2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) | 979-8-3503-8354-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC2PCT60090.2024.10486602

Lossless Image Compression using K-Means Clustering in Color Pixel Domain

Rimjhim Kumari
Department of Computer Science and Engineering
Galgotias University, Greater Noida, India
krrimjhim1445@gmail.com

Srinivasan Sriramulu
School of computing Science and Engineering
Galgotias University Greater Noida, India
ramuluvasan@gmail.com

Abstract— Compressing images is a method for shrinking the image's dimensions using a particular algorithm. Image compression is a solution associated with transmitting and storing large amounts of data for digital images. While image storage is necessary for medical images, satellite images, documents, and pictures, image transmission includes a variety of applications like television broadcasting, remote sensing via satellite, and other long-distance communication. These kinds of applications deal with image compression. Lossy compression and lossless compression are two separate methods for compressing images. While unneeded metadata is deleted in lossless compression to minimize file size, lossy compression permanently removes part of the original data to reduce file size. The terms "supervised learning" and "unsupervised learning" refer to two different machine learning techniques. As opposed to a non-supervised learning algorithm, which uses unlabeled input data, supervised learning uses labeled data. While clustering is a subtype of unsupervised learning, classification, and regression are additional subcategories of supervised learning. This paper discusses the application of clustering using K-means. This methodology explores the concept of lossless picture compression is the process of reducing the size of an image without sacrificing its quality which is achieved using k-mean clustering. The proposed methodology achieves higher compression ratios than other existing approaches to image compression algorithms.

Keywords—image, compression, k-mean, clustering, unsupervised, lossless

I. Introduction

Image compression is a method to lessen the number of bytes in a graphics without compromising the standard of images. By lowering the file size, more photographs may be saved in a given amount of RAM or disc space. The image needs less bandwidth when it is downloaded from a website or sent over the internet, which reduces network congestion and speeds up content delivery.

A method for reducing the size of an image is image compression file using a particular algorithm. Image compression is a solution associated with transmitting and storing large amounts of data for digital images. While image storage is necessary for medical images, satellite images, documents, and pictures, image transmission includes a variety of applications like television broadcasting, remote sensing via satellite, and other long-distance communication. These kinds of applications deal with image compression. Lossy compression and lossless compression are two separate methods for compressing images. While unneeded metadata is deleted in lossless compression to minimize file size, lossy compression permanently removes part of the original data to reduce file size. The terms "supervised learning" and "unsupervised learning" refer to two different machine learning techniques. As opposed to a non-supervised learning algorithm, which uses unlabeled input data, supervised learning uses labeled data. While clustering is a subtype of unsupervised learning, classification, and regression are additional subcategories of supervised learning.

The two most common forms of picture file compression are lossy and lossless methods. Lossyless compression shrinks an image file's size by permanently deleting unnecessary and unimportant data. Lossy compression may significantly reduce file size, but if used excessively, it can distort photos and cause a considerable loss in image quality. However, quality may be kept when compression is done properly.

Our paper aims to explain a method for reducing the information needed to represent an image, known as image compression. This technique is crucial in various fields, including digital cameras, smartphones, and low-resource devices, where it can save storage space. It is especially important in the medical field, where large amounts of image data, such as X-rays, MRIs, and CT scans, accumulate. Efficient storage and fast transmission over the network are essential for immediate diagnoses by experts. The primary goal of the suggested paper is:

1) To generate a lossless compressed image
2) To decrease the number of colors by averaging the colors closest to the original image.
3) To implement K-means clustering for achieving compressed image
4) To achieves higher compression ratios than other existing approaches of image compression algorithms.

This paper proposes an image compression method that uses K-means clustering to generate a lossless compressed image that achieves higher compression ratios than other existing approaches of image compression algorithms. The proposed document is organized as, in the beginning, this paper accurately describes initially what assumptions were made from the images in the order in which the method achieves the proposed results. Then this paper describes precisely how the unsupervised learning technique is used using k-means algorithm. In the end, this paper explains how the proposed compression method works as well as how k-mean is used to compress images.

II. Literature review

G.S.K.Krishna phani, B.Lavanya and Sanjay Kumar Singh proposed an idea of image comprecion using RLS-DLA ALGORITHM in their paper “STUDY AND ANALYSIS OFRLS-DLA ALGORITHMWITH NATURAL IMAGE COMPRESSION TECHNIQUES”[1]. The "Algorithm for Learning a Dictionary Using Recursive Least Squares " is the major focus of this paper. Learning over-complete dictionaries for sparse signal representation using RLS-DLA. While each training vector is processed, recursively updating the dictionary is done while using the training set.. The RLS-DLA technique was constructed using a convergence scheme that took the forgetting factor into account. The algorithm's core is simple and straightforward to use. Also, researches focused on a comparison between RLS-DLA with more traditional natural image compression methods like JPEG and JPEG 2000.

Lastly, the RLS-DLA approach's sparsity coefficients enable low bit rate picture compression. The experimental findings show that the RLS-DLA technique is excellent at compressing natural images at low bit rates.

Rishav Chatterjee et al propose an idea of Lossy Compression in their paper “Image Compression Using VQ for LossyCompression”[2]. Because of its fast rate of compression and straightforward decoding techniques, vector quantization is among the most widely used lossy compression methods. The design of the codebook is the primary VQ approach. In this study, the author used k-means clustering and VQ for lossy compression to compare and determine the compression ratios of jpg and tif photos.

Xin Yuan et al publish a paper “End-to-end evaluation of compressed sensing-based image compression vs JPEG”[3]. On the basis of compressive sensing, they describe a method for complete picture compression. The system that is being shown combines quantization and entropy coding for reconstruction with the traditional method on the complete picture of compressive sampling. It is demonstrated that the compression performance is comparable to JPEG and significantly superior at low rates with regard to decoded image quality and data rate. They research the variables that affect the system's performance, such as the reconstruction methods, ratio of quantization to compression trade-offs, and the selection of the sensing matrix. They offer a useful technique for choosing the compression ratio and the quantization step combinations that produce almost ideal for every given bit rate, quality out of all those that are conceivable.

Venkateswaran Narasimhan et al proposed a paper “Wavelet Domain K- Means Clustering Based Image Compression”[4]. Using the k-means algorithm with discrete wavelet transformation, this work proposes cutting-edge image compreion method. Wavelet coefficient clustering is used for each DWT band, greater compression rates are desired. In contrast to other approaches, this methodology applies DWT to only a portion of the original image. Here, 16x16 subblocks from a picture are decomposed on a single level using wavelets, then the coefficients are categorised by utilising the commonly used K-means clustering method. The centriods of each cluster are enumerated and arranged in a book. Only the index values are sent beyond the pale. It is possible to get high compression ratio by reducing the number of clusters. According to the simulation findings, changing the cluster size from 10 to 100 results in a compression ratio that ranges from 56.8 to 8.0 while keeping the image's quality adequate.

Nishant Kumar Singhai and Prateek Mishra published a paper[5]. The authors of this study suggested a brand-new hybrid approach for image compression. In this study, the authors lessen the impact of two transform methods and create a new one that is more effective than either one alone. Here, researchers compare their new strategy to the earlier one and suggest other approaches. The comparison made use of the peak signal-to-noise ratio, mean square error, and compression ratio. In terms of visual perception, the modern suggested hybrid technique i.e. the combination of WPT and DCT is superior. This article serves as a useful resource for choosing an image compression technique in the future. This work comes highly recommended for transmission and storage purposes.

Thamer Hameed proposed a paper “Image Compression Using Neural Networks: A Review”[6] . Technology for imaging and video coding has advanced considerably in recent years. Yet, taking into account the popularity of picture and video collecting systems, the expansion of picture data is far higher than that of the growth in compression ratio. It is broadly acknowledged, it will become harder to further increase the efficiency of coding inside the established hybrid coding scheme. The deep CNN that has recently revived the NN and achieved substantial success both in artificial intelligence domains and in signal processing also offers a new and intriguing approach to picture reduction. In this research, the author provides a thorough, up-to-date analysis of picture reduction methods based on neural networks.

“Design of Image Compression Algorithm Using MATLAB”[7] was another paper proposed by Abhishek Thakur et al. This work provides an overview of recent advancements in image security and recent developments in the field. In order to prevent unauthorised users from accessing the picture encryption techniques scramble the pixels of the image and lessen their correlation. This paper suggests a technique for applying a new algorithm called chaotic encryption method to encrypt the sender's messages. The messages that are exchanged between the two sides will be encrypted and decrypted using this key. The new cryptographic method is chaotic encryption. In this study, picture security is highlighted, and a better secure algorithm using chaotic encryption is designed to offer increased security and dependability.

A paper on a comparative study block incorporating wavelet, fractal image compression, embedded zero tree, and coding is presented.[8] was published by IJESRT Journal. The performance differences of several transform coding methods, including embedded zero tree image compression, wavelet, fractal, and block truncating coding, are examined in this research. This work focuses on key aspects of transform coding used for still image compression. The aforementioned methods have been applied successfully in numerous applications. Pictures produced using those methods produce excellent outcomes. The ratio of peak signal vs noise and the ratio of compression measurements are utilize to numerically analyse these techniques (CR). Researchers employ Matlab software's Image Processing Toolbox to carry out their proposed study.

The work "Lossy Image Compression using Discrete CosineTransform" by P. B. Pokle et al. [9]. The main issues with social media applications are the high data rates, high bandwidth, and enormous amounts of memory needed for compute and storage. Due to bandwidth restrictions, there are significant difficulties in sending such massive volumes of data via the network even with faster internet, throughput rates, and upgraded network infrastructure. This supports the requirement for creating compression methods in order to maximise available bandwidth. This article demonstrates how to use the discrete cosine transform to compress digital images and compares it to other techniques.

Surbhi Singh and Vipin kumar Gupta proposed a work on "Huffman Coding JPEG Image Compression and Decompression"[10]. The performance of picture compression is being improved by the authors in this research. Authors are enhancing the parameters of Normalized CrossCorrelation, Structural Content, MSE, Average Difference, PSNR, Maximum Difference, and Normalized Absolute Error to increase the performance of image compression. The authors can view from the outcomes discussion that NK and PSNR values are increasing while all other values are decreasing. Authors must lower the values of PSNR, AD, SC, MD, MSE, and NAE,and must raise NK esteems to increase picture compression's effectiveness.

Emy Setyaningsih and Agus Harjoko proposed a paper "Examination of Hybrid Image Compression Methods"[11]. Compression is the process of shrinking or condensing data while preserving the information's quality. This study provides a review of studies that highlight how various hybrid compression strategies have advanced over the past ten years. As with the JPEG compression approach, a hybrid compression strategy combines the best aspects of each group of technologies. In order to achieve an elevated ratio of compression bearing in mind the level of the image's reconstruction, this method mixes lossy and lossless methods of compression. While lossless compression results in high-quality data reconstruction and a comparatively high compression ratio because the data may subsequently be decompressed with the same outcomes as before the compression, lossy compression generates a reasonable ratio of compression. The concept undertaking additional research to enhance how well the picture performs compression is suggested talks about what is known about and problems with evolving hybrid methodology of compression.

Abdelhamid Mammeri, et al proposed a paper "A Survey ofImage Compression Algorithms for Visual Sensor Networks"[12]. The author of this survey study gave a brief summary of the current technology for algorithms for Visual Sensor Networks compression, and identified a modern categorization for present suggested method of compression, as well as a list of benefits, drawbacks, and unresolved investigation questions.

Andrea Vitali et al proposed an idea of "IMAGE COMPRESSION BY PERCEPTUAL VECTOR QUANTIZATION"[13]. This study describes a vector quantization-based picture compression method. Transforms and entropic coding, which are typically applied before and after quantization, are not employed in this process. The image's local attributes are tracked and exploited by the quantization stage using adaptive computation.

Bibhas Chandra Dhara proposed a paper "Block truncation coding and pattern fitting are used to compress images and videos for quick decoding."[14]. The goal of this thesis is to create an effective decoder for an image compression technique. It is obvious that spatial domain-based compression takes substantially less time to decode than sub-band compression approaches. Both BTC and VQ are frequently used techniques for compressing spatial domain data. The proposed paper suggests a blended compression technique that uses the concepts VQ and BTC to produce reasonable bit-rate per quality trade-off.

Sarang Bansod and Sweta Jain proposed a paper of "The Harmony Search Algorithm Has Been Improved For Colour Image Compression"[15]. The harmony search algorithm is effective for many different situations but is still somewhat new, particularly in the area of picture reduction. This study aims to increase the effectiveness of the method when used to compress different types of photos. Combining it with Bitcoin versions can yield a better outcome (Block Truncation Coding). Implementation outcomes further demonstrate the viability of this newly developed technique. Because the original image requires a lot of disc space and high transmission bandwidth, an efficient method for image compression is continually evolving. Although there are many different image compression methods, the main objective is always to provide the best method that meets the needs of the user and requires less memory, which is accomplished by achieving a strong compression ratio. To increase the compression ratio, certain earlier studies on lossless picture compression were suggested. The proposed methodology achieves higher compression ratios than other existing approaches to image compression algorithms.

III. Related Terminologies

Before going through the proposed paper, one should have a clear understanding of some related terminologies like compression, clustering, lossy and lossless compression, etc.

A. Image processing

Image processing is transferring an analogue image with a digital form and performing particular processes to it in order to enhance the image or extract more useful information from it. It functions like a signal time where an image, such a video frame or photograph, acts as the input and the output might be a picture or attributes related to that image.

Copyright © IEEE–2024 ISBN: 979-8-3503-8354-6

B. Compression

With aid of compression, the amount of storage space or bandwidth needed to display an image can be decreased. Image compression techniques entail shrinking the size of an image and adjusting it so that the quality isn't significantly compromised.

C. Lossy Compression

Lossy compression is a type of data compression that eliminates certain information in order to produce smaller files. After decompression, lossy compression does not reconstruct or restore the original data. Lossy compression is frequently applied to audio, video, and image files that include more data than is necessary.

D. Lossless Compression

A data compression technique called lossless compression allows for a flawless conversion of the compressed picture to the original image with absolutely no information loss. Lossless compression is possible because most real-world data displays statistical redundancy. Comparatively, lossy compression only allows for the approximate reconstruction of the original data, albeit typically having far higher compression rates.

E. Color Image Processing

An image processing application that uses a color-based method to extract information from an image.

F. Machine Learning

As the name indicates, the computer's capacity for learning is what gives it a more human-like character. In more locations than one may imagine, machine learning is currently being actively deployed. Learning by Machine is classified into three categories:

- Supervised Learning: As implied by the name, supervised learning entails a supervisor acting as a teacher. The technique of training a computer system using a set of labeled data is called supervised learning.
- Unsupervised learning: It is used to cluster and evaluate unfamiliar data sets. This approach runs independently, which means that no output is given to the model in order to find hidden patterns and information. By using solely the input parameter values, the training model automatically identifies groups or patterns.
- Reinforcement learning: It involves understanding of how to react in a circumstance to gain to the greatest extent. Data for RL is gathered from machine learning systems that employ a trial-and-error process. Input for either supervised or unsupervised machine learning does not include data.

G. Classification Algorithms

This sort of learning is supervised. Finding a model or function to help divide the data into several discrete values or category categories is a necessary step in the classification process. Data are classified using the input parameters under a variety of labels, and the labels are then forecasted for the data.

H. Regression Algorithms

It is an additional supervised learning method. Regression may be used as a substitute for employing classes or discrete values to find a model or function that divides the data into continuous real values. It also has the ability to detect distribution movement based on past data. Because it predicts a quantity, a regression predictive model's accuracy must be stated as an error in those predictions.

I. Clustering

This sort of learning is unsupervised learning. The machine model discovers that this strategy is generally used to group data based on various patterns, such as similarities or differences. These techniques are used to sort unclassified, unprocessed data objects.

J. Association

It is an additional type of unsupervised learning. This method is a rule-based machine learning strategy that identifies relationships between parameters in a huge data collection. This method mostly applies to market basket analysis, which aids in clarifying the connections between various products. To determine the relationship between the sales of specific item in relation to the sale of another focused on consumer conduct, for instance, shopping stores deploy algorithms based on this technique. For instance, if a consumer purchases milk, he might also purchase bread, eggs, or butter. Once properly taught, these models can be utilized to promote sales by devising various promotions.

IV. Materials and Methods

The subsequent materials - methods are utilised to evaluate the efficiency of lossless k-mean clustering image compression algorithms:

A. Materials

The following lossless compression techniques are taken into consideration for this study.

1) K-Means Clustering

An unsupervised technique, separates the set of unfamiliar data into different groups. In this case, K variable is used as an experimental parameter which indicates the number of pre-defined groups that are needed to be created during the procedure. For instance, if the value of K is two, this mean that only two clusters will exist, if the value of K is 3, there'll just be three clusters, and so on.

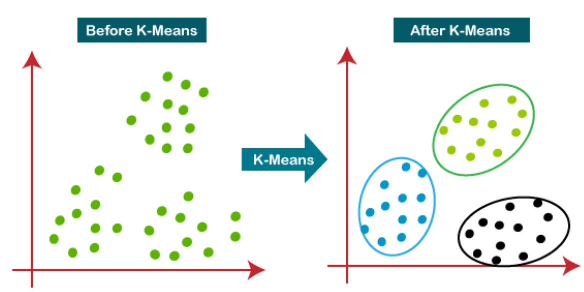

Fig. 1. K-MEAN CLUSTER

2) Compression Time:

It is important to take into account the compression times. This component might have very low values as high-speed computer accessories evolve, and those values might be influenced by how well computers work.

3) Measuring Compression Performances:

There are numerous metrics to gauge a compression algorithm's performance depending on the application. Space efficiency would be the primary consideration while evaluating performance. Another element is time effectiveness. The nature and structure of the input source affect performance. The type of compression algorithm—lossy or lossless—also affects the compression behavior. The space and time efficiencies would be higher than those of the lossless compression strategy if a particular source file was compressed using a lossy compression algorithm. Because it is challenging to measure overall performance, several measurements should be used to assess the performances of such compression families. The metrics used to assess the effectiveness of lossless algorithms are listed below.

The proportion of a compressed file's size to the size of the original file is known as the ratio of compression.

$$\text{Compression Ratio} = \frac{\text{size}_{\text{(after)}}}{\text{size}_{\text{(before)}}} \quad \text{... (1)}$$

The CR's opposite is the CF. It measures the ratio of a file's original size to its compressed size.

$$\text{Compression Factor, CF} = \frac{\text{size}_{\text{(before)}}}{\text{size}_{\text{(after)}}} \quad \text{... (2)}$$

The shrinkage of the source file is calculated using following:

$$\text{Saving \%} = \frac{\text{size}_{\text{(before)}} - \text{size}_{\text{(after)}}}{\text{size}_{\text{(before)}}} \% \quad \text{...(3)}$$

Using file sizes, each of the above-mentioned methods assesses how well compression algorithms work.

B. Methods

A series of image files are used to develop and test the k-mean algorithm, which is utilised to assess the effectiveness in lossless compression technique. The aforementioned factors are computed in order to evaluate performances.

1) Working of K-mean

The following steps show the phases of the K-Means algorithm:
Step 1: To get the total number of clusters, select the value of the K variable.
Step 2: Pick K's centroids or places at random. That data set might not have been utilised as input.
Step 3: Each data point should be allocated to the closest centroid, which will provide the K preset groupings.
Step 4: Determine the variance, then move each cluster's centroid.
Step 5: Re-assign each data point to the new centroid of each cluster by repeating Step 3.
Step 6: If there is a reassignment, proceed to step 4, otherwise, go to COMPLETE.
Step 7: Algorithm is complete.

2) Evaluating the performance of K-mean

Considering the total sum of clusters, there is no definitive answer because K-means is unable to learned from observations; it demands k as an input. Domain expertise and intuition may be helpful occasionally, even if it's not frequently the situation. Since groups are employed in the following-up simulation in the cluster predicted approach, on the basis of various K clusters, evaluation of the models' performance is possible.

In this paper, the elbow approach of indication is to understand K. Based on the sum of squared distance (SSE) between data points and the centroids that make up their assigned clusters, the elbow technique gives researchers an idea of what a suitable k number of clusters might look like. At the point when SSE begins to flatten out and form an elbow, the value of k is found. The geyser dataset will be used to evaluate SSE for different values of k to identify potential elbow and flattening points on the trajectory.

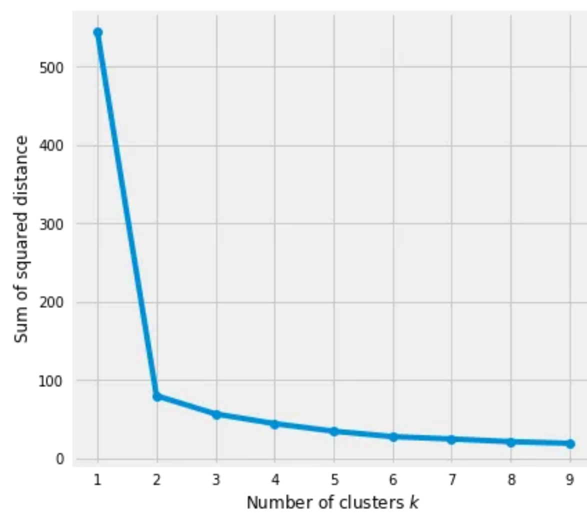

Fig. 2. Elbow Graph

The graph shows that k=2 is an acceptable option. It might be difficult to decide how many clusters to employ when the curve is monotonically dropping since there might not be an obvious elbow or point when the curve starts to flatten out.

V. Proposed Methodology

A. System Mode

K-means is an unsupervised machine-learning algorithm, as was indicated in the introduction. Simply put, the main distinction between supervised with unsupervised techniques that former learn by example (labels are already in the dataset), while the latter learn by trial and error.

There are two features in the figure below: x1 and x2.

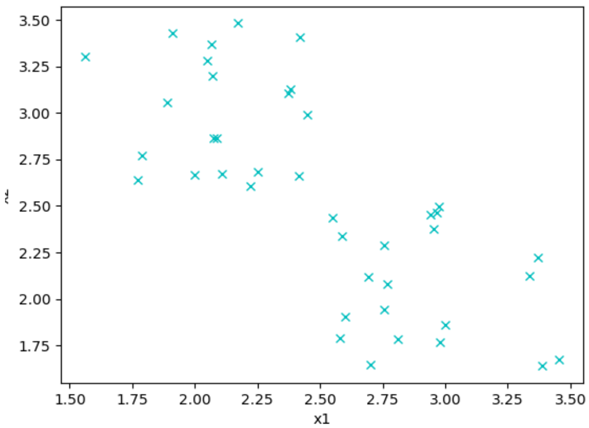

Fig. 3. Data Points

Each object needs to be sorted into one of two clusters. The most natural approach to go about it would be as follows:

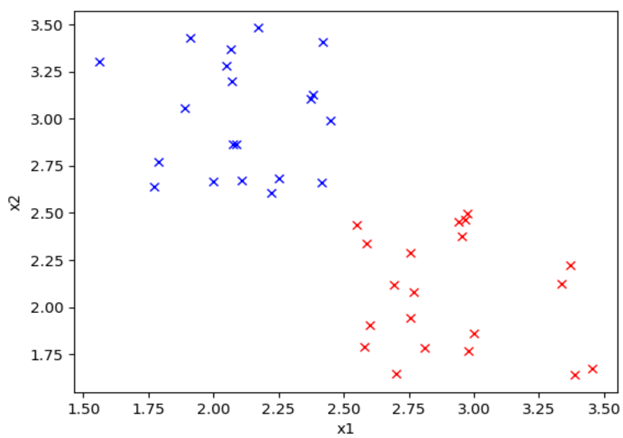

Fig. 4. Sorted Cluster

K-means is an algorithm to achieve exactly this. Recall that K means desired clusters number. Method start by randomly collecting K training instances and in µ so that the points are μ 1, μ 2, ..., μ K and $\mu \in R^n$ where is the number of features. At startup, the points could be quite close to one another, therefore it is needed to check if the outcome looks like how precise it appear because it might stuck be in a local optimum.

Then carry out the subsequent actions for a predetermined number of iterations:

- For every training example, assign $c^{(i)}$, to the closest centroid.
- For every centroid μ_k, set the location to be the average of examples assigned to it.

As the algorithm advances, normally the centroids will migrate to the center of the clusters and the overall distance of the examples to the clusters gets less

This technique utilised in proposed model. The pictures's pixels are made up of three values: R(ed), B(lue), and G. (reen). They can be thought of as the grid's 3D points for RGB. The main objective of the suggested work's picture compression is to decrease number of colors by taking the average K colors that are most reminiscent of the original picture. An image with fewer colors takes up less disc space, which is what proposed method desire.

Compressing the picture might also be a preparatory step for another method. Lowering the size of the input data often speeds up learning.

The effectiveness of K-means technique hinges on creates incredibly efficient clusters. It is difficult to determine the appropriate clusters number, though. Although there are various ways to figure out how many clusters are ideal, this post will concentrate on the optimal method. The steps are explained below:

The elbow approach is one of the most used techniques for determining the number of ideal clusters. Total variations inside a cluster are denoted by the acronym WCSS, or inside Cluster Sum of Squares.

$$WCSS = \sum_{C_k}^{C_n} \Big(\sum_{d_i \, in \, C_i}^{d_m} distance(d_i, C_k)^2 \Big) \quad \ldots(4)$$

Where,
C is the cluster centroids and d is the data point in each Cluster.

The following steps are followed by the elbow approach to determine the optimal cluster value:

- It applies K-means on the provided dataset, with K values ranging from 1 to 10.
- Determine the WCSS for each K value.
- Produces a graph that compares the computed WCSS values to each cluster's K value.
- A plot point or bend's sharp tip is considered to have the greatest value of K at that intersection.

The figure below represents the elbow method:

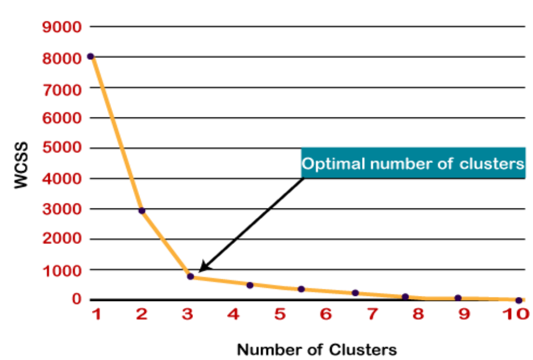

Fig. 5. WCSS VS No. of Cluster

B. Architecture

1) Algorithms

The K-Means cluster works as shown by the following algorithm:

Step 1: Using the elbow technique, select K and calculate the clusters' number.
Step 2: Randomly choose K's centroids or locations. It's possible that wasn't the supplied data set.
Step 3: Assign each information point to its nearest centroid, which will produce the required K groups.
Step 4: Determine the variance and move the centroid of each cluster.
Step 5: Repeat Step 3 to reassign each data point to the new centroid of each cluster.
Step 6: If a reassignment occurs, proceed to step 4; if not, go to COMPLETE.
Step 7: Algorithm completed

2) Flowchart

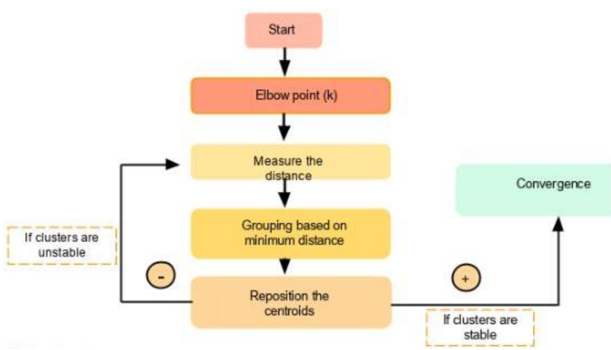

Fig. 6. Flowchart

3) Working

Initial step begin by implementing a set of function that creates initial points for the centroids. This function picks k unique points at random using the Elbow method from the input X, the training examples.

Then a method is construct to locate the nearest centroid for each training example. This is the first step of the algorithm. For each example in c, take X and the centroids as input and return the index of the closest centroid, an m-dimensional vector.

In the second step of the algorithm, calculate each example's distance from "its" centroid and take the average of those distances for each centroid, where k is the number of centroids. Transposing the examples is necessary because looping id done over the rows.

Finally, all the parameters are obtain to complete the K-means algorithm. The most iterations allowed, max_iter, is set to 10. Notably, if the centroids stop moving, return statement is implemented to return the results because there is no more room for optimization.

Now, in the next step, the image is obtained using already implemented k information. The picture is defined as the first (and final) command line option, so begin by attempting to open it. Model receive an Image object from Pillow, but the algorithm needs a NumPy array. So let's define a little helper function to convert them. For the pixels to be scaled to 0...1, observe how each value is divided by 255. Then model get the feature matrix X. Redesigning of the image is done because each pixel has the same meaning (color), thus they don't have to be presented as a grid. Finally, this model is able to use the algorithm to obtain K colors. These colors are chosen by the algorithm.

Output:
[336519.9994879626, 141611.69711384104,
86682.93995564576, 64686.94497731874,
50913.2632977519, 40686.22682230402,
32979.73628183274, 28492.671767520664,
25375.6265720664, 23094.491972583917,
20871.595122360202, 18868.510838939113,
17315.16554641331, 15905.816009267488,
14559.14441807909, 13297.934030623528,
12350.45380372168, 11492.53319826621,
10974.709382505742, 10364.74151021806]

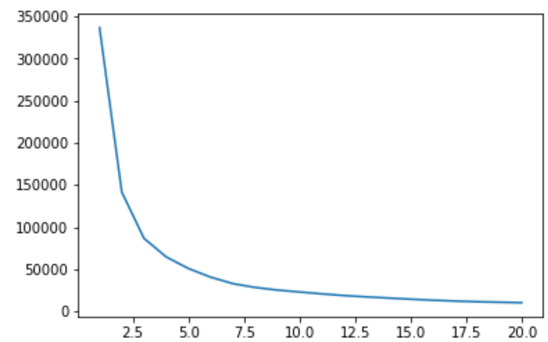

Fig. 7. Elbow Graph

Because the indexes supplied by the find k means function are 1 iteration behind the colors, the indexes is computed for the current colors. Each pixel has a value in the range of 0 to K, which naturally corresponds to its color.

When all the necessary information is obtained, reconstruction of the picture is perform by switching out the colour for the colour index and resizing it to match its original dimensions. Then transform the raw numbers back into an image using the Pillow method Image.fromarray. Additionally convert the indexes to integers because numpy only accepts those as indexes for matrices.

VI. Experimentation and Analysis

A. Experimental Setup

The image comparison K Means algorithm with Python, Numpy, and Pillow is set up as follows: First, python should be installed. Then, the next task is to install two python libraries i.e. OS, SYS, Numpy, and Pillow. Now, prepare the dataset using Python libraries such as Numpy and Pillow. After this, the setup will be ready to implement the proposed model of image compression

B. Experimental Parameters

1) Average K Color: K-means refers to the number of clusters that the algorithm was able to recognise from the data. The data points are clustered together in this approach to make the total squared distances between the data points and the centroid as short as feasible. The recommended model states that the size of the image is decreased by averaging the k colours that seem the closest to the original image.

2) Distance: It assesses every observation and places it in the nearest cluster. It defines the distances to the other centroids in order for it to qualify as the "closest" group. The K means the clustering algorithm updates a cluster's centroid whenever it adds or loses a data point.

3) Maximum Iterations: The algorithm iterates until the cluster centers have changed the fewest amount from the previous iteration. If the clusters are uniformly spherical in shape, K-means is particularly good at identifying the structure and drawing conclusions from the data. However, the method does a poor job of grouping the data if the clusters contain more intricate geometric features.

4) Compression Ratio

$$C\,R = \frac{\text{size}_{\text{after}}}{\text{size}_{\text{before}}} \qquad \ldots(5)$$

5) Compression Factor

$$C\,F = \frac{\text{size}_{\text{before}}}{\text{size}_{\text{after}}} \qquad \ldots(6)$$

6) Saving Percentage

$$S\,P = \frac{\text{size}_{\text{before}} - \text{size}_{\text{after}}}{\text{size}_{\text{before}}} \ \% \qquad \ldots(7)$$

7) Compression Time: If an algorithm's compression times are reasonable, it suggests that the algorithm is reasonable in relation to the time factor. This component might have very low values as high-speed computer accessories evolve, and those values might be influenced by how well computers work.

Compression Time = time before compression - time before compression

8) Within Cluster Sum of Squares: In the Elbow method, it is utilised. The acronym "WCSS" stands for "total variations inside a cluster.":

$$\text{WCSS} = \sum_{C_k}^{C_n}\Big(\sum_{d_i \, in \, C_i}^{d_m} distance(d_i, C_k)^2 \Big) \qquad \ldots(8)$$

Where,
C is the cluster centroids and d is the data point in each Cluster.

C. Performance Evaluation

K=5 resulted in a 71% reduction in file size, from 228kb to 65kb.

TABLE I. EVALUATION TABLE

Average K Color	20
Maximum Iterations	10
Compression Ratio:	0.285
Compression Factor:	3.507
Compression Time:	494
Within Cluster Sum of Squares	5
Saving Percentage	71.49%

D. Analysis

TABLE II. COMPARISIONS TABLE

S. No .	COMPARISIONS			
	Compression Technique	***K-mean***	***RLS-DLA***	***VQ***
1	Saving Percentage	71.49%	30%	-
2	Compression Type	Lossless	Lossy, lossless	Lossy
3	File extension	All image extention	-	JPG and TIF
4	Compression Ratio:	0.285	1.3	5.053101e − 01
5	Compression Factor:	3.507	0.769	0.0785
6	Compression Time	494 SEC	-	6 SEC

VII. CONCLUSION AND FUTURE SCOPE

It's clever to use image compression to reduce an image's size while keeping its resolution. The k-mean clustering algorithm is implemented for image compression in this study. WCSS value is calculated to get the appropriate value of k using the Elbow method. Using K-mean clustering, an unsupervised technique, and the proposed methodology achieves a lossless compressed image.

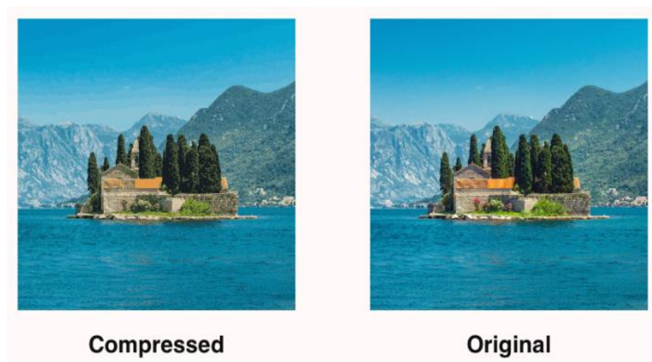

Fig. 8. Image Compresion

Each compression is ideal in its own context. Further compression techniques that can compress data at higher compression ratios while preserving image quality will be taken into consideration in the near future. Future developments in technology may allow for picture compression that takes into account both image symmetry and redundancy.

REFERENCES

[1] G.S.K.Krishna phani, B.Lavanya and Sanjay Kumar Singh proposed an idea of image compresion in their paper "Study And Analysis Ofrls-Dla Algorithmwith Natural Image Compression Techniques" 2010

[2] Rishav Chatterjee, Alenrex Maity and Rajdeep Chatterjee "Image Compression Using VQ for Lossy Compression" 2019

[3] Xin Yuan, Senior Member, IEEE and Raziel Haimi-Cohen, Senior Member, IEEE "Image Compression Based on Compressive Sensing:End-to-End Comparison with JPEG" arXiv : 1706.01000v3 [cs.CV]19Jan2020

[4] Venkateswaran Narasimhan and Y.V. Ramana Rao "K- Means Clustering Based Image Compression in Wavelet Domain" January 2007Information Technology Journal 6(1) DOI:10.3923/itj. 2007.148.153

[5] Nishant Kumar Singhai and Prateek Mishra "Image Compression And Performance Comparison Using Hybrid (WPT+DCT) Transform Method"

[ISinghai, 3(10), October 2016] ISSN: 2394-7659 IMPACT FACTOR-2.785

[6] Thamer Hameed "Image Compression Using Neural Networks: A Review" Publisher: International Association of Online Engineering (IAOE) Publication Name: International Journal of Online and Biomedical Engineering (iJOE) 2018

[7] Abhishek Thakur, Rajesh Kumar, Amandeep Bath, Jitender Sharma "Design of Image Compression Algorithm Using MATLAB" www.ijeee-apm.com International Journal of Electrical & Electronics Engineering 7 IJEEE, Vol. 1, Issue 1 (Jan-Feb 2014)e-ISSN: 1694-2310|p-ISSN: 1694-2426

[8] Er Kapoor, A Paper On A Comparative Study Block Truncating Coding, Wavelet, Fractal Image Compression & Embedded Zero Tree." IJESRT Journal ,2018

[9] Pravin B. Pokle and N. G. Bawane "Lossy Image Compression using Discrete CosineTransform" National Conference on Innovative Paradigms in Engineering & Technology (NCIPET-2012)Proceedings published by International Journal of Computer Applications® (IJCA)

[10] Surbhi Singh and Vipin kumar Gupta "JPEG Image Compression and Decompression by Huffman Coding" Volume 1 , Issue 5 , August

[11] – 2016 International Journal of Innovative Science and Research TechnologyISSN No: - 2456- 2165

[12] Emy Setyaningsih and Agus Harjoko "Survey of Hybrid Image Compression Techniques". International Journal of Electrical and Computer Engineering (IJECE) Vol. 7, No. 4, August 2017, pp. 2206~2214ISSN: 2088-8708, DOI: 10.11591/ijece.v7i4.pp2206-2214

[13] Abdelhamid Mammeri, Brahim Hadjou,and Ahmed Khoumsi "A Survey ofImage CompressionAlgorithmsfor Visual Sensor Networks" International Scholarly Research Network ISRN Sensor Networks Volume 2012, Article ID 760320, 19 pagesdoi:10.5402/2012/760320

[14] Andrea Vitali , Luigi Della Torre, Sebastiano Battiato, and Antonio Buemi "Image Compression By Perceptual Vector Quantization" 2000

[15] Bibhas Chandra Dhara "Image and video comperssion using block truncation coding and pattern fitting for fast decoding" 2008

[16] Sarang Bansod and Sweta Jain "Improved Harmony Search Algorithm For Color Image Compression" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169Volume: 2 Issue: 6 1669 – 1672

[17] S. Burak, G. Carlo, T. Bernd, and G. C. Beaulieu, "Medical image compression based on the region of interest, with application to colonct images," 2001.

[18] R. Begleiter, R. El-Yaniv, and G. Yona, "On prediction using variable order Markov models," J. Artif. Intell. Res. (JAIR), vol. 22,pp. 385–421, 2004

[19] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," J. Constr. Approx., vol. 13, pp. 57–98, 1997.

[20] ICLR 2021.2, 4, 6, and 9 R. Zhang, P. Isola, A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features asa perceptual metric. In CVPR, 2018. 8

[21] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine trans-form," IEEE Trans. Comput., vol. 23, no. 1, pp. 90–93, Jan. 1974.

[22] A. Krizhevsky, "Learning Multiple Layers of Features from TinyImages," Master's thesis.