Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 22 (2017) 75–84

Transportation Research Procedia

www.elsevier.com/locate/procedia

19th EURO Working Group on Transportation Meeting, EWGT2016, 5-7 September 2016, Istanbul, Turkey

Detecting change in road environment via analysis of marked point processes associated with traffic signs

Zoltán Fazekas*, László Gerencsér, Péter Gáspár

Institute for Computer Science and Control (MTA SZTAKI)
Kende u. 13-17, Budapest, H-1111, Hungary

Abstract

In the paper, logs of traffic signs – possibly recognized and recorded by an automatic traffic sign recognition system – are analyzed to detect change in some aspect of the road environment along a route and to locate the change point – also along the route – between the different environments. The logs considered here keep a record of the locations and types of the traffic signs installed and detected along a route. The traffic sign logs are seen as realizations of marked binomial processes and the minimal description length (MDL) approach is used for detecting change in the road environment along a route. In particular, the change detection problem associated with driving from one topographical area to another is addressed here as a simple illustrative example. In order to cater for an efficient solution of this task – and also for that of other road-environment change detection tasks – the on-the-fly minimization method used in the Page-Hinkley change detector has been adopted. Simulation results in respect of traffic sign data generated for test purposes corroborate the expected behavior of the detector. In respect of real traffic sign data, a good qualitative agreement was found between the GPS-based altitude-profile of the data collection trip – thresholded at some manually or automatically selected altitude after the trip – and the MDL-based topographical segmentation of the route. For this segmentation, the traffic sign locations, more precisely the path-lengths corresponding to these locations measured from the starting point of the route along the route, and the corresponding traffic sign types were used as input.

Keywords: Detection of road environment; Change detection; Advanced driving assistance systems; Marked point processes; Minimal description length approach; Page-Hinkley detector.

1. Introduction

Roads are planned for, constructed, operated, maintained and used within their environments. The environment of a road could refer to its natural environment, covering the geological, the geographical, the meteorological and the biological aspects, and to its human environment, covering the social, the political, the legal, the economic, and the socio-cultural issues. When driving a car, one must take into consideration the type, the geometry and the quality of the road used, the weather conditions around it, and also the actual and expected traffic over it, and many other

* Corresponding author. Tel.: +36-1-279-6000 ; fax: +36-1-466-7483.
E-mail address: zoltan.fazekas@sztaki.mta.hu

10.1016/j.trpro.2017.03.011

conditions and circumstances. Some of these conditions are determined by the country one is driving in (e.g., hand of traffic, speed-limits), some depend on the socio-cultural environment, while others depend on the natural environment of the road.

Usually, car drivers are assisted in many ways in recognizing, perceiving and understanding the conditions and circumstances of driving. The best engineering practices used in assisting drivers – driving their cars on different kinds of roads and in different viewing and traffic conditions – with regards to the above tasks are covered by Edquist et al. (2009). The means of assistance include e.g., road signs, traffic signs, information and directional signs, printed and electronic maps, navigational devices, various on-board sensors and nowadays also community-based information services, see e.g., Dennis et al. (2015). In case of driving a smart car, the driver is further helped by the car's various advanced driving assistance systems (ADAS), e.g., its anti-lock braking (ABS), lane-departure warning (LDW) and traffic sign recognition (TSR) systems, see Nikolič (2014). From this list, our paper will focus on the traffic signs, but the TSR systems will be also touched upon.

Undoubtedly, it is very important to perceive, to register, to understand and to obey each and every relevant traffic sign that one drives by, but perhaps it is of similar importance to register the natural/human/socio-economic environment of the road, understand its implied dangers (e.g., large number of pedestrians in a busy downtown area) and rules, and behave and drive in accordance with these. Presently, however, neither the TSR systems, nor any other common on-board camera-based systems assist the car drivers in this respect. Perhaps, in the not very distant days of even-smarter cars, a new type of ADAS will appear in the automotive market: the environment recognition systems (ERS). Again, the various environmental aspects mentioned above can be considered in this respect.

Hereinafter, a statistical inference approach – applied to traffic sign data – that could be utilized in the envisaged ERS is presented in the form of a simple instructive example. Consider therefore a log of geographical locations of traffic signs with the types of the traffic signs being also noted in it. Furthermore, it is assumed that the log is recorded along a loopless route taken by a car. Using statistical inference in the context of environment recognition is motivated by the expectation that certain traffic signs appear more frequently in some environments than they do in others. In town centers, signs indicating railway/bus station, restaurant, hotel/motel, cafeteria/refreshments, museum/historic building, parking places, particularly parking places against fee in the vicinity, and the ones warning the drivers of pedestrian traffic are likely to appear. Other traffic signs are installed more frequently in suburbs (e.g., traffic signs indicating industrial area, goods harbor, airport, low-flying aircraft/sudden aircraft noise) and still others in rural areas (e.g., traffic signs warning of cattle/wild animals, and falling rock).

In conjunction with the topographical environment of the road, one expects that the bend to left/right, the double bend and the steep ascent/descent traffic signs are installed in considerably bigger numbers in hilly areas than by the roads running over a flatland. In Fig. 1, a road segment – in the hilly area of the Mátra Mountains – that is marked with a bend to right traffic sign is shown as an example that supports this expectation.

Referring to the above environment detection examples, it is not claimed here that using statistical inference in respect of traffic sign logs is the best way to determine a particular environmental aspect of the road, e.g., whether

Fig. 1. A bending road in the Mátra Mountains. The drivers are forewarned of the sharp turn of the road by the bend to right traffic sign. In the inset, the two distant traffic signs appear enlarged.

one is driving in a hilly area, or over a flat land. Clearly, there are simpler and more precise methodologies for this purpose. However, some road environment detection tasks could be meaningfully addressed, partially, or even completely solved using the approach presented here. Also, detecting change in the road environment and locating environmental change points via statistical inference may overcome certain deficiencies of the present-day ADAS, and particularly of the present-day – still not quite reliable – TSR systems. For example, it could be used to compensate for some crucial traffic signs missed or incorrectly recognized by the on-board TSR system. Furthermore, it may serve as an additional measurement that can be incorporated into ADAS via sensor fusion.

The change detection in the topographical environment along a route based on logged occurrences of traffic signs and the associated task of detecting the topographical change point will be used as a running example in the paper. This target application and running example have been chosen despite the fact that change detection tasks in respect of other aspects of the road environment could be perhaps technically more relevant. The choice was motivated by the fact that a fairly simple stochastic model can be used in this case and the change detection results are easy to verify. These features were considered beneficial in this stage of the research. As it will be discussed in detail below, the traffic sign logs are seen as realizations of marked binomial point processes and the minimum description length (MDL) approach is used for change detection. This approach originates from the field of information theory and looks at how to create concise description of data generated by stochastic processes. Though it considers how to optimally encode data, but does it without the intention of actually compressing data (i.e., the traffic sign logs in the given case); rather the approach fits competing stochastic models to the data and thereby detects change. In order to cater for efficient implementations of the topographical and other road-environment change detection solutions, the on-the-fly minimization method utilized in the Page-Hinkley detector was adopted and tested in respect of artificial test data and on real traffic sign data. In Section 2, the aforementioned mathematical notions, methods and approaches are summarized and some references are provided for further reading.

2. Mathematical background

2.1. Marked point processes

A natural model for traffic sign data is a marked point process given by a increasing sequence of time-points, say T_n which may be defined in terms of traffic sign data, with time replaced by driving distance, or more precisely path-length. The points of a point process may be labeled with marks.

A marked point process then can be formalized as a pair (T_n, ρ_n), where ρ_n is the mark, see e.g., Baddeley (2007). For instance, in a log of traffic signs, a traffic sign location – along a route – may carry a label stating the type of the sign. A usual way of modeling logs of events (e.g., computer-related events, physical detections) is via a marked Poisson process, which is a genuinely continuous-time model. In developing an algorithm for detecting environmental changes – based on logs of traffic signs – a discrete-time approximation of the mentioned model have been used as discussed below.

2.2. Change detection

The problem of detecting abrupt changes in the dynamics of stochastic signals has been widely discussed in the literature, together with a number of applications, see Basseville and Nikiforov (1993) or Poor and Hadjiliadis (2008) for a recent excellent survey. This area of research has been initiated by problems of change detection for iid data, leading to the well-known Page-Hinkley detector, to be described below, see Page (1954), Hinkley (1971), Lorden (1971). The very same detector has been later applied and analyzed for dependent data, see Lai (1995), Gerencsér and Prosdocimi (2011), Gerencsér et al. (2013), and the references therein. The two performance criteria for a change detector is its average run length (ARL) between false alarms, which translates to false alarm probability, or false alarm rate (FAR), and the expected delay in detection.

2.3. Minimum description length approach in model selection and change detection

A novel approach to change detection based on the minimum description length (MDL) approach, originating in the work of Rissanen (1978), and extensively developed in Rissanen (1998), was proposed in Gerencsér and Baikovicius

(1989) and Gerencsér and Baikovicus (1992). The basic idea of the MDL approach is to choose between models for describing data on the basis of the minimal code-length by which we can encode our data, assuming the model – and only the model itself. The advantage of this methodology is its enormous flexibility. We will see that the widely used Page-Hinkley detector can be easily interpreted as a procedure relying on the MDL approach.

3. Methodology

A formulation of the Page-Hinkley change detector for detecting change in the mean of a step signal corrupted with additive white noise was presented in some detail by Basseville (1988). It was emphasized in his paper, and also numerous references – including Hinkley (1971) and Lorden (1971) which are also cited herein – were given in this respect, that the Page-Hinkley change detector can be used in a wider context, e.g., for detecting change between two known probability laws p_{θ_0} and p_{θ_0}. In the sequel, we briefly describe the Page-Hinkley change detector – using an MDL approach – for detecting the change point between two different known discrete probability laws.

Assume that we have a sequence of observations $\xi_1, ..., \xi_N$ such that for $n < \tau^*$ the ξ_n's form an independent identically distributed (iid) sequence of random variables taking discrete values with probabilities $P(\xi_n = x) = p(x, \theta_0)$, while for $n \geqq \tau^*$ the ξ_n'-s are iid and generated according to the probability law $p(x, \theta_1)$, so that the two probability laws are different. The problem is to estimate τ^* from observed data in real time.

An MDL approach to solve this problem is as follows: choose an arbitrary τ with $1 \leqq \tau \leqq N$, and, assuming $\tau^* = \tau$, encode the observed data optimally using the hypotheses on the data generating mechanism described above. Based on standard results of information theory, the overall optimal code-length $L_N(\tau)$ of the observed data, in an asymptotic sense – allowing block coding – is

$$L_N(\tau) = \sum_{n=1}^{\tau-1} - \log p(\xi_n, \theta_0) + \sum_{n=\tau}^{N} - \log p(\xi_n, \theta_1) \; . \tag{1}$$

Following the MDL principle, the estimator $\hat{\tau}$ of τ^* is then obtained by minimizing $L_N(\tau)$ in τ. In order to explore the shape of this function, let us compare successive values, such as, $L_N(\tau)$ and $L_N(\tau + 1)$, and define the score $\Delta L_N(\tau)$:

$$\Delta L_N(\tau) \triangleq L_N(\tau + 1) - L_N(\tau) = - \log p(\xi_\tau, \theta_0) + \log p(\xi_\tau, \theta_1) \; . \tag{2}$$

Note that $\Delta L_N(\tau)$ is independent of N, i.e., $\Delta L_N(\tau) = \Delta L(\tau)$. Furthermore, $\Delta L(\tau)$ can be interpreted as the difference of the optimal code-lengths encoding ξ_τ assuming the probability laws corresponding to θ_0 and θ_1, respectively. A remarkable property of $\Delta L(\tau)$ is that

$$E[\Delta L(\tau)] \leqq 0 \quad \text{for} \quad \tau < \tau^*, \quad \text{while} \quad E[\Delta L(\tau)] \geqq 0 \quad \text{for} \quad \tau \geqq \tau^*. \tag{3}$$

This property follows directly from the optimality of the codes, which is in turn based on the Kullback-Leibler inequality, see Cover and Thomas (1991). Thus τ^* is the minimizer of $E[L_N]$, thereby justifying the choice of $\hat{\tau}$ as a minimizer of L_N in yet another way.

A heuristic procedure for minimizing L_N in real-time is now obtained by identifying the time-point after which L_N has a definite upward trend. Thus we arrive at following signal to be monitored:

$$g_n = S_n - \min_{m \leqq n} S_m \; , \quad \text{where} \quad S_n = \sum_{k=1}^{n} \Delta L(k) \; . \tag{4}$$

It is then expected that g_n is typically 0 for $n < \tau^*$, while it is typically increasing for $n \geqq \tau^*$. Thus the following test can be devised: take a positive number $\lambda > 0$ and generate an alarm – for signaling the change – if $g_n > \lambda$; alternatively, define the estimator $\hat{\tau}$ as

$$\hat{\tau} = \min \langle n \; : \; g_n > \lambda \rangle \; . \tag{5}$$

This is the well-known Page-Hinkley change detector – see Page (1954), Hinkley (1971) and Lorden (1971) – for detecting the change in the parameter of the probability law that generates the observed data. The quality of the detector is controlled by the choice of the threshold λ: for its small values one may get higher FAR's, while for its large values the expected delays of detection may become too large. The trade-off between these two performance criteria is partially controlled by the Kullback-Leibler divergence between the probability laws $p(x, \theta_0)$ and $p(x, \theta_1)$.

3.1. Application for discrete time point processes

Consider now an discrete time approximation of a Poisson process obtained by taking an iid sequence of random variables ξ_n with binomial distribution, taking values 1 and 0, with probabilities θ and $1 - \theta$, respectively. An event occurs if $\xi_n = 1$. Then the optimal code-length for encoding ξ_n can be written as

$$-\xi_n \cdot \log\ \theta\ -\ (1 - \xi_n) \cdot \log\,(1 - \theta)\ . \tag{6}$$

Let us now consider a situation where each event has a mark, say color ρ, chosen independently of ξ_n. This is technically modeled by an iid sequence of pairs (ξ_n, ρ_n), so that the sequences (ξ_n) and (ρ_n) are independent. For notational simplicity assume that the range of ρ_n is $1, ..., m$, with $P(\rho_n = k) = p_k$.

Now, let us investigate the sequence of pairs $(\xi_n\,,\ \xi_n \cdot \rho_n)$. In order to express its optimal code-length, note that

$$P(\rho_n = k,\ \xi_n = 1)\ =\ P(\rho_n = k \mid \xi_n = 1) \cdot P(\xi_n = 1)\ =\ p_k \cdot \theta\ . \tag{7}$$

For notational convenience, let $\zeta_{n,1}, ..., \zeta_{n,m}$ be the indicator functions of detecting the marks $1, ..., m$, respectively, i.e., $\zeta_{n,k} = 1$, if $\rho_n = k$, while $\zeta_{n,k} = 0$, if $\rho_n \neq k$. Then the optimal code-length of encoding the pair $(\xi_n\,,\ \xi_n \cdot \rho_n)$ is as follows:

$$-\sum_{k=1}^{m} \xi_n \cdot \zeta_{n,k} \cdot \log(p_k \cdot \theta) - (1 - \xi_n) \cdot \log(1 - \theta)\ . \tag{8}$$

The θ-dependent part of the first term can be simplified by noting that

$$-\sum_{k=1}^{m} \xi_n \cdot \zeta_{n,k} \cdot \log\theta = -\,\xi_n \cdot \log\theta \cdot \sum_{k=1}^{m} \zeta_{n,k} = -\,\xi_n \cdot \log\theta\ \ .$$

Thus we arrive at the conclusion that the optimal code-length for the observed pair at time n is as follows:

$$-\xi_n \cdot \log\theta\ -\ (1 - \xi_n) \cdot \log(1 - \theta) - \sum_{k=1}^{m} \xi_n \cdot \zeta_{n,k} \cdot \log p_k\ \ . \tag{9}$$

Comparing the above expression to (6), it is easy readily seen that the additional code-length is due to the code-lengths of marks at actual events.

Now, let us assume that probability distribution for the overall rate of events is characterized by $\theta = \theta_0$ before the change, and by $\theta = \theta_1$ after the change. In addition, assume that probability distribution of the marks changes from $p_{0,1}, ..., p_{0,m}$ to $p_{1,1}, .., p_{1,m}$ in the same manner. The difference of the optimal code-lengths encoding the observed pair $(\xi_j\,,\ \xi_j \cdot \rho_j)$ is then, following (2), given by

$$\Delta L(j)\ =\ -\xi_j \cdot \log\frac{\theta_0}{\theta_1}\ -\ (1 - \xi_j) \cdot \log\frac{1 - \theta_0}{1 - \theta_1}\ -\ \sum_{k=1}^{m} \xi_j \cdot \zeta_{j,k} \cdot \log\frac{p_{0,k}}{p_{1,k}}\ \ . \tag{10}$$

Having defined this score, we can proceed by applying the Page-Hinkley detector – as described above – to find the estimator $\hat{\tau}$ of τ^*. Since the our observation sequence is iid and bounded, the results of Gerencsér and Prosdocimi (2011) apply, and we find that the empirical FAR for a time-homogeneous sequence is essentially the same as the expectation of FAR. The latter exponentially decays with the increase of the threshold λ chosen for the detector. It should be noted that the above procedure is also applicable if the exact probability distributions are not known. In this case, one needs to fall back on approximate models.

4. Traffic sign data collection

The log of traffic signs is viewed herein as a realization of a marked point process. In the log, the sequence of events is recorded; each event is characterized by its event-location (e.g., the path-length covered by the car along a known route to reach a specific traffic sign location) and its event-type (i.e., detecting a traffic sign of a given type). Initially, only three different event-types were distinguished by the model presented here: the detection of a bend to left/right traffic sign, that of a double bend starting left/right traffic sign, and that of a steep ascent/descend traffic sign. It should be noted that each of the traffic signs associated with these events is recognized by some present-day TSR systems. Later, however, a fourth event-type, namely the detection of a curve marker arrow, was included in the above event-type list. It should be noted that the curve marker arrows – see Fig. 5 – are not usual targets of the present-day TSR systems; though some experimentation with the recognition of traffic signs – including also the curve marker arrows – using an artificial neural network was reported in Rahmani (2012).

Prior to the data collection trip and the environmental change detection experiment described herein, some earlier data collection trips were carried out – in respect of other traffic sign types – in different regions of Hungary. Apart from the associated trajectory data and traffic sign logs, also the time-lapse photographs taken during these trips were utilized to estimate the spatial frequencies of the traffic signs that belong to the types targeted in this study. Spatial frequencies in hilly areas, as well as in flat lands were estimated. Then the marked binomial point process model – described in Subsection 3.1 – was tuned to these topographical environments resulting in two particular models; see Table 2 for the model parameters. In preparation for the evaluation of real traffic sign logs created for the purpose, simulations in respect of synthetically generated point processes – associated with certain virtual road environments – were carried out. The generation and the evaluation of the synthetic traffic sign data are summarized in Subsection 4.1. The details of the traffic sign data collection from real roads is presented in Subsection 4.2. While the associated environmental change detection results are presented in Section 5.

4.1. Generation and evaluation of synthetic traffic sign data

In order to carry out initial experiments with the Page-Hinkley detector applied to traffic sign data, simple marked point processes, namely marked binomial point processes, were synthetically generated for virtual routes traversing two different virtual road environments. The details of the Page-Hinkley detector applied to marked binomial point processes were elaborated in Subsection 3.1. It should be noted that in formulating the Page-Hinkley detector for the particular case described there, the point process was considered to have happened in time. In the present application, however, the traffic signs are better characterized with their locations along the route – than with the moment of their observation – therefore, technically, the path-length required to reach the considered locations will be used instead of the time-variable and each consecutive segments of the path, instead of the discrete times of the time domain, will be

Table 1. Stochastic model parameters for virtual road environments VRE0 and VRE1

Road environment	Traffic sign group	Group probability	Traffic sign type	Conditional probability
VRE0 – with many winding roads	Traffic signs signifying dangerous bends	$\theta_0 = 0.8$		$p_{0,1} = 0.5$
				$p_{0,2} = 0.1$
				$p_{0,3} = 0.1$
				$p_{0,4} = 0.3$
VRE1 – with few winding roads	Traffic signs signifying dangerous bends	$\theta_1 = 0.1$		$p_{1,1} = 0.2$
				$p_{1,2} = 0.1$
				$p_{1,3} = 0.1$
				$p_{1,4} = 0.6$

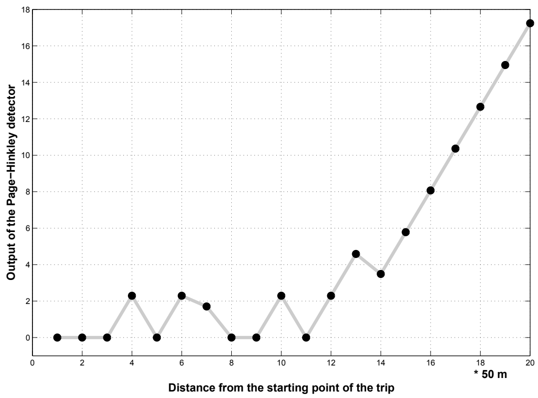


Fig. 2. The output of the Page-Hinkley change detector computed for a particular realization of the synthetically generated marked point process. The parameters of the two marked processes used in the data generation for virtual road environments VRE0 and VRE1 are given in Table 1. The change point between VRE0 and VRE1 was set between 13 and 14 in the data generation phase, i.e., the environmental change occurred after a path-length of 13 ∗ 50 m's having been covered during the virtual trip. In the detection phase – for this particular realization – the output of the change detector starts to shoot off – signaling the change – from 14, i.e., from the location corresponding to the path-length of 14 ∗ 50 m's.

looked at whether there is a traffic sign of a type mentioned above, or not. To mimic the real traffic sign data collection and to closely follow the change detection task outlined above – in respect of the topographical aspect of the road environment – the traffic sign types listed in the fourth column of Table 1 were used as marks. For the purpose of the synthetic traffic sign data experiments, the length of each route-segment was chosen to be 50 m.

The parameters for generating the marked binomial point processes – in a particular experiment – for virtual road environments VRE0 (a virtual road environment with many winding roads) and VRE1 (a virtual road environment with few winding roads) are given in Table 1. In the first column of the table, the identifiers of the road environments, while in the second column, the collective names for the traffic sign types that are shown in the fourth column are given. In the third column, the probabilities θ_0 and θ_1 are shown. The probabilities in the fifth column are the conditional probabilities of the individual traffic sign types, the condition being that a traffic sign from the group actually occurs within the given route-segment. Using the notation utilized in Subsection 3.1, these probabilities are denoted with $p_{0,1}, ..., p_{0,4}$ and $p_{1,1}, ..., p_{1,4}$ in Table 1.

Comparing the corresponding group probabilities and the corresponding conditional probabilities given in the table, one can see that the probability of the occurrence a traffic sign from the group is considerably higher in case of virtual road environment VRE0 than in case of VRE1. Also, the most probable and least probable traffic sign types are swaped between the two virtual road environments. For a particular realization of the model incorporating two marked binomial point processes – using the parameters given in Table 1 – with a particular choice of the change-location, the output of the Page-Hinkley detector is shown in Fig. 2. Though, threshold selection for the Page-Hinkley detector is far from being obvious, a pragmatic choice in this case could be 4.0, which is approximately the double of the amplitude values of the initial peaks in the detector output.

4.2. Gathering real traffic sign data

Several traffic sign data collection car trips within Hungary were organized and carried out in conjunction with this work. A tablet-based Android application that had been developed earlier for geo-tagged data recording was modified and used for entering traffic sign type data manually – via its graphical user interface – while the geo-tagging and the trajectory data collection was carried out in an automated fashion by the mentioned application. The trajectory and the traffic sign data were saved in a joint comma-separated values (CSV) file. Such files can be loaded into spreadsheet, statistical/mathematical and geographical information system (GIS) applications, such as MS Excel, MATLAB and QGIS, respectively, for statistical analysis and for the graphical presentation of the results.

The data collection personnel consisted of two persons for each car trip, namely a driver and a note-taker. The driver drove the car along the pre-defined route. Driving safely was her sole task. The note-taker was responsible

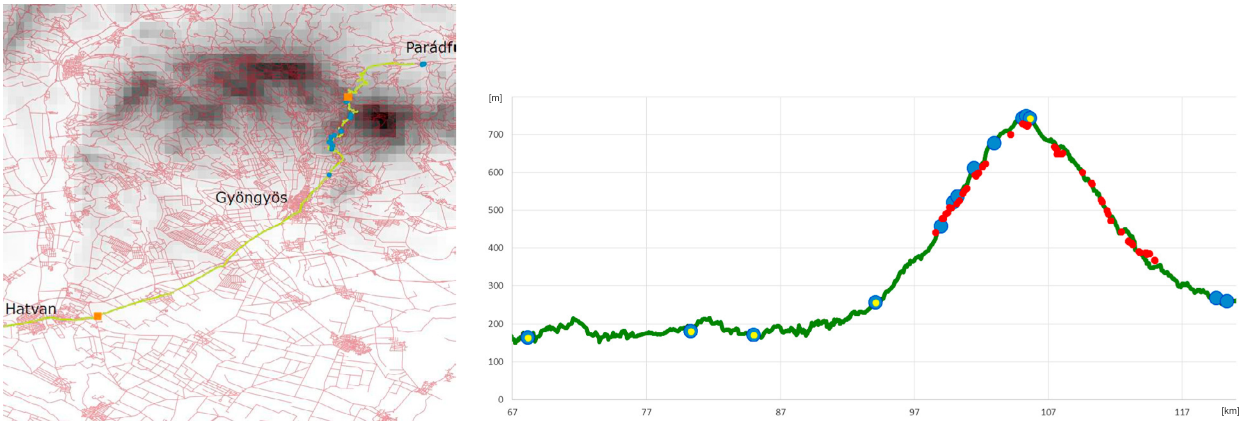


Fig. 3. The driving from the flat land of Jászság to Parádfürdő via the Mátra Mountains. The road network, as well as the surface elevation are indicated in the map. (left) The Mátra Mountains – north of Gyöngyös – appear as a dark grey blob. The blue dots mark the locations of the ”Bend to left/right” and ”Double bend starting left/right” traffic signs along the route. The orange squares mark the beginning and the end of the route section that was analyzed with the Page-Hinkley detector. The vertical profile – with the traffic sign locations marked on it – of the route section from the Jászság to Parádfürdő via the Mátra Mountains. (right) Note that the traffic sign locations for the return leg of the trip are also indicated in the same profile. Apart from the blue dots mentioned for the map, the yellow dots mark the ”Danger of skidding” traffic signs, while the red dots mark the ”Curve marker arrows”.

for the data-entries and used the mentioned tablet-based application to enter the type – and thereby via geo-tagging, the location – of the traffic sign seen en route. A forward viewing GoPro Hero3+ camera was installed onto the windscreen to record the road and its surroundings in time-lapse mode. The photograph shown in Fig. 1 was also taken the mentioned camera. Starting, stopping the camera and checking its proper functioning were also the note-taker's tasks.

The topographically more interesting part of a Budapest to Parádfürdő – via the Jászság and Mátra Mountains – and back to Budapest route is presented [1] in Fig. 3. The route section shown in the figure traverses two different topographical road environments: the relatively flat Jászság and the hilly region of the Mátra Mountains. The latter's highest point – at Mount Kékes – is 1014 m above sea level. For traversing the Mátra Mountains, the main road No. 24 was used. The geographically highest settlement and location along this road is Mátraháza, which lies about 700 m above sea level. Hence, the highest point in the vertical route profile shown in Fig. 3 corresponds to this settlement.

5. Detecting change in marked point processes associated with traffic signs

In the context of traffic signs set up along a route, the stochastic model will refer to the logged occurrences of traffic signs. The traffic signs are characterized with their types and their locations. An approach conceptually similar to ours was applied to creating the shortest possible summaries of large computer event logs in Kiernan and Terzi (2009). Herein, the change detection problem, i.e., the problem of finding a single transition between two different neighboring environments, associated with driving from one topographical area to another is considered. In order to detect environmental change along a route and to locate the environmental change point on the road, the Page-Hinkley detector is used in conjunction with the mentioned traffic sign data. The on-the-fly minimization method – used in the Page-Hinkley detector – of the overall code-length can be easily adapted to given environmental boundary detection problem. Having elaborated the necessary formulae of the particularized detector in Subsection 3.1, these are now put to work to analyze the synthetic and the gathered real traffic sign data. Fig. 4 shows the output of the Page-Hinkley detector for the route section shown in Fig. 3.

[1] The map presented in Fig. 3 was created using the QGIS Geographical Information System developed by QGIS Development Team (2015). The road data layer presented there comes from the map database contributed by OpenStreetMap contributors (2015), while the elevation data is from the Digital Elevation Model over Europe from GMES RDA project (2015)).

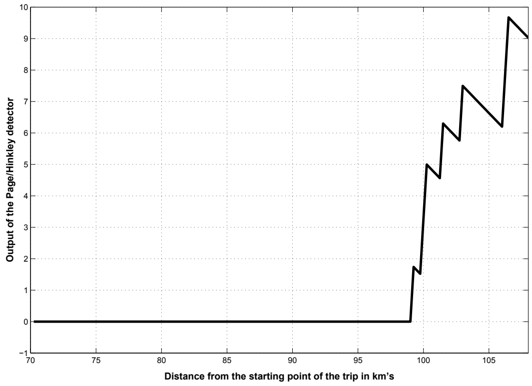
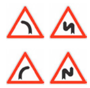
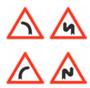


Fig. 4. The output of the Page-Hinkley detector computed for the traffic sign data shown in Fig. 3. The parameters of the two marked processes used for the flat land and for the hilly region are given in Table 2. The border point between the two topographical environments – on the route – is at 98 km (i.e., near Gyöngyös).

Table 2. Stochastic model parameters for a road on flat land and for a road in the hills.

Road environment	**Traffic sign group**	**Group probability**	**Traffic sign type**	**Conditional probability**
flat land	Traffic signs signifying dangerous bends	$\theta_0 = 0.03$		$p_{0,1} = 0.5$ $p_{0,2} = 0.5$
hilly region	Traffic signs signifying dangerous bends	$\theta_1 = 0.1$		$p_{1,1} = 0.5$ $p_{1,2} = 0.5$

The model parameters were selected based on the traffic sign data collected during earlier trips.The model parameters for the two road environments are summarized in Table 2. The length of each route-segment was chosen to be 250 m. A good qualitative agreement was found between the GPS-based altitude-profile of the data collection route – thresholded manually – and the step signal corresponding to the two topographical environments derived from the output signal of the Page-Hinkley detector also via thresholding. A more objective evaluation of the change detection could be achieved if the route was thresholded in an automatic fashion based on its altitude-histogram, see details of the thresholding method in Cseke and Fazekas (1990).

6. Some practical recommendations

During the data collection, it was noticed that the bend to the left/right, double bend starting left/right traffic signs were set up only in a small percentage of dangerous bends along the route shown in Fig. 3. The authors consider these omissions dangerous and call upon the road safety authorities to take appropriate measures to improve the situation.

In the driving direction towards Parádfürdő, the occurrences of the aforementioned traffic signs proved satisfactory to signal a change in the topographical environment, but during the back leg of the trip – driving on the same road – very few of the mentioned traffic signs were seen. Clearly, these omissions reduce the reliably of the change detection. However, the curve marker arrows – see Fig. 5 – were set up in abundance along the dangerous bends in both directions, so relying also on these traffic signs will result in a more reliable change detection. Furthermore, as such signs are set up near roundabouts and other dangerous abrupt bends, as well, the designers and the manufacturers of TSR systems should consider including these signs in the target-list of their systems.

Fig. 5. Curve marker arrow pointing right and left, respectively.

7. Conclusions

Logged occurrences of traffic signs were utilized for the purpose of detection of change in the topographical aspect of the road environment along a route, and of locating the topographical boundary. The logs are seen as realizations of marked binomial point processes and the MDL approach is used for change detection. In order to cater for efficient implementations of the topographical – and other – road-environment change detection solutions, the on-the-fly minimization method utilized in the Page-Hinkley detector was adopted. Change detection results – in respect of a stochastic model tuned to empirical traffic sign data collected beforehand – were presented in the paper. A good agreement was found between the thresholded altitude-profile of the presented flat land to hilly area route and its MDL-based topographical segmentation.

Acknowledgements

The authors acknowledge the financial support received for the research presented herein from National Research, Development and Innovation Office of Hungary in the frame of the VKSZ_14-1-2015-00125 project entitled ”Partially automated vehicle-platform with safety and eco-drive functions”.

References

Baddeley, A., 2007. Spatial point processes and their applications, in: Baddeley, A., Bárány, I., Schneider, R., Weil, W. (Eds.), Stochastic Geometry. Springer, Berlin. volume 1892 of *Lecture Notes in Mathematics*, pp. 1 – 75.

Basseville, M., 1988. Detecting changes in signals and systems – a survey. Automatica 24, 309 – 326.

Basseville, M., Nikiforov, I.V., 1993. Detection of abrupt changes: theory and application. volume 104. Prentice Hall, Englewood Cliffs.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. John Wiley & Sons.

Cseke, I., Fazekas, Z., 1990. Comments on gray-level thresholding of images using a correlation criterion. Pattern Recognition Letters 11, pp. 709–710.

Dennis, E.P., Wallace, R., Reed, B., 2015. Crowdsourcing transportation systems data. Michigan Department of Transportation & Center for Automotive Research. MI, USA.

Edquist, J., Rudin-Brown, C.M., Lenn, M., 2009. Road design factors and their interactions with speed and speed limits. Technical Report 298. Monash University Accident Research Centre. Melbourne, Australia.

Gerencsér, L., Baikovicius, J., 1989. Detection of change in ARMA parameters, in: Proceedings of the 28th IEEE Conference on Decision and Control, IEEE, Tampa, FL, USA.

Gerencsér, L., Baikovicus, J., 1992. Change-point detection as model selection. Informatica 3, pp. 3 – 20.

Gerencsér, L., Prosdocimi, C., 2011. Input-output properties of the Page-Hinkley detector. Systems and Control Letters 60, pp. 486 – 491.

Gerencsér, L., Prosdocimi, C., Vágó, Z., 2013. Change detection for finite dimensional Gaussian linear systems – a bound for the almost sure false alarm rate, in: Proceedings of the European Control Conference, EUCA. pp. 950 – 955.

GMES RDA project, 2015. Digital elevation model over Europe. URL: `http://www.eea.europa.eu/data-and-maps/data/eu-dem`.

Hinkley, D., 1971. Inference about the change-point from cumulative sum tests. Biometrika 58, pp. 509 – 523.

Kiernan, J., Terzi, E., 2009. Constructing comprehensive summaries of large event sequences. ACM Transactions on Knowledge Discovery from Data 3. Article no. 21.

Lai, T.L., 1995. Sequential changepoint detection in quality control and dynamical systems. Journal of the Royal Statistical Society 57, 613 – 658.

Lorden, G., 1971. Procedures for reacting to a change in distribution. The Annals of Mathematical Statistics 42, pp. 1897 – 1908.

Nikoliè, Z., 2014. Advances in Embedded Computer Vision. Springer International Publishing. chapter Embedded vision in advanced driver assistance systems. Advances in Computer Vision and Pattern Recognition, pp. 45 – 69.

OpenStreetMap contributors, 2015. Road network in Hungary. URL: `http://planet.openstreetmap.org`.

Page, E.S., 1954. Continuous inspection schemes. Biometrika 41, pp. 100 – 115.

Poor, H.V., Hadjiliadis, O., 2008. Quickest Detection. volume 40. Cambridge University Press.

QGIS Development Team, 2015. QGIS Geographic Information System. URL: `http://qgis.osgeo.org`.

Rahmani, S.B., 2012. Early model of traffic sign reminder based on neural network. Telkomnika 10, pp. 749 – 758.

Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, pp. 465 – 658.

Rissanen, J., 1998. Stochastic Complexity in Statistical Inquiry. volume 15 of *World Scientific Series in Computer Science*. WSP, NJ, USA.