

Statistical Interspace Models (SIMs): Application to Robust 3D Spine Segmentation

Isaac Castro-Mateos*, *Student Member, IEEE*, Jose M. Pozo, Marco Pereañez, Karim Lekadir, Aron Lazary, and Alejandro F. Frangi, *Fellow, IEEE*

Abstract—Statistical shape models (SSM) are used to introduce shape priors in the segmentation of medical images. However, such models require large training datasets in the case of multi-object structures, since it is required to obtain not only the individual shape variations but also the relative position and orientation among objects. A solution to overcome this limitation is to model each individual shape independently. However, this approach does not take into account the relative position, orientations and shapes among the parts of an articulated object, which may result in unrealistic geometries, such as with object overlaps. In this article, we propose a new Statistical Model, the Statistical Interspace Model (SIM), which provides information about the interaction of all the individual structures by modeling the interspace between them. The SIM is described using relative position vectors between pair of points that belong to different objects that are facing each other. These vectors are divided into their magnitude and direction, each of these groups modeled as independent manifolds. The SIM was included in a segmentation framework that contains an SSM per individual object. This framework was tested using three distinct types of datasets of CT images of the spine. Results show that the SIM completely eliminated the inter-process overlap while improving the segmentation accuracy.

Index Terms—Inter-process overlap, multi-object, statistical shape models, vertebral segmentation.

I. INTRODUCTION

STATISTICAL Shape Models (SSM) [1] are commonly employed to extract the shape characteristics of a certain population. However, the direct use of such models to characterize multi-object or complex structures requires a large training dataset to properly represent new cases. The reason is that not only the shape of each individual substructure must be

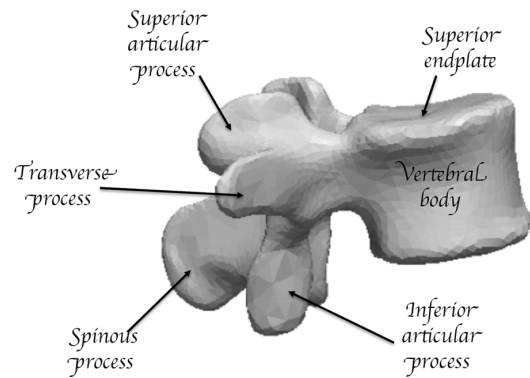


Fig. 1. Anatomy of a human vertebra.

modeled, but also the relative positions and orientations among all of them. A possible solution is to use an individual SSM per object. However, this strategy ignores the inter-relationships between them, which can involve both shape and pose, specially for the neighbour structures. This problem may be solved by the use of conditional or hierarchical models [2], [3] that contain information about the shape inter-relationship between objects. However, these approaches do not specifically model the variability of the regions and relative positions between near-contact structures in neighbour objects. This is reflected by the appearance of:

- Overlaps.
- Excessive separation of neighbouring structures.
- Unrealistic orientations between neighbouring objects.

In this article, we introduce new statistical model type coined statistical interspace model (SIM), introduced in Section IV, embedded into a new segmentation framework, Section V. SIM directly addresses the neighboring relationship between regions in different objects (or even in the same) of a multi-object structure by learning the statistical distribution of the interspace between them. SIM completely avoids overlaps, and imposes soft constraints on the acceptable poses and shapes, solving the three aforementioned issues.

This model may be applied in different applications concerning segmentation, registration, reconstruction, etc. For example, it may be employed to model the interspace between the two hemisphere of the brain, to restrain the separation between articulated bones, to control the relative position of the annulus and the nucleus of the Intervertebral discs (IVDs) or to maintain under control the separation between the two lungs to achieve realistic configurations, among possible applications.

As proof of concept, we tested the proposed framework on vertebra segmentation, since they are complex bodies (Fig. 1)

Manuscript received April 11, 2015; revised May 31, 2015; accepted June 02, 2015. Date of publication June 11, 2015; date of current version July 29, 2015. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant 269909, MySpine project. Asterisk indicates corresponding author.

*I. Castro-Mateos is with the Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Electronic and Electrical Engineering Department, The University of Sheffield, S1 3JD Sheffield, U.K. (e-mail: isaac.casm@sheffield.ac.uk).

J. Pozo and A. F. Frangi are with the Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Electronic and Electrical Engineering Department, The University of Sheffield, S1 3JD Sheffield, U.K. (e-mail: j.pozo@sheffield.ac.uk; a.frangi@sheffield.ac.uk).

M. Pereañez and K. Lekadir are with the Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Universitat Pompeu Fabra, 08018 Barcelona, Spain.

A. Lazary is with the National Center for Spinal Disorders, H-1126 Budapest, Hungary.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2015.2443912

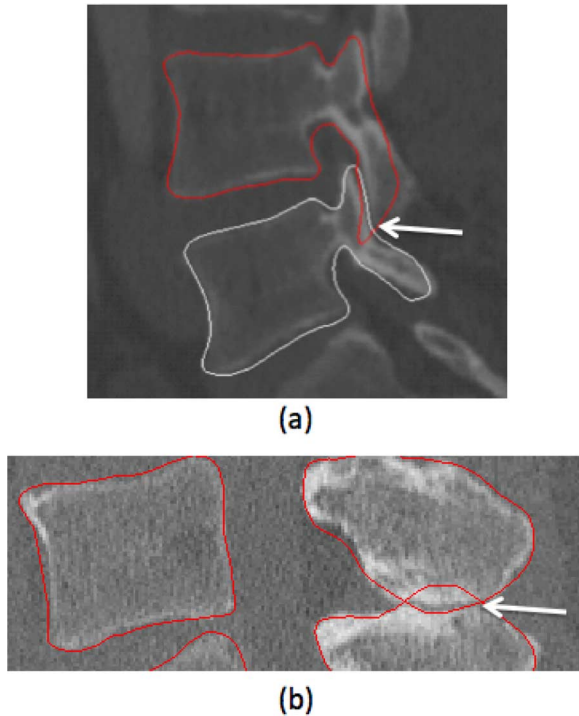


Fig. 2. The overlap in the vertebral process is a common problem in vertebral segmentation, mainly in the articular processes (a) where the separation is rather small but it can also happen in the spinous process as shown in (b).

that form a multi-object structure when combined to constitute the vertebral column. We aim to not only provide high accurate segmentations but also to handle the problem of inter-process or vertebral body overlap (Fig. 2) as well as to avert unrealistic configurations. This issue is important for a range of application such as patient-specific biomechanical modeling [4], [5] or computer-assisted spine surgery [6]. For biomechanical models the correct definition of the processes and its correct geometry are essentials to perform realistic simulations; even high accurate segmentation will not be acceptable, if they have overlaps. In spinal fusion surgery, the correct delineation of processes and correct definition of the interspace between them is crucial for the precise placement of the screws.

The remaining of the paper is structured as followed. Section II aims at providing a literature review of methods that employ overlap control or could be used for this purpose. Section III introduces the type of shape parametrization that will be used along this paper. These shapes will have an important role in each of the energies that will be used in the segmentation framework (Section V). The SIM is introduced in Section IV, where its construction procedure is depicted. The results of the method applied in the segmentation of vertebrae is presented in Section VII. The last two sections of this article are the Discussion and conclusions.

II. BACKGROUND

A. Zero Overlap Segmentation Methods

Overlap between objects may be avoided by the use of multi-label strategy classification, where each image pixel or voxel only has one value [7], [8]. These methods, however, do not

necessarily employ the information of the relative position between objects and therefore, the vertebral inter-process space might be wrongly segmented.

Another general methodology that produces zero overlap segmentations is labels propagation methods [9], [10], which performs the segmentation by warping a previously segmented image (source) into the space of the to-be-segmented image (target). This approach is generally employed with several sources and the results are combined to increase the accuracy. However, small regions such as the inter-process vertebral space might be ruled out in an image to image registration and a special treatment must be employed. In addition, these methods are usually computationally expensive. The same issues may be applied to level set segmentation frameworks [11], [12].

B. Vertebra Segmentation Methods

The segmentation of vertebrae is an important first stage for many applications as previously explained. Thus, there is a vast literature of segmentation from X-rays and Dual-energy X-ray Absorptiometry (DXA) images [13]–[16], [2], [17], [18], from CT images [19]–[28], [12], and from MR images [29]–[31]. However, in this section we are going to focus on the methods that directly or indirectly address the vertebral overlap.

Shen *et al.* [32] proposed a method to avert vertebrae from leaking to other structures using information extracted from detected planes between vertebrae. This information was included in a level-set framework. Kim *et al.* [33] employed a 3D fence to create a wall between the spine articular processes in order to be able to segment a vertebra using region growing. These two methods reduce the overlap between vertebrae by estimating the separation between them. However, they do not employ any prior information based on the training population, which could greatly improve the performance.

Prior information to avoid vertebral intersection may be acquired by constructing an SSM of the full spine as a single object. However, as mentioned earlier this model would require a large training dataset to achieve sufficient degrees of freedom. Although, it is possible the addition of relaxation steps, based on Active Contour Models [15] or Free Form Deformations [34], the segmentation will commonly leak into neighboring structures or will be deformed more than required when the best fit of the SSM do not closely resemble the real shape.

A more elegant solution was proposed by Kindler *et al.* [35]. They designed a vertebral segmentation algorithm based on a statistical pose model (SPM) of the spine, following the work of Boisvert *et al.* [36]. In this framework, shapes are deformed by displacing points in order to minimize an energy based on the image gradient and penalized by the distance to the pose given by the SPM. Since, inter-process overlap is reduced but not completely avoided, the authors added a new energy term that penalizes the overlap between vertebrae. However, it still does not prevent them completely. Our assumption is that in doing so the accuracy would decrease due to the SPM excessively restricting the possible outcomes. An analogous strategy to avoid overlaps and maintain under control the distance between structures was proposed in [37]. The approach was based on a modified spring energy employed in a mesh-to-mesh registration framework tested on wrist bones. A similar approach as

using the SPM was proposed by Kadoury *et al.* [38], in which the relationship between the individual shape variations are correlated to the global deformations with a series of rigid transformations and embedded into a Markov random field segmentation framework.

In [39] a model of the full spine was employed. However, the relative rotations and translations given by the spine articulated movement were removed and a SPM was trained for that purpose. Even though, this approach still requires a larger training dataset.

In an earlier work [40], we proposed a segmentation strategy for vertebral segmentation. The process employs an Active Shape Model approach [1], with a novel set of trained features for the shape to image adaptation step to partially address the overlap problem.

III. PARAMETERIZATION OF THE SHAPES

A. Point Correspondence

The point correspondence represents the idea that shapes are described by the same number of landmarks, \mathbf{s}_p , $p = 1, \dots, n$, and those landmarks are located in the same anatomical position.

The complete set of points concatenated in a $3 \times n$ matrix will be denoted as \mathbf{S} , assuming 3D shapes.

This concept is essential to build SSMs and SIMs. In this work, the point correspondence was achieved by an iterative registration approach, based on B-splines. This procedure began deforming a template mesh to all the shape meshes of the training set. Thereafter, the mean shape at each spinal position was computed. Subsequently, the template mesh was changed by the mean shapes and the process was iteratively repeated until the convergence of the means. The template mesh was created by smoothing and decimating a selected L3 vertebra and the ground-truth meshes were obtained using Marching cubes [41] in the binary masks achieved through manual segmentation.

B. Unconstrained Shape Instance

An unconstrained shape instance is created by the independent displacement of the landmarks that represents the shape. In the proposed segmentation framework, this type of shape instance is used only for the fitting to the image (\mathbf{S}^I), analogously to the usual ASM segmentation [1] (Section V-C).

C. Constrained Shape Instance by the Use of Statistical Shape Models (SSM)

An SSM describes the variability of a shape within a selected population. In general, apart from the point correspondence, the pose and scale of the cases are reduced by normalising to a reference shape with a similarity transformation. This normalisation is performed by generalized Procrustes alignment, which generates the mean shape, $\bar{\mathbf{S}}$, simultaneously with the aligned shapes.

Each aligned shape may be considered as being a sample from a $3n$ probability density function (pdf). Assuming this pdf takes the form of a Gaussian distribution [1], we can use Principal Component Analysis (PCA) to attain the main axes of the cloud of points. The SSM represents new shapes by the mean, $\bar{\mathbf{S}}$, and a

linear combination of a set of orthonormal principal components (PC), Φ_m , $m = 1, \dots, M$.

$$\hat{\mathbf{S}}^{\text{SSM}} = \bar{\mathbf{S}} + \sum_{m=1}^M b_m \Phi_m \quad (1)$$

where the coefficients b_m are called the *SSM parameters* and the relative influence of each PC is given by its corresponding standard deviations, σ_m . In addition, the b_m parameters are limited to 3 times σ_m ,

$$-3\sigma_m \leq b_m \leq 3\sigma_m, \quad \forall m = 1, \dots, M.$$

New instances, including pose and scaling, may be attained by applying a similarity transformation to the shape $\hat{\mathbf{S}}^{\text{SSM}}$:

$$\mathbf{S}^{\text{SSM}} = a \mathbf{R} \hat{\mathbf{S}}^{\text{SSM}} + \mathbf{c}. \quad (2)$$

Thus, the instance is a function of the SSM parameters, b_m , a scaling factor, a , a translation vector \mathbf{c} , and a rotation matrix, \mathbf{R} .

For multiple objects, each with an individual SSM and similarity transformation, the model is:

$$\begin{aligned} \hat{\mathbf{S}}_v^{\text{SSM}} &= \bar{\mathbf{S}}_v + \sum_{m=1}^M b_{v,m} \Phi_{v,m} \\ \text{and } \mathbf{S}_v^{\text{SSM}} &= a_v \mathbf{R}_v \hat{\mathbf{S}}_v^{\text{SSM}} + \mathbf{c}_v, \end{aligned} \quad (3)$$

where the index v runs along the set of objects, $v = 1, \dots, V$.

D. Constrained Shape Instance by the Use of B-Splines

In addition to the previous instances, we also consider, \mathbf{S}_v^{B} , which are given by smooth B-spline transformation from any chosen reference shape, \mathbf{S}_v^0 , parametrized by displacements, $\mathbf{z}_{v,f}$, at a grid of control points, $f = 1, \dots, F$. Each point in \mathbf{S}_v^{B} is then located at

$$\mathbf{s}_{v,p}^{\text{B}} = \mathbf{s}_{v,p}^0 + \sum_{f=1}^F B_f(\mathbf{s}_{v,p}^0) \mathbf{z}_{v,f} \quad (4)$$

where the B_f are the B-spline weights.

The output of the segmentation will be a B-spline instance. The advantage of this type of shape instance is twofold. First, it allows more flexibility than that of the SSM. In fact, this additional flexibility is required for the incorporation of the SIM into the segmentation framework, otherwise the SIM would restrict a large region of the SSM parameters values as incompatible. Second, the use of a B-spline instance ensures smooth segmentations, in contrast with the unconstrained type, that would require an extra regularization term [42].

IV. STATISTICAL INTERSPACE MODEL (SIM)

For an articulated multi-object structure, individual SSMs together with an SPM provide a complete set of parameters to generate instances of this object. However, using only these two models does not prevent unrealistic configurations, including overlaps, excessive separation between objects or unnatural orientations of neighbouring substructures. In the spine, for example, this affects the vertebral processes, which can intersect between them. Thus, low signal or high noise within the image

intensity may lead to incorrect segmentations including these artefacts.

A way to partially overcome these limitations is by the use of spring models, which are employed to maintain deformations under control by penalising large distance to a reference position. For instance, the approach of Klinder *et al.* [35] employs this type of model as part of the energy problem to handle vertebral overlaps.

However, the drawback of spring models for this purpose is twofold. First, they cannot guaranteed zero-overlap, in fact, they only reduce the amount of overlapped volume. Second, they do not model the statistical variation in size of the separation in the population.

An elegant solution to handle this limitation is to model the interspace between the objects. Therefore, the change in variation of an object O^1 affects the interspace between this object and its neighbour O^2 , which produces O^2 to change in shape. In order to model the interspace between object we employ the relative position vector between landmarks of neighbouring objects. Consider two points $(\mathbf{o}_d^1, \mathbf{o}_d^2)$ from two different objects or from the same one. The relative position vector δ_d is a vector that starts in \mathbf{o}_d^1 and finishes in \mathbf{o}_d^2 .

Therefore, the first step to construct the SIM is to select the points \mathbf{s}_d with $d = 1 \dots D$ being an index running all the points for which the SIM is going to be computed. In the mean shape, for each landmark \mathbf{s}_d , it is necessary to find the closest landmark that belongs to the object that is sharing the interspace with. These pair of landmark connection are propagated across all the training set and the relative position vector between them are employed for the SIM construction.

A. Individual Point Reference Frame

The computation of δ_d require a reference frame $(\hat{\mathbf{g}}_d^1, \hat{\mathbf{g}}_d^2, \hat{\mathbf{g}}_d^3)$ that is irrespective to the object shape and pose (Fig. 3). The reason for this is that we want to measure a vector that goes from one landmark to another and not only a magnitude.

In this work, we propose $\hat{\mathbf{g}}_d^3$ to be the normal vector of the object surface at point \mathbf{s}_d . For all cases, an edge is selected in the triangular mesh from \mathbf{s}_d to a neighbour point, defining the vector \mathbf{t} . The second vector is obtained by projecting \mathbf{t} in the perpendicular plane given by $\hat{\mathbf{g}}_d^3$:

$$\hat{\mathbf{g}}_d^2 = \mathbf{t} - (\mathbf{t} \cdot \hat{\mathbf{g}}_d^3) \hat{\mathbf{g}}_d^3, \quad \hat{\mathbf{g}}_d^1 = \frac{\hat{\mathbf{g}}_d^2}{\|\hat{\mathbf{g}}_d^2\|}. \quad (5)$$

And $\hat{\mathbf{g}}_d^1$ is their cross product $\hat{\mathbf{g}}_d^1 = \hat{\mathbf{g}}_d^2 \times \hat{\mathbf{g}}_d^3$.

Having the reference frame $(\hat{\mathbf{g}}_d^1, \hat{\mathbf{g}}_d^2, \hat{\mathbf{g}}_d^3)$ allows us to define the relative position vector between the different pair of points, in a coherent way among all the sample cases. Thus, the local frame provides local relative position correspondence, analogous to the point correspondence required for the landmarks in the SSM.

B. Separation Between Magnitude and Directional Vector

Having computed the relative position vector δ_d for each landmark \mathbf{s}_d in the SIM, a statistical interspace model might be created in the same fashion as the standard SSM. However, this approach would assume a Gaussian distribution (linear approach) on the relative position vectors of x , y and z . Since, a Gaussian shape is symmetric with respect to its mean position,

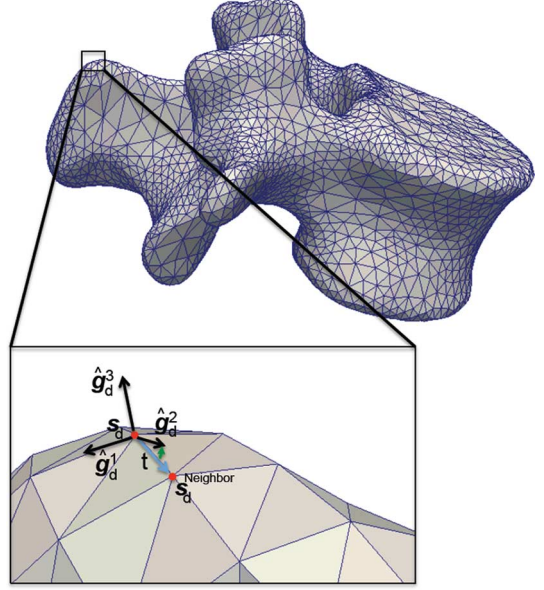


Fig. 3. Description of the local reference system for the point \mathbf{s}_d . The vector \mathbf{t} goes from \mathbf{s}_d to a previously selected neighbor ($\mathbf{s}_d^{\text{Neighbour}}$). The vector $\hat{\mathbf{g}}_d^3$ is the normal of the object surface at point \mathbf{s}_d , whereas $\hat{\mathbf{g}}_d^2$ is obtained by projecting \mathbf{t} onto the surface tangent plane perpendicularly to $\hat{\mathbf{g}}_d^3$. Finally, $\hat{\mathbf{g}}_d^1$ is the cross product of $\hat{\mathbf{g}}_d^2$ and $\hat{\mathbf{g}}_d^3$.

the consequence of this assumption is that negative distances are allowed, which results in the overlaps and other unrealistic geometries.

Thus, it is necessary to find a non-linear representation that properly models δ_d . Since, our aim is to control the magnitude and orientation of δ_d , a proper representation was the separation of δ_d into its magnitude γ_d and a unit vector ω_d that represents the orientation.

$$\delta_d = \gamma_d \omega_d. \quad (6)$$

This separation maps $\mathbb{R}^3 \rightarrow \mathbb{R}^+ \otimes S^2$. Thus, both factors \mathbb{R}^+ and S^2 are manifolds.

In particular, \mathbb{R}^+ is a Lie group for the real product, which endows it with a natural Riemannian geometry. Therefore, the tangent space of the magnitude is the natural logarithm of real positive numbers. The implication of using such model in an energy-driven segmentation is that the gap between the objects cannot be zero. Thus, two landmarks from different objects cannot share the same position. However, two objects could touch each other or even have an overlap with only this energy, which represents the logarithm of the distance between the points. This energy tries to maintain the points close to any position in the surface of a sphere with radius equal to the mean distance between the landmarks that were used for training the model.

The directional vector was modeled as the space S^2 , which may be seen as a unitary sphere when embedded into a Euclidean space. In this case, the tangent space used for computing all the measurements, was selected as the tangent space at the center position of the training samples in S^2 . This implies that we consider the arc-lengths from each training sample to the mean vector, to belong to a Gaussian distribution. The addition of this model ensures that each pair of points is closely aligned

to the direction given by the mean vector. However, this energy does not guaranteed a zero overlap in the segmentation process, since it measures the arc-length between two points in S^2 , whose maximum value is π . Notwithstanding, an overlap implies rotations close to π degrees which will increase the energy substantially. Although, we did not find any overlap in our database, a solution to a possible one is employing the same idea as in Active Shape Model (ASM) [1], where deformations were restricted to a maximum of $\pm 3 \times$ the standard deviation, in their case given by the eigenvalues. Another solution, is to extend the values so that π represents the infinity.

The separation between magnitude and unitary vector was tested by different Gaussianity tests. To prove that magnitude samples lie in a log-gaussian distribution, the ‘‘Lilliefors’ composite goodness-of-fit test’’ [43] and the ‘‘Jarque-Bera hypothesis test of composite normality’’ [44] were employed using as input the logarithm of the distances. Both of them could not reject the null hypothesis of the data being normally distributed at the 5% significance level. In the case of the unitary vector, our assumption is that the data in the tangent space at the mean position behaves as Gaussian. Since, we are facing a 2-dimensional distribution, the same tests were not available. Thus, we employed the test ‘‘Henze-Zirkler’s Multivariate Normality Test’’ [45], which could not reject the null hypothesis of the data being normally distributed at the 5% significance level.

Observe that SIM is not a generative model, meaning it cannot create an instance of the objects by itself. However, when it is employed with an SSM (or SSM and SPM) prevents impossible configurations and geometries in the instances.

C. Mean and Covariance Computations

In order to model the interspace between objects, the following four statistics are required: The mean of ω in $S^2(\bar{\omega}_d)$, the covariance matrix of ω in $T_{\bar{\omega}}(S^2)(C_d)$, the mean magnitude ($\bar{\gamma}_d$) and the variance of the magnitude l_d .

In order to compute the mean and covariance in a non-euclidean manifold, we adopted the methodology proposed by Fletcher *et al.* [46] to compute the Principal Geodesic Analysis (PGA). Their method assumes that the tangent space at the mean position of all the samples (μ) behaves as an Euclidean space. In this proposed method, the mentioned samples are going to be the individual magnitude (γ) and the individual directional vectors (ω) of each δ_d . Thus, $\mu = \{\bar{\omega}, \bar{\gamma}\}$ and it is defined as

$$\mu = \arg \min_u \sum_q D(u_v, u)^2 \quad (7)$$

where D denotes the geodesic distance between two points in the manifold. The subindex v runs along the different cases in the dataset. In this context, u would be either γ_d or ω_d .

The intrinsic mean (μ) was computed following the approach proposed by Fletcher *et al.* [46]. The process begins by selecting a random sample as the mean (μ) and projecting all the samples to the tangent space at this point ($T_{\mu}(M)$). In this space the mean of the projected samples (κ) is computed and transformed back to $S^2(Exp_{\mu}(\kappa))$ to obtain the new μ . This process is repeated until convergence.

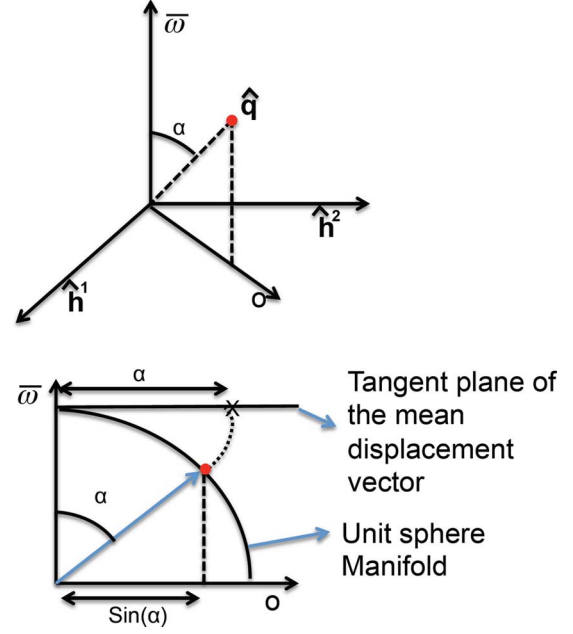


Fig. 4. Description of the projection of a point \hat{q} that belongs to the S^2 manifold to $T_{\bar{\omega}}(S^2)$. The plane defined by the vectors \hat{h}^1 and \hat{h}^2 is parallel to $T_{\bar{\omega}}(S^2)$. Then, $\text{Log}_{\bar{\omega}}(\hat{q})$ is the orthogonal projection of \hat{q} to $T_{\bar{\omega}}(S^2)$, rescaled to have the angle α as length, instead of $\sin \alpha$.

Since the magnitude belongs to the Lie group \mathbb{R}^+ , the distance function is defined as:

$$D(\gamma_d^1, \gamma_d^2) = \log \frac{\gamma_d^1}{\gamma_d^2}. \quad (8)$$

However, since we are facing a 1-dimensional problem, $\bar{\gamma}$ and l_d can be analytically computed as the standard mean and variance of all the log-magnitudes [47].

For two directions, the geodesic distance in S^2 is given by the arc length between them:

$$D(\omega_d^1, \omega_d^2) = a \cos(\omega_d^1 \cdot \omega_d^2) \quad (9)$$

where $\omega_d^1 \cdot \omega_d^2$ is the dot product of ω_d^1 and ω_d^2 .

For the sphere S^2 , it is necessary to define the manifold logarithm that projects a sample in S^2 into the tangent space, $T_{\bar{\omega}}S^2$. This operation requires two orthonormal vectors (\hat{h}^1, \hat{h}^2) to $\bar{\omega}$. Their orientation in the tangent plane is irrelevant, but they need to be consistent across the training dataset. We define the reference frame ($\hat{h}^1, \hat{h}^2, \hat{h}^3$) as the frame that results after rotating ($\hat{g}^1, \hat{g}^2, \hat{g}^3$) so that \hat{h}^3 coincides with $\bar{\omega}$. The logarithm is then given by:

$$\text{Log}_{\bar{\omega}}(\hat{q}) = (\hat{q} \cdot \hat{h}^1 \frac{\alpha}{\sin(\alpha)}, \hat{q} \cdot \hat{h}^2 \frac{\alpha}{\sin(\alpha)}) \quad (10)$$

with $\alpha = a \cos(\hat{q} \cdot \bar{\omega})$.

Observe that by identifying $T_{\bar{\omega}}S^2$ as the tangent plane to S^2 embedded in R^3 , the output of the logarithm is the projection of \hat{q} into $T_{\bar{\omega}}S^2$ rescaled so that its length is equal to α (Fig. 4).

Once each sample point in S^2 has been identified with its logarithm in the tangent space, the usual sample covariance matrix may be computed as that of an Euclidean space.

V. SEGMENTATION FRAMEWORK

A. Initialisation

As commonly required in segmentation procedures, our framework needs a rough initialization of the shape in the image. For a multi-object structure, this may be performed by locating each individual object independently or with some constrains. For our application to vertebral segmentation, we have applied the same initialization method as in [40]. The mean shape of each vertebra, $\bar{\mathbf{S}}_v$, is individually aligned to the image by means of a similarity transformation. This transformation is based on the approximate intervertebral discs (IVD) center points that were manually pinpointed in the image. During the training step, the upper and lower endplate middle landmarks, of the mean meshes at each vertebral position, were selected. These two points were employed to calculate the center of the vertebral body (VB), \mathbf{c}^1 and the directional vector from the lower to the upper endplate, \mathbf{u}^1 , of each initial vertebra.

In the segmentation phase, the IVD centers in the image were manually selected. These points define the centre of the VB, \mathbf{c}^2 , in the image and a new vector that goes from the lower IVD to the upper one in each vertebra \mathbf{u}^2 . Then, a translation vector is given by the difference between \mathbf{c}^2 and \mathbf{c}^1 , the scaling factor (e) is given by a scaled ratio of the magnitudes of \mathbf{u}^2 and \mathbf{u}^1 and the rotation matrix (\mathbf{R}^o) was computed using the Rodrigues' rotation formula [36]:

$$e = k \frac{\|\mathbf{u}^2\|}{\|\mathbf{u}^1\|} \quad (11)$$

$$\mathbf{R}^o = \mathbf{U} + \mathbf{B} + \frac{\mathbf{B}\mathbf{B}}{1 + \hat{\mathbf{u}}^1 \cdot \hat{\mathbf{u}}^2} \text{ and } \mathbf{B} = \hat{\mathbf{u}}^1 \hat{\mathbf{u}}^{2,T} - \hat{\mathbf{u}}^2 \hat{\mathbf{u}}^{1,T} \quad (12)$$

where \mathbf{U} is the identity matrix and the scaling factor k ($k = 0.9$) was used to compensate the difference between the real height of the VB and the distance between two consecutive IVD centers.

B. Statistical Shape Model Energy

The statistical shape model energy is commonly expressed as the squared Euclidean distance between the closest shape that can be represented with the SSM, \mathbf{S}^{SSM} , and the segmentation result, in our case \mathbf{S}^{B} .

$$E^{\text{SSM}}(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{SSM}}) = \frac{1}{2} \sum_{v,p} \|\mathbf{s}_{v,p}^{\text{B}} - \mathbf{s}_{v,p}^{\text{SSM}}\|^2 \quad (13)$$

where the squared distance is summed up for all points and objects.

C. Image Energy

The image energy is computed for each landmark independently following the work in [40]. Thus, the resultant shape \mathbf{S}^{I} is an unconstrained shape instance (Section III-B). This image energy is based on the oriented derivative from high to low intensity and the intensity itself, whose parameters were obtained using the training database, explained in Section VI. Thus, the energy function for these features is:

$$E_{v,p}^{\text{I}}(\mathbf{S}^{\text{I}}, \mathbf{I}) = (w^1 I(\mathbf{s}_{v,p}^{\text{I}}) + w^2 (\hat{\mathbf{n}} \cdot \nabla I(\mathbf{s}_{v,p}^{\text{I}}))) \quad (14)$$

where $\hat{\mathbf{n}}$ is the normal direction at point $\mathbf{s}_{v,p}^{\text{B}}$ from inside to outside the mesh, I is the image intensity.

The image energy has a second component that relates the result segmentation shape \mathbf{S}^{B} and \mathbf{S}^{I} :

$$E^{\text{I-B}}(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{I}}) = \frac{1}{2} \sum_{v,p} \|\mathbf{s}_{v,p}^{\text{B}} - \mathbf{s}_{v,p}^{\text{I}}\|^2. \quad (15)$$

D. Statistical Interspace Model Energy

Having computed the SIM model for a group of objects means that for each relative position vector δ_d there exist a $\bar{\gamma}_d$, l_d , $\bar{\omega}_d$ and \mathbf{C}_d (Section IV). Hence, the energy of a new δ_d may be achieved by:

$$E^{\text{SIM}} = \frac{N}{2D} \sum_d \epsilon_d \mathbf{C}_d^{-1} \epsilon_d^T + \log \left(\frac{\gamma_d}{\bar{\gamma}_d} \right)^2 l_d^{-1} \quad (16)$$

with $\delta_d = \gamma_d \omega_d$ and $\epsilon_d = \text{Log}_{\bar{\omega}}(\omega_d)$.

The coefficient $N/2D$ is used to compensate the difference in the number of points between E^{SIM} and the rest of energies. The value N is the total number of landmarks in all the whole structure and D is the number of relative position vectors (pair of points) in the structure that employs the SIM.

E. Segmentation Problem

The segmentation algorithm is based on minimising a linear combination of the previously presented energies:

$$\mathbf{S}^{\text{B}} = \arg \min_{\mathbf{S}^{\text{B}}} \min_{\mathbf{S}^{\text{I}}, \mathbf{S}^{\text{SSM}}} E(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{I}}, \mathbf{S}^{\text{SSM}}, \mathbf{I}) \text{ where } \quad (17)$$

$$E(\mathbf{S}^{\text{B}}, \mathbf{I}, \mathbf{S}^{\text{I}}, \mathbf{S}^{\text{SSM}}) = E^{\text{I}}(\mathbf{S}^{\text{I}}, \mathbf{I}) + w^{\text{B-I}} E^{\text{B-I}}(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{I}}) + w^{\text{SSM}} E^{\text{SSM}}(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{SSM}}) + w^{\text{SIM}} E^{\text{SIM}}(\mathbf{S}^{\text{B}}). \quad (18)$$

Each of the shapes represent an instance of all the objects that composed the multi-object structure. \mathbf{I} denotes the image to be segmented, which is a constant for the energy minimization. \mathbf{S}^{B} is the output segmentation, whereas \mathbf{S}^{I} and \mathbf{S}^{SSM} are auxiliary shapes. Nevertheless, all three shapes must be optimized to minimize the energy, even though only \mathbf{S}^{B} is of interest.

F. Optimisation

In order to solve the aforementioned problem, we propose an alternating optimization technique, where each of the three shapes involved \mathbf{S}^{B} , \mathbf{S}^{I} , and \mathbf{S}^{SSM} are iteratively optimized in turns.

1) *Shape Initialization*: The mean shape of each object, $\bar{\mathbf{S}}_v$, is scaled, translated and rotated (similarity transformation) according to some strategy (Section V-A), attaining \mathbf{S}_0 . This provides the initialization for the three shapes: $\mathbf{S}^{\text{I}} = \mathbf{S}^{\text{B}} = \mathbf{S}^{\text{SSM}} = \mathbf{S}_0$.

2) *Alternating Optimization*: Having initialized the three shapes, they are optimized as follows:

1) Only two energies involve \mathbf{S}^{I} . Thus, this optimization only needs to minimize the terms

$$E^{\text{I}}(\mathbf{S}^{\text{I}}, \mathbf{I}) + E^{\text{B-I}}(\mathbf{S}^{\text{B}}, \mathbf{S}^{\text{I}}), \quad (19)$$

since the others are constant. This optimisation is performed at each landmark independently, following the typical strategy used in ASM. Observe that because the

optimization is alternating between shapes, at each iteration the shape \mathbf{S}^I is obtained by displacing the landmarks of the \mathbf{S}^B from the previous step and not the previous \mathbf{S}^I .

- 2) **Optimization of \mathbf{S}^{SSM} .** The only term involving \mathbf{S}^{SSM} is
- $$E^{\text{SSM}}(\mathbf{S}^B, \mathbf{S}^{\text{SSM}}). \quad (20)$$

The shape \mathbf{S}^{SSM} contains two type of parameters: the similarity transformation of the shape and the SSM parameters for the PCs. These two sets are optimized also alternating:

- Exact optimization of similarity transformation.** Fixing the SSM parameters allows the optimization of the similarity transformation by a single algebraic operation using Procrustes.
- Exact optimization of the SSM parameters.** Fixing the similarity transformation allows E^{SSM} to be optimized in the SSM parameters by projecting the shape into the PC basis.

These two operations can be repeated several times until convergence of the full \mathbf{S}^{SSM} optimization.

- 3) **Gradient descent step for the shape \mathbf{S}^B .** This shape involves three terms of the energy:

$$w^{\text{I-B}} E^{\text{I-B}}(\mathbf{S}^B, \mathbf{S}^I) + w^{\text{SSM}} E^{\text{SSM}}(\mathbf{S}^B, \mathbf{S}^{\text{SSM}}) + w^{\text{SIM}} E^{\text{SIM}}(\mathbf{S}^B). \quad (21)$$

The optimisation is achieved through a gradient step procedure. The number of iterations in the gradient step was reduced to a maximum of 5 iterations, since the full optimisation process is repeated iteratively until the convergence of the \mathbf{S}^B .

The shape \mathbf{S}^B is a linear function of the control point displacements, $\mathbf{z}_{v,f}$, as was shown in (4). The gradient involved in the optimization of the energy with respect to \mathbf{S}^B is given by the derivatives

$$\frac{\partial E}{\partial \mathbf{z}_{v,f}} = \sum_p \frac{\partial \mathbf{s}_{v,p}^B}{\partial \mathbf{z}_{v,f}} \cdot \frac{\partial E}{\partial \mathbf{s}_{v,p}^B} = \sum_p B_f(\mathbf{s}_{v,p}^0) \frac{\partial E}{\partial \mathbf{s}_{v,p}^B} \quad (22)$$

where

$$\frac{\partial E}{\partial \mathbf{s}_{v,p}^B} = w^{\text{SSM}} \frac{\partial E^{\text{SSM}}}{\partial \mathbf{s}_{v,p}^B} + w^{\text{I-B}} \frac{\partial E^{\text{I-B}}}{\partial \mathbf{s}_{v,p}^B} + w^{\text{SIM}} \frac{\partial E^{\text{SIM}}}{\partial \mathbf{s}_{v,p}^B}. \quad (23)$$

The derivatives of the first two energies are simply

$$\frac{\partial E^{\text{SSM}}}{\partial \mathbf{s}_{v,p}^B} = \mathbf{s}_{v,p}^B - \mathbf{s}_{v,p}^{\text{SSM}} \quad \text{and} \quad \frac{\partial E^{\text{I-B}}}{\partial \mathbf{s}_{v,p}^B} = \mathbf{s}_{v,p}^B - \mathbf{s}_{v,p}^I. \quad (24)$$

The derivatives of the E^{SIM} are

$$\frac{\partial E^{\text{SIM}}}{\partial \mathbf{s}_{v,p}^B} = \sum_d \frac{\partial \mathbf{s}_{v,p}^B}{\partial \boldsymbol{\delta}_d} \frac{\partial E^{\text{SIM}}}{\partial \boldsymbol{\delta}_d} \quad (25)$$

where $\partial \mathbf{s}_{v,p}^B / \partial \boldsymbol{\delta}_d$ is a signed selector, whose value is -1 if the landmark \mathbf{s}_p is the origin of the relative position vector $\boldsymbol{\delta}_d$, $+1$ if it is the end of $\boldsymbol{\delta}_d$, and 0 if it is not contain in the SIM. The second factor can be computed from (16):

$$\frac{\partial E^{\text{SIM}}}{\partial \boldsymbol{\delta}_d} = \frac{N}{2D} \left[\left(\frac{\mathbf{I}}{\gamma_d} - \frac{\boldsymbol{\delta}_d^T \boldsymbol{\delta}_d}{\gamma_d^3} \right) \mathbf{R}_d^h \mathbf{R}_d^g \frac{\partial \boldsymbol{\epsilon}_d}{\partial \boldsymbol{\omega}_d} \mathbf{C}_d^{-1} \boldsymbol{\epsilon}_d^T + \frac{\boldsymbol{\delta}_d}{\gamma_d^2} L_d^{-1} \log \frac{\gamma_d}{\bar{\gamma}_d} \right] \quad (26)$$

where

$$\frac{\partial \boldsymbol{\epsilon}_d}{\partial \boldsymbol{\omega}_d} = \frac{\alpha}{\sin \alpha} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \beta \epsilon_{d,1} & \beta \epsilon_{d,2} \end{pmatrix} \quad \text{with } \beta = \frac{\cos \alpha}{\alpha \sin \alpha} - \frac{1}{\alpha^2}, \quad (27)$$

and $\epsilon_{d,1}$, $\epsilon_{d,2}$ denote the components of $\boldsymbol{\epsilon}_d$. The rotation matrices \mathbf{R}_d^g and \mathbf{R}_d^h , transform the world coordinate system to the individual reference system $(\hat{\mathbf{g}}^1, \hat{\mathbf{g}}^2, \hat{\mathbf{g}}^3)$, and from this reference frame to $(\hat{\mathbf{h}}^1, \hat{\mathbf{h}}^2, \hat{\mathbf{h}}^3)$, respectively (Section IV-C). Both matrices and both $(\hat{\mathbf{g}}^1, \hat{\mathbf{g}}^2, \hat{\mathbf{g}}^3)$ and $(\hat{\mathbf{h}}^1, \hat{\mathbf{h}}^2, \hat{\mathbf{h}}^3)$ must be recomputed at each iteration.

VI. DATA

Three datasets were employed for the validation of the proposed segmentation framework.

A. Dataset 1: Lumbar Spine of Healthy Subjects

It is composed of 30 CT scans from patients suffering from low back, collected at the National Center for spinal disorders (Budapest, Hungary). Patients were 13 males and 17 females with a mean age of 40 (age interval: 27–62 years). Those patients were selected for participating in the European Commission funded MySpine project [48].

Image resolution: Axial in-plane resolution of $0.608 \times 0.608 \text{ mm}^2$ with a spacing between slices of 0.625 mm .

Training and testing division: This dataset was divided into a testing set, compound of 25 CT images, and a training set of 5 for the feature training [40]. The SSM and SIM were evaluated in a leave-one-out manner. Therefore, for each vertebra in the testing set a model of 29 vertebrae was employed.

Ground-truth information: The ground truth meshes were manually created using the open source software ITK-SNAP [49].

B. Dataset 2: Lumbar Spine of Pathological Subjects

This database belonged to the spine web [50] and it was collected for vertebral detection purposes (DataSet 3 in the webpage). Thus, images did not have ground-truth, the resolution and image quality was poor, and it contains a large variation of diseases and regions of the spine. Ten images were selected from this database following the criterion of ensuring that the whole lumbar region was in the image and that patients show high degree of scoliosis. In addition, 4 out of the 10 images contain at least one fracture and 3 of them contain calcifications.

Image resolution: The axial in-plane resolution was 0.35×0.35 with a spacing between slices of 2.5 mm .

Training and testing division: The whole dataset was used as testing using the training information from Dataset 1. The features remained the same as in Dataset 1 and the SSM and SIM were constructed with the 30 cases from Dataset 1.

Ground-truth information: The ground truth meshes were manually created using the open source software ITK-SNAP [49].

C. Dataset 3: Thoracic and Lumbar Spine of Healthy Subjects

This database consists of 15 thoracic and lumbar CT scans used in the MICCAI challenge 2014 ‘‘Computational Methods

and Clinical Applications for Spine Imaging—Spine and Vertebrae Segmentation” [51]. This dataset is held in the spineWeb [50] as Dataset 2.

Image resolution: Images have an axial in-plane resolution of 0.5×0.5 and a spacing between slices of 1.0 mm.

Training and testing division: The MICCAI challenge organizers arranged a group of 10 cases for training purposes and 5 for the testing stage. Thus, features, SIM and SSM were trained with 10 cases.

Ground-truth information: The ground truth meshes were provided by the organizers. However, we corrected the training set to increase its accuracy. The testing remained the same for benchmarking purposes and owing to its higher quality.

VII. RESULTS

A. Parameters Selection

The proposed framework employs an optimization algorithm with four different energies to acquire the segmentation (Section V-F).

In our experiments the weights were kept to 1 for practical reasons, since the framework is not very sensitive to these values. Notwithstanding, other applications may require a further optimization of the weights.

The second parameter is the number of grid points for the B-splines shape instance (Section III-D). In the experiments, it was shown that set of values ranging from $5 \times 10 \times 10$ to $15 \times 30 \times 30$ per vertebra performed equally well. Smaller values did not have enough flexibility and results lacked accuracy and larger values allowed non-smooth shapes. For the experiments, we employed the smallest grid, $5 \times 10 \times 10$, meaning an averaged spacing of $9.3 \times 9.75 \times 8.25 \text{ mm}^3$.

Other parameters are:

- The number of landmarks in each vertebra was 4000.
- An SSM per vertebra was employed.
- The number of modes for each SSM was selected as the number that keeps 95% of the total variation. This results in 10,10,10,10,11 when trained with Dataset 1 and 8 from T1-T9 and 7 from T10-L5 when using Dataset 3.
- The weight of the features involved in E^I and E^{I-B} were 0.5, 0.3.

B. Evaluation Metrics

The segmentations were evaluated in two different manner. First, the accuracy was measured using Dice Similarity Index (DSI) [52] between the segmented shapes and ground truths the surface-to-surface distance from the segmented shape to the Ground Truth. It was computed the mean absolute surface distance (MASD), 95-percentil (95-p) and the maximum or Hausdorff distance (HD).

Second, the overlap between the different vertebrae across the spine was measured. The employed metric for this purpose was the mean intersection volume in voxels and mm^3 in the lumbar and thoracic spine, separately, for each Dataset.

The training databases were employed to create the features, weights, SSM and SIM as explained in Section VI.

TABLE I
VERTEBRA SEGMENTATION ACCURACY IN DATASET 1
(HEALTHY LUMBAR SPINE) WITH AND WITHOUT SIM

SIM	Error measure	L1	L2	L3	L4	L5	Global
No	Mean DSI(%)	93.3	93.7	93.6	93.6	93.8	93.8
	2xSTD of DSI(%)	3.2	3.2	3.3	3.2	3.2	3.2
	MASD(mm)	0.5	0.49	0.48	0.50	0.56	0.51
	95-p(mm)	1.11	1.09	1.09	1.09	1.24	1.10
	HD(mm)	4.4	4.3	4.7	4.8	5.6	4.8
Yes	Mean DSI(%)	94.7	95.1	94.9	94.8	93.7	94.8
	2xSTD of DSI(%)	1.8	1.7	1.8	2.1	4.5	2.8
	MASD(mm)	0.48	0.46	0.48	0.47	0.52	0.48
	95-p(mm)	1.02	0.99	0.98	1.03	1.19	1.03
	HD(mm)	4.1	4.1	4.3	4.4	5.2	4.5

TABLE II
VERTEBRA SEGMENTATION ACCURACY IN DATASET 2
(PATHOLOGIC LUMBAR SPINE) WITH AND WITHOUT SIM

SIM	Error measure	L1	L2	L3	L4	L5	Global
No	Mean DSI(%)	84.6	83.5	83.3	84.8	85.9	84.4
	2xSTD of DSI(%)	7.4	13.1	6.8	5.5	2.8	7.8
	MASD(mm)	1.36	1.5	1.6	1.41	1.31	1.44
	95-p(mm)	3.9	4.5	4.8	4.1	3.6	4.2
	HD(mm)	8.6	10.4	12.4	10.8	7.2	9.9
Yes	Mean DSI(%)	88.2	87.9	87.3	89.5	88.1	88.2
	2xSTD of DSI(%)	4.7	7.3	7.8	3.5	4.5	5.9
	MASD(mm)	0.94	0.99	1.0	0.89	1.06	0.98
	95-p(mm)	2.48	2.64	2.83	2.33	2.94	2.65
	HD(mm)	5.2	6.0	7.1	6.2	6.3	6.2

C. Quantitative Evaluation of the Method

The attained accuracy measurements for the three datasets are presented in Tables I–Table III. Each table contains the values with and without the use of the SIM in order to evaluate its importance.

Results show that the SIM increases the accuracy of the segmentations. The reason is that the SIM controls the relative distances of the vertebral bodies and spinous process, allowing the method to find better edges in the global optimisation. In addition, Table IV presents the overlap between adjacent vertebrae in number of voxels and in volume, for all the databases with and without the proposed SIM. It can be seen from the results that the proposed approach eliminates completely the overlap between processes. For all the datasets and vertebrae, we did not encounter a single case of overlap between the structures, which shows the robustness of the SIM. In contrast, the segmentations without this model induce significant overlap, in average over 25 mm^3 per vertebra in healthy spines going beyond 125 mm^3 in pathological cases. The better definition of the inter-vertebral relative positions also translates in an overall improvement of the segmentation accuracy as shown in Tables I–III.

D. Visual Results

A qualitative evaluation is presented in Figs. 5–7. Figs. 5 and 7 show the segmentation and ground-truth contours, superimposed to a sagittal and coronal slices of one patient from the Dataset 1 and another from the Dataset 3, respectively. On the other hand, Fig. 6 does not contain the ground-truth contour, it shows the segmentation of a patient from the Dataset 2 with and without the SIM, in order to better visualize the improvement given by this model. This patient contains a fracture in the

TABLE III
VERTEBRA SEGMENTATION ACCURACY IN DATASET 3 (HEALTHY THORACIC AND LUMBAR SPINE) WITH AND WITHOUT SIM

SIM	Error measure	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	L1	L2	L3	L4	L5	Lumbar	Thoracic
No	Mean DSI(%)	86.9	88.2	86.4	89.1	90.9	89.7	89.4	90.3	91.5	90.2	90.6	91.6	91.6	92.5	92.3	92.4	91.8	92.1	89.6
	2xSTD of DSI(%)	1.2	4.2	3.7	2.9	1.1	2.5	0.5	2.0	0.8	2.1	2.5	2.0	2.2	1.6	1.8	1.7	1.9	1.7	5.2
	MASD(mm)	1.52	0.54	0.63	0.53	0.46	0.54	0.58	0.54	0.53	0.69	0.72	0.63	0.69	0.52	0.54	0.59	0.61	0.57	0.66
	95-p(mm)	2.3	1.51	1.76	1.42	1.16	1.44	1.49	1.43	1.34	2.0	1.93	1.75	1.49	1.28	1.31	1.48	1.47	1.39	1.66
	HD(mm)	6.1	5.2	5.3	4.0	5.0	5.0	4.7	4.1	4.3	5.0	4.8	4.5	5.8	3.8	5.2	4.7	4.7	4.8	4.8
Yes	Mean DSI(%)	86.2	87.8	86.6	89.2	91.4	89.6	89.9	90.9	91.1	91.0	91.3	91.3	91.9	93.0	92.9	92.9	92.8	92.7	89.7
	2xSTD of DSI(%)	2.6	4.3	4.3	1.2	0.7	3.3	1.9	1.7	1.7	2.3	1.9	2.9	0.6	2.5	2.8	2.2	2.6	2.2	4.2
	MASD(mm)	0.72	0.58	0.6	0.51	0.43	0.52	0.55	0.52	0.55	0.59	0.62	0.69	0.50	0.47	0.50	0.50	0.50	0.49	0.58
	95-p(mm)	2.1	1.5	1.68	1.37	1.08	1.38	1.45	1.33	1.45	1.59	1.73	1.93	1.31	1.23	1.31	1.29	1.28	1.28	1.56
	HD(mm)	6.0	4.9	5.5	4.1	5.8	5.0	4.8	3.9	4.9	4.8	4.4	5.0	5.9	4.1	6.0	4.7	3.9	4.8	4.8

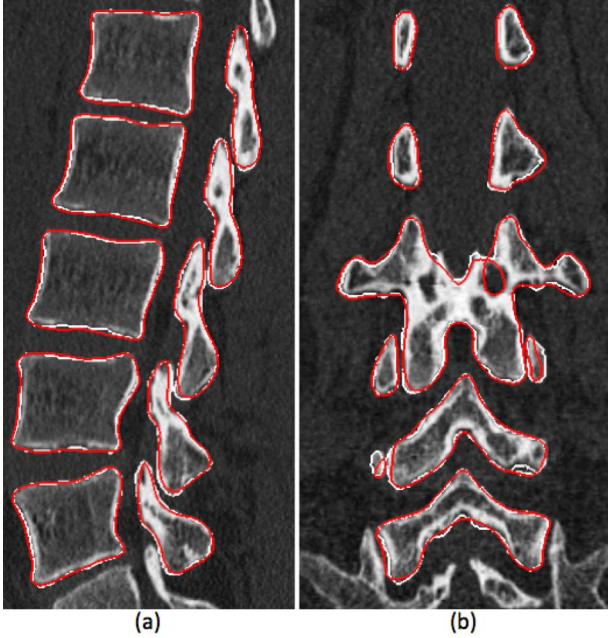


Fig. 5. Segmentation slices in sagittal and coronal view of one patient CT scan from the Dataset 1. The red contour represents the segmentation and the white contour the ground truth.

L4, which produces the spinous processes of the L3, L4 and L5 to collide. The use of the SIM maintains the separation between these processes under control ensuring not only zero-overlap but also high accuracy. Furthermore, the endplates of these three vertebrae are close to each other leading to overlaps when the SIM is not employed.

E. Sensitivity to the Initialisation

As explained in Section V-A, the segmentation process requires the center location of the IVDs. In order to test the sensitivity to this manual input, various patient CT images were segmented pinpointing the center in different positions within the IVD region. However, the evaluation measures resulted in similar results all the time with and without the SIM. The main difference was the location of overlap regions. An example of large overlap due to a complicated initialisation is presented in Fig. 8. The segmentation without using the SIM deforms the L1 vertebra unnaturally, allowing more overlap than in a normal setting, where the SIM still prevent the overlap and maintain the same accuracy than when the initial positions are closer to the real IVD centers.

F. Comparison With the State-of-the-Art

Although image resolution, type of database or the CT machine can influence the performance of different techniques, we are presenting, in Table V, a comparison of the accuracy between the state-of-the-art methods. In this table, the shown accuracy was obtained using CT images and evaluating the performance of the segmentation in the whole vertebra (VB + processes). The Dataset 2 was the only one employed for this comparison, since various state-of-the-art methods were tested in pathological cases. Therefore, the accuracy measurements from all the methods presented in Table V are only from the Lumbar region. An important difference with the results shown in Table II is that the SSMs and SIMs were created in a leave-one-out fashion using solely the Dataset 2 instead of being computed from the Dataset 1. It can be seen that the proposed technique compares favourably to the existing studies on the topic, while ensuring extra robustness to extract non-overlapping and anatomically meaningful segmentations of the vertebrae.

In addition, Table VI presents a comparison of the results of the methods participating in the MICCAI Challenge 2014 on spine imaging [51]. The dataset used for this challenge (Dataset 3) included healthy subjects only, without evident vertebral pathologies. Thus, the incorporation of the SIM does not improve greatly the mean accuracy (see Section VIII). Its main contribution is the reduction of large errors, which may be measured by the 95 percentile, Hausdorff (maximum) distance, or by specific-purpose measures as the overlapping volume. However, only mean measures (mean DSI and MASD) were computed in the Challenge. The issue of only using this type of measures is appreciated in Table VI, where the MASD and mean DSI of the proposed method are less accurate than those of our previous method [40], but the 95 percentile and Hausdorff distance are clearly improved. This effect is also observed when comparing the results with those of the method proposed by Korez *et al.* [53], since they provided additional measurements for the training set. The reported MASD (0.35) and mean DSI (93.1) were better than ours (mean DSI of 90.8, MASD of 0.52), but their Hausdorff distance (5.73 mm) was not as accurate (ours: 4.6 mm). Another example is the method introduced in [12] (Table V), which has a better DSI than ours but the Hausdorff distance is almost double. For future reference, we have added the 95 percentile and Hausdorff distance in the table.

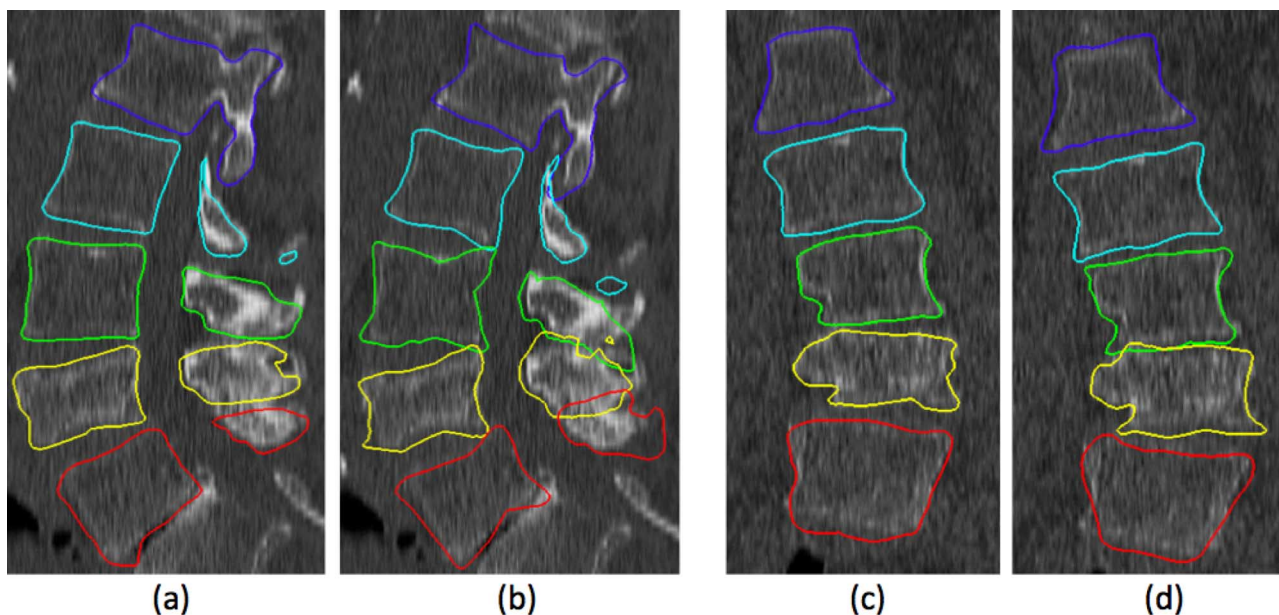


Fig. 6. Segmentation slices in sagittal and coronal view of one patient CT scan from the Dataset 2. Figures (a) and (c) represent the segmentation using the SIM and figures (b) and (d) without the SIM. It shows that the spinous processes of the L3, L4 and L5 vertebrae merged together because the L4 is fractured. This issue produces large errors and overlaps when the SIM is not employed.

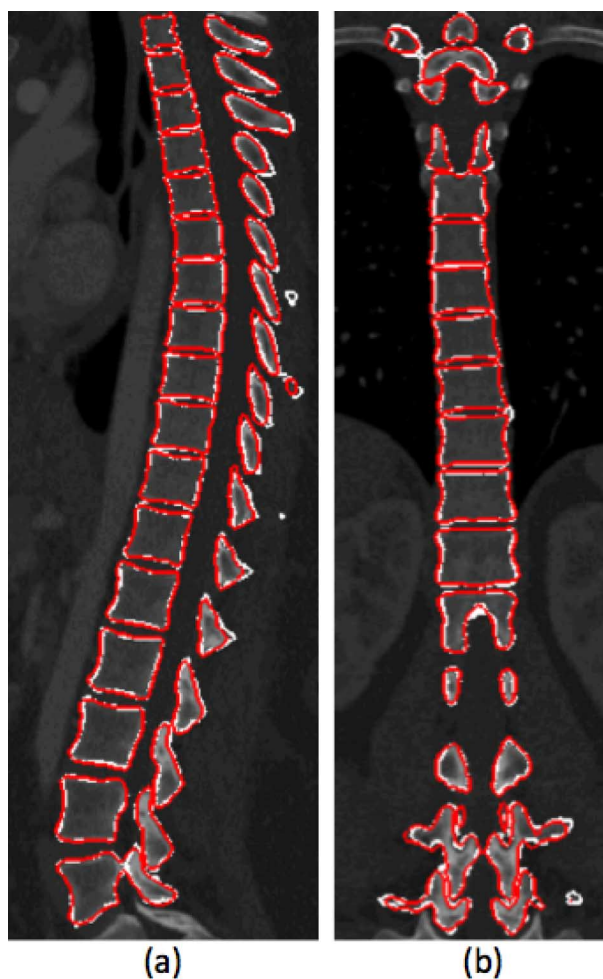


Fig. 7. Segmentation slices in sagittal and coronal view of one patient CT scan from the Dataset 3. The red contour represents the segmentation and the white contour the ground truth.

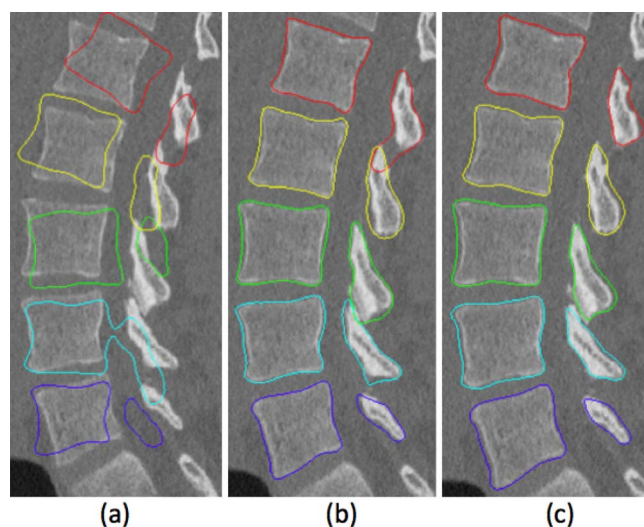


Fig. 8. Improvement obtained by employing the SIM in the segmentation scheme with a bad initialisation. The image (a) is the initial shape, the image (b) is the segmentation without using the SIM and the image (c) is when the SIM is employed. The main difference is the correction of the process in the L1.

TABLE IV
MEAN INTER-PROCESS OVERLAP IN VOXELS AND IN CUBIC MILLIMETERS
FOR THE CASE OF USING SIM AND WITHOUT FOR ALL THE DATABASES

Spine position	Mean Number of voxels		Mean Volume (mm^3)	
	No SIM	With SIM	No SIM	With SIM
D1: L1 - L5	897.2	0	200.26	0
D2: L1 - L5	2554.9	0	574.8	0
D3: T1 - T12	530	0	132.5	0
D3: L1 - L5	0	0	0	0

D1: Dataset 1 - Only Lumbar vertebrae of healthy subjects.
D2: Dataset 2 - Only lumbar vertebrae of pathological cases.
D3: Dataset 3 - Thoracic and lumbar vertebrae of healthy subjects.

TABLE V
COMPARISON OF THE STATE OF THE ART METHODS

Study	Image resolution	DSI (%)		Surface error (mm)			
		Mean	STD ¹	MASD ²	MPSD ³	95-p ⁴	HD ⁵
Weese <i>et al.</i> [23]	$0.48 \times 0.48 \times (0.72 - 2)mm^3$			0.93			6.14
Kadoury <i>et al.</i> [38]	$0.8 \times 0.8 \times (1 - 2)mm^3$			2.8 ± 1.9			
Rasoulouian <i>et al.</i> [39]	$0.6 \times 0.6 \times 0.6mm^3$ to $0.9 \times 0.9 \times 3.2mm^3$			1.38		3.8	8.91
Klinder <i>et al.</i> [35]	$0.32 \times 0.32 - 0.98 \times 0.98mm^2$ thickness: $0.8 - 2.8$				$0.73 - 0.79$		
Huang <i>et al.</i> [12]	$1.0 \times 1.0 \times 1.5mm^3$	94	2				10.06 ± 1.71
Proposed method	$0.35 \times 0.35 \times 2.5mm^3$	90	5.1	0.82	0.78	2.1	5.5

¹: standard deviation, ²: mean absolute surface distance, ³: mean point to surface distance, ⁴: percentile 95 of the surface to surface distance, ⁵: Hausdorff distance.

TABLE VI
COMPARISON OF THE METHODS THAT USED THE DATASET 3

Study	DSI (%)		Surface error (mm)		
	Mean	STD ¹	MASD ²	95-p ³	HD ⁴
Forsberg <i>et al.</i> [54]	92		0.47		
Seitel <i>et al.</i> [55]	87		1.1		
Hammernik <i>et al.</i> [56]	95		0.38		
Korez <i>et al.</i> [53]	95		0.37		
Castro-Mateos <i>et al.</i> [40]	93	4.1	0.55	1.52	6.2
Proposed Method	91	3.4	0.55	1.47	4.8

¹: standard deviation, ²: mean absolute surface distance, ³: percentile 95 of the surface to surface distance, ⁴: Hausdorff distance.

VIII. DISCUSSION

This paper presents a new method to model the interspace between objects in a multi-object structure by the use of a statistical interspace model (SIM). It was tested in the segmentation of the lumbar and thoracic spine, demonstrating its effectiveness to avoid segmentations with unnatural geometries, such as overlaps, that may cause simulations to result in unrealistic outcomes, even when the global accuracy is high.

Since normal metrics do not take into account this overlap effect, we explicitly measured the overlap with and without the use of the SIM, showing that this model completely avert overlaps.

However, As mentioned in Section IV, the SIM does not guaranteed zero overlap mathematically speaking. However, the likelihood of happening is so small that it never occurred in our experiments. The main difference with other approaches is that we propose a model of the interspace between objects to guaranteed that the shapes are not deformed in an unnatural manner as well as decreasing the likelihood of overlapping to almost 0. For instance, the probability of having an infinitesimal overlap is equally likely with or without the energy that blocks overlaps introduced in [35]. In contrast, this probability is much smaller when employing the SIM.

In addition, this model partially controls the possible configurations of the segmentation by restricting the shape of the separation between the objects. This is similar to Statistical Pose Models [36], however less restricted. As shown in Tables I–III,

the accuracy of the results does not change greatly by the addition of the SIM when segmenting healthy subjects. In general, the segmentation depends on the intensity, the SSM and the SIM energies. In the case of healthy subjects, the SSM contained enough flexibility to accommodate the shapes with a small error and the borders of the images were clear enough for the intensity energy to drive the shape to the right positions. Thus, the SIM did not influence the resultant shape in a great manner. However, in the case of pathological cases, the SSM energy tried to deform the shape to be similar to the closest healthy vertebra within its scope, which could have large differences. In addition, the borders were not clear and the intensity energy failed at finding the correct borders. Under these circumstances the SIM had a great impact at controlling the separation between the vertebral bodies and processes to find the most realistic shape. Observe that there are two factors for unclear borders: the first one is due to the pathology and not to the image quality (collision of vertebrae, decalcification, etc . . .), the second is due to the image protocol (voxel size, windows type, etc . . .). Dataset 1 and 3 did not have any issues whereas dataset 2 contain unclear borders due to both factors, large voxel size, fractures, etc . . .

The SIM and the SSM restrict the possible shapes of the segmentation. However, the intensity energy is the only one that deforms the shape within the image. The method employed in this article is based on the selection of a group of features that was proven to be accurate in the case of dataset 1 [40]. However, it may happen that other energies provide more accurate results for other datasets. Since, this energy is interchangeable for any other while maintaining the same framework with the SSM and SIM, it could be possible to use one of the proposed energies from the methods in Table VI to improve the border detection.

A. Limitations

The proposed method requires the manual selection of the IVDs center. To overcome this limitation an automatic center detection, such as the one proposed by [57], may be employed. Another issue of the method is that an SIM is created to model the variations of the space between two surfaces. Therefore, the external objects in the structure, which are vertebrae T1 and L5 in the thoracic and lumbar spine, may over-segment to external structures that are not considered in the model (Fig. 9). A feasible solution may be to include a part of objects that are outside the structure, vertebrae C6, S1, to model the interspace.

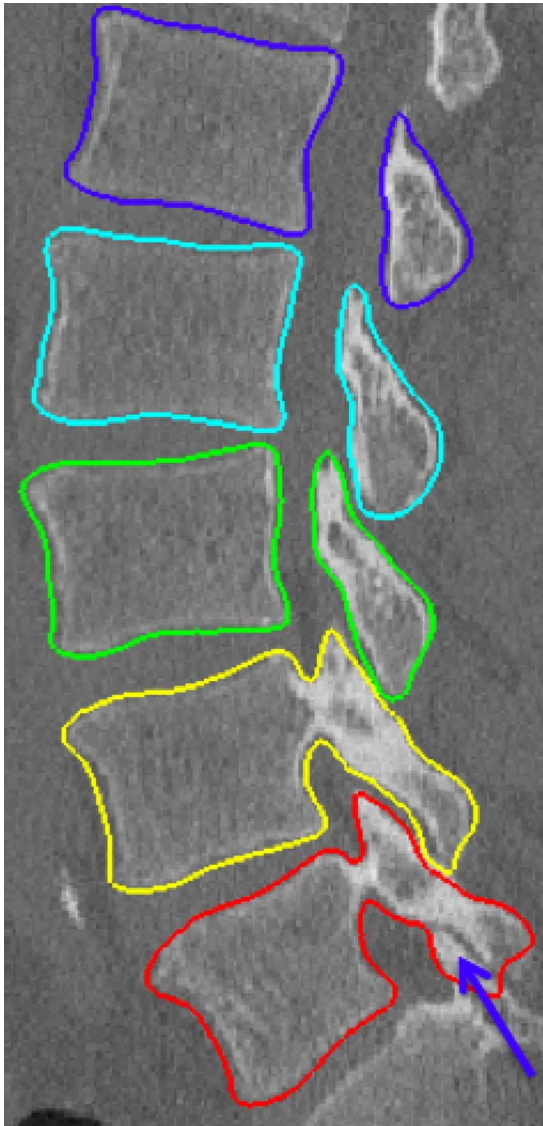


Fig. 9. The SIM requires two surfaces in order to model the space between them. Therefore, in the low spine segmentation paradigm the L1 and L5 may leak into their top and bottom vertebra, respectively. In this figure, this issue is shown with a blue arrow in the L5.

B. Computational Cost

The experiment were run using a PC with Intel(R) Xeon CPU E5620 at 2.40 Ghz with 8 GB of RAM and running a 64-bit Windows 8.1 operating system. All the code was written in C++ using the libraries from the Insight Toolkit [49] for basic manipulation of mesh and images. The proposed framework requires around 50 seconds per vertebra to perform the segmentation and only 1 second to initialize all the vertebra after manually selecting the center of the IVDs. In contrast, the standard Active Shape Model [1] or our previous vertebra segmentation approach [40] only requires 15 seconds per vertebra. The reason for the increase to 50 seconds is that the proposed framework employs a steepest gradient descent algorithm to do the optimisation because it does not have a closed-form solution. Observe that by removing the SIM but maintaining the same segmentation framework the process requires 49 seconds, it is only around 1s faster per vertebra.

IX. CONCLUSIONS

This article introduces the statistical interspace model (SIM), embedded in a framework for segmenting a multi-object structure. We tested the method with three databases of CT images of the spine. The SIM was employed to avoid the overlap in the vertebral and spinous processes and to maintain the correct separation between vertebral bodies.

The results show that SIM not only prevent vertebral overlaps but also increase the accuracy of the segmentation by maintaining the relative position of the vertebral bodies and the spinous processes. This control was found to be necessary in pathological cases, where vertebral bodies and/or spinous processes collapse with neighbouring structures. Furthermore, the proposed framework is not restricted to vertebra segmentation, it could be applied to other anatomical parts such as intervertebral disc (IVD), knees, heart or any other multi-object structure, where the control of the relative position among the individual objects is required to restrict rotations, translation and/or to prevent overlap between different tissues.

REFERENCES

- [1] T. Cootes, C. Taylor, and J. Graham, "Active shape models- their training and applications," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.
- [2] M. De Bruijne, M. T. Lund, L. B. Tanko, P. C. Pettersen, and M. Nielsen, "Quantitative vertebral morphometry using neighbor-conditional shape models," *Med. Image Anal.*, vol. 11, no. 5, pp. 503–512, 2007.
- [3] T. Okada, K. Yokota, M. Hori, M. Nakamoto, H. Nakamura, and Y. Sato, "Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images," in *MICCAI*. New York: Springer, 2008, LNCS, pp. 502–509.
- [4] T. H. Lim, J. C. Eck, H. S. An, J. H. Hong, J.-Y. Ahn, and J. W. You, "Biomechanics of transfixation in pedicle screw instrumentation," *Spine*, vol. 21, no. 19, pp. 2224–2229, 1996.
- [5] T. Jahng, Y. Kim, and K. Moon, "Comparison of the biomechanical effect of pedicle-based dynamic stabilization: A study using finite element analysis," *Spine J.*, vol. 13, no. 1, pp. 85–94, 2013.
- [6] D. Schlenszka, T. Laine, and T. Lund, "Computer-assisted spine surgery," *Eur. Spine J.*, vol. 9, no. 1, pp. S057–S064, 2000.
- [7] M.-H. Bae, R. Pan, T. Wu, and A. Badea, "Automated segmentation of mouse brain images using extended MRF," *Neuroimage*, vol. 46, no. 3, pp. 717–725, 2009.
- [8] X. Chen, J. Graham, C. Hutchinson, and L. Muir, "Automatic generation of statistical pose and shape models for articulated joints," *IEEE Trans. Med. Imag.*, vol. 33, no. 2, pp. 372–383, Feb. 2014.
- [9] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.
- [10] V. Jurcak *et al.*, "Atlas based automated segmentation of the quadratus lumborum muscle using non-rigid registration on magnetic resonance images of the thoracolumbar region," *Int. Symp. Biomed. Imag.*, pp. 113–116, May 2008.
- [11] M. S. Aslan, A. Shalaby, H. Abdelmunim, and A. Farag, "Probabilistic shape-based segmentation method using level sets," *Inst. Eng. Technol.*, pp. 1–13, 2013.
- [12] J. Huang, F. Jian, H. Wu, and H. Li, "An improved level set method for vertebra CT image segmentation," *Biomed. Eng. Online*, vol. 12, no. 1, p. 48, 2013.
- [13] P. P. Smyth, C. J. Taylor, and J. E. Adams, "Automatic measurement of vertebral shape using active shape models," *Image Vis. Comput.*, vol. 15, no. 8, pp. 575–581, 1997.
- [14] L. R. Long and G. R. Thoma, "Use of shape models to search digitized spine X-rays," in *Proc. 13th IEEE Symp. IEEE Comput.-Based Med. Syst.*, 2000, pp. 255–260.
- [15] G. Zamora, H. Sari-Sarraf, and L. R. Long, "Hierarchical segmentation of vertebrae from X-ray images," in *Proc. Int. Soc. Optics Photon. Med. Imag.*, 2003, pp. 631–642.

- [16] M. Roberts, T. Cootes, and J. Adams, "Automatic segmentation of lumbar vertebrae on digitised radiographs using linked Active appearance models," *Proc. Med. Image Understand. Anal.*, vol. 2, pp. 120–124, 2006.
- [17] S. Mahmoudi and M. Benjelloun, "A new approach for cervical vertebrae segmentation," *Progress Pattern Recognit., Image Anal. Appl.*, pp. 753–762, 2007.
- [18] P. Myslning, K. Petersen, M. Nielsen, and M. Lillholm, "A unifying framework for automatic and semi-automatic segmentation of vertebrae from radiographs using sample-driven Active shape models," *Mach. Vis. Appl.*, pp. 1–14, 2012.
- [19] A. Mastmeyer, K. Engelke, C. Fuchs, and W. A. Kalender, "A hierarchical 3D segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine," *Med. Image Anal.*, vol. 10, no. 4, pp. 560–577, 2006.
- [20] S. Tan, Y. Jianhua, M. Ward, L. Yao, and R. Summers, "3D multi-scale level set segmentation of vertebrae," *Int. Symp. Biomed. Imag.*, pp. 896–899, 2007.
- [21] M. S. Aslan, A. Ali, A. A. Farag, B. Arnold, D. Chen, and P. Xiang, "3D vertebrae segmentation in CT images with random noises," in *Proc. 2010 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2290–2293.
- [22] D. Stern, B. Likar, F. Pernus, and T. Vrtovec, "Parametric modelling and segmentation of vertebral bodies in 3D CT and MR spine images," *Phys. Med. Biol.*, vol. 56, no. 23, pp. 7505–7522, 2011.
- [23] J. Weese, M. Kaus, C. Lorenz, S. Lobregt, R. Truyen, and V. Pekar, "Shape constrained deformable models for 3D medical image segmentation," in *Information Processing in Medical Imaging*. New York: Springer, 2001, pp. 380–387.
- [24] J. Kaminsky *et al.*, "Specially adapted interactive tools for an improved 3D-segmentation of the spine," *Comput. Med. Imag. Graph.*, vol. 28, no. 3, pp. 119–127, 2004.
- [25] B. Naegel, "Using mathematical morphology for the anatomical labeling of vertebrae from 3D CT-scan images," *Comput. Med. Imag. Graph.*, vol. 31, no. 3, pp. 141–156, 2007.
- [26] J. Ma, L. Lu, Y. Zhan, X. Zhou, M. Salganicoff, and A. Krishnan, "Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model," in *MICCAI*. New York: Springer, 2010, LNCS, pp. 19–27.
- [27] J. Zhang, C. Yan, C. Chui, and S. Ong, "Fast segmentation of bone in CT images using 3d adaptive thresholding," *Comput. Biol. Med.*, vol. 40, no. 2, pp. 231–236, 2010.
- [28] M. Kirschner, M. Becker, and S. Wesarg, "3D Active shape model segmentation with nonlinear shape priors," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2011*. New York: Springer, 2011, pp. 492–499.
- [29] J. Carballido Gamio, S. J. Belongie, and S. Majumdar, "Normalized cuts in 3D for spinal MRI segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 1, pp. 36–44, Jan. 2004.
- [30] H. Szu-Hao, C. Yi-Hong, L. Shang-Hong, 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. .
- [31] A. Neubert *et al.*, "Automated detection, 3D segmentation and Analysis of high resolution spine MR images using statistical shape models," *Phys. Med. Biol.*, vol. 57, no. 24, p. 8357, 2012.
- [32] H. Shen, A. Litvin, and C. Alvino, "Localized priors for the precise segmentation of individual vertebrae from CT volume data," in *Medical Image Computing and Comput.-Assisted Intervention-MICCAI 2008*. New York: Springer, 2008, pp. 367–375.
- [33] 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, 2009.
- [34] H. Seim *et al.*, "Automatic segmentation of the pelvic bones from CT data based on a statistical shape model," *Spring Eurograph (VCBM)*, pp. 93–100, 2008.
- [35] T. Klinder *et al.*, "Automated model-based vertebra detection, identification, and segmentation in CT images," *Med. Image Anal.*, vol. 13, no. 3, p. 471, 2009.
- [36] J. Boisvert, X. Pennec, H. Labelle, F. Cheriet, and N. Ayache, "Principal spine shape deformation modes using Riemannian geometry and articulated models," in *Articulated Motion and Deformable Objects*. New York: Springer, 2006, pp. 346–355.
- [37] M. Van De Giessen *et al.*, "Constrained registration of the wrist joint," *IEEE Trans. Med. Imag.*, vol. 28, no. 12, pp. 1861–1869, Dec. 2009.
- [38] S. Kadoury, H. Labelle, and N. Paragios, "Spine segmentation in medical images using manifold embeddings and higher-order MRFs," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1227–1238, Jul. 2013.
- [39] A. Rasoulia, 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.
- [40] I. Castro-Mateos, J. Pozo, A. Lazary, and A. F. Frangi, "3D vertebra segmentation by feature selection Active shape model," in *Recent Advances in Computational Methods and Clinical Applicant. for Spine Imaging*. New York: Springer, 2015, pp. 241–245.
- [41] E. Lorensen William and E. Harvey, "Marching cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, pp. 163–169, Aug. 1987.
- [42] I. Mpiperis, S. Malassiotis, and G. Strintzis Michael, "Bilinear models for 3-D face and facial expression recognition," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 3, pp. 498–511, Sep. 2008.
- [43] H.-W. Lilliefors, "On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown," *J. Am. Stat. Assoc.*, vol. 64, no. 325, pp. 387–389, 1969.
- [44] C. Jarque and K. Bera Anil, "A test for normality of observations and regression residuals," *Int. Stat. Rev./Revue Internationale de Statistique*, pp. 163–172, 1987.
- [45] A. Trujillo-Ortiz, R. Hernandez-Walls, K. Barba-Rojo, and L. Cupul-Magana, "hZmvttest: Henze-Zirkler's Multivariate normality test," *A MATLAB*, 2007 [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do>
- [46] P. Fletcher, C. Lu, S. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 995–1005, Aug. 2004.
- [47] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998, vol. 4.
- [48] MySpine Patient-Specific Spinal Treatment Simulation MySpine, 2010 [Online]. Available: <http://www.myspineproject.eu>
- [49] L. Ibáñez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide* 2nd ed. Kitware, 2005 [Online]. Available: <http://www.itk.org/ItkSoftwareGuide.pdf>
- [50] Spineweb, 2013 [Online]. Available: <http://spineweb.digitalimaging-group.ca/>
- [51] *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, J. Yao, B. Glocker, T. Klinder, and S. Li, Eds. New York: Springer, 2015.
- [52] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 716–724, Dec. 1994.
- [53] R. Korez, B. Ibragimov, B. Likar, F. Pernus, and T. Vrtovec, "Interpolation-based shape-constrained deformable model approach for segmentation of vertebrae from CT spine images," in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*. New York: Springer, 2015, pp. 235–240.
- [54] D. Forsberg, "Atlas-based segmentation of the thoracic and lumbar vertebrae," in *Recent Advances in Computational Methods and Clinical Applicant. for Spine Imaging*. New York: Springer, 2015, pp. 215–220.
- [55] A. Seitel, A. Rasoulia, R. Rohling, and P. Abolmaesumi, "Lumbar and thoracic spine segmentation using a statistical multi-object shape + pose model," in *Recent Advances in Computational Methods and Clinical Applicant. for Spine Imaging*. New York: Springer, 2015, pp. 221–225.
- [56] K. Hammernik, T. Ebner, D. Stern, M. Urschler, and T. Pock, "Vertebrae segmentation in 3D CT images based on a variational framework," in *Recent Advances in Computational Methods and Clinical Applicant. for Spine Imaging*. New York: Springer, 2015, pp. 227–233.
- [57] S. Al-Helo, R. Alomari, V. Chaudhary, and M. Al-Zoubi, "Segmentation of lumbar vertebrae from Clinical CT using active shape models and GVF-snake," in *Proc. Annu. Int. Conf. IEEE EMBS*, 2011, pp. 8033–8036.