Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2014 March 21; 9034: 90341G–. doi:10.1117/12.2043715.

Robust Optic Nerve Segmentation on Clinically Acquired CT

Swetasudha Panda^{a,*}, Andrew J. Asman^b, Michael P. DeLisi^c, Louise A. Mawn^c, Robert L. Galloway^b, and Bennett A. Landman^{a,b}

^aElectrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^bBiomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^cOphthalmology and Neurological Surgery, Vanderbilt University, Nashville, TN, USA 37235

Abstract

The optic nerve is a sensitive central nervous system structure, which plays a critical role in many devastating pathological conditions. Several methods have been proposed in recent years to segment the optic nerve automatically, but progress toward full automation has been limited. Multi-atlas methods have been successful for brain segmentation, but their application to smaller anatomies remains relatively unexplored. Herein we evaluate a framework for robust and fully automated segmentation of the optic nerves, eye globes and muscles. We employ a robust registration procedure for accurate registrations, variable voxel resolution and image field-of-view. We demonstrate the efficacy of an optimal combination of SyN registration and a recently proposed label fusion algorithm (Non-local Spatial STAPLE) that accounts for small-scale errors in registration correspondence. On a dataset containing 30 highly varying computed tomography (CT) images of the human brain, the optimal registration and label fusion pipeline resulted in a median Dice similarity coefficient of 0.77, symmetric mean surface distance error of 0.55 mm, symmetric Hausdorff distance error of 3.33 mm for the optic nerves. Simultaneously, we demonstrate the robustness of the optimal algorithm by segmenting the optic nerve structure in 316 CT scans obtained from 182 subjects from a thyroid eye disease (TED) patient population.

Keywords

Multi-Atlas; Statistical Label Fusion; Optic Nerve; Segmentation; Computed Tomography

1. INTRODUCTION

The optic nerve (ON) plays a central role in processing and relaying visual information from the eye to the brain and is sensitive to pathological conditions including optic neuritis, multiple sclerosis (MS), glaucoma, optic neuritis and ischemic neuropathy. Understanding structural changes in the ON in the course of disease (e.g., inflammation, atrophy, axonal congestion) is essential to charting disease progression, planning interventions, and monitoring treatment efficacy. Hence, accurate and robust segmentation of the ON plays an essential role in studying the biophysical etiology, progression, and recurrence of these

* swetasudha.panda@vanderbilt.edu; <http://masi.vuse.vanderbilt.edu>; Medical-image Analysis and Statistical Interpretation Laboratory, Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235.

diseases. Automatic quantification of the location and volumetrics of the ON could increase sensitivity and specificity of pathology assessments as compared to coarse, manual region of interest approaches. Additionally, manual delineation of ON structures is time and resource consuming as well as susceptible to inter- and intra-rater variability.

Current segmentation procedures often require manual intervention due to anatomical and imaging variability and success of automated techniques is often dependent upon the application, modality and image quality. Bekes et al¹ proposed a geometric model-based method for semi-automatic segmentation of the eye balls, lenses, optic nerves and optic chiasm in CT images and reported quantitative sensitivity and specificity results from STAPLE of approximately 77%. Qualitatively, a lack of consistency was reported with the results they obtain for the nerves and chiasm. Noble and Dawant² proposed a tubular structure localization algorithm in which a statistical model and image registration are used to incorporate a priori local intensity and shape information and reported mean Dice similarity coefficient (DSC) of 0.8 when compared to manual segmentations over ten test cases. Atlas-based methods provide a model-free approach to segmentation. Multiple atlases significantly improve the accuracy compared to a single atlas^{3,4}. Although multi-atlas segmentation promises a robust and model-free approach to segment medical images from exemplar brain images, only limited success has been seen for segmentation of the ON with Dice coefficient ranging from 0.39 to 0.78⁵⁻⁷.

Herein, we are particularly interested in the development of a highly reliable technique for the segmentation of the ON, eye globe and muscles using multi-atlas segmentation on clinically acquired CT images. The dataset that is used contains a wildly varying range of field-of-view (ranging from whole head to more localized images of the orbit, as illustrated in Figure 1) and slice thicknesses (ranging from 0.4 to 5 mm)(Table 1). We initially employ a straightforward registration framework in which the bony structure obtained by thresholding the atlases and the target are affinely registered. The centroids of the left and right ON are estimated by k-means clustering of the voxels where more than 90% of the registered atlases agree on the presence of the ON. Using these centroids, a region-of-interest is determined for the ON structure for both the target and the atlases in a leave-one-out approach. We evaluate pairwise non-rigid registration procedures followed by (leave-one-out-cross-validation) LOOCV label fusion on the ON region-of-interest. We show that the optimal registration and label fusion pipeline results in robust and accurate segmentations of the ON anatomy despite wildly varying dataset by evaluating the pipeline on 316 CT scans obtained from 182 subjects from the thyroid eye disease (TED) patient population.

This manuscript is organized as follows. First, we explain the details of the bony structure rigid registration and region-of-interest estimation of the ON anatomy. Second, we explain the evaluation of non-rigid registration and multi-atlas label fusion pipeline for three state-of-the-art non-rigid registration algorithms (refactored version of the free form deformation algorithm in niftyreg package⁸, non-linear registration in the Automatic registration Toolbox (ART) package⁹, and the ANTS Symmetric Normalization (SyN) deformable registration in ANTS package¹⁰) and six label fusion algorithms (majority vote^{3,4,11,12}, STAPLE, Spatial STAPLE¹³, Local Weighted Vote^{14,15}, Non-Local STAPLE¹⁶ and Non-

Local Spatial STAPLE). For each method, we present quantitative and qualitative performance characteristics. Finally we evaluate the performance of the optimal pipeline on the large dataset to demonstrate its robustness.

2. METHODS AND RESULTS

The following is an overview of the framework we employ to evaluate the region-of-interest estimation, pairwise registration and multi-atlas label fusion methods as illustrated in Figure 1. Initially, pairwise affine registration of the bone structure A geometric deformation between an atlas and the target is computed for the bone threshold images. After registration, the set of labels in the coordinate system of each atlas are transformed to the target space to estimate a smaller region-of-interest. This is followed by pairwise non-rigid registration between the cropped target and the atlases. The propagated labels are then fused using multi-atlas label fusion algorithms in a leave-one-out cross validation approach. All studies were run on a 64-bit quad core 3.07 GHz desktop computer with 13 GB of RAM, running Ubuntu 12.04.

CT data

The empirical data consisted of 30 clinically acquired CT images of the brain, which were retrieved, in anonymous form under IRB supervision. The voxel size of the various images varied wildly, with in-plane resolution approximately 0.5 mm and slice thickness ranging from 0.4 mm to 5 mm for the various targets (Table 1). The “ground truth” segmentations were performed by an experienced rater using the MIPAV software package (<http://mipav.cit.nih.gov/>) for the full length of the left and the right optic nerves and the eye globes (with the 2 pairs of surrounding muscles for each globe) on all the subjects. Additionally, 316 CT scans were obtained from 182 subjects from the thyroid eye disease (TED) population.

Registration and multi-atlas label fusion pipeline

The bone structure for each image was first determined by thresholding each image at an intensity value of the minimum intensity increased by 30% of the range of intensities. This threshold was decided based upon visual inspection of representative scans. Pairwise affine registration is then performed between the bone threshold images using the Aladin algorithm^{17,18}, in the niftyreg package. In this implementation, normalized cross-correlation is used as the similarity metric and the block width is constant and equal to 4 voxels. The propagated labels are then averaged over the number of atlases to obtain a probability matrix for each target. The coordinates where the probability of ON labels is greater than 0.9 are identified as the regions where at least 90% of the raters agree on the location of the ON. This set of coordinates is then partitioned into 2 groups using k-means clustering. The centroids of the left and right clusters are then extended by 40 mm in all 3 dimensions to obtain the ON region-of-interest. The final registrations are then computed by performing non-rigid registration between the cropped target and the atlases.

We evaluate three state-of-the-art non-rigid registration algorithms:

1. refactored version of the free form deformation algorithm in niftyreg package with normalized mutual information and bending energy cost function,
2. non-linear registration in the Automatic registration Toolbox (ART) package with the k-means minimum variance cost function, and
3. the ANTS Symmetric Normalization (SyN) deformable registration in ANTS package with cross correlation as the cost function).

The source labels were transferred to the target coordinate space using the deformation field and nearest neighbor interpolation generating 29 propagated label volumes for each image considered as a target. The propagated labels are then fused using multi-atlas label fusion algorithms (majority vote (MV), STAPLE, Spatial STAPLE (spSTAPLE), Local Weighted Vote (LWV), Non-Local STAPLE (NLS) and Non-Local Spatial STAPLE (NLSS)) in a leave-one-out cross validation approach. For each statistical label fusion algorithms evaluated, identical parameter values were used whenever applicable. In NLS, the search neighborhood and the patch neighborhood were set to a $2 \times 2 \times 2$ mm and $1 \times 1 \times 1$ mm window respectively, centered at the target voxel of interest. The patch neighborhood was initialized to a $1 \times 1 \times 1$ mm window centered at the voxels of interest. The values of the standard deviation parameters σ_i and σ_d were set to 0.5 and 1.5 mm respectively. A combination of mean square difference and locally normalized cross correlation was used as an intensity similarity metric. In NLSS, the performance level parameters were calculated on a voxel-wise basis using a half-window size of $3 \times 3 \times 3$ mm in all cardinal directions

Quantitative accuracy is assessed using the DSC¹⁹, Hausdorff distance (HD)²⁰, and mean surface distance (MSD). The symmetric surface distance metrics are computed in both directions in terms of distance from the expert labels to the estimated segmentations and vice versa.

Figure 2 presents quantitative results for the three different structures considered are shown in for the 30 subjects. For both the structures, SyN ANTS registration followed by NLSS label fusion provided the most consistent results with a median DSC of 0.77, MSD 0.55 mm and HD 3.33 mm for the ONs and 0.86, 0.58 mm and 5.04 mm for the globes structure, respectively. Similar results were obtained for the eye globe structure.

The qualitative results presented in Figure 3 show reasonable segmentation of the ON structure validating the reliability of this method. Representative slices for 7 subjects are shown in the first two rows for comparison of manual and automatic segmentations. Slight over segmentation of the ONs is observed in subjects 4 and 7. The voxel-wise surface distance error calculated from the truth to the estimate is shown separately for the ONs and the eye globe structure. The automatic results are susceptible to larger distance errors at all boundaries and eye globe-muscle connection.

Performance analysis on the large dataset for validation of robustness

The above segmentation pipeline was used to segment the large dataset comprising 316 scans from 182 subjects. To access the quality of the results, the volumes of the automatic segmentations were first calculated for the ON and the eye globe structure and the outliers

were analyzed as shown in Figure 4(A). The volume histogram shows that the method works for a large number of the test scans. To isolate the outliers, we plot the label volumes against the slice thickness (which varies widely across scans) for both the manual labels on the initial 30 scans as well as the automatic results on the 316 test scans. Each outlier was manually examined. The failure cases are numbered as F1–5. These belong to 2 subjects with tumors in the ON region that resulted in over segmentation. The region of interest detection also failed in scans which included excess regions (back of the skull and extra background/the neck area as shown in F6–11), which might be due to inaccurate affine registration in those cases, leading to final segmentations in misplaced positions.

To further evaluate the accuracy of the results, we performed principal component analysis (PCA)²¹ on the images as shown in Figure 4(B). Firstly, all 316 test scans and the 30 initial scans were affine registered to a common quality analysis space chosen as one of the initial scans. The 316 automatic and 30 manual labels were transferred to the same space. Using the centroid of these labels, all the images were cropped around the ON ROI.

Each registered intensity image I_r was normalized by subtracting the mean intensity and dividing by the standard deviation.

$$I_{normalized} = \frac{I_r - \text{mean}(I_r)}{\text{std}(I_r)} \quad (1)$$

The normalized images were cropped around the ON ROI and 2 slices ($z=66,70$) were chosen. These two representative slices for each image were stored as a single $[295 \times 297 \times 2]$ column vector. All the images were similarly arranged to form an $L \times N$ matrix with $L=295 \times 297 \times 2$, and $N=316$.

The average image is then defined as, $I_{average} = \frac{\sum_{n=1}^N I_n}{N} \dots$

Each image differs from the average by the vector

$$\emptyset_i = I_i - I_{average} \quad (3)$$

Let $A = [\emptyset_1 \emptyset_2 \dots \emptyset_M]$, so the covariance matrix is given by

$$C = \frac{\sum_{n=1}^M \emptyset_n \emptyset_n^T}{M} = A A^T \quad (4)$$

For $N < L$, there will be only $N-1$ meaningful eigenvectors and the remaining will have associated zero eigenvalues. Considering the eigenvectors of $A A^T$,

$$A^T A v_i = \mu_i v_i \quad (5)$$

Pre-multiplying both sides by A ,

$$AA^T Av_i = \mu_i Av_i \quad (6)$$

Thus, eigenvectors of C are Av_i . These are normalized to obtain u_i , the eigenfaces. Let $\varepsilon = u_i(:, 1:2)$ be the first two eigenvectors. The images are projected into the first two principal components to obtain the weights ω_i .

$$\omega_i = \varepsilon^T A \quad (7)$$

The initial 30 CT images for which manual labels are available were also projected into the same space as above by normalization, subtraction of column means and multiplication with the first two columns of u . The automatic and manual projections were plotted and the outliers were manually examined for segmentation failure.

For analyzing the labels, an Euclidean distance transform was initially performed after separating the 4 different labels in each of the slices (same as for the intensity, $z=66$ and $z=70$) as binary images. This was followed by subtraction of column means, calculation of normalized eigenvectors and projection of the automatic and manual labels into the space of the first two principal components. Scans appearing as outliers in any of the plots were marked as failure cases in all of the plots. 15 scans from 11 subjects were detected as outliers. 2 subjects (16 scans) had optic nerve tumor and not all of them appear as outliers because the automatic segmentations are still of similar shape to the rest of the images. 8 subjects (8 scans) failed due to excess field of view (including the skull or the neck) and 1 subject (1 scan) had abrupt variation in the successive slices due to transmission failure of “classic” slice-wise DICOM. It is noted that not every tumor scan appears as an outlier in the PCA analysis because of similar shape of the overall intensity images and the segmentations, but all the scans have been marked as outliers.

One of the initial scans with slice thickness greater than 5 mm is poorly registered and appears like an outlier but is actually a working image as shown in the previous section (i.e., non-rigid registration works after the ROI cropping). A few of the working results are also shown. The similar distribution of the manual labels and the automatic labels projected into the principal components space shows the accuracy of the automatic segmentation on the large dataset even with wide variability in terms of subjects and scan slice thickness. This successfully demonstrates the robustness of the proposed pipeline.

3. DISCUSSION

The proposed fully automatic framework using multi-atlas segmentation provides consistent and reasonably accurate segmentations of the ON structure despite wildly varying clinically acquired datasets. The robust registration framework is able to consistently handle the high variability of the data in terms of field-of-view and voxel resolutions. The recently proposed NLSS label fusion algorithm is able to account for the small registration errors. We demonstrate that the optimal method is significantly more accurate compared to various state-of-the-art registration and fusion algorithms. Additionally, we demonstrate the robustness of the segmentation pipeline on a massive dataset with only 25 scans failing out

of the total 316, out of which 16 have tumors, 8 have excessive FOV, and 1 has corrupt data (missing slices).

Despite the success of the proposed framework, there is potential to improve performance through study of shape priors in the label fusion estimation framework, intensity-based refinement²², Markov Random Field regularization²³, or learning based correction of mislabeled voxels²⁴. Also, automatic detection and removal of unnecessary field-of-view for including only the head and the ON region could help resolve the outliers identified in the test dataset that were not due to excessive anatomical variability or corrupt data. Given that each step in the proposed pipeline is independent of the structures being segmented, and hence “model-free”, multi-atlas labeling provides a robust approach to segmentation of orbital anatomy, and is robust to the substantial variation present in clinically acquired data. Our automated segmentation approach could be used to provide analysis context (i.e., navigation), volumetric assessment or enable regional nerve characterization (i.e., localize changes).

Acknowledgments

This project was supported by NIH/NEI R21EY024036, NIH/NIBIB 1R03EB012461 and the National Center for Research Resources, Grant UL1 RR024975-01 (now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN.

References

1. Bekes G, Máté E, Nyúl LG, et al. Geometrical model-based segmentation of the organs of sight on CT images. *Medical physics*. 2008; 35:735. [PubMed: 18383695]
2. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*. 2004; 23:903–921.
3. Heckemann RA, Hajnal JV, Aljabar P, et al. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage*. 2006; 33:115–126. [PubMed: 16860573]
4. Rohlfing T, Russakoff DB, Maurer CR. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *Medical Imaging, IEEE Transactions on*. 2004; 23:983–994.
5. D’Haese, P-FD.; Duay, V.; Li, R., et al. Automatic segmentation of brain structures for radiation therapy planning. *Proceedings, Medical Imaging*; 2003; 2003. (available from International Society for Optics and Photonics)
6. Gensheimer, M.; Cmelak, A.; Niermann, K., et al. Automatic delineation of the optic nerves and chiasm on CT images. *Proceedings, Medical Imaging*; 2007. (available from International Society for Optics and Photonics)
7. Isambert A, Dhermain F, Bidault F, et al. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiother Oncol*. 2008; 87:93–99. [PubMed: 18155791]
8. Rueckert D, Sonoda LI, Hayes C, et al. Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on*. 1999; 18:712–721.
9. Ardekani BA, Braun M, Hutton BF, et al. A fully automatic multimodality image registration algorithm. *J Comput Assist Tomogr*. 1995; 19:615. [PubMed: 7622696]

10. Avants, BB.; Grossman, M.; Gee, JC. Biomedical Image Registration. Vol Springer; 2006. Symmetric diffeomorphic image registration: Evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe; p. 50-57.
11. Aljabar P, Heckemann RA, Hammers A, et al. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*. 2009; 46:726–738. [PubMed: 19245840]
12. Sabuncu MR, Yeo B, Van Leemput K, et al. A generative model for image segmentation based on label fusion. *Medical Imaging, IEEE Transactions on*. 2010; 29:1714–1729.
13. Asman AJ, Landman BA. Formulating Spatially Varying Performance in the Statistical Fusion Framework. *Medical Imaging, IEEE Transactions on*. 2012; 31:1326–1336.
14. Artaechevarria X, Munoz-Barrutia A, Ortiz-de-Solorzano C. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *Medical Imaging, IEEE Transactions on*. 2009; 28:1266–1277.
15. Isgum I, Staring M, Rutten A, et al. Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans. *Medical Imaging, IEEE Transactions on*. 2009; 28:1000–1010.
16. Asman AJ, Landman BA. Non-local statistical label fusion for multi-atlas segmentation. *Med Image Anal*. 2012
17. Ourselin S, Roche A, Subsol G, et al. Reconstructing a 3D structure from serial histological sections. *Image and vision computing*. 2001; 19:25–31.
18. Ourselin, S.; Stefanescu, R.; Pennec, X. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002*. Vol Springer; 2002. Robust registration of multi-modal images: towards real-time clinical applications; p. 140-147.
19. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945; 26:297–302.
20. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 1993; 15:850–863.
21. Turk M, Pentland A. Eigenfaces for recognition. *Journal of cognitive neuroscience*. 1991; 3:71–86. [PubMed: 23964806]
22. Lötjönen JMP, Wolz R, Koikkalainen JR, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage*. 2010; 49:2352–2365. [PubMed: 19857578]
23. Cardoso MJ, Clarkson MJ, Ridgway GR, et al. LoAd. A locally adaptive cortical segmentation algorithm. *Neuroimage*. 2011; 56:1386–1397. [PubMed: 21316470]
24. Wang H, Das SR, Suh JW, et al. A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. *Neuroimage*. 2011; 55:968–985. [PubMed: 21237273]

Figure 2.

Quantitative results of the evaluation of non-rigid registration and label fusion algorithms on the ONs and eye globe structure show that SyN diffeomorphic registration followed by Non-Local Spatial STAPLE label fusion is the most consistent performer across all 30 subjects

Figure 3.

Qualitative results for the optimal multi-atlas segmentation approach for 7 subjects are shown. For a typical subject, the top rows compare manual and automatic results for a representative 2D slice. The bottom rows show point-wise surface distance error of the label fusion estimate for the ONs and the eye globe structure. The proposed multi-atlas pipeline results in reasonably accurate segmentations for the ON structure. However, slight over segmentations of the ONs can be observed in certain cases (subjects 4 and 7).

Figure 4.

Performance analysis of the proposed segmentation pipeline on the dataset comprising 316 CT scans. (A) The top plots present histograms of the automatic segmentation volumes for ON and the eye globe structure. The lower plots show the volume plotted against the slice thickness. The red stars are for the 30 manual labels, and the blue stars are for the 316 test cases. Few of the failure (F) and working (W) cases (by manual examination of outliers) are shown. (B) Principal component analysis after registration to a common space on the 4 labels to isolate outliers. The 11 outlier subjects identified were either due to the presence of tumor (F1–2), due to excess field of view (including the skull or the neck) (F3–10) or due to discontinuous scans (F11).

Table 1

Variability in slice thicknesses for the dataset of 30 subjects.

Slice thickness (mm)	0.4	0.5	1.0	2.0	2.5	3.0	5.229
Number of images	2	8	1	4	12	2	1