# Clustering Algorithms for Image Segmentation and Dimensionality Reduction Techniques

Almansur Kakimov

January 26, 2025

## Objective

The objective of this assignment is to explore clustering algorithms for image segmentation and to apply dimensionality reduction techniques to visualize and simplify high-dimensional image data. Specifically, we focused on applying Principal Component Analysis (PCA) and K-Means clustering, alongside other techniques like Superpixel segmentation (SLIC), to process and analyze image data.

## 1. Dataset

In this report, we utilized a sample image from the sklearn.datasets module, specifically the "china.jpg" image, which is a suitable example for image segmentation tasks. The image was resized for efficiency, and the necessary preprocessing steps were applied.

## 2. Preprocessing

The image data was preprocessed as follows:

The image was resized by halving its dimensions to reduce computational complexity.
It was then converted to a floating-point format and normalized using img_as_float from the skimage library.
The image was flattened into a 2D array of pixels to facilitate feature extraction.

## 3. Feature Extraction

Each pixel of the image was treated as a feature vector with 3 values (RGB channels). This resulted in a 2D matrix where each row represents a pixel and each column represents a feature (R, G, and B values). The features were then normalized using StandardScaler to standardize them and ensure that the clustering algorithm works efficiently.

## 4. Dimensionality Reduction

We applied Principal Component Analysis (PCA) to reduce the dimensionality of the image data from the original 3D space (RGB channels) into a 2D feature space. This helped in visualizing the data more effectively while preserving the most important features of the original data. The PCA result was plotted using matplotlib, where the points were colored according to their original RGB values to demonstrate how the data points are distributed in the reduced feature space.

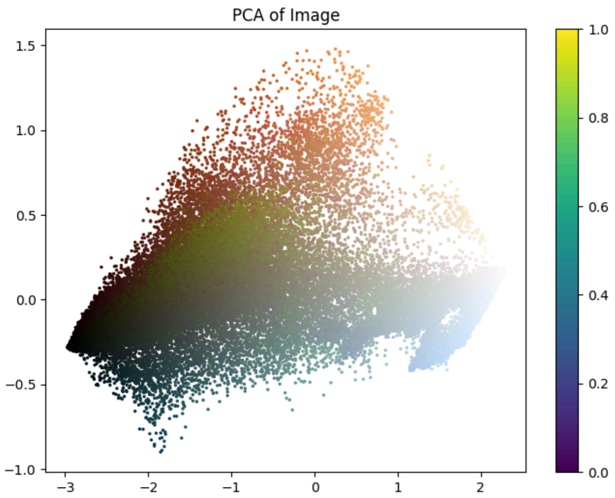


Figure 1: PCA of Image

### 5. Clustering Algorithms for Image Segmentation

#### K-Means Clustering

We implemented the K-Means clustering algorithm for image segmentation, setting the number of clusters to 5. The clustering labels were then reshaped back into the original image's dimensions, allowing us to visualize the segmented image. The result displayed the image divided into distinct clusters, each representing a different segment of the image with unique colors.

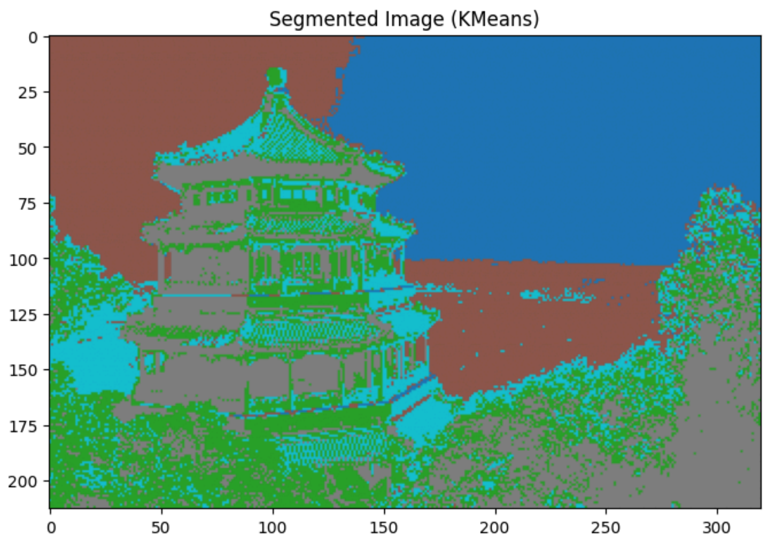


Figure 2: Segmented Image

#### SLIC Superpixel Segmentation

In addition to K-Means, we also employed SLIC (Simple Linear Iterative Clustering) for superpixel segmentation. SLIC segments the image into perceptually meaningful regions, and we used it with 200 segments and a compactness value of 10. The segmentation output was visualized to compare the quality of segmentation against the K-Means result.

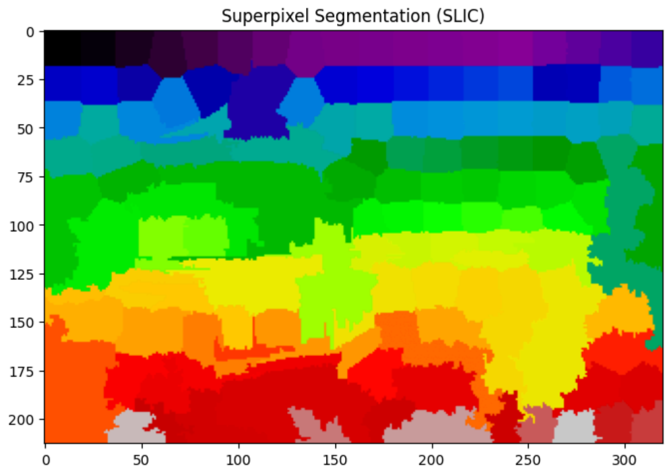


Figure 3: Superpixel Segmentation

**6. Results**

The following results were observed:

PCA Visualization: The reduced 2D feature space effectively captured the image's structure. The scatter plot showed how the color features of the image were distributed, with the RGB channels being distinguishable in the PCA projection.

K-Means Segmentation: The K-Means algorithm performed adequately, dividing the image into 5 clusters. The segmentation result displayed distinct regions with similar pixel intensities, although some boundary definitions were not as precise as desired.

SLIC Segmentation: SLIC segmentation was able to produce well-defined regions with clear boundaries, creating a superpixel segmentation that is visually more meaningful compared to the pixel-based K-Means clustering.

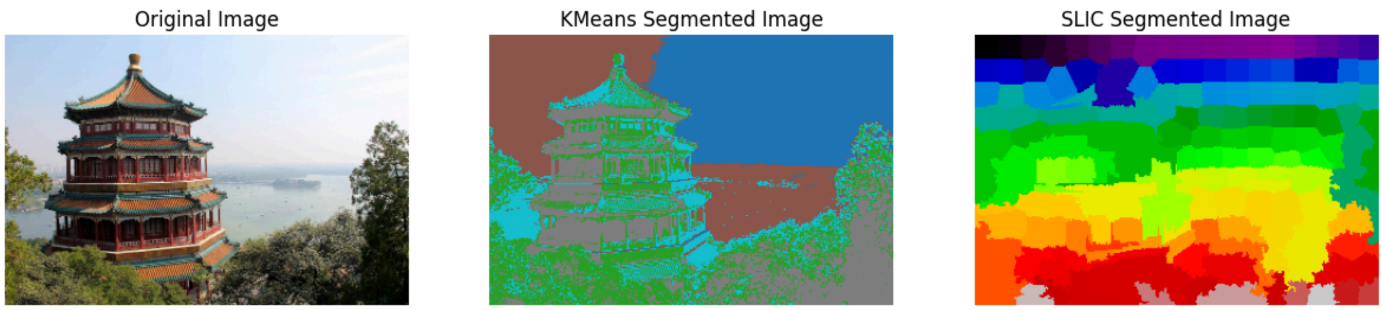


Figure 4: Comparison of Images

**7. Evaluation**

We compared the results from K-Means clustering and SLIC segmentation:

Clustering Performance: K-Means worked well in grouping pixels based on their color similarity but showed limitations in boundary sharpness. A higher number of clusters could improve the segmentation.

SLIC Superpixel Segmentation: SLIC outperformed K-Means in terms of producing meaningful segments with clear boundaries. It segmented the image into perceptually meaningful regions that aligned better with the natural structure of the image.

**8. Insights and Challenges**

Dimensionality Reduction: PCA was effective in reducing the feature space, but it does not always preserve the fine details necessary for segmentation, as it focuses on capturing variance rather than spatial coherence.

Clustering Algorithms: While K-Means is computationally efficient, it can struggle with images that have complex structures or noise. SLIC's superpixel approach provides a better solution for perceptual grouping of pixels.

Visualization: The dimensionality reduction (PCA) and segmentation results (both K-Means and SLIC) were effective in visualizing high-dimensional data in a way that is interpretable to humans.

**9. Conclusion**

In this assignment, we successfully implemented dimensionality reduction and clustering techniques to process and segment image data. PCA was useful for visualizing high-dimensional data, and K-Means clustering provided a basic approach for segmenting the image. However, SLIC segmentation proved to be more effective for creating meaningful image regions. Further improvements could include exploring different clustering algorithms and fine-tuning the parameters for optimal results.