IJESRT: 8(3), March, 2019 ISSN: 2277-9655

International Journal of Engineering Sciences &Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

Chief Editor
Dr. J.B. Helonde

Executive Editor
Mr. Somil Mayur Shah

Website: www.ijesrt.com Mail: editor@ijesrt.com

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

REDUCTION OF REDUNDANT PIXELS OF AN IMAGE USING EFFICIENT K-MEAN CLUSTERING ALGORITHM

A.Harika*[1],V.Revanth Narayana[1], Ch.Sagar Kumar[1] & Suresh Kurumalla[2]

[1]Student, Department of CSE, Anil Neerukonda Institute of Technology and Science

[2]Assistant Professor, Department of CSE, Anil Neerukonda Institute of Technology and Science

DOI: 10.5281/zenodo.2595866

ABSTRACT

Clustering means grouping of similar objects which is an important and essential part of Data Mining. Now a day's huge amount of idol data transferring over internet is an image data which consumes large memory size parallel takes more processing time.Similarlyfor transferring high quality image we require high bandwidth. Using some clustering algorithms redundancy pixels problem is reduced. Now in order to retrieve the image with reduction of redundant pixels there is a considerable amount of loss of data of an image. For this problem optimized method is chosen to provide a final image which is comparatively of smaller memory size than the original image and also visually similar to that of an original image. Hence we use K-means clustering algorithm which is essential. But existing K-means algorithm has some problems which include complex computations and also selection of centroid and cluster size should be initially declared. So we use improved k-means algorithm to generate the clusters with best accuracy and less computational time.

KEYWORDS: Data Mining, Clustering, Image segmentation.

1. INTRODUCTION

In today's world, due to the substantial growth of digital information, the data is being doubled for every 20 months and Transmission of raw images over different types of networks is required with extra demand of bandwidth because images data contains more information than simple text or documents databases are emerging in magnitudes, personalities and organizations. The computerization of business events generates ever-increasing streams of information because even simple communications, such as a telephone call, the usage of a credit card, or a medicated are usually recorded in a processor. With the widespread use of the Databases and the volatile progression of individual users and organizations are facing with the difficulty of usage of the huge data. Numerous traditional techniques for analysis and visualization have been introduced to face the rational usage of data, but due to increases in the volume of these data exponentially and randomly wit usage of successive amount of processing speed and time, traditional techniques had become in adequate. Data mining (DM) is the evolving techniques to extract models, outlines or knowledge of importance from enormous databases. Over the last twenty years, data mining is being developed enormously. These novel methodologies have their origins in approaches from statistics, pattern recognition, databases, and artificial intelligence and so on. Recent days, these are rising to a multi-disciplinary study. Data mining is a non-trivial process that can recognize the active, unidentified, potentially beneficial and eventually the apprehensible pattern from databases.

1.1 Clustering

Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities. A good clustering algorithm is able to identity clusters irrespective to their shapes. Other requirements of clustering algorithms are scalability, ability to deal with noisy data, insensitivity in the order of input records etc. Data mining is a multi-step process. It requires accessing and preparing data for a data mining algorithm, mining the data, analysing results and taking appropriate actions to achieve them. The data that is accessed is stored in one or more operational databases ,a data warehouse or a flat file. In data mining the data is collected in such a way to become a meaningful data by using two learning approaches i.e. supervised or unsupervised clustering.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

Clustering is a group of points which in term referred as group of pixel values which plays an important role in significant tasks in data mining (DM) that partitions the data samples into classes without any significant labelled samples. The illustrations belonging to the identical classes are homogenous or non-uniform and those among the classes are heterogeneous which a part of uniformed are. The process of assembling a group of physical or abstract entities into classes of similar entities is called clustering. A cluster is a collection of data items that are identical to one another with in the similar cluster and are dissimilar to the item in other clusters.

As a data mining task, cluster analysis can be used as a stand-alone tool to achieve in tuitions into the dissemination of information, to perceive the characteristics of every cluster and to emphasize on a specific group of clusters for additional study. Alternatively, it may serve as a pre-processing step for other algorithm, such as characterization and classification, which would then function on the identified clusters. Clustering is a challenging field of research where it's potential applications pure their own exceptional requirements.

1.2 Motivation

The mostly improvised conventional clustering algorithms in data mining like, k-means algorithm have complications in handling the challenges postured by the pool of regular statistics, which emphasis in ambiguity and unclear. The K-Means clustering is simple, but it has high time complexity, so it is not appropriate for enormous data set. Numerous approaches have been suggested to enhance the performance of k-means clustering algorithm as to discover healthier initial centroids in addition to further accurate clusters with less computational time. The regions obtained by K-means algorithm are also not close to human perception regions and has high computational time. The time complexity of the K-means algorithm is O (n k d t) that depends on the variables like n which is the size of the data objects in the database, k is the number of clusters of data, d is the dimensionality of the data objects and t is time required to execute the k-means algorithm. Apart from several proposed methodology, this paper proposed a novel technique as to decrease the computational time by minimizing the size of the data objects n. This methodology is known as G-Means clustering algorithm.

1.3. The Organization of Paper

In this paper there is a brief discussion of data mining, improvised clustering algorithms and clustering techniques. The further section describes the improvised algorithms and different clustering techniques used briefly in this paper so far. The suggested improved K-means algorithm clustering algorithm discussed in section 4 briefly. Section 5shows the experimental analysis of the proposed methodology on the image dataset. Section 6 shows the results of an input image and section 7 concludes the proposed Improved k-means Clustering Algorithm with high performance and less computational time.

2. BACKGROUND

2.1. Image compression

Image compression is the process of compact representation of an image, thereby minimizing the memory by reducing the image storage and transmission requirements by reducing the amount of information required to represent a digital image its worth that the redundancy is observed in every image data. Redundancy means the duplicate pixels of data in the image which may be occurred in either image or patterns. The image compression occurs by taking benefit of redundant information of an image. Reduction of redundancy provides to achieve saving of storage space of an image. Image compression mainly aims in achieving and obtaining the redundant pixels of an image when one or more of the redundant pixels are reduced or vanished. In an image compression, there are mainly three basic types of data redundancies which can be identified and exploited. Reduction of pixels by the method of compression is achieved by the identification and removal of one or more of the three basic data redundancies.

a. Inter Pixel Redundancy:

In image neighbouring pixels are not statistically independent they sometimes dynamically depends on other. That dependency occurs is due to correlation between the neighbouring pixels or alternate pixels of an image. This type of redundancy is called Inter-pixel redundancy. This type of redundancy is sometimes also called as spatial redundancy.

b. Coding Redundancy:

By using fixed or variable code words are selected to match the status of original source This type of coding is always reversible and usually implemented using lookup tables (LUTs).

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

c. Psycho Visual Redundancy:

By performing many experiments on the psycho physical aspects of human vision have made a excellence in responding proven that the human eye does not respond with equal sensitivity to all incoming visual information; some pieces of information are more important than others and unique. Most of the image coding algorithms in use today exploit this type of redundancy,such as the Discrete Cosine Transform (DCT) based algorithm at the heart of the JPEG encoding standard.

2.2.Types of Compression:

Compression can be of two types:

Lossless Compression and Lossy Compression.

a. Lossless Compression:

Lossless compression is a technique termed as if no data is lost and the exact replica of the original image can be retrieved by decompress the compressed image then the compression is of lossless compression type. Text compression is generally of lossless type.

b. Lossy Compression:

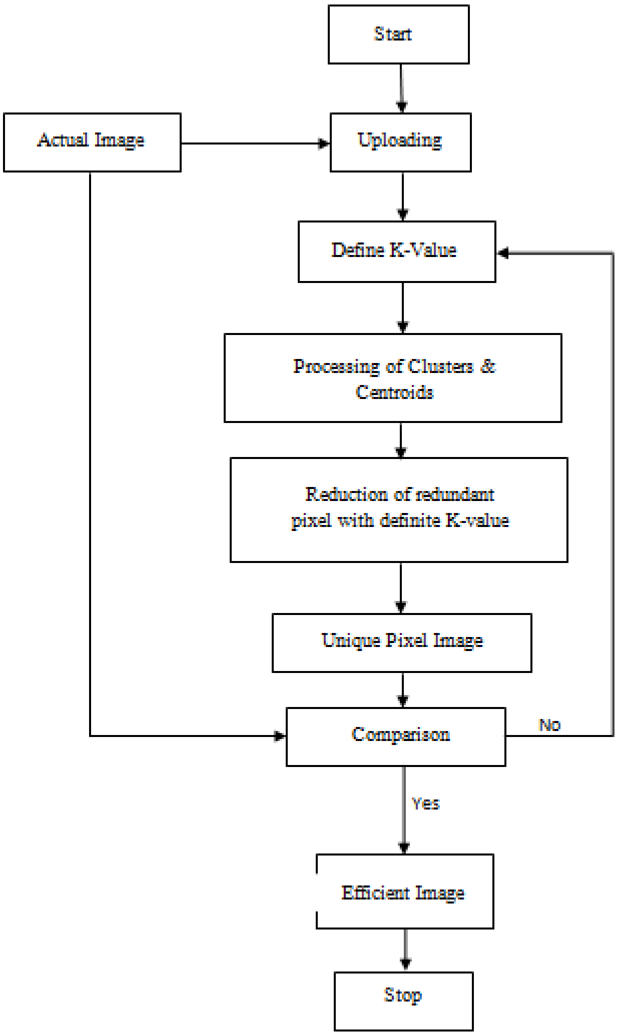
The lossy compression process neglects some less important data. The exact pattern replica of the original file with most intensive data and can't be retrieved from the compressed file. Lossy Compression is generally categorized and used for image, audio, video type of files. To decompress about the compressed data obtained then we can get a closer approximation without loss of the original file.

3. LITERATURE SURVEY

It aims in processing the technology of data mining and identifying their patterns and prospects from large amount of information. In Data Mining, Clustering is an important research topic and wide range of unverified classification application.[1]. Data mining is the analysis of extraction of data and the use of these algorithms particularly on a part of improved algorithms. software techniques for finding patterns and regularities in the set of data In this paper, an enhanced K-mean algorithm is introduced and compared with the basic K-mean algorithm [2].A new image clustering and compression method (DBIC) based on fuzzy logic and discrete cosine transform provides better compression ratio and performing time[3].K-means clustering is a method of cluster analysis which invokes to partition n number of observations into k clusters in which each observation belongs to the cluster with the nearest mean value. In this paper, we try to improve the performance of file using some changes in already existing algorithm by improvised algorithm[1]. In practical K-mean clustering algorithm any type of integer data or pixels values are used. The performance and complexity of basic K-mean clustering algorithm calculated in terms of number of iterations and time complexity is improved randomly. In future, this idea can be tested on text based clustering [2]. DBIC method can reach solution with only one iteration, and it preserves intelligibility of images together with this high compression ratio [3].

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7



Process for redundant pixels of an image.

4. PROPOSED METHODOLOGY

The main aim of this improved clustering algorithm is by removing the redundant data decreases the size and computation time while processing an image.We use large number of dataset of an images Improve the existing k-mean clustering algorithm is the main motto of this paper.

4.1Existing k-means clustering algorithm:

K-mean clustering algorithm is partition-based analysis method.The main goal of this algorithm is to form groups in the data, the number groups divided are based on a variable k. All the data points are clustered using these methodologies.In this K-mean clustering algorithm centroids of points are calculated which generates a new labelled data.

The pseudo code for k-mean clustering algorithm is given below:

```
1. The first step is to choose randomly k values from a dataset.
2. Choose initial centroid for each clustered group.
3. Calculate the distance between centroid and the data points of a data set.
4. Then add the point with minimum distance to that centroid to the appropriate          cluster.
5. Again calculate the centroids for new clusters and update the cluster groups with new assigned values.
6. Repeat the process until the required clusters are formed with minimum error.
```

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

INPUT:
N={d1,d2,…..}(number of data points)
K= Required number of cluster groups.

OUTPUT:
Set of K clusters are generated.

The following are the some of the characteristics of existing k mean clustering algorithm:
i)It is noise sensitive.
ii)There is no limitation in selection of number of clusters.
iii)It is efficient for large datasets. There are some disadvantages for existing k mean clustering algorithm:
i) The algorithm has large complexity O (n d k t) which is very large for huge data set.
ii) Regions obtained are not close to human eye perception.

4.2 Proposed K-mean clustering algorithm

In proposed algorithm, the main aim is to reduce the processing time and reduce the memory size. The Time complexity is significantly reduced. The regions produced are closed to human perception regions which is very useful for CBIR. In this approach we reduced the number of pixels that has to be processed for clustering by maintaining the count of pixels having the same values. Time Complexity is O (N d k t) Where N << n. Thus we reduce the time complexity significantly. The following are the steps for proposed k-mean algorithm.

Proposed algorithm:

- Read the image AND Obtain the Pixel data
- Calculate the frequency of every distinct pixel in the image.
- Pick K pixels randomly as Initial Cluster Centres
- Calculate the distance from each input pixel and assign it to the nearest cluster centre
- Update the centroids based on their mean values
- Repeat steps 2 & 3 Until no change in the cluster centres
- Represent each cluster with distinct colour
- Regenerate the Image using the clustered data

Distance functions:

- Simplest case: one numeric attribute A
  Distance (X, Y) = A(X) – A(Y)
- Several numeric attributes:
  Distance (X, Y) = Euclidean distance between X,Y
- Nominal attributes: distance is set to 1 if values are different, 0 if they are equal
- Are all attributes equally important
  Weighting the attributes might be necessary

The time complexity of the Basic K-means Algorithm is O(n d k t)
n = no of pixels
k = no of clusters
d = dimensions
t = iterations
n, d , k are constants.
Where t = no of iterations
t varies with the initial centres (random) and the methodology

- In this approach,

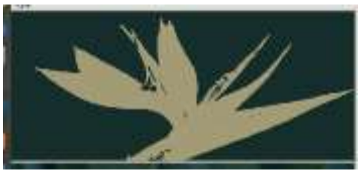
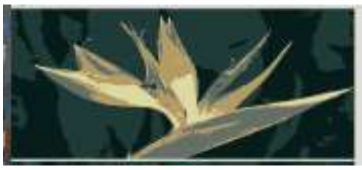
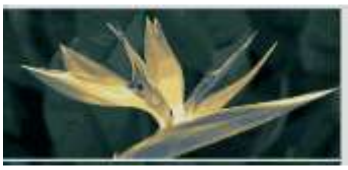
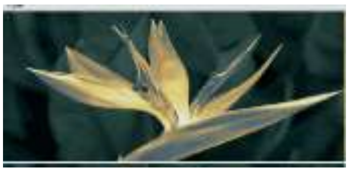
In this approach we reduced the number of pixels that has to be processed for clustering by maintaining the count of pixels having the same values.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

Time Complexity is O(N d k t) where N<< n
Thus we reduce the time complexity significantly.

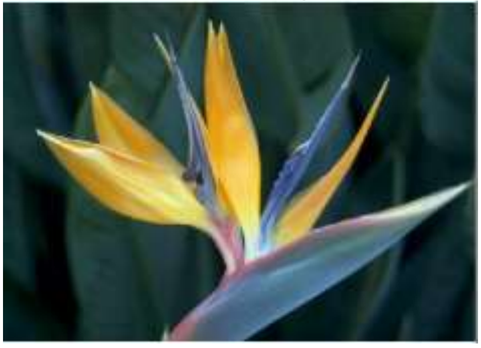
5. EXPERIMENTAL ANALYSIS

For improved K-means clustering algorithm of reduction of redundancy pixels the experimental analysis is based on different images.We take different data set of images and then compares the performance of existing k-mean algorithm and improved k-mean algorithm.In this we take clusters of size k=2,4,8,16.The analysis can be done on reduction of size of an image and computation time means processing time of an image.Improved algorithm takes less computational time than existing algorithm.

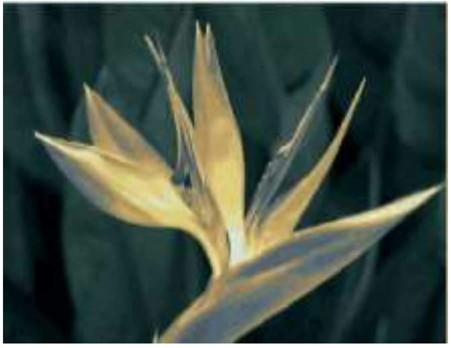
K value	picture
2	
5	
15	
25	

6. RESULT

Input image



Output K=25



IJESRT is licensed under a Creative Commons Attribution 4.0 International License.

7. CONCLUSION

The k-mean clustering algorithm plays a vital role in datamining. It is very efficient but in some cases it is not good. So, in this paper we proposed an improved k-mean clustering algorithm which solves the problems that do not solve by the existing k mean. In this paper the improved method helps to reduce the redundant data of an image of image dataset. Thus the memory size and the processing time can be reduced when compared to the existing k-mean algorithm.

REFERENCES

[1] Turk DemirdokumFabrikasıA.S¸ BozuyukBilecik-TURKEY :"An Algorithm for Image Clustering and Compression".

[2] Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity": Azhar Rauf, Sheeba, SaeedMahfooz, Shah Khusro and HumaJavedDepartment of Computer Science University of Peshawar Peshawar.

[3] "Enhanced K-Means Clustering Algorithm to Reduce Time Complexity for Numeric Values".BangoriaBhoomi M. PG Student [C.E.], Noble Engineering College, Junagadh, Gujarat.

[4] Xin Li and Michael T. Orchard "Edge-Directed Prediction for Lossless Compression of Natural Images", IEEE Transactions on Image Processing, vol. 10, NO. 6, Jun 2001

[5] Michael B. Martin and Amy E. Bell, "New Image Compression Techniques Using Multiwavelets and Multiwavelet Packets" ,IEEE Transactions on Image Processing, vol. 10, NO. 4, Apr 2001.

[6] Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J., (1999), "Blobworld: A system for region-based image indexing and retrieval," Third International Conference on Visual Information Systems, Springer.

[7] Kimia, B., (2001), "Shape Representation for Image Retrieval", Image Databases:Search and Retrieval of Digital Imagery, John Wiley & Sons, pp. 345- 358.

[8] M. C. Chiang, C. W. Tsai, and C. S. Yang, "A time-efficient pattern reduction algorithm for k-means clustering," Information Sciences, vol. 181, no. 4, pp. 716–731, 2011.

[9] H. Xiong, J. Wu, and J. Chen, "K-means clustering versus validation measures: a data-distribution perspective," IEEE Transactions on Systems, Man, and Cybernetics B: Cybernetics, vol. 39, no. 2, pp. 318–331, 2009.

[10] S. Guha, R. Rastogi, and K. Shim, "CURE: an efficient clustering algorithm for large databases," ACM SIGMOD Record, vol. 27, no. 2, pp. 73–84, 1998.

[11] M. C. Chiang, C. W. Tsai, and C. S. Yang, "A time-efficient pattern reduction algorithm for k-means clustering," Information Sciences, vol. 181, no. 4, pp. 716–731, 2011.

[12] .T. Elomaa and H. Koivistoinen, "On autonomous K-means clustering," in Proceedings of the International Symposium on Methodologies for Intelligent Systems, pp. 228–236, May 2005.

[13] de Amorim, R.C., Komisarczuk, P.: On initializations for the minkowski weighted k-means. Lecture Notes in Computer Science.

[14] de Amorim, R.C.: An empirical evaluation of different initializations on the number of k-means iterations. Lecture Notes in Computer Science 7629 (2013) 15–26.

IJESRT is licensed under a Creative Commons Attribution 4.0 International License.