International Journal of Communication Networks and Security

Volume 2 | Issue 3 Article 5

January 2014

COLOR IMAGE QUANTIZATION USING GDBSCAN

KUMAR RAHUL
Computer Science and Engineering Department, Indian School of Mines, Dhanbad, Jharkhand-826004, India, kumarrahul.icfai@gmail.com

SUMAN KUMAR BHATTACHARYA
Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand-826004, India, sumanbit33ster@gmail.com

ROHIT AGRAWAL
Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand-826004, India, agarwal0115@gmail.com

Recommended Citation
RAHUL, KUMAR; BHATTACHARYA, SUMAN KUMAR; and AGRAWAL, ROHIT (2014) "COLOR IMAGE QUANTIZATION USING GDBSCAN," *International Journal of Communication Networks and Security*: Vol. 2 : Iss. 3 , Article 5.
DOI: 10.47893/IJCNS.2014.1091
Available at: https://www.interscience.in/ijcns/vol2/iss3/5

COLOR IMAGE QUANTIZATION USING GDBSCAN

KUMAR RAHUL[1], SUMAN KUMAR BHATTACHARYA[2], ROHIT AGRAWAL[3]

[1,2,3]Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand-826004, India
E-mail: Kumarrahul.icfai@gmail.com, sumanbit33ster@gmail.com, agarwal0115@gmail.com

Abstract- Color image quantization is the most widely used techniques in the field of image compression. DBSCAN is a density based data clustering technique. However DBSCAN is widely used for data clustering but not very popular for color image quantization due to some of issues associated with it. One of the problems associated with DBSCAN is that it becomes expensive when used on whole image data and also the noise points been unmapped. In this paper we are proposing a new color image quantization scheme which overcomes these problems. Our proposed algorithm is GDBSCAN (Grid Based DBSCAN) where we first decompose the image data in grids and then apply DBSCAN algorithm on each grids.

Keywords- *Image Compression, Color Image Quantization, Clustering Process, DBSCAN, GDBSCAN.*

I. INTRODUCTION

When we use internet nowadays, the first priority is to fast accessing the multimedia data over the channel where the bandwidth is limited. Most of the multimedia data is in the form of image, video and audio. So for accessing high quality multimedia data over the limited channel we require some sort of data compression. Quantization [1-2] is one of the most widely used data compression techniques, where the whole data is represented using some representative data points. Clustering is a quantization technique where the group of data having similar kind of property is kept in one cluster and being represented by a cluster center. In this paper we will try to observe some quantization technique based on color image or color image quantization technique. We will be discussing a density based clustering approach for data clustering known as DBSCAN, and try to implement this technique to color image. We will observe the issued will be coming when DBSCAN will be applied on color image. Then we will propose our new algorithm GDBSCAN by taking DBSCAN as base to sort out the issues related with DBSCAN. In GDBSCAN we will use DBSCAN after decomposing the image into grids and apply to each grid. We will be observing the comparison when the conventional DBSCAN algorithm is applied on whole color image data with our proposed GDBSCAN algorithm.

The paper will be as follows: Section II will be the conventional DBSCAN algorithm. In Section III we will be discussing our proposed algorithm. Section IV will be result and Section V will conclude our paper.

II. DBSCAN

Density-based spatial clustering of applications with noise is a data clustering algorithm proposed by Martin et-al [3-4]. It finds a number of clusters starting from the estimated density distribution of corresponding nodes. Therefore it is called density-based clustering algorithm. DBSCAN clustering algorithm defines a cluster based on the notion of density reachability instead of mass or distance reachability.

We will see some of the definitions associated with DBSCAN clustering algorithm:

i) Eps-Neighborhood: The € (read as Eps)-neighborhood of a point p, denoted by $N_€$ (p), is defined by $N_€(p)= \langle q \in D \mid dist(p,q) \leq € \rangle$

ii) Directly Density-Reachable: Let there are two point p and q, then the point q is directly density-reachable from the point if it is not farther away than a given distance say € i.e. q is part of €-neighborhood of p.

iii) Density Reachable: If p is surrounding by many point say p_1, p_2, $p_3 \ldots p_n$, then the point q is density reachable from the point p if p_1=p and p_n=q and each point p_{i+1} is directly from point p_i for i = 1,2,3…n-1.

iv) Density-Connected: The relation of density-reachable is not symmetric. The point q might lie on the edge of a cluster, having insufficiently many neighbors to count as dense itself. This would halt the process of finding a path that stops with the first non-dense point. By contrast, starting the process with the point q would lead to p (though the process would halt there, being the first non-dense point). Due to this asymmetry, the notion of density-connected is introduced. Two points p and q are density-connected if there is a point o such that both the points p and q are density-reachable from o. Density connectedness is symmetric.

v) Cluster: Cluster is formed from the points which are reachable from one another with respect to a given distance and a minimum number of points within the distance. A cluster is a subset of the points of the database. A cluster should satisfies two properties: firstly, all points within the cluster are mutually density-connected and second, if a point p is a part of a cluster C and the point q is density-reachable from the point p with respect to a given distance and a minimum number of points within the distance, then the point q is also a part of the cluster C.

vi) MinPts: The minimum number of point required to form a cluster is known as MinPts.

vii) Noise: A point which is neither a core point nor density reachable to other point of cluster is term as noise. Let C_1 ,C_2,. . ., C_k be the clusters of the database D with respect to parameters $€_i$ and $MinPts_i$ for i = 1, . . ., k, Then the noise is define as the set of points in the database D that does not belonging to any cluster C_i , i.e. noise = $\langle p \in D \mid \forall\ i : p \notin C_i \rangle$.

Algorithm 1:

Step1: DBSCAN requires two parameters to form a cluster. The first parameter are € and the second parameter is the Minpts (minimum number of points).
Step2: To find a cluster, DBSCAN starts from an arbitrary point say starting point p that has not been visited and retrieves all points that are density-reachable from p with respect to € and MinPts.
Step3: If p is a core point, this point's €-neighborhood is retrieved, and if it contains sufficient points to form a cluster then a cluster is started otherwise the point is labeled as noise.
Step4: If p is a border point, no points are density-reachable from p and DBSCAN visits the next point of the database. As we use global values for € and MinPts, DBSCAN may merge two clusters into one (according to definition cluster), if two clusters of different density are “close” to each other.
Step5: If a point p_i is found to be a dense part of a cluster C_j, then p_i’s €-neighborhood will also be the part of the cluster C_j. Therefore, all points that are found within the €-neighborhood are added, as is their own €-neighborhood when they are also dense. This process continues until the density-connected cluster is completely found. Then, a new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.

When we applied DBSCAN clustering algorithm in true color images for color image quantization we encountered some problems:

When DBSCAN clustering algorithm was applied on whole image, it became computationally very expensive because for each unvisited point we need to find the Euclidean distance of that point with whole image data set. So for each point the computation becomes very high.

III. PROPOSED ALGORITHM

In our propose algorithm we tried to solve the problems associated with conventional DBSCAN clustering algorithm. Instead of applying the DBSCAN clustering algorithm on whole color image data set, we decomposed the image data into 16 × 16 grids. After decomposition we applied DBSCAN clustering algorithm on each grids. We got some set of cluster centers from each grid. Finally we mapped each pixel with the corresponding cluster center and reconstructed the image. The block diagram for the proposed algorithm is shown below in figure 1.

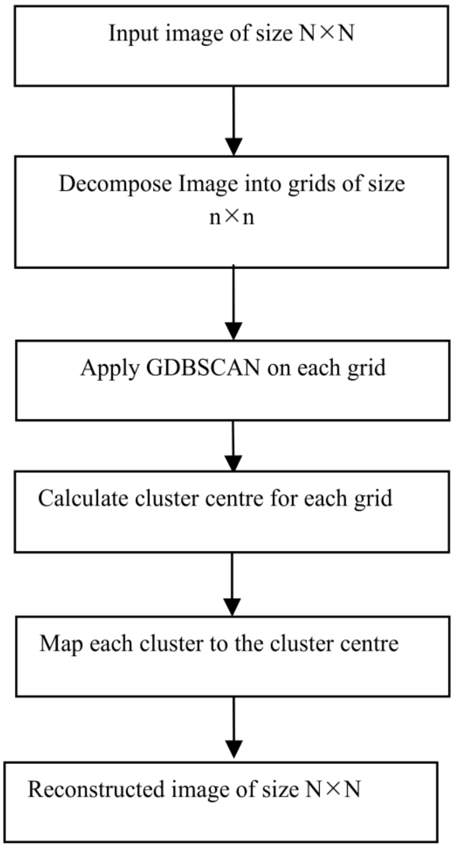

Figure 1: Block Diagram of Propose Algorithm

IV. RESULTS

We applied our proposed algorithm on various well known color images. We will compare in future our proposed GDBSCAN algorithm with conventional DBSCAN algorithm applied on whole image and see the computation time with respect to cluster formation for both methods. We have shown some of the results after applying GDBSCAN to the input image of Pepper (512 × 512) and Pepper (256 × 256) shown in figure 1 and figure 2 shows the result image after applying our proposed GDBSCAN algorithm for color image quantization. Table 1 shows the PSNR value and the computation time for cluster formation for our proposed GDBSCAN method.

V. CONCLUSION

In this paper we have seen a density based data compression technique DBSCAN, which is an efficient data clustering technique. We have applied DBSCAN clustering algorithm on color image for color image quantization and encountered some issues when DBSCAN clustering technique has applied on whole image. We proposed a new algorithm GDBSCAN by decomposing the color image data into grids and applied DBSCAN algorithm on each grid.

We will also see some future scope where we will try to efficiently map the noise points after formation of the cluster centers, which is now been ignored. We will try to reduce the total number of cluster center for the color image data.

(a) Pepper Color Image (512 × 512) (b) Pepper Color Image(256 × 256)

Figure 1: The Standard Test Image

(b) Pepper Color Image (512 × 512) (b) Pepper Color Image(256 × 256)

Figure 2: The Image after applying proposed GDBSCAN

Test Image (Pepper) Size	PSNR	Execution Time
512 × 512	43.4021	54.8405
256 × 256	36.2211	18.3285

Table 1: The PSNR and overall computation time (In Seconds) for our proposed GDBSCAN algorithm

REFERENCES

[1]. M.T Orchard and C.A Bouman, "Color quantization of images,", IEEE Transactions on Signal Processing, vol. 39, no. 12, pp. 2677-2690, 1991.

[2]. G. Gan, C. Ma and J. Wu, “Data Clustering: Theory, Algorithms, and Applications”, SIAM, 2007.

[3]. Martin Ester, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), 1996.

[4]. K. Mumtaz and K. Duraiswamy, “A novel density based improved k-means clustering algorithm-Dbkmeans”, International Journal on Computer Science and Engineering, vol.02, no. 02, pp. 213-218, 2010.