IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

Enhancing Color Image Clustering using K-Means Method

Prof. Ziad Alqadi[1], Dr. Ghazi M. Qaryouti[2], Prof. Mohammad Abuzalata[3], Prof. Yousf Eltous[4]

Albalqa Applied University, Faculty of Engineering Technology, Jordan, Amman[1,2,3,4]

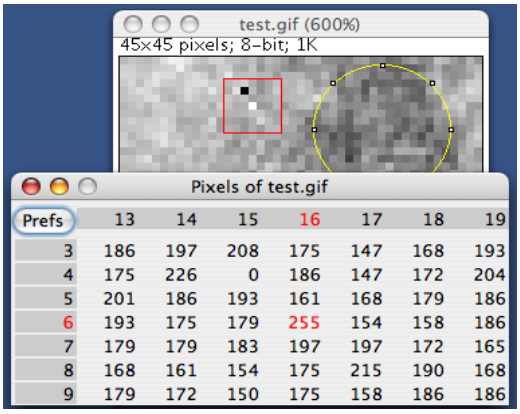
Abstract: Creating a color image features is an important and vital task, because these features can be used in a color image recognition system, so decreasing the effort of clustering, and decreasing the feature extraction time will lead to better and efficient color images recognition system. In this paper research we will investigate how using color image histogram will enhance the recognition system, without doing any changes in the accuracy and uniqueness of the k-means method of features extraction.

Keywords: Color image, histogram, resolution, features, k-means, cluster, centroid, extraction time, and speed up

I. INTRODUCTION

Digital image is one of the most important type of digital data, it is now widely used in many vital application such security [1], [2], [3] systems using cryptography [4] and steganography [5], [6], computer systems identification such as banking systems, robotics etc.

Digital image are usually represented by 2D matrix [7], [8] (gray image), or 3D matrix (color image) [9], [10]. The intersection of row and column is called a pixel and it takes the value between 0 and 255 for a gray image as shown in figure 1.



Prefs	13	14	15	16	17	18	19
3	186	197	208	175	147	168	193
4	175	226	0	186	147	172	204
5	201	186	193	161	168	179	186
6	193	175	179	255	154	158	186
7	179	179	183	197	197	172	165
8	168	161	154	175	215	190	168
9	179	172	150	175	158	186	186

Figure 1: Gray image and pixels

Color image matrix consists of three 2D matrix[11] : the first one represents the red color, the second one represents the green color, while the third one represent the blue color, each pixel has a 3 color value as shown in fig 2 and 3[12], [13].

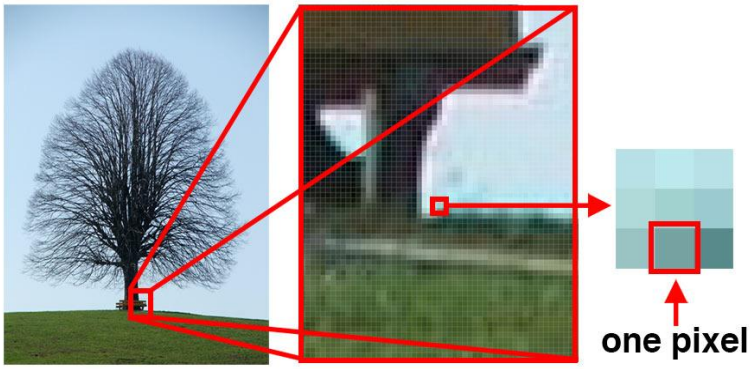


Figure 2: Color pixel

DOI 10.17148/IJARCCE.2020.9115

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

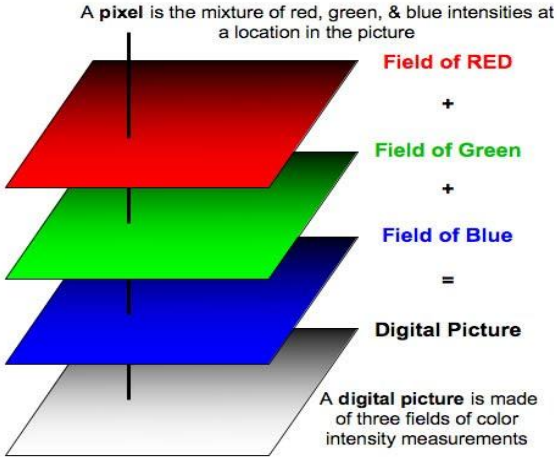


Figure 3: Mixing colors

Color images can be represented a histogram. An image histogram is a graph to show how many pixels are at each scale level, or at each index for the indexed color image [14], [15]. The histogram contains information needed for image equalization, where the image pixels are stretched to give a reasonable contrast.
For the gray image the histogram is a one column matrix with 256 elements, each index in this matrix points to the total repetition of the gray value in the image. For color images, each color channel can be represented by a histogram, and here we can add the tree histograms to find the total histogram, or we can covert the color image to gray one, and after that we can calculate the histogram as shown in figures 4 and 5.

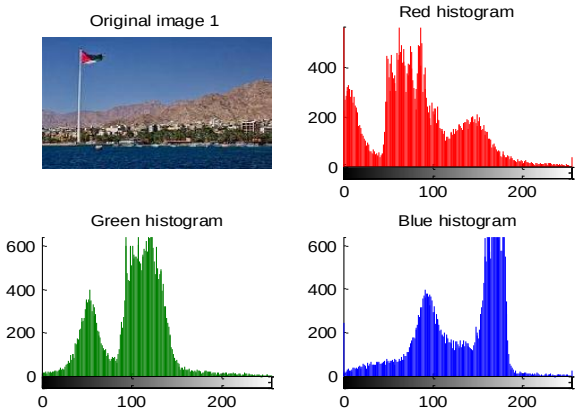


Figure 4: Color image and colors histogram

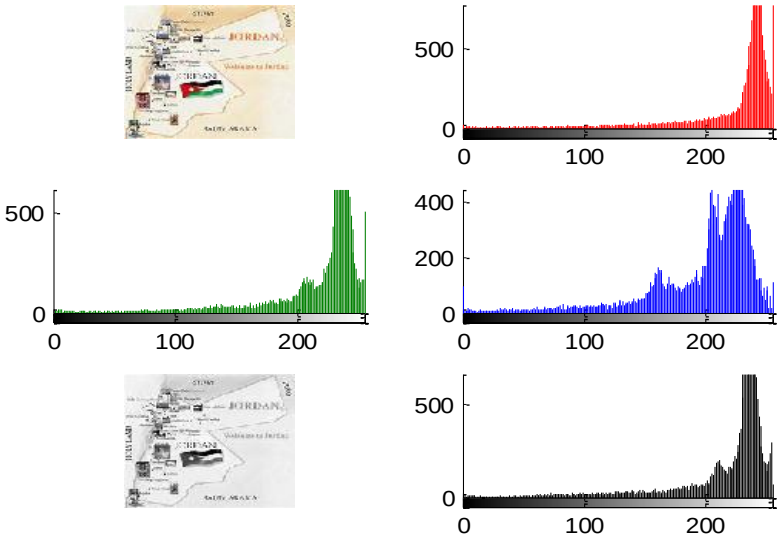


Figure 5: Equivalent gray image histogram

DOI 10.17148/IJARCCE.2020.9115

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

II. IMAGE FEATURES EXTRACTION

Image resolution refers to the number of pixels in an image. Resolution is identified by the number of rows and columns in the image matrix as well as the total number of pixels in the image, for example, an image that is 2048 pixels wide and 1536 pixels high (2048x1536) contains (multiply by 3) 9437184 pixels (or 9.4 Mega pixels), we can call it a 9.4 mega pixel image, so color images always have a huge number of pixels, this leads of adding difficulties in the process of matching the images pixel by pixel.

For the process of image identification we have to seek a method of image features creation, those features can be used later to identify the image.

Many methods were proposed for image features extraction [16-22], but here in this paper we will take the concentration on Kmeans clustering method [23], [24].

The image features extraction method must characterize by the following [25], [26], [27]:

- ✓ Features size must be significantly small (minimizing the number of elements in the features array) [17], [18].
- ✓ The features must be unique for each image [19], [20], [21], this allow us to identify only one image based on the identified features [22].
- ✓ The features extraction time must be very small.
- ✓ The method of features extraction must be simple to implement and flexible to suit any image with any size.
- ✓ The proposed method of features extraction must be adjustable to change the number of elements in the features array.

K-means clustering method takes the input data set of values (points) & group them into clusters as shown in fig 6 & 7:

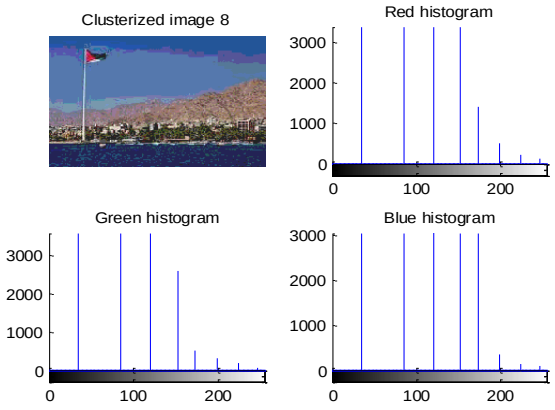


Figure 6: Clustering first image example to 8 clusters

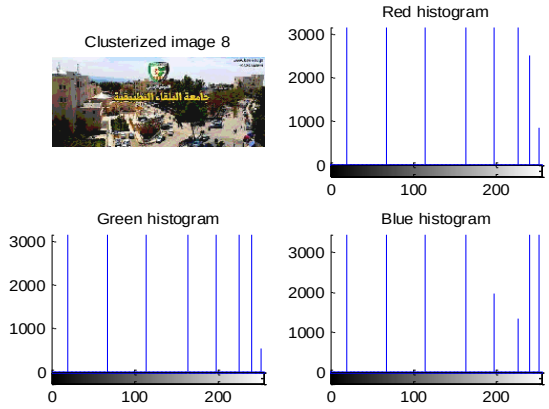


Figure 7: Clustering second image example to 8 clusters

Kmeans method is flexible method of features extraction because of the following:

- ✓ Number of clusters is variable (so the number of features is also variable).
- ✓ We can use the clusters centroids as features.
- ✓ We can use the number of points in each cluster as features.
- ✓ We can use the within clusters sums (WCS) as features.

The efficiency (minimum clustering time) depends on the size of the input data set to be Clusterized, table 1 shows the clustering time for various input data sets with various number of data values:

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

Table 1: Clustering time for various input data sets

Data size(elements)	256	500	1000	2000
Centroids (features)	**0.0448**	**0.0511**	**0.0850**	**0.0860**
	0.1528	**0.1726**	**0.2324**	**0.2637**
	0.2699	**0.2831**	**0.3701**	**0.4489**
	0.3924	**0.3919**	**0.4966**	**0.5706**
	0.5117	**0.5352**	**0.6144**	**0.6940**
	0.6384	**0.6847**	**0.7300**	**0.7891**
	0.7839	**0.8220**	**0.8329**	**0.8860**
	0.9295	**0.9421**	**0.9403**	**0.9682**
Extraction time(s)	0.016000	0.069000	0.076000	0.143000

From table 1 we can notice the following points:

- The extracted features for each data set are unique.
- The extraction time increases when the size of the input data set increases, figure 8 shows how the extraction time depends on the input data set size:

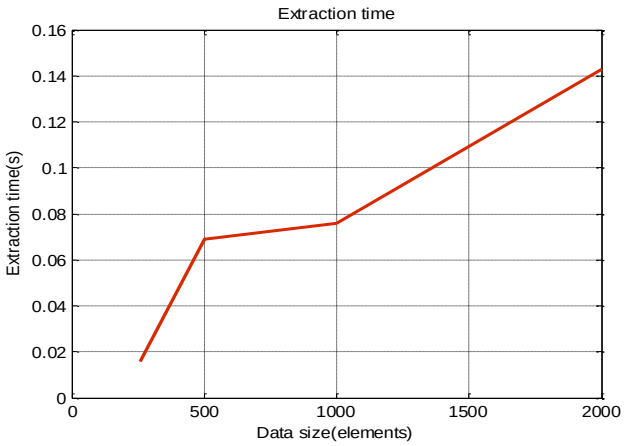


Figure 8: Relationship between extraction time and input data set size

III. IMPLEMENTATION AND EXPERIMENTAL RESULTS

To cluster the color image we have first must calculate the image histogram to be used as an input data set of the k-means method of clustering. Getting the color image histogram can be achieved by performing one of the following procedures:

1. Getting the histogram for each color, and then adding the three histograms to get the total image histogram.
2. Converting the color image to gray one, and then calculate the histogram for the gray image.
3. Reshaping the matrix of the color from 3D matrix to 2D matrix, and then calculate the histogram for the obtained 2D matrix.

Figures 9 and 10 show the histograms calculated by these procedures.

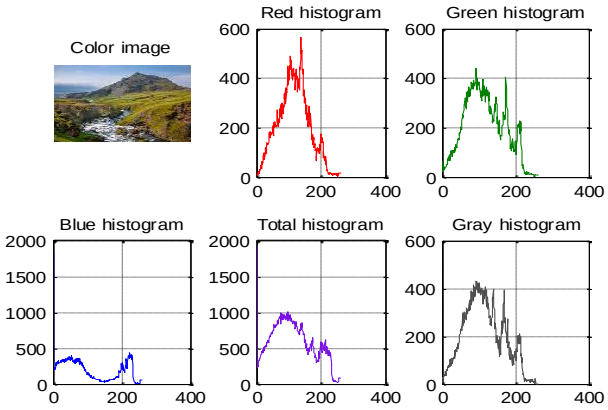


Figure 9: Colors histogram transformation

DOI 10.17148/IJARCCE.2020.9115

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

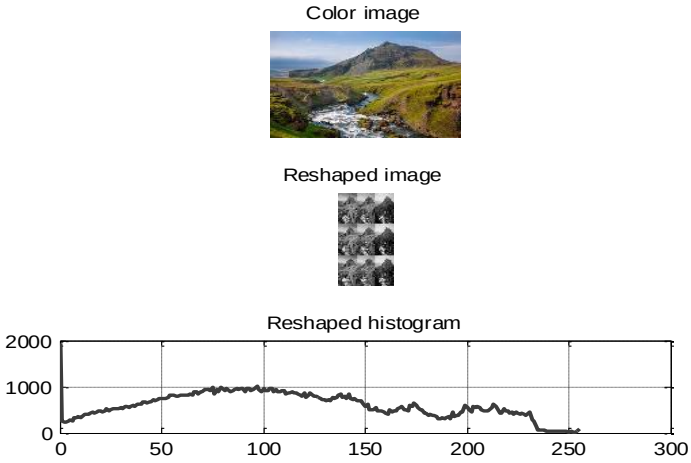


Figure 10: Reshaped histogram

Ten different color images were taken, for each color image we calculate the image histogram, the histogram then was used as an input data set for the clustering method, table 2 shows the results of implementation (8 clusters):

Table 2: Images features using image histogram

Image	1	2	3	4	5	6	7	8	9	10
Centroids (features)	49	10.3	39	653	186	9.8072	76	25.9	9.2432	0.0308
	94.6	30.9	97	2994	1548	31.3902	185	086.1	124.9444	0.1265
	134.7	72.1	321	4449	3412	71.1600	364	137.8	182.4082	0.2079
	192.5	179.3	550	7083	4959	114.5000	861	182.8	260.9231	0.2945
	300.6	341.4	750	10548	6399	203.2778	1245	234	311.9412	0.3891
	472.1	666.3	1130	14048	7433	310.9444	2364	282.5	372.8667	0.8586
	718.3	1051	2047	17126	9130	424.7222	4630	357.1	511.5000	2.1672
	1268	2493	14651	20272	11283	535.8788	35413	5066	854.0000	5.1928
Size (pixel)	**150849**	**77976**	**518400**	**5140800**	**4326210**	**122265**	**518400**	**150975**	**151353**	**6119256**
Extraction Time(s)	**0.1280**	**0.1120**	**0.1030**	**0.1280**	**0.1220**	**0.1130**	**0.1130**	**0.0910**	**0.0950**	**0.1380**

Table 3 shows the results of clustering using a color image as an input data set for k_means method of clustering:

Table 4: Images features using image as an input data set

Image	1	2	3	4	5	6	7	8	9	10
Centroids (features)	8.5521	25.4997	2.5774	9.1410	25.1968	11.9923	1.2240	18.4911	5.0550	38.4423
	31.2485	86.7116	28.9257	38.5558	55.6833	52.1211	31.8192	42.7221	24.3868	69.6645
	58.3165	133.6779	52.9561	63.2538	79.0278	76.5654	59.5256	71.0862	47.4036	95.5554
	89.2505	170.3380	76.5310	86.4431	102.9146	97.4980	85.6027	100.7797	76.0334	109.0075
	122.2610	204.1441	105.2039	113.4587	130.1602	116.9805	110.3198	133.1483	108.9245	124.7367
	163.5389	226.3049	136.3523	146.3849	161.7790	136.8949	139.5039	165.8910	143.0043	143.0428
	201.1614	239.9785	170.8464	182.6851	196.0312	164.9973	173.9882	211.4594	176.5531	164.4444
	236.8823	250.7539	212.9146	227.2600	239.1740	190.9049	241.6251	252.1470	231.0838	232.5371
Size (pixel)	**150849**	**77976**	**518400**	**5140800**	**4326210**	**122265**	**518400**	**150975**	**151353**	**6119256**
Extraction Time(s)	**1.6130**	**2.3120**	**9.5120**	**67.6070**	**47.2980**	**1.1360**	**5.7180**	**2.3050**	**2.3620**	**99.9350**

From tables 2 and 3 we can see the following facts:

- ✓ Both input data sets gave different features, but each of them is a unique for a certain image.
- ✓ Using image histogram increases the efficiency of clustering process by significantly decreasing the features extraction time, giving a big speed up as shown in table 3

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 1, January 2020

Table 3: Results comparisons

Image size	Average features extraction time(S)	Average features extraction time(S) using histogram	Speedup of using histogram
Small image size<500k pixels	3.5654	0.1143	31.1934
Greater than 1M pixels	71.6133	0.1143	626.5381

Kmeans clustering is a flexible method of data grouping, we can change the number of clusters to generate features array with different numbers of array elements, and also we can take the number of points in each cluster as a features or take within sums clusters as a features, table 4 and 5 shows the results of changing the number of clusters and the generated features using the number of points in each cluster.

Table 4: changing the number of clusters.

Image	1	2	3	4	5	6	7	8	9	10
Features Number of clusters=6	**51.4**	**16.8**	**81**	**697**	**298**	**14.5088**	**149**	**26.9**	**20.1429**	**9.2432**
	111.2	**66.5**	**331**	**3398**	**2724**	**64.4444**	**362**	**98.1**	**97.6905**	**136.4747**
	189.2	**210.8**	**575**	**6702**	**4969**	**134.0370**	**869**	**175.2**	**153.9750**	**231.6923**
	305.4	**637.6**	**799**	**10479**	**6796**	**248.3889**	**1306**	**247**	**222.5152**	**323.2759**
	592.7	**1051**	**1361**	**14606**	**8750**	**377.0000**	**3110**	**336.6**	**300.0465**	**511.5000**
	1268	**2493**	**1465.1**	**19429**	**11179**	**526.8421**	**35413**	**50660**	**384.3061**	**854.0000**
Features Number of clusters=4	**104.1765**	**25.5**	**232**	**2339**	**480**	**20.1791**	**229**	**65.4**	**32.1587**	**9.2432**
	196.5524	**230.6**	**686**	**7627**	**4783**	**121.4255**	**1000**	**183.2**	**142.8442**	**148.2149**
	338.1364	**899.9**	**1252**	**13628**	**7096**	**320.0606**	**2741**	**291.7**	**267.1207**	**299.7159**
	828.2000	**2493**	**14651**	**19329**	**10738**	**518.6190**	**35413**	**5066**	**375.9828**	**580.0000**

Table 5: Number of points in each cluster as a features

Image	1	2	3	4	5	6	7	8	9	10
Features (number of points in each cluster)	57	26	49	11	47	83	53	31	26	36
	49	1	33	24	49	18	1	1	27	57
	4	107	4	19	38	18	33	44	47	53
	32	6	36	58	31	18	5	31	30	32
	7	27	28	13	11	33	61	52	31	33
	27	7	83	30	26	25	72	43	19	8
	1	7	1	11	11	20	30	18	38	2
	79	75	22	90	43	41	1	36	38	35

IV. CONCLUSION

K-means clustering method is a suitable and flexible method used for color image features extraction, using color image histogram increases the clustering method of clustering by speeding up the process of color image features extraction.

Based on the obtained experimental results we can raise the following facts:

- Each obtained color image features are unique, and they can be easily used to identify the image.
- The number of elements in the features array can be variable, because it is very simple to change the number of clusters.
- Clustering is a flexible method, we can select the clusters centroids, or the number of points in each cluster, or the within clusters sums as a features.

REFERENCES

[1]. Musbah J. Aqel, Ziad A. Alqadi, Ibraheim M. El Emary, Analysis of Stream Cipher Security Algorithm, Journal of Information and Computing Science, Vol. 2, No. 4, 2007, pp. 288-298.

[2]. Ziad Al-Qadi, Musbah Aqel, Performance analysis of parallel matrix multiplication algorithms used in image processing, World Applied Sciences Journal, v. 6, issue 1, pp. 45-52, 2009.

[3]. Jamil Al-Azzeh, Bilal Zahran, Ziad Alqadi, Belal Ayyoub, Muhammed Mesleh, A Novel Based On Image Blocking Method To Encrypt-Decrypt Color image, International Journal on Informatics Visualization, v. 3, issue 1, pp. 86-93, 2019.

[4]. Khaled M Matrouk, Haitham A Alasha'ary, Abdullah I Al-Hasanat, Ziad A Al-Qadi, Hasan M Al-Shalabi, Investigation and Analysis of ANN Parameters, European Journal of Scientific Research, v. 121, issue 2, pp. 217-225, 2014.

[5]. Ziad A. AlQadi Amjad Y. Hindi, Majed O. Dwairi, A Novel Technique for Data Steganography, Engineering, Technology & Applied Science Research, v. 9, issue 6, pp. 4942-4945, 2019.

[6]. J Al-Azzeh M Abuzalata, Ziad Alqadi, Modified Inverse LSB Method for Highly Secure Message Hiding, International Journal of Computer Science and Mobile Computing, v. 8, issue 2, pp. 93-103, 2019.
[7]. Majed O Al-Dwairi, Ziad A Alqadi, Amjad A Abujazar, Rushdi Abu Zneit, Optimized true-color image processing, World Applied Sciences Journal, v. 8, issue 10, pp. 1175-1182, 2010.
[8]. Jamil Al Azzeh, Hussein Alhatamleh, Ziad A Alqadi, Mohammad Khalil Abuzalata, Creating a Color Map to be used to Convert a Gray Image to Color Image, International Journal of Computer Applications, v. 975, pp. 8887, 2016.
[9]. Jihad Nader, Ziad A. A. Alqadi, Bilal Zahran, Analysis of Color Image Filtering Methods, International Journal of Computer Applications, v. 174, issue 8, pp. 12-17, 2017.
[10]. Dr. Ziad Alqadi, Akram Mustafa, Majed Alduari, Rushdi Abu Zneit, True color image enhancement using morphological operations, International review on computer and software, v. 4, issue 5, pp. 557-562, 2009.
[11]. Bilal Zahran Belal Ayyoub, Jihad Nader, Ziad Al-Qadi, Suggested Method to Create Color Image Features Victor, Journal of Engineering and Applied Sciences, v. 14, issue 1, pp. 2203-2207.
[12]. Ziad Alqadi, A Modified Lbp Method to Extract Features from Color Images, Journal of Theoretical and Applied Information Technology, v. 96, issue 10, pp. 3014-3024.
[13]. Ziad AlQadi, Hussein M Elsayyed, Window Averaging Method to Create a Feature Victor for RGB Color Image, International Journal of Computer Science and Mobile Computing, v. 6, issue 2, pp. 60-66, 2017.
[14]. Qazem Jaber Ziad Alqadi, Jamil azza, Statistical analysis of methods used to enhance color image histogram, XX International scientific and technical conference, 2017.
[15]. Ziad Alqadi, Bilal Zahran, Qazem Jaber, Belal Ayyoub, Jamil Al-Azzeh, Ahmad Sharadqh, Proposed Implementation Method to Improve LSB Efficiency, International Journal of Computer Science and Mobile Computing, v. 8, issue 3, pp. 306-319, 2019.
[16]. Ziad Alqadi, Bilal Zahran, Qazem Jaber, Belal Ayyoub, Jamil Al-Azzeh, Enhancing the Capacity of LSB Method by Introducing LSB2Z Method, International Journal of Computer Science and Mobile Computing, v.8, issue 3, pp. 18, 2019.
[17]. Dr Rushdi S Abu Zneit, Dr Ziad AlQadi, Dr Mohammad Abu Zalata, A Methodology to Create a Fingerprint for RGB Color Image, IJCSMC, v. 6, issue 1, pp. 205-212, 2017.
[18]. Ziad Alqadi, Bilal Zahran, Jihad Nader, Estimation and Tuning of FIR Lowpass Digital Filter Parameters, International Journal of Advanced Research in Computer Science and Software Engineering, v. 7, issue 2, pp. 18-23, 2017.
[19]. Musbah J Aqel, Ziad ALQadi, Ammar Ahmed Abdullah, RGB Color Image Encryption-Decryption Using Image Segmentation and Matrix Multiplication, International Journal of Engineering and Technology, v. 7, issue 13.3, pp. 104-107, 2018.
[20]. Ziad AlQadi, Hussein M Elsayyed, Window Averaging Method to Create a Feature Victor for RGB Color Image, International Journal of Computer Science and Mobile Computing, v. 6, issue 2, pp. 60-66, 2017.
[21]. R Abu Zneit, Ziad AlQadi, M Abu Zalata, A Methodology to Create a Fingerprint for RGB Color Image, International Journal of Computer Science and Mobile Computing, v. 16, issue 1, pp. 205-212, 2017.
[22]. Mohammed Ashraf Al Zudool, Saleh Khawatreh, Ziad A. Alqadi, Efficient Methods used to Extract Color Image Features, IJCSMC, v. 6, issue 12, pp. 7-14, 2017.
[23]. Ahmad Sharadqh Naseem Asad, Ismail Shayeb, Qazem Jaber, Belal Ayyoub, Ziad Alqadi, Creating a Stable and Fixed Features Array for Digital Color Image, IJCSMC, v. 8, issue 8, pp. 50-56, 2019.
[24]. Ahmad Sharadqh Jamil Al-Azzeh , Rashad Rasras , Ziad Alqadi , Belal Ayyoub, Adaptation of matlab K-means clustering function to create Color Image Features, International Journal of Research in Advanced Engineering and Technology, v.5, iss.2, pp. 10-18, 2019.
[25]. Yousf Eltous Ziad A. AlQadi, Ghazi M. Qaryouti, Mohammad Abuzalata, Analysis of Digital Signal Features Extraction Based on Kmeans Clustering, International Journal of Engineering Technology Research & Management, v. 4, issue 1, pp. 66-75, 2020.
[26]. Prof. Yousif Eltous, Dr. Ghazi M. Qaryouti, Prof. Mohammad Abuzalata, Prof. Ziad Alqadi, Evaluation of Fuzzy and C_mean Clustering Methods used to Generate Voiceprint, IJCSMC, v. 9, issue 1, pp. 75 -83, 2020.
[27]. Jihad Nader Ismail Shayeb, Ziad Alqadi, Analysis of digital voice features extraction methods, International Journal of Educational Research and Development, v. 1, issue 4, pp. 49-55, 2019