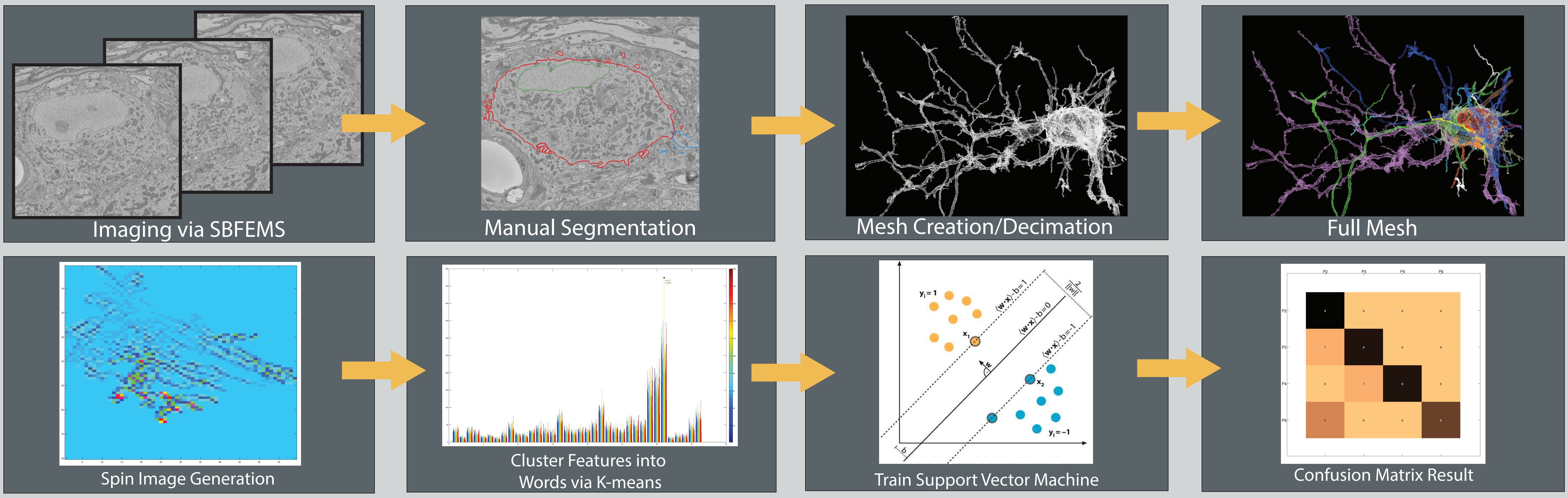
Automatic Classification of Neuron Morphology

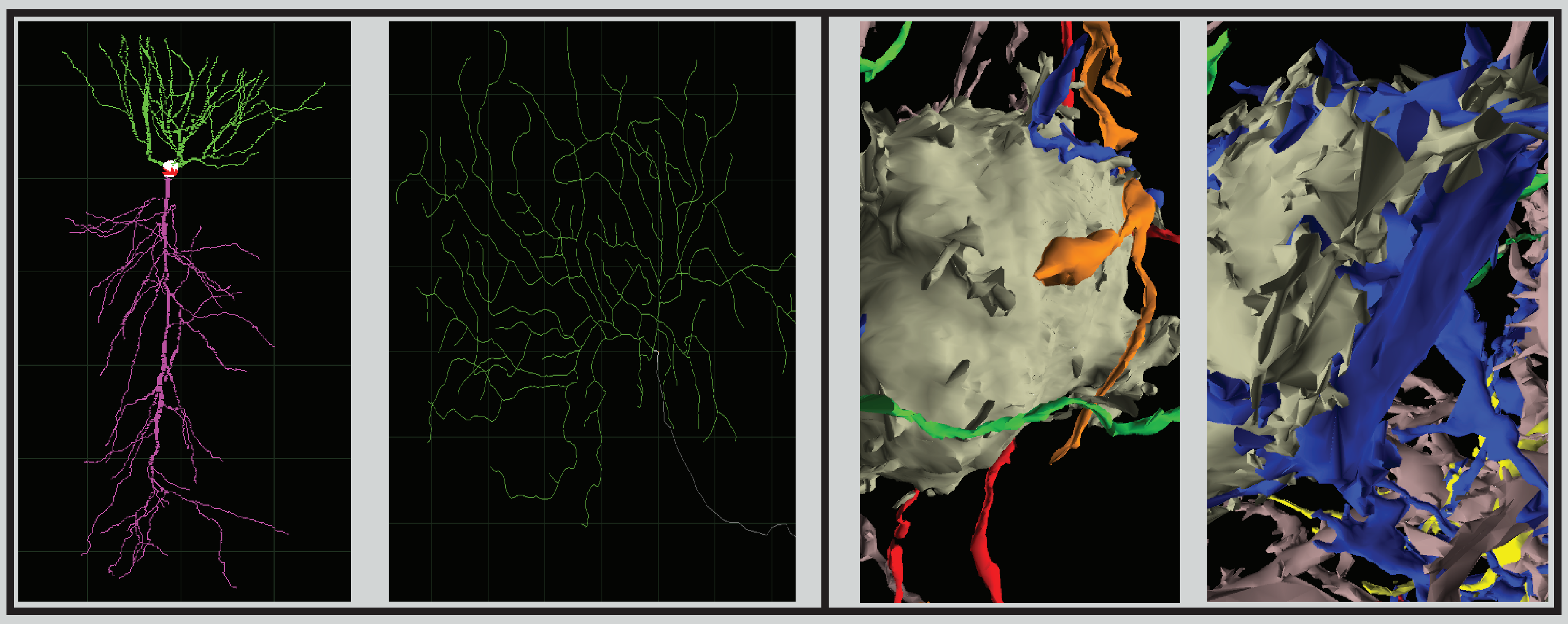
Quinn Jones[1], Michael Morehead[1], George A. Spirou[2], and Gianfranco Doretto[1]
[1]Lane Department of Computer Science and Electrical Engineering and [2]Sensory Neuroscience Research Center, West Virginia University, Morgantown, WV 26506



Problem Statement

With the advent of new microscopy techniques, the production time of high-resolution neuronal cell models is rapidly increasing. As more and more of these neurons are imaged, segmented, and meshed, the registration and classification of each cell will quickly outstrip the ability of human researchers. In 2013, President Barack Obama proposed the BRAIN Initiative, with the eventual goal of creating a functional map of the brain. The National Institutes of Health has recently requested submissions for a variety of topics, including a U01 grant entitled: *Transformative Approaches for Cell-Type Classification in the Brain* (U01) (RFA-MH-14-215).

As a preliminary study for the submission to this grant, and to further the ability to inventory and classify neuronal cells, we've developed a process leveraging techniques from computer vision and pattern recognition to quantify the abilities of a machine learning model for autonomous classification of high resolution cell models.



A comparison of cell types and age.
Far Left: Pyramidal cell morphology. Left: Ganglion cell morphology. Different cell types have drastically different morphologies.
Right: A principal cell at 3 days old (postnatal). Far Right: A principal cell at 6 days old (postnatal).
Notice the much larger synaptic surface area in the older cell.

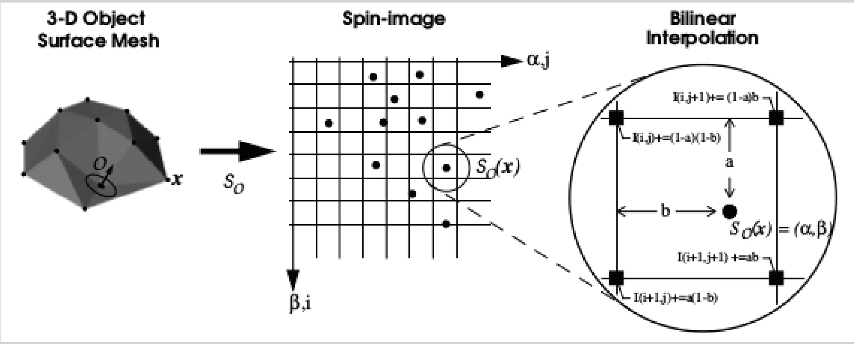
Hypothesis

The bag-of-words model, combined with a three-dimensional spin image and support vector machine, will classify high resolution models by typology.

Methodology

Model Creation

1. Extract necessary electron microscopy data via SBSEM
2. Manual segmentaion of neurons and subparts by hand tracing contours
3. Mesh the contours by connecting neighboring contours.
4. Export to a PCL File



A diagram of the spin image generation process.

Feature Extraction (spin images)

5. Load information from cell files into Point Cloud Library
6. Pick a uniform sample of points to create spin images for
   a. For every point in the model
      i. Use the Spin Map equation (Right) to transform coordinates into spin image space
      ii. Take list of distances and bilinearly interpolate onto a histogram
      iii.Save to a binary file for later use in dictionary learning

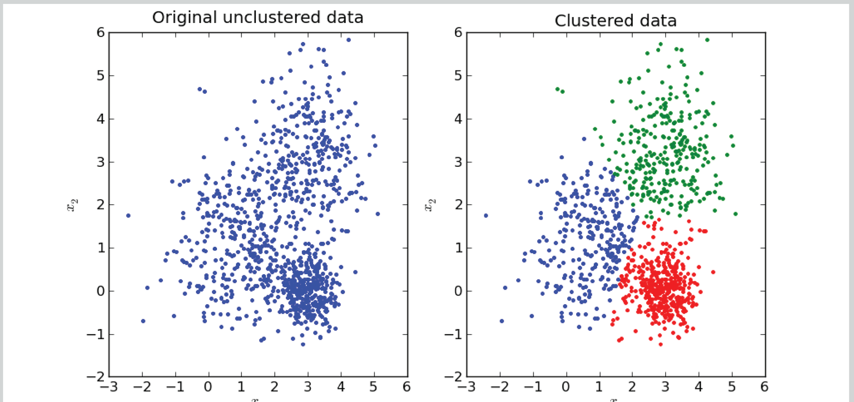
$$S_O: R^3 \to R^2$$

$$S_O(x) \to (\alpha, \beta) = \left(\sqrt{\|x-p\|^2 - (n \cdot (x-p))^2},\ n \cdot (x-p)\right)$$

Transformation of three-dimensional coordinates to cylindrical coordinates.

Dictionary Learning (using k-means)

7. Uniformly sample 1000 spin images from the set of all models
8. Select the number of centroids to be used
9. Use k-means to cluster the images into visual words
   a. Randomize initial centroid positions
   b. Group all data points by proximity to centroids
   c. Find mean of grouped data points
   d. Move centroid to the mean position
   e. Repeat until global minimum is found

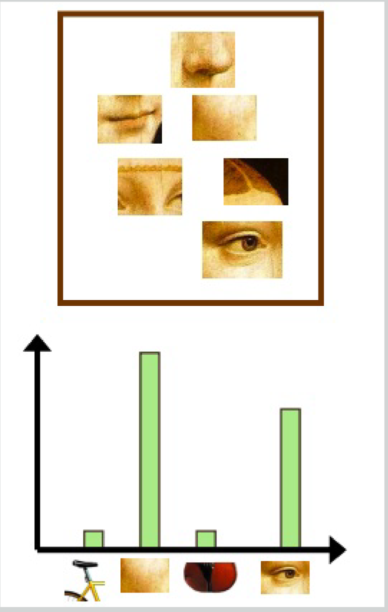


An example of k-means clustering on a generic dataset.

Methodology (continued)

Neuron Representation (bag-of-words model)

10. Count the amount of spin images that occur in each group for each cell
    a. Associate each spin image with the cell that produced it by using unique ID's
    b. Use the result from k-means to create a set of "words," i.e., centroid positions which represent similar cellular features
    c. Count the appearance of each cell's spin images in each feature
    d. Represent the cell as a histogram of the feature counts



Classification (Multiclass-SVM)

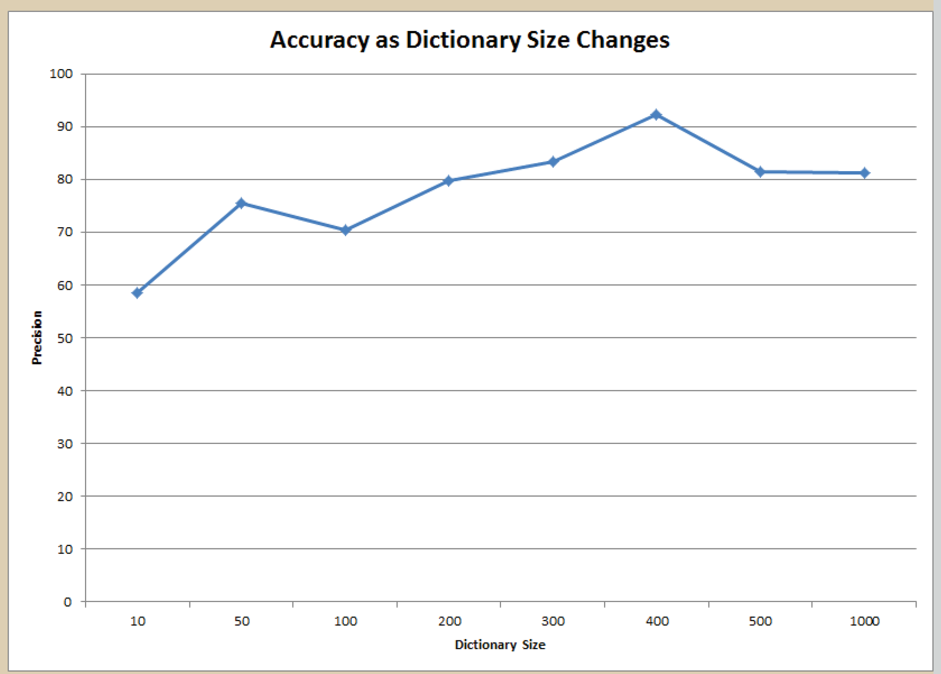
11. 10-Fold cross validation using Multiclass Support Vector Machine (SVM)
    a. Train an SVM using a polynomial kernel on a uniformly sampled tenth of the cells
    b. Use the SVM model to predict the classifications of the remaining 90%
    c. Cycle through each of the ten partitions
12. Average the accuracy across each of the folds to determine the total accuracy

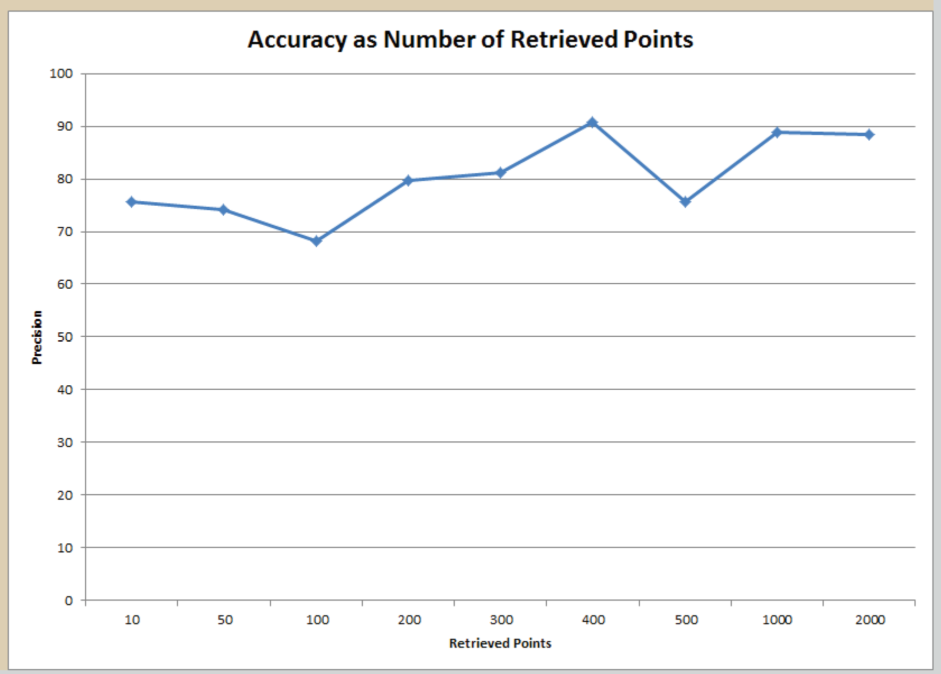
Results

We tested our method using 10-fold cross-validation. In addition there are a number of method paramaters that can be varied. Specifically, we varied the size of the dictionary in one trial and the number of received points in another.

The dictionary size was varied by increasing the number of centroids used in k-means clustering. During the variance testing, the number of retrieved points was fixed to 500. The best result being acheived at 400 visual words with accuracy exceeding **90%**.

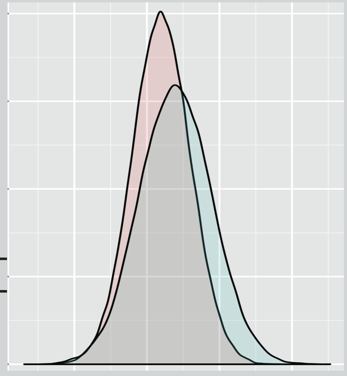
The amount of tested points was varied from 10 to 2000 points. The dictionary size was fixed at 400 visual words. The best result also exceeded **90%** at 400 retrieved points.





Discussion

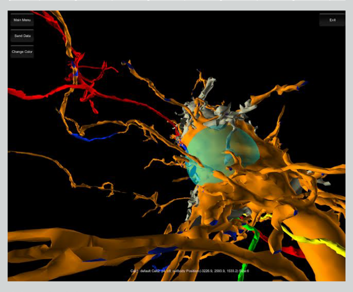
The accuracy results meet what would be expected for a suitable methodology for classifiying neurons automatically. It is unlikely that increasing the dictionary size or retrieved points quantity would yield better results due to the risk of overfitting. We highly recommend this methodology to classify nueron models when the volume of nueron models exceed that which can be analyzed manually.



Grey area indicates the intersection of red and green histograms.

Future Plans:

- Acquire more cells to improve prediction accuracy
- Implement Better Algorithms: Histogram Intersection or Generalized Histogram Intersection
- Include more data beyond morphology, e.g., synapse size, connection map, function attributes



A screenshot from the lab's synapse tagger a new research product featuring Oculus Rift technology. This software allows researchers to "tag" neural syanpses and aproximate their size. Data that could be used to fuel future analysis.

Photo References: Rocca, Luigi, et al. "Fast Neighborhood Search on Polygonal Meshes." Eurographics Italian Chapter Conference. 2011.
Morehead, Michael, Jones, Quinn, et al. BrainTrek – An Immersive Environment for Investigating Neuronal Tissue. In Proceedings of the IEEE Symposium on 3D User Interfaces, pp. 157–158, March 2014.
Weis, Martin, et al. "Comparison of different classification algorithms for weed detection from images based on shape parameters." Bornimer Agrartechn. Ber 69 (2009): 53-64.
Johnson, Andrew E., and Martial Hebert. "Using spin images for efficient object recognition in cluttered 3D scenes." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21.5 (1999): 433-449.
Bellot, Patrice, and Marc El-Bèze. "A clustering method for information retrieval." Technical ReportIR-0199, Laboratoire d'Informatique d'Avignon, France (1999).