ARTICULATED ATLAS FOR SEGMENTATION OF THE SKELETON FROM HEAD & NECK CT DATASETS

Sebastian Steger[⋆]*, Matthias Kirschner*[†]*, Stefan Wesarg*[†]

[⋆] Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[†] Technische Universität Darmstadt, Graphisch-Interaktive Systeme, Darmstadt, Germany

ABSTRACT

In this paper a novel articulated atlas for the fully automated segmentation of the skeleton from head & neck CT datasets is presented. An individual atlas describing the shape and appearance is created for each individual bone. Principal Component Analysis is used to learn spatial relations between those atlases resulting in a unified articulated atlas. Transformations are parameterized using the matrix exponential to enable linear combinations required for learning. The adaptation to test images considers appearance, distance to bone structures and the trained articulation space.

For evaluation, an atlas created from 10 manually labeled training images has been applied to 46 clinically acquired head & neck CT datasets. Visual inspection showed that in 74% of the cases, the adaptation process was successful. In a second experiment leave-one-out validation was used to quantify the segmentation accuracy. The successfully adapted cases resulted in an average volume overlap error of 30.67 and an average symmetric surface distance of 0.76 mm.

Index Terms— Bone Segmentation, Articulated Atlas, Statistical Shape Models

1. INTRODUCTION

Due to its high contrast, bone can easily be segmented in CT images. Sometimes however, it is not only required to know if a voxel belongs to either *bone* or *not bone*, but also to which bone (e.g. skull, mandible, C3) the voxel belongs. An example for that is a landmark based image registration scheme based on corresponding segmented bones in both images. As the shapes of bones can become quite complex and different bones are close to each other (e.g. the vertebrae in the spine), it is not an easy task to distinguish neighboring bones from each other. Thus model knowledge is required.

One possibility to separate bone structures is to incorporate prior knowledge of an organ shape into the segmentation process: As an example, Kirschner et al. [1] used Statistical Shape Models (SSMs) to segment single vertebrae. However, this approach is too local for our scenario, as we are interested in segmenting multiple structures simultaneously. Relative pose of mulitple organs can be modeled by means of articulated models. One way to build articulated models is to explicitly model valid transformations between structures (e.g. the joint between femur and pelvis [2] or various types of joints for the atlas based segmentation of mice form CT data [3]). Alternatively, relative pose can be learned from training examples [4]. In this paper, we follow the latter approach, and learn the space of relative pose – the *articulation* – of bones in the head & neck region. Our articulated model is combined with atlases for each individual bone, which is used to segment the bones with high accuracy.

This work is partially funded within the NeoMark project (FP7-ICT-2007-2-224483) by the European Commission.

2. ARTICULATED ATLAS

The articulated atlas (AA) consists of several *items* i, one for each individual bone, i.e. the skull, the mandible and the vertebrae. For each item a probability image, an average gray value image and a relative transform with respect to the center of the atlas is given. Principal component analysis (PCA) is used to create a model describing possible item articulations.

2.1. Parameterization of the transformations

The transformation of each item has 7 degrees of freedom, 3 for translation, 3 for rotation and 1 for isotropic scaling, i.e. a *similarity transformation*. To enable the PCA, the transformations should be parameterized with a unique and linear 7-dimensional vector. Whereas this is straightforward for translation and scaling, commonly used parameterizations for rotations – such as euler angles or versors – are non linear. Alexa [5] tackled this problem by using the matrix exponential to describe transformations:

$$T_{\mathbf{a}}(\mathbf{x}) = exp\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_5 & a_6 & a_7 & a_8 \\ a_9 & a_{10} & a_{11} & a_{12} \\ a_{13} & a_{14} & a_{15} & a_{16} \end{pmatrix} \cdot \mathbf{y}\,, \qquad (1)$$

where $\mathbf{a}$ is the parameter vector, $\mathbf{y}$ is the point in homogeneous coordinates that is to be transformed and the matrix exponential is given by $exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}$. Note that this

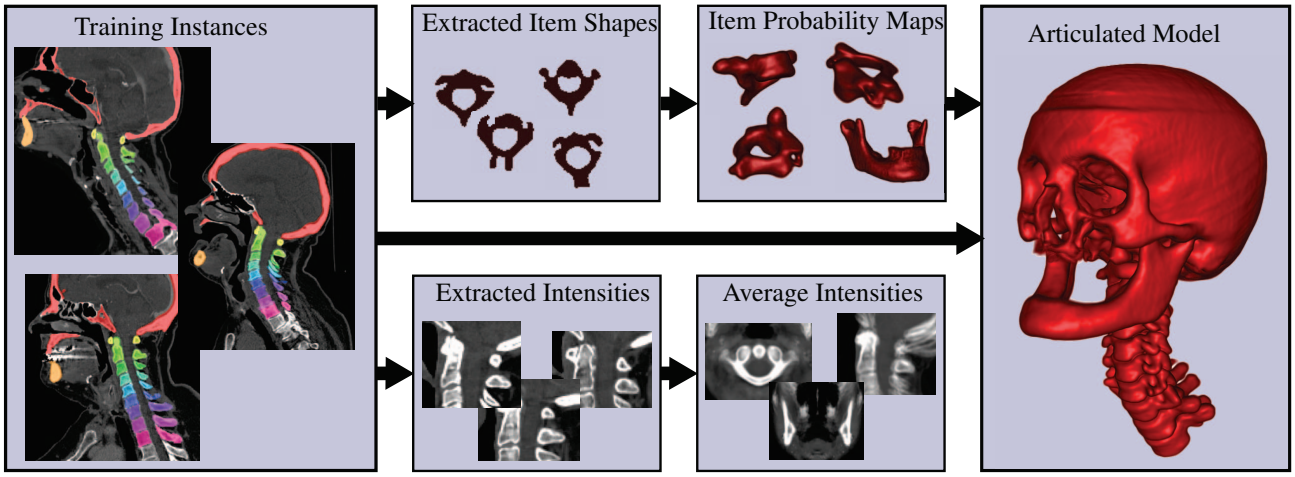


Fig. 1. Training of the articulated atlas.

kind of parameterization has the properties we would expect from a linear parameterization: $T_{\mathbf{0}}$ corresponds to the identity transform; $T_{-\mathbf{a}}$ is $T_{\mathbf{a}}$'s inverse; $T_{n\cdot\mathbf{a}}$ executes $T_{\mathbf{a}}$ n times successively (e.g. if $T_{\mathbf{r}}$ is a rotation of 1 degree around a certain axis, $T_{n\cdot\mathbf{r}}$ is a rotation of n degrees around the same axis).
We exploit these properties to restrict the possible transformations to translations, rotations and scaling only:

$$\begin{aligned} T_{\mathbf{p}}(\mathbf{y}) = exp(p_0 S \quad &+ \quad p_1 R_x + p_2 R_y + p_3 R_z \\ &+ \quad p_4 T_x + p_5 T_y + p_5 T_z) \cdot \mathbf{y} \ , \end{aligned} \quad (2)$$

where T_d corresponds to a small translation in direction d, R_a to a small rotation around axis a and S to a small scaling. The 7 dimensional vector $\mathbf{p}$ describes the transformation of a single item. The whole articulated atlas with m items can thus be characterized with $7m$-dimensional parameter vector $\mathbf{x}$.

2.2. Articulated Atlas Creation

The creation of the articulated atlas (see Fig. 1) is based on n annotated CT datasets. In each training instance $j \in [0..n-1]$ all voxels of each item were manually assigned to a unique label.
Item shape training: Each model item i is extracted from each training instance. All extracted instances of each model item are transformed to a common coordinate system using a similarity transform in order to maximize the volume overlap. An average image is created from the aligned binary images, approximating a *probability image* $p_j : D \subset \mathbb{R}^3 \mapsto [0,1]$ for an item.
Training vector description: For each training instance j, first the global location and orientation is obtained. This is done by rigidly aligning a reference label image to the training label instance, thereby maximizing the overlap. Then the probability image of each item is aligned to the corresponding item in the label image. The resulting absolute transformations are set into relation to the instance's global location and orientation and a 7-dimensional parameter vector p_i can be obtained for each of the m items. The resulting $7m$ dimensional vector $\mathbf{x_j}$ uniquely describes the articulation of the training instance j, independent from its global position and orientation.
Articulation training: The space of all possible articulations can be described by applying the PCA on the training articulation matrix $X = (\mathbf{x_1}, .., \mathbf{x_n})$, like in a SSM. Note that the fundamental difference between the AA and a SSM is that correlations between linearly parameterized transformations are exploited instead of correlations between model points. An arbitrary articulation $\mathbf{x}$ can then be described with a $k \ll 7m$ dimensional vector $\mathbf{b}$:

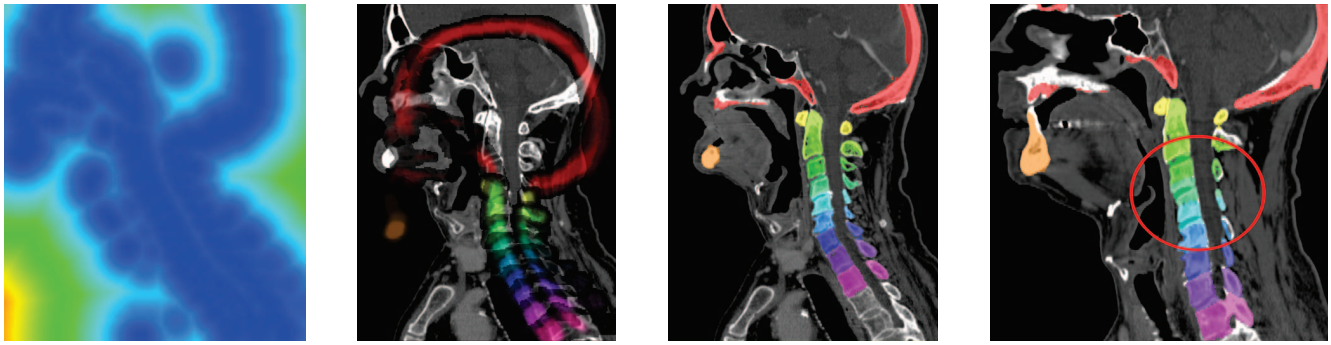
$$\mathbf{x} = \bar{\mathbf{x}} + A \cdot \mathbf{b} + \mathbf{r} \ , \quad (3)$$

where $\bar{\mathbf{x}}$ is the average articulation of all training instances x_j, A is a matrix consisting of the eigenvectors of the covariance matrix $C = \frac{1}{n-1}\bar{X}\bar{X}^T$, and $\bar{X} = (\mathbf{x}_0 - \bar{\mathbf{x}}, \ldots, \mathbf{x}_{n-1} - \bar{\mathbf{x}})$. The more likely an articulation $\mathbf{x}$ is, the smaller the magnitude of $\mathbf{b}$ and $\mathbf{r}$ will be.
Intensity training: The absolute transformation of each model item in each training instance is known. Based on that, a box around each item is extracted from all training CT datasets and back transformed to the item's probability image coordinate system. An average intensity image $g_j : D \subset \mathbb{R}^3 \mapsto \mathbb{R}$ is then created for each item.

2.3. Atlas adaptation

The atlas adapted to a CT image $I : \mathbb{R}^3 \mapsto \mathbb{R}$ by minimizing an energy functional depending on the 6 dimensional external parameter vector $\mathbf{t}$ describing the global translation and rotation and the $7m$ dimensional articulation parameter vector $\mathbf{x}$:

$$E(\mathbf{x}, \mathbf{t}) = E_{\text{i}}(\mathbf{x}, \mathbf{t}) + \alpha E_{\text{bd}}(\mathbf{x}, \mathbf{t}) + \beta E_{\text{md}}(\mathbf{x}) + \gamma E_{\text{ar}}(\mathbf{x}) \ . \quad (4)$$



(a) Distance to bone voxels (blue=small, red=large)

(b) Average pose atlas in the image center before adaptation

(c) Segmentation obtained by a successfully adapted atlas

(d) Incorrect segmentation due to adaptation to wrong vertebrae

Fig. 2. Adaptation of the articulated atlas to head & neck CT datasets.

E_{i} ensures gray value similarity between the test image and the trained intensities of the atlas items, E_{bd} ensures that the atlas items are close to high CT intensities (i.e. bones), E_{md} reduces overlap of neighboring items and E_{ar} ensures that the atlas is within or at least close to the trained articulation space. α, β and γ are weights controlling the relative impact of each term. Energy optimization is done with a simple gradient descent optimizer, first finding the optimal external parameters $\mathbf{t}$ and then jointly finding articulation and external parameters $\mathbf{x}, \mathbf{t}$. Initially the center of the model is moved to the center of the test image.

Intensity Term: Gray value similarity is measured using a weighted mean squared error:

$$E_{\text{i}}(\mathbf{x},\mathbf{t}) = \sum_{\forall j} \sum_{\forall \mathbf{y} \in D} (G_\sigma * p_j)(\mathbf{y}) \left(g_j(\mathbf{y}) - I((T_t \circ T_{\mathbf{x}_j})(\mathbf{y}))\right)^2$$

where G is the Gaussian kernel. The Gaussian smoothing of the item's probability image ensures that not only gray values within the model item but also in the neighborhood have an impact on the energy functional.

Bone Distance Term: The test image is thresholded with a threshold of 200 HU resulting in a rough segmentation of all bony structures. Then the largest connected component is extracted and the distance image $BD : \mathbb{R}^3 \mapsto \mathbb{R}_0^+$ is created (see Fig. 2). Based on that, the bone distance term is computed:

$$E_{\text{bd}}(\mathbf{x},\mathbf{t}) = \sum_{\forall j} \sum_{\forall \mathbf{y} \in D} p_j(\mathbf{y}) \cdot DB((T_t \circ T_{\mathbf{x}_j})(\mathbf{y}) \,. \qquad (5)$$

This term pushes the model to bony structures in the image ensuring convergence even if the model's initial position is far away from its final position.

Model Distance Term: The overlap between neighboring atlas items is measured by:

$$E_{\text{md}}(\mathbf{x}) = \sum_{\forall j} \sum_{\forall \mathbf{y} \in D} \sum_{\forall k \in N(j)} p_j(\mathbf{y}) p_k(T_{\mathbf{x}_k}^{-1} \circ T_{\mathbf{x}_j}(\mathbf{y})) \,, \quad (6)$$

where $N(j)$ is the set of all neighboring atlas items of j. This term supports the articulation restriction by preventing that neighboring items erroneously converge to the same entity during the adaptation process. This risk is particularly high for neighboring items with similar appearance (e.g. vertebrae).

Articulation Restriction Term: The range of possible transformations is restricted to transformations within or close to the learnt space of transformation. We model the probability density of the shapes by an approximate Gaussian distribution using the approach of Moghaddam and Pentland [6]. The articulation restiction term then becomes

$$E_{\text{ar}}(\mathbf{x}) = \sum_{i=1}^{k} \frac{b_i^2(\mathbf{x})}{\lambda_i} + \frac{n-k-1}{\sum_{j=k+1}^{n-1} \lambda_j} \|\mathbf{r}(\mathbf{x})\|^2 \,, \qquad (7)$$

where the λ_i, $\lambda_1 \geq \ldots \geq \lambda_{n-1} > 0$ are the non-zero eigenvalues of C. The first term is the Mahalanobis distance in the linear subspace $\mathcal{U}$ spanned by the first eigenvectors A, and the second term approximates the Mahalanobis distance in the subspace complementary to $\mathcal{U}$. The latter term ensures that articulations are not completely restricted to the linear subspace $\mathcal{U}$, which increases the flexibility of the model.

2.4. Label Extraction

Once the model is adapted to the test image, each of its voxels needs to be assigned to either background or to a unique atlas item label l. This is done based on the atlas items' probabilities at the voxel's location $\mathbf{z}$ and on the voxel's image intensity $I(\mathbf{z})$:

$$l(\mathbf{z}) = \underset{\forall j}{\text{argmax}} \; p_b(I(\mathbf{z})) \cdot p_j(T_{\mathbf{x}_j}^{-1} \circ T_t^{-1}(\mathbf{z})) \,, \qquad (8)$$

where p_b is the probability of being a bone voxel for a given image intensity. In case the probability product $p_b \cdot p_i$ drops below a certain threshold, $l(\mathbf{z})$ is instead set to special label reserved for background.

	all cases		successful adaptions	
	VOE	ASD	VOE	ASD
Skull	24.71	0.84	22.49	0.71
Mandible	30.72	1.16	31.36	1.16
Atlas	41.51	1.16	35.30	0.85
Axis	34.13	1.01	28.30	0.78
C3	43.64	1.33	30.82	0.71
C4	47.86	1.79	29.57	0.65
C5	50.51	2.15	31.01	0.70
C6	49.76	2.43	29.76	0.66
C7	47.51	2.53	28.20	0.69
Th1	48.08	2.74	29.68	0.83
Th2	50.24	3.26	33.40	0.99
Average	45.92	2.04	30.67	0.76
Std. Dev.	8.88	0.81	3.26	0.16

Table 1. Volume overlap error (VOE) and average symmetric surface distance (ASD) in mm for all test cases (left column) and only the successfully adapted cases (right column)

3. EXPERIMENTS & RESULTS

The evaluation was based on a set of 56 clinically acquired head & neck CT datasets from 40 oral cancer patients. The inter slice resolution ranged between 1.0 and 2.0 mm and some of the images had strong imaging artifacts. Some patients had an abnormal anatomy due to surgery and/or osteoporosis. For 10 representative instances of this set the skull, the mandible, the cervical vertebrae and the upper two thoracic vertabras were manually labeled. Those images along with their right/left flipped versions form the training set from which the articulated atlas was created.

This atlas was then applied to the remaining 46 CT scans using the same set of adaptation parameters. α, β and γ were choosen so that all energy terms have the same order of magnitude. As ground truth segmentation was not available, only qualitative results can be reported. Visual inspection of the segmentation result shows that in 34 cases (74%) all atlas items had proper location and rotation. In 7 cases (15%) on atlas vertebrae covered two vertebrae in test image or vice versa, so that all subsequent atlas items were shifted by one vertebra (see Fig. 2(d)). In the remaining 5 cases (11%) the adaptation process failed completely. Note that some of the erroneous cases could be recovered by either changing the adaptation parameters or by manually changing the initial position of the model.

In a second experiment leave-one-out cross validation was performed. 10 different atlases from only 9 of the 10 items were built and applied to the remaining item. The segmentation result was compared to ground truth for each atlas item individually using the volumetric overlap error (VOE) and the average surface distance (ASD). For all items combined the VOE was 45.92 and the ASD was 2.04 mm on the average. As in 3 cases some atlas vertebrae converged to the wrong vertebra in the test image, these results are not meaningful. Thus the results for the correct adaptations only were measured separately (30.67 average overall VOE and 0.76 mm average overall ASD). See Table 1 for details.

4. CONCLUSION & OUTLOOK

In this paper we presented a novel articulated atlas for the joint segmentation of multiple rigid structures. Shape and intensities of individual items are learnt along with the spatial relation (articulation) to each other. With an average ASD of 0.76 mm for correctly adapted instances, the articulated atlas turns out to be suitable for the fully automated segmentation of individual bones from head & neck CT datasets.

In contrast to deformable models, the dimensionality of the optimization problem is a lot lower and the initialization is less critical. The biggest challenge remaining is that in some cases, atlas items may adapt to wrong structures. In the future this shall be addressed by either modifying the adaptation process or the inclusion of further structures to the atlas (e.g. the ribs, the lungs) to provide a more global context. Furthermore we plan to rebuild the head & neck skeleton atlas using more training images, thereby hoping to increase the accuracy. Finally we aim at applying the articulated atlas to other parts of the body, e.g. the thorax.

5. REFERENCES

[1] M. Kirschner, M. Becker, and S. Wesarg, "3D active shape model segmentation with nonlinear shape priors," in *Proc. MICCAI 2011*, 2011, pp. 492–499.

[2] D. Kainmueller, H. Lamecker, S. Zachow, and H.-C. Hege, "An articulated statistical shape model for accurate hip joint segmentation," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, 2009, pp. 6345–6351.

[3] M. Baiker et al., "Atlas-based whole-body segmentation of mice from low-contrast micro-ct data," *Medical Image Analysis*, vol. 14, no. 6, pp. 723 – 737, 2010.

[4] J. Boisvert, F. Cheriet, X. Pennec, H. Labelle, and N. Ayache, "Geometric variability of the scoliotic spine using statistics on articulated shape models," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 557–568, 2008.

[5] M. Alexa, "Linear combination of transformations," New York, NY, USA, 2002, SIGGRAPH '02, pp. 380–387, ACM.

[6] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans Pattern Anal Mach Intell*, vol. 19, pp. 696–710, 1997.