Processing Volumetric Data for Correlative Analysis: An Anecdote from a Core Facility

Connon I. Thomas[1] and Naomi Kamasawa[1*]

[1.] Electron Microscopy Core Facility, Max Planck Florida Institute for Neuroscience, Jupiter, FL, USA.
* Corresponding author: naomi.kamasawa@mpfi.org

Advancements in electron microscope (EM) imaging have allowed us to rapidly acquire datasets on the order of terabytes or more, allowing even small labs to produce massive amounts of data in reasonable time frames. However, this data holds little value before it exits the processing pipeline, widely considered to be greatest bottleneck of EM workflows. Image processing and analysis are complicated, with countless algorithms and programs available to perform any one step of the pipeline [1]. Each must be performed carefully yet efficiently to result in a useful and interpretable output.

Prominent labs devoted to EM study often have teams of annotators, coding staffs, and computer clusters to create and apply advanced data processing tools. These teams have made great strides in the field scientifically, even involving high profile companies such as Google [2]. However, many of these advancements remain inaccessible to labs which have few staff and limited resources. Most new scripts and programs require knowledge of coding and software such as Python, Matlab, or C++ to implement, and have yet to be adopted by the greater EM community. The underlying workings of new techniques are also often shrouded by complex mathematics and terminology which require a theoretical background that many microscopists do not possess.

Thankfully, until new high-throughput techniques of image processing are made accessible, there are numerous free or affordable options for data processing available. This paper will describe the pipeline we applied to correlate *in vivo* light microscopy (LM) data of pyramidal neurons in mammalian visual cortex to volumetric EM data from the same neurons, to illustrate how small labs can utilize available software (Fig. 1A). The pipeline begins after collecting large tiled datasets using serial block-face scanning EM which contain the soma and basal dendrites of a cell, and involves necessary steps for extraction of that cell from the raw data and re-identification of previously imaged features. If performed by a single researcher, this process can require several weeks to two months of work—depending on the complexity of the data—for one neuron with several dendrites and dozens of spines (Fig. 1B, C). For this type of project, assignment of steps to several in people in parallel greatly increases throughput.

Following volume acquisition, image data was sent to an institutional server location as it was captured, then was downloaded to a workstation with 64 GB RAM and sufficient graphics capabilities. Images were loaded into TrakEM2 through Fiji for tile montaging, alignment, image inversion, and contrast normalization, which often requires several days. This software works better than many when handling large datasets (~1 TB); nevertheless, it requires large amounts of memory to operate and has a less-than-intuitive UI. ROIs containing dendrites of interest (several per neuron) were cropped from the dataset and exported as image stacks, since few segmentation programs can handle full datasets [3].

Cropped ROIs were imported to Microscopy Image Browser (MIB) [4], where the dendrite and spines of interest were segmented semi-manually from the soma outward utilizing interpolation, superpixels, watershed, and graph cut segmentation. These tools help to shorten this longest step of analysis. The

benefits of this software are that it is free, has built in tools for image processing and model statistics, has an easy-to-learn UI, and has extensive documentation; however, it seems to be limited in the size of data that it can handle (<200 GB). As a work-around, data can be chopped into manageable pieces and worked on separately, then recombined using MIB. Dendritic segmentations could be exported as various different file types and could even be directly rendered in MIB, or in Fiji, Imaris, or Matlab through the software.

For reconstruction and quantification, Amira (Thermo Fisher Scientific) was used to create 3D meshes from cell segmentations and calculate measurements of spine volume, PSD length, and bouton volume. Unfortunately, this software is costly, lacks community support, and is prone to freezing when working with large datasets. We also used open source program named Blender to smooth models, create high quality renderings, movies, and even perform the morphological measurements listed above (Fig. 1C). The application of this software in scientific analysis is still under active development, as its main use is in the graphic arts industry [5].

To correlate cells between EM and LM we used a 3D virtual reality program called syGlass (IstoVisio, Inc.), which allowed for intuitive data exploration and cross-platform feature comparison. It has proved useful for matching individual dendritic spines from LM data stacks to the EM model, which informed final analyses and quantifications. It also contains numerous useful tools and is under active development. Nonetheless, this software is not free and requires a virtual reality workstation.

In summary, this paper hopes to serve three purposes: 1) describe an example workflow that we used to process and output useful biological data; 2) encourage conversation of already available data analysis tools within the EM community that can widen the bottle neck of image processing/analysis, including aspects of different software which work well and those which are lacking; and 3) serve as a call to action to demystify cutting edge techniques in data processing and make them more accessible to the community.

References:

[1] M Kittelmann et al., Journal of Microscopy **263** (2016), p. 200.
[2] M Januszewski et al., Nature Methods **15** (2018), p. 605.
[3] DR Berger et al., Frontiers in Neural Circuits **12** (2018).
[4] I Belevich et al., PLoS Biology **14** (2016), p. e1002340.
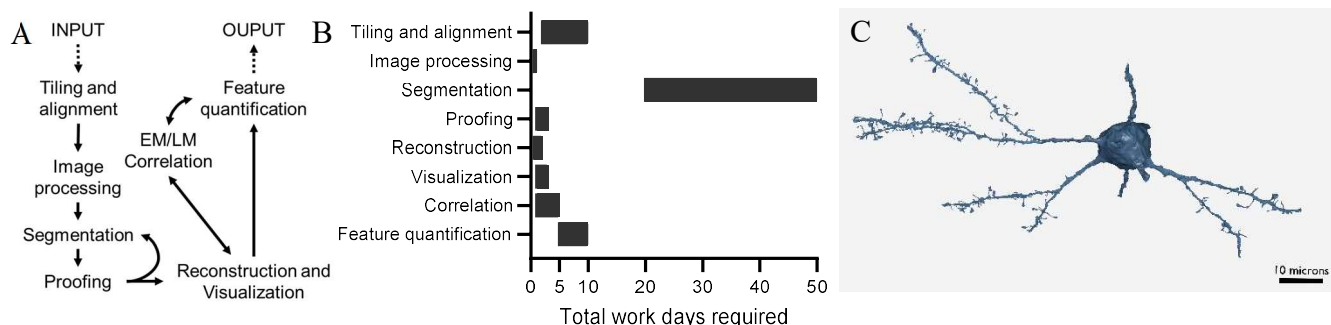[5] C Lee et al., bioRxiv (2019), p. 534479.

Figure 1. A) Flowchart of an example data processing pipeline for correlative analysis. B) Total time required for different steps of the pipeline for one complete neuron. C) Example of a single reconstructed neuron from EM image data, with parts of basal dendrite segments shown.