Validation of 3-D-Segmentation Results

Frank Schmitt

fschmitt@uni-koblenz.de

University of Koblenz-Landau (Germany)

Institute for Computational Visualistics (http://www.uni-koblenz.de/icv)

Image Recognition Lab

Lutz Priese, Frank Schmitt, Patrick Sturm, Haojun Wang

- Segmentation: Dividing an image into homogeneous, spatially connected regions

- Good segmentation (relative to aim):
  - Only few, big segments should overlap the “object of interest”
  - No segment should overlap with several different objects of interest
- How to judge quality of segmentation?
  - Visual inspection
  - Comparison with a “ground truth”

- Only possible if the objects of interest are clearly recognizable for humans, e.g. traffic signs in color images
- In our group: Manual analysis of 15.000 color images of traffic scenes
- Visual inspection is hard for 3d images
- In MR brain images, white matter and gray matter are hard to distinguish

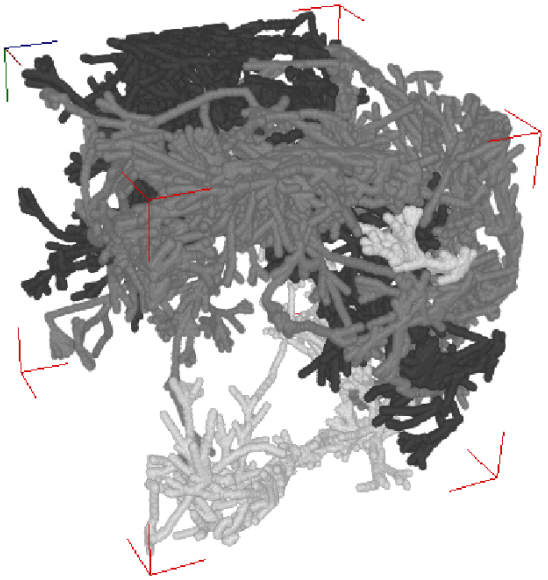
- Normally there’s no ground truth available for real 3d images
- We artificially generated test images $\Longrightarrow$ ground truth known
- Use Lindenmayer system to generate 2-3 fractal-like forms similar to neuronal plexuses
- Give every plexus a distinct gray level (only small difference between plexuses)
- Give the background a common gray level similar to plexuses

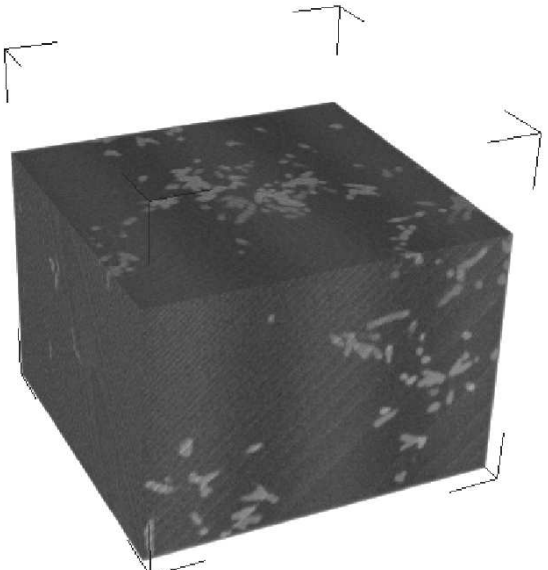
Artificial degradation of images:

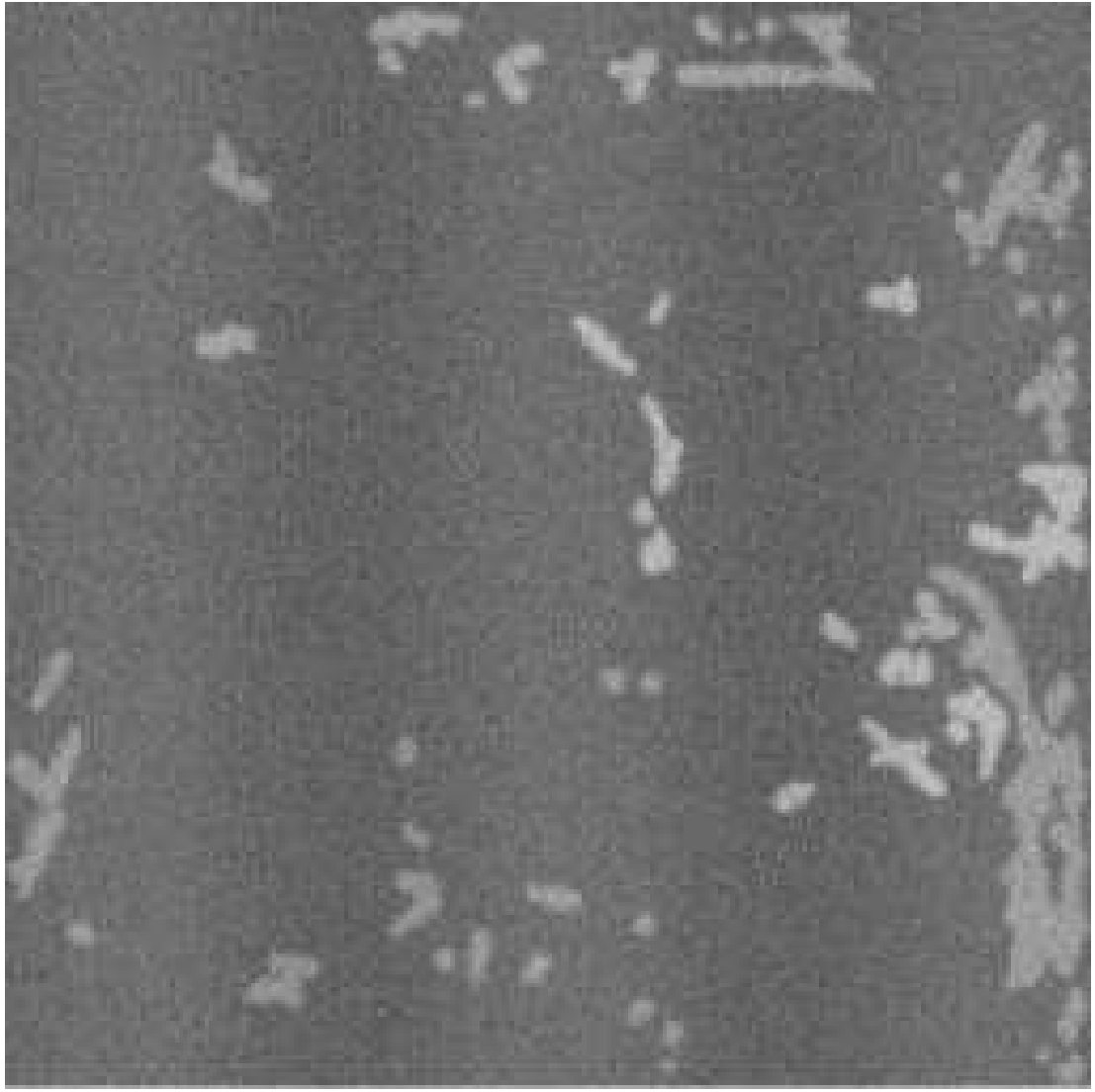
- Add a sinus curve to each image to introduce inhomogeneities
- Blur images by a binomial filter to unsharpen edges
- Add Gaussian noise which is common in MR images

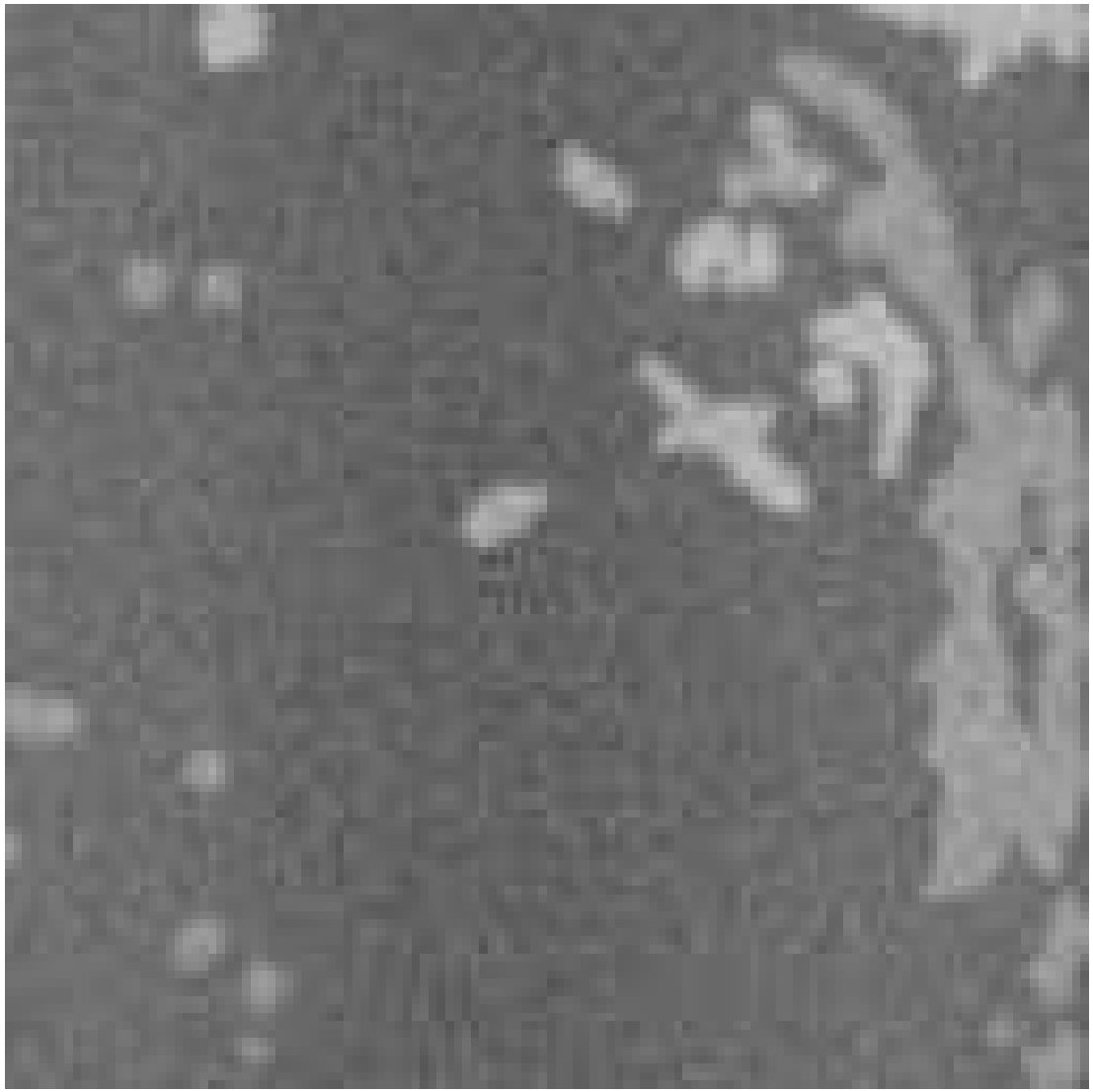
http://www.uni-koblenz.de/∼lb

Download-Section: 3d-Plexus-Images

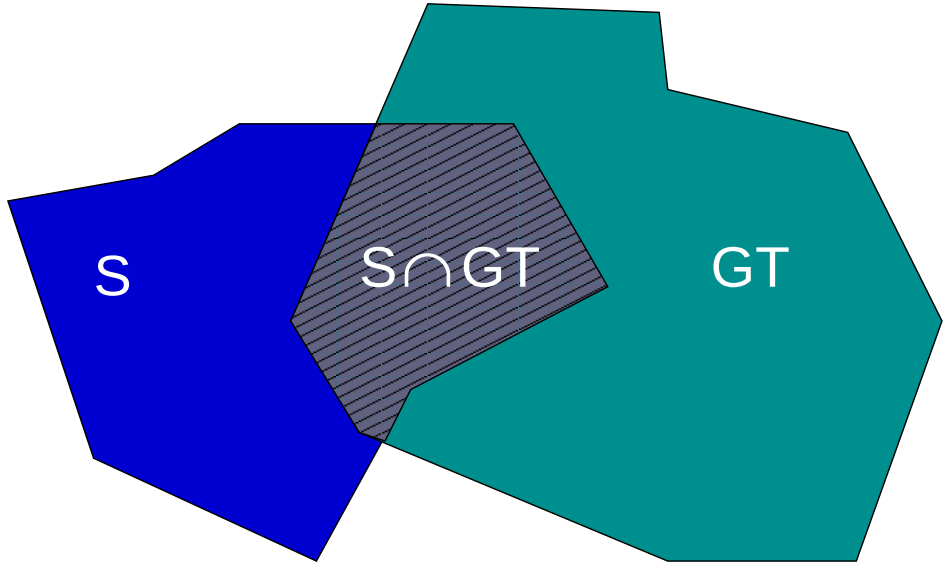






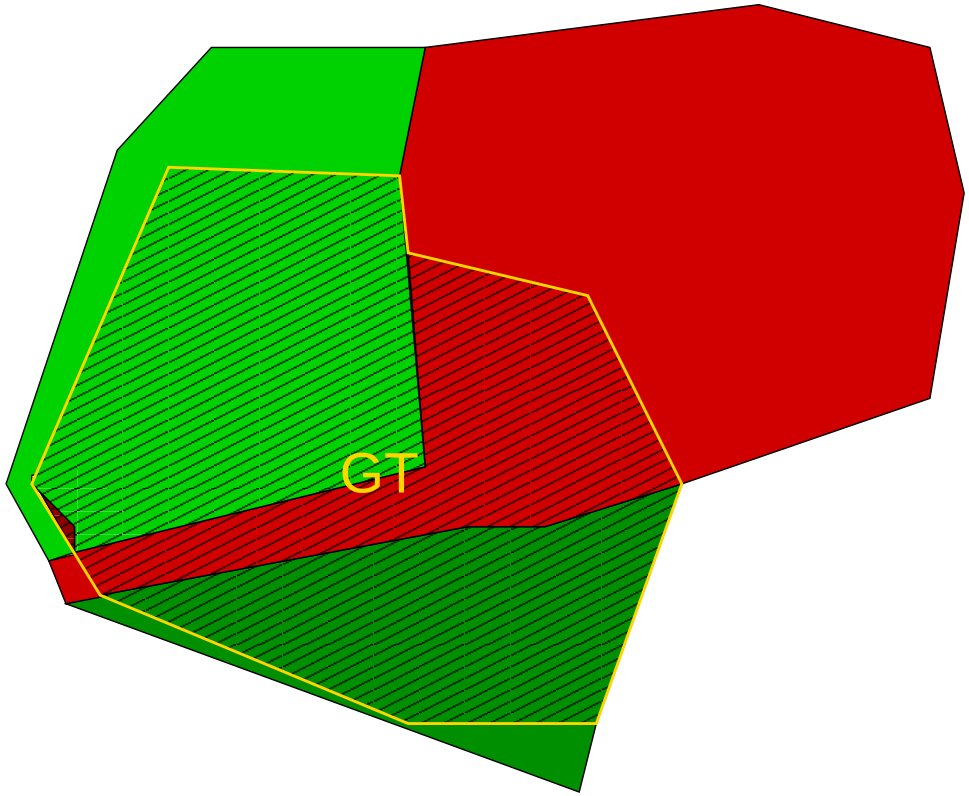


Quality of a computed segment S with respect to a ground truth segment GT:



- Coverability Rate: $CR_{GT}(S) := \frac{Volume(S \cap GT)}{Volume(GT)}$
- Error Rate: $ER_{GT}(S) := \frac{Volume(S \setminus GT)}{Volume(S)}$

- True segments aren’t detected as a whole, we have to aggregate several segments to match them
- Aggregate segments which are not too small and overlap mostly with the “true segment”

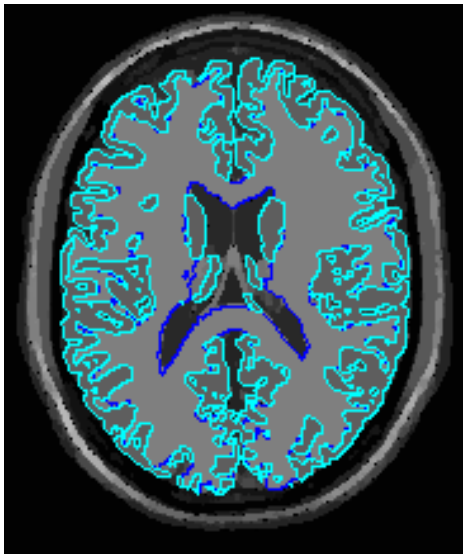


Brainweb project offers Simulated Brain Database:

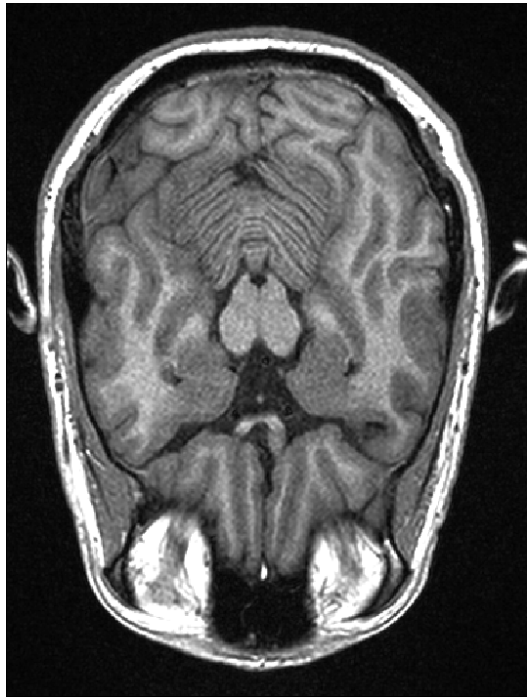
- One image with different noise and bias field levels
- To our knowledge only source of MR brain image with known ground truth

Drawbacks:

- Only one image
- Pure numerical analysis of voxels not sufficient to judge quality

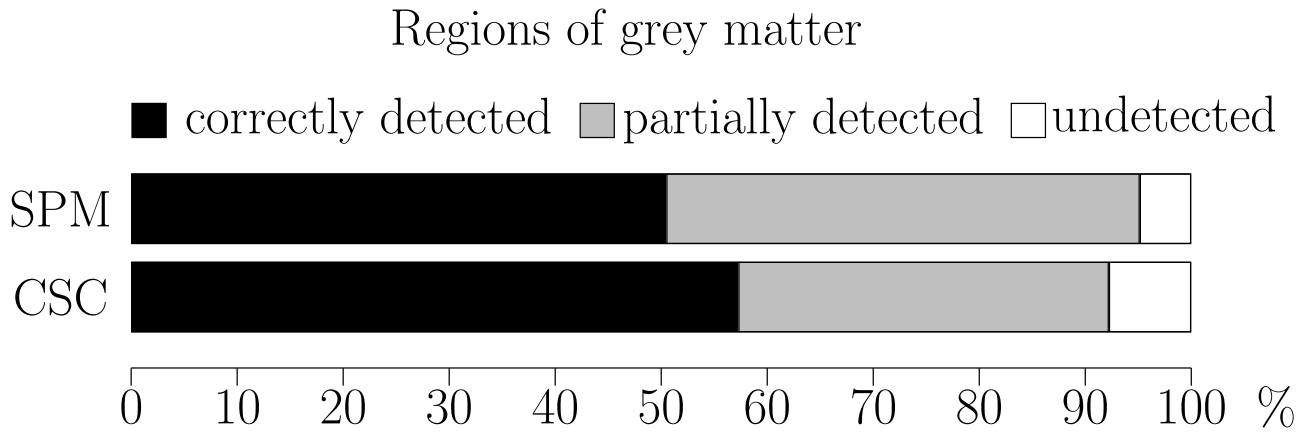


- Acquired 8 images of healthy human brains
- Cooperation with German Armed Forces Central Hospital (GAFCH), Koblenz, Germany
- Manual evaluation of characteristic slices by experts
- Catalog of criteria allows quality rating
- Comparison with standard method SPM



- 8 Criteria
- Criteria 1-3 count how many distinguishable solitary regions belonging to grey matter (which have been observed by the experts in the current slice) are correctly, partly and not detected respectively
- Criteria 4-6 count the same for observable CSF spaces in the current slice
- Criterion 7 counts the absolute number of errors in the borderline between medulla and cortex
- Criterion 8 counts the absolute number of errors in the borderline between cerebrum and subarachnoid space

- Absolute number of errors in the borderline between cerebrum and subarachnoid space:
  - SPM: 15.58 errors
  - CSC: 12.17 errors



- Current work: Analysis of MR brain images including tumors
- Evaluation until now by visual inspection
- New evaluation criteria have to be developed