Pakistan Journal of Applied Sciences 1 (4): 508-511, 2001

Color Quantization by Modified K-Means Algorithm

Songül Albayrak
Computer Engineering Department
Yildiz Technical University 80750 Istanbul, Turkiye

Abstract: In this study, K-means algorithm, a distance-based clustering algorithm, is modified depending on the color quantization application. Each color cluster center is calculated as its weighted mean by using histogram value. This algorithm uses average distortion optimization strategy to improve the perceived image quality on quantized image. In this application we have also used two different color spaces like RGB and CIELab to examine the effect of color spaces on clustering. Our application supports mapping from 256-color to 16-color images to show the results.

Keywords: Clustering, Color quantization, K-Means algorithm, Color spaces

Introduction

The application of an algorithm which is used to reduce the number of colors used in an image is called color quantization. The ultimate goal of color quantization is the mapping of each color in the original image to a new color in a representative color map, while maintaining the appearance of the original image.

Color quantization is very popular subject in the field of computer graphics, image processing and data compression. After the quantization process of an image from 24-bit color to 8-bit color, the required space to store and also transmission bandwidth can be decreased. In this application we use a quantization form 256-color to 16-color to show the efficiency of the algorithm.

We present an improved clustering method on color clustering in 3-D color spaces. In this application, we have used a modified K-Means algorithm which is an iterative optimization method and called it as weighted K-Means algorithm. The proposed method distinguishes itself from other by improving the perceived image quality on quantized image.

In this paper we will give some formulation about the quantization process and color spaces like RGB and CIELab which is used in this application.

Color Spaces

RGB: Images must be represented on RGB space to display them on a monitor. This is the cause of the fact that all CRT displays have the red, green and blue phosphors. The result of that RGB color space becomes the most used color space. It is a simple cube and has a 3-D simple geometry (Jain 1989).

The components of color space represent the brightness of corresponding color. If all the axis are zero, it represents black and if all colors are the maximum the result is white. The points on the diagonal line between the black and white point are gray. The disadvantage of the RGB space is it is not **perceptually uniform.**

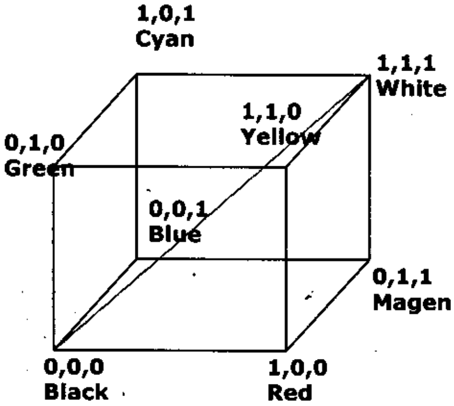

Fig. 1: The RGB cube with binary components

CIE; In 1931, the **Commission Internationale de l'Eclairege (CIE)** developed a device dependent color model based on human perception. The CIE XYZ model, as its known, defines three primaries called X,Y and Z that can be combined to match any color humans see. This relates to the **tristimulus theory** of color perception, that states that the human retina has 3 kinds of cones with peak sensitivities to 580nm (red), 545nm (green), and 440nm (blue). These primaries are combined in an additive manner.

The CIE XYZ color space and the CIE chromaticity diagram are not perceptually uniform. The advantage of the CIELab color space which is introduced in 1976 is that the distance between points is now approximately proportional to the perceived color difference, something definitely not true in the 1931 CIE XYZ diagram (Albayrak 2001).

Table 1: Standard White Points with chromaticity and Color Temperature

	Color Temperature [K]	CIE 1931		CIE 1964		Source
		x_w	y_w	x_w	y_w	
Illuminant A	2856	0.44757	0.40745	0.4512	0.4059	[Wyszecki 82] [Aguston 87]
Illuminant B	4874	0.34842	0.35161	0.3498	0.3527	[Wyszecki 82] [Aguston 87]
Illuminant C	6774	0.310006	0.31616	0.3104	0.3191	[Wyszecki 82] [Aguston 87]
Illuminant D65	6504	0.3127	0.3291	0.3138	0.3310	[Walker 98] [Aguston 87]
Illuminant D65	6504	0.312713	0.329016	-	-	[CICA 98]
Direct Sunlight	5335	0.3362	0.3502	-	-	[Aguston 87]
Light from overcast sky	6500	0.3134	0.3275	-	-	[Aguston 87]
Illuminant E	5400	1/3	1/3	1/3	1/3	[Wyszecki 82] [Aguston 87]

Eq (1) is used to transform from RGB color space to CIELab color space. Table 2 gives the chromaticity coordinates of CIE illuminant D65 which is used to compute (X_r,Y_r,Z_r), (X_g,Y_g,Z_g), and (X_b,Y_b,Z_b).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Table 2. Chromaticity Coordinates of ITU-R BT.709 and CIE standard Illuminant D65

	Red	**Green**	**Blue**	**White-D65**
x	0.6400	0.3000	0.1500	0.3127
y	0.3300	0.6000	0.0600	0.3290
z	0.0300	0.1000	0.7900	0.3583

By using CIE standard Illuminant D65 reference white as $[X_w\ Y_w\ Z_w] = [0.95045\ \ 1.0\ \ 1.088754]$ and ITU-R BT.706 camera standard, the transformation matrix of Eq. (2) is obtained to transform from RGB to CIEXYZ color space.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2)$$

Brightness (L^*), red-green component (a^*) and yellow-blue component (b^*) can be written as follows (Jain 1989);

$$L^* = 116(Y/Y_w)^{1/3} - 16, \qquad 1 \le 100Y \le 100 \quad (3)$$

$$a^* = 500[(X/X_w)^{1/3} - (Y/Y_w)^{1/3}] \quad (4)$$

$$b^* = 200[(Y/Y_w)^{1/3} - (Z/Z_w)^{1/3}] \quad (5)$$

Definition of Color Quantization Problem: Suppose $C = \{c_i,\ i = 1,2,\ldots,K\}$ is a set of all colors which are 3 dimensional vector in the color imege. N is the number of all colors in the color image I. Color quantization is performed by clustering the color space into a given number K of cluster S_k. $\overline{C} = \{\overline{c}_j, j = 1,2,\ldots,K\}$ is a set of quantization levels and $\overline{I}$ represents the quantized image (Verevka 1995).

- In this representation, quantization process can be called a mapping process;

$$q: C \to \overline{C} \quad (6)$$

- Each color c is mapped to the nearest quantization level $\overline{c}$;

$$d(c,\overline{c}_i) \le d(c,\overline{c}_j)\ \ j = 1,2,\ldots,K \ \Rightarrow q(c) = \overline{c}_i \quad (7)$$

d represents the Euclid distance between three dimensional vectors c and $\overline{c}$.

- $S_k,\ k = 1,2,\ldots,K$ is a set of clusters which is obtained after a quantization process in the image color space.

$$S_k = \{c \in C:\ \ q(c) = \overline{c}_k\} \quad (8)$$

(a)

(b)

(c)

Fig. 2: Original RGB Lenna image (a) is quantized by Weighted K-Means (b) and K- Means (c)

(a)

(b)

(c)

Fig. 3: Original RGB Pepers image (a) is quantized by Weighted K- Means (b) and K-Means (c)

The distortion occurs when we place a color c by its quantized value $q(c)$. This is called as average quantization error of a color and represented by $\varepsilon_q\,(c,q(c))$;

$$\varepsilon_q\,(c,q(c)) = \frac{1}{N}\sum_{i=1}^{N} d(c_i - q(c_i)) \qquad (9)$$

- A measure of quantization error must take into account not only the quantization error of each color but also the frequency of the occurrence of the color on the image. Therefore a distortion measure over an image must be defined. The quantization error of an image I is given by

$$\varepsilon_q\,(c,I) = \frac{1}{\sum_{i=1}^{N} F(c_i)} \sum_{j=1}^{N} F(c_j).d(c_j,\bar{c}_j) \qquad (10)$$

F(c) is the frequency of the color c on the image.

Weighted K-Means Algorithm: In the RGB and CIELab color spaces, modified K-Means algorithm is performed to obtain color clusters and representative color of this clusters. Representative color of a cluster should be nearby place of dominant color which means the most used color. In weighted K-Means algorithm, weighted means are computed at each iteration by the way of histogram values. Weighted K-Means algorithm is performed as follows;

- Randomly select K color centers $m_1(1), m_2(1), \ldots m_K(1)$ and compute the color distances of each color from the color centers. Color distance between the points of **a** and **b** that's in n-dimensional color space is computed by Euclid distance as follow (Duda and Hart 1973);

$$d(\mathbf{a},\mathbf{b}) = \sqrt{\sum_{i=1}^{n} (\mathbf{a}_i - \mathbf{b}_i)^2} \qquad (11)$$

- At the t'th iterative step distribute the samples {c} among the K cluster domains, using the relation $c \in S_j(t)$ *if* $\|c - m_j(t)\| < \|c - m_i(t)\|$ for all i=1,2,...,K, i≠j (12)
  where $S_j(t)$ denotes the set of samples whose cluster center is $m_j(t)$

- With the new members of clusters, compute new cluster centers as weighted means given by Eq (13)

$$m_j(t+1) = \frac{1}{T_j}\sum_{c_i \in S_j(t)} F(c_i).c_i,\ j = 1,2,\ldots,K \quad \forall c_i \mid c_i \in S_j(t) \qquad (13)$$

$$T_j(t) = \sum_{c_i \in S_j(t)} F(c_i),\quad \forall c_i \mid c_i \in S_j(t)\ i = 1,2,\ldots,256 \qquad (14)$$

- If $m_j(t+1)=m_j(t)$ for j=1,2,...,K the algorithm is converged and the procedure is terminated. Otherwise go to Eq (12).

Experimental Results: To demonstrate the results of color image quantization algorithm we have chosen a set of 8-bit RGB images like Lenna, Mandrill and Peppers. We quantize the color image from 256-color to 16-color by K-Means and Weighted K-Means algorithms. Table 3 gives the image resolution, average quantization error of a color and quantization error of each image. Weighted K-Means algorithm gave a smaller quantization error of each images.

Table 3: Average quantization error of a color and quantization error of each images for K-Means and Weighted K-Means algorithms

		Lenna	Mandrill	Peppers
K-Means	$\varepsilon_q(c,q(c))$	21.38	27.59	21.36
	$\varepsilon_q(c, I)$	17.94	25.34	19.08
Weighted K-Means	$\varepsilon_q(c,q(c))$	22.99	27.83	21.41
	$\varepsilon_q(c, I)$	14.55	24.19	18.04

Fig.2 and Fig.3. (a) show original 256-color RGB Lenna and Peppers images. (b) show quantization results of 16-color for Weighted K-Means and (c) show quantization results for traditional K-Means clustering.

Conclusion

In this study we proposed modified K-Means algorithm which is a distance based method with histogram values in 3-D color spaces like RGB and CIELab . We called this method as weighted K-Means algorithm. In this way we have improved the perceived image quality on quantized image. In addition, CIELab color space is efficient on the applications of distance based classification because the distance between the points in CIELab color space is approximately proportional to the perceived color difference.

References

Aguston G.A.,(1987) Color Theory and It's Application in Art and Design, Springer-Verlag

Albayrak S., (2001), Renk Özelliği ile İçerik Tabanlı Görüntü Erişimi, PhD Thesis, Yildiz Technical University, Istanbul,2001

Duda R.O. and Hart P.E.,(1973) Pattern Classification and Scene Anaysis, A Wiley-Inter Science Publication

Jain A.K., (1989) Fundamentals of Digital Image Processing, Prentice Hall International Editions

Tou J.T. and Gonzalez R.C., (1974) Pattern Recognition Principles, Addison-Wesley Publishing Copmany

Walker J.,(1976), Color Rendering of Spectra, April

Wyszecki G. and Stiles W.S.,(1982), Color Science Concept and Methods, Quantative Data and Formula, John Wiley

Verevka O., (1995) "Local K-Means Algorithm for Colour Image Quantization", Proceeding of Graphics Interface, pages 128-135