Contour Assessment for Quality Assurance and Data Mining

Tom Purdie, PhD, MCCPM

Objective

Understand the state-of-the-art in contour assessment for quality assurance including data mining-based techniques

Outline

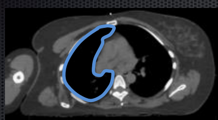
Applications of automated contour assessment
- classification for data mining
- error detection for quality assurance plan review process

Gathering and using data
- supervised vs unsupervised learning contour models
- class imbalance

Methods for assessing contour quality
- machine learning
- segmentation

Clinical examples and potential pitfalls

Contour Quality Assurance
Classification for Data Mining



left lung
left_lung
Left
LT Lung
Lung
LEFT
Left Lung
LUNG
LLUNG

Automated contouring error detection based on supervised geometric attribute distribution models for radiation therapy: A general strategy

Hsin-Chen Chen, Jun Tan, Steven Dolly, and James Kavanaugh
Department of Radiation Oncology, Washington University, St. Louis, Missouri 63110

Mark A. Anastasio
Department of Biomedical Engineering, Washington University, St. Louis, Missouri 63110

Daniel A. Low
Department of Radiation Oncology, University of California Los Angeles, Los Angeles, California 90095

H. Harold Li, Michael Altman, Hiram Gay, Wade L. Thorstad, Sasa Mutic, and Hua Li[a]
Department of Radiation Oncology, Washington University, St. Louis, Missouri 63110

(Received 18 June 2014; revised 24 November 2014; accepted for publication 2 January 2015; published 29 January 2015)

Medical Physics, 2015

Automatically detect radiation therapy organ at risk contouring errors using novel geometric attribute distribution (GAD) models

Contour Quality Assurance
Error Detection

- Delineated structures are subject to errors from a number of sources:
  - limitations of current imaging techniques in visualizing human anatomy (e.g., insufficient contrast, resolution, or both)
  - inherent anatomical variability among individuals for manual contouring
  - inaccurate automated contouring

Contour Quality Assurance

Error Detection

- Delineated errors are a result of:
  - Mislabeling
  - Missed slices
  - Overlapped structures
  - Isolated voxels or small regions
  - Extremely large or small slices
  - Structure volume and location distortions

Geometric Attribute Distribution (GAD) models

Inputs:

- Geometric attributes (centroid, volume, and shape)
- Spatial relationship of neighboring structures
- Anatomical similarity of individual contours among patients

Models characterize:

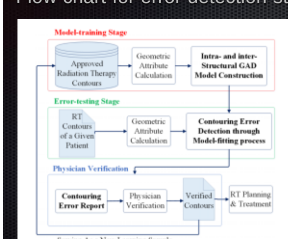
- Inter-structural centroid and volume variations
- Intra-structural shape variations

Model attributes:

- Scalable and deformable
- Constrained by the respective principal attribute variations calculated from the training sets based on clinically verified contours

Geometric Attribute Distribution (GAD) models

Flow chart for error detection strategy



Geometric Attribute Distribution (GAD) models

Iterative weighted model
- Receiver Operating Characteristic (ROC) curve used to select optimal system parameters

Data
- 44 patient datasets (29 training and 15 testing)
- 1350 testing data-based attribute samples (1123 correct and 227 incorrect attributes samples) were synthesized from accurate contours

Geometric Attribute Distribution (GAD) models

Quantitative analysis for contouring error detection
- Balanced accuracy (BA) → avoiding false classification
- Recall → identify positive samples
- Specificity → identify negative samples
- F-score → harmonic mean of Precision and Recall

Results
- BAs of 0.911 (centroid) and 0.842 (volume)
- BAs of 0.881 (brainstem shape) and 0.856 (left parotid shape)

Advantages of the method
- Training of incorrect contours is not required
- Enables incorrect slice identification

A framework for automated contour quality assurance in radiation therapy including adaptive techniques

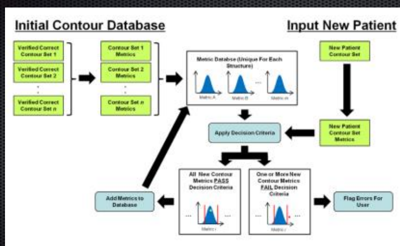
M B Altman, J A Kavanaugh, H O Wooten, O L Green, T A DeWees, H Gay, W L Thorstad, H Li and S Mutic

Department of Radiation Oncology Washington University in St. Louis 4921 Parkview Place Campus Box 8224 St. Louis, MO 63110, USA

Physics in Medicine and Biology, 2015

Automated contour QA methodology

Verified contours in database used for decision metrics



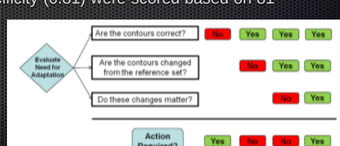
Population Derived Metrics

Four fundamental types of metrics are used by the method:

- Size/Shape metrics (example: volume) describe the geometric size and shape of a contour or structure in 2- or 3-dimensions,
- Positional metrics describe the relative position of one contour to another (example: centroid- to-centroid (C-to-C) distance),
- Image-based metrics relate to some property of the pixel values underlying the structure in the image (example: average pixel intensity in a structure)
- Binary-Type metrics which have only two possible outcome states (example: are gaps (missing planes) present in a structure, yes or no).

Decision Criteria

- Heuristically determined window of ±1.96σ to determine "passing' structures where σ is the standard deviation of a given metric
- Method tested using a double-blind approach
- Users generated "engineered" contouring errors from verified accurate contours
- Sensitively (0.95) and specificity (0.81) were scored based on 81 structures with 42 errors



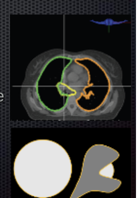
Groupwise Conditional Random Forests for Automatic Contour Classification and Quality Assessment for Radiotherapy Planning

Chris McIntosh*, Igor Svistoun, and Thomas G. Purdie

IEEE Transactions on Medical Imaging, 2013

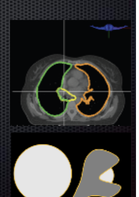
Terminology

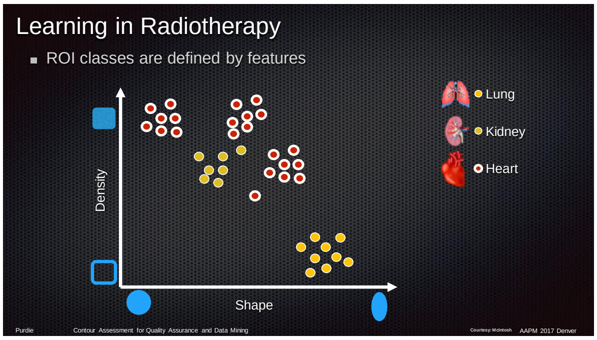
- Region of interest (Segmentation)
  - Delineation of targets and organs at risk, characterized as a set of voxels with binary labels indicating object or background
- Features
  - Measurements and descriptors of shape or intensity
  - Examples: Volume, smoothness, intensity, etc.
- Observed variables
  - Measurements or features that can be calculated from the data
- Latent variables
  - Unobserved variables that we seek to infer from the observed variables
  - Examples: Anatomical label for an ROI, volume of a particular ROI

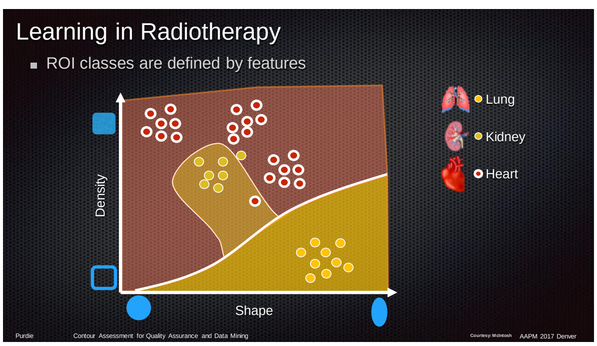


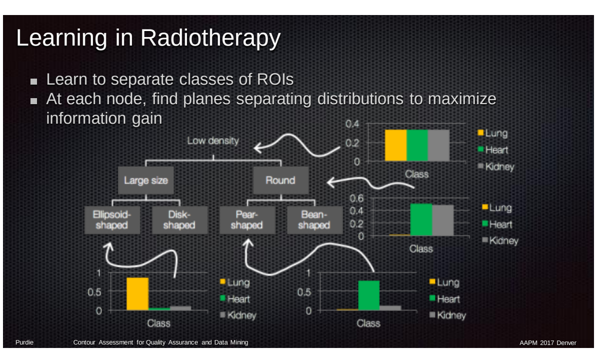
Terminology

- Conditional probability distribution
  - Distribution that estimates how well a latent variable is predicted by observed variables
  - How well does an ROI's volume predict the lung class label?
- Learning
  - The act of learning the probability distribution using training examples
  - If it's a very large ROI, it's probably a lung
- Inference
  - Selecting the most likely outcomes given the observed data and the learned probability distributions
  - Since its large, and dark, it is a lung



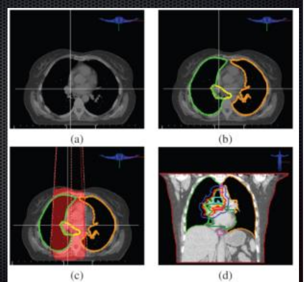

Learning in Radiotherapy
ROI classes are defined by features
Lung
Kidney
Heart
Density
Shape
Purdie
Contour Assessment for Quality Assurance and Data Mining
Courtesy: McIntosh
AAPM 2017 Denver


Learning in Radiotherapy
ROI classes are defined by features
Lung
Kidney
Heart
Density
Shape
Purdie
Contour Assessment for Quality Assurance and Data Mining
Courtesy: McIntosh
AAPM 2017 Denver


Learning in Radiotherapy
Learn to separate classes of ROIs
At each node, find planes separating distributions to maximize information gain
Low density
Large size
Round
Ellipsoid-shaped
Disk-shaped
Pear-shaped
Bean-shaped
Lung
Heart
Kidney
Class
Purdie
Contour Assessment for Quality Assurance and Data Mining
AAPM 2017 Denver

Groupwise Information

- Classes of ROIs commonly co-occur in plans
  - Lungs with hearts
- Groups of ROIs have contextual features
  - Heart is near the lungs
  - Left and right lung are of similar size
- Build novel algorithm: Groupwise Conditional Random Forests to model this information



Validation Results

	Overall	Normal Tissue	Targets
Number of Treatment Plans	1574		
Number of Plan Classes	7		
Number of ROI Classes	77		
Number of ROIs	17579	9125	8454
Accuracy	91.6 %	97.4 %	85.3 %

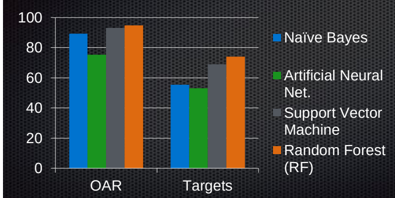
Validation Results

	Overall	Overall Updated
Number of Treatment Plans	1574	15000 +
Number of Plan Classes	7	121
Number of ROI Classes	77	331
Number of ROIs	17579	52000 +
Accuracy	91.6 %	89.4 %

Validation Results

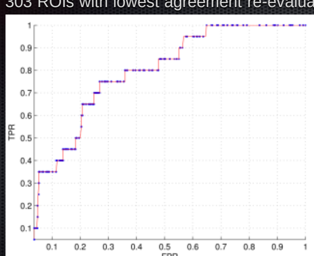
17,579 ROIs from 1574 clinically delivered plans

- 77 distinct ROI classes
- Train 60%, test 40%; stable for less training data



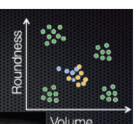
Error Detection

- 303 ROIs with lowest agreement re-evaluated



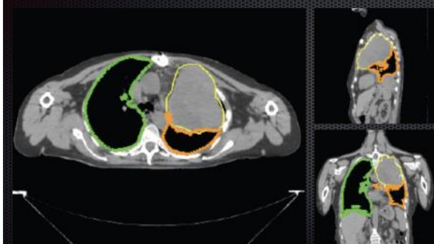
Target ROI Classification

- Need contextual information
- Lung target appears with lung ROI



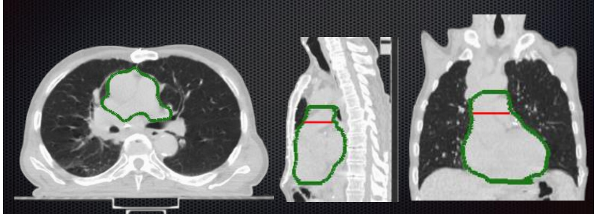
False Positive Organ Classification

- Supervised vs Unsupervised Learning



False Positive Organ Classification

- Supervised vs Unsupervised Learning



Summary

- Automated contour assessment will have important implications in routine QA and data mining activities
- Proof of principle methods are being developed with promising preliminary results
- Error detection and ROI classification validation results can be well described using common metrics that we employ already
- There is tremendous potential to apply these automated methods to labor intensive treatment approaches i.e. adaptive, and may ultimately be required for these treatment approaches to be successful