International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 10 | Oct 2019 www.irjet.net p-ISSN: 2395-0072

Finding Dominant Color in the Artistic Painting using Data Mining Technique: A Survey

Piyush. S. Atram[1], Prof. Pramila M. Chawan[2]

[1]M.Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *In this article aims to discover the color which has more impact in a painting via opencv library with python programming language. To be able to find dominant color I'll use k means clustering as opposed to try and find histogram and other clustering methods in the data mining techniques for every pixel. I'll use numpy and sklearn libraries for clustering. In this article, the focus will be on improving and implementing k means algorithm 2-3 times faster than usual k means clustering algorithm besides finding the dominant color in the input images.*

Key Word: Data mining, k means, hierarchical, histogram, opencv, artistic painting, clustering

1. INTRODUCTION

Painting is the task of applying paint, color, conceal or other medium to a strong surface. Painting is a significant structure inside the visual expressions, acquiring components including drawing, motion (as in gestural painting), arrangement or reflection (as in synopsis work of art). By making the work of art digitized it has been seen that we can recognize and control picture features. In the painting, it is beyond the realm of imagination to expect to recognize dominant color physically. The most dominating color pixels are dictated by clustering the image pixels. The dominant color of pictures can likewise be utilized in applications outside of looking. Systems that show images on a PC show or TV may naturally create a matte that encompasses each image. Such systems may choose a color pixels for the dominant color.

Anyway crude information can't be utilized straightforwardly. Its genuine worth is anticipated by removing data valuable for choice help. In many zones, information investigation was generally a manual procedure. At the point when the size of information control and investigation goes past human capacities, individuals search for registering advances to computerize the procedure. Few of them discussed below:

1.1 K Means

K mean is unsupervised learning algorithm that take care of the clustering problem. The method defines a basic and simple approach to group a given data points through a specific number of clusters (let k clusters). Basically, thought is to define k centers, one for each cluster. These centers ought to be put in a path due to various location May causes distinctive outcome.

Clustering is significant and fundamental idea of data mining field utilized in different systems. In Clustering, data points are partitioned onto different clusters. These clusters represents to some significant target. Means, clusters are the block of comparative objects.

1.2 Histogram

Histogram is considered as a diagram or plot which is related to recurrence of pixels in a Gray Scale Image. With pixel regards (reaching out from 0 to 255). Grayscale picture is a picture wherein the estimation of each pixel is an example model, that is, it passes on just power information where pixel worth vacillates from 0 to 255. Pictures of this sort, generally called high differentiate, are made uniquely out of shades of diminish, moving from dim at the weakest power to white at the most grounded where Pixel can be considered as an each point in an image.

1.3 Hierarchical clustering

Various hierarchical clustering includes making groups that have a top to bottom approach. For instance, all documents and envelopes on the hard circle are composed in a chain of importance. There are two kinds of various hierarchical clusters, Divisive and Agglomerative.

2. LITERATURE REVIEW

2.1 Data mining techniques

In data mining, there are a few procedures which are utilized for clustering the data points. For example k means, hierarchical, Gaussian (EM) grouping, Fuzzy C-means, Density based clustering.

Based on exactness and running time the performance of k means and hierarchical clustering calculation is determined utilizing numerous different tools.

This work results that exactness of k means is higher than the various clustering strategies. So for huge data points k means algorithm is great.

2.2 Clustering with k means:

The generally utilized strategy in data mining is well known as k means clustering. This is straightforward and implementable so as to group the data points in various clusters according to the class necessity.

K means algorithm needs to calculate the distance between every data object and all group points in every iteration. This tedious procedure influences the productivity of clustering algorithm. In any case, it can be improved by thinking about different measures.

Consequently different further research expected to concentrate on issues that debasing the algorithm execution.

2.3 Retrieval of color from painting

The research work on the finding the color from the painting which might contains the combination of multi-color may result with the help of image processing techniques

"Kodituwakku and Selvarajah, (2004) investigated the retrieval efficiency of color histograms, color moments and color coherent vectors (CCV) by means of precision and recall."

In the recent work, open source library for image processing is implemented to generate the features accurately from the digital image. The authors concluded that combination of color descriptors produced better retrieval rate compared to individual color descriptors with the help of machine learning techniques.

3. PROPOSED SYSTEM

3.1 Problem statement

"To find the dominant color in an artistic painting using data mining technique via opencv."

3.2 Problem Elaboration

This has been observed that finding maximum color used from the painting image manually is bit difficult. This may leads to excessive use of colors from pallets. But its solution can be addressed digitally with the help of image processing and data mining techniques. In the data mining, there are several clustering techniques which can address the given problem statement.

To carried out the process of finding the dominant color from the image is done by performing the python script with image processing module i.e. opencv which help program to manipulate with the image format. Furthermore the library which is used in the proposed system like numpy and sklearn libraries are used to perform the mathematical modeling and clustering purposes.

Above solution will help the artists to know which color has been used more to draw the painting.

3.3 Proposed Methodology

There are different methods in the data mining to find the dominant color from the input image of artistic painting. But each of them have their pros and cons which might affect the system in terms of space and time complexity. Techniques like k means clustering, hierarchical clustering, Histogram analysis etc.

To work efficiently with the propose system and after understanding from the given literature survey k means clustering technique from the data mining helps a lot as compare to other clustering data mining techniques.

Here's how the algorithm implemented:

1. Select K points as initial centroids.

2. Repeat this.

3. Form K cluster by assigning each point to its closest centroids.

4. Recomputed the centroids of each cluster until means remains unchanged.

In the above algorithm the traditional working of the k means clustering can be implemented easily with the he k means algorithm is known to have a time complexity of O (n^2), where n is the input data size.

In the proposed system to work with fast k means algorithm the straightforward changes of k means clustering strategy to run k means faster by having few of the following steps:

1. The first stage is a fast distance calculation using only a small set of the data to derive the best possible area of the centers.

2. The second stage is a slow distance calculation in which the initial centers used are taken from the first stage.

3. The fast and slow stages represent the speed of the movement of the centers. In the slow stage, the whole data points can be used to get the exact location of the centers

The complexity of the k means is (KQN) where K is the number of clusters, Q is the number of iteration required to get to the stopping criteria and N is the input.

3.4 Proposed System Architecture

The proposed system workflow is as given as

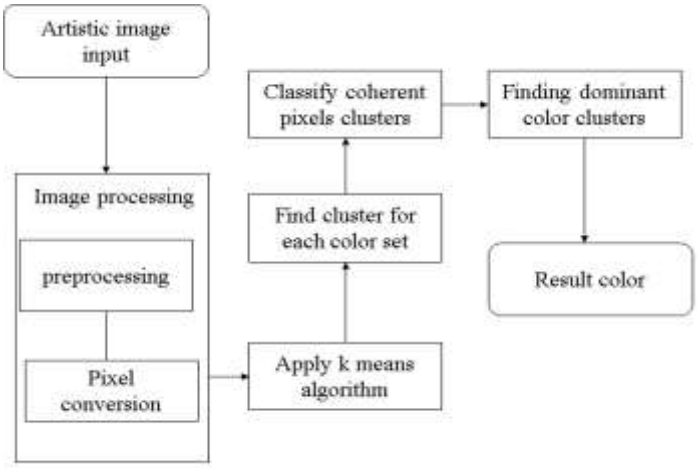

Fig -1: Work flow

As shown in above block diagram the painting image will be taken as the input for the process and with the help of opencv library in python to preprocess the input image and converted into the pixels vector to be used as data sample points for clustering with k means algorithm. By grouping the pixels of three main different colors i.e. RGB and finding the dominant color cluster among the present clusters with the fast k means clustering technique discussed in the proposed methodology.

4. CONCLUSION

In this Papers, k means clustering techniques with corresponding method and image feature extraction are reviewed. K means being most efficient among data research. Need further improvement in different part of that clustering algorithm. Above examined writing study, proposed k means algorithm is proficiently utilized for clustering the individual colors. This clustering algorithm can be improved with the proposed system and can able to find the dominant color from the artistic painting.

REFERENCES

[1] Gaurav Shrivastav and Piyush Singh. International Journal of Engineering Research & Technology (IJERT), "A Review: Fast Image Retrieval Based on Dominant Color Feature", Volume 03, Issue 01 (January 2014).

[2] Tapas Kanungo, Nathan S. Netanyahu, Angela Y. Wu. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, "An Efficient k Means Clustering Algorithm: Analysis and Implementation", VOL. 24, NO. 7, JULY 2002.

[3] Serkan Kiranyaz, Stefan Uhlmann, and Moncef Gabbouj. Department of Signal Processing, Tampere University of Technology, Tampere, Finland. 2009 Seventh International Workshop on Content-Based Multimedia indexing, "Dominant Color Extraction based on Dynamic Clustering by Multi-Dimensional Particle Swarm Optimization".

[4] E. A. Khadem, E. F. Nezhad, M. Sharifi, "Data Mining: Methods & Utilities", Researcher2013; 5(12):47-59. (ISSN: 1553-9865).

[5] K. A. Abdul Nazeer, M. P. Sebastian,îImproving the Accuracy and Efficiency of the k means Clustering Algorithm, Proceedings of the World Congress on Engineering 2009 Vol I WCE 2009, July 1 - 3, 2009, London, U.K .

[6] Arhter, D. and Vassilvitskii, S.: How Slow is the kMeans Method? SCG'06, Sedona, Arizona, USA. (2006).

[7] Elkan, C.: Using the Triangle Inequality to Accelerate K – Means. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, (2003).

[8] http://www.aishack.in/tutorials/kmeans-clustering/

[9] https://zeevgilovitz.com/detecting-dominant-colours-in-python

[10] https://matthewragan.com/2018/05/17/touchdesigner-finding-dominant-color/