# Traffic congestion detection in a city using clustering techniques in VANETs

# Anita Mohanty<sup>1</sup>, Sudipta Mahapatra<sup>2</sup>, Urmila Bhanja<sup>3</sup>

<sup>1</sup>Department of Electronics & Instrumentation Engineering, SIT, India <sup>2</sup>Department of Electronics & Electrical Communication Engineering, IIT, India <sup>3</sup>Department of Electronics & Telecommunication Engineering, IGIT, India

#### Article Info

Article history:

# ABSTRACT

Received May 9, 2018 Revised Nov 27, 2018 Accepted Dec 7, 2018

#### Keywords:

Fuzzy C-means Clustering Fuzzy K-means Clustering K-means Clustering Traffic Congestion Vehicular Ad Hoc Network

Road traffic congestion, a serious illness in developing regions, is one of the biggest problems in our day-to-day life, resulting in delays, wastage of fuel and money. In this paper, a new model is developed using Simulation of Urban Mobility (SUMO) simulator for simulating a realistic traffic scenario for a large city like Bhubaneswar where, traffic congestion is a critical issue. In a city, traffic congestion is characterised by many parameters such as rapid growth of population, number of four wheelers, inadequate and poor road infrastructures and shortage of physical plan to govern the developments, which are focused on enhancing the volume of the roads by raising the number of lanes, over-passes, underpasses and over-bridges at many junctions. However, for the success of these master plans to fully overcome the congestion issues, it is necessary to transmit the congestion information to vehicles coming towards a congestion area by using a Vehicular Ad-hoc Network. This paper analyzes clustering techniques in Vehicular Ad-hoc Networks to detect congestion in roads with the minimal infrastructural support. The raw data from vehicles are classified using cluster analysis. Out of a number of algorithms that are used to solve the congestion detection problem, three important algorithms such as Centroid based K-means, object based FCM and FKM algorithms are compared in this work on the basis of data points and number of clusters. The results of the algorithms are close to each other, but fuzzy techniques are preferable as the traffic situations are dynamic in nature.

> Copyright © 2019 Institute of Advanced Engineering and Science. All rights reserved.

#### **Corresponding Author:**

Anita Mohanty, Department of Electronics & Instrumentation Engineering, SIT, Bhubaneswar, 751024, India. Email: anita@silicon.ac.in

# 1. INTRODUCTION

Transportation traffic control is a critical problem in this advanced era. A lot of time and fuel are wasted everyday by vehicles facing congestion around the world [1]. The reason behind it is the increase in population and the number of vehicles in large cities. Because of this, an automated traffic control system is required to manage the congestion problem smoothly and on a continuous basis [2]. Traffic congestions occur either due to some external factors such as road maintenance, rush hours, heavy rain, fog and bottleneck condition, etc., which are predictable or unpredictable incidents created due to the behaviour of drivers, accidents, etc. In large cities traffic congestion is becoming worse due to the rise in population, the number of four wheelers, a rapid development of business centres and an increase in social and economic activities.

As day by day the number of vehicles is increasing, traffic congestion becomes a typical scenario in large cities which waste a lot of time as well as fuel. In large cities, we observe certain huddles like the

behavior of outsiders, construction/repair work of roads/pavements, weather conditions and the behavior of street hawkers that slow down the traffic and create problems for people travelling from one place to another. Also, the construction of roads/pavements or repair of pot holes cause more amount of delay and in turn lead to severe traffic congestion. Drainage systems in cities on rainy days work badly and introduce delays of about 30 to 45 minutes in travel. Sometimes, congestions take place due to U-turn of vehicles during peak hours. The city administrators are trying to work out the problem by widening the roads, by constructing over-bridges, over-passes, underpasses etc. at the junctions. Such steps are not sufficient enough to minimize trafficnow-a-days.

Hence, efficient intelligent systems are required to be implemented to control the current traffic jams in big cities, where vehicles can communicate with each other using Vehicular Ad Hoc Networks (VANETs) to overcome the congestion problem [3-5]. These congestion detection mechanisms implemented in an intelligent system can be categorized into two types: traffic management control unit based congestion detection and VANET based congestion detection. In this first approach a plenty of sensors are installed to gather traffic information and the control unit is used to govern the event of road congestion by scrutinizing the data collected from the sensors [6-9]. In the second approach vehicles in moving condition are used to collect the information of vehicles in close proximity and take decision about congestion and exchange the road condition due to congestion with the other vehicles. According to a number of congestion detection methods such as automatic traffic congestion identification based on gain amplifier theory [10], congestion recognition using wavelet technique [11] and congestion detection by pattern recognition [12], the amount of messages received from individual vehicles are more. As the bandwidth availability in a VANET is finite, a message aggregation scheme is required to be considered. A structure-free message aggregation scheme is described in [13] where individual vehicle with the other vehicles may behave as a message aggregator, but the disadvantage is that it should be within a predefined area of the event. Cooperative Traffic Congestion Detection (CoTEC) is a method [14], which handles a message aggregation technique based on fuzzy logic to detect the congestion on the road. But, none of these methods are able to reduce the bandwidth requirement.

As the movement of vehicles is dynamic in nature, a fuzzy logic based clustering technique is preferable. At a junction, when vehicles are facing congestion, nearer vehicles have their parameters very close to each other. So, by using appropriate clustering algorithm a cluster can be formed with parameters more similar to each other. When their cluster centres are close to each other, it means that the vehicles are in a close proximity leading to congestion.

One of the popular techniques *K*-means clustering (a hard clustering technique) is fast, robust, easier to implement and better computational time. But it is unsuccessful in getting overlapping clusters [15]. In fuzzy clustering techniques, an object is not only the member of a cluster but member of many clusters. But this system is relatively costlier than hard computing techniques. Our paper compares three different clustering algorithms to detect congestion by taking some of the vehicle parameters like speed, fuel consumption and  $CO_2$  emission into consideration in a VANET environment. In our research work, *K*-means, Fuzzy C-means and Fuzzy *K*- means clustering algorithms are analysed based on their execution time.

These above clustering techniques are tested to control the traffic problem in a big city like Bhubaneswar, an administrative, information technology, education and tourism driven city. Although these techniques exist, these were applied for traffic congestion detection earlier. In this paper the real world vehicle data are extracted and simulated in mat lab using these techniques. Section 2 gives an overview of a VANET. The congestion at Kalinga Hospital Junction of Bhubaneswar City is created using SUMO simulator and various parameters are extracted for clustering as explained in Section 3. The methodologies are explained in Section 4. These are illustrated taking one example scenario with the results are reported in Section 5. Finally, Section 6 concludes the paper.

#### 2. VANET: AN OVERVIEW

VANETs are cost effective, distributed traffic congestion detection systems. These generally require a set of inexpensive devices, which can be incorporated into vehicles and which communicate with a satellite to accumulate the data and transfer it to the system. In an Intelligent Transport System (ITS), each device works as a sensor, receiver and router to broadcast the information throughout the network, for a safe and comfortable driving and travelling experience. This continuous exchange of information between vehicles includes data about the speed of the vehicles and their locations. VANETs are used in an ITS to improve the driving efficiency, traffic safety and comfort and also detect road congestion [16, 17].

The main components of an ITS, as shown in Figure 1 are: Application Units (AUs), vehicle Board Units (BUs) and Road Side Units (RSUs), installed separately or integrated with BUs. AUs are sophisticated devices which provide applications related to vehicle safety. BUs are installed on board of a vehicle and communicate with BUs installed in other vehicles or with Road Side Units. They also communicate with

AUs. RSUs are fixed units installed along the side of the road to provide the coverage and connectivity to all vehicles [19].

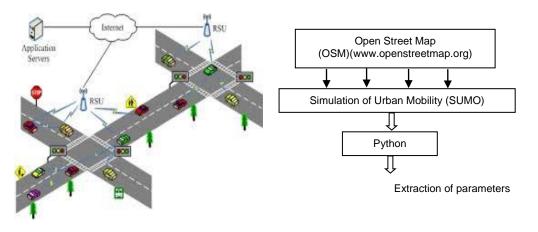


Figure 1. Architecture of an ITS [18]

Figure 2. Process Flow for extracting Real Time data from vehicles

# 3. EXTRACTION OF PARAMETERS USING SUMO SIMULATOR

Simulation of Urban Mobility (SUMO) simulator is used to create a traffic scenario from which the parameters of vehicles are extracted to detect the congestion in a particular junction. The process flow for extracting real time data from vehicles is shown in Figure 2. The real map of Kalinga Hospital Junction, the most crowded junction of Bhubaneswar, is taken from Open Street Map and given to SUMO for simulation of a real time traffic scenario. SUMO is an open source, highly portable, microscopic and continuous road traffic simulation package designed to handle large road networks. The scenarios in SUMO simulator has two parts: road network (map) including roads, streets, traffic lights junctions etc. and traffic demand expressing the details of vehicles like speed of vehicles, direction, departure time and arrival time, position etc. Figure 3(a) shows the import network of Kalinga Hospital Junction taken from http:// openstreetmap.org [20].

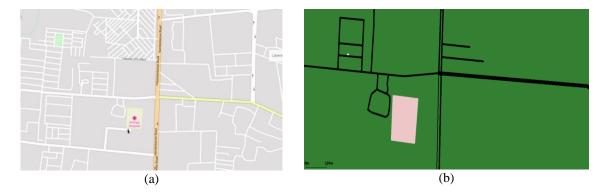


Figure 3. (a) Original Open Street Map of Kalinga Hospital Junction, Bhubaneswar [20] and (b) Imported map from OSM in SUMO

This downloaded map saved in .osm file format is imported to SUMO to create traffic environment which is saved in .cfg file as shown in Figure 3(b) with the help of Netconvert, Polyconvert and randomTrips.py tools. The congestion on a road in SUMO simulator is created by delaying a vehicle on a lane, which can be assumed as an accident on a road in real world. At Kalinga Hospital Junction the congestion is created in SUMO as shown in the Figure 4. Then the raw outputs, which contains lane id, CO,  $CO_2$ ,  $NO_{x_n}$ ,  $PM_{x_n}$  noise, fuel consumption, maximum speed, mean speed etc. are extracted from the simulator for simulation.

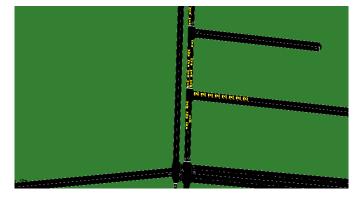


Figure 4. Congestion created at the Kalinga Hospital Junction

#### 4. CLUSTERING ALGORITHMS FOR DETECTION OF CONGESTION

Mostly, vehicles installed with a variety of on-board sensors generate plenty of messages that yield the issue of channel competition and exhaust the limited available bandwidth. In a VANET, on board units are installed in vehicles to accumulate the outputs of the sensors such as the vehicles speed, fuel consumption and  $CO_2$  emission into a single message and transmit to all vehicles out of which one behaves as a node to process them using clustering technique. Using clustering techniques, the dataset is precisely partitioned into clusters such that the data in each cluster has the same distinguished attribute. The vehicles with the same or nearly same speed are grouped together into a single cluster. The minimum distance between the centers of the clusters decides the closeness between the clusters and ultimately the congestion in a lane.

#### 4.1. K-means Clustering

*K*-means clustering is a partitioning algorithm where *m* objects of a data set *D* are organized into *k* partitions ( $k \le m$ ) where the partitions are represented as a cluster. Here, the objects belong to a cluster are said to be "similar" to each other and "dissimilar" to objects in other clusters in terms of the attributes of the data set [21-22]. In the Centroid-based *K*-means clustering technique the Centroid of a cluster is the center point and is differentiated from data points by Euclidean distance between the two objects (or points)  $x_i$  and  $a_i$ .

#### 4.2. Fuzzy C- means clustering

The Fuzzy C-means (FCM) clustering is an unsupervised clustering algorithm which creates k clusters by taking the data points having a high degree of belongingness to that cluster. The distance from any given data point to the cluster center is expressed as minimum objective function [23-24].  $O(mk^2i)$  is used to calculate the time complexity of the FCM algorithm, here the total number of objects is m, the number of clusters is k and the number of iteration is i.

#### 4.3. Fuzzy K- means clustering

In fuzzy *K*-means clustering, a given group of feature vectors x converted into an improved one through partitioning *N* data points. This process starts with a group of introductory cluster centers and reruns this process till it satisfies a stopping criterion. It is expected that two clusters don't have the same cluster centers. If they are same, then a cluster center comes out of the process to avoid coincidence [25]. Here the fuzzy relationship between a data point and cluster centers is represented by a membership  $p_{ij} \in [0,1]$  value which represents the degree of belongingness of data point  $x_i$  and cluster center  $a_i$ .

# 5. COMPARATIVE ANALYSIS OF THE CLUSTERING APPROACHES

# 5.1. The Data Set

We have used the data set D asshown in Table 1 to characterize the messages with attributes of speed (km/hr), fuel consumption (ml/s) and CO<sub>2</sub>(mg/s) emission which are taken from SUMO Simulator. The messages collected for detection of congestion are generated by the on board units installed in the vehicles. For the detection of traffic congestion, a total of 27 vehicles are taken to form the data set D. These samples are grouped to form different clusters and then are compared.

Table 1. The Data Set					
Attributes 🔶	Speed (Km/Hr)	Fuel Consumption	CO <sub>2</sub> Emission		
Sample Number		(ml/sec)	(mg/sec)		
1	0	1.13	2624.72		
2	1.70	1.37	3180.87		
3	2.54	1.18	2743.50		
4	3.12	1.57	3655.93		
5	2.86	1.49	3460.67		
6	9.30	3.60	8382.62		
7	10.43	3.00	6983.71		
8	11.56	3.86	8971.36		
9	15.64	3.64	8475.07		
10	16.48	3.87	9001.59		
11	17.19	5.39	12536.11		
12	17.33	4.46	10371.44		
13	22.34	7.74	18006.88		
14	21.98	7.60	17680.25		
15	21.73	5.73	13322.30		
16	27.17	2.59	6020.83		
17	26.11	7.75	18035.66		
18	26.44	6.47	15044.55		
19	27.30	2.28	5297.24		
20	5.84	2.40	5584.19		
21	6.76	1.89	4407.81		
22	27.64	3.67	8549.24		
23	19.88	7.02	16330.06		
24	18.22	4.79	11147.94		
25	12.9	4.54	10553.88		
26	6.59	2.74	6385.81		
27	8.26	2.83	6578.98		

# 5.2. Experimental Results and Observations

The K-means Clustering, Fuzzy C-means Clustering and Fuzzy K-means clustering techniques are implemented in Matlab 2015.

# 5.2.1. Implementation of K-Means Clustering

The messages from vehicles in the  $m \times p$  data matrix where m is the number of data messages and p is the number of attributes of those messages are grouped into k clusters. The K-means graph for the vehicle data set (speed, fuel consumption and CO<sub>2</sub> emission) represents three clusters. The graphical representation of the scattered vehicles having three attributes: speed, fuel consumption and CO<sub>2</sub> emission as mentioned in the dataset is shown in the Figure 5(a).

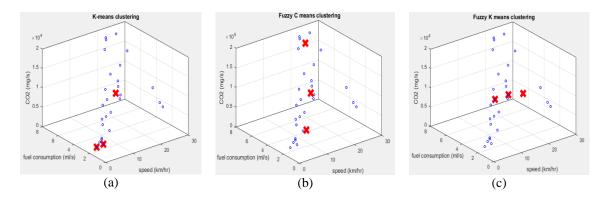


Figure 5. (a) *K*-means clustering of data messages, (b) Fuzzy C-Means clustering of data messages and (c) Fuzzy K-Means clustering of data messages

#### 5.2.2. Implementation of Fuzzy C-Means Clustering

The Fuzzy C-means Clustering (FCM) is used to perform clustering of different messages received from various vehicles at a junction. The function FCM takes the data set from the vehicles and a desired

**D** 889

number of clusters are generated. Figure 5(b) is a three dimensional plot of the three attributes, speed, fuel consumption and CO<sub>2</sub> emission for each of the vehicles and red X marks shows the centers of clusters.

#### 5.3.3 Implementation of Fuzzy K-means Clustering

Sometimes most of the vehicles do not have clear attributes. Hence an intermediary nature in quality and type exists between the vehicles for which a soft division is required. The fuzzy *K*-means (FKM) clustering technique is a best method to work on the above said problem. The Fuzzy *K*-means graph with the vehicle data set (speed, fuel consumption and  $CO_2$  emission) represents three clusters. Figure 5(c) shows a scattered Fuzzy *K*-means graph of vehicle dataset with three attributes: speed, fuel consumption and  $CO_2$  emission.

#### **5.3.4 Experimental Results**

The efficiency of FCM, *K*-means and Fuzzy *K*-means techniques are tested in Matlab [26]. All computations are performed on HP Intel(R) Core(TM) i3-4000M CPU @ 2.40GHz with 4GB RAM. In our experiment, the data are the messages coming from vehicles moving towards a congested area. 27 messages are received with attributes of speed, fuel consumption and  $CO_2$  emission. That means the data set is consisting of 27 data points. The average computing time (in seconds) for *K*-means, FCM and Fuzzy *K*-means are listed in the Table 2 with 50 numbers of iterations. It is observed from the Table 2 that *K*-means clustering technique consumes less average computing time than FCM and Fuzzy *K*-means clustering technique. The distances between the cluster centers for different techniques are listed in Table 3, Table 4 and Table 5. From our results, it is shown that all the distances measured between the clusters in Fuzzy *K*-means are very less. That means Fuzzy *K*-means technique is better to use to detect road congestion.

The comparison between these techniques in terms of average computing time is shown in the Figure 6. By seeing these comparison results, it may be safely stated that the cluster formation speed of K-means clustering algorithm is more than FCM algorithm and Fuzzy K-means. But in FCM and Fuzzy K-means techniques each point has a probability of belonging to each cluster rather than belonging to just one cluster as in K-means. Because of this only we can prefer fuzzy techniques to our problem of traffic congestion detection as vehicles are dynamic in nature.

		the Data Set, D		
Methods		k (	Number of clusters)	
	2	3	4	5
k means	0.0664	0.0682	0.1046	0.1201
FCM	0.0671	0.0691	0.1787	0.2169
Fuzzy k-means	0.5410	0.6497	0.6558	0.6838

Table 2. The Average Computing Time (in Seconds) for K-Means, FCM and Fuzzy K-Means Using

Table 3. The Distance etween the Centres in *K*-Means Technique

Centres of	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
clusters					
Cluster1	0	556.2	118.8	7432.8	836
Cluster2	556.2	0	437.4	6876.7	279.8
Cluster3	118.8	437.4	0	7314.1	717.2
Cluster4	7432.8	6876.7	7314.1	0	6596.9
Cluster5	836	279.8	717.2	6596.9	0

Table 4.	The Distance between	the Centres in	FCM Technique	
Cluster1	Cluster	Cluster?	Cluster/	

Centres of clusters	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0	5720.0	8478.1	3766.5	2846.8
Cluster2	5720.0	0	14198	9486.5	2873.2
Cluster3	8478.1	14198	0	4711.6	11325
Cluster4	3766.5	9486.5	4711.6	0	6613.4
Cluster5	2846.8	2873.2	11325	6613.4	0

Ta	Table 5. The Distance Between the Centres in Fuzzy K-Means Technique					
Centres of clusters	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	
Cluster1	0	60.3805	0.0269	30.5972	0.0095	
Cluster2	60.3805	0	60.4070	30.7033	60.3898	
Cluster3	0.0269	60.4070	0	30.6240	0.0174	
Cluster4	30.5972	30.7033	30.6240	0	30.6066	
Cluster5	0.0095	60.3898	0.0174	30.6066	0	

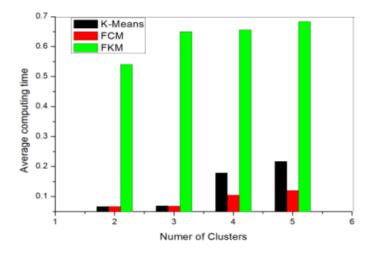


Figure 6. Comparison between number of clusters and average computation time for different techniques

#### 6. CONCLUSION

From our results, we conclude that FCM and Fuzzy *K*-means produce close results to *K*-means clustering in the process of detection of congestion on a busy road but still they require more execution time than *K*-means clustering because of the involvement of fuzzy measures calculations in the algorithm. And, out of these fuzzy techniques, Fuzzy *K*-means is better as the distance between cluster centers is lesser than FCM technique giving the idea that congestion is more prominently detected in Fuzzy *K*-means.

#### REFERENCE

- [1] Dhanya P. M, Ananth S. Efficient Traffic Congestion Detection Method in Vanet. International Journal for Technological Research in Engineering. 2013; 1(3).
- [2] Joshi A, Mishra D. Review of Traffic Density Analysis Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*. 2015; 4 (7).
- [3] Ahmed H, Pierre S, Quintero A. A Flexible Testbed Architecture for VANET. *Vehicular Communication*. 2017; 9: 15-126.
- [4] Yoo H, Kim D. Repetition-based cooperative broad casting for vehicular ad-hoc networks. *Comput.Commun.* 2011; 34 (5):1870-1882.
- [5] Fußler H, Mauve M, Hartenstein H, Kasemann M, Vollmer D. *Location-based routing for vehicular ad-hoc networks.* Mobile Computing and Communications Review, SIGMOBILE, NewYork, USA, 2002.
- [6] Wei Q, Yang B. Adaptable vehicle detection and speed estimation for changeable urban traffic with anisotropic magneto resistive sensors. *IEEE Sensors Journal*. 2017; 17(7): 2021-2028.
- [7] Sen R, Siriah P, Raman B. RoadSoundSense: Acoustic sensing based road congestion monitoring in developing regions. Sensor, Mesh and Ad Hoc Communications and Networks, SECON'2011, Salt Lake City, Utah, USA, 2011: 125–133.
- [8] Roy A, Gale N, Hong L. Automated traffic surveillance using fusion of Doppler radar and video information. *Mathematical and Computer Modelling*. 2011; 54: 531–543.
- [9] Mandal K, Sen A, Chakraborty A, Roy S, Batabyal S, Bandyopadhyay S. *Road traffic congestion monitoring and measurement using active rfid and gsm technology*. IEEE Intelligent Transportation Systems, ITSC'11, Washington DC, USA. 2011: 1375–1379.
- [10] Jiang G, Niu S, Chang A, Meng Z, Zhang C. Automatic traffic congestion identification method of expressway based on gain amplifier theory. Advanced Computer Control, ICACC'10, Shenyang, China, 2010;2: 648–651.

- [11] Zhu W, Barth M. Vehicle trajectory-based road type and congestion recognition using wavelet analysis. IEEE Intelligent Transportation Systems Conference, ITSC'06, Paris, France, 2006: 879-884.
- [12] Vaqar S, Basir O. Traffic pattern detection in a partially deployed vehicular ad hoc network of vehicles. *IEEE Wireless Communications*. 2009; 16 (6): 40–46.
- [13] Liu C, Chigan C. Structure-less message aggregation (SLMA): Reliably and efficiently improve information precision and certainty for VANETs. IEEE GLOBECOM'10, Miami, FL, USA. 2010: 1–6.
- [14] Bauza R, Gozalvez J. Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. *Journal of Network and Computer Applications*. 2013; 36: 1295-1307.
- [15] EmreCelebi M, Hassan Kingravi A, Patricio Vela A.A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*. 2013;40 (1): 200-210.
- [16] Kumar Vishal, Mishra Shailendra, Chand Narottam. Applications of VANETs: Present & Future. Communications and Network, 2013; 5: 12-15.
- [17] Terroso-Saenz F, Valdes-Vela M, Sotomayor-Martinez C, Toledo-Moreo R, Gomez-Skarmeta A. A cooperative approach to traffic congestion detection with complex event processing and VANET. *IEEE Transactions on Intelligent Transportation Systems*. 2012; 13 (2): 914–929.
- [18] https://www.semanticscholar.org/paper/Reliable-Routing-in-Vehicular-Adhoc-Network-Using-(-Rao-Karyemsetty /a0d63df28b68bb9ebd46e0311a14dc1958c09173/figure/0
- [19] Bauza R, Gozalvez J, Sanchez-Soriano J. Road traffic congestion detection through cooperative vehicle-to-vehicle communications. IEEE Conference on Local Computer Networks, LCN'10, Denver, Colorado, USA, 2010: 606– 612.
- [20] https://www.openstreetmap.org/#map=16/20.3147/85.8161
- [21] Ghosh S, Dubey S K. Comparative Analysis of K-Means and Fuzzy CMeansAlgorithms. *International Journal of Advanced Computer Science and Applications*. 2013; 4(4): 35-39.
- [22] Har-Peled S, Sadri B. How fast is the k-means Method. ACMSIAM Symposium on Discrete Algorithms, Vancouver. 2005.
- [23] Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. New York: Plenum Press, 1981.
- [24] Yong Y, Chongxun Z, Pan L. A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding. *Measurement Science Review*. 2004; 4(1).
- [25] Chang Chih-Tang, Lai Jim Z C, Jeng Mu-Der. A Fuzzy K-means Clustering Algorithm using Cluster Center Displacement. *Journal of information Science and Engineering*. 2011; 27: 995-1009.
- [26] Math works http://www.mathworks.com