Patient surface model and internal anatomical landmarks embedding

Xia Zhong[1], Norbert Strobel[2,4], Annette Birkhold[2], Markus Kowarschik[2], Rebecca Fahrig[2], Andreas Maier[1,3]

[1]Pattern Recognition Lab, FAU Erlangen-Nürnberg
[2]Siemens Healthcare GmbH, Forchheim Germany
[3]Erlangen Graduate School in Advanced Optical Technologies(SAOT)
[4]Fakultät für Elektrotechnik, Hochschule für angewandte Wissenschaften Würzburg-Schweinfurt
xia.zhong@fau.de

Abstract. The patient surface model has shown to be a useful asset to improve existing diagnostic and interventional tasks in a clinical environment. For example, in combination with RGB-D cameras, a patient surface model can be used to automate and accelerate the diagnostic imaging workflow, manage patient dose, and provide navigation assistance. A shortcoming of today's patient surface models, however, is that, internal anatomical landmarks are not present. In this paper, we introduce a method to estimate internal anatomical landmarks based on the surface model of a patient. Our method relies on two major steps. First, we fit a template surface model is to a segmented surface of a CT dataset with annotated internal landmarks using keypoint and feature descriptor based rigid alignment and atlas-based non-rigid registration. In a second step, we find for each internal landmark a neighborhood on the template surface and learn a generalized linear embedding between neighboring surface vertices in the template and the internal landmark. We trained and evaluated our method using cross-validation in 20 datasets over 50 internal landmarks. We compared the performance of four different generalized linear models. The best mean estimation error over all the landmarks was achieved using the lasso regression method with a mean error of 12.19 ± 6.98 mm.

1 Introduction

Patient modeling has shown great potential in medical applications. One reason is that RGB-D cameras have become more easily available. For example, in the diagnostic environment, Singh et al. [1] demonstrated that a patient surface model can be fitted to the patient using a RGB-D camera ahead of CT scans to automate and accelerate the workflow. In the interventional environment, a patient surface model was used to estimate the skin dose [2], monitor the breathing motion, or provide navigation assistance [3]. The patient model used in these applications is a surface model. Related activities described in the literature derive an articulated skeleton [4], model the shape of the surface model

and compare different estimation methods [5]. Unfortunately, clinical relevant internal anatomical landmarks are not included in the surface model, as it only models the outer shape of a patient.

In this paper, we propose a method for finding internal landmarks of the human body e.g. vertebra, based on a surface model and evaluate its accuracy. Once internal landmarks are found, we can embed an appropriately adjusted spine model into the surface model. As organs have a fixed position relative to the spine, fitting a spine model to a surface model can, for example, be used for organ-specific positioning of a C-arm system.

2 Materials and Methods

Our learning-based method embeds the internal landmarks into a patient surface model. We use CT datasets with annotated landmarks to train our method. For each CT data set, the surface is segmented and represented as a surface mesh $\boldsymbol{x}$. We register a template surface model $\hat{\boldsymbol{x}}$ to this segmented surface. Afterwards, we learn the mapping function $\mathcal{M}_i$ to establish the relation between model $\hat{\boldsymbol{x}}$ and associated landmark $\boldsymbol{l}$ such that

$$\mathcal{M}_i = \arg\min_{\mathcal{M}_i} \sum_t \|\boldsymbol{l}_{t,i} - \mathcal{M}_i(\hat{\boldsymbol{x}}_t)\|_2^2 \tag{1}$$

2.1 Surface model registration and fitting

We first trained an Atlas for the patient model using the Civilian American and European Surface Anthropometry Resource Project (CAESAR) data. Using an active shape model (ASM), a patient surface model can be described as $\hat{\boldsymbol{x}}_t = \bar{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{b}_t$, where the $\bar{\boldsymbol{x}}$ is the mean surface model in the atlas, $\boldsymbol{D}$ describes a matrix comprising the modes of variation, and $\boldsymbol{b}_t$ is the associated weighting vector. Then, for each segmented surface model $\boldsymbol{x}_t$, we need to find the best fitting weighting vector $\boldsymbol{b}_t$ and associated rigid transformation $\mathcal{T}_t$ by minimizing

$$\boldsymbol{b}_t, \mathcal{T}_t = \arg\min_{\boldsymbol{b},\mathcal{T}} \mathcal{D}\left(\boldsymbol{x} - \mathcal{T}(\bar{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{b})\right)^2 \tag{2}$$

In Eq. (2), the function $\mathcal{D}$ calculates the minimal distance between two meshes. To solve this minimization problem efficiently, we rely on alternating minimization. We first estimate the transformation $\mathcal{T}$ based on a given $\boldsymbol{b}$. This is a well-known surface mesh rigid registration problem and can be solved using feature descriptors e.g. HoG, SPIN, and LDSIFT [6] and feature matching. The challenge in our case is that the two meshes may partially overlap and the deformation between two meshes may be non-rigid. As patients usually are positioned head first, supine, i.e., are on their back, the outer shape of the patient may potentially lack distinctive key points. To establish correspondences between feature points robustly, we first sample the key points as propsed by Sahillioğlu et al. [7], such that the sampling is almost uniform. In this way, we

avoid accumulation of feature points in high curvature regions when identifying key points. For each key point, we calculate the LDSIFT feature descriptor, which is robust to non-rigid deformations [6]. We further incorporate the prior knowledge, that the transformation $\mathcal{T}$ comprises mainly translations and that rotation is very limited. In our actual implementation, we divided the surface mesh into eight primary districts using the principal axis of the mesh, and rejected any cross-district matching. Due to the non-rigid deformation between two meshes, we may find ambiguity in rigid transformation. To counteract this effect, we consider all feasible results using RANSAC as motion cluster and use mean shift algorithm to search the cluster center of the biggest cluster. This cluster center is used as our motion estimation $\mathcal{T}_t$. Knowing the transformation $\mathcal{T}_t$, the weighting factor $\boldsymbol{b}_t$ can be estimated by minimizing

$$\boldsymbol{b}_t = \arg\min_{\boldsymbol{b}} \mathcal{D}\left(\boldsymbol{x} - \mathcal{T}_t(\bar{\boldsymbol{x}} + \boldsymbol{D}\boldsymbol{b})\right)^2 + \alpha_1 \mathcal{R}\left(\boldsymbol{x}, \hat{\boldsymbol{x}}\right) \tag{3}$$

where the scalar α is a parameter, and the function $\mathcal{R}$ calculates the overlap ratio between the segmented and the estimated surface mesh. This function regularizes the minimization problem such that the estimated surface will not be degenerated. After this step, the residual error between estimated and segmented surface is minimized using non-rigid mesh registration, as proposed by Allen et al. [8].

2.2 Surface and internal landmark embedding

After fitting the surface model $\hat{\boldsymbol{x}}$ to the segmented CT surface $\boldsymbol{x}$, we look for the mapping $\mathcal{M}_i$ between $\hat{\boldsymbol{x}}$ and the annotated internal landmarks $\boldsymbol{l}_i$ in CT data sets. For the i^{th} landmark in the t^{th} CT data $\boldsymbol{l}_{t,i}$, we assume that the landmark position (coordinates) can be expressed as a generalized linear combination of neighboring vertices $\hat{\boldsymbol{x}}_{t,j} \in \mathcal{N}(\boldsymbol{l}_i)$. The neighborhood is defined using k-nearest neighbor with Mahalanobis metric. We used the covariance matrix of the surface model $\boldsymbol{x}$ to calculate the Mahalanobis distance. We select a large neighborhood such that the residual error due to non-linearity in mapping is minimized. The number and indicies of neighborhood vertices can vary in each data set. We use the unions according to the indicies as the neighborhood for each internal landmark. To describe this more formally, we introduce matrix $\boldsymbol{X}_{\mathcal{N}(\boldsymbol{l}_i)} = [X^T_{1,\mathcal{N}(\boldsymbol{l}_i)}, \cdots, X^T_{N,\mathcal{N}(\boldsymbol{l}_i)}]^T$ and matrix $\boldsymbol{L}_i = [\boldsymbol{l}^T_{1,i}, \cdots, \boldsymbol{l}^T_{1,i}]^T$ where matrix $X_{t,\mathcal{N}(\boldsymbol{l}_i)} = [\boldsymbol{x}_{t,1}, \cdots, \boldsymbol{x}_{t,M}]$ comprises all vertices in an associated neighborhood for the t^{th} data set and the landmark $\boldsymbol{l}_{t,i}$. The scalar N denotes the number of dataset and the scalar M denotes the number of neighboring vertices. At this point, we can formulate a generalized linear mapping with a cost function for the linear mapping weighting vector $\boldsymbol{w}_i$ for landmarks $\boldsymbol{l}_{t,i}$.

$$\boldsymbol{w}_i = \arg\min_{\boldsymbol{w}_i} \|\boldsymbol{X}_{\mathcal{N}(\boldsymbol{l}_i)}\boldsymbol{w}_i - \boldsymbol{L}_i\|^2 + \lambda\mathcal{P}(\boldsymbol{w}_i) \tag{4}$$

In Eq. (4), the scalar λ is a Lagrange multiplier for the penalty function $\mathcal{P}$. The penalty function $\mathcal{P}$ is different for each generalized model. In ridge regression, the

L2 norm of $\boldsymbol{w}$ is used while in lasso, the L1 norm is used respectively. The elastic net regression using a mixture of L1 and L2 norm while in linear embedding the sum of $\boldsymbol{w}$ is used. We investigated the linear embedding [9], ridge regression, lasso regression and elastic net regression method to estimate the landmark position. Noted, that in case of linear embedding regression, the minimization problem can be solved in closed form.

3 Evaluation and Results

We trained our algorithm using Anatomy3 from the Visceral dataset [10]. This dataset comprises 20 full-body CT scans with 50 associated internal landmarks in each dataset. An example of the dataset can be found in Fig. 2. We evaluated our algorithm using leave-one-out cross-validation and the estimation error is given in mean $\pm$ standard deviation in mm. We compared the estimation result using linear embedding (LLE), ridge regression (Ridge), lasso regression (Lasso) and elastic net regression (Elastic). The results are shown in Fig. 1. As we can see from the results, in most of the cases, the ridge regression has the worst performance with an overall estimation error of 22.19 ± 12.36 mm. The closed-form solution of LLE method outperformed ridge regression and reduced the mean error to 13.43 ± 12.00 mm. Using elastic regression method, the standard deviation was reduced further and the mean estimation error was 12.71 ± 7.32 mm. The best performance was achieved using lasso regression. The overall landmark estimation error then was 12.19 ± 6.98 mm.

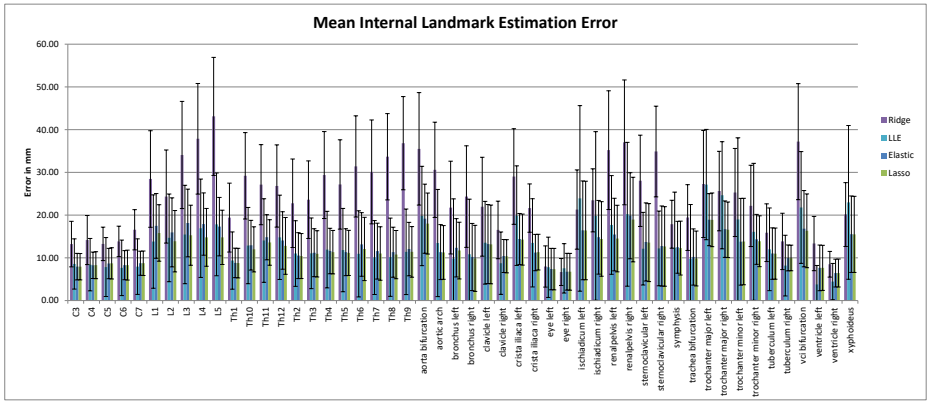

Fig. 1. Mean internal landmark estimation error using linear embedding (LLE), ridge regression (Ridge), elastic net regression (Elastic) and lasso regression (Lasso)

4 Discussion

The goal of our work has been to position anatomical landmarks inside a surface based on surface vertices close to associated landmarks. We found that our approach based on a linear mapping relative to a template surface yielded good matches between predicted landmark positions and their actual positions. The results also indicate that our template fitting method is robust as the accuracy for the template fitting is crucial for the subsequent internal landmark estimation. We can also see from the results that the ridge and LLE methods have a higher estimation error than Elastic and Lasso regression. One of the reasons can be that by introducing L1 norm as a regularizer, the Elastic and Lasso regression impose the sparsity of weighting vector $\boldsymbol{w}$. This might be an indication that we could improve our neighborhood selection method and only include vertices which are significant for the estimation. In this work, we do not investigate the mapping using non-linear methods as the number of our datasets is limited. So far, we estimated internal landmarks independently and have not yet considered their joint estimation. As human anatomy follows certain rules there should be a potential to improve our results by exploiting any mutual relationships between them. Therefore, for certain groups of landmarks, e.g., the spine, an active shape model of the spine could potentially further improve the estimation result. We see our work of embedding the internal landmarks to the surface model as adding prior information to our datasets. In Fig. 2, we show an example of fitting an adapted spline model using the estimated internal landmarks. In future work, we will look into the refinement method of this initial estimation when pre-operative image data is available, i.e., we will try to calculate an a posteriori estimate based on our a-priori estimate.

Acknowledgments We gratefully acknowledge the support of Siemens Healthineers, Forchheim, Germany. We also thank Siemens Corporate Technology for providing the avatar database. Note that the concepts and information presented in this paper are based on research, and they are not commercially available.

References

1. Singh V, Chang Y, Ma K, et al. Estimating a Patient Surface Model for Optimizing the Medical Scanning Workflow. In: Med Image Comput Comput Assist Interv; 2014. p. 472–479.
2. Johnson PB, Borrego D, Balter S, et al. Skin dose mapping for fluoroscopically guided interventions. Med Phys. 2011;38(10):5490–5499.
3. Bauer S, Wasza J, Haase S, et al. Multi-modal Surface Registration for Markerless Initial Patient Setup in Radiation Therapy using Microsoft's Kinect Sensor. In: Proc IEEE Int Conf Comput Vis; 2011. p. 1175–1181.
4. Anguelov D, Koller D, Pang HC, et al. Recovering articulated object models from 3D range data. In: Uncertain Artif Intell; 2004. p. 18–26.
5. Zhong X, Strobel N, Kowarschik M, et al. Comparison of Default Patient Surface Model Estimation Methods. In: Proc BVM; 2017. p. 281–286.

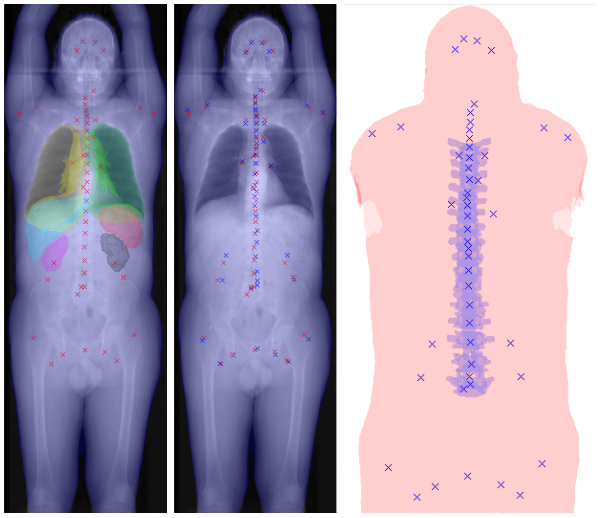

Fig. 2. A sample of the Anatomy3 training set is shown on the left. The annotated internal landmarks are indicated using red crosses and the segmented internal organs are shown in different colors. The middle figure shows the output of our algorithm, i.e., internal landmarks derived based on the surface. The result on the right demonstrates the same idea, but this time a patient surface model was used which was generated based on patient meta data such as height and weight.

6. Darom T, Keller Y. Scale-invariant features for 3-D mesh models. vol. 21; 2012. p. 2758–2769.
7. Sahillioğlu Y, Yemez Y. Minimum-distortion isometric shape correspondence using EM algorithm. IEEE Trans Pattern Anal Mach Intell. 2012;34(11):2203–2215.
8. Allen B, Curless B, Popović Z. The space of human body shapes: reconstruction and parameterization from range scans. In: ACM Trans Graph. vol. 22; 2003. p. 587–594.
9. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. science. 2000;290(5500):2323–2326.
10. del Toro OAJ, Goksel O, Menze B, et al. VISCERAL–VISual Concept Extraction challenge in RAdioLogy: ISBI 2014 challenge organization. Proc VISCERAL Challenge at ISBI. 2014; p. 6–15.