 s, the approach is sufficiently fast to be used in various clinical applications, including automatic scan alignment, computer-aided diagnosis and decision support systems.

References

Adams, M.A., Dolan, P., 2005. Spine biomechanics. J. Biomech. 38, 1972–1983.

Adams, M.A., Dolan, P., 2011. Biomechanics of vertebral compression fractures and clinical application. Arch. Orthop. Trauma Surg. 131, 1703–1710.

Alomari, R.S., Corso, J.J., Chaudhary, V., 2011. Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. IEEE Trans. Med. Imag. 30, 1–10.

Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., Ayache, N., 2008. Geometric variability of the scoliotic spine using statistics on articulated shape models. IEEE Trans. Med. Imag. 27, 557–568.

Boykov, Y., Funka-Lea, G., 2006. Graph cuts and efficient N–D image segmentation. Int. J. Comput. Vis. 70, 109–131.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Chen, T., Zhang, W., Good, S., Zhou, K.S., Comaniciu, D., 2009. Automatic ovarian follicle quantification from 3d ultrasound data using global/local context with database guided segmentation. In: Proc. ICCV, pp. 795–802.

Corso, J.J., Alomari, R.S., Chaudhary, V., 2008. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Proc. MICCAI, pp. 202–210.

Coulon, O., Hickman, S.J., Parker, G.J., Barker, G.J., Miller, D.H., Arridge, S.R., 2002. Quantification of spinal cord atrophy from magnetic resonance images via a B-spline active surface model. Magn. Reson. Med. 47, 1176–1185.

Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E., 2011. Regression forests for efficient anatomy detection and localization in CT studies. In: Proc. 2010 MICCAI workshop on Medical Computer Vision. Springer-Verlag, Berlin, Heidelberg, pp. 106–117.

Dietterich, T.G., Hao, G., Ashenfelter, A., 2008. Gradient tree boosting for training conditional random fields. J. Mach. Learn. Res. 9, 2113–2139.

Donner, R., Birngruber, E., Steiner, H., Bischof, H., Langs, G., 2011. Localization of 3D anatomical structures using random forests and discrete optimization. In: Proc. 2010 MICCAI Workshop on Medical Computer Vision. Springer-Verlag, Berlin, Heidelberg, pp. 86–95.

Emrich, T., Graf, F., Kriegel, H.P., Schubert, M., Thoma, M., Cavallaro, A., 2010. CT slice localization via instance-based regression. In: Proc. SPIE Medical Imaging, p. 762320.

Feng, S., Zhou, S., Good, S., Comaniciu, D., 2009. Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information. In: Proc. CVPR, pp. 2488–2495.

Georgescu, B., Zhou, X.S., Comaniciu, D., Gupta, A., 2005. Database-guided segmentation of anatomical structures with complex appearance. In: Proc. CVPR, pp. 429–436.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York.

Huang, S.H., Chu, Y.H., Lai, S.H., Novak, C.L., 2009. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. IEEE Trans. Med. Imag. 28, 1595–1605.

Huang, S.H., Lai, S.H., Novak, C., 2008. A statistical learning approach to vertebra detection and segmentation from spinal MRI. In: Proc. ISBI, pp. 125–128.

Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Hornegger, J., Navab, N., Comaniciu, D., 2009. Personalized modeling and assessment of the aortic-mitral coupling from 4D TEE and CT. In: Proc. MICCAI, pp. 767–775.

Jäger, F., Hornegger, J., Schwab, S., Janka, R., 2009. Computer-aided assessment of anomalies in the scoliotic spine in 3-D MRI images. In: Proc. MICCAI, pp. 819–826.

Karney, C.F., 2007. Quaternions in molecular modeling. J. Molec. Graph. Modell. 25, 595–604.

Kelm, B.M., Pal, C., McCallum, A., 2006. Combining generative and discriminative methods for pixel classification with multi-conditional learning. In: Proc. ICPR, Hong Kong. pp. 828–832.

Kelm, B.M., Zhou, S.K., Suehling, M., Zheng, Y., Wels, M., Comaniciu, D., 2010. Detection of 3D spinal geometry using iterated marginal space learning. In: Proc. 2010 MICCAI Workshop on Medical Computer Vision. Springer-Verlag, Berlin, Heidelberg, pp. 96–105.

Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C., 2009. Automated model-based vertebra detection, identification, and segmentation in CT images. Med. Image Anal. 13, 471–482.

Koller, D., Friedman, N., 2009. Probabilistic Graphical Models. Principles and Techniques. Adaptive Computation and Machine Learning. The MIT Press.

Kolmogorov, V., Zabin, R., 2004. What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26, 147–159.

Kschischang, F.R., Frey, B.J., Loeliger, H.A., 2001. Factor graphs and the sum-product algorithm. IEEE Trans. Inf. Theory 47, 498–519.

Lafferty, J., McCallum, A., Pereira, F., 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proc. ICML, pp. 282–289.

Lasserre, J.A., Bishop, C.M., Minka, T.P., 2006. Principled hybrids of generative and discriminative models. In: Proc. CVPR. IEEE Computer Society, Washington, DC, USA, pp. 87–94.

McIntosh, C., Hamarneh, G., 2006. Spinal crawlers: deformable organisms for spinal cord segmentation and analysis. In: Proc. MICCAI, pp. 808–815.

Pekar, V., Bystrov, D., Heese, H.S., Dries, S.P.M., Schmidt, S., Grewer, R., den Harder, C.J., Bergmans, R.C., Simonetti, A.W., van Muiswinkel, A.M., 2007. Automated planning of scan geometries in spine MRI scans. In: Proc. MICCAI, pp. 601–608.

Peng, Z., Zhong, J., Wee, W., Lee, J.H., 2006. Automated vertebra detection and segmentation from the whole spine MR images. In: Proc. IEEE Int. Conf. Eng. Med. Biol. Soc., pp. 2527–2530.

Pennec, X., 2006. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. J. Math. Imag. Vis. 25, 127–154.

Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S., Bystrov, D., Schnörr, C., 2007. Spine detection and labeling using a parts-based graphical model, in: Proc. IPMI, pp. 122–133.

Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D., 2009. Hierarchical parsing and semantic navigation of full body CT data. In: Proc. SPIE Medical Imaging, pp. 725–732.

Seifert, S., Dillmann, R., 2007. Biomechanical modeling of the cervical spine on the basis of tomographic data. Biomed. Tech. 52, 337–345.

Seifert, S., Kelm, M., Möller, M., Mukherjee, S., Cavallaro, A., Huber, M., Comaniciu, D., 2010. Semantic annotation of medical images. In: Proc. SPIE Medical Imaging.

Sofka, M., Ralovich, K., Birkbeck, N., Zhang, J., Zhou, S.K., 2011. Integrated detection network (IDN) for pose and boundary estimation in medical images. In: Proc. IEEE Int. Biomed. Imag.: From Nano to Macro Symp., pp. 294–299.

Tschirhart, C.E., Finkelstein, J.A., Whyne, C.M., 2007. Biomechanics of vertebral level, geometry, and transcortical tumors in the metastatic spine. J. Biomech. 40, 46–54.

Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A., 2008. Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans. Med. Imag. 27, 495–508.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: Proc. CVPR.

Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D., 2008. A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-D MRI. In: Proc. MICCAI, pp. 67–75.

Wels, M., Kelm, B.M., Tsymbal, A., Hammon, M., Soza, G., Sühling, M., Cavallaro, A., Comaniciu, D., 2012. Multi-stage osteolytic spinal bone lesion detection from CT data with internal sensitivity control. In: van Ginneken, B., Novak, C.L. (Eds.), Proc. SPIE Medical Imaging, pp. 831513–1.

Wels, M., Zheng, Y., Carneiro, G., Huber, M., Hornegger, J., Comaniciu, D., 2009. Fast and robust 3-D MRI brain structure segmentation. In: Proc. MICCAI, pp. 575–583.

Yao, J., O'Connor, S.D., Summers, R.M., 2006. Automated spinal column extraction and partitioning. In: Proc. ISBI, pp. 390–393.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2008. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imag. 27, 1668–1681.

Zheng, Y., Georgescu, B., Ling, H., Zhou, S.K., Scheuering, M., Comaniciu, D., 2009a. Constrained marginal space learning for efficient 3D anatomical structure detection in medical images. In: Proc. CVPR, pp. 194–201.

Zheng, Y., Lu, X., Georgescu, B., Littmann, A., Mueller, E., Comaniciu, D., 2009b. Robust object detection using marginal space learning and ranking-based multi-detector aggregation: application to left ventricle detection in 2D MRI images. In: Proc. CVPR, pp. 1343–1350.

Zhou, S.K., Guo, F., Park, J.H., Carneiro, G., Jackson, J., Brendel, M., Simopoulos, C., Otsuki, J., Comaniciu, D., 2007. A probabilistic, hierarchical, and discriminant framework for rapid and accurate detection of deformable anatomic structure. In: Proc. ICCV, pp. 1–8.