Advances and Challenges in Contour QA for Adaptive RT

Kristy K Brock, PhD, DABR, FAAPM
Professor, Department of Imaging Physics
Director, Image Guided Cancer Therapy Program
University of Texas MD Anderson Cancer Center

Disclosures

• I have a licensing agreement for deformable image registration technology with RaySearch Laboratories.

Objective

• Learn about specific advances and challenges in contour QA which relate to adaptive radiation therapy (ART)
Motivation

• The mandatory role of auto-segmentation in adaptive RT
  - Workflow
  - Time constraints
  - Competing risks of additional time on the table (leading to motion) and uncertainty in contours
• The migration of contouring errors from systematic to random in real-time, daily, adaptive RT
• Examples of QA and evaluation for adaptive RT

Adaptive Tools

Adaptive Tools
Impact of Errors in Contours

**Standard Treatment**
- Error in contours → error in treatment plan for the entire course
- Random and systematic errors in treatment impact this error by blurring or shifting the dose
  - Potentially making the error worse or better

**Adaptive Treatment**:
- If DIR-based contour propagation: Error in contours → error in treatment plan for the entire course
- If Model-based segmentation: Error in contours → error in treatment plan for the fractions the plan is used

Contour-Based Validation

- Dice Similarity Coefficient (DSC)
- Mean Distance to Agreement (MDA)
- Hausdorff Distance (HD) = max
- Doesn’t relate the contour error to dose!
Effect of DIR Uncertainty Reduction on Lung Dose Accumulation

• Dose accumulation:
  – Summation of the radiation dose by taking into account tissue motion (e.g., breathing)
• Clinical Importance:
  – Discrepancy between the planned dose (static) vs the accumulated dose
  – >1 Gy differences in dose parameters could potentially be clinically significant considering 48-60 Gy plans

Effect of DIR Uncertainty Reduction on Lung Dose Accumulation

• Question: How are geometric uncertainties translated to dosimetric uncertainties?
• Data: 10 Lung SBRT patients Tx with 48-60 Gy in 3-4 Fr
• No tumor response over the short course of treatment
• Dose accumulation is a multi-step process (inhale to exhale, planning to each fraction)

Dosimetric Impact: Initial Step in Dose Accumulation

• Dose Heterogeneity Index
  \[ \text{DHI} = \frac{D_{20} - D_{80}}{D_{Rx}} \]
Differences in $D_{\text{min}}$ were significant ($p = 0.05$).

- 5/12 cases with $> 1$ Gy $D_{\text{min}}$ difference exhibit the following:
  - DHI $> 15$
  - DSC differences $> 0.08$

**Dosimetric Impact:**

Initial Step in Dose Accumulation

- Differences in $D_{\text{min}}$ were significant ($p = 0.05$)
- 5/12 cases with $> 1$ Gy $D_{\text{min}}$ difference exhibit the following:
  - DHI $> 15$
  - DSC differences $> 0.08$

**Summary: Dosimetric Impact**

- 1.5 mm reduction of DIR error translated to $>1$ Gy differences in $D_{\text{min}}$ in up to 50% of a patient population with the following characteristics:
  1. DHI $> 15$
  2. DIR-induced Dice differences $> 0.08$
- These characteristics were specific criteria but not highly sensitive since there were cases that met the criteria without resulting in $>1$ Gy differences (in accumulated dose).

On-board magnetic resonance (MR) image guidance during radiation delivery offers the potential for more accurate treatment delivery. To utilize the real-time image information, a crucial prerequisite is the ability to successfully segment and track regions of interest (ROI). The purpose of this work is to evaluate the performance of different segmentation algorithms using motion images (4 frames per second) acquired using a MR image-guided radiotherapy (MR-IGRT) system.
All methods were able to successfully segment the bladder and the kidney, but only FKM, KHM, and VR-TPDS were able to segment the liver tumor and the duodenum.

The performance of the thresholding, FKM, KHM, and RD-LSE algorithms degraded as the local image contrast decreased, whereas the performance of the VP-TPDS method was nearly independent of local image contrast due to the reference registration algorithm.

For segmenting high-contrast images (i.e., kidney), the thresholding method provided the best speed (< 1 ms) with a satisfying accuracy (Dice = 0.95).
Summary: MR-IGRT

- When the image contrast was low, the VR-TPDS method had the best automatic contour.
- Results suggest an image quality determination procedure before segmentation and a combination of different methods for optimal segmentation with the on-board MR-IGRT system.

7 female and 14 male patients diagnosed as nasopharyngeal cancer (NPC) were included in this study.

OARs including the parotid gland (PG), the submandibular gland (SMG), the cervical vertebra (VTB) and the vertebral foramen (VF), on both PCT and CBCT, were manually delineated by an experienced physician using a commercial treatment planning system (TPS, Eclipse 10.0, Varian) and double checked by this same physician three months later to ease intra-observer variations.

### Algorithms

<table>
<thead>
<tr>
<th>Class</th>
<th>Algorithm</th>
<th>Tool/Implementation Environment</th>
<th>Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optical Fate</td>
<td>Original Non-Subtract method</td>
<td>DIRMITMedia</td>
<td>HS</td>
</tr>
<tr>
<td></td>
<td>Contracted from Schuch and Luna-Karalev method</td>
<td>DIRMITMedia</td>
<td>HS_K</td>
</tr>
<tr>
<td>Press Form Artefact method</td>
<td>DIRMITMedia</td>
<td></td>
<td>FPO</td>
</tr>
<tr>
<td>Demons</td>
<td>Original Demons method</td>
<td>DIRMITMedia</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>Modified Demons method</td>
<td>DIRMITMedia</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Symmetric Press Demons method</td>
<td>DIRMITMedia</td>
<td>SPF</td>
</tr>
<tr>
<td></td>
<td>Double Press Demons method</td>
<td>DIRMITMedia</td>
<td>SPF</td>
</tr>
<tr>
<td></td>
<td>Deformation with Intensity Smoothening Corrected</td>
<td>CUDAFEN</td>
<td>DRE</td>
</tr>
<tr>
<td>Liver-cirr</td>
<td>Original-used Contourletion method</td>
<td>DIRMITMedia</td>
<td>LS</td>
</tr>
<tr>
<td>Spline</td>
<td>S-Spline method</td>
<td>Elekta/C+</td>
<td>SSpine</td>
</tr>
</tbody>
</table>

[Table 1: Overview of Ten DEM Algorithms.](#)
Summary: DIR evaluation

- It was found that the evaluated DIRs in this work did not necessarily outperform rigid registration.
- DIR performed better for bony structures than soft-tissue organs, and the DIR performance tended to vary for different ROIs with different degrees of deformation as the treatment proceeded.
- Generally, the optical flow-based DIR performed best, while the demons-based DIR usually ranked last except for a modified demons-based DIR used for CT-CBCT DIR.
- These experimental results suggest that the choice of a specific DIR algorithm depends on the image modality, anatomic site, magnitude of deformation and application.
- Careful examinations and modifications are required before accepting the auto-propagated contours, especially for automatic re-planning ART systems.

Ten head and neck cancer patients who received weekly CBCT imaging as part of a previous study at our institution were included in the study. Two observers (GT1 and GT2) independently contoured the parotids on the pCT and each weekly CBCT for each patient; these structures were taken as the ground truth. ADMIRE was used to propagate these ground truth parotids from the pCT onto each CBCT, and the accuracy of the propagations was measured with DSC and mean DTA. In addition to the accuracy of the propagated contours, the inter-observer variation was estimated from the concordance of the two sets of ground truth structures.
Methods

- The ability of the automated workflow to detect gross propagation errors was tested by copying contours to incorrect images for a subset of patients for a single observer (GT1). Propagated contours on CBCT’s 3–6 were copied onto CBCT2, such that the contours on CBCT2 originated from a different image set.
- The automated workflow was performed on these structures and the consistency metrics were measured. The ability of the uncertainty metrics to identify these errors was investigated.
- For the second error scenario, Gaussian noise was added to the CBCT images (CBCT2-6)

Accuracy

| Table 1. Mean accuracy and standard deviations of the propagated structures relative to the ground truth structures, and inter-observer variation for the CBCT images and pCT. Note that inter-observer variations calculated excluding the three patients with large discrepancies between observers are denoted with *. |
|-----------------|---------------|-----------------|
| Propagation     | DSC           | Mean DTA (mm)   |
| Propagation accuracy (GT1) | 0.82 ± 0.02  | 1.64 ± 0.26     |
| Propagation accuracy (GT2) | 0.79 ± 0.06  | 1.96 ± 0.43     |
| Inter-observer variation (CBCT) | 0.74 ± 0.05  | 3.52 ± 1.49     |
| Inter-observer variation (CBCT)* | 0.75 ± 0.06  | 3.05 ± 1.13     |
| Inter-observer variation (pCT) | 0.84 ± 0.03  | 2.20 ± 1.18     |
| Inter-observer variation (pCT)* | 0.86 ± 0.02  | 1.57 ± 0.21     |
Consistency Metrics

Accuracy as a Function of Noise

Figure 3. Plot of the consistency metrics for error scenario 1. The errors are shown as red dots, and the standard deviations as green circles. The clear separation between the standard deviations and errors means that these errors could be identified from the consistency metrics.

Figure 4. ROC curves for the three-fold cross validated logistic regression model for GT1 (A) and GT2 (B). The solid black line shows the average curve, and the red dotted line shows the random guess line. The AUC was 0.90 for GT1 and 0.85 for GT2.
Summary: Automated Workflow

- Contour propagation is an essential component of ART, but unreliable propagation limits its routine clinical implementation.
- There are currently no tools to aide patient-specific QC of contour propagation.
- An automated workflow for patient-specific QC of contour propagation, based on consistency metrics calculated from multiple registrations, has been presented and tested on a set of ten head and neck patients with simulated propagation errors.
- This work shows potential as a tool for quality control of contour propagation, and could help facilitate the clinical implementation of adaptive radiotherapy.

Conclusions

- Auto-segmentation is key for adaptive RT
- We are still limited by comparison to manual contours, which have uncertainties as well.
- DIR-based auto-segmentation is not necessarily better than rigid
- Image quality plays a role in auto-segmentation
- Accuracy of auto-segmentation really needs to be assessed in terms of dosimetric impact.