University of the Philippines, Diliman
College of Engineering
Department of Computer Science

CS 180 Machine Problem 2
Segmentation using K-means Clustering

Introduction

Clustering is the process of grouping objects in such a way that objects in a group are more similar to one another than those from other groups. One application of clustering is in the automatic segmentation of regions of interest in images. For this machine problem, you will apply kmeans clustering to isolate parasites in blood smear images.
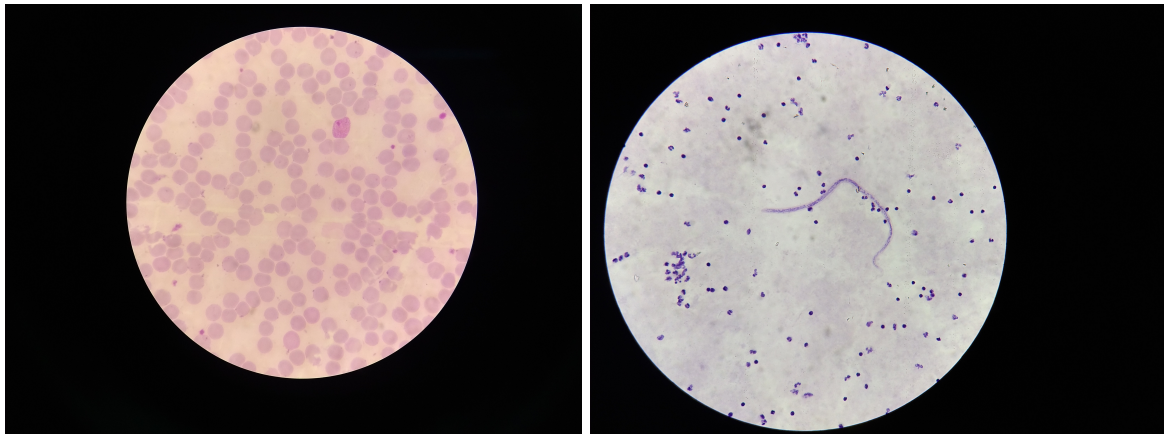
Dataset

The dataset is categorized based on the specimen:

- *Plasmodium*
- *Schistosoma*
- *Filarioidea*

Each category is composed of 10 images.

Below are a sample images from the dataset.

Program

Apply kmeans clustering to segment the parasites in the images. For this machine problem, you may use built-in libraries implementing the kmeans algorithm. If you are using Python, you can find the *KMeans* class in the *scikit-learn* module.

Analysis

For each of the specimen, train the algorithm using kmeans and 2 images. Discuss the results of the segmentation (on all images).

- Running kmeans multiple times, each time with 2 random centroids, what would the segmentation look like?

- Manually selecting specific pixel locations, one for the parasite and one for the others, as centroids, how would the segmentation change?

- If the centroids are increased to more than 2, how would the segmentation change? Are there specific regions other than the parasite segmented?

- Try the segmentation using different color spaces (HSV,CIELa*b*). How would the segmentation change? Which colorspace produced the best segmentation?

- Which of the specimens is the easiest to segment? the hardest?

Deliverables

Submit the following via email to `kedelaspenas@up.edu.ph` with subject *CS180 MP2 <SECTION> <Surname>*:

- Source Code (Filename: *CS180MP2_<SECTION>_<Surname>.<ext>*)
- Analysis (Filename: *CS180MP2_<SECTION>_<Surname>.pdf*)

Deadline

The deadline of this machine problem is on 5PM 14 March 2018.

Notes

Some of the images contain a black background enclosing the specimen region due to the circular nature of the microscope view lenses. You might need to add another centroid to capture this group of pixels.