T.K. Kozai

**MATLAB problem set VIII: Histology Image Processing**

In this assignment, you will learn to load image sets, prepare images for analysis, and conduct some basic analysis techniques.

**Part 0: Loading the dataset**

The first thing we need to do is to load the image and variables.

```
help imread
```

We also need to enter the scale in microns/pixel. This can be found in the .txt file ( 0.621 um/pixel)
Since these are 12bit images (2^4096), each pixel is represented by a value between 0 and 4095.

Question: What are some example labels that should be used with cell count type analysis?

**Part 1: Cell counting based analysis**

Cell counting based analysis should be conducted for markers that generally have a constant concentration level and/or when using polyclonal antibodies as they do not accurately represent a concentration level. Variability in shape size and slicing geometry (during sectioning) can dramatically impact the quantification and standard deviation of the results

1) We first need to pick the number of bins and the size of each bin. Given that neurons have a radius of ~25 microns
   Pick bin sizes between 25-50 microns and number of bin sizes such that 200<(bin size)*(# of bins)<350

2) Load the image, and obtain the center of the probe as in Part 2

3) Similar to part 2, obtain the edges of the probe.

4) display the image using image or imagesc
   Consider using gray scale color scheme

4) Start a 'for' loop for each bin.

5) For each bin, draw 4 lines using plot and 'hold on' onto the image from step 4, plot the lines in red, cyan, or black (choose a color that least conflicts with your image).
   i) line for minimum radius from the left edge to the right edge.
   ii) line for maximum radius from left edge to right edge.
   iii) line for left edge from minimum radius to maximum radius
   iv) line for right edge from minimum radius to maximum radius

6) Save/print figure to turn in with the .m . Ensure location, number of bins, and bin sizes are included.

**Part 2: Single Value Decomposition (SVD): PCA for 2D arrays**

With electrophysiology we used principal component analysis to separate waveforms into clusters. For PCA, we plotted voltage with respect to time. With images, we will plot intensity levels with respect to time. With a 2D matrix, instead of generating individual principal components for each waveform, we will generate a diagonal matrix. The diagonal elements will be non-negative and decreasing order. In another words, the biggest signals will be near the top of this matrix, and noise will be at the bottom. We want to do this, since unlike with spike waveforms, the intensity of each pixel was acquired at the same time (virtually the same time) for each frame. We want to remove noise and artifacts that equally affect all pixels.

First, lets look at our image stack. Use a image viewer such as ImageJ https://**imagej**.nih.gov/ij/ (ImageJ is a free software sponsored by the National Institute of Health). In imageJ load the tif using File>Import>Image Sequence Here, you see Layer II/III neurons in mouse V2 cortex. When a horizontal bar appears at the top of the image, a light is being shined into the animal's contralateral eye. You can play the image by clicking on the arrow at the bottom left of the window or using '/'. You can also use '<' or '>' to advance frames manually. In the top left corner, note how many frames the TIFF stack has. The movie was acquired at 4Hz.

1) We need to load the XYZ (XY-T) images into Matlab. There are a number a ways you can do this depending on your MATLAB version. A sure fire way is to use a FOR loop for each Z {eg. (:,:,Z) }

```
help for
help imread
```

2) Find the height, width, and # of frames of the image stack

```
help size
help stksize
```

3) In order to process the image, we need to convert the XYZ stack into vectors of intensity vs time for each pixel. Convert the 3D matrix into a 2D matrix.

```
help reshape
```

Note: you may need to convert the uint to double.

4) Now let's decompose the data using singlar value decomposition

```
help svd
```

5) Now plot the singular value into a XY plot (X should be the value and Y should be the coefficient number (plot all))

```
help diag
help plot
help semilogy
```

How many coefficients should there be and why?

6) The singular values decrease with increasing coefficient numbers. Large singular values represent "distinct intensity signal" while low singular values represent "noise". Find the coefficient # cutoff for when the singular value contains the least amount of "unique signal" and when it contains "mostly noise" (when the singular values becomes parallel with the X axis)

Save/print the best plot from #5 to turn in. Indicate coefficient number you will use as the cutoff somewhere on the figure.

7) Now we want to reconstruct the vectors using only the first however many coefficients you indicated in step #6, using the formula indicated in the SVD function.

```
help SVD
    [U,S,V] = svd(X) produces a diagonal matrix S, of the same
    dimension as X and with nonnegative diagonal elements in
    decreasing order, and unitary matrices U and V so that
```

X = U*S*V'.

Note: for the U, S, and V matrices, note that the singular values has a maximum coefficient equal to the number of frames. In another words, you want to multiply all of your pixel dimension for each matrix, but only the row/column up to the number of coefficients you deemed necessary in step #6

8) Lastly, we need to convert the vectors back into a 3D movie.

reshape(reconstruction,height,width,frames);

9) Try playing the original and SVD processed movie side by side in matlab

```
figure(1)
for k=1:1000
subplot(1,2,1)
imagesc('original movie variable'(:,:,k))
subplot(1,2,2)
imagesc('SVD movie variable' (:,:,k))
pause(0.25)
end
```

Optional: you can also reconstruct the movie of the "noise" (:,n:1000) to make sure you chose a good coefficient cutoff

Turn in the labeled plot with the .m

## Figures Checklist

1. What are some example labels that should be used with cell count type analysis?
2. 1-6 implant location (X, Y), # of bins, bin size.
3. How many coefficients should there be and why?
4. 2-6 Coeff #

Optional

1. Cell Count vs Distance plot (probe vs control)
2. 4-9 movie