Evaluation of Segmentation Algorithms in CT Scanning

Seemeen Karimi, Xiaoqian Jiang, Pamela Cosman
University of California, San Diego
La Jolla, USA 92093

Harry Martz
Lawrence Livermore National Laboratories
Livermore, USA 94551

***Abstract*—We developed a method to evaluate the accuracy of segmentation algorithms. Oversegmentation, undersegmentation, missing and spurious labels may all appear concurrently in machine segmented images. Segmentation algorithms make systematic errors and have different optimal operating ranges. Existing methods of segmentation evaluation do not evaluate these details. Our method, based on multiple feature recovery, reports systematic errors and indicates optimal operating ranges of features, besides measuring overall errors. A knowledge of the magnitude and type of errors can be used for tuning or selecting segmentation algorithms. Although our method was developed for CT scanning for security, it is applicable to other fields, including medical imaging, where multi-object feature recovery, non-uniform costs and a knowledge of optimal operating ranges are helpful.**

***Keywords*-segmentation; evaluation; feature recovery;**

I. Background

Quantitative evaluation aids the evolution of segmentation algorithms. Many methods evaluate machine segmentation (MS) against a ground truth (GT) using a distance between the sets of edge pixels [1]. However, edge distance is a poor indicator of feature retrieval of complex shapes in a cluttered setting with variable numbers of labels. Other methods interpret the different labeled regions as clusters, and measure distance between clusterings, or are based on volume overlap [2]. None of these methods evaluate systematic errors for multiple segments, or operating ranges, or allow objects to be prioritized.

II. Methods

Our method compares features between the GT and MS labels. The GT and MS may have different numbers of labels, which are arbitrarily numbered. An optimal one-to-one correspondence between label pairs is established using the well-known Hungarian algorithm, which maximizes the total intersection in voxels of all GT and MS labels. Features are computed within each label. Feature measurements of all the labels in an image are placed into a feature descriptor (vector). A feature recovery scatter (FRS) plot is generated from the MS and GT feature descriptors. A line is fit to the FRS data using a robust fit. A slope $S < 1$ indicates oversegmentation, $S > 1$ indicates undersegmentation, and $S = 1$ indicates random errors. A residual error is computed using the L1 norm divided by the total of the feature in the GT. Moving averages created from the FRS indicate optimal operating ranges. We show volume and mass features here, but the method can use any pointwise or regional features. Mass is computed from the original CT images, and yields different results than volume because luggage objects are heterogenous. We also weight the feature values in the descriptor to prioritize objects, either by user-defined values or by the inverse of standard-deviation, which prioritizes uniform objects. Mass and uniformity features are more relevant to threat assessment than volume. Our data comprises five 3D CT images of suitcases with their GT images, and 24 MS label images from five research groups [3].

III. Results and Discussion

Figure 1 shows FRS plots of volume and uniformity-weighted mass for one of the five MS algorithms along with a robust fit. This algorithm had the smallest volume residual error (0.37), and smallest systematic error (best slope). This algorithm had poor recovery of uniformity-weighted mass, a property that could not be detected by volume overlap. The residual error was smallest (0.33), but not random. The moving average (not shown) showed that this algorithm's feature retrieval improved with object mass.

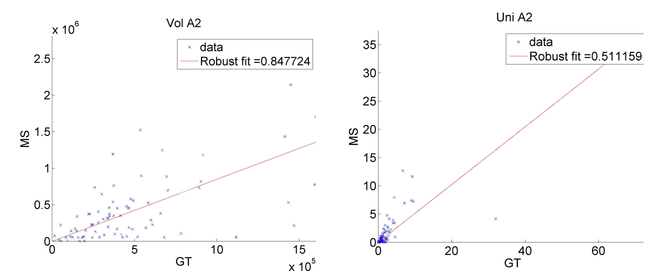

Figure 1. FRS plots for one algorithm (called A2) for volume (left) and uniformity-weighted mass (right). There were 81 ground truth labels.

References

[1] T. Heimann et al, "Comparison and evaluation of Methods for Liver Segmentation from CT datasets," IEEE Transactions on Medical Imaging, vol 28, pp. 1251-65, 2009.

[2] X. Jiang, C. Marti, C. Irniger, and H. Bunke, "Distance measures for Image Segmentation Evaluation," EURASIP J. on Applied Signal Processing, vol. 1, pp. 209, 2006.

[3] Final report on Algorithm Development for Security Applications 6, ALERT, A DHS Center of Excellence, 2011

978-0-7695-4921-7/12 $26.00 © 2012 IEEE
DOI 10.1109/HISB.2012.64

IEEE computer society