Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours

Karl D. Fritschera)
Department for Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts 02114

Marta Peroni
Paul Scherrer Institut, Villigen 5232, Switzerland

Paolo Zaffino
Department of Experimental and Clinical Medicine, Magna Graecia University, Catanzaro 88100, Italy

Maria Francesca Spadea
Department of Experimental and Clinical Medicine, Magna Graecia University, Catanzaro 88100, Italy

Rainer Schubert
Institute for Biomedical Image Analysis, Private University of Health Sciences, Medical Informatics and Technology, Hall in Tirol 6060, Austria

Gregory Sharp
Department for Radiation Oncology, Massachusetts General Hospital, Boston, Massachusetts 02114

(Received 9 August 2013; revised 21 February 2014; accepted for publication 5 April 2014; published 24 April 2014)

Purpose: Accurate delineation of organs at risk (OARs) is a precondition for intensity modulated radiation therapy. However, manual delineation of OARs is time consuming and prone to high interobserver variability. Because of image artifacts and low image contrast between different structures, however, the number of available approaches for autosegmentation of structures in the head-neck area is still rather low. In this project, a new approach for automated segmentation of head-neck CT images that combine the robustness of multiatlas-based segmentation with the flexibility of geodesic active contours and the prior knowledge provided by statistical appearance models is presented.

Methods: The presented approach is using an atlas-based segmentation approach in combination with label fusion in order to initialize a segmentation pipeline that is based on using statistical appearance models and geodesic active contours. An anatomically correct approximation of the segmentation result provided by atlas-based segmentation acts as a starting point for an iterative refinement of this approximation. The final segmentation result is based on using model to image registration and geodesic active contours, which are mutually influencing each other.

Results: 18 CT images in combination with manually segmented labels of parotid glands and brainstem were used in a leave-one-out cross validation scheme in order to evaluate the presented approach. For this purpose, 50 different statistical appearance models have been created and used for segmentation. Dice coefficient (DC), mean absolute distance and max. Hausdorff distance between the autosegmentation results and expert segmentations were calculated. An average Dice coefficient of $DC = 0.81$ (right parotid gland), $DC = 0.84$ (left parotid gland), and $DC = 0.86$ (brainstem) could be achieved.

Conclusions: The presented framework provides accurate segmentation results for three important structures in the head neck area. Compared to a segmentation approach based on using multiple atlases in combination with label fusion, the proposed hybrid approach provided more accurate results within a clinically acceptable amount of time. © 2014 American Association of Physicists in Medicine. [http://dx.doi.org/10.1118/1.4871623]

Key words: radiation therapy planning, atlas-based segmentation, geodesic active contours, InShape models

1. INTRODUCTION

In the past two decades intensity modulated radiotherapy (IMRT) became the state of the art method for the treatment of head-neck cancer by improving the sparing of organs at risk (OARs). For an efficient and safe use of IMRT, however, an exact delineation of OARs in 3D CT volumes has to be performed. This task is also referred to as segmentation and is traditionally performed by medical experts following anatomical definitions. According to Harari et al., the segmentation of a single head-neck case takes 2.7 h on average.\(^1\) In addition, manual segmentation is prone to inter- and intraobserver variability,\(^2\) which can directly influence the quality of the treatment plan and especially the dose distribution for OARs.\(^3\)
Consequently, the development of fully automatic segmentation approaches for radiotherapy planning in the head neck area became a very active field of research. However, due to low contrast of soft tissue in head and neck CT images and also due to image artifacts (e.g., caused by dental implants), the number of applicable automatic segmentation approaches is still rather low. Most of the presented approaches are either based on the usage of (multiple) atlases or statistical models of shape and/or appearance (SSM, SAM).

Atlas-based segmentation turned out to be a good choice in a number of segmentation scenarios due to its robustness and its potential to perform segmentation without user interaction. Using a reference (=atlas) image, in which structures of interest are presegmented, the optimal transformation between the atlas image and the image to be segmented has to be identified. By applying this transformation to the segmented structures of interest, a segmentation for all labeled structures in the atlas image can be obtained. When using atlas-based approaches the challenges are twofold: First, a (deformable) registration method that is capable of aligning patient and atlas images accurately—also in the presence of noise—has to be found. For this purpose, different approaches like Demons registration,\textsuperscript{4} block-matching,\textsuperscript{5} and B-Spline registration\textsuperscript{6} have been used. Second, due to physiologic or pathologic anatomical variations of certain organs of interest, establishing exact correspondences between a given atlas image and a new patient CT is not always feasible. Consequently, approaches based on the usage of atlases that reflect an average patient anatomy\textsuperscript{7} or the usage of multiple atlases in combination with different atlas selection strategies\textsuperscript{8, 9} have been proposed. In addition, voting schemes that mark single voxels as being inside or outside of OARs (Refs. 10 and 11) have been introduced. However, approaches based on voxel-wise voting do not necessarily provide closed surfaces, which can result in isolated voxels in the final segmentation result. Such anatomically incorrect labels can affect subsequent treatment planning and require time consuming manual editing. Consequently, hybrid approaches that postprocess the result of the atlas-based segmentation by using active contours\textsuperscript{6} and graph cuts\textsuperscript{12, 13} have been introduced.

In contrast to atlas-based segmentation, approaches based on the usage of statistical shape or appearance models\textsuperscript{14, 15} for automatic delineation typically provide closed and anatomically correct surfaces by restricting the final segmentation results to anatomically plausible shapes described by the statistical model. In a (separate) training phase, characteristic variations of shape or appearance of one or multiple structures of interest are learned by using a training set of images in combination with correct delineations of the structures of interest. As a result, anatomically correct shape or appearance variations of structures of interest can be represented in a compact form by using, e.g., landmarks,\textsuperscript{14, 15} surface meshes,\textsuperscript{16} or vector fields\textsuperscript{17} in combination with statistics on voxel intensities and principal component analysis (PCA).\textsuperscript{18}

In the actual segmentation process, the best fitting model instance for a new patient has to be found for each structure of interest in order to obtain a final segmentation result. A common problem of traditional segmentation approaches based on SAM/SSM, however, is that they are often limited to specific shapes described by the statistical model, which makes them less flexible and highly dependent on large training sets. In addition, they require proper prealignment of the model to the image. In order to overcome this problem, the model often has to be prealigned to the image by an expert user or by using different types of model to image registration approaches. Commonly, the average shape of an organ is used as a starting point for an evolutionary process that provides a final segmentation by identifying the best fitting model instance for a new image. In order to prevent the model-based segmentation approaches to run into local minima, SAMS and SSMS have been combined with atlas-based segmentation\textsuperscript{4} in order to provide a better initialization for the model-based segmentation. Recently, Wachinger \textit{et al.} have presented a hybrid-approach for the segmentation of the parotid glands, which is based on the usage of Gaussian Processes in combination with random forests.\textsuperscript{19}

The purpose of the presented study was to develop a novel hybrid approach for the fully automatic segmentation of structures in the head-neck area that combines multiatlas-based segmentation (MABS) with geodesic active contours\textsuperscript{20} and statistical appearance models. By this means, the strengths of each of the single methods should be efficiently combined in one hybrid approach. In the presented approach, a multiatlas-based segmentation approach is used to define an initial surface. Subsequently, an anatomically correct approximation of this initial surface is calculated. This approximation is used to initialize a segmentation process, in which statistical appearance models and geodesic active contours\textsuperscript{20} mutually guide each other, in order to obtain a final segmentation result. By using this combination of three approaches, the robustness of atlas-based segmentation shall be combined with the flexibility of geodesic active contours and the prior information about anatomically plausible appearances of a structure provided by the appearance models. The model building step, as well as the segmentation itself can be performed without any user interaction. The general work-flow of this segmentation framework is also illustrated in Fig. 1.

The proposed algorithm will be used to segment the brainstem as well as both parotid glands. Both of these structures are highly critical for RT planning and also highly challenging for autosegmentation approaches, since they are susceptible to images noise and partly show very low intensity differences to neighboring organs. In addition, the shape of the parotid gland is highly variable, which is posing a major challenge for the creation of SAMs especially if only a limited number of training data-sets is available. In the remainder of this paper, the technical background of the model creation and segmentation pipelines will be presented (Sec. 2). Thereafter, a detailed description of the performed experiments (Sec. 3) and evaluation of the new hybrid approach (Sec. 4), including a discussion of the results (Sec. 5) will be provided.

2. METHODS

Notational remark: In the following, an upper case letter (e.g., $S$) in the subscript of a variable name (e.g., $L_S$ or $I_S$)
depicts a whole set (of images, labels, or transformations), whereas a lower case letter (e.g., $s$) in the subscript depicts one single instance (of an image, label, or transformation) (e.g., $L_s$ or $I_s$).

2.A. Preprocessing of atlas images

In a preprocessing step, all 18 available atlases (17 for right parotid gland) were pre-aligned to one reference data-set in order to bring all atlas volumes into one common reference system. Using normalized mutual information metric, all data-sets were registered rigidly onto one (randomly chosen) data-set.

2.B. Multiatlas-based segmentation

The MABS approach used in this project is based on the algorithm presented by Peroni. In general, this MABS approach can be divided into a pairwise registration step and a label fusion step, in which a modified version of the method presented by Sabuncu et al. was used.

Peroni also performed an atlas-selection step in order to improve the speed of the multiatlas-based registration. However, atlas selection also leads to a potential decrease in segmentation accuracy. Consequently, no atlas selection was performed in this project.

2.B.1. Pairwise registration

In pairwise registration, each image of the whole set of prealigned atlas images $I_S$ is registered nonrigidly to a test image $I$ using a multiresolution, cubic B-Spline registration approach. In general, B-spline registration methods use cubic B-splines to define a displacement field that maps the voxels in the moving image to those in a reference image (also referred to as fixed image). In combination with a metric that quantifies the similarity between the fixed and moving images, the registration problem can be solved using a gradient descent optimization approach. In order to improve the speed of the registration a highly-efficient multicore implementation of multiresolution B-Spline registration algorithm presented by Sharp et al. was used in this project. In this approach, a grid alignment scheme is used in order to speed up necessary B-Spline interpolation and gradient computation. In this project, B-Spline registration was used in combination with a quasi-Newton optimizer. As proposed by Han et al., a metric based on mutual information was used for pairwise registration in order to deal with potential changes in image contrast and to add robustness in the presence of image artifacts.

2.B.2. Label fusion

After the pairwise alignment of the set of atlas images $I_S$ to a new test image $I$, the resulting set of transformations $\Phi_S$ can be used to deform the set of labels $L_S$ in the set of atlas images $I_S$. By this means $n$ different labelings, corresponding to $n$ potential segmentation results for test image $I$ can be obtained. In order to get one final labeling for each structure of interest, different strategies like averaging or majority voting have been proposed. In this project, the label fusion approach proposed by Peroni, which is based on the method presented by Sabuncu et al. was applied. Assuming image voxels to be independent, the value for $C$, which represents a
score that is used to assign each voxel of the image to either a structure or background [see also Eq. (4)] can be calculated as follows:

\[ C_I(x) = \sum_{s=1}^{n} p(L(x)=l, I(x)|L_s, \Phi_s) p(I(x)|L_s, \Phi_s), \]

where \( p(L(x)=l, I(x)|L_s, \Phi_s) \) is called label likelihood and \( p(I(x)|L_s, \Phi_s) \) is the image likelihood term. \( \Phi_s \) denotes the transformation resulting from pairwise registration described in Sec. 2.B.1 and \( n \) refers to the number of images in \( S \). Potential values for \( l \), which represents the labeling of voxel \( x \), are 0 (=outside label) or 1 (=inside label). The values for the label likelihood are assumed to be proportional to the values of the signed Mauer distance map \( D^s \) of atlas label \( L_s \) for atlas image \( I_s \) for each voxel of the test image, so that

\[ p(L(x)=l, I(x)|L_s, \Phi_s) \propto e^{-D^s(I_s(x))}, \]

where \( \rho \) is a weighting factor with values between [0; 1].

In this project, a truncated quadratic function as proposed by Sharp et al.\(^{27} \) was used in order to formulate the image likelihood, which is based on the intensity differences between voxels of a warped atlas images \( I_s \) and test image \( I \),

\[ p(I(x)|L_s, \Phi_s) = \begin{cases} \sqrt{2\pi \sigma^2} e^{-\frac{1}{2\sigma^2}(I(x)-L_s(\Phi_s(x)))^2} & \text{if } e^{-\frac{1}{2\sigma^2}(I(x)-L_s(\Phi_s(x)))^2} > \rho, \\ 0 & \text{else,} \end{cases} \]

where \( \rho \) is a threshold parameter defining the minimum image similarity. By using a truncated quadratic function for the calculation of the probability, outliers get down weighted compared to using a nontruncated quadratic function.

The final assignment \( \hat{L} \), which determines if a voxel of test image \( I \) is part of a structure \( S \) or not, is calculated as\(^{27} \)

\[ \hat{L}(x) = \begin{cases} 1 & \text{if } \frac{C_I(x)}{C_I(x)+C_0(x)} > \tau \\ 0 & \text{else,} \end{cases} \]

where \( \tau \) is a constant between [0, 1].

### 2.C. Creation of a statistical appearance model

#### 2.C.1. Removing nonrelevant pose variations from data-sets

In order to create statistical appearance models for brain-stem and parotid glands, the InShape approach\(^ {18} \) was used. InShape models are using deformation fields in combination with intensity information to model appearance variations of a structure of interest.\(^ {18} \) For the creation of InShape models, the whole set of training subjects consisting of \( n \) images \( I_s \) and \( n \) labels \( L_s \) (depicting one structure of interest in \( I_s \)) have to be provided. One of these subjects acts as a reference subject \( r \) (consisting of reference image \( I_r \) and corresponding label \( L_r \)). The basis for the creation of InShape models is the computation of an optimal transformation \( T_{G_r} \); \( x \rightarrow x' \), which maps any point \( x \) of reference subject \( r \) onto the corresponding point \( x' \) of a training subject. \( T_s \) is created by concatenating global transformation \( T_G \) and respective local nonrigid transformation \( T_s \). The pose variations of a structure in different data-sets represented by \( T_G \) have to be compensated, in order to assess shape and appearance differences without any bias caused by differences in pose. This can be achieved by globally prealigning all training data-sets to a common reference volume. In this project, \( T_G \) was assessed by performing a rigid prealignment. The computation of \( T_G \) is initialized by aligning the center of mass of all training labels \( L_s \) to the center of mass of \( L_r \). Subsequently, signed distance maps are created for the reference label \( L_r \) and all training labels \( L_s \). Using this representation, all training labels \( L_s \) are aligned to \( L_r \) in a rigid manner based on Mattes Mutual Information metric.\(^ {25} \) The resulting transformations \( T_G \) are applied to \( L_s \) and the respective training images \( I_s \). The remaining differences among the images and labels of the training-set are now assumed to represent the appearance variations that shall be modeled using the InShape approach.

#### 2.C.2. Creation of a stable reference and calculation of deformation fields

Initially, reference subject \( r \) was chosen randomly. Because of this, \( r \) could potentially represent a rare anatomical appearance variation, which would have negative influence on the nonrigid registration and consequently the creation of the InShape model. In order to avoid this bias, the following procedure is applied: First, the signed distance map of the reference label \( L_r \) is registered to each rigidly prealigned training label \( L_s \) using diffeomorphic demons registration.\(^ {28} \) This results in a set of \( n \) nonrigid transformations \( T_s \). Each transformation \( T_s \) is represented as a displacement field and maps each voxel of \( L_r \) to the respective voxel in \( L_s \). Using the set of transformations \( T_s \), an average displacement field \( T_m \) was calculated and used to warp \( L_r \) resulting in a modified reference label \( L_{r_{\text{mod}}} \). Using the updated reference label \( L_{r_{\text{mod}}} \) nonrigid registration and subsequent warping (using newly calculated set \( T_s \)) are repeated until the DICE coefficient between \( L_{r_{\text{mod}}} \) and \( L_{r_{\text{mod}}} \) in two subsequent iterations is higher than 0.99. The final set of displacement fields \( T_s \) is representing shape variations among the training set and are used to create the final InShape representations as described in Sec. 2.C.3.

#### 2.C.3. Creating InShape representation

Since not only shape differences, but also appearance differences represented by varying voxel intensities among the training images shall be described by the InShape model the displacement fields obtained in the previous step have to be combined with intensity variations in one representation. For this purpose, all displacement fields \( T_s \) resulting from the final nonrigid demons registration step (described in Sec. 2.C.2) get inverted using an approach presented by Chen et al.\(^ {29} \) For each training subject the inverted displacement field \( T_{s_{\text{inv}}} \) is used to warp the respective greyscale image \( I_s \). By this means, the structure that shall be modeled gets “shape normalized” in all training images. This means that the structure of interest has the same outer shape (the shape of the stable reference \( L_{s_{\text{mod}}} \)) in the set of warped training images \( I_{s_{\text{warped}}} \), but maintains its individual spatial greyscale distribution.
In a final step, all voxels in $I_{\text{warped}}$ that are not representing the structure of interest are set to zero. By this means, voxels that shall not be used to assess appearance variations are removed from the InShape representation. As a result, the individual appearance of a training subject is represented by the nonzero voxels in $I_{\text{warped}}$ and the respective displacement fields $T_l$. This combined information can be summed up in a vector $\bar{a}$ for a subject $n$ of the training set

$$\bar{a}_n = \begin{bmatrix} \bar{x}, \bar{y}, \bar{z}, \bar{i} \end{bmatrix},$$

where $\bar{x}, \bar{y}, \bar{z}$ are vectors containing $x, y, z$ components of the displacement field of subject $n$ and $\bar{i}$ is the vector containing the intensity values of all nonzero voxel in $I_{\text{warped}}$. This vector $\bar{a}$ is also referred to as InShape representation. The size of $\bar{a}$ is number of pixels for the structure of interest times 4.

### 2.D. Creating a statistical InShape model

After generating an InShape representation for each subject of the training set, a training matrix $M$, where

$$M = [\bar{a}_1 \ldots \bar{a}_n]$$

can be defined and used as input for the principal component analysis (PCA). Since $\bar{x}, \bar{y}, \bar{z}$, and $\bar{i}$ have different units, a standardized version $M_s$ of $M$ is created. In $M_s$, each variable in $\bar{s}$ gets standardized by subtracting its mean and dividing by its standard deviation.

Thereafter, the covariance matrix $\Sigma$ of $M_s$ can be decomposed as

$$\Sigma = U \Lambda U^T$$

using singular value decomposition (SVD), where $\Lambda$ is a diagonal matrix and contains the eigenvalues of $\Sigma$. The columns of $U$ contain the corresponding eigenvectors.

A new instance of an InShape model $a_{\text{new}}$ can be calculated by using

$$a_{\text{new}} = ((\mu + U_k \alpha) \odot \varphi) + \lambda,$$

where $U_k$ represents the first $k$ eigenvectors of $\Sigma$. $\alpha$ is a parameter vector of length $k$, which defines the location of the model instance in PCA space. Because of the usage of $M_s$ instead of $M$ as input for the PCA, $\mu$ is a vector of zeros. $\varphi$ and $\lambda$ are vectors containing the standard deviations and mean values of the variables forming the InShape representations.

The whole process for the creation of an InShape model is summarized in Fig. 2.

### 2.E. Segmenting images using InShape models and geodesic active contours

Apart from using InShape models for diagnostic purposes, they can also be used for segmentation tasks in combination with geodesic active contours (GAC). In contrast to most other segmentation approaches, which are based on the usage of statistical shape or appearance models, this approach is less restrictive, since it is not forcing the final contour to be exactly determined by a model instance. Instead, the final result is also influenced by the geodesic active contours, which are evolving based on image information (in terms of image gradient) in combination with surface curvature and a balloon force. By this means, prior anatomical information provided by statistical appearance models can be combined with the flexibility and accuracy provided by geodesic
active contours. This can be achieved by using the following equation:

\[ u(t + 1) = u(t) + (\omega_1 g (c + \omega_2 \kappa))|\nabla u(t)| \\
+ \gamma(u^*(t) - u(t)), \]  

(9)

where \( u \) is the embedding function, \( g \) is a function of the image gradient, \( \kappa \) is the curvature of the evolving front, \( c \) is a balloon force which forces the surface to evolve outward, and \( \omega_1, \omega_2 \) are weighting factors for \( g \) and \( \kappa \). \( g \) is defined as follows:

\[ g = \frac{1}{1 + e^{-(\frac{\alpha g - \beta}{\gamma})}}, \]

(10)

where \( \alpha = -0.1 \) and \( \beta = 0.2 \) and

\[ I_{GN} = \frac{I_G - \mu_L}{\sigma_L^2}, \]  

(11)

where \( I_G = |\nabla I_T \ast G| \) and

\[ I_T(x) = \begin{cases} 
I(x) & \text{if } t_1 < I(x) < t_2 \\
0 & \text{else,} 
\end{cases} \]  

(12)

where \( t_1 = \mu_L - \sigma_L \) and \( t_2 = \mu_L + \sigma_L \). Thresholds are applied in order to weaken some strong edges (e.g., skin-air interface), which are irrelevant for the segmentation. This has large influence on the calculations of \( g \) [as defined in Eq. (10)] (see also Fig. 3). \( \mu_L \) and \( \sigma_L \) are mean value and standard deviation of all voxels \( x \) of input image \( I \), for which \( \tilde{L}(x) = 1 \) . \( \tilde{L}(x) \) is the final assignment for each voxel of \( I \), which represents the MABS results [see also Eq. (4)]. \( u^* \) represents the estimation of the final contour and corresponds to the set of nonzero voxels of an InShape model instance \( a_{new} \), which can be calculated using Eq. (8). A signed distance map of \( u^* \) can be calculated using a binarized version of \( a_{new} \). The optimization of \( u^* \) is obtained by performing a deformable (InShape) model to image registration task, which aims at finding the best fitting InShape model instance for a given \( u \). This is achieved by optimizing parameter vector \( p \), which is defining the underlying nonrigid transformation:

\[ p = [r_1, r_2, r_3, t_1, t_2, t_3, \alpha_1, \ldots, \alpha_n], \]  

(13)

where \( r_1, r_2, r_3 \) and \( t_1, t_2, t_3 \) define rotation and translation, whereas \( \alpha_1, \ldots, \alpha_n \) refer to the model parameters, which define the appearance of a specific model instance using \( n \) principal components [Eq. (8)]. Normalized mutual information (NMI) (Ref. 22) in combination with a simultaneous perturbation stochastic approximation (SPSA) scheme\(^{32} \) is used to optimize \( u^* \). NMI calculation is restricted to the voxels within the bounding boxes around \( u \) and \( u^* \). Due to the fact that the dimensions of the bounding boxes, which are recalculated after each iteration, changes with \( u \) (and consequently also the voxels used for NMI calculation), \( u^* \) is dependent on \( u \).

As illustrated in Fig. 4, the final segmentation result is not necessarily equivalent to \( u^* \) (as in most other model-based segmentation approaches). Instead, \( u^* \) is used to drive the evolving level sets close to edges that potentially represent the structure of interest and prevents the contour from leaking based on anatomical information provided by the InShape model.

### 2.F. Combining atlas-based segmentation and model-based segmentation approach

As pointed out in Sec. 1, the main objective of this project was to combine atlas-based segmentation, InShape models and GAC in a segmentation pipeline, that is more flexible than purely model-based approaches and provides exact results that are based on anatomically plausible shapes. For this purpose, the result of the atlas-based segmentation was used for the initialization of the InShape segmentation algorithm in order to minimize the risk of being trapped into local

---

**Fig. 3.** Effect of using \( I_T(x) \) [Eq. (12)] for the calculation of \( g \), left: one sample slice of the original CT image, center: resulting \( g \) values for sample slice without using thresholds, right: resulting \( g \) values for sample slice using thresholds as defined in Eq. (12).

**Fig. 4.** Illustrating evolution of \( u \) and \( u^* \) for a sample data-set of the left parotid gland (left: axial sample slice of test image, center: initialization, right: final result: \( u^*(t) \): overlay, \( u(t) \): contour).
minima, which is a common problem of model-based segmentation approaches. The following procedure has been developed to combine MABS with InShape models and geodesic active contours:

First, the surface \( S_0 \) resulting from atlas-based registration, the new test image \( I_{\text{new}} \) and the InShape model have to be brought to a common reference system. For the alignment, the signed distance map representation of \( S_0 \) gets registered to the signed distance representation of the reference label \( L_r \) (see also Sec. 2.C.2) using a rigid transformation in combination with normalized mutual information metric\(^{22}\) resulting in a rigid transformation \( T_{\text{fin}} \). As a result, the transformed version of \( S_0 \) can be used as initial surface for the InShape segmentation step and \( T_{\text{fin}} \) provides the initial values for \( r_1, r_2, r_3 \) and \( l_1, l_2, l_3 \) in parameter vector \( p \) [Eq. (13)].

Frequently, the values for \( \alpha_1, \ldots, \alpha_n \) of parameter vector \( p \) are set to zero when starting the level-set segmentation process so that \( u^* \) and \( u \) would both be initialized using the average shape/appearance of the structure to be segmented. In this project, however, \( S_0 \) should be utilized to initialize the level-set segmentation. Whereas the signed distance map (level set representation) of \( S_0 \) can directly be used to initialize \( u \) in this case, the following approach was applied in order to obtain the corresponding model parameter vector \( \alpha \). After calculating \( T_{\text{fin}} \), a new InShape representation \( R_{\text{fin}} \) for each structure to be segmented is created by using \( S_0 \) in combination with \( I_{\text{new}} \). For this purpose, the same steps as described in Sec. 2.C.3 are applied in order to calculate \( R_{\text{fin}} \). Thereafter, initial values for \( \alpha \) can be obtained by projecting \( R_{\text{fin}} \) into the InShape model space following the equation:

\[
\alpha = U_k^T (R_{\text{mod}} - \mu), \tag{14}
\]

where \( U_k^T \) and \( \mu \) are equivalent to Eq. (8) and \( R_{\text{mod}} = (R_{\text{fin}} - \lambda) \odot \varphi \) (\( \odot \) denotes componentwise division) as pointed out in Sec. 2.D.

Using this approach, \( S_0 \) can directly be used as the initial surface for the geodesic active contour in combination with the anatomically correct approximation of \( S_0 \) defined by model parameter vector \( \alpha \). By this means, weighting factor \( \gamma \) does not have to be decreased during initialization and the full amount of information provided by the InShape model as well as the full potential of the geodesic active contours can be used. Especially in situations where a good estimate for \( u^* \) can be provided (in this case by MABS), this initialization approach is more stable and reliable compared to using the average InShape instance to initialize the segmentation. This is especially true for cases, where \( u \), as defined by the MABS result, initially is very different from the average shape of the structure to be segmented.

After this initialization step, the same approach as described in Sec. 2.F is used to perform InShape segmentation.

3. EXPERIMENTS

3.A. Data

For this project 18 CT images with a spatial resolution between approx. 0.5 \( \times \) 0.5 \( \times \) 2.5 mm and approx. 1.2 \( \times \) 1.2 \( \times \) 2.5 mm have been used. The images have been acquired at the Massachusetts General Hospital, Boston, MA using a General Electric RT16 CT scanner. Scans were acquired in helical mode at 140 kVp with 2 cm collimation. Automatic current modulation was used for dose reduction. In all 18 CT images, left parotid gland and the brainstem were segmented manually by a medical expert. The right parotid gland was available in 17 CT images since it was consumed by a tumor in one of the images. The guidelines used for segmentation of the parotid glands are equivalent to the guidelines provided by Water et al. in Ref. 33. The superior border of the posterior clinoind was used to define the superior border of the brainstem.

In order to define the inferior border of the brainstem the top of the first cervical vertebra C1 was used.

3.B. InShape model creation

In a first step, InShape models for the brainstem, left and right parotid glands have been created separately as described in Sec. 2.C. For each of the three structures, one data-set was chosen randomly as an initial Ref. 3 iterations of deformable demons registration were required in order to create a stable reference label for all structures as pointed out in Sec. 2.C.2. This reference label was used to create InShape representations for all subjects. For evaluation purposes, a leave-one-out cross validation approach has been used. So instead of splitting the data only in test- and training-data, a single data-set was used as the validation data, and the remaining data-sets as the training data. This is repeated such that each of the data-sets is used once as the validation data-set. By this means, the number of training sets, as well as the validity of the validation can be maximized. In order to avoid any bias, the data-set that was used as a reference for the model creation was excluded from the tests. Consequently, 50 different InShape models had to be calculated [17 for left parotid (LP) and brainstem (BS) and 16 for the right parotid (RP)]. Figure 5 shows the appearance variations described by the first three principal components for BS and LP of one sample InShape model.

3.C. Training multiatlas-based segmentation

In order to evaluate potential improvements of the segmentation results provided by the new hybrid approach compared to MABS in a fair manner, different configurations for MABS were tested. For this purpose, three different parameter configurations for the multilevel B-spline registration have been tested. The different settings are summed up in Table I. The main differences among the three settings are the number of registration levels and the parameter restricting the maximum number of registration iterations (which was only used in setting 1). Moreover, four different values for \( \tau \) (0, 0.2, 0.3, 0.4, 0.5) have been tested. The parameters \( \omega, \sigma, \) and \( \rho \) were set to \( \omega = 0.25, \sigma = 1.5, \rho = 0.5 \) according to evaluations in a previous project.\(^{27}\)

On the whole, 12 different parameter configurations have been tested. Again a leave-one-out scenario has been used for the determination of the best working parameters resulting in...
192 test runs for RP and 204 test runs for BS and LP. The median of the Dice scores between MABS result and ground truth for RP and BS where calculated for each of the 12 parameter combinations and used to determine the best working parameter configuration.

4. RESULTS

In this section, the newly presented approach will be referred to as “MABSInShape.” The term “InShape segmentation” refers to the combination of model-based segmentation and geodesic active contours.

For evaluation purposes three different metrics have been calculated as follows:

\[
D = \frac{2|A \cap B|}{|A| + |B|}
\]

(A and B refer to the set of nonzero voxels in the two images).

\[
MD(A, B) = \max\{md(a, b), md(a, b)\}
\]

where \(md(a, b) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} d(a, b)\).

\[
DH_{\text{max}}(A, B) = \max \{ \max_{a \in A} \min_{b \in B} d(a, b), \min_{b \in B} \max_{a \in A} d(a, b) \}.
\]

The validation was performed using a PC with an Intel Core i5-3210M CPU with 2.50 GHz and 12 GB RAM.

4.A. Evaluation of MABS using optimized parameters

According to the tests described in Sec. 3, parameter setting 1 (see Table I) in combination with parameter values \(\tau = 0.4\) (BS) and \(\tau = 0.3\) (LP and RP) for the label fusion step have been used to obtain MABS segmentation results for all three structures of interest. Using these settings, the average time for the multiatlas-based segmentation was approx. 15 min per data-set. The median DICE value that could be achieved using MABS was \(D = 0.82\) for LP and RP and \(D = 0.86\) for BS.

The resulting labels were used for the initialization and evaluation of the combined segmentation approach (MABSInShape) as pointed out in Sec. 4.B.

4.B. Evaluation of the performance using MABSInShape segmentation approach

In order to evaluate if the MABS segmentation accuracy could be improved by using InShape segmentation the parameter settings summed up in Table II were used. Min. and max. intensity values for NMI were calculated using intensity values of the respective structures in the training data. \(\omega_1\) and \(\omega_2\) were set empirically based on our experience in former
TABLE II. Parameter settings for InShape segmentation.

| Parameter                          | BS   | LP, RP |
|------------------------------------|------|--------|
| Number of histogram bins used for NMI | 100  | 100    |
| Min. and max. intensity values used for NMI | −100,150 | −50,100 |
| Number of PCs for InShape segmentation | 10 (>95% of variance) | 12 (>95% of variance) |
| \(c\) [Eq. (9)]                  | −0.3 | −0.3   |
| \(\omega_2\) [Eq. (9)]          | 0.5  | 0.8    |
| \(\omega_1\) [Eq. (9)]          | 1    | 1      |
| \(\gamma\) [Eq. (9)]            | 0.6  | 0.1    |

projects. Systematic tests were performed to identify best parameter settings for \(c\) and \(\gamma\) using a leave-one-out approach. The tested parameter ranges were: 0–0.5 for \(c\) and 0–1 for \(\gamma\). Most striking difference between the parameter values for BS and parotids is the significantly higher \(\gamma\)-value for brainstem segmentation. This is due to the fact that LP and RP are showing larger appearance variations than the brainstem. In combination with the fact that only a rather low number of training-sets was available less weight was put on the model for parotid segmentation. The calculation of the InShape representations to initialize the segmentation took around 3 min per data-set and the final segmentation process itself finished after approx. 1 min.

Figure 6 shows the validation results for the segmentation of LP, RP, and BS using MABS and MABSInShape approach. The notation *-M (e.g., RP-M) in Fig. 6 represents results obtained using MABS, whereas *-MIS (e.g., RP-MIS) stands for the respective results using MABSInShape. Dots indicate significant differences between *-M and respective *-MIS results. For each of the three structures median and average DICE score has improved when using MABSInShape compared to MABS, whereas all median values for mean absolute distance and max. Hausdorff distance decreased. Dots on top of the box plots indicate significant differences between MABS and respective MABSInShape results for each structure based on a paired Wilcoxon rank test (\(p < 0.05\)).

5. DISCUSSION AND CONCLUSIONS

In this paper, a fully automated hybrid segmentation approach, that combines multiatlas-based segmentation using label fusion, statistical appearance models and geodesic active contours in one approach has been presented. In a first step, statistical InShape models have been created for BS, LP, and RP. The model creation process is fully automated and provides a compact representation of appearance variations for the structures of interest. The MABS acts as basis for the new hybrid approach and provides high quality segmentation results even in the presence of noise and low image contrast. However, MABS is lacking the ability to provide...
anatomically plausible segmentation results and partly also shows less accurate results near boundaries. In contrast to this, the InShape approach offers the possibility to automatically create statistical models that combine the description of shape variability and appearance variations reflected by intensity differences in one compact representation. Using the MABS result in combination with InShape models and geodesic active contours has two main benefits: On the one hand, the final MABSInShape result is based on providing the most similar plausible appearance of a structure based on the variations learned from a training set given the MABS result. This is performed totally independent from image noise and image contrast. Geodesic active contours (GAC) on the other hand are based on using image information and provide high accuracy, in areas where reliable image information (in terms of image boundaries) is present. The combination of both, InShape and GAC, as described in Sec. 2.F combines the strengths of both methods in one approach. The typical problem of geodesic active contours, which is the leaking of the evolving contour into neighboring structures in the absence of strong boundaries can be avoided by combining GAC with InShape models. On the other hand, the typical downside of model-based segmentation approaches, which is their susceptibility to local minima, has been overcome by using MABS as an initialization providing the ability to restrict the final optimization to a rather local search space.

The presented pipeline has been tested for the segmentation of brainstem and parotid glands, which are two crucial structures in RT treatment planning. Moreover, the segmentation of both structures is a very challenging task due to low soft tissue contrast and the high amount of variabilty concerning shape and appearance of these structures. Both structures have also acted as test structures in several head and neck auto-segmentation challenges. Although MABS already provided highly qualitative results, the evaluated metrics could be improved for all three structures using the combined MABSInShape approach. All quantitative tests were carried out using a leave-one-out cross-validation scenario in order to maximize the validity of the evaluation.

The approach most closely related to this work has been introduced by Fortunati and is also based on using a combination of three methods (atlas-based segmentation, statistical intensity models, and graph cuts). However, in contrast to Fortunati, a statistical model that is calculated offline (independently from the segmentation process) was used in this project and the type of model is largely different. Moreover, in the presented approach GAC and statistical appearance model are mutually influencing each other, while in Fortunati’s approach statistical models and atlas-based segmentation act as an input for the graph-cuts, but not vice versa. In addition, our approach seems to be faster (3 h per data-set using Fortunati’s approach). However, it has to be considered that the number of segmented structures in Fortunati’s work is higher compared to our work.

Due to the lack of common validation data and partly also due to the usage of different validation metrics, comparison to other methods is a difficult task. However, when comparing the obtained DICE (DS) score (which is the most commonly used metric in the literature) to the results of other studies it turns out that they are at least in the same range or higher. Recently published results for BS segmentation are: $DS = 0.78$,$^{13} DS = 0.83$,$^5$ $DS = 0.91$,$^4$ $DS = 0.58$,$^{36}$ $DS = 0.76$ (Ref. 37) vs $DS = 0.86$ using our approach. Average DICE for parotid glands were: $DS = 0.85$ (left) and $DS = 0.86$ (right),$^5$ $DS = 0.83$ (left and right),$^4$ $DS = 0.69$ (left), and $DS = 0.66$ (right),$^{36}$ $DS = 0.72$ (left PG) (Ref. 37) vs. $DS = 0.84$ (left PG) and $DS = 0.81$ (right PG) using our approach. Wachinger et al. obtained a median DICE score of $DS = 0.80$ (left and right parotid)$^{19}$ using a subset (10 patients) of the data-sets that were used in this project. It also has to be mentioned that the number of test cases in our study (16 for RP and 17 for LP and BS) was larger than in all the other studies (6–13 data-sets) cited above. However, especially for model-based approaches, the number of available training data is still quite low considering the large appearance variations of the segmented structures among different patients.

Ongoing and future work aims at testing different atlas selection strategies in combination with MABSInShape in order to further improve the speed of the segmentation. Moreover, the approach shall be tested on additional structures in the head-neck area.

Unfortunately, it was not possible to assess interobserver variability for the segmentation of our data sets, since only one manual delineation was available for each structure. Consequently, the initiation of a multicenter study in order to create a larger test data-set, for which interobserver variability for different structures in the head neck region can be obtained, is another major objective for the future.

ACKNOWLEDGMENTS

This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant No. U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.
head and neck CT images,” Med. Image Comput. Comput. Assist. Interv. 11(Pt 2), 434–441 (2008).

6X. Zhang, J. Tian, Y. Wu, J. Zheng, and K. Deng, “Segmentation of head and neck CT scans using atlas-based level set method,” MIDAS J. (2009).

7O. Commowick, S. K. Warfield, and G. Malandain, “Using frankenstein’s creature paradigm to build a patient specific atlas,” Med. Image Comput. Comput. Assist. Interv. 12(Pt 2), 993–1000 (2009).

8T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, “Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains,” Neuroimage 21(4), 1428–1442 (2004).

9O. Commowick and G. Malandain, “Efficient selection of the most similar image in a database for critical structures segmentation,” Med. Image Comput. Comput. Assist. Interv. 10(Pt 2), 203–210 (2007).

10B. M. Dawant, S. L. Hartmann, J. P. Thirion, F. Maes, D. Vandermeulen, and P. Demaille, “Automatic 3-d segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. I. Methodology and validation on normal subjects,” IEEE Trans. Med. Imaging 18(10), 909–916 (1999).

11M. R. Sabuncu, B. T. Thomas Yeo, K. v. Leemput, B. Fischl, and P. Goland, “A generative model for image segmentation based on label fusion,” IEEE Trans. Med. Imaging 29(10), 1714–1729 (2010).

12F. van der Lijn, M. de Bruijne, S. Klein, T. den Heijer, Y. Y. Hoogendam, A. van der Lugt, M. M. B. Breiter, and W. J. Niessen, “Automated brain structure segmentation based on atlas registration and appearance models,” IEEE Trans. Med. Imaging 31(2), 276–286 (2012).

13V. Fortunati, R. F. Verhaart, F. van der Lijn, W. J. Niessen, J. F. Veenland, M. M. Paulides, and T. van Walsum, “Tissue segmentation of head and neck CT images for treatment planning: A multiasia approach combined with intensity modeling,” Med. Phys. 40(7), 071905 (14pp.) (2013).

14T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” IEEE Trans. Pattern Anal. Machine Intell. 23(6), 681–685 (2001).

15T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models—their training and application,” Comput. Vis. Image Underst. 61(1), 38–59 (1995).

16S. Zhang, Y. Zhan, X. Cui, M. Gao, J. Huang, and D. Metaxas, “3d anatomical shape atlas construction using mesh quality preserved deformable models,” Comput. Vis. Image Underst. 117(9), 1061–1071 (2013).

17D. Rueckert, A. F. Frangi, and J. A. Schnabel, “Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration,” IEEE Trans. Med. Imaging 22(8), 1014–1025 (2003).

18K. D. Fritscher, A. Grunerbl, and R. Schubert, “3d image segmentation using combined shape-intensity prior models,” Int. J. Comput. Assist. Radiol. Surg. 1(6), 341–350 (2007).

19C. Wachinger, G. Sharp, and P. Golland, “Contour-driven regression for label inference in atlas-based segmentation,” Med. Image Comput. Comput. Assist. Interv. 16(Pt 3), 211–218 (2013).

20V. Caselles, R. Kimmel, and G. Sapiro, “Geodesic active contours,” Int. J. Comput. Vis. 22(1), 61–79 (1997).

21M. Peroni, “Methods and algorithms for image guided adaptive radio- and hardon-therapy,” Ph.D. thesis (Politecnico di Milano, 2011).

22J. V. Hajnal, D. Hill, and D. Hawkes, Medical Image Registration, Biomedical Engineering Series (CRC Press, Boca Raton, FL, 2001).

23G. C. Sharp, M. Peroni, R. Li, J. Shackleford, and N. Kandasamy, “Evaluation of preplasmic b-spline registration on the empire10 data set,” Medical Image Analysis for the Clinic: A Grand Challenge (2010), pp. 99–108.

24P. E. Gill and W. Murray, “Quasi-newton methods for unconstrained optimization,” IMA J. Appl. Math. 9(1), 91–108 (1972).

25D. Mattes, D. R. Haynor, H. Vesselle, Th. K. Lewellyn, and W. Eubank, “Nonrigid multimodality image registration,” Proc. SPIE 4322, 1609–1620 (2001).

26R. C. Maurer, R. Qi, and V. Raghavan, “A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions,” IEEE Trans. Pattern Anal. Machine Intell. 25(2), 265–270 (2003).

27G. C. Sharp, M. Peroni Peroni, N. Shusharina, J. Shackleford, P. Golland, and G. Baroni, “A robust intensity similarity measure for multi-atlas segmentation,” Med. Phys. 40(6), 536–537 (2013).

28T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, “Diffeomorphic demons: Efficient non-parametric image registration,” NeuroImage 45(Supplement 1), S61–S72 (2009).

29M. Chen, W. Lu, Q. Chen, K. J. Ruchala, and G. H. Olivera, “A simple fixed-point approach to invert a deformation field,” Med. Phys. 35(1), 81–88 (2008).

30K. Fritscher, A. Grunerbl, M. Hanni, N. Suhm, C. Hengg, and R. Schubert, “Trabecular bone analysis in CT and X-ray images of the proximal femur for the assessment of local bone quality,” IEEE Trans. Med. Imaging 28(10), 1560–1575 (2009).

31B. Schuler, K. D. Fritscher, V. Kuhn, F. Eckstein, T. M. Link, and R. Schubert, “Assessment of the individual fracture risk of the proximal femur by using statistical appearance models,” Med. Phys. 37, 2560–2571 (2010).

32J. C. Spall, Stochastic Optimization, Stochastic Approximation and Simulated Annealing (John Wiley and Sons, Inc., New York, 2001).

33A. van de Water, H. P. Bijl, H. E. Westerlaan, and J. A. Langendijk, “De- lineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia,” Radiother. Oncol. 93(3), 545–552 (2009).

34B. G. Ginniken, K. Murphy, T. Heinmann, V. Pekar, and X. Deng, Medical Image Analysis for the Clinic—A Grand Challenge (CreateSpace Independent Publishing Platform, 2010).

35V. Pekar, S. Allaire, J. Kim, and D. Jaffray, “Head and neck auto-segmentation challenge,” MIDAS J. 11 (2009).

36R. Sims, A. Isambert, V. Gregoire, F. Bidaud, L. Fresco, J. Sage, J. Mills, J. Bourhis, D. Lefkopoulos, O. Commowick, M. Benkebil, and G. Malandain, “A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck,” Radiother. Oncol. 93(3), 474–478 (2009).

37J.-F. Dainsie and A. Blunhofer, “Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: A clinical validation,” Radiat. Oncol. 8(1), 154 (2013).