

Comparison and Evaluation of Image Segmentation Metrics

Motivation

Segmentation algorithms partition an image or volume (i.e. 2D or 3D image) into multiple homogeneous regions, typically to distinguish between different objects. There is a multitude of segmentation algorithms, so it is an ongoing challenge to evaluate the result of such segmentation algorithms in terms of how well they perform their task in a specific scenario. The typical way to perform such an evaluation is to let an expert user manually segment the given input image. This manual segmentation is then taken as reference, as so-called *Ground Truth*. Subsequently, different segmentation algorithms are run on the same input image. Then, the difference between the results from these algorithms and this Ground Truth is calculated. For this difference, a multitude of measures has been proposed, all with specific advantages and disadvantages.

Goal

The goal in this project is to create a tool that can calculate multiple segmentation quality measures; ideally some voxel-based methods such as Jaccard Index, Segmentation Accuracy, Precision/Recall, but also contour-based methods, see for example Csurka et al. [2]. For this, an image/volume processing toolkit such as ITK [6] can be employed. The tool should be able to calculate the measure for a the whole image and for patches of the volume to be able to assess the local segmentation quality. Datasets, for example volume datasets from CT scans, can be supplied. The tool should be able to answer these questions:

- What correlations are there between the different measures?
- When comparing the qualities of the different measures: What drawbacks do the measures have? Are there specific segmentation issues that are especially well reflected by a certain measure? Are there problems that do not show up in the measures?

Starting Literature

1. T. A. Lampert, A. Stumpf and P. Gañarski, "An Empirical Study Into Annotator Agreement, Ground Truth Estimation, and Algorithm Evaluation," Image Processing, IEEE Transactions on, Vol. 25 (6), pp. 2557-2572, June 2016, doi: [10.1109/TIP.2016.2544703](https://doi.org/10.1109/TIP.2016.2544703).
2. G. Csurka, D. Larlus, F. Perronnin, "What is a good evaluation measure for semantic segmentation?," Proceedings of the British Machine Vision Conference 2013, pp. 32.1-32.11, September 2013, doi: [10.5244/C.27.32](https://doi.org/10.5244/C.27.32).
3. Y. J. Zhang, "A review of recent evaluation methods for image segmentation," Proceedings of the Sixth International Symposium on Signal Processing and its Applications, Kuala Lumpur, August 2001, pp. 148-151, doi: [10.1109/ISSPA.2001.949797](https://doi.org/10.1109/ISSPA.2001.949797).
4. H. Zhang, J. E. Fritts, S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," Computer Vision and Image Understanding, Vol. 110 (2), May 2008, pp. 260-280, doi: [10.1016/j.cviu.2007.08.003](https://doi.org/10.1016/j.cviu.2007.08.003).
5. A. Taha, A. Hanbury, O. Jimenez del Toro, "A formal method for selecting evaluation metrics for image segmentation," IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 932-936, doi: [10.1109/ICIP.2014.7025187](https://doi.org/10.1109/ICIP.2014.7025187).
6. The Insight Segmentation and Registration Toolkit - itk.org

Kontakt

Bernhard Fröhler (bernhard.froehler@fh-wels.at)