Pattern Assisted Cluster Editing & Validation of Automated Cell Segmentation Results

by

Isaac Morse Abbott

An Abstract of a Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the degree of
MASTER OF SCIENCE
Major Subject: ELECTRICAL ENGINEERING

Approved:

Badrinath Roysam, Thesis Adviser

Rensselaer Polytechnic Institute
Troy, New York

November 2008
(For Graduation December 2008)

ABSTRACT

Fluorescent imaging techniques provide biologists and researchers with the ability to do real-time imaging of living organisms and systems due to their non-invasive and non-destructive properties. Today fluorescent microscopes can output huge five-dimensional datasets that occupy whole hard-drives. The task of processing and analyzing this data has become increasingly complicated and the need for smart, automated software tools has never been greater.

Even the best-available automated 2-D and 3-D segmentation algorithms for processing fluorescent images of cell nuclei and other biological objects of widespread interest have a non-zero error rate. Given the large and growing size of these data sets, there is a compelling need to develop efficient methods to identify and correct the automated segmentation errors. There is an equally compelling need to assess the performance of the automated algorithms. Finally, there is a need to analyze the errors made by the automated segmentation algorithms, and to identify opportunities to modify the algorithms to prevent such errors in the future.

The concept of edit-based validation has been proposed to achieve the above objectives simultaneously, recognizing the fact that the manual effort to correct automated system errors should be roughly proportional to the error rate of the automated algorithms. The goal of this thesis is to develop a software system that goes one step further. Specifically, we demonstrate that kernel based pattern analysis algorithms can help identify automated segmentation errors efficiently, allowing “group” or “cluster” editing of multiple errors simultaneously. This dramatically reduces the amount of manual effort required compared to unassisted edit-based methods. We term this methodology PACE (Pattern Assisted Cluster Editing).

To demonstrate the effectiveness of PACE, we developed a software system named FARSIGHT that allows the user to view the image data, and the automatically generated multivariate metadata (mainly object features) using multiple data visualization tools that are all actively linked. This linkage is important, and allows the data to be viewed in

multiple spaces simultaneously, and a cluster of objects to be selected based on any combination of operations in any of the multiple spaces. We term this method “Actively linked multiple spaces architecture (ALISA)”.

The proposed FARSIGHT toolkit will meet this need. The concept of biological objects has been formalized. Biological objects can be generated from any segmentation result type and subsequently visualized and edited via a simple graphical user interface (GUI). Group editing allows biologists to perform edits and edit-based validation on subsets of biological objects. These groups can be constructed using plotting tools, feature queries, and outlier or pattern recognition libraries. In addition to this, the software toolkit incorporates a linked-view architecture where objects can be visualized in image or geometric views, table views, and plot views. Operations performed in one view are immediately visible in all of the views.

The FARSIGHT toolkit provides a scalable, modular, and expandable architecture necessary for the next generation of biological research. The size of datasets will continue to increase and biologists will rely more heavily on statistics, group editing and edit-based validation. The biological object concept will incorporate new types of objects. Methods for detecting outliers and groups will improve and be applied to the system.