

Atlas-Based Auto-segmentation of Head and Neck CT Images

Xiao Han¹, Mischa S. Hoogeman², Peter C. Levendag², Lyndon S. Hibbard¹, David N. Teguh², Peter Voet², Andrew C. Cowen¹, and Theresa K. Wolf¹

¹ CMS, Inc., 1145 Corporate Lake Drive, St. Louis, MO 63132, USA

² Erasmus Medical Center – Daniel den Hoed, Rotterdam, The Netherlands

Abstract. Treatment planning for high precision radiotherapy of head and neck (H&N) cancer patients requires accurate delineation of many structures and lymph node regions. Manual contouring is tedious and suffers from large inter- and intra-rater variability. To reduce manual labor, we have developed a fully automated, atlas-based method for H&N CT image segmentation that employs a novel hierarchical atlas registration approach. This registration strategy makes use of object shape information in the atlas to help improve the registration efficiency and robustness while still being able to account for large inter-subject shape differences. Validation results showed that our method provides accurate segmentation for many structures despite difficulties presented by real clinical data. Comparison of two different atlas selection strategies is also reported.

1 Introduction

Head and neck (H&N) cancer is the sixth most common cancer in the world and accounts for more than 40,000 new cases annually in the United States and about half a million cases worldwide. Due to the complex geometry of the various target volumes and organs at risk (OARs), it is desirable to apply intensity-modulated radiotherapy (IMRT) for their treatment. The successful implementation of IMRT, however, requires accurate delineation of target volumes and OARs on planning computed tomography (CT) images. Manual contouring, although still the de facto standard in clinical radiotherapy planning, is tedious, time-consuming, and suffers from large intra- and inter- rater variability. Tools for automated segmentation are thus needed.

In this work we develop an atlas-based method for automatic segmentation of critical structures and lymph node regions in a given H&N CT image. By definition, atlas-based segmentation is the process of performing segmentation on novel data using the knowledge of a prior segmentation - a dataset that has had the structures of interest already labeled. Many atlas-based segmentation methods have been proposed in the literature for medical imaging applications (cf. [1,2] and references therein), but most of the developments were driven by the labeling of MR brain images and only few exist for H&N CT image segmentation [2,3,4].

The most critical component of an atlas-based segmentation method is arguably the atlas registration that is used to map the atlas image to a novel subject image. Different atlas-based methods most often differ by the registration methods they apply. Image registration in atlas-based segmentation is a difficult problem due to the presence of large inter-subject anatomical variations as well as image noise and other artifacts. To achieve accurate and robust atlas registration, in this work we designed a hierarchical image registration framework that comprises a linear registration and two non-linear registration steps. A major consideration in our specific design is to incorporate object shape information from the atlas to help constrain the registration process. The advantage is at least twofold. First, such information can be used to effectively reduce the degrees of freedom of the registration problem, thus improving the robustness of the method. For example, to segment objects in a novel image, we only need the mapping information for object boundary points in the atlas image. Correspondence for object boundary points are also more reliable to find than interior ones. Second, using the existing object shape as a constraint can help design a better registration regularization method. Typical isotropic smoothing has difficulty dealing with incoherent deformation of neighboring structures. In this work, we propose a surface-based regularization approach, which ensures smooth and coherent deformation of object boundary points in the atlas image. Such a regularization is much more efficient to compute than volumetric anisotropic smoothing. The same framework also allows the use of statistical shape information if desired.

In the following, we first present the proposed atlas registration framework, with emphasis on a novel “poly-smooth” non-linear image registration scheme. We then discuss different atlas selection strategies. Finally, we present validation results based on real clinical data.

2 Hierarchical Atlas Registration

The hierarchical atlas registration framework we developed consists of three major steps: a linear registration, an object-driven “poly-smooth” non-linear registration, and a shape-constrained dense deformable registration. Each step has increased degrees of freedom, and is used to provide initialization for the next step. To deal with possible image contrast changes, we use mutual information and local correlation coefficients as the image matching metric. In the following, we use I to denote the atlas image, and J a novel subject image to be segmented.

2.1 Linear Registration

The linear registration step aims to correct for global differences in position, orientation, and size between an atlas and a subject image. We use a 9 degrees of freedom linear registration model where the free parameters include three translation, three rotation, and three scaling factors. To find the optimal transformation parameters, we maximize the global mutual information (MI) between

the atlas and subject images using a multi-resolution stochastic gradient-descent optimization scheme [5].

2.2 Poly-Smooth Nonlinear Registration

The term ‘‘poly-smooth’’ registration is analogous to the ‘‘poly-affine’’ (or locally-affine) registration proposed by Arsigny et al. [6]. Like in [6], the whole image registration is driven by the alignment of major structures defined in the atlas image, i.e., the objects to be segmented. The basic idea is to reduce the degrees of freedom of the non-linear registration, in order to achieve a more robust registration. The poly-affine framework, however, restricts each object to follow an affine transformation, which is insufficient to model large inter-subject shape variations. An ad hoc solution was introduced in [6] that splits a large object into multiple smaller components. Here, we only force points belonging to the same object to deform coherently, i.e., the deformation field should be smooth along the boundary surface of each object. Thus, the model is more flexible and can naturally handle large shape changes.

The poly-smooth registration is computed iteratively. At each iteration, the optimal displacement $\mathbf{u}(\mathbf{x})$ at each object boundary point (as defined in the atlas) is first estimated through a block-matching approach [7]. Similar to [7], we use a block-wise mutual information as the local image similarity measure:

$$\text{BMI}_{\mathbf{x}}(I, J, \mathbf{u}(\mathbf{x})) = \int_{B(\mathbf{x})} \log \frac{p(I(\tilde{\mathbf{x}}), J(\tilde{\mathbf{x}} + \mathbf{u}(\mathbf{x})))}{p_I(I(\tilde{\mathbf{x}}))p_J(J(\tilde{\mathbf{x}} + \mathbf{u}(\mathbf{x})))} d\tilde{\mathbf{x}} , \quad (1)$$

where $B(\mathbf{x})$ denotes the local neighborhood of \mathbf{x} in the atlas image I . $p_I(\cdot)$ and $p_J(\cdot)$ denote the marginal intensity probability density functions, and $p(\cdot, \cdot)$ is the joint intensity probability density function. The three density functions can be estimated based on the joint intensity histogram of the two images.

The initial estimates $\mathbf{u}(\mathbf{x})$ from block-matching can be noisy. To enforce smooth deformation of each object, we regularize $\mathbf{u}(\mathbf{x})$ by minimizing the following energy:

$$E(\mathbf{u}^{\text{opt}}) = \int_S \|\mathbf{u}^{\text{opt}} - \mathbf{u}\|^2 + \lambda \|\nabla_S \mathbf{u}^{\text{opt}}\|^2 dS , \quad (2)$$

where the integral and the gradient operators are both defined with respect to the object boundary surface. In practice, \mathbf{u}^{opt} can be approximated by smoothing $\mathbf{u}(\mathbf{x})$ over the object surface using a surface Gaussian filter. The regularization is performed for each object separately. We note that if a statistical shape model is available for an object, we can also smooth $\mathbf{u}(\mathbf{x})$ by projecting it onto the eigenspace of the learned shape deformations.

Once \mathbf{u}^{opt} is computed for all object surface points, we extrapolate it to the whole image domain following the nearest-neighbor rule. We then smooth this volumetric deformation field using a weighted Gaussian smoothing, where a high weight value (e.g., 1) is assigned to object surface points, and a low value (e.g., 0.0001) for others. The smoothing leads to a regular volume deformation field.

The reason that we compute a full volume deformation field is because the evaluation of the block-wise mutual information requires correspondence information of off-boundary points as well. To ensure the invertibility of the computed registration, we further apply the Lie group filtering method of [8] to the estimated volume deformation field to get the final field at the current iteration. Note that the volumetric deformation field smoothing and regularization also leads to natural coupling of the deformation of neighboring objects.

The output deformation field of the current iteration is applied to the subject image, which is then used to start the next iteration. At convergence, the consecutive deformation fields are composed to get the final poly-smooth registration between the atlas and a given subject image. To further improve computational speed, a multi-resolution implementation is applied.

2.3 Shape-Constrained Dense Deformable Registration

In the final stage of atlas registration, we compute a full free-form deformable registration to further refine the results produced by the previous step. Unlike most existing free-form registration methods, we again make use of the prior atlas segmentation to ensure smoothness of mapped object surfaces.

Like in the previous step, this dense volumetric registration alternates between optimizing the matching of the atlas and subject images and regularization of the estimated deformation field. To align object boundaries better while still being robust to image contrast changes, we choose the sum of Gaussian-windowed local correlation coefficients (LCC) as proposed in [9] as the image matching metric, which is defined as follows:

$$E(I, J, \mathbf{T}) = \int_{\Omega} (I, J \circ \mathbf{T})(\mathbf{x}) d\mathbf{x} = \int_{\Omega} \frac{\text{COV}(I, J \circ \mathbf{T})}{\sqrt{\text{Var}(I) \cdot \text{Var}(J \circ \mathbf{T})}} d\mathbf{x} , \quad (3)$$

where Ω denotes domain of the image I . $\text{Var}(I) = G * (I - \bar{I})^2$ is the local variance as weighted by a Gaussian window G , $\bar{I} = G * I$ is the local mean, and $\text{COV}(I, J \circ \mathbf{T}) = G * [(I - \bar{I})(J \circ \mathbf{T} - \overline{J \circ \mathbf{T}})]$ is the local covariance between image I and image $J \circ \mathbf{T}$. We found that this similarity measure provides better alignment of image edges than the mutual-information metric. On the other hand, free-form registration using this metric is sensitive to initialization because the metric cannot differentiate well neighboring edges with similar orientation. This problem is largely solved by using the poly-smooth registration result to provide a good initialization.

Let $\mathbf{U}(\cdot)$ denote the deformation field such that $\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{U}(\mathbf{x})$. A gradient-descent minimization of (3) leads to an update of $\mathbf{U}(\cdot)$ at each iteration step by:

$$\mathbf{U} \leftarrow \mathbf{U} + \alpha \left((I - \bar{I}) - (J \circ \mathbf{T} - \overline{J \circ \mathbf{T}}) \cdot \frac{(\overline{I \cdot J \circ \mathbf{T}} - \bar{I} \cdot \overline{J \circ \mathbf{T}})}{\sqrt{\text{Var}(J \circ \mathbf{T})}} \right) \nabla J \circ \mathbf{T} , \quad (4)$$

where α is a user-defined scaling factor.

Unlike [9] where an isotropic Gaussian smoothing is used to regularize the deformation field \mathbf{U} , we propose to incorporate object boundary information in

the atlas image to help preserve the shape of deformed objects and to achieve spatially adaptive regularization. To this end, we formulate the deformation field regularization as the minimization of the following energy functional:

$$E(\mathbf{U}^{\text{opt}}) = \int_{\Omega} \|\mathbf{U}^{\text{opt}} - \mathbf{U}\|^2 d\mathbf{x} + \lambda_1 \int_S \|\nabla_S \mathbf{U}^{\text{opt}}\|^2 dS + \lambda_2 \int_{\Omega \setminus S} \|\nabla \mathbf{U}\|^2 d\mathbf{x} . \quad (5)$$

In (5), S now denotes the union of all object surfaces, and Ω is the whole image domain. Similar to (2), the second term enforces coherent deformation of surface points belonging to the same object, which helps to preserve a smooth shape for each deformed object. The third term enforces smooth deformation over the rest of the volume, which by itself is similar to the isotropic regularization used in [9]. We choose weights $\lambda_1 \gg \lambda_2$ to emphasize the shape-constraint term. With this choice, the optimal solution to (5) can be approximated by two consecutive smoothings of the input field \mathbf{U} . First, we perform a surface-based smoothing of \mathbf{U} for voxels belonging to an object's boundary surface. We then perform a weighted Gaussian smoothing of \mathbf{U} over the whole image volume with weights chosen to be high at object surface locations and low otherwise.

After the final LCC registration step, we can then map the structures' boundary points from the atlas image to the subject image to get an automated segmentation of corresponding structures in the subject image.

3 Atlas Selection

Atlas selection or construction is another important part of an atlas-based segmentation method. Four different atlas-selection strategies are typically found in the literature, as summarized in [1]. The most straightforward strategy is to use one individual segmented image as the atlas. A major drawback, however, is that segmentation accuracy degrades with increasing dissimilarity between the atlas and the image to be segmented. When more than one labeled image are available, a better strategy is to adaptively select an optimal atlas for each individual subject. The third strategy is to compile an average atlas based on all labeled training data. Nevertheless, cross-subject averaging may remove detailed shape information for individual structures, thus limiting the accuracy that can be achieved. A more popular approach now is to apply multiple atlases to generate multiple segmentations of the same image, which can then be combined systematically in a multi-classifier framework to get a final unique segmentation.

In this work, we evaluate and compare two atlas selection strategies, which are the use of an optimal single atlas and the multi-atlas-based segmentation. For the former, the atlas image with the highest MI similarity to the subject image after the global linear registration is selected for the final segmentation. The linear registration is fast to compute, and adds little overhead to the segmentation workflow. For the fusion of multiple single-atlas segmentations, we apply the STAPLE algorithm [10] as the decision fusion method.

4 Results

4.1 Data

H&N CT images of 10 patients were randomly selected from a clinical database and used as the test data. All images have a voxel size of $0.9375 \times 0.9375 \times 2.5 \text{mm}^3$. The following critical structures were manually delineated by experts for each image: the mandible (MAND), the parotids (PAR), the submandibular glands (SMG), the masseter muscles (MASS), the pterygoid muscles (PTERY), the brainstem (B-STEM) and the spinal cord (CORD) (the shortened notations are used in Fig. 2 below). These normal structures plus the patient skin contours (manually labeled as well) serve as the object models in the poly-smooth non-linear registration and in the shape-constrained dense LCC registration. Most datasets also have the lymph node regions (levels II, III, and IV) manually labeled according to the published guidelines of [11].

This group of test data is very challenging for the atlas-based segmentation. Not only do the images differ largely in patient head pose and anatomical shape, the patients also differ in tumor types and stages. For example, there are two nasopharyngeal tumor patients, two oropharyngeal patients, and three base-of-tongue patients. Four patients are node-negative (N0) and the other six are node-positive (N+). In addition, some of the images are contrast-enhanced whereas the others are not.

4.2 Validation Results

We use a leave-one-out strategy to evaluate the proposed segmentation method and the atlas-selection strategies: for each subject, the remaining subjects are considered as possible atlas candidates. The Dice overlapping coefficient (cf. [1]) is used as the accuracy measure.

The automatic segmentation result for one subject is shown in Fig. 1, together with the manual labeling for the same patient. A single atlas was used in this experiment. The manual segmentation shows large differences in node level shapes for the left and the right sides, which are due to differences in the contouring protocols for N+ and N0 patients and also for N+ and N0 sides. These differences are one major source of error for the auto-segmentation of node levels. The agreement for normal structures is visually very good.

The Dice coefficients are summarized in the box plots of Fig. 2, where the left plot summarizes the results with the single optimal atlas strategy and the right summarizes the multi-atlas results. Due to lack of space, the Dice coefficients for bilateral structures such as the parotids are averaged together for each subject.

The single atlas results are less satisfactory, especially for the node level regions. This is partially due to inherent challenges presented by the test datasets. For example, an atlas subject that is close in overall shape to a novel subject may differ in node-involvement (N+ versus N0); thus their manual node level contouring follows different protocols. In addition to the effect of the N0 and N+ differences, another source of error is the disagreement in the separation between

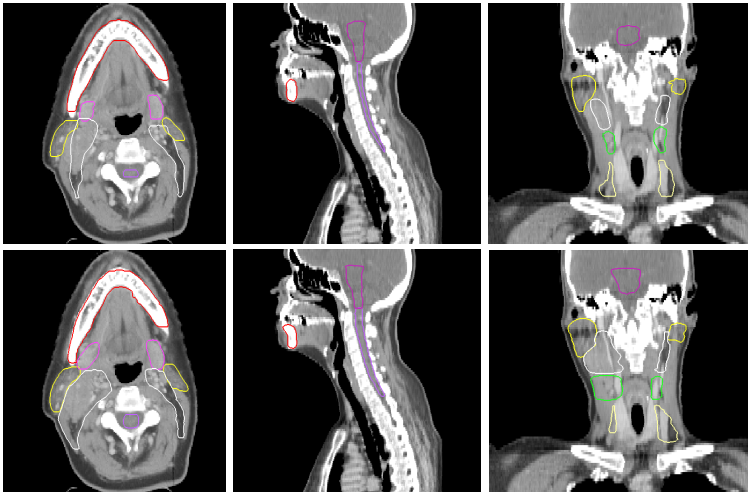


Fig. 1. Comparison of atlas-based segmentation of a N+ patient (top row) with the corresponding manual labeling (bottom row). A single atlas was used. The manual labeling of node levels differs significantly for the N+ (viewer's left) and the N0 (viewer's right) sides.

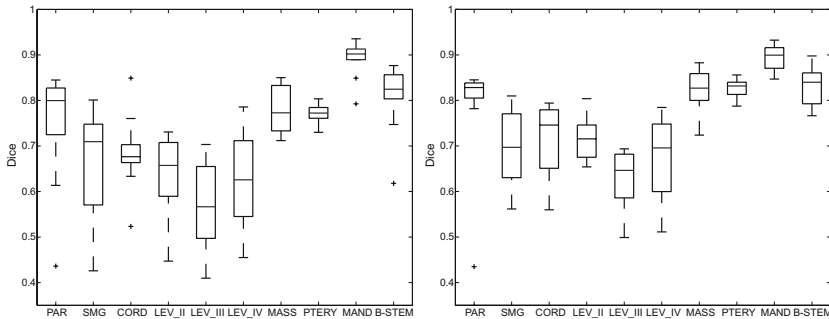


Fig. 2. Box-and-whisker plots for the Dice coefficients for 10 structures. (a) The single atlas results; (b) the multi-atlas results. The boxes show the 25th and 75th percentiles and the center line inside each box shows the median value over 10 subjects.

adjacent node levels, such as between levels II and III and between levels III and IV. The use of multiple atlases significantly improves the accuracy, as shown in the right plot (the improvements were found to be statistically significant at the 0.05 level using the paired-*t* test for all but three structures: parotid, spinal cord, and brain-stem). The median Dice coefficients are over 0.8 for five of the seven normal structures, and are over 0.65 for all structures considered. The results compare favorably with other reported methods in the literature [3,2].

The whole segmentation method takes approximately seven minutes for a single atlas on a quad-core 2.66GHz PC running the Windows XP operating

system. Thus, even with nine atlases, the total computation time is just over an hour. We note that the current code is not fully optimized and there is still much room for further improvement in the computational speed.

5 Conclusion

In this work, we developed an atlas-based method for automatic segmentation of critical structures and lymph node levels for H&N CT images. The method is robust and performs very well on real clinical data, especially when a multi-atlas segmentation strategy is applied. Although manual editing is still needed for most structures before the results can be used for treatment planning, we believe that with the assistance of an automated tool, overall contouring efficiency and consistency can be significantly improved. Future work includes validation on a larger set of data and comparison with other atlas construction strategies.

References

1. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer Jr., C.: Quo vadis, atlas-based segmentation? In: *The Handbook of Medical Image Analysis*, pp. 435–486. Kluwer Academic / Plenum Publishers (2005)
2. Commowick, O., Grégoire, V., Malandain, G.: Atlas-based delineation of lymph node levels in head and neck CT images. *Radiother. Oncol.* (2008)
3. Isambert, A., Grégoire, V., Bidault, F., et al.: Atlas-based automatic segmentation (ABAS) of head and neck (H&N) structures in conformal radiotherapy: Atlas development and preliminary results in clinical context. In: *ESTRO 2007* (2007)
4. Teng, C., Shapiro, L., Kalet, I.: Head and neck lymph node region delineation using a hybrid image registration method. In: *Proc. ISBI 2006*, pp. 462–465 (2006)
5. Viola, P., Wells, W.: Alignment by maximization of mutual information. *Int. J. Comput. Vision* 24(2), 137–154 (1997)
6. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A fast and log-Euclidean polyaffine framework for locally rigid or affine registration. In: *Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057*, pp. 120–127. Springer, Heidelberg (2006)
7. Suárez, E., Santana, J., Rovaris, E., Westin, C., Ruiz-Alzola, J.: Fast entropy-based nonrigid registration. In: *Moreno-Díaz Jr., R., Pichler, F. (eds.) EUROCAST 2003. LNCS, vol. 2809*, pp. 607–615. Springer, Heidelberg (2003)
8. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric diffeomorphic image registration with the Demons algorithm. In: *Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792*, pp. 319–326. Springer, Heidelberg (2007)
9. Pennec, X., Roche, A., Cachier, P., Ayache, N.: Non-rigid MR/US registration for tracking brain deformations. In: *Blum, R., Liu, Z. (eds.) Multi-sensor Image Fusion and Its Applications. Marcel Dekker Inc., New York* (2005)
10. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Trans. Med. Imag.* 23(7), 903–921 (2004)
11. Grégoire, V., Levendag, P., et al.: CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHAHCA, EORTC, GORTEC, NCIC, RTOG consensus guideline. *Radiother. Oncol.* 69, 227–236 (2003)