Color Quantization Application Based on K-Means in Remote Sensing Image Processing

Guojian Cheng¹a, Junjie Wei²b

¹School of Computer Science
Xi’an Shiyou University
Xi’an City in China
00-86-15029188968

²School of Computer Science
Xi’an Shiyou University
Xi’an City in China
00-86-18829030343

¹gjcheng@xsyu.edu.cn  b18829030343@163.com

ABSTRACT: In this paper, four kinds of RGB remote sensing images are processed using color quantization algorithm based on K-Means to reduce the number of colors in the image. The color quantization algorithm is to select the most representative color and reduce the useless color in the image as much as possible. This paper assumes that the RGB remote sensing image is composed of multiple pixels. Using the K-Means algorithm to perform unsupervised clustering on these pixels with specific colors, color quantization can be realized. The use of K-Means for color quantization of remote sensing images can reduce the number of colors in those images, so that remote sensing images can be reproduced well in lower performance computer equipment. At the same time, color quantization reduces the size of remote sensing images and improves the efficiency of remote sensing image processing.

CCS Concepts
• Theory of computation → Theory and algorithms for application domains → Machine learning theory → Unsupervised learning and clustering.

1. INTRODUCTION
Remote sensing image is a professional RGB color image that provides a lot of meaningful spatial data information. In recent years, with the development of artificial intelligence and big data, many advanced remote sensing image processing technologies have emerged. However, limited by the performance of the computer, remote sensing images with too many colors can’t be processed in a low-configuration computer. Therefore, it is necessary to use a color quantization algorithm to reduce the number of colors in the remote sensing image while reproducing the original image 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 of remote sensing images not only enables low-performance devices to perform remote sensing image research, but also makes remote sensing images more usable. It can also reduce the size of images, and improve the efficiency of remote sensing image processing.
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 [1], agriculture [2], energy exploration [3], 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 $2^{(8+8+8)} = 16777216$[4].

However, not all computers are able to process these high quality images. In low-performance computers, we need to use images that are small in number of colors but that reflect the target features in the image. At this time, it is necessary to color quantize the image with a large number of 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 clustering method and the other is segmentation method. In this paper, RGB remote sensing 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 remote sensing image into the clustering problem of pixels. The mean square error of the final clustering result is shown in formula (2) [5].

$$MSE_{RGB} = \frac{1}{N_p} \sum_{p=1}^{N_p} \sum_{i \in C_p} (R_p - R'_i)^2 + (G_p - G'_i)^2 + (B_p - B'_i)^2$$ (2)

![Figure 1 Three-dimensional representation of the RGB remote sensing image of the aircraft](image)

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

3. K-Means

The K-Means algorithm is an unsupervised clustering algorithm that automatically clusters based on the similarity of individual data points. The K-Means algorithm is simple and easy to implement, and it still has 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 shown in formula (3):

\[
d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}
\]  
(3)

The basic idea of K-Means is introduced as follows:

(1) Select k data points from the data set containing n data points as the initial cluster center \(C_1\), where the data points are the pixels in the image;

(2) Calculate the distance \(d\) between each data point and the cluster center separately. Suppose the data point coordinates are \((x_1, y_1, z_1)\) and the cluster center coordinates are \((x_2, y_2, z_2)\). 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;

\[
d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}
\]  
(4)

(3) Calculate the mean according to the existing data points of each category, and re-select the cluster center of each category according to formula (5);

\[
C_i = \frac{1}{n_j \in C_i} \sum x_j
\]  
(5)

(4) This is repeated until the cluster center \(C_i\) does not change.

In summary, the purpose of K-means clustering is to divide the raw data into k classes \(S = \{S_1, S_2, \ldots, S_k\}\) given the number of classification groups \(k (k \leq n)\), on the numerical model, which is to find the minimum value of the Formula (6):

\[
\arg \min_{\mu} \sum_{i=0}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2
\]  
(6)

Where \(\mu_i\) represents the average of \(S_i\).

This paper mainly analyzes the clustering of pixels in RGB remote sensing images. The cluster center \(k\) represents the number of colors. This parameter needs to be set by yourself. The value of \(k\) must be smaller than the number of colors of the original RGB remote sensing 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 this paper is four kinds of RGB remote sensing images, which was established by Wuhan University for remote sensing image target detection[6][7]. According to the different detection targets, the images are divided into four categories: aircraft, overpass, playground, and oil tank, as shown in Figure 2. In the original RGB remote sensing image, the aircraft image has 955,260 colors, the overpass image has 992,225 colors, the playground image has 414,447 colors, and the oil tank image has 992,225 colors.

![Figure 2. Four different kinds of remote sensing images.](image)

The experimental environment for this article is Python 3.6, NumPy 1.14, Scikit-Learn 0.19, and matplotlib 2.2. Numpy is an open source scientific computing package implemented by Python. It
supports a large number of dimensional arrays and matrix operations, and also provides a large number of mathematical function libraries for array operations. This experiment introduces the Numpy package for three-dimensional array operations. Matplotlib is a Python-implemented drawing package that generates simple visual data and images. This experiment introduces the matplotlib drawing package, which can display the remote sensing image and related description of the output in a visual form. Scikit-Learn, is a machine learning algorithm library based on NumPy, SciPy, and matplotlib. The K-Means algorithm used in the experiment can be directly called from the Scikit-Learn algorithm library.

The basic idea of the experiment is to cluster the pixels of the RGB remote sensing image, group the pixels with similar colors into one class, and finally reconstruct the remote sensing image with a larger number of colors into the remote sensing image with less color. In this paper, four RGB remote sensing images with different targets are tested. The number of colors set for each of the two experiments was 64 and 32, respectively. To illustrate the advantages of the method chosen in this paper, a Random-based color quantization algorithm was used in the experiment to compare with the K-Means-based color quantization algorithm. Throughout the process, it is necessary to keep the visual error to a minimum to achieve the best results of this experiment.

The specific steps of the experiment are designed as follows:

1. Set the parameter $k=64$, where $k$ is the number of cluster centers. When $k=64$, the color quantization algorithm quantizes the number of colors of the remote sensing image into 64.

2. Load input data. The input data is four kinds of RGB remote sensing image: aircraft (955260 colors), playground (414447 colors), overpass (992225 colors), and oil tank (992225 colors).

3. Mark out the pixels in the remote sensing image. In this paper, the RGB remote sensing 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 and the Random algorithm are respectively used to create a color palette, and the four kinds of remote sensing images are color quantized and reconstructed into a remote sensing image with 64 colors.

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

6. Set the parameter $k=32$ again. When $k=32$, the color quantization algorithm quantizes the number of colors in the remote sensing image into 32 colors.
(7) Repeat the above steps (2) to (5) to obtain the color quantization results of the K-Means and Random algorithms, as shown in Figure 4.

(8) Deriving the quantized map and comparing the size of the original image with the quantized image, and Table 1 is obtained.

![Figure 4. Color quantification results of four kinds of remote sensing images (32 colors).](image)

| Image method | Aircraft | Overpass | Playground | Oil tank |
|--------------|----------|----------|------------|----------|
| Original image | 109.0K | 155.0K | 100.0K | 115.0K |
| Quantized image (64) K-Means | 84.9K | 122.0K | 70.4K | 80.9K |
| Quantized image (32) K-Means | 64.4K | 99.8K | 55.1K | 63.3K |
| Quantized image (64) Random | 88.3K | 123.0K | 77.0K | 83.8K |
| Quantized image (32) Random | 65.2K | 101.0K | 57.4K | 68.4K |

4.2 Result Analysis

Figure 3 is the result of 64-color remote sensing image obtained by color quantization of the original RGB remote sensing image, and Figure 4 is the result of 32-color remote sensing image obtained by color quantization of the original RGB remote sensing image. Both figure 3 and figure 4 show the color quantization results of the remote sensing images of the aircraft, overpass, playground and oil tank from left to right. The first line is the original image. The second line is the quantized result obtained using the color quantization algorithm based on K-Means. The third line is the quantized result obtained using the Random-based color quantization algorithm.

Observing the results of the two experiments, it can be clearly 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 remote sensing images, which can better reflect the color characteristics of remote sensing images.

As can be seen from Table 1, a major advantage of color quantization is that the size of the RGB remote sensing image is reduced, which greatly saves the storage space of the RGB remote sensing image in the computer. When the computer processes the image of RGB remote sensing image, the RGB remote sensing image with less memory can improve the processing speed and improve the research efficiency.
5. CONCLUSION
In this paper, K-Means-based and Random-based color quantization algorithms are applied to RGB remote sensing images. In the two experiments, the number of cluster centers was set to 64 and 32, respectively, and two sets of experimental results were obtained. A comprehensive analysis of the results of two experiments yielded the following conclusions:

(1) The color quantization algorithm based on K-Means has better effect on color quantization of remote sensing images.

(2) Color quantization of RGB remote sensing 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 remote sensing images to be reproduced in more low-performance devices, improves image usability, but also saves computer storage space and improves image processing efficiency.

REFERENCES
[1] Zhou Yue, Shen Li, Yang Jie. Analysis of Tongue Image Characteristics of Traditional Chinese Medicine Based on Image Processing [J]. Infrared and Laser Engineering, 2002, 31(6): 490-494 (in Chinese). DOI = 10.3969/j.issn.1007-2276.2002.06.007
[2] 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:1002-6819.2002.04.038
[3] 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
[4] 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
[5] 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
[6] 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
[7] 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