Generation and Initialization of Stable 3D Mass-Spring Models for the Segmentation of the Thyroid Cartilage

Jana Dornheim[1], Lars Dornheim[1], Bernhard Preim[1], Ilka Hertel[2], and Gero Strauss[2]

[1] Otto-von-Guericke-Universität, Postfach 4120, D-39106 Magdeburg, Germany
[2] Hals-Nasen-Ohren-Universitätsklinik, Universitätsklinikum Leipzig, Liebigstr. 18a, D-04103 Leipzig, Germany

Abstract. The preoperative planning of primary tumor resections in the larynx region shall be supported by a 3D visualization of the patient-specific anatomy and pathological situation. This requires a segmentation of the larynx cartilage structures from computed tomography (CT) datasets.
In our work, we use 3D Stable Mass-Spring Models (SMSMs) for this segmentation task. Thereto, we create a specific 3D deformable shape model for the segmentation of the thyroid cartilage. A new concept for elastic initialization of this model is presented, allowing the deformable model to adapt specifically to patient-specific shape variations and pathological deformations.
We show that using our generation and initialization method, prototypical 3D deformable shape models can be adapted to very different patients without any prior training and prior knowledge about new patients' data.

1 Introduction

In the case of tumor affections in the larynx and lower hypopharynx, the patient's life expectancy and further life quality depend strongly on the required surgical treatment. The parts of the larynx which need to be resected, determine the patient's ability to breathe, swallow and speak. For the decision on a surgical strategy, the extent of the tumor must be evaluated with respect to infiltration of the following structures:

- the vocal chords and muscles (usually judged by laryngoscopy),
- the glottis, subglottic and supraglottic space, and
- the larynx cartilages, in particular the epiglottis, thyroid cartilage, cricoid cartilage, and the two arytenoid cartilages (often judged by CT [1]).

For the assessment of the air and cartilage structures, a 3D visualization of the patient-specific anatomy and pathological situation is desirable to reduce uncertainties in the chosen surgical procedure. This requires a precise segmentation of the larynx and its substructures from neck CT datasets. These images are very

rich of different small structures of high signal intensity (Figure 4 in the evaluation section gives an impression) making any segmentation task very difficult.

The inhomogenous nature of the cartilage itself makes its segmentation a challenging task, for which neither simple edge-based techniques, such as LiveWire, nor gray-value-based segmentation techniques are appropriate. Since profound anatomical knowledge is needed to bridge areas of weak signal in the cartilage wall, we target at a 3D model-based segmentation of the thyroid cartilage.

2 State of the art

For the segmentation of the thyroid cartilage, no specific approaches exist to our knowledge. General methods for the segmentation using 3D deformable models are known [2]: *3D Active Contours* or *Balloons* [3] incorporate rough shape knowledge by means of a viscousity condition. The use of an inflation force prevents them from shape collapse and drives them towards the target structure's contour. This global representation of shape does however not allow to model complex shape information. *Implicit 3D deformable models* [4] do not bear the problem of instability and need no inflation force. However, they are restricted to describing regular geometric shapes, that can be described by a simple equation. *Active Shape Models (ASM)* and *Active Appearance Models (AAM)* [5] provide support for segmenting more complex shapes by means of a statistical analysis of training data. They require large amounts of training data and a very good correspondance of 3D points, which makes model creation laborious and segmentation results potentially imprecise. For the segmentation of pathological shape variations (e.g. caused by a tumor), ASMs and AAMs are principally inappropriate, because these shape variations are very individual and cannot be trained.

Stable Mass-Spring Models (SMSMs, [6]) are prototypical 3D shape models, that need no training, but an initial model describing the expected shape. These models are especially appropriate for tracking and searching as well as for segmentation, if the individual structure is known in general. They are very robust to noise and gaps in the data, as [7] show for the segmentation of the left ventricle in 3D SPECT.

In our application, such specific deformations need to be modeled. Therefore it is not enough to create an SMSM like in [8] that prototypically models the target structure. A further adaption to the patient-specific pathological shape variation is necessary, which will be introduced in this work. This way, a segmentation model is created, that is directly tailored to an individual patient and does not represent unnecessary shape knowledge about other patient's like statistical models would.

3 Method

In our work, we construct an SMSM of the thyroid cartilage semi-automatically from a manual sample segmentation. We thereto adapt and refine the model

creation (section 4) introduced in [8], which proposes a model topology consisting of two parts created independantly from each-other:

1. an (outer) surface submodel containing masses with *gradient sensors* and the contour faces representing the modeled object contour, and
2. an (inner) volumetric submodel containing *intensity sensors*.

Both submodels are connected afterwards to an overall model.

To adapt this prototypical model to each individual patient's data, an elastic initialization technique is introduced (section 5.1) to translate, rotate, scale and deform the constructed general model nonlinearly and model-consistently to fit the segmentation target structure by means of key masses. Because of these starting conditions, the segmentation of the thyroid cartilage becomes possible in this difficult data.

4 Model generation for the thyroid cartilage

The semi-automatic model generation for the thyroid cartilage is based on a manual segmentation created from a dataset with a visually average-shaped larynx. The sample segmentation is available as a binary volume dataset. For our examination, two different models were created and evaluated:

1. One *volumetric* model, according to the model creation of [8], consisting of a surface submodel and a volumetric submodel connected by 1:1 connections.
2. For comparison, a pure *surface model* is created, consisting of the surface submodel only. This model was employed to evaluate whether segmentation based on edge detection alone is more appropriate than using gray value information.

4.1 Generation of the surface model

The sample segmentation was resampled down to an isotropic voxel size for efficient model generation. From this resampled segmentation, an isosurface was generated using the marching cubes method (Fig. 1(a)), smoothed and simplified with Quadric Error Metrics [9] to a certain degree (Fig. 1(b)), that is adjustable for different abstraction levels. The resulting number of triangles (in our case 50–200) provides an appropriate modelling of the cartilage shape while still allowing for realtime model simulation (Fig. 1(c)).

The resulting surface was used to create the surface model: For each vertex of the mesh, a mass point was created. For each edge in the mesh, a spring connecting the incident vertices, rsp. mass points, was created. All masses and spring constants throughout the model were set to 1.0. A *direction-weighted gradient sensor* [10] was attached to each mass point of the surface model, ensuring that contour masses are only attracted by image gradients of the same direction as the incident surface normal. This prevents the model from being distracted by neighbouring but irrelevant gradients, which is a common problem in neck CT datasets.

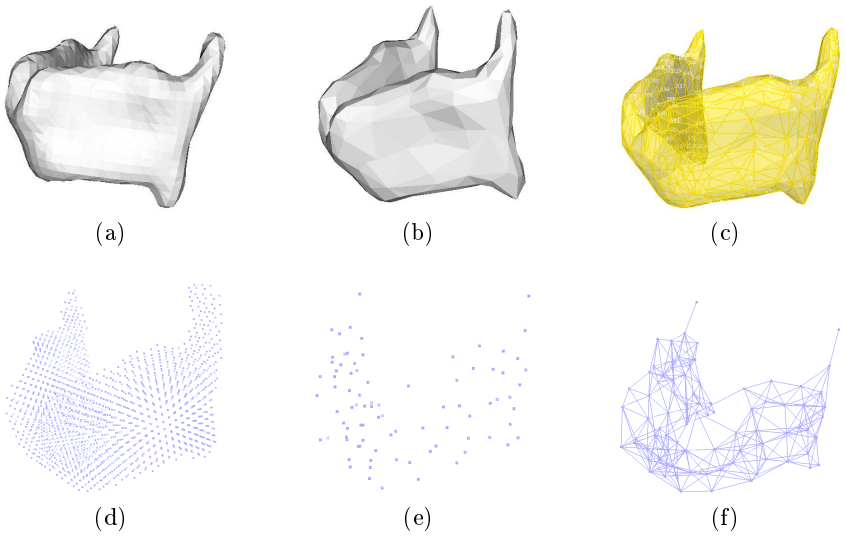


Fig. 1. Stages of model creation (surface submodel: (a)-(c), volumetric submodel: (d) - (f)) for the thyroid cartilage: **(a)** Isosurface, **(b)** Surface simplification, **(c)** Surface submodel, **(d)** Mass point initialization, **(e)** Mass point reduction, **(f)** Volumetric submodel

4.2 Generation of the volumetric model

For the generation of the volumetric model, inner masses with attached *intensity sensors,* dragging these masses towards neighbouring voxels of a certain gray value interval, have to be created and combined with the surface submodel. Thereto, an initial set of mass points is created by placing one mass point at each voxel of the resampled manual segmentation (Fig. 1(d)). Then, the initial point set is reduced iteratively by the following relaxation.

Reduction of inner masses. For each mass point, all mass points inside a neighbourhood of radius r are moved to their common center of mass and merged. By iterating this relaxation, the initial point set is reduced considerably, and fills the manual segmentation evenly (Fig. 1(e)).

The convergence of the relaxation towards a reasonable point set depends on the choice of r. According to our experiments, a radius of half the desired minimum distance of two mass points leads to a convergence representing the original shape well. A dense placement of the volume masses has the advantage that the inner properties of the segmentation target structures are measured at more positions, which is equivalent to a higher sampling rate. We therefore always chose r to be within $[voxelsize; \sqrt{2} \cdot voxelsize]$.

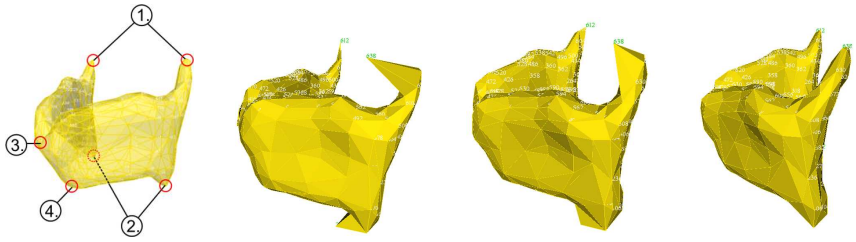


Fig. 2. Elastic model initialization: **(a)** Key masses, **(b)** Key masses placed, **(c)** Model during adaptation to key positions, **(d)** Model after adaptation to key positions

Cross-linking of the inner masses. Each of the resulting volume mass points is linked with each neighbouring mass point within a user-defined radius p. In our model of the thyroid cartilage, a radius of $p = 10mm$ (for a voxel size of 2.148 mm) led to good results in all cases (Fig. 1(f)).

4.3 Connecting inner and outer submodel

Both the volume and the surface submodel are interconnected by springs and merged to one volumetric model. A 1:1 interconnection (each point of the surface submodel is connected to the closest mass point of the volume submodel) has proved to be appropriate for good model stability.

5 Segmentation process

5.1 Elastic, model-consistent initialization

A good initial adjustment of the deformable model to the individual patient's larynx shape is needed, so that the adaptation of the model to the dataset will not be distracted by 'wrong' gray value information of adjacent structures.

For this initialization, the model's position, rotation, scaling and expected shape have to be adjusted for our application. Classical initialization methods correcting only the model's position, rotation and/or scaling are not sufficient according to our tests (see section 6.2 for details).

We therefore introduce a new initialization method, which exploits the model specification already available: In the created deformable model, a mass point at each of the most prominent landmarks is marked as a *key mass* at the end of the model creation process. The user can specify the positions of these key masses by clicking into the dataset. The 6 key masses for the model of the thyroid cartilage are positioned at the *cornu superius left* and *right* (1.), *cornu inferius left* and *right* (2.), as well as the *upper (Adam's apple)* (3.) and *lower* end of the *larynx front side* (4.) (Fig. 2(a)).

These key masses are then fixed and the model simulation is started with only the spring and torsion forces active, but all sensor input turned off[3]. The internal forces, normally representing the model's shape knowledge during a regular segmentation, adapt the model's shape to the key positions (Fig. 2(b)-(d)), while keeping the model as consistent as possible to the shape knowledge it represents. With this method, the model adjusts itself flexibly and nonlinearly to the specified key positions. Guided by the key masses, it adjusts position, size, orientation and shape during this process to the individual patient's anatomy.

After complete adaptation, the rest lengths and rest directions of all springs are set to their current (deformed) length and direction values. This way, the new expected shape is anchored in the model.

5.2 Model adaptation

After initialization, the precise model adaptation is started[4] . All key masses are left fixed, so that the model adaptation occurs within their frame of reference. This way, the model is kept at the correct position. Besides, the lengths of the cornu superius and inferius vary widely among different patients. By keeping the top of the cornu fixed, we can ensure that the whole cornu is found. The simulation is stopped, when the model speed falls below a certain threshold, which always happens because of the damping.

6 Evaluation

6.1 Data material and ground truth

12 CT datasets of the neck were acquired for preoperative planning, containing the larynx. The slice thickness of the datasets ranged from 1.5 mm to 6.0 mm. The datasets varied significantly w.r.t. signal-to-noise ratio, contrast and motion artifacts. In 3 datasets, the larynx was displaced or partially destroyed due to tumor affection. On all 12 datasets, a manual segmentation of the thyroid cartilage was created by an experienced user and controlled by a radiologist. These verified manual segmentations were used as a ground truth for the evaluation.

6.2 Model initialization

To evaluate the single effect of our elastic initialization method from section 5.1, we compare it to the classical initialization methods of positioning and positioning with independent scaling for each axis, where always the optimal initialization is computed with regard to the ground truth. Rotation correction

[3] simulation parameters: spring force weighting $w_f = 5.0$, torsion force weighting $w_t = 10.0$, damping factor $d = 0.001$, simulation time step $\triangle t = 0.05$

[4] simulation parameters: sensor force weighting $w_s = 0.05$ for gradient sensors, $w_s = 0.001$ for intensity sensors, spring force weighting $w_f = 1.0$, torsion force weighting $w_t = 2.0$, damping factor $d = 0.001$, simulation time step $\triangle t = 0.05$

Evaluation Measure	Position	Position / Scale	Elastic
Hausdorff Distance	30.54 mm	29.16 mm	20.06 mm
Average Distance	4.73 mm	4.09 mm	2.90 mm

Table 1. Average initialization results of the standard initialization methods (position and position / scale) and our elastic, model-constistend initialization method compared to the ground truth, measured for 11 CT datasets of the neck using a model created on the 12th dataset (leave-one-out-test)

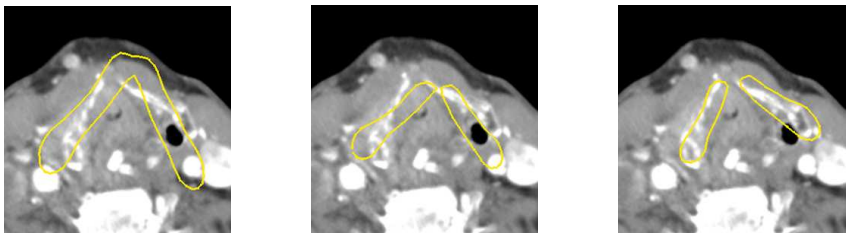
did not make sense here, since the datasets have because of the same imaging process all the same principal direction.

In the optimal case, a "perfect" model initialization would be equal to the ground truth. We therefore calculated the shape (border) distances (Hausdorff and average distance) of both classical initialization methods and our elastic, model-constistent initialization method to the ground truth for evaluation. Table 1 shows, that the new elastic, model-based initialization technique places the model roughly 30 % - 40 % closer to the ideal segmentation result than the classical initialization methods. This is an important improvement for every segmentation method using a local search technique, such as the SMSM approach used for our application.

Furthermore, these numbers show, that the elastically initialized model's shape approximates already the individual shape of the segmentation target structure for a single patient. Otherwise, the shape distances from table 1 would not be so much lower than the ones from the classic initialization methods, which optimally match the ground truth using models without deformation allowed, but only scaling (Fig. 3 illustrates this fact).

6.3 Segmentation experiments

From a dataset with an average-shaped larynx, a volumetric model (consisting of a surface and a volumetric submodel) and a pure surface model were generated



(a) (b) (c)

Fig. 3. Enhancement by elastic model initialization: **(a)** Manual initialization using only translation, **(b)** Manual initialization using translation and scaling, **(c)** Elastic initialization with 6 landmarks

Evaluation Measure	Volumetric model	Surface model
Hausdorff Distance	11.11 mm	9.84 mm
Average Distance	1.20 mm	1.06 mm

Table 2. Average segmentation results of the 2 models and 1 experienced user compared with the ground truth, measured for 11 CT datasets of the neck

as described above (section 3). The key masses were marked manually and used throughout all experiments. The two models were then applied to the remaining 11 datasets (leave-one-out-test) in the following manner:

1. The user marks the key positions in the dataset (6 markers).
2. The model is automatically positioned and scaled according to the bounding box of the key positions.
3. The model is adapted to the key positions (1 click for stopping this phase).
4. The newly adapted shape is automatically learned by the model.
5. The model adaptation to the dataset is performed with the key masses still fixed. (1 click for model stopping).

The segmentation results for both models, as well as the manual segmentation results of an experienced user were compared with the given ground truth by different evaluation measures (Tab. 2).

6.4 Results

In all 11 datasets, the thyroid cartilage could be robustly segmented with an average border distance of 1.064 mm to the ground truth. No significant loss of segmentation quality could be found in the cases of pathological larynx shapes (Fig. 4(a)). Weak-signal holes in the cartilage were successfully bridged by the model's intrinsic shape knowledge (Fig. 4(b)). The model adaptation time needed for elastic initialization was 0.5–1.5 minutes for all datasets, the model adaptation to the datasets took 2–4 minutes per dataset (all measures performed on a standard PC: Pentium M, 1,7 GHz, 512 MB RAM).

Our results show, that the volumetric model is not superior to the pure surface model. In fact, with intensity sensor weighting $w_s > 0.001$, the intensity sensors tend to be attracted by false gray values in neighbouring structures. This leads to strong model instability and significantly worse segmentation results. We therefore recommend using a pure surface model for the segmentation of the thyroid cartilage, which will be less affected by gray value inhomogenities.

The model's adaptation to the datasets is significantly better in the lower part of the thyroid cartilage (i.e. below the adam's apple) than near the upper border, which may cause up to 50 % of the observed undersegmentation. This can be attributed to several circumstances:

– The lower part of the thyroid cartilage tends to be signal-intensive, while the model often lacks signal support in the upper part. This may lead to

strong undersegmentation of this area (Fig. 4(d)). This causes the relative high values the Hausdorff distance. A simple user interaction might prevent this behaviour.

– In some cases where the hyoid bone is very close to the thyroid cartilage, the upper sensors of the model are attracted by its strong signal. In order to prevent leaking, the os hyoideum may be subtracted from the dataset first. Another possibility is to integrate the hyoid bone into the model to ensure that the masses for the thyroid cartilage are kept at appropriate distance.

Except for the hyoid bone, no other neighbouring structures distracted the model. This must be attributed to the new method for initialization followed by shape learning. Without these techniques, the thyroid cartilage could not be separated from the thyroid gland, blood vessels and the trachea robustly.

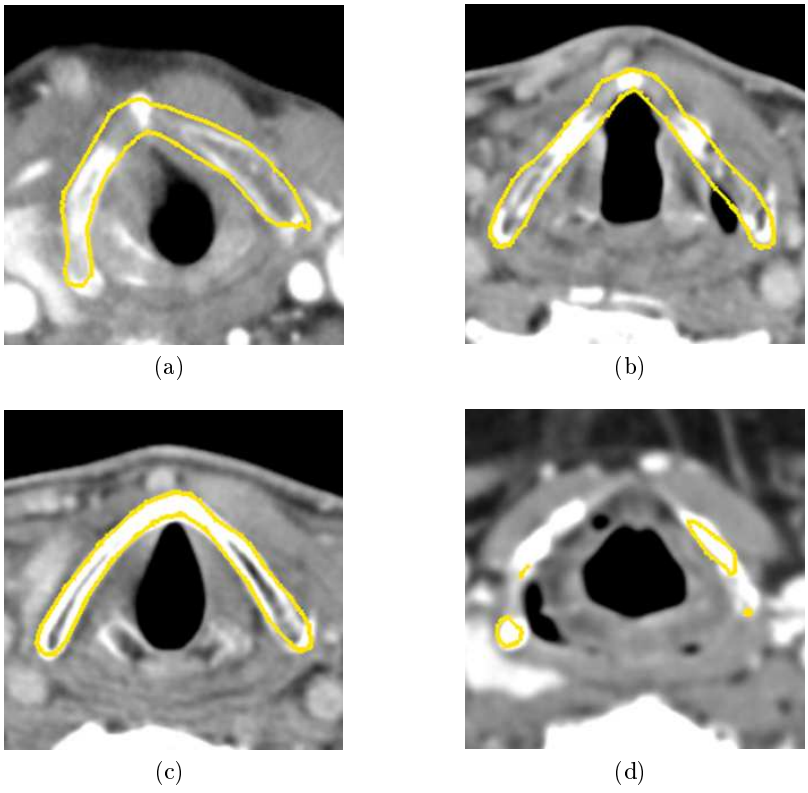


Fig. 4. Results of the model adaptation to the datasets.

7 Discussion

A deformable 3D model (Stable Mass-Spring Model) has been constructed and adapted for the segmentation of the inhomogeneous and complex-shaped thyroid cartilage. We introduced model-consistent (position, orientation, rotation and shape) adaptability of the model to individual patient shape variations by means of an elastic initialization. In contrast to statistical shape models however, our method is not limited to a pre-learned range of shape variations. Instead, it is always initialized to represent the shape information it needs by means of a few key masses. This makes it especially suited for segmenting pathological structures.

Compared with a manual or LiveWire segmentation, the model offers a drastic reduction of interaction effort. Already now, the model can be used at least as a presegmentation of the cartilage, which needs only be corrected at 2–3 positions by the user. In contrast to other 3D models, such as implicit models and ASMs, interaction is intuitively supported by the explicit shape representation of our model.

References

1. Myers, E.N., ed. In: Operative Otolaryngology: Head and Neck Surgery. Volume 1. W. B. Saunders Company (1997) 403–443
2. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: A survey. Medical Image Analysis **1** (1996) 91–108
3. Cohen, I., Cohen, L.D., Ayache, N.: Using deformable surfaces to segment 3d images and infer differential structures. CVGIP: Image Understanding **56** (1992) 242–263
4. Bardinet, E., Cohen, L.D., Ayache, N.: A parametric deformable model to fit unstructured 3D data. CVIU **71** (1998) 39–54
5. Cootes, T., Edwards, G., Taylor, C.: Comparing active shape models with active appearance models. In: BMVC. (1999) 173–182
6. Dornheim, L., Tönnies, K.D., Dornheim, J.: Stable dynamic 3D shape models. In: ICIP. (2005) III–1276–1279
7. Dornheim, L., Tönnies, K.D., Dixon, K.: Automatic segmentation of the left ventricle in 3D SPECT data by registration with a dynamic anatomic model. In: MICCAI. (2005) 335–342
8. Dornheim, L., Dornheim, J., Tönnies, K.D.: Automatic generation of dynamic 3d models for medical segmentation tasks. In: SPIE: Medical Imaging. (2006)
9. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: SIGGRAPH. (1997) 209–216
10. Dornheim, L., Dornheim, J., Seim, H., Tönnies, K.D.: Aktive Sensoren: Kontextbasierte Filterung von Merkmalen zur modellbasierten Segmentierung. In: Bildverarbeitung für die Medizin. (2006)