

# A Framework for Automated Spine and Vertebrae Interpolation-Based Detection and Model-Based Segmentation

Robert Korez<sup>†\*</sup>, Bulat Ibragimov<sup>†</sup>, Boštjan Likar, Franjo Pernuš, and Tomaž Vrtovec

**Abstract**—Automated and semi-automated detection and segmentation of spinal and vertebral structures from computed tomography (CT) images is a challenging task due to a relatively high degree of anatomical complexity, presence of unclear boundaries and articulation of vertebrae with each other, as well as due to insufficient image spatial resolution, partial volume effects, presence of image artifacts, intensity variations and low signal-to-noise ratio. In this paper, we describe a novel framework for automated spine and vertebrae detection and segmentation from three-dimensional (3D) CT images. A novel optimization technique based on interpolation theory is applied to detect the location of the whole spine in the 3D image and, using the obtained location of the whole spine, to further detect the location of individual vertebrae within the spinal column. The obtained vertebra detection results represent a robust and accurate initialization for the subsequent segmentation of individual vertebrae, which is performed by an improved shape-constrained deformable model approach. The framework was evaluated on two publicly available CT spine image databases of 50 lumbar and 170 thoracolumbar vertebrae. Quantitative comparison against corresponding reference vertebra segmentations yielded an overall mean centroid-to-centroid distance of 1.1 mm and Dice coefficient of 83.6% for vertebra detection, and an overall mean symmetric surface distance of 0.3 mm and Dice coefficient of 94.6% for vertebra segmentation. The results indicate that by applying the proposed automated detection and segmentation framework, vertebrae can be successfully detected and accurately segmented in 3D from CT spine images.

**Index Terms**—Object detection, image segmentation, interpolation theory, deformable models, spine, vertebra, computed tomography.

## I. INTRODUCTION

ACCURATE and robust segmentation of vertebrae from medical images is an essential tool in many clinical applications of spinal imaging. Knowledge of the detailed shape of individual vertebrae can considerably aid early diagnosis, surgical planning and follow-up assessment of a number of spinal pathologies, such as degenerative disorders, spinal deformities (e.g. scoliosis), trauma and tumors. Two-dimensional (2D) and three-dimensional (3D) images of the

spinal anatomy are usually acquired using plain radiography or computed tomography (CT) to obtain a highly detailed representation of bony structures, and magnetic resonance (MR) to obtain a high soft tissue contrast representation. Currently, conventional radiography is frequently used for screening purposes (e.g. to check for fractures), but since superimposition of other (bony) structures may prevent the identification of spinal structures of interest and therefore conceal eventual pathologies, an accurate and authoritative diagnosis can only be established by correlating the findings of a physical exam to CT or MR images. The advantage of MR is its ability to visualize the spinal anatomy without using ionizing radiation, however, in the resulting image there are no sharp margins between many tissue types, including edges within vertebrae. On the other hand, continuing advances in CT, such as faster scanning times, higher image resolution and decreased radiation exposure, make CT the modality of choice for assessing the spine and vertebrae.

In recent years, several automated and semi-automated methods focusing on vertebra segmentation have been developed for CT images. Leventon et al. [1] introduced a segmentation process that used prior shape information to estimate the maximum a posteriori high-dimensional surface, for which the zero level set converged on vertebra boundaries. Kim and Kim [2] proposed a fully automated method that was based on constructing 3D fences to separate vertebrae from valley-emphasized Gaussian images, and then the region growing algorithm was applied within the constructed 3D fences to obtain the final segmentation. Klinder et al. [3] progressively adapted tube-shaped segments to extract the spine curve, performed vertebra detection on curved-planar reformatted images using the generalized Hough transform, identified vertebrae by rigid registration of appearance models to the detected candidates, and obtained the final segmentation by adapting shape-constrained deformable models [4] to individual vertebrae. Kadoury et al. [5], [6] built an articulated shape manifold by embedding the vertebrae from a database into a low-dimensional sub-space, and applied the Markov random field optimization to infer between the shape manifold and the shape of an unseen vertebra. A level set vertebra segmentation framework was described by Lim et al. [7], who extracted local geometrical features using the Willmore flow and prior shape knowledge using kernel density estimation. Ma and Lu [8] introduced a hierarchical coarse-to-fine deformable surface-based segmentation that relied on response maps of a trained bone structure edge detection algorithm. A statistical

Manuscript received November 28, 2014. This work was supported by the Slovenian Research Agency under grants P2-0232 and L2-4072. <sup>†</sup>Both authors contributed equally to this work. Asterisk indicates corresponding author.

R. Korez and B. Ibragimov are with the Laboratory of Imaging Technologies, Faculty of Electrical Engineering, University of Ljubljana, Slovenia (e-mail: robert.korez@fe.uni-lj.si, bulat.ibragimov@fe.uni-lj.si).

B. Likar, F. Pernuš and T. Vrtovec are with the Laboratory of Imaging Technologies, Faculty of Electrical Engineering, University of Ljubljana, Slovenia (e-mail: boštjan.likar@fe.uni-lj.si, franjo.pernus@fe.uni-lj.si, tomaz.vrtovec@fe.uni-lj.si).

multi-vertebral model that treated the shape and pose of each vertebra independently was described by Rasoulian et al. [9], who also proposed a novel iterative expectation maximization registration technique to align the multi-vertebral model to CT spine images. Recently, Ibragimov et al. [10] presented a segmentation framework, in which a novel landmark-based shape representation of vertebrae was combined with game-theoretic landmark detection augmented by the strategy dominance concept.

Irrespective of the applied approach, segmentation of vertebrae strongly depends on a relatively accurate and robust initialization, which can be considered as a vertebra detection task. However, this task is hampered by the fact that neighboring vertebrae are of similar shape, and therefore they cannot be appropriately distinguished without observing the whole spine or a larger spine section, and without observing their neighboring structures, such as the sacrum, ribs or specific internal organs. Semi-automated vertebra detection methods require manual identification of such an anchor structure [11], or manual annotation of the observed vertebra in image of one modality followed by its propagation to images of different modalities of the same subject [12]. On the other hand, automated vertebra detection methods split the task into the detection of the whole spine and of candidate vertebrae, which is often based on inter-vertebral spatial relationships. Feature-based methods treat vertebra centers or all image voxels belonging to the observed vertebra as points of interest, and model them by intensity-based features that are combined into landmark detectors by machine learning classification techniques [13]–[17], using either full or marginal space learning [18]. Alternatively, detection can also rely on vertebral shape, often described by active shape and appearance models [19]–[21], the generalized Hough transform [3], [22] or features based on anatomical symmetry and cylindricity of the vertebral body [23]. Although most approaches assume that the whole spine or a predefined part of the spine is present in the image, several frameworks also deal with more challenging images of arbitrary fields of view [3], [15], [16].

Automated and semi-automated detection and segmentation of vertebrae from CT images is still a challenging task due to a relatively high degree of anatomical complexity (i.e. vertebrae consisting of the vertebral body, pedicles, laminae and spinous process), presence of unclear boundaries (e.g. at the location of basivertebral veins), and articulation of vertebrae with each other (e.g. due to the degenerative disc and/or joint disease) (Fig. 1). Moreover, the detection and segmentation processes may also be hampered by insufficient image spatial resolution, partial volume effects, presence of image artifacts, intensity variations and low signal-to-noise ratio. Nevertheless, detection and segmentation of any object of interest can always be represented as a specific optimization problem, which is usually non-analytically defined and often computationally demanding. In the worst case, when the objective function that is being optimized is not smooth, i.e. its values computed for neighboring points do not give enough useful information about the target point, brute force algorithms can be used. On the other hand, in the case of smooth objective functions, computationally more efficient approaches based on multi-

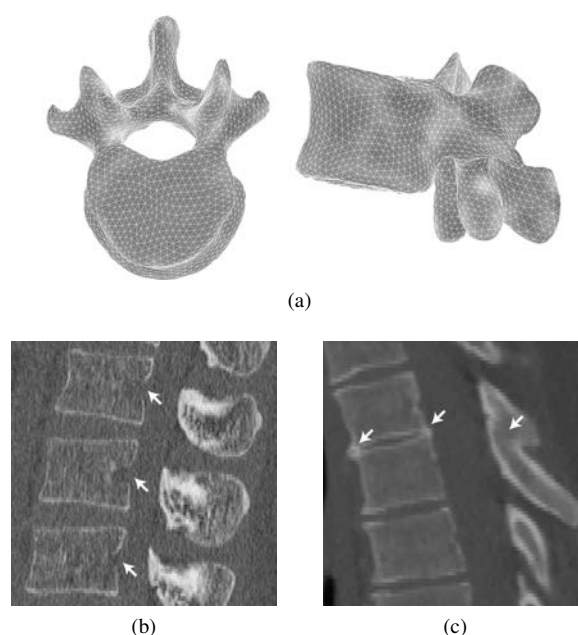


Fig. 1. Detection and segmentation of vertebrae from CT images is a challenging task, for example, due to (a) high anatomical complexity (shown by a rendered 3D face-vertex surface mesh), (b) presence of unclear boundaries at the location of basivertebral veins (indicated by arrows in a sagittal cross-section), and (c) articulation of vertebrae with each other due to the degenerative disc and joint disease (indicated by arrows in a sagittal cross-section).

scale pyramids [24] or graph search algorithms [25] can be applied. However, finding the global optimum of the objective function does not always guarantee a correct detection and segmentation, especially if the optimization problem is ill-posed, which may cause that the optimal detection or segmentation result is not associated with the global but with a local optimum of the devised problem. Such lack of correspondence is clearly observed in vertebra detection, where neighboring vertebrae are of similar appearance and shape. As standard optimization techniques, such as the Powell's optimizer [26] and covariance matrix adaptation evolution strategy (CMA-ES) [27], detect only a single local optimum, it can be concluded that there is a need for an optimizer that will consider not only the global optimum but also local optima of the problem that represent candidate locations for each vertebra, and that can be further used for accurate and robust vertebra detection and segmentation.

To overcome the above mentioned limitations and ensure the detection of all local optima that correspond to candidate locations of vertebrae, we propose a novel approach that is based on interpolation theory, which predicts the location of the object of interest by computing a specific detector response for a sparse set of points, i.e. possible transformations of the object of interest. The obtained vertebra detection results represent a robust and accurate initialization for segmentation of individual vertebrae, which is performed by an improved shape-constrained deformable model approach [4], based on a vertebra surface mesh deformation technique that moves mesh vertices to their optimal locations while preserving the underlying vertebral shape. The automated interpolation-based vertebra detection and the improved model-based vertebra

segmentation are combined into a novel framework that was evaluated on two publicly available CT spine image databases of lumbar and thoracolumbar vertebrae.

## II. SPINE AND VERTEBRA DETECTION

Detection of the object of interest (e.g. spine, vertebra) in an unknown image  $I$  can be, in general, performed by optimizing an objective function  $f = f(\mathbf{p}, I)$ , which maps the model of the object of interest, represented by a vector  $\mathbf{p}$  of  $n$  parameters describing its geometric properties (e.g. position, rotation, scaling etc.), to image  $I$  in  $n$ -dimensional parameter space  $\mathbb{R}^n$ . In practice, optimization is performed on a bounded domain  $\Omega^n$ ;  $\Omega^n \subset \mathbb{R}^n$ ;  $\mathbf{p} \in \Omega^n$ , as a plausible range of parameters can be usually defined from the imaging system and/or nature of imaged objects. The most straightforward but computationally demanding approach for optimizing  $f$  is to apply brute force to compute its response for every  $\mathbf{p} \in \Omega^n$ , while alternative approaches [26], [27] compute the response of  $f$  in the neighborhood of an initial  $\mathbf{p}'$ ;  $\mathbf{p}' \in \Omega^n$ , and then predict and iteratively move towards the optimal  $\mathbf{p}^*$ . Although such optimization is computationally less demanding and may, especially if  $\mathbf{p}'$  is relatively close to  $\mathbf{p}^*$ , lead to the globally optimal solution, it may still converge to a local optimum because only a part of  $\Omega^n$  is observed. To obtain the globally optimal solution, the whole  $\Omega^n$  has to be taken into consideration, however, the actual computations of  $f$  have to be minimized to reduce computational complexity. To achieve such a setting, we propose a novel optimization scheme that is based on *interpolation theory* [28], and we apply it for spine and vertebra detection in 3D images. Depending on  $f$ , the scheme searches for optima in the form of either minima or maxima, however, in the following text we refer to them as maxima and to the problem as maximization.

### A. Interpolation Theory

The main concept of interpolation theory [28], often referred to as approximation theory, is to select the interpolation function, in our case  $\bar{f} = \bar{f}(\mathbf{p}, I)$ , from a given class of functions so that its response for a limited set of nodes  $\mathcal{P}$ ;  $\mathcal{P} \subset \Omega^n$ , which form the interpolation grid over the observed domain  $\Omega^n$ , passes through the known response of the objective function  $f = f(\mathbf{p}, I)$  at the same nodes in  $\mathcal{P}$ , and that its response is similar to the response of  $f$  for  $\Omega^n \setminus \mathcal{P}$ :

$$\min_{\bar{f}} \sum_{\mathbf{p} \in \Omega^n} |f(\mathbf{p}, I) - \bar{f}(\mathbf{p}, I)|; \quad \forall \mathbf{p} \in \mathcal{P} : f(\mathbf{p}, I) = \bar{f}(\mathbf{p}, I), \quad (1)$$

assuming that  $f$  is smooth and its values are similar for similar parameters  $\mathbf{p}$ ;  $\mathbf{p} \in \Omega^n$ . If  $\bar{f}$  and  $\mathcal{P}$  are selected properly, the difference between  $f$  and  $\bar{f}$ , i.e. the interpolation error, is small, and the global maximum of  $\bar{f}$  corresponds to the global maximum of  $f$  at the same  $\mathbf{p}^*$ ;  $\bar{f}_{\max} = \bar{f}(\mathbf{p}^*, I) \approx f_{\max} = f(\mathbf{p}^*, I)$ .

In the case of *linear interpolation*,  $\bar{f}$  is a sequence of piecewise linear polynomials (i.e. line segments) connecting nodes in  $\mathcal{P}$ . However, local maxima of  $\bar{f}$  are always located at  $\mathbf{p} \in \mathcal{P}$ , therefore obtaining the global maximum of  $\bar{f}$  is only

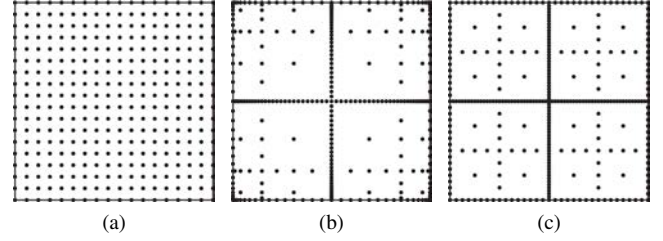


Fig. 2. The (a) equidistant, (b) Chebyshev and (c) Clenshaw-Curtis interpolation grids, defined for two dimensions ( $n = 2$ ) with a similar number of nodes over domain  $\Omega^2$ .

possible if  $\mathcal{P}$  at random contains  $\mathbf{p}^*$ . In the case of *polynomial interpolation*,  $\bar{f}$  is described by a polynomial of degree  $|\mathcal{P}| - 1 > 2$  that passes through all nodes in  $\mathcal{P}$ , where  $|\mathcal{P}|$  is the number of nodes. In most cases, polynomial interpolation provides a better approximation for  $f$  than linear interpolation, however, its general disadvantage is the sensitivity to the selection of  $\mathcal{P}$ . In the case of *spline interpolation*, nodes in  $\mathcal{P}$  are connected by low-degree piecewise polynomials (i.e. splines) that smoothly transit from one to another. Spline interpolation is less sensitive to the selection of  $\mathcal{P}$  than polynomial interpolation. As polynomial and spline interpolation usually result in the smallest interpolation error, we select polynomials and splines for the interpolation function  $\bar{f}$  that is used in our framework to detect the spine and vertebrae in 3D images.

To minimize the interpolation error, the nodes in set  $\mathcal{P}$  that form the interpolation grid over the observed domain  $\Omega^n$  have to be selected according to the selected  $\bar{f}$ . Polynomial interpolation is very sensitive to the selection of  $\mathcal{P}$ , as  $\bar{f}$  can oscillate if  $\mathcal{P}$  is not adequate to approximate  $f$ . Moreover, by increasing the size of  $\mathcal{P}$ , these oscillations may be intensified, which is called the Runge's phenomenon [28]. Nevertheless, according to the Weierstrass approximation theorem [28], any smooth function can be approximated as closely as desired by a polynomial function, if  $\mathcal{P}$  is optimally defined. On the other hand, the Runge's phenomenon is avoided in the case of spline interpolation, as low-degree piecewise polynomials do not oscillate as much as high-degree polynomials. If  $\zeta$  is the maximal distance between any two adjacent nodes in  $\mathcal{P}$ , the interpolation error is proportional to  $\zeta^2$  for linear splines and to  $\zeta^4$  for cubic splines. To minimize the interpolation error,  $\zeta$  should be therefore minimized, which implies that the *equidistant grid* (Fig. 2a) is optimal in the case of spline interpolation. In the case of polynomial interpolation, the interpolation error as well as the effects of the Runge's phenomenon can be considerably reduced by the *Chebyshev grid* (Fig. 2b) [28], where the nodes are distributed more densely towards the limits of  $\Omega^n$ . An expansion of the Chebyshev grid is the *Clenshaw-Curtis grid* (Fig. 2c) [28], and although it is associated with a larger interpolation error, its nodes are distributed more densely in the center of  $\Omega^n$ , where the maximum of  $f$  is usually located. This property can be particularly advantageous in the case of medical image analysis, where  $\Omega^n$  usually represents the volume of interest that is centered at and spans around the object of interest.

Irrespectively of the selection of  $\bar{f}$  and  $\mathcal{P}$ , the interpolation

of high-dimensional functions is computationally challenging, as the number of nodes in the interpolation grid increases exponentially with the increasing number of dimensions  $n$ . To reduce the computational complexity, various strategies can be applied. For example, iterative optimization methods [29] reduce the complexity by performing optimization dimension-wise, while in the case of dimension-adaptive optimization, the number of nodes per dimension decreases with the increasing number of dimensions [30] or a certain number of nodes is randomly distributed over the observed domain [31]. In our framework, to reduce the computational complexity we propose the following interpolation-based dimension-wise algorithm that determines a set of local maxima of the objective function  $f$  with known response at nodes of the given interpolation grid  $\mathcal{P}$ , defined on domain  $\Omega^n$  of  $n$  dimensions, assuming a given interpolation function  $\bar{f}$  (Fig. 3):

- 1) Initialize set  $\mathcal{S} = \{\}$  of optimized dimensions and set  $\bar{\mathcal{S}} = \{1, 2, \dots, n\}$  of the remaining dimensions.
- 2) Interpolate  $f$  with  $\bar{f}$  over  $\mathcal{P}$  against the first  $d \ll n$  dimensions.
- 3) Find  $L$  local maxima of  $\bar{f}$  and define a corresponding set  $\mathcal{L}$  of their locations;  $\mathcal{L} = \{p_1, p_2, \dots, p_L\}$ , then insert the observed dimensions into  $\mathcal{S}$ ;  $\mathcal{S} \leftarrow \mathcal{S} \cup \{1, 2, \dots, d\}$ , and remove them from  $\bar{\mathcal{S}}$ ;  $\bar{\mathcal{S}} \leftarrow \bar{\mathcal{S}} \setminus \{1, 2, \dots, d\}$ .
- 4) For each  $l$ -th element  $p_l$  in  $\mathcal{L}$ , i.e.  $l = 1, 2, \dots, L$ :
  - a) If  $\bar{\mathcal{S}} \neq \{\}$ , interpolate  $f$  with  $\bar{f}$  over  $\mathcal{P}$  against the next  $d$  available dimensions by considering dimensions from  $\mathcal{S}$  to be optimized.
  - b) Find the global maximum  $\bar{f}_{\max}$  of  $\bar{f}$  and replace the  $l$ -th element in  $\mathcal{L}$  with the location  $p_{\max}$  of  $\bar{f}_{\max}$ ;  $p_l \leftarrow p_{\max}$ , then insert the observed dimensions into  $\mathcal{S}$  and remove them from  $\bar{\mathcal{S}}$ .
  - c) If  $\bar{\mathcal{S}} = \{\}$  return to 4), otherwise return to 4a).
- 5) The resulting set  $\mathcal{L} = \{p_1, p_2, \dots, p_L\}$  represents the approximate locations of  $L$  local maxima of  $f$  across  $n$  dimensions.

For the proposed algorithm, the optimization dimensions  $n$  have to be ordered according to the decreasing number of corresponding interpolation nodes, while the number of dimensions  $d$  observed in each iteration may vary through different iterations. Nevertheless, by taking into account only  $d$  dimensions at once, the computational complexity of the high-dimensional optimization problem is reduced. The algorithm was conceived by taking into account the properties of spine and vertebra detection. As the standard spine image acquisition procedure limits the orientation of the imaged subject, the relative inclination of the spinal column in the image is more constant than its location, and similarly the orientation of each individual vertebra is restricted by the location of neighboring vertebrae. By considering such setting, it is therefore recommended to associate dimensions related to translation with a larger number of nodes than dimensions related to rotation.

In the case no prior information is available about which dimensions are more important, we propose to use an accelerated version of the dimension-adaptive optimization algorithm [31],

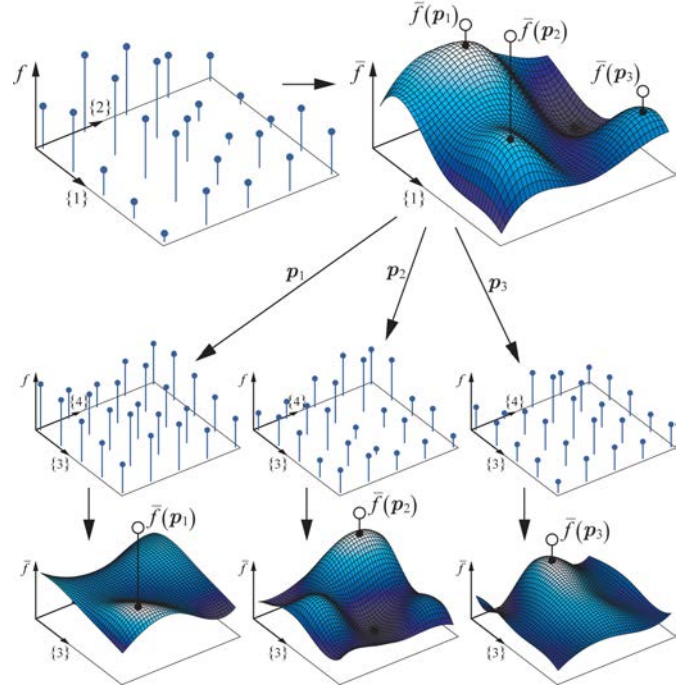


Fig. 3. An illustration of the proposed interpolation-based dimension-wise algorithm for the reduction of computational complexity. For the first  $d$  out of  $n$  dimensions (e.g.  $d = 2$ ,  $n = 4$ ), the objective function  $f$  is evaluated at nodes of the interpolation grid on domain  $\Omega^d$ , and then interpolated by  $\bar{f}$ . Each of the resulting  $L$  local maxima (e.g.  $L = 3$ ) at location  $p_l$  initializes the interpolation against the next  $d$  dimensions, and the location of the resulting global maximum is assigned to replace  $p_l$ . This procedure is continued until all  $n$  dimensions are analyzed, and location  $p_l$  that corresponds to the largest of the resulting values is selected as the location of the global maximum of  $f$  over domain  $\Omega^n$ .

[32] to reduce the computational complexity of the performed interpolation. The algorithm performs the interpolation on a sparse grid by adaptively increasing the number of nodes for dimensions with the largest interpolation error until the estimated interpolation error is smaller than a predefined threshold or until a predefined number of nodes is used. The dimension-adaptive algorithm is an option for optimization of high-dimensional problems, where the number of dimensions does not allow to reliably use the proposed dimension-wise algorithm.

### B. Interpolation-Based Spine and Vertebra Detection

Let  $\mathcal{M} = \{\mathcal{V}, \mathcal{F}\} = \{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_Q\}$  represent the 3D face-vertex surface mesh that corresponds to the mean shape model of the observed spinal region of  $Q$  vertebrae, where  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$  represents the 3D face-vertex surface mesh of the  $q$ -th vertebra, and  $\mathcal{V} = \{\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_Q\}$  and  $\mathcal{F} = \{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_Q\}$  are, respectively, sets of mesh vertices and mesh faces. The objective function  $f$  that measures the agreement between mesh  $\mathcal{M}$  (or  $\mathcal{M}_q$ ) and image  $I$  is defined as [33]:

$$f(p, I) = \sum_{i=1}^{|\mathcal{F}|} \langle g_H(c_i|p, I), n(c_i|p) \rangle, \quad (2)$$

where  $\langle \cdot, \cdot \rangle$  denotes the dot product,  $|\mathcal{F}|$  (or  $|\mathcal{F}_q|$ ) is the number of mesh faces,  $c_i$  is the location of the centroid of the  $i$ -th mesh

face (i.e. its geometric center of mass),  $\mathbf{g}_H(\mathbf{c}_i, I)$  is the Haar-like gradient [34] of image  $I$  at  $\mathbf{c}_i$ , and  $\mathbf{n}(\mathbf{c}_i)$  is the face normal at  $\mathbf{c}_i$ . Notation  $\mathbf{c}_i|_p$  represents the geometric transformation of centroid  $\mathbf{c}_i$  according to vector  $\mathbf{p}$  of  $n$  transformation parameters on domain  $\Omega^n$ . If mesh  $\mathcal{M}$  (or  $\mathcal{M}_q$ ) is correctly aligned with the observed spinal region (or vertebra) in image  $I$ , then mesh normals pointing outwards of the mesh are in maximal agreement with the Haar-like gradients pointing outwards of spinal structures, resulting in the maximum of the objective function  $f$ . As Haar-like gradients are described as a difference between two neighboring cuboids instead of voxels, they are not sensitive to local intensity fluctuations [34].

Although the detection of the observed spinal region can be performed by finding the global maximum of  $f$  that describes the maximal agreement of mesh  $\mathcal{M}$  and image  $I$  on domain  $\Omega^n$  (Eq. 2), it is not adequate for robust detection of individual vertebral levels due to natural anatomical differences in the size and/or curvature of the spine. If the detection error is usually low in sagittal and coronal directions (i.e. anatomical left-to-right and anterior-to-posterior directions, respectively), the detection error in the axial direction (i.e. anatomical cranial-to-caudal direction) can be considerable, as neighboring vertebrae are of similar appearance and shape, and therefore the mean shape model of the observed spinal region can be shifted along the axial direction for one or even two vertebral levels. To correctly detect individual vertebrae,  $\Omega^n$  has to be defined separately for each observed vertebral level so that it also captures its neighboring levels. As a result, however, several local maxima of  $f$  exist on  $\Omega^n$ , usually corresponding to all vertebral levels captured within  $\Omega^n$ . Moreover, the global maximum of  $f$  may not correspond to the location of the observed vertebral level. It is therefore required to extract all (or several) local maxima of  $f$  for the observed vertebral level and consider them in combination with local maxima extracted for neighboring vertebral levels. Let set  $\mathcal{L} = \{\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_Q\}$  represent the locations of local maxima of  $f$  for all  $Q$  vertebral levels, where  $\mathcal{L}_q = \{\mathbf{p}_{q,1}, \mathbf{p}_{q,2}, \dots, \mathbf{p}_{q,L}\}$  is the set of  $L$  locations of local maxima for the  $q$ -th vertebral level, and each location  $\mathbf{p} \in \mathcal{L}$  is represented by a vector of  $n$  transformation parameters on domain  $\Omega^n$ . To detect the location of each vertebral level, we find the optimal path  $\mathcal{R}^*$  that passes through  $Q$  locations, each corresponding to a different vertebral level:

$$\mathcal{R}^* = \arg \max_{\mathcal{R}} \left( \sum_{q=1}^{Q-1} f_s(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}}, I) g(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}}) \right), \quad (3)$$

where set  $\mathcal{R} = \{r_1, r_2, \dots, r_Q\}$  is an arbitrary combination of indices of local maxima, each corresponding to a different vertebral level so that  $\forall r_q \in \mathcal{R} : \exists! \mathbf{p}_{q,r_q} \in \mathcal{L}_q$ , i.e. exactly one index corresponds to each set  $\mathcal{L}_q$ ;  $q = 1, 2, \dots, Q$ . Function  $f_s(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}}, I) = f(\mathbf{p}_{q,r_q}, I) + f(\mathbf{p}_{q+1,r_{q+1}}, I)$  measures the appearance (related to image  $I$ ), while function  $g(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}})$  measures the agreement in shape of locations  $\mathbf{p}_{q,r_q}$  and  $\mathbf{p}_{q+1,r_{q+1}}$  corresponding to neighboring  $q$ -th

and  $(q+1)$ -th vertebral levels, respectively:

$$g(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}}) = \prod_{d=1}^n G(\Delta^d(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}}), \mu_{q,q+1}^d, \sigma_{q,q+1}^d), \quad (4)$$

where  $\Delta^d(\mathbf{p}_{q,r_q}, \mathbf{p}_{q+1,r_{q+1}})$  is the difference in  $d$ -th dimension between locations  $\mathbf{p}_{q,r_q}$  and  $\mathbf{p}_{q+1,r_{q+1}}$ . Function  $G(\Delta^d(\cdot, \cdot), \mu_{q,q+1}^d, \sigma_{q,q+1}^d) = \exp(-\frac{1}{2}((\Delta^d(\cdot, \cdot) - \mu_{q,q+1}^d)/\sigma_{q,q+1}^d)^2)$  is the Gaussian function of  $\Delta^d(\cdot, \cdot)$ , where  $\mu_{q,q+1}^d$  and  $\sigma_{q,q+1}^d$  are, respectively, the estimated mean and standard deviation (SD) of  $\Delta^d(\cdot, \cdot)$  that can be obtained, for example, from a training repository. The obtained optimal path  $\mathcal{R}^* = \{r_1^*, r_2^*, \dots, r_Q^*\}$  is used to determine the set of locations of local maxima  $\mathcal{L}^* = \{\mathbf{p}_1^*, \mathbf{p}_2^*, \dots, \mathbf{p}_Q^*\} = \{\mathbf{p}_{q,r_q^*}, \mathbf{p}_{q,r_q^*}, \dots, \mathbf{p}_{q,r_q^*}\}$ , which contains the optimal location  $\mathbf{p}_q^*$  of each vertebral level in the observed spinal region;  $\forall r_q^* \in \mathcal{R}^* : \exists! \mathbf{p}_q^* = \mathbf{p}_{q,r_q^*} \in \mathcal{L}_q \cap \mathcal{L}^*$ .

### III. VERTEBRA SEGMENTATION

Model-based segmentation of the object of interest is, in general, performed by initializing a model of the object of interest relatively close to the observed anatomical structure in the image, and then deforming the model so that it adapts to object boundaries. Let  $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_Q\}$  represent the spine and vertebra detection repository that consists of  $Q$  correctly detected vertebrae in the 3D spine image  $I$  (Sec. II), where  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$  is the 3D face-vertex surface mesh of the  $q$ -th vertebra that represents the model of the vertebra as the observed anatomical structure. The results of the interpolation-based spine and vertebra detection therefore initialize vertebra segmentation, which is achieved by adapting  $\mathcal{M}_q$  to vertebra boundaries in image  $I$ . For this purpose, we propose an iterative segmentation framework that is built upon the existing *shape-constrained deformable model approach* [4]. In contrast to the original approach, we introduced several modifications that considerably improved the accuracy of the resulting segmentation. The following two steps are executed in each iteration of the proposed segmentation framework:

- *boundary detection* searches for characteristic vertebra boundaries that correspond to mesh face centroids;
- *mesh deformation* attracts the mesh to vertebra boundaries and penalizes the deviations of the mesh from the training repository.

Both steps of this iterative framework are described in detail in the following subsections.

#### A. Boundary Detection

In each  $p$ -th iteration of the proposed segmentation framework, each mesh face centroid  $\mathbf{c}_i$ ;  $i = 1, 2, \dots, |\mathcal{F}_q|$  is displaced along its corresponding mesh face normal  $\mathbf{n}(\mathbf{c}_i)$  to find a new centroid  $\mathbf{c}_i^*$ :

$$\mathbf{c}_i^* = \mathbf{c}_i + \delta j_i^* \mathbf{n}(\mathbf{c}_i), \quad (5)$$



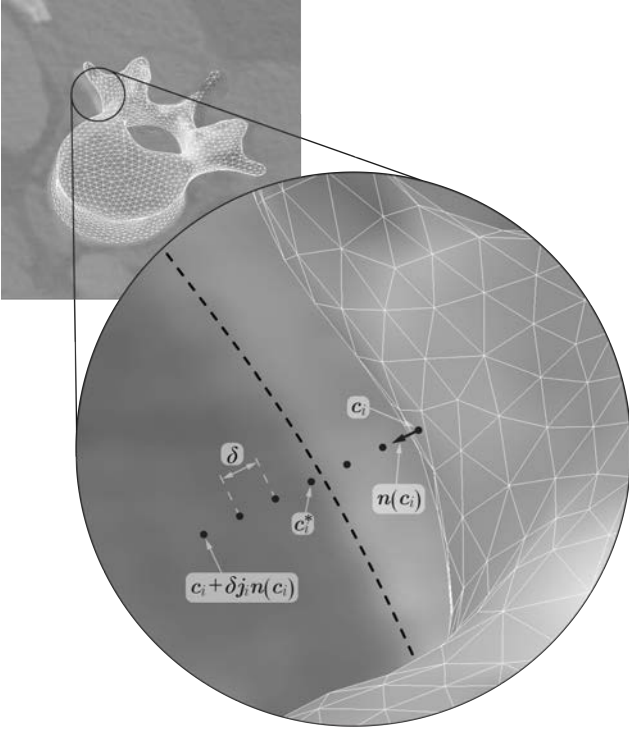


Fig. 4. Each  $i$ -th face centroid  $c_i$  of the 3D face-vertex surface mesh of the observed vertebra is displaced for  $\delta j_i$  along the sampling parcel in the direction of its face normal  $\mathbf{n}(c_i)$ . In an iterative framework, where the length of the sampling parcel is gradually reduced, each centroid moves to the location  $c'_i$  that best corresponds to vertebra boundaries (dashed curve).

where  $\delta$  is the length of the unit displacement, and  $j_i^*$  is an element from set  $\mathcal{J}$ ;  $j_i^* \in \mathcal{J}$ , that represents the search profile along  $\mathbf{n}(c_i)$ , called the sampling parcel (Fig. 4):

$$\mathcal{U} = \{-j, -j+1, \dots, j-1, j\}; \quad j = J - \lceil \alpha p \rceil, \quad (6)$$

where  $\lceil \cdot \rceil$  denotes the ceiling operator (i.e.  $\lceil a \rceil$  is the smallest integer not less than  $a$ ), and  $\alpha$  regulates the size of the sampling parcel that decreases from  $2(J - \lceil \alpha \rceil) + 1$  at initial iteration  $p = 1$  to  $2(J - \lceil \alpha P \rceil) + 1$  at final iteration  $p = P$ . The element  $j_i^*$  that defines the location of  $c'_i$  is determined by detecting vertebra boundaries:

$$j_i^* = \arg \max_{j_i \in \mathcal{J}} \{F(c_i, c_i + \delta j_i \mathbf{n}(c_i)) - D \delta^2 j_i^2\}, \quad (7)$$

where  $c'_i = c_i + \delta j_i \mathbf{n}(c_i)$  is the candidate location for  $c'_i$  (Eq. 5), and parameter  $D$  controls the tradeoff between the response of the boundary detection operator  $F$  (Eq. 8) and the distance from  $c_i$  to  $c'_i$ . Apart from reducing the length of the sampling parcel, within which new mesh face centroids are searched for, in comparison to the original approach [4] we also propose an improved boundary detection operator  $F$  that is based on image intensity gradients, weighted by two image appearance operators:

$$F(c_i, c'_i) = \frac{g_{\max} (g_{\max} + \|\mathbf{g}_{w1}(c'_i)\|)}{g_{\max}^2 + \|\mathbf{g}_{w1}(c'_i)\|^2} \langle \mathbf{n}(c_i), \mathbf{g}_{w2}(c'_i) \rangle, \quad (8)$$

where  $\|\cdot\|$  denotes the vector norm,  $g_{\max}$  is the estimated mean magnitude of intensity gradients at vertebra boundaries that is

used to suppress the weighted gradients, which may occur if the gradient magnitude at vertebra boundaries is considerably smaller than of another object in its neighborhood (e.g. disc implants, pedicle screws), and  $\mathbf{g}_{w1}$  and  $\mathbf{g}_{w2}$  are the two image appearance operators at candidate mesh centroid location  $c'_i$ :

$$\mathbf{g}_{w1}(c'_i) = (1 + C(c'_i) + R(c'_i)) \mathbf{g}(c'_i), \quad (9)$$

$$\mathbf{g}_{w2}(c'_i) = (1 + R(c'_i)) \mathbf{g}(c'_i), \quad (10)$$

where  $\mathbf{g}(c'_i)$  is the intensity gradient at  $c'_i$ ,  $C(c'_i) \in [0, 1]$  is the continuous response to the Canny edge operator [35], and  $R(c'_i) \in [-1, 1]$  is the continuous response to the random forest regression model  $R = R(X, \mathbf{y})$  [36] for a given feature matrix  $X$  and associated label vector  $\mathbf{y}$  that form the training data. Let  $\{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_M\}$  represent the training repository of  $M$  aligned 3D face-vertex surface meshes of the observed  $q$ -th vertebral level, where each mesh  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$ ;  $m = 1, 2, \dots, M$ , consists of  $|\mathcal{V}_q|$  vertices and  $|\mathcal{F}_q|$  faces. For each  $\mathcal{M}_q$ , an appearance feature matrix  ${}^m X_k$  of size  $2 \times |\mathcal{F}_q|$  is first generated for each element  $k$  from set  $\mathcal{K} = \{-K, -K+1, \dots, K-1, K\}$ ;  $k \in \mathcal{K}$ :

$${}^m X_k = \left[ I({}^m c'_i), \langle \mathbf{g}({}^m c'_i), \mathbf{n}({}^m c_i) \rangle \right]_{i=1,2,\dots,|\mathcal{F}_q|}^T, \quad (11)$$

where  $I({}^m c'_i)$  and  $\mathbf{g}({}^m c'_i)$  are, respectively, the image intensity and intensity gradient at centroid  ${}^m c'_i = {}^m c_i + \delta k \mathbf{n}({}^m c_i)$ , which represents the displacement by  $\delta k$  of centroid  ${}^m c_i$  along its corresponding face normal  $\mathbf{n}({}^m c_i)$ . The feature matrix  $X$  of size  $|\mathcal{K}| |\mathcal{F}_q| \times 2$  and the associated label vector  $\mathbf{y}$  of size  $|\mathcal{K}| |\mathcal{F}_q| \times 1$  are then defined as:

$$X = \left[ X_{-K}, X_{-K+1}, \dots, X_{K-1}, X_K \right]^T, \quad (12)$$

$$\mathbf{y} = \left[ e(-K), e(-K+1), \dots, e(K-1), e(K) \right]^T, \quad (13)$$

where matrix  $X_k$ ;  $k \in \mathcal{K}$ , is the mean matrix of  $\{X_k, {}^2 X_k, \dots, {}^M X_k\}$ , and  $e(k)$ ;  $k \in \mathcal{K}$ , represents an all-equal  $(-2|k|/K+1)$ -value vector of size  $1 \times |\mathcal{F}_q|$  so that  $e(0)$  is a vector of all-ones and  $e(-K) = e(K)$  is a vector of all-minus-ones (i.e. corresponding to the limits of the response). The resulting random forest regression model  $R(c'_i)$  therefore generates a continuous prediction for a query candidate centroid  $c'_i$ , with  $R(c'_i) = 1$  indicating that the location of  $c'_i$  corresponds to the appearance characteristics of vertebra boundaries, and  $R(c'_i) = -1$  indicating the opposite.

### B. Mesh Deformation

Once the new mesh face centroids  $c_i^*$  are detected, mesh  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$  is reconfigured in each  $p$ -th iteration of the proposed segmentation framework by minimizing the weighted sum  $E$  of energy terms:

$$\min_{\mathcal{M}_q} \{E\} = \min_{\mathcal{M}_q} \{E_{\text{ext}} + \beta E_{\text{int}}\}, \quad (14)$$

where  $\beta$  is the weighting parameter. The external energy  $E_{\text{ext}}$  (Eq. 14) attracts mesh  $\mathcal{M}_q$  to new face centroids  $c_i^*$ ;  $i = 1, 2, \dots, |\mathcal{F}_q|$  (Eq. 5), that are located on vertebra boundaries:

$$E_{\text{ext}} = \sum_{i=1}^{|\mathcal{F}_q|} w_i^* \left\| \text{proj} \left( c_i^* - c_i, \frac{\mathbf{g}_{w1}(c_i^*)}{\|\mathbf{g}_{w1}(c_i^*)\|} \right) \right\|^2, \quad (15)$$

where  $\text{proj}$  denotes the vector projection,  $|\mathcal{F}_q|$  is the number of mesh faces,  $\mathbf{g}_{w1}$  is the image appearance operator (Eq. 9), and  $w_i^*$ ;  $i = 1, 2, \dots, |\mathcal{F}_q|$ , are weights that are defined according to the obtained  $j_i^*$  (Eq. 7) to give a greater influence to more promising centroid locations:

$$w_i^* = \max \left\{ 0, F(\mathbf{c}_i, \mathbf{c}_i^*) - D \delta^2 j_i^{*2} \right\}. \quad (16)$$

By projecting  $\mathbf{c}_i^* - \mathbf{c}_i$  onto  $\mathbf{g}_{w1}(\mathbf{c}_i^*)$ , the external energy becomes invariant to movements of mesh centroids within the plane perpendicular to the intensity gradient, which in turn reduces the risk of mesh centroids being trapped by false surrounding structures. The internal energy  $E_{\text{int}}$  (Eq. 14) penalizes the deviations of mesh  $\mathcal{M}_q$  from the training repository, represented by set  $\{\mathcal{M}_q^1, \mathcal{M}_q^2, \dots, \mathcal{M}_q^M\}$  of  $M$  aligned 3D face-vertex surface meshes of the observed  $q$ -th vertebral level, where each mesh  $\mathcal{M}_q^m = \{\mathcal{V}_q^m, \mathcal{F}_q^m\}$ ;  $m = 1, 2, \dots, M$ , consists of  $|\mathcal{V}_q|$  vertices and  $|\mathcal{F}_q|$  faces. In comparison to the original approach [4], we incorporate a shape prior constraint into the computation of  $E_{\text{int}}$  by approximating the input mesh as a linear combination of meshes from the training repository. Let the vertices in the training repository be further squeezed into a matrix  $S_q = [\mathbf{s}_q^1, \mathbf{s}_q^2, \dots, \mathbf{s}_q^M]$  of size  $3|\mathcal{V}_q| \times M$ , where  $\mathbf{s}_q^m = \Phi(\mathcal{V}_q^m)$  represents the concatenation of vertices from the  $m$ -th mesh in the training repository according to transformation  $\Phi: \mathbb{R}^{|\mathcal{V}_q| \times 3} \rightarrow \mathbb{R}^{3|\mathcal{V}_q| \times 1}$  (i.e.  $\Phi$  concatenates triplets of mesh vertices into a column vector). The deformation of the observed mesh  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$  is formulated as an optimization problem [37] that incorporates the shape prior constraint into mesh deformation by approximating  $\Phi(\mathcal{V}_q)$  with  $S_q$ :

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^{M \times 1}} \|S_q \mathbf{w} - T(\Phi(\mathcal{V}_q))\|, \quad (17)$$

where  $\mathbf{w}$  is a vector of size  $M \times 1$  that represents the linear combination coefficients for vertices in the training repository, and  $T$  is the rigid transformation matrix that aligns the input shape  $\Phi(\mathcal{V}_q)$  to the common canonical space of training shapes in  $S_q$ . By estimating  $T$  using Procrustes superimposition [38], Eq. 17 is reduced to a basic convex optimization problem that is solved by the conjugate gradient method [39], resulting in optimal coefficients  $\mathbf{w}^*$ . The shape prior constraint then results in reconfigured vertices  $\mathcal{V}_q^*$ :

$$\mathcal{V}_q^* = \Phi^{-1}(T^{-1}(S_q \mathbf{w}^*)), \quad (18)$$

that form a new mesh  $\mathcal{M}_q^* = \{\mathcal{V}_q^*, \mathcal{F}_q\}$  by preserving the original mesh topology, i.e.  $|\mathcal{V}_q| = |\mathcal{V}_q^*|$  and  $\mathcal{F}_q = \mathcal{F}_q^*$ . As a result, the internal energy  $E_{\text{int}}$  is defined as:

$$E_{\text{int}} = \sum_{i=1}^{|\mathcal{V}_q|} \sum_{j \in \mathcal{N}_i} \|(\mathbf{v}_i - \mathbf{v}_j) - (\mathbf{v}_i^* - \mathbf{v}_j^*)\|^2 \quad (19)$$

where  $\mathbf{v}_i$  and  $\mathbf{v}_i^*$  are vertices from sets  $\mathcal{V}_q$  and  $\mathcal{V}_q^*$ , respectively, and  $\mathcal{N}_i$  is the set of vertices neighboring to  $\mathbf{v}_i$  (or  $\mathbf{v}_i^*$ , since the topology is preserved). The internal energy therefore restricts the flexibility of mesh  $\mathcal{M}_q$  by penalizing the deviation between deformation vertices  $\mathcal{V}_q$  and constraint vertices  $\mathcal{V}_q^*$ .

## IV. EXPERIMENTS AND RESULTS

In this section, we describe the details of image databases (IV-A), implementation (IV-B) and performed experiments (IV-C), and we present the obtained evaluation results (IV-D).

### A. Spine Image Databases

The proposed automated framework for spine and vertebra detection and segmentation was evaluated on two publicly available databases of CT spine images that are part of the *SpineWeb*,<sup>1</sup> a collaborative platform for research on spine imaging and image analysis. The first database<sup>2</sup> consists of 10 axially reconstructed CT images of the lumbar spine with a total of 50 lumbar vertebrae from 10 subjects (7 males and 3 females; mean age  $\pm$  SD of  $46.0 \pm 13.5$  years), in-plane voxel size of  $0.282 - 0.791$  mm and cross-sectional thickness of  $0.725 - 1.530$  mm [10]. The second database<sup>3,4</sup> consists of 10 axially reconstructed CT images of the thoracolumbar spine with a total of 120 thoracic and 50 lumbar vertebrae from 10 subjects (3 males and 7 females; mean age  $\pm$  SD of  $23.0 \pm 3.7$  years), in-plane voxel size of  $0.313 - 0.361$  mm and cross-sectional thickness of 1 mm [41]. For both databases, a reference segmentation binary mask was available for each vertebra in each image.

### B. Implementation Details

The proposed framework was implemented in C# and Matlab without code optimization. After constructing the mean shape model for each vertebral level and for the observed spinal region, spine and vertebra detection was performed and followed by vertebra segmentation.

1) *Mean Vertebra Shape Models*: Let set  $\mathcal{I}$  contain 3D images of the spine, where each image  $I \in \mathcal{I}$  is assigned a set  $\mathcal{B} = \{B_1, B_2, \dots, B_Q\}$  of  $Q$  binary masks representing reference segmentations of vertebrae in  $I$ . To extract the shape of the  $q$ -th vertebra, the marching cubes algorithm [42] was applied to binary mask  $B_q$ , resulting in a 3D face-vertex surface mesh of genus 1 (i.e. the number of holes was 1, as expected for vertebrae) with  $|\mathcal{V}_q| = 31.542 - 161.790$  vertices and  $|\mathcal{F}_q| = 2|\mathcal{V}_q|$  faces (i.e. triangles). The dependency of the number of vertices on image voxel size and on the size of the observed vertebra was removed by isotropic remeshing [43] with mean edge length of 2.25 mm, resulting in a mesh with  $|\mathcal{V}_q| = 1.353 - 5.642$  vertices. To establish pointwise vertex correspondences among meshes of the same vertebral level, the coherent point drift algorithm [44] was applied to obtain the nonrigid transformation among corresponding sets of vertices. Finally, the generalized Procrustes alignment [38] was used to remove translation, rotation and scaling from corresponding meshes, yielding the mean shape model for

<sup>1</sup> Accessible via <http://spineweb.digitalimaginggroup.ca>.

<sup>2</sup> Accessible via <http://lit.fe.uni-lj.si/tools.php>.

<sup>3</sup> Accessible via <http://spineweb.digitalimaginggroup.ca/dataset.html>.

<sup>4</sup> The proposed framework is an improved version of our winning method [40] of the Spine and Vertebrae Segmentation Challenge that ran on this database at the "2nd MICCAI Workshop & Challenge on Computational Methods and Clinical Applications for Spine Imaging - CSI 2014" (<http://csi-workshop.weebly.com>).

each vertebral level, represented by a 3D face-vertex surface mesh  $\mathcal{M}_q = \{\mathcal{V}_q, \mathcal{F}_q\}$ . The mean shape model of the observed spinal region  $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_Q\}$  was constructed by concatenating the obtained mean vertebra shape models.

2) *Spine and Vertebra Detection*: According to the selected interpolation function  $\bar{f}$ , interpolation grid  $\mathcal{P}$  and optimization algorithm for the reduction of computational complexity (Sec. II-A), we devised five interpolation schemes for spine and vertebra detection (Sec. II-B):

- S-E-W: (cubic) spline functions (S) – equidistant grid (E) – dimension-wise algorithm (W);
- P-C-W: polynomial functions (P) – Chebyshev grid (C) – dimension-wise algorithm (W) [32];
- P-C-A: polynomial functions (P) – Chebyshev grid (C) – dimension-adaptive algorithm (A) [31], [32];
- P-CC-W: polynomial functions (P) – Clenshaw-Curtis grid (CC) – dimension-wise algorithm (W) [32];
- P-CC-A: polynomial functions (P) – Clenshaw-Curtis grid (CC) – dimension-adaptive algorithm (A) [31], [32].

As the equidistant grid was identified as the optimal interpolation grid for spline interpolation (Sec. II-A), it is in such case reasonable not to apply the dimension-adaptive optimization algorithm that results in a non-equidistant grid. Each interpolation scheme from the list above was independently applied for spine and vertebra detection in each image  $I \in \mathcal{I}$ . The initial location of the mean shape model of the observed spinal region  $\mathcal{M}$  was each time set to the average location of the spine obtained from all images in  $\mathcal{I}$ .

To detect the spine in each image  $I \in \mathcal{I}$ , the pose of  $\mathcal{M}$  was optimized, according to the selected interpolation scheme, against translation  $\mathbf{p} = \mathbf{t} = (x, y, z)$ , i.e. coordinates  $x$ ,  $y$  and  $z$  representing sagittal, coronal and axial anatomical directions, respectively, in the 3D Cartesian coordinate system. The nodes in  $\mathcal{P}$  were distributed across the interval of  $[-30 \text{ mm}, +30 \text{ mm}]$  for  $x$ ,  $y$  and  $z$ , and the resulting number of nodes was around 1350 for every interpolation scheme. The dimensionality of the optimization problem was therefore  $n = 3$  on the corresponding domain  $\Omega_s^3$ , meaning that complexity reduction (Sec. II-A) was in this case not required (i.e.  $d = n = 3$ ). Haar-like gradients, used for the objective function  $f$  (Eq. 2), were computed in  $15 \times 15 \times 15 \text{ mm}^3$  large voxel neighborhoods. The global maximum of  $\bar{f}$  represented the location of the spine in the 3D image, and was further used to initialize the detection of each individual vertebra.

To detect each  $q$ -th vertebra in  $I \in \mathcal{I}$ , the pose of  $\mathcal{M}_q$  was optimized, according to the selected interpolation scheme, against translation  $\mathbf{p} = \mathbf{t} = (x, y, z)$ . The nodes in  $\mathcal{P}$  were distributed across the interval of  $[-25 \text{ mm}, +25 \text{ mm}]$  for  $x$  and  $y$ , and across the interval of  $[-70 \text{ mm}, +70 \text{ mm}]$  for  $z$ , resulting in around 3540 nodes for every interpolation scheme. The dimensionality of the optimization problem was therefore again  $d = n = 3$  on the corresponding domain  $\Omega_v^3$ , meaning that complexity reduction was also not required. However,  $\Omega_v^3$  was defined separately for each observed vertebra, moreover, it captured several neighboring vertebrae to take into account the eventual longitudinal shift of the mean shape model of the

observed spinal region. Haar-like gradients (Eq. 2) were again computed in  $15 \times 15 \times 15 \text{ mm}^3$  large voxel neighborhoods, and in contrast to spine detection,  $L = 20$  local maxima of  $\bar{f}$  were extracted, corresponding to vertebrae within  $\Omega_v^3$ . By considering all vertebrae in the observed spinal region, set  $\mathcal{L}^*$  (Sec. II-B) of optimal locations for each vertebra in the observed spinal region was determined and further used to initialize the alignment of mean vertebra shape models.

To perform a more accurate alignment of  $\mathcal{M}_q$  to the corresponding  $q$ -th vertebra in  $I \in \mathcal{I}$ , the length of the observed spinal region and the relative orientation of each vertebra were first calculated from set  $\mathcal{L}^*$  (Eq. 3). Next,  $\mathcal{M}_q$  was scaled according to the ratio between the obtained length and the length of the mean shape model of the observed spinal region  $\mathcal{M}$ , and rotated according to the location of its neighboring mean vertebra shape models in  $\mathcal{L}^*$ . According to the selected interpolation scheme, the pose of  $\mathcal{M}_q$  was then optimized against  $\mathbf{p} = (\mathbf{t}, s, \varphi)$  consisting of translation  $\mathbf{t} = (x, y, z)$  scaling  $s$  and rotation  $\varphi = (\varphi_x, \varphi_y, \varphi_z)$ , i.e. angles  $\varphi_x$ ,  $\varphi_y$  and  $\varphi_z$  of rotation about axes  $x$ ,  $y$  and  $z$ , respectively, of the 3D Cartesian coordinate system. The nodes in  $\mathcal{P}$  were distributed across the interval of  $[-7 \text{ mm}, +7 \text{ mm}]$  for  $x$ ,  $y$  and  $z$ , across the interval of  $[0.9, 1.1]$  for  $s$ , and across the interval of  $[-9^\circ, +9^\circ]$  for  $\varphi_x$ ,  $\varphi_y$  and  $\varphi_z$ . The dimensionality of the optimization problem was therefore  $n = 7$  on the corresponding domain  $\Omega_v^7$ , and as an exhaustive search through such a number of dimensions is computationally expensive, the proposed dimension-wise and dimension-adaptive complexity reduction algorithms (Sec. II-A) were in this case applied (for the dimension-wise algorithm,  $d = 4$  dimensions consisting of translation  $\mathbf{t}$  and scaling  $s$  were optimized first, and then the remaining  $d = 3$  dimensions consisting of rotation  $\varphi$  were optimized second). After complexity reduction, the resulting number of nodes was around 2000 for every interpolation scheme. Haar-like gradients (Eq. 2) were computed in  $7 \times 7 \times 7 \text{ mm}^3$  large voxel neighborhoods, and after extracting the location of  $L = 5$  local maxima of  $\bar{f}$ , the location of the global maximum was selected to generate a binary mask  $B_q^{\text{det}}$  [45] representing the final detection result of the  $q$ -th vertebra in image  $I$ .

3) *Vertebra Segmentation*: The final vertebra detection results were used to initialize the framework for model-based segmentation of each  $q$ -th observed vertebra (Sec. III). The framework parameters were, similarly as for the original shape-constrained deformable model approach [4], set to  $J = 15$  for thoracic and  $J = 25$  for lumbar vertebrae (Eq. 6),  $\alpha = 0.9$  (Eq. 6),  $D = 1.2 \text{ mm}^{-2}$  for thoracic and  $D = 0.6 \text{ mm}^{-2}$  for lumbar vertebrae (Eqs. 7 and 16),  $\delta = 0.3 \text{ mm}$  (Eqs. 5, 7 and 16),  $g_{\text{max}} = 100 \text{ HU}$  (Eq. 8) and  $\beta = 33$  (Eq. 14). For the random forest regression model,  $K = 2$  was used (Eqs. 11, 12 and 13). The iterative procedure of repeated boundary detection (Sec. III-A) and mesh deformation (Sec. III-B) steps consisted of  $P = 15$  iterations for thoracic and  $P = 25$  iterations for lumbar vertebrae. Since the total energy (Eq. 14) is a sum of quadratic terms, the minimization problem was efficiently solved by the conjugate gradient method [39]. The resulting mesh, which was optimally positioned according to



the proposed segmentation framework, was used to generate a binary mask  $B_q^{\text{seg}}$  [45] representing the segmentation of the  $q$ -th vertebra in image  $I$ .

### C. Experiments

All experiments were performed independently for the database of lumbar spine images and for the database of thoracolumbar spine images by applying leave-one-out cross-validation on each database, and executed on a personal computer with Intel Core i5 processor at 3.2 GHz and 16 GB of memory without a graphics processing unit. The detection of all five vertebrae in the database of lumbar spine images (i.e. levels from L1 to L5) took on average around 3 min, while the segmentation of each individual vertebra took on average around 110 s. For the database of thoracolumbar spine images, the detection of all 17 vertebrae (i.e. levels from T1 to L5) took on average around 8 min, while the segmentation of each individual vertebra took on average around 90 s. The total computational times can be therefore estimated to, on average, around 12 min for lumbar spine images and around 33 min for thoracolumbar spine images. The performance of vertebra detection was evaluated by computing the detection error  $\epsilon$  and Dice coefficient  $\kappa$  for the final vertebra detection result. The detection error  $\epsilon$  was computed as the centroid-to-centroid distance, i.e. for the  $q$ -th observed vertebra in image  $I$ , the Euclidean distance between the center of mass of the reference segmentation binary mask  $B_q$  and the center of mass of the obtained detection binary mask  $B_q^{\text{det}}$  was determined. The performance of vertebra segmentation was evaluated by computing the segmentation error  $\eta$  and Dice coefficient  $\kappa$  for the final vertebra segmentation result. The segmentation error  $\eta$  was computed as the mean symmetric surface distance, which estimates the error between the reference segmentation binary mask  $B_q$  and the obtained segmentation binary mask  $B_q^{\text{seg}}$  using Euclidean distances between their boundary voxels:

$$\eta = \frac{1}{N_q + N_q^{\text{seg}}} \left( \sum_{i=1}^{N_q} d_i + \sum_{j=1}^{N_q^{\text{seg}}} d_j^{\text{seg}} \right), \quad (20)$$

where  $N_q$  and  $N_q^{\text{seg}}$  are, respectively, the numbers of boundary voxels in  $B_q$  and  $B_q^{\text{seg}}$ ,  $d_i$  is the distance between the  $i$ -th boundary voxel in  $B_q$  and its closest boundary voxel in  $B_q^{\text{seg}}$ , while  $d_j^{\text{seg}}$  is the distance between the  $j$ -th boundary voxel in  $B_q^{\text{seg}}$  and its closest boundary voxel in  $B_q$ .<sup>5</sup>

### D. Results

Vertebra detection was initialized at the average location of the spine obtained from all images in the database that, in terms of mean  $\pm$  SD of the detection error  $\epsilon$  and Dice coefficient  $\kappa$ , corresponded to  $\epsilon = 21.0 \pm 11.0$  mm and  $\kappa = 22.6 \pm 16.3\%$  (Tab. I), and was considered successful only when the resulting Dice coefficient was above 50% (i.e.  $\kappa > 50\%$ ). The ratio of successfully detected vertebrae against all

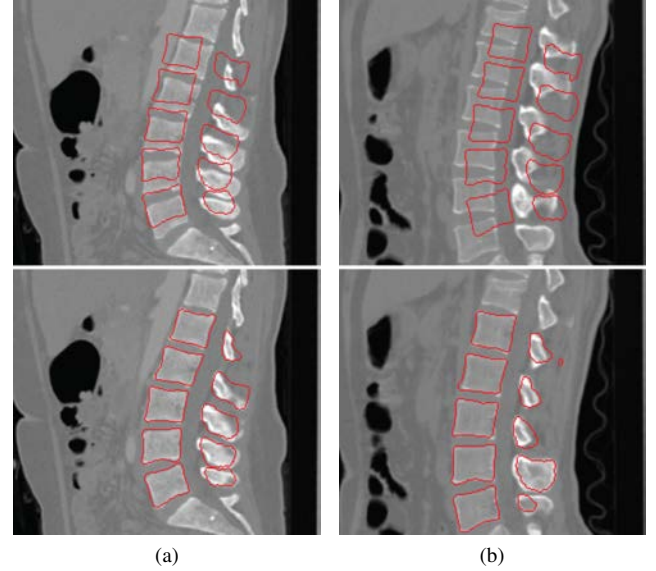


Fig. 5. Initialization of the mean shape model of the lumbar spine (top) and final vertebra detection results (bottom), obtained after applying the S-E-W scheme (interpolation by spline functions on the equidistant grid with dimension-wise optimization), showed in mid-sagittal cross-sections for two ((a) and (b)) randomly selected images from the lumbar spine database.

vertebrae in the database, referred to as the success rate (SR), can be therefore considered as an indicator of the detection reliability. For vertebrae from both lumbar and thoracolumbar spine image databases, the overall detection performance was  $\epsilon = 1.1 \pm 0.7$  mm and  $\kappa = 83.6 \pm 4.7\%$  for the S-E-W (97% SR),  $\epsilon = 1.8 \pm 1.8$  mm and  $\kappa = 79.1 \pm 8.0\%$  for the P-C-W (91% SR),  $\epsilon = 1.9 \pm 0.8$  mm and  $\kappa = 75.3 \pm 4.9\%$  for the P-C-A (95% SR),  $\epsilon = 1.6 \pm 1.4$  mm and  $\kappa = 80.0 \pm 6.6\%$  for the P-CC-W (94% SR), and  $\epsilon = 2.2 \pm 1.1$  mm and  $\kappa = 74.1 \pm 5.8\%$

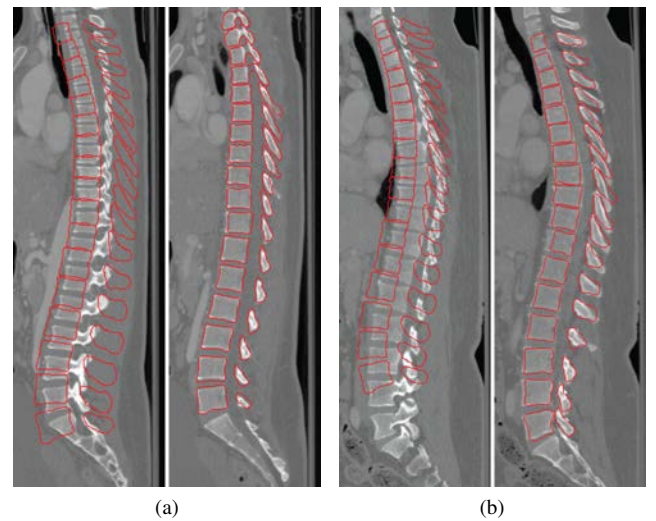


Fig. 6. Initialization of the mean shape model of the thoracolumbar spine (left) and final vertebra detection results (right), obtained after applying the S-E-W scheme (interpolation by spline functions on the equidistant grid with dimension-wise optimization), showed in mid-sagittal cross-sections for two ((a) and (b)) randomly selected images from the thoracolumbar spine database.

<sup>5</sup>The source code for evaluating segmentation results in terms of segmentation error  $\eta$  and Dice coefficient  $\kappa$  is accessible via <http://mbi.dkfz-heidelberg.de/grand-challenge2007> and was written for the workshop "3D Segmentation in the Clinic: A Grand Challenge (in conjunction with MICCAI 2007)".

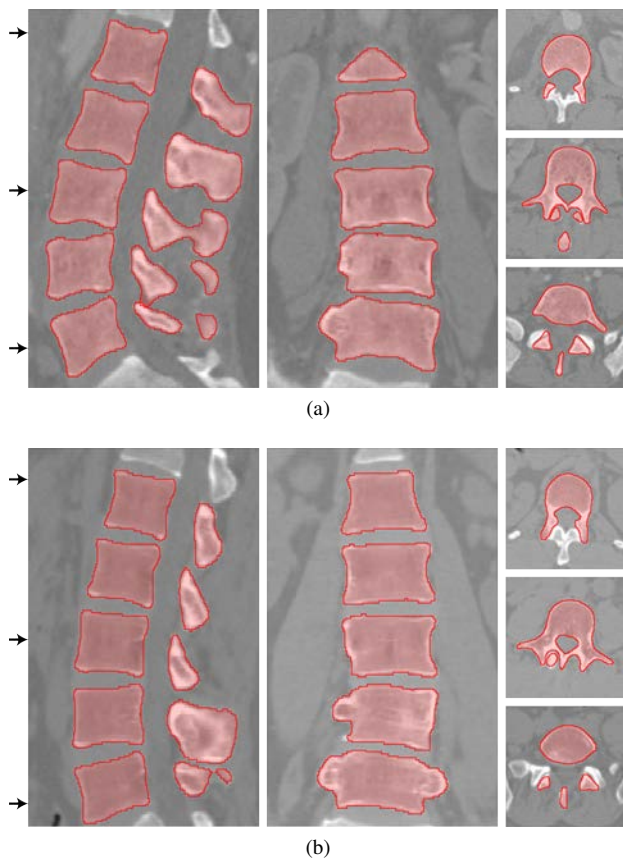


Fig. 7. Segmentation results for two randomly selected images from the lumbar spine database ((a) and (b), corresponding to Fig. 5), showed in mid-sagittal cross-sections (left), mid-coronal cross-sections (middle) and selected axial cross-sections (right) at locations indicated by arrows.

for the P-CC-A (94% SR) optimization scheme. Examples of the initialization and final vertebra detection results are shown in Figure 5 for the database of lumbar spine images, and in Figure 6 for the database of thoracolumbar spine images. The main reason for unsuccessful vertebra detection was in the longitudinal shift of the mean shape model of the observed spinal region for one or more vertebral levels after the spine detection step. In the case of the S-E-W optimization scheme, the shift occurred for only one image in the lumbar spine image database. As the S-E-W optimization scheme resulted in the most accurate and reliable vertebra detection (Tab. I), it was used to initialize vertebra segmentation. Table II shows, separately for each vertebral level and image database, detailed vertebra detection results for the S-E-W optimization scheme.

Vertebra segmentation was initialized by the S-E-W detection results that, in terms of mean  $\pm$  SD of the segmentation error  $\eta$  (Eq. 20) and Dice coefficient  $\kappa$ , corresponded to  $\eta = 1.0 \pm 0.3$  mm and  $\kappa = 83.6 \pm 4.7\%$  (Tab. I). For vertebrae from both lumbar and thoracolumbar spine image databases, the overall segmentation performance was  $\eta = 0.3 \pm 0.1$  mm and  $\kappa = 94.6 \pm 2.0\%$ . Examples of the final vertebra segmentation results are shown in Figure 7 for the database of lumbar spine images, and in Figure 8 for the database of thoracolumbar spine images. Although detected correctly, two vertebrae at level T1 from the thoracolumbar spine image database were not segmented, because the corresponding original images

were cropped so that the superior articular facets at T1 were not visible. Table II shows, separately for each vertebral level and image database, detailed vertebra segmentation results for the proposed approach.

## V. DISCUSSION

In this paper, we described a novel framework for automated spine and vertebra detection and segmentation in 3D from CT spine images. To detect the spine and vertebrae in the 3D image, we applied a novel interpolation-based optimization approach first to detect the location of the whole observed

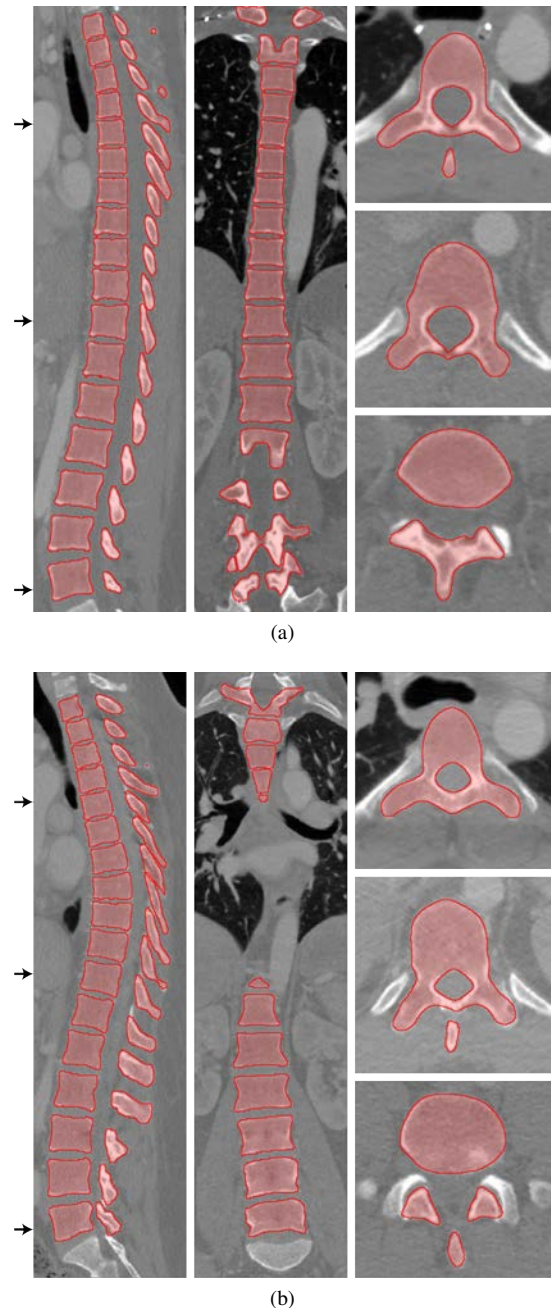


Fig. 8. Segmentation results for two randomly selected images from the thoracolumbar spine database ((a) and (b), corresponding to Fig. 6), showed in mid-sagittal cross-sections (left), mid-coronal cross-sections (middle) and selected axial cross-sections (right) at locations indicated by arrows.

TABLE I

SUMMARY AND COMPARISON OF VERTEBRA DETECTION AND SEGMENTATION RESULTS IN TERMS OF MEAN  $\pm$  STANDARD DEVIATION OF THE DETECTION ERROR  $\epsilon$ , SEGMENTATION ERROR  $\eta$  AND DICE COEFFICIENT  $\kappa$  ( $N$  DENOTES THE NUMBER OF VERTEBRAE). DETECTION RESULTS ARE REPORTED SEPARATELY FOR THE PROPOSED SPLINE INTERPOLATION ON THE EQUIDISTANT GRID WITH THE DIMENSION-WISE OPTIMIZATION (S-E-W), POLYNOMIAL INTERPOLATION ON THE CHEBYSHEV GRID WITH THE DIMENSION-WISE (P-C-W) AND DIMENSION-ADAPTIVE OPTIMIZATION (P-C-A), POLYNOMIAL INTERPOLATION ON THE CLENSHAW-CURTIS GRID WITH THE DIMENSION-WISE (P-CC-W) AND DIMENSION-ADAPTIVE OPTIMIZATION (P-CC-A), AND ITERATED MARGINAL SPACE LEARNING APPROACH [17] WITH RUSBOOST (IMSL-RB) AND RANDOM FOREST (IMSL-RF) CLASSIFICATION ( $\epsilon < 10$  MM FOR IMSL-RB AND IMSL-RF, AND  $\kappa > 50\%$  FOR OTHER APPROACHES WAS CONSIDERED AS A SUCCESSFUL DETECTION). SEGMENTATION RESULTS ARE REPORTED SEPARATELY FOR THE PROPOSED IMPROVED AND ORIGINAL SHAPE-CONSTRAINED DEFORMABLE MODEL APPROACH (DM-SC) [4]

Detection approach	Lumbar spine images ( $N = 50$ )			Thoracolumbar spine images ( $N = 170$ )			All images ( $N = 220$ )		
	$N$	$\epsilon$ (mm)	$\kappa$ (%)	$N$	$\epsilon$ (mm)	$\kappa$ (%)	$N$	$\epsilon$ (mm)	$\kappa$ (%)
Initialization	50	$26.0 \pm 16.3$	$25.3 \pm 20.1$	170	$19.5 \pm 9.4$	$21.8 \pm 15.0$	220	$21.0 \pm 11.6$	$22.6 \pm 16.3$
S-E-W (applied)	45	$1.1 \pm 0.8$	$85.1 \pm 5.1$	169	$1.1 \pm 0.6$	$83.2 \pm 4.6$	214	$1.1 \pm 0.7$	$83.6 \pm 4.7$
P-C-W (tested)	34	$2.8 \pm 3.2$	$77.3 \pm 12.6$	167	$1.6 \pm 1.3$	$79.5 \pm 6.6$	201	$1.8 \pm 1.8$	$79.1 \pm 8.0$
P-C-A (tested)	40	$1.5 \pm 1.1$	$85.2 \pm 5.4$	170	$2.0 \pm 0.7$	$74.2 \pm 4.0$	210	$1.9 \pm 0.8$	$75.3 \pm 4.9$
P-CC-W (tested)	37	$2.5 \pm 2.5$	$76.2 \pm 8.8$	170	$1.4 \pm 1.0$	$80.8 \pm 4.7$	207	$1.6 \pm 1.4$	$80.0 \pm 6.6$
P-CC-A (tested)	38	$2.6 \pm 2.1$	$75.8 \pm 9.8$	168	$2.1 \pm 0.7$	$73.7 \pm 4.3$	206	$2.2 \pm 1.1$	$74.1 \pm 5.8$
iMSL-RB [17]	45	$1.7 \pm 0.9$	–	170	$2.2 \pm 1.6$	–	215	$2.1 \pm 1.5$	–
iMSL-RF [17]	45	$1.5 \pm 0.8$	–	170	$2.0 \pm 1.5$	–	215	$1.9 \pm 1.4$	–

Segmentation approach	Lumbar spine images ( $N = 50$ )			Thoracolumbar spine images ( $N = 170$ )			All images ( $N = 220$ )		
	$N$	$\eta$ (mm)	$\kappa$ (%)	$N$	$\eta$ (mm)	$\kappa$ (%)	$N$	$\eta$ (mm)	$\kappa$ (%)
Initialization (S-E-W)	45	$1.1 \pm 0.4$	$85.1 \pm 5.1$	169	$1.0 \pm 0.3$	$83.2 \pm 4.6$	214	$1.0 \pm 0.3$	$83.6 \pm 4.7$
DM-SC (applied)	45	$0.3 \pm 0.1$	$95.3 \pm 1.4$	167	$0.3 \pm 0.1$	$94.4 \pm 2.1$	212	$0.3 \pm 0.1$	$94.6 \pm 2.0$
DM-SC [4]	45	$0.6 \pm 0.2$	$92.2 \pm 2.1$	167	$0.4 \pm 0.1$	$91.5 \pm 2.0$	212	$0.5 \pm 0.1$	$91.6 \pm 2.0$

spinal region, then to detect the location of individual vertebrae within the observed spinal region, and finally to align the vertebra mean shape model to each vertebra in the observed spinal region. In general, the objective function used for vertebra detection suffers from convergence to wrong maxima, i.e. the location of the observed vertebra can be wrongly detected at the location of its neighboring vertebrae. Standard optimization techniques, such as the gradient descent, Powell's optimizer and CMA-ES, are usually limited by initialization and detect only a single maximum, and can therefore not be used to extract several local maxima of the objective function. Multi-scale optimization techniques that bring initialization closer to the global maximum are also not effective, as the actual location of the observed vertebra is not necessary associated with the global maximum of the corresponding objective function. On the other hand, optimizers that analyze the complete domain where the object of interest can be located and where the objective function is defined, such as the brute force approach, are also not feasible due to inherent computational costs. It can be therefore concluded that a robust and efficient spine and vertebra detection algorithm should observe the complete domain where the spine and vertebrae as the objects of interest can be located and, at the same time, it should minimize the eventual computational costs. To tackle these problems, we applied the concept of interpolation theory to compute the objective function on a sparse set of points, distributed over the observed domain,

and then approximated the objective function on the observed domain. From the methodological perspective, we studied the performance of spline and polynomial interpolation functions, and equidistant, Chebyshev and Clenshaw-Curtis interpolation grids, which were combined with a dimension-wise or a dimension-adaptive optimization algorithm for complexity reduction into five interpolation schemes. In the case of polynomial interpolation, the performed experiments (Tab. I) show that both the Chebyshev and Clenshaw-Curtis grids, and also both algorithms for complexity reduction produced similarly accurate and reliable vertebra detection results, although it has to be noted that the proposed dimension-wise algorithm slightly outperformed the applied dimension-adaptive algorithm. However, polynomial interpolation was, especially in terms of detection accuracy, outperformed by spline interpolation on the equidistant grid with the dimension-wise optimization algorithm for complexity reduction, indicating that a gradient-based objective function cannot be adequately approximated by a high-degree polynomial but by a combination of low-degree polynomials, i.e. splines, and that the proposed dimension-wise complexity reduction algorithm ensures accurate detection even for a limited number of interpolation nodes, which can be nevertheless increased for a specific application. In fact, the proposed interpolation-based approach is not limited to vertebra or arbitrary object detection, but can be used in various image analysis problems to speed up optimization processes and smooth the potentially

TABLE II  
VERTEBRA DETECTION AND SEGMENTATION RESULTS IN TERMS OF MEAN  $\pm$  STANDARD DEVIATION OF THE DETECTION ERROR  $\epsilon$ , SEGMENTATION ERROR  $\eta$  AND DICE COEFFICIENT  $\kappa$  FOR THE APPLIED DETECTION (SPLINE INTERPOLATION ON THE EQUIDISTANT GRID WITH THE DIMENSION-WISE OPTIMIZATION) AND SEGMENTATION (IMPROVED SHAPE-CONSTRAINED DEFORMABLE MODEL) APPROACH ( $N$  DENOTES THE NUMBER OF VERTEBRAE;  $\kappa > 50\%$  WAS CONSIDERED AS A SUCCESSFUL DETECTION)

Database	Level	Applied detection approach			Applied segmentation approach		
		$N$	$\epsilon$ (mm)	$\kappa$ (%)	$N$	$\eta$ (mm)	$\kappa$ (%)
Lumbar spine images ( $N = 50$ )	L1	9	$0.87 \pm 0.43$	$87.36 \pm 1.16$	9	$0.28 \pm 0.08$	$95.55 \pm 0.79$
	L2	9	$0.65 \pm 0.24$	$87.38 \pm 1.92$	9	$0.23 \pm 0.07$	$96.18 \pm 0.63$
	L3	9	$0.83 \pm 0.17$	$86.33 \pm 1.50$	9	$0.28 \pm 0.12$	$95.82 \pm 1.05$
	L4	9	$1.07 \pm 0.40$	$85.11 \pm 1.90$	9	$0.35 \pm 0.10$	$94.92 \pm 1.33$
	L5	9	$2.13 \pm 1.21$	$79.34 \pm 8.91$	9	$0.40 \pm 0.15$	$94.13 \pm 1.74$
	L1-L5	45	$1.1 \pm 0.8$	$85.1 \pm 5.1$	45	$0.3 \pm 0.1$	$95.3 \pm 1.4$
Thoracolumbar spine images ( $N = 170$ )	T1	9	$0.95 \pm 0.59$	$81.10 \pm 3.45$	7	$0.36 \pm 0.07$	$91.78 \pm 0.96$
	T2	10	$1.57 \pm 0.86$	$79.01 \pm 8.03$	10	$0.29 \pm 0.06$	$93.18 \pm 0.81$
	T3	10	$1.25 \pm 0.76$	$80.39 \pm 7.82$	10	$0.30 \pm 0.08$	$92.95 \pm 1.30$
	T4	10	$1.00 \pm 0.69$	$82.50 \pm 3.18$	10	$0.39 \pm 0.22$	$91.87 \pm 3.05$
	T5	10	$1.42 \pm 0.84$	$82.08 \pm 6.16$	10	$0.39 \pm 0.21$	$92.08 \pm 2.97$
	T6	10	$0.84 \pm 0.53$	$83.99 \pm 3.93$	10	$0.34 \pm 0.15$	$93.41 \pm 1.52$
	T7	10	$0.92 \pm 0.59$	$84.66 \pm 3.45$	10	$0.32 \pm 0.17$	$93.91 \pm 1.73$
	T8	10	$0.90 \pm 0.46$	$84.61 \pm 4.03$	10	$0.27 \pm 0.15$	$94.83 \pm 1.40$
	T9	10	$1.22 \pm 0.88$	$83.84 \pm 5.05$	10	$0.24 \pm 0.11$	$95.39 \pm 1.17$
	T10	10	$0.85 \pm 0.54$	$85.54 \pm 2.51$	10	$0.31 \pm 0.14$	$94.68 \pm 1.74$
	T11	10	$1.37 \pm 0.34$	$82.12 \pm 3.23$	10	$0.25 \pm 0.06$	$95.53 \pm 0.68$
	T12	10	$0.91 \pm 0.23$	$85.69 \pm 1.86$	10	$0.23 \pm 0.07$	$95.78 \pm 0.69$
	L1	10	$1.16 \pm 0.43$	$83.74 \pm 2.79$	10	$0.21 \pm 0.04$	$96.14 \pm 0.47$
	L2	10	$1.01 \pm 0.47$	$82.96 \pm 3.27$	10	$0.22 \pm 0.04$	$96.08 \pm 0.31$
	L3	10	$0.84 \pm 0.25$	$83.71 \pm 3.20$	10	$0.25 \pm 0.07$	$95.93 \pm 0.58$
	L4	10	$0.98 \pm 0.39$	$83.34 \pm 3.16$	10	$0.25 \pm 0.06$	$95.92 \pm 0.63$
	L5	10	$1.25 \pm 0.71$	$84.55 \pm 3.59$	10	$0.30 \pm 0.06$	$94.95 \pm 0.97$
	T1-L5	169	$1.1 \pm 0.6$	$83.2 \pm 4.6$	167	$0.3 \pm 0.1$	$94.4 \pm 2.1$
	All levels	214	$1.1 \pm 0.7$	$83.6 \pm 4.7$	212	$0.3 \pm 0.1$	$94.6 \pm 2.0$

harmful fluctuations of objective functions.

For an accurate segmentation of vertebrae and vertebral structures, accurate and reliable initialization is indispensable. In the proposed framework, the final vertebra detection results proved to be an adequate initialization for the subsequent vertebra segmentation, which was achieved by applying an improved shape-constrained deformable model approach. The main disadvantage of the original shape-constrained deformable model approach introduced by Weese et al. [4] is that it can easily detect boundaries of neighboring vertebrae instead of the observed vertebra, which is especially frequent for the thoracic region, where strong articulation among vertebrae is present, or when there is advanced calcification between vertebrae. Klinder et al. [3] overcame such situations by accounting for spatial relationships between vertebrae within a collision detection scheme. In this case, the computational cost for vertebra boundary detection increased considerably but could be reduced by an efficient implementation of the collision detection. Regarding mesh deformation, the flexibility of the mesh that is regulated by the internal energy was penalized either by encouraging the distribution of mesh vertices to stay close to the canonical space spanned by the point distribution

model [4], or by maintaining the distribution of mesh vertices according to the input mesh [3]. However, the drawback of such approaches is in shape variations that are usually complex and therefore difficult to model by a point distribution model (i.e. mean shape with variations), and in a limited vertex displacement that further results in a lower model variance. In contrast to the original approach [4], we introduced several modifications that improved the accuracy of the resulting segmentation. To overcome the problem of boundary detection on neighboring vertebrae, we gradually reduced the length of the sampling parcel, within which new mesh face centroids are searched for, and proposed an improved boundary detection operator that incorporates prior knowledge about the observed vertebra by generating a random forest regression model for the corresponding image intensities and intensity gradients. To overcome the limits of mesh flexibility, we incorporated a shape prior constraint into the computation of the internal energy by approximating the input mesh as a linear combination of meshes from the training repository.

From the point of view of the obtained detection and segmentation results, the performance of the proposed framework is superior or comparable to the results reported by existing



studies of vertebra detection and segmentation from 3D CT spine images, however, it has to be noted that most studies address the problem of detection and segmentation separately. Glocker et al. [15], [16] reported an overall success rate of 81% for detecting 2595 vertebrae from a heterogeneous database of CT images of different fields of view, and an overall success rate of 76% and 70% for detecting vertebrae from, respectively, a database of 200 images of the abdomen and 224 images of highly pathological spines. On the other hand, Schwier et al. [14] reported an average success rate of 96% for detecting 418 vertebrae from images of 20 subjects. Using the iterated marginal space learning approach [18], Kelm et al. [17] reported a mean detection error of 3.2 mm and 2.4 mm for detecting intervertebral discs from, respectively, 30 CT and 42 MR whole spine images, both with a success rate of around 98%. For the segmentation of vertebrae from different spinal regions, Kim and Kim [2] reported a success rate of 80% for vertebrae from CT images of 50 subjects, which was determined according to a subjective classification based on visual assessment. For segmenting vertebrae from the thoracolumbar spinal region, Klinder et al. [3] reported a mean point-to-surface error (i.e. one surface is represented by a set of surface points and the other by a surface mesh model) of 1.1 mm for 353 vertebrae, while Kadoury et al. [5], [6] reported a mean surface-to-surface error (i.e. both surfaces are represented by surface mesh models) of 2.0 mm and 2.4 mm, and a Dice coefficient of around 94.3% and 92.1% for 204 and 357 vertebrae, respectively. For vertebrae from the lumbar spinal region, a Dice coefficient of 89.3% was reported by Lim et al. [7] for segmenting 100 vertebrae, and a Dice coefficient of 93.6% was reported by Ibragimov et al. [10] for segmenting 50 vertebrae. Furthermore, Ma et al. [8] reported a mean point-to-surface error of 0.9 mm for 480, and Rasoulian et al. [9] reported a mean surface-to-surface error of 1.4 mm for 135 lumbar vertebrae. By applying the proposed vertebra detection with the S-E-W interpolation scheme, we achieved a mean detection error of 1.1 mm and Dice coefficient of 83.6% for 214 out of 220 thoracolumbar vertebrae, meaning that on average 97% of vertebrae were successfully detected. It has to be noted that the ratio of successfully detected vertebrae was higher for the thoracolumbar than for the lumbar spine image database (mean of 99% vs. 90%), which is most probably due to a younger population in the database of thoracolumbar spine images (mean of 23 years vs. 46 years), as aging is associated with bone demineralization that makes vertebrae less pronounced in CT images and therefore more difficult to detect. By using the obtained vertebra detection results to initialize vertebra segmentation, we achieved a mean segmentation error of 0.3 mm and Dice coefficient of 94.6% for 212 out of 214 thoracolumbar vertebrae (two not completely visible vertebrae at level T1 were excluded).

The results of the proposed method can not be directly compared to the results reported by existing studies because of different evaluation methodologies and data collection techniques, as well as because of different databases that include different spinal regions and a different number of observed vertebrae. However, for the purpose of an objective comparison, we implemented two state-of-the-art detection

and segmentation algorithms, applied them to the same two publicly available databases of lumbar and thoracolumbar CT spine images, and used the same validation metrics to evaluate the obtained results. To detect vertebrae, we adapted the iterated marginal space learning approach for intervertebral disc detection of Kelm et al. [17] and implemented it with the random under-sampling with boosting (RUSBoost) [46] and random forest [36] classification, which resulted (Tab.I) in a mean detection error of 2.1 mm and 1.9 mm, respectively, for 215 out of 220 thoracolumbar vertebrae, meaning that on average 98% of vertebrae were successfully detected in both cases (the Dice coefficient does not derive from the framework). To segment vertebrae, we implemented the original shape-constrained deformable model approach of Weese et al. [4], which resulted (Tab.I) in a mean segmentation error of 0.5 mm and Dice coefficient of 91.6% for 212 out of 214 thoracolumbar vertebrae (two vertebrae at level T1 were again excluded). According to the obtained results and comparison to existing methods, we can conclude that the results obtained by the proposed framework for automated spine and vertebrae interpolation-based detection and model-based segmentation are highly accurate.

## VI. CONCLUSION

In this paper, we presented an automated framework for spine and vertebra detection and segmentation, in which detection was performed by a novel optimization scheme based on interpolation theory, and segmentation was performed by applying an improved shape-constrained deformable model approach. The performance was evaluated on two publicly available databases of CT spine images that are part of the collaborative platform *SpineWeb*, resulting in highly accurate detection and segmentation. Future directions of our work will include the extension of the proposed framework to different spinal structures (e.g. intervertebral discs), different image modalities (e.g. MR) and images of different fields of view, as well as validation of the proposed framework on larger image databases that will contain also various spinal pathologies.

## REFERENCES

- [1] M. Leventon, W. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition – CVPR 2000*, vol. 1, Hilton Head, SC, USA, Jun. 2000, pp. 316–323.
- [2] Y. Kim and D. Kim, "A fully automatic vertebra segmentation method using 3D deformable fences," *Comput. Med. Imag. Graph.*, vol. 33, no. 5, pp. 343–352, Jul. 2009.
- [3] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz, "Automated model-based vertebra detection, identification, and segmentation in CT images," *Med. Image Anal.*, vol. 13, no. 3, pp. 471–482, Jun. 2009.
- [4] J. Weese, M. Kaus, C. Lorenz, S. Lobregt, R. Truyen, and V. Pekar, "Shape constrained deformable models for 3D medical image segmentation," in *Proc. 17th International Conference on Information Processing in Medical Imaging – IPMI 2001*, ser. Lect. Notes in Comput. Sci., G. Goos, J. Hartmanis, and J. van Leeuwen, Eds., vol. 2082. Davis, CA, USA: Springer Berlin Heidelberg, Jun. 2001, pp. 380–387.
- [5] S. Kadoury, H. Labelle, and N. Paragios, "Automatic inference of articulated spine models in CT images using high-order Markov random fields," *Med. Image Anal.*, vol. 15, no. 4, pp. 426–437, Aug. 2011.
- [6] S. Kadoury, H. Labelle, and N. Paragios, "Spine segmentation in medical images using manifold embeddings and higher-order MRFs," *Med. Image Anal.*, vol. 32, no. 7, pp. 1227–1238, Jul. 2013.



- [7] P. Lim, U. Bagci, and L. Bai, "Introducing Willmore flow into level set segmentation of spinal vertebrae," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 115–122, Jan. 2013.
- [8] J. Ma and L. Lu, "Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1072–1083, Sep. 2013.
- [9] A. Rasoulzadeh, R. Rohling, and P. Abolmaesumi, "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+pose model," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1890–1900, Oct. 2013.
- [10] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "Shape representation for efficient landmark-based segmentation in 3-D," *IEEE Trans. Med. Imag.*, vol. 33, no. 4, pp. 861–874, Apr. 2014.
- [11] K. Weiss, J. Storrs, and R. Banto, "Automated spine survey iterative scan technique," *Radiology*, vol. 239, no. 1, pp. 255–262, Apr. 2006.
- [12] Y. Otake, S. Schafer, J. Stayman, W. Zbijewski, G. Kleinszig, R. Graumann, A. Khanna, and J. Siewerdsen, "Automatic localization of vertebral levels in X-ray fluoroscopy using 3D-2D registration: a tool to reduce wrong-site surgery," *Phys. Med. Biol.*, vol. 57, no. 17, pp. 5485–5508, Sep. 2012.
- [13] S. Huang, Y. Chu, S. Lai, and C. Novak, "Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI," *IEEE Trans. Med. Imag.*, vol. 28, no. 10, pp. 1595–1605, Oct. 2009.
- [14] M. Schwier, T. Chitiboi, T. Hülhagen, and H. Hahn, "Automated spine and vertebrae detection in CT images using object-based image analysis," *Int. J. Num. Meth. Biomed. Eng.*, vol. 29, no. 9, pp. 938–963, Sep. 2013.
- [15] B. Glocker, J. Feulner, A. Criminisi, D. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Proc. 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*, ser. Lect. Notes Comput. Sci., N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., vol. 7512. Nice, France: Springer Berlin Heidelberg, Oct. 2012, pp. 590–598.
- [16] B. Glocker, D. Zikic, E. Konukoglu, D. Haynor, and A. Criminisi, "Vertebrae localization in pathological spine CT via dense classification from sparse annotations," in *Proc. 16th International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2013*, ser. Lect. Notes Comput. Sci., K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., vol. 8150. Nagoya, Japan: Springer Berlin Heidelberg, Sep. 2013, pp. 262–270.
- [17] B. Kelm, M. Wels, S. Zhou, S. Seifert, M. Suehling, Y. Zheng, and D. Comaniciu, "Spine detection in CT and MR using iterated marginal space learning," *Med. Image Anal.*, vol. 17, no. 8, pp. 1283–1292, Dec. 2013.
- [18] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1668–1681, Nov. 2008.
- [19] L. Long and G. Thoma, "Use of shape models to search digitized spine X-rays," in *Proc. 13th IEEE Symposium on Computer-Based Medical Systems - CBMS 2000*. Houston, TX, USA: IEEE, Jun. 2000, pp. 255–260.
- [20] B. Howe, A. Gururajan, H. Sari-Sarraf, and L. Long, "Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized Hough transform and extensions to active appearance models," in *Proc. 6th IEEE Southwest Symposium on Image Analysis and Interpretation - SSIAP 2004*. Lake Tahoe, NV, USA: IEEE, Mar. 2004, pp. 182–186.
- [21] S. Koopman, K. Hua, and C. Bhadrakom, "Automatic classification system for lumbar spine X-ray images," in *Proc. 19th IEEE International Symposium on Computer-Based Medical Systems - CBMS 2006*. Salt Lake City, UT, USA: IEEE, Jun. 2006, pp. 213–218.
- [22] M. Larhmam, M. Benjelloun, and S. Mahmoudi, "Vertebra identification using template matching model and K-means clustering," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 9, no. 2, pp. 177–187, Mar. 2014.
- [23] D. Štern, B. Likar, F. Pernuš, and T. Vrtovec, "Automated detection of spinal centrelines, vertebral bodies and intervertebral discs in CT and MR images of lumbar spine," *Phys. Med. Biol.*, vol. 55, no. 1, pp. 247–264, Jan. 2010.
- [24] T. Cootes, C. Taylor, and A. Lanitis, "Active shape models: evaluation of a multi-resolution method for improving image search," in *Proc. 5th British Machine Vision Conference - BMVC 1994*, E. Hancock, Ed. York, UK: BMVA Press, Sep. 1994, pp. 327–336.
- [25] C. Lampert, M. Blaschko, and T. Hofmann, "Beyond sliding windows: object localization by efficient subwindow search," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2008*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
- [26] M. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Comput. J.*, vol. 7, no. 2, pp. 155–162, 1964.
- [27] N. Hansen, "The CMA evolution strategy: a comparing review," in *Towards a New Evolutionary Computation: Advances in the Estimation of Distribution Algorithms*, J. Lozano, P. Larrañaga, I. Inza, and E. Bengioetxea, Eds. Springer Berlin Heidelberg, 2006, pp. 75–102.
- [28] E. Cheney and W. Light, *A Course in Approximation Theory*. Providence, RI, USA: American Mathematical Society, 2009.
- [29] M. Griebel and M. Holtz, "Dimension-wise integration of high-dimensional functions with applications to finance," *J. Complex.*, vol. 26, no. 5, pp. 455–489, Oct. 2010.
- [30] H.-J. Bungartz, *Finite Elements of Higher Order on Sparse Grids*. Aachen, Germany: Shaker Verlag GmbH, 1998.
- [31] T. Gerstner and M. Griebel, "Dimension-adaptive tensor-product quadrature," *Computing*, vol. 71, no. 1, pp. 65–87, Aug. 2003.
- [32] A. Klimke and B. Wohlmuth, "Algorithm 847: Spinterp: piecewise multilinear hierarchical sparse grid interpolation in MATLAB," *ACM Trans. Math. Softw.*, vol. 31, no. 4, pp. 561–579, Dec. 2005.
- [33] R. Kimmel and A. Bruckstein, "Regularized Laplacian zero crossings as optimal edge integrators," *Int. J. Comput. Vis.*, vol. 53, no. 3, pp. 225–243, Jul. 2003.
- [34] P. Viola and M. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
- [35] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
- [36] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [37] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou, "Towards robust and effective shape modeling: sparse shape composition," *Med. Image Anal.*, vol. 16, no. 1, pp. 265–277, Jan. 2012.
- [38] I. Dryden and K. Mardia, *Statistical Shape Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 1998.
- [39] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bur. Stand.*, vol. 49, no. 6, pp. 409–436, Dec. 1952.
- [40] R. Korez, B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "Interpolation-based shape-constrained deformable model approach for segmentation of vertebrae from CT spine images," in *Proc. 2nd MICCAI Workshop & Challenge on Computational Methods and Clinical Applications for Spine Imaging - CSI 2014*, J. Yao, B. Glocker, T. Klinder, and S. Li, Eds., Boston, MA, USA, Sep. 2014, pp. 208–213.
- [41] J. Yao, J. Burns, H. Munoz, and R. Summers, "Detection of vertebral body fractures based on cortical shell unwrapping," in *Proc. 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012*, ser. Lect. Notes Comput. Sci., N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., vol. 7512. Nice, France: Springer Berlin Heidelberg, Oct. 2012, pp. 509–516.
- [42] W. Lorensen and H. Cline, "Marching cubes: a high resolution 3D surface construction algorithm," *Comput. Graph.*, vol. 21, no. 4, pp. 163–169, Jul. 1987.
- [43] M. Botsch and L. Kobbelt, "A remeshing approach to multiresolution modeling," in *Proc. 2nd Eurographics/ACM SIGGRAPH Symposium on Geometry Processing - SGP '04*, Nice, France, Jul. 2004, pp. 185–192.
- [44] A. Myronenko and X. Song, "Point set registration: coherent point drift," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
- [45] T. Yoo, M. Ackerman, W. Lorensen, W. Schroeder, V. Chalana, S. Aylward, D. Metaxas, and R. Whitaker, "Engineering and algorithm design for an image processing API: a technical report on ITK - the Insight Toolkit," in *Proc. of 10th Conference on Medicine Meets Virtual Reality 02/10*, ser. Stud. in Health Technol. Informat., J. Westwood, H. Miller Hoffman, R. Robb, and D. Stredney, Eds., vol. 85. Amsterdam, The Netherlands: IOS Press, Jan. 2002, pp. 586–592.
- [46] C. Seiffert, T. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: a hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.