REVIEW ARTICLE

Automated delineation of radiotherapy volumes: are we going in the right direction?

¹G A WHITFIELD, FRCR, PhD, ²P PRICE, FRCR, ³G J PRICE, PhD and ³C J MOORE, PhD

¹Radiotherapy Related Research, University of Manchester, Manchester, UK, ²Department of Surgery and Cancer, Imperial College, London, UK, and ³North West Medical Physics, Christie NHS Foundation Trust, Manchester, UK

ABSTRACT. Rapid and accurate delineation of target volumes and multiple organs at risk, within the enduring International Commission on Radiation Units and Measurement framework, is now hugely important in radiotherapy, owing to the rapid proliferation of intensity-modulated radiotherapy and the advent of four-dimensional image-guided adaption. Nevertheless, delineation is still generally clinically performed with little if any machine assistance, even though it is both time-consuming and prone to interobserver variation. Currently available segmentation tools include those based on image greyscale interrogation, statistical shape modelling and body atlas-based methods. However, all too often these are not able to match the accuracy of the expert clinician, which remains the universally acknowledged gold standard. In this article we suggest that current methods are fundamentally limited by their lack of ability to incorporate essential human clinical decision-making into the underlying models. Hybrid techniques that utilise prior knowledge, make sophisticated use of greyscale information and allow clinical expertise to be integrated are needed. This may require a change in focus from automated segmentation to machine-assisted delineation. Similarly, new metrics of image quality reflecting fitness for purpose would be extremely valuable. We conclude that methods need to be developed to take account of the clinician's expertise and honed visual processing capabilities as much as the underlying, clinically meaningful information content of the image data being interrogated. We illustrate our observations and suggestions through our own experiences with two software tools developed as part of research council-funded projects.

Received 15 September 2011

Revised 20 August 2012

Accepted 3 September 2012

DOI: 10.1259/bjr.20110718

© 2013 The British Institute of Radiology

In the last two decades radiotherapy has become multidimensional and dynamic. The innovations of three-dimensional conformal radiotherapy and intensity-modulated radiotherapy (IMRT) have allowed radiotherapy dose distributions to be shaped and reshaped much more closely to targets, allowing better avoidance of organs at risk (OARs) and hence reduced normal tissue toxicity, and facilitating dose escalation to the target volume(s). However, such precise treatments require that the target volumes and OARs are accurately outlined in line with the constructs introduced by the International Commission on Radiation Units and Measurement (ICRU) with ICRU-50 and subsequently revised guidelines [1–3]. Several decades have elapsed since the objective of computerised segmentation was

conceived, and we are still a long way from successfully achieving “automation”. Indeed, most radiotherapists still have no option but to delineate manually, which is time-consuming, particularly for multiple disease target volumes and OARs (e.g. for IMRT to head and neck tumours). Reducing greyscale images to ICRU compatible structures is not so easy given that they do not necessarily have a continuously identifiable “boundary”. Hence, what can actually take place is “outlining” to define an acceptable region of interest. The result is dependent on the operator/clinician and so susceptible to interobserver variation and error. To some extent this can be masked by the considerable convenience of the ICRU “tolerance” margins. Nevertheless, these errors can be the greatest source of geometric uncertainty in radiotherapy [4].

There is an assumption that the image data being worked on are always fit for purpose, regardless of the circumstances of the acquisition. This is an important issue given that most delineations are performed at soft-tissue sites where structure visibility is already severely limited by the physical laws of X-ray–tissue interactions and the potentially diffuse, irregular nature of structures of interest. Additionally, this is exacerbated by the

Address correspondence to: Professor Pat Price, Department of Surgery and Cancer, Imperial College Division of Surgery, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK. E-mail: p.price@imperial.ac.uk

Development of early versions of GeoCut and GeoCut3D was supported by EPSRC grant EP/D077702/1. SCULPTER was supported by EPSRC grant GR/S41340/01. GAW was supported by a Cancer Research UK McElwain clinical research training fellowship.

concern to constrain X-ray exposures to patients, and the inherent limitations of gathering fan-beam CT (FBCT) data in a slice-by-slice fashion in the presence of physiological motion. Four-dimensional (4D) image-guided adaptive radiotherapy, which uses analysis of on-treatment imaging to adapt the treatment plan, is expected to be the future standard of care [5]. This may require cone-beam CT (CBCT) volumes to be routinely delineated on several acquisitions during treatment, as well as on the initial radiotherapy planning (RTP) FBCT scan. Consequently, speed of delineation will assume greater importance. On-treatment CBCT images may be of still lower quality than the RTP FBCT, making delineation for comparative purposes even more difficult and error-prone. Hence, fast and accurate segmentation tools, not only operating *de novo* on image volumes such as RTP FBCT but also for propagating segmentation between closely related but poorer quality images (e.g. the different motion phases of a sorted 4D FBCT scan, or on-treatment CBCT images), are potentially invaluable. It is therefore not surprising that more automated methods of delineation or segmentation, with the potential to reduce interobserver variation and/or speed delineation, are the subject of intense research, with some tools beginning to find application in areas of radiotherapy. Occasionally some of these approaches can get significant parts of a delineation right, but that does not mean much when the non-trivial parts are wrong and need extensive editing. We note that such editing can be no more than a palliative for the deficiencies of a segmentation approach, since by definition they do not address the root causes of failure where it matters most. The reality is that a significant divergence between expert performance (which remains the “gold standard”) and machine-assisted approaches to segmentation persists. Likely reasons for this are explored in this paper, with a view to stimulating advances via a shared understanding of the potential obstacles to progress.

Conceptual approaches to medical image segmentation

It is some 20 years since ICRU structures were systematically introduced and in practice the clinical expert remains the gold standard defining them. From this we must deduce that for the moment the trained clinician remains an indispensable image-processing “component” for applications that are non-trivial. Failure to incorporate broad-based prior knowledge is likely to be one reason for the lack of progress; another is the failure to appreciate how expert image interrogation proceeds when looking for embedded structure in what after all are simply greyscale arrays reflecting very narrow physical properties. Hence, we illustrate our thoughts by considering basic computerised greyscale interrogation with no prior knowledge and atlas-based approaches, where extensive prior statistical knowledge of greyscale content assists segmentation—both functioning, at least in principle, without expert assistance. To this we add expert image interrogation operating synergistically with shape model assistance to embody subjective and objective prior knowledge of structure, intentionally without greyscale processing, which is

delegated to the observer. Some of the operational assumptions that underpin these approaches will be considered in more detail in the following sections, drawing on our own particular experiences:

- image content or greyscale alone
- atlas-based segmentation tools
- shape model-based segmentation.

Image content or greyscale alone—segmentation without prior knowledge

Tools that segment based purely on greyscale content of the image use a number of different approaches. Where tissue changes are so large that we can have confidence in the specificity of the imaging modality, simple intensity-based segmentation may be suitable for delineation of very high- or low-contrast areas, such as the lungs or bony structures, where boundaries are essentially so well defined they are trivial. Some algorithms may incorporate edge detection to cope with less obvious scenarios; for example, approaches that we have developed include initial boundary seeding followed by adaptive gradient characterisation at four points either side, which is particularly useful for head and neck contour generation plagued by cavities with severe partial volume effects. Some algorithms need multiple *seeds* as definitive starting points to indicate the regions to segment. For example, “random walker” developed by Leo Grady at Siemens’ Princeton Research Laboratory (Siemens AG, Munich, Germany) uses *k* seeds to initialise *k* different objects, and then segments the image according to which of the seeds a random walker starting at each unlabelled pixel has the highest probability of reaching [6]. Random walker is biased to avoid crossing high-intensity gradients (boundaries) and respects “weak” boundaries (boundaries with sections missing).

Unfortunately, even weak organ or tumour boundaries may be lacking. What to the surgeon might be palpable disease or an obvious tissue boundary has to be interpreted purely in terms of the image data—in the case of a CT image, a map of X-ray attenuation coefficient estimates through the patient. The differences in X-ray attenuation between various soft tissues are small, as are those between soft tissues and many tumours comprising rogue cells with the same or similar physical attributes.

Atlas-based segmentation—greyscale with prior knowledge

In atlas-based segmentation, deformable image registration is used to register the image volume to be segmented to a reference image volume (the *atlas*) that has already been segmented, or vice versa. A particular voxel in the new image is then simply assigned to the structure, which it is deformed onto in the atlas.

The segmentation of the reference image volume could be done by any method, including manual delineation; the accuracy with which the reference image is segmented affects the results, so that commercial products may

include segmentation by “expert” clinicians. The success of a particular deformable registration will depend on the algorithm used and the selected atlas, and can be assessed by means of a *similarity measure*, for example the popular *mutual information* (MI) measure. However, the MI measure is a global measure of similarity and does not give any spatial information to indicate image regions where the deformable registration may be less than satisfactory. (In fact all three of the segmentation approaches discussed provide deterministic output without reporting confidence measures in the points said to form or lie within a boundary.) Anatomical variation, which is very common, and vanishing volumes (e.g. rectal gas in prostate radiotherapy) may be problematic for atlas-based segmentation tools. For this reason, selection of the best atlas is important; it may be selected based on some criterion, such as the amount of deformation needed to match the image volumes, or the similarity measure after deformable registration. Rohlfing et al [7] provide a readable account of atlas-based segmentation.

Commercial offerings are beginning to appear. For example, to reduce segmentation inaccuracies, multiple subatlases have been explored within the recent European Union Methods and Advanced Equipment for Simulation and Treatment in Radio Oncology (MAESTRO) project and implemented in the French “Dosisoft” Atlas Based Automated Segmentation (ABAS) RTP package. Elekta’s ABAS and Varian’s IKOE (Varian Medical Systems, Inc., Palo Alto, CA) knowledge-based segmentation are also available. A number of papers show some success at propagating outlines from one image volume to another, for example between the phases of 4D CT, such as shown by Speight et al [8] using Elekta’s ABAS, or between RTP and on-treatment images; in this case the “atlas” is patient specific and is more likely to be a close match to the patient’s anatomy.

Shape model-based segmentation

Shape model-based segmentation tools use prior knowledge to perform constrained segmentation of individual structures. All shape models make a significant assumption that there is a generic underlying structure to be gleaned from the available data. The presence of abnormalities can present a challenge in this context. The approach to model construction can take varying forms; for example, it may be based on hierarchical addition of detail through principal components analysis (PCA) of prior delineations. Adaptation of a model to the particular image volume is usually greyscale-based.

For example, in “Model-Based Segmentation” in Philips Pinnacle³ TPS v8.0 (Philips BV, Eindhoven, the Netherlands), the starting point for a delineation is a mesh surface model based on PCA of prior delineations. The adaptation algorithm extends an *active snake* approach into an *active mesh*, in which mesh evolution is driven by the greyscale and constrained by the shape model [9, 10]. The algorithm uses a patently artificial but nonetheless conveniently tractable concept of energy, which is drawn from mainstream physics, minimising

the sum of a construct termed “external energy”, dependent on a notional “attraction” of the boundary to image features, and an “internal energy”, which penalises deviation from the statistical shape model. The “attraction” can be driven by slope, which has an intuitive appeal, as well as a simple mathematical representation through the greyscale gradient. Editing tools are provided to correct the segmentation.

RaySearch (RaySearch Laboratories, Stockholm, Sweden) and Siemens are also developing model-based tools based on PCA of prior delineations and greyscale adaptation. Siemens uses initial model-based segmentation of prostate, bladder, rectum or femoral heads as a starting point for its random walker. Varian (Varian Medical Systems Inc., Palo Alto, CA) is developing its equivalent, Smart SegmentationTM. Accuray’s AutoSegmentationTM product (Accuray Inc., Sunnyvale, CA) uses model-based segmentation to delineate prostate, seminal vesicles, rectum, bladder and femoral heads. Each of these companies is gradually adding to the body sites covered and working to improve the quality of segmentation, to reduce the amount of editing required.

Examples of difficulties in automated segmentation

Rather than discussing the limitations of different conceptual approaches to segmentation only in the abstract, it will be helpful to give some concrete illustrations. In this section, therefore, we will use examples from our own experiences with software tools developed at the Christie NHS Foundation Trust (Manchester, UK) and by our research partners at the University of Central Lancashire (Preston, UK) to illustrate some practical barriers and difficulties, which are by no means limited to these particular software tools. The first example—GeoCut, a purely intensity-based segmentation tool (*i.e.* of the first type mentioned above)—illustrates limitations that may be encountered when any prior knowledge about the structures to be segmented is lacking. The second example—SCULPTER, which, to our knowledge, is unique in combining subjective and objective shape modelling without greyscale input (*i.e.* of the third type above)—demonstrates limitations that could be improved by incorporating greyscale input.

Intensity-based segmentation

GeoCut and GeoCut3D are the two-dimensional (2D) and three-dimensional (3D) versions of image intensity-based (greyscale-based) delineation software developed by Dr Matuszewski’s group within the Applied Digital Signal and Image Processing laboratory at the University of Central Lancashire. GeoCut works with single bitmap images (e.g. individual CT slices), while GeoCut3D uses Digital Imaging and Communications in Medicine (DICOM) image sets (e.g. an entire CT scan) treated as a contiguous volume; it was hoped that GeoCut3D would avoid the need for laborious slice-by-slice delineation.

After specifying *N* structures and assigning each a colour, the user marks areas of the image belonging to

the N objects by means of brush strokes. "Run" is then selected to see the segmentation results as a colour mask over the base image. The ratio of base image to colour mask can be varied by a slider. The segmentation can be refined by adding brush strokes and rerunning until satisfied with the result. The principle underlying GeoCut and GeoCut3D is as follows. After the user has marked parts of the N objects, for each point (pixel or voxel) x in the image—where x is a vector representing (x, y) in 2D or (x, y, z) in 3D—a "distance" $D_i(x)$ ($i=1, \dots, N$) of x from each of the N specified objects is calculated. Point x is considered to be a part of object j if and only if $D_j(x) < D_i(x)$ for $i=1, \dots, N$ ($i \neq j$). The "distance" between x and a particular object depends on the gradient of image intensity along a path connecting the two. Thus, if there is a curve that connects point x and object j and does not traverse any areas of high image gradient (typically organ boundaries), then point x will be considered part of object j , if this "distance" is less than the distance to all the other objects. The novelty of GeoCut and GeoCut3D lies in the intuitive way in which the user interacts with the segmentation. Most other greyscale-based methods rely on fully automatic segmentation after the user specifies an initial contour; if the segmentation is unsatisfactory, the user can generally only try again using a different starting contour or tweak some parameter, which does not have any direct intuitive meaning, in the optimisation algorithm, without any certainty of improvement.

We tested GeoCut on 12-bit planning CT images and GeoCut3D on a public DICOM image set (ENTERIX abdo arterial from <http://pubimage.hcuge.ch:8080>), initially using every fifth slice, and subsequently every second slice, from the 1-mm slice-spaced FBCT sequence. GeoCut was good at segmenting high-contrast

areas with few brush strokes, for example distinguishing body from background, and for segmenting the lungs and bony structures such as the vertebrae and ribs (Figure 1a). (Such structures are also relatively easily thresholded on RTP and CBCT scans using established commercial software.) GeoCut3D also allowed rapid segmentation of the background (air), body and lungs, but difficulties were encountered with the vertebrae; parts of vertebrae were not included, and there were regions where the segmentation "spilled" into adjacent lower density areas (Figure 1b). Segmentation of soft-tissue structures was much more laborious. In 2D, multiple brush strokes were needed to segment the oesophagus and aorta (Figure 1c). Spicular errors (needle-like projections) and "spill" from the intended area into connected areas of similar intensity occurred and were difficult to limit. In order to segment the left kidney and the liver, numerous marks were required to separate them from adjacent structures of similar density (Figure 1d). In 3D, the "spill" into adjacent regions of similar density was much worse than in 2D as the number of paths for leakage increased. For example, the aorta, which was visually well demarcated, was poorly segmented. By placing multiple brush strokes on a particular slice, it was possible to rectify the segmentation on that slice, but on other slices the segmentation still spilled into adjacent tissues, necessitating further intervention on those slices. GeoCut3D did perform somewhat better with more closely spaced CT sections (effectively 2 mm) than with more coarsely spaced sections (effectively 5 mm), but the results were still clearly unacceptable for clinical use; moreover 2-mm slice spacing would not generally be used for RTP scans of the chest and abdomen owing to the higher radiation dose.

Figure 1. (a) High-contrast areas such as the body (green), lungs (blue) and bones (yellow) were easily and accurately segmented in GeoCut. (b) Vertebrae were incompletely segmented in GeoCut3D and the segmentation "spilled" into adjacent areas. (c) Soft tissues such as the ascending and descending aorta (brown) and oesophagus (orange) were less accurately segmented in GeoCut, requiring many brush strokes. (d) Many brush strokes were needed to segment the liver (orange-brown) and the left kidney (purple) in GeoCut, especially where they abutted tissues of similar density.

In summary, GeoCut and GeoCut3D were unsuitable for delineating all but the highest contrast structures, owing to insufficient intensity variation between abutting areas to be segmented. Dr Matuszewski's group has further developed the GeoCut3D approach, particularly by applying smoothing to ensure that segmented objects do not suffer from spiky protrusions, and claim some success in prostate and bladder delineation [11]. However, these examples illustrate that differences in greyscale between tumours or OARs and adjacent tissues are often limited or absent; it seems likely that any greyscale-based tool for automating delineation would need to incorporate a level of shape knowledge.

Shape model-based segmentation

We illustrate the shape model-based segmentation approach using SCULPTER (structure creation using limited point topology evidence in radiotherapy), a novel tool for computer-assisted delineation, developed at the Christie NHS Foundation Trust (EPSRC grant GR/S41340/01) [12]. SCULPTER was devised to be faster and more accurate than manual delineation (*e.g.* eliminating "crinkle-cut" slice-to-slice jaggedness) and is intended especially for large, effectively contiguous image volumes, such as spiral CT or CBCT. SCULPTER embodies prior knowledge about the organ to be delineated in a shape model. Uniquely, adaptation of the model is based on sparse points marked by the observer rather than dense greyscale data, the philosophy being that a clinician can better decide a few very definite boundary points than a solely mathematically based greyscale algorithm, especially where image quality is less than perfect, as with CBCT.

A prototype SCULPTER version used a simple primitive structure as the shape model (*e.g.* an ellipsoid for a bladder, or a cylinder for the rectum). This would be deformed onto 10–20 anatomically clear organ boundary points specified by the clinician. Using this prototype, SCULPTER bladder volumes drawn by a single observer familiar with SCULPTER on either MRI or CT corresponded closely to manually drawn volumes, while SCULPTER was statistically significantly faster on both CT (2.5×) and MRI (4.2×) [13]. In a multiobserver study using CT imaging, bladder and prostate volumes corresponded closely whether drawn manually or by SCULPTER, while SCULPTER was statistically significantly faster for prostate volumes [13]. Although this version accelerated contouring, it did not reduce interobserver variation.

The current SCULPTER version uses as a starting point a PCA-based statistical shape model describing the shape and principal modes of shape variation of previously outlined organs of that type [14]. The model is able to interpolate or extrapolate from the delineations used to build it to new cases. The user selects 10–20 points representing boundary points of which he/she is very certain. SCULPTER calculates a surface passing through these and uses the shape model to fill in the rest. Importantly, the user-defined points can be selected in any order and in any orthogonal plane, which is particularly useful for CBCT data, which are acquired as a continuous volume, rather than as slice data for conventional CT. By contrast, for conventional manual

outlining, points usually have to be placed in a connected fashion in one plane (typically axial) and a contour closed on each slice before moving to another slice, forcing the outliner to commit—even if some points are little more than guesses. The current version of SCULPTER was tested for bladder delineation on 10 serial CBCT scans for a prostate cancer patient. The shape model was primed using 37 bladder delineations over 16 serial CBCT scans of a second patient. 3–7 observers performed both manual and SCULPTER delineations on the 10 serial CBCTs. The interobserver variability was shown to be less for SCULPTER than for manual delineations [14]. We also tested the current SCULPTER version for delineation of oesophageal and oesophagogastric junctional (OGJ) tumours. In this case SCULPTER was primed using oesophageal and OGJ tumour outlines drawn on one RTP and three CBCT delineations per patient for each of nine patients by each of two clinicians. SCULPTER was unsuitable for this purpose; usually parts of the SCULPTER outline differed by several millimetres from the obvious organ boundary. An example of the typical difficulty encountered is shown in Figure 2. Many points had to be added to try to force the boundary to an acceptable position; adding points in the area where the boundary was poor resulted in new deviations in other areas. Particular difficulties were encountered in trying to close the superior and inferior extents of the volume; very many points had to be placed to try to force the boundary.

Thus the SCULPTER approach appears to be suitable for delineating whole structures, such as the bladder or prostate, which are usually treated in their entirety even though the tumour occupies only a portion. Similarly, it should be suitable for OARs such as the bladder, kidney, or liver, which again would be entirely delineated. It may also be suitable for outlining OARs that form part of an organ, provided that is a constant portion. For example, it may be suitable for delineating the rectum as an OAR for prostate RTP, provided that constant superior and inferior borders are chosen for delineations (both those used to prime it and those drawn using SCULPTER). In all these cases, it should be feasible to build a shape model describing the modes of shape variation. Similarly, a shape model for the whole of the *normal* oesophagus is possible, given the typical curving path as it skirts organs in the chest. However, the shape modelling approach generally is unsuitable for partial organ delineation where the part delineated is inconstant, or a focus on segmentation of diseased areas, since the assumption of an underlying shape model is then flawed.

Since SCULPTER's shape model fills in areas of the boundary of which the clinician is uncertain, it might be expected to reduce interobserver variation. However, even if the filled in areas are less variable between observers, they will not necessarily be more accurate if the prior delineations were also consistently based on guesses in these areas. SCULPTER may therefore be particularly valuable for delineating on images of poor quality, such as CBCT, where the prior knowledge is of higher quality, for example based on CT. Similarly, SCULPTER might prove of value for delineating OARs on CT where more reliable prior knowledge is available from MRI.

As mentioned, the philosophy of the SCULPTER approach is that the clinician can determine boundary points more reliably than a greyscale algorithm,

Figure 2. Example of delineation of an oesophageal tumour via SCULPTER (structure creation using limited point topology evidence in radiotherapy), a tool for computer-assisted delineation developed at the Christie NHS Foundation Trust (Manchester, UK). The clinician-marked points are shown as (a, b) spheres and (c–e) + signs. (a) The initially fitted surface and (b) the refined surface, passing through all clinician-marked points, are shown. (c) Axial, (d) sagittal and (e) coronal views on the planning CT scan show that the SCULPTER surface clearly deviates from the oesophageal boundary.

particularly where image quality is poor. However, it is apparent from Figure 2 that even a relatively simple greyscale algorithm might improve on the way parts of the boundary between high-contrast regions (e.g. lung/soft tissue, or soft tissue/bone) are filled in between the very definite user-defined points, and might reduce the user interaction required. Hybrid approaches that utilise prior knowledge, make sophisticated use of greyscale information and allow clinical expertise to be integrated are needed.

Human image interrogation

It is instructive to consider aspects of how humans analyse images, since image analysis by expert clinicians remains the gold standard in critical medical applications, ranging from microscale histological assessment of tumours to macroscale image-guided delivery of

therapeutic radiation sufficient to sterilise tumours. The output of such image interrogation is often objective, in the form of measurements recorded in a precise manner and, so far as radiation therapy is concerned, direct overlays onto the digital image(s) in question. Some 30 years after digital imaging facilities became widely available in medical practice in the form of CT X-ray reconstruction sequences, machine-assisted feature extraction and segmentation algorithms are still judged against or even modelled from expert clinician output, even though the operational processes and content basis for that output are still poorly understood. Of course clinical experience can also be wrong—we cannot say “clinicians know best” without supporting that with evidence—but it is instructive to consider the degree of sophistication present in the human visual system, especially in relationship to the trained expert.

Human image interrogation involves optical processing coupled with retinal transduction, neural analysis and

interpretation by the observer. In other words, human image analysis is not merely a question of optics; the visual system includes its own image processing, historically referred to as a “psychophysical” component (*i.e.* the brain processes visual images for optimum analysis) [15, 16]. For delineation of radiotherapy structures, the input to the visual system is a digital image array, with some degree of noise, viewed using a display device. Expert clinicians do not give equal attention to all parts of an image; in fact the human visual system is specifically adapted for the detection of edge detail and discontinuities, which is one form of a larger set of features we recognise as “pattern”. The Mach band edge effect [17] is the basis of many optical illusions and direct evidence of edge enhancement performed for vision-related tasks. Edge enhancement is largely performed in the retina owing to a process called “lateral inhibition”. Retinal neurons that are most strongly stimulated fire, while sending out inhibitory signals to adjacent neurons, thus increasing the perceived image contrast at the boundary. However, edges are often less well defined than we would like. In outlining OARs for RTP, the concept of a definite organ boundary may be a reasonable one, although delineation is hampered by the fact that abutting tissues or organs may have very similar CT densities. The situation is more complex for a tumour, which grows by infiltration into and degradation of the surrounding tissues, making its periphery appear diffuse. Here it might be useful to draw an analogy with a drop of milk entering a glass of water, starting with a very definite boundary then progressing to an ever more diffuse boundary as it disperses into the main bulk of the water. Hence, it is possible for a tumour to be entirely indistinct in CT scans because diseased and healthy tissues share the same physical and chemical properties. Similar comments apply to other imaging modalities, which highlights the fact that there is still no entirely cancer-specific imaging technique. Even functional imaging like positron emission tomography is lacking in specificity. However, in the situation where a distinct edge is difficult to discern, or poorly defined, humans are able to analyse image data in a much more sophisticated way than simple edge detection. For example, the clinician may detect that the texture of one region looks different to another. In homing in on suspect features, she/he may iteratively alter the greyscale window width and level until subtle features of interest are identified. By observing the effect of changing greyscales, the clinician may even utilise the temporal part of the human visual response [18]. Finally, internally generated noise in the human visual system could be playing a role in expert image interrogation. It has been known for some time that evidence for an underlying signal in an otherwise variable data set can be produced by the addition of moderate levels of internal noise, which boosts parts of the signal above the visibility threshold in a process referred to as stochastic resonance (SR) [19]. Indeed, the basic SR approach has recently been explored for tumour segmentation in X-ray mammograms [20, 21]. Recent work by Aihara et al [22] suggests that some observers benefit more from internal noise SR than others, which may help to explain interobserver variation.

In summary, humans can recognise pattern in all its contexts to determine the probable organ or tumour

boundary. We are equipped to be able to see pattern even in what would commonly be termed noise, but is in fact chaotic, with threshold levels of structure rather than pure randomness. Use of digital window controls may help to isolate pattern evidence from which we can deduce the probable boundary position between regions that barely differ in terms of their simple greyscale content.

Where are mathematical approaches lacking?

Greyscale-based, shape model-based and atlas-based approaches all work best where there are clear differences in greyscale between abutting tissues or organs; this is often not so. As mentioned, automated tools that segment based on greyscale intensity generally focus on some method of edge detection. Model-based segmentation methods also often focus on edge detection for the adaptation of the model to the particular image data. This focus on edge detection may be in part due to the origin of current image-processing techniques in engineering applications, where edges are usually very definite, rather than to medical applications. Often other local image information, such as differences in texture between regions, is not optimally used in inferring boundary position. Where boundaries are indistinct or absent, manual delineation can operate indirectly, by inference (*e.g.* by noting unusually displaced normal anatomy), rather than by direct observation of the OAR or target disease itself. Such interacting organs and tissue ensembles, particularly those that are diseased, are embedded in the expert via extended training that cannot yet be adequately translated computationally. As mentioned, the shape model-based approach is only suitable for delineation of whole organs (which might be tumour-bearing; *e.g.* the prostate) or constant parts of an organ, and is not suitable for delineating only a tumour or diseased portion of a tissue. The same in fact applies to the atlas-based approach, since the reference atlas would in general lack a corresponding tumour volume in that particular location.

What can be done now?

The ability to delineate accurately is clearly trainable to some degree. Radiologists typically use multiple imaging modalities in interpreting the extent of disease. In RTP, clinicians typically outline on only one or two types of imaging (*e.g.* a CT and a particular MRI sequence), although a greater range of imaging may be referred to. There is scope to improve manual delineation by education regarding normal anatomy and different imaging modalities, and possibly also by better software to integrate information from different modalities. Furthermore, inter-observer variation in manual delineation may be reduced to some extent by the use of delineation protocols, which may specify a number of factors such as additional imaging modalities and the CT window level and width to be used for delineation (*e.g.* [23]).

Current methods of automated delineation are useable in certain situations, which ideally would be more clearly defined by clinicians. Clinicians might define that a certain technique yields clinically acceptable results on

images of a certain kind for OAR delineation or for the delineation of particular tumour types. For example, segmentation tools that are useable with high-quality CT images may not yield acceptable results on CBCT images. The limitations should be clearly understood, and may provide stimulus for development of new techniques.

Future directions

Fundamentally, there is a need for greater development and exploration of new methods in image processing suited to medical image segmentation. To date, there has been relatively little diversity in the mathematical approaches explored, which have often focused on techniques developed for engineering applications. These approaches have not given significant weight to clinical experience, which ideally would be captured and extended to reduce interobserver variation, as well as to speed delineation. Semi-automated techniques, which allow for expert user interaction, may prove more efficient in the medium term than fully automated segmentation, which all too often leaves the user to correct a deficient segmentation using only basic editing tools such as are used for manual delineation. Such research may benefit from involvement of researchers from a variety of disciplines, including those using signal and image processing in other fields, as well as from clinician input.

Two examples of techniques that could be further explored are "phase congruency" and "regularity statistics" such as "approximate entropy" [24]. The fundamental importance of image phase content has long been known, and was clearly demonstrated in 1981 [25]. In 1996 Kovessi [26] showed that at points along edges the Fourier transform phase components are coherent or "congruent". That is to say, when the image is examined in the frequency domain (*i.e.* represented as a superposition of sinusoidal waves), the waves are in phase (*i.e.* their phase coincides or is "congruent") at image boundaries. One advantage of the phase congruency method over gradient-based methods of edge detection is that the phase congruency is dimensionless and therefore invariant to changes in scale, such as image magnification. This means that a universal threshold value can be set to denote a boundary, applicable to many image types. In 2009 a development of Kovessi's phase congruency algorithm was reported by Szilagyi and Brady [27], and used to emphasise already relatively clear boundaries in microscale tumour vessel and macroscale pre-clinical pancreatic cancer ultrasound images. However, at blurred or diffuse edges Kovessi's phase congruency is diminished, the significance of which is unclear. The same is true for the output from the classic image gradient method. In the presence of noise both are problematic. This is important, since medical images are invariably noisy, not least because of ethical and legal constraints on the level of imaging agent that can be used on human subjects—radiation being the most emotive. Yet these are the less than ideal images used daily by experts, who find signs of structure even when it is embedded in noise. This is not a search for the simplistic edge, but for structure that exhibits enough regularity to be distinctive.

"Regularity statistics" might provide an appropriate metric of image structure for practical use. For regularity

statistics, the spatial ordering of the data matters. "Approximate entropy" is one such statistic, which measures the regularity or pattern within data at a local level, producing low values where internal structure is strong and high values where it tends towards being random. One of us (CM) used this for one-dimensional voice signal processing to establish a reference standard for normal voicing, quantified using a single metric, and deployed this to objectively monitor voicing recovery in larynx cancer patients after radiation therapy [28, 29]. One of us (CM) extended the concept of approximate entropy to two and three dimensions in order to assist clinical experts to interrogate 2D and 3D image volumes (EPSRC grant EP/H024913/1 "Technology in Radiotherapy Feasibility Studies") [30]. An approximate entropy value is calculated at each pixel (or voxel) location in an image, to give a measure of the regularity or pattern of image data within a defined distance of that pixel or voxel. Image data that are absolutely regular and predictable (*e.g.* uniform grey levels or data repeating spatially in a predictable pattern) have low approximate entropy; less regular data (*e.g.* greyscale data containing a boundary) have higher approximate entropy; and uncorrelated data (noise) have very high approximate entropy. When tested on CBCT images, of lower quality than typical RTP images, this appears to have the potential to provide clinicians with calibrated evidence of structural importance at points in an image. We envisage that this could provide supporting evidence to guide clinicians in semi-automated delineation, thereby potentially reducing interobserver variation.

Ideally a certain minimum standard of image quality should be insisted upon for delineation purposes. Although expert clinicians can readily appreciate image quality, it would be valuable to develop metrics to quantify this. Such metrics might be used, for example, to assure the quality of every clinical image, or in clinical trials to specify a certain minimum image quality for delineation, as well as specifying a delineation protocol and/or segmentation tools which should be used. We have explored the use of the 2D and 3D approximate entropy concepts mentioned above to measure image quality, which is feasible at a local level, allowing for example local comparison of the quality of CT scans with different qualities of CBCT. However, there is a need for metrics of global image quality.

Conclusion

Research is needed to develop new image-processing techniques and software tools that are better suited to medical image segmentation than currently available methods, in order to speed delineation and reduce interobserver variation. Methods are needed which are useable even when boundary information is barely discernible or absent, and which are suitable for use with on-treatment images, as well as with higher quality images used for radiotherapy treatment planning. Since delineation by expert clinicians remains the gold standard in terms of accuracy, the assumptions underlying automated segmentation approaches should be revised from the standpoint of clinical expertise rather than taking their lead from mathematical or scientific expediency. Techniques should aim to capture expert

clinical experience, as well as utilising prior knowledge and making sophisticated use of greyscale information. Semi-automated rather than fully automated techniques may be best suited for this purpose. Ideally, such techniques would graphically indicate confidence limits on the calculated boundary position. Metrics of image quality would be valuable to help assure the quality of images for delineation purposes.

Acknowledgments

We thank Dr Matuszewski's group in the Applied Digital Signal and Image Processing laboratory at the University of Central Lancashire for supplying early versions of GeoCut and GeoCut3D.

References

- International Commission on Radiation Units and Measurement. Prescribing, recording and reporting photon beam therapy. ICRU report 50. Bethesda, MD: ICRU; 1993.
- International Commission on Radiation Units and Measurement. Prescribing, recording and reporting photon beam therapy. ICRU report 62 (supplement to ICRU report 50). Bethesda, MD: ICRU; 1999.
- International Commission on Radiation Units and Measurement. Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). ICRU report 83. Oxford, UK: Oxford University Press; 2010.
- British Institute of Radiology working party. Geometric uncertainties in radiotherapy: defining the planning target volume. London, UK: British Institute of Radiology; 2003.
- National Radiotherapy Advisory Group. Radiotherapy: developing a world class service for England. [Cited 13 May 2011.] Available from: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_074576.pdf
- Grady L. Random walks for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2006;28:1768–83.
- Rohlfing T, Brandt R, Menzel R, Russakoff DB, Maurer CR, Jr. Quo vadis, atlas-based segmentation. In: Suri JS, Wilson DL, Laxminarayan S, eds. *Handbook of medical image analysis*. Vol. III. Registration models. New York, NY: Kluwer Academic/Plenum Publishers; 2005. pp. 435–86.
- 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.
- Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. *Comput Vis Image Underst* 1995;61:38–59.
- 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.
- Zhang Y, Matuszewski BJ, Histace A, Precioso F, Kilgallon J, Moore C. Boundary delineation in prostate imaging using active contour segmentation method with interactively defined object regions. In: Madabhushi A, Dowling J, Yan P, Fenster A, Abolmaesumi P, eds. *MICCAI '10. Proceedings of the 2010 International Conference on Prostate Cancer Imaging: computer-aided diagnosis, prognosis, and intervention*; 20–24 September 2010; Beijing, China. Berlin, Germany: Springer; 2010. pp. 131–42.
- Moore C, Sykes J, MacBain C, Khoo V, Sharrock P. SCULPTER: fast evidence-based 3D anatomy sculpting in radiotherapy by computerised topological self-mapping. In: *CARS 2003. Computer assisted radiology and surgery. Proceedings of the 17th International Congress and Exhibition*. Int Congr Seri 2003;1256:124–9.
- McBain CA, Moore CJ, Green M, Price G, Sykes JS, Amer A, et al. Early clinical evaluation of a novel three-dimensional structure delineation software tool (SCULPTER) for radiotherapy treatment planning. *Br J Radiol* 2008;81:643–52.
- Price GJ, Moore C. Comparative evaluation of a novel 3D segmentation algorithm on in-treatment radiotherapy cone beam CT images. In: Pluim J, Reinhardt J, eds. *Medical Imaging 2007: image processing, proceedings of SPIE*, vol. 6512. SPIE Medical Imaging 2007; 18–20 February 2007; San Diego, CA. Bellingham, WA: SPIE; 2007. pp. 38–1–11.
- Fechner GT. *Elemente der psychophysik*. Leipzig, Germany: Breitkopf und Härtel; 1860.
- Pantle AJ, Sekuler RW. Velocity-sensitive elements in human vision: initial psychophysical evidence. *Vision Res* 1968;8:445–50.
- Wikimedia Foundation Inc. Mach bands. [Cited 10 September 2011.] Available from: http://en.wikipedia.org/wiki/Mach_bands
- Manahilov V. Spatiotemporal visual response to supra-threshold stimuli. *Vision Res* 1995;35:227–37.
- Simonotto E, Riani M, Seife C, Roberts M, Twitty J, Moss F. Visual perception of stochastic resonance. *Phys Rev Lett* 1997;78:1186–9.
- Peng R, Chen H, Varshney PK. Noise-enhanced detection of micro-calcifications in digital mammograms. *IEEE J Sel Top Signal Process* 2009;3:62–73.
- Peng R, Chen H, Varshney PK, Michels JH. Stochastic resonance: an approach for enhanced medical image processing. In: *Proceedings of the 3rd IEEE-NIH Life Science Systems and Applications workshop*; 8–9 November 2007; Bethesda, MD. New York, NY: IEEE; 2008. pp. 253–6.
- Aihara T, Kitajo K, Nozaki D, Yamamoto Y. Internal noise determines external stochastic resonance in visual perception. *Vision Res* 2008;48:1569–73.
- Steenbakkers RJ, Duppen JC, Fitton I, Deurloo KE, Zijp LJ, Comans EF, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. *Int J Radiat Oncol Biol Phys* 2006;64:435–48.
- Pincus SM. Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 1991;88:2297–301.
- Oppenheim AV, Lim JS. The importance of phase in signals. *Proc IEEE* 1981;69:529–41.
- Kovesi P. Invariant measures of image features from phase information [PhD]: University of Western Australia; 1996. [cited 8 October 2012.] Available from: http://repository.uwa.edu.au:80/R/-?func=dbin-jump-full&object_id=5454&silo_library=GEN01
- Szilagy T, Brady M. Feature extraction from cancer images using local phase congruency: a reliable source of image descriptors. In: *Proceedings of 2009 6th IEEE International Symposium on Biomedical Imaging: from nano to macro (ISBI)*; 28 June to 1 July 2009; Boston, MA. Piscataway, NJ: IEEE Signal Processing Society; 2009. pp. 1219–22.
- Moore C, Manickam K, Willard T, Jones S, Slevin N, Shalet S. Spectral pattern complexity analysis and the quantification of voice normality in healthy and radiotherapy patient groups. *Med Eng Phys* 2004;26:291–301.
- Moore CJ, Manickam K, Slevin N. Collective spectral pattern complexity analysis of voicing in normal males and larynx cancer patients following radiotherapy. *Biomed Signal Process Control* 2006;1:113–19.
- Marchant T, Murphy M, Madden G, Moore C. Quantifying structure regularity in fluorescence microscopy cell images using a novel multidimensional approximate entropy metric. In: *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP)*; 11–14 September 2011; Brussels, Belgium. Piscataway, NJ: IEEE Signal Processing Society; 2011. pp. 3085–8.