

Reflections on the current status of commercial automated segmentation systems in clinical practice

Jonathan Sykes, BA, MSc, PhD, MIPEM

Leeds Cancer Centre – Medical Physics and Engineering, St James's University Hospital, West Yorkshire, United Kingdom

Journal of Medical Radiation Sciences **61** (2014) 131–134

doi: 10.1002/jmrs.65

Delineation of multiple anatomical structures, both target and organs at risk, is a requirement for the planning of modern radiotherapy techniques such as intensity modulated radiation therapy (IMRT) and volumetric modulated arc radiation therapy (VMAT) as well as for the reporting of dose to structures for correlation with treatment outcome. Manual delineation of the many anatomical structures, particularly target structures which require a radiologist or radiation oncologists input, is time-consuming and therefore expensive. As such, it is one of the major obstacles to meeting the increasing demand for IMRT and VMAT. Automatic segmentation has for a long time held the promise of decreasing the requirement for manual delineation but has been the reserve of research institutions using 'in-house' software solutions.^{1–3} With the recent availability of commercial systems for automatic image segmentation (at least 11 now available⁴), there has been a rapid increase in published articles evaluating the performance of these algorithms for the automatic segmentation of structures in various anatomical sites.^{5–9}

In this issue of the *Journal of Medical Radiation Sciences*, Greenham et al.¹⁰ evaluate the performance of the ABAS system (Elekta AB, Stockholm, Sweden) for delineation of the prostate and pelvic organs. As with other automatic segmentation studies, they report varying degrees of success with some structures being clinically acceptable while others require considerable manual editing. Variation in performance across patients is also observed. Indeed, this is the finding of the author's own institution (not published) where the ABAS system has been implemented for head and neck and prostate patients since 2011. Typically, there is a net reduction in the amount of time taken to delineate the contours if they are first automatically created and then manually edited. In this editorial, the different methods of automatic segmentation are introduced briefly before discussing some of their fundamental limitations.

Methods of Auto-Segmentation

For a good review of automatic segmentation methods the reader should refer to the article by Sharp et al.⁴ To summarise, there are two methods employed in commercial systems. These are either atlas-based segmentation (ABS) alone or in combination with model-based segmentation (MBS). How these are implemented varies from one system to another. The ABAS system used in the study by Greenham et al.¹⁰ uses a combination of both ABS and MBS. The atlas data used by ABAS consist of a number of CT scans each with their own validated set of structures which can be developed for the local institution – a set of atlases. After initial alignment with rigid body registration, deformable image registration is used to deform the CT images of each of the atlas patients to the target patient. The deformation is further refined using a final deformation stage which incorporates some knowledge (or model) of the organs being segmented. The contours from each atlas are then deformed to the target patient using their respective deformation map. In the final stage, the STAPLE algorithm¹¹ is used to combine the multiple structures (for each organ) into a single structure by maximisation of likelihood. This combined contour is, in effect, a consensus of opinion of the contours originating from each of the atlases in the atlas set.

Limitations of Automated Segmentation Methods

There are a number of limitations to this approach which are in common with other commercial systems. The deformable image registration is never perfect for at least the following reasons. (1) Deformable image registration algorithms tend to be constrained (regularised) to avoid locally severe deformations and ensure the global

integrity of the image registration. However, this can mean that any locally severe deformation that does exist in the patient does not get matched well. (2) The contrast in the images between boundaries of organs may also be indistinct leading to errors in the deformation and this can vary from patient to patient, organ to organ and even across an organ. (3) Image registration algorithms have a 'capture range'. If a deformation is over a large distance, for example, deformation between the superior wall of a full bladder and an empty bladder, the image registration algorithm is unlikely to correctly account for this deformation because it is beyond the capture range.

In an effort to reduce the impact of a poor image registration between a single atlas and a target patient, multiple atlases can be employed. Selecting the best atlas would require human intervention and so algorithms such as the STAPLE algorithm¹¹ or majority vote have been introduced. However, these algorithms make their decision based on the set of contours alone, without reference to the underlying image data. This has two effects. If there is a random error between the deformed atlases at a point on the surface of an organ to be segmented, the combined contour may not fall exactly on the edge of the underlying image data. The larger the random error is, then the greater the likelihood of a systematic error. In this case, a larger number of atlases should help minimise the error. However, having more atlases does not always mean better segmentation. Using fill state of the bladder as an example, consider the scenario where the majority of the atlases have a small or medium-sized bladder with just one atlas having a large bladder and with the target patient having a large bladder. For the small/medium bladder atlases the deformation is beyond the capture range of the deformable image registration algorithm, so only the atlas with the large bladder is likely to give a good estimate of the true deformation. A STAPLE or majority vote algorithm will assume the grossly incorrect deformations, the majority, are correct and therefore ignore or give low weighting to the outlier that has good deformation. This illustrates the need for alternative solutions such as intelligent pre-selection of atlases, for example, sub-dividing the atlases into small, medium and large bladder atlas sets. This was alluded to by Greenham et al.¹⁰

Another potential pitfall of the STAPLE and majority vote algorithms, depending on how they are implemented, is that the consensus structures for two abutting structures may be based on a different selection of atlases. This can lead to overlapping of the two structures when the atlases are combined. A final limitation, as described by Sharp et al.,⁴ is that the

combined structures can lead to small islands disconnected from the main structure.

Evaluating Performance of Automated Segmentation

To evaluate the performance of automatic segmentation it is common to compare the automated segmentation with manually delineated contours. A description of metrics for comparing contours is given by Sharp et al. and also in the review by Jameson et al.^{4,12} The dice similarity coefficient (DSC)¹³ is used widely but is very difficult to interpret because (1) its value is dependent on the volume being compared and (2) it gives no indication of the distance between the two contours. One relatively easy to implement solution to this is to use the normalised dice similarity coefficient nDSC,^{14,15} which simultaneously removes the dependence on volume and also attaches a clinical significance to the discrepancy.

Another common metric used to compare contours is the mean distance to agreement. This gives a better indication of the magnitude of the error in terms of distance but can also be misleading in terms of clinical impact. For instance, a contour may be within 1 mm of the reference contour for 90% of its surface area but for 10% of the surface there could be a clinically unacceptable 10 mm discrepancy. The overall mean distance would be 1.9 mm, which could be considered as clinically acceptable even when the 10 mm discrepancy has clinical significance.

Defining whether a segmentation is acceptable is not trivial. Not all anatomical structures are equal and furthermore, not all regions on the surface of an anatomical structure are equal. For instance, the impact of a 2 mm error in the definition of the femoral heads will not have as much impact on the optimisation of a treatment plan for the prostate as it will for the rectum. Furthermore, an error in delineation of the anterior wall of the rectum, where the dose gradients are highest, will have more impact on the treatment plan than an error in delineation of the posterior wall. Defining acceptability on whether the contour is within inter-delineator variation is also problematic. Inter-delineator studies will also demonstrate that the random error can be spatially variant, and this may or may not coincide with where the greatest accuracy is required for dose planning.^{16,17}

The method used in the article by Greenham et al.¹⁰ could be criticised for not using any of the more complex metrics for comparing structures. However, their method of subjectively scoring the discrepancy of the contours inherently compensates for the variation of clinical acceptability of the discrepancy both between organs and across the surface of an organ.

Clinical Implementation of Automated Segmentation Systems

It is worth spending some time validating the atlas contours. An ABS/MBS system will never work if local contouring practices differ from those used to define the atlas or model in the first place. Implementation can be further complicated if there are a variety of contouring practices within an institution. At the author's hospital, ABAS was primed with atlas head and neck contours closely following international consensus guidelines¹⁸; but none of the clinical oncologists were following these guidelines at the time. However, all agreed that they wanted to conform to the guidelines. Consequently, ABAS was used as a tool which facilitated a change in practice. ABS or MBS systems could also be used to facilitate conformance to trial-contouring protocols where there were no previous guidelines.

Automatic Image Segmentation Systems: Are They Worth It?

If an ABS or MBS system is to be effective, then it will need to reduce the time to delineate the required anatomy considerably. Any significant time saving is likely to have a high perceived value by the individual oncologists and radiation therapists who would normally undertake this task. However, in monetary terms, it would take many patients to offset the software and hardware costs of an ABS/MBS system and the cost of the considerable time and effort required to implement the system. For example, at the author's institute, there are ~300 head and neck cases per year which is atypically large compared to most cancer care centres. If ABS/MBS is able to save 50% of the total contouring time (estimated to be 90 min) and the hourly salary cost is \$80 (AUS), then the total saving will be ca. \$18,000. So without accounting for the cost of the resources required to implement the system, it could take up to 2 years before the investment starts to pay dividends. If ABS/MBS can be implemented for multiple treatment sites, then its efficiency savings may be realised sooner.

The Future of Automatic Image Segmentation

There is no doubt that automatic image segmentation is going to play a critical role in future radiotherapy. Accurate and efficient auto-segmentation is going to be a requirement if adaptive techniques, where the plan is adapted to changing anatomy on one or more occasions during treatment or on a daily basis¹⁹ or if 'plan of the day' type techniques, are going to become mainstream.²⁰

In the early days of IMRT, optimisation algorithms in treatment planning software were in their infancy. Consequently, early implementation of IMRT tended to be the domain of large research-oriented centres. Over the years, the software has improved considerably with a corresponding increase in utilisation. Expectations are that automatic segmentation will also improve as deformable image registration algorithms become more refined and tailored to particular anatomical sites. The increase in use of magnetic resonance imaging as a medium for target and normal tissue delineation and in future for image guidance may also increase the accuracy and reliability of auto-segmentation due to its increased soft tissue definition. With these improvements, a corresponding uptake of the use of auto-segmentation techniques is expected.

References

1. Pekar V, McNutt TR, Kaus MR. Automated model-based organ delineation for radiotherapy planning in prostatic region. *Int J Radiat Oncol Biol Phys* 2004; **60**: 973–80.
2. Commowick O, Grégoire V, Malandain G. Atlas-based delineation of lymph node levels in head and neck computed tomography images. *Radiother Oncol* 2008; **87**: 281–9.
3. Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. *Int J Radiat Oncol Biol Phys* 2007; **68**: 522–30.
4. Sharp G, Fritscher KD, Pekar V, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med Phys* 2014; **41**: 050902.
5. Larrue A, Gujral DM, Nutting CM, Kadir T, Gooding M. Impact of the number of available atlases on the performance of adaptive multi-atlas contouring. *Radiother Oncol* 2014; **111**(Suppl. 1): 62.
6. Voet PWJ, Dirx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. *Radiother Oncol* 2011; **98**: 373–7.
7. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F. Performance of an atlas-based autosegmentation software for delineation of target volumes for radiotherapy of breast and anorectal cancer. *Radiother Oncol* 2012; **102**: 68–73.
8. Pirozzi S, Horvat M, Nelson A, Piper J. Atlas-based segmentation: evaluation of a multi-atlas approach for prostate cancer. *Int J Radiat Oncol Biol Phys* 2012; **84**: S799.
9. Sjöberg C, Lundmark M, Granberg C, Johansson S, Ahnesjö A, Montelius A. Clinical evaluation of multi-atlas based segmentation of lymph node regions in head and

- neck and prostate cancer patients. *Radiat Oncol* 2013; **8**: 229.
10. Greenham S, Dean J, Fu K, et al. Evaluation of atlas-based auto-segmentation software in prostate patients. *J Med Radiat Sci* 2014; **61**: 151–58.
 11. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 2004; **23**: 903–21.
 12. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. *J Med Imaging Radiat Oncol* 2010; **54**(5): 401–10.
 13. Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945; **26**: 297–302.
 14. Ezhil M, Choi B, Starkschall G, Bucci MK, Vedam S, Balter P. Comparison of rigid and adaptive methods of propagating gross tumor volume through respiratory phases of four-dimensional computed tomography image data set. *Int J Radiat Oncol Biol Phys* 2008; **71**: 290–6.
 15. Speight R, Sykes J, Lindsay R, Franks K, Thwaites D. The evaluation of a deformable image registration segmentation technique for semi-automating internal target volume (ITV) production from 4DCT images of lung stereotactic body radiotherapy (SBRT) patients. *Radiother Oncol* 2011; **98**: 277–83.
 16. Meijer GJ, Rasch C, Remeijer P, Lebesque JV. Three-dimensional analysis of delineation errors, setup errors, and organ motion during radiotherapy of bladder cancer. *Int J Radiat Oncol* 2003; **55**: 1277–87.
 17. Fuller CD, Nijkamp J, Duppen JC, et al. Prospective randomized double-blind pilot study of site-specific consensus atlas implementation for rectal cancer target volume delineation in the cooperative group setting. *Int J Radiat Oncol Biol Phys* 2011; **79**: 481–9.
 18. Grégoire V, Levendag P, Ang KK, et al. CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC RTOG consensus guidelines. *Radiother Oncol* 2003; **69**: 227–36.
 19. Speight R, Harding R, Lindsay R, et al. Evaluation of atlas based auto-segmentation for head and neck target volume delineation in adaptive/replan IMRT? *ICCR* 2013; **2013**: 0–13.
 20. Burrige N, Amer A, Marchant T, et al. Online adaptive radiotherapy of the bladder: small bowel irradiated-volume reduction. *Int J Radiat Oncol Biol Phys* 2006; **66**: 892–7.

Jonathan Sykes

Leeds Cancer Centre – Medical Physics and Engineering

St James's University Hospital

West Yorkshire, United Kingdom

Tel: +44 113 2067921

Fax: +44 113 2068941

E-mail: sykesjonr@gmail.com