

Robust Detection and Segmentation for Diagnosis of Vertebral Diseases using Routine MR Images

Dženan Zukić^{1†}, Aleš Vlasák^{2,5}, Jan Egger³, Daniel Hořínek^{4,5}, Christopher Nimsky³, and Andreas Kolb¹

¹Computer Graphics and Multimedia Systems Group, University of Siegen, Germany

²Department of Neurosurgery, University Hospital Motol, Prague, Czech Republic

³Department of Neurosurgery, University Hospital Gießen and Marburg, Marburg, Germany

⁴Department of Neurosurgery, Jan Evangelista Purkyně University, Ústí nad Labem, Czech Republic

⁵International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic

Abstract

The diagnosis of certain spine pathologies, such as scoliosis, spondylolisthesis and vertebral fractures, are part of the daily clinical routine. Very frequently, MRI data are used to diagnose these kinds of pathologies in order to avoid exposing patients to harmful radiation, like X-ray.

We present a method which detects and segments all acquired vertebral bodies, with minimal user intervention. This allows an automatic diagnosis to detect scoliosis, spondylolisthesis and crushed vertebrae. Our approach consists of three major steps. First, vertebral centers are detected using a Viola-Jones like method, then the vertebrae are segmented in a parallel manner and, finally, geometric diagnostic features are deduced in order to diagnose the three diseases.

Our method was evaluated on 26 lumbar datasets containing 234 reference vertebrae. Vertebra detection has 7.1% false negatives and 1.3% false positives. The average Dice coefficient to manual reference is 79.3% and mean distance error is 1.76 mm. No severe case of the 3 illnesses was missed, and false alarms occurred rarely – 0% for scoliosis, 3.9% for spondylolisthesis and 2.6% for vertebral fractures.

The main advantages of our method are high speed, robust handling of a large variety of routine clinical images, and simple and minimal user interaction.

Categories and Subject Descriptors (according to ACM CCS): I.4.6 [Image processing and computer vision]: Segmentation—Pixel classification

1. Introduction

Lower back pain for adults is rather common and its prevalence is rising [FHA*09]. The most common causes involve spinal cord, such as narrowing (stenosis) of the spinal canal, and that has been well investigated. For diseases involving the vertebrae 2D X-ray is frequently used for screening, but authoritative diagnosis is made based on Computed Tomography (CT) or Magnetic Resonance Image (MRI). The cancer risk from radiation exposure in CT imaging [RGMB10] makes MRI preferable in the clinical routine. Furthermore, 2D X-ray and CT cannot reveal all pathologies, and frequently the usage of MRI is diagnostically required.

Pathologies such as scoliosis (curvature in anatomical left-right direction), vertebral fracture (crushed vertebra) and spondylolisthesis (misaligned vertebra) can be diagnosed from vertebral shapes, positions and orientations, so their segmentations are a step required for diagnosis.

Most 3D approaches focus on CT datasets only, such as [KOE*09, ML13]. However, these methods rarely transfer to MRI because of additional challenges. Routine MRI has a lower and strongly anisotropic resolution. Unlike CT's Hounsfield unit (HU), MRI does not have standardized units of measurement. In spine CT, bone edges are the only high-intensity edges. In MRI there are strong edges between many tissue types including edges within vertebral bodies, e.g. Fig. 11. MRI also has a non-homogeneous intensity across

† Contact author, dzenan.zukic@uni-siegen.de

the image, e.g. the central region has higher intensity and better contrast than the marginal areas. Lastly, MRI has many different parameters which can be changed to emphasize different tissues, resulting in many different scanning sequences which sometimes produce radically different image intensities. All these facts are detrimental to the automation of segmentation procedures for MRI datasets.

Spinous processes are poorly seen on MR images of low inter-slice resolution, which are the routine in clinical practice. Also, transverse processes are usually not seen at all because they are outside of the acquisition volume. Therefore, we focus on the segmentation of vertebral bodies instead of the whole vertebrae.

In this paper we focus on a segmentation system of practical clinical usefulness. This kind of system is required to work on a large variety of routine clinical datasets containing pathologies. It has to be reasonably fast and should not require cumbersome initialization. Finally, it must support diagnosis of diseases of interest.

We present a novel method for the segmentation of vertebral bodies in routine MRI datasets. These datasets have highly anisotropic voxels (e.g. $0.6 \times 0.6 \times 4.4 \text{ mm}^3$) and our evaluation set consists of a large variety of MRI sequences, see Tab. 2. Our three-stage approach first detects vertebral centers, then segments the vertebrae in a parallel manner and, finally, deduces geometric diagnostic features in order to diagnose scoliosis, spondylolisthesis and vertebral fractures.

The major motivation for extending our previous approach [ZVD*12] was to further increase robustness, which is the key feature for a reliable diagnosis of vertebral pathologies. This paper presents the following contributions:

- Automatic detection of vertebral body centers and sizes based on a Viola-Jones [VJ01] method with novel candidate filtering, which enables
 - a global prediction of all vertebra centers, which improves the overall segmentation and diagnosis process,
 - parallelized computation of per-vertebra features, classified image merging and constrained inflation, and
 - a novel size-goal force, which guides the segmentation to detected vertebral body size, thus improving segmentation accuracy.
- Optimization of algorithm parameters through training on manually segmented datasets.
- Lastly, our center+mesh representation of segmentation allows easy diagnostic measurements of the vertebral shape in order to detect scoliosis, spondylolisthesis and crushed vertebrae.

Our approach is efficient because it is globally single-pass, and iterative only on the level of segmenting individual vertebral bodies. Moreover, initial vertebral body detection enables parallelization of estimating per-vertebra features.

Therefore segmentation of any individual vertebral body is not dependent on segmentation results of any other vertebral body, so there is no possibility that a failure in segmentation will lead to misdetections or bad segmentations of the rest of the sequence. The robustness is thus improved.

2. Related Work

Much research has already been done on spine segmentation. There is a large number of 2D methods due to their relative simplicity and low computational requirements [MCP*09, HCLN09, CGBM04, EKD*12]. As 2D approaches process individual slices they can miss important information, such as a curvature or a positional shift in the anatomical left-right direction, so 3D approaches are preferable.

We are aware of only five prior fully 3D segmentation methods applied to MRI: Hoard and Martel [HM02], Davatzikos et al. [DLSH02], Štern et al. [ŠLPV11], Ayed et al. [APM*12], and Neubert et al. [NFS*11, NFE*12]. Some are concerned with detection only [RET13, KWZ*13, SMB11] and could be plugged into our method replacing initial vertebra detection. Kelm et al. [KWZ*13] optionally do segmentation of intervertebral disks, but report neither precision nor running time. He et al. [HPE*08] give a general survey of deformable models not specifically tailored to the medical (spine) segmentation.

Hoard and Martel [HM02] have developed a segmentation algorithm that combines thresholded region-growing with morphological filtering and masking using set shapes. Their algorithm is initialized by one or two ellipses per vertebra, and can be split into three steps: segmentation of the vertebral bodies, segmentation of the posterior structures, and lastly manual corrections. Their datasets have isotropic voxels ($1 \times 1 \times 1 \text{ mm}^3$), and they combined two images of a double echo FISP acquisition sequence. Overall, it is a method suited to assist spine surgery, using images quite different from routinely acquired ones. They tested their method on 30 vertebrae. Surface registration error for vertebral bodies was $1.25\text{mm} \pm 0.28\text{mm}$. They also calculated average percentage of “good” points to be 79.4%, with 3 rejected registrations out of 30 experiments. Running time of their algorithmic part was 5-10 minutes (not counting the manual steps).

Davatzikos et al. [DLSH02], with a long line of research in hierarchical deformable models [SHD01], were mainly interested in the registration of different spine images to a manually segmented template image. An isosurface was extracted from this template image resulting in 837 vertices (vertebral bodies L1 to S1 and a corresponding portion of the spinal canal). This deformable model was trained on 13 additional images. In order to determine the transformation which registers a test image with the template image, the surface model is initially placed in the test image overlapping the true position of the spine segment in the test image, and hierarchically deformed to conform to the edges of the test

image. The transformation between the stereotaxic space of the template image and space of the test image is used to determine the correlation between patient symptoms and image lesions. The evaluation was done using the leave-one-out method on routine images ($0.93 \times 0.93 \times 3 \text{ mm}^3$) of healthy volunteers. The average overlap was $81.5\% \pm 3.6\%$. They require only one initialization per dataset but do not state the execution time.

Štern et al. [ŠLPV11] perform the segmentation by optimizing 29 parameters of a 3D deterministic model of the vertebral body. They maximized dissimilarity between inside and outside intensities, and steered their method by image gradients. The method is initialized with one point per vertebra and an accompanying size, depending on the anatomical position, i.e. upper thoracic, lower thoracic or lumbar. They evaluated their method on 75 vertebral bodies from nine T_2 -weighted images. Three of their images were of a routine type ($0.4 \times 0.4 \times 3 \text{ mm}^3$), the others were isotropic ($1 \times 1 \times 1 \text{ mm}^3$). The mean radial Euclidean distance between the segmentation surface and ground truth points was $1.85\text{mm} \pm 0.47\text{mm}$. It takes 1-15 minutes per vertebra.

Ayed et al. [APM*12] formulate segmentation as a distribution-matching problem. They split it into a series of sub-problems, each of which can be solved via a convex relaxation and the augmented Lagrangian method. This results in a parallel method, and they implement CPU and GPU variants. They used T_2 -weighted MR images of the lumbar spine. They evaluated their results only on 2D mid-sagittal slices, where they reach 85% DSC – a precision which corresponds to 78% for 3D case. The GPU variant runs for 3 minutes, and the CPU variant for 75 minutes.

Neubert et al. [NFS*11, NFE*12] localize 3D spine curve and extract approximate positions of vertebral bodies using active rectangles. This serves as initialization for segmentation using active shape models with shape deformation using gray level models. Vertebral bodies and intervertebral discs are segmented. They use a high resolution SPACE MRI sequence with in-plane pixel size of $0.34 \times 0.34 \text{ mm}^2$ and slice thickness of $1 - 1.2 \text{ mm}$. They used 14 healthy volunteers with 132 vertebral bodies for quantitative evaluation. They achieve 91% DSC, 0.67mm mean absolute shape distance and 4.08mm Hausdorff distance. They also classified intervertebral discs into degenerate and healthy classes. Their method takes an average of 35 minutes per vertebra [NFE*12], not counting some 3 minutes spent on intensity normalization per dataset. The method is completely automatic. This means that the user has no chance to correct mis-detections, and this is especially important if running time (5h) is taken into account.

Our system addresses the stated drawbacks (complex initialization, long running time, special MRI sequences) of prior systems and allows for comparably fast and robust segmentation of pathological spine and vertebra shapes on a large variety of datasets (both T_1 and T_2 weighted). Com-

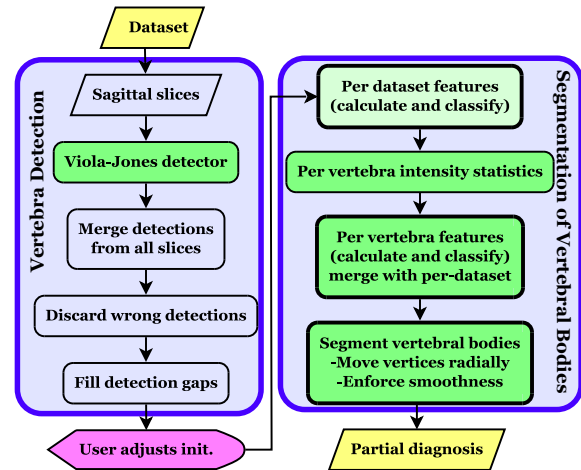


Figure 1: High level diagram of the full system. Boxes with thick black outlines are computationally intensive. Green boxes employ parallel computation on the CPU. Light-green is only partly parallelized.

pared to our prior method [ZVD*12], which has been based on a sequential detection of the vertebral bodies, our new approach detects all vertebral centers in an initial step. This increases the robustness of the detection and segmentation of the vertebral bodies and, thus, makes a reliable diagnosis possible.

3. Overview

The three major steps of our method are depicted in Fig. 1.

Vertebra Detection: In this stage, the vertebral centers are detected using a Viola-Jones method [VJ01]. The initially detected centers are filtered in order to remove false positive and false negative detections and thus to minimize the required user intervention. Finally, the sizes of the user-picked vertebral bodies are estimated (see Sec. 4).

Segmentation of Vertebral Bodies: The segmentation incorporates a multi-feature boundary classification and a balloon inflation based segmentation approach. The balloon inflation operates under a constrained subdivision hierarchy, which allows a smooth and robust segmentation with respect to the (still noisy) combined boundary probability map. The size parameter estimated in the detection stage is utilized to increase robustness (see Sec. 5). The methods used in this step, beside some minor modifications and the parallel execution, are directly taken from [ZVD*12].

Diagnosis: Based on the segmentation result, position and volume of the vertebral bodies are used to deduce disease-specific parameters which drive the diagnosis (see Sec. 6).

Please note, that in contrast to our prior work [ZVD*12],

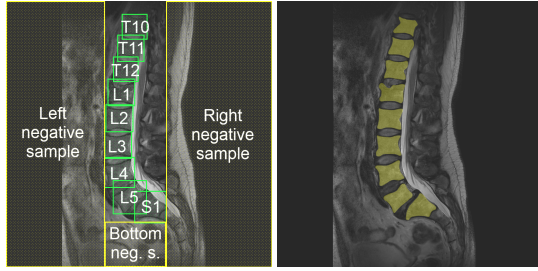


Figure 2: Left: training samples for Viola-Jones detector. Green boxes are positive samples, yellow boxes are negative samples. Right: ground truth overlaid over the original slice (dataset F02, slice 9).

our new approach is detecting all vertebral centers in a first step. This allows the parallel execution of several downstream processing steps (dark green components in Fig. 1).

4. Initial Vertebra Detection

This section describes the detection of all vertebral body centers, used as initialization for the segmentation step. After applying the Viola-Jones method to detect candidate vertebrae (Sec. 4.1), the candidates are filtered (Sec. 4.2) and minimal human intervention is needed to complete this step (see Sec. 4.3). The Viola-Jones method was chosen for its speed (a few seconds), because other approaches like the Generalized Hough Transform or that of Vrtovec et al. [VLP05] are slow (minutes to hours [ŠLPV10]).

In the following, we use the coordinate system induced by the acquisition of sagittal MR images, i.e. with respect to the human body, x is front-to-back, y is top-to-bottom and z left-to-right.

4.1. Viola-Jones Detector

The training data for the (single) Viola-Jones detector is derived from the reference segmentations utilizing all training datasets jointly. A bounding box is determined for each vertebral body on each slice of all datasets used for training. Expanding this bounding box to a square serves as a positive sample for detector training. All positive samples are resized to 16×16 pixels and have 10 additional, slightly rotated versions ($\pm 15^\circ$). The Viola-Jones detector has a built-in scaling, and we targeted it to adults (20-50 mm squares).

Left, right and bottom rectangles of each slice, which are not covered by positive samples, are added to negative samples (Fig. 2), along with all slices which contained no part of reference vertebrae. Additionally some hand-chosen rectangles very near the vertebral bodies are added to negative samples, otherwise some nearby structures such as aorta or spinous processes get mistaken for vertebral bodies.

A boosted cascade with 40 stages, each stage with a minimum hit rate of 0.998 and target maximum false alarm rate

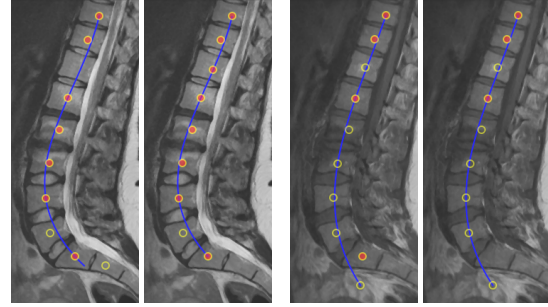


Figure 3: Vertebrae detection (red-filled points are candidates on the current slice, empty points are lying on other slices, the fitting curve is shown in blue; see Tab. 2 for information about the datasets): Left, the unfiltered center candidates of dataset AKs5 and, middle left, a fully successful filtering (the T12 gap was filled and the unwanted S2 vertebra detection was removed). Unfiltered center candidates of dataset AKs6 (middle right) with wrong result – 1 false negative and 1 false positive after filtering (right). In 38% (10/26) of datasets no user intervention is required.

of 0.5 (limited by maximum stage size of 100 tree stumps) has been applied. The number of positive and negative samples used was 7500 each. We use the OpenCV implementation of the Viola-Jones detector [VJ01].

The initial vertebra detection based on the Viola-Jones detector [VJ01] is done on all sagittal slices independently. If a dataset is not acquired as sagittal slices, it is reformatted into sagittal slices. The candidates' in-slice positions provide the (x, y) coordinates and approximate size (vertebral body diameter) of the vertebral body candidates, the z -coordinate of the center is derived from weighted average of slice-adjacent candidate centers.

Note, that the result of the Viola-Jones detector is imperfect, containing spurious detections (false positives) and missing vertebral bodies (false negatives); see Fig. 3.

4.2. Detection Filtering

In order to remove wrong detections and to fill missed vertebrae, a spine center curve is fitted to all detected vertebra candidates in 3D (see Fig. 4). The S-shaped lower spine has two distinct curvatures and therefore a third order polynomial is an appropriate curve model. At this stage misdetections along the x -axis are more important than precise fitting along the z -axis. Since we need robustness to false positives we sacrifice precise fitting to the scoliotic-shaped spine which requires a fourth order polynomial.

The distances of all vertebra candidate centers to the spine center curve are calculated and the one with the largest distance is removed, if it is above the approximate radius of that vertebral body candidate. Then a new spine center curve is

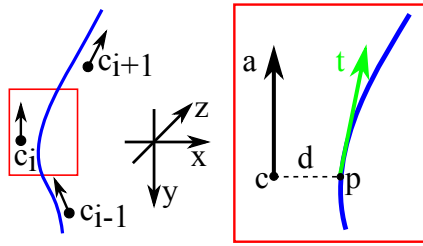


Figure 4: Spine center curve fitting. Left: vertebra centers, axes, and the fitting curve. Inset on the right: vertebra i shown in detail (index i omitted). We use y axis as abscissa and two polynomials dependent on it. We minimize the distances d_i between vertebra candidate centers c_i and their projections onto the abscissa p_i using L_1 -norm.

fitted and the procedure is repeated for as long as there are candidates to be removed.

The topmost and bottommost vertebra detections are frequently fitted well by the polynomial, even if they are unwanted false positives. Therefore, we remove them in case they are located far from the closest vertebra.

Furthermore, the frequency of candidates along the fitted spine center curve is examined. The Theil-Sen linear estimator [The92] is used to determine robustly the linear function of expected distance between vertebral body centers – the distances increase towards the lower end of the vertebral column. If a particular distance is below 75% of the expectation from this fitting, meaning that an extraneous vertebra is detected there, the corresponding candidate is removed. Similarly, gaps larger than 150% of expected distance are filled with the best-fitting number of evenly distributed new centers. Still missing vertebrae, e.g. the top- or bottommost, need to be added manually in the next step.

4.3. Manual Correction and Labeling

Since a correct setting of the initial vertebral body centers is crucial for the following segmentation and diagnosis, the result of the filtered detections is presented to the user for verification. The user can reposition (mouse drag), add (left click) and remove points (right click). This is done on sagittal cross-sections, even if the dataset was acquired along some other body axis (e.g. coronal). The user chooses cross-section with a slider, and middle slice of the detected vertebral body centers is preselected by the program.

New vertebral body centers should be placed on a slice which is approximately central to the vertebral body. The size of these new vertebral bodies is linearly interpolated from the sizes of neighboring vertebral bodies. The user also chooses a label for the bottommost vertebra (usually S1) from which other vertebra labels are calculated, in order to have a correct human-readable diagnosis result.

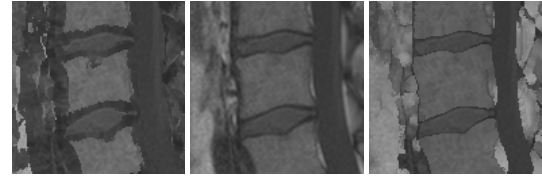


Figure 5: Left: low image. Middle: original slice. Right: high image. The difference between low and high values indicates proximity to a boundary.

5. Segmentation of Vertebral Bodies

The segmentation of the vertebral bodies consists of a multiple features based boundary model, which is used to steer the segmentation of the individual vertebral bodies.

5.1. Feature-Based Boundary Classification

The vertebral body boundaries are estimated using multiple features, which are classified into probabilities of the voxel v being at a boundary, and combined to a final probability $p(v)$ using the weighted average (see Fig. 6). Using multiple classifiers and then combining them is known to improve results and robustness [KHDM98]. We apply three edge-based and two intensity-based features. As intensity-based features require local, per-vertebra intensity statistics, we examine a circle with 2.8 mm radius around each of the detected vertebral body centers. After removing outliers (intensities appearing only once), minimum and maximum are used for further processing.

Due to the per-vertebra statistics, the boundary probability map needs to be calculated for each vertebra separately. This is done inside a bounding box which is twice as large as the detected diameter of the vertebra.

The edge features are based on LH (low-high) values [ŠBSG06] (Fig. 5), Canny edges, and thresholded gradient magnitudes. The difference between low and high values, L, H indicates proximity to a boundary. The boundary probability is deduced from these values and the current voxel v intensity I by $p_{LH}(v) = ((H - I) - (I - L)) / (I_{max} - I_{min})$, where I_{max} and I_{min} are the maximum and minimum intensity. Note that we shift intensities so they start at 0 if it is not the case, before any other processing is done.

Gradient magnitudes and Canny edges are multiplicatively enhanced using the structure tensor as described by Fernández and Li [FL03] to improve the detection of 2-manifold edges and suppress one-dimensional features. For both features, the respective probability is derived by applying a distance field (DF) and a linear transform with clipping (LTwC), see Fig. 6.

The two intensity-based features directly work on the thresholded MRI intensities. The thresholds are taken from the intensity statistics taken in the proximity of the detected

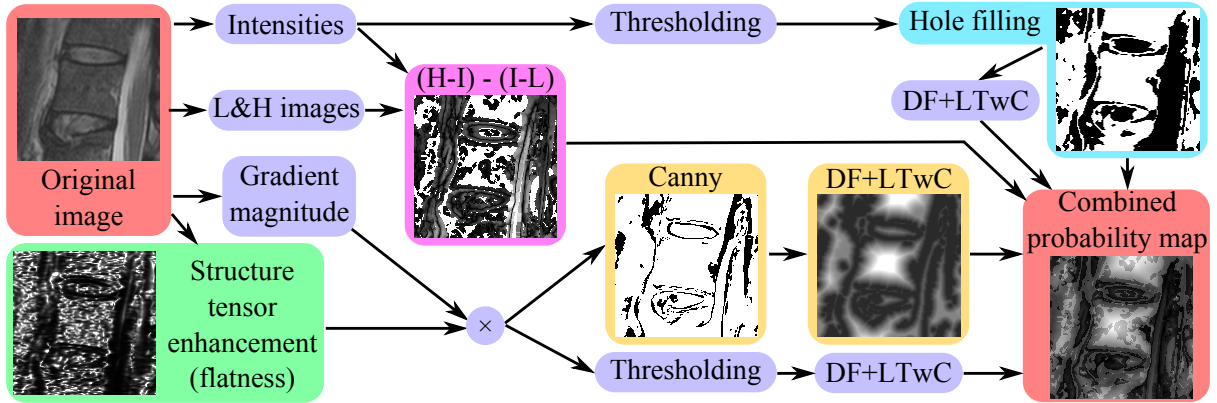


Figure 6: Features explained on T12 vertebra from F02 dataset in Tab. 2. In all images except original black indicates edges, whereas white represents more homogenous regions like a vertebral body's interior.

vertebral body centers, i.e. the minimum and maximum values after outlier removal. Holes are filled by morphological closing using a 3-voxel-diameter ball structuring element.

We use the binarized intensities (each voxel either 0% or 100%) as one feature. The second feature is the distance field constructed on this binary image, effectively treating 0→1 transition as another edge feature (therefore using DF+LTwC transformation). We thus incorporate both sharp edges (binary 0/1) and a smooth edge approach (normalized distance field) for mesh inflation. After combining the probabilities, true edges usually end up having boundary probability around 90%.

The distance field constructed on binarized intensities is the most informative feature. After parameter optimization (Sec. 7) it gets a weight of around 30-35% while the other four features have around 10-20%. Some other features, such as σ -weighted deviation from mean initialized intensity, have been investigated and discarded due to small benefit because of redundancy.

5.2. Segmenting Individual Vertebral Bodies

Each vertebral body is segmented using an iterative inflation algorithm. The algorithm starts with a small triangular subdivision surface mesh at the detected vertebral body center. Its orientation is approximated by relative positions of the neighboring centers. This mesh is enlarged using balloon inflation forces [TF88, Coh91], constrained by smoothness, steered towards a star-convex geometry and approximate size.

Smoothness is enforced by using a constrained subdivision surface scheme (see Sec. 5.2.2). Star-convexity is stimulated by allowing the inflation only along center-vertex direction. This deters self-intersections of the inflated surface. There is also a size-goal force, which drives the segmentation towards the estimated size detected by the Viola-Jones method (see Sec. 5.2.1).

The following steps are performed iteratively:

1. Move the mesh vertices (outwards) towards the boundary with no regard to subdivision rules.
2. Normalize the subdivision hierarchy by moving vertices so that they comply with subdivision rules.
The subdivision hierarchy is normalized using a combination of least squares optimal fitting and a heuristic rule which includes smoothing as a side-effect. Smoothing is required to overcome noisy voxels.
3. Recalculate the polyhedron center from the polyhedron shape to account for different inflation speed in different directions.
4. Stop when convergence is detected.

5.2.1. Inflating the Mesh

We do the inflation by examining each vertex separately. Depending on the boundary probability and the size-goal force, we either inflate or deflate the vertex along the radial (center-vertex) direction. The step size is equal to the minimum voxel spacing, i.e. 0.5–1.1 mm for our datasets.

Initially, a vertex v is in the interior and will inflate as long as the boundary probability $p(v) < 0.5 + f_{sg}$, where f_{sg} is the size-goal force. When $p(v) \geq 0.5 + f_{sg}$, the sign of the probability derivative (adjusted by the size-goal force) $\frac{\delta(p+f_{sg})}{\delta \vec{r}}$ in inflation direction \vec{r} decides on whether to inflate or to deflate, thereby moving the vertex towards the maximum boundary probability. By using the boundary probability gradient only when the probability is high (vertex near the boundary), noise inside the vertebral body does not block the inflation. Convergence is achieved when the average center-surface distance (“radius”) stops increasing.

$$f_{sg} = \kappa \cdot \tanh\left(\frac{r_{detected} - r_i}{r_{detected}}\right) \cdot \frac{s_{min} - (r_i - r_{i-1})}{s_{min}}$$

is the size-goal force designed to overcome local minima which sometimes occur at smaller sizes than anatomically

possible. $r_{detected}$ is the approximate radius detected by the Viola-Jones algorithm – the goal radius, r_i is the radius in the current iteration and s_{min} is the minimum voxel spacing. κ (≈ 0.1) is a parameter steering the force influence and is optimized for segmentation accuracy (see Sec. 7). f_{sg} only gives preference to a certain size, it does not guarantee it.

When $r_i < r_{detected}$, f_{sg} increases inflation pressure, and when $r_i > r_{detected}$, f_{sg} creates deflation pressure. Out of precaution this factor is limited to $(-1,1)$ range by the \tanh function.

The last factor is there to suppress the size-goal force during normal inflation. When the radius growth between the last two iterations is large ($r_i - r_{i-1}$ close to s_{min}), this factor lowers the influence of f_{sg} . This is done in order to respect the edge probabilities, and to give influence to f_{sg} only when the inflation enters a local minimum (small $r_i - r_{i-1}$) which is far from the desired size. This factor is always positive, because $r_i - r_{i-1}$ cannot be bigger than s_{min} which is the maximum possible step size.

If the initial center estimate for the segmentation is very close to a boundary, the surface will inflate much more in the opposite direction. This will result in a highly uneven distribution of vertices over the surface, which is detected using the standard deviation of the edge lengths ($\sigma > 1$ mm). In this case, the segmentation is restarted at the current center; this sometimes occurs with S1 vertebra or with some thoracic vertebrae with low contrast, when the segmentation leaks outside of the true vertebral body.

5.2.2. Constrained Subdivision Hierarchy

The butterfly algorithm is the simplest interpolating subdivision scheme working on triangle meshes, i.e. a subdivision scheme that retains vertices of coarser levels in the refinement. The modified butterfly was presented by Zorin et al. [ZSS96], and it avoids problems with irregular vertices, i.e. vertices with valence $\neq 6$.

Our base mesh consists of a closed triangular polyhedron with 32 vertices and 60 triangles. We create new mesh levels through the subdivision rules until the average edge length l is comparable to the voxel size s : $s \leq l < 2s$, where $s = \sqrt[3]{s_x s_y s_z}$ is the geometric mean of the voxel spacings.

The inflation step repositions the vertices with no regard to the subdivision rules. Thus, we need to enforce the subdivision hierarchy for all the dependent levels afterwards. We use a global least squares optimization (Jacobi SVD), minimizing the vertex position correction and edge length deviation. The latter stimulates an even vertex distribution.

The number of control levels (number of levels, the vertices of which are independent) is defined by how many levels exist overall: $n_{control} = \lfloor \frac{n_{total}-1}{2} \rfloor$.

If the control levels contain many vertices, overshooting effects start to appear, which is frequent in interpolat-

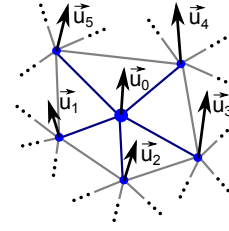


Figure 7: Heuristic hierarchy normalization adjusts positions of coarser level vertices, and then calculates positions of dependent vertices using the subdivision rules. u_i – position updates from inflation step. u_0 – control vertex. u_1 – u_5 – directly dependent vertices.

ing schemes. This can be reduced by using a low or moderate number of vertices in the base mesh, or by applying slight smoothing after each iteration, thus fighting noise and small ambiguities in the data. Thus, we combine the least squares fitting scheme with the heuristic hierarchy normalization which implicitly smoothes the mesh. The influence of the heuristic scheme is controlled by a weight parameter which is part of the optimization (see Sec. 7).

The heuristic hierarchy normalization is based on a local rule, which adjusts the positions of the vertices in coarser levels until the base level has been reached. This adjustment is based on the position update which has been applied to each vertex during the inflation step. Basically, coarser-level update vectors are averages of finer-level update vectors, influenced by the modified butterfly subdivision rules. The averaging is formulated in a way to approximate the (non existing) inverse of the subdivision rule.

Position update vectors are propagated from finer to coarser levels as follows (also see Fig. 7):

$$\vec{u}_{adjusted} = w \cdot \vec{u}_0 + \frac{1-w}{n} \sum_{i=1}^n \vec{u}_i, \quad w = \frac{1}{2 \cdot 3^\lambda},$$

where λ is the level above the finest level. The weight w is related to the number of vertices which have contributed to the adjustment of directly dependent vertices; it decreases exponentially. 3 is chosen as the base because in the regular case (valence 6), the number of vertices increases 3-fold with each level.

6. Diagnosis

We focus on three diseases which can be diagnosed from segmentation of vertebral bodies alone, i.e. on scoliosis, spondylolisthesis and vertebral fractures.

6.1. Scoliosis

Normally, the vertebral column has curves in front-back (sagittal) plane, having a shape of two concatenated letters “S”. In the left-right direction (coronal plane), healthy spine

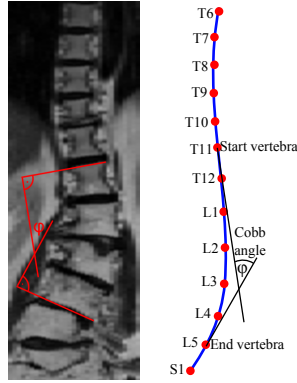


Figure 8: Scoliosis is visible on coronal slices, looking at a person from front or back. On the left is a real image (dataset C002) with manual measurement lines. On the right is how we calculate scoliosis using the fitted centerline and its tangents at each vertebral body center. In this case: $\varphi = 32.6^\circ$.

is generally straight with only minor asymmetry – similar to facial asymmetries. If there is a curve in the coronal projection of the spine, that condition is known as scoliosis.

According to standard clinical practice the severity is determined by spinal curvature measured in degrees, called Cobb angle [Aeb05]. Fig. 8 (left) depicts how the Cobb angle φ is measured. If the Cobb angle is above 20° tracking is warranted, and above 30° treatment is considered.

For determining the degree of scoliosis, we fit the spine center curve (fourth-order polynomial) to segmented vertebral body centers (see Fig. 4). We evaluate the Cobb angle [Cob48] between all possible pairs of vertebrae, using tangents of the fitted curve taken at vertebra center positions. We report the biggest Cobb angle found along with the vertebrae between which it is occurring.

6.2. Spondylolisthesis

Spondylolisthesis is misalignment of vertebrae which results in change of posture and can include pain. It is usually divided into low-grade which does not usually require treatment, and high-grade (above 25% [Sar87]).

Measurement of the degree of spondylolisthesis in clinical practice $S\% = \frac{m}{d}$ is adapted for manual measurement on a single sagittal slice or 2D X-ray (Fig. 9 top-left) [WW83].

We chose to measure the degree of spondylolisthesis by calculating the distance of vertebral body centers to the fitted spine center curve and dividing it by vertebral body diameter: $S\% = \frac{m_1 + m_2}{d}$ (Fig. 9 right). This measure is more robust and stable with respect to the segmentation imperfections and does not require the analysis of the shape of each vertebral body so it is also computationally efficient.

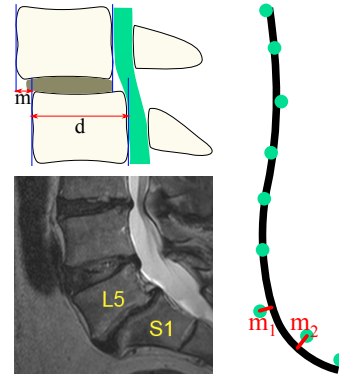


Figure 9: Spondylolisthesis schematic on a sagittal slice. The usual medical calculation method (top left) is not always easy to apply, especially for L5/S1 pair (bottom left, dataset S01). Right: the way we calculate misalignment ($m_1 + m_2$) using the fitted centerline and vertebral body centers (L5/S1 misalignment=14%).

6.3. Vertebral Fracture

Crushed vertebral bodies are usually due to mechanical stress applied to a weakened bone (typically a result of osteoporosis). In clinical practice diagnosis is usually established based on qualitative impression [LRDG04]. Several different approaches have been used in medical research [LRDG04], but they are tailored for manual measurements based on 2D sagittal X-rays or single sagittal slices [WLJG95].

To diagnose crushed vertebrae we fit a third-order polynomial to vertebral body volumes using the L_1 -norm. That way outliers do not influence the fitting, so the detection is robust. Besides the actual volume of each vertebral body we now also have a volume expectation $V(i) = P(i)$, by evaluating the fitted polynomial at each vertebra index. Analyzing the difference between them results in diagnosis. If the actual segmented volume is below 80% of the polynomial-derived volume expectation, the vertebra is indicated as crushed.

The 80% of expected volume parameter (20% crushed) was obtained empirically. Crush-measure routinely reaches 10% in non-crushed (fracture-less) datasets. Average crushing over all our datasets is 2.46% with standard deviation $\sigma = 3.87\%$. We chose $5\sigma \approx 20\%$ as the threshold.

7. Parameter Optimization

The Viola-Jones detector clearly relies on training data. The rest of our segmentation system is not dependent on any kind of training. However, since the whole system has some $N = 16$ parameters p_i , such as thresholds or weights (see Tab. 1), which we do not want to expose to the user, we determine their values via an optimization process.

We define the objective function via the Dice Similarity

Table 1: *Optimized parameters.*

| Parameter Type | Used in/for |
|---------------------------------|---------------------|
| 4 image filtering radii | Feature calculation |
| 4 image filtering thresholds | Feature calculation |
| 4 feature weights | Boundary estimation |
| 4 constraint importance factors | Inflation |

Coefficient: $f(p_1, \dots, p_N) = \sum_{x \in T} DSC(x; p_1, \dots, p_N)$ where T is the set of datasets used for training.

A single evaluation of the function consists of segmenting all the training datasets with the given parameter values and returning the sum of all DSC values. We use the local, derivative-free COBYLA algorithm [Pow94] to maximize the function f . This optimization increases the average DSC by $\approx 1\%$ and reduces distance error by $\approx 5\%$. The increase comes primarily from datasets with T1 and TIRM sequences. T2 datasets had a mix of increases and decreases.

8. Results

8.1. Evaluation Setup

We tested our method on 22 pathological datasets and 4 datasets from healthy volunteers for a total of 234 vertebrae. 13 are female and 13 male patients. 17 of these datasets are publicly released together with this submission. Data came from 7 different hospitals and 9 scanning stations (2 hospitals had 2 each). Reference segmentations were produced by manually tracing the vertebral body edges in the primary acquisition plane (23 sagittal, 2 axial, 1 coronal). 10 of the reference segmentations were done by neurosurgeons, 14 by an experienced user (the primary author) and 2 by an experienced user under neurosurgeon supervision. The important property of these datasets is the high anisotropy of voxel size between $2.7\times$ and $8.2\times$ (see Tab. 2 and Fig. 11). Manual segmentation time was 3-6 minutes per vertebra.

9 of those datasets, with 81 vertebrae, were used for training the Viola-Jones detector and later to optimize the parameters. They were chosen from the datasets we could publicly release. We wanted to have in both, the training set and the test set, varied MRI sequences, anisotropy factors and at least one each of the three diagnosable pathologies. Other than that, the choice was arbitrary. From Tab. 2 and Figs. 12, 13 and 14 (AvgTrain and AvgTest) it can be seen that both of these sets have very similar characteristics and results.

The variety of datasets the Viola-Jones detector can handle depends on its training set. Ours has a lot of variety – T1, T2 and TIRM sequences, and a range of TE, TR and other parameters. But if new MRI sequences need to be handled with low amount of errors, the detector needs to be re-trained with the expanded training set.

To check segmentation expertise of the primary author, one dataset (F02) was manually segmented both by the primary author and a neurosurgeon (the second author). Dice

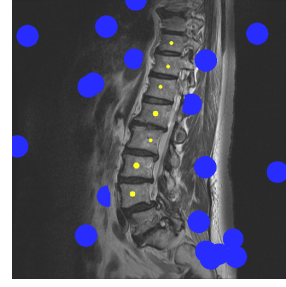


Figure 10: *Power Watersheds initialization. One cross-section of seeds overlaid on image (dataset Ble). Blue are background seeds, yellow are seeds of vertebral bodies. It is noticeable on the vertebral body seeds that they have different Z-positions (they appear to be of different sizes on this cross-section).*

coefficient is 91%, which is on par with segmentation correlations between neurosurgeons [EBK* 11] (86-96% DSC).

Testing was done on a machine with Intel Core i7-920 2.67GHz quad-core processor. The average dataset processing time is about 75 seconds: several seconds for the initialization and median filtering (between selecting the dataset and manual initialization correction) and the rest for feature calculation and segmentation. The times varied from 20 seconds (AKa3) to 400 seconds (case_2). The execution time grows approximately linearly with the number of voxels.

Besides comparison to methods developed specifically for spine MRIs, we also subjected all our datasets to the power watersheds of Couprie et al. [CGNT09] as one representative of a general segmentation method, taking the publicly available source code from <http://powerwatershed.sourceforge.net/>. The execution time of power watersheds on a typical dataset ($512 \times 512 \times 16$) is about 20 seconds. The manual initialization of this method in 3D is quite painstaking, so we opted for automatic seed creation, derived from ground truth data, for a thorough comparison.

One seed was placed into each vertebral body on a random position near the center, and twice as many seeds into the background (Fig. 10). Since background seeds were large (20 voxel radius), they were clipped by a safety region of interest around vertebral bodies.

Power watersheds crashed on a few datasets (with ≥ 10 million voxels) due to hitting 2GB memory limit (Win32). Our old method [ZVD* 12] also crashed on Neubert's case_* datasets. We measured surface distance errors using the Metro mesh comparison tool [CRS98], which incorporates an approximate mean distance error between two meshes.

The overview of results is given in Figs. 12, 13 and 14. Datasets marked with (*) and written in italic font are used for training the detector and optimizing parameters, the rest only for testing. **AvgTest** is average of datasets used for testing, and these numbers are **reported** through-

Table 2: Information about the datasets used for the quantitative evaluation. *Seq.* – sequence, $S_{x,y}$ – voxel spacing in x - and y -directions (millimeters), S_z – spacing along z -axis. xyR – resolution of image along x - and y -axes, zR – z -resolution. *AF* – anisotropy factor $\frac{S_z}{S_{x,y}}$. *Path* – pathologies (*Sco*–scoliosis, *VF*–vertebra fracture, *SL*–spondylolisthesis, +–other pathologies not diagnosable from vertebra segmentation). N_{VB} – number of vertebral bodies in a dataset. *SB* – segmented by (*NS*–neurosurgeon, *EU*–experienced user). *MVox* – millions of voxels. A_{iso} – spacing of isotropic image with equivalent voxel volume. Datasets after the thick horizontal line were also used in our previous work [ZVD*12].

| Dataset | MRI Seq. | Sex | Age | $S_{x,y}$ | S_z | xyR | zR | AF | Path. | N_{VB} | SB | MVox | A_{iso} |
|----------------|----------|-----|-----------|-------------|-------------|------------|-------------|-------------|-------|------------|----|--------------|-------------|
| AKa2 | T2 frFSE | F | 21 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKa3 (*) | T1 FSE | F | 21 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKa4 | TIRM | F | 21 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKs3 | T2 frFSE | F | 22 | 0.7 | 4 | 512 | 25 | 5.69 | + | 10 | EU | 6.55 | 1.26 |
| AKs5 (*) | T2 frFSE | F | 22 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKs6 | T1 FSE | F | 22 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKs7 (*) | TIRM | F | 22 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| AKs8 (*) | T1 FSE | F | 22 | 0.7 | 4 | 512 | 15 | 5.69 | + | 9 | EU | 3.93 | 1.26 |
| DzZ_T2 | T2 TSE | M | 27 | 0.55 | 4.4 | 640 | 12 | 8.05 | None | 9 | EU | 4.92 | 1.1 |
| F04 | T2 TSE | F | 69 | 1.12 | 3 | 448 | 23 | 2.69 | VF+ | 17 | EU | 4.62 | 1.55 |
| Hoe | T2 frFSE | M | 58 | 0.63 | 4.4 | 512 | 14 | 7.04 | + | 8 | EU | 3.67 | 1.2 |
| Lan | T2 TSE | M | 79 | 0.78 | 4.4 | 384 | 13 | 5.63 | + | 9 | EU | 1.92 | 1.39 |
| LanII | T1 TSE | M | 79 | 0.78 | 4.4 | 384 | 13 | 5.63 | + | 9 | EU | 1.92 | 1.39 |
| case_2 | T2 SPACE | M | 40 | 0.34 | 1.2 | 641 | 296 | 3.49 | None | 9 | AN | 121.62 | 0.52 |
| case_10 | T2 SPACE | F | 47 | 0.34 | 1 | 636 | 299 | 2.91 | None | 8 | AN | 120.94 | 0.49 |
| F02 (*) | T2 SE | M | 51 | 0.5 | 3.85 | 768 | 18 | 7.7 | VF,SL | 9 | NS | 10.62 | 0.99 |
| StI (*) | T2 SE | M | 71 | 0.5 | 3.85 | 704 | 20 | 7.7 | St | 8 | NS | 9.91 | 0.99 |
| Ble | T2 frFSE | F | 64 | 0.63 | 4.4 | 512 | 14 | 7.04 | + | 10 | NS | 3.67 | 1.2 |
| C002 (*) | T2 TSE | F | 74 | 1.12 | 3.3 | 448 | 31 | 2.96 | Sco | 13 | NS | 6.22 | 1.6 |
| DzZ_T1 | T1 TSE | M | 27 | 0.68 | 4.4 | 512 | 12 | 6.44 | None | 9 | EU | 3.15 | 1.27 |
| F03 | T2 TSE | M | 72 | 1.19 | 3.3 | 320 | 25 | 2.77 | VF | 7 | NS | 2.56 | 1.67 |
| Geh | T2 frFSE | M | 25 | 0.63 | 4.4 | 512 | 10 | 7.04 | + | 7 | NS | 2.62 | 1.2 |
| LC | T2 SE | M | 47 | 0.73 | 4.4 | 384 | 14 | 6.03 | + | 7 | NS | 2.06 | 1.33 |
| S01 (*) | T2 SE | M | 65 | 0.47 | 3.85 | 640 | 16 | 8.19 | SL | 7 | NS | 6.55 | 0.95 |
| S02 (*) | T2 SE | F | 55 | 0.47 | 3.85 | 640 | 16 | 8.19 | SL | 8 | NS | 6.55 | 0.95 |
| Sch | T2 frFSE | M | 42 | 0.63 | 4.4 | 512 | 16 | 7.04 | SL+ | 7 | NS | 4.19 | 1.2 |
| Average | | | 45 | 0.68 | 3.8 | 527 | 38 | 5.85 | | 9.0 | | 13.53 | 1.19 |
| StdDev | | | 22 | 0.21 | 0.88 | 104 | 76.5 | 1.69 | | 2.1 | | 31.79 | 0.27 |

out the paper. AvgTrain is average of datasets used for training and AvgZVD is average of datasets used in our previous work [ZVD*12].

8.2. Vertebra Detection Accuracy

The vertebra detection rate of our approach is 92.9% – in other words 7.1% of vertebral bodies are not detected. Spurious detections are rare – false positive rate is 1.3%. This is a vast improvement over our old method [ZVD*12], which missed 21.2% vertebrae and falsely detected 2.5% on the same test-set.

Detection of upper thoracic vertebrae was deficient, because only one of the datasets used for detector training had vertebrae above T10 (dataset C002 also had T6-T9). Most false positives occurred for the S1 vertebra and the upper thoracic vertebrae, which is influenced by the low contrast

for the upper thoracic region. The S1 vertebra simply has a significantly different geometry than lumbar and thoracic vertebrae, which is not taken into account by our system.

8.3. Segmentation Accuracy

The mean distance of segmentation from the reference surface is $1.76\text{mm} \pm 0.38\text{mm}$. In our case of anisotropic voxels, a relative measure is deduced by dividing the distance by the edge-length of the isotropic voxel $A_{iso} = \sqrt[3]{S_x S_y S_z}$. The average A_{iso} for these 26 datasets is $1.19\text{mm} \pm 0.27\text{mm}$, and relative distance is 1.59 ± 0.79 .

The average DSC of 79.3% is close to Davatzikos et al. with 81%, but they used healthy individuals. Healthy datasets are easier to segment and achieve better DSC. The datasets with more severe pathologies (e.g. crushed vertebra) have lower DSC than less noticeable pathologies (e.g. steno-

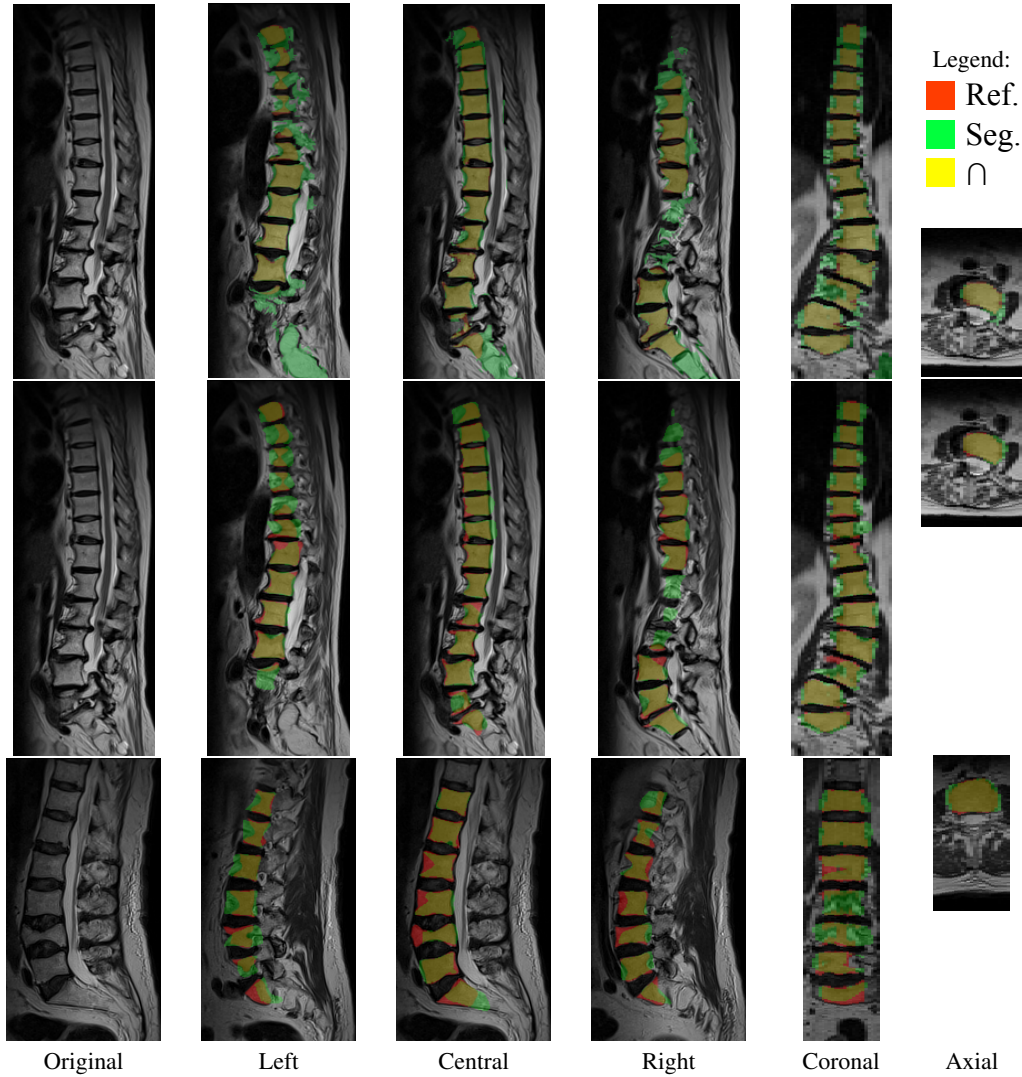


Figure 11: Overlay of reference segmentation and the one produced by power watersheds (top row) and our method (bottom two rows). Images were cropped to save space (unimportant parts were cut off). See Tab. 2 for dataset properties. Top and middle: dataset C002. Bottom: F03. Dice coefficients are 65%, 75% and 76%.

sis). Pathologies obviously and naturally reduce the detection rate. Using the relative distances error of 1.59 ± 0.79 , our method can be compared to Štern et al., who report an error of $1.85\text{mm} \pm 0.47\text{mm}$ on a 1mm^3 grid. When taking into account only the mid-sagittal slice, our method achieves $\approx 86\%$ DSC, which is similar to Ayed et al. at 85%. Hoard and Martel’s surgery-oriented method with thorough initialization and manual correction of segmentation at $1.25\text{mm} \pm 0.28\text{mm}$ mean distance error remains the most precise.

Neubert kindly provided two datasets with manual segmentations (case_2 and case_10), so we could directly compare our method to theirs [NFS*11, NFE*12].

In order to have a fair comparison to Neubert et

al. [NFE*12] their distance results should be converted into relative distances. A_{iso} of 0.5mm transforms distance error of 0.67mm into a relative distance error of 1.34. This is better than ours (1.59). Neubert et al. also have higher DSC (91% vs our 78%). However this high quality comes at a cost of average execution time per dataset of 5.5 hours.

When they ran their method on datasets of higher anisotropy, their DSC was two percentage points lower: 83% vs 85% (see [NFS*11], dataset group I vs II). If less time is allotted to the iterative optimization, DSC is lower: ten-fold reduction in execution time (5.5h \rightarrow 35min) lowers DSC from 91% [NFE*12] to 85% [NFS*11].

As our method was tailored for lower resolution routine

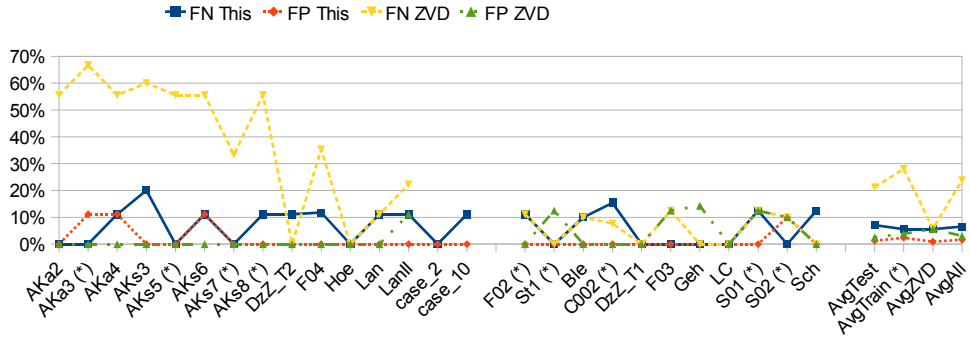


Figure 12: Vertebra detection results (False Positives and False Negatives) for this and our previous work [ZVD* 12]. The new datasets (AKa2-case_10) are more challenging.

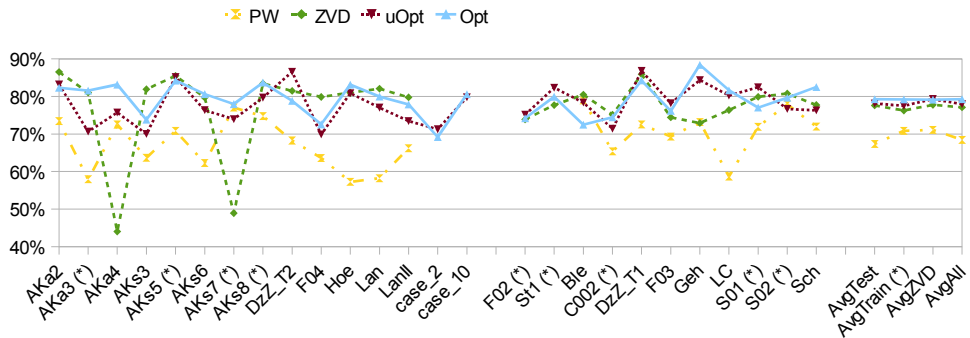


Figure 13: Dice scores for Power Watersheds (PW), our previous work [ZVD* 12], and current paper before (uOpt) and after parameter optimization (Opt).

datasets, it had low precision on Neubert's high-resolution datasets, case_2: 2.29mm and case_10: 1.09mm. Unfortunately, cross-validation with the method from Neubert et al. [NFS* 11, NFE* 12] was not possible.

Our method is vastly superior to power watersheds for the purpose of vertebral body segmentation in MRI. Both methods have similar execution time, but ours has higher DSC (67% vs 79%), lower Hausdorff distance (12mm vs 36mm) and half an order of magnitude lower relative distance error (1.59 vs. 6.85).

In comparison with our previous work [ZVD* 12] on old datasets (AvgZVD in Figs. 12,13,14), we now have slightly higher DSC (79.2% vs. 77.8%) and noticeably better mean distance error: 1.76 mm vs. 2.17 mm and relative distance 1.50 vs. 1.83. On new datasets the old method mostly fails to segment the pathological vertebrae, and as it uses the detection by segmentation approach (Fig. 12) it means that the hardest-to-segment vertebral bodies are not taken into account for Dice coefficients (Fig. 13) and distance errors (Fig. 14), data points AvgTest, AvgTrain and AvgAll.

Due to parallelization, execution time is about the same in spite of additional processing (Viola-Jones detector, heuristic subdivision hierarchy normalization, size-goal force).

Moreover, the user can now intervene in case of mis-detection of vertebral bodies at the very beginning of the processing by adding or removing vertebra center initialization points. This ensures that all the vertebral bodies are segmented, which is a significant contributor to robustness.

The majority of discrepancies between manual and automatic segmentations stem from lateral slices and upper thoracic vertebrae (see Fig. 11). Lateral slices make it harder to algorithmically discern a boundary due to significant partial volume effects. Power watersheds mostly fail on lateral edges of vertebral bodies, too (Fig. 11).

8.4. Diagnosis Accuracy

Our datasets had no examples of high-grade vertebral fracture. Two instances of high-grade spondylolisthesis were classified correctly, as was the only high-grade scoliosis. There were 6 false alarms for severe crushed vertebra and 9 for severe spondylolisthesis, or 2.6% and 3.9% respectively. There were no false alarms for scoliosis.

Overall, our system errs on the side of caution, falsely drawing attention to non-diseased vertebrae (5.5%) and never missing a diseased one, leading to a conservative and robust behavior.

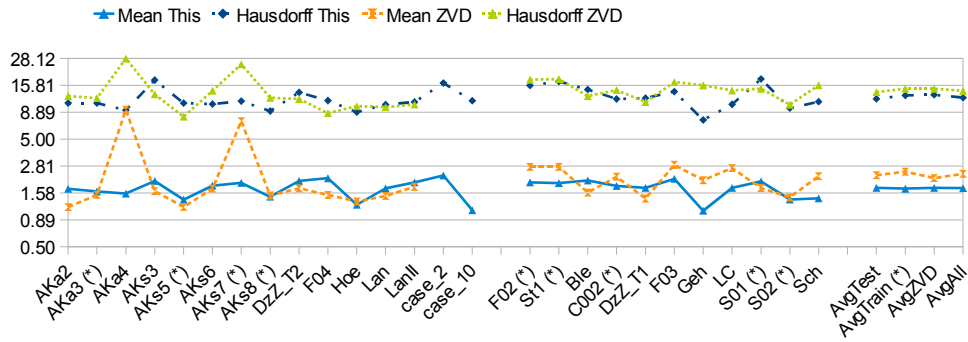


Figure 14: Distance errors (in millimeters) for our previous work [ZVD*12] and this paper. Mean – mean distance of points in the automatic segmentation mesh from the reference mesh. Hausdorff – Hausdorff distance between segmented and reference meshes. Scale is logarithmic.

Diagnostic measurement errors for crushed percent, spondylolisthesis percent and Cobb angle are 4.3 percentage points, 4.6 percentage points and 3° , respectively. Relative errors (55%, 52% and 31%) are quite large because the ground truth measurements are low (2.5%, 4% and 9.2°).

This proves once again that high reliability of diagnosis is not easy to achieve. Therefore the diagnosis results established using our system can only be used as a help to the medical professionals, and cannot replace their diagnosis.

9. Conclusion

We presented a novel method for spine segmentation in routine MRI. Our segmentation is fast and robust with respect to the low and anisotropic resolution of routine MRI datasets and to pathological spine and vertebra shapes. Our vertebral body segmentation method is inflation-based and incorporates a novel constrained subdivision surface approach for smoothness control. A multiple features boundary estimator along with size-goal force provide robustness. Automatic vertebra center detection reduces initialization time and allows parallel segmentation. The segmentation takes about 70 seconds for the whole dataset.

Our method was tested on a larger set of 234 vertebral bodies (199 from pathological, 35 from healthy datasets) than prior work (Hoad and Martel: 30 vertebrae, Štern et al.: 75 vertebral bodies; Ayed et al.: 75 vertebral bodies, Neubert et al.: 132 vertebral bodies from healthy volunteers, Davatzikos et al.: 84 vertebral bodies from healthy volunteers). Most importantly our method works on a large variety of datasets, whereas others restrict themselves to just one type of datasets: Štern et al. – T_2 weighted, Davatzikos et al. – T_1 weighted, Ayed et al. – T_2 weighted, Neubert et al. [NFE*12] – T_2 SPACE sequence, and Hoad and Martel – T_1 FISP.

Open science

Together with this submission we make our program code publicly available under an open source license, including

17 anonymized datasets with corresponding segmentations. <http://www.cg.informatik.uni-siegen.de/en/spine-segmentation-and-analysis>
A written consent was collected for every MRI study. Please cite this paper if you use any of these in your work.

Acknowledgments

We are grateful to Aleš Neubert from The University of Queensland, Brisbane, Australia for providing datasets case_2 and case_10 and their manual segmentations. He received financial support from Australian Research Council's Linkage Projects funding scheme LP100200422. We are also grateful to Thomas Dukatz and Miriam Bauer from University Hospital Marburg for some of the manual segmentations. Aleš Vlasák and Daniel Hořínek were supported by European Regional Development Fund – Project FNUSA-ICRC (No. CZ.1.05/1.1.00/02.0123).

References

- [Aeb05] AEBI M.: The adult scoliosis. *European Spine Journal* 14, 10 (2005), 925–948. 8
- [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 *MICCAI* (2012), Springer, pp. 520–527. 2, 3
- [CGBM04] CARBALLIDO-GAMIO J., BELONGIE S., MAJUMDAR S.: Normalized cuts in 3-D for spinal MRI segmentation. *IEEE Trans. Medical Imaging* 23, 1 (2004), 36–44. 2
- [CGNT09] COUPRIE C., GRADY L. J., NAJMAN L., TALBOT H.: Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In *Proceedings of ICCV* (2009), pp. 731–738. 9
- [Cob48] COBB J.: Outline for the study of scoliosis. *Instructional Course Lectures* 5 (1948), 261–275. 8
- [Coh91] COHEN L. D.: On active contour models and balloons. *CVGIP: Image Understanding* 53, 2 (1991), 211–218. 6
- [CRS98] CIGNONI P., ROCCHINI C., SCOPIGNO R.: Metro: Measuring error on simplified surfaces. *Computer Graphics Forum* 17, 2 (1998), 167–174. 9

- [DLSH02] DAVATZIKOS C., LIU D., SHEN D., HERSKOVITS E. H.: Spatial normalization of spine MR images for statistical correlation of lesions with clinical symptoms. *Radiology* 224, 3 (2002), 919–926. 2
- [EBK*11] EGGER J., BAUER M. H. A., KUHN D., CARL B., KAPPUS C., FREISLEBEN B., NIMSKY C.: Evaluation of a novel approach for automatic volume determination of glioblastomas based on several manual expert segmentations. *CoRR abs/1103.1474* (2011). 9
- [EKD*12] EGGER J., KAPUR T., DUKATZ T., KOŁODZIEJ M., ZUKIĆ DŽ., FREISLEBEN B., NIMSKY C.: Square-cut: A segmentation algorithm on the basis of a rectangle shape. *PLoS ONE* 7, 2 (2012), e31064. 2
- [FHA*09] FREBURGER J., HOLMES G., AGANS R., JACKMAN A., DARTER J., WALLACE A., CASTEL L., KALSBECK W., CAREY T.: The rising prevalence of chronic low back pain. *Arch Intern Med* 169, 3 (2009), 251–258. 1
- [FL03] FERNÁNDEZ J.-J., LI S.: An improved algorithm for anisotropic nonlinear diffusion for denoising cryo-tomograms. *J. Structural Biology* 144, 1-2 (2003), 152–161. 5
- [HCLN09] HUANG S.-H., CHU Y.-H., LAI S.-H., NOVAK C.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. *IEEE Trans. Medical Imaging* 28, 10 (2009), 1595–1605. 2
- [HM02] HOAD C. L., MARTEL A. L.: Segmentation of MR images for computer-assisted surgery of the lumbar spine. *Physics in Medicine and Biology* 47, 19 (2002), 3503. 2
- [HPE*08] HE L., PENG Z., EVERDING B., WANG X., HAN C. Y., WEISS K. L., WEE W. G.: A comparative study of deformable contour methods on medical image segmentation. *Image and Vision Computing* 26, 2 (2008), 141–163. 2
- [KHDM98] KITTLER J., HATEF M., DUIN R., MATAS J.: On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20, 3 (1998), 226–239. 5
- [KOE*09] 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, 3 (2009), 471–482. 1
- [KWZ*13] KELM B. M., WELS M., ZHOU S. K., SEIFERT S., SUEHLING M., ZHENG Y., COMANICIU D.: Spine detection in CT and MR using iterated marginal space learning. *Medical Image Analysis* 17, 8 (2013), 1283–1292. 2
- [LRDG04] LENCHIK L., ROGERS L. F., DELMAS P. D., GENANT H. K.: Diagnosis of osteoporotic vertebral fractures: importance of recognition and description by radiologists. *American Journal of Roentgenology* 183, 4 (2004), 949–958. 8
- [MCP*09] MICHPOULOU S., COSTARIDOU L., PANAGIOTOPOULOS E., SPELLER R., PANAYIOTAKIS G., TODD-POKROPEK A.: Atlas-based segmentation of degenerated lumbar intervertebral discs from MR images of the spine. *IEEE Trans. Biomedical Engineering* 56, 9 (2009), 2225–2231. 2
- [ML13] MA J., LU L.: Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Computer Vision and Image Understanding* 117, 9 (2013), 1072–1083. 1
- [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* 9 (2012), 8357–8376. 2, 3, 11, 12, 13
- [NFS*11] NEUBERT A., FRIPP J., SHEN K., SALVADO O., SCHWARZ R., LAUER L., ENGSTROM C., CROZIER S.: Automated 3D segmentation of vertebral bodies and intervertebral discs from MRI. In *IEEE DICTA* (2011), pp. 19–24. 2, 3, 11, 12
- [Pow94] POWELL M.: A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in Optimization and Numerical Analysis*. Springer, 1994, pp. 51–67. 9
- [RET13] RAK M., ENGEL K., TÖNNIES K.-D.: Closed-form hierarchical finite element models for part-based object detection. In *Vision, Modeling and Visualization* (2013), Eurographics, pp. 137–144. 2
- [RGM10] RICHARDS P. J., GEORGE J., METELKO M., BROWN M.: Spine computed tomography doses and cancer induction. *Spine* 35, 4 (2010), 430–433. 1
- [Sar87] SARASTE H.: Long-term clinical and radiological follow-up of spondylolysis and spondylolisthesis. *Journal of Pediatric Orthopaedics* 7, 6 (1987), 631–638. 8
- [ŠBSG06] ŠEREDA P., BARTROLÌ A., SERLIE I., GERRITSEN F.: Visualization of boundaries in volumetric data sets using LH histograms. *IEEE Trans. Visualization and Computer Graphics* 12, 2 (2006), 208–218. 5
- [SHD01] SHEN D., HERSKOVITS E. H., DAVATZIKOS C.: An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures. *IEEE Transactions on Medical Imaging* 20, 4 (2001), 257–270. 2
- [ŠLPV10] ŠTERN D., LIKAR B., PERNUŠ F., VRTOVEC T.: Automated detection of spinal centrelines, vertebral bodies and intervertebral discs in CT and MR images of lumbar spine. *Physics in Medicine and Biology* 55, 1 (2010), 247. 4
- [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, 23 (2011), 7505. 2, 3
- [SMB11] SCHULZE F., MAJOR D., BÜHLER K.: Fast and memory efficient feature detection using multiresolution probabilistic boosting trees. *Proc. WSCG* 19, 1 (2011), pp. 33–40. 2
- [TF88] TERZOPOULOS D., FLEISCHER K.: Deformable models. *The Visual Computer* 4, 6 (1988), 306–331. 6
- [The92] THEIL H.: A rank-invariant method of linear and polynomial regression analysis. In *Henri Theil's Contributions to Economics and Econometrics*. Springer, 1992, pp. 345–381. 5
- [VJ01] VIOLA P., JONES M. J.: Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Computer Vision & Pattern Recognition* (2001), vol. 1, pp. 511–518. 2, 3, 4
- [VLP05] VRTOVEC T., LIKAR B., PERNUŠ F.: Automated curved planar reformation of 3D spine images. *Physics in Medicine and Biology* 50, 19 (2005), 4527. 4
- [WLJG95] WU C., LI J., JERGAS M., GENANT H.: Comparison of semiquantitative and quantitative techniques for the assessment of prevalent and incident vertebral fractures. *Osteoporosis International* 5, 5 (1995), 354–370. 8
- [WW83] WILTSE L. L., WINTER R. B.: Terminology and measurement of spondylolisthesis. *The Journal of Bone and Joint Surgery* 65, 6 (Jul 1983), 768–772. 8
- [ZSS96] ZORIN D., SCHRÖDER P., SWELDENS W.: Interpolating subdivision for meshes with arbitrary topology. In *Proc. ACM SIGGRAPH* (1996), pp. 189–192. 7
- [ZVD*12] ZUKIĆ DŽ., VLASÁK A., DUKATZ T., EGGER J., HOŘÍNEK D., NIMSKY C., KOLB A.: Segmentation of vertebral bodies in MR images. In *Vision, Modeling and Visualization* (2012), Eurographics, pp. 135–142. 2, 3, 9, 10, 12, 13