MULTIMODAL INFERENCE OF ARTICULATED SPINE MODELS FROM HIGHER ORDER ENERGY FUNCTIONS OF DISCRETE MRFS

Samuel Kadoury, Nikos Paragios

Laboratoire MAS, Ecole Centrale de Paris, Chatenay-Malabry, France
GALEN Group, INRIA Saclay, Ile-de-France, Orsay, France

ABSTRACT

In this paper, we introduce a novel approach based on higher order energy functions which have the ability to encode global structural dependencies to infer articulated 3D spine models to CT volume data. A personalized geometrical model is reconstructed from biplanar X-rays before spinal surgery in order to create a spinal column representation which is modeled by a series of intervertebral transformations based on rotation and translation parameters. The shape transformation between the standing and lying poses is then achieved through a Markov Random Field optimization graph, where the unknown variables are the deformations applied to the intervertebral transformations. Singleton and pairwise potentials measure the support from the data and geometrical dependencies between neighboring vertebrae respectively, while higher order cliques are introduced to integrate consistency in regional curves. Optimization of model parameters in a multi-modal context is achieved using efficient linear programming and duality. A qualitative evaluation of the vertebra model alignment obtained from the proposed method gave promising results while the quantitative comparison to expert identification yields an accuracy of 1.8 ± 0.7 mm based on the localization of surgical landmarks.

Index Terms— Articulated 3D spine, multimodal inference, Markov random fields, high order cliques, discrete optimization

1. INTRODUCTION

Deformable anatomical models are powerful tools for recovering the shape of a patient's anatomy when only partial information or sparse image data is available. For example in orthopedic surgery, 3D computer-generated models have assisted specialists in surgical planning and instrument navigation during the intervention [1]. It offers a unique advantage to visualize the anatomy during surgery and localize anatomical regions without segmenting operative images. By fusing these images such as CT, C-arm CT, MR or ultrasound with an accurate preoperative model, the surgeon can see the position and orientation of the instrumentation tools on precise anatomical models. While these emerging technologies were successful for knee or hip replacement applications, corrective spinal surgery is particularly challenging due the complex three-dimensional (3D) deformations of the spine combined with asymmetric deformation of the vertebrae, high variability of the articulated structure and required precision for pedicle screw insertion.

Registration of intraoperative fluoroscopic images and preoperative CT/MR images has been proposed to aid interventional and surgical orthopedic procedures [2]. In some cases, 3D models were registered to 2D X-ray and fluoroscopic images using gradient amplitudes for optimizing the correspondence of single bone structures [3]. Objective functions using surface normals from statistical PDMs were applied for the femur [4] or pelvis [5]. In spine registration however, one important drawback is that each vertebra of the spine is treated individually instead of as a global shape which hinders surgeons to exploit virtual fluoroscopy imaging.

To tackle this issue, articulated models allow to account for the global geometrical representation [6] by incorporating knowledge-based intervertebral constraints. These 3D intervertebral transformations were transposed in [7] to accomplish the segmentation of the spinal cord from CT images, but multi-modal registration has yet to be solved. Optimization is also based on gradient-descent, prone to non-linearity and local minimums. These methods require segmentation of 3D data or fluoroscopic image, which itself is a challenging problem and has a direct impact on registration accuracy. In [8], we propose a discrete optimization method for articulated spine models, that effectively infers a preoperative model to intraoperative data using only singleton data terms and pairwise constraints. Although intervertebral orientations and translations are nicely captured, the method fails to encode high level geometrical representation of spine changes between the initial exam and during surgery.

In this work, we propose a novel framework which incorporates statistics of regional spinal curves represented by higher order cliques for registering preoperative 3D articulated spine models in a standing position to lying intraoperative 3D CT images. We use a personalized 3D spine reconstructed from biplanar X-rays to derive an articulated model represented with intervertebral transformations. Inference is achieved through a Markov Random Field (MRF) graph which incorporates three optimality components: modular image data-terms which avoids image segmentation, dual potentials to constrain the adjustment of intervertebral links between neighboring objects and global priors of the regional spine components encoded as higher order functionals which are solved with a primal-dual discrete optimization scheme. We therefore infer a new annotated spine representation from data gathered during surgery based on a high-resolution personalized 3D model.

2. METHODS

2.1. Personalized 3D Reconstruction of Articulated Spines

The preoperative 3D spine model is obtained from biplanar X-ray images using a statistical embedding and image-based models based on the work of [9]. A non-linear manifold embedding $\mathcal{M}$ from a dataset of 3D pathological spines is created to predict an initial model with 17 vertebrae, 6 points per vertebra, based on the patient's centerline. This manifold establishes the patterns of legal variations

Research was supported in part by a FQRNT grant. Thanks to Dr. H. Labelle for the clinical data.

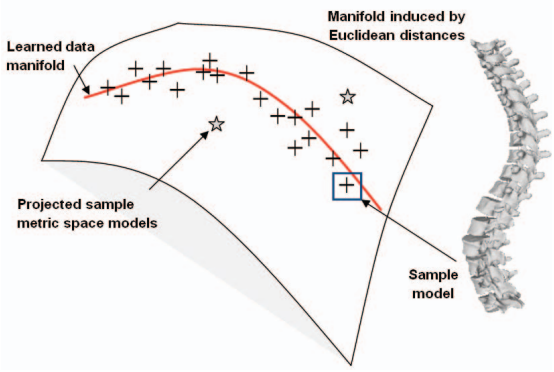


Fig. 1. Illustration of spine distribution embedded onto on a low-dimensional manifold. Stars represent new sample projected onto the manifold.

of spine shape changes in a low-dimensional sub-space based on locally linear embeddings as illustrated in Fig 1.

Once the global shape is obtained, the 3D model is then refined locally for each vertebra via a segmentation approach which extends 2D geodesic active regions in 3D [10], in order to evolve prior deformable 3D surfaces. An atlas of vertebral meshes $S_i = \{\mathrm{v}_{i1}, \ldots, \mathrm{v}_{iN}\}$ with vertices v_j are initially positioned and oriented from their respective 6 precise landmarks s_i composing the entire spine $\mathbf{S}$, and the surface is evolved so that the projected silhouettes of the morphed 3D models would therefore match the 2D information on the biplanar X-rays. The 3D landmark coordinates s_i and corresponding polygonal vertebral meshes $S_i \in \mathbf{S}$ are optimal with regards to the image and statistical distribution. The 3D landmarks s_i are used to rigidly register each vertebra to its upper neighbor, and the resulting rigid transforms are optimized in the registration problem. Hence, the spine is represented by a vector of local inter-vertebral rigid transformations $A = [T_1, T_2, \ldots, T_N]$. To perform global anatomical modeling of the spine, we convert A into an absolute representation:

$$\mathbf{A}_{\text{abs}} = [T_1, T_1 \circ T_2, \ldots, T_1 \circ T_2 \circ \ldots \circ T_N] \tag{1}$$

using recursive compositions. The transformations are expressed in the local coordinate system of the lower vertebra, defined by vectors v_x, v_z and $v_y = v_x \times v_z$, where v_x and v_z are the vectors linking pedicle and endplate midpoints respectively. Center of transformation is located at the midpoint of all 4 pedicle tips. The rigid transformations described in this paper are the combination of a rotation matrix R and a translation vector t. We formulate the rigid transformation $T = \{R, t\}$ of a vertebral mesh triangle as $y = Rx + t$ where $x, y, t \in \Re^3$. Composition is given by $T_1 \circ T_2 = \{R_1R_2, R_1t_2 + t_1\}$, while inversion as $T^{-1} = \{R^T, -R^Tt\}$.

2.2. Transformation Inference from Higher Order MRFs

A successful inference between the spine model $\mathbf{S}$ controlled by the articulations (denoted as $\mathbf{A}_{\text{abs}}$) and the image $\mathcal{I}$ must be accomplished by establishing similarity criterions which will drive the model deformation towards the optimal solution. We search the optimal displacement points $\vec{\mathbf{D}} = (\vec{\mathbf{d}}_1, \ldots, \vec{\mathbf{d}}_n)$ of the articulation vectors T that give a good compromise between the encoded prior constraints established by manifold statistics and the fidelity to the image information. Formally, the inference of the model $\mathbf{A}_{\text{abs}}$ to the image $\mathcal{I}$ is given by:

$$(\vec{\mathbf{d}}_1, \ldots, \vec{\mathbf{d}}_n) = \underset{\vec{d}_i}{\operatorname{argmin}}\ E(\mathbf{S}^0, \mathcal{I}, (\vec{d}_1, \ldots, \vec{d}_n)). \tag{2}$$

The energy E of inferring the spine model $\mathbf{S}$ in the image $\mathcal{I}$ is a function of the displacement vectors $\vec{D} = (\vec{d}_1, \ldots, \vec{d}_n)$ in the transformation space applied to the articulation vector $\mathbf{A}_{\text{abs}}$. This influences the data-related term $V(\mathbf{A}^0_{\text{abs}} + \vec{D}, \mathcal{I})$ expressing the image cost, a local prior term $V(\mathbf{N}, \vec{D})$ measuring deformation between neighboring vertebrae and a global higher order term $V(\mathbf{H}, \vec{D})$ which models the global deformation of a regional curve. The energy function E is modeled by an MRF graph G, composed of transformation nodes i with pairwise neighborhoods $\mathcal{N}$:

$$\begin{aligned} E\left(\mathbf{S}^0, \mathcal{I}, \vec{D}\right) &= V\left(\mathbf{A}^0_{\text{abs}} + \vec{D}, \mathcal{I}\right) + V\left(\mathbf{N}, \vec{D}\right) + V\left(\mathbf{H}, \vec{D}\right) \\ &= \sum_{i \in G} V_i(T_i^0 + \vec{d}_i, \mathcal{I}) \\ &+ \lambda \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(T_i^0 + \vec{d}_i, T_j^0 + \vec{d}_j) \\ &+ \alpha \sum_{c \in \mathcal{C}} V_c(\mathbf{T}_c^0 + \vec{d}_c) \end{aligned} \tag{3}$$

where $\mathbf{A}^0_{\text{abs}} + \vec{D} = \{T_1^0 + \vec{d}_1, \ldots, T_n^0 + \vec{d}_n\}$ are the equivalence for articulated components. The data term in the image domain seeks to minimize the distance between model and $\mathcal{I}$:

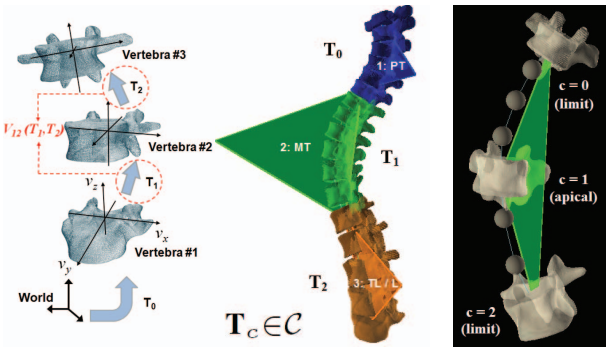
$$V_i(T_i^0 + \vec{d}_i, \mathcal{I}) = \int_\Omega \eta_{\mathrm{X}}(\mathcal{I}, S_i(T_i^0 + \vec{d}_i))dT \tag{4}$$

$$\text{with: } \eta_{\mathrm{X}} = \sum_{\mathrm{v}_{ij} \in S_i} (\gamma^2 + \gamma \|\nabla \mathcal{I}(\mathrm{v}_{ij})\|)/(\gamma^2 + \|\nabla \mathcal{I}(\mathrm{v}_{ij})\|^2).$$

This attracts mesh triangles to target high-intensity voxels in the gradient CT volume without segmentation. The term γ is defined as a dampening factor. The second term of Eq.(3) is a local prior term which are pairwise potentials representing the smoothness term between two consecutive vertebrae and help to constrain the vertebrae main direction in the optimization step by assigning a binary clique value for $V_{ij} = \{0, 1\}$ with $|(T_i^0 + \vec{d}_i) - (T_j^0 + \vec{d}_j)| \leq \epsilon_1$ for $\Re(i) \equiv \Re(j)$ and $||\vec{d}_i|| - ||\vec{d}_j|| \leq \epsilon_2$ for $\Re(i) \neq \Re(j)$ (Fig. 2(a)). The final term in the energy function represents the higher order potentials. We parameterize the potentials with clique variables $\mathbf{T}_c$ taking on corresponding costs θ_q if the cliques are assigned to the displacement vectors $\vec{d}_c$. To encode anatomical coherence of the spine, three cliques representing each of the spine's regions are composed of three vertebrae, linking the inflexion (change of curvature) and the apical (most deviated) vertebrae as illustrated in Fig. 2(b)-(c). The potential functions are defined as:

$$V_c(\mathbf{T}_c^0) = \min\{\min_{q \in \{1,2,\ldots,t\}} \theta_q + \Delta_q(\mathbf{T}_c^0), \theta_{\max}\} \tag{5}$$

where $\theta_q = ||\psi(\mathbf{T}_c) - \Pi_{\phi(\mathcal{M})}(\psi(\mathbf{T}_c))||$ is a geodesic distance calculated from the low-dimensional mapping of the clique variable onto the manifold using $\psi : \Re \rightarrow \mathcal{M}$ to the projection point Π on the prior distribution $\phi(\mathcal{M})$, and $\Delta_q(\mathbf{T}_c^0) = \sum_{i \in c} w_{il}^q \delta(T_i = l)$ is a



(a) (b) (c)

Fig. 2. (a) Pairwise link between neighboring vertebrae; (b) Regional higher order cliques based on (c) vertebral triplets.

deviation function. The Kronecker delta function δ generates binary variables, while the weights are assigned such that:

$$w_{il}^{q} = \begin{cases} 0 & \text{if } T_q = l \\ \theta_{\max} & \text{otherwise} \end{cases} \tag{6}$$

depending on wether the clique variable is given the appropriate label (see sec. 2.3). We develop an optimization procedure which minimizes Eq.(3) by means of an efficient discrete optimization algorithm using an MRF which is explained in the next section.

2.3. Energy Minimization

The optimization strategy for Eq.(3) of the resulting MRF is based on a discrete labeling principle where we seek to assign the optimal labels $\mathcal{L} = \{l_1, \ldots, l_i\}$, defined in the quantized space $\Theta = \{\vec{d^1}, \ldots, \vec{d^i}\}$ of displacements, to the vertebral transformations represented by nodes T_i so that the total energy of the graph is minimum. If we consider that displacing an intervertebral transformation vector by $\vec{d^n}$ is equivalent to assigning label l and that the current solution is given by $T_i^t = T_i^0 + \sum_t \vec{d}^{l_i t}$, which adopts the pyramidal coarse-to-fine quantization approach in a temporal minimization, the energy Eq.(3) can be re-written as a labeling problem:

$$\begin{aligned} E^t(l_1, \ldots, l_n) = & \sum_{i \in G} V_i(T_i^{t-1}, l_i) \\ & + \lambda \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(T_i^{t-1}, T_j^{t-1}, l_i, l_j) \\ & + \alpha \sum_{\mathbf{T}_c \in \mathcal{C}} V_c(\mathbf{T}_c^{t-1}, l_c). \end{aligned} \tag{7}$$

We solve the minimization of the higher order cliques in Eq.(7) by transforming them into quadratic functions [11] using a $(t+1)$-state switching variable which finds the deviation function which assigns the lowest cost to the labeling:

$$\min V_c(\mathbf{T}_c^0) = \min_{\mathbf{T}_c^0, z \in \{1,2,\ldots,t+1\}} f(z) + \sum_{i \in c} g(z, T_i) \tag{8}$$

where $f(z) = \{\theta_q, \theta_{\max}\}$ is a cost assigning function depending on the state variable z and $g(z, T_i) = w_{il}^q$ when $z = q$ and $T_i = l \in \mathcal{L}$, while $g(z, T_i) = 0$ when $z = t+1$. We apply a Primal-Dual algorithm called FastPD [12] which can efficiently solve the registration problem in a discrete domain by formulating the duality theory in linear programming. The advantage of such an approach lies in its generality, efficient computational speed, and guarantees the global optimum without the condition of linearity.

3. RESULTS

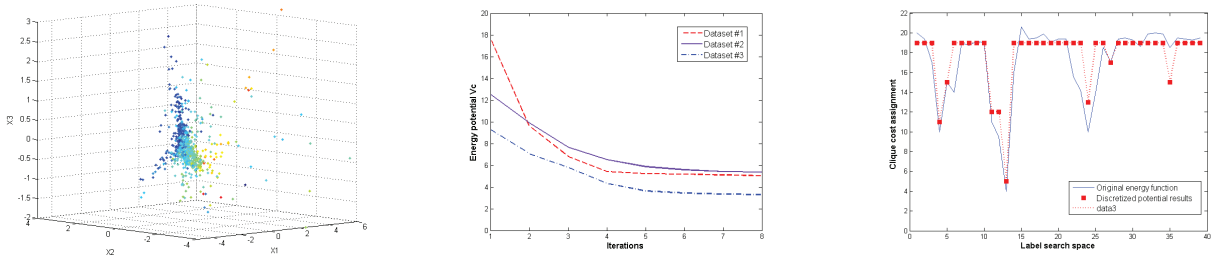
We experimented the articulated inferences of preoperative spine models to intraoperative data by confronting the obtained registered landmark accuracy to expert identification. A dataset of 12 separate CT volumes of the lumbar and main thoracic regions were obtained from different patients ($512 \times 512 \times 251$, resolution: 0.8×0.8 mm, thickness: $1 - 2$ mm), acquired for operative planing purposes. Preoperative X-rays of patients were obtained for initial 3D reconstruction. The CT data was manually annotated with 3D landmarks, corresponding to left and right pedicle tips as well as midpoints of the vertebral body. The smoothness term was set at $\lambda = 0.4$, while the clique variable was $\alpha = 0.3$. Tests were performed in C++ on a 2.8 GHz Intel P4 processor and 4 GB memory.

We first experimented the manifold-based higher order energy functions based on the search of the label space. Since the functional potentials are directly linked to the costs assigned by the geodesic distances to the manifold, we can therefore assess the performance of this metric as an efficient parametrization of the higher order cliques. Fig. 3 displays the obtained embedding for a dataset of 711 spine models in $\mathcal{M}$, as well as selected energy potential distributions. Results show the search for the minimum energy is able to obtain cliques that fall near the manifold, thus corresponding anatomically coherent configurations.

Then for each case of the CT dataset, registration is performed to automatically align the CT volume with $\gamma = 0.05$ to the given preoperative model $\mathbf{S}$ and quantitative assessment consisted of measuring the RMS distance with the manually segmented landmarks. Table 1 presents the results from this experiment with 3D landmark RMS differences and surface errors based on DICE mesh overlap scores, comparing the proposed method with cost functions excluding pairwise and higher order energy terms. Results for overall vertebral landmark errors have improved by 0.38 mm compared to the previous approach [8], which is significant for the required guidance accuracy. These results seem to confirm that exploiting higher order geometrical constraints on the whole shape prior does help to converge towards a global minimum. Visual registration results of the 3D model on selected multi-planar CT slices is shown in Fig. 4.

4. CONCLUSION

Statistical deformable model methods are often dedicated to single anatomical structures. Shape analysis of articulated models on the other hand has been sparsely investigated due to the difficulty in constraining the higher number of transformation variables. The method we propose not only allows to infer shape deformations of object constellations using discrete optimization techniques, but offers the possibility to learn the variations of spinal shape in complex corrective procedures. Hence we progress to whole body deformation by introducing novel high order cliques in the optimization to impose global shape constraints, while integrating the advantages of image-based registration. Increased accuracy of pedicle landmark localization was achieved towards this end.



(a) (b) (c)

Fig. 3. (a) Low-dimensional manifold of the spine dataset used to determine costs for clique variables. (b) High order energy potential distributions. (c) Parametrization of manifold clique variables in the label search space corresponding to the original energy function.

	Image-term		Image + pairwise terms		**Higher order method**	
	RMS (mm)	DICE (%)	RMS (mm)	DICE (%)	RMS (mm)	DICE (%)
Thoracic vertebrae	11.4	82	2.2	91	1.7	94
Lumbar vertebrae	5.2	80	1.9	94	2.0	93
Total vertebrae	9.6	81	2.1	92	**1.8**	**94**

Table 1. Comparison between optimization schemes with cost functions integrating only image related terms, singleton-pairwise potentials, and the proposed method with additional higher order functions. Evaluation is made on root-mean-square (RMS) differences (inferred vs. annotated landmarks) and with DICE scores from segmented vertebrae.

Future work will look at adapting local shape variations using mesh relaxation techniques based on bone density fields which may increase the accuracy of the geometrical alignment. Furthermore, a real-time feasibility evaluation of the approach during corrective surgery is planned for clinical use.

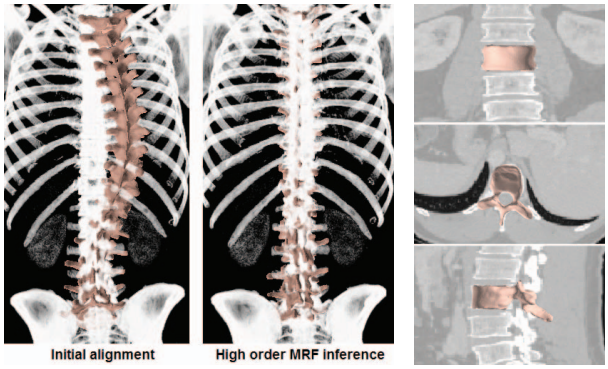


Fig. 4. Qualitative assessment of multimodal registration results of spine model $\mathbf{S}$ with CT images $\mathcal{I}$.

5. REFERENCES

[1] L.-P. Amiot and F. Poulin, "Computed tomography-based navigation for hip, knee, andm spine surgery," *Clinical Ortho. Related Res.*, vol. 421, pp. 77–86, 2004.

[2] K. Foley, D. Simon, and Y. Rampersaud, "Virtual fluoroscopy: computer-assisted fluoroscopic navigation," *Spine*, vol. 26, pp. 347–51, 2001.

[3] P. Markelj et al., "Robust gradient-based 3-D/2-D registration of CT and MR to X-ray images," *IEEE Trans. Med. Imag.*, vol. 27, pp. 1704–14, 2008.

[4] G. Zheng and X. Dong, "Unsupervised reconstruction of a patient-specific surface model of a proximal femur from calibrated fluoroscopic images," in *MICCAI*, 2007, pp. 834–41.

[5] G. Zheng, "Statistically Deformable 2D/3D Registration for Accurate Determination of post-operative cup orientation from single standard X-ray radiograph," in *Proc. MICCAI*, 2009, pp. 820–27.

[6] J. Boisvert, F. Cheriet, X. Pennec, H. Labelle, and N. Ayache, "Geometric variability of the scoliotic spine using statistics on articulated shape models," *IEEE Trans. Med. Imag.*, vol. 27, pp. 557–68, 2008.

[7] T. Klinder, R. Wolz, C. Lorenz, A. Franz, and J. Ostermann, "Spine segmentation using articulated shape models," in *Proc. MICCAI*, 2008, pp. 227–34.

[8] S. Kadoury and N. Paragios, "Surface/volume-based articulated 3D spine inference through Markov Random Fields," in *Proc. MICCAI*, 2009, pp. 92–99.

[9] S. Kadoury, F. Cheriet, and H. Labelle, "Personalized X-ray 3D reconstruction of the scoliotic spine from statistical and image models," *IEEE Trans. Med. Imag.*, vol. 28, pp. 1422–35, 2009.

[10] N. Paragios and R. Deriche, "Geodesic active regions: New paradigm to deal with frame partition problems in CV," *Visual Comm. Image Repre.*, vol. 13, pp. 249–68, 2002.

[11] C. Rother, P. Kohli, W. Feng, and J. Jia, "Minimizing sparse higher order energy functions of discrete variables," in *Proc. CVPR*, 2009, pp. 1382–89.

[12] N. Komodakis, G. Tziritas, and N. Paragios, "Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies," *CVIU*, vol. 112, pp. 14–29, 2008.