Tumour auto-contouring on 2d cine MRI for locally advanced lung cancer: A comparative study

Martin F. Fast a,⇑,1, Björn Eiben b,1, Martin J. Menten a, Andreas Wetscherek a, David J. Hawkes b, Jamie R. McClelland b, Uwe Oelfke a

a Joint Department of Physics, The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London; and b Centre for Medical Image Computing, Department of Medical Physics and Biomedical Engineering, University College London, United Kingdom

Article info

Article history:
Received 31 May 2017
Received in revised form 11 August 2017
Accepted 13 September 2017
Available online xxxx

Keywords:
Lung tumour tracking
Auto-contouring
MRI-guided radiotherapy
Locally advanced lung cancer

Abstract

Background and purpose: Radiotherapy guidance based on magnetic resonance imaging (MRI) is currently becoming a clinical reality. Fast 2d cine MRI sequences are expected to increase the precision of radiation delivery by facilitating tumour delineation during treatment. This study compares four auto-contouring algorithms for the task of delineating the primary tumour in six locally advanced (LA) lung cancer patients.

Material and methods: Twenty-two cine MRI sequences were acquired using either a balanced steady-state free precession or a spoiled gradient echo imaging technique. Contours derived by the auto-contouring algorithms were compared against manual reference contours. A selection of eight image data sets was also used to assess the inter-observer delineation uncertainty.

Results: Algorithmically derived contours agreed well with the manual reference contours (median Dice similarity index: ≥ 0.91). Multi-template matching and deformable image registration performed significantly better than feature-driven registration and the pulse-coupled neural network (PCNN). Neither MRI sequence nor image orientation was a conclusive predictor for algorithmic performance. Motion significantly degraded the performance of the PCNN. The inter-observer variability was of the same order of magnitude as the algorithmic performance.

Conclusion: Auto-contouring of tumours on cine MRI is feasible in LA lung cancer patients. Despite large variations in implementation complexity, the different algorithms all have relatively similar performance.

© 2017 The Author(s). Published by Elsevier Ireland Ltd. Radiotherapy and Oncology xxx (2017) xxx–xxx

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Radiotherapy guidance based on magnetic resonance imaging (MRI) is currently becoming a clinical reality. Several hybrid MRI/radiotherapy devices are undergoing advanced testing or are already at the point of clinical service [1–4]. The arrival of these novel treatment machines could usher in a new era in image-guided radiotherapy. For moving and deforming tumours such as seen in lung cancer patients, real-time adaptive radiotherapy can benefit from repeatedly contouring the tumour and organs-at-risk in an automated and reliable fashion. Hybrid MRI/radiotherapy devices can acquire repeated 2d cine MR images of the internal patient anatomy during treatment [5] thus facilitating a number of suitable motion management strategies for lung cancer patients. These include gating [6], margin assessment and reduction [7], dynamic multi-leaf collimator (MLC) tracking [8], respiratory coaching based on audiovisual feedback [9], and combinations thereof. Through reductions in planning target volumes and better beam-target alignments [8,10,11], these techniques are expected to improve treatments for lung cancer by sparing more healthy tissue and allowing for dose escalation when needed.

A number of studies on auto-contouring of lung tumours based on 2d cine MRI have been published (Table 1). Unfortunately, none of these studies explicitly mentions lung tumour staging. Two of the studies resorted to healthy volunteers using normal lung structures as surrogate for tumour motion [12,13]. This study focusses on tumour auto-contouring in locally advanced non-small cell lung cancer or limited stage inoperable small cell lung cancer (subsequently referred to collectively as locally advanced lung cancer). This patient cohort has a 5-year overall survival of only ~20% [14] and needs better treatment options. Tumour auto-contouring is expected to be challenging for these patients as lesions tend to be large and in close proximity to critical structures.

Most previous studies used a balanced steady-state free precession (bSSFP) sequence [13,15–20]. The dynamic bSSFP sequence is widely available and captures T2/T1-weighted contrast, highlighting fat and fluids. Slice orientations were limited to the sagittal plane, or in case of interleaved (orthogonal) acquisitions of two planes, or in case of poor tumour visibility at an oblique angle (perpendicular to the axial plane).

This study compares four auto-contouring algorithms for the task of delineating the primary tumour in six locally advanced lung cancer patients. The auto-contouring algorithms were selected to cover the broad range of previously published algorithms: multi-template matching, a pulse-coupled neural network, intensity-driven deformable image registration, and a feature-driven registration. Besides the ubiquitous bSSFP sequence, this study also uses a spoiled gradient echo sequence more akin in appearance to the familiar CT images [21]. All algorithmically derived contours are compared against manual reference contours to benchmark their validity. We also assessed the inter-observer delineation uncertainty.

**Material and methods**

**Patient cohort**

Six locally advanced lung cancer patients were included in this study (Table 2). Although patient #5 suffered from a malignant pleural effusion and was treated with palliative intent, a radical treatment approach might be feasible in the future [22], especially with online MRI guidance.

**Image acquisition**

A total of twenty-two cine MR image series were acquired on a Siemens MAGNETOM Aera (Siemens Healthcare GmbH, Erlangen, Germany) 1.5 T scanner using either a bSSFP (TrueFISP), or a spoiled gradient echo (GRE, FLASH) acquisition technique. Patients were scanned during free breathing in head-first-supine position with arms down. Informed consent was obtained from the patients according to institutional standards. Acquisition parameters are summarised in the supplement. Each image series was acquired for approximately 1 min. A stack of three 2d imaging planes was manually positioned to ensure it intersected with the time-averaged tumour as measured on a 3d golden angle radial VIBE [24] MR image (pixel spacing: 1.5 × 1.5 × 3 mm, field of view: 384 × 384 × 288 mm, TR = 2.51 ms, TE = 1.24 ms, bandwidth = 1085 Hz/Px, acquisition time = 4.58 min, fat-saturation) acquired at the beginning of each imaging session. After scanning, the image plane which intersected best with the tumour (usually the central slice) was selected for further processing, resulting in 1.8 Hz image sampling. Images were positioned in the sagittal and coronal planes, or in case of poor tumour visibility at an oblique angle (perpendicular to the axial plane).

**Manual contouring and inter-observer variability**

In each cine MR sequence primary tumour contours were manually generated by one of four expert observers, all following the same delineation procedural protocol. For a selection of image data sets two additional secondary tumour contour sets were generated independently by expert observers (giving a total of three manual contour sets) to assess the inter-observer delineation variability. The selection consisted of eight image data sets representing all combinations of imaging sequences, slice orientation, and small or large tumour motion amplitude (threshold: 5 mm). The observers generating secondary contours used the same window/level setting as was used to generate the primary contours. Only the first 30 images in each series were contoured by the additional observers to lighten the work load. A STAPLE delineation [25] approximating the hidden, unobservable ground truth delineation was generated from the three manual contours. The inter-observer variability was assessed by measuring the distance from each individual contour to the STAPLE contour.

### Table 2: Previous lung auto-contouring studies based on 2d cine MRI.

<table>
<thead>
<tr>
<th>Reference</th>
<th>B₀ [T]</th>
<th>Sequence</th>
<th>Orientation</th>
<th>Algorithm</th>
<th>Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cervino et al. [12]</td>
<td>3.0</td>
<td>n.p.</td>
<td>SAG slice</td>
<td>TM, ANN</td>
<td>5 healthy volunteers</td>
</tr>
<tr>
<td>Tryggestad et al. [13]</td>
<td>1.5</td>
<td>bSSFP</td>
<td>SAG/COR</td>
<td>TM</td>
<td>2 healthy volunteers</td>
</tr>
<tr>
<td>Shi et al. [15]</td>
<td>1.5</td>
<td>bSSFP</td>
<td>SAG stack</td>
<td>TM</td>
<td>12 NSCLC</td>
</tr>
<tr>
<td>Yun et al. [16]</td>
<td>0.35</td>
<td>bSSFP</td>
<td>SAG slice</td>
<td>PCNN</td>
<td>4 NSCLC</td>
</tr>
<tr>
<td>Paganeli et al. [17]</td>
<td>3.0</td>
<td>bSSFP</td>
<td>SAG/COR</td>
<td>SIFT</td>
<td>9 lung patients</td>
</tr>
<tr>
<td>Seregni et al. [18]</td>
<td>3.0</td>
<td>bSSFP</td>
<td>SAG/COR</td>
<td>SIFT</td>
<td>6 lung patients</td>
</tr>
<tr>
<td>Mazur et al. [19]</td>
<td>0.35</td>
<td>bSSFP</td>
<td>SAG slice</td>
<td>SIFT</td>
<td>4 lung patients</td>
</tr>
<tr>
<td>Bourque et al. [20]</td>
<td>1.5</td>
<td>bSSFP</td>
<td>SAG slice</td>
<td>Part. filt.</td>
<td>4 NSCLC</td>
</tr>
</tbody>
</table>

**Abbreviations:** ANN = artificial neural network, bSSFP = balanced steady-state free precession, COR = coronal, n.p. = not provided, NSCLC = non small cell lung cancer. Part. filt. = partial filter.

* a Interleaved acquisition.
* b Non-lung subjects and phantoms excluded.
* c Derived from 3.0 T image.

**Table 2:** Patient characteristics. Staging was performed according to the AJCC recommendations [23]. Average 2d GTV size and maximum centroid motion extent (2nd to 98th motion percentile along the principle axis of motion following principle component analysis) were derived from manual contours on the sagittal slice.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age (a)</th>
<th>GTV (cm²)</th>
<th>Motion (mm)</th>
<th>Pathology</th>
<th>TNM</th>
<th>Stage</th>
<th>Tumour position</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>F</td>
<td>70</td>
<td>16.4</td>
<td>4.2</td>
<td>NSCLC</td>
<td>T2bN1M0</td>
<td>IIIB</td>
<td>Right upper lobe</td>
</tr>
<tr>
<td>#2</td>
<td>M</td>
<td>76</td>
<td>14.0</td>
<td>6.1</td>
<td>NSCLC</td>
<td>T4N0M0</td>
<td>IIIA</td>
<td>Left hilar</td>
</tr>
<tr>
<td>#3</td>
<td>M</td>
<td>50</td>
<td>27.9</td>
<td>8.4</td>
<td>SCLC</td>
<td>T4N1M0</td>
<td>IIIB</td>
<td>Left lower lobe</td>
</tr>
<tr>
<td>#4</td>
<td>M</td>
<td>60</td>
<td>22.1</td>
<td>22.0</td>
<td>SCLC</td>
<td>T4N1M0</td>
<td>IIIB</td>
<td>Right middle lobe</td>
</tr>
<tr>
<td>#5</td>
<td>M</td>
<td>68</td>
<td>27.1</td>
<td>6.6</td>
<td>NSCLC</td>
<td>T4N0M1a</td>
<td>IV</td>
<td>Left lower lobe</td>
</tr>
<tr>
<td>#6</td>
<td>M</td>
<td>70</td>
<td>86.1</td>
<td>1.5</td>
<td>NSCLC</td>
<td>T4N2M0</td>
<td>IIIB</td>
<td>Right upper lobe</td>
</tr>
</tbody>
</table>

**Abbreviations:** GTV = gross tumour volume. (N)SCLC = (non) small cell lung cancer.

Auto-contouring

Each image data set was divided into a training data set (the first ten images) and a test data set of 90–120 images (all subsequent images). Manual contours from the test data set were only used for evaluation purposes and not for algorithm training. Ten training images were chosen to capture at least one complete respiratory cycle.

The following auto-contouring algorithms were implemented.

Multi-template matching (MTM)

This algorithm was adapted from previous work [26,27]. A rectangular template is cropped from each training image using the smallest tumour-enclosing (axis-aligned) bounding box extended by three pixels in all directions. The test image pixel with the highest normalised cross-correlation score from any of the ten templates is then selected as best match. The contour from the respective best training image is subsequently copied to the test image.

Pulse-coupled neural network (PCNN)

This algorithm was previously presented by Yun et al. [16]. It attempts to improve grey-value contrast between tumour and lung tissue by means of a pulse-coupled neural network (PCNN) approach. Within the PCNN algorithm each image pixel is treated as a neuron. The internal activity of each neuron is influenced by the image intensity and other adjacent neurons. Neurons with similar levels of activity pulse together yielding tumour contours if PCNN parameters are selected appropriately. PCNN parameters are optimised based on the training images and associated manual contours.

Intensity-driven deformable image registration (DIR)

This algorithm uses a two-stage approach to propagate a contour from the training image set to a test image. First, the image similarity is calculated between all training images and the current test image by the means of local normalised cross-correlation [28]. Second, the most similar training image is then transformed by a B-spline deformable image registration (DIR [29,30]) to match the test image. The resulting displacement vector field is subsequently applied to deform the contour from the corresponding training image to the current test image.

Feature-driven registration (SIFT)

Similar to earlier approaches (e.g. [17]) this algorithm uses image features to select, deform and propagate a training contour.
onto each test image. Feature extraction [31] and scale invariant feature transform (SIFT) description [32] is performed for all training images and the test image within a region of interest. The best feature set combination is selected and used to generate a transformation from the corresponding training image to the test image. The training contour is then propagated according to the established transformation.

Contour comparison and statistical tests

Automatically generated contours were compared with manual ground-truth contours for all test images using the Dice similarity index, Hausdorff distance, centroid distance, and mean contour distance (c.f. supplement). The centroid distance is especially important for applications such as dynamic tumour tracking [8]. The Hausdorff and mean contour distances are relevant when treatment is adapted to tumour deformations [33].

Differences between contouring methods were tested for significance in terms of contour similarity using a linear mixed effects model. The model intends to separate the random variation due to the finite number of image sets, from the potentially systematic variation introduced by the auto-contouring algorithm. It was constructed in R (R Core Team, v3.4.0) using the lme4 package [34] with the contouring algorithm as predictor. A second model was applied to the data from each contouring algorithm to separate out the impact of different predictors: imaging sequence, slice orientation, maximum motion, and observer. A third model was applied to the inter-observer data to investigate the impact of imaging sequence, slice orientation, and maximum motion on contouring consistency.

Results

Fig. 1 shows examples of the manually and automatically generated contours for the eight cine MR sequences that were selected for inter-observer analysis.

Inter-observer variability

The eight inter-observer cases were evaluated by comparing the STAPLE delineation with each observer using all similarity metrics (Fig. 2 and supplement). For the Hausdorff distance and the mean contour distance, the bSSFP sequence resulted in significantly lower values \( p < 0.05 \). No other significant differences were found.

Auto-contouring

The overall performance of all auto-contouring algorithms is shown in Fig. 3 and the supplement. Contour creation per image took about 1 ms (MTM), 25 ms (PCNN), 150 ms (SIFT), and 500 ms (DIR) on standard workstation computers. Absolute differences

Fig. 2. Results of inter-observer variability study for the complete dataset (all) and selected subsets. The centre-line of the box corresponds to the median, the box signifies the first and third quartiles, the whiskers extend the box by 1.5 times the interquartile range, and crosses are outliers. Significance levels are defined as * \( p < 0.05 \), ** \( p < 0.01 \), and *** \( p < 0.001 \). The vertical axes are matching with Fig. 3.

in median quantities are generally small when comparing different algorithms. The DIR algorithm performs significantly better than the SIFT or PCNN algorithm for all metrics. The MTM algorithm is significantly better than the PCNN or SIFT algorithm for a subset of metrics. No significant difference could be detected between MTM and DIR.

Fig. 3. Algorithmic auto-contouring performance is evaluated for the complete dataset (all) and selected subsets. See caption of Fig. 2 for a description of the quantities visualised by the plots.
The complete dataset was then split according to image acquisition sequence, slice orientation, and motion amplitude (Fig. 3). For PCNN, more motion resulted in reduced algorithmic performance for all metrics ($p < 0.001$). The same correlation was observed for the centroid distance using SIFT ($p < 0.05$). Using the bSSFP sequence resulted in an improved mean contour distance for the SIFT algorithm ($p < 0.05$). In addition, the observer was a significant predictor for the centroid distance of DIR and the mean contour distance of PCNN (both $p < 0.05$). The magnitude of observer influence was estimated to be 0.87 mm (centroid distance, DIR) and 0.70 mm (mean contour distance, PCNN). No other significant differences were found.

Discussion

Lung tumour patients with locally advanced disease are expected to benefit from online MRI guidance. An essential component for adapted treatment on the basis of continuously acquired cine MR images is the ability to automatically delineate the tumour. To the best of our knowledge, systematic evaluations of auto-contouring algorithms for cine MRI from patients with locally advanced lung cancer were not previously reported in the literature. We thus implemented a broad range of auto-contouring algorithms and compared their performance on image data sets from a cohort of six locally advanced lung cancer patients.

In our study, the MTM and DIR algorithms outperformed the PCNN and SIFT algorithms for all investigated metrics. For the two best performing algorithms median values were 0.94 (Dice similarity index), 4.43 mm (Hausdorff distance), $< 1.52$ mm (centroid distance), and $< 1.38$ mm (mean contour distance). Centroid and mean contour distances approach a point where pixel resolution is the limiting factor. The MTM algorithm only compensates for translations but not for previously unseen rotations or deformations. The conceptually higher flexibility of the other algorithms might be preferable for longer treatment durations. In this study, scanning for only one minute allowed us to investigate different MR acquisition sequences and slice orientations for the same patient without exceeding the patient-tolerable total scanning time.

Although the bSSFP sequence has been almost exclusively used in previous cine MRI studies for lung (Table 1), we could not find conclusive evidence showing that bSSFP is preferable over GRE for the task of auto-contouring lung tumours. Previous studies reported that bSSFP images are preferable over GRE [21,35] for a variety of reasons. It should be noted that our study investigates the primary disease seems feasible. Based on the results of this study, automated tracking of bSSFP sequence resulted in more similar manual contours with the bSSFP sequence was a significant predictor for observer similarity. Using bSSFP resulted in more similar manual contours with observer similarity. Using bSSFP resulted in more similar manual contours with the bSSFP sequence resulted in an improved mean contour distance, DIR) and 0.7 mm (mean contour distance). No other significant differences were found.

Inter-observer variability provides perspective on the achieved auto-contouring performance. The median values for the contouring metrics indicate that inter-observer variability is of similar magnitude as the algorithmic performance: 0.97 (Dice similarity index), 3.96 mm (Hausdorff distance), 0.76 mm (centroid distance), and 0.85 mm (mean contour distance). Given the consistently high performance of all algorithms, future gains in targeting accuracy may only be achieved from improvements in manual contouring or from making the auto-contouring independent from manual contours. In the inter-observer variability study, only the choice of imaging sequence was a significant predictor for observer similarity.

Acknowledgements

Research at The Institute of Cancer Research is supported by Cancer Research UK under Programme C33589/A19908. Research at the Centre for Medical Image Computing, UCL, is supported by the CRUK grant C33589/CRC521. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research. We would like to thank Hannah Bainbridge, David Collins and Cynthia Eccles (all ICR/RMH) for assisting in the MR image acquisition. Special thanks to Kyle Bromma (University of Victoria, Canada) for assisting in the implementation of MTM and PCNN.
References