UNIVERSITY
OF
CALIFORNIA

New Methods for Monitoring Spatial Truck Travel Patterns in California Using Existing Detector Infrastructure

A Research Report from the University of California Institute of Transportation Studies

Stephen G. Ritchie, Kate Kyung Hyun, and Andre Tok, Institute of Transportation Studies, UC Irvine

August 2017

PROJECT ID: UC-ITS-2017-36 | DOI:10.7922/G29P2ZVJ

ITS·IRVINE
INSTITUTE *of* TRANSPORTATION STUDIES

Technical Report Documentation Page

1. Report No. UC-ITS-2017-36	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle New Methods for Monitoring Spatial Truck Travel Patterns in California Using Existing Detector Infrastructure		**5. Report Date** August 2017
		6. Performing Organization Code ITS-Irvine
7. Author(s) Stephen G. Ritchie, PhD Kate Kyung Hyun, PhD https://orcid.org/0000-0001-7432-8058 Andre Tok, PhD		**8. Performing Organization Report No.** N/A
9. Performing Organization Name and Address Institute of Transportation Studies, Irvine 4000 Anteater Instruction and Research Building Irvine, CA 92697		**10. Work Unit No.** N/A
		11. Contract or Grant No. UC-ITS-2017-36
12. Sponsoring Agency Name and Address The University of California Institute of Transportation Studies www.ucits.org		**13. Type of Report and Period Covered** Final Report
		14. Sponsoring Agency Code UC ITS
15. Supplementary Notes DOI:10.7922/G29P2ZVJ		

16. Abstract

This study developed a methodology to accurately estimate network-wide truck flows by leveraging existing point detection infrastructure, namely inductive loop detectors. The tracking model identifies individual trucks at detector locations using advanced inductive signatures and matches vehicle pairs at detector locations, using an extended form of the Bayesian classification model to estimate matching and non-matching probabilities of the vehicle pairs Several vehicle feature selection and weighting methods including Self Organizing Map and K-means clustering were applied to better identify individual vehicles from signature data. It was shown that the proposed extensive feature processing enhanced vehicle identification performance even among vehicle pools sharing similar physical configurations. The developed model was tested along an approximately 5.5-mile freeway segment on I-5 and CA-78 in San Diego, California where only 67 percent of the total trucks were observed at both up- and down-stream detector sites. Results showed balanced performances in exactness and completeness of matching with 91 percent of correct outcomes for multi-unit trucks.

17. Key Words Trucks, vehicle detectors, loop detectors, freight traffic, traffic surveillance, spatial analysis, data collection, data processing, vehicle classification		**18. Distribution Statement** No restrictions.	
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 29	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the UC Institute of Transportation Studies

The University of California Institute of Transportation Studies (UC ITS) is a network of faculty, research and administrative staff, and students dedicated to advancing the state of the art in transportation engineering, planning, and policy for the people of California. Established by the Legislature in 1947, the UC ITS has branches at UC Berkeley, UC Davis, UC Irvine, and UCLA.

Acknowledgements

This study was made possible through funding received by the University of California Institute of Transportation Studies from the State of California via the Public Transportation Account and the Road Repair and Accountability Act of 2017 (Senate Bill 1). The authors would like to thank the State of California for its support of university-based research, and especially for the funding received for this project. The authors would also like to thank Kevin Lombard of Caltrans District 7, Horatius Petreaca of Caltrans District 8, and Leticia Vega and Tim Padilla of Caltrans District 11 for their assistance in setting up field equipment and providing access to detector locations used in this research.

Disclaimer

New Methods for Monitoring Spatial Truck Travel Patterns in California Using Existing Detector Infrastructure

UNIVERSITY OF CALIFORNIA INSTITUTE OF TRANSPORTATION STUDIES

August 2017

Stephen G. Ritchie, Kate Kyung Hyun, and Andre Tok
Institute of Transportation Studies, UC Irvine

[page intentionally left blank]

TABLE OF CONTENTS

List of Tables

List of Figures

Executive Summary

In response to the significant growth of freight movement in recent years, transportation agencies are increasingly aware of the need for detailed information on truck flows within highway network systems for traffic monitoring and operations, freight demand analysis and environmental impact studies. However, the limited availability of truck data sources complicates the capture of truck travel patterns which significantly vary by season, time of day and location. This study introduces an anonymous truck tracking method over a complex network system leveraging existing traffic detection infrastructure, namely inductive loop detectors (ILDs). The tracking model identifies individual trucks at detector locations using advanced inductive signatures and matches vehicle pairs at detector locations. Several vehicle feature selection and weighting methods including Self Organizing Map and K-means clustering were applied to better identify individual vehicles from signature data. It was shown that the proposed extensive feature processing enhanced vehicle identification performance even among vehicle pools sharing similar physical configurations. An extended form of the Bayesian classification model was applied to estimate matching and non-matching probabilities of the vehicle pairs. The developed model was tested along an approximately 5.5-mile freeway segment on I-5 and CA-78 in San Diego, California where only 67 percent of the total trucks were observed at both up- and down-stream detector sites. Results showed balanced performances in exactness and completeness of matching with 91 percent of correct outcomes for multi-unit trucks. As a case study, the tracking algorithm was implemented over a larger network system with six detector locations and showed that the algorithm was able to capture the route choice patterns of different truck types. These results indicate that the network-wide tracking model can facilitate further understanding of spatial and temporal truck flow patterns, which are of importance in freight transportation analyses and planning.

Introduction

According to a national freight strategic plan (NFSP) (1), the U.S. population is expected to increase 321 million in 2015 to 389 million by 2045, with economic growth doubling in size. Consequently, freight movement is expected to increase approximately 42 percent by 2040, which is equivalent to an average of 1.3 percent increase per year. Among various modes in freight transportation, trucks show the largest expected increase in flows by 2040 since they handle the most ton-miles in the US. Specifically, 70 percent and 64 percent of the tonnage and value, respectively, of goods are shipped by trucks (1). Increasing freight demand will consequently yield substantial impacts on road networks. The NFSP (1) reported that assuming no capacity changes, truck and passenger vehicle traffic will increase recurring peak-period congestion by 34 percent in 2040, compared to 10 percent in 2011. Hence, freight transportation agencies need effective policies and regulations to successfully operate existing highway systems and reduce negative impacts from trucks such as greenhouse gas emissions, pavement maintenance, and noise. Along this vein, the most recent surface transportation authorization act, the Moving Ahead for Progress in the 21st Century Act (MAP-21) (2), was enacted in 2012 with an establishment of the national freight policy (NFP) (1) that aims to maintain competitiveness and efficiency of the freight transportation system. The main objectives of these programs include: i) reducing congestion and improving performance of the freight transportation, ii) identifying and monitoring major freight corridors to facilitate freight flows, and iii) developing data collection and analytic tools for freight modeling to allow public and private sectors to better assist in the decision-making process.

A number of researchers have pointed out that more efficient data collection methods and planning strategies are needed for more accurate truck movement estimations (3, 4). One of the unique travel patterns of truck traffic compared to general traffic is that the truck flow significantly varies by location and time. Typically, passenger cars and local service trucks show heavy traffic during the morning and afternoon peak hours in weekdays. On the other hand, through (i.e., passing) trucks show constant traffic throughout during 24 hours, seven days a week (1, 3, 5). Moreover, truck travel patterns closely link to service industry and commodity types that trucks are associated with (6, 7). For example, truck volumes observed in urban areas are likely constant for weekdays and weekends as they mainly perform localized service. However, truck volumes near port areas are much higher during weekdays than weekends because ports often do not operate on weekends. Therefore, when estimating and collecting truck flows, it would be important to consider their travel behaviors because volume or travel pattern of sample truck populations may not represent the total population, which would eventually result in inadequate restrictions of freight corridors and facilities, or displacements of freight activities (1).

However, due to a lack of data availability, there have been difficulties in understanding truck travel patterns. The main sources of truck data are either truck surveys conducted for limited durations in particular seasons or truck counts at permanent detection facilities such as automated count stations and weigh-in-motion (WIM) sites, or temporary installed active sensors. Specifically, surveys such as the national vehicle inventory and use survey (VIUS) (8), or State level intercept surveys, provide detailed information on truck type, Origin-Destination (OD), weight and VMT throughout the US or by State. However, the surveys typically obtain partial data from sampled populations as representative pools. Therefore, if survey sampling is biased or inaccurate responses are collected, survey outcomes may not represent the total population, as assumed. In addition, surveys tend to be implemented for a limited period of time, therefore the outcomes may not reflect seasonality in truck flows. To complement such survey approaches, passive sensor technologies such as WIM or inductive loop detectors (ILDs) have been widely used. Since such detection systems already typically installed along truck corridors in the U.S., and are capable of providing temporally continuous data, the full measures of truck volumes can be obtained by this technology. However, since the detection systems only provide point observations such as volume and occupancy, additional modeling efforts are required to obtain path flows or travel time estimations. Recently, active sensors technologies such as Automatic Vehicle Identification (AVI), Global Positioning System (GPS), and Bluetooth have been used to provide vehicle flow data. Since these sensors capture vehicles at multiple locations and match them with their unique IDs, vehicle flow and travel time can be obtained directly from the sensors. However, this data collection method requires sensor installments in vehicles and the network to collect vehicle ID information. Besides these additional installation efforts, data collection may cause temporal and spatial sampling biasness because only sampled populations registered as probe vehicles provide the information.

Hence, this study developed a methodology to accurately estimate network-wide truck flows by leveraging existing point detection infrastructure. Recently published work by the authors (6) proposed a framework for tracking truck flows at a link level utilizing WIM systems. Defined as vehicle tracking or re-identification, vehicles detected at passive detectors are matched between detection locations using data features from the detection systems. Compared to the current data sources that provide truck flow data, the tracking approach is advantageous for measuring spatial and temporal variations in truck flows because the total volumes of trucks are temporally continuously tracked and monitored along the truck corridors. This is an improvement over the prior link-based tracking approach, which had several limitations in capturing dynamic truck activity in a complex road network. Since the tracking was focused on matching vehicles between sparsely located WIM stations at the corridor level, actual implementation of the tracking would not be effective at a smaller scale such as capturing complex travel behavior of trucks within a city. In addition, since distance between adjacent WIM sites vary significantly, tracking accuracies may be compromised along corridors where the distance between WIM sites is large (9).

Considering the long travel distance of many freight trucks, tracking should be accomplished over a complex network even across different freeways. Hence, this study developed a network-wide tracking approach solely based on ILD systems, which is the most prevalent detection system in the US. The dense and widespread installation of these sensors ensure that spatial and temporal variations of truck activity can be recorded by detectors and used for the network-wide tracking framework. Therefore, this study expects to facilitate understanding of spatial and temporal truck flow pattern over a large network and provide valuable insights for policy makers and planning agencies to identify primary truck routes and to estimate path flows of trucks along different truck corridors. As an illustration, an application of truck route choice behavior by truck type is presented in this report.

Literature Review

General Vehicle Tracking Studies

The concept of vehicle re-identification or tracking was first introduced in late 1980s with primary focus on general traffic. As ILDs have been the most prevalent detection systems in the US, waveform signatures from the ILD have been commonly utilized in anonymous vehicle tracking studies. Bohnke and Pfannerstil (10) first introduced the use of inductive waveforms to re-identify vehicle sequences. Kuhne (11) followed by developing a freeway vehicle re-identification technique using dynamic traffic flow models. Sun et al. (12) proposed a multi-objective optimization approach to formulate vehicle re-identification problem using inductive loop signatures on a 1.2-mile freeway section. Oh and Ritchie (13) developed an anonymous vehicle tracking algorithm focusing on passenger vehicles at signalized intersections using probabilistic pattern recognizer. Tawfik et al. (14) adapted the lexicographic methods developed by Sun et al. (12) with a heuristic decision tree algorithm and showed 89 percent accuracy. Jeng et al., (15) developed a real-time vehicle re-identification algorithm with different loops configurations and showed the matching accuracies ranging from 50.7 to 54.2 percent by detector configurations. Abdulhai and Tabib (16) identified a new distance measure to improve accuracy of re-identification algorithm including conventional statistical measures (i.e., Euclidean, Correlation, Lebesgue, First derivative, MSE, and FFT) and neural network (i.e., back propagation neural network, time delay neural network and probabilistic neural network). The result selected the neural network as the best distance measure with 56% accuracy. Coifman and Krishnamurthy (17) used individual occupancy measures as vehicle features. Based on loop occupancy, long vehicles were identified as distinct vehicles to track arrival sequence, which showed about 40% of long vehicles re-identified.

As all the passenger vehicles have two axles with similar metallic compositions, the previous vehicle tracking studies could only be performed along short distance (i.e., approximately 1 mile) focusing on performance measure such as a travel time estimation, rather total flow estimation. Travel time estimation requires simpler vehicle matching algorithm than total flow estimation because only a portion of vehicles needs to be successfully tracked. In other words, only a portion of vehicles, which are distinct in its feature thus have higher chance to be correctly matched, could be used for the tracking.

Truck Tracking Studies

For truck tracking, Cetin et al. (9, 18) and Jeng and Chu (19) developed a long-distance truck tracking algorithm. Cetin et al. (18) used WIM data to track long distance truck traffic over 100 miles in Oregon and Cetin et al. (9) extended the original work to multiple locations with different corridors. These studies applied a Bayesian approach using two-step algorithms where the first step finds matched vehicle pairs while the second step screens out false matches using the posterior probability from the Bayesian model. Jeng and Chu (19) utilized inductive loop signatures and WIM for truck tracking. Specifically, the inductive loop signatures were the main source to match vehicles and the WIM data were subsequently used to filter out mismatching vehicles. Vehicles were matched based on proximity measures such that a vehicle pair with the minimum distance between signatures obtained at up- and down-stream stations was selected as the matching pair. Hyun et al. (6) recently applied a Bayesian approach with extensive feature selection and weighting techniques using WIM and loop signatures. Results showed that tracking was successful when features from WIM and signatures are simultaneously utilized for tracking. Overall, over 80 percent of tracking accuracy was shown on 26-mile freeway corridor in Southern California.

Background of Inductive Loop Detector Systems

As the predominant detection system in the US, ILDs are one of the most common data sources in various applications in transportation analysis such as traffic operation and monitoring, planning, safety, and environmental studies. In California, ILD measures such as vehicle volume, occupancy, and estimated instantaneous speed are publicly open through the Performance Measurement System (PeMS) (20) for over 25,000 ILDs on highway mainlines, ramps, and local arterials.

A conventional presence loop detector detects a presence of vehicle in a bivalent mode with zero or one pulses. However, an advanced ILD – which simply replaces the conventional ILD card in the traffic cabinet – has the capability to generate inductive waveforms signatures for individual vehicles. The advanced ILDs records disturbances to the inductance field caused by the metallic composition of a vehicle as a waveform signature at up to 1200 samples per second. With high sampling rates, detailed vehicle types and body configurations can be identified from inductive signatures using advanced models while the conventional binary output can only measure vehicle presence (21). Since advanced ILD systems do not require modifications to existing in-pavement sensors, they are a sustainable and cost-effective method for collecting and distinguishing individual truck observations.

Methodology

The tracking process comprises seven main steps as shown in Figure 1. When a vehicle (i.e., target) is detected at a downstream location, an algorithm aims to match the vehicle at a corresponding upstream location or to declare that the vehicle did not pass the upstream location. As the upstream candidate sets examined for every target vehicle can get impractically large, the first and second steps of the algorithm focus on effectively narrowing down potential candidates prior to the matching process. Then, the signature features are processed to enable waveform data to be better utilized to distinguish vehicles even those in similar signature patterns. The final vehicle tracking processes adopted the two-steps Bayesian algorithms from the previous study of corridor level tracking (6). The algorithms perform two consecutive steps of best match selection and filtering where the best match selection finds the best candidate for each target by comparing matching probabilities to the all candidates. The filtering determines whether the best candidate is indeed a true match or the if vehicle did not pass the upstream location. The latter indicates that the vehicle entered the corridor midway through another ingress, and dd not arrive through the upstream location.

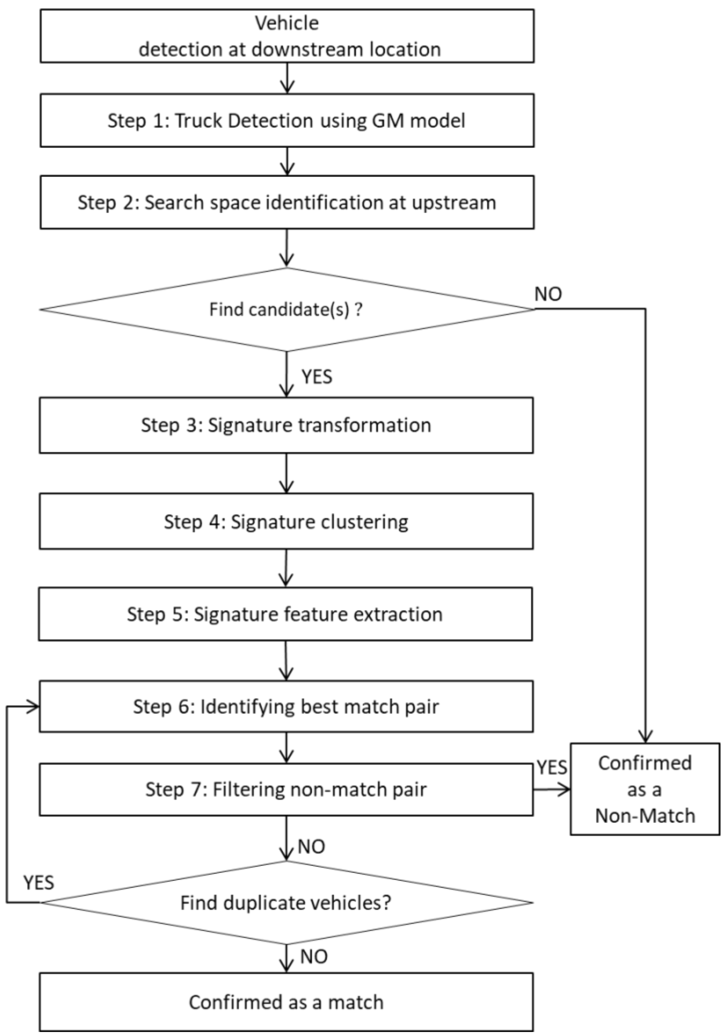


Figure 1 Flowchart of network-wide tracking model.

Truck Detection Algorithm

Since the ILD has no capability of classifying vehicles, a truck detection algorithm was applied to categorize trucks into two groups (i.e., Single-units and Multi-units) and to exclude passenger vehicles from the tracking process. The truck detection algorithm is based on Gaussian Mixture (GM) model that linearly composes multiple (m) Gaussian distributions, $N(\mu_m, \Sigma_m)$ with a mixing proportion of p_m (22). Hyun et al., (23) presented a binary GM model for distinguishing trucks from non-trucks using ILD data. Specifically, the binary model focuses on differences in duration over the loop by vehicle type. This study applied three duration distributions for passenger vehicles, Single-unit trucks and Multi-unit trucks using tri-modal GM distributions. Simply, the algorithm captures the differences in length of stay over a loop, which depends on length of vehicle. Details on truck detection algorithm refer to Hyun et al. (23).

$f(x) = \sum_{m=1}^{M} \quad p_m \cdot N(x; \mu_m, \Sigma_m)$ (eq. 1)

where m is number of mixture components, $N(\mu_m, \Sigma_m)$ is a Gaussian distribution with mean μ and covariance matrix Σ, and p_m is the mixing proportion.

The GM model was tested with the tracking population including passenger vehicles and showed 99 percent, 75 percent, and 95 percent of Correct Classification Rate (CCR) for passenger vehicle, single-unit truck, and multi-unit truck, respectively. While passenger vehicles and multi-unit trucks showed high classification rates, approximately 19 percent of single-unit trucks were classified as passenger vehicles. This is because trucks with short body length such as utility and service trucks show similar duration of longer passenger vehicles like pick-up trucks.

Search Space Identification

Once a target is identified at a downstream, the tracking algorithm searches potential candidates at upstream locations using minimum and maximum travel time. The minimum travel time is estimated based on the speed limit. In order to consider truck travel behavior of slow speed with multiple stops for rest break, the search space should not be constrained from narrow temporal window. In this study, maximum travel time is defined as a relaxed threshold, for example three times of the minimum travel time. However, this relaxed search space may significantly increase the candidate vehicles especially on a heavy traffic corridor, and consequently burden to the tracking process to identify the exact match. On the other hand, the increased number of vehicles more likely contains a matching vehicle in the candidate pool, therefore the tracking performance would be eventually higher if proper matching and filtering processes are followed.

$\{minimum\ travel\ time\ (tt_{ff}) \leq Temporal\ window < maximum\ travel\ time\ (e.g., \alpha_{ts} * tt_{ff}\ )\}$ (eq. 2) where α_{ts} can be any value greater than 1 and this study uses three

Signature Transformation

A signature waveform is a representation of a vehicle metallic shape. Ideally, signatures from the same vehicle would be identical regardless of the detected locations. However, multiple factors such as vehicle speed and vehicle's lateral position on the loop, loop geometry, and sensor calibration and sensitivity may warp the signatures and produce different shapes by location even from the same vehicle. Therefore, this study applied a signature transformation step to minimize such effects prior to extracting the signature features for tracking. The transformation step initiates with a normalization step that standardizes signatures with the same scale from zero to one for time and magnitude axis. This normalization step allows the signature features to be extracted with the same scale and to remove particular impacts caused from different sensor calibration by detector locations.

Although the signature normalization step is effective way to reduce different magnitude scales from the loop geometry and sensor calibration, there are still a number of factors that could affect signature shapes, for instance, sensor sensitivity, vehicle's lateral position, and vehicle acceleration or deceleration on the loop. However, it is assumed that the signatures from the same vehicle would have similar waveform patterns as well as distinct signature structures such as peak location and relative magnitudes of the peaks although not every part of the signature could be exactly identical. In this regard, the previous study (6) developed a heuristic algorithm that transforms signatures to minimize distances between signatures from the same vehicle. The algorithm transforms the candidate signature to fit the target signature along a horizontal (i.e., time) axis by linearly shifting and stretching the signature.

However, since the approach only focuses on the horizontal transformation, distortions in magnitude, which are caused by loop sensitivity and vehicle speed, were not successfully corrected. Since the distances between signatures are used as vehicle features, large magnitude differences even from few points could significantly decrease the matching probability. Therefore, in this paper, vertical transformation is introduced along with the horizontal fitting as follows.

Horizontal transformation (shift and stretch)

$$S_i^{shift}(t) = S_i^{shift}(t) + \beta^{shift} \qquad [e.g., -0.20 \le \beta^{shift} \le 0.20] \text{ (eq. 2)}$$

$$S_i^{stretch}(t) = S_{shift}(\beta^{stretch} \cdot t) \qquad [e.g., 0.8 \le \beta^{stretch} \le 1.2] \text{ (eq. 3)}$$

Vertical transformation

$$S_i^{vertical}(t) = S_i^{stretch}(t) + \beta^{shift} \qquad [e.g., -0.20 \le \beta^{shift} \le 0.20] \text{ (eq. 4)}$$

where t represents time, β^{shift} and $\beta^{stretch}$ represent the shifting and stretching coefficient, respectively.

These horizontal and vertical transformation steps are iteratively performed until the minimum difference between the signature pairs is obtained or the iteration is reached the maximum threshold. Figure 2 illustrates candidate and target signatures from the same vehicle with horizontal and vertical transformation steps. Although the overall shapes of original signatures are similar, the magnitude distances between candidate and target are huge especially the time between 0.5 to 1 second; however, the transformation steps closely fitted the candidate signature to the target and successfully reduced distance between two signatures.

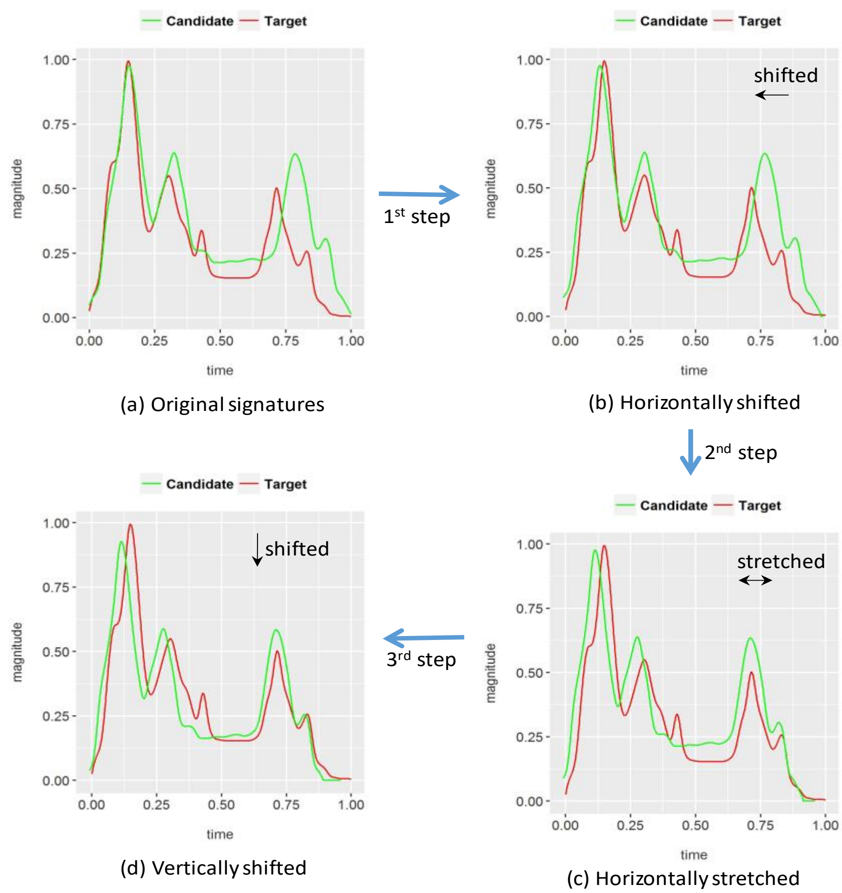
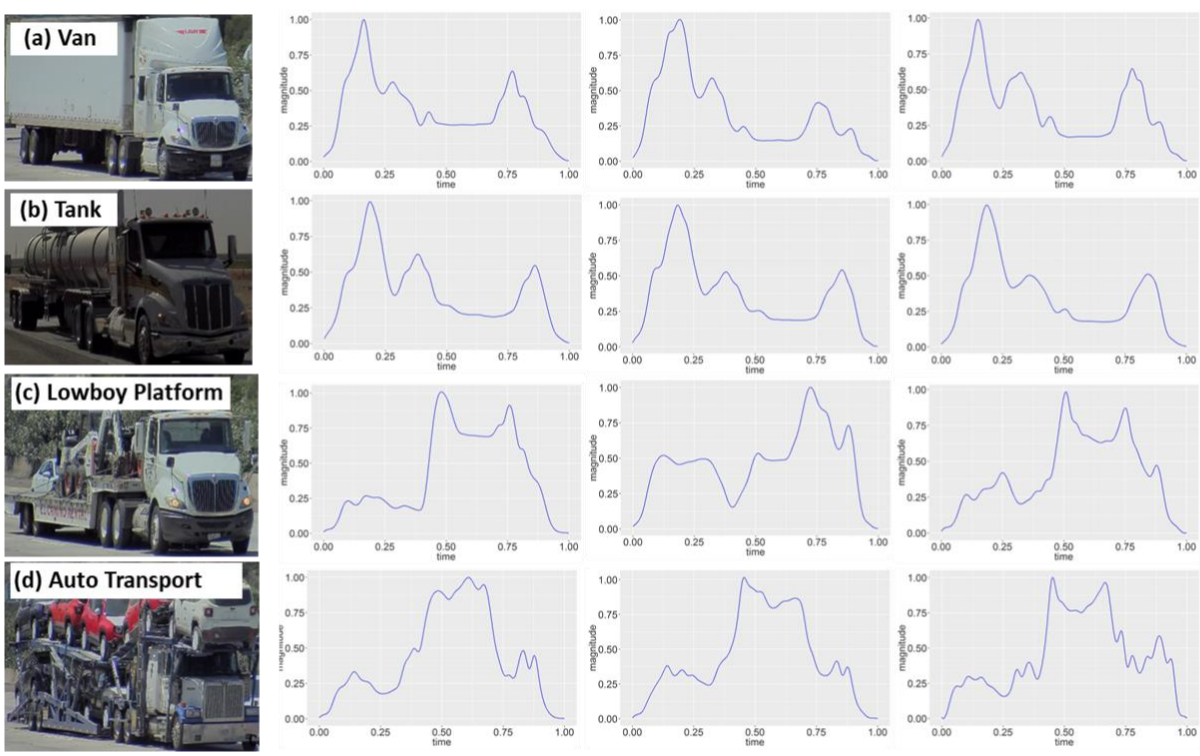


Figure 2 Horizontal and vertical signature transformation.

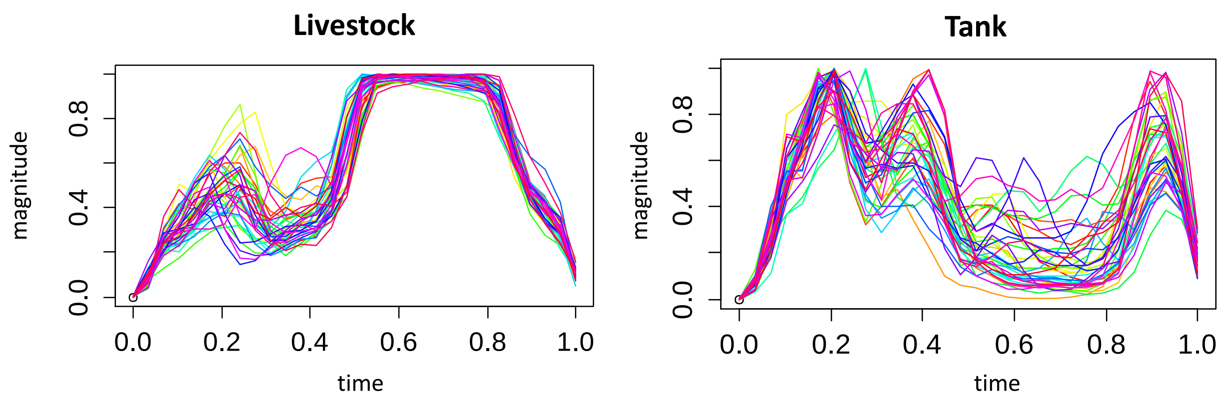
Signature Clustering Approach

One challenge in vehicle tracking is to select the exact match from several candidates that have the same vehicle type due to their similar signature waveforms. Figure 3(a) compares signatures of four truck types in Multi-Unit FHWA class 9 (3S2) generated from three different trucks. It is found that while the signatures between trailer types are distinguishable, the signatures within the same body configuration are much less distinct to each other. Notably, three signatures from the tank show almost identical waveforms because the trailers have the most similar configuration and physical shape. In addition, the signatures from the dropped deck types, lowboy platform and automobile transport trucks, have similar pattern of high magnitude in a trailer part. Since the distance between dropped deck and loop sensor is close, the signatures show high magnitudes in those parts.

An effort was extended to capture even small distinction in individual signature and found that most trucks tend to have unique devices or accessories such as a tire chain, a tool box or a particular metal frame. This study focused on capturing those small distinctions in a signature to distinguish individual vehicles. Moreover, it was also empirically learned that different truck types have the particular objects at designated areas such as under the driver unit or rear part of trailer due to their physical configuration and loading/unloading characteristics. Figure 3(b) depicts the randomly chosen fifty signatures of livestock and tank trailers. Although the overall signature patterns from the same trailer types are very similar, there is a particular section that has more variations in individual signatures. For example, all the signatures of the livestock trailers have high magnitudes from 0.5 to 1 second; however, the front part (up to 0.4 second) of the signatures showed larger variations in magnitude by trucks. Similarly, the middle parts of the tank (between 0.3 to 0.8 second) highly vary by vehicle. These parts are expected to have greater capability of identifying differences among vehicles and better distinguish matching vehicles based on their unique magnitude patterns.



(a) Individual signatures

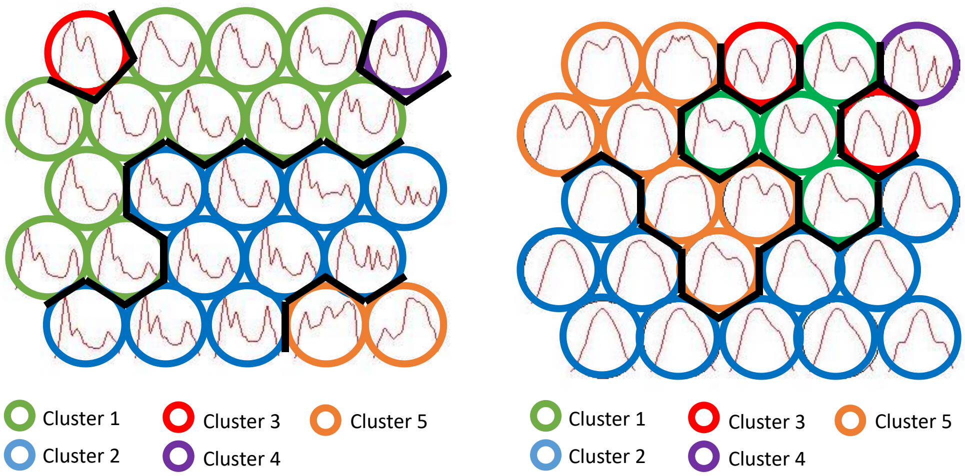


(b) Fifty randomly chosen signatures

Figure 3 Comparison of (a) individual and (b) fifty signatures in the same body configuration.

However, it should be noted that these features would be hardly noticeable if signatures from different type of trailers are compared altogether. Therefore, this study developed a signature clustering model to group trucks with similar signature patterns using Self Organizing Map (SOM) algorithm, which is an unsupervised clustering method based on a neural network (24). Since this method is unsupervised approach where data are not labeled for modelling, signatures can be clustered by their overall patterns, not classified by any type of labels. In this vein, the algorithm categorizes signature waveforms solely depending on their inherent shapes. Specifically, the SOM is consisted with two components of training and mapping. The training step builds a map using given input data set and the mapping procedure clusters a new input vector based on the map. The map contains several nodes that are connected to the input, not to each other. All the nodes have topological positions and change their positions along with the neighbors' positions. In other words, the nodes adapt themselves to the input and increase differences to the other vectors (i.e., unselected inputs) to form clusters.

This study trained a SOM with a large dataset that contain varieties of signature patterns. The data set includes 28,328 single and multi-unit trucks collected at four ILD sites in California (21). A total of 25 nodes are initially used for the map and K-means clustering (25) was used to find the optimal number of clusters. Five clusters are selected as optimal as shown in Figure 4 for (a) Multi- and (b) Single-unit. Each color in the figure represents a cluster and each node shows a representative signature pattern. For example, a total of 11 green nodes and one red node in Figure 4(a) represent a cluster 1 and cluster 3, respectively. It is also visually confirmed that the nodes with the same color show similar signature shapes, as two nodes in cluster 5 of Multi-unit show high magnitudes on the trailer part.



(a) Multi-unit unit Clusters (b) Single-unit Clusters

Figure 4 Clusters and signature patterns in (a) Multi-Unit and (b) Single Unit.

Feature Selection

Feature Extraction

Signature feature vectors represent differences between the signature attributes of a target and a candidate vehicle, where the attributes include 50 normalized magnitude measurements obtained at evenly distributed points along the temporal axis of the inductive signature.

$V_{ij}^{l} = V_{i}^{l} - V_{j}^{l}$ (eq. 5)

where i represents candidate, j represent target, and l represent signature magnitude ($l \in \{1, \dots 50\}$)

Feature Distribution

To apply Bayesian inference to the vehicle features, probability distributions of match and non-match cases should be initially estimated for each feature. This study applied the distributions from the previous corridor level tracking model (6) because the estimated feature distributions should be spatially and temporally transferable to the new dataset. The initial model determined the match and non-match based on the sum of Euclidean distance of fifty features and a visual validation from groundtruth process confirmed the matching outcome. The model found that the features were fitted to Gaussian distributions centered at zero for both of match and non-match cases (see Figure 5). This is because a feature represents the distance of vehicle attribute (i.e., signature magnitude) and the distance should be ideally zero for matching cases. Even non-match cases are centered at zero but show larger variance in distributions because incorrectly matched pairs were still initially matched due to their smallest total distance.

The Chapter 4.4 confirmed that each signature cluster has different waveform patterns. This implies that the clusters may have different important significant parts that have more abilities to differentiate vehicles. Therefore, the parametric density functions were estimated by cluster and the signature features that have more power to distinguish vehicles are selected and weighted by cluster. Figure 5 shows examples of feature distributions for three clusters. The first feature (i.e., feature #13) shows that the cluster 1 has larger variances for both match and non-match while the cluster 2 shows smaller variances for match. These differences ensure the assumption that the features have varying ability in distinguishing match and non-match by clusters. In other words, the feature #13 has the most influence for the second cluster in distinguishing match and non-match because differences in probabilities of match and non-match are the largest in this cluster.

The feature weights were estimated and categorized into four labels – critical, signature, insignificant, and inverse – based on their importance in distinguishing match and non-match vehicle pairs in descending orders (6). If match and non-match distributions are not statistically different, and the variance of match distribution is smaller than that of non-match, the feature is categorized as critical since it plays critical role in distinguishing two distributions. On the other hand, a feature's match and non-match distributions are statistically the same, the feature is labeled as an insignificant.

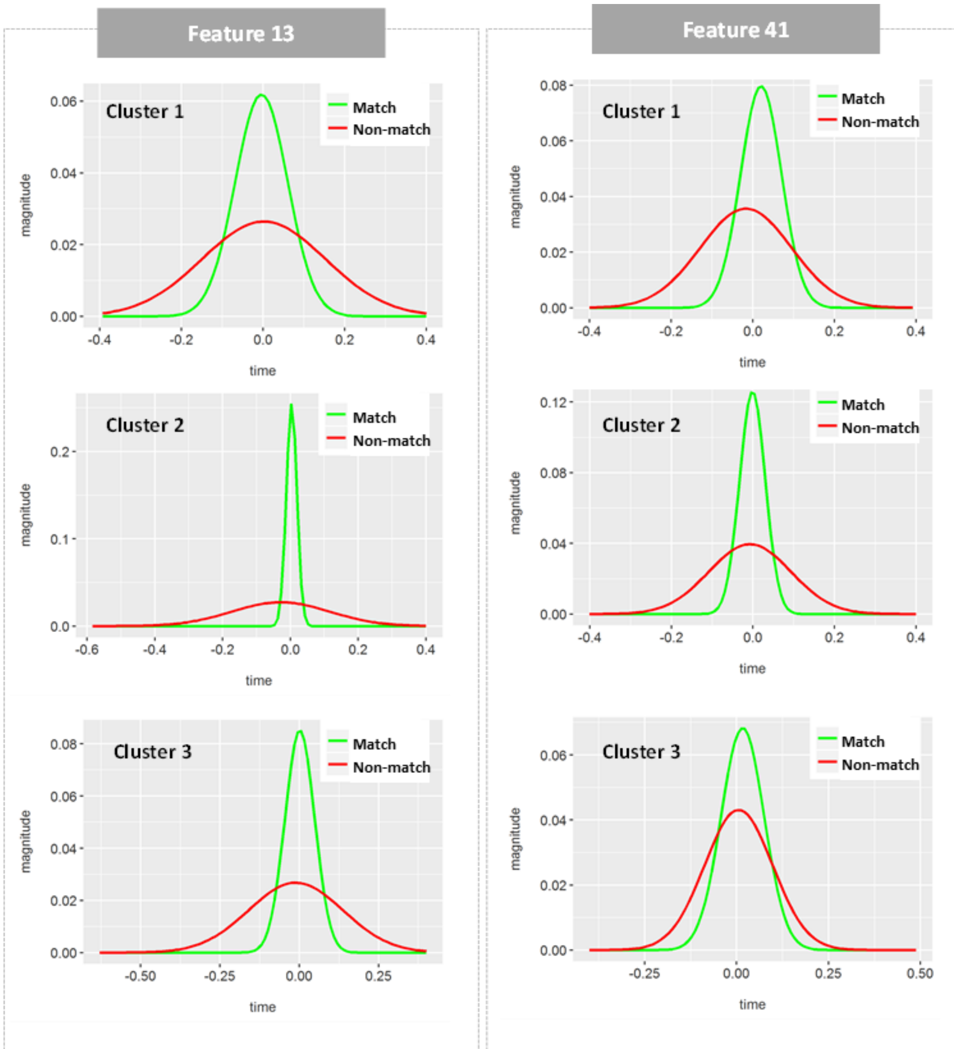


Figure 5 Comparisons of parametric density functions by cluster.

Bayesian Framework for Vehicle Matching and Filtering

This study applied an extended form of a naïve Bayes classifier. As a family of probabilistic classifier in machine learning techniques, a naïve Bayes classifier is a supervise learning that requires labeled training data. In other words, a pair of example should consist of input attributes and output value. In this study, vehicle pairs (i.e., examples) of match and non-match (i.e., outputs) can be distinguished by vehicle attributes (i.e., feature vectors) by a mapping function of Bayesian classifier.

The goal of Bayesian framework is to identify a matching vehicle at upstream location for every target vehicle detected at downstream location. The first model, defined as a vehicle matching model, determines the best candidate among multiple candidates for all target trucks. Only matching pairs' distributions are used in this process as the candidate with the highest matching probability ($p(\mu_{ij})$) is chosen as the best candidate. However, to consider nonpassing trucks, there is a need for the second model, filtering model, to filter out vehicles that did not pass both of upstream and downstream detector locations. The filtering model therefore compares the matching ($p(\theta_{ij} = 1)$) and non-matching ($p(\theta_{ij} = 0)$) probabilities for every potential matching pairs obtained from the first model. Simply, the matching probability is higher than 0.5 (or non-matching probability), the matching outcome is declared as a match. The final step examines duplicates in matching pairs, and if the duplicate candidate is identified, the target re-searches its best matching vehicle (refer back to Figure 1).

Vehicle matching model

$p(\mu_{ij} \mid V_{ij}^1, V_{ij}^2 \dots V_{ij}^L) \propto \frac{\alpha_l \cdot \Pi_{l=1}^{L} p(\mu_{ij})}{\alpha_l \cdot \Pi_{l=1}^{L} p(\mu_{ij})}$, where α_l represents a feature label (eq. 6)

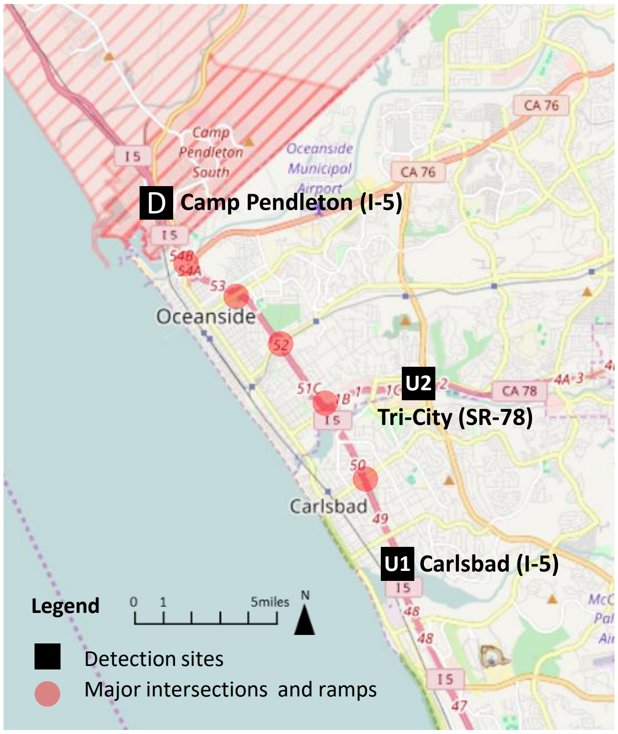
Vehicle filtering model

$$p(\theta_{ij} = 1 \mid V_{ij}^1, V_{ij}^2 \dots V_{ij}^L) = \frac{\Pi_{l=1}^{L} \alpha_l \cdot p(\theta_{ij}=1)}{p(\theta_{ij}=1) \cdot \Pi_{l=1}^{L} \alpha_l \cdot p(\theta_{ij}=1) + p(\theta_{ij}=0) \cdot \Pi_{l=1}^{L} \alpha_l \cdot p(\theta_{ij}=0)} \text{ (eq. 7)}$$

$$p(\theta_{ij} = 0 \mid V_{ij}^1, V_{ij}^2 \dots V_{ij}^L) = 1 - p(\theta_{ij} = 1 \mid V_{ij}^1, V_{ij}^2 \dots V_{ij}^L) \quad \text{(eq. 8)}$$

Data

The tracking is implemented on a major truck network containing two upstream locations and one downstream location as shown in Figure 6. The distances from each upstream to downstream ILDs are 5.2 and 5.5 miles, spanning five major freeway intersections and ramps. Signature data from the northbound truck lanes were collected from the upstream #1 (U1) and the downstream sites (D), and from the westbound truck lanes at the upstream #2 site (U2) on July 7th, 2016. A total of 424 vehicles were collected at the downstream locations where 58 percent of trucks are multi-unit trucks (Table 1). There were 284 trucks observed at both up- and down-stream location in this network, referred to as common vehicles, which is 67 percent of the total vehicle captured at downstream location. Inductive signature data and side-fire images for trucks were collected at these sites for visually validating the model, which were manually linked through a groundtruth data processing.



Site Location	Distance	Site Description	Collection Dates
Camp Pendleton (I5) to Carlsbad (I5)	5.2 miles	D-U1	July 7th 2016
Camp Pendleton (I5) to Tri-City (SR 78)	5.5 miles	D-U2	11:20AM – 12:40PM

Figure 6 Data collection site.

Table 1 Data Collected for Network-wide Tracking

Dataset	# of trucks collected at downstream	% of Multi-unit at downstream	# of trucks collected at upstream	% of Multi-unit at upstream	# common vehicles	% of common vehicle from the total vehicle detected at downstream
D-U1			421	54%	222	
D-U2	424	58%	118	49%	62	67%
Total			539	53%	284	

Results

Tracking outcomes are results of a binary classification problem whether the declared pair is a match or a non-match. To evaluate classification problems, a confusion matrix is generally used to show a number of actual and predicted outcomes in correct and incorrect consequences. In this study, actual and predicted match and non-match vehicles are presented in the table for Multi- and Single-unit, respectively (see Table 2). Performance measures are also introduced to evaluate the classification problem. First, an accuracy considers both correct match and non-match pairs across the total pairs. However, the accuracy is known to show biased results when a problem has a large imbalance in two outcomes because the problem tends to send more outcomes to majority case to achieve a high performance. For example, if majority vehicles in a tracking network are not common vehicles, significantly more non-matches would be expected than matches. If the algorithm successfully classifies non-match, high accuracy is obtained regardless of the match results, although more attention should be given to the match results. To overcome this problem and show trade-offs of the actual and predicted correct outcomes, especially for a desired outcome (i.e., match), additional measures, recall and precision, are considered. The precision is defined as the number of positive (i.e., matching) prediction divided by the total number of positive class predicted while the recall is the number of positive predictions divided by the number of positive class actual values. In other words, precision indicates a measure of exactness, showing how many actual matches are predicted. The recall is a measure of completeness because it shows how many of predicted matches are indeed correct. Since the precision and recall contains different information in match cases, a performance measure that includes a balanced outcome between the precision and recall is introduced as a F1 score.

$$F1\ score = 2 * \frac{recall * precision}{recall + precision} \quad \text{(eq. 9)}$$

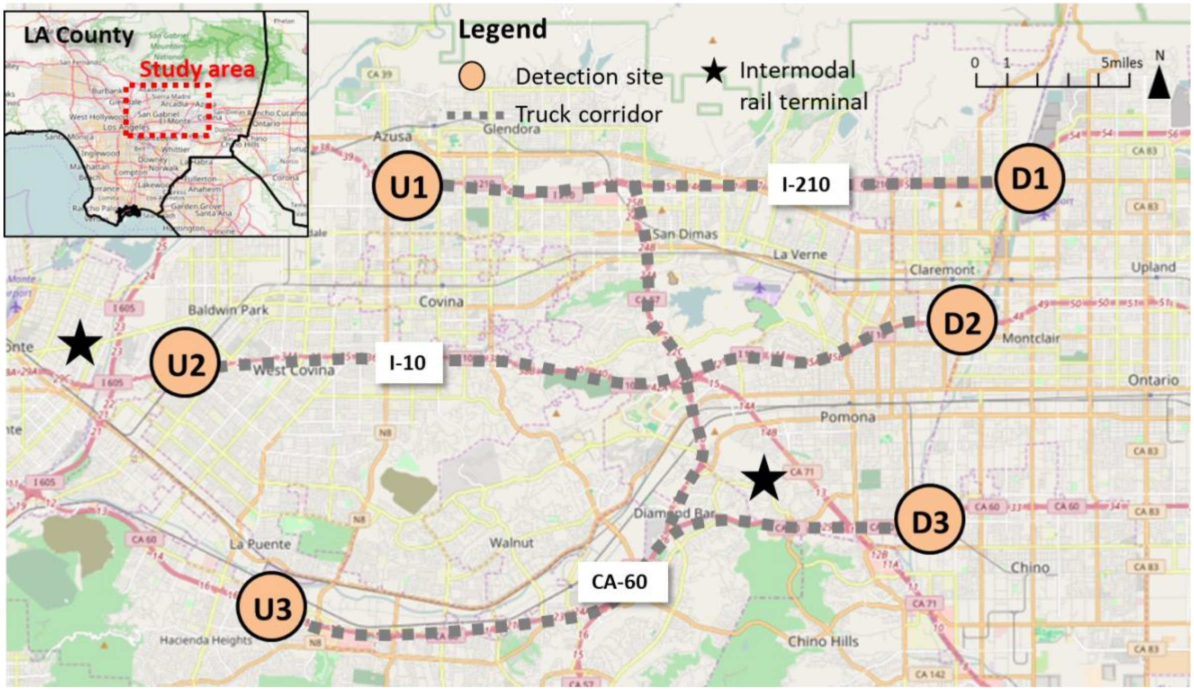
Multi-unit showed 91 percent of F1 score while Single-unit showed 60 percent. In Multi-unit, most of vehicles were common vehicles that were observed both up- and down-stream locations and the matching and filtering algorithms successfully declared them as match. Compared to the accuracy, recall and precision show better performance in Multi-unit, which indicate the algorithm performs better in identifying match pairs than filtering non-common vehicles. However, the filtering algorithm tends to keep non-common Single-unit trucks as match. Although 54 percent of Single-unit vehicles were not common vehicles, only half of them were filtered as non-matches. This is because the Single-unit has less variety in its truck configuration, which makes it challenging distinguishing vehicles in both matching and filtering processes.

Table 2 Confusion Matrix and Performance Measure

<table>
<tr><th rowspan="2" colspan="2">Multi-unit</th><th colspan="3">Predicted</th><th rowspan="2" colspan="2">Single-unit</th><th colspan="3">Predicted</th></tr>
<tr><th>Match</th><th>Non-match</th><th>Total</th><th>Match</th><th>Non-match</th><th>Total</th></tr>
<tr><td rowspan="3">Actual</td><td>Match</td><td>182</td><td>19</td><td>201</td><td rowspan="3">Actual</td><td>Match</td><td>56</td><td>27</td><td>83</td></tr>
<tr><td>Non-match</td><td>19</td><td>25</td><td>44</td><td>Non-match</td><td>48</td><td>48</td><td>96</td></tr>
<tr><td>Total</td><td>201</td><td>44</td><td>245</td><td>Total</td><td>104</td><td>75</td><td>179</td></tr>
<tr><td colspan="2">Performance Measures</td><td colspan="3">Accuracy= (182+25)/245 = 84%
Recall = 91%
Precision = 91%
F1 score = 91%</td><td colspan="2">Performance Measures</td><td colspan="3">Accuracy= 56%
Recall = 67%
Precision = 54%
F1 score = 60%</td></tr>
</table>

Application to Truck Monitoring with Route Choice Behavior by Truck Types

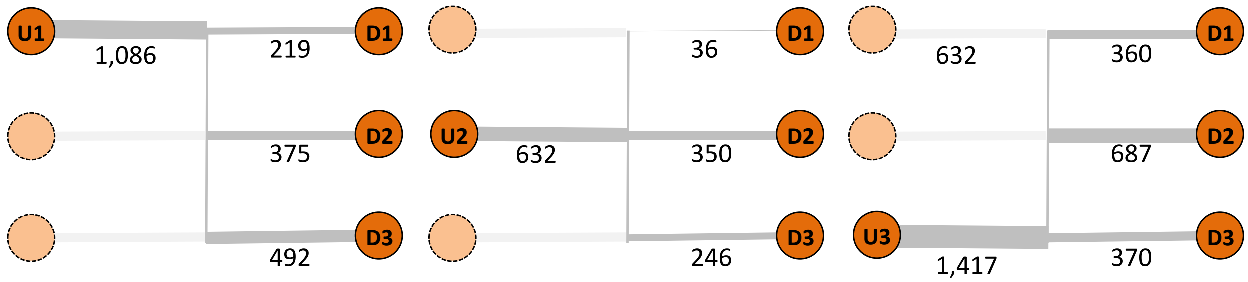
Since the Bayesian model tracks vehicles among multiple detection locations, trucks can be monitored with their route choice over a large network. As a use-case, truck monitoring is implemented in a larger network with an integration of the recently developed truck body classification model (21 Hernandez et al., 2016). This case study chose six ILD locations located on I-210, I-10 and SR-60 in Southern California. These corridors are the major routes connecting San Bernardino county and Los Angeles county as shown in Figure 7. These three highways run parallel and serve as alternative routes for each other. Signature data and body configuration model estimates were collected on August 3rd, 2016 from UCI-TAMS (26). This application only considers Multi-unit trucks in the tracking process. It should be noted that the truck monitoring case study shows an application of network-wide tracking and the tracking results presented in this section were the estimates from the tracking model and not manually confirmed with visual groundtruth process. Although the truck tracking results are not validated with visual confirmation of their license plates, this study is expected to show valuable insights in truck route choice and travel patterns in a large and complex network system.



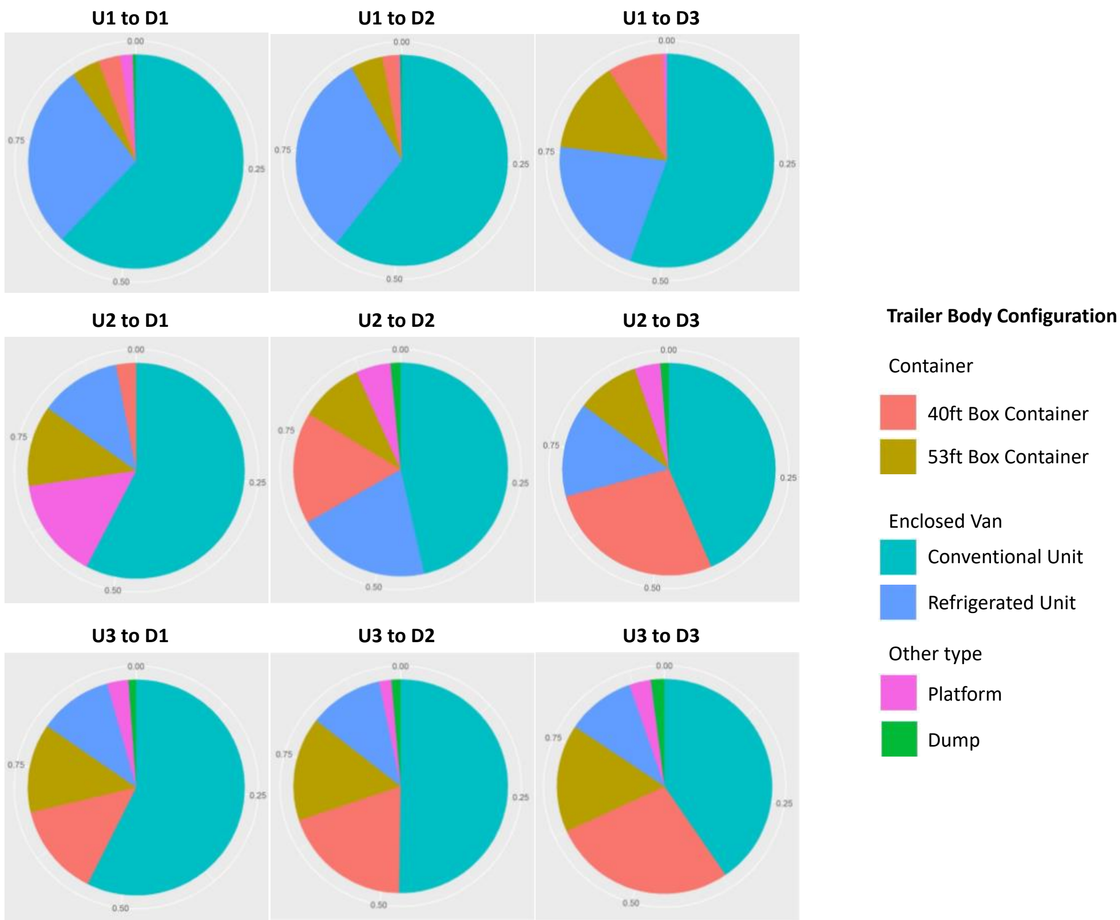
Site id	Site Description	Site Location (highway)	Vehicle Observed	Multi-unit Estimated
U1	Upstream Site #1	Azusa (I-210)	46,747	8,657
U2	Upstream Site #2	West Covina (I-10)	33,086	3,886
U3	Upstream Site #3	La Puente (SR-60)	48,714	11,183
D1	Downstream Site #1	Claremont (I-210)	29,696	4,097
D2	Downstream Site #2	Montclair (I-10)	54,342	9,394
D3	Downstream Site #3	Chino (SR-60)	53,970	8,416

Figure 7 Site map for network-wide truck monitoring.

Figure 8(a) presents route choice patterns by upstream location. The truck corridors were simplified and the thickness of the line indicates the relative flows between sites in this figure. For example, trucks observed at U1 passed to the downstream locations in descending order of D3, D2, and D1. Although U1 and D1 are located on the same highway, the least proportion of trucks passed through the both locations. Overall, the routes mostly used are from I-210 (U1) to CA-60 (D3), and from CA-60 (U3) to I-10 (D2). Figure 8(b) compares the truck body configuration by routes by integrating the tracking outcomes to the truck body classifications (16). A total of 9 combinations of upstream-downstream sites are presented with six body configurations. The most common body type is enclosed van followed by either 40ft port container or refrigerated unit of enclosed van. Specifically, a large amount of port container trucks was travelled to D3 (SR-60) from all the upstream locations, which could be because D3 is closely located to two transcontinental rail terminals.



(a) Route choice patterns in total trucks



(b) Route choice patterns by truck body configurations

Figure 8 Route choice patterns (a) in total trucks and (b) by truck body configurations.

Conclusion

This study developed a network-wide tracking model using ILD waveform signatures of trucks based on a Bayesian framework. As the tracking algorithm solely utilizes signature features for vehicle matching, the tracking could be implemented for temporally continuous tracking along any major truck network. However, as ILD signature profiles highly depend on the ferrous shape of vehicles, trucks with similar body configurations inevitably provide similar signature shapes. Therefore, this study applied extensive signature processing steps, including signature transformation and clustering approaches, to better identify vehicles and to increases tracking accuracy. Specifically, SOM was used to group signatures by their overall shapes, which allowed the signature features to become more salient and distinguishable by individual vehicles.

Tracking results showed 91 and 67 percent correct matching rates for multi-unit and single-unit trucks. Specifically, multi-units showed high performance in both exactness and completeness in finding matching pairs. The tracking model implemented a truck monitoring application over a complex network with six detector locations at port adjacent cities in Southern California. Although six sites were fairly closely co-located, distinct travel patterns were monitored by truck type, which showed the ability of the tracking model to analyze temporal and seasonal variations of truck activities by affiliated industry.

References

1. USDOT, “National Freight Strategic Plan.” Department of Transportation, US DOT, www.transportation.gov/freight/nfsp

2. Federal Highway Administration Office of Policy and Governmental Affairs, “A Summary of Highway Provisions.” Moving Ahead for Progress in the 21st Century Act (MAP-21), 17 July 2012, www.fhwa.dot.gov/map21/summaryinfo.cfm.

3. Ogden, K. W. "Truck movement and access in urban areas." Journal of Transportation Engineering 117.1 (1991): 71-90.

4. Tok, A., Hyun, K., Hernandez, S., Jeong, K., Sun, Y. Rindt, C., and Ritchie, S.G., (2017), Truck activity monitoring system (TAMS) for freight transportation analysis, Journal of the Transportation Research Board (In Press)

5. Hallenbeck, M. E., and Kim., G. Final Technical Report for Task A: Truck Loads and Flows. No. WA-RD 320.3. 1993.

6. Hyun, K., Tok, A., and Ritchie, S.G., (2017), Long Distance Truck Tracking from Advanced Point Detectors using a Selective Weighted Bayesian Model, Journal of Transportation Research Part C, Vol. 82, pp 24-42

7. Hyun, K., and Ritchie, S.G., (2017), A sensor location decision model for truck flow measurement, Journal of the Transportation Research Board, (2644), pp.1-10

8. U.S. Census Bureau, “Vehicle Inventory and Use Survey.” 2004, www.census.gov/svsd/www/vius/products.html.

9. Cetin, M., Monsere, C., Nichols, A. and Ustun, I., 2011b. Key factors affecting the accuracy of reidentification of trucks over long distances based on axle measurement data. Transportation Research Record: Journal of the Transportation Research Board, (2243), pp.1-8.

10. Böhnke, P., and Pfannerstill, E., "System for the automatic surveillance of traffic situations." ITE journal 56.1 (1986): 41-45.

11. Kuhne, R. D. Freeway control using a dynamic traffic flow model and vehicle reidentification techniques. No. 1320. 1991.