Automated Contour Segmentation for Treatment Planning - Fasten Your Seatbelt

Minsong Cao, Ph.D.

Department of Radiation Oncology
University of California, Los Angeles
UCLA

Learning objectives

• Recognize the current and future role of auto-contouring in RT

• Understand the strengths and limitations associated with various automation methods

• Understand practical considerations in implementing automated contour segmentation
Disclosure

• None

Background

• Clinical contouring is critical for treatment planning:
  • time consuming and labor intensive
  • large variations among different institutions, operators and tools
  • directly impact dosimetry quality and clinical decision

Factors significantly associated with longer contour times:
• Setup (prone or supine)
• Operator (junior or senior)
• Planning system
• Part of the day
• Day of the week

203 breast cases:
• Mean contouring time: 34min
• Mean contouring structure: 4.7

V.A. Andrianarison et al. / Cancer/Radiothérapie 22(2018)
Background

• Contour is one of the largest sources of dosimetric uncertainty
  • contour error (small structures: optic chiasm, cochlea etc.)
  • contour variability (intra- and inter-operator)
• quality of contouring:
  • spatial accuracy
  • dosimetric accuracy


Automated contour segmentation

• Seek to reduce time and inter-operator variability
• Clinical applications:
  • Standard treatment planning
  • Adaptive treatment planning
  • Motion tracking and gating
• Commercial products available, but not frequently used in clinical practice
• Conflict findings reported on contour accuracy and time saving
Automated segmentation methods

• Non prior-knowledge
  • Directly based on image voxel intensities and/or gradient
  • High contrast structures e.g. lung, bone, air cavity
  • May require post-processing
  • Various tools available in contouring software

• Prior-knowledge
Automated segmentation methods

- Prior-knowledge
  - Atlas based segmentation
  - Statistical model based segmentation:
    - Shape (SSM) or
    - Appearance (SAM)
  - Machine learning based segmentation
  - Hybrid segmentation

Atlas based segmentation – Single Atlas
Atlas based segmentation – Multi-atlas

- Multi-Atlas Selection
- Voting schemes:
  - Majority voting
  - Intensity weighting
  - STAPLE

Performance of atlas based segmentation

- Heavily depends on quality of atlas images and reference contours
- Atlas selection strategy:
  - Robust metric
  - Image registration
- database size: the more the merrier?
  - No consensus on database size
  - Computing time limitation

Single verse multiple atlas

- Multi-atlas seeks to improve robustness of segmentation
  - Not biased by a single subject
  - Prone to topological error
  - Voting scheme and parameter selection are crucial

D. Teguh et al. IJROB., Vol. 81(4), 2011

Atlas based segmentation – image registration

- Quality of segmentation highly relies on image registration method
- Deformable image registration (DIR) plays an important role
- Challenge: Ground-truth is not available
- Many different approaches and transformation modes
- Quality assurance and validation is important, but usually not accessible to end-users
Atlas based segmentation - DIR

<table>
<thead>
<tr>
<th>Class</th>
<th>Transformation</th>
<th>Maximum dimensionality of transformation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric</td>
<td>Rigid</td>
<td>6</td>
<td>Allows translation in 3 directions and rotations about 3 axes.</td>
</tr>
<tr>
<td></td>
<td>Affine</td>
<td>12</td>
<td>In addition to translation/rotation, allows uniform scaling and shear (e.g., parallel lines stay parallel).</td>
</tr>
<tr>
<td></td>
<td>Free-form</td>
<td>3N</td>
<td>Local, voxel-based deformation, often regulated by a smoothing parameter.</td>
</tr>
<tr>
<td></td>
<td>Global spline-based methods (e.g., thin plate splines)</td>
<td>3N</td>
<td>Parameters deformation using a parametric grid of basis function control points with constrained global influence (e.g., deformation is global).</td>
</tr>
<tr>
<td></td>
<td>Local spline-based methods (e.g., B-spline)</td>
<td>3N</td>
<td>Parameters deformation using a weighted grid of control points of basis functions with local influence (e.g., deformation is local).</td>
</tr>
<tr>
<td>Physical</td>
<td>Viscous/elastic/optical flow (e.g., demons)</td>
<td>3N</td>
<td>Spatially variant voxel displacement voxel displacement by a vector field in a deforming medium, by intensity gradients (deformation is local).</td>
</tr>
<tr>
<td></td>
<td>Finite element methods (FEM)</td>
<td>3N</td>
<td>Spatially variant voxel displacement voxel displacements governed by biomechanical tissue properties (deformation is local).</td>
</tr>
</tbody>
</table>

Statistical model based segmentation

- Confine the segmented contours to anatomically plausible shape or appearance
- Require training dataset to characterize variation of shape or appearance of structure
- Fit the test image to the model based on image intensities, gradients, features etc.
Machine-learning based segmentation

• Outstanding performance in classification, detection, pattern recognition
• Automatically learn priors for structures or image context and tissue appearance
• Require training and significant computational resource
• Usually combined with shape model or atlas based methods

Adaptive radiation therapy (ART)

• Re-optimize the treatment plan to account for:
  • Systematic changes (off-line ART)
    • Tumor regression/progression
    • Weight loss
  • Stochastic changes (on-line ART)
• Efficiency is essential
  • Off-line ART: between fractions
  • On-line ART: within a single fraction
Segmentation for ART

- Intra-object segmentation for anatomy at two different time points
- Deformable image registration is the most popular method
- Usually requires labor-intensive manual correction

Source Target Deformed

Segmentation for ART

- Time constraints require more robust and accurate segmentation
- Advanced auto-segmentation approaches

• Literature review of segmentation and registration methods for adaptive cervical cancer treatment planning:
  • Landmark, rigid, B-spline, shape constrained B-spline registration
  • A average of 0.85 Dice similarity and mean surface distance of 2-4mm
  • The use of shape priors significantly improved segmentation accuracy

Motion management – Tx gating

• High speed real time imaging allows direct gating based on anatomy (e.g. GTV, OAR)
• Require automatic detection of target and OARs
• Deformable image registration between baseline and real time images
Motion management – Tx gating

Evaluation of segmentation performance

- Geometric based
  - Moment based
    - Center/Volume of structure
  - Overlap based
    - Dice similarity coefficient
  - Distance based
    - Average/maximum distance

Sharp et al. Medical Physics, Vol. 41, No. 5, 2014
Geometry based evaluation

- **Overlap based metric**
  - Dice similarity coefficient (DSC)
    - Intuitive and quantitative
    - Insensitive to large structure
    - Insensitive to fine details
  - Jaccard Index (JAC)
    \[
    JAC = \frac{|X \cap Y|}{|X \cup Y|} = \frac{DSC}{2 - DSC}
    \]

\[DSC = \frac{2|X \cap Y|}{|X| + |Y|}\]

- **Distance based metric**
  - Hausdorff distance (HD)
    - distance between the points on the boundaries of two structures
    - sensitive to small regions
    - usually use 95% percentile
Geometry based evaluation

- Each of metrics has its own strength and weakness
- May not correlate with each other
- Multiple metrics to be reported together
- Do not directly relate to plan dosimetry!

DSC = 0.92
HD = 8mm

Inter-observer and Intra-observer contour variations always exist

Who has the ground truth?

“Right Answer”
Inter-observer variability

- A single manual contour may not truly represent the ground truth
- Consensus on contour definition is not always available
- Inter-observer variability should be used as benchmark to assess the accuracy and robustness of auto-segmentation

<table>
<thead>
<tr>
<th>Parotid gland</th>
<th>Dice</th>
<th>$0.85$ (Ref. 59)</th>
<th>$0.66 \pm 0.1$ (Ref. 52)</th>
<th>$0.76 \pm 0.08$ (Ref. 49)</th>
<th>$0.60$ (Ref. 53)</th>
<th>$0.26 \pm 0.73$ (Ref. 60)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaccard</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evaluation of segmentation performance

- **Geometric based**
  - Do not directly reflect plan dosimetry!

- **Dosimetric based**
  - Dose optimization
  - Dosimetric metrics (DVHs)
  - Clinical decision

**From geometry to dosimetry**

Stiehl B et al. AAPM 2017
From geometry to dosimetry

Are we ready?
Auto-segmentation Challenge

• Allows assessment of state-of-the-art segmentation methods under unbiased and standardized circumstances:
  • The same datasets (training/testing)
  • The same evaluation metrics

• Head & Neck Auto-segmentation Challenge at MICCAI 2015 conference

• Lung CT Segmentation Challenge 2017 at AAPM Annual Meeting

Auto-segmentation Challenge

• Head & Neck Auto-segmentation Challenge at MICCAI 2015 conference
  • Date from RTOG 0522 clinical trial
  • 25 datasets as training data
  • 10 datasets for off-site and 5 for on-site (2 hours) testing
  • 9 anatomical structures (brainstem, optical chiasm, mandible, parotid glands and submandibular glands)
**Table III.** Comparison of the main features of the participants’ segmentation approaches.

<table>
<thead>
<tr>
<th>Team</th>
<th>Segmentation approach</th>
<th>Nonrigid registration</th>
<th>Initialization</th>
<th>Similarity measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>FH</td>
<td>Model-based (SSM)</td>
<td>Clamped plate spline warp&lt;sup&gt;58&lt;/sup&gt;</td>
<td>Atlas-based</td>
<td>Mutual information</td>
</tr>
<tr>
<td>IM</td>
<td>Model-based (AAM)</td>
<td>Groupwise image registration&lt;sup&gt;46&lt;/sup&gt;</td>
<td>Alignment of center of gravity/scale</td>
<td>Minimum description length</td>
</tr>
<tr>
<td>UB</td>
<td>Atlas- and model-based (ASM)</td>
<td>DEEDs algorithm&lt;sup&gt;49&lt;/sup&gt;</td>
<td>Atlas-based</td>
<td>Self similarity context</td>
</tr>
<tr>
<td>UC</td>
<td>Atlas-based</td>
<td>Elastic transformation (ELAST)&lt;sup&gt;32&lt;/sup&gt;</td>
<td>Atlas-based</td>
<td>Mutual information</td>
</tr>
<tr>
<td>UW</td>
<td>Basic image processing</td>
<td>-</td>
<td>Landmark detection</td>
<td>Sum of squared distances (SSD)</td>
</tr>
<tr>
<td>VU</td>
<td>Atlas-based</td>
<td>Adaptive bases algorithm (ABA)&lt;sup&gt;52,56&lt;/sup&gt;</td>
<td>Atlas-based</td>
<td>Normalized mutual information</td>
</tr>
</tbody>
</table>

---

**Off-site**

(a) [Graph showing Dice values for various structures horizontally]

**On-site**

(b) [Graph showing Dice values for various structures horizontally]
### Mandible:
- Exclusion of teeth
- Image artifacts from dental implant

<table>
<thead>
<tr>
<th>Team</th>
<th>IM</th>
<th>FH</th>
<th>UW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>AAM</td>
<td>SSM</td>
<td>Image Processing</td>
</tr>
<tr>
<td>Dice</td>
<td>0.93</td>
<td>0.785</td>
<td>0.728</td>
</tr>
<tr>
<td>95%HD(mm)</td>
<td>2.041</td>
<td>5.919</td>
<td>29.458</td>
</tr>
</tbody>
</table>

ASM - active appearance model
SSM - statistical shape model

### Parotid glands:
- Large shape variation
- Poor soft tissue contrast
- Heterogeneous tissue including vessels and ducts

<table>
<thead>
<tr>
<th>Team</th>
<th>Dice</th>
<th>Max HD [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>UB</td>
<td>0.814</td>
<td>12.000</td>
</tr>
<tr>
<td>IM</td>
<td>0.826</td>
<td>30.490</td>
</tr>
</tbody>
</table>

ASM
AAM
More on segmentation challenge

- AAPM 2017 Thoracic Auto-segmentation Challenge
- RTOG 1106 contouring atlas
- 36 training sets, 12 offsite test and 12 live competition cases
- Inter-observer contour variability considered

Table 1. Summary of specific implementation details of various segmentation methods used in the contest. Training time is the segmentation time for one patient. DLC, deep-learning contouring; MAC, multi-atlas contouring.

<table>
<thead>
<tr>
<th>Method</th>
<th>Approach</th>
<th>Unique implementation features</th>
<th>Training time</th>
<th>Testing time</th>
<th>Run-time GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DLC</td>
<td>- Hierarchical segmentation; use long to constrain location of other structures; 2D/3D (lung) and 3D (other) residual Units - Training from scratch</td>
<td>3 days</td>
<td>30 s</td>
<td>Time X 12GB</td>
<td></td>
</tr>
<tr>
<td>2 DLC</td>
<td>- Two-step segmentation: first step to locate structures and second step to segment structures - 3D Unet</td>
<td>2 days</td>
<td>10 s</td>
<td>Time Xp 12GB</td>
<td></td>
</tr>
<tr>
<td>3 DLC</td>
<td>- 2D Multiclass network to reduce the demand for a high spc graphics card at run time - Fine-tuning of pretrained network - Loss function penalizing small structures</td>
<td>&gt;7 days</td>
<td>6 min</td>
<td>GTX 1080 2GB</td>
<td></td>
</tr>
<tr>
<td>4 MAC</td>
<td>- Structure-specific label fusion</td>
<td>-</td>
<td>8 h</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>5 DLC</td>
<td>- 2D ResNet</td>
<td>14 days</td>
<td>2 min</td>
<td>K40</td>
<td></td>
</tr>
<tr>
<td>6 MAC</td>
<td>- Use uncertain and unprocessed training data from an operating clinic; no preprocessing - Organ-based STAPLE fusion</td>
<td>-</td>
<td>5 min</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>7 DLC</td>
<td>- 3D (lung, heart) and 2D (other) Uints - Google Inception layers for convolution</td>
<td>4 h</td>
<td>2 min</td>
<td>Pascal</td>
<td></td>
</tr>
</tbody>
</table>

DLC – Deep Learning Contouring  MAC – Multi-atlas Contouring
Segmentation challenge - Champions

- Xiao Han (Elekta Inc.)
  *Automatic Thoracic CT Image Segmentation using Deep Convolutional Neural Networks*

- Xue Feng (University of Virginia)
  *A 3D UNet based thoracic segmentation framework using cropped images*

- Bruno Oliveira (University of Minho)
  *Automatic Multi-organ Segmentation in 3D Computed Tomography*

---

**Table 1. Intertester differences in segmentation of organs at risk (OARs) for the analyzed metrics. HD95, 95% Hausdorff distance; MSD, mean surface distance.**

<table>
<thead>
<tr>
<th>OAR</th>
<th>Dice</th>
<th>HD95 (mm)</th>
<th>MSD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung left</td>
<td>0.956 ± 0.019</td>
<td>5.17 ± 2.73</td>
<td>1.51 ± 0.67</td>
</tr>
<tr>
<td>Lung right</td>
<td>0.955 ± 0.019</td>
<td>6.71 ± 3.91</td>
<td>1.87 ± 0.87</td>
</tr>
<tr>
<td>Heart</td>
<td>0.931 ± 0.015</td>
<td>6.42 ± 1.82</td>
<td>2.21 ± 0.59</td>
</tr>
<tr>
<td>Esophagus</td>
<td>0.818 ± 0.039</td>
<td>3.33 ± 0.90</td>
<td>1.07 ± 0.25</td>
</tr>
<tr>
<td>Spinal cord</td>
<td>0.862 ± 0.038</td>
<td>2.38 ± 0.39</td>
<td>0.88 ± 0.23</td>
</tr>
</tbody>
</table>
Fig. 4. The Dice values achieved by the seven methods for the evaluated organs in the in vivo context. The reference: Dice value computed from the intra-subject variability in manual segmentation, for which the normalized score is 1.0, is shown as the dashed line. (a) Lungs (left & right), (b) heart, (c) spinal cord, and (d) esophagus. (Color figure can be viewed at wileyonlinelibrary.com)
Summary

• Automated segmentation has shown promising performance in contouring for treatment planning

• Improvement on robustness, accuracy and throughput is still needed:
  • Consensus on contouring and benchmark database
  • Standardization of imaging acquisition; improvement of image quality; combination of multiple image modalities
  • Advancement in model and machine-learning based algorithms
  • Quality metrics and QA tools for spatial and dosimetric uncertainties
  • Effective translation from research to clinic with sufficient user training
Acknowledgement

- Dr. Dan Ruan
- Dr. Ke Sheng
- Brad Stiehl