ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Segmentation of Images using Density-Based Algorithms

Atrayee Dhua[1], Debjani Nath Sarma[2], Sneha Singh[3], Bijoyeta Roy[4]

Under Graduate Student, SMIT, Sikkim Manipal University, Gangtok, India[1, 2, 3]

Assistant Professor, SMIT, Sikkim Manipal University, Gangtok, India[4]

Abstract: Image segmentation of noisy image has nowadays gained popularity in the field of computer vision. This paper presents comparison between two density based clustering algorithms DBSCAN and Mean Shift. Considering an image as a dataset of pixels we firstly remove salt and pepper noise from an image using median filtering technique followed by applying DBSCAN algorithm to cluster it. Next, the implementation of Mean Shift algorithm is seen followed by the comparison of both the outputs. Density based clustering algorithms are used to find spatial connectivity and colour similarity of the pixels, which is used to discover clusters of arbitrary shape leading to the partitioning of pixels and further isolating the noise points. Experimental results using proposed method demonstrate encouraging performance.

Keywords: Image segmentation, Cluster, DBSCAN, Median Filter, Salt and Pepper noise.

I. INTRODUCTION

Clustering is the process of grouping data into classes or clusters so that objects within a cluster have high similarity between them, but are very different from objects in other clusters. An image can be regarded as a spatial dataset that contains a huge amount of data which needs to be processed to make it understandable. Image Segmentation is implemented so as to partition an image into a collection of connected set of pixels so as to cluster it into non-overlapping and homogeneous groups based on intensity and texture of the image. The segmentation algorithm employs pre-processing, that involves denoising the given image to pass through an appropriate filter such as median filter.

In this paper, we present an image containing salt and pepper noise or impulse noise to be segmented using density-based approach. Salt and pepper noise are those that appear bright in dark parts of an image and dark in the bright parts of an image. The method used for the removal of this kind of noise is median filter. Median filtering techniques are much effective in removing salt and pepper noise as compared to other filtering techniques. In median filtering, the neighbouring pixels are ranked according to brightness (intensity) and the median value becomes the new value for the central pixel. Denoising is followed by using DBSCAN (Density Based Spatial Clustering of Applications with Noise) to integrate the spatial connectivity and colour similarity simultaneously in the segmentation process. Using clustering algorithms involve some problems. For example-if the user does not have much knowledge about the domain, it may become difficult to know which input parameters to be used for specific database. Furthermore, spatial database may contain an awful lot of data in which, trying to find clusters of data may become costly. Shapes of cluster may also be arbitrary and very complex. There are some well-known clustering algorithms like K-means, K-medoid, Hierarchical Clustering and Self-Organized Maps. Nevertheless, none of these algorithms can handle all these three mentioned problems in a good way. Another density-based clustering algorithm, Mean Shift is also implemented which is capable of clustering images of arbitrary shape but has more time complexity and less accuracy as compared to that of DBSCAN. A tabular representation showing the difference in the run-time for different images using both the clustering algorithms is presented in the paper.

The remainder of the paper is organised as follows: Introduction to density based clustering method, introduction to Median filtering technique, segmentation method, and experimental results and lastly conclude the paper with summary of lessons learned and future research.

II. DBSCAN

Clustering is widely used in many areas such as data mining, statistics, biology, pattern recognition, computer vision and machine learning. It was proposed by Ester in 1996.In this study we apply a density–based clustering approach, DBSCAN, to segment an image. A brief introduction of DBSCAN algorithm is discussed below: DBSCAN locates regions of high density that are separated from one another by regions of low density. For a given spatial dataset and objects in two-dimensional space, we need two input parameters for forming a cluster: Eps and MinPts. Eps gives the radius of a single cluster and MinPts give the minimum number of points that we intend to have within the given cluster. DBBSCAN is based on seven rules or definitions.

Definition 1**:** (Eps-neighbourhood of a point) - The Eps-neighbourhood of a point p, denoted by N_{Eps} (p) is given by N_{Eps} (p) = $\{q \in D \mid dist(p, q) \leq Eps\}$ where, D=

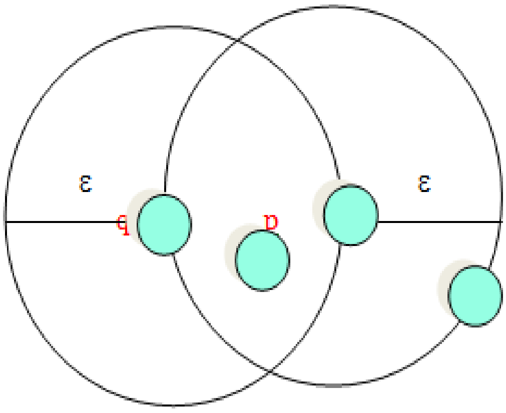
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

number of points within a specified radius (Eps). (Fig.1) [1]

Definition 2: (Core points) - Eps-neighbourhood of an object contains at least MinPts of objects, then the object is called core object and the points within that cluster are called core points. In the given figure below p is referred to as the core point. (Fig. 1) [1]

- A border point has fewer than MinPts within Eps, but is in the neighbourhood of a core point and lies on the border of the cluster.
- Any two core points are close enough within a distance Eps of one another are put in the same cluster.

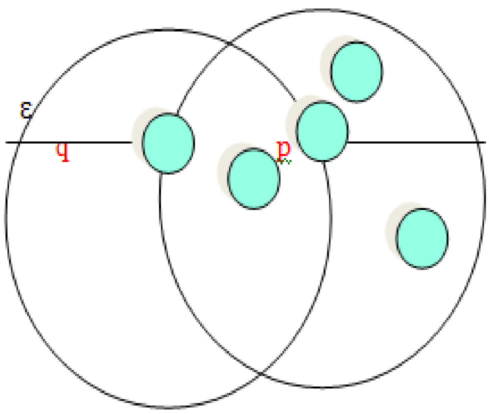


ε- Eps Neighbour of p and q
p is a core point (MinPts = 4) and q is not a core point

Fig. 1: Eps-neighbourhood of a point and Core points

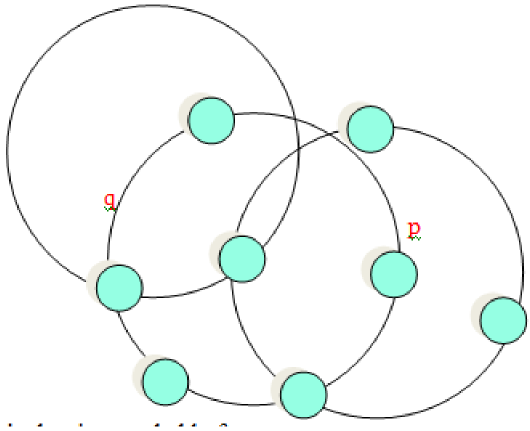
Definition 3: (Directly density-reachable) - An object p is directly density-reachable from object q if p is within the Eps-neighbourhood of q and q is a core object. (Fig. 2)

- $p \in N_{Eps}(q)$ and[1]
- $|N_{Eps}(q)| \geq MinPts$ [1]



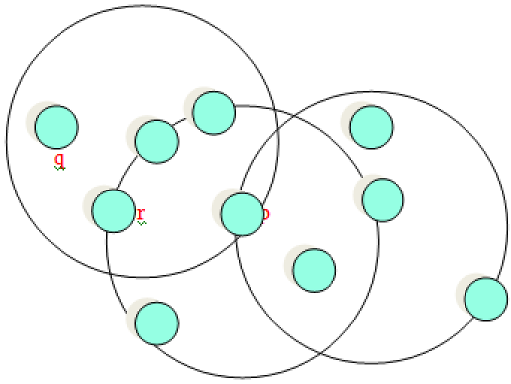
q is directly density-reachable from p
p is not directly density-reachable from q
Fig. 2: Directly density-reachable

Definition 4: (Density-reachable) - An object p is density-reachable from q w.r.t Eps and MinPts if there is a chain of objects $p_1, \ldots, p_n$ with $p_1 = q$, $p_n = p$ such that p_{i+1} is directly density-reachable from p_i w.r.t Eps and MinPts for all $1 \leq i \leq n$. (Fig.3). [1]



q is density-reachable from p
p is not density-reachable from q
Fig. 3: Density-reachable

Definition 5: (Density-connectivity) - Object p is density-connected to object q w.r.t Eps and MinPts if there is an object o such that both p and q are density-reachable from o w.r.t Eps and MinPts. (Fig. 4)



p and q are density-connected to each other by r
Density-connectivity is symmetric in nature
Fig. 4: Density-connectivity

Definition 6: (Cluster) - A cluster C, in a set of objects D w.r.t Eps and MinPts is a non-empty subset of D satisfying:

- Maximality: For all p, q if $p \in C$ and if q is density-reachable from p w.r.t Eps and MinPts then also $q \in C$.
- Connectivity: For all p, $q \in C$, p is density-connected to q w.r.t Eps and MinPts in D.
- Note: Cluster contains core objects as well as border objects.

Definition 7: Objects which are not directly density-reachable from at least one core object. A noise point is any point that is not core point or a border point.

III. MEDIAN FILTER

The pre-processing step involves the removal of extraneous artifacts that affects the quality of an image. Median filter is a non-linear filtering technique used to remove noise from an image while preserving the edges. It is most effective for the removal of salt and pepper noise

 DOI 10.17148/IJARCCE.2015.4561

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

or impulse noise. An image can be regarded as an array of pixel values. Median filter works as follows:

Each pixel is traversed in sequential manner and each of them is replaced by median value of neighbouring pixels. The image is considered to be a 3 × 3 matrix. It is then converted into a size 1 × 9 matrix. The pattern of neighbours is called window. To calculate the median values sort all the pixel values into numerical order and then replacing the considered pixel by the median (middle) pixel value.

IV. METHODOLOGY

In this section, a description of the proposed segmentation method used is given. (Fig. 5) The aim is to use a Median Filter to remove noise from a given noisy image and further by using a density-based approach the clustering of the image is done which results in the segmentation of the image.

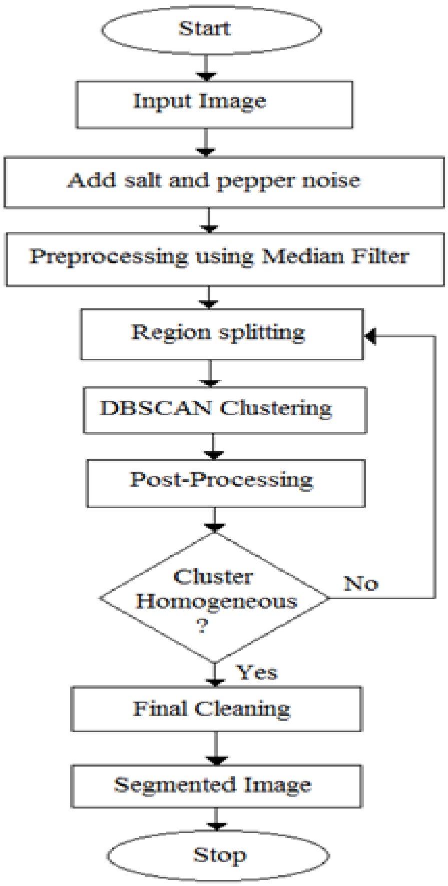


Fig.5: Flowchart of image segmentation procedure

The flowchart shows the steps for the segmentation procedure that are explained as follows:

A. Add salt and Pepper noise
The first step is the adding of salt and pepper noise to a given input image. The noise particles are

B. Pre-processing
Noise points present in the image can degrade the quality of image so it is necessary make the image devoid of noise. The image is smoothened by using a 3×3 median filter which sorts the pixel values according to its ascending order into a window of matrix 1×9, thereby selecting the mean value as the first value of the new matrix for a noiseless image. Using this filter helps in preserving the edges of the image for border detection purpose.

C. Region Splitting
The basic idea of region splitting is to divide the whole image into four sub-regions and thereby further breaking each sub-region into smaller sub-regions until no further splitting can occur. It follows a divide and conquers method or top-down approach. Finally merging of the adjacent regions takes place. The resultant image has better accuracy.

D. DBSCAN Clustering
The sub-regions we get after applying region-splitting in the given image is used for further processing steps. Next step in this algorithm is DBSCAN Clustering in which clustering process keeps homogeneous structures of the split regions of the image into same clusters or region.

This process requires two input parameters i.e. Eps and Minpts. Here, Eps denotes the radius of cluster we want to have within the region of which the minimum number of points (Minpts) to form a cluster will be checked. Generally, for images with larger pixel value, larger radius values are taken to improve the accuracy of the output.

The process starts with selection of an arbitrary point in the image. Taking the point and the value of input radius, Minpts within the region formed will be checked. If the number of points in neighbourhood of the given arbitrary point is equal to or exceeds the input Minpts value, then it will form a cluster. In the same way, we proceed to form further clusters. Points lying outside the cluster form noise points.

E. Post-Processing
After DBSCAN clustering is applied once on the given image, to improve the accuracy of the output produced the process of DBSCAN clustering is iterated more than once in the way given below:

Size of clusters formed is compared with the size of image. If size of any of the clusters formed is greater than 1/10th of total size of image another iteration of DBSCAN is performed on it. After the size of cluster becomes less than 1/10th of image the iteration stops.

F. Final Cleaning
This is the last step in the algorithm. Here, if the size of any of the clusters is less than 1/100th of the size of the image, that cluster is merged with the nearest cluster.

V. MEAN SHIFT

Mean shift is a non-parametric density-based clustering algorithm which is used to discover maxima of a density function. It does not require prior information about the

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

number of clusters and can be used for the clustering of any arbitrary shaped object. Mean shift transforms the given space into probability density function (PDF) and for each data point, t defines a window around it and computes the mean of the data point. Then it shifts the centre of window to the mean and repeats the algorithm till its convergence.

A given image can be regarded as 3-dimensional space with its pixels as n number of objects. Firstly, the Kernel density function is calculated as follows:

$f(x) = \frac{1}{nh^3}\sum_{i=1}^{n} K\left(\frac{\mathbf{x-x^i}}{\mathbf{h}}\right)$, where h is the kernel parameter.

Now, to find the kernel function:

$K(x) = \exp^{-\frac{x^2}{2\sigma^2}}$

In the next step, we find the gradient of probability density function which is set to zero for finding the peaks of density function.

The resultant mean shift function m(x) is given by :

$$m(x) = \frac{\sum_{i=1}^{n} x_i \exp^{-\frac{(x-x_i)^2}{2\sigma^2}}}{\sum_{i=1}^{n} \exp^{-\frac{(x-x_i)^2}{2\sigma^2}}}$$

m(x) denotes the new location of point x and the mean shift distance is given by the distance travelled by the point from x to m(x).

VI. EXPERIMENTAL OUTPUTS

Fig. 1. Input Image

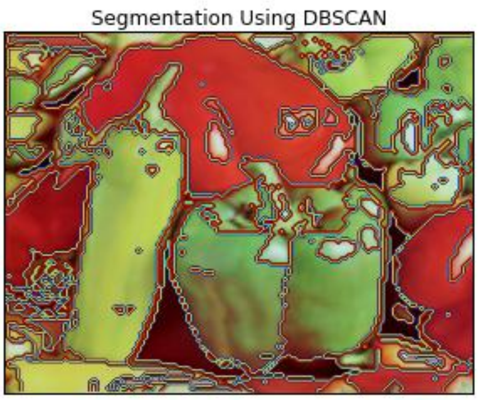


Fig. 2. Segmented image using DBSCAN

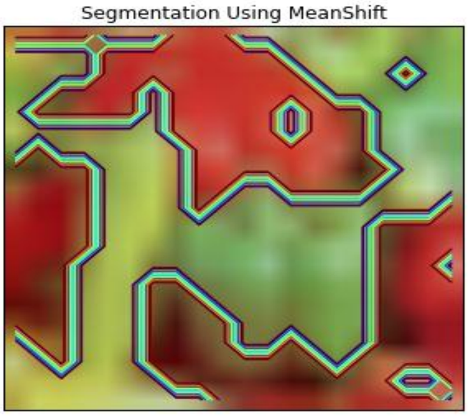


Fig. 3. Segmented image using Mean-shift

Fig. 4. Input Image

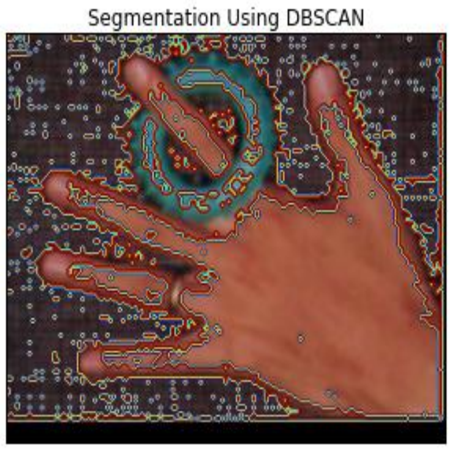


Fig. 5. Segmented image using DBSCAN

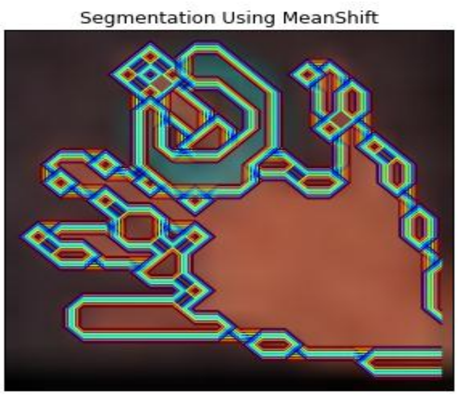


Fig.6. Segmented image using Mean-shift

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Fig. 7. Input Image

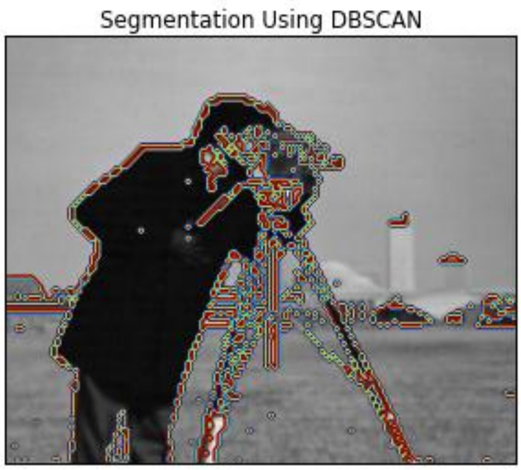


Fig.8. Segmented image using DBSCAN

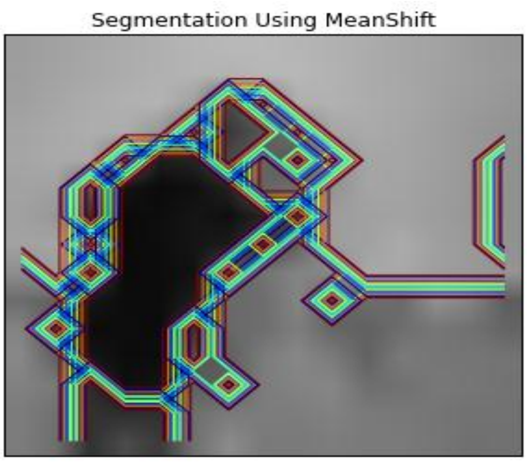


Fig.9.Segmented image using Mean-shift

TABLE I. RUN TIME IN SECONDS

Image	DBSCAN	Mean Shift
1.	0.2609	1.254
2.	0.491	0.862
3.	0.209	0.401
4.	0.376	2.038
5.	0.681	1.097

VII. CONCLUSION

In this paper, a clustering algorithm DBSCAN (Density Based Clustering Algorithm) is presented which is used to cluster images into different homogeneous structures. Furthermore, a Median Filtering technique for removing salt and pepper noise is also implemented which is effective in preserving edges of a given image and can also be applied for edge detection technique. Here, two Density based algorithm DBSCAN (Density Based Clustering Algorithm) and Mean shift algorithm are compared which are used to cluster images into different homogeneous structures. The resultant output shows that DBSCAN is having better accuracy and time complexity as compared to that of Mean shift algorithm. DBSCAN is very effective in isolating noise points from a given image whereas Mean shift algorithm does not work well for noisy images.

REFERENCES

[1]. Ester M., Kriegel H.-P., Sander J., and Xu X. (1996) "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland: Oregon, pp. 226-231.

[2]. Celebi, M.E.; Aslandogan, Y.A.; Bergstresser, P.R., "Mining biomedical images with density-based clustering," Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on, vol.1, no., pp.163,168 Vol. 1, 4-6 April 2005.

[3]. Qixiang Ye; Wen Gao; Wei Zeng, "Color image segmentation using density-based clustering," Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on, vol.2, no., pp.II, 401-4 vol.2, 6-9 July 2003.

[4]. James Church, Dr. Yixin Chen, and Dr. Stephen Rice, "A Spatial Median Filter for Noise Removal in Digital Images", Computer Science and Information System, University of Mississippi, Southeastcon, 2008. IEEE, vol., no., pp.618,623, 3-6 April 2008,doi: 10.1109/SECON.2008.4494367.

[5]. Pinaki Pratim Acharjya, Soumya Mukherjee, Dibyendu Ghoshal,"Digital Image Segmentation Using Median Filtering and Morphological Approach", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1, January 2014.

[6]. Oncel Tuzel, Fatih Porikli, Peter Meer "Kernel Methods for Weakly Supervised Mean Shift Clustering", IEEE 12th International Conference on Computer Vision, 2009.

[7]. Yizong Cheng "Mean Shift, Mode Seeking, and Clustering", IEEE Transactions on pattern analysis and machine intelligence, vol. 17, no. 8, August 1995.