Running the Phenocam Matlab code

by Michael Toomey, mtoomey@fas.harvard.edu

December 11, 2012

Scripts

Below is a list of the necessary code which should be placed in the Matlab search path (>> addpath *dir*). The files in bold are the ones that will be modified and executed directly.

ROICreation.m
date2jd.m
isleapyear.m
myquantile.m
phenotimeseries.m
turbopheno.m
turbophenomaster.m

Step 1: Make Region of Interest (ROI) Guide

In an ideal site, the field of view would be identical for every photograph and one could simply make one ROI that can be applied to the entire time series. But due to camera maintenance, changes in the site, and simple unintended movement, the field of view can change multiple times throughout an image archive. Hence, we use multiple ROIs for some sites, necessitating an orderly system for handling multiple inputs and ensuring that the proper ROI is used for each photo.

In the site folder, ***harvard***, you will see that there are three folders: one folders with *2009* data, one for outputs (*outputs*), and one for the ROIs (*ROI*). The ROI folder contains the image mask files in TIFF format (*.tif) and the ROI guide, a comma-separated-file (*csv). There can be multiple ROIs in this folder, but there should be only one ROI guide for each mask type (e.g. deciduous trees, full canopy). The *harvard* folder currently has one ROI guide that encompasses the entire canopy (*harvard_canopy_0001_roi.csv*). It is best to describe the type of ROI (i.e. *deciduous or canopy*) that is being used as it will be relevant in the processing step. Open the CSV in a text editor and you will see the following:

```
#
# ROI file for Harvard EMS site
#
# site: harvard
# description: "entire canopy"
#
start_date,start_time,end_date,end_time,maskfile,sample_image
2009-01-01,00:00:00,9999-01-01,00:00:00,harvard_canopy_0001_01.tif,harvard_2009_05_03_113138.jpg
```

The comments, marked by #, are primarily for reference. The main two lines of interest are the second to last and last line, which describe the input fields and reference the harvard deciduous ROI, respectively. As the first line indicates, the inputs are starting date and time, ending date and time, the mask filename, and the sample image. The mask file name should be a TIFF file and have the *tif extension; it will be saved in the *ROI* folder. Lastly, the sample image will be helpful for making the ROI in the next step. There is no single best image – but some tips are to use spring or fall images for mixed

canopies, where deciduous trees will stand out from conifer trees. Of course, also make sure the sample has adequate lighting for viewing.

Luckily, harvard is a very steady site, so there is only need for one ROI. Thus, the beginning time is the first day of imagery, 2009-01-01 and since the whole time series is covered by this ROI, we put in the arbitrary ending date of 9999-01-01. But let's say that we saw a shift in the field of view halfway through 2009 – specifically after the photo, harvard_2009_06_01_050138.jpg. Then, we would want to use the time stamp from the photo, 2009-06-01 05:01:38, as the ending time for the first ROI and make a second ROI, changing the ROI guide as follows:

start_date,start_time,end_date,end_time,maskfile,sample_image
2009-01-01,00:00:00, 2009-06-01 05:01:38,harvard_canopy_0001_01.tif,harvard_2009_05_03_113138.jpg
2009-06-01,05:31:38, 9999-01-01,00:00:00,harvard_canopy_0001_02.tif,harvard_2009_08_01_090138.jpg

Above, you can see that we used the time stamp from the following photo as the starting time for the second ROI, harvard_canopy_0001_02.tif, and also inserted a new sample image that covers that time range.

Note: Be sure to follow the file naming conventions for proper operation of the code. These were developed for optimal use of the Phenocam archive and undesirable results could occur if different file naming conventions are used.
For the ROI guide: *sitename_ROItype_XXXX.csv*
For the ROI image: *sitename_ROItype_XXXX_yy.tif*
where *XXXX* is the ordinal number of the set of ROIs for that cover type (e.g. 0001, 0002) and yy is the ordinal number of the ROI TIFF file for the ROI guide – this is relevant only if there are camera shifts, as in the above example.

Step 2: Make ROIs
To make ROIs we will use the Matlab script, **ROICreation.m**, and the ROI guide. Check the help section of the function to see the inputs and syntax.

If we are to create an ROI for the Harvard site, run the following line:

ROICreation('C:\harvard\2009\05','harvard_2009_05_03_113138.jpg','harvard_canopy_0001_01.tif',1)

The first input argument is the directory for the site, year and month containing the sample image, which is the second argument – note that the latter has simply been copied and pasted from the ROI guide. Similarly, the mask name, bartlettdecidmask.tif, has also been copied and pasted from the ROI guide. The last argument represents the type of ROI to be made, of which there are three options: I chose option #1, which is the free hand option.

When you run this line, it will open a window with the sample image; use the cursor and left mouse buttons to place nodes for the free hand polygon and then right click to finish. For the rectangle and ellipse ROIs, simply hold down the left button to drag out the polygon, which can then be moved until right clicking. The ROI will be saved in the same folder as the photograph, although you may choose to use relative pathnames to store elsewhere (e.g. '..\..\harvard_canopy_0001_01.tif').

Step 3: Run image processing

The script *turbophenomaster.m* is the wrapper script that will execute the relevant scripts and run the processing. Note that for the code to run the data must be in the proper file structure. That is, for each site, there must be separate year folders, each with 12 monthly sub-folders, each of which contain the images. In addition, the site folder must also contain an *ROI* folder (with the CSV and ROIs) and an *outputs* folder. It is designed to handle all years of data for all sites, and so there are some outputs that won't be as important for single sites. You will need to modify the User Defined Variables section, lines 31-48, as copied below:

```
%%%%%%%%%%%%%%%%%%%%%%%% USER DEFINED VARIABLES %%%%%%%%%%%%%%%%%%%%%%%%
% base directory, the one that contains all site directories
basedir = 'C:\PhenoCam\';
% filter for maskguide files, which should have the endings, "*roi.csv"
maskguidefilter = '*deciduous*roi.csv';
% sites to process, in a cell array
sites = {'harvard'};
% size of window for smoothing - default is 3
windowsize = 3;
% start and ending time (military time/24h), default is 6 and 18
sttime  = 6;
endtime = 18;
% minimum brightness threshold, in percentage, default is 15
threshold = 10;
% smoothing technique, where 1 = 90th percentile, 2 = mean, 3 = median,
% default is 1
smoothtechnique = 1;
%%%%%%%%%%%%%%%%%%%%%%%% END OF USER DEFINED VARIABLES %%%%%%%%%%%%%%%%%%%%%%%%
```

- *basedir* indicates the directory that contains all site folders.
- *maskguidefilter* is the search filter for the type of ROIs that we want to process
- *sites* is a list of sites to run
- *windowsize* is the size of the window, in days, for data smoothing – 3 is advised
- *sttime* and *endtime* are the earliest and latest hours of the day that will be used for processing images – hours should be in local time, unless your site's file naming convention is different
- *threshold* is the minimum picture brightness, in percent, needed to process a photo; this is done to exclude dark and nighttime photos that might not be picked up by hour range filters. The advised value is 10-15% brightness.
- *smoothtechnique* is the smoothing method – 90th percentile is advised

After specifying these variables you can now execute *turbophenomaster* at the command line. Users with a large number of sites and who have the Parallel Computing Toolbox may want to also modify lines 53-55 and 71 to allow parallel processing. The code will also report some status messages in the Command Window.

The script will output three CSV files for each year to the *outputs* folder. For example,
raw-harvard_deciduous_0001_2009_names.txt
raw-harvard_deciduous_0001_2009.txt
smooth90th-harvard_deciduous_0001_2009.txt

The first CSV includes the following fields:
Year, decimal day of the year, ROI red brightness mean, ROI green brightness mean, ROI blue brightness mean, ROI green chromatic coordinate (Green/(Red+Green+Blue)), file name of the photo

The second CSV is identical but omits the filenames.

The third CSV includes the following fields:
Year, day of the year, ROI green chromatic coordinate