International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:5, 2013

A Hybrid Approach for Color Image Quantization Using K-means and Firefly Algorithms

Parisut Jitpakdee, Pakinee Aimmanee, and Bunyarit Uyyanonvara

***Abstract*—**Color Image quantization (CQ) is an important problem in computer graphics, image and processing. The aim of quantization is to reduce colors in an image with minimum distortion. Clustering is a widely used technique for color quantization; all colors in an image are grouped to small clusters. In this paper, we proposed a new hybrid approach for color quantization using firefly algorithm (FA) and K-means algorithm. Firefly algorithm is a swarm-based algorithm that can be used for solving optimization problems. The proposed method can overcome the drawbacks of both algorithms such as the local optima converge problem in K-means and the early converge of firefly algorithm. Experiments on three commonly used images and the comparison results shows that the proposed algorithm surpasses both the base-line technique k-means clustering and original firefly algorithm.

***Keywords*—**Clustering, Color quantization, Firefly algorithm, K-means.

I. Introduction

COLOR image quantization or color quantization (CQ) is an image processing technique for reduction the number of colors in image, useful in the limitations of image display, data storage and transmission [1]. There are two types of color quantization algorithms [2]; first is the splitting algorithm such as: median-cut [3], center-cut [4], oc-tree [5], and the second is clustering algorithms. Color quantization by clustering can be done by grouping color points into small clusters and then finding a representative for each cluster. This clustering-based quantization problem is an optimization problem because it wants to minimize the error of quantization by minimizes the sum of distance between the center of each cluster and its members. The maximum inter-cluster distance is minimized to find the global solution of quantization image in [6]. The most popular clustering method, K-means algorithm is used in color quantization [7] and [8].

Many optimization algorithms such as swarm intelligence algorithms or nature-inspired optimization algorithms are applied for performing color quantization such as; Particle Swarm Optimization (PSO) [9], a modified artificial fish swarm algorithm [10], Bacteria Foraging Optimization [11], Honey Bee Optimization [12] A hybrid approach of a hybrid of fuzzy c-means (FCM), particle swarm optimization (PSO), and genetic algorithms (GA) is proposed for color quantization [13].

One of famous bio-inspired optimization algorithm that will be used mainly in this paper is the Firefly algorithm (FA) introduced by Xin She Yang [14]. This algorithm based on the flash producing behavior of fireflies. Yang used the FA for nonlinear design problems [15] and multimodal optimization problems [16] and showed the efficiency of the FA for finding global optima.

In this study, the Firefly Algorithm (FA) is applied together with K-means algorithm to cluster colors in images for color quantization. To study the performance of the our method, we used 3 difference images from The USC-SIPI Image Database [17] and evaluate results by calculating the Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR).

The paper is organized as follows. Section 2 briefly describes a firefly algorithm. Section 3 presents a detail of proposed method, FA+K. The experimental results are showed in Section 4. Finally, Section 5 is a conclusion of this study.

P. Jitpakdee is with Sirindhorn International Institute of Technology, Thammasat University, 131 Moo 5, Tiwanont Road, Bangkadi, Muang, Pathumthani, 12000, Thailand (e-mail: parisut@gmail.com).

P. Aimmanee is with Sirindhorn International Institute of Technology, Thammasat University, 131 Moo 5, Tiwanont Road, Bangkadi, Muang, Pathumthani, 12000, Thailand (e-mail: pakinee@siit.tu.ac.th).

B. Uyyanonvara is with Sirindhorn International Institute of Technology, Thammasat University, 131 Moo 5, Tiwanont Road, Bangkadi, Muang, Pathumthani, 12000, Thailand (e-mail: bunyarit@siit.tu.ac.th).

II. Firefly Algorithm

A Firefly Algorithm (FA) is an optimization algorithm that simulates the flash pattern and characteristics of fireflies. It first proposed by Yang [14], [15], and [16]. The Firefly Algorithm is a population-based algorithm to find the global optima of objective functions based on swarm intelligence. Each firefly is attracted by the brighter glow of other neighboring fireflies. When a couple of fireflies are father away, the attractiveness is decreasing.

In Firefly algorithm, there are three idealized rules defined by Yang [14]: 1) All fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex; 2) Attractiveness is proportional to their brightness. Thus, for any two fireflies, the less bright one will move towards the brighter one. If there is no brighter one than a particular firefly, it will move randomly; 3) the brightness of a firefly is from the objective function. For a maximization problem, the brightness can simply be proportional to the value of the objective function. The pseudo code of Firefly algorithm [14] is shown in Fig. 1.

The attractiveness function β calculate from the distance $r_{i,j}$ of the firefly is determined by:

$$\beta(r_{i,j}) = \beta_0 e^{-\gamma r_{i,j} r^2} \quad (1)$$

where, β_0 is the attractiveness and γ is the light absorbtioncoefficient at the source. It should be noted that the

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:5, 2013

$r_{i,j}$which is the euclidence distance between any two fireflies i and j at $x_i$and x_j, where x_i and x_j are the spatial coordinate of the fireflies i and j, respectively.

The movement of a firefly i, which is attracted to another more attractive firefly j is determined by:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{i,j}^2}\left(x_i^t - x_j^t\right) + \alpha\left(rand - \tfrac{1}{2}\right) \qquad (2)$$

where x_i^t is the firefly i at the iteration t, α is the randomization parameter.

Firefly Algorithm: FA

Objective function $f(x), x = (x_1, \ldots, x_d)^T$
Define the parameters
Generate initial population of fireflies x_i $(i = 1,2, \ldots, n)$
Calculate Light intensity L_i at x_i is determined by $f(x_i)$
Repeat:
for$i = 1{:}n$ all n fireflies
for$j = 1{:}n$ all n fireflies
if (L_j>L_i),
Move firefly i towards j in d-dimension
Attractiveness varies with distance r via $exp[-\gamma r]$
Evaluate new solutions and update light intensity
if no one of firefly brighter than L_i, L_i move randomly
Rank the fireflies and find the current best
Until: Maximum iteration or minimum change of objective function

Fig. 1 The pseudo code of Firefly Algorithm (FA)

III. Proposed Algorithm

A. Clustering Based Color Quantization

Data Clustering is a kind of unsupervised learning. The aim of clustering is to allocate dataset into subgroups or clusters with homologous properties. Clustering is applied to a variety of problems, including data mining, text mining, image analysis, and pattern recognition. The k-means clustering and is the most popular clustering techniques because their simplicity of implementation and convergence speed but the drawback is they are sensitive to initialization that easily adjoined in local optima [18]. Clustering-based color quantization clusters the colors in images to small groups of color with minimum distortion. In this study we use RGB color space; including red channel, green channel, and blue channel. Every pixel in image is represented by three values; red level, green level, and blue level. Let I is an original image, represented as

$$I = [p_1, p_2, \ldots, p_m] \qquad (3)$$

where$m$ is the number of pixels in image I, and p_i is the i^{th}pixel of image, given by

$$p_i = [R_i, G_i, B_i] \qquad (4)$$

whereR_i, G_i, B_i are the red level, green level, and blue level of pixel p_i.

To quantize image to k colors, every pixel in original image are clustered to k groups, and then the pixel members of each group are replaced by its centroid. The set of centroid result represented by

$$C = [c_1, c_2, \ldots, c_k] \qquad (5)$$

wherec_j is the centroid of cluster j , calculated asthe mean of cluster:

$$c_j = [\frac{\sum_{i\in c_j}^{m_j} R_i^j}{m_j}, \frac{\sum_{i\in c_j}^{m_j} G_i^j}{m_j}, \frac{\sum_{i\in c_j}^{m_j} B_i^j}{m_j}] \qquad (6)$$

wherem_j is the number of members in cluster j, and R_i^j, G_i^j, B_i^j is the red, green, and blue level of pixel $p_i$member of cluster c_j.

K-means algorithm is famous clustering algorithm; it divided data into k clusters. The initial centroids are random selected. Each data point is grouped to the nearest centroid, usually calculate the distance by the Euclidean distance. The new centroid is the mean of data in each cluster. The data point is re-grouped to the nearest centroid and the new centroids are calculated again and repeat this until convergence.

Fig. 2 shows the example result of quantization by K-means clustering. Fig. 2 (a) is the original image and the quantized image in 4 colors is showed in Fig. 2 (b). Fig. 2 (c).shows the scatter plot of every pixel in original image in three dimensions, and compare with the Fig. 2 (d) the same scatter plot but the color of each point are replaced by the color of its centroid.

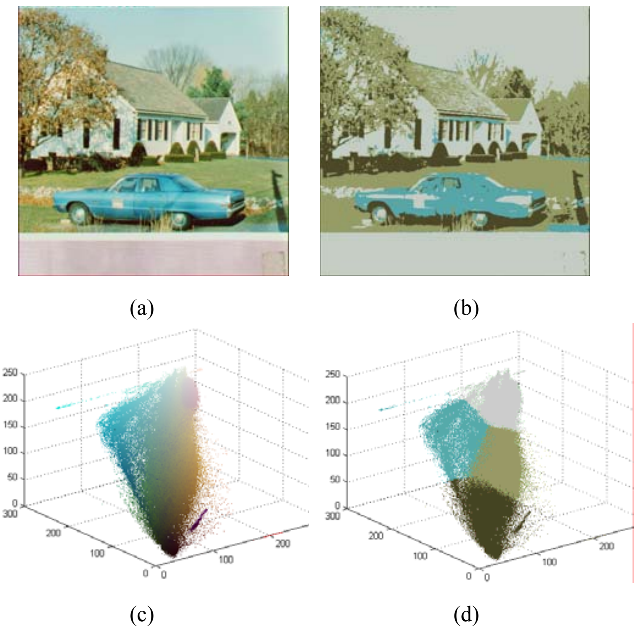


Fig. 2 The example result of clustering by K-means Algorithm (a) an original image (b) 4 colors quantized image (c) the scatter plot of original image (d) the scatter plot of quantized image

B. Hybrid Approach of Firefly Algorithm and K-Means

We proposed 2 stages of clustering; Pre-clustering and Post-clustering.

In first stage, which is called pre-clustering stage, we used firefly algorithm for clustering the pixels in image to sub groups. The drawback of firefly algorithm is that the firefly may prematurely converge, or in the other hand it takes a long times to converge, depend on the problem and the setting of the parameters such as the attractiveness and the light absorption coefficient. To handle with this problem,we set number of iterations of firefly algorithm to a fix small number, *maxF*. We run the algorithm to find the set of centroids. The centroid results theoretically will be close to the correct solution. Each firefly x_i is the candidate colors of quantized image, represented as the set of centroid pixels in red, green, and blue channel, represented by:

$$x_i = \left[c_1^i, c_2^i, \dots, c_k^i\right] \quad (7)$$

where x_i is the i[th] firefly,*i*= 1,2,…*n*, *n* is the number of fireflies and c_j^i is the centroid of cluster *j* in firefly *i*, *j*= 1,2,…*k*, and *k* is the number clusters.

The main purpose of quantization is to reduce colorsin image with minimum distortion. The objective function in firefly algorithm is the summation of the distance between each firefly member and its centroids, also called compactness, which can be formulated as below:

$$Compactness = \frac{1}{np}\sum_{j=1}^{k}\sum_{p_i \in C_j}\left(p_i - z_j\right)^2 \quad (8)$$

where, np is number of image pixels,k is number of cluster, p_i is the pixel *i*member of cluster C_j.andz_j is the center of the cluster C_j.

The centroids results are adjusted by k-means clustering in the second stage. The drawback of k-means algorithm is it sensitive to initialization that easily close to local optima. To solve this problem, we use the initial centroids obtained from firefly algorithm in previous stage. Fig. 3 shows the flowchart of our approach.

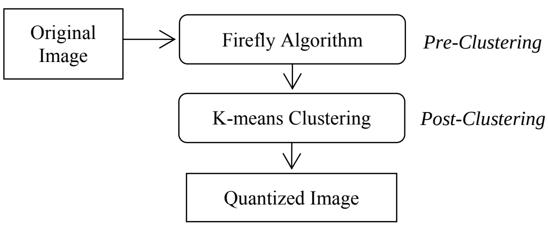


Fig. 3 Flowchart of our color quantization approach (FA+K)

IV. Experiments and Results

For evaluate the advantage of our approach, we have tested in three popular images for image processing testing from USC-SIPI Image Database [17], named “Lena”, “Peppers”, and “Mandrill”. The images are 512x512 sized. Experiments are implemented using MATLAB Version 7.10 on an Intel CORE[TM] i5 1.6 GHz with 3.9 GB RAM.

We compared our algorithm (FA+K) with the k-means clustering and the original firefly algorithm. The parameter setting show in Table I.

Color quantization algorithms efficiency is measured by Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) [19]. MSE measures the distortion between the original image and its quantized image. If the MSE is small, the resulted image from CQ closely is similar to the original. MSE is computed by (9):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(I_i - \hat{I}_i\right)^2 \quad (9)$$

where, I_i is the pixel *i*of theoriginal image, $\hat{I}_i$ is the pixel *i*of thequantized image, and *n* is the number of pixels.

TABLE I
Parameters Setting

Parameter	K-means	FA	FA+K
Maximum iteration	200	100	20+50
β_0	-	1	1
γ	-	1	1
α	-	1	1
n	-	10	10

PSNR calculates by (10). The higher the number, the more correct the Quantization.

$$PSNR = 10\ \log_{10}\left(\frac{M^2}{MSE}\right) \quad (10)$$

where, M is the maximum value of the original image.

The quantization results of our hybrid approach compared with the original methods k-means and firefly algorithm with different number of quantized color *k*; 8, 16, 32 are shown in Table II, III, and IV, respectively. The original images and quantized images results in 8, 16, and 32 colors, are shown in Fig. 4, 5, and 6, respectively.

TABLE II
Comparison PSNR and MSE for Lena and Peppers and Mandrill images with K= 8

Method	Lena		Peppers		Mandrill	
	PSNR	*MSE*	*PSNR*	*MSE*	*PSNR*	*MSE*
K-means	26.810	135.533	24.227	245.688	22.416	372.768
FA	26.831	134.900	23.944	262.235	22.408	373.344
FA+K	**26.886**	**133.195**	**24.307**	**241.198**	**22.425**	**372.029**

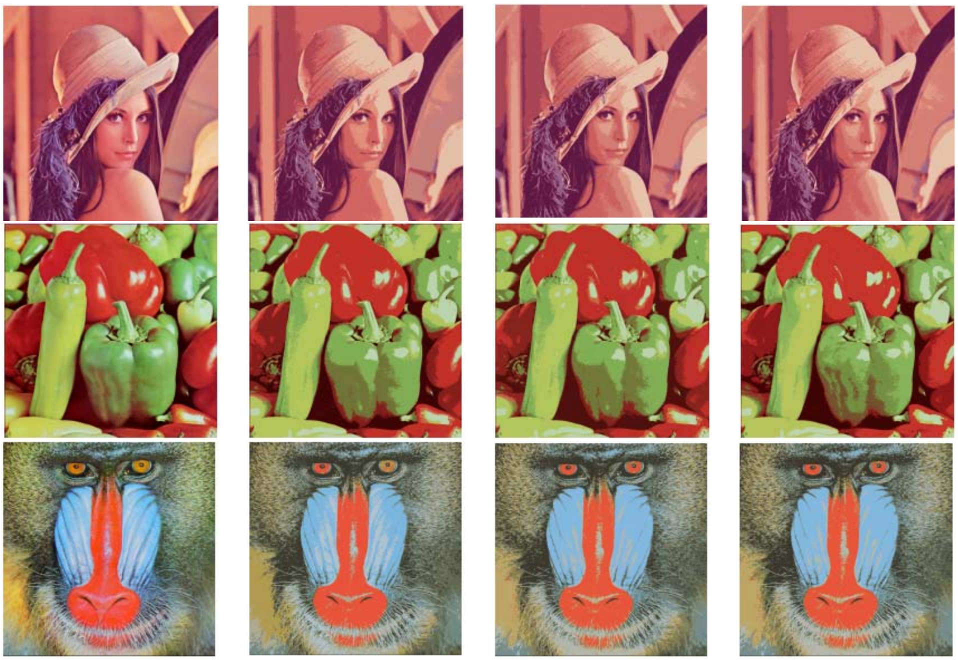
TABLE III
Comparison PSNR and MSE for Lena and Peppers and Mandrill images with K= 16

Method	Lena		Peppers		Mandrill	
	PSNR	*MSE*	*PSNR*	*MSE*	*PSNR*	*MSE*
K-means	29.578	71.659	26.695	139.205	24.907	210.076
FA	29.583	71.586	26.366	150.154	24.905	210.198
FA+K	**29.700**	**69.671**	**26.853**	**134.220**	**24.923**	**209.302**

TABLE IV
COMPARISON PSNR AND MSE FOR LENA AND PEPPERS AND MANDRILL IMAGES WITH K=32

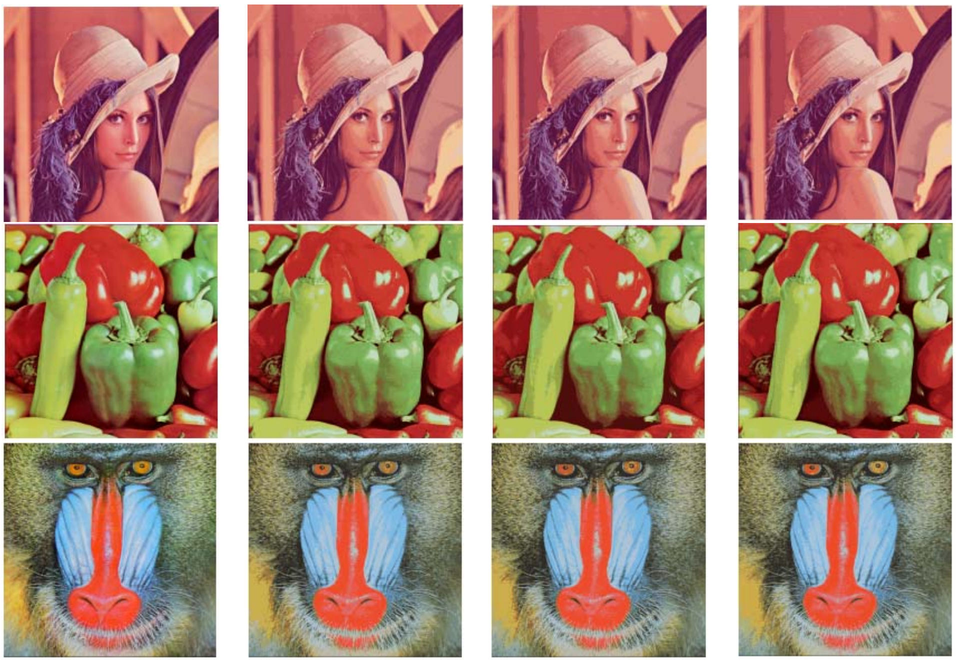
Method	Lena		Peppers		Mandrill	
	PSNR	*MSE*	*PSNR*	*MSE*	*PSNR*	*MSE*
K-means	32.178	39.384	29.254	77.209	27.153	125.264
FA	32.078	40.282	28.937	83.059	27.160	125.048
FA+K	**32.199**	**39.192**	**29.321**	**76.023**	**27.191**	**124.155**

Fig. 7, 8, and 9 show the "Lena" quantized images, 8 colorsin each channel, red, green, and blue, respectively, and show the histograms with the centroid lines of each clusters from by K-means algorithm (dot-dash lines), FA algorithm (dash line) and our approach FA+K (straight line).



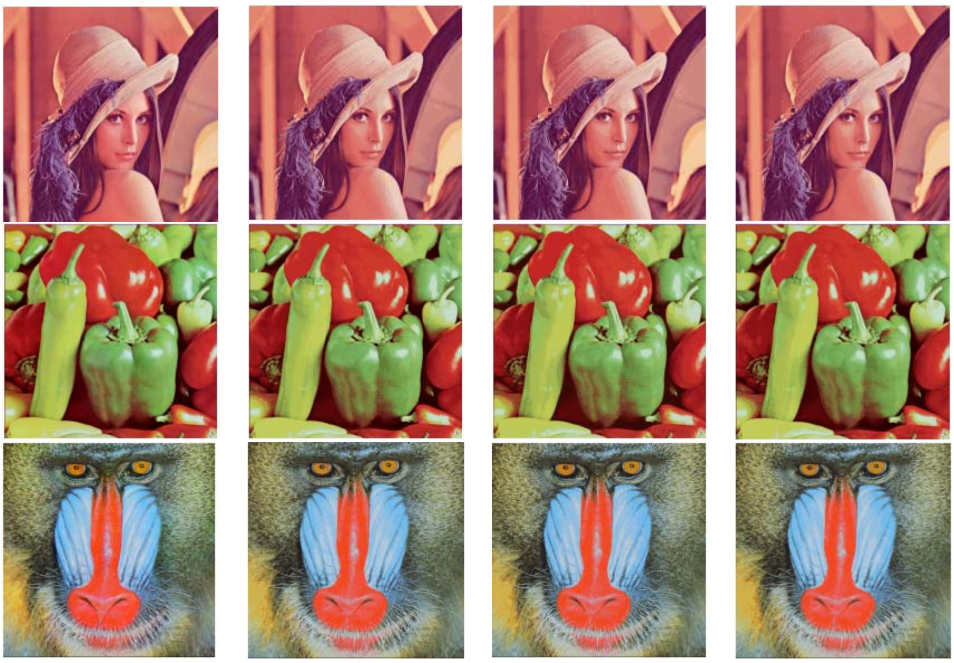
Original Image K-means FA FA+K

Fig. 4 Original image and quantized image in 8 colors from 3 methods; K-means, Firefly algorithm and our approach



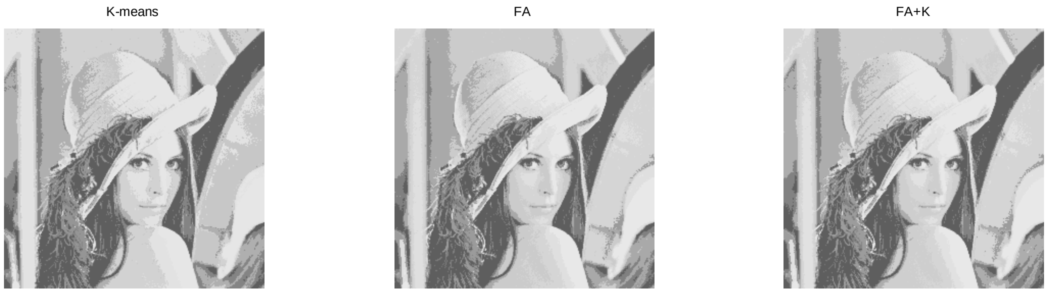
Original Image K-means FA FA+K

Fig. 5 Original image and quantized image in 16 colors from 3 methods; K-means, Firefly algorithm and our approach

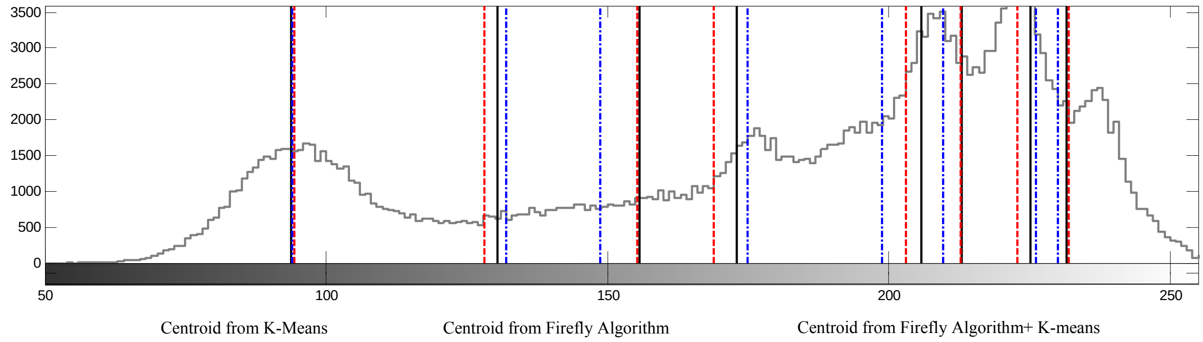


Original Image K-means FA FA+K

Fig. 6 Original image and quantized image in 16 colors from 3 methods; K-means, Firefly algorithm and our approach



(a) Red channel Images

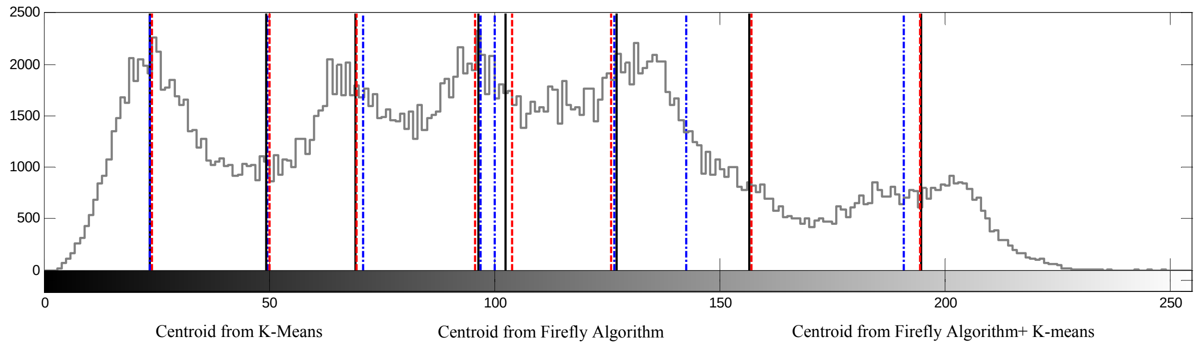


(b) Red channel Histogram

Fig. 7 (a) the red channel quantized images of "Lena" image in 8 colors from 3 methods; K-means, Firefly algorithm and our approach(b)The histograms of Lena Image in red channel and the centroids of each clusters from quantization results by K-means algorithm (dot-dash lines), FA algorithm (dash line) and our approach FA+K (straight line)

(a) Green channel Images

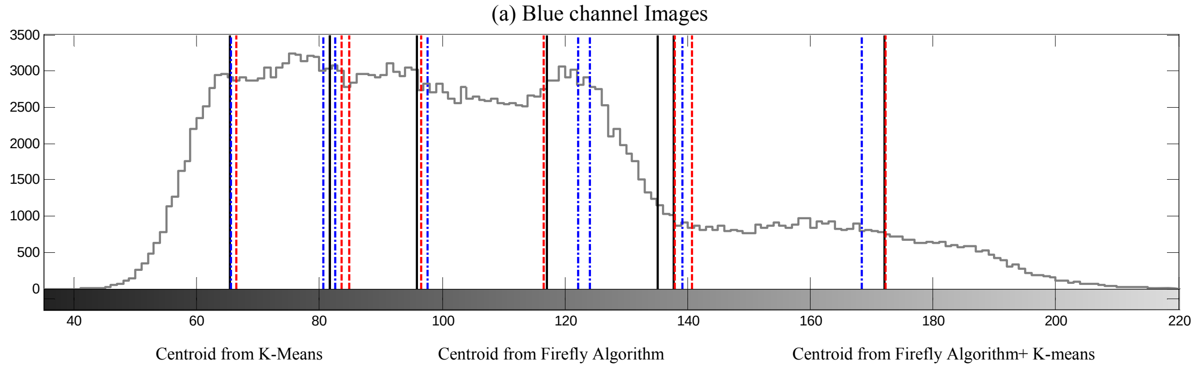
International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:5, 2013



(b) Green channel Histogram

Fig. 8 (a) the green channel quantized images of "Lena" image in 8 colors from 3 methods; K-means, Firefly algorithm and our approach(b)The histograms of Lena Image in red channel and the centroids of each clusters from quantization results by K-means algorithm (dot-dash lines), FA algorithm (dash line) and our approach FA+K (straight line)

(a) Blue channel Images



(b) Blue channel Histogram

Fig. 9(a) the blue channel quantized images of "Lena" image in 8 colors from 3 methods; K-means, Firefly algorithm and our approach(b)The histograms of Lena Image in red channel and the centroids of each clusters from quantization results by K-means algorithm (dot-dash lines), FA algorithm (dash line) and our approach FA+K (straight line)

V. Conclusion

This paper proposes a new hybrid approach for color image quantization based on firefly algorithm and K-means clustering method. The potential centroids of the clusters from Firefly algorithm are adjusted by K-means algorithm. The combination approach of two algorithms can solve the problem of each other. The performance of proposed algorithms is measured using the mean square error (MSE) and the peak signal-to-noise ratio (PSNR). The experimental results over 3 images show that the proposed algorithm is efficient. The MSE is lower than both K-means clustering algorithm and the normal Firefly clustering algorithms.

References

[1] J. P. Braquelaire and L. Brun, "Comparison and optimization of methodsof color image quantization," *IEEE Transactions on Image Processing*, vol. 6, no. 7, 1997, pp. 1048–1052.

[2] P. Scheunders, "A comparison of clustering algorithms applied to color image quantization,",*Pattern Recognition Letters*, New York, vol.18, 1997, pp. 1379-1384.

[3] P. Heckbert, "Color Image Quantization for Frame Buffer Display,"*Computer Graphics(Proc. Siggraph)*, vol. 16, no. 3 July 1982, pp. 297-307.

[4] G. Joy, Z.Xiang, "Center-cut for color image quantization," *The Visual ComputerI, Berlin,* vol.10, January 1993,pp. 62-66.

[5] Gervautz, M., Purgtathofer, W., "A Simple Method for Color Quantization: Octree Quantization," CA: Academic, San Diego, 1990.

[6] Zhigang Xiang,"Color image quantization by minimizing the maximum intercluster distance," *ACM Trans. Graph.*,vol.16, no.3,July 1997, pp. 260-276.

[7] M. EmreCelebi,"Effective initialization of k-means for color quantization,"*In Proc, IEEE int. conf. on Image processing, 2009* IEEE Press, Piscataway, NJ, USA, pp. 1629-1632.

[8] M. EmreCelebi, "Improving the performance of k-means for color quantization,"*Image Vision Comput*., vol.29, no.4, March 2011, pp. 260-271.

[9] Ziqiang Wang, Xia Sun, and Dexian Zhang, "A swarm Intelligence Based Color Image Quantization Algorithm," *Int. Conf. Bioinformatics and Biomedical Engineering: ICBBE 2007*, July, 2007, pp. 592-595.

[10] DanialYazdani, HadiNabizadeh, ElyasMohamadzadehKosari, and Adel NadjaranToosi, "Color quantization using modified artificial fish swarm algorithm,"*in proc. 24th int. conf. on Advances in Artificial Intelligence (AI'11)*, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 382-391.

[11] KaurRajinder, GirdharAkshay, Gupta Surbhi, "Color Image Quantization based on Bacteria Foraging Optimization." *Int. J. Computer Application*, vol. 25, no. 7, 2011, pp. 33-42.

[12] RupinderKaur, Vikas Gupta, "Proposed Method for Color Image Quantization: Honey Bee," *Int. J. Comp. Sci. & Comm. Eng. (ijcsce)*,vol.1, no.2, 2012, pp. 19-22.

[13] Mahamed G. Omran , Andries P. Engelbrecht,Ayed Salman, "A color Image Quantization Algorithm Based on Particle Swarm Optimization," *Informatica*, vol. 29, 2005, pp. 261–269.

[14] X. S. Yang, Nature-Inspired metaheuristic Algorithms. Luniver Press, 2008.

[15] X. Yang, Firefly algorithm, stochastic test functions and design optimization. ;In Proceedings of IJBIC. 2010, 78-84.

[16] X. Yang, Firefly Algorithms for Multimodal Optimization. ;In Proceedings of SAGA. 2009, 169-178.

[17] The USC-SIPI image database, University of Southern California. http://sipi.usc.edu/database/.

[18] Jain Ak, Murty MN, Flynn PJ. "Data Clustering: a review,"*ACM ComputSurv*, vol.31, no.3,, 1999, PP 256-323.

[19] Thung, Kim-Han,"A survey of image quality measures,"*in Proc. of Int. Conf. for Technical Postgraduates*,2009,pp.1-4.