Computer Methods and Programs in Biomedicine 155 (2018) 93–99

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Vertebral body segmentation in wide range clinical routine spine MRI data

Georg Hille[a,*], Sylvia Saalfeld[a], Steffen Serowy[b], Klaus Tönnies[a]

[a] Department of Simulation and Graphics, University of Magdeburg, Universitätsplatz 2, Magdeburg 39106, Germany
[b] Department of Neuroradiology, University Hospital of Magdeburg, Leipziger Straße 44, Magdeburg 39120, Germany

ARTICLE INFO

Article history:
Received 29 June 2017
Revised 27 November 2017
Accepted 11 December 2017

Keywords:
Clinical spine MRI
Vertebral body
Segmentation
Hybrid level-sets
Various MRI sequences

ABSTRACT

Background and objective: In this work we propose a 3D vertebral body segmentation approach for clinical magnetic resonance (MR) spine imaging. So far, vertebrae segmentation approaches in MR spine imaging are either limited to particular MR imaging sequences or require minutes to compute, which can be hindering in clinical routine. The major contribution of our work is a reasonably precise segmentation result, within seconds and with minimal user interaction, for spine MR imaging commonly used in clinical routine. Our focus lies on the applicability towards a large variety of clinical MR imaging sequences, dealing with low image quality, high anisotropy and spine pathologies. **Methods:** Our method starts with a intensity correction step to deal with bias field artifacts and a minimal user-assisted initialization. Next, appearance-based vertebral body probability maps guide a subsequent hybrid level-set segmentation. **Results:** We tested our method on different MR imaging sequences from 48 subjects. Overall, our evaluation set contains 63 datasets including 419 vertebral bodies, which differ in age, sex and presence of spine pathologies. This is the largest set of reference segmentations of clinical routine spine MR imaging so far. We achieved a Dice coefficient of 86.0%, a mean Euclidean surface distance error of 1.59 ± 0.24 mm and a Hausdorff distance of 6.86 mm. **Conclusions:** These results illustrate the robustness of our segmentation approach towards the variety of MR image data, which is a pivotal aspect for clinical usefulness and reliable diagnosis.

1. Introduction

Various evaluations in orthopedic and neuroradiological diagnosis, ranging from scoliosis, stenosis, osteoporosis or vertebral fractures to bone metastases, rely on non-invasive medical imaging like computed tomography (CT) and magnetic resonance imaging (MRI) [1]. Most of the related spine segmentation approaches focussing on CT or 2D X-rays [2,3] benefit from the high contrast of bone tissue as well as the mostly isotropic high spatial resolutions. Addressing the cancer risk from radiation exposure in CT and X-ray, diagnostic MRI became an indispensable technique in clinical decision-making. Besides, CT and X-ray imaging cannot adequately deal with some pathologies like bone tumors and metastases. Therefore, MRI is often essential for diagnosis. However, some characteristics of routine spine MRI tremendously hamper the automation of segmentation approaches. Firstly, anisotropic spatial resolution often results in partial volume effects, resulting in blurred delineation between different tissue types especially to the lateral ends. Furthermore, bias field artifacts cause non-homogenous intensities between central and marginal areas. Lastly, the image quality and emphasis of different tissues is affected by various imaging parameters since standarized measurement units like Hounsfield units (HU) in CT do not exist in MRI.

A robust segmentation of vertebral bodies is a major step towards a precise and reliable diagnosis. Intervention and radiotherapy planning and navigation could be enhanced by combining pre-interventionally segmented vertebral bodies with intra-interventionally acquired image data. However, different diagnostic MRI sequences are used in clinical practices. A segmentation method that is relevant in a clinical setting will have to deal with this large variety of MRI sequences and parameter settings and should be reasonably fast. Previous works often shifted away from the challenges of clinical settings by almost solely applying to only a particular MRI sequence or requiring minutes to compute [4–7]. Domain knowledge for identifying and segmenting the vertebrae may be requested from the medical expert during segmentation, but user input should be minimal for not unduly interfering with the clinical workflow during intervention.

* Corresponding author.
E-mail address: georg.hille@ovgu.de (G. Hille).

https://doi.org/10.1016/j.cmpb.2017.12.013
0169-2607/

The main motivation of this paper is to present an approach, which supports wide-ranging clinical applicability referred to the large variety of clinical routine spine MRI datasets. Typically, they consist of different image acquisition parameters and sequences, spatial resolution, spine section and healthy vertebrae as well as pathologies like fractures and metastases. Therefore, we propose a hybrid level-set-based approach and assembled an evaluation set with clinical routine datasets as well as datasets for research purposes.

Some research has been carried out on spinal segmentation, though most approaches were applied to mid-sagittal 2D images [8]. Their main disadvantage lies in processing only discrete slices. Thus, important information supporting reliable and precise measurements and diagnoses are omitted. A few 3D segmentation approaches were presented for MRI, which we will discuss below.

Hoad et al. [9] presented a combination of a thresholded region-growing algorithm with morphological filtering and shape masking for segmenting vertebral bodies and posterior structures in isotropic ($1 \times 1 \times 1$ mm^3) steady state precession acquisition sequence images. Their method was designed for this particular case rather than for anisotropic clinical routine spine MRI. Their evaluation set contained 30 vertebrae, achieving an Euclidean surface registration error of 1.25 ± 0.28 mm compared to a thresholded CT segmentation of the same subject.

Another segmentation strategy is based on deformable models, e.g., active contour models (ACM) [10,11], active shape models (ASM) [12] or finite element models (FEM) [13]. Davatzikos et al. [14] trained a deformable shape model to register image data with template images. A deformable model of the lumbar spine was initially placed in the test images and subsequentely deformed to match image gradients. They achieved an average Dice similarity coefficient (DSC) of $81.5 \pm 3.6\%$ on routine images of young healthy volunteers with a spatial resolution of $0.93 \times 0.93 \times 3$ mm^3.

Štern et al. [5] also applied a model-based approach, while optimizing 29 shape parameters by maximization of the dissimilarity between inner and outer object intensities, guided by image gradients. Their approach was initialized with one point per vertebra and by estimating the size by specification of the spine segment (upper/lower thoracic and lumbar). The evaluation set contained 75 vertebral bodies of nine subjects, three of them with resolution of $0.4 \times 0.4 \times 3$ mm^3 and six with isotropic voxels ($1 \times 1 \times 1$ mm^3). Their approach resulted in a radial Euclidean distance between segmented object surface and ground truth points of 1.85 ± 0.47 mm. Processing time ranged from 1 to 15 min per vertebra.

Neubert et al. [4,15] used ASM to segment vertebral bodies and intervertebral discs alike. They tested their fully automatic approach on 14 healthy volunteers with 132 vertebrae, acquired with high resolution MRI ($0.34 \times 0.34 \times 1$ to 1.2 mm^3) They obtained a mean DSC of 91% and a mean Hausdorff distance of 4.08 mm. However, the average run time per vertebra of 35 min [15] was high. Hence, segmentation of an entire dataset required approximately 5 h computing time, although no pathologic data was included.

Ayed et al. [16] pursued the idea of formulating the segmentation in MRI as a distribution-matching problem with a convex relaxation solution. For efficient computation, they split their problem into various sub-problems, where each one could be solved via convex relaxation and the augmented Lagrangian method. A mean DSC of 85% was achieved, but was only determined on 2D mid-sagittal slices.

Zukić et al. [17] combined edge and intensity-based features, i.e. Canny edges and thresholded gradient magnitudes to a multiple-feature-based model. Their approach was initialized by a previous vertebral center detection step using a Viola-Jones detector. The surface mesh of their model was enlarged by balloon forces and constrained by smoothness and the approximated vertebral body size. They achieved an average DSC of 79.3% and a mean surface-to-surface distance of 1.76 ± 0.38 mm. The method was evaluated on clinical routine datasets consisting of a large variety of MRI sequences including both healthy and pathological vertebrae.

Athertya et al. [18] proposed a fuzzy C-means clustering for vertebral body segmentation in T_1-weighted MR images. They assessed their method on 16 cases resulting in a mean DSC of 86.7% and a Hausdorff distance of 5.40 mm. The fuzzy C-means clustering was followed by various morphological operations including a shape ratio criteria to extract the vertebrae from surrounding tissues.

Chu et al. [6] fully automatically localized vertebral bodies to define ROIs for a subsequent segmentation step, where they were using random forest classification for estimating the fore- or background likelihood of each pixel within the produced ROIs. The results were combined with a learned probability map to segment each vertebral body via thresholding. Chu et al. tested their approach on 23 T_2-weighted images, without stating any pathologies, achieving an overall DSC of 88.7%, a mean absolute surface distance of 1.5 ± 0.2 mm and an average Hasudorff distance of 6.4 ± 1.2 mm. The average computational time per data set was about 1.3 min.

More recently Korez et al. [7] introduced a convolutional neural network (CNN)-based approach in spine MRI segmentation. Their method linked active shape models with likelihood maps of the vertebral bodies and achieved an overall DSC of 93.4%, an average Hausdorff distance of 3.83 mm and a mean symmetric surface distance of 0.54 mm. Korez et al. trained and tested their methods on the 23 T_2-weighted images made publicly available from Chu et al. [6].

Goankar et al. [19] presented a machine learning-based system for vertebral body segmentation on clinical MR images of the lumbar spine. In contrast to Chu et al. and Korez et al. they examined the applicability of their method to different MRI sequences, though they trained only on T_2-weighted images. The implementation of superpixels based multi-parameter ensemble learning was followed by some morphological post-processing to increase segmentation scores. Goankar et al. had in total 48 sagittal T_2 and 15 T_1 MR scans and randomly selected 6 T_2 scans for training procedure. The spatial resolution varied in-plane from 0.34×0.34 to 1.1×1.1 mm and slice thickness was between 0.5 to 5.0 mm. While training and segmenting vertebrae on T_2-weighted images their resulting mean DSC was 83%. Vertebrae segmentation on T_1-weighted images after training on T_2-weighted images resulted expectably in lower DSC scores (average 75%).

2. Materials and method

2.1. Image data

We assembled an evaluation set, which includes image data of both various clinical and research purposes. It consists of four different databases, including 63 sagittal datasets with overall 419 vertebral bodies of the thoracic and lumbar spine, which differ in age, sex and presence of spine pathologies. The evaluation data was acquired in different hospitals with various MRI scanners, by which means robustness of our method w.r.t. the diverse clinical routine settings in spine MRI is assessable (see Table 1). Regarding this, a key characteristic of most datasets is the high anisotropy factor (slice spacing divided by in-plane pixel spacing), ranging from 1.6 to 8.19 (see Fig. 1).

The first of our databases consists of pre-interventionally acquired MRI data before radiofrequency ablations (RFA). This data includes vertebral bodies with metastases from different primary tumours. Dependent on the origin, these metastases are of osteoplastic or osteolytic type and therefore differently affect vertebrae

Table 1
Characterization of all datasets used for our evaluation. MRI_{Seq} - MRI sequence (SE - spin echo, TSE - turbo spin echo, TIRM - turbo inversion recovery magnitude), $P_{x,y}$ - pixel spacing in mm, S_z - slice thickness in mm, M - acquisition matrix, F_A - anisotropy factor, $\#_V$ - number of segmented vertebral bodies, SpS - spine segment (C - cervical, T - thoracic, L - lumbar), Path. - predence of pathology (n.s. - not stated), R - reference segmentation by N (neuroradiologists resp. neurosurgeon) or T (trained field expert). The horizontal lines categorize the datasets according to their origin (first section - pre-interventionally acquired before RFAs, second section - publicly released by Zukić et al. [17], third section - publicly released by Chu et al. [6], fourth section - part of the SHIP study [20], T_1/T_2 means, we evaluated both sequences from the same subject.

Dataset	MRI_{Seq}	$P_{x,y}$	S_z	M	F_A	$\#_V$	SpS	R	Path.	Age	Sex
preRFA_1	T_1 TSE	0.5	3.3	$640 \times 640 \times 20$	6.6	5	C7–T4	T	–	54	F
preRFA_2	T_1 TSE	0.78	3.3	$512 \times 512 \times 20$	4.23	7	T1–T7	T	+	70	M
preRFA_3	T_1 TSE	0.68	3.3	$512 \times 512 \times 20$	4.85	6	T12–L5	T	+	61	M
preRFA_4	T_1 TSE	0.49	3.3	$528 \times 528 \times 17$	6.73	7	T3–T9	T	+	76	M
preRFA_5	T_1 TSE	0.49	3.3	$528 \times 528 \times 15$	6.73	8	T7–L2	T	+	74	M
preRFA_6	T_1 TSE	0.46	3.3	$640 \times 640 \times 17$	7.17	5	T12–L4	T	+	76	M
Aka2	T_2 FSE	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	21	F
Aka3	T_1 FSE	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	21	F
Aka4	TIRM	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	21	F
Aks5	T_2 FSE	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	22	F
Aks6	T_1 FSE	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	22	F
Aks7	TIRM	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	22	F
Aks8	T_1 FSE	0.70	4	$512 \times 512 \times 15$	5.69	8	T10–L5	T	+	22	F
C002	T_2 TSE	1.12	3.3	$448 \times 448 \times 31$	2.96	12	T6–L5	N	+	74	F
DzZ_T2	T_2 TSE	0.55	4.4	$640 \times 640 \times 12$	8.05	8	T10–L5	T	–	27	M
DzZ_T1	T_1 TSE	0.68	4.4	$512 \times 512 \times 12$	6.44	8	T10–L5	T	–	27	M
F02	T_2 SE	0.5	3.85	$768 \times 768 \times 18$	7.7	8	T10–L5	N	+	51	M
F03	T_2 TSE	1.19	3.3	$320 \times 320 \times 25$	2.77	6	T12–L5	N	+	72	M
F04	T_2 TSE	1.12	3	$448 \times 448 \times 23$	2.69	12	T6–L5	N	+	69	F
S01	T_2 SE	0.47	3.85	$640 \times 640 \times 16$	8.19	6	T12–L5	N	+	65	M
S02	T_2 SE	0.47	3.85	$640 \times 640 \times 16$	8.19	7	T11–L5	N	+	55	F
St1	T_2 SE	0.5	3.85	$704 \times 704 \times 20$	7.7	7	T11–L5	N	+	71	M
Chu (1–23)	T_2 TSE	1.25	2.0	$305 \times 305 \times 39$	1.6	7	T11–L5	T	n.s.	n.s.	F, M
SHIP (1–9)	T_1/T_2 TSE	1.12	4.4	$448 \times 448 \times 15$	3.67	5	L1–L5	N, T	n.s.	29–65	F, M

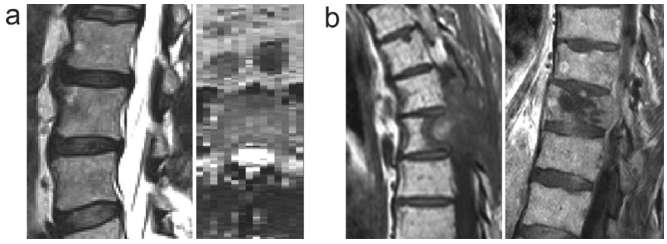

Fig. 1. A sagittal (a, left) and reconstructed axial (a, right) slice demonstrate how partial volume effects caused by high anisotropy could hamper the distinction between bony and surrounding structures. Furthermore, various types of metastases affect the signal intensities different, depending on the particular MRI sequence (b).

in their shape and visibility. While bony structures emit similar signals in T_1- and T_2-weighted MRI sequences, metastases could considerably differ in image intensities. This tremendously complicates automatic segmentation methods.

Since the comparison of segmentation approaches and their results between entirely different datasets must be considered as indirect, our evaluation set furthermore consists of overall 39 datasets made publicly available together with the related work of Zukić et al. [17] and Chu et al. [6]. Hence, we could match our results directly with those works.

The database from Zukić et al. [17] includes both healthy and pathologic datasets, e.g., with scoliosis, spondylolisthesis and vertebral fractures. Their data was acquired in different hospitals, consisting of various MRI sequences and parameter settings.

The third database comprises 23 T_2-weighted turbo spin echo MR images of thoracolumbar spine of volunteers and is publicly available [6]. It is not stated, whether it contains spine pathologies.

The fourth database comprises of epidemiological image data from the Study of Health in Pomerania (SHIP) [20]. It features spine MR images of volunteers, including T_1- and T_2-weighted sequences. This cohort study provides high anisotropic image data solely for research purposes. Using this data in our study serves the purpose to understand the limits of our method regarding spatial resolution and image quality, since MRI acquisition time in SHIP was limited leading to images of lower quality.

The reference segmentation of the ground truth was performed manually either by neuroradiologists respectively neurosurgeons or by trained field experts (see Table 1). For both, the pre-interventionally acquired and the SHIP datasets a second reader produced a reference segmentation to assess the impact of inter-observer variability on segmentation quality measures.

2.2. Methodology

The major steps of our proposed method are as followed (see also Fig. 2):

1. Initially a Gaussian filter-based intensity correction was implemented as a pre-processing step to deal with bias field artifacts. We set the filter kernel size to $120 \times 120 \times 30$ mm^3 and σ to 20 mm to estimate the bias field of each image volume. In order to remove it, the original image was divided by the bias estimation. Subsequently, each image was laterolateral upsampled to the in-plane resolution to provide spatial isotropy.
2. We initialized our method with three points in a selectable mid-sagittal cross-section to approximate the size, center and sagittal orientation of each vertebral body. For this purpose, both corners of the superior endplate as well as the posterior corner of the inferior endplate were marked. The lateral flection angle could be deduced from interpolating the landmarks' z-coordinates of consecutive vertebral bodies.
3. Intensity-based features, e.g., median and variance, were obtained from a cube within the vertebral center and with variable edge length, i.e., two fifths of the specific vertebral body height and length.

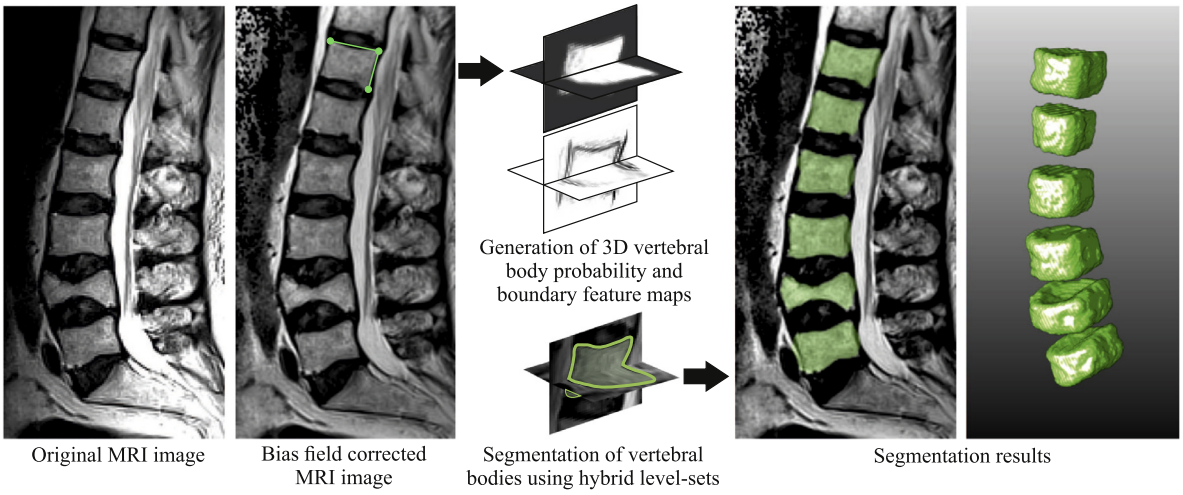

Fig. 2. The pipeline of the presented methodology.

4. An abstracted vertebral body shape model was placed upon each vertebral center with the approximate vertebral body length, height and orientation.
5. Within this shape, a pre-segmentation was performed based on adaptive thresholding. The previously gained intensity-based features ensure patient independence, as well as imaging sequence independence and therefore, avoid common difficulties regarding thresholding in MR images. Subsequent the result was morphologically filtered, at first by hole filling and dilating with a 3 mm-diameter ball structering element and removing objects smaller than 1 cm^3. To yield the vertebral body probability map P, the resulting binary image was distance-transformed by a Gaussian convolution (kernel size of 10 mm^3 and $\sigma = 2$) and multiplied with the source image. This smoothing weakens local constrains at the boundaries of the pre-segmented object and enables level-set convergence away from disadvantageous placed shape models.
6. Boundary feature maps G of each vertebral body were computed via dilating the extracted boundaries of the fitted vertebral body shape model, using a 3 mm-diameter ball structering element, subsequently distance transform them likewise the probability map and multiplying them with the gradient magnitude images. This feature ensured level-set convergence towards object boundaries within the range of the model contours. The probability map P and the boundary feature map G define both terms of the hybrid level-set formulation (1).

The 3D hybrid level-sets approach is based on the work presented by Zhang et al. [21] and combines regional intensity and boundary features with an approximative geometry of the target object for steering and constraining the curve towards vertebral body boundaries. The region information form a counterweight to attenuate leakage problems frequently emerged in boundary-based methods. The level-set-functional to be minimized is defined as:

$$\mathcal{E}(\phi) = -\alpha \int_{\Omega} P \cdot H(\phi) d\Omega + \beta \int_{\Omega} G \cdot |\nabla H(\phi)| d\Omega, \tag{1}$$

where $H(\phi)$ represents the Heaviside function, Ω is the image domain and the weights α and β were used to balance both terms. We empirically determined the ratio of α to β with 4: 3. The probability map P encourages the level-set contours to enclose regions of a specific per-vertebra intensity range and is defined as:

$$P = g(I_s(\underline{x}), \sigma) \cdot I(\underline{x}), \tag{2}$$

where $I(\underline{x})$ is the pre-processed image from step 1 and $g(I_s(\underline{x}), \sigma)$ is the result of the distance transformed and morphologically filtered pre-segmentation $I_s(\underline{x})$ in step 5. The boundary feature map G is the functional of the geodesic active contour term in the hybrid level-set formulation and is defined as:

$$G = g(S_c(\underline{x}), \sigma) \cdot |\nabla I(\underline{x})|, \tag{3}$$

where $|\nabla I(\underline{x})|$ is the gradient magnitude image and $g(S_c(\underline{x}), \sigma)$ is the result of the distance transformation via Gaussian convolution of $S_c(\underline{x})$, which represents a binary image of the extracted and dilated shape model contour.

With ϕ_t defined as a signed distance function:

$$\phi_t = \alpha P + \beta div(G \nabla \phi), \tag{4}$$

it could be derived as an simplified iterative approximation of a partial differential equation (PDE) from the gradient flow applied to (1). Like Zhang et al. [21] initally proposed, each iteration step i starts with a re-initialization of ϕ^i, subsequent the embedded function ϕ evolves as an intermediate step with a predefined time step using:

$$\bar{\phi}^i = \phi^i + \triangle_t \alpha P. \tag{5}$$

After re-initializing $\bar{\phi}^i$, it is updated to ϕ^{i+1} by solving the PDE:

$$\phi_t = \beta div(g \nabla \bar{\phi}^i). \tag{6}$$

2.3. Evaluation

Evaluation was done on an AMD Phenom II X4 955 processor. Ground truth segmentations were available for each dataset, created by radiologists or trained field experts. To match our results with as many related approaches as possible, we provide the average DSC, the Euclidean surface distances and Hausdorff distances.

3. Results

Inter-observer variability was assessed for the first and fourth database, where two reference segmentations for each dataset were available. The average DSC between two reference segmentations is 88.4% and therefore, similar to the result stated by Zukić et al. [17]. The mean Euclidean surface distance is 0.76 ± 0.4 mm. Furthermore, there is no essential difference regarding the overall DSC between two trained field experts and a neuroradiologist with a field expert (88.7% to 88.1%), though the mean Euclidean surface distance (1.33 ± 0.47 mm to 1.70 ± 0.97 mm) is somewhat higher for the latter. These results are comparable with inter-observer

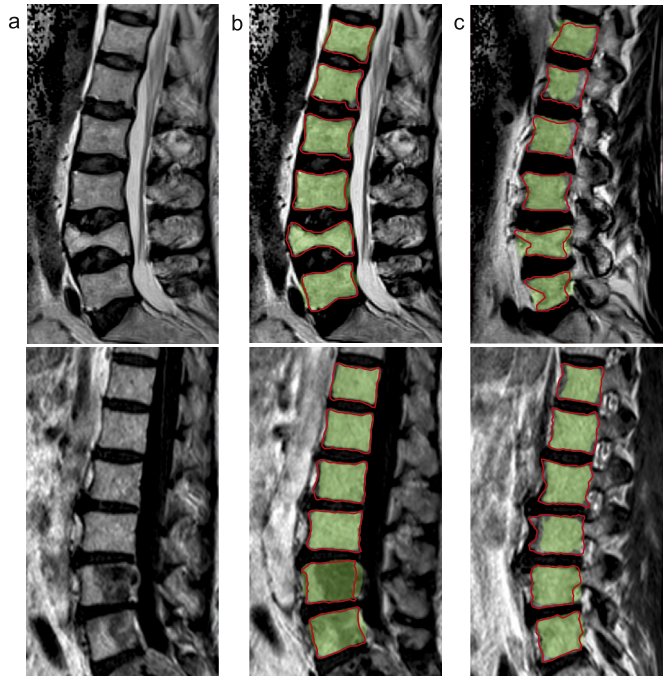

Fig. 3. Overlay of our results and the reference segmentation of dataset *F03* (top row) and *preRFA_3* (bottom row). Red contours correspond to the reference segmentations and green overlays illustrate the segmentation results produced by our method. Mid-sagittal (a and b) and a random sagittal cross-section are shown (c). The mean DSC of those datasets were 84.3% and 86.7%, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variabilities of segmentation tasks regarding other organs or structures [22,23]. Additionally, they must kept in mind, when assessing results of segmentation methods, e.g., are DSC scores well over 90% still a quality increase or merely indications of overfitting, especially in learning based methods?

The overall average DSC of our evaluation set is 86.0%, the mean Euclidean surface distance is 1.59 ± 0.24 mm and the mean Hausdorff distance is 6.86 mm. Most discrepancies between the reference and our semi-automatic segmentation arose from lateral slices, caused by the impact of partial volume effects, complicating the algorithmic detection of object boundaries (see Figs. 1 and 3). Segmentation is typically a trade-off between data- and model-driven terms, which is why we strengthened the latter at locations of weak image boundaries to avoid leakage problems, common for level-set methods [21]. However, this increased the dependence of segmentation on model assumptions, for example the spatial extent of the vertebral body model, promoting under- or oversegmentation at such locations. The correlation of image data and model knowledge is a fundamental issue for any segmentation method based on model knowledge [5,15,17]. To attenuate level-set convergence towards disadvantageously placed model boundaries, we implemented distance-based features in both, the probability and the boundary feature map, in order to enable convergence towards image gradients within the near surrounding. Additionally, pathologies like vertebral fractures or metastases hamper segmentation approaches, according to deformations and atypical intensities within and round vertebral bodies, especially if the cortical layers are affected. The latter manifests itself as weak or discontinuous delineations towards surrounding tissues, promoting leakage problems, which could be compensated via strengthened local model terms.

There is a slightly difference in the segmentation quality between T_1- and T_2-weighted sequences, where the latter suffers especially from oversegmentation problems due to similar intensities within the vertebral body and the cerebrospinal fluid of the adjacent spinal cord. Low spatial resolutions, as commonly used in clinical routine spine MR imaging promote this effect.

Average per-vertebra computational time was 5.4*s*, whereby the execution time strongly depends on the spatial resolution. This impact is apparent from the computational time differences between datasets of Zukić et al. with partially twice the spatial resolution of those datasets of Chu et al. (see Table 2).

4. Discussion

In this study a segmentation approach for vertebral bodies in clinical routine MRI was presented. We particularly placed importance on clinical requirements and settings, a focus most previous works shifted away from. Therefore, the major contribution of the proposed work is a precise segmentation result, computed within seconds, despite the large variety of datasets and the presence of pathologies.

With regard to clinical application the accuracy and precision of segmentation approaches are of primary importance. In respect of accuracy the proposed method is comparable to the state-of-the-art (see Table 2), achieving mainly superior results [5,17,19,24,25], even though few works showed more precise segmentation results [7,15]. Most of the related work focused only on one particular MR sequence, however, clinical settings require an applicability for various imaging sequences according to the diagnostic purposes. Those works from Korez et al. [7] and Neubert et al. [15] with a significantly higher segmentation accuracy than ours, considered only T_2-weighted images in their evaluation and therefore, missing out an important aspect of clinical settings. Furthermore, the high quality of the work of Neubert et al. comes at a cost of a considerably longer processing time per vertebral body of 35 min (vs. our 5.4 s) on recent hardware. A 10-fold processing time reduction decreases their average DSC from 90.8% [15] to 85% [4], which is on a par with ours. Athertya et al. [18] achieved a mean DSC of 86.7%, which is comparable to our method (86.0%), however, their average Hausdorff distance is slightly lower than ours (5.40 mm vs. our 6.86 mm). They solely tested their method on T_1-weighted images and the post-processing step including an area criterion to extract the vertebra from oversegmented surrounding tissue raises doubts about applicability to pathologic altered vertebrae due to fractures, metastases etc. The authors did not state any spatial resolution of their MR images, although it has been found to be of crucial importance for segmentation approaches in clinical settings.

In addition to our comprehensively composed evaluation set, only Zukić et al. [17] and Goankar et al. [19] tested their methods on multiple MR sequences, though both achieving worse segmentation results than our proposed method. Goankar et al. [19] proofed robustness of their fully automated learning-based method despite variation of scanning parameters. Their mean DSC scores with 83% (trained and segmented on T_2-weighted scans) and 75% (trained on T_2-weighted scans, segmented on T_1-weighted scans) were lower than ours though. Futhermore, they did not state any computational times or distance measures.

Nevertheless, a comparison of different segmentation approaches must be considered as indirect, since the evaluated datasets differ from each other and therefore, should be interpreted cautiously. Though, we could compare our approach directly with the work of Zukić et al. [17], since they publicly provide a large part of their evaluation data sets and corresponding results. Following results refer solely to these shared data. Both, in mean DSC (79.4% vs. our 84.1%) and distance measures like mean surface-to-surface distance (1.81 mm vs. our 1.68 mm) and Hausdorff distance (12.36 mm vs. our 7.89 mm) our method is more precise. Nonetheless, their method did not inevitably require man-

Table 2
Previous and the presented work in comparison. MRI_{Seq} - MRI sequence, $\#_{DS}$ - number of datasets, $\#_V$ - number of segmented vertebral bodies, SpS - spine segment (T - thoracic, L - lumbar), DSC - Dice similarity coefficient, ED - average Euclidean surface distance, HD - Hausdorff distance, t_C - computational time, ^{AD} - all databases, ^{Z} - datasets publicly provided by Zukić et al. [17], ^{C} - datasets publicly provided by Chu et al. [6].

Works	MRI_{Seq}	$\#_{DS}$	$\#_V$	SpS	DSC [%]	ED [mm]	HD [mm]	t_C
Štern et al. [5]	T_2	9	75	T, L	–	1.85 ± 0.47	–	1–15 min
Kadoury et al. [25]	T_1	8	136	T, L	–	2.95 ± 1.85	–	–
Neubert et al. [15]	T_2	14	132	T, L	90.8 ± 1.8	0.67 ± 0.17	4.08 ± 0.94	35 min
Zukić et al. [17]^{Z}	T_1, T_2	17	153	T, L	79.3 ± 5.0	1.76 ± 0.38	11.89 ± 2.56	8.3s
Schwarzenb. et al. [24]	T_2	2	10	L	81.3 ± 5.1	–	–	19s
Chu et al. [6]^{C}	T_2	23	161	T, L	88.7 ± 2.9	1.5 ± 0.2	6.4 ± 1.2	–
Korez et al. [7]^{C}	T_2	23	161	T, L	93.4 ± 1.7	0.54 ± 0.14	3.83 ± 1.04	–
Athertya et al. [18]	T_1	16	–	T, L	86.7 ± 4.1	–	5.40 ± 1.12	5.6s
Goankar et al. [19]	T_1, T_2	57	–	T, L	79 ± 5.0	–	–	–
Ours^{AD}	T_1, T_2	63	419	T, L	86.0 ± 3.9	1.59 ± 0.24	6.86 ± 1.06	5.4s
Ours^{Z}	T_1, T_2	17	153	T, L	84.1 ± 2.5	1.68 ± 0.24	7.89 ± 1.12	14.0s
Ours^{C}	T_2	23	161	T, L	88.2 ± 1.9	1.66 ± 0.28	6.01 ± 1.01	1.3s

ual user input, but is based on the detection accuracy of a Viola-Jones detector. Their mean execution time per vertebra is shorter than ours ($\approx$8.3 s vs. our $\approx$21.8 s), but one must consider that we tested our approach on standard rather than on high-end hardware likewise in clinical settings.

Publicly provided datasets of Chu et al. [6] enabled the direct comparison with both their results and those of Korez et al. [7]. Following results refer solely to these shared datasets. Our results are on a par with the work of Chu et al. [6] differing only in decimal place of DSC or distance measures. Korez et al. [7] and their deep learning-based method achieved superior results, but arouse doubts about applicability in clinical settings concerning computational time and the variety of MR sequences. While not stating any performance measures, we re-implemented their network to review their approximate processing time. To predict a medium sized patch under idealized conditions (with NVIDIA GeForce GTX 970) it took about 3.75 min, while our approach required only 1.26 s per vertebra. Furthermore, clinical applicability of CNN-based spine MRI segmentation techniques still needs to be verified, with regard to the variety of MRI sequences and pathologically altered vertebrae. For a thorough assessment of segmentation results, inter- and intra-rater variability should be taken into account, since the variability of manually produced ground truth data highly depends on numerous factors including, e.g., the complexity of the anatomical structure, the image quality and the rater's expertise. Inter-rater variability is known to range between mean DSCs of 88.4% and 91% [17] for thoracolumbar spine MR data. Our method has proven robustness w.r.t. clinical settings while achieving encouraging segmentation quality within reasonable computing time.

5. Conclusion

We comprehensively evaluated our hybrid level-set approach for vertebral body segmentation in clinical routine spine MRI. It combines regional intensity and boundary features to steer and constrain level-set curves towards vertebral body boundaries. The semi-automatic initialization with approximate vertebral body center and size determination increases the robustness of segmentation w.r.t. the spine section, imaging sequence and deforming pathologies. Overall our evaluation set contains 63 datasets, differing in age, sex and presence of spine pathologies, including in total 419 vertebrae. Compared to related works this was the largest test set so far. The average DSC of the evaluation set is 86.0%, while a mean Euclidean surface distance 1.59 ± 0.24 mm and a Hausdorff distance of 6.86 mm was achieved. The major contribution of the presented approach is its applicability to a large variety of MRI sequences and parameter settings, while requiring only minimal user input and providing results within seconds. Additionally, the proposed method is suitable for image data with the presence of pathologies like fractures, scoliosis or spinal metastases. These are essential requirements for the clinical applicability. While most related works shifted away from the limitations of clinical settings, our method proofs robustness and precision. In future work, the segmentation should be combined with automatic initial vertebrae detection and can be also employed as a prerequisite for co-registration of multimodal images.

Acknowledgments

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

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. For this type of study formal consent is not required.

Funding

This work was funded by the Federal Ministry of Education and Research within the Forschungscampus *STIMULATE* (grant number 13GW0095A).

Conflict of interest

None declared.

References

[1] M. Rak, K.D. Tönnies, On computerized methods for spine analysis in MRI: a systematic review, Int. J. Comput. Assist. Radiol. Surg. 11 (8) (2016) 1445–1465.
[2] A.A. Darwish, M.A.-M. Salem, D. Hegazy, H.M. Ebeid, Vertebrae segmentation techniques for spinal medical images, in: Proceedings of ICICIS, IEEE, 2015, pp. 110–115.
[3] K. Hammernik, T. Ebner, D. Stern, M. Urschler, T. Pock, Vertebrae segmentation in 3D CT images based on a variational framework, in: Proceedings of MICCAI, Springer, 2015, pp. 227–233.
[4] A. Neubert, J. Fripp, K. Shen, O. Salvado, R. Schwarz, L. Lauer, C. Engstrom, S. Crozier, Automated 3d segmentation of vertebral bodies and intervertebral discs from MRI, in: Proceedings DICTA, 2011, pp. 19–24.
[5] D. Štern, B. Likar, F. Pernuš, T. Vrtovec, Parametric modelling and segmentation of vertebral bodies in 3d CT and MR spine images, Phys. Med. Biol. 56 (23) (2011) 7505.
[6] C. Chu, D.L. Belavỳ, G. Armbrecht, M. Bansmann, D. Felsenberg, G. Zheng, Fully automatic localization and segmentation of 3D vertebral bodies from CT/MR images via a learning-based method, PLoS ONE 10 (11) (2015) e0143327.

[7] R. Korez, B. Likar, F. Pernuš, T. Vrtovec, Model-based segmentation of vertebral bodies from MR images with 3D CNNs, in: Proceedings of MICCAI, Springer, 2016, pp. 433–441.
[8] S.-H. Huang, Y.-H. Chu, S.-H. Lai, C.L. Novak, Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI, IEEE Trans. Med. Imag. 28 (10) (2009) 1595–1605.
[9] C.L. Hoad, A.L. Martel, Segmentation of MR images for computer-assisted surgery of the lumbar spine, Phys. Med. Biol. 47 (19) (2002) 3503.
[10] V. Caselles, R. Kimmel, G. Sapiro, Geodesic active contours, Int. J. Comput. Vis. 22 (1) (1997) 61–79.
[11] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 1 (4) (1988) 321–331.
[12] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image 61 (1) (1995) 38–59.
[13] M. Rak, K. Engel, K.D. Tönnies, Closed-form hierarchical finite element models for part-based object detection, in: Vision, Modeling and Visualization, 2013, pp. 137–144.
[14] C. Davatzikos, D. Liu, D. Shen, E.H. Herskovits, Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms 1, Radiology 224 (3) (2002) 919–926.
[15] A. Neubert, J. Fripp, C. Engstrom, R. Schwarz, L. Lauer, O. Salvado, S. Crozier, Automated detection, 3d segmentation and analysis of high resolution spine MR images using statistical shape models, Phys. Med. Biol. 57 (24) (2012) 8357.
[16] I.B. Ayed, K. Punithakumar, R. Minhas, R. Joshi, G.J. Garvin, Vertebral body segmentation in MRI via convex relaxation and distribution matching, in: Proceedings of MICCAI, Springer, 2012, pp. 520–527.
[17] D. Zukić, A. Vlasák, J. Egger, D. Hořínek, C. Nimsky, A. Kolb, Robust detection and segmentation for diagnosis of vertebral diseases using routine MR images, Comput. Graph Forum 33 (6) (2014) 190–204.
[18] J. Athertya, G.S. Kumar, et al., Fuzzy clustering based segmentation of vertebrae in T1-weighted spinal MR images, Int. J. Fuzzy Log Intell. Syst. 6 (2016) 23–34.
[19] B. Goankar, Y. Xia, D. Villaroman, A. Ko, M. Attiah, J. Beckett, L. Macyszyn, Multi-parameter ensemble learning for automated vertebral body segmentation in heterogeneously acquired clinical MR images, IEEE J. Transl. Eng. Health Med. 5 (2017) 1–12.
[20] H. Völzke, D. Alte, C.O. Schmidt, D. Radke, R. Lorbeer, et al., Cohort profile: the study of health in pomerania, Int. J. Epidemiol. 40 (2011) 294–307.
[21] Y. Zhang, B.J. Matuszewski, L.K. Shark, C.J. Moore, Medical image segmentation using new hybrid level-set method, in: Proceedings of MEDIVIS, 2008, pp. 71–76.
[22] X. Geets, J.-F. Daisne, S. Arcangeli, E. Coche, M. De Poel, T. Duprez, G. Nardella, V. Grégoire, Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI, Radiother Oncol 77 (2005) 25–31.
[23] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index 1: scientific reports, Academic Radiol. 11 (2004) 178–189.
[24] R. Schwarzenberg, B. Freisleben, C. Nimsky, J. Egger, Cube-cut: vertebral body segmentation in MRI-data through cubic-shaped divergences, PLoS ONE 9 (2014) e93389.
[25] S. Kadoury, H. Labelle, N. Paragios, Spine segmentation in medical images using manifold embeddings and higher-order MRFs, IEEE Trans. Med. Imag. 32 (7) (2013) 1227–1238.