Transportation Research Part C 74 (2017) 150–167

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A new methodology for vehicle trajectory reconstruction based on wavelet analysis

Mehdi Rafati Fard, Afshin Shariat Mohaymany *, Matin Shahri

Traffic Research Laboratory, School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran

ARTICLE INFO

Article history:
Received 8 April 2016
Received in revised form 9 November 2016
Accepted 10 November 2016
Available online 25 November 2016

Keywords:
Vehicle trajectory
Noise removal
Wavelet analysis
Outliers
NGSIM

ABSTRACT

Vehicle trajectories with high spatial and temporal resolution are known as the most ideal source of data for developing innovative microscopic traffic models. Aside from the method applied for collecting the vehicle trajectories, such data are more or less error-infected. The ever-increasing noise amplitude during the process of deriving the data (such as speed and acceleration) required for developing models, might change or even hide the structure of data and lead to useful information being overlooked. This highlights the importance of presenting the efficient methods which are adequate to remove noise and enhance the quality of vehicle trajectory data. Accordingly, in this paper a simple two-step technique based on wavelet analysis has been recommended for filtering errors and reconstructing trajectory data. Primarily, by using wavelet transform a special treatment was employed to identify and modify the outliers. Next, the noise in trajectory data was eliminated by applying the wavelet-based filter. The results of applying the proposed method to the synthetic noise-infected trajectory and the NGSIM dataset reveal how appropriate its performance is compared with other methodologies in terms of quantitative criteria.

1. Introduction

Progress in technology and the possibility of collecting vehicle trajectory data with high frequency have widened new horizons in understanding the driving behavior and developing, calibrating and validating the microscopic traffic flow models (e.g. car-following, lane-changing and gap-acceptance). As a consequence, a more realistic understanding of the driver's behavior and traffic flow dynamics is apparent. Traffic trajectories are temporally and spatially more accurate and precise compared with other types of traffic data and enable calculating many traffic flow variables. With respect to the fact that the real world data is scarcely noiseless (or negligible with little noise), extracting useful information from raw data often requires the theoretically reasonable and robust noise removal methods.

Vehicle trajectory data constitutes a series of discrete positional location of vehicles in consecutive time intervals. The existing procedures of collecting vehicle trajectory data can be categorized into two main classes: (1) extended floating car (ExFC) and (2) aerial photography. In the former, the relative speed, acceleration and spacing of equipped vehicle to surrounding ones are recorded continuously with high frequency as the vehicle moves in traffic flow. The recorded data are then used to obtain the speed, acceleration and positional location of leader and follower vehicle (Brackstone et al., 1999; Ma and Andreasson, 2005). An alternative extended floating car approach exists in which, the vehicles are equipped with highly accurate GPS and move as platoons in traffic flow. Giving the positional location of vehicles through GPS raw data, the speed,

* Corresponding author.
E-mail address: shariat@iust.ac.ir (A. Shariat Mohaymany).

http://dx.doi.org/10.1016/j.trc.2016.11.010

acceleration and inter-vehicle spacing are calculated (Gurusinghe et al., 2002; Punzo and Simonelli, 2005). The possibility of long-term data collection, covering various traffic conditions, and providing accessibility to supplementary data, i.e. driver's age or gender as well as the possibility of recording data related to interaction between driver and vehicle (addressed as Nanodata) using CAN-bus is regarded as the substantial distinction of this approach. Notwithstanding the adequacy of data obtained through the mentioned approach to compare the microscopic traffic models as well as develop the nanoscopic ones, the data might be influenced by the experimental conditions.

The second approach, which has been employed extensively to collect the trajectory data, is through video recording of a segment and extracting the trajectories using image processing and tracking algorithms. Aerial photography is basically conducted by mounting the camera on a fixed position (FHWA, 2004; Wei et al., 2005; Xin et al., 2008) or on-board aerial platform (Hoogendoorn et al., 2003; Smith, 1985; Treiterer, 1975). The principle advantage of this method is it allows a large amount of trajectory data to be extracted and consequently, macroscopic and microscopic description of traffic flow is obtained. Nevertheless, short length of segment and accordingly, the short temporal period of each trajectory can be regarded as the limitations of this method.

Apart from the methodology applied to collect vehicle trajectories, imposing the measurement and processing errors to raw data is inevitable. While measurement errors are produced as random noise on positional location of vehicles (or other variables which are directly recorded), processing errors occur due to process of numerical derivative of trajectory data which is required to extract the speed and acceleration. The derivative is essential because most of the microscopic traffic models such as car-following models (e.g. stimuli-response (Gazis et al., 1961), intelligent drive model (Treiber et al., 2000), safety distance (Gipps, 1981) and optimal velocity (Bando et al., 1995)) and lane-changing ones (e.g. discrete choice (Toledo et al., 2003) and rule-based (Gipps, 1986)) require speed and acceleration data in addition to the positional location of vehicles. Depending on the required accuracy, different methods of numerical derivation have been suggested. Approximation to the first and second-order derivative of f(x) and the related errors using five-point stencil approach have been indicated in Eqs. (1) and (2) respectively. It can be inferred that, the first and second derivative errors are four-order. The resulting numerical derivative will be negligible because the trajectory data is commonly recorded with high temporal resolution (0.1 s).

$$f'(x) = \frac{f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)}{12h} + O(h^4) \tag{1}$$

$$f''(x) = \frac{-f(x+2h) + 16f(x+h) - 30f(x) + 16f(x-h) - f(x-2h)}{12h^2} + O(h^4) \tag{2}$$

The main complexity while applying the trajectory data is the ever-growing amplitude related to random noise of positional location which occurs due to numerical derivative (Punzo et al., 2011; Toledo et al., 2007; Treiber and Kesting, 2013). Consider two consecutive observations of the positional location of the same vehicle, including the actual values plus relatively little noise (X1 and X2 are the location of first and second location). The corresponding noise values (ε_1 and ε_2) are independent with probability density function f(x). In order to obtain speed, derivative approximation through forward differencing technique is implemented as Eq. (3).

$$V(t) = \frac{[X_2(t+\Delta t) + \varepsilon_2(t+\Delta t)] - [X_1(t) + \varepsilon_1(t)]}{\Delta t} \tag{3}$$

The numerator of Eq. (3) constitutes a deterministic $(X_2(t+\Delta t) - X_1(t))$ and a random component $(\varepsilon_2(t+\Delta t) - \varepsilon_1(t))$. If the value related to random variables, ε_1 and ε_2 with corresponding probability density function, $f_{\varepsilon_1}(x)$ and $f_{\varepsilon_2}(x)$ is known, the probability density function of a new random variable $(z = \varepsilon_2 - \varepsilon_1)$ is acquired based on Eq. (4) which is the convolution of $f_{\varepsilon_2} * f_{-\varepsilon_1}$.

$$P_Z(z) = (f_{\varepsilon_2} * f_{-\varepsilon_1})(z) = \int_{-\infty}^{+\infty} f_{\varepsilon_2}(\varepsilon) f_{-\varepsilon_1}(z-\varepsilon) d\varepsilon \tag{4}$$

Assuming Gaussian noises with mean zero and standard deviation σ, the integration of Eq. (4) follows a Gaussian distribution with standard deviation equal to $\sqrt{2}\sigma$ (Eq. (5)). The result is then divided by the denominator (Δt) and the resulting noise will follow a stochastic Gaussian distribution with mean zero and standard deviation $\sqrt{2}\sigma \times \Delta t$.

$$P(z) = \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} \exp\left(\frac{-z^2}{2\left(\sqrt{2}\sigma\right)^2}\right) \tag{5}$$

Accordingly, it is evident that the noise amplitude grows noticeably by the sequential derivative of positions. For instance, a Gaussian noise with the mean and standard deviation corresponding to 0 and 5 cm in trajectory position would impose an extreme error to speed (with mean 0 and standard deviation, 0.7 m/s) and acceleration (with mean 0 and standard deviation, 10 m/s^2).

The negative effects of noise in the calibration of microscopic traffic models has thoroughly been addressed in literature. The amount of effects differ depending on methodologies applied in the calibration process (direct or indirect). According to