IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Cramming Identification through Spatiotemporal Data

Srikanth Lakumarapu[1*], Rashmi Agarwal[2]

[1] Department of Computer Science and Engineering, Madhav University, Rajasthan, India
[2] Department of Computer Science and Engineering, Madhav University, Rajasthan, India

[*]Corresponding Author: lsrikanth.ap@gmail.com, Tel.: +91- 94414-56336

Available online at: www.ijcseonline.org

Accepted: 10/Jun/2018, Published: 30/Jun/2018

***Abstract* -** Indian roads carry almost 90 per cent of the country's passenger traffic and around 65 per cent of its freight. In India sales of automobiles and movement of freight by roads is growing at a rapid rate along with the increasing rate of traffic. Geo-spatial temporal data with geographical information explodes as the development of GPS-devices using mobiles. To dig out the video patterns behind the video data efficiently in huge spatial temporal data, using an OPTICS algorithm on gpsdata from the traffic video footage introduced. Through above cluster types provides number of cluster groups with identifying the video information from the existing archives video footages. This work deals with the clustering of the video data from the large geospatial temporal traffic videos using TRAFFICOPTICS algorithm organizing archives through information. In-order to identify the vehicles from the video footages in large traffic network that to identify the congestion through the spatiotemporal data mining method.

Keywords – GPS, GEO, OPTICS, SPATIAL, GEO-SPATIAL, spatiotemporal, clustering

I. Introduction

India has the second largest road network across the world at 5.4 million km. This road network transports more than 60 per cent of all goods in the country and 85 per cent of India's total passenger traffic. Road transportation has gradually increased over the years with the improvement in connectivity between cities, towns and villages in the country. Cluster analysis is a most important method for database mining. Clustering can considered the most important *unsupervised learning* problem; so, as every other problem of this kind, it deals with finding a *structure* in a collection of unlabelled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way". [4]

In spatial data mining, analysts use geographical or spatial information to produce business intelligence or other results. This requires specific techniques and resources to get the geographical data into relevant and useful formats. The core goal of a spatial data-mining project is to distinguish the information in order to build real, actionable patterns to present, excluding things like statistical coincidence, randomized spatial modelling or irrelevant results. One way analysts may do this is by combing through data looking for "same-object" or "object-equivalent" models to provide accurate comparisons of different geographic locations. Temporal dimension is described about time period attached to the data which expresses when it was valid or stored in the database. [1]

A temporal database stores data relating to time instances. It offers temporal data types and stores information relating to past, present and future time. The temporal database has two major notions or attributes.

1. Valid time.

2. Transaction time.

More specifically the temporal aspects usually include valid time and transaction time. These attributes can combined to form bitemporal data.

- **Valid time** is the time period during which a fact is true in the real world.
- **Transaction time** is the time period during which a fact stored in the database was known.
- **Bitemporal data** combines both Valid and Transaction Time.

A spatiotemporal database is a database that manages both space and time information. Traditional methods in spatiotemporal database always incur a data search in a large range while users query about trajectories in a particular spatiotemporal region in Web based applications. Non-spatial measures (also called attribute or characteristic data) used to characterize non-spatial features of objects. Major requirements of Spatio Temporal Clustering technique are :Discovery of clusters with arbitrary shape, Ability to handle high dimensional data, Ability to deal with spatial, non-

© 2018, IJCSE All Rights Reserved

spatial and temporal attributes, Independent of input data order, Good Interpretability and usability, Ability to deal with nested clusters moreover is should be scalable too.[3, 15]

Cluster-ordering contains media backup information which is equivalent to the density-based clustering corresponding to a broad range of parameter settings. It is a versatile basis for both automatic and interactive cluster analysis. We show how to automatically and efficiently extract not only 'traditional' clustering information (e.g. representative points, arbitrary shaped clusters), but also the intrinsic clustering structure. For medium sized data sets, the cluster-ordering can be represented graphically and for very large data sets, we introduce an appropriate visualization technique. Both are suitable for interactive exploration of the intrinsic clustering structure offering additional insights into the distribution and correlation of the data. Ordering points to identify the clustering structure (OPTICS) is an algorithm for finding density-based clusters in spatial data. Its basic idea is similar to DBSCAN, but it addresses one of DBSCAN's major weaknesses: the problem of detecting meaningful clusters in data of varying density. In order to do so, the points of the database are (linearly) ordered such that points, which are spatially closest, become neighbours in the ordering. Additionally, a special distance is stored for each point that represents the density that needs to be accepted for a cluster in order to have both points belong to the same cluster. OPTICS represented as a dendrogram. [2]

Section I deals with the Introduction to Spatiotemporal mining and its algorithm. Section II deals with the related works of Spatiotemporal mining. Section III deals with the existing implementation of DBSCAN and Optics algorithm. Our proposal found in Section IV is to cluster the large spatial temporal traffic cctv video footage to identify the vehicle from the large video footage database using OPTICS algorithm. Using the OPTICS algorithm the proposal system are design and the result are to be generate as graph found to in Section V. Section VI provide an conclusion idea over identifying the traffic using the algorithms.

II. Related Work

A. Methodology on Spatial mining

Spatial Data Mining is a new and rapidly developing area of data mining concerned with the identification of interesting spatial patterns from data stored in spatial databases and geographic information systems. Geographic Information Systems (GIS) enable capturing, storing, analysing and managing data and associated attributes which are spatially referenced to the Earth. GIS are used in various areas such as environmental impact assessment, urban planning, cartography, criminology, traffic analysis, etc. Here, we have undertaken the process of traffic analysis based on information available in GIS. With increasing traffic volumes on urban roads, particularly in the large cities, existing roundabouts and priority-controlled junctions are being replaced with traffic signals to address capacity, efficiency and safety issues or to provide better amenity for pedestrians and cyclists. Daily traffic jams reflect the fact that the capacities of the road network are not satisfied or even exceeded. It is therefore crucial to investigate new technologies and alternative methods of traffic management to reduce congestion without increasing road space [1].

Spatial Data Mining has been shown to significantly help improving traffic safety, and has been used in many traffic related works [4]. In traffic data, traffic density patterns on hourly, weekly, and monthly scales can be obtained from density plots. Such plots identify traffic peaks, and can be of help to traffic specialists in planning routes and safety measures, as well as to individual drivers. Traffic control systems for large traffic networks have attracted much attention, recently. One challenge of traffic controlling is the prediction of the traffic. Traffic flow prediction, i.e. conditionally, forecasting the traffic conditions in the network, given prevailing traffic conditions, the predicted traffic demands, and the candidate control scenarios. What we need are efficient and effective methods that are able to estimate the traffic for any point of time in the future. Traffic predictions are very important as they enable to detect potential traffic jam spots. Based on the information provided from a traffic prediction system we could initiate certain traffic control methods to avoid the traffic jams. One of the most important applications of traffic control systems is the control of road network traffic. In particular at rush-hour when the risk of the occurrence of traffic jams is very high traffic control systems would be very important.

B. System on Temporal mining

Temporal data mining is often carried out with one of the following intentions; predictions, classification, clustering, search and pattern identification. Early predictive models assumed a linear combination of the sample values. But later, neural networks and AI modelling were employed to develop non-linear temporal modelling. Clustering in time series data provides an opportunity to understand it at a higher level of abstraction by studying the characteristics of the grouped data. Temporal data clustering has numerous applications ranging from understanding protein structure to learning and characterizing financial transactions. The interest for pattern identification in large time series data is comparatively recent and was originated from data mining itself. Pattern recognition from data mining perspective. The sequential pattern analysis was then used to identify the features after which they are input into classifiers (like Nave Bayes Classifier) for data processing. The spatial dimension describes whether the objects considered are associated to a fixed location (e.g., the information collected by sensors fixed to the ground) or they can move, i.e., their location is dynamic and can change in time.

© 2018, IJCSE All Rights Reserved

Temporal analysis of the traffic accidents is performed using methods of short time series analysis. A time series is a very common type of data, and there are many available algorithms and methods for time series analysis. In this work, time series clustering was used to identify groups of similar time series obtained for all Slovenian municipalities. We analyze two types of time series: the number of accidents by month and the number of accidents over the 11 years included in the database. Both types of time series considered are of the type referred to as short time series (consisting of 4-20 measurements), where a different approach to clustering is required than the approaches used for clustering of long time series. Short time series clustering has recently become popular due to its practical applications in biology (DNA microarray analysis) and economics. An approach to qualitative clustering of short time series of traffic accidents is one of the contributions of this paper. [14]

C. Procedure on Spatiotemporal mining

Explosive growth in geospatial and temporal data as well as the emergence of new technologies emphasize the need for automated discovery of spatiotemporal knowledge. Spatiotemporal data mining studies the process of discovering interesting and previously unknown, but potentially useful patterns from large spatial and spatiotemporal databases. Figure 1 shows the process of spatiotemporal data mining. Given input spatiotemporal data, the first step is often preprocessing to correct noise, errors, and missing data and exploratory space-time analysis to understand the underlying spatiotemporal distributions. Then, an appropriate spatiotemporal data mining algorithm is selected to run on the preprocessed data, and produce output patterns. [3][11]

Common output pattern families include spatiotemporal outliers, associations and tele-couplings, predictive models, partitions and summarization, hotspots, as well as change patterns. Spatiotemporal data mining algorithms often have statistical foundations and integrate scalable computational techniques. Output patterns are post-processed and then interpreted by domain scientists to find novel insights and refine data mining algorithms when needed. [10][6]

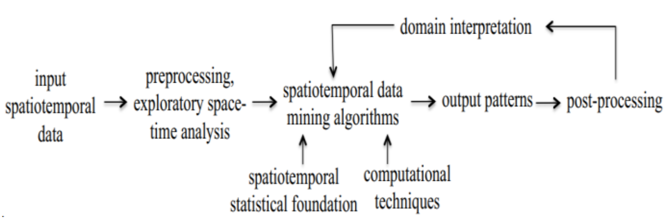


Figure 1. The process of spatiotemporal data mining

D. Existing Spatiotemporal clustering algorithm

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)DBSCAN is a density-based algorithm which is robust to outliers. It requires two parameters

- Eps – Maximum radius of the neighborhood
- MinPts – Minimum number of points in the Eps-neighborhood of a point

Algorithm: DBSCAN

Input:

- D : Dataset
- Eps: radius for neighbouring points
- minPts: minimum points within Eps distance for a point to become a core-point.

Output:

Clusters based on spatial closeness.

1. DBSCAN(D, eps, minPts)
2. C=0
3. For each point p in D
4. If p is visited
5. Continue next point
6. Mark p as visited
7. neighborPts = regionQuery(P,eps)
8. If sizeof(NeighborPts) < minPts
9. Mark p as NOISE
10. Else
11. C = next cluster
12. expandCluster(p, NeighborPts, C, eps, minPts)

Algorithm: expandCluster(p, NeighborPts, C, eps, minPts)

1. Add p to cluster C
2. For each point p' in NeighborPts
3. If p' is not visited
4. Mark p' as visited
5. NeighborPts' = regionQuery(p', eps)
6. If sizeof(NeighborPts') >= minPts
7. NeighborPts = NeighborPts joined with NeighborPts'
8. If p' is not yet member of any cluster
9. Add p' to cluster C

Algorithm: regionQuery(p, eps)

1. Return all points within p's eps-neighbourhood

E. Monitoring Traffic via Spatiotemporal Mining

Research on traffic incidents in transportation science has focused on driving behaviors and physical models. For example, cellular transmission modeling (CTM) has been applied to simulate the formation and dissipation of traffic jams at the microscopic level. This model considers driving

© 2018, IJCSE All Rights Reserved

behaviors, such as lane changing, acceleration, and deceleration. Similarly, car-following models describe the processes by which drivers follow each other in the traffic stream to better simulate the congestion and dissipation caused by traffic incidents. These microscopic models focus on typical road structures, such as intersections, freeways, and rectangular grid networks, to achieve precise results. [7]

In contrast, macroscopic network traffic simulation models derived from the LWR kinematic wave theory, have recently been proposed to simulate and predict various traffic behaviors, including incidents on large road networks. Despite the comprehensive results obtained using these methods, the long computation times have made these techniques non-optimal for implementation within TIM. Moreover, the spatial transferability of these microscopic methods is limited, and they cannot be applied to real road networks. In particular, methods focusing on secondary incident identification also consider predicting the spatiotemporal influence of incidents. These approaches mostly utilize statistical and physical algorithms to calculate the influence scope of primary incidents, therefore, leading to a better detection of secondary incidents.

For example, applied a spatiotemporal speed evolution method to imprint the dynamic of the influence scope, taking advantage of detector data. Similarly, methods using Bayesian learning approach, deterministic queuing diagrams and regression models can also determine the extent of an incident. These approaches can determine the spatiotemporal influence of incidents; however, they rely on historical data and their implementation is limited within freeways. The research of closely related to ours; they proposed a dynamic approach to determine the spatiotemporal thresholds of incidents in a large-scale road network. Their estimation was based on shockwave theory and validated to have over 70% accuracy. On the other hand, although this approach considered queuing on freeways or arterial roads, more specific behaviors at the intersection remained disregarded.[8]

As the nature of transportation issues is dynamic and spatial, a macroscopic data-driven method cannot avoid large training sets. Static analytical methods of GIS ignore the dynamic nature of transportation, and microscopic transportation models have limited spatial transferability.

In other words, the relevance of geospatial information for transport modeling is significant, but not yet adequately considered in most cases, which has even been unsettled in recent years. Thus, transportation models should be integrated with spatiotemporal GIS analysis techniques to accommodate the dynamics of traffic incidents. With the ability to predict traffic effects dynamically, specific TIM can be implemented to mitigate the loss caused by traffic incidents. [13]

F. Congestion identification using cluster identification

Traffic congestion has been characterised as a clustering phenomenon, from both theoretical and empirical aspects. Theoretical aspects focus on how a cluster of vehicles moving at low speeds at critical traffic densities could cause congestion, which is referred to as a‘ phantom jam’

- These theoretical findings are developed mainly on motorways where traffic flow is uninterrupted for long distances
- Investigation of such traffic congestions in an urban road network still remains as a research challenge, due to regular and irregular disturbances (e.g. traffic lights, pedestrian crossings) in traffic flow, as well as the higher rate of interaction between drivers. Therefore, empirical research has investigated how spatio-temporal clustering can be used to detect congestion patterns in an urban road network
- Examples of such research include clustering links having increased LJTs or increased traffic density
- These pieces of empirical evidence mostly focus on congestion detection and their application to NRC detection requires further research. Nevertheless, these research findings illustrate the effectiveness of spatio-temporal clustering on congestion detection.

Traffic congestion, as most other spatial phenomenon, is dynamic. Therefore, a coherent combination of both spatial and temporal aspects is required when clustering spatio-temporal data.The wide usage of the term ‘clustering’, however, results in different meanings attached to the term cluster and there are two main strategies for clustering.

The first strategy uses a similarity function to group similar observations into clusters

- The aim of a similarity based clustering algorithm is to maximise the similarity of observations within a cluster (i.e. minimise the distance between its observations) and maximise the dissimilarity between different clusters (i.e. maximise the distance between clusters).
- Some of the representative algorithms in this strategy include k-means and DBSCAN.
- The second strategy applies a statistical significance testing procedure to detect statistically significant clusters
- Methods developed within this strategy use the statistical properties of the data and aim to detect ‘anomalous’ clusters.

The detected clusters often correspond to unusual increases in the observed 44 phenomenon, like a region having unusually high rates of disease or crime.

© 2018, IJCSE All Rights Reserved

The representative methods of this strategy include spatial scan statistics and space-time scan statistics

- An important point should be highlighted regarding the linkage between these two strategies
- It has been suggested that hypothesis testing should be conducted to validate the detected clusters from the first strategy
- This, indeed, is very similar to the idea behind the second strategy, as it involves a significance testing process to determine whether the clusters are observed by chance alone or not. However, it has been observed that such hypothesis testing has not been applied in the case of most clustering algorithms based on the first strategy.

Therefore, it is reasonable to distinguish the two strategies of clustering.

G. Traffic congestion in Automobiles

Traffic congestion is a condition on transport networks that occurs as use increases, and is characterized by slower speeds, longer trip times, and increased vehicular queueing. When traffic demand is great enough that the interaction between vehicles slows the speed of the traffic stream, this results in some congestion. While congestion is a possibility for any mode of transportation, this article will focus on automobile congestion on public roads. As demand approaches the capacity of a road (or of the intersections along the road), extreme traffic congestion sets in. When vehicles are fully stopped for periods of time, this is colloquially known as a traffic jam or traffic snarl-up. Traffic congestion can lead to drivers becoming frustrated and engaging in road rage.

Traffic congestion occurs when a volume of traffic or modal split generates demand for space greater than the available street capacity; this point is commonly termed saturation. There are a number of specific circumstances which cause or aggravate congestion; most of them reduce the capacity of a road at a given point or over a certain length, or increase the number of vehicles required for a given volume of people or goods. About half of U.S. traffic congestion is recurring, and is attributed to sheer weight of traffic; most of the rest is attributed to traffic incidents, road work and weather events. Traffic research still cannot fully predict under which conditions a "traffic jam" (as opposed to heavy, but smoothly flowing traffic) may suddenly occur. It has been found that individual incidents (such as accidents or even a single car braking heavily in a previously smooth flow) may cause ripple effects (a cascading failure) which then spread out and create a sustained traffic jam when, otherwise, normal flow might have continued for some time longer.

At present, IT is entering all industry verticals with the latest technologies such as IoT, connected devices, machine learning, deep packet assessment, and augmented and virtual reality. As these technologies continue to advance, integration of IT technology in traffic management systems will lead to the development of advanced systems with far more capabilities and enhanced performance, drastically enhancing the safety of individuals. Furthermore, market players now provide traffic management software integrated with machine learning and deep packet examination technologies, and have witnessed heavy demand for these software, thereby leading to healthy adoption.

Use of so-called intelligent transportation system, which guide traffic[16]:

- Traffic reporting, via radio, GPS and mobile apps, to advise road users
- Variable message signs installed along the roadway, to advise road users
- Navigation systems, possibly linked up to automatic traffic reporting
- Traffic counters permanently installed, to provide real-time traffic counts
- Convergence indexing road traffic monitoring, to provide information on the use of highway on-ramps
- Automated highway systems, a future idea which could reduce the safe interval between cars (required for braking in emergencies) and increase highway capacity by as much as 100% while increasing travel speeds.
- Parking guidance and information systems providing dynamic advice to motorists about free parking
- Active Traffic Management system opens up UK motorway hard shoulder as an extra traffic lane; it uses CCTV and VMS to control and monitor the traffic's use of the extra lane.

III. Existing System

The Density-based notion is a common approach for clustering. Density-based clustering algorithms are based on the idea that objects which form a dense region should be grouped together into one cluster. They use a fixed threshold value to determine dense regions. They search for regions of high density in a feature space that are separated by regions of lower density.

Density-based clustering algorithms such as DBSCAN, OPTICS, DENCLUE, CURD are to some extent capable of clustering databases. One drawback of these algorithms is that they capture only certain kinds of noise points when clusters of different densities exist. Furthermore, they are adequate if the clusters are distant from each other, but not satisfactory when clusters are adjacent to each other.

© 2018, IJCSE All Rights Reserved

In chosen DBSCAN algorithm, because it has the ability in discovering clusters with arbitrary shape such as linear, concave, oval, etc. Furthermore, in contrast to some clustering algorithms, it does not require the predetermination of the number of clusters. DBSCAN has been proven in its ability of processing very large databases. In the literature, DBSCAN algorithm was used in many studies. For example, the other popular density based algorithm OPTICS (Ordering Points To Identify the Clustering Structure) is based on the concepts of DBSCAN algorithm and identifies nested clusters and the structure of clusters. Incremental DBSCAN algorithm is also based on the clustering algorithm DBSCAN and is used for incremental updates of a clustering after insertion of a new object to the database and deletion of an existing object from the database.

The OPTICS algorithm generates the augmented cluster-ordering consisting of the ordering of the points, the reachability-values and the core-values. However, for the following interactive and automatic analysis techniques only the ordering and the reachability-values are needed. To simplify the notation, we specify them formally:

Let *DB* be a database containing *n* points. The OPTICS algorithm generates an ordering of the points $o:\{1..n\} \rightarrow DB$ and corresponding reachability-values $r:\{1..n\} \rightarrow R\geq 0$. The visual techniques presented below fall into two main categories. First, methods to get a general overview of the data. These are useful for gaining a high-level understanding of the way the data is structured. It is important to see most or even all of the data at once, making pixel-oriented visualizations the method of choice. Second, once the general structure is understood, the user is interested in zooming into the most interesting looking subsets. In the corresponding detailed view, single (small or large) clusters are being analyzed and their relationships examined. Here it is important to show the maximum amount of information which can easily be understood. Thus, we present different techniques for these two different tasks.

Because the detailed technique is a direct graphical representation of the cluster-ordering first and then continue with the high-level technique. A totally different set of requirements is posed for the automatic techniques. They are used to generate the intrinsic cluster structure automatically for further (automatic) processing steps.

The cluster-ordering of a data set can be represented and understood graphically. In principle, one can *see* the clustering structure of a data set if the reachability-distance values are plotted for each object in the cluster-ordering. The generating distance influences the number of clustering levels, which can be seen in the reachability-plot. The smaller we choose the value, the more objects may have an UNDEFINED reachability-distance. Therefore, we may not see clusters of lower density, i.e. clusters where the core objects are core objects only for distances larger than ε. In order to identify the clusters contained in the database, we need a notion of "clusters" based on the results of the OPTICS algorithm. As we have seen above, the reachability value of a point corresponds to the distance of this point to the set of its predecessors. From this (and the way OPTICS chooses the order) we know that clusters are dents in the reachability-plot.

IV. Proposed System

Planned system deals with identifying the density of clusters based on spatiotemporal data as metadata of files along with the gps (global positioning system) data to identify the congestion in automobile traffic. Mostly commonly used congestion monitoring in the road were monitored through manual counting and toll based. Next level was to monitor the vehicle monitored via gps system can be possible by the vehicle owner only. Government of India came into next era of providing 13 digit number to the sim card used in GPS like IoT devices.

Our proposal is to incorporate the mutual data like gps which will be a good specific example for spatiotemporal data as an one part with Closed-Circuit Television (CCTV) footage which deployed in the city can be taken as a another part of identifying the traffic congestion caused by commercial vehicle in a particular timing can be detected. By using these two part data on the TRAFFICOPTICS algorithm techniques effective and efficient congestion controlling mechanism can be resulted. Approaching of TRAFFICOPTICS works with hierarchical technique, but instead of maintaining a set of known facts on the GPS monitoring of vehicles. But, unprocessed cluster members, a priority queue (e.g. using an indexed heap) is used.

Algorithm for TRAFFICOPTICS on GPSDATA as follows,

TRAFFICOPTICS (DB, eps, MinPts)

1. Each vehicle v in LiveMovementVechicles
    a. Each point p (longitude, latitude, timestamp, v) of GPSDATA
        i. p.reachabilityDistance = UNDEFINED
    b. Each unprocessed point p of GPSDATA
        i. N = getNeighborVechicle(p, eps)
        ii. mark p as processed based on timestamp
        iii. output p to the ordered list of timestamp
        iv. if (core-distance(p, eps, Minpts) != UNDEFINED)
            1. Seeds = empty priority queue
            2. RECOLLECT(N, p, Seeds, eps, Minpts)
            3. for each next q in Seeds were the neighborVehicleTimestamp
                a. N' = getNeighbors(q, eps)
                b. mark q as processed
                c. output q to the ordered list
                d. if (core-distance(q, eps, Minpts) != UNDEFINED)
                    i. update(N', q, Seeds, eps, Minpts)

© 2018, IJCSE All Rights Reserved

In RECOLLECT (), the priority queue Seeds is updated with the ε - neighborhood of p and q respectively based in the timestamp.

Algorithm for Recollecting of Vehicle with respect to timestamp as follows,

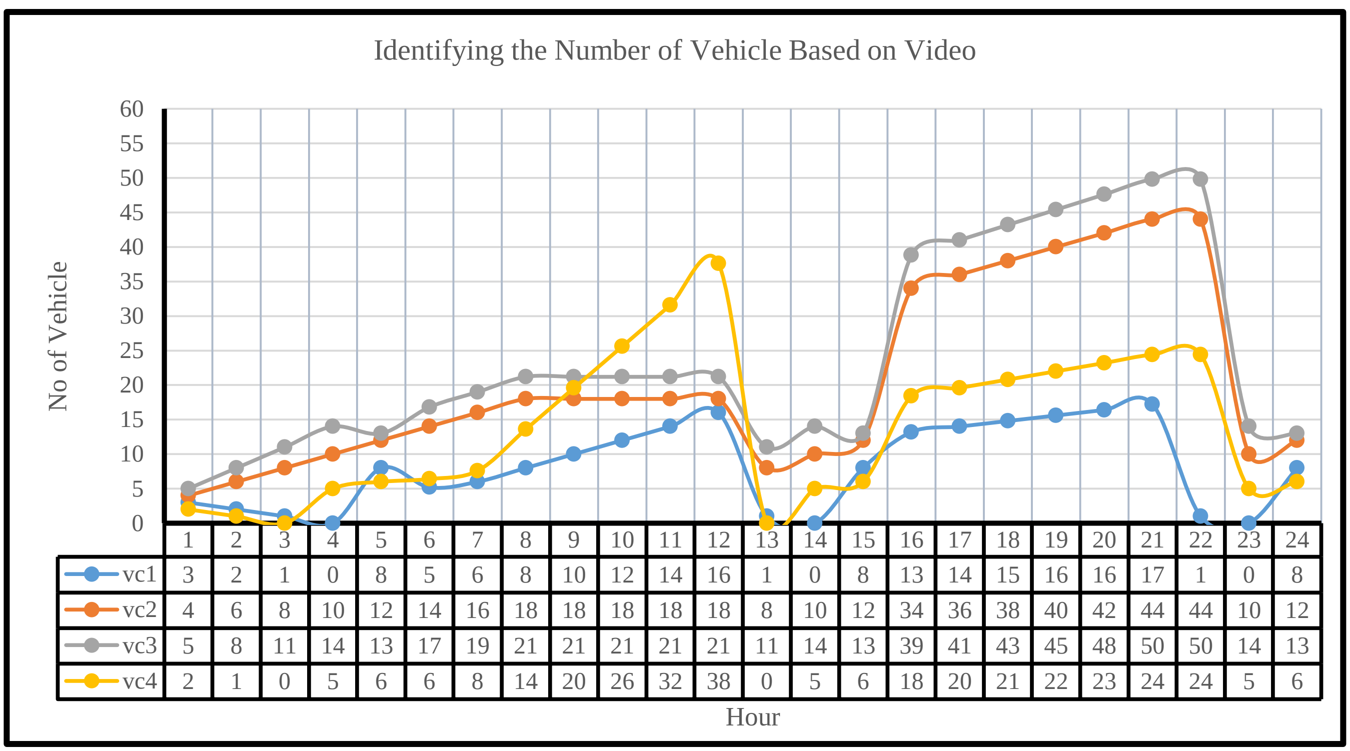
RECOLLECT (N, p, Seeds, eps, Minpts)

1. coredist = core-distance(p, eps, MinPts)
2. Each o in N was a NeighborVehicleTimestamp
   a. if (o is not processed)
      i. new-reach-dist = max(coredist, dist(p,o))
      ii. if (o.reachability-distance == UNDEFINED)
         1. o.reachability-distance = new-reach-dist
         2. Seeds.insert(o, new-reach-dist)
      iii. else if (new-reach-dist < o.reachability-distance)
         1. o.reachability-distance = new-reach-dist
         2. Seeds.move-up(o, new-reach-dist)
3. Each c in GLS (GeoLocationSurvilance) found in Spatial are taken
   a. NumberofDirections N = c has number of camera directions on road
   b. Each DirectionofVideo, vd in N
      i. Extract frame, f of vd(timestamp)
      ii. In every f trace the vehicles
         1. Count the vehicle, v in the f
         2. Map the each count with direction refers as density and closeness among the vehicles found

TRAFFICOPTICS hence outputs the points in a particular ordering, annotated with their smallest reachability distance in the original algorithm, the core distance relation to cluster formation. As a part of first level of identifying the particular vehicle at location and time aggregation data and consecutively the next level is to monitor the surveillance footages sorted in temporal manner for past days on the particular time to detect the congestion. Applying the cluster formation on the CCTV footage recording to extract the particular timestamp frame. These, aggregation proves the theory of identifying the cluster and make a suggestion to provide or look for an alternative solutions analogy in traffic management console.

V. Results

In this proposed above algorithm consist of two levels to identify the traffic in roadways. For the sampling experiment in this analysis consist of a four road junction in a city was taken live footage and gpsdata of commercial vehicles are considered. Based on the TRAFFICOPTICS algorithm the number of vehicles were identified in four video camera VC1, VC2, VC3, VC4. These camera footages consider to find the number of vehicle identified based on the video of each camera, clusters are constructed on hourly basis. In accordance with that the table holds the detail of number of vehicle identified based on the video footage of each camera in each junction and their comparative statements are represented in the graph of the Figure 2.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
vc1	3	2	1	0	8	5	6	8	10	12	14	16	1	0	8	13	14	15	16	16	17	1	0	8
vc2	4	6	8	10	12	14	16	18	18	18	18	18	8	10	12	34	36	38	40	42	44	44	10	12
vc3	5	8	11	14	13	17	19	21	21	21	21	21	11	14	13	39	41	43	45	48	50	50	14	13
vc4	2	1	0	5	6	6	8	14	20	26	32	38	0	5	6	18	20	21	22	23	24	24	5	6

Figure 2: Identifying the Number of Vehicle based on Video

© 2018, IJCSE All Rights Reserved

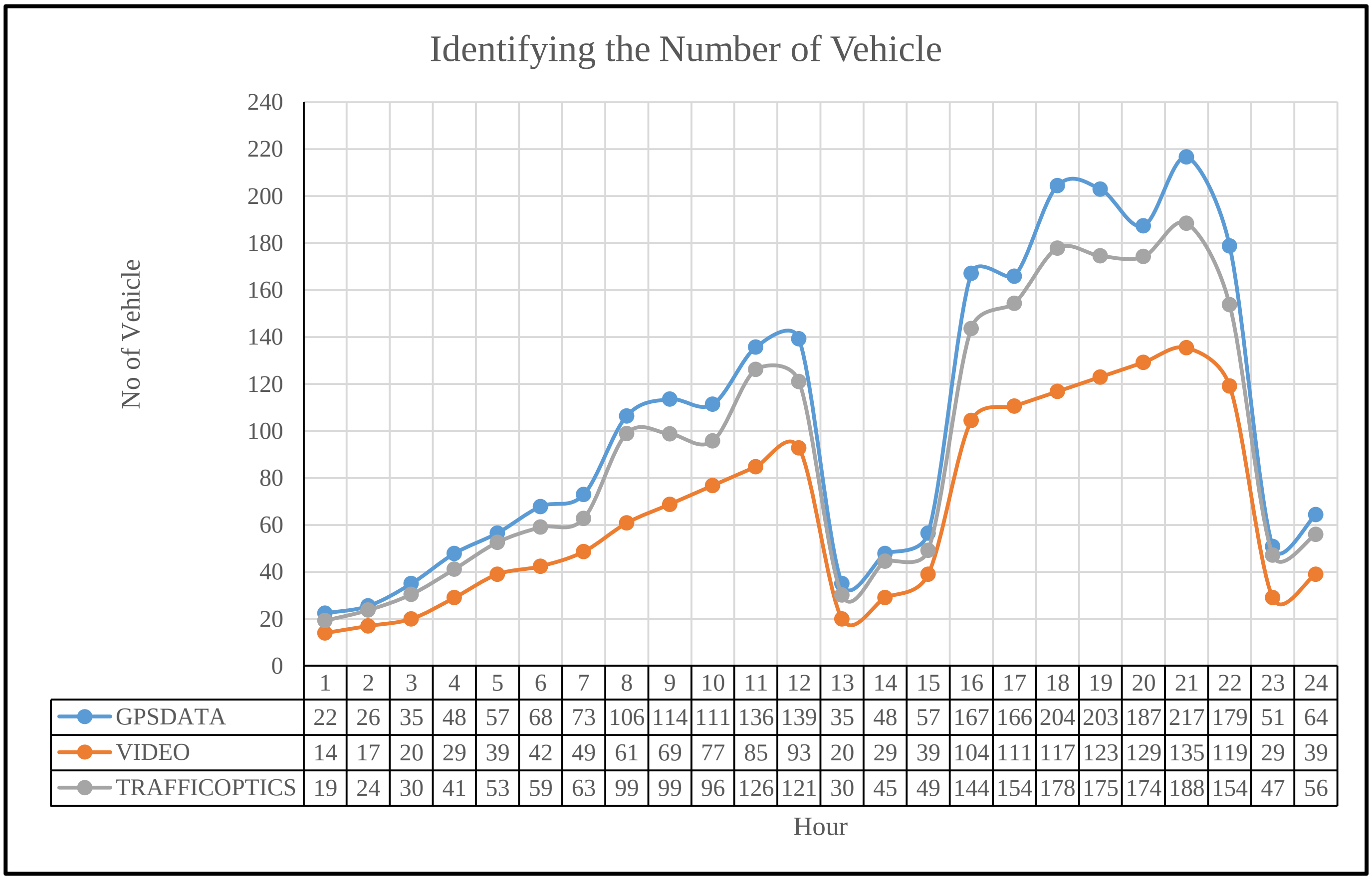


Figure 3: Identifying the Number of Vehicle based on Video, GPSDATA and Combing video & GPS (TRAFFICOPTICS)

Table with number of vehicle identified in each video camera footage stored along with the GPSDATA based on the identified vehicle on hourly basis are noted. From the GPSDATA and video camera footages data about the number of vehicles identified on hourly basis are processed under the TRAFFICOPTICS are generated are compared with individual GPSDATA and VIDEO alongside of proposed algorithm to show the comparison shown in Figure 3.

VI. Conclusion

Spatiotemporal information of automobiles are found huge due to Internet of Things as trendy era to percept the real world values. Through our proposed work, able to identify the number of vehicles from the video footages and along with GPSDATA to identify the congestion through the cluster formation based on spatiotemporal. The vehicle identification from the Video footage and the GPSDATA through TRAFFICOPTICS algorithm was efficient cluster density estimation than comparable to the generic GPSDATA and number of vehicle detection from video footage.

References

[1]. K.P. Agrawal, Sanjay Garg, Shashikant Sharma, and Pinkal Patel, “Development and validation of OPTICS based spatio-temporal clustering technique”, Inf. Sci. 369, C (November 2016), pp.388-401, 2016.

[2]. Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, Jörg Sander , OPTICS: Ordering Points To Identify the Clustering Proc. ACM SIGMOD’99 Int. Conf. on Management of Data, Philadelphia PA, pp.1-12, 1999.

[3]. Derya Birant, Alp Kut, “ST-DBSCAN: An algorithm for clustering spatial–temporal data”, Data & Knowledge Engineering 60, pp. 208–221, 2007.

[4]. Latika Sharma, Nitu Mehta,” Data Mining Techniques: A Tool For Knowledge Management System In Agriculture”, International Journal of Scientific & Technology Research, VOLUME 1, ISSUE 5, JUNE 2012.

[5]. Manjula Aakunuri, Dr.G.Narasimha, Sudhakar Katherapaka,”Spatial Data Mining: A Recent Survey and New Discussions”, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 2 (4), 1501-1504, ISSN: 0975-9646, 2011.

[6]. Shashi Shekhar, Zhe Jiang, Reem Y. Ali, Emre Eftelioglu, Xun Tang, Venkata M. V. Gunturi and Xun Zhou , “Spatiotemporal Data Mining: A Computational Perspective”, ISPRS International Journal of Geo-Information 2015, *4*, 2306-2338, 2015.

© 2018, IJCSE All Rights Reserved

[7]. Tanvi Jindal, Prasanna Giridhar, Lu-An Tang, Jun Li, Jiawei Han, “Spatiotemporal Periodical Pattern Mining in Traffic Data”, UrbComp’13, Chicago, Illinois, USA, August 11–14, 2013.
[8]. Khamphao Sisaat, Hiroshi Ishii, Masato Terada, Chaxiong Yukonhiatou, Hiroaki Kikuchi, Surin Kittitornkun , “A Spatio-Temporal malware and country clustering algorithm: 2012 IIJ MITF case study”, International Journal of Information Security, Volume 16 Issue 5, Pages 459-473, October 2017.
[9]. Liu, C., Gong, Y., Laflamme, S., Phares, B., & Sarkar, S., “Bridge damage detection using spatiotemporal patterns extracted from dense sensor network”, Measurement Science and Technology, 28(1), 014011, 2017.
[10]. Wu, L., Liu, C., Huang, T., Sharma, A., & Sarkar, S., “Traffic sensor health monitoring using spatiotemporal graphical modelling” Proceedings of the 2nd ACM SIGKDD Workshop on Machine Learning for Prognostics and Health Management, August 13-17, Halifax, Nova Scotia, Canada, 2017.
[11]. Jiang, Z., Liu, C., Akintayo, A., Henze, G., & Sarkar, S.,” Energy prediction using spatiotemporal pattern networks” Applied Energy, 206, 1022-1039, 2017.
[12]. Xiaojing Wu, Raul Zurita-Millab, Menno-Jan Kraakb, Emma Izquierdo-Verdiguier, “CLUSTERING-BASED APPROACHES TO THE EXPLORATION OF SPATIO-TEMPORAL DATA”, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W7, 2017
ISPRS Geospatial Week 2017, 18–22 September 2017, Wuhan, China, 1387-1391.
[13]. D.Gokula priya, Dr.C.Saravanabhavan, K.Suvitha "Detecting Congestion Patterns in Spatio Temporal Traffic Data using Frequent Pattern Mining", SSRG International Journal of Computer Science and Engineering (SSRG - IJCSE), V4 (10), ISSN: 2348 – 8387, 11-14 October 2017.
[14]. Nada Lavra, Domen Jesenovec, Nejc Trdin, and Neža Mramor Kosta,“Mining Spatio-temporal Data of TrafficAccidents and Spatial Pattern Visualization “, Metodološki zvezki, Vol. 5, No. 1, pp. 45-63, 2008.
[15]. Pawan S. Wasnik, S.D.Khamitkar, Parag Bhalchandra, S. N. Lokhande, Ajit S. Adte, “An Observation of Different Algorithmic Technique of Association Rule and Clustering”, International Journal of Scientific Research in Computer Science and Engineering, Vol.06, Issue.01, pp.28-30, 2018.
[16]. S. Kshirsagar, M. Jadhav, C. Bhagwat, “Real Time Road Inspection System with Distance Intimation Technique”, International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.1, pp.16-19, 2017.

© 2018, IJCSE All Rights Reserved