



## Automated model-based vertebra detection, identification, and segmentation in CT images

Tobias Klinder<sup>a,b,\*</sup>, Jörn Ostermann<sup>a</sup>, Matthias Ehm<sup>b</sup>, Astrid Franz<sup>b</sup>, Reinhard Kneser<sup>c</sup>, Cristian Lorenz<sup>b</sup>

<sup>a</sup> Institut für Informationsverarbeitung, Leibniz University of Hannover, Appelstrasse 9A, 30167 Hannover, Germany

<sup>b</sup> Philips Research Europe – Hamburg, Digital Imaging Systems, Röntgenstrasse 24–26, 22335 Hamburg, Germany

<sup>c</sup> Philips Research Europe – Aachen, X-ray Imaging Systems, Weisshausstrasse 2, 52062 Aachen, Germany

### ARTICLE INFO

#### Article history:

Received 5 August 2008

Received in revised form 3 February 2009

Accepted 9 February 2009

Available online 20 February 2009

#### Keywords:

Geometric modelling

Deformable models

Vertebra identification

Vertebra labelling

Vertebra segmentation

### ABSTRACT

For many orthopaedic, neurological, and oncological applications, an exact segmentation of the vertebral column including an identification of each vertebra is essential. However, although bony structures show high contrast in CT images, the segmentation and labelling of individual vertebrae is challenging.

In this paper, we present a comprehensive solution for automatically detecting, identifying, and segmenting vertebrae in CT images. A framework has been designed that takes an arbitrary CT image, e.g., head-neck, thorax, lumbar, or whole spine, as input and provides a segmentation in form of labelled triangulated vertebra surface models. In order to obtain a robust processing chain, profound prior knowledge is applied through the use of various kinds of models covering shape, gradient, and appearance information. The framework has been tested on 64 CT images even including pathologies. In 56 cases, it was successfully applied resulting in a final mean point-to-surface segmentation error of  $1.12 \pm 1.04$  mm.

One key issue is a reliable identification of vertebrae. For a single vertebra, we achieve an identification success of more than 70%. Increasing the number of available vertebrae leads to an increase in the identification rate reaching 100% if 16 or more vertebrae are shown in the image.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

A reliable segmentation of the vertebral column is essential for numerous orthopaedic, neurological, and oncological applications. In the context of computer aided spinal surgery, exact knowledge about the shape of individual vertebrae is of great importance, e.g., for spinal biopsies, implants, or the insertion of pedicle screws. Additional interest in the vertebral column arises from the fact that it serves as a reference structure, e.g., supporting the localisation of other organs. In most cases not only a precise segmentation but also an identification of individual vertebrae is needed, e.g., to assure that the correct vertebra is involved in the surgical procedure.

Since bony structures show high contrast in computed tomography (CT), this imaging technique is in many cases the modality of choice. However, in spite of high contrast, an exact segmentation of vertebrae still remains challenging. Mostly, the segmentation is impeded by, e.g., sampling or reconstruction artefacts, unclear object boundaries, similar structures in close vicinity, or pathologies. Another key issue is the shape complexity of the objects to be

segmented and their inter-patient variability affecting an automated image processing. Furthermore, variability of image acquisition between scanners results in different resolutions and noise.

Although several vertebrae show typical characteristics (e.g., thoracic vertebrae have an articulation to the ribs), the shapes of neighbouring vertebrae are typically very similar so that an automatic identification is generally difficult to obtain. Even physicians usually solve the identification problem by searching for one characteristic vertebra, e.g., the first cervical or first thoracic vertebra, and subsequently label the individual vertebrae iteratively. In many cases, the labelling can be facilitated by the use of reference structures. However, two main problems typically arise when working on spine CT scans (see Fig. 1). On the one hand, most volumes only cover a limited portion of the spine containing a few vertebrae which impedes distinguishing characteristics. On the other hand, images can be reduced in-plane showing only a small region around the vertebrae and thus provide little structural context.

In this paper, we address the automatic detection, identification, and segmentation of individual vertebrae in arbitrary CT scans. A framework has been designed that requires a CT image as input and provides segmented and labelled vertebrae in form of triangulated surface meshes as output. To circumvent the problems mentioned above, prior knowledge is involved in the processing chain

\* Corresponding author. Address: Institut für Informationsverarbeitung, Leibniz University of Hannover, Appelstrasse 9A, 30167 Hannover, Germany. Tel.: +49 (0) 40 5078 2592; fax: +49 (0) 40 5078 2510.

E-mail address: [tobias.klinder@philips.com](mailto:tobias.klinder@philips.com) (T. Klinder).



**Fig. 1.** Disparity of CT images: In (a) only a small portion of the spine is shown, while in (b) the axial view is very limited so that surrounding structures are mostly not visible.

by using various kinds of models covering shape, gradient, and appearance information. In a first step, the spine curve is extracted and a curved planar reformation is performed to visually remove the curvature of the spine. After detecting the vertebrae in the reformatted image, the candidates are identified in the original data set using specific models. Finally, a 3-D deformable model approach is applied to precisely segment the vertebrae.

This paper is organized as follows: Initially, in Section 2, related work dealing with vertebra segmentation and labelling is discussed. In Section 3, differentiating contributions of our approach are outlined and an overview of our framework is given. Section 4 introduces the generation of the different kinds of models involved in the framework. In Section 5, the particular components of the framework are presented in detail. Success rates and accuracy of the framework as well as some further experiments concerning the identification of vertebrae are evaluated in Section 6. A conclusion of the presented methods including a discussion finishes this paper in Section 7.

## 2. Previous work

Several methods have been published aiming at an automated segmentation and labelling of the vertebral column for all common medical image data. However, different modalities, like magnetic resonance imaging (MRI) or computed tomography (CT) require different approaches.

In MR images, bony structures appear as dark regions with sparse detail and low contrast whereas the intervertebral disks often appear much brighter. Thus, most approaches focus at first on the automatic detection of the intervertebral disks. In contrast to MRI, bony structures show high contrast in CT images while the intervertebral disks can be hardly distinguished from surrounding soft tissue.

Vrtovec et al. (2005, 2007) presented a method for spine curve detection in CT and MR images described by a polynomial model to provide a curved planar reformation (CPR) of the spine column. In CT, the spine curve was fitted to a set of points extracted from a distance map that emphasized the vertebral bodies. In MR images, vertebral body centres were located by searching for circular regions of low homogeneous intensity. Although both approaches dealt with spine curve detection and spine reformation, neither detection nor segmentation of individual vertebrae was considered.

A method for an automatic extraction and partitioning of the spinal cord in CT images by locating the vertebral body surfaces was presented by Yao et al. (2006). The spinal cord was detected by using a watershed algorithm whereas a directed graph search discriminated erroneous detections. The intervertebral disks were detected by an evaluation of the grey value profile along the vertebral bodies. The method was successfully performed on 69 of 71 data sets, but leakages into the pelvis occurred in 14 cases. Admit-

tedly, identification and segmentation of the vertebrae were not carried out.

Localisation and identification of the lower spinal anatomy in MR lumbar scans was tackled by Chwialkowski et al. (1991). Starting from the lumbosacral junction, labelling was performed by counting the detected vertebrae. However, since the sacrum had to be at least in parts in the field of view, arbitrary CT images could not be processed. Furthermore, segmentation was not considered.

Probabilistic graphical models for automatically locating the vertebral column and labelling the intervertebral disks were recently proposed by Schmidt et al. (2007) focusing on whole spine and Corso et al. (2008) dealing with lumbar spine MR images. In each case, appearance information of the disks as well as spatial relationships between disks were incorporated in the model. Segmentation was in both cases not addressed.

Two other approaches working on whole MR images that provided a detection and segmentation of vertebral bodies were published by Peng et al. (2005) and Jerebko et al. (2007). Due to performance reasons, Peng et al. extracted a 2D sagittal slice with the largest number of visible intervertebral disks, whereas missing disks, e.g., in the case of a scoliotic spine, were detected in adjacent slices. Segmentation of the vertebral bodies was achieved through a Canny edge operator and gap filling in a 2D slice. Jerebko et al. extracted the spine curve by polynomial fitting and afterwards segmented the vertebral bodies using a model that was based on geometric primitives. Finally, a classification algorithm was applied to detect vertebra metastases.

It has to be stressed that the previous four approaches either worked on whole spine or lumbar spine MR scans, so that the number of expected vertebrae could be taken into account.

In contrast to the above mentioned methods, Herring and Dawant (2001) automatically identified lumbar vertebrae in CT images with an unknown field of view by performing a surface-based registration. The lumbar spine was first coarsely segmented with simple thresholding and triangulated by marching cubes. Earlier generated vertebra models that were built from phantoms were registered to the extracted patient mesh. Identification was successfully carried out on two patients by evaluating the standard deviation of the registered and the reference mesh. Nevertheless, manually placed landmarks in the patient image were necessary for the registration.

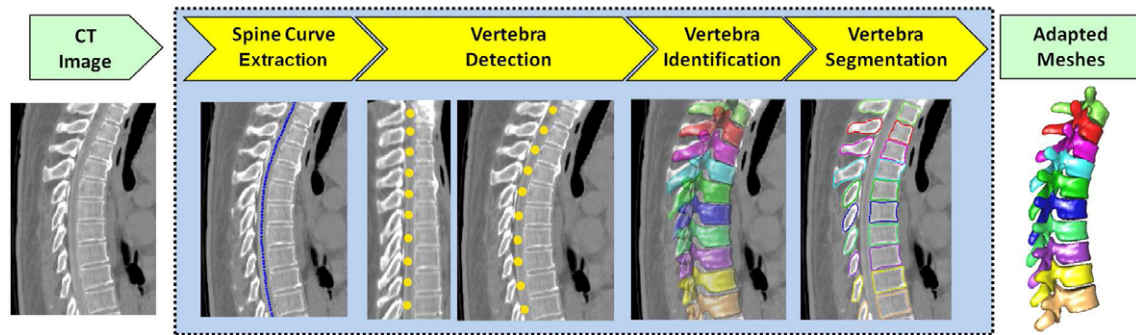
Precise segmentation of lumbar vertebrae in CT images was proposed by Mastmeyer et al. (2006). However, landmarks had to be manually placed in the vertebral bodies. For thorax CT images, Shen et al. (2008) recently presented an automatic vertebra segmentation. Admittedly, vertebra labelling was not addressed. It has to be noted that both methods were developed for specific spine regions.

## 3. Contributions and framework overview

Although various work has been presented aiming at vertebra detection, identification, and segmentation as introduced in the previous section, all of them lack in some key issues due to at least one of the following reasons:

- focus on vertebra detection, identification or segmentation but not addressing all tasks in a complete processing chain,
- process of specific scans, e.g., lumbar or whole spine images,
- reference structures were taken into account and thus had to be shown in the images, e.g., sacrum,
- user interaction was required.

In this paper, we overcome the above mentioned limitations to present a completely automatic framework that processes arbitrary CT scans (head-neck, thoracic, and lumbar but also whole spine images) even with a limited in-plane field of view, since it



**Fig. 2.** Block diagram of framework. Based on the spine curve extraction, a CPR is applied and the vertebrae are detected in the reformatted image. Found candidates are transformed back to the original image – visualized as yellow spheres located between each vertebra. Afterwards, candidates are identified – visualized as prepositioned models; colour coding illustrates labels of vertebra meshes. Finally, meshes are adapted to the respective image structures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

does not require any reference structures. As output, it provides a segmentation represented as labelled vertebrae in form of triangulated surface meshes. Evaluation is performed on a considerably large amount of 64 test volumes in order to draw reliable conclusions. The individual components of our framework are outlined as follows (see Fig. 2):

- (1) *Spine curve extraction:* Based on a progressive adaptation of small tube-shaped segments represented as triangulated surface meshes, the spinal canal is extracted.
- (2) *Vertebra detection:* In order to simplify the detection, a curved planar reformation (CPR) is at first applied on the volume based on the extracted spine curve. Detection is carried out in the reformatted image using generalised Hough transform (GHT) models of all vertebra.
- (3) *Vertebra identification:* For identification, appearance models are rigidly registered to the detected candidates and the resulting similarity measure is evaluated. The models contain average intensity information inside a bounding box around each vertebra.
- (4) *Vertebra segmentation:* The final segmentation is carried out by adapting triangulated shape models of the individual vertebrae. The original adaptation algorithm is thereby extended by additional features designed for the segmentation of model ensembles.

#### 4. Overview of vertebra models

As mentioned in the introduction, several issues impede an automatic vertebra segmentation and labelling. To circumvent most of the problems, we make use of a-priori knowledge in form of different types of models explained in detail as follows.

##### 4.1. Training data

Statistical information about the individual vertebrae is learned from selected reference images. Ten volumes showing the spine entirely have been chosen as training images, so that each model can be generated based on the same amount of samples from the training set. The images were acquired from Philips Brilliance 40 and Brilliance 16P CT scanners with an in-plane resolution ranging from 0.36 mm to 0.96 mm and a slice thickness of 0.5–3.0 mm.

##### 4.2. Vertebra shape models

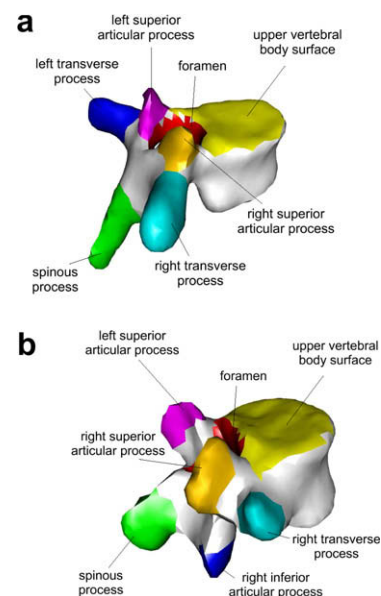
###### 4.2.1. Model geometry

One common way to obtain a first surface representation of the anatomical object of interest is to manually segment the structure from medical image data. In order to circumvent this time-con-

suming approach and be sure to cover all particularities of all vertebrae, e.g., bifurcation of spinous process of cervical vertebrae, we rather based our model generation on the scanning of commercially available plastic phantoms with a Philips Brilliance 40 CT scanner. Out of the scanned data set, the individual vertebrae are firstly segmented using simple thresholding and afterwards triangulated using the marching cubes algorithm (Lorensen and Cline, 1987). Finally, mesh operations are applied like surface smoothing, vertex insertion in the case of too large triangles, or edge deletion in the case of too small edges to obtain a smooth surface with a roughly uniform distribution of vertices. Each final vertebra model consists of the same amount of triangles  $T = 3000$ . Depending on the topology, the number of vertices varies between  $V = 1496$  and  $V = 1500$ . The number of triangles was chosen as a trade-off between model resolution and computational complexity.

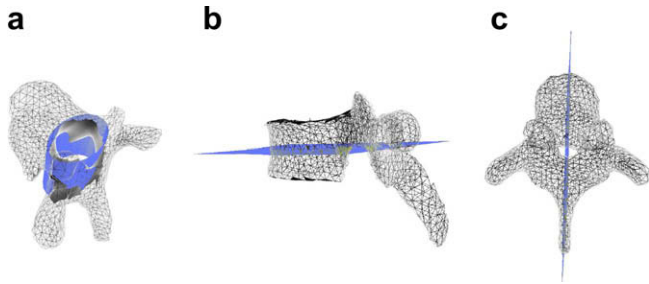
In addition to the vertebrae, a sacrum surface model is generated by manually segmenting the sacrum in one of the available test images. Given the obtained binary image, the same processing is performed as introduced for the vertebrae obtaining a sacrum mesh consisting of  $V = 3425$  vertices and  $T = 6928$  triangles.

In order to easily derive certain vertebra properties directly from the mesh, e.g., distance between vertebral body surfaces, we additionally introduce a sublabelling for the triangles of the



**Fig. 3.** Illustration of vertebra surface models showing the sixth thoracic (a), and fifth lumbar vertebra (b). Labelled regions are coloured.





**Fig. 4.** Invariant features of vertebrae are used to define the VCS: Cylinder fit to vertebral foramen (a), middle plane of upper and lower vertebral body surface (b) as well as sagittal symmetry plane (c) are calculated.

meshes. Except for the atlas (C1) and the axis (C2) which show deviant structures, the following regions have been interactively defined for the third cervical (C3) down to the fifth lumbar vertebra (L5) comparably: upper and lower vertebral body plane, left and right superior articular process, left and right inferior articular process, left and right transverse process, spinous process, and foramen. Exemplarily, Fig. 3 shows two sublabelled vertebra meshes.

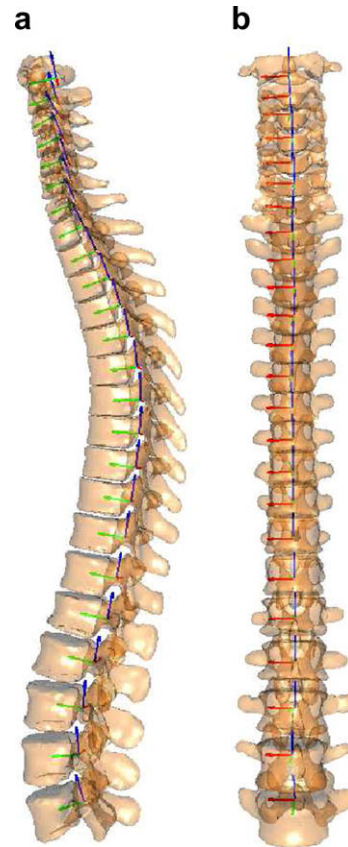
#### 4.2.2. Vertebra coordinate system

In addition to shape information, a local vertebra coordinate system (VCS) is defined for each individual vertebra to express its location and orientation. Since the shape of the vertebrae changes along the spine, a VCS is derived from typical invariant object characteristics based on the above introduced labelling. Moreover, by using object specific features, the VCS is robust against inter-subject changes in shape. In the context of spine curvature analysis, e.g., for scoliosis treatment, the representation of the spine as a consecution of VCSs has proven to be beneficial (see, e.g., Petit et al., 2004; Boisvert et al., 2008). However, in both approaches, the landmarks had to be set manually by clinical experts to derive the respective coordinate systems.

Again, for the vertebrae C3 to L5, the VCS can be defined in the same way since these objects show similar shape characteristics. The definition of the VCS is based on three object-related simplified representations: a cylinder fit to the vertebral foramen, the middle plane of the upper and lower vertebral body surfaces, and the sagittal symmetry plane of the vertebra (see Fig. 4). For the cylinder fit, a Downhill–Simplex algorithm is applied to find the minimum of the sum of squared distances between foramen vertices and cylinder surface. The middle plane is found by fitting two planes to upper and lower body surfaces given by the labelled triangles and finally averaging the plane parameters. The symmetry plane is calculated by searching for a plane perpendicular to the middle plane where the absolute of the sum of signed distances between vertices and plane is minimal.

The origin of the VCS located in the mid-vertebral plane at the centre of the vertebral foramen. The normal vector of the middle plane defines the  $z_{VCS}$ -axis, the  $x_{VCS}$ -axis is defined by the orthogonal component of the normal vector of the symmetry plane and the  $z_{VCS}$ -axis. The cross product of  $z_{VCS}$ -axis and  $x_{VCS}$ -axis results in the  $y_{VCS}$ -axis (Klinder et al., 2007).

To define also a unique coordinate system for C1 and C2, other characteristics are chosen. For the definition of the respective VCSs, a vertex subset is chosen containing the vertices belonging to the inferior articular faces and the anterior tubercle. The axes of the specific VCSs are given by the covariance analysis of the subset. The eigenvector corresponding to the largest eigenvalue of the covariance matrix points to the lateral direction defining the  $x_{VCS}$ -axis, the second eigenvector pointing towards the anterior tubercle gives the  $y_{VCS}$ -axis, and finally, the third eigenvector



**Fig. 5.** Mean vertebra models. Each shape model carries its own local coordinate system. For visualization, spine is expressed as a consecution of VCSs. Mean transformations between neighbouring vertebrae are calculated and mean models are placed at respective location using the VCSs. Lateral view of spine (a) and ventral (b).

points towards the direction of the spinal canal denoting the  $z_{VCS}$ -axis. The barycentre of the subset defines the origin of the VCS.

#### 4.2.3. Mean model generation

One key issue in geometric modelling is to establish surface correspondences between individuals. Many work is still done in the automatic determination of corresponding landmarks between surfaces (see, e.g., Heimann et al., 2005; Hufnagel et al., 2007). In this paper, we tackle the correspondence problem by preserving the established vertex distribution of the generated vertebra shape models. The created reference meshes of the 24 presacral vertebrae and the sacrum are adapted to the ten test images by manually positioning the models and automatically adapting the surface meshes to the respective image structure. By using the adaptation algorithm introduced later in Section 5.4.1, the topology remains unchanged and the vertex distribution is constrained to be similar to the respective reference mesh. In the case of vertebra shape models, a reliable one-to-one correspondence can be obtained as shown in Lorenz and Krahnstöver (2000) when working with template triangulations that are “coated” to the respective individuals. Thus, the introduced sublabelling of characteristic areas remains valid throughout the individualization of the respective reference meshes. In the case of misadaptations, manual corrections were interactively performed (Timinger et al., 2003).

After segmentation of all vertebrae in the test images, a Procrustes co-registration is performed for corresponding vertebrae meshes yielding for each vertebra a respective mean surface model (see Fig. 5). As vertex correspondence are preserved between corre-

sponding vertebra meshes, a point-based registration that accounts for isotropic scale, rotation, and translation is applied (Horn, 1987).

#### 4.3. Generalised hough transform models

The generalised Hough transform (GHT) (Ballard, 1981) is a robust and powerful method to detect arbitrary shapes in an image undergoing geometric transformations. During GHT learning, description of the shape is encoded into a reference table also called *R*-table. Its entries are vectors pointing from the shape boundary to a reference point being commonly the gravity centre of the shape. During detection, the gradient orientation is measured at each edge voxel of the new image yielding an index for an entry of the *R*-table. Then, the positions pointed by all vectors under this entry are incremented in an accumulator array. Finally, the shape is given by the highest peak in the accumulator array.

Usually, the GHT is trained for one single reference shape of an object class. However, in this case, more sophisticated GHT models are built that do not only represent the shape of one object but contain information about an amount of objects. For vertebra detection, we build a comprehensive vertebra GHT model covering C1 to L5 as well as cervical, thoracic and lumbar specific GHT models. The comprehensive model will be applied if the field of view of the available image is completely unknown while the more specific models can be used if some indication for a certain area of the spine is given. For model generation, the respective shapes of the training set are superimposed based on the VCSs and the obtained aligned shapes are used to build up the *R*-table. Note that the orientation is not considered during GHT learning since all shapes were aligned via the respective VCS. To accelerate the GHT which usually requires high memory demand and high computational efforts, we restrict the transformations during detection to isotropic scale and translations not regarding any rotational components. Before applying the GHT, we moreover downsample each image to a voxel resolution of  $3.0 \times 3.0 \times 3.0$  mm using nearest neighbour interpolation. This resolution proved to be sufficient throughout our experiments. For more implementation details concerning the GHT, we refer to Schramm et al. (2006).

In contrast to the above mentioned shape models, the GHT models do not only contain shape information but also gradient information.

#### 4.4. Appearance models

Appearance models in our case make use of the average volumetric intensity information inside a bounding box around the object while not explicitly modelling the object's shape. For each vertebra, mean intensity volumes of the local neighbourhood of the respective object are generated by aligning volumes around corresponding vertebrae and averaging the intensity values. Alignment is performed by registration of vertebra shape models adapted to the

training data to their corresponding mean models which are used as reference structures. Since point correspondences were preserved between corresponding vertebra shape models, the transformation between adapted and mean models can be calculated in a closed-form solution. The obtained transformations can then be applied on the respective voxel positions. For registration, we allow seven degrees of freedom including translation, rotation, and isotropic scaling (Horn, 1987) in order to compensate not only differences in poses but also different sizes of the vertebrae. Each appearance model is represented as a volumetric image with an isotropic sampling distance of 0.5 mm. The size of volume is defined so that all registered vertebrae are completely contained with an additional offset of 3.0 mm in each direction. Location and pose of the object in the volumes are expressed by the VCSs of the mean vertebra models that were used for alignment. Exemplarily, the appearance model of the sixth thoracic vertebra is shown in Fig. 6. Inter-patient variability becomes obvious by blurry object boundaries whereas the particular objects are still clearly delimited.

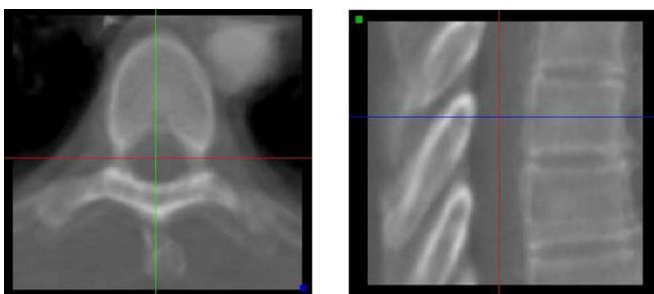
### 5. Framework

#### 5.1. Spine curve extraction

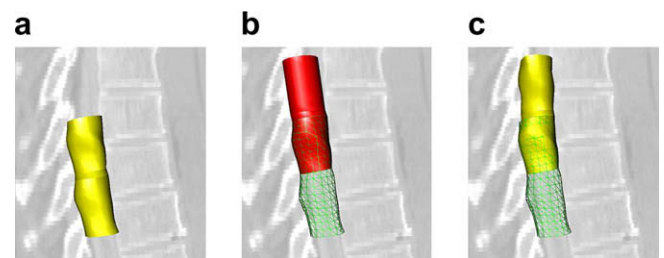
The extraction of the spine curve starts with the detection of a seed point inside the spinal cord. Due to the fact that the spinal cord is surrounded by the cortical bone of the vertebral foramen, we simply apply a GHT using a model which is learned with foramen meshes derived from the labelled vertebra models. By applying the GHT, we always obtained a global maximum in the Hough space that corresponds to a location inside the spinal cord.

Starting from the initial seed point, we apply a progressive adaptation algorithm to trace along the spinal canal. Although there are several other possibilities to extract the centreline of the spine (as e.g., Rangayyan et al., 2006), we have chosen this approach since it enables the tracing and segmentation of highly variable tube-like structures in volumetric images even in the case of low resolution or presence of high noise. Its main idea is to progressively adapt and concatenate small tube-shaped segments, so-called *tubelets*, to trace along any tubular image structure (Peters et al., 2008). Each tubelet is thereby represented as a triangulation of consecutive ellipse shaped contours.

As soon as the first tubelet is positioned based on the detected seed point and adapted, a newly generated tubelet is concatenated and both tubelets are adapted. After adaptation, a new tubelet is again concatenated. However, in all subsequent steps, solely the newly generated tubelet and the adapted tubelet of the last step are adapted to the image structures while all other tubelets are deactivated. The tubelet adapted in the last iteration is thereby utilised to fix the new tubelet. Adaptation is carried out as described in Section 5.4.1.



**Fig. 6.** Appearance model of sixth thoracic vertebra in axial and sagittal view generated from ten training images. The volume has a sample distance of 0.5 mm in each direction.



**Fig. 7.** Progressive adaptation: Recently adapted tube consisting of two tubelets is shown as yellow surface mesh (a). Last adapted tubelet is extracted and a new tubelet is concatenated resulting in one mesh shape shown in red (b) which is finally adapted (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Progressive adaptation is performed in both directions of the spinal canal and finally the centreline is extracted by calculating the barycentres of all contours of the tubular mesh. As tubelet parameters, we used for each tubelet  $N_V = 12$  vertices per contour,  $N_c = 4$  number of contours, and a distance between two contours of  $d = 3$  mm. If the length of the tubelet is too large, the tubelet will not be flexible enough to follow strong spine curvatures especially in the case of pathologies. If the tubelet is too short, the tubelet gets easily trapped in gaps between the vertebrae like in the lumbar spine. Exemplarily, Fig. 7 visualises the progressive adaptation.

One problem of spinal cord extraction is the possible progression of the trace into the sacrum and the cranium, respectively. Since the field of view is unknown but the spine curve of the vertebral column is restricted by the sacrum as a lower and the atlas as an upper boundary, the existence and location of these structures are moreover determined. Detection of the objects is performed using respective GHT models trained for sacrum and atlas. While the sacrum is detected in the original image, a search for the atlas is done in the reformatted image (see next section) since it is very small and its orientation varies more between patients due to the pose of head. The criterion for the objects being in the patient image is the exceeding of a certain threshold concerning the amount of votes in the Hough space as well as the compliance of the inequality  $v_{\text{GHT},1} \geq 1.5v_{\text{GHT},2}$  in the Hough space, where  $v_{\text{GHT},1}$  and  $v_{\text{GHT},2}$  are the votings that correspond to the locations of the two highest maxima. Due to inter-patient variations a constant threshold has turned out to be not sufficient for a successful detection.

## 5.2. Vertebra detection

Before identifying the individual vertebrae, they have to be detected in the image. In order to make the detection invariant of the spine curvature, a curved planar reformation (CPR) similar to Vrtovec et al. (2005) is at first applied. Calculation of the CPR is thereby based on the previous extraction of the spinal cord. Image planes perpendicular to the spine curve within a small region of interest around the spinal canal are extracted and stacked up resulting in the reformatted image. By reducing the region of interest, the performance of all subsequent steps is enhanced while it is ensured that the vertebrae are still completely contained in the image. Furthermore, intersections of the extracted image planes are prevented.

Detection is carried out in the reformatted image using the GHT models introduced in Section 4.3. Since the previously applied CPR removes the spine curvature, a GHT can be applied that only accounts for translation making the GHT computationally feasible. If the atlas or sacrum detection found a reliable maximum so that it can be assumed that the respective object is shown in the image, corresponding specific GHT models are used for detection. In this case, the reformatted image is divided into cervical, thoracic, and lumbar part depending on the size of the shown extract. Based on average lengths for the respective areas, the sizes of the sections are determined. For each section of the reformatted image, detection is then carried out using corresponding specific GHT models meaning, e.g., a GHT model trained with cervical vertebrae for the cervical section. If neither atlas nor sacrum has been found, GHT models incorporating information of all vertebrae are applied, since no indication for the field of view is given. The obtained Hough space usually provides more local maxima than vertebrae contained in the field of view. Candidates that obtain a large distance to the detected spinal cord are removed as well as candidates being too close to neighbouring maxima.

**Table 1**

Evaluation of similarities of vertebra models to detected candidates. The identification is carried out by averaging each diagonal which represents a configuration.

	C1	C2	C3	...	L5
Cand 1	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	...	$s_{1,24}$
Cand 2	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	...	$s_{2,24}$
Cand 3	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	...	$s_{3,24}$
...	...	...	...	...	...
Cand $N_c$	$s_{N_c,1}$	$s_{N_c,2}$	$s_{N_c,3}$	...	$s_{N_c,24}$

## 5.3. Vertebra identification

Extracted candidates are identified by registering the appearance models to the candidates and measuring the similarity of the detected objects to a given model. Identification is carried out in the original image, only considering a small region of interest around the detected candidate to avoid computational effort. The maximum of the similarity determines the final position of the object. The similarity measures between all candidates and models are finally evaluated as illustrated in Table 1.

For each diagonal of this table, denoted as a *configuration*, the average similarity measure is calculated and the diagonal with the highest mean similarity value is supposed to correspond to be the true configuration. Thus, so far we are always assuming that the detection provides consecutive candidates without missing or false detections. By evaluating the whole table, which means that, e.g., also the fifth lumbar vertebra is compared to the atlas, not only false detections beyond the vertebral column can be detected, but also the reliability of the identification of vertebra candidates are evaluated. As a similarity measure for the registration, we apply local correlation which turned out to provide the best results compared to cross correlation and sum of squared differences. Although local correlation is a spatial filtering technique, it is computationally attractive since only small neighbourhoods of six adjacent voxels and a small fraction of image voxels, are sufficient for a correct alignment (Netsch et al., 2000). For the local correlation, we define the squared correlation coefficient

$$C_i(I^{\text{src}}, I^{\text{trg}}) = \frac{\left(\sum_{l \in S_i} X_{i,l}^{\text{src}} X_{i,l}^{\text{trg}}\right)^2}{\sum_{l \in S_i} \left(X_{i,l}^{\text{src}}\right)^2 \sum_{l \in S_i} \left(X_{i,l}^{\text{trg}}\right)^2} \quad (1)$$

where  $X_{i,l}^{\text{src}} = I_i^{\text{src}} - \overline{I_i^{\text{src}}}$  and  $X_{i,l}^{\text{trg}} = I_i^{\text{trg}} - \overline{I_i^{\text{trg}}}$ . The respective average grey values in the neighbourhood  $S_i$  are denoted as  $\overline{I_i^{\text{src}}}$  and  $\overline{I_i^{\text{trg}}}$  while  $I_i^{\text{src}}$  and  $I_i^{\text{trg}}$  are the grey values of voxel  $l$  of reference and target image. The number of voxel neighbourhoods in the image are chosen among those that have the largest variance of image grey values in the corresponding 6-neighbourhoods being anatomically most relevant (Netsch et al., 2000). The local correlation is obtained by summing up all contributions

$$\text{LC}(I^{\text{src}}, I^{\text{trg}}) = \frac{1}{M} \sum_{i \in M} C_i(I^{\text{src}}, I^{\text{trg}}) \quad (2)$$

where  $M$  is the number of voxels in the overlap region of source image and transformed target image. Note that the local correlation  $\text{LC}(I^{\text{src}}, I^{\text{trg}})$  with  $0 \leq \text{LC}(I^{\text{src}}, I^{\text{trg}}) \leq 1$  is a measure of the similarity and can be interpreted as a probabilistic value. The optimal alignment of reference and transformed target image which corresponds to the maximum of the similarity measure is finally obtained using a Gauss–Newton optimisation. For alignment, we allowed an affine transformation with twelve degrees of freedom.



## 5.4. Vertebra segmentation

### 5.4.1. Mesh adaptation

Adaptation of the triangulated meshes is carried out using a shape-constrained deformable models approach (Weese et al., 2001) where an external force attracts the mesh triangles to image features while an internal constraint assures the model shape. Using a physical metaphor, the iterative procedure of mesh deformation is performed by minimizing the summed energy term

$$E = E_{\text{ext}} + \alpha E_{\text{int}}. \quad (3)$$

The parameter  $\alpha$  balances the influence of both energy terms. Mesh adaptation is accomplished by carrying out surface detection for the barycentre of each triangle  $\hat{\mathbf{x}}_i$  in the direction of its normal  $\mathbf{n}_i$

$$\mathbf{c}_j = j\delta\mathbf{n}_i \quad \text{with } j = -l, \dots, l \quad (4)$$

which results in  $(2l+1)$  sampling points  $\mathbf{c}_j$  with a distance of  $\delta$ .

A feature function, yielding a high response at object boundaries, is evaluated at every sampling point. The new candidate point  $\tilde{\mathbf{x}}_i$  is that point  $\hat{\mathbf{x}}_i + \mathbf{c}_j$  that best combines feature value  $F(\cdot)$  and distance  $j\delta$  to the current point  $\hat{\mathbf{x}}_i$

$$\tilde{\mathbf{x}}_i = \hat{\mathbf{x}}_i + \delta\mathbf{n}_i \operatorname{argmin}_{j=-l, \dots, l} \{Dj^2\delta^2 - F(\hat{\mathbf{x}}_i + \mathbf{c}_j)\} \quad (5)$$

with  $D$  controlling the influence of feature strength and distance to the triangle barycentre.

In order to detect object boundaries, the function  $F(\cdot)$  is based on the image gradient. Due to the fact that the gradient can be very high in local areas, e.g., metal boundaries like screws or wires for bone fixation after surgery, the gradient can be optionally bounded by some threshold  $g_{\text{max}}$ . A parameter  $\sigma \in \{1, -1\}$  is moreover introduced accounting for the gradient direction. With these forethoughts, the feature function is formulated as

$$F(\mathbf{x}) = \begin{cases} \sigma \cdot G(\mathbf{x}) & : I_{\min} < I(\mathbf{x}) < I_{\max} \\ 0 & : \text{otherwise} \end{cases} \quad (6)$$

with

$$G(\mathbf{x}) = \mathbf{n}^T \nabla I(\mathbf{x}) \frac{g_{\text{max}}(g_{\text{max}} + \|\nabla I(\mathbf{x})\|)}{g_{\text{max}}^2 + \|\nabla I(\mathbf{x})\|^2} \quad (7)$$

where  $\mathbf{n}$  is the normal of the triangle for which the feature search is carried out. The restriction of image intensity values to  $I_{\min} < I(\mathbf{x}) < I_{\max}$  makes the adaptation less vulnerable to adjacent false attractors.

With the most promising surface points  $\tilde{\mathbf{x}}_i$  and the current triangle centres  $\hat{\mathbf{x}}_i$ , the external energy is given as

$$E_{\text{ext}} = \sum_{i \in T} w_i \left\| \frac{\nabla I(\tilde{\mathbf{x}}_i)^T \nabla I(\hat{\mathbf{x}}_i)}{(\|\nabla I(\hat{\mathbf{x}}_i)\|)^2} (\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i) \right\|^2 \quad (8)$$

The projection of  $(\tilde{\mathbf{x}}_i - \hat{\mathbf{x}}_i)$  onto the gradient normal at the target point makes the energy invariant to movements of the triangle within the object tangent plane, preventing the triangle from becoming stuck at the target position.

The weights  $w_i$  in Eq. (8) are chosen according to

$$w_i = \max \{0, F_i(\tilde{\mathbf{x}}_i) - Dj^2\delta^2\} \quad (9)$$

to give the most promising surface points the largest influence during mesh reconfiguration.

Penalizing a deviation of the models shape regularises the image forces acting on it. Thus, only mesh shapes consistent with defined constraints are possible and the attraction of individual triangles to false image structures decreases. The internal energy

**Table 2**

Overview of parameters involved in mesh adaptation.

Parameter	Description
$\alpha$	Balances influence of ext. and int. energy
$l$	Defines number of sampling points on search profile
$\delta$	Distance between sampling points on profile
$D$	Balances feat. strength and distance to triangle
$\sigma$	Considered gradient direction for feat. search
$I_{\min} / \max$	Min/max grey value considered for feat. search
$g_{\text{max}}$	Gradient threshold for feat. search

$$E_{\text{int}} = \sum_{j \in V} \sum_{k \in N(j)} \|(\hat{v}_j - \hat{v}_k) - s\mathbf{R}(v_j - v_k)\|^2 \quad (10)$$

preserves shape similarity of all adapted vertices  $v_j$  to the model vertices  $\hat{v}_j$  with  $N(j)$  being the set of neighbours of vertex  $j$ . The scaling factor  $s$  and the rotational matrix  $\mathbf{R}$  are determined by a closed-form point-based registration method based on a singular value decomposition prior to calculation of Eq. (10). The introduced energy term penalises deviations from the initial model that can not be expressed by a similarity transformation on the vectors between neighboured vertices of the adapted mesh and corresponding vertices of the model.

In the final optimisation scheme, represented by the minimization of both energy terms (Eq. (3)), the vertex positions of the surface mesh are indeed the parameters to be varied. As only interdependencies between neighbouring vertices exist and the energy terms are of a quadratic form, the conjugate gradient method is used for minimization of the final equation system with a sparsely filled matrix.

As the adaptation depends on a high number of parameters, Table 2 gives a final overview of all variables with a short description.

### 5.4.2. Multi-object segmentation

In the case of images with, e.g., high noise or low resolution, or strong calcifications between vertebrae, misadaptation frequently occurred that were manifested as an overlapping of the vertebral body surfaces of neighbouring vertebrae. One way to improve the segmentation is to account for spatial relations between the involved objects. Thus, each vertebra is not adapted individually but a simultaneous segmentation of the ensemble is applied where the objects are interacting on each other throughout the adaption. For this purpose, we modified the decision for the best point for each triangle during the feature search from Eq. (5) by including a collision detection

$$\tilde{\mathbf{x}}_i = \hat{\mathbf{x}}_i + \delta\mathbf{n}_i \operatorname{argmin}_{j=-l, \dots, l} \{cd + Dj^2\delta^2 - F(\hat{\mathbf{x}}_i + \mathbf{c}_j)\}. \quad (11)$$

The additional term  $cd$  accounts for the collision with other objects where  $c$  is a weighting factor and  $d$  gives the depth of the collision. For each triangle, collision detection is carried out along the search ray defined by  $\mathbf{r}_i = \hat{\mathbf{x}}_i + t\mathbf{n}_i$ . The triangle barycentre is denoted as  $\hat{\mathbf{x}}_i$  and  $\mathbf{n}_i$  is its surface normal. If a sampling point lies inside another mesh,  $d$  is equal to the distance between the sampling point and the closest intersection point of the search ray and the neighbouring mesh. If no collision is present,  $d$  will be equal to zero. Note that Eq. (11) does not generally prohibit collision but makes points that lie inside other meshes less attractive. For higher values of  $c$ , collision is more punished. In Klinder et al. (2008), we have shown that this modification of the feature search improves the segmentation.

An effective implementation of the collision detection is achieved by using the sublabelling of the meshes so that adjacent surface patches of neighbouring meshes can be assigned, i.e. in our case patches that belong to the vertebral body surface planes of neighbouring vertebrae. Only search rays belonging to triangles

of a patch that potentially shows an intersection with a surface of a neighbouring mesh are tested for collision. Thus, the computational complexity is significantly reduced. Adaptation using collision detection is not applied for the lumbar spine as in this area a larger distance between the individual vertebrae is present.

## 6. Results

### 6.1. Image data

All experiments were carried out on 64 patient images with a voxel size in-plane ranging from 0.316 mm to 0.976 mm and a slice thickness of 0.8 to 2.79 mm. Each image plane had  $512 \times 512$  voxels with varying voxel numbers in *z*-direction of about 20–300. The image data consisted of

- 18 head-neck scans,
- 33 thorax scans including 3 scans showing also the lumbar spine,
- 9 abdominal scans showing a small part of the lower thoracic and upper lumbar spine,
- 4 scans that exclusively show the lumbar spine.

The data was inspected by a clinical expert who categorised the images into

- 5 scoliosis cases,
- 1 kyphosis cases,
- 5 cases had one or more collapsed vertebra,
- 3 images with inserted pedicle screws,
- 1 image with contrast agent in the spinal canal,
- 35 images with strong calcifications or degenerative changes to the endplates (spondylophytes), in some cases almost ubiquitous and bridging,
- 2 very noisy images,
- 21 cases without significant abnormalities.

The images came from several institutions and were acquired with different scanners. The ten training images were not included in the test data.

### 6.2. Ground truth generation

For the evaluation of our framework, especially for the final segmentation, we were faced with the problem how to effectively generate a valid ground truth on such a large test set with an amount of objects of interest (>400 vertebrae). One possibility was to adapt the respective vertebra meshes and then interactively correct misadaptations. While this introduced a slight bias since the ground truth generation made partially use of the same algorithm, a ground truth for each vertebra took less than 10 min.

In order to analyse the effect of the bias, an independent ground truth was generated for three exemplarily selected cases (one normal head-neck, one scoliotic thoracic, and one normal abdominal scan) by working slice by slice and labelling each voxel belonging to a certain vertebra. Segmentation of each vertebra took thereby about 30–40 min. In order to compare the two ground truths, a voxel volume was calculated from the corrected meshes and the distance of all edge voxels of one ground truth to the closest edge voxels of the other and reverse was calculated. Comparing the two ground truths, we observed an average difference of  $0.76 \pm 0.62$  mm. Finally, the segmentation results for the three cases were evaluated using both ground truths. An evaluation using the voxel-wise ground truth gives an average voxel difference of  $0.96 \pm 1.21$  mm while  $0.85 \pm 1.13$  mm using the ground truth obtained

from a corrected mesh. As the ground truth generation by mesh corrections was much faster but the difference to an independent ground truth did not seem to be very different, we interactively corrected all vertebrae in the 64 test cases providing our ground truth for further evaluation.

It has to be noted that the generated ground truths were verified by a clinical expert.

### 6.3. Framework performance

For each component of the designed framework, runtime and success rates were analysed. Since the runtime performance depended on the size of the image and the amount of shown vertebrae, all values concerning time measurements were calculated only for thoracic scans showing twelve vertebrae and ran on a Pentium 4 with 2.4 GHz.

#### 6.3.1. Spine curve extraction

The initial seed point detection in the centre of a vertebral foramen, represented by the global maximum in the Hough space, was carried out reliably for all patient images. The average detection time on thoracic scans was 7.6 s.

Spine curve extraction using progressive adaptation of tubelets was successfully performed on all data sets with an average processing time on thoracic scans of 7.0 s. In all cases the atlas and the sacrum were detected so that the end of the spinal cord was also found correctly.

Two exemplarily chosen cases for the spinal cord extraction are shown in Fig. 8.

#### 6.3.2. Vertebra detection

The detection of consecutive vertebrae was successfully performed on 59 out of 64 data sets resulting in a success rate of 92%. Detection was defined as a success if one candidate was found for each vertebra that was completely shown in the image. Furthermore, the candidate had to be located within the volume given by the vertebral body planes of the corresponding vertebra. The average detection time of all vertebra candidates in thoracic scans amounted to 13.3 s. As an example, Fig. 9 shows the result for a thoracic CT scan of a patient with an extreme spine curvature. The reformatted image with the corresponding overlaid Hough space is illustrated in (a). The local maxima of the Hough space were determined and the found candidates were transformed into the original patient image shown as local coordinate systems (b).

The detection succeeded for all images categorised as normal cases but also for all available cases with inserted pedicle screws or contrast agent in the spinal canal as well as for all scoliotic or kyphotic spines. It is worth mentioning that even a very difficult

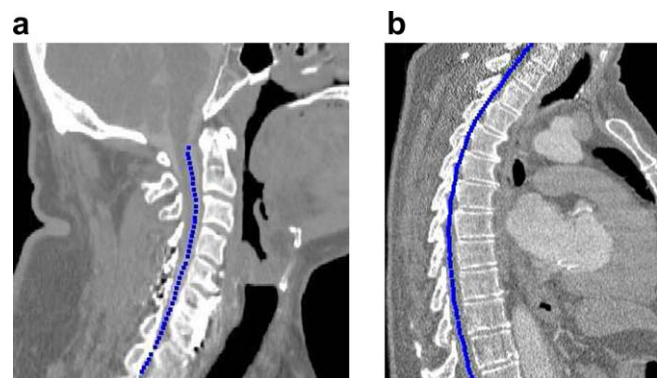
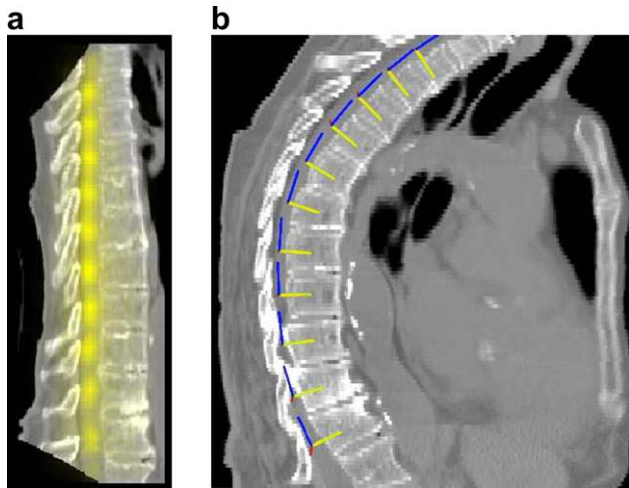


Fig. 8. Exemplarily chosen cases for successful spine curve extraction: Image with metal artefacts (a) and noise (b).





**Fig. 9.** Example of the candidate detection. Using the extracted centreline, the spine column is reformatted and local maxima on the Hough space are determined by a generic GHT model (a). The determined candidates are transformed back into the patient image and visualized as local coordinate systems (b).

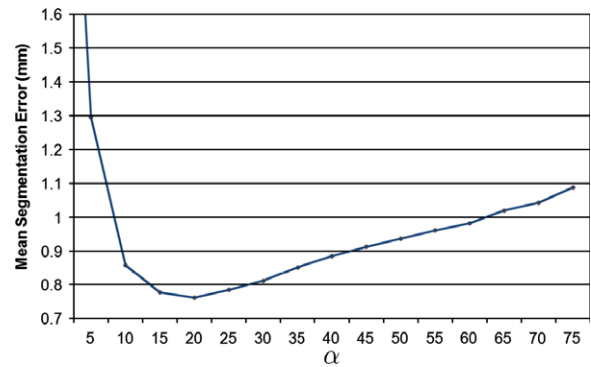
case with six collapsed vertebrae was successfully processed. However, especially in the head-neck region where the vertebrae are very small, pathologies heavily impeded an automatic detection. Thus, three head-neck cases did not succeed due to strong calcifications and degenerative changes to the endplates. Moreover, two thoracic images failed, one had one collapsed vertebra and the other had again strong calcifications between the vertebrae resulting in very unclear object boundaries. In all five failed cases, one vertebra was not detected.

Although GHT models training was based on the original images, whereas vertebra detection was carried out on reformatted images, detection did not improve when using models trained on CPR images.

We furthermore investigated vertebra detection based on a grey value profile analysis along the spinal cord similar to Yao et al. (2006). However, to our experience that approach worked well for lumbar and thoracic spines but could be hardly adapted for cervical cases or thoracic images with strong calcifications where the boundary between neighbouring vertebra bodies is often not visible.

### 6.3.3. Vertebra identification

In 56 out of the remaining 59 cases (95%), identification succeeded, where in each failure case a shift by one vertebra position



**Fig. 11.** Influence of parameter  $\alpha$  on segmentation accuracy based on training images. When  $\alpha$  gets too small, attraction to misleading structures occurs, when  $\alpha$  gets too large the model stays similar to the initial shape mainly undergoing rigid transformations.

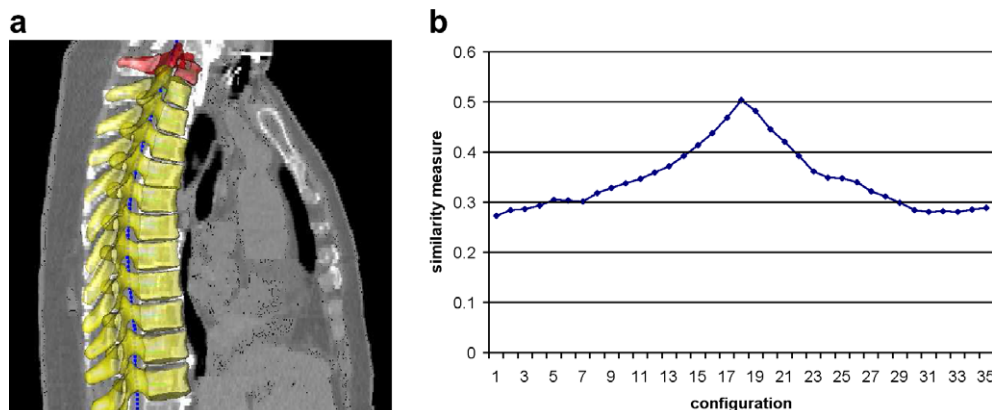
occurred. Two normal cases failed, one thoracic and one lumbar, showing only a small portion of the spine, and one pathological thoracic case with collapsed vertebrae. However, the registration even truly identified one pathological case with only four lumbar vertebrae.

Considering the table of similarities (Table 1), we determined the mean value of each diagonal. Thus, the number of possible configurations amounted to  $N_{\text{conf}} = N_{\text{cand}} + N_{\text{mod}} - 1$  where  $N_{\text{cand}}$  and  $N_{\text{mod}}$  denote the number of candidates and models, respectively, e.g., with 24 appearance models and 12 detected candidates, the number of possible configurations amounts to 35. Exemplarily, Fig. 10 shows the identification of one case.

Although a multi-resolution approach was applied, the affine registration was quite time-consuming, lasting for the identification of twelve vertebrae in thoracic scans 2192 s (36.5 min) on average. Since the detection curves always showed one global maximum (see Fig. 10), more efficient search strategies could be implemented than testing out each combination. However, in this study, we were rather interested in analysing the problem in detail than finding already a computationally attractive solution.

### 6.3.4. Vertebra segmentation

Free-form deformation of the positioned meshes was carried out as introduced in Section 5.4.1 and Section 5.4.2. The affine registration performed for candidate identification already provided a good initialization with a mean vertex-to-surface error between adapted and ground truth meshes of 1.83 mm.



**Fig. 10.** Identification of vertebrae: Truly identified candidates with positioned mean meshes (left). The seventh cervical vertebra is coloured in red, while the thoracic vertebrae are shown in yellow. Note that the meshes have not yet been adapted, but the registration found already a good position of the meshes in the image. The corresponding similarity measures for each configuration (right) shows a global maximum for the true configuration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

**Table 3**  
Accuracy of segmentation for individual vertebrae concerning sublabelled regions (SU, upper and lower vertebral body planes; SP, spinous process; TP, transverse process; All, entire mesh). Results concerning sublabelled regions are not presented for the cervical vertebrae if the region is not clearly defined or only few triangles belong to it.

	C1	C2	C3	C4	C5	C6	C7	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	L1	L2	L3	L4	L5
All $\mu$	0.92	0.81	0.94	0.94	1.02	1.03	1.13	1.02	1.36	1.35	1.28	1.28	1.34	1.23	1.30	1.25	1.01	1.01	1.10	0.74	0.79	0.76	0.76	0.73
All $\sigma$	0.99	0.97	1.04	1.07	1.07	1.08	1.09	1.02	1.03	1.04	1.13	1.09	1.21	1.23	1.09	1.00	0.90	0.95	1.03	0.84	0.94	0.89	1.01	0.82
SU $\mu$	-	-	1.07	0.85	0.88	1.09	1.03	1.03	1.52	1.43	1.38	1.37	1.60	1.37	1.54	1.54	1.36	1.39	1.43	0.93	0.93	0.86	0.82	0.74
SU $\sigma$	-	-	0.99	1.05	1.04	1.06	1.10	1.02	1.03	1.07	1.11	1.09	1.22	1.09	1.23	1.11	1.03	1.02	1.00	0.97	0.94	0.96	0.90	0.88
SP $\mu$	-	-	1.33	0.90	0.90	0.96	0.76	0.95	1.31	1.32	1.33	1.34	1.52	0.94	1.09	1.02	0.77	0.96	1.09	0.75	0.90	0.70	0.69	0.73
SP $\sigma$	-	-	0.98	1.04	1.04	1.11	1.02	1.01	1.01	1.05	1.09	1.21	1.31	1.01	0.96	0.97	0.89	0.91	1.01	0.89	0.91	0.72	0.75	0.77
TP $\mu$	-	-	-	-	-	-	-	1.36	1.60	1.72	1.54	1.47	1.83	1.58	1.57	1.53	1.23	1.12	1.42	0.91	1.03	1.13	1.27	0.63
TP $\sigma$	-	-	-	-	-	-	-	1.03	1.03	1.13	1.19	1.09	1.21	1.18	1.16	1.02	1.01	0.98	1.05	0.93	0.96	1.03	0.81	0.87

One crucial aspect of the adaptation was its dependence on several parameters (see Table 2). For the analysis of the sensitivity of the adaptation with respect to the settings, we evaluated the vertex-to-surface error between adapted and ground truth meshes for the ten training images. The parameters  $l$  and  $\delta$  defining the sampling profile for feature search were inspected carefully. When the distance between sampling points on the profile was too large, the exact borders could not be found thus increasing the segmentation accuracy significantly. For that purpose, the sampling distance  $\delta$  was set to 0.3 mm being the smallest resolution contained in the test images. By defining a number of sampling points per profile, it had to be ensured that the object borders were within the capture range of the adaptation. A value of  $l = 25$  turned out to be appropriate. Keeping the sampling distance but defining a longer search profile did not improve the segmentation but slowed down the segmentation because more points were taken into account. To limit the number of possible false attractors but still ensure to take into account all voxels belonging to a vertebra, the minimum hounsfield value considered during feature search  $I_{\min}$  was set to 0 HU. Furthermore, the maximal hounsfield value  $I_{\max}$  was fixed to 2000 HU to prevent adaptation to metal artefacts. Due to the fact that the gradient at the boundary of a vertebra always showed a transition from bright to dark, the parameter  $\sigma$  was set to one.

Since the parameter  $\alpha$  influences the weighting of the two energy terms, its impact on the segmentation result was rather large. To analyse its influence on the final segmentation, we varied its value but keeping all other parameters fixed with a plot given in Fig. 11.

However, as  $g_{\max}$  limits the response of the external energy, its value has implicitly an influence on the trade-off of external and internal energy, too. An increase of  $g_{\max}$  had to some extent the same effect as a decrease of  $\alpha$ . Thus, we moreover calculated a  $g_{\max}$  resulting in a minimal segmentation error in dependence of  $\alpha$ . The higher  $\alpha$  gets, the higher the optimal values for  $g_{\max}$  becomes while when decreasing  $\alpha$ , the optimal values become smaller. Interestingly, the optimal values for  $g_{\max}$  remained almost constant around the minimum shown in Fig. 11. Based on the ten training images, optimal values were found as  $g_{\max} = 121$  HU/mm and  $\alpha = 18$ . We moreover investigated how  $\alpha$  varied between the training images when its optimal value was determined for each image individually while the other parameters remained unchanged. The optimal value varied between 15 and 22.

Finally, we analysed the sensitivity of the algorithm against changes in distance weight  $D$  and collision detection weight  $c$ . However, the results turned out to be not very sensitive against changes of these parameters. By keeping all other parameters but varying both values around empirically found settings, optimal weights were found as  $D = 0.6 \frac{1}{\text{mm}^2}$  and  $c = 2.2 \frac{1}{\text{mm}}$ .

Vertebra segmentation was performed on the test images with  $\alpha = 18$  and  $g_{\max} = 123$  HU/mm,  $\sigma = 1$ ,  $l = 25$ ,  $\delta = 0.3$  mm,  $D = 0.6 \frac{1}{\text{mm}^2}$ ,  $I_{\min} = 0$  HU,  $I_{\max} = 2000$  HU and  $c = 2.2 \frac{1}{\text{mm}}$ . Adaptation stopped when convergence was reached.

To analyse the dependence of the number of vertices, meshes with different number of vertices were adapted to the training images. However, the mesh resolution seemed not to be very crucial regarding segmentation accuracy unless the number of vertices was not too low providing only a rough surface approximation.

For the evaluation of the final segmentation error, we determined the distance of a vertex of the adapted mesh to the closest surface point (not necessarily a vertex) of the ground truth mesh. Additionally, the error for selected subregions that were considered to be critical from the segmentation perspective was calculated as presented in Table 3. With an overall final mean error of

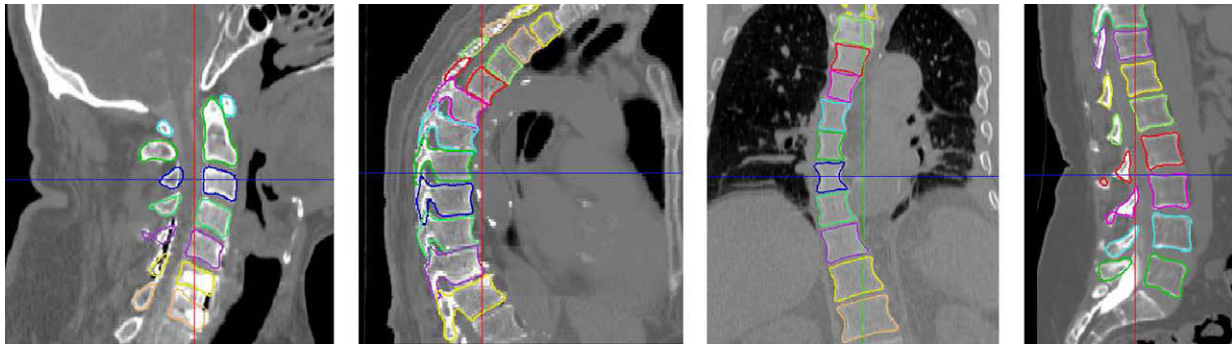


Fig. 12. Segmentation results for four chosen volumes. The colour coding illustrates the vertebra labelling.

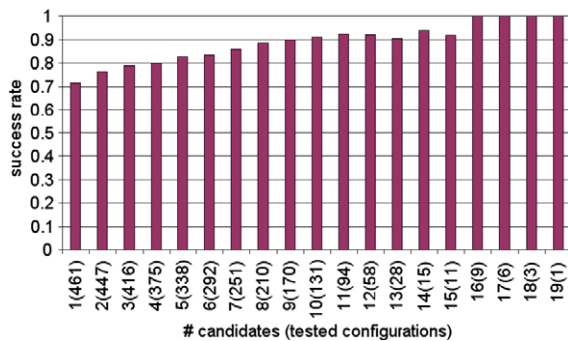


Fig. 13. Identification success rates for different amounts of vertebrae. Success rates increase as more candidates were detected. The values in parentheses on the x-axis denote the number of configurations for each case.

$1.12 \pm 1.04$  mm, we strongly believe that a reliable segmentation was obtained. Selected example cases are shown in Fig. 12. Even for the areas expected to be critical, the segmentation results for the three chosen areas did not extremely differ compared to the overall mean. The segmentation of the thoracic vertebrae was slightly worse than for cervical and lumbar scans mainly caused by two facts. On the one hand, the transverse processi adapted in some cases to the ribs, which can be also seen from a slightly higher segmentation error concerning the transverse processi. On the other hand, most of the thoracic scans showed a low resolution. As the lumbar vertebrae were mostly clearly separated from surrounding bony structures, the segmentation results were indeed better than for the other regions. Regarding thoracic CT scans, the adaptation process required an entire computation time of 179.5 s for all objects in the image.

#### 6.4. Identification experiments

One key contribution of our processing chain is the vertebrae identification as neighbouring vertebrae show similar characteristics. To analyse the robustness of the identification, further experiments were performed.

##### 6.4.1. Dependence on the amount of candidates

By analysing each row of every table of similarities (see Table 1), the success rate for an arbitrary amount of detected candidates could be evaluated. Fig. 13 shows the identification success rate for one up to 19 candidates (largest amount of visible vertebrae in image). Even if only one single vertebra is given, the registration successfully identified the object in more than 70% of the cases. Increasing the number of available vertebrae lead to an increase in identification rate reaching 100% if 16 or more vertebrae were shown in the image since this reduced the number of possible configurations. The slight decrease for the cases of 13 and 15 shown vertebrae was probably caused by the fact that the more vertebrae were shown, the smaller the sample size was.

##### 6.4.2. Identification rate for individual vertebrae

Another key aspect is how reliable certain vertebrae can be identified. From a visual inspection, there are some vertebrae that can be identified more easily than others. The identification rates for individual vertebrae is shown in Fig. 14a. The values were obtained by evaluating each row in the table of similarities (see Table 1) meaning that for each candidate all vertebrae were registered. As one would expect, the identification rate was quite low in the middle of the thoracic spine since in this area the neighbouring vertebrae were very similar and could also be hardly distinguished by human observers. However, the large differences in success rates between the neighbouring vertebrae, e.g., fourth cervical

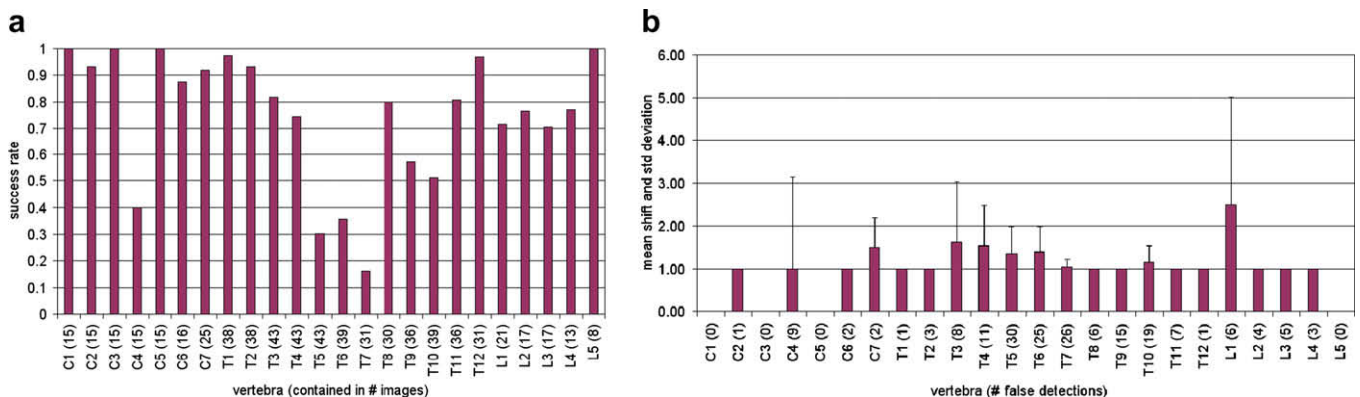


Fig. 14. Vertebra identification rates (a), mean transposition and standard deviation in the case of failed identification (b).



vertebra (C4), were surprising and probably caused by the fact that the appearance models were created only from a small sample size of ten images. Another reason could be that neighbouring objects – besides of the cervical vertebrae – were extracted in parts from different samples. The shift in vertebrae in the case of a failure in identification is given in Fig. 14b. In 84% a shift by one vertebra, and in 13% by two vertebrae occurred. It has to be noted that each row always showed a global maximum in the table of similarities.

## 7. Discussion and conclusion

To the best of our knowledge, this is the first time that a complete framework for detection, identification, and segmentation of the vertebral column in arbitrary CT images is reported. For that purpose, models of all presacral vertebrae were generated covering shape, gradient, and appearance information. Although all models were trained on a relatively small sample of ten images, useful results were achieved for a considerable amount of 64 CT images even including various kinds of pathologies as, e.g., scoliotic spines or inserted pedicle screws. In contrast to most of the earlier proposed work, we presented results for data showing different regions of the spine, e.g., head-neck, thorax, or whole spine images and not only for one specific type of data. The framework showed its robustness by working successfully in 56 cases and its precision by obtaining a mean point-to-surface error of  $1.12 \pm 1.04$  mm. In five cases, the detection failed while in three data sets the candidates were not truly identified.

One difference of our framework to most other approaches is the possibility to identify vertebrae shown in the images but still coping with arbitrary data and not limiting the processing chain to specific data, as, e.g., whole spine images. Since individual vertebrae show similar characteristics, various experiments dealing with vertebra identification were carried out. Even if only a single vertebrae was given, the identification was successful in more than 70% cases. In the case of a failed identification, mostly a shift by only one vertebra occurred.

Besides the identification, the computational effort of the proposed method is acceptable. However, for a clinical setup various improvements might be taken into account to speed up the identification. In our experiments all possible combinations were evaluated in order to analyse the possibility of identifying vertebrae. Even in images with a limited in-plane field of view always show the onsets of the ribs in the case of thoracic vertebrae. Hence a classification of thoracic and non-thoracic vertebrae which reduces the number of possible configurations of the object constellation is a promising approach for overcoming these extensive computations. Furthermore, the fact that the similarity measures showed a global maximum when testing out all constellations might be taken into account to search more efficiently for the true combination.

Future work will investigate these possibilities to reduce the computational effort so that the system can be integrated in a clinical workflow. Moreover, we will further increase the robustness of the framework. So far, most failure cases were caused by missing detections. To improve the detection rate, one option might be to integrate a more global model, as, e.g., proposed by Schmidt et al. (2007). Another point to be addressed is a comprehensive analysis of the accuracy of clinical parameters derived from the output of our framework, as, e.g., spine curvature.

## Acknowledgements

We thank the Princess Margaret Hospital, Toronto, Canada for providing all image data, PFL-A for the Generalized Hough Transform software, and Dr. med. S. Dries for categorising the images and inspecting the ground truth.

## References

- Ballard, D., 1981. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition* 13 (2), 111–122.
- Boisvert, J., Chéret, F., Pennec, X., Labelle, H., Ayache, N., 2008. Geometric variability of the scoliotic spine using statistics on articulated shape models. *IEEE Transactions on Medical Imaging* 27 (4), 557–568.
- Chwialkowski, M., Shile, P., Pfeifer, D., Parkey, R., Peshock, R., 1991. Automated localization and identification of lower spinal anatomy in magnetic resonance images. *Computers and Biomedical Research* 24 (2), 99–117.
- Corso, J., Alomari, R., Chaudhary, V., 2008. Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: *Proceedings of Medical Imaging Computing and Computer Assisted Intervention*, vol. LNCS 5241, pp. 202–210.
- Heimann, T., Wolf, I., Williams, T., Meinzer, H., 2005. 3D active shape models using gradient descent optimization of description length. In: *Proceedings of Image Processing in Medical Imaging*, pp. 566–577.
- Herring, J., Dawant, B., 2001. Automatic lumbar vertebral identification using surface-based registration. *Computers and Biomedical Research* 34 (2), 74–84.
- Horn, B., 1987. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A* (4), 629–642.
- Hufnagel, H., Pennec, X., Ehrhardt, J., Handels, H., Ayache, N., 2007. Shape analysis using a point-based statistical shape model built on correspondence probabilities. In: *Proceedings of Medical Imaging Computing and Computer Assisted Intervention*, vol. LNCS 4791, pp. 959–967.
- Jerebko, A., Schmidt, G., Zhou, X., Bi, J., Anand, V., Liu, J., Schoenberg, S., Schmucking, I., Kiefer, B., Krishnan, A., 2007. Robust parametric modeling approach based on domain knowledge for computer aided detection of vertebrae column metastases in MRI. In: *Proceedings of Image Processing in Medical Imaging*, pp. 713–725.
- Klinder, T., Lorenz, C., von Berg, J., Dries, S., Bülow, T., Ostermann, J., 2007. Automated model-based rib cage segmentation and labeling in ct images. In: *Proceedings of Medical Imaging Computing and Computer Assisted Intervention*, vol. LNCS 4792, pp. 195–203.
- Klinder, T., Wolz, R., Lorenz, C., Franz, A., Ostermann, J., 2008. Spine segmentation using articulated shape models. In: *Proceedings of Medical Imaging Computing and Computer Assisted Intervention*, vol. LNCS 5241, pp. 227–234.
- Lorensen, W., Cline, H., 1987. Marching cubes: a high resolution 3D surface construction algorithm. *Computer Graphics* 21, 163–169.
- Lorenz, C., Krahstöver, N., 2000. Generation of point-based 3D statistical shape models for anatomical objects. *Computer Vision and Image Understanding* 77, 175–191.
- Mastmeyer, A., Engelke, K., Fuchs, C., Kalender, W., 2006. A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. *Medical Image Analysis* 10 (4), 560–577.
- Netsch, T., Rösch, P., Weese, J., van Muiswinkel, A., Desmedt, P., 2000. Grey value-based 3-D registration of functional MRI time-series: comparison of interpolation order and similarity measure. In: *Proceedings of SPIE Medical Imaging*, vol. 3979, pp. 1148–1159.
- Peng, Z., Zhong, J., Wee, W., Lee, J., 2005. Automated vertebra detection and segmentation from the whole spine mr images. In: *Proceedings of Engineering in Medicine and Biology*, pp. 2527–2530.
- Peters, J., Ecabert, O., Lorenz, C., von Berg, J., Walker, M., Ivanc, T., Vembar, M., Olszewski, M., Weese, J., 2008. Segmentation of the heart and major vascular structures in cardiovascular ct images. In: *Proceedings of SPIE Medical Imaging*, vol. 6914, pp. 691417–691429.
- Petit, Y., Aubina, C., Labelle, H., 2004. Spinal shape changes resulting from scoliotic spine surgical instrumentation expressed as intervertebral rotations and centers of rotation. *Journal of Biomechanics* 37, 173–180.
- Rangayyan, R., Deglint, H., Boag, G., 2006. Method for the automatic detection and segmentation of the spinal canal in computed tomographic images. *Journal of Electronic Imaging* 15 (3), 033007.
- Schmidt, S., Kappes, J., Bergtholdt, M., Pekar, V., Dries, S., Bystrov, D., Schnörr, C., 2007. Spine detection and labeling using a parts-based graphical model. In: *Proceedings of Image Processing in Medical Imaging*, vol. LNCS 4584, pp. 122–133.
- Schramm, H., Ecabert, O., Peters, J., Philomin, V., Weese, J., 2006. Towards fully automatic object detection and segmentation. In: *Proceedings of SPIE Medical Imaging*, vol. 6144, pp. 11–20.
- Shen, H., Litvin, A., Alvino, C., 2008. Localized priors for the precise segmentation of individual vertebrae from CT volume data. In: *Proceedings of Medical Imaging Computing and Computer Assisted Intervention*, vol. LNCS 5241, pp. 367–375.
- Timinger, H., Pekar, V., von Berg, J., Dietmeyer, K., Kaus, M., 2003. Integration of interactive corrections to model-based segmentation algorithms. In: *Proceedings of Bildverarbeitung Für die Medizin*, pp. 11–15.
- Vrtovec, T., Likar, B., Pernus, F., 2005. Automated curved planar reformation of 3D spine images. *Physics in Medicine and Biology* 50 (19), 4527–4540.
- Vrtovec, T., Ourselin, S., Gomes, L., Likar, B., Pernus, F., 2007. Automated generation of curved planar reformations from MR images of the spine. *Physics in Medicine and Biology* 52 (10), 2865–2878.
- Weese, J., Kaus, M., Lorenz, C., Lobregt, S., Truyen, R., Pekar, V., 2001. Shape constrained deformable models for 3D medical image segmentation. In: *Proceedings of Image Processing in Medical Imaging*, pp. 380–387.
- Yao, J., O'Connor, S., & Summers, R., 2006. Automated spinal column extraction and partitioning. In: *Proceedings of IEEE International Symposium Biomedical Imaging: Nano to Macro*, pp. 390–393.