Transportation Research Part C 66 (2016) 99–118

Contents lists available at ScienceDirect

Transportation Research Part C

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

# Real-time traffic state estimation in urban corridors from heterogeneous data

Alfredo Nantes [a,*], Dong Ngoduy [b], Ashish Bhaskar [a], Marc Miska [a], Edward Chung [a]

[a] Smart Transport Research Centre, Queensland University of Technology, QLD-4001 Brisbane, Australia
[b] Institute for Transport Studies, The University of Leeds, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

*Article history:*
Received 19 December 2014
Received in revised form 23 June 2015
Accepted 7 July 2015
Available online 21 July 2015

*Keywords:*
Urban traffic estimation
Incremental extended Kalman filter
Multi-source data fusion
First order traffic model

ABSTRACT

In recent years, rapid advances in information technology have led to various data collection systems which are enriching the sources of empirical data for use in transport systems. Currently, traffic data are collected through various sensors including loop detectors, probe vehicles, cell-phones, Bluetooth, video cameras, remote sensing and public transport smart cards. It has been argued that combining the complementary information from multiple sources will generally result in better accuracy, increased robustness and reduced ambiguity. Despite the fact that there have been substantial advances in data assimilation techniques to reconstruct and predict the traffic state from multiple data sources, such methods are generally data-driven and do not fully utilize the power of traffic models. Furthermore, the existing methods are still limited to freeway networks and are not yet applicable in the urban context due to the enhanced complexity of the flow behavior. The main traffic phenomena on urban links are generally caused by the boundary conditions at intersections, un-signalized or signalized, at which the switching of the traffic lights and the turning maneuvers of the road users lead to shock-wave phenomena that propagate upstream of the intersections. This paper develops a new model-based methodology to build up a real-time traffic prediction model for arterial corridors using data from multiple sources, particularly from loop detectors and partial observations from Bluetooth and GPS devices.



## 1. Introduction

Model-based traffic state estimation addresses the problem of estimating quantities from sensor data that are not directly observable, but that can be inferred, such as density and flow. Sensors carry only partial information about those quantities, and their measurements are corrupted by noise. Model-based traffic state estimation seeks to recover state variables from the data and effectively capture various traffic phenomena such as stop-and-go waves, synchronized flow and oscillatory flow. Probabilistic state estimators are typically implemented through Kalman Filters (KFs) and their extended versions such as Extended Kalman Filters (EKFs) and the Unscented Kalman Filters (UKFs). Other approaches to traffic state estimation are based on Monte Carlo methods, such as Mixture KFs (Sun et al., 2003) which is embedded in a simple Cell Transmission Model (CTM) developed by Daganzo (1994), Ensemble KFs (Work et al., 2010) which is embedded in a modified (discrete) version of the first order traffic model developed by Lighthill–Whitham–Richards, hereafter LWR for short (Lighthill and

* Corresponding author. Tel.: +61 (0)7 3138 1461.
*E-mail addresses:* a.nantes@qut.eud.au (A. Nantes), d.ngoduy@leeds.ac.uk (D. Ngoduy).

http://dx.doi.org/10.1016/j.trc.2015.07.005
0968-090X/

Whitham, 1955; Richards, 1956), Particle filters (PFs) adopted in second order macroscopic traffic models (Mihaylova and Boel, 2004; Hegyi et al., 2007; Ngoduy, 2011a) as well as first order models (Nantes et al., 2013), etc. These techniques form a family of estimators known as *on-line Bayesian observers* that have been extensively used for traffic applications. Wang and Papageorgiou (2005) and Wang et al. (2006) showed how a (linearized) second order macroscopic model, termed METANET (Messmer, 2001), can be incorporated into the EKF algorithm. This EKF-METANET framework has been extensively applied in large freeway network (Wang and Papageorgiou, 2007). The EKF algorithm has also been used with LWR-type models for freeway network estimation problems in which the Lagrangian discretization method has been embedded (Yuan et al., 2012). Other research has been devoted to the application of EKFs and UKFs for the estimation of multi-class traffic in freeways (van Lint et al., 2008; Ngoduy and Sumalee, 2010; Ngoduy, 2011b; Yuan et al., 2014). Hegyi et al. (2006, 2008) compared the performance of different filters such as the UKF, EKF and PFs for freeway traffic state estimation and reported that the accuracy of these methods is equivalent.

Although a number of methods have been proposed to fuse data from mobile sensors, most recent research in this area has been conducted within the freeway context and has not been applied to the estimation of the urban flow. In general, traffic state estimation problems in arterial corridors are substantially more difficult than in freeways, due to the lack of dedicated data collection systems and the complexity of the traffic dynamics at intersections. Traditionally, inductive loop detectors were used to collect traffic data. However, installing and maintaining a dense network of such detectors is costly. The recent rapid advances in information technology have led to various cost-effective data collection systems which have enriched the sources of empirical data for the traffic state estimation problems. Such new data are usually collected from floating cars, cell-phones, video cameras and other remote sensing techniques. As indicated in van Lint and Hoogendoorn (2009,), fusing data from multiple sources could yield better accuracy, increased robustness and confidence about the estimation. Yet, most approaches to traffic state estimation have considered a single source of data. One of the challenges of fusing multi-source data stems from the fact that different sources usually provide information at different spatio-temporal resolutions. A few approaches have been recently proposed to reconstruct the traffic state from heterogeneous data but most of these methods belong to the class of data-driven assimilation techniques and are limited to the freeway context only (van Lint and Hoogendoorn, 2009; Treiber and Kesting, 2009). The main disadvantage of the data-driven approach is that, in general, it fails to generalize outside the context within which the algorithm has been trained. Recently, Deng et al. (2013, 2014) have devoted to the development of an efficient Newell-type model for the multiple source data fusion problems in freeways. As such, the measurement equations linking density to measurements are different from the ones proposed in our work. There is a fundamental difference between Deng et al. (2013, 2014) and our work, that is, the former uses an offline estimator and, as such, the problem of time synchronization between EKF time steps and sampling period (e.g. of the Bluetooth scanners) of the scanners disappears. This is so because all pairs of expected-measured values can be treated independently from each other. Moreover, in Deng et al. (2013, 2014) the estimation can only be carried at the end of the observation horizon. A few studies have been conducted towards the estimation/prediction of arterial traffic but not yet to the real-time estimation problems using data from different sources (Wu and Liu, 2011; Lu et al., 2014; Bhaskar et al., 2014; Feng et al., 2014; Wu and Liu, 2014; Seo et al., 2015).

This research departs from the current state-of-the-art by developing an efficient data fusion technique, namely Incremental EKF, which incorporates a (discrete) LWR model for urban corridors, in order to deal with the problem of estimating traffic from multiple data sources. The GPS technology enables the collection of instantaneous speeds information at any location, whereas the Bluetooth technology (BT) scanners allow measuring the travel times or average speeds along any link connecting two scanned locations. These data sources are typically available at different rates. For instance, the data from inductive loops is often aggregated in fairly long periods of time (e.g. 1–5 min). On the other hand, the BT scanners go through scanning cycles of at least six seconds (Peterson et al., 2006). Finally, the frequency at which the GPS data is sampled may vary depending on the settings of the GPS device. Recently, there has been some attempts to build up efficient model based solutions to deal with such emerging data, for example Seo and Kusakabe (2015) have used a second order and a first order traffic model, respectively, to estimate traffic states from GPS data (collected from probe vehicles or mobile sensors). Nevertheless, few attempts have been undertaken to build up a successful model based approach to fuse data from heterogeneous sources. We refer the reader to Bachmann et al. (2013) for more details.

In this paper, we propose a new version of the EKF which allows incorporating the heterogeneous measurements incrementally, whenever they become available. The idea behind such incremental EKF is simple and is based on the fact that all measurements can be considered independent, given the state of the system and the system inputs. This assumption is discussed in details in Section 4. This newly proposed method is tested using synthetic data generated from a realistic microscopic traffic simulator (AIMSUN[1]). The validation is based on a similarity measure which uses a simple image-processing technique in order to compare true and estimated spatio-temporal density diagrams.

The rest of this paper is organized as follows. Section 3 describes the arterial traffic model used, which accounts for the flow behavior at the intersections. In Section 4, we present a state-space form of the whole system and derive the transition and sensor models to be used for the implementation of the Bayesian estimator. In this section, we show how traffic data with different spatio-temporal resolutions can be incorporated into a single estimator. Section 5 pertains to the derivation

[1] http://www.aimsun.com/.