Estimating the Number of Lane Change Maneuvers on Congested Freeway Segments

Chao Wang and Benjamin Coifman

Abstract— **It has long been recognized that lane change maneuvers (LCMs) can influence the relationships underlying traffic flow theory or even disrupt the relationships if lane change maneuvers at critical locations are not properly accounted for. However, the research in this area has been limited by the fact that currently there is no efficient method to collect the data on the times and locations of LCMs since it requires both spatial and temporal coverage. Our research seeks to estimate the number of LCMs in a given time-space region in a manner that is compatible with existing vehicle detectors. Employing Vehicle Reidentification (VRI), matching a vehicle observation at one detector station to the observation from a same vehicle at another station, this work develops a model to estimate the number of entering vehicles (N_{EN}) and the number of exiting vehicles (N_{EX}) between the trajectories of two reidentified vehicles. In effect, the work promises to allow for a direct estimate of the number of LCMs. Complete trajectory data sets are used to validate the model since these data include information on all of the LCMs. Although there are some limitations and more research is needed, the approach could eventually be used to continually estimate the number of LCMs from conventional detector stations; thereby providing new insight into travel patterns between lanes and the associated impacts.**

I. Introduction

It has long been recognized that lane change maneuvers (LCMs) can influence the relationships underlying traffic flow theory or even disrupt the relationships if lane change maneuvers at critical locations are not properly accounted for. Recent empirical research by several groups has shown that LCMs are one of the mechanisms that reduce capacity at bottlenecks, e.g., [1]. LCMs are also closely related to traffic accidents, e.g., [2-3]. Despite the large impact from LCMs, the research in this area has been limited by the fact that currently there is no efficient method to collect data on the times and locations of LCMs. Such measurement requires extensive coverage both spatially and temporally. Most studies that use LCMs from field data rely on film or video, and require labor-intensive efforts to extract the necessary information. Image processing technologies are starting to help in this task, but for accurate LCM data the labor demands remain high. Meanwhile, recognizing the need to understand LCMs, several researchers have developed linear models, nonlinear models and stochastic models in an attempt to capture the interrelations between traffic conditions and the frequency of LCMs, e.g., [4-7]. However, the outputs from these models are coarse and generally in need of large-scale validation and calibration.

Our research seeks to estimate the number of LCMs in a given time-space region in a

C. Wang is a Graduate Research Assistant in the Department of Civil and Environmental Engineering and Geodetic Science, at the Ohio State University, Columbus, OH, 43210 USA (e-mail: wang.730@osu.edu).

B. Coifman is an Associate Professor with a joint appointment in the Department of Civil and Environmental Engineering and Geodetic Science; and the Department of Electrical and Computer Engineering at the Ohio State University, Columbus, OH 43210 USA (phone: 614-292-4282; fax: 614-292-3780; e-mail: Coifman.1@OSU.edu).

manner that is compatible with existing vehicle detectors. This work develops a model to estimate the number of entering vehicles (N_{EN}) and the number of exiting vehicles (N_{EX}) between the trajectories of two non-successive vehicles. The method requires data from conventional inductive loop detectors, i.e., this method utilizes data collected at discrete points in space to estimate the number of LCMs between the detector stations. Unlike the previous work on LCMs, this estimation employs Vehicle Reidentification (VRI), the process of matching a vehicle observation at one detector station to the observation from a same vehicle at another station. Based on VRI results, [8] showed an approach to yield inflow between reidentified vehicles, i.e., N_{EN}-N_{EX}, which is a constraint on the possible values of N_{EN} and N_{EX} utilized in the present study. As shown herein, an upper bound and lower bound on N_{EN} and N_{EX} can be obtained and used to estimate N_{EN} and N_{EX}.

In effect, the work promises to allow for a direct estimate of the number of LCMs. Complete trajectory data sets are used to validate the model since these data include information on all of the LCMs. Although there are some limitations and more research is needed, the approach could eventually be used to continually estimate the number of LCMs from conventional detector stations; thereby providing new insight into travel patterns between lanes and the associated impacts.

The remainder of this paper is organized as follows. First, the VRI algorithm is briefly reviewed in Section II. Next, the method to estimate the number of LCMs is introduced in Section III. Then the data set used to validate the proposed method is presented in Section IV, followed by an illustration of the proposed method in Section V. Finally, in Section VI the paper closes with a summary and conclusions.

II. Background

Vehicle Reidentification (VRI) can provide information about the times when a given vehicle passed two (or more) locations, e.g., for vehicle a3 in Figure 1 in the form that the upstream vehicle observation at t_1 and the downstream vehicle observation at t_3 came from the same vehicle. VRI can be used to capture traffic conditions over extended links using conventional vehicle detection equipment at discrete locations, i.e., points along the roadway. Different emerging technologies have been applied for VRI, such as video and Automatic Vehicle Identification (AVI); however, the application of such VRI systems typically requires the deployment of new hardware. Utilizing the existing infrastructure, [9] uses vehicle length measurements from existing loop detector hardware to reidentify vehicles and is modeled in this study. The basic idea of the algorithm is to match vehicle sequences since a sequence of vehicle lengths is much more unique than the constituent measurements and hence provides more information than the individual vehicle lengths. The algorithm consists of three basic steps: measure vehicle length, identify possible matches, and address minor lane change maneuvers.

When a vehicle passes over a loop detector buried under the pavement, the resulting measurements can be used to estimate the length of the vehicle. However, measurement uncertainty exists because of resolution constraints arising from the relatively low sampling frequency of 60 Hz. The uncertainty is inversely proportional to velocity and true vehicle length. So, vehicle length estimate from the loop detector is expressed as a range, instead of a

single discrete value.

In the algorithm, vehicles are assigned successive arrival numbers as they pass a detector station. These numbers are assigned independently for each lane at each station (upstream and downstream). Stations are taken in adjacent pairs and lanes are examined independent of other lanes beyond the fact that vehicles may enter or leave the subject lane at any intervening location. A set of possible matches is found for each downstream observation, i.e., the set of feasible upstream observations is identified whose length ranges intersect that of the given downstream observation. The set of possible matches for each downstream vehicle is stored as a row in a matrix. The rows of the matrix are indexed by downstream vehicle number and the columns are indexed by the upstream offset, i.e., the difference between the upstream vehicle number and the downstream vehicle number. Thus, in this matrix, if there were no lane change maneuvers the true matches would all have the same upstream offset and fall in a single column since the upstream and downstream vehicle numbers would increase at the same rate. Finding the true matches is complicated by the fact that lane change maneuvers disrupt the sequences and cause the true matches to switch columns, some vehicles do not have true matches, and most rows have many possible but incorrect matches distributed randomly across the row. Provided LCMs are infrequent enough that a given sequence of true (but unknown) matches is much longer than the sequences of incorrect matches that arise, the true matches can be found by retaining matches that form long sequences. When too many LCMs occur the algorithm stops reidentifying vehicles until another lengthy sequence is observed. As a result of this process, the algorithm will identify platoons of vehicles and many non-reidentified vehicles will pass between such platoons.[1]

III. Method To Quantify LCMs

Consider the hypothetical example in Figure 1, showing the time space diagram in a single lane spanning a pair of adjacent detector stations, denoted upstream and downstream. The bold lines labeled from a1 to a6 in the figure represent six vehicles that passed both stations and are reidentified by a VRI algorithm such as presented in Section II. Since the VRI algorithm matches vehicle platoons, the reidentified vehicles will tend to fall in sequences of successive vehicles, with non-reidentified vehicles between them. Figure 1 shows two sequences of reidentified vehicles, a1 to a3 are in one sequence and a4 to a6 in another. Between the two sequences are non-reidentified vehicles labeled from b1 to b5, some of which are not reidentified because they left or entered the subject lane between detector stations while others simply are not reidentified although they passed both stations in the subject lane. Our research seeks to estimate the number LCMs in the time-space region between sequences of reidentified vehicles, that is, between the trajectories of a3 and a4 in Figure 1.

[1] For details on one process of finding the true matches see [9]. The present paper only emulates the reidentification results, the details of the reidentification process are not critical, and the only key fact is that vehicle reidentification is possible in the first place.

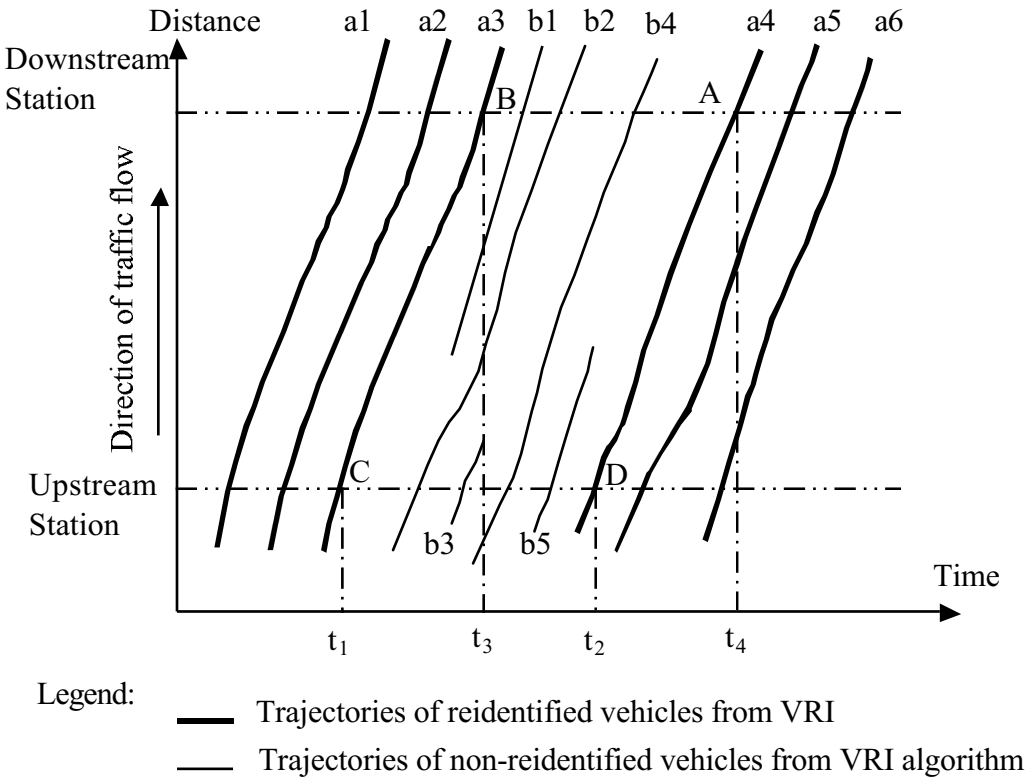

Figure 1. Time space diagram showing the method to quantify LCMs

Inflow can be estimated as equation (1) based on the algorithm presented in [8] and the t_i shown in Figure 1.

$$inflow(t_3, t_4) = (N_d(t_4) - N_d(t_3)) - (N_u(t_2) - N_u(t_1)) \quad (1)$$

Where,

$inflow(t_3, t_4)$ is the inflow between the trajectories of two reidentified vehicles that pass the downstream at time t_3 and t_4 respectively,

$N_d(t)$ is the arrival (sequential) number of the vehicle that passes the downstream station at time t, and

$N_u(t)$ is the arrival (sequential) number of the vehicle that passes the upstream station at time t.

Inflow is the difference between N_{EN} and N_{EX} between two successive reidentified vehicles and given the inflow, the set of possible N_{EN} and N_{EX} is greatly reduced. Figure 2 shows the set of feasible values when the inflow is 1.

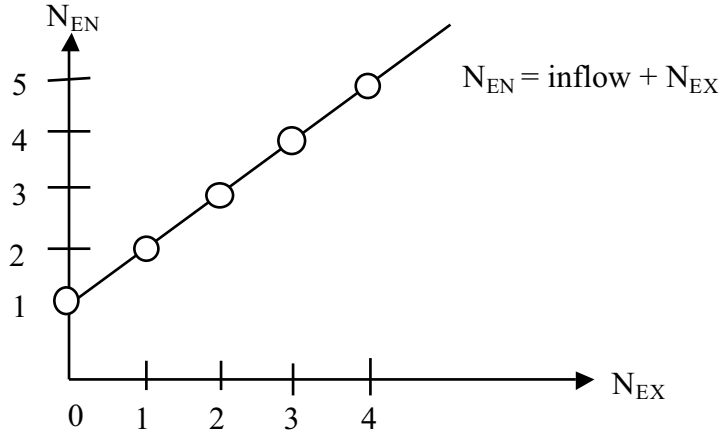

Figure 2. The relationship between inflow and feasible LCMs when the inflow is 1

Figure 2 illustrates that equation (1) constrains all the possible pair-wise solutions to the integer values, shown with circles on the straight line. Since neither N_{EN} nor N_{EX} can be negative, the lower bound for N_{EN} is $\max(0, \text{inflow})$ and for N_{EX} is $\max(0, -\text{inflow})$. Although reduced from two dimensions to one, the number of possible pair-wise solutions is still infinite since there are no upper bounds for N_{EN} and N_{EX}; but, subject to reasonable constraints upper bounds can be established. If no vehicle both enters and leaves the subject lane between the trajectories of two reidentified vehicles, then N_{EX} and N_{EN} are bounded by the number of vehicles to pass the upstream and downstream stations, respectively, between reidentified platoons. This assumption does not always hold, as some vehicles do pass through lanes quickly. But provided that the number of vehicles making such pairs of LCMs is relatively small, the assumption on the upper limits greatly simplifies the problem without omitting the true N_{EN} and N_{EX} pair from the feasible region. For example, consider the inflow between the trajectories of reidentified vehicles a3 and a4 in Figure 1. At the upstream station, there are four unidentified vehicles between a3 and a4. Based on the assumption, these are the only vehicles that can leave the lane between a3 and a4, so N_{EX} can be no larger than four. Similarly, the upper bound of N_{EN} is three because only three vehicles passed downstream between a3 and a4. As a result the solution space of N_{EN} and N_{EX} is finite based on the constraints,

$$\max(0, \mathit{inflow}) \le N_{EN}(t_3, t_4) \le N_d(t_4) - N_d(t_3) - 1 \tag{2.a}$$

$$\max(0, -\mathit{inflow}) \le N_{EX}(t_3, t_4) \le N_u(t_2) - N_u(t_1) - 1 \tag{2.b}$$

$$\mathit{inflow}(t_3, t_4) = N_{EN}(t_3, t_4) - N_{EX}(t_3, t_4) \tag{2.c}$$

Where,

$N_{EN}(t_3, t_4)$ is the number of entering vehicles between the trajectories of two reidentified vehicles that pass the downstream station at time t_3 and t_4 respectively, and

$N_{EX}(t_3, t_4)$ is the number of exiting vehicles between the trajectories of two reidentified vehicles that pass the downstream station at time t_3 and t_4

respectively.

As the VRI algorithm reidentifies more of the passing vehicles (i.e., a higher reidentification rate), constraint (2) will yield tighter solution bounds. For example, if b4 in Figure 1 were reidentified, the solution bounds between b4 and a4 are [0,0] for N_{EN} and [1,1] for N_{EX}, which leads to the true number of LCMs. Similarly, reducing matching errors in the VRI algorithm will also help to improve the reliability of constrains (2).

Where N_{EN} and N_{EX} fall within constraints (2) depends on many factors, such as traffic conditions, the presence of on or off ramps, and whether the lane is an inside lane, middle lane or outer lane. Identifying exactly where N_{EN} and N_{EX} fall within the constraints is the subject of ongoing research. For simplicity in this proof of concept study we use the mid-point of the bounds as a point estimate. So, the equations to estimate the number of LCMs are as follows,

$$N_{EN}(t_3,t_4) = \frac{1}{2}[\max(0, inflow) + N_d(t_4) - N_d(t_3) - 1] \tag{3.a}$$

$$N_{EX}(t_3,t_4) = \frac{1}{2}[\max(0, -inflow) + N_u(t_2) - N_u(t_1) - 1] \tag{3.b}$$

From equation (1), it can be shown that constraint (2.c) is met by definition if equation (3) is used to estimate N_{EN} and N_{EX}.

IV. Data Description

The present study employs a vehicle trajectory dataset provided by Federal Highway Administration's (FHWA) Next Generation SIMulation (NGSIM) project [10]. The primary reason to employ a vehicle trajectory dataset is that it includes the true LCM information, which will be compared with the estimation from the proposed method. The vehicle trajectory data were collected on a segment of interstate I-80 just north of Oakland, California between 4:00 p.m. and 4:15 p.m. on April 13, 2005. Figure 3 shows schematic diagram of the study site.

The data were collected using seven video cameras mounted on a 30-story building adjacent to I-80. Vehicle trajectory data were transcribed from the video data using a semi-automated process that automatically detected and tracked vehicles from the video images. The data include vehicle length, speed, lateral position, longitudinal position, lane identification, preceding vehicle ID and following vehicle ID for every 1/10th of a second.

The site was approximately 1,650 feet long with an on-ramp near the start at Powell St. and an off-ramp to Ashby Ave. just downstream of the study area. The segment is congested for the 15 minutes when the data were collected. The space mean speed across all 6 lanes is 19 mph, ranging from a space mean speed in lane 1 of 30 mph (a high occupancy vehicle, HOV, lane) and in lane 6 of 14 mph.

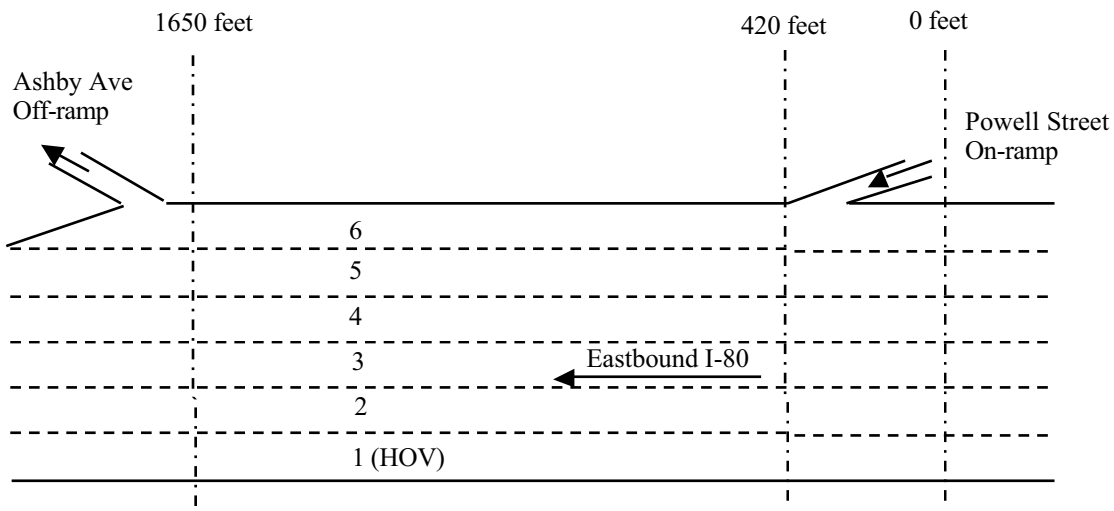

Figure 3. Schematic of study segment on I-80, not to scale (adapted from [11])

The dataset provided by NGSIM [10] also includes dual loop detector data. However, the detector stations are located beyond the segment in which the trajectory data were extracted and the loop detector data is aggregated to 30-seconds, which is incompatible with the VRI algorithm.

Ground truth LCM information is critical to evaluate the performance of the proposed method to estimate LCMs and such rich trajectory data is one of the few means that can provide such information. As a publicly available data set, the vehicle trajectory of NGSIM is uncommon due to the difficulty both in achieving an unobstructed view of a non-trivial length of freeway, and more importantly the extensive labor needed to reduce the data with sufficient precision. Once these trajectories are in hand, it is easy to establish virtual detector stations at fixed points along the road and simulate the results of the VRI algorithm. Because the true matches are already known for these data, rather than employing the VRI algorithm from Section II, for this study we emulate the algorithm by specifying which vehicles are reidentified in a manner consistent with the performance of the VRI algorithm, as explained in the following section.

V. Pilot Study Based On Field Data

Because this pilot study seeks to demonstrate the ability to estimate N_{EN} and N_{EX} and the true matches are known for each vehicle, the analysis emulates the VRI algorithm. Rather than reidentifying vehicles from the virtual detectors the VRI results were simulated based on the assumption that all platoons with size of three or more will be reidentified. This simplified process represents a VRI performance better than presently achievable by dual loop detectors, as it eliminates all incorrect matches and guarantees other matches are found. But the VRI algorithm is not the topic of this paper and these assumptions eliminate confounding factors that could otherwise obscure the estimation results in this proof of concept work.

The virtual upstream station was located at 200 feet into the study segment and the virtual downstream station was located at 1,400 feet. All 6 lanes were studied. The estimation results for lane four are shown in Figure 4. A total of 21 platoons were observed passing both

detectors in this lane over the 15 minute sample, yielding 20 pairs of N_{EN} and N_{EX}. Figure 4(a) shows the true number of N_{EN} obtained from trajectory data, the estimated N_{EN} based on the method presented in Section III, and its upper and lower bounds from constraint (2) for each of these 20 pairs (each corresponding to a time space region like ABCD in Figure 1). Figure 4(b) shows the same information for N_{EX}. The sum of all 20 estimations for N_{EN} will yield an estimation of the total number of entering vehicles in lane 4 during the 15 minutes (from 4:00 p.m. to 4:15 p.m.) between the virtual upstream and downstream stations. The spikes in Figure 4 come from the fact that the number of vehicles between two reidentified platoons (e.g., $N_d(t_4) - N_d(t_3)$ in Figure 1) varies over a large range. Figure 4(c-d) show the difference between the estimated and true value for N_{EN} and N_{EX} for each pair of platoons.

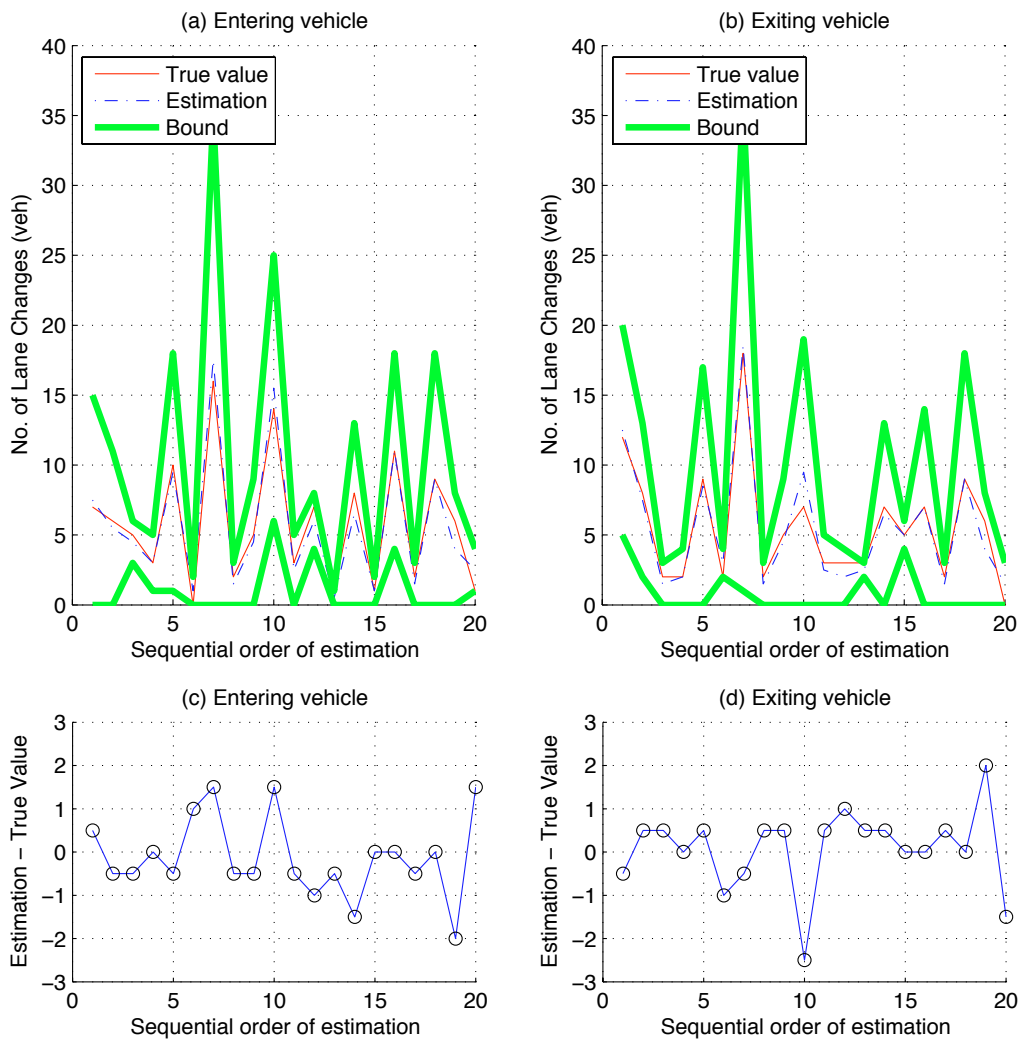

Figure 4. Estimation of the number of LCMs in lane 4

Figure 4 shows that in this case the proposed method provides a good estimation of the true number of LCMs. It also indicates that the upper bound and lower bound defined in Section III do in fact bound the true values of N_{EN} and N_{EX}.

Figure 5 shows the corresponding plots for lane 3 with the bounds removed for clarity. In this case only 20 platoons were observed, yielding 19 pairs of N_{EN} and N_{EX}. Note that three true values of N_{EN} are zero and the corresponding true values of N_{EX} are non-zero, which

means that there are only exiting vehicles to separate vehicle platoons in the corresponding time space region.

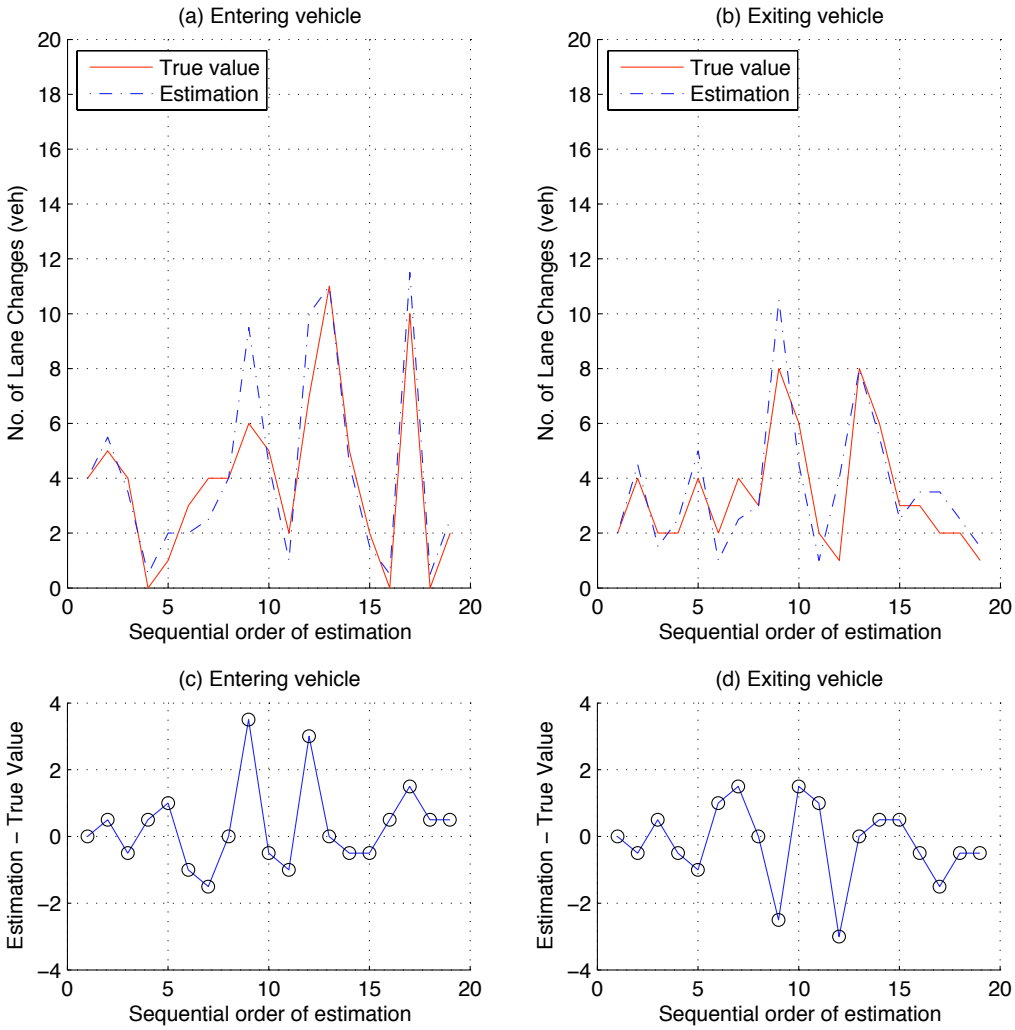

Figure 5. Estimation of the number of LCMs in lane 3

To quantify and evaluate the differences between the true and the estimated values, the following measures are applied.

$$MAE = \frac{\sum_{i=1}^{I} | \hat{N}_i - N_i |}{I} \tag{4.a}$$

$$MARE = \frac{\sum_{i=1}^{I} \frac{| \hat{N}_i - N_i |}{N_i}}{I} \tag{4.b}$$

Where,

MAE is the Mean Absolute Error,

$MARE$ is the Mean Absolute Relative Error,

$\hat{N}_i$ is the ith estimation of N_{EN} or N_{EX}, and

N_i is the ith true number of N_{EN} or N_{EX}.

In equation (4.b), when N_i is zero, $\hat{N}_i$ and N_i are removed and I is adjusted accordingly. Table 1 shows the estimation results for all six lanes from these two performance measures.

Table 1. Estimation results for all lanes

	lane 1 (HOV)	lane 2	lane 3	lane 4	lane 5	lane 6	lane 3 (VRI 5)
No. of through vehicle	335	299	197	204	163	184	197
No. of reidentified vehicles	318	250	123	96	57	72	65
Reidentification rate	95%	84%	62%	47%	35%	39%	33%
Absolute error (MAE) NEN [vehicles]	1.60	0.98	0.89	0.73	2.00	3.00	2.83
Absolute error (MAE) NEX [vehicles]	1.60	0.98	0.89	0.68	1.92	2.88	2.50
Relative error (MARE) NEN	75%	33%	27%	21%	21%	30%	32%
Relative error (MARE) NEX	50%	48%	41%	16%	26%	46%	61%
Number of estimations	5	21	19	20	13	17	9

Increasing the minimum platoon length to five vehicles rather than three, the last column of Table 1 shows the estimation results for lane 3, simulating a lower reidentification rate.

Table 1 shows that estimation results in lane 2, lane 3 and lane 4 all have MAE within 1 vehicle. Some MAREs in these three lanes are high because of infrequent LCMs. Lane 5 also yields low MARE, though the MAE is higher. Both MAE and MARE for lane 6 are high because of traffic from the on-ramp and to the off-ramp just beyond the study region. Lane 1 is a HOV lane with very few LCM and high reidentification rate, so it is not surprising to see the high MAREs in Lane 1.

VI. Conclusions

Lane change maneuvers (LCMs) are important to traffic flow theory, traffic delays, and safety. However, relatively little research has been done to study LCMs because of the difficulty in collecting LCM data. Field collection is time and labor intensive, even with the help of advanced image processing technology. Previous research has studied the relation between traffic condition and frequency of LCMs, but the model output is coarse and generally in need of large-scale validation and calibration. In response to the need for quantification of LCMs, the present study seeks to estimate the number of LCMs between the trajectories of two reidentified vehicles using the existing loop detector infrastructure. The primary advance of this work is the fact that it can estimate N_{EN} and N_{EX} while earlier work could only estimate inflow, the difference between the two parameters.

The proposed method emulates an existing Vehicle Reidentification (VRI) algorithm and estimates inflow between two reidentified vehicles. Based on the assumption that no vehicle will both enter and leave the studied lane between the trajectories of two reidentified vehicles, an upper and lower bound of N_{EN} and N_{EX} can be obtained. For simplification in this proof of concept, the mid-point of the bounds is used as an estimation of N_{EN} and N_{EX}. The proposed method was validated using vehicle trajectory data collected during congested conditions. Since no loop detector data were available, virtual detector stations were used and VRI results

were simulated based on the assumption that all platoons with size of three or more vehicles will be reidentified. The pilot study provided a direct estimate of the number of LCMs with reasonable accuracy.

The present study has some limitations, for example, the VRI algorithm was not implemented because of the lack of loop detector data for the studied segment. Another example is that the mid-point of the upper and lower bound is taken as an estimation of N_{EN} and N_{EX}, discounting the influence of traffic conditions on the frequency of LCMs. Our group has also studied the impacts of LCMs on congested freeway segment delays [12], which has the potential to be used to improve the estimation method proposed in this paper. Although there are some limitations and more research is needed, the approach could eventually be used to continually estimate the number of LCMs from conventional detector stations; thereby providing new insight into travel patterns between lanes and the impacts.

ACKNOWLEDGEMENTS

This material is based upon work supported in part by the National Science Foundation under Grant No. 0133278 and by the California PATH (Partners for Advanced Highways and Transit) Program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation. The Contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data and results presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification or regulation.

REFERENCES

[1] Coifman, B., Krishnamurthy, S., and Wang, X., (2003), Lane change maneuvers consuming freeway capacity, *Proc. of Traffic and Granular Flow 2003*, October 3, 2003, Delft, Netherlands, pp. 3-14.

[2] Golob, T.F., Recker, W, Alvarez, V., (2004), Safety aspects of freeway weaving sections, *Transportation Research A*, 38(1), pp. 35-51.

[3] Golob, T.F., Recker W.W., (2003), Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions, *Journal of Transportation Engineering*, Vol. 129, No.4, pp. 342-353.

[4] Worrall R. D. and Bullen A.G., (1970), An empirical analysis of lane changing on multiple highways, *Highway Research Record 303*, pp. 30-43.

[5] Pahl, J., (1970), Lane-change frequencies in freeway traffic flow. *Highway Research Record 409*, pp. 17-25.

[6] Sheu J-B., (1999), A stochastic modeling approach to dynamic prediction of section-wide inter-lane and intra-lane traffic variables using point detector data. *Transportation Research Part A*, 33(2), pp. 79-100.

[7] Chang G.L. and Kao Y.M., (1991), An empirical investigation of macroscopic lane-changing characteristics on uncongested multilane freeways. *Transportation Research Part A*, 25(6), pp. 375-389.

[8] Coifman, B., (2003), Estimating density and lane inflow on a freeway segment. *Transportation Research Part A*, 37(8), pp. 689-701.

[9] Coifman, B. and Cassidy, M., (2002), Vehicle reidentification and travel time measurement on congested freeways. *Transportation Research Part A*, 36(10), pp. 899-917.

[10] Federal Highway Administration, (2006), Next Generation Simulation Community (NGSIM), http://www.ngsim.fhwa.dot.gov, accessed on November 26, 2006.

[11] Cambridge Systematics, Inc., (2005), Summary report: NGSIM I-80 Data Analysis (4:00 p.m. to 4:15 p.m.)*, Technical report*.

[12] Coifman, B., Mishalani, R., Wang, C., Krishnamurthy, S., (2006), Impacts of lane change maneuvers on congested freeway segment delays: pilot study, *Transportation Research Record*, No. 1965, pp. 152-159.