International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 8 Issue 7, July -2019, Impact Factor: 4.059

Color Image Compression Based on Fuzzy Clustering

Dr. Shahbaa Ibrahim Khaleel

Software Engineering Dept., College of Computer Science and Mathematics, Mosul University, Iraq

ABSTRACT

This research presents four ways to compress color images using traditional clustering technique and fuzzy clustering. These methods reduce the storage space required for file storage and facilitate data transfer by reducing transport costs. In order to obtain the best image quality and to improve the performance of the compress system, the first method TKMR was established by applying the conventional cluster algorithm, kmeans, to the color image data, and to obtain a cluster image and then using the lossless compressed method represented by run length encoding on the data generated from the cluster to obtain compress data. The second method, FCMR, was used a fuzzy clustering as fuzzy cmeans method, for the cluster process and then to compress the data. In the third method FGKR, the Gustafson Kessel algorithm was used for the clustering of color image data and then the RLE algorithm was implemented to obtain efficient data. The fourth Method PCMR created the process of clustering using the Possibilistic c-means algorithm, Which was adopted in its work on the principle of fuzzy cmeans algorithm. This method was the best of all methods used in the research, although the four methods were efficient in applying them to color images.

Keyword: color images compression, kmeans clustering, fuzzy clustering, run length algorithm.

HOW TO CITE THIS ARTICLE

Dr. Shahbaa Ibrahim Khaleel, "Color Image Compression Based on Fuzzy Clustering", International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 7, July-2019.

1- INTRODUCTION

The importance of image compression methods has increased in image processing. This was a result of the rapid development and progress of computer capability, the corresponding development in the multimedia market as well as the emergence of WWW, which facilitated Internet access for everyone In addition, the advancement of video technology, including television, has made new and better algorithms needed for image compression. Storing and transferring such data requires a large size and a large package width, which may be very expensive. Image compression techniques are related to reducing the increase in data representation to reduce data storage requirements and thus reduce the cost of communications. The development of efficient compression technologies will continue to be a challenge for future communications systems and advanced multimedia applications.[1]

In this research, the traditional cluster, k-means were used, and fuzzy clustering methods used here, which represented by a fuzzy c-mean algorithm, gustafson kessel algorithm and Possibilistic c-means to clustering the data of images and then used compression algorithm to compress these data, here it is run length encoding.

2-TECHNIQUES OF COMPRESSION

Digital image representation requires a large number of binary bits in many applications or programs. It is important to provide techniques to represent an image or information contained in an image using as little space as possible when

adopting a copy of an image of an acceptable quality, It is referred to as "source encoding" or "image encoding" or "image compression". [2].

The advantages of images compression techniques are that the compressed files are small in size when compared to the original files. They occupy less space in memory, reduce the time required to transfer the original image by transferring them after compression, and some algorithms help to encrypt the sent files which leads to increased confidentiality of information transmitted that are used only by the stakeholders, as well as reduce the cost when it comes to the transfer of a few data across the lines of communication, and sometimes the compressed files especially the image of them are of higher clarity, which helps in such cases to improve the image.[3]

The compression can be accomplished in many ways and methods, depending on the quality of the data to be pressed and according to its composition. The resulting image is affected by the method used to compress it. On this basis there are two types of compressing methods that are classified according to the method of saving data: Lossless and loosy compression. In lossless compression, it is possible to return the compressed data formats identical to the original data. The reconstructed image is exactly the same as the original image. The space needed by the image file is minimized and without any loss of information, the algorithms of this type are widely used with text files that do not allow the loss of any information during the compression process. It is also used with medical image applications and in archival documents. One of the methods used in this type is the Huffman coding method, the mathematical coding method and the Run Length Encoding RLE method. [2, 4]

2.1-RUN LENGTH ENCODING

A method to compress the image works by calculating the number of adjacent pixels that have the same gray level value. This account is then encrypted and stored, called run length, RLE is a simple way to encrypt the source when there is a long string of the same data, in a sequential way in a set of data. For example, data S contains strings: S = 6 6 6 6 6 6 6 2 2 2 2 2 2 2 2 2 2 2 2 2 4 4 4 4 4 4 4 491 1 1 1 1 1, Where S contains long strings of the 6's, 2's, 4's,1's ... etc. Instead of encrypting each model in a series separately, all the data can be represented together and by simply pointing to the value of the model and the length of its string when it appears. In this way data S can be encrypted as follows: (6 7) (2 12) (4 8) (9 1) (1 6).To facilitate understanding of this method, we show a pair of numbers in parentheses where the first value represents the intensity of the light for the pixel, and the second indicates the length of its chain. In some cases the appearance of symbol strings may not be obvious, but it may be possible to process the data to help encrypt string length [3,5,6].

In this research, a classical clustering algorithm, a fuzzy clustering algorithms, were used. Then, after obtaining the cluster centers, a clustered image is formed and encoded by the string-length cryptography REL algorithm in one dimensions.

3-CLUSTERIN METHODS

Clustering operation is a method of aggregating similar data into different groupsor clusters and each group must be similar and the different aggregates must be non-containerized on shared data. This technique is classified as an unsupervised method of data classification. The choice of homogeneity or similarity criteria plays an important role in the accuracy of the result of division.[7]

Gray-scale image segmentation are represented by applying the clustering algorithm to this type of image. The result is that the gray level is split into a number of groups or clusters, and the level within the cluster is similar to the gray level in another cluster. Clustering is a useful method in analyzing many patterns, aggregation, decision making, data processing, data mining, document retrieval, image segmentation, classification of patterns, K-means, C-means, split and merge, all these methods used to clustering images. In addition to these classical methods of fragmentation, the logical fragmentation methods appear very effective for coding.[8]

3.1-KMEANS CLUSTERING METHOD

Traditional kmeans algorithm is the most frequently used algorithm for its simplicity and is an unsupervised classification method, where a set of values is given and tries to divide them into K clusters. A redundancy algorithm is used to reduce the sum of square spaces of all points in the cluster space. The behavior of the kmeans algorithm is influenced by the number of assigned cluster centers, the choice of primary cluster centers, the order in which the samples are taken, as well as the geometric characteristics of the data. Although there is no general evidence of a convergence of this algorithm, it can be expected that there will be acceptable results when data are presented which are relatively distant from one another. In most applied cases, the application of this algorithm will require experimenting with different values for K, ie the number of clusters in addition to different options from the start settings.[8]

3.2-FUZZY CMEANS METHOD

The basic principles of the fuzzy clustering are that each element has a "point of light" that is a partial membership degree in the fuzzy clusters. The grade of membership is selected within the period [0,1] for each sample in each cluster, and the selection is based on specific measurements. In the fuzzy cluster, the membership of the model is determined by membership degree to each cluster. Here we will present a very common fuzzy cluster, "most common and used" Fuzzy algorithm known as the fuzzy c-means algorithm.[9]

The clusters centers V (t) are calculated based on the values of the membership function U (t) according to the following equation: $v_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} x_k}{\sum_{k=1}^{n} u_{ik}^{m}}$, i = 1,2,3,.... .., c

The values of the membership function are updated to obtain U (t + 1) based on the value of V (t) according to the following equation: $u_{ik} = \frac{\left[\frac{1}{(x_k - v_i)^2}\right]^{\frac{1}{m-1}}}{\sum_{j=1}^{c}\left[\frac{1}{(x_k - v_j)^2}\right]^{\frac{1}{m-1}}}$, i = 1,2,..... ....., c ; k = 1,2,..... ....., n

The fuzzy c-means algorithm is applied to the image and the data points or sample points x_1, x_2, ..., x_n represent the grayscale values of the pixels. Thus, n represents the total number of pixels in the image.[10]

3.3-FUZZY GUSTAFSON KESSEL METHOD

Each cluster in the data sets contained in the Gustafson Kessel algorithm is linked to each point and the matrix denotes cluster center and covariance of it. While the original cmeans fuzzy algorithm makes the undeclared hypothesis that the clusters are spherical, the Gustafson Kessel algorithm not limited to the restriction and can identify oval groups.[11]
The membership function is calculated as the following equation:[12]

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{D_{ij}}{D_{kj}}\right)^{\frac{2}{q-1}}}$$

The cluster center v_i of this algorithm is updated as: $v_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^q S_j}{\sum_{j=1}^{n} (\mu_{ij})^q}$

3.4-POSSIBILISTIC CMEANS METHOD

To solve the problems and constraints found in the fuzzy cmeans algorithm, a new algorithm based on the principle of the fuzzy cmeans algorithm was developed and named as Possibilistic c-means. In this method the fuzzy center v_i is computed by using the following equation:[13] $v_i = \frac{\sum_{j=1}^{N} (\mu_{ij})^m x_j}{\sum_{j=1}^{N} (\mu_{ij})^m}$

and the fuzzy membership μ_{ij} is calculated by using this equation: $$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^{2}}{\eta_i} \right)^{\frac{1}{m-1}}}$$

4-EFFICIENCY MEASURE OF COMPRESSION

For color images, their quality after processing is calculated by calculating the value of the retrieved color image of the three colors separately, ie, a value is found, according to the following equation:[6]

$$RMSE = \sqrt{\frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1}^{N} \left[f(x,y) - \hat{f}(x,y) \right]^2}$$

And then find the value $RMSE$ of the full color image ie $RMSE_T$ according to the following formula:

$$RMSE_T = \sqrt{\frac{1}{3}(RMSE_R^2 + RMSE_G^2 + RMSE_B^2)}$$

The Peak Signal to Noise Ratio (PSNR) value for the three color components is also calculated as follows:

$PSNR_{(color)} = 10 \log_{10} \left(\frac{255^2}{MSE_{(color)}} \right) dB$ Where: $MSE_{(color)} = \frac{1}{N}(x_{jk} - y_{jk})^2$

The x_{jk} it represents the color value of the original optical point and y_{jk} represents the color value of the optical point after the compression and decompression process. The final value $PSNR$ is calculated by taking the mean of the three values $PSNR$ calculated for red, green and blue according to the following equation:

$$PSNR = \frac{PSNR_{(red)} + PSNR_{(green)} + PSNR_{(blue)}}{3}$$

To make sure the compression process is performed, the compression ratio is calculated by dividing the size of the uncompressed file “the original file size” on the size of the compressed file. As well as the percentage or number of bits per optical point, which in turn represents the number of bits required to represent each optical point in the compressed image as shown in the following equations:

$$compressio\ nratio\ (CR) = \frac{Uncompress\ edFileSize}{Compressed\ FileSize}$$

$$BitsPerPix\ el = \frac{NumberofBi\ ts}{NumberofPi\ xels} = \frac{(8)(\ Numberofby\ tes\ )}{N * N}$$

5-EXPERIMENTAL AND RESULTS

In this research, the fuzzy compression system was programmed using Matlab 2018. This language is one of the modern languages in that it is used in the technical fields because of its characteristics and advantages and provides the user with ease of handling and processing data accurately and includes a large number of processes and functions which dealing with mathematical and logical operations, and dealing with images and sound and processing of it. Here the color images were processed using four clustering algorithms where the color of the image is divided into a number of clusters. These methods are: the traditional method k-means, the fuzzy c-means and gustafson kessel methods, and the possibilistic cmeans, the image will then be obtained is clustered image, then compressed it with run length encoding to obtain an compressed image in effectively form and with fewer bits.

The research programs were implemented and runs on various images with different colors and dimensions, and obtained an efficient results for the methods of compression based on the cluster mentioned in the search to reduce the size of the compress file and increased the compression ratio in images.

At first, a colored car image was used with a dimaintion of 275 * 183 pixels. The image data was clustered using the traditional kmeans algorithm. It is easy and efficient but contains many calculations. The image has been obtained is clustered image, and then used lossless compression method run length encoding to obtained compress image, this first

method that same TKMR. But the image after decoding is not very clear, so the clustering methods are used by clustered the color image data using a fuzzy cmeans algorithm that allows for easy, flexible division based on fuzzy group theory. The performance in FCMR was better than the conventional or traditional kmeans algorithm, and after the fuzzy clustering process, the cluster data was compressed with run length encoding lossless method. The image was produced after the decoding was better than the previous traditional method. The gustafson kessel method and the possibilistic cmeans method were used to collate the image data. Thus, a clusterd image was obtained and then compressed with a run length encoding method "FGKR, PCMR", which was better than the previous methods, as shown in Table 1 and Figure 1 of the car image used in the research.

Table 1: Explains the results of the methods used in the research for Car image with resolution 275 * 183 and 150975 number of pixels

Methods	*BPP*	*Points Number*	*Comp. File Size*	*CR*	*PSNR*
TKMR	1.573770	29700	69.6 KB	2.13:1	18.874982 dB
FCMR	0.892651	16846	34.8 KB	4.25 : 1	30.066614 dB
FGKR	0.460050	8682	18.56 KB	7.97 : 1	54.115699 dB
PCMR	0.014837	280	2.694 KB	54.94 : 1	63.440441 dB


Original Image


Reconstruct Image TKMR


Reconstruct Image FCMR


Reconstruct Image FGKR


Reconstruct Image PCMR

In addition, all methods of research were applied to flowers image of dimensions 215 * 235 and cake image size 225 * 225 and standard images with dimensions 256 * 256, lena image and peppers image. The results are presented in the following tables and figures:

Table 2: Explains the results of the methods used in the research for Flowers image with resolution 215 * 235 and 151575 number of pixels

Methods	*BPP*	*Points Number*	*Comp. File Size*	*CR*	*PSNR*
TKMR	1.914089	36266	85.3 KB	1.74 : 1	20.115464 dB
FCMR	1.417226	26852	55 KB	2.69 : 1	30.066806 dB
FGKR	0.848478	16076	33.84 KB	4.37 : 1	46.492861 dB
PCMR	0.162665	3082	7.03 KB	21.05 : 1	64.069178


Original Image


Reconstruct Image TKMR


Reconstruct Image FCMR


Reconstruct Image FGKR


Reconstruct Image PCMR

Table 3: Explains the results of the methods used in the research for Cake image with resolution 225 * 225 and 151875 number of pixels

Methods	*BPP*	*Points Number*	*Comp. File Size*	*CR*	*PSNR*
TKMR	1.707931	32424	79.2 KB	1.87 : 1	18.964664 dB
FCMR	1.313080	24928	51.6 KB	2.87 : 1	30.066776 dB
FGKR	1.052234	19976	41.7 KB	3.55 : 1	46.990003 dB
PCMR	0.345863	6566	14.35	10.31 : 1	65.370793 dB


Original Image


Reconstruct Image TKMR


Reconstruct Image FCMR


Reconstruct Image FGKR


Reconstruct Image PCMR

Table 4: Explains the results of the methods used in the research for lena image with resolution 256 * 256 and 196608 number of pixels

Methods	*BPP*	*Points Number*	*Comp. File Size*	*CR*	*PSNR*
TKMR	1.569010	38560	90.3 KB	2.13 : 1	19.522414 dB
FCMR	0.922119	22662	47.6 KB	4.03 : 1	30.066446 dB
FGKR	0.395345	9716	21.56 KB	8.91 : 1	54.862105 dB
PCMR	0.020915	514	3.291 KB	58.34 : 1	63.384674 dB


Original Image


Reconstruct Image TKMR


Reconstruct Image FCMR

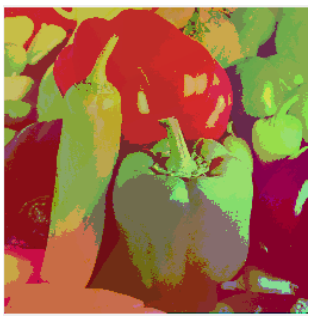

Reconstruct Image FGKR


Reconstruct Image PCMR

Table 5: Explains the results of the methods used in the research for Peppers image with resolution 256 * 256 and 196608 number of pixels

Methods	*BPP*	*Points Number*	*Comp. File Size*	*CR*	*PSNR*
TKMR	1.237956	30424	71.9 KB	2.67 : 1	17.249454 dB
FCMR	0.787598	19356	41.4 KB	4.64 : 1	30.063231 dB
FGKR	0.315186	7746	17.37 KB	11.05 : 1	54.788869 dB
PCMR	0.020304	499	3.259 KB	58.9 : 1	64.872267 dB


Original Image


Reconstruct Image TKMR


Reconstruct Image FCMR


Reconstruct Image FGKR


Reconstruct Image PCMR

For all the previous results, efficient results were obtained for the compression methods based on the fuzzy clustering to reduce the size of the compressed files and increase the compression ratio of the colored images. The best method is the run length encoding that based on clustering using the PCMR method, This method obtained highest compression ratio, because the size of the composite file had the lowest size of all methods, as well as the value of PSNR was the best value in this method for all images and the proportion of bits per optical point "pixel" had this method has the lowest value means that the PCMR method is the best method among all the methods used to search.

CONCLUSIONS

After applying the traditional clustering methods and fuzzy clustering methods on a set of color images in different dimensions and then applying the RLE method to obtain images efficiently compressed, the following were observed:For the traditional clustering method using the k-means algorithm, results were not good in terms of evaluation of reconstructed images, ie, after decompression by looking or by calculating PSNR values and other measures calculated in this research that explained in tables.As for the fuzzy clustering method Represented by the fuzzy c-mean algorithm it gave good results with the compression process, which did not loss of color for the colored images, but was not the most efficient method.Since the size of the compressed file did not decrese as required, so other methods were used to conduct the clustering.The FGK method was used to perform the clustering process and then to compress the clustering data by RLE. This method gave the best results compared to the two methods above. It did not lead to loss of colors or distortion of the image milestones. On the contrary, it gave brilliant images but in some cases the bright images are undesirable, In order to improve the performance, PCM was used in clustering. Finally, the PCM clustering method was used after obtaining the clustering data, RLE was applied to this clustering data. This method gave the best results and did not lead to loss of colors or distortion of image milestones or loss of details or entered noise on the image. It is from the problems associated with some images compression operations, Also, the reconstructed images were not more glamorous than their original nature, as was the case with the FGKR method. Where the best results were given by all the measures used, in addition to their evaluation by way of consideration. The images of this method were the best reconstructed images.

REFERENCES

[1] Wei Y. W., (2009), An Introduction to Image Compression, Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, ROC.
[2] Shivashetty V. and Rajput G.,(2015), ADAPTIVE LIFTING BASED IMAGE COMPRESSION SCHEME USING INTERACTIVE ARTIFICIAL BEE COLONY ALGORITHM , © CS & IT-CSCP 2015, pp. 09–21.
[3] Prasetyo K. E. et al.,(2017),Comparison of Text Data Compression Using Run Length Encoding, Arithmetic Encoding, Punctured Elias Code and Goldbach Code, / International Journal of Engineering and Technology (IJET), Vol 9 No 5 Oct-Nov 2017.
[4] Singh M., Kumar S. Et al., (2016), Various image compression techniques: Lossy and lossless, International Journal of Computer Applications (0975-8887), Vol. 142 , No. 6.
[5] Annangi S., (2017), RTL DESIGN OF EFFICIENT MODIFIED RUNLENGTH ENCODING ARCHITECTURES USING VERILOG HDL, International Journal of Electronics and Communication Engineering and Technology (IJECET), Volume 8, Issue 1, January - February 2017, pp. 52–57, Article ID: IJECET_08_01_006.
[6] Kaur K. et al., (2017), Image Compression Using Run Length Encoding (RLE), International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 5.
[7] Allahyari M., Pouriyeh S. Et al. , (2017),A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques, KDD Bigdas, August 2017, Halifax, Canada.
[8] Chitra K., Maheswari D. , (), A Comparative Study of Various Clustering Algorithms in Data Mining, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.8, August- 2017, pg. 109-115.
[9] GOKTEN P. O. and BASER F. et al. (2017), Using fuzzy c-means clustering algorithm in financial health scoring, Audit Financiar, vol. XV, nr. 3(147)/2017, pp. 385-394.

[10] Zang W. , Zhang W. et al., (2017), A Kernel-Based Intuitionistic Fuzzy C-Means Clustering Using a DNA Genetic Algorithm for Magnetic Resonance Image Segmentation, Entropy 2017, 19, 578.
[11] Krishna P. C. and Venkateswari S.,(2017), Application of Gustafson-Kessel-like clustering algorithm in Delineation of management Zones in precision Agriculture, International Journal of Applied Agricultural Research ISSN 0973-2683 Volume 12, Number 3 (2017) pp. 279-293.
[12] Serir L., Ramasso E., Zerhouni N., (2012), Evidential Evolving Gustafson-Kessel Algorithm For Online Data Streams Partitioning Using Belief Function Theory, International Journal of Approximate Reasoning., 2012, 53 (5), pp.747-768.
[13] Gosaina A. and Dahiyab S. , (2016), Performance Analysis of Various Fuzzy Clustering Algorithms: A Review, Published by Elsevier: Procedia Computer Science79(2016)100–111.