

# SuMAc 2013 Project

The goal for this project is to use what you learned this week about image processing and analysis to solve a classification problem and then make a 5 to 10 minute long presentation on Friday at the banquet.

Each group will be given a data set consisting of 64 color images and a vector containing labels that describe what group each image belongs to. Many of the images are noisy, so you may want to apply a noise reduction technique to them. Alternatively, you may choose not to do so and work directly with the noisy images. While the images are in color, you can decide whether you want to convert them to grayscale or even edge maps. You can compute distances between images using either Euclidean or Manhattan distances and can decide whether you want to work with the images themselves or their image histograms.

**All of these choices are completely up to you!**

While we would like for you to try performing classification, you can focus on your favorite aspects of what you learned this week. If noise reduction and edge detection were your favorite concepts, you can devote most of your time to performing those techniques on the data, including applying multiple filters to reduce noise. If you really enjoyed the ideas of distance and classification, you can see what types of distances and classification methods work best for this data.

The most important part of the project is to experiment with the data and try your ideas out. If your ideas don't work well, that is perfectly fine. Often times, the best way to find what works is to first find what **doesn't** work! Or, as Jules Verne said it in *A Journey to the Center of the Earth*, "Science, my lad, has been built on many errors; but they are errors which it was good to fall into, for they led to the truth."

## Data

The data set consists of 64 color images stored in a vector of cells and has been provided to you as what is called a .mat file. This file can be read into Matlab using one of the following commands at the prompt, depending on your data:

```
load dataset1.mat    load dataset2.mat    load dataset3.mat
```

Make sure that your Matlab directory is set to the same one that contains the data file. Otherwise, Matlab won't know what to read in. Since this project will take you many days, you will need to save your work throughout the week. While you can certainly choose your own name for the saved work, I recommend using something like the following:

```
save project_day1.mat    save project_day2.mat    etc.
```

I suggest this, so that you can always return to a previous day's work in case something goes wrong with your data while you work. You can then read in your new files in the same way as your original data set.

The images are broken up into 8 groups, each containing 8 images. To simplify the programming, the images are grouped within the cell vector according to group. The groups are as follows:

- Group 1: A park, images 1 through 8
- Group 2: The sky, images 9 through 16
- Group 3: A tile floor, images 17 through 24
- Group 4: An office wall, images 25 through 32
- Group 5: Stairs, images 33 through 40
- Group 6: A brick sidewalk, images 41 through 48
- Group 7: A brick wall, images 49 through 56
- Group 8: A tiled sidewalk, images 57 through 64

## **Presentations**

Each group will make a short presentation between 5 and 10 minutes in length at the banquet about the project and what they learned during the camp. For the presentation, groups may present their work in a number of ways. You can create a Powerpoint presentation, use a Word document, or demonstrate what you did using Matlab, itself. Just like with the rest of the project, the choice is yours!