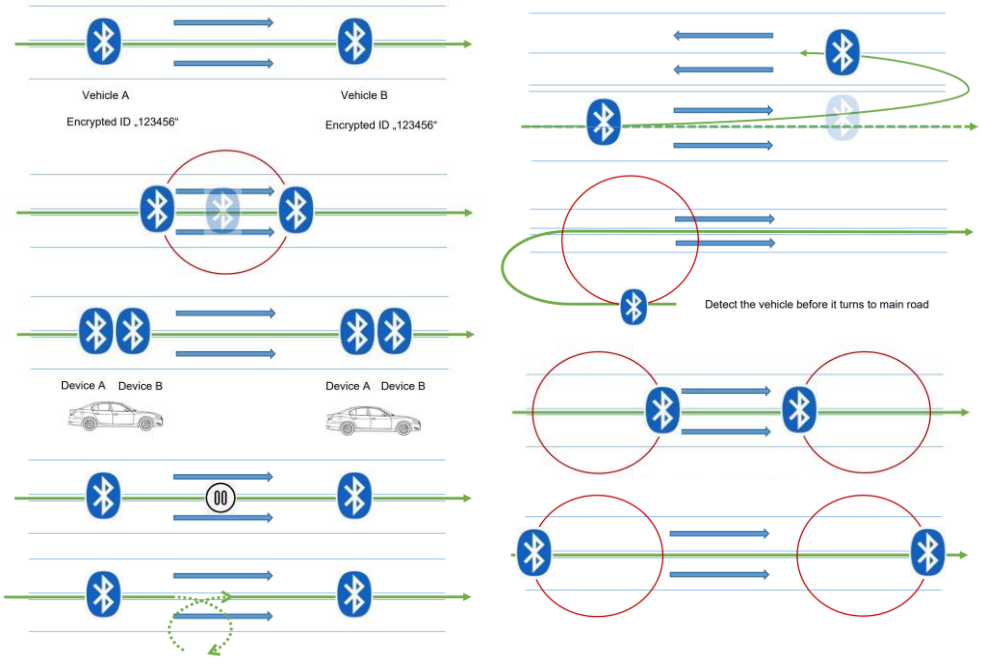
Filter Algorithms for Low-Resolution Travel Time Data

Master‘s Thesis of Youjie Sun

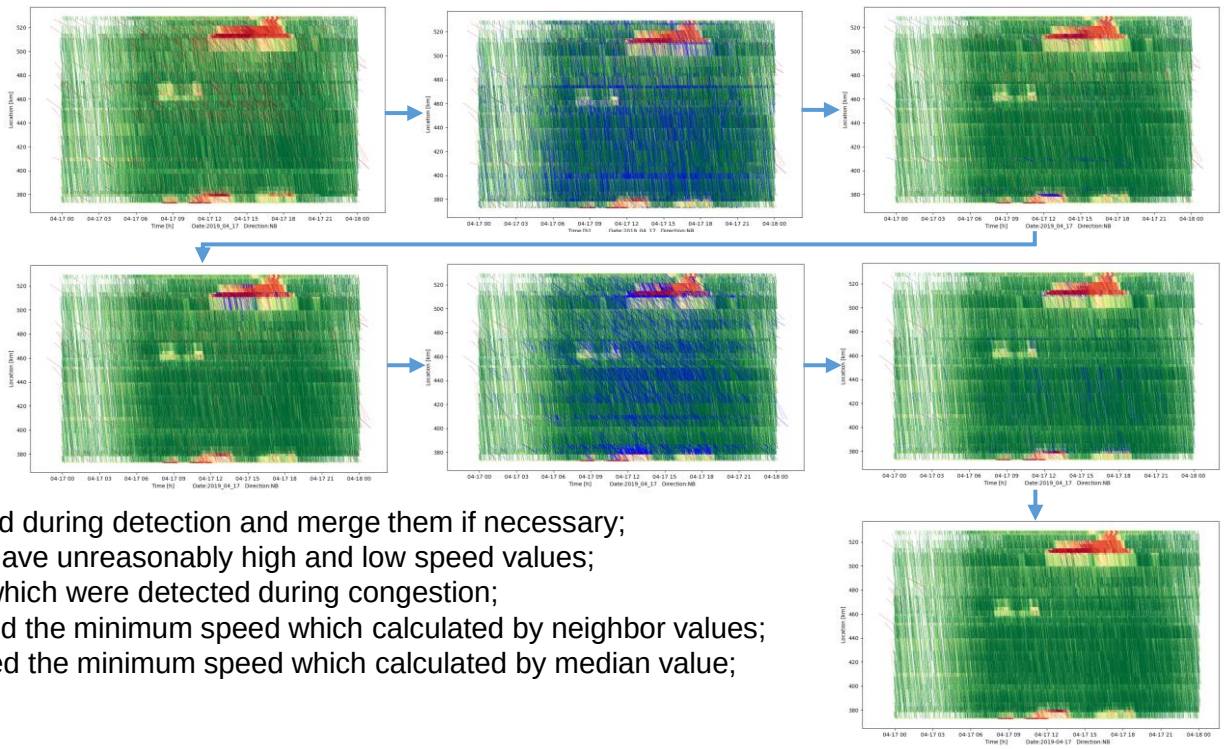
Mentoring:
Dipl.-Math. Lisa Kessler
Dr.-Ing. Matthias Spangler



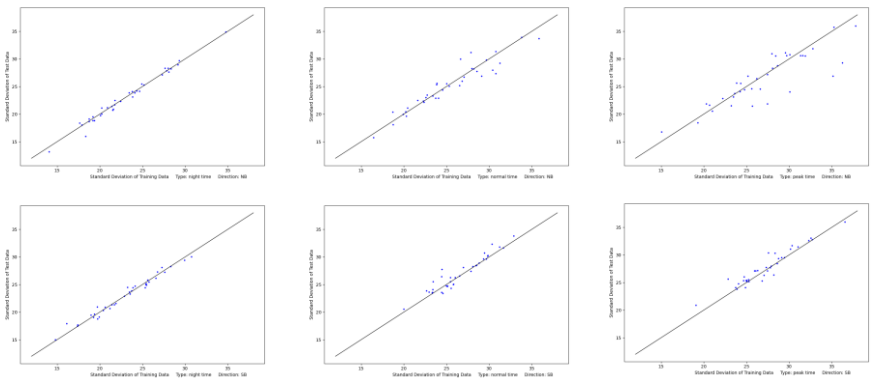
The low-resolution travel time data discussed in this thesis is collected by Bluetooth sensor. Due to the characteristics of Bluetooth sensor, these data usually contain outliers, which may influence the application. The reasons that may cause outliers are listed in the left pictures. After analysis, the outliers have the following characteristics:

- Some data have the same encrypted ID and were detected within a short time interval, and they need to be merged.
- Some outliers were detected within a short time interval but have different encrypted IDs. Besides, they were also detected by the same successive detector within a short time interval. These data need to be merged into one data.
- Some outliers have lower speed value than the other data detected by the same sensor around its point in time and need to be eliminated.
- Some outliers are higher speed than the other vehicles around. Since the high-speed values are not concentrated for application, only high-speed values during congestion need to be eliminated.

In this thesis, 6 filters were used, 2 pre-filters and 4 main filters. The pre-filters are to eliminate outliers which were caused during detection and the main filters eliminate outliers which have unrealistic speed values. After implementation of filters, 8 % of data on average were eliminated, including 5 % of them by pre-filters and 3 % of them by main filters. On the right, the implementation of filters on northbound data of April 17th is showed.



1. Raw data: The raw data of that day;
2. Pre-filters: Eliminate outliers which were caused during detection and merge them if necessary;
3. Fixed threshold filter: Eliminate outliers which have unreasonably high and low speed values;
4. TiDeHiCo filter: Eliminate high-speed outliers which were detected during congestion;
5. TiDeCoNe filter: Eliminate outliers which exceed the minimum speed which calculated by neighbor values;
6. TiDeCoMe filter: Eliminate outliers which exceed the minimum speed which calculated by median value;
7. Filtered data: The data after all filters.



R2 values	night	normal	peak
northbound	0.98	0.89	0.66
southbound	0.97	0.89	0.88

The evaluation of filters needs to choose training and test data set. In this thesis, the training data set is every two days beginning from April 1st, and the test data set is every two days beginning from April 2nd. The outliers usually have speed values that are different from realistic data. Thus, to compare the performance of filters on the training set and the test set, it is possible to compare the standard deviation of all speed value of each segment to see how similar or different the training set and the test set is. Due to data characteristic, the whole data were divided into two directions, northbound and southbound, and three time types, nighttime (21-5), peak time (7-9 and 16-19), and normal time (other time). The picture on the left shows the relationship between standard deviation of training data set and test data set. The black line is y=x. The R^2 value prove the similarity between training data set and test data set. The universality of filters on German freeway A9 has been proved by data from April to May in 2019.