Robust MR Spine Detection Using Hierarchical Learning and Local Articulated Model

Yiqiang Zhan[1], Dewan Maneesh[1], Martin Harder[2], and Xiang Sean Zhou[1]

[1] Siemens Medical Solutions USA, Inc., Malvern, USA
[2] Siemens Healthcare Imaging MR, Erlangen, German

Abstract. A clinically acceptable auto-spine detection system, i.e., localization and labeling of vertebrae and inter-vertebral discs, is required to have high robustness, in particular to severe diseases (e.g. scoliosis) and imaging artifacts (e.g. metal artifacts in MR). Our method aims to achieve this goal with two novel components. *First*, instead of treating vertebrae/discs as either repetitive components or completely independent entities, we emulate a radiologist and use a *hierarchial* strategy to learn detectors dedicated to anchor (distinctive) vertebrae, bundle (non-distinctive) vertebrae and inter-vertebral discs, respectively. At run-time, anchor vertebrae are detected concurrently to provide redundant and distributed appearance cues robust to local imaging artifacts. Bundle vertebrae detectors provide candidates of vertebrae with subtle appearance differences, whose labels are mutually determined by anchor vertebrae to gain additional robustness. Disc locations are derived from a cloud of responses from disc detectors, which is robust to sporadic voxel-level errors. *Second*, owing to the non-rigidness of spine anatomies, we employ a *local articulated* model to effectively model the spatial relations across vertebrae and discs. The local articulated model fuses appearance cues from different detectors in a way that is robust to abnormal spine geometry resulting from severe diseases. Our method is validated by 300 MR spine scout scans and exhibits robust performance, especially to cases with severe diseases and imaging artifacts.

1 Introduction

As one of the major organs in the human body, spine relates to various neurological, orthopaedic and oncological studies. Magnetic resonance imaging (MR) is often preferred for spine imaging due to the high contrast between soft tissues. However, MR imaging quality is highly dependent on the position and orientation of the slice group. For example, a high-res transversal slice group should be positioned in parallel to inter-vertebral disc and centered at the junction of spinal cord. In current MR workflow, high-res slice group positioning is performed manually in a 2D/3D scout scan. Compared to 2D scout, 3D scout provides comprehensive anatomical context, which facilitates slice group positioning even in strong scoliotic cases (c.f., Fig. 3a). However, the manual positioning in 3D scout also takes more time due to cross slice navigation. Therefore,

N. Ayache et al. (Eds.): MICCAI 2012, Part I, LNCS 7510, pp. 141–148, 2012.

automatic spine detection in 3D scout becomes very desirable to improve MR spine workflow.

Automatic spine detection work in MR can be traced back to the 1980's [1], where a heuristic algorithm is designed to detect lumbar discs in 2D MR slices. Alomari et.al. [2] proposed a 2D lumbar vertebrae labeling system incorporating appearance and geometrical priors. However, more complicated spine geometry in 3D (especially for disease cases), and smaller/challenging appearance of cervical vertebrae, would make this approach limiting for 3D MR whole spine labeling. One of the first 3D whole spine detection methods was proposed by Schmidt et. al. [3]. Local appearance cues learned by random trees are combined with non-local geometrical priors modeled by a parts-based graphical model. Another interesting method presented in [4] focuses on learning disc location in a nine dimensional transformation space. Iterative marginal space learning is proposed to generate candidates comprising position, orientation, and scale, which are further pruned by an anatomical network. In general, state-of-the-art methods did achieve certain robustness by combining low-level appearance and high-level geometry information. However, in the presence of severe imaging artifacts or spine diseases (see Fig.3a), which are more common in 3D MR scout scans, none of existing methods provides evidence of handling these cases robustly. (Note that spine detection algorithms for other imaging modalities [5] may not be borrowed to MR owing to the intrinsically different appearances.)

In fact, two unique characteristics of spine anatomies are mostly ignored in previous works. First, although spine is composed of repetitive components (vertebrae and discs), these components have different distinctiveness and reliability in terms of detection. Second, spine is a non-rigid structure, where local articulations exist in-between vertebrae and discs. This articulation can be quite large in the presence of certain spine diseases. An effective geometry modeling should not consider vertebrae detections from scoliotic cases as errors just because of the abnormal geometry. Building upon these ideas, in this paper, we propose a spine detection method by exploiting these two characteristics. Instead of learning a general detector for vertebrae/discs or treating them as completely independent entities, we use a hierarchical strategy to learn "distinctiveness adaptive" detectors dedicated to anchor vertebrae, bundle vertebrae and inter-vertebral discs, respectively. These detectors are fused with a local articulated model to propagate information from different detectors handling abnormal spine geometry. With the hallmarks of *hierarchical learning* and *local articulated model*, our method becomes highly robust to severe imaging artifacts and spine diseases.

2 Method

2.1 Problem Statement

Notations: Human spine usually consists of 24 articulated vertebrae, which can be grouped as cervical (C_1-C_7), thoracic (T_1-T_{12}) and lumbar (L_1-L_5) sections. These 24 vertebrae plus the fused sacral vertebrae (S_1) are the targets of spine labeling in most clinical practices.

We define vertebrae and inter-vertebral discs as $V = \{v_i | i = 1 \cdots N\}$ and $D = \{d_i | i = 1 \cdots N-1\}$, where v_i is the *i-th* vertebra and d_i is the inter-vertebral disc between the *i-th* and *i+1-th* vertebra. Here, $v_i \in \mathbb{R}^3$ is the vertebra center and $d_i \in \mathbb{R}^9$ includes the center, orientation and size of the disc. It is worth noting that i is not a simple index but bears anatomical definition. In this paper, without loss of generality, v_i is indexed in the order of vertebrae from head to feet, e.g., v_1, v_{24}, v_{25} represent C_1, L_5 and S_1, respectively.

Formulation: Given an image I, spine detection problem can be formulated as the maximization of a posterior probability with respect to V and D as:

$$(V^*, D^*) = arg \max_{V,D} P(V, D|I) \tag{1}$$

Certain vertebrae that appear either at the extremity of the entire vertebrae column, e.g., C_2, S_1, or at the transition regions of different vertebral sections, e.g., L_1 , have much better distinguishable characteristics (red ones in Fig. 1(a)). The identification of these vertebrae helps in the labeling of others, and are defined as "*anchor vertebrae*". The remaining vertebrae (blue ones in Fig. 1(a)) are grouped into a set of continuous "bundles" and hence defined as "*bundle vertebrae*". Vertebrae characteristics are different across bundles but similar within a bundle, e.g., C_3-C_7 look similar but are very distinguishable from T_8-T_{12}.

Denoting $V_{\mathcal{A}}$ and $V_{\mathcal{B}}$ as anchor and bundle vertebrae, the posterior in Eq. 1 can be rewritten and further expanded as:

$$P(V, D|I) = P(V_{\mathcal{A}}, V_{\mathcal{B}}, D|I) = P(V_{\mathcal{A}}|I) \cdot P(V_{\mathcal{B}}|V_{\mathcal{A}}, I) \cdot P(D|V_{\mathcal{A}}, V_{\mathcal{B}}, I) \tag{2}$$

In this study, we use Gibbs distributions to model the probabilities. The logarithm of Eq. 2 can be then derived as Eq. 3.

$$\begin{aligned} \log[P(V, D|I)] = & A_1(V_{\mathcal{A}}|I) && \Longleftarrow P(V_{\mathcal{A}}|I) \\ & + A_2(V_{\mathcal{B}}|I) + S_1(V_{\mathcal{B}}|V_{\mathcal{A}}) && \Longleftarrow P(V_{\mathcal{B}}|V_{\mathcal{A}}, I) \\ & + A_3(D|I) + S_2(D|V_{\mathcal{A}}, V_{\mathcal{B}}) && \Longleftarrow P(D|V_{\mathcal{A}}, V_{\mathcal{B}}, I) \end{aligned} \tag{3}$$

Here, A_1, A_2 and A_3 relate to the appearance characteristics of anchor, bundle vertebrae and inter-vertebral discs. S_1 and S_2 describe the spatial relations of anchor-bundle vertebrae and vertebrae-disc, respectively. It is worth noting that the posterior of anchor vertebrae solely depends on the appearance term, while those of bundle vertebrae and inter-vertebral discs depend on both appearance and spatial relations. This is in accordance to the intuition: while anchor vertebrae can be identified based on its distinctive appearance, bundle vertebrae and inter-vertebral discs have to be identified using both appearance characteristics and the spatial relations to anchor ones.

Fig. 1(b) gives a schematic explanation of Eq. 3. Our framework consists of three layers of appearance models targeting to anchor, bundle vertebrae and discs.

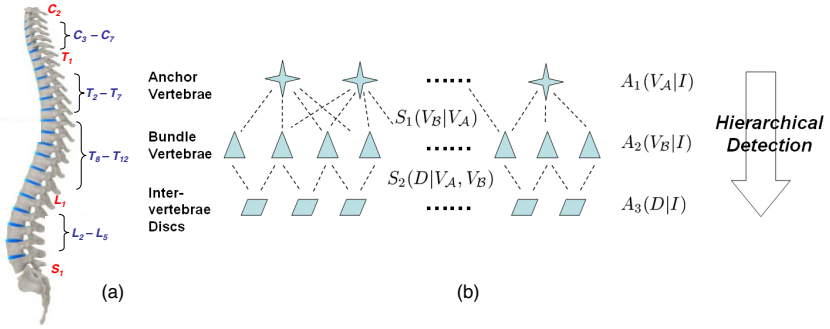

Fig. 1. (a) Schematic explanation of anchor(red) and bundle(blue) vertebrae. (b) Proposed spine detection framework.

The spatial relations across different anatomies "bridge" different layers (lines in Fig. 1). Note that this framework is completely different from the two-level model of [2], which separates pixel- and object-level information. Instead, different layers of our framework target to anatomies with different appearance distinctiveness.

2.2 Hierarchical Learning Framework

Building upon the the success of learning-based anatomy detection work [6], we employ the *Adaboost* cascade classification framework along with over-complete Haar wavelets features to model appearance characteristics of vertebrae and discs. However, in order to model the different characteristics of vertebrae and discs, we employ different training strategies for each scenario as discussed below.

Anchor Vertebrae: Distinctive characteristics of anchor vertebrae (red ones in Fig.1(a)) warrant that their detectors be trained in a very discriminative way with high response only around the center of that specific vertebra.

Bundle Vertebrae: Bundle vertebrae look similar to their neighbors but different from remote ones. Therefore, both the extremes of training a general detector for *all* bundle vertebrae (including distal ones), or specific detectors for neighboring vertebrae, would adversely affect the detector robustness and reliability. For example, consider a scenario where two specific detectors are trained to differentiate similar appearance T_9 and T_{10} vertebrae. If local imaging artifacts are present around T_9, T_9 detector might have highest response at T_{10}, since T_{10} is more salient than T_9 in this situation. This problem is also observed in [4], where "(standard) MSL approach may end up with detections for the most salient disks only". To avoid these issues, we employ a strategy in the middle to group similar neighboring vertebrae as several "bundles" (blue ones in Fig.1(a)). Each bundle has one detector that learns the commonality of corresponding vertebrae and distinguishes them from other bundles.

Inter-Vertebral Discs: Compared to vertebrae detection, disc detection has a high dimensional configuration space with 9 parameters. Different from [4],

Table 1. Training scheme of detectors for anchor vertebrae, bundle vertebrae and inter-vertebral discs

Detector	Positive Samples	Negative Samples	Image Alignment
Anchor vertebrae	Voxels close to the center of the ***specific*** vertebrae	Remaining voxels in the ***entire*** volume image	No alignment
Bundle vertebrae	Voxels close to the centers of ***any*** vertebrae within the bundle	Remaining voxels in the ***local*** volume image covering neighboring bundles	Aligned by anchor vertebrae
Inter-vertebral Discs	Voxels located on the disc	Remaining voxels in the ***local*** volume image covering the two neighboring vertebrae	Aligned by two neighboring vertebrae

which learns/detects a disc as a whole, we treat each voxel on the disc as an individual sample. Disc locations are derived by fitting disc response maps with principal component analysis. In this way, disc detection becomes robust to sporadic classification errors at voxel-level. Since voxels on the same disc are almost indistinguishable, similar to bundle vertebrae, all of them are"bundled" in the training stage.

To summarize, the differences in training strategies primarily exist in the selection of positive/negative samples and image alignment before feature extraction, which are outlined in Table 1. Moving down the table from anchor vertebrae to inter-vertebral discs, as the targeted anatomies become less and less distinctive, more positive samples are extracted in a more local fashion, and the image alignment becomes more and more sophisticated.

Using the above strategy, we train detectors for anchor vertebrae, bundle vertebrae and inter-vertebral discs as $\mathcal{A}_i(\mathfrak{F}(p))$, $\mathcal{B}_j(\mathfrak{F}(p))$, and $\mathcal{D}_k(\mathfrak{F}(p))$. Here, $\mathfrak{F}(p)$ denotes the over-complete Haar features extracted around voxel p, and $\mathcal{A}_i$, $\mathcal{B}_j$ and $\mathcal{D}_k$ are the trained Adaboost classifiers, which select and combine a small proportion of $\mathfrak{F}(p)$ to achieve best anatomy detection. The appearance terms in Eq. 3 are eventually concretized as $A_1(V_\mathcal{A}|I) = \sum_{v_i \in V_\mathcal{A}} \mathcal{A}_i(\mathfrak{F}(v_i))$, $A_2(V_\mathcal{B}|I) = \sum_{v_j \in V_\mathcal{B}} \mathcal{B}_j(\mathfrak{F}(v_j))$ and $A_3(D|I) = \sum_{d_k \in D} \sum_{p \in d_k} \mathcal{D}_k(\mathfrak{F}(p))$.

2.3 Local Articulated Spine Model

Recall the definition of Eq. 3, $S_1(V_\mathcal{B}|V_\mathcal{A})$ and $S_2(D|V_\mathcal{A}, V_\mathcal{B})$ model the spatial relations between anchor-bundle vertebrae and vertebrae-discs, respectively. Spine is a flexible structure where each vertebra has freedom of local articulation (see Fig. 2). The local rigid transformation can be quite large in the presence of certain spine diseases, e.g., scoliosis. Shape/gemetry modeling methods [7] that treats the object as a whole can not effectively model these local variations of the spine geometry. In our study, we employ a local articulated spine model [8][9] to describe the spatial relations across vertebrae. Assume v_i is an anchor vertebra and $\{v_{i+1}, \cdots, v_{i+M}\}$ are the subsequent bundle vertebrae. As shown in Fig. 2, the spatial relations between anchor and bundle vertebrae are modeled as

$[T_i, T_i \circ T_{i+1}, \ldots, T_i, \circ T_{i+1} \circ \ldots \circ T_{i+M-1}]$, where T_i defines a local similarity transformation between v_i and v_{i+1}. $S_1(V_\mathcal{B}|V_\mathcal{A})$ is defined as:

$$S_1(V_\mathcal{B}|V_\mathcal{A}) = \sum_i e^{-(\psi(T_i)-\mu_{T_i})^T \Xi_{T_i} (\psi(T_i)-\mu_{T_i})} + 2/(1+e^{\gamma \|\psi(T_i)-\psi(T_{i+1})\|^2}) \quad (4)$$

Here, $\psi(.)$ is an operator that converts T_i to a vector space, i.e., the rotation part of T_i is converted to its quaternion. μ_{T_i} and Ξ_{T_i} are the Frechet mean and generalized covariance of local transformation T_i, calculated as [8]. The first term contains the prior information of local transformations across *population*. The second term evaluates the difference between local T_i across the same *spine*. These two terms complement each other, such that a scoliotic spine still gets a high value of S_1 , due to the continuity of its local transformations.

Spatial configurations between vertebrae and discs, $S_2(D|V_\mathcal{A}, V_\mathcal{B})$, is modeled with two assumptions: 1) A vertebral disc is roughly perpendicular to the line connecting its neighboring vertebrae centers; and 2) Center of an inter-vertebral disc is close to the mid point of the two neighboring vertebrae centers. $S_2(D|V_\mathcal{A}, V_\mathcal{B})$ is then defined in the similar fashion as Eq. 4.

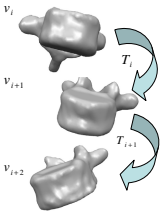

Fig. 2. Local articulation model

2.4 Hierarchical Spine Detection

As a high-dimensional and non-linear function, Eq. 3 is optimized using a multi-stage algorithm. Different stages target to anchor vertebrae, bundle vertebrae and inter-vertebral discs, respectively. In each stage, we alternatively optimize the appearance terms and spatial terms.

Fig. 1 (a) gives a more schematic explanation of the optimization procedure. This hierarchial detection scheme emulates a radiologists and guarantees the robustness in three aspects: 1) Anchor vertebrae are detected *concurrently* to provide redundant and distributed appearance cues. Even when some anchor vertebrae are missed due to severe local imaging artifacts, others still provide reliable clues for spine detection. 2) Detectors of bundle vertebrae and discs provide support cues. More specifically, instead of trying to directly derive vertebrae labels, bundle vertebrae detectors provide a set of candidates whose labels are *mutually* assigned according to relative positions to anchor vertebrae. Note that labels assigned by different anchor vertebrae might be different, and are fused through the maximization of S_1. Disc detectors return a cloud of responses for disc localization, which is robust to individual false classifications as well. 3) Local articulated model propagates these appearance cues in a way robust to abnormal spine geometry resulting from severe diseases.

Table 2. Evaluations of spine detections. LS: LSpine, CS: CSpine, WS: WholeSpine

	W/O Hierarchy			W/O Articulation			Proposed Method		
	Perfect	Accept	Reject	Perfect	Accept	Reject	Perfect	Accept	Reject
LS	85	10	5	93	5	2	98	2	0
CS	65	11	4	76	3	1	79	0	1
WS	97	16	7	106	8	6	116	2	2
All	247	37	16	275	16	9	**293**	**4**	**3**
	82.4 %	12.3%	5.3%	91.7%	5.3%	3.0%	**97.7%**	**1.3%**	**1.0%**

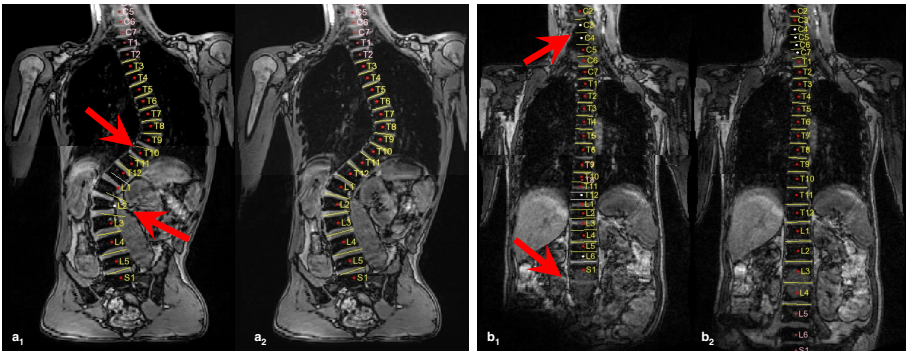

Fig. 3. Comparisons of spine detection using different methods. Curved coronal MPRs are shown for better illustration. (a): A scoliotic case using Method2 (a1) and the proposed method (a2). (b): An artifact case using Method1 (b1) and the proposed method (b2).

3 Results

Our experimental data includes 405 LSpine, CSpine and WholeSpine scout scans with isotropic resolution 1.7mm. (105 for training and 300 for testing). These datasets come from different clinical sites and were generated by different types of Siemens MR Scanners (Avanto 1.5T, Verio 3T, Skyra 3T, etc.). Quantitative evaluation is carried on 355 discs and 340 vertebrae from 15 WholeSpine scans. The average translation errors of discs and vertebrae are $1.91mm$ and $3.07mm$. The average rotation error of discs is 2.33°.

A larger scale evaluation is performed on 300 scans (80 CSpine, 100 LSpine and 120 WholeSpine), including 43 (14.3%) with severe pathology and 36 (12.0%) with strong imaging artifacts. Three experienced radiologists rated spine detection results as "perfect" (no manual editing required), "acceptable" (minor manual editing required) and "rejected" (major manual editing required). For comparison, we also evaluate results from two adapted versions of the proposed method, **Method1**: without hierarchical learning and **Method2**: without local articulated model. As shown in Table 2, the proposed method generates "perfect" results in more than **97%** cases, which is significantly better than the

others (Two examples are shown in Fig. 3.). In general , Method2 is better than Method1, since the lack of articulated model mainly affects cases with abnormal spine geometry, e.g., scoliosis, which has a small proportion in our datasets. Another interesting observation is that Method1 has larger impacts on CSpine than LSpine, but Method2 is in the other way around. This phenomenon in fact results from the different sizes of cervical and lumbar vertebrae. Due to the smaller size of cervical vertebrae, it is prone to error detections using non-hierarchical detectors. On the other hand, the larger size of lumbar vertebrae makes the detection more sensitive to abnormal spine geometry, which can only be tackled with the local articulated model.

4 Conclusion

In this paper, we proposed a robust method to detect spine in 3D MR scout scans. Using hierarchical learning framework and local articulated model, our method exhibits accurate and robust performance on 300 testing datasets.

References

1. Chwialkowski, M., Shile, P., Peshock, R., Pfeifer, D., Parkey, R.: Automated detection and evaluation of lumbar discs in mr images. In: IEEE EMBS, pp. 571–572 (1989)
2. Alomari, R., Corso, J., Chaudhary, V.: Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. IEEE Trans. Med. Imaging 30, 1–10 (2011)
3. Schmidt, S., Kappes, J.H., Bergtholdt, M., Pekar, V., Dries, S.P.M., Bystrov, D., Schnörr, C.: Spine Detection and Labeling Using a Parts-Based Graphical Model. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
4. Kelm, M., Zhou, S., Sühling, M., Zheng, Y., Wels, M., Comaniciu, D.: Detection of 3d spinal geometry using iterated marginal space learning. In: MCV, pp. 96–105 (2010)
5. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in ct images. Medical Image Analysis 13, 471–482 (2009)
6. Zhan, Y., Dewan, M., Harder, M., Krishnan, A., Zhou, X.S.: Robust automatic knee mr slice positioning through redundant and hierarchical anatomy detection. IEEE Trans. Med. Imaging 30, 2087–2100 (2011)
7. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Towards robust and effective shape modeling: Sparse shape composition. Medical Image Analysis 16, 265–277 (2012)
8. Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., Ayache, N.: Geometric variability of the scoliotic spine using statistics on articulated shape models. IEEE Trans. Med. Imaging 27, 557–568 (2008)
9. Kadoury, S., Labelle, H., Paragios, N.: Automatic inference of articulated spine models in ct images using high-order markov random fields. Medical Image Analysis 15, 426–437 (2011)