FULLY AUTOMATIC SHAPE CONSTRAINED MANDIBLE SEGMENTATION FROM CONE-BEAM CT DATA

Sebastian T. Gollmer, Thorsten M. Buzug

Institute of Medical Engineering, University of Lübeck,
Ratzeburger Allee 160, 23562 Lübeck, Germany

ABSTRACT

Cone-beam CT images are useful in operative dentistry but suffer from a comparatively bad image quality with regard to the signal-to-noise ratio. Therefore, we use a statistical shape model (SSM) for robust segmentation of the mandible. In contrast to previous approaches, our method (i) is fully automatic in terms of both, the establishment of correspondence and the segmentation itself, and (ii) allows for leaving the learned principal subspace. By this means, we attain a segmentation accuracy equal to the current reference work on SSM based mandible segmentation whereas our training population is 3.5 times smaller. An important reason therefor is the establishment of correspondence by optimizing a model-based cost function. Our results indicate that SSMs with optimized correspondence can help to improve segmentation accuracy compared to alternative approaches, thus accounting for the first time for the importance of correspondence optimization in an application for image segmentation.

***Index Terms*—** Statistical shape model, correspondence, automatic segmentation, shape prior, mandible

1. INTRODUCTION

Dental cone-beam CT (CBCT) devices are more and more used for generating overview scans prior to dental surgeries. The planning of those interventions on the basis of a patient individual model of bone anatomy requires an appropriate segmentation in the first place. At this, only automatic methods are feasibly for time and financial reasons. However, a poor signal-to-noise ratio and the frequent occurrence of artifacts, which are inherent problems of low-dose scanners, makes fully automatic segmentation a challenging task. Moreover, the field of view (FOV) is restricted to a volume tightly enclosing the jaw and the image quality becomes even worse at its fringe. Therefore, the bone anatomy (e.g. condyles) is often hardly discernible in this very area.

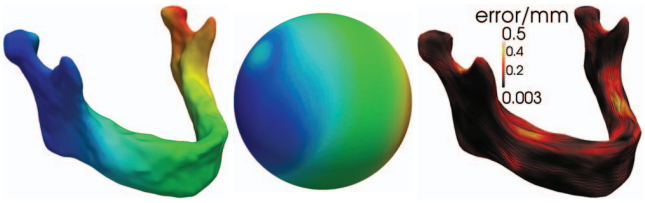
The aforementioned problems demand for segmentation strategies that are not based solely on the underlying image information but also do incorporate prior knowledge about the shape to be segmented. Statistical shape models (SSMs) integrate the to be expected shape variability of a particular object class [1], thus providing means for separating adjacent structures with similar intensity values, increasing the robustness against noise as well as for extrapolating the object's shape in case of occluded image information.

The adaptation of the SSM to the object of interest in the image is done based on local image features. At this, elementary prerequisites are the knowledge of the model points at which should be searched for local features and the initial global position of the SSM in the image. Previous works on SSM based mandible segmentation require manual interaction in at least one of these components. In [2] correspondences are determined semi-automatically in 2D cross-sectional images. Kainmueller et al. [3] establish correspondence in 3D, but do as well need to resort to manual interaction. In contrast, we determine correspondence fully automatically by optimizing a model-based cost function, the current quasi-standard for establishing correspondence (cf. [4]). Yet another aspect that has not been considered so far in mandible segmentation is to allow deviations from the principal shape prior since it acts often too restrictive on the allowed deformations, thus limiting the attainable segmentation accuracy. Our approach follows closely the one proposed in [5]. However, we propose an energy function that (i) is less restrictive and (ii) features an elegant balancing between shape constraint and image adaption that is highly intuitively interpretable.

2. STATISTICAL SHAPE MODELS

Let $\{S_i \subset \mathbb{R}^3,\ i = 1, \ldots, n_s\}$ be a set of $n_s \in \mathbb{N}^+$ training shapes. By sampling those at corresponding positions using $n_p \in \mathbb{N}^+$ landmarks we obtain the shape vectors $\mathbf{x}_i \in \mathbb{R}^{3n_p}$. These are aligned in a common, normalized coordinate frame, in order to calculate the mean shape $\bar{\mathbf{x}} = {n_s}^{-1} \sum_{i=1}^{n_s} \mathbf{x}_i$ as well as the $n_m \in \mathbb{N}$ eigenvalues $\{\lambda_m,\ m = 1, \ldots, n_m\}$ and eigenvectors $\mathbf{P} = (\mathbf{p}_1 \ldots \mathbf{p}_{n_m})$ of the shape vectors' covariance matrix. At this, $\mathbf{P}$ forms the principal subspace (or feature space) $F \subset \mathbb{R}^{3n_p}$ and the linear shape model represents the shapes $\{\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \mid \mathbf{b} \in \mathbb{R}^{n_m}, b^{(m)} \in \left[-3\sqrt{\lambda_m}, 3\sqrt{\lambda_m}\right]\}$.

Send correspondence to {gollmer,buzug}@imt.uni-luebeck.de



(a) (b) (c)

Fig. 1. Exemplary shape S_i (a), its spherical parameterization $\boldsymbol{\omega}_i$ (b) (same colors identify correspondence between S_i and $\boldsymbol{\omega}_i$) and by means of equidistant remeshing of $\boldsymbol{\omega}_i$ reconstructed shape vector $\mathbf{x}_i$ (c). Here, colors indicate the reconstruction error compared to S_i. The mean reconstruction error of the training shapes $\{S_i\}$ is $0.10 \pm 0.01\,\mathrm{mm}$.

2.1. Establishing Correspondence

The main challenge in SSM construction is the establishment of corresponding points. A common approach is to map the shapes $\{S_i\}$ into an appropriate parameter domain and sample all parameterizations equidistantly using n_p landmarks. In this work we, in contrast to [3], opt for the unit-2-sphere $\mathbb{S}^2$ as parameter domain (Fig. 1). More important, given such an $\mathbb{S}^2$-parameterization, there exist well-known algorithms for establishing improved correspondence by iteratively optimizing a model-based cost function (cf. [4]).

We use the *distmin*-approach [6] in order to generate initial parameterizations $\{\boldsymbol{\omega}_i : \mathbb{S}^2 \rightarrow S_i,\ i = 1, \ldots, n_s\}$. This is motivated by the fact that, as it has been shown recently [6], a trade-off between the minimization of the distortions of a single parameterization $\boldsymbol{\omega}_i$ on the one hand (Fig. 1(c)) and a highly consistent alignment of all parameterizations $\{\boldsymbol{\omega}_i\}$ on the other hand, is especially suited for subsequent correspondence optimization. This is done based on the algorithm in [7], but using another cost function: We replace the *simplified MDL* objective by the *detcov* cost function [8], which performed superior in our experiments. However, in contrast to [8] we do not set the regularization parameter ϵ_{detcov} to a predefined small value, but instead calculate it as in [7] depending on the mean radius r_m of the training shapes $\{S_i\}$. That is, $\epsilon_{\text{detcov}} = (\sigma_n / r_m)^2$, where $\sigma_n = 0.3$ is the expected standard deviation of the pixel noise (here: $r_m = 48\,\mathrm{mm}$).

3. SEGMENTATION WITH SHAPE PRIOR

3.1. Initialization

Our automatic procedure for localizing the mandible is based on the analysis of the distribution of the intensity values and their gradients in the image volume $I(\boldsymbol{v}), \boldsymbol{v} \in \Omega \subset \mathbb{R}^3$, making a costly search method as proposed in [3] obsolete. We start by defining a region of interest (ROI) Ω^*, which contains only the body of the mandible but not its rami. Thereby, we take advantage of the fact that the set of voxels that contribute to the teeth, $V_t = \{I(\boldsymbol{v}) \mid I(\boldsymbol{v}) \in [t_{\text{min}}, t_{\text{max}}], \boldsymbol{v} \in \Omega\}$, can be identified quite robustly on the basis of an appropriate intensity interval $[t_{\text{min}}, t_{\text{max}}] \in \mathbb{Z}^2$. Subsequently, we estimate the extent of the body of the mandible in y- and z-direction by projecting V_t onto the y-z-plane and accumulating the pixels over these two axes. Within the in this way defined ROI Ω^*, all voxels are classified according to their intensity values and their gradient, in order to identify those that contribute to the cortical bone of the mandible, $V_c = \{I(\boldsymbol{v}) \mid I(\boldsymbol{v}) \in C \wedge \partial I(\boldsymbol{v})/\partial \boldsymbol{v} \geq g_{\text{max}}/10,\ \boldsymbol{v} \in \Omega^*\}$. Here, $C = [c_{\text{min}}, c_{\text{max}}] \in \mathbb{Z}^2$ is an appropriate intensity interval and $g_{\text{max}} \in \mathbb{R}$ the maximum intensity gradient, which is calculated using the Sobel operator. Finally, we perform a dilation on the binary volume that contains V_c. This yields a connected representation of the cortical bone, which is randomly sampled using approx. 10,000 points. The resulting point set is rigidly registered with the in the center of Ω positioned $\bar{\mathbf{x}}$ using the Iterative Closest Point algorithm [9].

3.2. Image Features

Let $\mathbf{y} \in \mathbb{R}^{3n_p}$ be an approximation of the sought segmentation that has been projected into the tangent space at $\bar{\mathbf{x}}$, $\mathbf{Y}$ be the representation of $\mathbf{y}$ in Ω, and $\mathbf{x}$ be the closest valid shape instance. We are interested in finding a segmentation $\mathbf{Y}$ for which each landmark position $\mathbf{Y}^{(j)}, j = 1, \ldots, n_p$ best matches the local image gradient. To this end, we analyze 1-D intensity profiles by sampling the image volume along the outwards pointing surface normal $\mathbf{n}_j$ at positions $\mathbf{a}_j^{(k)}$, $\{\mathbf{a}_j^{(k)} = \mathbf{Y}^{(j)} + k\,\Delta\,\mathbf{n}_j,\ k = -n_l, \ldots, n_l,\ n_l \in \mathbb{N}^+\}$, with distance Δ. We perform a directional smoothing along $\mathbf{n}_j$ using a Gaussian kernel which results in the intensity values $I'(\mathbf{a}_j^{(k)})$, calculate the directional gradient $g_j(\mathbf{a}_j^{(k)})$ using finite differences, and a deformation vector field $\mathbf{u} \in \mathbb{R}^{3n_p}$ with

$$\mathbf{u}^{(j)} = \begin{cases} \arg\min g_j(\mathbf{a}_j^{(k)}) & \text{if } I'(\mathbf{a}_j^{(k)}) \in C, g_j(\mathbf{a}_j^{(k)}) < g_{\text{min}} \\ \mathbf{0} & \text{else.} \end{cases}$$

Here, g_{min} is a gradient threshold and $\mathbf{0} = [0,0,0]^T$. The fitness value $w_j \in [0,1]$ equals 1 if $\mathbf{u}^{(j)} \neq \mathbf{0}$ and 0 otherwise.

3.3. Model Adaption

The standard approach for determining the pose parameters $\mathbf{T}$ and the shape parameters $\mathbf{b}$ is by iteratively minimizing $\mathcal{L} = \sum_{j=1}^{n_p} ||\hat{\mathbf{Y}}^{(j)} - \mathbf{T}\left(\bar{\mathbf{x}}^{(j)} + (\mathbf{P}\mathbf{b})^{(j)}\right)||_2^2$ [1]. In each iteration we firstly calculate $\mathbf{T}$ [10], thereby ignoring feature positions $\hat{\mathbf{Y}}^{(j)}$ with $w_j < 1$ and penalizing those with distance $(d_j)|_{w_j=1} = ||\hat{\mathbf{Y}}^{(j)} - \mathbf{T}\mathbf{y}^{(j)}||_2$ greater than $1/n_p \sum_{j=1, w_j=1}^{n_p} d_j$ with $0.5\, d_j^{-2}$. Subsequently we determine the shape parameters of $\mathbf{y}$ as in [1].

The obtained segmentation $\mathbf{y}$ is inherently constrained to lie within the principal subspace F. This confines the achievable segmentation accuracy in case of small training

populations and/or large shape variability, as it occurs frequently in medical applications. The *distance from feature space* (DFFS) concept [11] on the other hand provides means for quantifying deviations from F, and was recently successfully applied in the context of SSM based segmentation [5]. The principal idea is to find the segmentation $\mathbf{y}$ that minimizes the (weighted) sum of a shape energy E_{shape} and an image energy E_{image}. We calculate E_{image} similar to [5] as $E_{\text{image}}(\mathbf{y}) = \sum_{j=1}^{n_p} ||\mathbf{y}^{(j)} - \hat{\mathbf{y}}^{(j)}||^2$, whereas $\hat{\mathbf{y}}$ is the projection of the deformed shape $\hat{\mathbf{Y}}$ into the tangent space at $\bar{\mathbf{x}}$. However, our shape energy is the DFFS itself. That is,

$$E_{\text{shape}}(\mathbf{y}) = \sum_{j=1}^{n_p} ||\mathbf{y}^{(j)} - \bar{\mathbf{x}}^{(j)}||^2 - ||\mathbf{b}||^2, \qquad (1)$$

whereas the shape energy in [5] combines DFFS and DIFS (*distance in feature space*. Minimizing Eq. (1) should provide more flexibility in terms of the allowed deviation from the learned shape prior since it drives $\mathbf{y}$ towards the nearest plausible shape $\mathbf{x} \in F$ whereas E_{shape} in [5] is designed to also keep the deviation from $\bar{\mathbf{x}}$ small. Moreover, the proposed E_{shape} has the same characteristic as E_{image}, namely that of an Euclidean distance in $\mathbb{R}^{3n_p}$, whereas E_{shape} in [5] is basically the sum of a Mahalanobis and an Euclidean distance, thus demanding for carefully balancing energies of different magnitude. Accordingly we propose to solve

$$\arg\min_{\mathbf{y}} \left(\alpha E_{\text{image}}(\mathbf{y}) + (1-\alpha) E_{\text{shape}}(\mathbf{y})\right) \qquad (2)$$

in order to allow deviations from the principal shape prior. At this, we want to point out that the parameter $\alpha \in [0,1]$ that balances E_{image} and E_{shape} has a quite intuitive interpretation. We also want to mention that a similar E_{shape} was employed in [12] in the context of 2D segmentation using levelsets and kernel PCA. The minimization problem is solved using the L-BFGS method [13] of the vxl-library (`http://vxl.sourceforge.net/`).

4. RESULTS

The training population consists of 30 shapes (13/17 female/male, age: 18-72), which have been segmented semi-automatically from conventional CT data thereby neglecting the teeth region. These are used for building a SSM as detailed in Sec. 2.1 which we denote as *detcov*-SSM and also for building an SSM using the *distmin* [6] and the *spharm* [14] method. All models are built using $n_p = 4002$ landmarks and n_m is selected individually for each SSM such that it contains 98 % of the observed shape variability. We use different SSM since we also want to shed light onto the question whether (i) optimized correspondence, which typically come along with better generalization and specificity values [6, 15], perform better in an application for image segmentation, and (ii) in how far the DFFS concept depends on the correspondence.

We compared our results with manually generated reference segmentations using the average symmetric surface distance (ASD), the root mean square symmetric surface distance (RMSD), the maximum symmetric surface distance (MSD), and the volumetric overlap error (VOE). We also performed two-sample T^2 tests to test for differences between different SSM and shape priors, with the null hypothesis that those perform in average equally well. Qualitative and quantitative results are given in Fig. 2, and Tab. 1, respectively.

Table 1. Deviation from the reference segmentation after automatic initialization and after segmentation using different SSMs with standard shape prior (middle) and with DFFS shape prior (bottom). Values indicate mean $\pm$ std. dev.

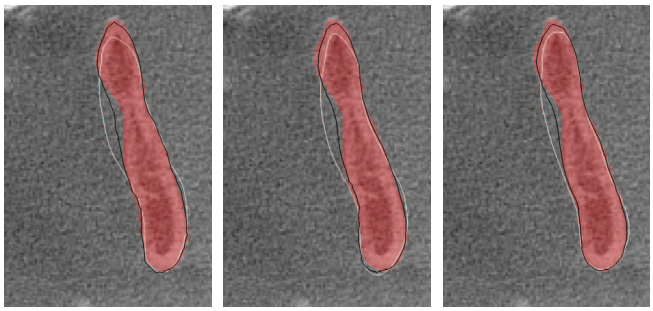
	ASD/mm	RMSD/mm	MSD/mm	VOE/%
initial	2.0 ± 0.4	2.7 ± 0.5	13.2 ± 2.6	47.5 ± 8.7
distmin	0.9 ± 0.2	1.4 ± 0.3	$\mathbf{9.6} \pm 2.5$	24.3 ± 6.4
spharm	0.9 ± 0.1	1.4 ± 0.2	10.9 ± 3.7	23.3 ± 4.4
detcov	0.8 ± 0.1	1.2 ± 0.2	10.0 ± 3.2	20.9 ± 3.9
distmin	0.6 ± 0.2	1.2 ± 0.3	10.4 ± 2.6	16.5 ± 5.6
spharm	0.6 ± 0.2	1.2 ± 0.2	11.0 ± 3.1	15.6 ± 4.5
detcov	$\mathbf{0.5} \pm 0.0$	$\mathbf{1.1} \pm 0.2$	11.3 ± 3.5	$\mathbf{14.0} \pm 1.9$

We evaluated our automatic segmentation on six dental CBCT images, whereas $t_{\text{min}}/t_{\text{max}} = 1500/2700$, $\Delta = 1.0/0.5/0.25\,\text{mm}$ for the 3 different resolutions of our image pyramid, $n_l = 4$, $c_{\text{min}}/c_{\text{max}} = 400/1200$, $g_{\text{min}} = \frac{-100}{\text{mm}}$. For the standard shape prior we used 20/30/40 on the different resolution levels. The segmentation with DFFS prior is initialized using the standard prior with 20/15 iterations on the 2nd/1st resolution. Then, we solve Eq. (2) using 2/4 iterations on the 1st/0th level and allowing for at most 20 function evaluations per iteration, whereas convergence is usually achieved after less than 10 iterations. We tested different values for α and use the best-performing $\alpha = 0.7$ throughout this work.

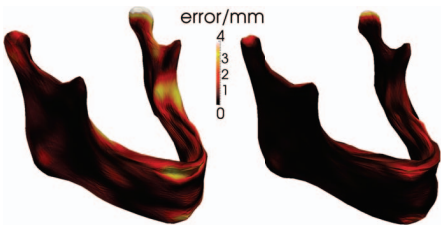
5. DISCUSSION AND CONCLUSION

The figures in Tab. 1 indicate that our algorithm provides feasible segmentation results. As for the initialization (Sec. 3.1) we have to admit that only the mandible body is considered. However, this very subpart allows to localize the mandible bone with an accuracy that is sufficient for our purposes (cf. Tab. 1). Moreover, we observed by experiment that more distinct local errors that may persist at the other parts of the bone like the rami or condyles are corrected by the image feature based pose estimation during the model adaption (Sec. 3.3).

Concerning the segmentation results using different SSMs and shape priors we make the following observations (cf. Tab. 1 and Fig. 2): Firstly, the *detcov*-SSM performed better in our experiments than the SSMs without optimized correspondence. This holds for both, the standard and the DFFS shape prior, indicating that correspondence optimization can help to improve results of SSM based segmentation applications. However, we have to admit that our test population is too small to prove significant superiority of the *detcov*-SSM



(a) *distmin*-SSM (b) *spharm*-SSM (c) *detcov*-SSM



(d) Segmentation for one data set using the *detcov*-SSM with standard (left) and DFFS shape prior. Colors indicate the distance to the reference.

Fig. 2. Results for different SSM and shape priors. In (a)-(c) results for the standard/DFFS shape prior are shown in white/black and the reference segmentation is tagged red.

at the 5 % significance interval. Secondly, using the DFFS shape prior the ASD/VOE can be improved by up to 38/33 % compared to the standard shape prior. This improvement can clearly be seen in Fig 2 and is of statistical significance for all three SSMs (p-values between 0.003 and 0.047).

On the other hand we must acknowledge that the MSD is slightly worse in case the DFFS prior is used. Clearly, this is due to the permitted deviation from the learned principal subspace that is more prone to outliers as they occur frequently in the noisy low-dose images. A more intelligent image feature selection might help to eliminate these outliers. However, our intensity intervals were intentionally chosen relatively liberal and less specific than in [3] to facilitate the segmentation of images from different CBCT devices.

Despite this limitation we want to point out that we have devised an approach for SSM based mandible segmentation which, in contrast to previous works [2, 3], (i) does not demand for any manual interaction—neither for establishing correspondence nor for localization or segmentation, and (ii) allows for leaving the learned principal subspace in an elegant and mathematically sound way. By this means we achieve a mean / rms / max. segmentation-to-reference distance of 0.5 ± 0.1 / 0.9 ± 0.3 / $6.4 \pm 1.9\,\mathrm{mm}$, which is considerably smaller than the values reported in [2] and equal to that in [3], but our training population is 3.5 times smaller. At this it should also be noted that our experimentally confirmed selection of α allows considerable deviations from the principal subspace F of our moderately sized training population. In our opinion this is a reasonable choice that bears potential for further improvement if used in combination with learned intensity/gradient profiles. Our next steps will be into that direction. Moreover, we plan to extend our training population and the number of test cases in order to strengthen our results.

6. REFERENCES

[1] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active Shape Models – Their Training and Application," *CVIU*, vol. 61, no. 1, pp. 38–59, 1995.

[2] S. Rueda, J.A. Gil, R. Pichery, and M. Alcaniz, "Automatic Segmentation of Jaw Tissues in CT Using Active Appearance Models and Semi-automatic Landmarking," in *MICCAI*, 2006, pp. 167–174.

[3] D. Kainmueller, H. Lamecker, H. Seim, M. Zinser, and S. Zachow, "Automatic Extraction of Mandibular Nerve and Bone from Cone-Beam CT Data," in *MICCAI*, 2009, pp. 76–83.

[4] T. Heimann and H.-P. Meinzer, "Statistical Shape Models for 3D Medical Image Segmentation: A Review," *Med Image Anal*, vol. 13, no. 4, pp. 5431–563, 2009.

[5] M. Kirschner and S. Wesarg, "Active Shape Models Unleashed," in *SPIE 7962*, 2011.

[6] M. Kirschner, S.T. Gollmer, S. Wesarg, and T.M. Buzug, "Optimal Initialization for 3D Correspondence Optimization: An Evaluation Study," in *IPMI*, 2011, pp. 308–319.

[7] T. Heimann, I. Wolf, T. Williams, and H.-P. Meinzer, "3D Active Shape Models Using Gradient Descent Optimization of Description Length," in *IPMI*, 2005, pp. 566–577.

[8] A.C.W. Kotcheff and C.J. Taylor, "Automatic Construction of Eigenshape Models by Direct Optimization," *Med Image Anal*, vol. 2, no. 4, pp. 303–314, 1998.

[9] P.J. Besl and N.D. McKay, "A Method for Registration of 3-D Shapes," *IEEE TPAMI*, vol. 14, no. 2, pp. 239–256, 1992.

[10] B.K.P. Horn, "Closed-form Solution of Absolute Orientation Using Unit Quaternions," *J Opt Soc Am A*, vol. 4, no. 4, pp. 629–642, 1987.

[11] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE TPAMI*, vol. 19, no. 7, pp. 696–710, 1997.

[12] S. Dambreville, Y. Rathi, and A. Tannenbaum, "A Framework for Image Segmentation Using Shape Models and Kernel Space Shape Priors," *IEEE TPAMI*, vol. 30, no. 8, pp. 1385–1399, 2008.

[13] D.C. Liu and J. Nocedal, "On the Limited Memory BFGS Method for Large Scale Optimization," *Math Prog B*, vol. 54, no. 3, pp. 503–528, 1989.

[14] C. Brechbühler, G. Gerig, and O. Kübler, "Parametrization of Closed Surfaces for 3-D Shape Description," *CVIU*, vol. 61, no. 2, pp. 154–170, 1995.

[15] M.A. Styner, K.T. Rajamani, L.-P. Nolte, G. Zsemlye, G. Székely, C.J. Taylor, and R.H. Davies, "Evaluation of 3D Correspondence Methods for Model Building," in *IPMI*, 2003, pp. 63–75.