Vision, Modeling, and Visualization (2017)
M. Hullin, R. Klein, T. Schultz, A. Yao (Eds.)

Star Convex Cuts with Encoding Swaps for Fast Whole-Spine Vertebra Segmentation in MRI

M. Rak[1] and K. D. Tönnies[1]

[1]University of Magdeburg, Germany

Abstract

We propose an automatic approach for fast vertebral body segmentation in three-dimensional magnetic resonance images of the whole spine. Previous works are limited to the lower thoracolumbar section and often take minutes to compute, which can be problematic in clinical routine or for data sets with numerous subjects. We address these limitations by a graph cut formulation. Our formulation involves appearance and shape information as well as star-convexity constraints to ensure a topologically correct segmentation for each vertebra. For close targets such as adjacent vertebrae, implementing star-convexity without fusing targets (naive binary formulations) or increasing run time/loosing optimality guarantees (multi-label formulations) is challenging. We provide a solution based on encoding swaps, which preserve optimality and ensure topological correctness between vertebrae. We validated our approach on two data sets. The first contains T_1- and T_2-weighted whole-spine images of 64 subjects. The second comprises 23 T_2-weighted thoracolumbar images and is publicly available. Our results are competitive to previous works (or better) at a fraction of the run time. We yielded Dice coefficients of 85.1 ± 4.4 *% and* 89.7 ± 2.3 *% with run times of* 1.65 ± 0.28 *s and* 2.73 ± 0.36 *s per vertebra on consumer hardware.*

Categories and Subject Descriptors (according to ACM CCS): I.4.6 [Image Processing and Computer Vision]: Segmentation—Pixel classification I.5.4 [Pattern Recognition]: Applications—Computer vision

1. Introduction

Magnetic resonance imaging has become a valuable non-invasive tool for the clinical assessment of the spine, including measurements of Cobb angles for rating of kyphosis / scoliosis or the identification of compression fractures such as crushed / wedged vertebrae. The rising clinical interest in magnetic resonance-based analysis has led to a number of works on automatic segmentation of vertebral bodies (simply called vertebrae hereafter) both model- and data-driven.

A first model-driven approach was presented in [ŠLPV11], who use a superquadrics-based parameteric shape model that adapts to a nearby vertebra based on intensity information. Alternatively, [ZVE*14] employ balloon forces to inflate a surface mesh with smoothness constraints directly inside the vertebra. To avoid ambiguities between vertebrae, [RET13] arranged multiple vertebrae into an elastic finite element model, which adapts to the data via image-derived forces.

Statistical modeling with active shape models was used in [NFE*12] to fit each vertebra individually. The concept was generalized to part-based models by [KLP13] and [SRF*14], who include shape and pose relations between multiple vertebrae. Both used non-linear mappings to improve the shape space representation. Recently, [KLPV16] showed encouraging results by linking active shape models with vertebra likelihood maps generated from convolutional neural networks.

Data-driven techniques are usually patch-based, performing segmentation in a small neighborhood around each vertebra. To this end, [SFNE14] match a cubically-shaped template to a nearby vertebra via graph cuts. In [HGT16] geodesic active contours and the Chan-Vese intensity model are combined into a level set segmentation. A learning-based approach is presented in [CBA*15], combining appearance learned via random forests with shape information estimated via Parzen windows into a vertebra probability map, which is then thresholded.

In purely two-dimensional mid-sagittal settings, data-driven techniques are typically applied to the whole image. For instance, [GMCD14] use decision trees to combine appearance, shape and pose information into a statistical inference task, which is solved by Gibbs sampling. An augmented Lagrangian method is present in [APM*12], who match distributions of vertebrae appearance features to a known reference. To ensure compactness for all vertebrae, normalized cuts with spatial smoothness were used in [ZLF*13].

To support a wide range of applications, segmentation techniques should apply to different sequences and to the whole spine. They should be reasonably fast, because time may be critical in clinical routine or for data sets with numerous subjects. These chal-

DOI: 10.2312/vmv.20171270

lenges are often overlooked in previous works, which are limited to the lower thoracolumbar section of the spine, easily take minutes to compute and - except for [ZVE*14] - only apply to a single sequence. To the best of our knowledge, we are the first to address all named challenges.

We contribute a novel binary graph cut formulation, which fuses patch-based star convex segmentation with whole-image constraints, ensuring topologically correctness for each and between vertebrae. Akin to previous works, we integrate appearance and shape information as well as boundary attraction priors to drive the segmentation on the patch level.

2. Method

2.1. Preprocessing

Vertebrae segmentation is typically applied only after a preliminary initialization based on either user-interaction [ŠLPV11, KLP13, RET13, SRF*14, HGT16, KLPV16] or automatic detectors [NFE*12, ZVE*14, CBA*15]. We interpret vertebrae localization as a pre-processing step, for which many valuable techniques exist, cf. the recent surveys of [AGKC15, RT16b]. In particular, we use the fast whole-spine detector of [RT16a], which applies to T_1-weighted and T_2-weighted images alike. They combine geometrical and appearance relations between adjacent vertebrae into a second-order graphical model, which is then used to infer the most likely combination of vertebra locations. Please see [RT16a] for further details on vertebra localization.

Based on the detection, we extract cubically-shaped vertebra-centered patches for the whole spine as outlined in Figure 1. Patch-based strategies significantly reduce the problem size compared to whole-image segmentation and ease modeling by focusing on individual vertebrae. However, ambiguities can arise for close targets when patch-wise results are re-embedded into the image domain, which is especially true if vertebrae are not well-separated by intervertebral discs. In what follows, we first detail our patch-wise formulation and show how to combine the patch-wise tasks into a joint ambiguity-free formulation afterwards.

2.2. Patch-wise Formulation

For each extracted vertebra patch, we interpret its segmentation as an energy minimization problem. In particular, we seek a binary labeling $l \in \{0,1\}^{|\mathcal{P}|}$ of the voxels $p \in \mathcal{P}$ of the patch into foreground ($l_p = 1$), i.e. voxels inside the central vertebra, and background ($l_p = 0$), i.e. voxels outside of it, that minimizes

$$E(l) = \underbrace{\sum_{p\in\mathcal{P}} A_p(l_p)}_{\text{Appearance}} + \underbrace{\sum_{p\in\mathcal{P}} D_p(l_p)}_{\text{Shape}} + \underbrace{\sum_{(p,q)\in\mathcal{B}} B_{pq}(l_p,l_q)}_{\text{Boundaries}} + \underbrace{\sum_{(p,q)\in\mathcal{C}} C_{pq}(l_p,l_q)}_{\text{Star-Convexity}}. \quad (1)$$

Our model involves soft priors for appearance, shape and boundary attraction as well as hard constraints that ensure a star-convex

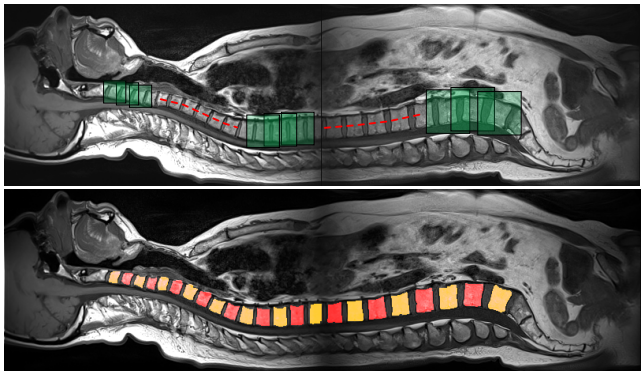

Figure 1: *Our patch-based framework for whole spine vertebra segmentation. Image patches (green) are extracted for each vertebra (top). Then, patch-based vertebra segmentation is applied and the results are re-embedded into the image domain (bottom). The vertebra segmentation is color-coded to ease the differentiation between neighboring vertebrae. Please note that there is considerable overlap between the neighboring image patches.*

segmentation. Edge sets $\mathcal{B}$ and $\mathcal{C}$ comprise the ordered voxel pairs (p,q) that are linked by adjacency on the patch and by star-convexity constraints, respectively.

We will design each term such that the resulting energy is graph-representable, in which case the minimization of Equation 1 takes $\mathcal{O}(\#\text{voxels} \cdot \#\text{edges}^2)$, cf. [BK04]. To be graph-representable, all pairwise terms $T(l_i,l_j)$, i.e. boundary attraction and star-convexity, have to obey $T(0,0)+T(1,1) \leq T(0,1)+T(1,0)$ [KZ04]. This essentially means that assignments of different labels should not be cheaper than the assignment of similar ones. We now discuss each term in greater detail.

2.2.1. Appearance

Recent works encode appearance information by advanced techniques like decision trees [GMCD14], random forests [CBA*15] or convolutional neural networks [KLPV16]. These require training data, which may become outdated as soon as the imaging sequence or acquisition parameters change. We use a rather simple but more case-specific prior, which is estimated directly from the intensities of the image patch. Our appearance prior reads

$$A_p(l_p) = [l_p = 1] \cdot \underbrace{\left(\frac{|I_p - \mu|}{\alpha\sigma} - 1\right)}_{i_p - 1}, \quad (2)$$

whereby Iverson brackets $[\cdot]$ select only the foreground, which implies zero costs for the background. The foreground costs range between minus one and positive infinity. They depend on the difference of the voxel intensity I_p to a known reference vertebra intensity μ. To become invariant to intensity shifts and scaling, the difference is normalized by a known dispersion σ of the vertebra intensities.

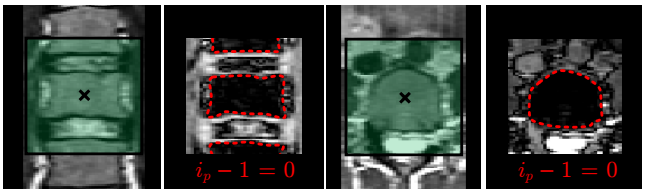

Figure 2: *Our appearance prior on coronal (left) and axial (right) slices of a particular vertebra patch (green). Ideally, the appearance costs $i_p - 1$ (second and fourth image) are negative in the foreground, i.e. inside the central vertebra, and positive for most of the background. If our appearance prior would be considered on its own, then the energy minimization would essentially be a voxel-wise thresholding at the pivot point $i_p - 1 = 0$. Please note that we oversimplified the thresholding (red) for illustration purposes; the actual result will be leaky and contain spurious components.*

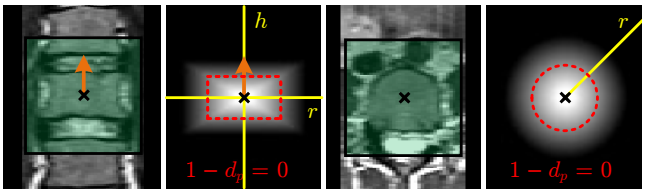

Figure 3: *Our shape prior on coronal (left) and axial (right) slices of a particular vertebra patch (green). To compute the shape costs (second and fourth image), we utilize a cylindrical coordinate system that is aligned with the vertebra's up-vector (orange). The coordinate axes (yellow) are defined along the cylinder (height h) and away from it (radius r). Ideally, the shape costs $1 - d_p$ are positive in the foreground, i.e. inside the central vertebra, and negative for the background. If our shape prior would be considered own, then the energy minimization would essentially be a voxel-wise thresholding at the pivot point $1 - d_p = 0$.*

Our appearance prior is illustrated in Figure 2. It favors foreground when Equation 2 becomes negative and background otherwise. The pivot point is at an intensity difference of $\alpha\sigma$, which means that parameter α should be chosen such that it differentiates between the intensity ranges of the foreground and background. We estimate reference intensity μ and dispersion σ from a small central part of the patch via the median intensity and the median absolute deviation, respectively. Both estimators are robust to varying types of noise and corruption with up to fifty percent of non-vertebra tissue, which makes them reliable even under misalignment due to the preliminary vertebra localization.

2.2.2. Shape

Since appearance alone cannot differentiate between the central vertebra of a patch and the neighboring ones, information about shape is necessary. Recent work uses advanced techniques to implement shape information, e.g. a probabilistic atlas [GMCD14, CBA*15] or deformable statistical models [KLPV16]. We found that such a detailed shape description is not advantageous per se. It biases the segmentation considerably, which becomes problematic when the shape prior is not well aligned with the central vertebra or does not well reflect its particular shape.

Taking this into account, we use a rather coarse description of shape. We encode shape information by a cylindrical prior that is aligned with the vertebra's up-vector. We let the prior decrease gradually with distance to the center of the patch instead of introducing a steep cut-off at the expected location of the vertebra boundary like in previous work. Hence, our method is more robust to misalignment due to the preliminary vertebra localization. Our shape prior reads

$$D_p(l_p) \;=\; [l_p = 0] \cdot \underbrace{\left(1 - \max\left(s_h \frac{2\,|h_p|}{d}, s_r \frac{2\,|r_p|}{d}\right)\right)}_{1 - d_p}, \qquad (3)$$

whereby Iverson brackets $[\cdot]$ select only the background, which implies zero costs for the foreground. The background costs range between one and negative infinity (for an unbound patch). They depend on the distance d_p, in cylindrical coordinates (h_p, r_p), of the voxel p to the center of the patch. To become invariant to size changes along the spine, the distance is normalized by the inter-vertebra distance d.

Our shape prior is illustrated in Figure 3. It favors foreground whenever Equation 3 becomes positive and background otherwise. The pivot point depends on $\frac{2s_h}{d}$ and $\frac{2s_r}{d}$, which means that scaling factors s_h and s_r should be chosen such that they reflect the relative vertebra height and depth/width, respectively. The particular choice of s_h and s_r is uncritical, because our shape costs decrease gradually with distance to the center without a steep cut-off. We will derive a reasonable choice for s_h and s_r before the experiments. The inter-vertebra distance d and the vertebra's up-vector are estimated from the location of the neighboring vertebrae.

2.2.3. Boundaries

To attract the segmentation towards tissue boundaries, we introduce a binary term for each pair of neighoring voxels (under Von Neumann neighborhood), if they do not share the same label. Such terms cause additional costs along the boundary of the segmentation, which may lead to a "shrinking bias" if the relative importance is too high. To deal with this issue, we use the following prior

$$B_{pq}(l_p, l_q) \;=\; [l_p \neq l_q] \cdot \beta \max\left(0, 1 - \frac{|I_p - I_q|}{\alpha\sigma}\right), \qquad (4)$$

whereby Iverson brackets $[\cdot]$ select only cases where adjacent voxels do not share the same label. Costs range between zero and one. They depend on the difference between the intensities I_p and I_q of both voxels. To become invariant to intensity shifts and scaling, the difference is normalized by the known dispersion σ. Parameter $\beta \in [0, \infty)$ the controls importance of our boundary term.

Our boundary prior applies zero costs if the intensity difference is large enough. Hence, shrinking cannot occur in these cases. The rationale for truncation is that above a certain threshold all intensity

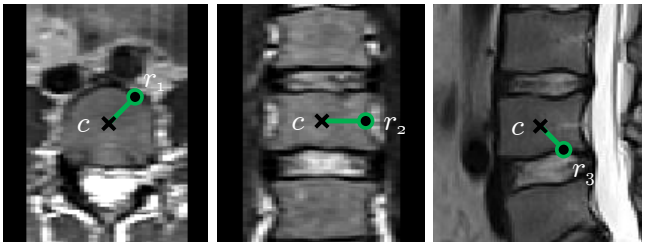

Figure 4: *Star-convexity constraints on axial (left), coronal (middle) and sagittal (right) slices. If voxels r_i (dots) would be assigned to the foreground, then any other voxel on the line segment (green) to the center voxel c (cross) would be foreground too. The result is a single connected component without holes.*

changes can be safely classified as tissue boundaries. The truncation occurs at an intensity difference of $\alpha\sigma$, which is a reasonable choice as α will be chosen such that it differentiates between the intensity ranges of the foreground and background.

2.2.4. Star-Convexity

We employ the star-convexity constraints of [Vek08] to ensure a topologically correct segmentation of the vertebra. In particular, we will restrict the segmentation to a single connected component. To this end, we enforce that for any foreground voxel r, every other voxel p on the line segment between r and the center of the patch c is foreground too. The idea is illustrated in Figure 4.

Star-convexity may be implemented by rasterization, transforming the line segment between the center c and every voxel r into a sequence of voxels $(c,\dots,p,q,\dots r)$ and "tying" together each pair of neighboring voxels of the sequence by hard constraints. Given two such voxels p and q the star-convexity constraints read

$$C_{pq}(l_p,l_q) \;=\; [l_p = 0 \wedge l_q = 1]\cdot\infty, \tag{5}$$

where Iverson brackets $[\cdot]$ assign infinite costs when foreground shall be assigned after some background voxel. The rasterization need not be computed "online" during segmentation, because it is independent of the appearance of the patch and the voxel size.

We precompute the rasterization and all pairs of voxels on a sufficiently large reference patch via Bresenham's line algorithm [Bre65]. The pairs are loaded before the segmentation and cropped to each particular image patch during segmentation, leading to an efficient implementation of star-convexity.

2.3. Joint Formulation

For close targets such as neighboring vertebrae, ensuring star-convexity is challenging, as outlined in Figure 5. Binary whole image formulations will fuse targets and multi-label settings are less performant / do not share optimality guarantees of binary tasks. Our binary patch-based problem formulation has neither problem, but ambiguities may arise when patch-wise results are re-embedded into the image domain. This is especially true when vertebrae are

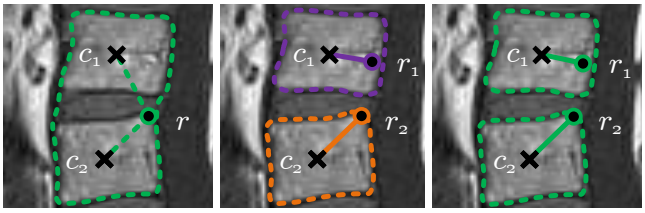

Figure 5: *Star-convexity implementation for neighboring vertebrae. A binary whole image formulation (left) will fuse both vertebrae (green region). The rationale is that any voxel r will be "tied" to both centers c_1 and c_2 simultaneously. Multi-label settings (middle) avoid this issue by imposing star-convexity for each label individually, but optimality guarantees are lost. Our binary patch-wise formulation (right) has neither problem, because the vertebra patches are treated individually.*

not well-separated by large intervertebral discs, because the patch energy could then favor to bridge the thin gap to the neighboring vertebra.

To circumvent the ambiguities and preserve optimality, we combine the patch-wise formulations into a joint binary minimization problem with additional topological constraints that guarantee a non-overlapping vertebra segmentation. In particular, we seek a combination of patch-wise labelings that minimizes

$$E\left(l^1,\dots,l^{\#\text{vertebrae}}\right) \;=\; \underbrace{\sum_{i=1}^{\#\text{vertebrae}} E^i\left(l^i\right)}_{\text{Patch Energy}} + \underbrace{\sum_{i=2}^{\#\text{vertebrae}} O^{i-1,i}\left(l^{i-1},l^i\right)}_{\text{Non-Overlap}}, \tag{6}$$

where the introduced superscripts enumerate all vertebra patches from head to foot (or vice versa). The first sum pools the already introduced energies of the individual vertebra patches and the second sum handles the regions where the neighboring patches overlap. We now discuss the implementation of the latter in greater detail.

2.3.1. Direct Solution

For any such overlap region, only one of the two involved vertebra patches shall be allowed to assign foreground to a shared voxel, because otherwise ambiguities can occur. Hence, we seek to establish binary terms between the shared voxels of both vertebra patches to prohibit such situations. The non-overlap requirement of the patch-wise labelings is equivalent to the hard constraints

$$O^{ij}\left(l^i,l^j\right) \;=\; \sum_{(p,q)\in\mathcal{O}^{ij}} \underbrace{\left[l_p^i = 1 \wedge l_q^j = 1\right]\cdot\infty}_{T^{ij}\left(l_p^i,l_q^j\right)}, \tag{7}$$

where Iverson brackets $[\cdot]$ assign infinite costs when both patches

assign foreground to a shared voxel. Set $\mathcal{O}^{ij}$ comprises the voxels shared by both patches. In particular, it contains the ordered pairs (p,q) of voxels p from patch i and voxels q from patch j that represent the same voxel after re-embedding into the image domain.

The constraints of Equation 7 cannot be realized with graph cuts directly, because it is not graph-representable in its current form. In particular, the $T^{ij}(\cdot)$ violate $T(0,0)+T(1,1) \leq T(0,1)+T(1,0)$ [KZ04], which essentially means that the assignment of different labels should not be cheaper than the assignment of similar ones. We can, however, derive an equivalent formulation that is indeed graph-representable.

2.3.2. Encoding Swaps

We go back to our representation of foreground and background, which are implemented by $l_p^i = 1$ and $l_p^i = 0$, respectively. Noticing that this is a convention, we could just as well have swapped the meaning of labels. In particular, we could have encoded foreground with $l_p^{\tilde{i}} = 0$ and background with $l_p^{\tilde{i}} = 1$, where the tilde differentiates between the swapped and the standard encoding.

Obviously, swapping the encoding of every vertebra patch does not solve the problem, because the resulting non-overlap constraints are not graph-representable either. However, when we swap the encoding only for every other vertebra patch from head to food, then the hard constraints change to

$$T^{i\tilde{j}}\left(l_p^i, l_q^{\tilde{j}}\right) \;=\; \left[l_p^i = 1 \wedge l_q^{\tilde{j}} = 0\right] \cdot \infty, \tag{8}$$

where standard encoded patch i overlaps with encoding-swapped patch $\tilde{j}$. It is easy to verify that both the $T^{i\tilde{j}}(\cdot)$ and their counterparts $T^{\tilde{i}j}(\cdot)$ obey $T(0,0)+T(1,1) \leq T(0,1)+T(1,0)$. The patch energies remain unaffected by the encoding swap if appearance and shape priors as well as star-convexity constraints are adjusted accordingly. Please note that the argument does not contradict the graph-representability of [KZ04]; it rather exploits the available degrees of freedom. Eventually, we compute the optimal labeling by the algorithm of [BK04]. In particular, we use the implementation provided by the Darwin framework [Gou12].

The concept of encoding swaps is not limited to our particular application. Let each image patch be a node in a graph and let each overlap region between two patches be an edge between their associated nodes. In this notation, every bipartite graph, i.e. every graph that has a two-coloring, can benefit from an efficient implementation of non-overlap constraints via encoding swaps. Specifically, chain-like overlap layouts like ours are covered, but also all forms of tree-like layouts and certain grid-like overlap layouts too.

3. Experiments

3.1. Data Sets and Preprocessing

We carried out experiments on two data sets. The first data set (DS1) comprises T_1- and T_2-weighted whole-spine images of 64 subjects from the "Study of Health in Pomerania" [VAS*11]. For DS1 ground truth segmentations are available from C3 to L5. The second data set (DS2) comprises 23 T_2-weighted thoracolumbar images and is publicly available [CBA*15]. For DS2 ground truth segmentations are available from T11 to L5. Both data sets were acquired by turbo spin echo sequences on Siemens 1.5 Tesla imagers and reconstructed sagittaly at $1.12 \times 1.12 \times 4.4$ mm and $1.25 \times 1.25 \times 2.0$ mm, respectively. To simplify processing, we upsampled all images in mediolateral direction to yield isotropic voxels.

The vertebrae localization of [RT16a] correctly detected 96.0 % of the vertebrae in DS1 at an accuracy of 3.45 ± 2.20 mm to the ground truth centers. For DS2 the detection rate was 98.1 % with 3.07 ± 1.78 mm to the ground truth centers. The quality difference is mainly due to the finer laterolateral resolution (2.0 vs. 4.4 mm). For both data set, the localizations took around one second per vertebra (Intel Core i5 @ 4×3.30 GHz). All falsely detected vertebrae were corrected manually before the actual segmentation, i.e. a user-specified vertebra center was used instead of the found location.

During experiments we used a fixed set of parameters. Since our segmentation method requires only a rough estimate of the expected horizontal and radial scaling of the shape prior, we simply took the first subject of both data sets as reference, yielding $s_h = \frac{2}{3}$ and $s_r = 1$. The radial scaling needs to be larger to reflect that the vertebra height is smaller than its width / depth. The parameters for appearance and boundary attraction were found empirically. We set $\alpha = 4\frac{1}{2}$ and $\beta = 1$, but the parameters may be varied $\pm\frac{1}{2}$ with little effect on most cases. This is due to the fact that we designed each energy by linear priors without a steep cut-off rather than non-linear priors that are used more frequently. For comparison to previous works, we assessed the Dice coefficient (DC), the average inter-surface distance (AD) and the Hausdorff distance (HD).

3.2. Results and Discussion

A quantitative comparison to previous works is given in Table 1. From the results on DS2 we see that our approach surpasses most previous works on thoracolumbar images. Only the active shape models of [NFE*12] and [KLPV16] showed slightly and significantly better results, respectively. Since the inter-reader variation for thoracolumbar vertebrae is reported between 90.3 % [HGT16] and 91 % [ZVE*14], considerably better results should be interpreted with caution, because reader adaptation is not unlikely. This is especially true when machine learning is involved like in [KLPV16]. Moreover, according to [NFE*12] the active shape model fitting takes several minutes per vertebra (Intel Core 2 Duo @ 2×2.0 GHz), casting doubts about its applicability in practice. Performance was not reported in [KLPV16], but we expect even longer run times compared to [NFE*12] due to their particular convolutional neural network. We verified this by implementing their network in TensorFlow with cuDNN. Under idealized conditions, the prediction of a medium-sized $51 \times 51 \times 51$ vertebra patch already took 3.75 ± 0.02 minutes (NVIDIA GeForce GTX 970).

Our approach took only 19.1 ± 2.5 seconds per image on DS1 (Intel Core i5 @ 4×3.30 GHz). This is a considerable speed-up compared to all previous works, where the segmentation of a section of the whole spine may take minutes. For comparison, [CBA*15] reported 1.3 minutes (unknown multi-core system @ 3.0 GHz) for thoracolumbar images with seven vertebrae and [SRF*14] segments lumbar images in "less than two minutes" (Intel Core i5 @ 4×2.5 GHz). The difference becomes most

Table 1: *Quantitative comparison to previous works on magnetic resonance-based vertebra segmentation. Work is categorized into 2d (mid-sagittal) and 3d (volumetric) analysis. For each category, work is sorted chronologically. Please note that we included the 2d analysis techniques for completeness, their results are not comparable to the 3d setting, which is considerably more challenging. Underlining indicates a common data set. Abbreviations: DC - Dice coefficient; AD - average inter-surface distance; HD - Hausdorff distance; C - cervical; L - lumbar; TL - thoracolumbar; W - whole spine; T_1w - T_1-weighted; T_2w - T_2-weighted; ^{P} provided by author; ^{R} recalculated from results.*

2/3d	Works	Section	Sequence	#Images	#Vertebrae	DC [%]	AD [mm]	HD [mm]
2d	Huang [HCLN09]	C, L, W	T_2w	?	52	96±?	-	-
	Ayed [APM*12]	L	T_2w	15	75	85 ± 5.1	-	-
	Zheng [ZLF*13]	L	T_1w, T_2w	5	?	96.6 ± 0.3	-	1.7 ± 0.2
	Ghosh [GMCD14]	L	T_2w	13	?	84.4 ± 3.8	-	-
3d	Stern [ŠLPV11]	TL	T_2w	9	75	-	1.85 ± 0.47	-
	Neubert [NFE*12]	TL	T_2w	14	132	90.8 ± 1.8^R	0.67 ± 0.17^R	4.08 ± 0.94^R
	Kadoury [KLP13]	TL	T_1w	8	136	-	2.93 ± 1.83^R	-
	Schwarzenberg [SFNE14]	L	T_2	2	10	81.3 ± 5.1	-	-
	Suzani [SRF*14]	L	T_1w	9	45	-	3.02 ± 0.82^R	9.20 ± 2.43^R
	Zukic [ZVE*14]	TL	T_1w, T_2w	17	153	79.3 ± 5.0^P	1.76 ± 0.38	11.89 ± 2.65^P
	Chu [CBA*15]	TL	T_2w	23	161	88.7 ± 2.9	1.5 ± 0.2	6.4 ± 1.2
	Hille [HGT16]	TL	T_1w	6	34	84.8±?	1.29 ± 0.42	6.55±?
	Korez [KLPV16]	TL	T_2w	23	161	93.4 ± 1.7	0.54 ± 0.14	3.83 ± 1.04
3d	Ours DS1	W	T_1w, T_2w	128	1412	85.1 ± 4.4	1.41 ± 0.36	5.36 ± 1.48
	Ours DS2	TL	T_2w	23	161	89.7 ± 2.3	1.31 ± 0.27	5.30 ± 1.12

clear when considering the run time per vertebrae. For instance in [HGT16] the computation "never exceeded 60 seconds" and [ŠLPV11, NFE*12] report several minutes per vertebra (Intel Core 2 Duo @ 2×2.0 GHz and 2.83 GHz, respectively). In general, all previous works state run times well above ten seconds per vertebra, while our's can be computed in 2.73 ± 0.36 seconds.

The results in Table 1 indicate that DS1 is more challenging than DS2. This is due the lower laterolateral resolution (4.4 vs. 2.0 mm) and the presence of cervical and upper thorarcic vertebrae, which are more affected by partial volume effects. Exemplary results for both data sets are depicted in Figure 6. Despite these challenges, our results are better than most previous works (see Table 1), whereby only the mediocre results of [SFNE14, SRF*14, ZVE*14] are based on images with a similarly coarse laterolateral resolution. Considering T_1- and T_2-weighted subsets, we observe that the latter gave better results (84.5 ± 4.9 vs. 85.6 ± 4.0 % DC). Presumably, this is due to the increased constrast to intervertebral discs. As could be expected, the result quality increased from cervical (83.9 ± 5.0 % DC) over thorarcic (84.9 ± 3.7 % DC) to lumbar vertebrae (87.2 ± 3.3 % DC). Since the average patch is smaller in DS1, the run time per vertebra decreased to 1.65 ± 0.28 seconds (Intel Core i5 @ 4×3.30 GHz) on average.

4. Conclusion

We proposed an automatic approach for fast vertebral body segmentation in three-dimensional magnetic resonance images. Our work is novel, because it applies to T_1- and T_2-weighted images and to the whole spine, which we demonstrated on image data of 64 subjects (128 images in total). Compared to previous work, we are the first to show a working solution on a reasonably large data set. Moreover, our results on a publicly available benchmark data set with 23 T_2-weighted thoracolumbar images are competitive to previous works (or better) at a fraction of the run time.

Our approach has two limitations. First, the segmentation quality may decrease when the preliminary vertebra localization produces too large misalignments, which was not the case for our data but might be elsewhere. To deal with such cases, we suggest to improve the estimate of each vertebra center and each up-vector by a quick symmetry-based self-registration of each vertebra, the details of which are given in Appendix A.

Second, in case severe pathologies alter the appearance or shape of the vertebrae, e.g., strong metastases or burst fractures, the segmentation may be poor. Any available appearance or shape modeling technique will struggle to capture the large diversity accompanied with such cases. To be precise, we are not aware of any work addressing this challenge. Machine learning techniques may be an option, but only after a large pathological training data set becomes available, which is - as of today - not the case. In the future, we want to address this challenge.

Acknowledgments

We thank all parties and participants of the Study of Health in Pomerania. This research was partially funded by the the German Research Foundation (TO 166/13-2).

Appendix A: Symmetry-based Vertebra Self-Registration

When the vertebra locations or up-vectors are too inaccurate, we suggest to improve it by establishing a vertebra-centric coordinate system via a symmetry-based self-registration. To this end, we seek to find a transformation between the hypothesized vertebra-centric coordinates and the image domain such that the vertebral symmetry

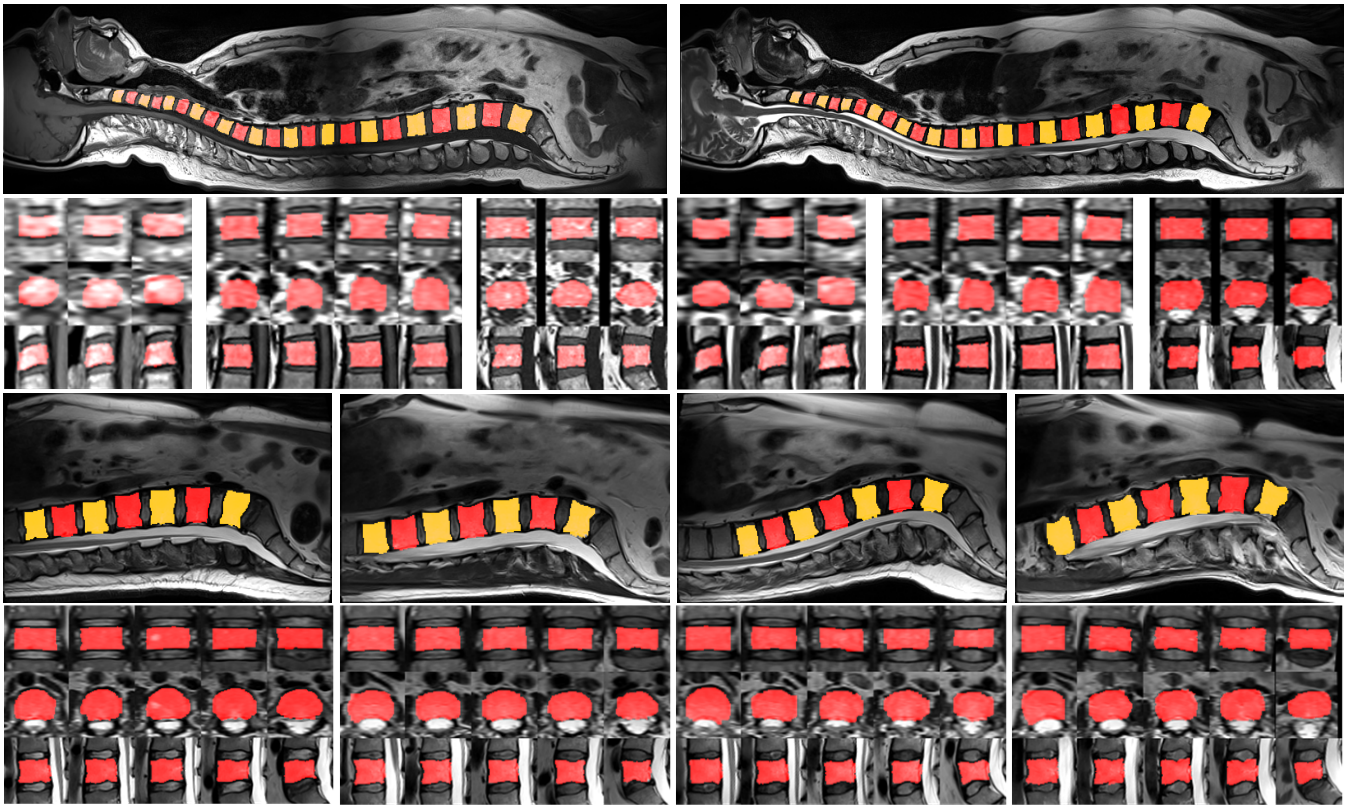

Figure 6: *First two rows: segmentation results on T_1-weighted (left) and T_2-weighted (right) images of a subject from DS1. Last two rows: segmentation results on T_2-weighted images of different subjects from DS2. First and third row: mid-sagittal slices after re-embedding of patch-wise segmentation results into the image domain. Vertebra coloring reflects encoding swaps. Second and fourth row: coronal, transverse and sagittal views of the individual cervical, thoracic and lumbar vertebrae.*

planes are aligned with the axes of the coordinate system. The idea is illustrated in Figure 7.

The transformation involves six degrees of freedom: three for translation and three for rotation. These are determined completely if all three symmetry relations (top-bottom, left-right and front-back symmetry) are utilized. Hence, we establish the vertebra-centric coordinate system (x, y, z) by minimizing

$$E(T) = \underbrace{C\left(I_{T(x,y,z)}, I_{T(-x,y,z)}\right)}_{\text{Top-Bottom}} + \underbrace{C\left(I_{T(x,y,z)}, I_{T(x,-y,z)}\right)}_{\text{Left-Right}} + \underbrace{C\left(I_{T(x,y,z)}, I_{T(x,y,-z)}\right)}_{\text{Front-Back}}, \tag{9}$$

where $T(\cdot)$ is the sought transformation into the image domain and $I_{T(\cdot)}$ is the image expressed in vertebra-centric coordinates. $C(\cdot)$ is an appearance-based cost function, which compares $I_{T(x,y,z)}$ with the x-, y- and z-mirrored copies of itself to quantify the symmetry.

To implement the cost function, virtually any intensity-based measure of alignment quality may be used. More important than that is the distance-based weighting of the quality measure, because the symmetry assumptions are less reliable the farther away we are from the vertebra boundary. We found that a Gaussian weighting with a standard deviation of roughly one half of the inter-vertebra distance is an adequate choice. Depending on the quality measure, a suitable non-linear optimization method should be chosen.

Some self-registration results based on mean squared intensity differences and a quasi-Newton method are shown in Figure 8. Even for the lumbar spine section, the self-registration took below one second per vertebra (Intel Core i5 2500, 4×3.3 GHz) for all tested images. Based on visual inspection, the rate of failure is below two percent for both data set and imaging sequences.

References

[AGKC15] ALOMARI R. S., GHOSH S., KOH J., CHAUDHARY V.: Vertebral column localization, labeling, and segmentation. In *Spinal Imaging and Image Analysis*. 2015, pp. 193–229. 2

[APM*12] AYED I. B., PUNITHAKUMAR K., MINHAS R., JOSHI R., GARVIN G. J.: Vertebral body segmentation in MRI via convex relaxation and distribution matching. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (2012), pp. 520–527. 1, 6

[BK04] BOYKOV Y., KOLMOGOROV V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE*

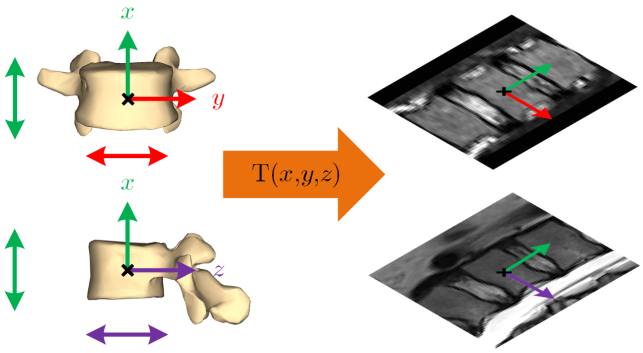

Figure 7: *A vertebra-centric coordinate system (left) may be obtained by finding a transformation $T(\cdot)$ into the image domain (right) such that the top-bottom (green double-headed arrow), the left-right (red double-headed arrow) and the front-back (purple double-headed arrow) symmetry is maximized. The vertebra renderings are provided by BodyParts3D [MFT*09] and licensed under Creative Commons Attribution - Share Alike 2.1 Japan.*

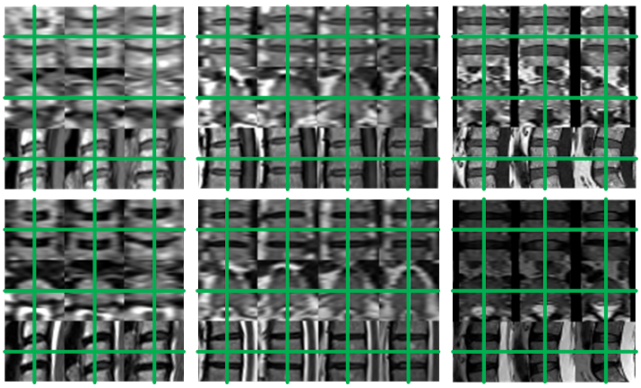

Figure 8: *Self-registration results for different cervical (left), thoracic (middle) and lumbar (right) vertebrae in T_1-weighted (top) and T_2-weighted (bottom) images of the same subject. Each result comprises the obtained mid-coronal (top), mid-axial (middle) and mid-sagittal (bottom) plane of the vertebra. The assumed symmetry axes are shown in green.*

Transactions on Pattern Analysis and Machine Intelligence 26 (2004), 1124–1137. 2, 5

[Bre65] BRESENHAM J. E.: Algorithm for computer control of a digital plotter. *IBM Systems Journal 4* (1965), 25–30. 4

[CBA*15] CHU C., BELAVÝ D. L., ARMBRECHT G., BANSMANN M., FELSENBERG D., ZHENG G.: Fully automatic localization and segmentation of 3d vertebral bodies from CT/MR images via a learning-based method. *PloS one 10* (2015), e0143327. 1, 2, 3, 5, 6

[GMCD14] GHOSH S., MALGIREDDY M. R., CHAUDHARY V., DHILLON G.: A supervised approach towards segmentation of clinical MRI for automatic lumbar diagnosis. In *Proceedings of the Workshop on Computational Methods and Clinical Applications for Spine Imaging* (2014), pp. 185–195. 1, 2, 3, 6

[Gou12] GOULD S.: DARWIN: A framework for machine learning and computer vision research and development. *Journal of Machine Learning Research 13* (2012), 3533–3537. 5

[HCLN09] HUANG S. H., CHU Y. H., LAI S. H., NOVAK C. L.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. *IEEE Transactions on Medical Imaging 28* (2009), 1595–1605. 6

[HGT16] HILLE G., GLASSER S., TÖNNIES K.: Hybrid level-sets for vertebral body segmentation in clinical spine MRI. *Procedia Computer Science 90* (2016), 22–27. 1, 2, 5, 6

[KLP13] KADOURY S., LABELLE H., PARAGIOS N.: Spine segmentation in medical images using manifold embeddings and higher-order MRFs. *IEEE Transactions on Medical Imaging 32* (2013), 1227–1238. 1, 2, 6

[KLPV16] KOREZ R., LIKAR B., PERNUŠ F., VRTOVEC T.: Model-based segmentation of vertebral bodies from MR images with 3d CNNs. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), pp. 433–441. 1, 2, 3, 5, 6

[KZ04] KOLMOGOROV V., ZABIN R.: What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence 26* (2004), 147–159. 2, 5

[MFT*09] MITSUHASHI N., FUJIEDA K., TAMURA T., KAWAMOTO S., TAKAGI T., OKUBO K.: Bodyparts3d: 3d structure database for anatomical concepts. *Nucleic Acids Research 37* (2009), D782–D785. 8

[NFE*12] NEUBERT A., FRIPP J., ENGSTROM C., SCHWARZ R., LAUER L., SALVADO O., CROZIER S.: Automated detection, 3d segmentation and analysis of high resolution spine MR images using statistical shape models. *Physics in Medicine and Biology 57* (2012), 8357–8376. 1, 2, 5, 6

[RET13] RAK M., ENGEL K., TÖNNIES K. D.: Closed-form hierarchical finite element models for part-based object detection. In *Proceedings of the International Workshop on Vision, Modeling and Visualization* (2013), pp. 137–144. 1, 2

[RT16a] RAK M., TÖNNIES K. D.: A learning-free approach to whole spine vertebra localization in MRI. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (2016), pp. 283–290. 2, 5

[RT16b] RAK M., TÖNNIES K. D.: On computerized methods for spine analysis in MRI: a systematic review. *International Journal of Computer Assisted Radiology and Surgery* (2016), 1–21. 2

[SFNE14] SCHWARZENBERG R., FREISLEBEN B., NIMSKY C., EGGER J.: Cube-cut: vertebral body segmentation in MRI-data through cubic-shaped divergences. *PloS one 9* (2014), e93389. 1, 6

[ŠLPV11] ŠTERN D., LIKAR B., PERNUŠ F., VRTOVEC T.: Parametric modelling and segmentation of vertebral bodies in 3d CT and MR spine images. *Physics in Medicine and Biology 56* (2011), 7505–7522. 1, 2, 6

[SRF*14] SUZANI A., RASOULIAN A., FELS S., ROHLING R. N., ABOLMAESUMI P.: Semi-automatic segmentation of vertebral bodies in volumetric MR images using a statistical shape+pose model. In *Proceedings of SPIE Medical Imaging* (2014), p. 90360P. 1, 2, 5, 6

[VAS*11] VÖLZKE H., ALTE D., SCHMIDT C. O., RADKE D., LORBEER R., FRIEDRICH N., AUMANN N., LAU K., PIONTEK M., BORN G., ET AL.: Cohort profile: The Study of Health in Pomerania. *International Journal of Epidemiology 40* (2011), 294–307. 5

[Vek08] VEKSLER O.: Star shape prior for graph-cut image segmentation. In *Proceedings of the European Conference on Computer Vision* (2008), pp. 454–467. 4

[ZLF*13] ZHENG Q., LU Z., FENG Q., MA J., YANG W., CHEN C., CHEN W.: Adaptive segmentation of vertebral bodies from sagittal MR images based on local spatial information and Gaussian weighted chi-square distance. *Journal of Digital Imaging 26* (2013), 578–593. 1, 6

[ZVE*14] ZUKIĆ D., VLASÁK A., EGGER J., HOŘÍNEK D., NIMSKY C., KOLB A.: Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images. *Computer Graphics Forum 33* (2014), 190–204. 1, 2, 5, 6