Slicer3 Validation Description

Release 3.0

Brad Davis, Will Schroeder

June 7, 2007

KITWARE Inc.
Clifton Park, NY

Contents

0.1   Testing and Validation

The core EMSegment algorithm code was ported, with only minor changes, from Slicer2 to Slicer3. However, differences in segmentation results between these versions could still result from programming errors, bug fixes, alternate methods of preprocessing data, incompatible parameter settings, or user error. In order to encourage Slicer2 users to transition to Slicer3, it is necessary to critically asses and minimize these differences.

We plan to address this issue by testing and validating the Slicer3 EMSegment implementation in three phases. In the first phase we will incorporate regression testing into the EMSegment module; these tests will determine whether or not the sourcecode is executing properly and guard against the introduction of bugs. In the second and third phases and we will validate the segmentation results produced by the EMSegment module. The second phase will focus on the segmentation algorithm itself, without any data preprocessing. The third phase will focus on the entire segmentation pipeline, including preprocessing and segmentation. The Slicer2 EMSegment module, which is known to produce good results, will provide the 'ground truth' basis for these validation studies. These phases are described in detail in sections 0.1.1, 0.1.2, and 0.1.3.

The validation studies will be carried out by Sylvain Bouix and others at the Psychiatry Neuroimiging Laboratory (PNL) at Brigham and Women's Hospital. This collaboration will be especially effective since members of the PNL have extensive experience using and assessing the performance of the EMSegment algorithm.

For these tasks it is important to define exactly how segmentation results will be compared. Results will be compared byte-by-byte for regression tests. However, minor bug fixes made recently to the EMSeg-

ment algorithm make this comparison method impractical and unproductive for comparison with Slicer2 results. Instead, we will use the same criterion used by the PNL in 2006 when they compared the results of their command-line-based EMSegment pipeline with the then new Slicer2 EMSegment module. For both implementations, they analyzed the distribution of several segmentation-based measures: volume $(V(A))$, normalized volume difference $\left(\frac{V(A)-V(B)}{V(A)+V(B)}\right)$, and Jaccard coefficient $\left(\frac{V(A\cap B)}{V(A\cup B)}\right)$. Distributions were generated for grey matter, white matter, and CSF.

After analyzing these distributions and investigating outliers, they concluded that the implementations produced sufficiently similar results. In particular, they found absolute volume differences of less than 5% for grey matter and white matter, and less than 10% for CSF. For the Jaccard coefficient they found overlap of grater than 90% for white matter, greater than 80% for grey matter, and greater than 70% for CSF. The different numbers are related to the inherent shape and imaging variability of these structures. For all outliers, a human expert decided if the new segmentation was acceptable or not.

0.1.1   Regression testing

Purpose: Improve the stability and dependability of the EMSegment module source-code by adding built-in regression tests that determine if the segmentation logic is functioning properly.

Explanation: At this time the EMSegment module code is tested manually using a particular testing dataset. Our goal is to add tests to the main Slicer3 ctest testing suite. This will enable users, as well as computers that report to the Slicer3 dashboard, to detect and report errors in the EMSegment code. These tests will provide an important measure of overall functionality. However, they will not provide full coverage of the module's code, or test the GUI itself. The tests must be fast and must run without requiring the user to download a separate testing dataset.

Work Plan: Kilian will provide a means of testing with synthetic images and a known, ground-truth segmentation result. Kitware will generate a separate test based on real medical imagery, and incorporate both tests into the Slicer3 testing suite and Dart dashboard reporting mechanism.

0.1.2   Validation: Core EMSegment algorithm

Purpose: Validate results of the EMSegment *algorithm implementation* in Slicer3.

Explanation: For this first phase of the validation process, we will focus directly on validation of the segmentation algorithm, separate from any data preprocessing (e.g., intensity normalization, atlas-to-target registration).

Work Plan: We expect this process to be iterative. Kitware will provide the PNL with a Slicer3-based command-line implementation of the EMSegment algorithm. The PNL will batch-process their data and analyze the segmentation results. They will compare these results to results produced by the Slicer2 module on identical input data. Any abnormalities will be jointly investigated until their source is understood, fixes are made, and the validation process rerun.

If necessary, Kitware will fix programming errors related to the Slicer3 module or command-line implementation and provide the PNL with an updated version of the software. On a limited basis Kitware will collaborate with Kilian to fix minor bugs within the segmentation algorithm code. Kilian is responsible for major bug fixes or updates to the segmentation algorithm code.

0.1.3   Validation: Segmentation pipeline

Purpose: Validate results of the EMSegment *segmentation pipeline implementation* in Slicer3.

Explanation: For this final stage of validation, we will analyze the results of the entire segmentation pipeline: intensity normalization, atlas-to-target registration, target-to-target registration, and segmentation.

Work Plan: Same as 0.1.2 except the PNL will utilize the entire segmentation pipeline.