Christopher Micek
Dr. Russel Taylor
Computer Integrated Surgery II
4/18/2017

Paper Seminar:

Validation of automatic landmark identification for atlas-based segmentation for radiation treatment planning of the head-and-neck region

Authors: Claudia Leavens, Torbjørn Vik, Heinrich Schulz, Stéphane Allaire, John Kim, Laura Dawson, Brian O'Sullivan, Stephen Breen, David Jaffray, and Vladimir Pekar

Introduction:

The aim of our project is to identify corresponding regions of non-contoured soft tissue in the head and neck for patients in a database of run-length encoded binary masks of contoured anatomy from CT images and the associated radiotherapy dose data, with the hope that refined dose-toxicity analysis could then be performed with reasonable accuracy on these areas. This would make the avoidance of critical structures easier in the absence of ground-truth data. Our current approach involves using a subset of these patient masks to make shape atlases including the regions of interest we wish to have in our final analysis. We begin by using coherent point drift (CPD) to deformably register each component patient to a random template patient including all the anatomy of interest via iterative bootstrapping, and then taking the mean of the resulting point clouds to use as an atlas. Any bias incurred by using the initial template patient can be eliminated by repeating the process with the previous atlas as the template.

CPD can easily be used to register contoured anatomy from the atlas to contoured anatomy in another patient, since information already exists about point-to-point correspondences in the target anatomy. However, for situations where the desired anatomy does not exist in the target, we use thin-plate splines (TPS) to register the atlas to our target patient, which yields a deformation field that can be used to bring any arbitrary region from atlas space into the volume of the patient. Our approach currently is satisfactory, but in the future it is worth considering how we can improve accuracy and decrease runtime to increase its usability by clinicians (TPS takes ~120 seconds to deform a volume of 250 points for us). The paper I am reviewing by Leavens et al. uses a landmark-based method for creating atlases of and

segmenting the head and neck using gray-valued CT images, but there are some techniques presented therein that we may be able to apply to our approach.

Technical Approach:

The authors aim to eliminate the need for manual contouring of the head and neck anatomy by developing an automatic segmenting algorithm, and identify three key steps in the process: i) automatic landmark identification in the image dataset of interest, ii) automatic landmark-based initialization of deformable surface models to the patient, and iii) the adoption of these models to patient-specific anatomical boundaries of interest. This paper focuses on the implementation and validation of the first step.

27 landmarks were manually selected by a radio oncologist from 10 patient CTs that met the following criteria:

i) The landmarks could be reliably defined in all patients
ii) They could be distinguished from surrounding structures based on their gray value representations
iii) They could be considered 3D points
iv) They were distributed in relevant regions of interest

In addition to requiring the size, shape, location and orientation of target objects, the expected/permitted amount of variation in these parameters was required to limit the amount of possible deformation in atlas creation and registration. This was accomplished by defining the principle modes of variation in the landmarks using principal component analysis (PCA). The landmarks were first rigidly aligned using the Procrustes method, and then the covariance matrix $\mathbf{C} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$, where $\mathbf{x}_i$ is the coordinate vector of the i-th landmark, and N is the total number of landmarks. Then the eigenvectors $q_j, j = 1, \dots, M$ of $\mathbf{C}$ are the M principle modes of geometric variation in the data. An atlas is created by registering 10 training patient landmarks to a template via TPS, and then averaged.

Once an atlas is created, to transfer the atlas landmarks to a new patient the landmarks must be identified in the patient anatomy, and a registration from the atlas to the patient performed. This is done by finding the optimal transformation of the landmarks on to the patient volume by minimizing the cost function

$$T_{opt} = \arg\min_{T} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{G_i} |g_k - \hat{g}_k|$$

using a stochastic controlled random search (CRS). Here G_i is the number of voxels around the grey-value template around the *i*-th landmark (here a $30 \times 30 \times 30$ mm^3 volume centered about the landmark), and g_k and $\hat{g}_k$ are the grey values of the *k*-th voxels of the patient and atlas template volumes, respectively. The transformations are of the form

$$T(\mathbf{x}, \mathbf{p}) = \mathbf{R}\mathbf{x} + \sum_{j=1}^{M} w_j \mathbf{q}_j,$$

where **R** is a rigid isotropic scaling transformation, and the latter sum is a non-rigid transformation taking into account the principle modes of variation (the authors use $M = 5$). Once the transformation is found, the atlas is registered to the patient with TPS.

A set of 20 patient datasets were available for training and testing (only 16 were considered viable for atlas creation, because four were edentulous). To validate the landmark identification, 20 atlases were created, using different combinations of 10 patients. The landmark identification technique was applied to the remaining 10 patients, and the RMS distances from each registered landmark to ground truth data were calculated and compared. Furthermore, it was hypothesized that atlases with larger eigenvolumes (the product of the non-zero eigenvalues from PCA of their landmarks) would have better performance than those with smaller volumes, since this is indicative of greater atlas variability. The eigenvolumes of all 8008 possible combinations of atlas training datasets were calculated, and the performance of the set of 10 with the largest eigenvolume was compared with the set of 10 lowest.

Results:

The results of the automatic landmark identification for the patient sets above were compared to ground truth data. The average RMS distance of landmarks from the sets with higher eigenvolumes was 9.5 ± 0.6 mm, and that of the smaller was 11.0 ± 0.9 mm, confirming the hypothesis that training datasets with larger eigenvolumes performed better. The CRS to find the optimal transformation from atlas to patient took 60 seconds on a 3.4 GHz Intel PC, with 1 GB RAM.

Discussion:

The authors tout the use of PCA as an effective method of reducing the dimensionality of their data, increasing runtime, and constraining the possible positions of landmarks. In addition, they also denote the landmarks with the best performance: the posterior aspect of the nasal septum, the mental foramen, the mandibular foramen, the inner ear, and the coronoid process, all of which have greater differences in contrast relative to their local surroundings. By contrast, the left and right coracoid process exhibited the poorest performance, due to the nonrigid nature of the anatomy and more variability in ground-truth data. The authors also mention the variability in identifying landmarks for ground truth data as a potential source of error, and that increasing the number of iterations for CRS would improve the performance of landmark identification for larger structures. However, there is no data given to corroborate any of these assertions (How much ground truth variability was there? Is there a measurement of the error this caused? How does the algorithm improve with the number of iterations?), although they seem reasonable.

Although the authors perform their analysis using CT image data using a small number of points, the problem of creating and registering an atlas of the head and neck in a reasonable amount of time is present in both their paper and our project. I think it would be useful for us to incorporate and evaluate some of the optimizations made here for our purposes. For example, we can look at using PCA to reduce the dimensionality of and potentially speed up our registration algorithm, and we can look at the eigenvolumes of our sets of patient points to pick an optimal set of patients to use for atlas creation. I am curious to see how this approach generalizes to larger datasets, as the authors only use a sample of 27 points per patient, whereas a single organ in our patient data can contain several hundred points. Would the CRS runtime still be acceptable? Also, the paper shows that patient data with larger eigenvolumes yield more accurate results, but accuracy of within 9.5 mm seems fairly low to me. There is no mention of the voxel size of the CT data used, so it is difficult to tell if this number is related to the algorithm used or the voxel spacing. It remains to be seen if this error would be higher or lower for a greater number of landmark points.

References

1. Leavens, C., Vik, T., Schulz, H., Allaire, S., Kim, J., Dawson, L., ... & Pekar, V. (2008, March). Validation of automatic landmark identification for atlas-based segmentation for radiation treatment planning of the head-and-neck region. In *Medical Imaging* (pp. 69143G-69143G). International Society for Optics and Photonics.