

MA3111: Mathematical Image Processing

Homework 4

To submit, zip the followings into `mipXXXXXX.zip` where `XXXXXX` is your student ID:

- **report file:** contains answers to all questions and the requested figures.
- `my_kmeans.m`

In this homework, you will implement k-means clustering algorithm on color image `iris-RGB.tif`. **Do not call or copy the code of MATLAB's built-in function `kmeans` or anything similar.**

Given $k \in \mathbb{N}$, vectors $Z = \{z_1, \dots, z_Q\}$, and cluster centers $M = \{m_1, \dots, m_k\}$, the **intra-cluster variance** is defined as

$$E_k(Z, M) = \sum_{i=1}^k \sum_{z \in C_i} \|z - m_i\|^2,$$

where $z_q \in C_i$ if and only if z_q is closer to m_i than any other cluster centers.

For this problem, $z = (r, g, b, \gamma x, \gamma y)^\top$, i.e., 5-dimensional. Your `my_kmeans.m` should be general enough to handle data of any dimensionality.

1. Finish `my_kmeans.m` by following its comments. Feel free to ignore them.
2. Run `test.m`. Set `num_trial=2`. Try $(k, \gamma) \in \{3, 5\} \times \{0, 0.2, 0.5\}$. **There are a total of 6 combinations to run.** For each combination, record the centers, intra-cluster variance, and show the final segmented images in your report.

Questions:

1. which k gives lower intra-cluster variance?
2. what happens as γ goes up, and why?