RESEARCH ARTICLE   OPEN ACCESS

Image Color Quantization Algorithm Based on CFSFDP

R.L. Zhang[1]

[1] Department of Computer Science,Jinan University, Guangzhou, China

---------------------------------------************************--------------------------------

Abstract:

In the color quantization method, the clustering quantization method is widely used because it retains the accurate image subject information, and the color clustering method usually uses k-means as its clustering method, but the k-means algorithm can not adapt to the irregular clustering shape and sensitive to the initial clustering center. As consequence, the image color value in the RGB space clustering effect is not good. In this paper, a color quantization algorithm based on CFSFDP is proposed. The algorithm has better performance on clustering the irregularity of color value distribution. The experiment shows that error of the quantization image and the original image 99% smaller than k-means, when using our algorithm. And the algorithm can improve the image retrieval performance based on the color histogram. The CFSFDP quantization algorithm improves the retrieval recall by 4 ~ 6%.

Keywords — **CFSFDP, color quantization, cluster, image retrieve.**
---------------------------------------************************--------------------------------

I. INTRODUCTION

1) Color Quantization

Color quantification refers to the reduction of the number of image color types by a certain method, usually by implementing a color mapping table. Image of a true color image extract its three-channel histogram as an image feature with large number of bits. So the image feature needs to be compressed, so as to reduce the complexity of calculation.

2) CFSFDP

CFSFDP(Clustering by fast search and find of density peaks) [1] is a probability-based clustering algorithm.The idea of the algorithm is that the density of the points around the cluster center is high and at the same time from other high density points.The key point of the algorithm is calculatin the point density and the distance between points.With data set $S = \{x_1 \ldots x_N\}$, let $d_{ij} = dist(x_i, x_j)$ which means the distance of x_i and x_j, let $I_s = \{1..N\}$ be subscript of S. For each point, define its local density ρ_i as the number of points within a range.

$$\rho_i = \sum_{j \in I_s} X(d_{ij} - d_c) \qquad (1)$$

$X(x)$ is defined as

$$X(x) = \begin{cases} 1, x < 0 \\ 0, x \geq 0 \end{cases} \qquad (2)$$

Define δ_i as the shortest distance between a point to another point whose local density is higher:

$$\delta_i = \min_{j: \rho_j > \rho_i}(d_{ij}) \qquad (3)$$

For a point that has highest density, its δ_i is set as $\max(d_{ij})$. In the literature [1], decision graph is introduced to find the cluster center. In figure 1, 28 points are divided into two categories. It can be seen that the maximum density points are points 1 and 10, Is also the center of the cluster; figure 2 is called the decision graph, where the Y axis represents the distance of the point, and the X axis represents the local density of the point. It can be seen from figure 2 that the cluster center 10 and 1 are relatively special. Distance and local density of point 10 and 1 are relatively high, while the points 26,27,28 have a high distance, but their local density is low.We can see that these three points are outliers, so the high distance and low local density points should be ignored before clustering. The method of classification is to input the cluster

centers $M = \{m_1 \dots m_j\}, j \in \{1 \dots n_c\}$, which n_c is the number of cluster centers, initialize the classification mark $C = \{c_1 \dots c_N\}$

$$c_i = \begin{cases} c_i & \text{when } x_i \text{ is cluster center} \\ c_{n_i} & c_i = -1 \end{cases} \quad (4)$$

Then cluster the non-clustering center points. For point x_i, let n_i as the denote number of the point whose local density is higher then x_i while is closest to x_i. Then cluster the points.

$$c_i = \begin{cases} c_i & \text{when } x_i \text{ is cluster center} \\ c_{n_i} & c_i = -1 \end{cases} \quad (5)$$
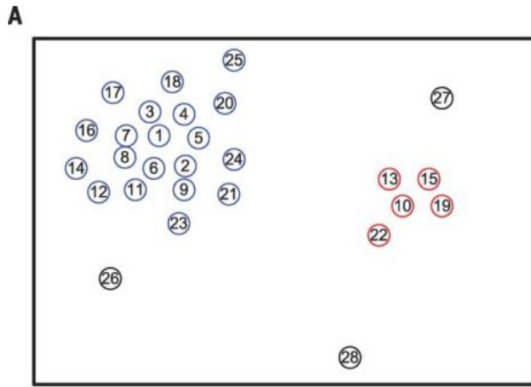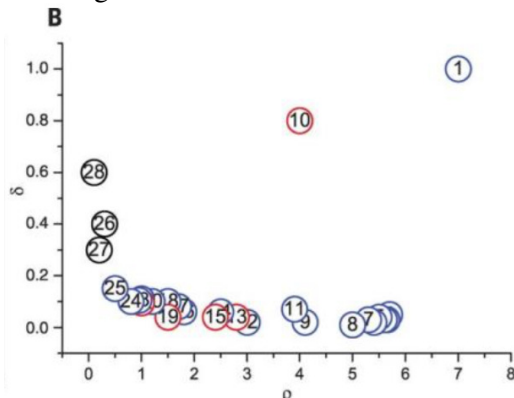

figure 1 data distribution of clusters

figure 2 decision graph of clusters center

II. IMAGE COLOR QUANTIZATION ALGORITHM BASED ON CFSFDP

This section describes a CFSFDP-based image color quantization method that describes how CFSFDP is applied to color quantization, including preprocessing color values, selecting clustering centers, and detailed processes for quantizing algorithms.

1) Color Value Preprocessing

For the input image I_s, the color histogram of its RGB space is obtained, and the histogram feature statistics are sorted in descending order to obtain the characteristic statistics sequence.

$$H = \left\{ h_{c_1}, \; h_{c_2}, \; \dots, \; h_{c_k} \mid \sum h_{c_j} = 1 \right\} \quad (6)$$

which $h_{c_1} > h_{c_2} > \dots > h_{c_k}$, $h_{c_j} = \dfrac{n_{c_j}}{N}$, and n_j is the number of pixels with eigenvalues c_j in the image, N is the total pixels number of image. In order to remove the isolated point and the noise, the top 90% of the characteristic statistics are intercepted, that is, there is a constant which makes

$$\sum_{j=1}^{l} h_{c_j} = 0.9$$

A denoised histogram is obtained. $\tilde{H} = \left\{ h_{c_1}, \; h_{c_2}, \; \dots, \; h_{c_l} \mid \sum h_{c_j} = 0.9 \right\}$

2) Select the cluster center

Through the above description, we can see that CFSFDP is more efficient than k-means, has better clustering effect, and has better clustering performance for irregular edge clustering. Therefore, we choose CFSFDP as color clustering algorithm. In the above description, CFSFDP selects the clustering center by drawing the decision graph, and it is clear that the clustering center can not be selected through the decision graph in the color quantization process. In the application of image color characteristics, such a clustering center can be obtained by counting the main color values of the histogram of the color values, the main color values can be obtained by counting the first 70% color values, and then merging the selected color values For the approximate color values, the weighted average is obtained by averaging the color values as the cluster center.

3) Quantification method

For the input image F_s, the color values are clustered by the CFSFDP algorithm to obtain the clustering set $S = \{S_1 \dots S_k\}$ and the cluster center set $C = \{C_1 \dots C_k\}$, where k is the number of clustering centers, S_i is the set of color values for the i-th cluster, and C_i the color value of i-th cluster center. For the each color values of the images, the color values in each category are replaced with the clustering center color values.

III. EXPERIMENT

The main experimental data in this paper is cloth material. The data set contains 2149 cloth pictures including solid color, simple texture, complex texture, are captured by professional cameras in the same stable light source. will get the image collection for Farbic-MC. The data set is divide into solid color set(Farbic-PC), simple texture set(Farbic-SC) and complex texture set(Farbic-CC).

This section compares the experimental k-means clustering quantization with the proposed CFSFDP-based quantization method. The experimental conditions are as follows:

（1）The k-means algorithm uses the method of random initialization to cluster the color values of the RGB space in the image. For each class cluster, use the color values of the clusters at the center of the cluster.

（2）Experimental method: Farbic-PC, Farbic-SC, Farbic-CC image collection, respectively, with the image collection of each picture as an input image to retrieve, each return 10 pictures, the average recall rate.

（3）Evaluation standard:

recall rate:

$$E = \frac{m}{t}$$

Chromaticity error:

$$E_d = \sqrt{(r_a - r_e)^2 + (g_a - g_e)^2}$$

Figure 3 shows that the CFSFDP algorithm is better than the k-means algorithm, and the quantized image is smaller than the original image and has better quantization effect in the three types of fabrics. Figure 4 can also be seen that the CFSFDP quantization algorithm effectively improves the performance of the retrieval system. The recall rate in the Farbic-PC image set is about 6% higher than that of the k-means and about 4% in the Farbic-SC image set. Which is related to the characteristics of the two algorithms. The clustering formed by the distribution of the image color values in the RGB space is not spherical or even irregular. The density-based CFSFDP can better adapt this kind of irregular shape. The color values are better assigned to the clusters, resulting in better

quantification. It should be noted that this paper uses the color histogram as a feature extraction method, so both method quantificating Farbic-CC, in the recall rate comparison has low performance.

	Farbic-PC	Farbic-SC	Farbic-CC
k-means	10.4183	16.4342	30.2156
CFSFDP	0.4465	0.9437	3.4477

figure 3 Comparison of Chromaticity Errors in Different Quantization Methods

	Farbic-PC	Farbic-SC	Farbic-CC
k-means	84.4183%	82.9976%	73.3035%
CFSFDP	90.4465%	86.7432%	72.3987%

figure 4 Comparison of the results of different quantitative methods

IV. CONCLUSIONS

In this paper, CFSFDP is applied to color quantization, and the deficiency of k-means quantization algorithm is improved. In this paper, we mainly study the application of CFSFDP in color quantization, and introduce the statistical assistant selection of color histogram for the problem that CFSFDP can not automatically select the quantization center. The method of removing isolated points and noise in CFSFDP application is proposed. Through the experimental comparison of k-means and CFSFDP clustering algorithm to quantify the results, this paper proposed based on CFSFDP clustering algorithm is indeed more effective than k-means clustering algorithm.

REFERENCES

[1] RODRIGUEZ, Alex; LAIO, Alessandro. Clustering by fast search and find of density peaks. *Science*, 2014, 344.6191: 1492-1496.

[2] VEREVKA, Oleg. Color image quantization in windows systems with local k-means algorithm. In: *Proceedings of VI Western Computer Graphics Symposium*. 1995. p. 74-79.