SAGE: An Approach and Implementation Empowering Quick and Reliable Quantitative Analysis of Segmentation Quality

Danna Gurari¹, Seule Ki Kim¹, Eugene Yang¹, Brett Isenberg², Tuan A. Pham³, Alberto Purwada³, Patricia Solski³, Matthew Walker³, Joyce Y. Wong², Margrit Betke¹
¹ Department of Computer Science; ² Department of Biomedical Engineering; ³ Department of Biology

Scope: Segmentation Analysis
Select measure to produce score indicating how similar a query segmentation is to a gold standard segmentation.


Motivation

Computer Vision
• Demonstrate effective algorithm design with good scores...
  Which evaluation measure?

Biology/Biomedical Engineering
• Trust analyses derived from segmentations...

Data Acquisition (Videos/Images) → Segmentation → Classification → Tracking → Shape/Behavior Analysis

Which gold standard?

1 2 3 4
RAW A B C D E

Acknowledgments NSF HCC Grant IIS-0910908

Key Contribution #1: Segmentation Analysis Model
Observing that the performance score depends on the gold standard segmentation, we propose an analysis approach that introduces the consideration of how to establish the gold standard. It links annotation collection approaches with gold standard generation methods and evaluation algorithms into a unified framework we call SAGE (Segmentation Annotation Collection, Gold Standard Generation, and Evaluation).

Key Contribution #3: How to Model Gold Std Segmentation
Case studies showing the impact of annotation collection and gold standard generation on establishing trusted (i.e., high-consensus) gold standard segmentations in practice.

<table>
<thead>
<tr>
<th>ID</th>
<th># of Images</th>
<th>Imaging Modality</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35</td>
<td>Phase Contrast</td>
<td>Neonatal rat smooth muscle cells</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>Phase Contrast</td>
<td>Fibroblast cells of the Balb/c 3T3 mouse strain</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>Phase Contrast</td>
<td>Vascular smooth muscle cells from rabbit aortas</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>MRI</td>
<td>Left renal artery and the iliac bifurcation of a New Zealand White Rabbit</td>
</tr>
</tbody>
</table>

Gold Standard Generation
User can select an original or fused annotation to represent the gold standard. Supported fusion algorithms are STAPLE [Warfield 2004] and TPM [Meyer 2006]

Evaluation

Gold Standard Generation - ImageJ annotations with each person having their annotations calculated against the gold standard. The system was used to calculate the following six evaluation measures:
- Intersection over Union (IoU)
- Jaccard Index
- Dice Coefficient
- Hausdorff Distance
- Surface Distance
- Symmetric Surface Distance

Gold Standard Generation - Used ImageJ

Evaluation Score: 1.0

1) Which annotation tool? Preference for Amira over ImageJ
2) Who annotates? Education and experience not important
3) Should fusion methods be used? Preference for original over fused annotations

Key Contribution #2: Toolbox
Freely available implementation of SAGE: http://www.cs.bu.edu/~betke/SAGE

Annotation Collection

Image Acquisition → Image Collection → Segmentation → Generation → Query Evaluation → Score

Overview of SAGE (yellow) within the context of analyzing a query segmentation.