

Q1-4. Image Color Segmentation using K-Means Clustering

```
import numpy as np
import os
import urllib

import matplotlib as mpl
import matplotlib.pyplot as plt
from matplotlib.image import imread

import sklearn
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans

#####
### Image Color Segmentation using K-Means Clustering ###
#####

### Prepare an image #####
images_path = os.path.join('.', "dataset")
os.makedirs(images_path, exist_ok=True)
DOWNLOAD_ROOT = "https://raw.githubusercontent.com/rickiepark/handson-ml2/master/"
filename = "ladybug.png"
print("Downloading", filename)
url = DOWNLOAD_ROOT + "images/unsupervised_learning/" + filename
urllib.request.urlretrieve(url, os.path.join(images_path, filename)) # Download an input image

image = imread(os.path.join(images_path, filename)) # Load an input image
print(image.shape)

X = image.reshape(-1, 3) # Reshape 3D image data structure into 2D data list (list of X Y Z coordinates)
print(X.shape)

### K-Means Image Color Segmentation #####

segmented_imgs = [] # List for saving segmented images

# Conduct K-Means clustering with a different number of clusters
n_colors = (10, 8, 6, 4, 2)
for n_clusters in n_colors:
    kmeans = KMeans(n_clusters=n_clusters, random_state=42).fit(X) # Cluster the data

    segmented_img = kmeans.cluster_centers_[kmeans.labels_] # Each point in the image will have its label that corresponds to its cluster.
    # Map each point in the image to the coordinates of its cluster/label
    # => Segment the image by mapping each image point to the cluster

    segmented_imgs.append(segmented_img.reshape(image.shape)) # Reshape back to original 3D image structure for display

### Show the results of image segmentation using K-Means with different number of clusters #####
plt.figure(figsize=(10,5))
plt.subplots_adjust(wspace=0.05, hspace=0.1)

plt.subplot(231)
plt.imshow(image)
plt.title("Original image")
plt.axis('off')

for idx, n_clusters in enumerate(n_colors):
    plt.subplot(232 + idx)
    plt.imshow(segmented_imgs[idx])
    plt.title("{} colors".format(n_clusters))
    plt.axis('off')

plt.show()
```



- 이미지에서 각 Pixel 의 RGB 값을 Feature 로 사용하여 K-Means Clustering 을 수행함 .
- 이를 통해 이미지의 각 Pixel 의 RGB 값이 Cluster 를 중심으로 Labeling 됨 .
- 특정 범위의 RGB 값 범위에 Cluster 가 배치되기 때문에 , 각 Cluster 는 특정 종류의 색상을 나타낼 수 있음 . 이 때 , 특정 Cluster 를 많이 사용할수록 많은 색상들이 수용되기에 이미지의 색상이 풍부해짐 . 반면 Cluster 를 적게 사용할수록 적은 색상들이 수용되기에 이미지의 색상이 적게 표현됨 .
- 이를 통해 각 이미지 Pixel 의 RGB 값을 Clustering 하여 색상 범위를 나뉘낼 수 있음 . 그리고 사용하는 Cluster 중 일부만 사용하거나 Cluster 개수를 변경함으로써 색상 분리 (Color Segmentation) 을 구현할 수 있음 .