**KMeans Image Color Quantization for Social Impact in kenya**

**Penina Wanyama**

**SDS6/44238/2023**

**MSc Data Science**

**School of Computing & Informatics**

[**Penina.wanyama@students.uonbi.ac.ke**](mailto:Penina.wanyama@students.uonbi.ac.ke)

**ABSTRACT**

In this paper, an RGB image is processed using a color quantization algorithm based on K-Means to reduce the number of colors in the image. The color quantization algorithm selects the most representative color and reduces the ineffective color in the image as much as possible. This paper assumes that the RGB image is composed of multiple pixels. Color quantization can be realized using the K-Means algorithm to perform unsupervised clustering on these pixels with specific colors. The use of K-Means for color quantization of the images can reduce the number of colors in those images so that they can be reproduced well in lower-performance computer equipment. At the same time, color quantization reduces the size of the images and improves the efficiency of image processing such as uploading over the internet.

**1**. **INTRODUCTION**

Kenya faces many challenges with low bandwidth, resource-constrained healthcare, and digital preservation of cultural heritage. These challenges stem partly from the large files produced from high-color images. True-color images typically contain thousands of colors, which makes their display, storage, transmission, and processing problematic. With the advent of artificial intelligence and big data, many advanced image-capturing and processing equipment have emerged in recent years. However, limited by the low computer performance and storage, and low bandwidth speeds in Kenya too many colors can be efficiently stored, processed, or transmitted over the internet. Therefore, it's necessary to use a color quantization algorithm to reduce the number of colors in an image while reproducing the original features as well as possible.

Color quantization is the process of reducing the amount of color in an image in the case where the visual error between the original image and the quantized image is low. Color quantization in Kenya can help reduce internet costs and optimize bandwidth through adaptive image compression leading to fast web browsing and access to educational materials online even for mobile data users in rural areas. Smaller-sized images resulting from color quantization will also help improve healthcare imaging and foster efficient utilization of storage space in healthcare facilities and hence allow for easy adoption of telemedicine. Color quantization can help digitize cultural artifacts with a minimal storage footprint. This will facilitate online access to Kenya’s heritage by the global audience fostering education, awareness, and appreciation of Kenya’s culture.

# 2. COLOR QUANTIZATION

With the development of computer technology and the improvement of hardware performance, the collection of high-quality images is no longer an issue. Image processing technology is also widely used in various fields such as medicine, agriculture, energy exploration, etc. The color digital image in the computer is generally obtained by mixing three basic colors of red (R), green (G), and blue (B) according to a certain ratio. The colors of these images are discrete and are represented by a set of binary values between 0 and 255. The maximum number of colors stored in the computer is 2(8+8+8) = 16777216.

However, not all computers can process these high-quality images. In low-performance computers, we need to use images that are small in color but that reflect the target features in the image. Currently, it is necessary to color quantize the image with many colors. Color quantization is the merging of less important colors in an image into a relatively important color and minimizes the visual error between the original image and the quantized image. In this process, the color value of the RGB image is changed from *R*, *G*, *B* to *R*', *G*', *B*', and the formula (1) represents the color error *E* after color quantization.

*E*= (*R*−*R*')2 + (*G*−*G*')2 + (*B*−*B*')2 (1)

There are two main types of color quantization algorithms, one is the clustering method and the other is the segmentation method. In this project, RGB images are treated as pixels with different colors in space, as shown in Figure 1, the clustering-based color quantization algorithm is used to convert the color quantization problem of RGB image into the clustering problem of pixels. The mean square error of the final clustering result is shown in formula (2) [5].

(2)

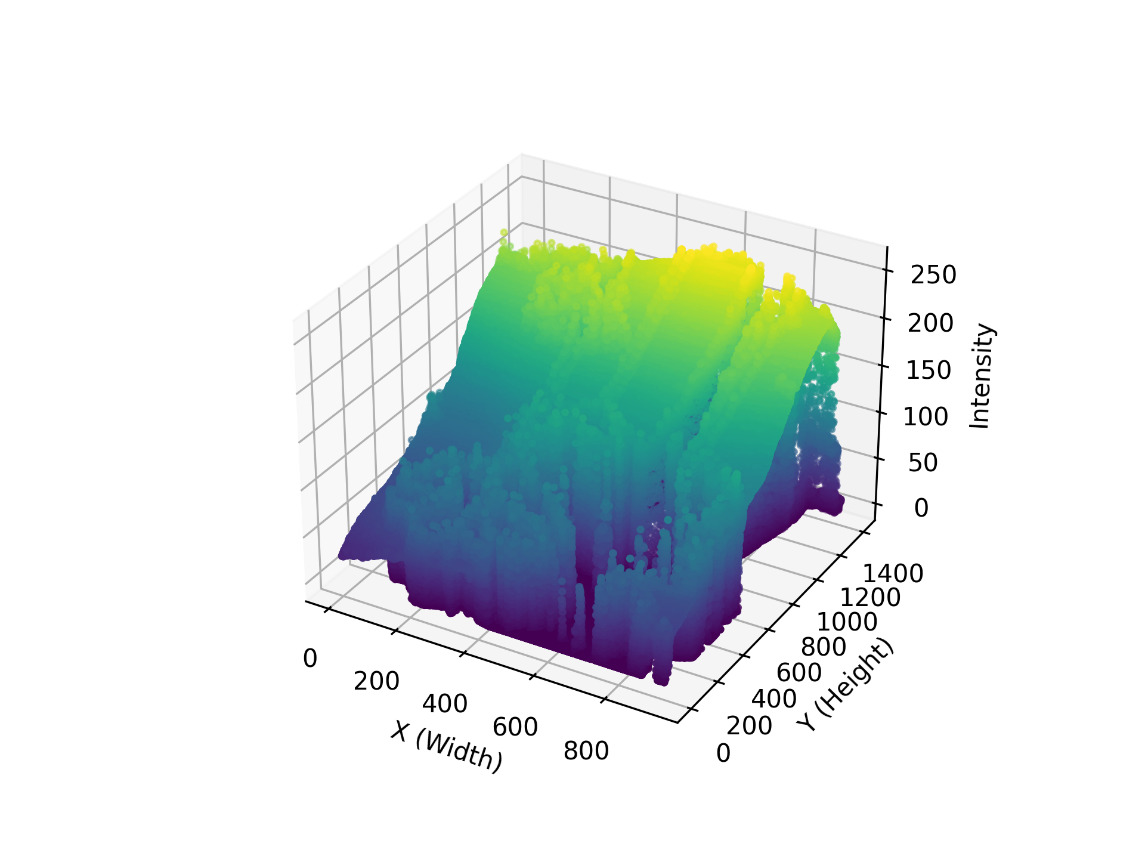


Figure 1 Three-dimensional representation of the RGB image of a coconut tree.

However, the difference in color quantization for human vision is not the same as the difference in its value. The requirements of color quantization in this paper are mainly to reduce visual error and retain the most representative color so that the quantized remote sensing image is visually as similar as possible to the original image.

# 3. K-Means

The K-Means clustering algorithm is an unsupervised clustering algorithm that automatically clusters based on the similarity of individual data points. It's simple and easy to implement, and it still has a good clustering effect and high processing efficiency for large data sets.

In the clustering process, the similarity between any two data points is measured by distance. Common distance measurement methods include Euclidean distance, Chebyshev distance, Manhattan distance, etc. The distance metric used in this paper is the Euclidean distance. The Euclidean distance is the linear distance between two points in the metric space. The calculation method is formula (3):

*d*=

= (3)

**K-Means steps**

1. Select K data points from a data set containing n data points as the initial cluster center C1, where the data points are the pixels of the image.
2. Calculate the distance *d* between each data point and the cluster centers separately. Suppose the data point coordinates are (x1, y1 , z1) and the cluster centre coordinates are (x2, y2 , z2). The calculation method is shown in formula (4), and the data point with the smallest distance from the cluster center is classified as the same category as the cluster center.

*d* = (4)

1. Calculate the mean according to the existing data points of each category, reselect the cluster centers of each category according to formula (5)

(5)

1. This is repeated until the cluster center Ci doesn’t change.

In an art shell, the purpose of K-Means clustering is to divide the raw data in to K classes S = {S1, S2, ….Sk} given the number of classification groups k (k≤n), on the numerical model, which is to find the minimum value of the formula (6):

Where µi represents the average o S*i*

This paper mainly analyses the clustering of pixels in RGB images. The cluster center K represents the number of colors. This parameter needs to be set by oneself. The value of K must be less than the number of colors of the original RGB image. In the two experiments in this paper, K is set to 64 and 32 respectively.

# 4. EXPERIMENT

*4.1 Experiment procedure*

The dataset used in the paper is a coconut tree, sourced from [unsplash.com](https://unsplash.com/). In the original RGB image, the coconut image has 1308534 colors.

The experiment environment for this article is 3.11.5, NumPy 1.24.3, Scikit-Learn 1.3.0, and matplotlib 3.7.2. NumPy is an open-source scientific computing package implemented by Python. It supports multidimensional arrays and matrix operations. This experiment introduces the NumPy package for three-dimensional array operations. Matplotlib is a Python-implemented graphing/drawing package, which can display the coconut tree image and related description of the output in the visual form. Scikit-Learn is a machine learning algorithm library based on NumPy, SciPy, and matplotlib. The K-Means algorithm used in this experiment can be directly called from the Scikit-Learn library.

The basic idea of the experiment is to cluster the pixels of the RGB image, group pixels with similar colors into one class, and finally reconstruct the original image with a larger number of colors into an image with less color. In this paper, two images with different targets are tested. The number of colors set for the experiment is 16,32,64,128 and 256 respectively. Throughout the experiment, it’s important to keep the visual error to a minimum to achieve the best results for this experiment.

The specific steps for the experiment are designed as follows:

1. Set the parameter *k*= [16,32,64,128,56] where *k* is the number of cluster centers. When *k*=16, the color quantization algorithm quantizes the number of colors of the image into 16, then 32 and so on until we get to k=256.

2. Load input data. The input data has one RGB coconut tree image with 1308534 colors.

3. Mark out the pixels in image. In this paper, the image is transformed into an image composed of pixels, which is convenient for clustering pixels with similar colors in the next step.

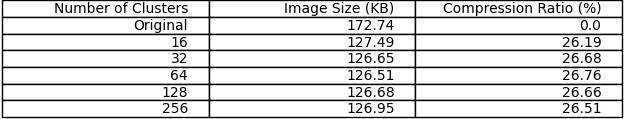
4. The K-Means algorithm is used to create a color palette, and the image is color quantized and reconstructed into images with [16,32,64,128,256] clusters.

5. Using the “pyplot” module of the matplotlib package to visualize the output, as shown in Figure 3.



6. Derive the quantized map and comparing the size of the original image with the quantized image, and Table 1 is obtained.

**Image Size and Compression Ratio Comparison**



***4.2 Results Analysis***

The results (Figure 3) of the 16,32,64,128 and 256 color clusters in the image colour quantization were obtained. We note that the size of the reproduced image is generally smaller than the original image after the compression process. An increase in the number of clusters increases the size of the reproduced image. The first line is the original image.

From the two experiments, it can be found that the quantized image of the second row contains more color features than the third row. Therefore, the color quantization algorithm based on K-Means has a small visual error in the color quantization of images, which can better reflect the color characteristics of remote sensing images.

As can be seen from the table (Table 1) and image visualizations above, a major advantage of color quantization is that the size of the RGB image is reduced, which greatly saves the storage space of the RGB image in the computer. When the computer processes the image, the RGB image with less memory can improve the processing speed and improve the research efficiency.

**5. CONCLUSION**

K-Means-based algorithms are applied to RGB coconut tree. In the experiment, the number of cluster centres was set to 16,32,64,128 and 256, respectively, and the experimental results were obtained. A comprehensive analysis of the results of experiment yielded the following conclusions:

1. The color quantization algorithm based on K-Means has a better effect on the color quantization of images.

2. Colour quantization of RGB images can reduce image size.

Therefore, the result of using the color quantization algorithm based on K-Means has a smaller visual error and consumes less memory space. Color quantization not only enables RGB images to be reproduced in more low-performance devices, improves image usability, but also saves computer storage space and improves image processing and transmission over the internet efficiency.

**REFERENCE**

1. Yang Jinfeng, Chen Qing, Han Xiaori et al. Application of Digital Image Processing Technology in Leaf Area Measurement of Vegetables [J]. Transactions of the Chinese Society of Agricultural Engineering, 2002, 18(4): 155-158 (in Chinese). DOI = 10.3321/j.issn:10026819.2002.04.038
2. Cheng Guojian, Ma Wei, Wei Xinshan, et al. Rock Fabric Recognition Based on Image Processing and Neural Network [J]. Journal of Xi'an Shiyou University: Natural Science Edition, 2013, 28(5): 105-110 (in Chinese).DOI = 10.3969/j.issn.1673-064X.2013.05.021
3. REN Zhibin, ZHAI Yongxin, YANG Yinghui, YANG Huaijiang.Color quantization of color

images in uniform color space[J].Optics and Precision Engineering,2002(04):340-345 (in Chinese). DOI = 10.3321/j.issn:1004-924X.2002.04.004

1. Ozturk C , Hancer E , Karaboga D . Color Image Quantization: A Short Review and an Application with Artificial Bee Colony Algorithm [M]. IOS Press, 2014. DOI = 10.15388/Informatica.2014.25
2. Long Y , Gong Y , Xiao Z , et al. Accurate Object Localization in Remote sensing images Based

on Convolutional Neural Networks[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 55(5):2486-2498. DOI = 10.1109/TGRS.2016.2645610

1. Xiao Z , Liu Q , Tang G , et al. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2):618-644. DOI = 10.1080/01431161.2014.999881