

Atlas-based Auto-contouring – Balancing Accuracy with Efficiency in OnQ rts®

Susan Barley,¹ Clare Antoine,² Gareth Webster,² Marie Tiffany,² Navinah Nundlall,² Rosemary Simmons³ and Andrew Hartley⁴

1. Lead Clinical Scientist, Oncology Systems Limited, Shrewsbury, UK; 2. Medical Physicist; 3. Specialist Radiographer;
4. Clinical Oncologist, Queen Elizabeth Hospital, Birmingham, UK

Abstract

Atlas-based auto segmentation with OnQ rts® has been shown to deliver time-savings for the delineation of organs at risk in head and neck patients being treated with intensity-modulated radiotherapy. However, as the initial time to set up atlases can be high in busy departments the optimal number of atlas cases needed for auto-contouring was investigated. Using conformity index and mean distance to conformity to compare automatically generated with gold standard clinical contours, it was found that the majority of contours were unaffected by reducing the number of atlas cases from 30 to 10. The optimum number of atlas cases, however, was considered to be 20 due to the reduction in accuracy of the mandible, larynx and brain, below this level.

Keywords

Radiation therapy, atlas-based segmentation, radiotherapy planning, deformable image registration, OnQ rts

Disclosures: Susan Barley is an employee of Oncology Systems Limited. Clare Antoine, Gareth Webster, Marie Tiffany, Navinah Nundlall, Rosemary Simmons and Andrew Hartley have no conflicts of interest to declare.

Received: 13 October 2014 **Accepted:** 2 December 2014 **Citation:** *European Oncology & Haematology*, 2014;10(2):98–101 DOI: 10.17925/EOH.2014.10.2.98

Correspondence: Susan Barley, Oncology Systems Limited (OSL), 14 Longbow Close, Shrewsbury, SY1 3GZ, UK. E: susan.barley@osl.uk.com

Support: The publication of this article was supported by Oncology Systems Limited.

Delineation of contours remains the one part of radiotherapy that is completely manual in the majority of centres and therefore the potential for time-saving is huge,¹ being estimated to range from 23 %² to 41 %.³ One of the additional benefits of auto-contouring is the reduction in inter-observer variation, which in general is not quantified in radiotherapy. Balik et al.,⁴ showed that, when comparing the demons and small deformation inverse consistent linear elastic (SICLE) algorithm for propagating contours between cone beam computed tomography (CBCT) scans of non-small cell lung carcinoma (NSCLC) patients taken weekly, over 7 weeks that the Dice similarity coefficient (DSC) for the two algorithms was similar to the DSC seen for manual contouring between different trained observers. This shows that all contours should be independently checked by a separate clinician.^{3,5} However, if atlas-based contours are computer generated, the need for the second clinician is replaced by one clinician checking the contours produced by the software:³ another potential manpower saving of an auto-contouring system.

Sharp et al.⁶ performed an overview of the current status of auto segmentation in clinical radiotherapy. One of the main findings of the paper was that the atlases used should be customised and department specific. The slow uptake in clinical use by centres that have purchased commercially available software products that perform atlas-based segmentation is in many cases linked to the need for atlases to be customised and department specific. The initial input of time to create customised atlases, knowing what will work and how many atlases are needed in the library is too much of a challenge to already

over-stretched centres, despite the time savings of a fully operational system being well cited in the literature.

One cancer centre that has embraced the use of atlas-based segmentation to outline the normal anatomy for all radical head and neck radiotherapy treatment plans is Queen Elizabeth Hospital (QEB), Birmingham, UK. Following on from research that showed that the use of OnQ rts (Oncology Systems Limited, UK) (one of a number of commercially available software products that can be used for atlas-based auto segmentation) could reduce outlining time by 36 minutes – from 90 down to 54 minutes – work has now been performed to establish the optimal number of atlases required in the atlas library for the contouring of head and neck structures and which structures are most affected by the number of atlases used. This work will be reported in this paper.

Background

OnQ rts is a stand-alone software product that has been designed for the display, evaluation, co-registration and fusion of medical images, auto-contouring of anatomical structures and display of radiation therapy dose distributions to aid in radiation therapy planning. The auto-contouring protocol uses atlas-based segmentation of CT image data and is available for the following anatomical sites: head and neck, male pelvis, and thorax.

The concept of atlas-based auto-contouring is to populate the software with a 'library' of prepared, contoured CT cases that act as a typical

range of gold standard template cases, from which the CT images and contours are mapped from an atlas case onto the new patient case. Any structures that exist in an atlas case will be mapped onto the new patient case. This usually includes organs at risk only, but may include target lymph node anatomy, and possibly standardised tumour volumes. The atlas library can be expanded and customised by saving approved volumes of interest (VOIs) from previous patient cases as new atlas files.

Several different approaches are commercially available, varying in precisely how data are chosen from the atlas library, how it is mapped to the patient case and how image processing is applied during the process. OnQ rts automatically selects the ‘best fit’ atlas case using a mutual information algorithm and the lateral and anterior/posterior digital reconstructed radiographs (DRRs) generated when a case/atlas is imported into the system. It then applies rigid image registration (RIR) and deformable image registration (DIR), plus an additional series of processes, to deform the atlas CT case and copy the deformed contours onto the new patient images.

The RIR algorithm geometrically aligns the coordinate system of one dataset to the coordinate system of another by changing its transformation parameters (translation and rotation). The algorithm is based on an iterative technique, which calculates a global similarity measure between the two datasets.⁷ In OnQ rts, the similarity measure used is mutual information. Mutual information is a statistical measure that finds its roots in information theory. It is a measure of how much information one random variable contains about another. The mutual information (I) of two random variables A and B is defined in equation 1:

$$I(A,B) = \sum_{a,b} \mathcal{P}_{A,B}(a,b) \log \frac{\mathcal{P}_{A,B}(a,b)}{\mathcal{P}_A(a) \times \mathcal{P}_B(b)} \quad (1)$$

Where $\mathcal{P}_{A,B}(a,b)$ is the joint probability of the random variables A and B, and $\mathcal{P}_A(a)$ and $\mathcal{P}_B(b)$ are the marginal probabilities of A and B, respectively. For images, the joint and the marginal probabilities can be approximated by the normalised intensity histograms of each volume. The algorithm is based on a multi-resolution iterative technique where each iteration attempts to maximise the mutual information between the two volumes and thus their co-alignment.⁸

The DIR algorithm is a procedure where the moving image (MI) is transformed to match the static image (SI) by local warping of the MI. In this way, local differences between SI and MI can be eliminated and the two datasets can be elastically matched. It is again based on a multi-resolution iterative scheme where during each iteration a similarity measure is evaluated to estimate the co-alignment of the two volumes. The demons algorithm is used in the DIR procedure and it is an intensity-based DIR technique.

The demons algorithm was introduced by Thirion⁹ who proposed that non-parametric non-rigid registration be considered as a diffusion process. The forces are inspired from optical flow equations and the method alternates between computation of the forces and regularisation by a simple Gaussian smoothing. This results in a computationally efficient algorithm compared with other non-rigid registration procedures. In OnQ rts, two variations of the Demons algorithm have been implemented. One deals with single modality problems and the other with multi-modality and is based on the point-wise mutual information criterion.^{9–12}

Table 1: Table Showing the Atlas Library Size Used and the 12 Different Combinations of Atlases

Atlas Library Size	Atlas Case numbers					
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
1	1					
5	1–5	6–10	11–15	16–20	21–25	26–30
10	1–10	11–20	21–30			
20	1–20					
30	1–30					

Method

At QEB, the OnQ rts atlas library is populated with 30 anonymised head and neck CT scans, each having 20 organs at risk (OARs) contoured and reviewed according to local protocols. OARs contoured in the atlases include: mandible, larynx, right and left parotid, brainstem, brain, spinal cord, right and left cornea, chiasm, pituitary, right and left cochlea, right and left optic nerve, right and left orbit, right and left lens. Eleven further CTs from previously treated patients were anonymised for use as test cases.

Atlas cases were numbered 1–30. All OAR contours were automatically generated 12 times on each test case using a range of different atlas library sizes and combinations as shown in Table 1. For example, for an atlas library with five atlases, six combinations of CT datasets (sets) were used to include all 30 atlas cases. The aim was to identify a point at which increasing atlas library size no longer improved agreement between the automatically generated contours and the clinical contours, which were used as a reference.

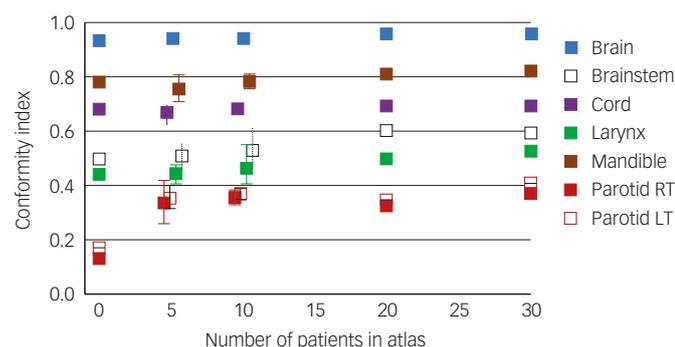
The contours generated by each atlas library were compared against the reference clinical outlines using the analysis module of OnQ rts. The two contour comparison metrics available in this module were conformity index (CI) and mean distance to conformity (MDC). The CI is the ratio of the volume of overlap between two outlines to the volume encompassing the full extent of both outlines. Therefore, two volumes that entirely overlap will have a CI equal to 1 and two volumes that are completely separate will have a CI equal to 0.¹³ The MDC is defined as the mean distance of each outlying voxel from the reference contour. This new metric was developed by Jena et al.¹³ as CI does not provide any information on differences in shape.

Results

For small, low-contrast structures, including the corneas, chiasm, pituitary and cochleae, there was no apparent trend with atlas library size, with a Pearson’s coefficient of 0.2 for the pituitary (minimum value) to 0.6 for the right cornea (maximum value) for CI. The relation of atlas number of MDC was in a similar range. The agreement for these contours was poor in all cases, with a mean CI of 0.11 ± 0.015 (mean \pm SD) seen for the pituitary gland (range 0.09–0.13). The MDC indicates slightly better agreement with a mean of 4.4 ± 0.9 mm (range 1.9–5.3 mm).

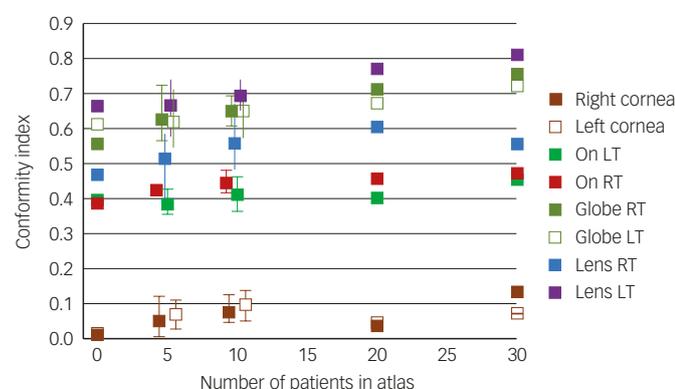
The mandible, brain, parotids and spinal cord shown an improvement in accuracy of the automatically generated contours as the size of the atlas library increases, although there is little difference in the results when going from an atlas library of 20 to 30 patient scans. In fact for the left and right parotids, there is a large improvement in accuracy between one and five atlases and then only minimal improvements

Figure 1: Variation in Mean Conformity Index with Atlas Library Size



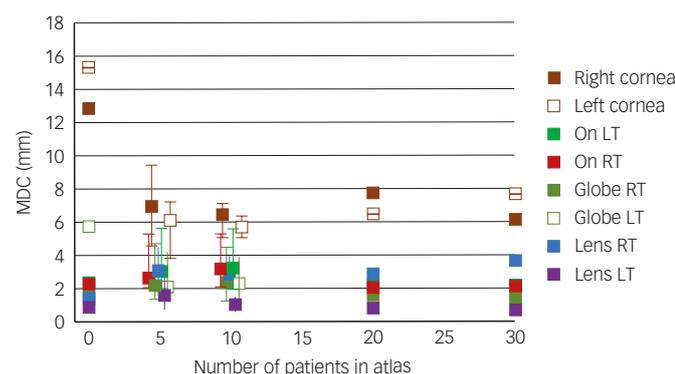
The error bars indicate the range of conformity index seen when different atlas sets were used. LT = left; RT = right.

Figure 2: Variation of Mean Conformity Index with Atlas Library Size for Optical Structures



The error bars indicate the range of conformity index seen when different atlas sets were used. LT = left; RT = right.

Figure 3: Variation of Mean Distance to Conformity with Atlas Library Size for Optical Structures



The error bars indicate the range of mean distance to conformity (MDC) seen when different atlas sets were used. RT = left; RT = right.

beyond that (see Figure 1). The Pearson coefficient is greater than 0.85 for all the organs at risk in Figure 1, apart from the right parotid for

which the value is 0.63. Similar results were seen when comparing the MDC for these structures.

The variation in CI and MDC for optical structures are shown in Figures 2 and 3. Figure 2 appears to show an improvement in accuracy of the automatically generated contours as the size of the atlas library increases for lens, globes and optic nerves, with the Pearson's coefficient ranging from 0.71 (right lens) to 0.99 (left globe) for these structures. However, the plot of MDC (Figure 3) shows no trend between number of atlases and improved conformity, with Pearson's coefficients all below 0.7.

Discussion

The change in number of atlases had the most influence on the automated contours for the brain, larynx and mandible, with the metrics showing improvements in accuracy as atlas size was increased up to 20 cases. Initial atlas selection matches the DRR from the current case to the DRRs of each atlas case, identifying the best match using a mutual information algorithm. Since the position of the brain and larynx are closely related to local bony anatomy, it could be expected that the range of the available atlas cases may affect these structures. Good results were seen for the automated cord contours, regardless of the atlas chosen, most likely due to the effectiveness of the relevant post processing tasks.

For small, low-contrast structures, the results were not as good. This is in agreement with Isambert et al.¹⁴ who found that for atlas-based auto segmentation in the brain, the DSC was low for structures with a volume smaller than 7cm³. In some cases, the metrics indicated a better result when a smaller range of atlases was available. This suggests that optimal atlas selection is not made for these structures. The accuracy for optical structures with atlas size varies. Automated contours for the lens and globes were good, possibly reflecting the strength of the post-processing step for these structures.

The number of test cases used in this study was relatively low due to the time constraints of performing the work while commissioning the system for clinical use. This is a limitation of the study, although we believe it provides useful knowledge when commissioning similar systems.

Conclusion

These results indicate that 20 is an optimal number of atlases to have in the library for auto-contouring head and neck anatomy. If the number of atlases was reduced to 10, most structures would remain unaffected but it may impact on the quality of contours generated for the larynx and mandible.

The selection of atlas cases plays an important role in the quality of contours produced, as the impact of increasing the numbers of atlas cases appears to focus primarily on structures whose position is closely linked to the bony anatomy that drives atlas selection. The use of post processing reduces variation in the accuracy of the contours produced, regardless of the original atlas case selected. Additional work is planned to evaluate the optimal atlas size for pelvic anatomy. ■

- Gambacorta MA, Valentini C, Dinapoli N, et al., Clinical validation of atlas-based auto-segmentation of pelvic volumes and normal tissue in rectal tumours using auto-segmentation computed system, *Acta Oncol*, 2013;52:1676-81.
- Sjöberg C, Lundmark M, Granberg C, et al., Clinical evaluation of multi-atlas based segmentation of lymph node regions in head and neck and prostate cancer patients, *Radiat Oncol*, 2013;8:229-35.
- Mattiucci G, Boldrini L, Chiloiro G, et al., Automatic delineation for replanning in nasopharynx radiotherapy: what is the agreement among experts to be considered as benchmark?, *Acta Oncol*, 2013;52:1417-22.
- Balk S, Weiss E, Jan N, et al., Evaluation of 4-dimensional computed tomography to 4-dimensional cone-beam computed tomography deformable image registration for lung cancer adaptive radiation therapy, *Int J Radiat Oncol Biol Phys*, 2013;86:372-9.
- Fritscher KD, Peroni M, Zaffino P, et al., Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours, *Med Phys*, 2014;41:051910.
- Sharp G, Fritscher KD, Pekar V, et al., Vision 20/20: perspectives on automated image segmentation for radiotherapy, *Med Phys*, 2014;41:050902.

7. Maes F, Collignon A, Vandermeulen D, et al., Multimodality image registration by maximization of mutual information, *IEEE Trans Med Imaging*, 1997;16:187–98.
8. Pluim JP, Maintz JB, Viergever MA, Mutual-information-based registration of medical images: a survey, *IEEE Trans Med Imaging*, 2003;22:986–1004.
9. Thirion JP, Image matching as a diffusion process: an analogy with Maxwell's demons, *Med Image Anal*, 1998;2:243–60.
10. Lu H, Cattin PC, Reyes M, A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons, *Conf Proc IEEE Eng Med Biol Soc*, 2010;5951–4.
11. Vercauteren T, Pennec X, Perchant A, et al., Diffeomorphic demons: efficient non-parametric image registration, *Neuroimage*, 2009;45:S61–72.
12. Wang H, Dong L, O'Daniel J, et al., Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy, *Phys Med Biol*, 2005;50:2887–905.
13. Jena R, Kirby N, Burton K, et al., A novel algorithm for the morphometric assessment of radiotherapy treatment planning volumes, *Br J Radiol*, 2010;83:44–51.
14. 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, *Radiat Oncol*, 2008;87:93–9.