

USING PROBE VEHICLE DATA FOR AUTOMATIC EXTRACTION OF ROAD TRAFFIC PARAMETERS

Maria Alexandra ROMAN (POPESCU) – PhD student, Technical University of Civil Engineering, Faculty of Geodesy

Abstract: Through this paper the author aims to study and find solutions for automatic detection of traffic light position and for automatic calculation of the waiting time at traffic light. The first objective serves mainly the road transportation field, mainly because it removes the need for collaboration with local authorities to establish a national network of traffic lights. The second objective is important not only for companies which are providing navigation solutions, but especially for authorities, institutions, companies operating in road traffic management systems. Real-time dynamic determination of traffic queue length and of waiting time at traffic lights allow the creation of dynamic systems, intelligent and flexible, adapted to actual traffic conditions, and not to generic, theoretical models. Thus, cities can approach the Smart City concept by boosting, efficienting and greening the road transport, promoted in Europe through the Horizon 2020, Smart Cities, Urban Mobility initiative.

Keywords: Intelligent Transportation System, Probe Vehicle Data, mobility

1. Introduction

“we have given up trying to make drivers intelligent, so now we are making the vehicles and roads intelligent!” [1]

It is easy to identify that within any modern society transportation and energy represent the two main pillars for a sustainable economy; the citizens and goods are transported using all transportation means, but mostly using road transportation.

New communications, IT and computer science technologies have been included more and more often in all domains, and transportation made no exception. Intelligent Transportation Systems (ITS) do not refer only to roads and vehicles. These systems began to be highly used in air transportation, maritime transportation and also railway transportation. As time passes, premises such as increased mobility towards globalization, increasing fuel economy to stabilize pollution and passenger safety are becoming critical and stand at the bases of all the research projects carried recently for ITS domain.

When approaching a new domain, the natural starting point is to study the available standards. After only a quick analysis of all ITS' definitions published by the time this paper was written, the conclusion which can be redraw is that ITS can serve large number of functionalities for different purposes; the systems can interact with a single user or vehicle, or can influence an entire transportation network. Yet, no matter the chosen expression within these definitions, it is clear that ITS can increase safety in road transportation, can improve traffic conditions and increase the transportation capacity, can increase the efficiency and productivity of public transport and commercial vehicles and can reduce fuel consumption and CO₂ emissions.

2. ITS Services

ITS services are developed based on international and national taxonomies. These architectures are based on the implementation of ITS systems nationally, but all are defined to provide the

same ITS services defined by ISO 14813-1 standard. Within this paper, the interest is focused specifically on travel information services and traffic management services.

According to [2], Road Traffic Management Services are defined as "*a motion control system for vehicles, passengers and pedestrians within the transport network*". Road traffic management services include many different working groups, such as road traffic, road traffic incidents, requests for public road transport, maintenance works for road transport infrastructure or enforcement of traffic regulations. In fulfilling the tasks of these working groups, road traffic management services are based on operation of two subsystems: traffic monitoring and event / traffic incidents management.

The *Traffic Monitoring Subsystem* is designed to collect and process traffic information used by traffic management and road traffic incidents management service. Methods for collecting traffic indicators are divided into two categories: intrusive methods and non-intrusive methods. Basically, intrusive methods are based on the existence of a data recording device and a sensor placed in the structure of the transport network (road) or in its vicinity: pneumatic road tube, piezoelectric sensors or magnetic loops. Non-intrusive methods are based on observations and measurements made from distance: manually countdown, active or passive infrared technology, passive magnetic sensors, microwave radar, ultrasound technology, video processing technology or location data (Floating Car Data).

The FCD system represents an alternative to existing technologies, being a qualitative and robust data source. The principle of this system is to collect real-time traffic data by locating the vehicles in the transportation network, using GSM / GPRS or GNSS signals.

Currently, the GNSS experimental data tend to be widely used as a real-time data source for many ITS service providers; the main draw-backs of this approach is the limited number of vehicles equipped with such sensors and the highly acquiring costs, compared with GSM / GPRS based systems. Because nowadays most vehicles are equipped with at least one mobile phone, this aspect can be harnessed as anonymous traffic data source. This approach is especially useful in urban areas where positioning is more accurate compared with GNSS systems, due to the smaller distance between GSM antennas.

The main objectives of this subsystem *Event / Traffic Incidents Management Subsystem* are incidents classification and supply response when they actually occur. In this subsystem traffic events are classified in three categories: possible, planned and current. Functions of traffic incidents management subsystem include archiving, management and static data classification of traffic incidents. This subsystem is also handling the provision of interfaces for operating the traffic events and reports for analyzing traffic events.

According to [2], Travel Information Services represent "*the collection and processing of static and dynamic information, in real time, about travel conditions, weather conditions or other isolated incidents, and providing this information to road users.*"

Geographic Referencing represents the ability geo-locating (set of coordinates) and conventional definition (address) of the event traffic that needs to be broadcasted within a travel information service. In terms of geographic referencing, several standards have been issued by this time and all of them are in use currently in different applications.

Traffic Message Channel (TMC) Standard is defined and maintained by Traveler Information Services Association (TISA). As described in [3], TMC is a specific application for the efficient delivery of traffic and travel messages. TMC messages are intended for special decoders which can select messages relevant and can transform them into understandable messages to the end user, such as: digital display, voice synthesis, etc.

Transport Protocol Experts Group (TPEG) Standard is also defined and maintained by TISA. According to [4], TPEG technology is already recognized as providing a "toolbox" and different methods / algorithms for location, necessary for the provision of specific services to different types of transport. Further description of the TPEG standard and encoding algorithm can be consulted in [5], [6] and [7].

Open Location Referencing (OpenLR) Standard is defined and maintained by TomTom Company. OpenLR standard assumes the existence of a digital map, both on the service provider side (that provides coded location-based services) and on the user's terminal memory side (which is used to decode and display the broadcasted event). As described in [8], the geographic referencing method of OpenLR' standard meets the location communication requirements between systems which use different digital maps. In the digital maps, locations are considered as points, lines or polygons objects. As highlighted in [9], the main condition is that both the service provider's map (on which a location is determined and coded) and the user's map to meet the standards required for navigation applications (in terms of accuracy and content).

3. Using FCD for Automatic Extraction of Road Traffic Parameters

The main component in terms of shaping the reality of the transport network is the data FCD (Floating Car Data). Lately though, the term "Probe Vehicle Data" (PVD) or simplest expression, "probe data" is more commonly used for this type of data.

Connected Vehicle Systems, gradual integration of smart phones into the road structure, the improvement of Web 2.0, but especially the progressive emergence of automated vehicles have created a productive ground for an unprecedented gathering of PVD. The impact of technological development has made its mark in a special way in this area, mainly because PVD are now considered low-cost methods to obtain parameters of traffic such as: travel time, speed, volume indicators, and direction of travel.

When building an efficient navigation / travel assistance system there are some parameters which are critical for assuring a high quality service level. Besides a current and accurate digital transport network, the average speed of travel on each segment of the transport network and the delays relative to average speed of travel, for each segment of the transport network must be as updated as possible. Also, lately most of the solutions offered to the end consumer are using real-time information regarding the current traffic conditions.

An important project for developing algorithms for estimating the waiting time at traffic light is carried out by the Shanghai Jiao Tong University, in collaboration with Hong Kong University of Science and Technology (HKUST). Further information about the result of the project can be consulted in [10]. PVD have been successfully used in studies for local estimation of the pollution level. In December 2013, the University of Portland has released the results of the research [11]. Other researches and projects carried out at universities around the world, on the same topic as the one approached in this paper, can be consulted at [12], [13] and [14].

Through this paper the author aims to study and find solutions for **automatic detection of traffic light** and for **automatic calculation of the waiting time at traffic light**.

The first objective serves mainly the road transportation field, because it removes the need for collaboration with local authorities to establish a national network of traffic lights.

The second objective is important not only for companies which are providing navigation solutions, but especially for authorities, institutions, companies operating in road traffic management systems.

Real-time dynamic determination of traffic queue length and of waiting time at traffic lights allow the creation of dynamic systems, intelligent and flexible, adapted to actual traffic

conditions, and not to generic, theoretical models. Thus, cities can approach the Smart City concept by boosting, efficienting and greening the road transport, promoted in Europe through the [15] initiative.

To achieve these two objectives, we require access to a real probe database. From experience acquired up to the time of writing this paper, there are several criteria to be considered in the selection of the data used for this purpose:

- The data collected must come from different categories of users of the road network (taxi companies, transport companies, courier companies, preferably regular participants to urban road traffic);
- To increase the accuracy of determinations made on the basis of these data it is recommended that the interval between two consecutive records do not exceed 30 seconds;
- Depending on the purpose of the study conducted, it is recommended that data be divided into time slots specific local traffic.

There are critical aspects concerning the privacy of the users from which the PVD originate. The Technical Group TC204 / WG16 of ISO has published the international standard "*Basic Principles for Personal Data Protection in evidence Vehicle Information Service*", which deals with personal data protection in the field of ITS. For data used in this research, all the principles described in [16] are met. These parameters do not allow user identification and there is no additional parameter to send to other databases that could identify the user of these services.

3.1 Automatic Detection of Traffic Light Position

The technologies used in developing and applying the algorithm to extract the locations of traffic lights, testing and presentation of results are:

- PostgreSQL database, with PostGIS extension;
- Open Source GIS Software - Quantum GIS;
- ArcGIS Desktop GIS ESRI Software, 60 days trial license;
- Spatial database of the road transportation network for Bucharest [17];
- PVD database, for May and June 2013, for entire Bucharest territory, provided by Arobs Transilvania Company [18];

The data used in this research was collected between 05/01/2013 – 06/31/2013 for the entire area of Bucharest, Romania. The data collected originates from taxi companies, commercial fleet monitored by GPS and civilian users of the transport network.

To present the results of the algorithm for extracting the locations of traffic lights, the Iuliu Maniu Boulevard was chosen, direction of travel from the ring road to the city center. For quantitative evaluation (number of semaphores determined) and quality (positioning accuracy), locations of traffic lights were collected using mobile data collection technologies.

The principle underlying this algorithm is: a vehicle approaching a traffic light displaying red will decelerate. The proposed algorithm tries to identify these deceleration sequences, which will be associated with the related road network segment - called candidate segment. The probable position of the traffic light will be considered at the end of the candidate segment, in the direction of travel of the candidate segment.

In the context of this research, the acceleration / deceleration curves represent sequences of minimum 3 successive records, non-interpolated, for which the speed of travel increases or

decreases constantly. As will be presented in the next section, these curves have an important role in the algorithm of extracting the probable position of traffic lights.

Being at the bases of the algorithm proposed in this paper, firstly it should be studied the dependence of these sequences on the traffic time slot. A brief analysis of the issue mentioned above was carried out for ten users traveling on Iuliu Maniu Boulevard, between crossroads with Virtuții Street and with Grozăvești Street, on May 22, 2013 between 7:30 to 8:00 time slot (five users) and between 12:30 - 13 : 00 time slot (five users). The reason for choosing this boulevard section was that the existing traffic lights (three) are placed at pedestrian crossing and not at crossroads; this way the driving behavior could be studied and considered in the algorithm. The purpose of this analysis is to compare rush hour (morning or evening) traffic versus the free traffic hours to understand which intervals are better to be used in the first iteration of the proposed algorithm for acceleration / deceleration curves extraction. The results are shown in the following graphs.

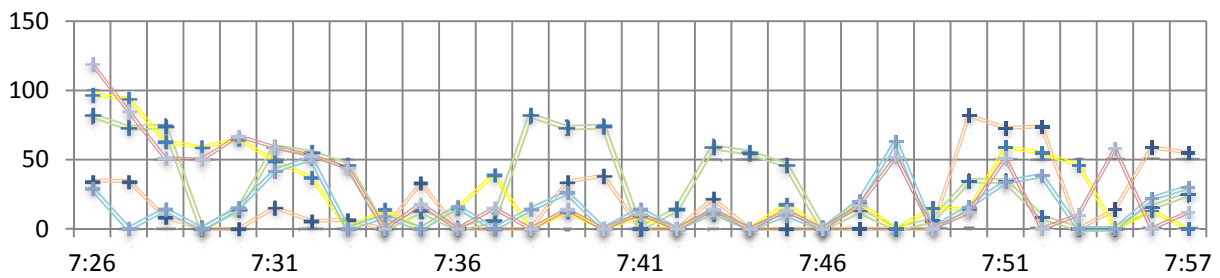


Fig 1 - Acceleration / deceleration curves, 22 May, 2013 between 7:30-8:00 time slot, on Iuliu Maniu Boulevard (between crossroads with Virtuții Street and with Grozăvești Street respectively)

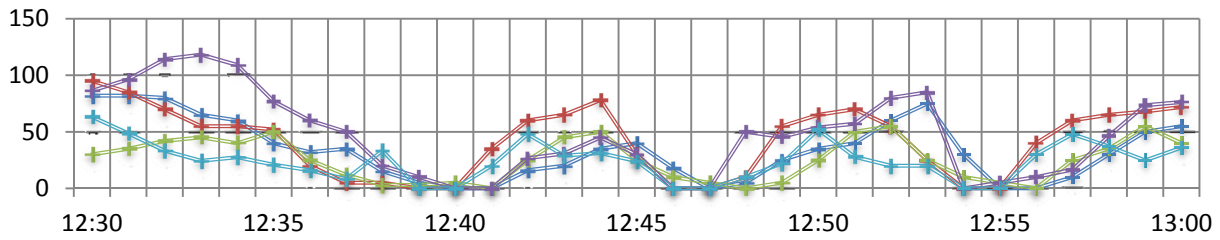


Fig 2 - Acceleration / deceleration curves, 22 May, 2013 between 12:30-13:00 time slot, on Iuliu Maniu Boulevard (between crossroads with Virtuții Street and with Grozăvești Street respectively)

The algorithm used to determine the location of traffic lights consists of three main modules:

- **Mobility Model** - based on PVD the initial trajectories are generated. Using the Kalman filter these paths are improved by interpolating data when sampling interval exceeds 30 seconds. Moreover, it was concluded that for two consecutive positions with speeds above 60 km / h and sampling interval of 30 seconds the vehicle travels at least 500 meters, so additional interpolation is done in order to achieve a higher accuracy of the extracted trajectories. These conditions are applied iteratively to create the paths vector.
- **Extracting the acceleration / deceleration sequences** - starting from the principle that a vehicle approaching a red traffic light will reduces it's speed of travel, the deceleration curves will be used to select the candidate segments.
- **Extracting the probable positions of traffic lights** – for the deceleration sequences previously extracted the algorithm will calculate the bounding box. The probable position of the traffic light will be considered to be the end node of the segment in the direction of travel of the traffic.

To illustrate the algorithm and to presenting the results, the chosen road sector was Iuliu Maniu Boulevard, between crossroads with Apusului Street and with Virtuții Street. The reason for choosing this boulevard section is that, beside traffic lights placed at pedestrian crossing there are traffic lights placed at crossroads as well; this adds an additional degree of difficulty in the algorithm, because it was determined that the driving behavior when facing a traffic light at pedestrian crossing is slightly different than when facing a traffic light at crossroads: the driver starts slowing down sooner for the crossroads than for pedestrian crossing.

To evaluate the quality of the obtained results, the estimated positions of traffic lights were compared with their actual positions. Standard deviations of the estimated positions were calculated after applying this algorithm, which were represented in the graph of Figure 3. The absolute standard deviation of the positions of the seven traffic lights considered was between 8-14m.

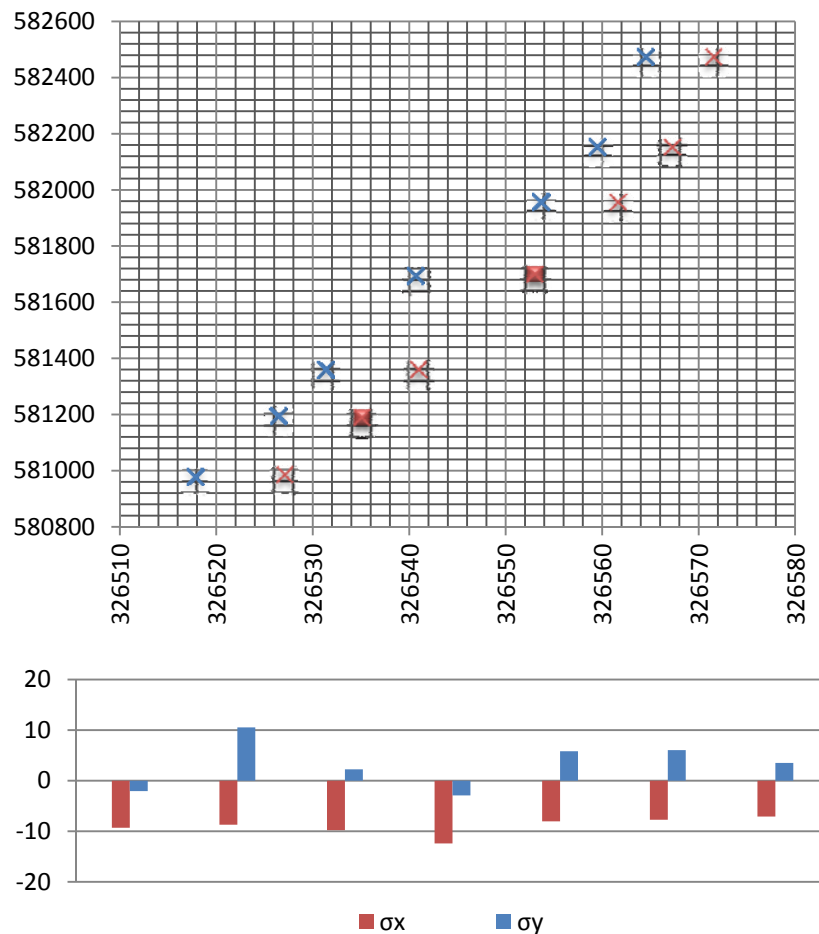


Fig. 3 - Comparative study between the estimated position (blue) and actual position (red) and standard deviation (meters) of the estimated position for traffic lights on Iuliu Maniu Boulevard, Bucharest, between crossroads with Apusului Street and with Virtuții Street

The algorithm presented and the obtained results are dependent on the digital transport network. Before applying the algorithm a preparatory work of the road network is necessary or data may be purchased from specialized companies. This aspect can be considered the weak point of this algorithm, since establishing a digital transport network suitable to apply this algorithm for a wider area (several cities or nationwide) involves high costs. However, the algorithm provides good results, determining the right number of traffic lights and also a qualitative probable position. For companies owning geographical databases and also big probe data (e.g. TomTom, Nokia, AND) the cost of applying such algorithms is down to minimum and the benefits are great, since they can upgrade the navigation experience by providing smaller and more precise travel time.

3.2 Automatic Calculation of the Waiting Time at Traffic Light

As highlighted in [19], finding out the travel time within the transport network involves, at a great extent, knowing the waiting time at traffic lights. Bucharest benefits from an intelligent management and control system of the urban traffic, known as BTMS (Bucharest Traffic Management System). According to [20], this system is composed of an adaptive traffic control subsystem (UTC), a management subsystem for the vehicles of public transport operators (PTM) and a CCTV subsystem for vehicles monitoring at intersections.

Based on this adaptive subsystem, the waiting time at the traffic lights is dynamically changed based on the current state of local traffic. This aspect actually makes more difficult to know the values of waiting time by the entities concerned. However, based on a high coverage with PVD of the urban environment, it can be developed an algorithm for estimating the waiting time at traffic lights.

The proposed algorithm for determining the waiting time at traffic lights is based on the results obtained previously, using the algorithm for identifying the locations of traffic lights, in the sense that the acceleration / deceleration curves determined are correlated with selected segments of the road network for which the existing of a traffic light is determined.

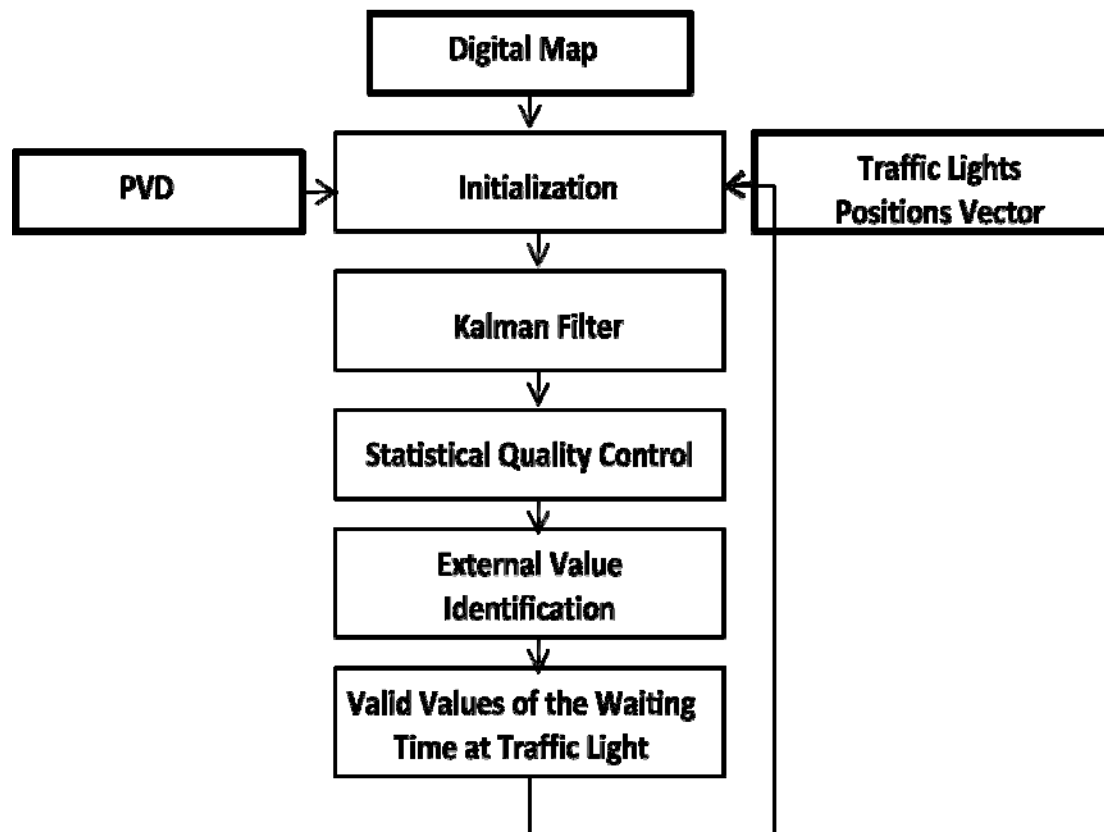


Fig 4 - Modeling the waiting time at traffic lights

Using the Kalman filter, the algorithm will identify the sequences of deceleration curve - stationary - acceleration curve to determine the stationary time and then, to correlate it with the transport network in order to be applied to the correspondent traffic light. The flowchart of the algorithm is shown in Figure 4. The variation of stationary time for a particular traffic light is modeled as the result of an arbitrary process. Over a certain number of epochs, it is determined the period of time between the last value of the deceleration curve and the first value of the acceleration curve for each sequence of deceleration - stationary – acceleration.

Following the algorithm described, to present the results a traffic light from Iuliu Maniu Boulevard was selected (placed between crossroads with Apusului Street and Virtuții Street). Given that Bucharest benefits from BTMS, data were processed per traffic slots as, for example, the value obtained for between 7:00 - 8:59 may differ significantly from the value obtained for 15:00 – 16:59 time slot.

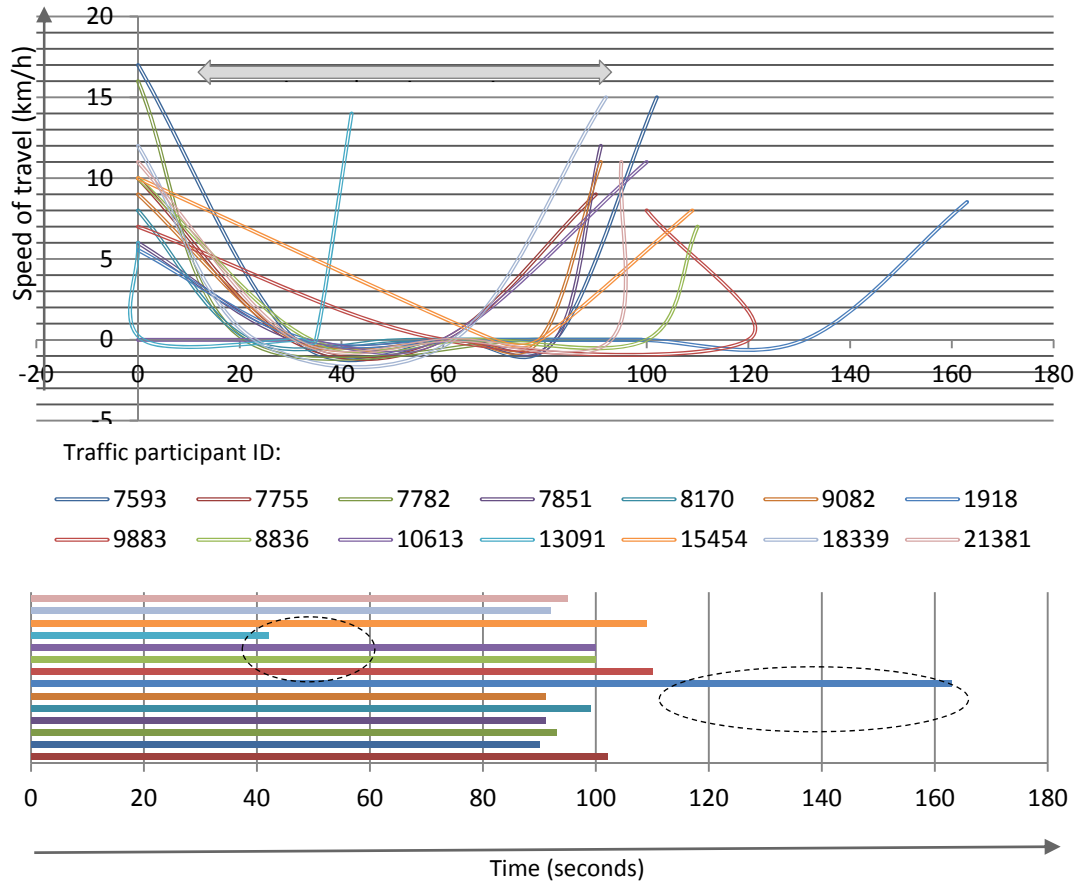


Fig 5 - Estimation of stationary time at traffic light (crossroad between Iuliu Maniu with Virtuții Street)

The chart above reproduces the reasoning by which the algorithm runs. Once identifying a deceleration – stationary time – acceleration sequence, the algorithm calculates, for each traffic participant ID, the time interval between the last value of deceleration and the first value of acceleration (figure 1, chart 1). External values are identified and removed (indicated in figure 5, chart 2). The time interval obtained represents the waiting time at traffic light, determined with a resolution of ± 30 seconds. The results presented cannot be validated against a real value, since there are no measures that can be introduced as benchmarks. These results can be validated only by rigorous statistical analysis based on the results of each iteration. For the traffic light selected for exemplifying results, for the 15:00 - 17:00 time slot and with a resolution of ± 30 seconds, waiting time at traffic light was estimated to 90 seconds.

The algorithm presented in this chapter is used to determine a posteriori the waiting time at traffic light. For near-real time determination of this parameter, it is necessary to include a Kalman filter prediction loop. Because of the high sampling interval (30 seconds), the algorithm is based on a linear interpolation procedure which affects the accuracy of the results. The data suitable for applying such algorithm would have to have a sampling interval of 5 seconds, or even 1 second.

4. Conclusions and Future Research

The reasons underlying the idea of developing Intelligent Transport Systems are among the most varied, especially considering that in recent years, ITS increasingly more found their application in various fields.

Long-term vision in this field is to produce truly multi-modal Intelligent Transport Systems, capable of providing real-time traffic information. According to a press release from IBM in 2010, *"the current trend in transportation management is to use data to make predictions about traffic conditions and allow private agencies to implement strategies and technologies for providing travel information in advance, in accordance with the predictions made"*[21].

The challenges regarding geographical referencing methods are still numerous. Although TMC method delivers the most reliable and accurate results, location tables are not stable and are not maintained in a way to ensure compatibility with previous or future tables. To assure services that rise to the expectations of users, providers must manage numerous location tables, with different versions, at the same time.

OpenLR standard, developed by a private company, involves intellectual property restrictions. Moreover, geographic referencing complexities are difficult to reproduce in the terminal, the most optimistic reproduction percent reaching 80-90%. There is a potential error in the reproduction of each route and this can lead to endless requests for technical support to explain why the route was not well recreated.

Up to this point, the two issues addressed in this paper have been studied before at other universities in the world. The approach presented in this paper is different in that it relies, to a high degree, on the existence of a spatial databases specific in determining this type of parameters. It is agreed that, to create this database, is required additional investment and powerful processing resources, but the obtained results have a high confidence.

The aim of this paper is to expose a methodology for understanding the rules of urban traffic - by determining the location of traffic lights and the waiting time - without approaching the local authorities and without working with databases, stored in various forms (excel files, paper, various lists, etc), containing data regarding the traffic system. The presented method is universal, and it can be applied in any location as long as they used the same input data: digital road network and PVD.

Dynamical computation of traffic parameters (such as stationary time at traffic light, spontaneous traffic incidents), in real-time, can help building truly intelligent, adaptive and dynamic transport and monitoring systems, which can bring cities closer to the 'Smart City' concept, promoted in Europe through the Horizon 2020 initiative.

The documentation and the investigation phases within this research highlighted a number of other directions to be explored:

- Transforming the methodology into an Artificial Neural Network (ANN) structure;
- Adapting the algorithm (including prediction loops) for near-real-time estimations;
- Expanding and testing the algorithm using data at regional or national level;
- Adapting the algorithm to identify other traffic events (crosswalks, spontaneous traffic incidents, etc).

References

- [1] Williams, B. (2008). *Intelligent Transport Systems Standards*. Boston, US: ARTECH HOUSE INC, ISBN-13: 978-1-59693-291-3.
- [2] CEN/ISO (2014). Intelligent transport systems (ITS) - Traffic and Travel Information Messages via Traffic Message Coding – Part 1: Coding protocol for Radio Data System. CEN/ISO TS 14813-1. Online.
- [3] TMC-TL. (03 May 2012). TMC Location Tables - Certified by the TMC Forum and TISA. Version 29: Traveller Information Services Association.
- [4] TISA. (2011). Intelligent Transport Systems (ITS) — Traffic and Travel Information (TTI) via Transport Protocol Experts Group, Generation 2 (TPEG2) - Part 21: Geographic Location Referencing, Version 01. SP10038. Online: Traveller Information Services Association.
- [5] TISA, SP10012. (08 March 2010). Guidelines for ‘ZIP’ Compression of TPEG Service Data Frames Using the ZLIB format, Version 04. Online: Traveller Information Services Association.
- [6] TISA, SP10037. (12-May-2011). Intelligent transport systems (ITS) - Traffic and Travel Information (TTI) via Transport Protocol Experts Group, Generation 2 (TPEG2) – Part 20: Extended TMC Locations for applications, Version 01, . Online: Traveller Information Services Association.
- [7] TISA, SP10038. (20 January 2011). Intelligent Transport Systems (ITS) — Traffic and Travel Information (TTI) via Transport Protocol Experts Group, Generation 2 (TPEG2) - Part 21: Geographic Location Referencing, Version 01, . Online: Traveller Information Services Association.
- [8] TomTom International B.V (2012). OpenLR™. White Paper. Version: 1.5 revision 2. Deutsches Zentrum fuer Luft- und Raumfahrt e. V. DLR, German Aerospace Center
- [9] OpenLR Organization. (2013). Online: <http://www.openlr.org/>
- [10] Fujimoto S. (2011). *On the subject of probe data, Messages from Departments and Centers of NILIM* . MLIT, JAPAN: National Institute for Land and Infrastructure Management.
- [11] Katherine Eleanor Bell. (2013). Identification and Characterization of Fine Particulate Matter Hot Spots on an Urban Arterial Corridor Integrating Probe Vehicle, Traffic and Land Use Data. Portland State University.
- [12] Lee J. & Rakotonirainy A. (2009). *Use of Probe Vehicles to Increase Traffic Estimation Accuracy in Brisbane*. Queensland University of Technology: Centre for Accident Research and Road Safety - Queensland (CARRS-Q).
- [13] Basnayake C. (2004). *Automated Traffic Incident Detection Using GPS-Based Transit Probe Vehicles*. University of Calgary, Alberta: Department of Geomatics Engineering.
- [14] Sagar S.G. & Saha N. (2013). Preparation of Congestion Index of Two Major Arterials of Dhaka City Based on Floating Car Method-an Effective Tool for Transportation Planning. International Journal of Urban Planning and Transportation, Vol.27, Issue.1: ISSN:2051-5820.
- [15] Horizon 2020, Smart Cities, Urban Mobility . (2014, September). *Market Place of the European Innovation Partnership on Smart Cities and Communities*. Online: <http://www.eu-smartcities.eu/>.
- [16] ISO. (2010). Intelligent transport systems — Basic principles for personal data protection in probe vehicle information services. ISO 24100. Online: <https://www.iso.org/obp/ui/#iso:std:iso:24100:ed-1:v1:en>.
- [17] Open Street Map. (2014, August). *Open Street Map*. Online: <https://www.openstreetmap.org>.
- [18] Arobs Transilvania Software (2014). Online: http://www.arobs.ro/proiecte/localizare-prin-gps/?project_id=12&cHash=5c97532991cf3f5b5c0fddff7d4372e9
- [19] Comert G. & Cetin M. (2008). *Queue length estimation from probe vehicle location and the impacts of sample size*. European Journal of Operational Research: DOI:10.1016/j.ejor.2008.06.024.
- [20] UTI. (2014, August). *Sistemul de management al traficului pentru Municipiul București*. Online: <http://uti.eu.com/print.php?pageid=902>.
- [21] IBM (2010). The Case for Smarter Transportation. United States of America. IBM Corporation