

GE103

Dominant color in image

Prateek Reddy , Navrinder, Priyanka pawar, Arun Kumar

2021eeb1224-2021eeb1224@iitrpr.ac.in

2021meb1302-2021meb1302@iitrpr.ac.in

2021mmb1373-2021mmb1373@iitrpr.ac.in

2021mmb1370-2021mmb1370@iitrpr.ac.in

I. INTRODUCTION

An image contain many colour and every colour contain different shades so we are clustering(it can be viewed a subspace of dataset with similar properties) different shades of colour with similar RGB(for example green contains hundred's of different RGB's with nearly same values) values to same cluster . Detecting the dominant color in any image by using k- means algorithm is the fundamental notion of our project . Rudimentary required for our project is concepts of python like for loops, matplotlib , PIL library functions, basic concepts of co-ordinate geometry to understand this concept .

II. LITERATURE REVIEW

Clustering is the process of grouping or dividing large data set into smaller data sets of similarity so that the objects in the same clusters are more similar to each other and more different from the objects in the group .it is a non supervised machine learning model.

clustering is important analysis techniques that is employed to large datasets and finds its application in the recommendation system data mining knowledge discovery , bioinformatics and generated is not only huge in volume ,but around the world we need to process this data in parallel to reduce the cost of process

But Clustering is not free from weaknesses such that the user has to specify the number of clusters to be generated before the start of the algorithm ,which can be very difficult. Another problem which may arise when applying k means clustering algorithms is the resolution problem which is over resolution or under - resolution .Over resolution is a state when the final number of clusters generated is more than the actual number of clusters and under resolution is the converse of over resolution .

III. OBJECTIVE

There are three main parts to find the dominant colour .Firstly we read the image and get the RGB values of image using PIL library, this is the dataset for the k-mean clustering algorithm .

Then by using k-mean clustering algorithm ,we divide the data set into k sets, the question here arises is how many clusters is optimal, we can give algorithm any number of clusters .

Here comes the elbow method it finds the optimal number of clusters from the graph ,in this graph we plot number of clusters vs wcss (wcss is the sum of squares of the distances of each data point in all clusters to their respective centroids). It find the point where it look like a elbow joint (we are entering it manually) this is the point of optimum clusters.

Then we calculate centroid of the clusters formed from k which we got from elbow plot, we take centroid of the cluster with most number of elements as final RGB for most dominating colour.

IV. CONCLUSION

The k-mean clustering algorithm can find dominating colour to a good accuracy .we can resize/decrease the resolution to run the code fast.

Some times centroid of the cluster is little deviated from the original colour because in cluster data is distributed across the volume/area. The result of dominate color dependent on number of clusters we are taking ,from elbow curve we can can optimum number of clusters but some times we cannot find the exact point in the curve. These are the boundary conditions of using k-mean clustering.

V. ACKNOWLEDGEMENT

We all would like to thank our mentor Shahid sir and Prof.Sudarshan sir who gave the us the opportunity to work on this project of finding dominant colour by k-mean clustering and elbow plot.

VI. REFERENCES

<https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>
https://www.w3schools.com/python/matplotlib_intro.asp
https://www.tutorialspoint.com/python_pillow/python_pillow_quick_guide.htm