Analysis of GPS Data Interpolation Methods for Driver's Behavior Detection

Tin Lai Lai Mon
GIS Lab
University of Computer Studies
Yangon, Myanmar
tinlailaimon@ucsy.edu.mm

Thin Lai Lai Thein
GIS Lab and FIS
University of Computer Studies
Yangon, Myanmar
tllthein@ucsy.edu.mm

***Abstract*— Interpolation of GPS points is one of the biggest challenges in the world of today. GPS signals may be lost because of many reasons including signal loss and battery loss of the transceiver. Our research work is intended to make an online detection and alert for driver's behavior. In order to do so, we need to find the current segment of the trajectory of the driving path, first. The segment is formed by joining the GPS points of the movement of the vehicle. If some between GPS points are lost, it is difficult to make a precise segment. Therefore, we need to find missing points between two consecutive points using interpolation methods. In this research work, erroneous GPS points of datasets are removed before doing any interpolation. After that, 7 different interpolation methods are analyzed and simulated using MATLAB.**

Keywords*— *Interpolation, Driver's behavior, GPS trajectory cleaning

I. Introduction

There are many reasons for happening traffic accidents. It may be because of the bad planning, design, construction and operation of roads, lack of maintenance of roads, lack of traffic rules and regulations, bad weather conditions, insufficient light along roads. The deadliest factor is human error. This includes unawareness of traffic rules and roadway condition; lack of driving skills; poor judgment; failure to interact and adjust to prevailing roadway conditions; and most importantly, aggressive driving [1]. Driver's behavior detection is urgently needed in the world so that accidents due to drivers' careless driving behavior will be reduced.

To detect whether driver is driving aggressively or not precisely, timestamp, current trajectory path, velocity and acceleration of driver's vehicle must be accurately recorded. To get current trajectory path correctly, each and every GPS point must be recorded precisely too. Nevertheless, sometimes GPS points may be lost. Consequently, there will be gaps between some GPS points and it will be difficult or not accurate to make a trajectory path. In order to solve this problem, it is necessary to find missing segment. Missing segment is a portion of the interest trajectory which should be there according to a given GPS sampling rate and vehicle's movement direction, but is missing.

A simple way to fill those kinds of gaps is to make a link between two closest GPS points which are before and after the gap. But it must be considered on the GPS sampling rate too. GPS sampling rate is the period at which GPS point will be recorded and sent to the server.

In [2], Agzam Idrissov and Mario A. Nascimento tried to fill the gap of GPS points by linking the closest two points before and after the perceived gap. In their research work, Distance Time Warping (DTW) method was used for interpolation. There also is a disadvantage using DTW on the quality of the distance measure because trajectories are compared using piecewise. But it had the problems for ignorcurring the GPS sampling rate and it could occur missing sub-trajectory.

In [3], Milan Horemuz and Johan Vium Andersson wrote about polynomial interpolation of GPS satellite. In their works, researchers described an algorithm for polynomial interpolation of GPS satellite coordinates and it is implemented in MATLAB for real-time processing. In the paper, the position and velocity of GPS satellites from both broadcast and precise ephemerides are shown. They tested their algorithm with different polynomial orders and with different timestamps. Nevertheless, no exact solution is found for interpolation of GPS points problem.

In [4], Marcio Geovani Jasinski and Fabiano Baldo showed a method for identification of aggressive driver behavior. They showed how to clean GPS points, how to remove stop points and how to cluster points for noise removing. But, there is a lack of description about interpolation.

In this research work, different kinds of interpolation methods are applied on finding the missing GPS points. Moreover, removing inaccurate points before interpolation is also done so that the accuracy of the interpolation gets higher. Our research methodology is described in section II. In section II, it is also described about the interpolation methods, removing unnecessary points before interpolation. Experimental results are shown in section III. In this section, we did our research based on Pyay Road, Myanmar using 7 different interpolation methods. Our research work is concluded in the last section.

II. Research Methodology

Overall system design of the research work is shown in Fig. 1. First of all, GPS points will be collected along the way of the vehicle. After that, missing GPS points will be found using interpolation. And then, GPS stop points will also be detected and removed. After cleaning the GPS points, data will be enriched by considering the other external effects such as road speed limits, current weather conditions and nearby important places. In this paper, it will be emphasized only on solving the interpolation problem.

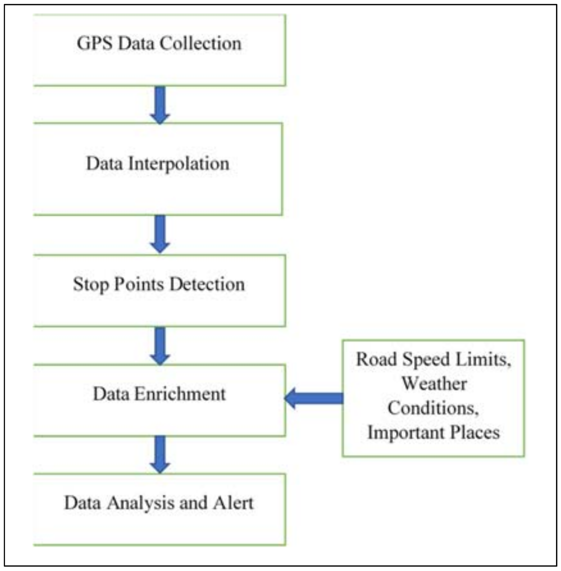

Fig. 1. Overall System Design

A. Interpolation

Interpolation is finding some unknown information from a given set of known information. In our case, interpolation is finding missing GPS points among recorded GPS points. There are different kinds of interpolant methods and the most useful methods are

- Linear Interpolation,
- Nearest Neighbor Interpolation,
- Next Neighbor Interpolation,
- Previous Neighbor Interpolation,
- Shape-preserving Piecewise Cubic Interpolation,
- Akima Cubic Hermite Interpolation and
- Spline Interpolation.

Linear interpolation tries to find a match for a given linear polynomial between each previous and next data points for curves, or between sets of three points for surfaces.

Neighbor interpolation methods set the value of an interpolated point to the value of the nearest or next or previous data point accordingly.

Cubic spline interpolations work for fitting a different cubic polynomial between each previous and next data points for curves, or between sets of three points for surfaces. Shape-preserving is also called as Piecewise Cubic Hermite Interpolation (PCHIP) and this method maintains monotonicity and the shape of the data and it can be used for curves only. Biharmonic interpolation is used for 2-D or 3-D scattered data. Thin-plate spline interpolation fits smooth surfaces that also extrapolate well and it is used only for surfaces.

In order to choose the required interpolation method, it is necessary to consider based on the characteristics of the data for fitting, the smoothness of the curve, calculation time, post-fit analysis requirements, and so on. The linear and nearest neighbor methods have lower calculation time, but they are not enough to be smooth. The cubic spline and shape-preserving and biharmonic methods need much more processing time, but they give the results with smooth.

Suppose that we have two consecutive points, P1 and P2, on a segment with timestamps t1 and t2 respectively. Let I and T be the threshold values of the interpolation and segmentation of trajectory. If the timestamp between t_2 and t_1 is larger than the interpolation threshold, I, it is considered that this missing segment is complete interpolation. However, if the difference between t2 and t1 is not only larger than the interpolation threshold value but also the trajectory breaking threshold, that segment is not needed to interpolate. But, the segment is necessary to break up into separate trajectories. Breaking up the segments into separate trajectories is also necessary for clustering.

Segmentation of trajectory condition occurs when GPS signal is lost for a limited amount of time (suppose that limited time is 5 minutes) between two consecutive GPS points, this trajectory is needed to separate into two sub trajectories. Such kind of trajectory segmentation is performed before GPS stop points detection process. Therefore, our research work is that if there is a missing segment on the trajectory, it will be filled in with generated and evenly distributed points. Our interpolation is performed according to the previous point P_k and next P_k points of the missing point.

Suppose that P is a set of points along the trajectory path and P_{pre} and P_{next} are the endpoints of the missing trajectory segment. If we know the distance between those two points, the number of trajectory subsegments can be estimated and it is symbolled as n. The number of subsegments is needed for even distribution of points across the whole missing trajectory segment:

$$n = \frac{2kDist(P_{pre}, P_{next})}{\sum_{j=pre-k}^{pre-1} Dist(P_j, P_{j+1}) + \sum_{j=next}^{next+k-1} Dist(P_j, P_{j+1})}$$

Where Dist(P_{pre},P_{next}) is the Euclidean distance between P_{pre} and P_{next}. As the points before P_{pre} and after P_{next} are known points, it can be easily calculated the distance between consecutive points P_j and P_{j+1} before P_{pre} and after P_{next}.

Then, the distance between two consecutive generated points P_i and P_{i+1} (where pre<i<next) to fill in a missing segment is defined as:

$$Dist(P_i, P_{i+1}) = \frac{Dist(P_{pre}, P_{next})}{n}$$

Basically, the distance between k consecutive previous (next) points of endpoints of the missing point is taken. The sample points are shown in Fig. 2.

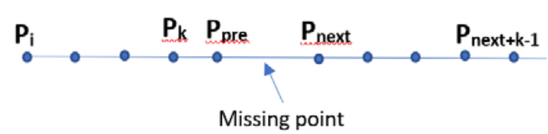

Fig. 2. Sample points

Then, N-1 points are created from P_{pre} towards P_{next} according to calculated distance $Dist(P_i, P_{i+1})$. When all required points are generated, a timestamp is added for each point based on the number of segments and on the time difference between two endpoints of the missing segment. This is necessary for grouping criterion of clustering algorithms.

B. *Removing Inaccurate GPS Points before Interpolation*

In this work, before doing the interpolation process, inaccurate GPS points are removed which can be found in almost all the datasets. Those inaccurate GPS points will make an effect on the accuracy of interpolation. Those points may be like having the same GPS points with different timestamps which may also be stop points. Although they have their respective timestamp, it will affect the accuracy of distance between GPS points too. Moreover, existing stop detection algorithms still have problems in detecting such gatherings because they depend on the minimum threshold stop time of the vehicle and the choice of correct minimum threshold stop time is also a big challenge.

Moreover, there are null speeds in some GPS points that were randomly changing their location with time. It is assumed that this is because of the inaccuracy of GPS receivers (sometimes GPS points can be falsely identified when an object remains at the same point) and those points are also necessary to eliminate.

Finally, when the raw GPS points are segmented into different trajectories, some trajectories are found to be erroneous for other steps of processing. Therefore, trajectories that did not match with predefined threshold value on the minimum number of pints per trajectory are removed (in this research work threshold value is set to 10 points).

III. Experimental Results

In this research work, 50 GPS points along the Pyay Road from A.D. bus stop to Hanthawaddy Roundabout are collected, first. The distance between those two places is 5 km long. Along this path, some GPS points are lost and those points will be found by interpolation. The sampling rate is 1 second, that is GPS point will be recorded for each and every 1 second. Recorded points of Pyay Road are shown in Fig. 3.

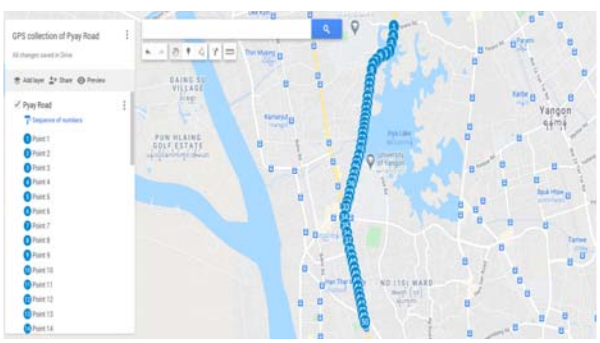

Fig. 3. 50 GPS points along the Pyay Road (from A.D to Hanthawaddy Roundabout)

Firt of all, we tried to interpolate the first 7 GPS points with some missing points (they are intentionally removed). It is shown in Fig. 4.

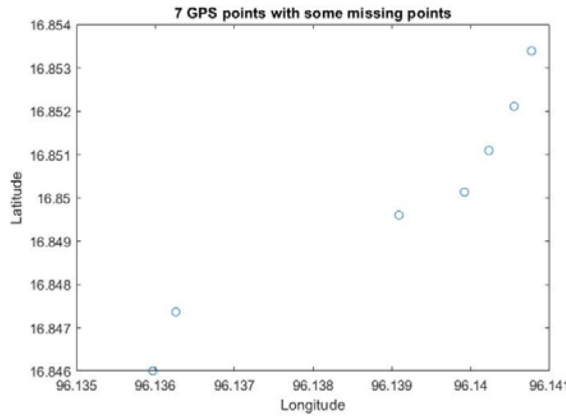

Fig. 4. 7 GPS points with some missing points

After interpolation, 3 GPS points are found between P5 and P9 and they are shown in Fig. 5.

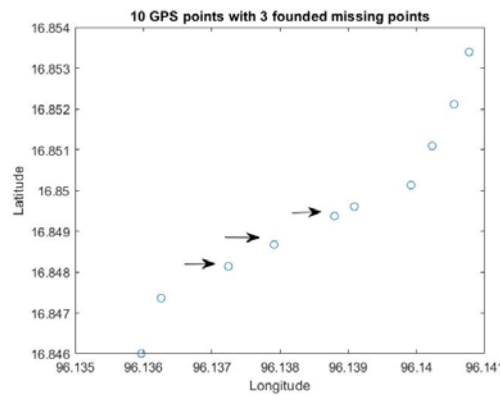

Fig. 5. 10 GPS points with 3 founded missing points

After finding the missing points, they are compared with the real GPS points and the comparison result is shown in Fig. 6. It can be seen that there is no error in most of the points. But, there is a little bit Longitude value error in some points and it is necessary to solve this problem.

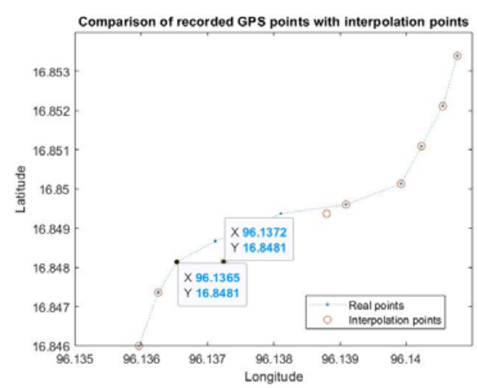

Fig. 6. Comparison of recorded GPS points with interpolation points

In order to solve the inaccuracy problem, 7 interpolation methods are applied and compared. After testing the interpolation with 10 GPS points, the experiment is done using 50 GPS points from AD to Hanthawaddy. First of all 50 recorded GPS points are plotted on the coordinates and it is shown in Fig. 7.

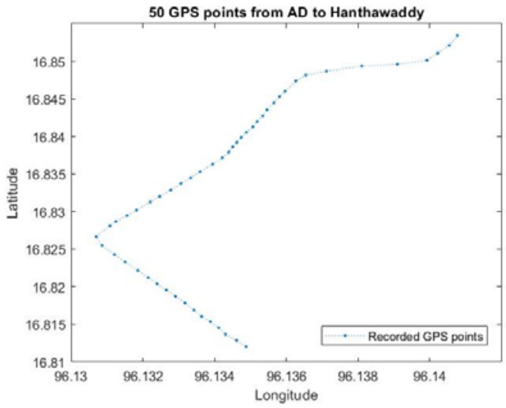

Fig. 7. Recorded 50 GPS points

After that some points are removed (it is assumed that they are missing points) from the recorded points and they are assumed to be missing points. Recorded GPS points with some missing points is shown in Fig. 8. There are 4 missing points located between P10 and P12, P20 and P22, P30 and P32, P40 and P42.

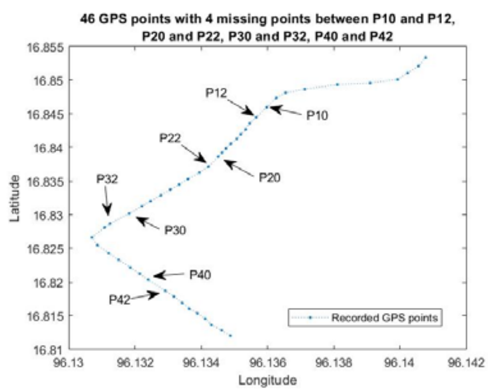

Fig. 8. Recorded GPS points with 4 missing points

First of all, 4 points are interpolated using linear interpolation method. In this method the missing point is found using the values at and the results are shown in Fig. 9. At the minimum, 2 points are required, more memory utilization and processing time are demanded than nearest neighbor method. According to the results, it is obviously seen that all the interpolated points are almost the same as the recorded points.

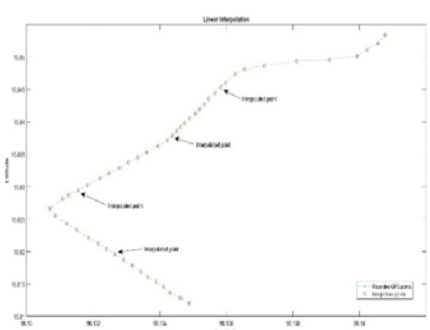

Fig. 9. Linear Interpolation

Results by nearest neighbor interpolation is shown in Fig. 10. This method is based on the value of the query point which is located at the nearest sample grid point. This method needs 2 points at the minimum, modest memory utilization and it gives the results with fastest processing time. As shown in the figure, it can be seen that interpolated points are not exactly the same as the recorded points and there are some errors.

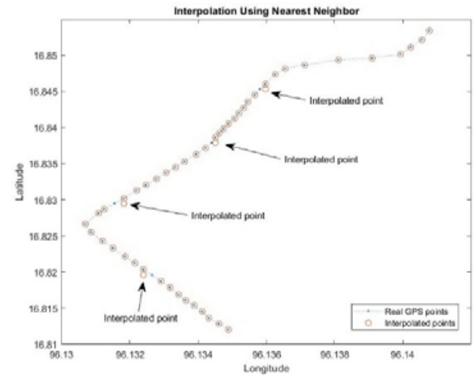

Fig. 10. Nearest Neighbor Interpolation

Fig. 11 shows the interpolation results of next neighbor interpolation. In this method, the query point is the value at the next sample grid point. At the minimum, 2 points are needed for this method. The requirement for memory and processing time is the same as the nearest neighbor method. The results are erroneous as nearest neighbor interpolation and interpolated missing points are not concentric with the recorded GPS points.

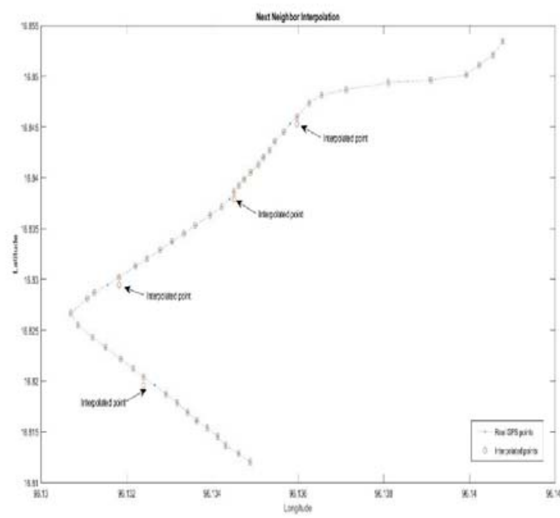

Fig. 11. Next Neighbor Interpolation

Next results of previous neighbor interpolation method are shown in Fig. 12. This interpolation method finds the interpolated point using the value at the previous sample grid point. It requires 2 points at the minimum for interpolation, requires more memory usage and processing time than the linear interpolation. It is more erroneous than the linear interpolation too.

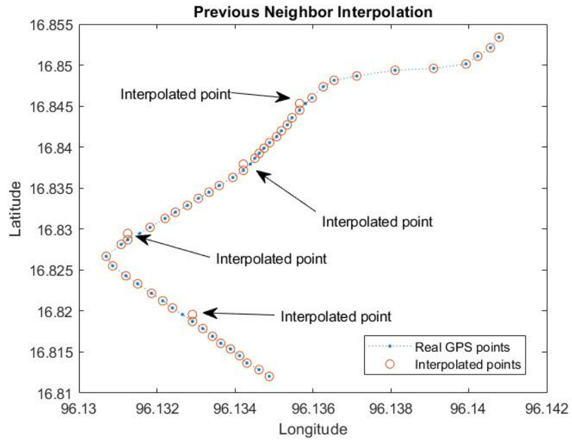

Fig. 12. Previous Neighbor Interpolation

Results of shape-preserving piecewise cubic interpolation are shown in Fig. 13. This interpolation method finds the interpolated pints based on a shape-preserving piecewise cubic interpolation of the values at neighboring grid points. At the minimum 2 points are needed for this interpolation and it requires more memory utilization and processing time than the linear interpolation.

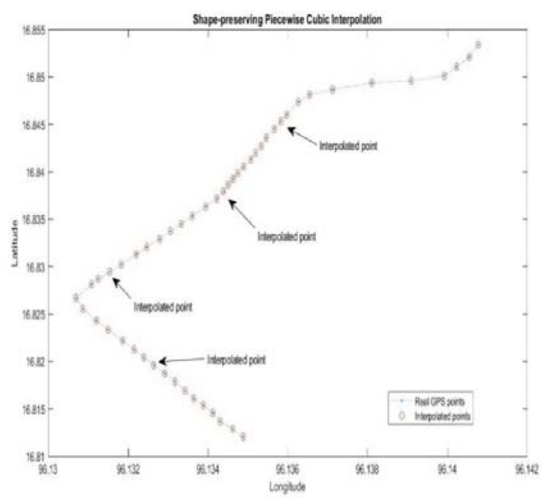

Fig. 13. Shape-preserving Piecewise Cubic Interpolation

Modified Akima cubic Hermite interpolation is applied for the next results and they are shown in Fig. 14. This method uses a piecewise function of polynomials with degree at most three. In this method, the Akima formula is modified for removing overshoots. It also needs not less than 2 points. It has fewer undulations than Spline interpolation, but does not flatten as aggressively as shape-preserving method.

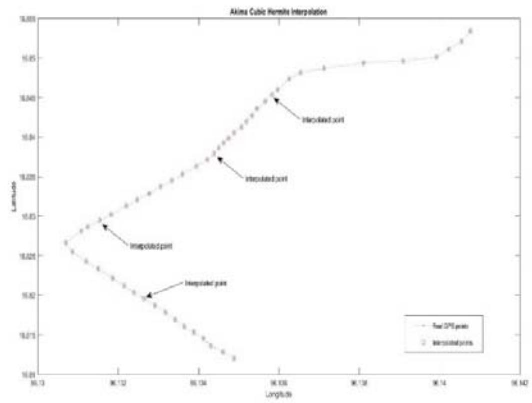

Fig. 14. Akima Cubic Hermite Interpolation

Results shown in Fig. 15 are achieved by using Spline interpolation. This interpolation uses a query point at neighboring grid points in each respective dimension. It requires 4 points at the minimum, more memory utilization and processing time than shape-preserving piecewise cubic interpolation method.

Table I shows the errors of 4 missing points "P11", "P21", "P31" and "P41" for 7 different interpolation methods. From Table I, it is clearly seen that all the error values are under 4 decimal points which can be negligible from the point of view of accuracy. Nevertheless, it is noticed that nearest neighbor interpolation method and next neighbor interpolation method have much more error than the other methods.

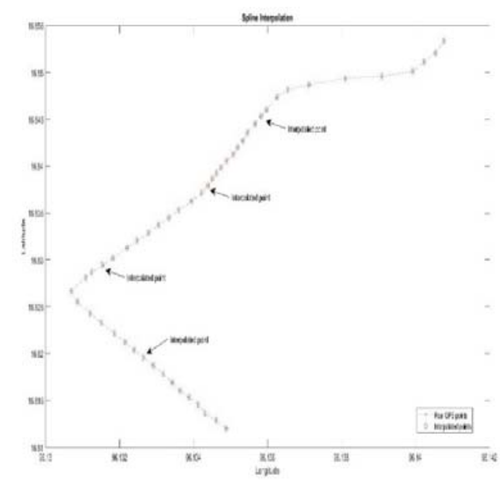

Fig. 15. Spline Interpolation

TABLE I. ERROR TABLE OF DIFFERENT INTERPOLATION TYPES

	P11	P21	P31	P41
Linear	0.0000174	-0.0000301	-0.0000056	-0.0000095
Nearest Neighbor	0.00016	0.00011	0.00027	-0.00026
Next Neighbor	0.00016	0.00011	0.00027	-0.00026
Previous Neighbor	-0.00015	-0.00018	-0.0003	0.00026
Shape-preserving Piecewise Cubic	0.0000177	-0.0000187	-0.000011	-0.0000126

Akima Cubic Hermite	0.0000177	-0.0000196	-0.0000098	-0.0000122
Spline	0.0000155	-0.0000111	-0.0000157	-0.0000161

CONCLUSION

In this paper, we tried to solve the problem of inaccurate and missing GPS points which is related with our research work, driver's behavior detection. It is expected that if we can find the missing GPS points with high accuracy, our trajectory will also have higher accuracy. As a consequence, detection of driver's behavior will be hopefully precise. According to the error value of different interpolation methods, no interpolation method can find the exact missing point. But, all the error values are under 4 decimal points. By comparing the methods to each other, it can be concluded that nearest neighbor interpolation and next neighbor interpolation have much more error than others. Therefore, our system will be more compatible with other 5 methods with high accuracy.

REFERENCES

[1] M. Abojaradeh, B. Jrew, H. Al-Ababsah, A. Al-Talafeeh, "The effect of driver behavior mistakes on traffic safety," in Civil and Environmental Research, Vol. 6, No. 1, 2014, pp. 39-54.

[2] Agzam Idrissov, Mario A. Nascimento, "A trajectory cleaning framework for trajectory clustering," Mobile Data Challenge (by Nokia) Workshop, Newcastle, UK, June 2012.

[3] Milan Horemuz, Johan Vium Andersson, "Polynomial interpolation of GPS satellite coordinates," in Springer-Verlag, January 2006, pp.67–72.

[4] M. G. Jasinki., F. Baldo, "A method to identify aggressive driver behaviour based on enriched GPS data analysis," in The ninth international conference on advanced geographic information systems, Nice, France, 2017, pp. 97–102.