Transportation Research Part C 72 (2016) 254–271

Contents lists available at ScienceDirect

Transportation Research Part C

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

Collecting ambient vehicle trajectories from an instrumented probe vehicle
High quality data for microscopic traffic flow studies

Benjamin Coifman[a,*], Mo Wu[b], Keith Redmill[b], Douglas A. Thornton[c]

[a] The Ohio State University, Joint Appointment with the Department of Civil, Environmental, and Geodetic Engineering, and the Department of Electrical and Computer Engineering, Hitchcock Hall 470, 2070 Neil Ave, Columbus, OH 43210, United States
[b] The Ohio State University, Department of Electrical and Computer Engineering, United States
[c] Battelle Memorial Institute, United States

ARTICLE INFO

Article history:
Received 29 February 2016
Received in revised form 11 June 2016
Accepted 1 September 2016

Keywords:
Highway traffic
Empirical data
Microscopic data set
Traffic flow theory
Car following
Lane change maneuvers
Congested traffic
LIDAR

ABSTRACT

This paper presents the methodology and results from a study to extract empirical microscopic vehicular interactions from a probe vehicle instrumented with sensors to monitor the ambient vehicles as it traverses a 28 mi long freeway corridor. The contributions of this paper are two fold: first, the general method and approach to seek a cost-effective balance between automation and manual data reduction that transcends the specific application. Second, the resulting empirical data set is intended to help advance traffic flow theory in general and car following models in particular. Generally the collection of empirical microscopic vehicle interaction data is either too computationally intensive or labor intensive. Historically automatic data extraction does not provide the precision necessary to advance traffic flow theory, while the labor demands of manual data extraction have limited past efforts to small scales. Key to the present study is striking the right balance between automatic and manual processing. Recognizing that any empirical microscopic data for traffic flow theory has to be manually validated anyway, the present study uses a "pretty good" automated processing algorithm followed by detailed manual cleanup using an efficient user interface to rapidly process the data. The study spans roughly two hours of data collected on a freeway during the afternoon peak of a typical weekday that includes recurring congestion. The corresponding data are being made available to the research community to help advance traffic flow theory in general and car following models in particular.

1. Introduction

This paper presents the methodology and results from a study to extract empirical microscopic vehicular interactions from an instrumented probe vehicle. The contributions of this paper are two fold: first, the general method and approach to seek a cost-effective balance between automation and manual data reduction that transcends the specific application. Second, the resulting empirical data set is intended to help advance traffic flow theory in general and car following models in particular.[1]

* Corresponding author.
E-mail address: Coifman.1@OSU.edu (B. Coifman).
[1] As of publication the data will be posted to Coifman (2016).

http://dx.doi.org/10.1016/j.trc.2016.09.001

Like many data reduction problems this work seeks to find a balance between automated and manual processing. All too often in problems like these the data reduction approach is chosen a priori to be either strictly automated or strictly manual. Usually manual approaches are inexpensive to develop but the recurring labor costs are expensive. Automated systems are usually inexpensive to run but only perform well when conditions meet expectations. In general the marginal cost becomes progressively more expensive for each unit increase in performance from an automated system. A robust automated system can prove to be quite expensive and may still yield a non-negligible error rate. So to ensure top quality data even the best automated system would need a human in the loop to validate the results. Herein lies the key methodological insight of our work: if the human is already in the loop to validate the data, provided care has been taken to develop the right time-efficient tools for the user, the marginal costs to have this human to actively clean the results should be small compared to the savings that can be realized in the automated system. If done right, a "pretty good" automated system to do the majority of the processing followed by supplemental manual cleaning (i.e., over and above simple validation) can produce a high quality data set that is beyond the capabilities of a "superior" automated system on its own while only encumbering a fraction of the labor costs from a "purely manual" approach.[2] The generic approach to seek a cost effective balance between the power of the automated system and demands on the *human in the loop* transcends the particular application to instrumented probe vehicle data.

This generic methodological approach allows us to overcome many of the constraints that have limited previous efforts in the specific area of collecting empirical microscopic vehicle interaction data. Historically automatic data extraction has not provided the precision necessary to advance traffic flow theory,[3] while the labor demands of manual data extraction have limited past studies to small scales.[4]

The data for this study come from a probe vehicle that is instrumented with positioning sensors for localization and ranging sensors to perceive the ambient vehicles and measure inter-vehicle relationships over time and space as the probe travels through the traffic stream. Specifically, this work develops the process of extracting vehicle trajectories from data collected by three perception sensors, consisting of a pair of forward and rearward facing LIDAR sensors that each scan an arc of 180° in a plane parallel to the roadway and a single forward facing radar sensor that has greater range but only a 12° field of view. The automated processing consists of segmenting vehicle and non-vehicle returns in the LIDAR data, grouping the vehicle returns into distinct vehicles, tracking vehicles over time, and reporting the vehicle tracks in world coordinates. The radar sensor already tracks vehicles and only reports tracked targets, eliminating the post-processing based tracking for the radar data. In either case numerous errors remain in the automatically tracked target data that need to be cleaned by the human reviewer. The study data were collected over roughly two hours on a freeway during a typical weekday afternoon rush hour that includes recurring congestion. The probe vehicle's tour consists of two round trips over a 28 mi route along urban freeways. This tour passes through four separate major freeway interchanges involving: I-70, I-71, I-270, I-670, and SR-315; as well as many more interchanges with arterial roadways.

To place the work in context, the bold curve in Fig. 1a shows a hypothetical example of a trajectory from a conventional probe vehicle. It includes the speed and acceleration of the conventional probe vehicle, but no information about the environmental stimuli that the vehicle responded to, e.g., the unobserved trajectories of the other vehicles that are shown with faint curves. With the addition of the perception sensors on the probe, as per the current study, it is now possible to also collect the trajectories of the leading and following vehicles too, giving rise to the three tracked trajectories in the lane of travel, shown with bold curves in Fig. 1c. Thereby yielding a lengthy sample of the instrumented probe vehicle's response to its leader and concurrent shorter samples of many different followers behind the probe vehicle. This plot only shows the probe vehicle's lane of travel, the perception sensors are also used to track vehicles in the adjacent lanes, typically yielding at least two trajectories per immediately adjacent lane and often several trajectories in further lanes (but the far lanes are often occluded by vehicles in the intervening lanes, so coverage of the far lanes is lower than the immediately adjacent lanes).

This data collection is in response to a longstanding need for empirical microscopic traffic flow data, e.g., Haight (1963) observed that most of the literature on traffic flow theory was conceived in purely mathematical terms, with observational studies limited to supporting particular theories. While exceptions exist both then and now, most contemporary traffic flow theory is still built upon models that ultimately have purely mathematical origins, e.g., hydrodynamic models (Lighthill and Whitham, 1955; Richards, 1956) and car following models (Chandler et al., 1958; Gazis et al., 1961). Even though modern models are much more sophisticated than those when Haight made his observations over 50 years ago, the field remains limited by the quantity and quality of empirical traffic data. Plausible but incorrect hypotheses perpetuate in the absence of accurate empirical microscopic data for model development. While there have been a handful of true empirical microscopic freeway data sets collected, a great need remains for more data in general and larger varieties of conditions in particular. This need remains because the collection and reduction are extremely expensive, which in turn constrains the length of roadway monitored, the location observed, and duration of the study.

[2] Since the automated system used in our data reduction is unique to the specific raw data set and this raw data set is of finite size, it does not make sense to find the optimally efficient balance. So throughout this paper we use "pretty good" to clearly denote that while there is almost certainly a "better" approach, the chosen automated approach is good enough to ensure that indeed the subsequent manual cleaning is far less demanding than a purely manual approach.

[3] For example, the unrealistic relationships in the NGSIM data set discussed in Section 1.1 that arose from the automated processing with only cursory manual validation, or the tracking errors exhibited by a more sophisticated vehicle tracking algorithm in Fig. 6 of Coifman et al. (1998).

[4] For example, Treiterer and Myers (1974) took several years to manually track roughly 70 vehicles over 4 min and 3.3 miles.