DOI 10.20544/TTS2021.1.1.21.p55
UDC••656.11.08:[519.676:004.65(437.3)
Professional paper

Traffic Accident Prediction Based on Comprehensive Data Comparison

Pavel Vrtal[1] and Jakub Nováček[2]

Abstract – **The project aims to improve the road safety by identification of potentially risky locations on roads that show similarities to known accident locations. This contribution presents an algorithm that is combining the commonly used methods and the road characteristics identified through the road network analysis and road passport. It is able to search for potentially risky locations on the road network and alert the road managers to the potentially risk prone sites.**

Keywords – **Traffic accident prediction, Road safety, Comparison algorithm, Accident sites**

I. Introduction

The issue of the road accidents is a long-standing and ongoing issue that requires constant attention. The Czech Republic (CR), like any other European Union country, is obliged to adopt a plan to reduce the consequences of the road accidents, the so-called White Paper [1]. This document serves as the base for a number of national strategic documents, including the National Road Safety Strategy in the CR [2], [3]. One of the main objectives of the current practices is to promote the idea of "Vision Zero". This idea ideally represents a transport system without fatalities or serious injuries. Achieving such a goal is so far only utopia, but it is still assumed that it will be possible to achieve or at least come very close to this result in the future.

It can be said that there is a number of factors by which this objective can be met. One of the potential solutions is the development of modern technologies and systems that will not only contribute to reduction of accidents, but also help to prevent them. Essentially, these elements are already being used as support systems for vehicles, where the last stage is completely autonomous driving without human assistance. Progress in this issue is considerable and a lot of time and money was invested in the development. However, to effectively achieve Vision Zero, another key condition must be met. A shift or a transformation of the transport infrastructure to such a state that it does not pose a risk by itself. A number of procedures are currently used to proactively help to improve this condition. Road safety inspections can be seen as a typical example [4], where the acquired knowledge directly improves the quality of the existing roads. Another effective tool is the implementation of road safety audits, which in turn highlights the road design deficiencies and risks already at the design stage. Although the quality of the road network is constantly being improved in this way, the overall progress is not sufficient. There is still a number of places that are risky in design and dangerous for road users, however, not addressed.

The current approach of identification of accident sites in the CR is based on certified methodologies that use various classifications. Currently, the most widely used approach is to identify the accident sites by three key criteria. By a definition, the accident site is a site where occurred at least 3 accidents with personal consequences within1year, at least 3 accidents with personal consequences of the same type in 3 years, or at least 5 accidents of the same type during 1-year period [2]. Another approach is based on identification of the accident sites through mathematical models that predict the locations by the traffic volumes [5]. It is also possible to determine the accident sites by the relevant road managers on the basis of their practical experience. Nevertheless, independently on the chosen approach, the identification of accident site is usually only retrospective and the approach is therefore not proactive.

The aim of this contribution is to highlight the fact that there are a number of locations on the road network that show a number of similar characteristics to already identified accident sites and could therefore present a potential risk for the road user. The only difference is that no significant accident event or a larger number of accidents has yet taken place. By identifying such potentially risky sites, it would be possible to achieve the desired proactive approach and avoid unnecessary injuries. Therefore, an algorithm for identification of new, potential accident-prone locations based on predefined input information is presented. The algorithm is based on the principle of conditional probability.

The following chapters set out how this outcome can be achieved. The first part of the text describes the way in which the input parameters concerning the directional and elevation characteristics of the road can be appropriately identified. The next section focuses on the creation of a database of the individual features necessary to identify a match to already known accident locations. Last but not least, a procedure how to apply it in practice is outlined.

II. Methods

The initial idea was to create a test algorithm that would be able to detect and search for similar sections in other parts of the road network based on predefined road parameters (directional and elevation guidance). All outputs are designed to correspond with a particular probability match to a pre-identified location. However, it is necessary

[1] Pavel Vrtal, CTU in Prague, Faculty of Transportation Sciences, Konviktská 20, Prague, Czech Republic, e-mail: vrtalpav@fd.cvut.cz
[2] Jakub Nováček, CTU in Prague, Faculty of Transportation Sciences, Konviktská 20, Prague, Czech Republic, e-mail: jakub.novacek@fd.cvut.cz

to consider the non-uniform characteristics of the road section. At the same time, there is also a problem with the different orientation of the cardinal points of the individual roads. Furthermore, it is also not possible to simply find the mathematical properties of a road curve defined by previously known equations. For this reason, it was necessary to use a procedure that could effectively identify a specific road section with high accuracy without any additional information about the specific properties. The ideal way was to use a common geodetic coordination system, such as the single trigonometric cadastral network (S-JTSK) used in the CR, to determine the position and elevation of any point of the road in the terrain. Additionally, it was necessary to obtain approximate information on the curvature of the analysed roads using the high quality data from the map portal of the Czech Office of Surveying and Cadastre [6] and the web portal Mapy.cz [7]. The current research is in its initial and testing phase, so only the road centerline position and elevation of the road has been utilized. In the future, it will be necessary to use the axes of individual lanes. However, the export of the position coordinates of road axis points directly from map applications is not entirely usable. The output of individual points is not with a constant spacing and the recorded position of points changes dynamically depending on the curvature of the road.

Therefore, in order for the comparison algorithm to work properly, it was necessary to split the roads into fixed lengths. In this way, it was possible to obtain a sufficiently detailed coordinate description of the selected roads (Fig 1). The road segment length should be short enough to enable the determination of low curvature radius, yet, not too small due to the overall length of the road network.

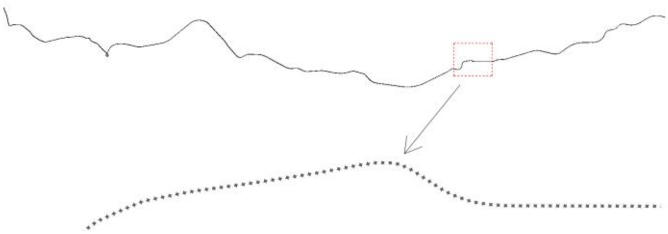

Fig. 1. Splitting the road into fixed length elements

The position point comparison algorithm is constructed in such a way that the individual section lengths between adjacent points are taken as a specific parameter. The difference between the previous and subsequent X and Y coordinates is then calculated. It enables to obtain the positional differences of the X and Y coordinates and thus the road spatial layout. An illustrative example is shown in Fig 2.

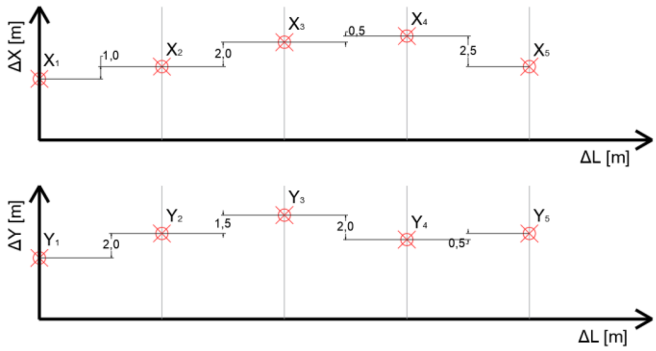

Fig. 2. Relative displacement of road segment end points

Fig 3 further defines the relative displacements of the individual points of the road segments. At the same time, a subsequent calculation of the area under the road curve, enables to identify similar road segments within the road network in high accuracy. The identification is done through comparison of the change in the area beneath individual road sections. In this way, the algorithm is able to find similar sections of road on any length of section based on the acceptable difference threshold. The amount of surface change is almost constant in sections that show small changes in geometry. Typical examples would be straight road sections or directional curves with a constant radius. The match is no longer such at the transitional point, however, these points can still be located. It is also able to distinguish the direction of the road (left-hand, right-handed curve).

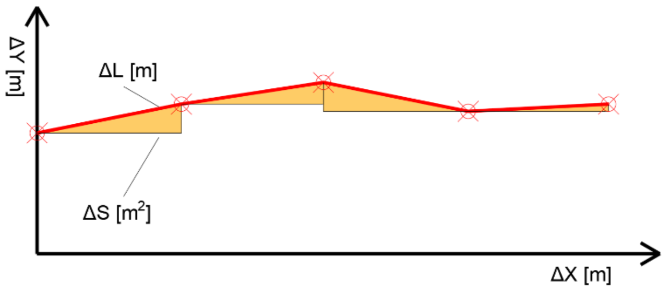

Fig. 3. Calculation of the area under the road curve based on the relative displacement

Completing the elevation map for the road sections in question is considerably more complicated. The freely available map documentation is not sufficiently accurate to determine an adequate match. In order to achieve the most accurate result, it was necessary to use models of aerial laser scanning of the elevation in the CR. These models are digital terrain models which captures the natural or human activity-modified surface in digital form. The models are in the form of heights of discrete points in an irregular triangular area network (TIN). Digital models used for the purposes of this study are the 1st and 5th generation models (DMR 1G and 5G). According to the authors' assessment, these models contain the most suitable data for the purposes. The mean error of DMR 1G is 0.4 m in open terrain, without obstacles, and 0.7 m in forested landscape [8]. It can be seen that the achieved accuracy is not quite ideal, however, the advantage of using this scanning is in the possibility to identify bridge objects, which are not so accurate or completely missing in the following generations. The reason is that subsequent generations have a more aggressive error filtering approach, which lead to omission of these objects. The use of the latest, 5th generation surface scanning, on the other hand, have a significantly smaller mean error. The DMR 5G is stated to have an accuracy of 0.18 m in open terrain and 0.3 m in forested landscapes [9]. By combining the aforementioned models together with directional location of road network, it was possible to determine the vertical alignment. However, due to the differences in heights of the DMR, it was necessary to develop an algorithm that estimated the correct heights. The approach works by the prediction of subsequent points through calculation of the trend of the height values. As a base information was taken the DMR 5G model. If the subsequent height from the model corresponded with the predicted value, it was possible to consider that the subsequent height was correct. If the prediction of the next point height significantly differed, it was evaluated as an incorrect by the algorithm and replaced by the 1st generation height information. In case, that both models' heights showed values which differed significantly, the predicted height information was used. The model of the particular road vertical alignment was thus created by iterative evaluation of the most suitable points. The road was then divided into sections of constant length, similar to the directional alignment. In order to further minimize the effects of local errors, the resulting vertical alignment was smoothed with use of the simple moving average. The idea is that the changes in the vertical alignment of the road are gradual and continuous. At the same time, to check the error rate, maximum and mean errors were evaluated and this information was used for new iteration to improve the resulting quality of the model. The test output from the vertical alignment of road can be seen in the following Fig 4 and 5. In the first case the road is routed under the bridge structure, in the second case the road is routed over the bridge.

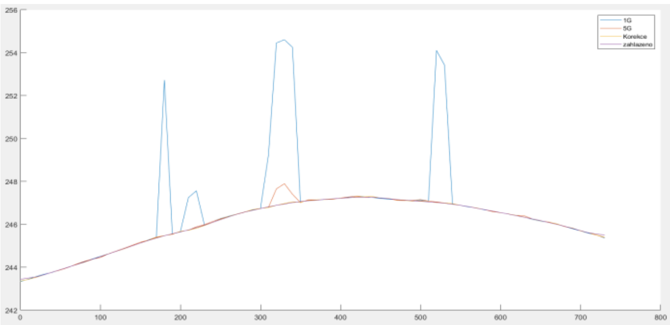

Fig. 4. Vertical alignment model – under the bridge (DMR 1G – blue line, DMR 5g – orange line, 1st iteration – yellow line, final model – violet line)

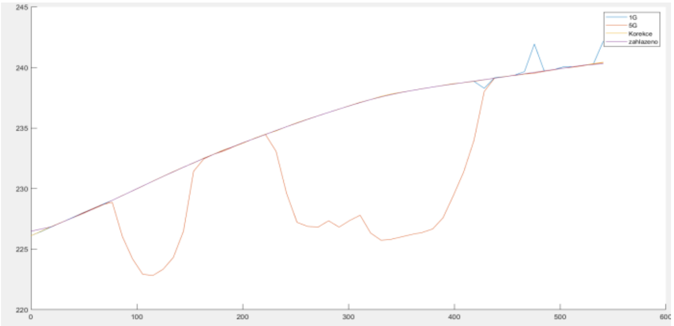

Fig. 5. Vertical alignment model – over the bridge (DMR 1G – blue line, DMR 5g – orange line, 1st iteration – yellow line, final model – violet line)

After obtaining general information on the directional and vertical alignment, a detailed assessment of the individual elements on and near the road must be provided. The principle for the identification of typologically similar sections within the road network is based primarily on a sufficiently detailed database containing various segments. Fig 6 shows the initial layout of the database structure, which can subsequently be used for the filtering. This part is still under development and it is likely that there will be subsequent modifications to evaluate all aspects crucial for an accurate determination. The database is designed in such a way that each road section "Road element" is supplemented with different groups of attributes that modify the resulting correspondence of the selected section with other road sections. In general, the sub-categories of the "Road element" can be divided into elements that will contain basic information on geometric parameters, information related to the traffic characteristics, road infrastructure or facilities in close proximity to the road. At the same time, each section is supplemented with information related to traffic accidents statistics. This database will also include the use of data obtained in the context of road safety inspections aggregated in the CEBASS (Central Evidence of Road Safety Analysis) database [10].

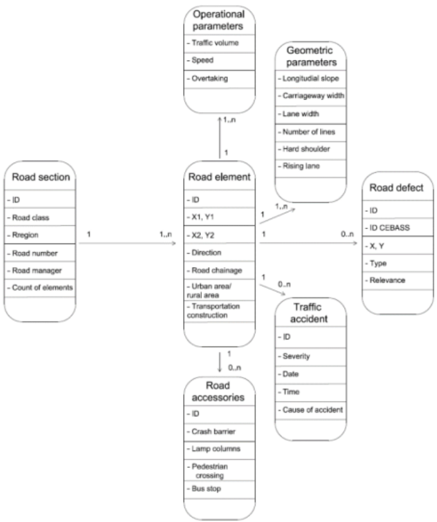

Fig. 6. Database layout and its relations

III. Possible limitations

The accuracy of the resulting forecast of potentially risky road sections will depend mainly on the quality of the input data, which significantly influence the result. At the same time, it is necessary to establish a scale of significance for individual attributes, which can be used to prioritize sub-elements of the database over others. If the thresholds of the matching is set too high in the final stage, the algorithm will not be able to find any similar results. For this reason, it is necessary to test the possibilities of how accurate the matching performs and to find a compromise between identical and errant solutions.

If the results are sufficiently valid, desirable step is a creation of a simple and intuitive web interface that will contain the above-mentioned principles. Through filter windows or by directly marking of an area on the map, it will be possible to find places with similar character from a pre-selected location. The graphical representation in the map base will then be discussed as to how best to interpret these results in order to make them as user-friendly as possible for wide range of potential users (such as road administrators, safety experts, road managers).

IV. Future direction of the research

The future steps lies mainly in adding new elements to the database, which will characterize the roads in more detail. At the same time, in this way, elements of a different nature will be collected that can develop the passport of the roads in more detail. Another way to use the new solution is, for example, to implement information in car navigation systems in such a way that the system would inform the driver that he is in areas where increased caution is needed. Last but not least, it is possible to look at the result of this project as a database that can be used to search for very similar places on the roads and thus suggest, for example, the reconstruction of selected elements or sections of the road.

V. Conclusion

The search for new ways to effectively address and prevent traffic accidents is a necessity for the future development and improvement of the traffic safety. This article illustrates one possible step to make a comprehensive use of the available data and to look at accident rates in a different way than has been done so far. If the full functionality of the proposed solution is achieved, it will be possible to effectively predict a potentially hazardous locations where significant socio-economic loses could occur. It also has the benefit of informing the road managers that there are areas on the road network that need to be addressed as a priority and enables to take a proactive approach.

Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS21/138/OHK2/2T/16.

References

[1] White Paper on transport – Roadmap to a single European transport area – Towards a competitive and resource- efficient transport system, European Comission, Luxembourg, ISBN 978-92-79-18270-9, 2011.

[2] Strategy BESIP 2021-2030, Ministry of Transport, Prague, 2020.

[3] Czech Road Safety Strategy 2011-2020, CDV – Transport Research Centre, Brno, 2017

[4] Road safety inspection, methodology of implementation, CDV – Transport Research Centre, Brno, ISBN 978-80-86502-49-6, 2013.

[5] Identification of the critical accident locations using GIS analysis, methodology, CDV – Transport Research Centre, Brno, ISBN 978-80-86502-76-2, 2014.

[6] ČÚZK Available from: https://www.cuzk.cz/

[7] Mapy.cz Available from: https://www.mapy.cz/

[8] Digital surface model of the Czech Republic 1st Generation Available from: https://geoportal.cuzk.cz/(S(5du3ksqqvjx5wzxlx4c1cc0n))/Default.aspx?lng=CZ&mode=TextMeta&side=vyskopis&metadataID=CZ-CUZK-DMP1G V&mapid=8&menu=303

[9] Digital surface model of the Czech Republic 1st Generation Available from: https://geoportal.cuzk.cz/(S(h5swv3qhfw4rewqdrbahvqan))/Default.aspx?mode=TextMeta&side=vyskopis&metadataID=CZ-CUZK-DMR5G-V&head_tab=sekce-02-gp&menu=302

[10] CEBASS Available from: https://cebass.rsd.cz/