Automated contour segmentation for treatment planning: challenges and potentials

Minsong Cao, Ph.D.

Department of Radiation Oncology
University of California, Los Angeles
UCLA

Learning objectives

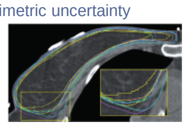
- To review the current state-of-the-art methods on automated contour segmentation for treatment planning
- To understand the challenges and potentials on automated contour segmentation

Disclosure

- None

Background

- Clinical contouring is critical for treatment planning:
  - directly impact dosimetry quality and clinical decision
  - time consuming and labor intensive
- Contour is one of the largest sources of dosimetric uncertainty
  - contour error and variation
  - quality of contouring:
    - spatial accuracy
    - dosimetric accuracy



Jameson M. et al. J Medl Img Radit Onc 54 (2010)

David Geffen School of Medicine

UCLA Health

Automated contour segmentation

- Seek to reduce time and inter-observer 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

David Geffen School of Medicine

UCLA Health

Automated segmentation methods

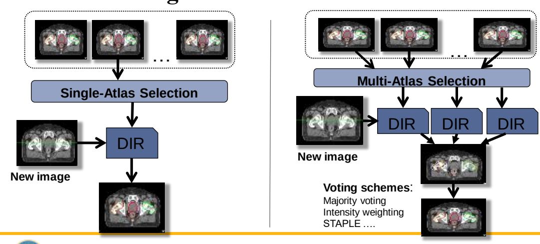
- Non prior-knowledge
  - Directly based on image voxel intensities and/or gradient
  - High contrast structures e.g. lung, bone, air cavity
- Prior-knowledge
  - Atlas based segmentation
  - Statistical model based segmentation: Shape (SSM) or Appearance (SAM)
  - Machine learning based segmentation
  - Hybrid segmentation

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

David Geffen School of Medicine

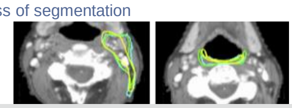
UCLA Health

Atlas based segmentation



Performance of atlas based segmentation

- Quality of atlas images and reference contours
- Atlas selection strategy: robust metric
- No consensus on database size
- Multiatlas can improve robustness of segmentation
  - Prone to topological error
  - Voting scheme is crucial
- Combination of Multimodality images (MRI and CT)



Single atlas - yellow Multiatlas - turquoise Reference -green

D. Teguh et al. Int. J. Rad Onc Biol. Phys., Vol. 81(4), 2011

Atlas based segmentation - DIR

- Quality of segmentation highly relies on deformable image registration (DIR)
- Ground-truth is not available
- Many different approaches and transformation modes

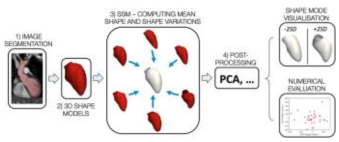
TABLE II. Commonly used transformation models. (N = number of voxels in an image).

Class	Transformation	Maximum dimensionality of transformation	Description
Geometric	Rigid	6	Allows translation in 3 directions and rotations about 3 axes
	Affine	12	In addition to translation/rotation, allows uniform scaling and shear (e.g., parallel lines stay parallel)[44-49]
	Free-form	3N	Local, voxel-based deformation, often regularized by a smoothing parameter
	Global spline-based methods (e.g., thin plate splines)	3N	Parameterizes deformation using a parametric grid of basis function control points with constrained global influence (e.g., deformation is global)
	Local spline-based methods (e.g., B-spline)	3N	Parameterizes deformation using a weighted grid of control points of basis functions with local influence (e.g. deformation is local)
Physical	Viscous/elastic/optical flow (e.g., demons)	3N	Spatially variant voxel displacement voxel displacement by a vector field in a deforming medium, by intensity gradients (deformation is local)
	Finite element methods (FEM)	3N	Spatially variant voxel displacement voxel displacements governed by biomechanical tissue properties (deformation is local)

Brock et al. Med. Phys. 44 (7), July 2017

Statistical model based segmentation

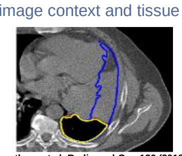
- Confine the segmented contours to anatomically plausible shape or appearance
- Require training dataset to characterize variation of shape or appearance of structure
- Fit the test image to the model based on image intensities, gradients, features etc.



Biglino et al. Heart, 2016 0:1–6

Machine-learning based segmentation

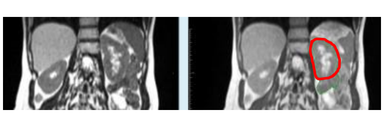
- Outstanding performance in classification, detection, pattern recognition
- Automatically learn priors for structures or image context and tissue appearance
- Require training and significant computational resource
- Usually combined with shape model or atlas based methods

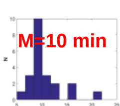


Lustberg et al. Radio and Onc 126 (2018)

Segmentation for adaptive planning

- Intra-object segmentation for anatomy at two different time points
- Deformable image registration is the most popular method
- Time constraints require very robust and accurate segmentation





Lamb J, Cao M, Kishan A, et al. Cureus 9(8): e1618


Contents lists available at ScienceDirect
Artificial Intelligence in Medicine
journal homepage: www.elsevier.com/locate/aiim
ELSEVIER
Methodological Review
A review of segmentation and deformable registration methods applied to adaptive cervical cancer radiation therapy treatment planning
Soumya Ghose, Lois Holloway, Karen Lim, Philip Chan, Jacqueline Veera, Shalini K. Vinod, Gary Liney, Peter B. Greer, Jason Dowling

- 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

David Geffen School of Medicine
UCLA Health
13

Evaluation of segmentation performance

- Geometric
  - Moment based
    - Center/Volume of structure
  - Overlap based
    - Dice similarity coefficient
  - Distance based
    - Average/maximum distance
  - Intra-observer variability
- Dosimetric
  - Dose optimization
  - Dosimetric metrics (DVHs)
  - Clinical decision

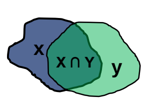
Sharp et al. Medical Physics, Vol. 41, No. 5, 2014

David Geffen School of Medicine
UCLA Health
14

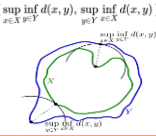
Geometry based evaluation

- Dice similarity index (DSI)
  - Insensitive to large structure
  - insensitive to fine details
- Hausdorff distance (HD)
  - Sensitive to small regions
  - Usually use 95% percentile
- May not correlate with each other
- Do not relate to dosimetry!

$$D = \frac{2|X \cap Y|}{|X| + |Y|}$$



$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\},$$



David Geffen School of Medicine
UCLA Health

Inter-observer variability

- A single manual contour may not truly represent the ground truth
- Inter-observer and Intra-observer contour variations exist
- 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

Medical Physics

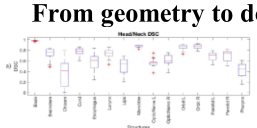
Vision 20/20: Perspectives on automated image segmentation for radiotherapy

Gregory Sharp, Karl D. Fritscher, Vladimir Pekar, Marta Peroni, Nadya Shusharina, Harini Veeraraghavan, Jinzhong Yang

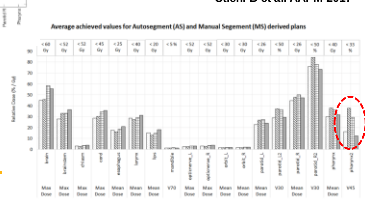
Table I. Inter-rater variability estimates and automatic segmentation accuracy of selected organs.

Site	Evaluation	Inter-rater variability	Automatic segmentation accuracy
Bladder	Dice	0.77 ± 0.06 (Ref. 45) 0.90 (Ref. 46) 0.94 (Ref. 47)	[0.72,0.93] (Ref. 48) 0.86 ± 0.17 (Ref. 47)
Bowel	Dice		[0.83,0.88] (Ref. 48)
Brain	Dice	0.98 ± 0.01 (Ref. 49)	0.96 (Ref. 50)
Brainstem	Dice CL	0.83 ± 0.10 (Ref. 51) 0.58 ± 0.2 (Ref. 52) 0.66 ± 0.17 (Ref. 49) 0.23 (Ref. 53)	0.86 ± 0.03 (Ref. 54) [0.77,0.81] (Ref. 48) 0.84 (Ref. 50)
Breast	Jaccard	0.81 ± 0.14 (Ref. 55)	
Esophagus	Dice Jaccard	0.7 ± 0.1 (Ref. 56)	[0.62,0.68] (Ref. 48)
Eyes	Dice	0.83 ± 0.09 (Ref. 51)	0.84 (Ref. 50)
Femoral heads	Dice		[0.92,0.94] (Ref. 48)
Heart	Dice Jaccard	0.87 ± 0.17 (Ref. 55)	[0.87,0.88] (Ref. 48)
Parotid gland	Dice Jaccard CL	0.85 (Ref. 59) 0.66 ± 0.1 (Ref. 52) 0.76 ± 0.08 (Ref. 49) [0.43,0.86] (Ref. 57) 0.60 (Ref. 53) [0.26,0.73] (Ref. 60)	0.85 ± 0.03 (Ref. 61) [0.73,0.79] (Ref. 48) 0.74 (Ref. 50)
Neck lymph nodes (stations)	Dice		0.74 (Ref. 50) [0.49,0.75] (Ref. 58)
Optic nerve	Dice	0.50 ± 0.17 (Ref. 59)	
Parotid gland	Dice Jaccard CL	0.85 (Ref. 59) 0.66 ± 0.1 (Ref. 52) 0.76 ± 0.08 (Ref. 49) [0.43,0.86] (Ref. 57) 0.60 (Ref. 53) [0.26,0.73] (Ref. 60)	0.85 ± 0.03 (Ref. 61) [0.73,0.79] (Ref. 48) 0.74 (Ref. 50)

From geometry to dosimetry



Stiehl B et al. AAPM 2017



Auto-segmentation Challenge

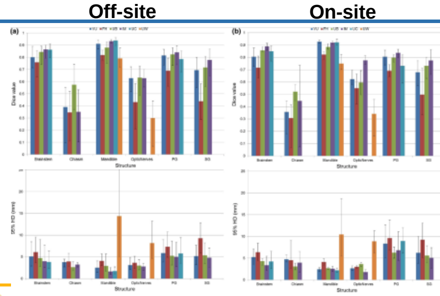
- Allows assessment of state-of-the-art segmentation methods under unbiased and standardized circumstances:
  - The same datasets (training/testing)
  - The same evaluation metrics
- Head & Neck Auto-segmentation Challenge at MICCAI 2015 conference
  - Date from RTOG 0522 clinical trial
  - 25 datasets 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.

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

Raudaschl et al.: Medical Physics, 44 (5), 2017

Off-site **On-site**



Raudaschl et al.: Medical Physics, 44 (5), 2017

Mandible:

AAM - active appearance model
SSM - statistical shape model

- Exclusion of teeth
- Image artifacts from dental implant

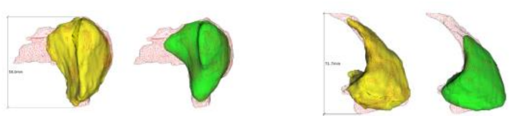
Team	IM	FH	UW
Method	AAM	SSM	Image Processing
Dice	0.93	0.785	0.728
95%HD(mm)	2.041	5.919	29.458

Raudaschl et al.: Medical Physics, 44 (5), 2017

Parotid glands:

- Large shape variation
- Poor soft tissue contrast
- Heterogeneous tissue including vessels and ducts

Team	Dice	Max HD [mm]
UB	0.814	12.000
IM	0.826	30.490



Raudaschl et al.: Medical Physics, 44 (5), 2017

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
- Intra-observer contour variability considered
    - 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

http://autocontouringchallenge.org

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

David Geffen School of Medicine

UCLA Health

Acknowledgement

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

David Geffen School of Medicine

UCLA Health