K-Means Spectral Umixing on CIRC

This is a user guide for running the k-means spectral unmixing algorithm on CIRC. It will explain what code you need to download, where you need to put it and how to run it in order to perform spectral unmixing using the k-means algorithm.

Assembling the necessary files

You can find the code for running k-means spectral unmixing on box here: https://rochester.app.box.com/folder/60387154437

and on GitHub here: https://github.com/tristan-mcrae-rochester/Multiphoton-Image-Analysis/tree/master/Spectral%20Unmixing/Code/k_means_unmixing_circ

You will need to download the entire folder labeled k_means_unmixing_circ. The main Matlab file you will be running is KMeansUnmixing.m and the main sbatch file you will use to run Matlab on CIRC will be spectral_unmixing.sbatch. You will need to move these files to your CIRC workspace to run them using CIRC. For help with this, you can visit https://info.circ.rochester.edu/Getting_Started/Connecting_using_Windows/Transferring_Files.html

For simplicity, it can be easiest to put the file(s) you want to unmix in the same folder as the code is in. This way you will not have to deal with telling Matlab in what path to look for your image.

Preparing the code

When running k-means spectral unmixing on CIRC, you will need to modify the file spectral_unmixing.sbatch to have it fit the specific file you want to unmix. The sbatch file will look like this:

```
#!/bin/sh
#SBATCH -p standard
#SBATCH --mem=50GB
#SBATCH -t 5-0:00:00
#SBATCH -J spectral_unmixing_job
#SBATCH -o spectral_unmixing.out

module load matlab

matlab -nodesktop -nodisplay -nosplash -singleCompThread -r "
num_channels = 3; num_timesteps = 1; num_fluorophores_to_unmix = 3;
channels_to_unmix = [1, 2, 3]; load_file = '112018 D7 A Ca lung NT c
d31.oif'; save_method = 'single_tiff'; save_file = 'testing_sbatch_l
arge'; KMeansUnmixing(num_channels, num_timesteps, num_fluorophores_
to_unmix, channels_to_unmix, load_file, save_method, save_file); exi
t"
```

The first 6 lines in blue that start with a '#' are all instructions to CIRC that deal with the computational resources it is requesting and where the command line output saves. You shouldn't have to change any of these.

The lines

Module load matlab

and

matlab -nodesktop -nodisplay -nosplash -singleCompThread -r "

set up CIRC to run Matlab and shouldn't need to be modified at all.

The red lines that follow are what you will have to edit when you are running spectral unmixing. They are a series of Matlab commands that will be executed when the sbatch file is run. Each command is separated by a semicolon. The first several commands are setting variables and are what you will have to modify when you run this code. The variables you will have to edit are as follows:

num_channels – The number of spectral channels in the image you are reading in. This parameter has to be set to exactly the number of channels that the data you load in has.

num_timesteps – The number of timesteps in the image you are reading in. This parameter has to be set to exactly the number of timesteps that the data you load in has.

num_fluorophores_to_unmix – The number of fluorophores you want to unmix. The program will automatically create one more cluster than the number of fluorophores you ask it to unmix (attempting to create a background cluster). This number will be one fewer than the total number of clusters that are found through k-means. You have some freedom in what this parameter is set at that can give you different results. When and why you would do that is described in the next section.

channels_to_unmix – The indices of the channel(s) you want to be unmixed. Do not necessarily need to have the same number of channels here as num_fluorophores_to_unmix. These should always be enclosed in brackets and if there is more than one channel, the channels should be separated by commas. E.g. [1] or [1, 2, 3].

load_file – The full relative path of the file you want to load. Should be enclosed in quotations ("") and include the file extension (e.g. .tif). For an oif file, have the name of the oif file here rather than the name of the folder that contains the individual slices. For ease of use, the file you want to load (and the folder containing individual slices in the case of oif files) should be in the same directory as your code and sbatch file. The unmixing code can read in files of type .tif, .oir and .oif.

save_method – One of either "single_tiff" or "multiple_tiffs". Choosing "single_tiff" will save all slices to as single tiff file. Choosing "multiple_tiffs" will save each 2D slice as its own tif file in an output folder whose name is determined by the save_file parameter.

save_file – Name of the file where you want to save your outputs. If your save method is "single_tiff", the output will be saved to <save_file>.tif. If your save method is "multiple_tiffs" a folder with the name <save_file> will be created and individual slices will be saved inside of it.

Strategy for selecting parameters

A good starting point is to select every channel for unmixing and ask it to unmix however many fluorophores were present during imaging.

These settings will not always provide the best unmixing results and there are reasons where you would want to deviate from that.

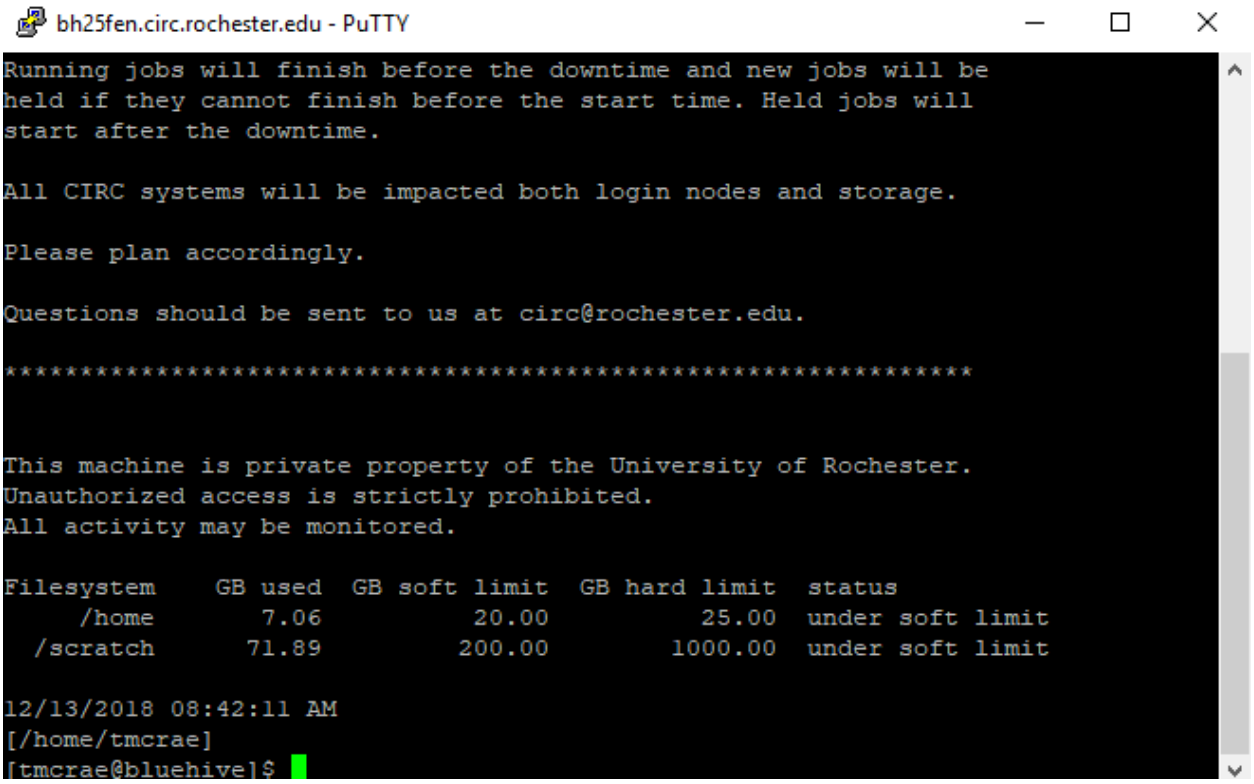
If there are some channels that you know already only include one fluorophore, you can uncheck their box and reduce the number of fluorophores you want to unmix by one.

If, after unmixing, it seems like there are still multiple fluorophores contained in some of the output channels, you can increase the number of fluorophores you ask it to unmix. This will have the effect of adding more clusters to k-means. If done successfully, this will result in two or more clusters that represent the same fluorophore but no clusters that represent multiple fluorophores.

Running k-means spectral unmixing on CIRC

Once you have modified spectral_unmixing.sbatch to your liking and both the image you want to analyze and the unmixing code are in the same folder on CIRC, you can run the spectral_unmixing.sbatch file. To do this, perform the below steps. For more information about using CIRC in this way, you can visit https://info.circ.rochester.edu/BlueHive/Running_Jobs.html

Open a command line interface with CIRC and log in (perhaps using PuTTY)



Navigate to the folder where you saved your code and image to be unmixed.

```
12/13/2018 08:42:11 AM
[/home/tmcrae]
[tmcrae@bluehive]$ cd /scratch/tmcrae/channel_separation/

12/13/2018 08:44:15 AM
[/scratch/tmcrae/channel_separation]
[tmcrae@bluehive]$
```

Start the unmixing code with the command

sbatch spectral_unmixing.sbatch

```
12/13/2018 08:47:18 AM
[/scratch/tmcrae/channel_separation]
[tmcrae@bluehive]$ sbatch spectral_unmixing.sbatch
Submitted batch job 6907659
```

You can check that it is running with the command

squeue –u <username>

```
[tmcrae@bluehive]$ squeue -u tmcrae
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           6907659  standard spectral   tmcrae  R       0:01      1 bhc0065
```

You can also cancel a job with the command *scancel <jobid>*

You can view the printouts of the program in real time to check on progress with the command

tail –f spectral_unmixing.out

```
[tmcrae@bluehive]$ tail -f spectral_unmixing.out

                          < M A T L A B (R) >
                Copyright 1984-2018 The MathWorks, Inc.
                 R2018b (9.5.0.944444) 64-bit (glnxa64)
                            August 28, 2018


To get started, type doc.
For product information, visit www.mathworks.com.

Unknown IlluminationType value 'null' will be stored as "Other"
Unknown IlluminationType value 'null' will be stored as "Other"
Unknown IlluminationType value 'null' will be stored as "Other"
Unknown LaserMedium value 'Alexa Fluor 647' will be stored as "Other"
Unknown LaserMedium value 'EGFP' will be stored as "Other"
Unknown LaserMedium value 'SHG' will be stored as "Other"
Reading series #1
    ...
Scaling pixel intensities
Running K-Means. This may take a few minutes for large datasets.
  iter   phase       num                sum
     1       1     65536            240.334
     2       1      6623            198.307
     3       1      3039            176.902
     4       1      1682            169.467
     5       1       873            167.494
     6       1       460            166.741
     7       1       243            165.835
     8       1       677            161.637
     9       1       789            155.664
    10       1       702            152.343
```

And you can exit from this real-time view with ctrl+c

Once the unmixing program finishes running, your unmixed results will be saved in either a single tiff or a folder of one tiff for each 2D slice depending on what you specified for save_method. This file/folder will be in the same directory you ran the code from.

Additional Questions

For any other questions, concerns or requests, contact Tristan McRae at Tristan_mcrae@urmc.rochester.edu.