Final Technical Project Report

Automated Plate Recognition and Truck Trip Tracking

Project #: *IG1752329*

RES#2016-32

UT#R01-1313-556

Submitted to

TDOT Research Office in Long Range Planning

by

Lee D. Han, Zhihua Zhang, Stephanie Hargrove, and Mark Burton

of

Center for Transportation Research
The University of Tennessee

Submitted: August 2019
Revised: May 2020
2nd Revision: January 2021

Technical Report Documentation Page

1. Report No. RES2016-32	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle *Automated Plate Recognition and Truck Trip Tracking*		5. Report Date June 2019
		6. Performing Organization Code
7. Author(s) Han, L., Zhang, Z., Hargrove, S., Burton, M.		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Tennessee Knoxville 525 John Tickle Engineering Building 815 Neyland Dr Knoxville, TN 37996		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 37B355
12. Sponsoring Agency Name and Address Tennessee Department of Transportation Research Office 505 Deaderick Street, Suite 900 Nashville, TN 37243		13. Type of Report and Period Covered Transportation Research August 2016 to June 2019
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study sought to apply automated license plate recognition (ALPR) technology to track trucks trips. ALPR does not work perfectly in the U.S. because of the thousands of different designs, colors, shapes, fonts, etc. of license plate from different states. To overcome this, a class of machine learning algorithms were developed to help track trucks by matching license plates read, correctly or incorrectly, by ALPR devices. While these unsupervised machine learning algorithms worked great for short distance (<10 miles) scenarios, they have never been tested for long distance scenarios, which was the main challenge of this study.

Three sets of field studies were conducted at strategically selected Interstate sites in Tennessee using mobile ALPR stations. The first study tracked trucks on I-75 from Georgia to Kentucky and to Virginia via I-81. The second study tracked trucks from Georgia and Alabama to Kentucky via I-24 and I-65. The third study tracked westward trucks through Nashville via I-40 and around Nashville via I-840. The tracking distance was between 50 and 250 miles. In general, the total matching percentage ranged from 14% to 48%. This is common and largely due to spatial and temporal leakages between stations far apart.

17. Key Words	18. Distribution Statement
ALPR, LICENSE PLATE TRACKING, FREIGHT TRACKING, AUTOMATED LICENSE PLATE RECOGNITION, TRUCK TRIPS	No restrictions. This document is available through the National Technical Information Service, Springfield, VA 22161; www.ntis.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	71	$83,024.16

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

DISCLAIMER

This research was funded through the State Planning and Research (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under **RES #: 2016-32, Research Project Title: Automated Plate Recognition and Truck Trip Tracking**

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

CONTENTS

List of Figures

List of Tables

List of Acronyms

AL	Alabama
ALPR	Automated License Plate Recognition
AM	Ante Meridiem
ATRI	American Transportation Research Institute
BTS	Bureau of Transportation Statistics
Co	County
DMS	Dynamic Message Sign
EB	Eastbound
FAF	Freight Analysis Framework
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
Fri	Friday
GA	Georgia
GPS	Global Positioning System
ID	Identification
KY	Kentucky
LPR	License Plate Recognition
MM	Mile Marker
Mon	Monday
mph	Miles Per Hour
NB	Northbound
OCR	Optical Character Recognition
OD	Origin-Destination
ORNL	Oak Ridge National Laboratory
PrePass	A weigh station bypass service for truckers and fleets
PM	Post Meridiem
RFID	Radio Frequency Identification
Sat	Saturday
SB	Southbound
Sun	Sunday
TDOT	Tennessee Department of Transportation
TDSHS	Tennessee Department of Safety and Homeland Security
THP	Tennessee Highway Patrol
Thu	Thursday
TMC	Traffic Management Center
TN	Tennessee
TOD	Time of Day
Tue	Tuesday
UT	University of Tennessee
UTK	University of Tennessee at Knoxville
VA	Virginia

VMS	Variable Message Sign
WAZE	A cellphone-based navigation app
WB	Westbound
Wed	Wednesday

EXECUTIVE SUMMARY

This study sought to apply automated license plate recognition (ALPR) technology to track trucks trips. ALPR does not work perfectly in the U.S. because of the thousands of different designs, colors, shapes, fonts, etc. of license plate from different states. To overcome this, a class of machine learning algorithms were developed to help track trucks by matching license plates read, correctly or incorrectly, by ALPR devices. While these unsupervised machine learning algorithms worked great for short distance (<10 miles) scenarios, they have never been tested for long distance scenarios, which was the main challenge of this study.

Three sets of field studies were conducted at strategically selected Interstate sites in Tennessee using mobile ALPR stations. The first study tracked trucks on I-75 from Georgia to Kentucky and to Virginia via I-81. The second study tracked trucks from Georgia and Alabama to Kentucky via I-24 and I-65. The third study tracked westward trucks through Nashville via I-40 and around Nashville via I-840. The tracking distance was between 50 and 250 miles. In general, the total matching percentage ranged from 14% to 48%. This is common and largely due to spatial and temporal leakages between stations far apart.

The key findings of the study suggest:

- License plate data collected at weigh stations by Tennessee Department of Safety and Homeland Security (TDSHS) are free to Tennessee Department of Transportation (TDOT) and can be quite useful without additional infrastructural investments. They could be used to establish travel patterns, trip frequency, travel time, stops, etc. between weigh stations.
- Permanent installation of additional ALPR stations at strategic locations, such as state border crossing points, could be quite useful for tracking trucks.
- The mobile ALPR units for research purposes in this study were limited in hours of operation and couldn't track many of the trucks that travelled, at least partially, outside of the study periods. The restrictive hours of operation, from sunrise to sunset, were implemented in the interest of student researchers' safety and security in remote field locations near high-speed roadways. A network of permanently installed ALPR stations could have helped with tracking a higher percentage of trucks.

CHAPTER 1 INTRODUCTION

The objective of this study is to investigate freight mobility patterns at the state/metropolitan level in Tennessee, which could shed some light on the challenging effort towards calibrating the freight routing algorithm in Federal Highway Administration's Freight Analysis Framework (FAF) at the national level. To that end, the Automated License Plate Recognition (ALPR) technology and the high accuracy plate-matching algorithms developed at the University of Tennessee, were used in this project.

Freight mobility and commodity flow information is of increasing importance to transportation planning and management agencies. Only a small number of studies have investigated freight mobility patterns due to the lack of observed data on truck route choices, which are difficult to obtain via traditional travel surveys. Transportation planners often have to make assumptions on truck route choice behaviors that are hard to verify in the context of freight movement. Alternative means towards studying truck route choices, therefore, point to GPS tracking and license plate tracking technologies. Of these, the GPS-based truck tracking method requires the instrumentation of individual trucks for active tracking. It can be quite involved, intrusive, and expensive with limited representation of the truck population. The license plate tracking requires only passive observation of existing identification means on every truck and is thus chosen for this study.

To gain a better understanding of freight mobility patterns in the transportation network, including origin-destination (O-D) freight flows, and truck travel routes, two license plates datasets were used in this project: mobile ALPR data collected in situ by the study team and Tennessee Department of Safety and Homeland Security (TDSHS) weigh stations that routinely collect ALPR data. Originally, the truck GPS data from the American Transportation Research Institute (ATRI) was requested, which would have provided the GPS location and time information of trucks through the State of Tennessee. This plan did not work out for reasons beyond the purview of this study. The weigh station license plate data were subsequently requested from the Tennessee Highway Patrol (THP), which is under TDSHS. The field study license plate data were collected through three separate field visits with ALPR equipment at strategic locations along the primary trucking corridors in Tennessee.

This study aims to have a better understanding of the freight mobility pattern on the Interstate highways of Tennessee, primarily through tracking truck license plates. As presented in this report, the license plate data collected in this study did shed some light on truck route choice and freight mobility patterns. When a wide area of deployment of the ALPR technology is realized and a large amount of weigh station data are obtained, various aspects related to freight movement, such as air quality, safety, and fuel efficiency, could be better assessed.

The remainder of this report is organized as follows. Chapter 2 presents a synopsis resultant from the literature review effort on trucking tracking technology. Chapter 3 presents the field studies and empirical results. In a similar fashion, the weigh station ALPR data and the analysis results are presented in Chapter 4. Lastly, Chapter 5 closes the report with conclusions and recommendations.

CHAPTER 2 TRUCK TRACKING TECHNOLOGY REVIEW

Trucks are just vehicles with much larger dimensions and different operational characteristics than average passenger cars. To track trucks, or any vehicles for that matter, one could monitor the trajectory of the target vehicle if a signal were actively emitted from the vehicle with uniquely identifiable ID and location. A less intrusive approach is to observe the target vehicle using identifiable means, such as a license plate, along the path of travel. A general comparison of these approaches is presented in this chapter.

2.1 ALPR-based Technology

Automatic License Plate Recognition, or ALPR, technology was first deployed in 1979 when the British Police Scientific Development Branch prototyped a system for trial deployment in Wokingham, UK. The ALPR technology can read the vehicle license plate with various algorithms, like optical character recognition (OCR) algorithm. LPR Technology primarily consists of six algorithms:

1. Plate locating.
2. Plate orientation and sizing.
3. Normalization.
4. Character segmentation.
5. Optical character recognition (OCR), and
6. Syntactical/geometrical analysis.

During the past decade, ALPR technology saw wider real-time deployment as computer, communication and video technologies matured. The ALPR technology has been predominantly used in combating auto theft and law enforcement, but it also holds great potential in investigating freight mobility patterns since ALPR can collect and store the information of trucks, such as time, date, location and license plate information.

Figure 2-1 Several portable ALPR devices in deployment in Tennessee

ALPR technology has three kinds of formats: fixed systems that are installed at roadside or strategic locations, mobility systems that are installed on police cars or other vehicles, and portable systems that can be deployed at selected locations. The portable LPR units have the capability to be deployed quickly for short-term study and special event data collection purposes. The downside of such systems include 1) the potential lack of reliable power source in the field, 2) the safety of the operation personnel, 3) the security of the unit out in the field, 4) the network communication means for data transmission, 5) the need to recalibrate the unit for each use and installation, and 6) the extensive training needs for the operation personnel when large scale study requires multi-site and multi-unit deployment.

Vendors of various ALPR systems often make claims of an over-95% accuracy of their LPR device. While this could be true in the controlled lab environment on license plates from states the LPR units are pre-calibrated for, the actual field results can be much lower when hundreds of different plate designs are encountered by the LPR units. In the previous studies by Han et al, brand new uncalibrated units could have accuracy as low as 30%-60%. The manufacturers tend to attribute the lower than claimed accuracy to the following factors:

- Motion blur and camera vibration,
- Poor lighting and visibility,
- Bent, damaged, dirty, and modified plates,
- Lack of unified fonts and colors,
- Reflectivity of plates and paints, and
- Plate designs "unfriendly" to LPR.

License plate reading accuracy, however, is not the whole story on vehicle tracking. Even with less than desirable individual plate reading accuracy, an artificial intelligence based self-learning algorithm developed by Han et al. (Oliveira-Neto, Han, & Jeong 2013) could "learn" to match license plates read, even if incorrectly, at multiple locations. Basically, the tracking process consists of collecting vehicle license plate numbers and arrival times at various checkpoints, matching the license plates between consecutive checkpoints, and computing travel times from the difference in arrival times (Oliveira-Neto, Han, & Jeong, 2013; Turner, Eisele, Benz, & Holdener, 1998; Wagner & Fischer, 1974). Further details are presented in Appendix A. The algorithms have previously achieved high matching accuracies in the range of 97%, for distances of up to 10 miles and were used in this project.

Since all vehicles are legally required to have license plates installed and displayed, no additional onboard technology, new identification devices, or user consent are required. The ALPR approach here is a passive and non-intrusive system.

2.2 Non-LPR based Technology

Unlike the passive system, one could actively track vehicles/persons using GPS, cell triangulation, and Bluetooth if a cellular unit is on the vehicle. Even technology like RFID could be used to track vehicles. These, however, are intrusive, in that some sort of owner consent or device deployment is needed, and they do not cover the entire population of vehicles.

2.2.1 Dedicated Global Positioning System (GPS)

GPS technology is widely implemented to identify vehicle position information (latitude and longitude, time). The GPS device can locate the object at any time with reasonably high accuracy via satellite trilateration. GPS technology is mature and proven to be effective. While for mobile phones the accuracy of GPS is in the range of 5-8 meters. Dedicated GPS units can be accurate to 3 meters or less than 10ft. Although, there are several limitations associated with GPS technology:

- GPS could lose signals at places like dense forest, canyon walls, skyscrapers, and bridges,
- GPS could lose signals under heavy cloud, rain, and other inclement weather conditions,
- GPS may not work well in indoor and underground spaces, and
- GPS may have coverage gaps due to satellite maintenance, radio interference, and solar storms.

American Transportation Research Institute (ATRI) has been collecting the truck GPS data of key national corridors since 2002, and, according to ATRI, billions of truck GPS traces in North America are obtained annually. This data can provide insights into understanding freight activity and freight mobility for public agencies at both the federal and regional level. For this project, the truck GPS data from ATRI was requested, but not available.

2.2.2 Bluetooth

Bluetooth devices are commonly embedded in mobile phones, in-vehicle navigation systems, and increasingly in other wireless peripherals, such as headphones and smart watches, for the purpose of secure short-range data communication. The Bluetooth device can broadcast a unique hardware ID when enabled and can be detected and uniquely identified, hence tracked, by roadside monitoring sensors. Thus, Bluetooth data can be obtained via an in-vehicle Bluetooth device while maintaining user anonymity. The data can be used in tracking vehicles since it provides the location and time information of the vehicle.

A major problem with Bluetooth data is that the adequate sample size is not always guaranteed since Bluetooth devices need to be enabled to be detected. In addition, Bluetooth has a lower geolocation accuracy (~75 feet) in comparison with that of GPS (~10 feet). In a low-density roadway system, like those in rural areas, one might be able to perform some map-matching post-processing to fix minor geolocation errors. In the dense roadway network common in urban areas, this would be more challenging as a Bluetooth device (and the vehicle carrying it) could be geolocated at a nearby, but erroneous, location due to the lower accuracy of the technology and the complexity of the environment.

2.3 Other Data Sources

Several other data sources are also commonly implemented to track trucks, including mobile probe data, and toll tag data.

2.3.1 Mobile Probe Data

Mobile probe data is increasingly used in real-time traffic monitoring. Cell phones can be considered as probes to collect mobile probe data. The data contains information, such as vehicle location and time, which can be used to obtain vehicle speed and travel time.

The mobile probe data gains its popularity due to its cost-effectiveness, real-time capabilities, and relative accuracy. But some issues are associated with the mobile probe data, like

- Mobile probe data may not have a large enough sample size in sparsely populated areas, or in off-peak times,
- Mobile probe data, such as GPS, has inherent errors under inclement atmospheric or weather conditions, and
- Mobile probe data has privacy concerns.

It should be noted that cell phone GPS/location data are different from those from dedicated GPS units install on some of the truck fleets and on some of the American-made passenger vehicles specifically for vehicle tracking purposes. While dedicated GPS units are tracked at all times, typically with high temporal resolution, cell phone GPS tracking can only be accessed anonymously or through subscription/consent.

2.3.2 Toll tag data

Toll tags data have been considered as an important alternative data source to determine travel times along the roadway. The toll tag data is especially useful in an area with a large number of toll roads, like Florida. The toll tag data can be used to determining traffic speed, travel time and O-D matrices, through matching the toll tags over a known distance.

The toll tags data holds potential in tracking vehicles, but it has a primary issue, namely, the availability of accurately recognized toll tags. It can be partly attributed to the following reasons:

- The coverage of toll tag readers is limited
- Not all vehicles have toll tags
- Toll tags can be read duplicate times, or misread

Since tolls and, hence, toll tags are not common in Tennessee and surrounding states, the technology was not considered for this study.

2.4 Established Difficulty with Long-Distance License Plate Matching

License plate matching is easy if the license plate is read correctly. This, however, is not the case in the real-world, especially for the U.S., where thousands of different designs, colors, dimensions, fonts, and reflective materials are used in different states. When a license plate is misread at an ALPR location, even if only by one or two characters, it becomes very difficult to match and, hence, track the plate at another location. As mentioned in Section 2.1, Dr. Han's research group at the University of Tennessee published a series of primary literatures on how to overcome this problem with text-mining and machine-learning techniques (Oliviera-Neto, et al. 2012, 2013) Their success, primarily limited to short-distance (< 10 miles), demonstrated an algorithm successfully learned to match plates at two different locations several miles apart. The correct matching rate was reported in the range of over 97% with a false positive rate of less than 1%. While this is very promising, plate matching for longer distance is much harder.

A major challenge is in the way the algorithm learns. Unlike supervised learning where a person, a dog, or a computer is told by an outside agent, a teacher perhaps, whether a response was correct or otherwise, unsupervised learning does not have a teacher. The algorithm must use some basic rules to figure out for itself if a match is correct or not. This is the essence of Dr. Han's matching algorithm. By using some text-mining techniques and calculating the probably of matches, Dr. Han's algorithm deduces and continually updates the likelihood two plates are a match. This learning process requires many pairs of readings, correct or not, from the same plate at two different ALPR stations. It works well with short-distance tracking scenarios where each plate is read at both locations, which automatically creates many training samples. It becomes more difficult for longer distance scenarios where insufficient training samples may result, preventing the algorithm from learning efficiently.

Another issue with long distance matching is the simple fact that a significant number of trucks that happened to pass one station would not pass the other station hundreds of miles away because of the wide ranges of possible origins, destinations, and route choices. For two ALPR stations a short distance apart, say 5 miles, on a straight stretch of Interstate, it is highly likely a truck would be captured at both stations within a few minutes. But when that distance is extended to 50 miles with some stops and crossroads in between, the likelihood the same truck would be captured at both stations declines and requires a longer "time window" to allow for the variability in travel time. The bigger the time window, the more plates need to be compared to make sure the potential correct match is not omitted inadvertently. At 5 miles apart, the algorithm may only need a time window of a few minutes while at 50 miles apart, the algorithm will need a time window of hours. When that distance is extended to 250 miles with many Interstate junctions and even more potential stops in between, the likelihood a truck would pass though the two stations within a few hours drops down severely and the time window is measured in days.

CHAPTER 3   FIELD DATA COLLECTION AND ANALYSIS

3.1 ALPR-Based Field Data Collection

Three field studies were conducted in October 2016, April 2017, and May 2017 to collect truck travel data along selected portions of the State's Interstate highways. These activities are shown in Figure 3-1 and listed below.

- Field Study 1 – Northbound trucks in all lanes at TN/GA border on I-75 were tracked to TN/VA border on I-81 near Bristol and to TN/KY border on I-75. This was conducted on October 22, 2016.
- Field Study 2A – Northbound trucks in all lanes north of TN/GA border on I-24 were tracked to TN/KY border near Clarksville on I-24 and to TN/KY border on I-65. This was conducted on April 8, 2017.
- Field Study 2B – Northbound trucks in all lanes north of TN/AL border on I-65 were tracked to TN/KY border near Clarksville on I-24 and to TN/KY border on I-65. This was conducted on April 9, 2017.
- Field Study 3 – Westbound trucks on I-40 west of I-40/I-840 split on the west side of Nashville were tracked to the east side of Nashville on I-40 and I-840 right before the two roads merge again. This was conducted on May 8, 2017.

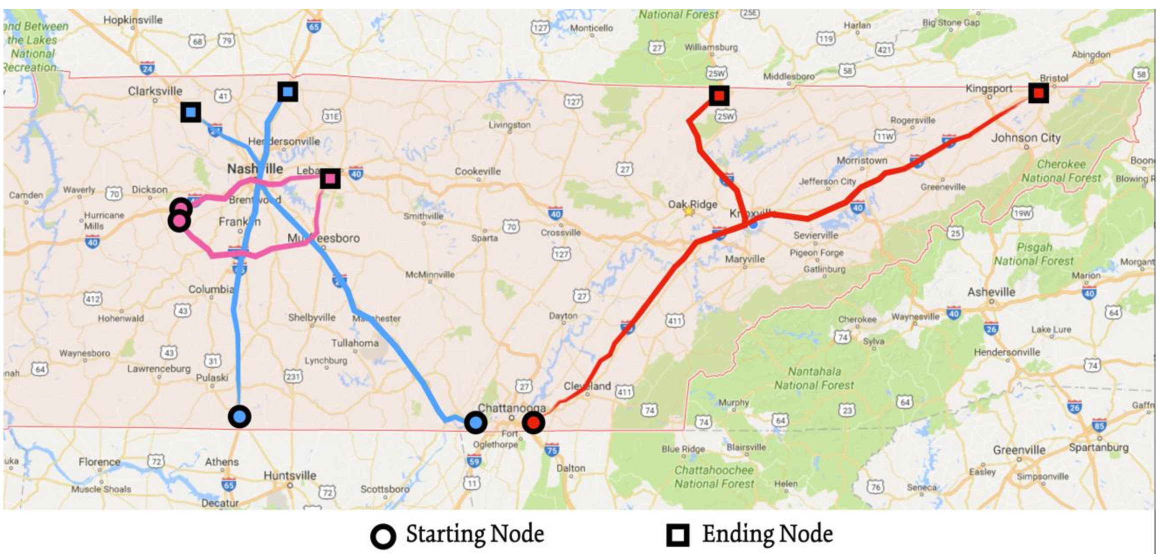


Figure 3-1 Map of locations of the three field studies

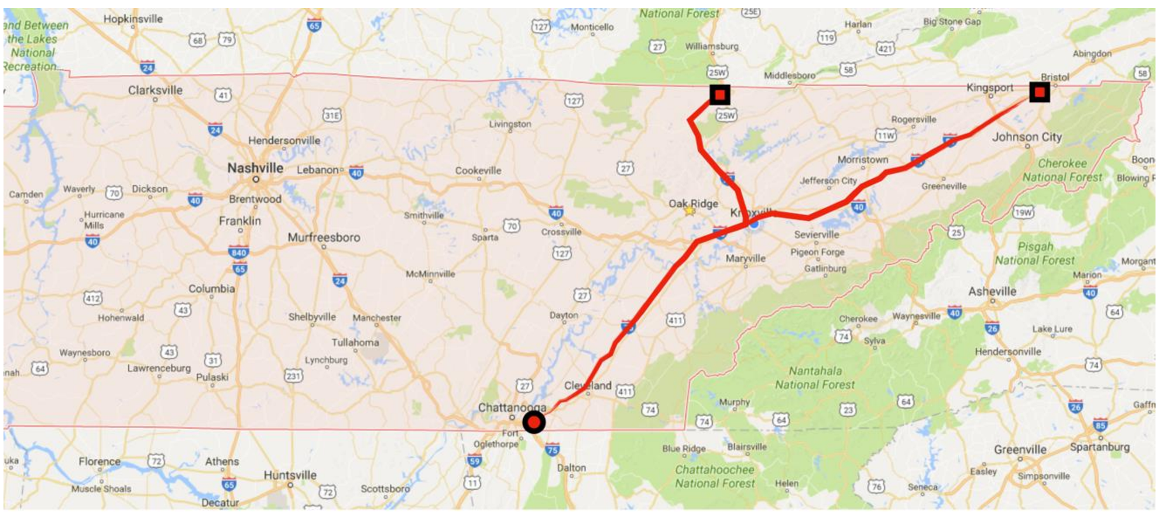
The remainder of this chapter will present the details of each of these studies and the analysis results of the data from these studies.

3.2 Case 1 – Northbound I-75 to I-75/I-81

3.2.1 Field Study Setup

The first field study was conducted on October 22, 2016, tracking the northbound trucks in all lanes from TN/GA border on I-75 to TN/VA border on I-81 near Bristol and to TN/KY border on I-75 (Figure 3-2). The distance from TN/GA border on I-75 to I-81 near Bristol is about 223 miles, and the distance from TN/GA border on I-75 to TN/KY border on I-75 is about 167 miles. These distance figures are important for the estimation of the "time window" when the tracked truck may arrive at the destination based on the average operational speed of the trucks.

The LPR devices were set up by a group of well-trained University of Tennessee - Knoxville (UTK) graduate and undergraduate students at strategic locations, shown in Figure 3-2. Three groups set up the LPR devices at the three locations from 5 AM to 7 PM.



Starting Node   Ending Node

Figure 3-2 The location map of the first field study

3.2.2 Field Data Assessment

Images of trucks at the above three locations were captured with ALPR devices, and then the license plates of trucks were extracted with the ALPR software. A total of 657; 1,662; and 1,026 license plates were collected at TN/GA border on I-75, at I-81 near Bristol, and at TN/KY border on I-75, respectively. The obvious lower number of 657 at the Chattanooga station near TN/GA border was due to an unfortunate incident where a student accidentally "fried" one of the ALPR units deployed at that locations together with the data after the field visit. The incident would not have occurred for permanently mounted ALPR units. Mobile ALPR units that are transported, assembled, set up and calibrated on site, are then disassembled, and transported back to our lab after each field study, which made them subject to many potential hazards in that process. They could have been dropped by accident, hit by lightning in rain, experienced a power outage in the field, or gotten damaged in transport, etc. The incident was due to reversing the polarity of a cable when trying to download the data. In a previous field study for FMCSA, we used permanently mounted ALPR units where all data were automatically transmitted via cellular network, which eliminated a lot of such hazards. But for research work with temporary setups, mishaps do occur. This incident further underscores the challenge for portable ALPR deployment for research/study purposes.

3.2.3 Plate Matching Results

After running the license plate matching algorithm, the results, shown in Table 3-1, suggest 10.4% of the trucks went through I-75 North and 3.7% went through I-81 East. This may indicate that for all the northbound trucks from I -75 South, about three-fourths stayed on I-75 while one-fourth switched to I-81. While there is certain "leakage" of trucks not tracked, perhaps due to the established difficulty of long-distance license plate matching, one might assume this I-75/I-81 truck split ratio holds true for freight mobility/pattern purposes. However, these matches only reflect same-day day-time trips. Early morning trips that passed the starting node before 5 AM or crossed the ending nodes after 7 PM would not be matched here.

Table 3-1 License plate matching results of the first field study

October 22nd	Truck Path	Distance (miles)	# of Matches	Starting Node	Ending Node	Matching Percentage
	I-75 → I-75	167	68	657	1,026	10.4%
	I-75 → I-81	223	24	657	1,662	3.7%

The low matching percentages of trucks in this study can be attributed to several reasons. First, the self-learning algorithm depends on many samples (a high number of trucks traversing both the starting and the ending ALPR stations) to learn efficiently. It worked well with short-distance tracking scenarios but did not have enough training samples under long distance tracking scenarios to be proficiently trained.

Second, the field studies were performed only from sunrise to sunset for most cases, in the interest of the safety/security of the researchers in the field after dark. With hundreds of miles to travel for the subject trucks, only a portion of them were able to complete the journey and be tracked at both the start and the end stations. Many of the trucks may have stopped along the way for a variety of reasons including, for example, refueling, resting, pick-up/drop-off, and so on, which can lead to much longer journey times. The longer the journey time, the less likely the trip could be captured within our field study time window. The simple fact that a significant number of trucks that happened to pass one station would not pass the other station hundreds of miles away because of the wide ranges of origins, destinations, and route choices.

Finally, for two ALPR stations a short distance apart, say 5 miles, on a straight stretch of Interstate, it is highly likely a truck would be captured at both stations within a few minutes. But when that distance is extended to 50 miles with some stops and crossroads in between, the likelihood the same truck would be captured at both stations declines quickly and requires a longer "time window" to allow for the variability in travel time. When that distance is extended to 250 miles with many Interstate junctions and even more potential stops in between, the likelihood a truck would pass though the two stations within a few hours drops down to single digits.

3.2.4 Data Analysis

Since the matched results are relatively small, the implications of these results may be limited. The mean travel time from I-75 South to I-75 North is 2 hours and 36 minutes (156 minutes), and the mean travel time from I-75 South to I-81 East is about 3 hours and 12 minutes (292 minutes). This leads to an average journey speeds of 64 mph for Georgia-Tennessee-Kentucky trucks and 46 mph for Georgia-Tennessee-Virginia trucks. It is reasonable to speculate that the Virginia-bound trucks likely took a break, perhaps in the Knoxville area, had some delay at a weigh station, or had to stop due to company policies or Federal Motor Carrier Safety Administration (FMCSA) hours of service guidelines (FHCSA 2015). In hindsight, another set of ALPR stations along I-40 west of the I-40/I-75 split would have been helpful to provide more insights. However, this was not possible at the time due to the fact the research team had only six ALPR units.

3.2.5 Findings

This first field study investigated the truck travel patterns from TN/GA border on I-75 to TN/VA border on I-81 near Bristol and to TN/KY border on I-75. The results showed that for all the northbound trucks from I -75 South, about three-fourths stayed on I-75 while one-fourth switched to I-81.

The first field study was conducted as an initial pilot to also work out the kinks in the data collection process. The number of matched vehicles were only 14.1% of the traffic, which was somewhat disappointing. Given that we did lose about a half of the data at the starting node due to the aforementioned manual error during the data downloading process, better matching results could have resulted. Due to the challenges of short field study window (from sunrise to sunset) and the lack of enough of matches, the heuristic learning algorithm was not implemented. All matches reported herein are, thus, with all characters correctly matched.

Compared to the aggregated statistics in Freight Analysis Framework (FAF) 2012 long-haul truck traffic on major US Interstate corridors (BTS 2017), the general volume of trucks for I-75 and I-81, at the destination nodes, appear to be reasonable. Since I-81 tends to have more trucks than I-75, to reconcile the 1-to-3 ratio of trucks going from Chattanooga to I-81 vs I-75, a significant amount of truck volume on I-81 must have come from I-40, at least for the day-time traffic.

An interesting research question did rise about the approximately 75-25 split (10.3% vs 3.1%) on I-75 vs I-81. The I-81 route is 33% longer than the I-71 route. If there are equal amounts of "leakage," or trucks exiting before the end node or not finishing within the same time window, per mile, the I-81 route could have lost an extra 33% of trucks going in that direction. One could take this into consideration and adjust the split accordingly. The validity of this consideration will have to wait for further studies.

3.3 Case 2 – Northbound I-65/I-24 to I-24/I-65

3.3.1 Field Study Setup

For the second field data collection effort, two separate configurations were used on April 8, 2017 and April 9, 2017. Case 2A was conducted on April 8, 2017 tracking northbound trucks in all northbound lanes from the TN/AL border on I-65 to TN/KY border near Clarksville on I-24 and to TN/KY border on I-65. On the next day, Case 2B was conducted tracking northbound trucks in all northbound lanes from the TN/GA border on I-24 to TN/KY border near Clarksville on I-24 and to TN/KY border on I-65 (Figure 3-3).

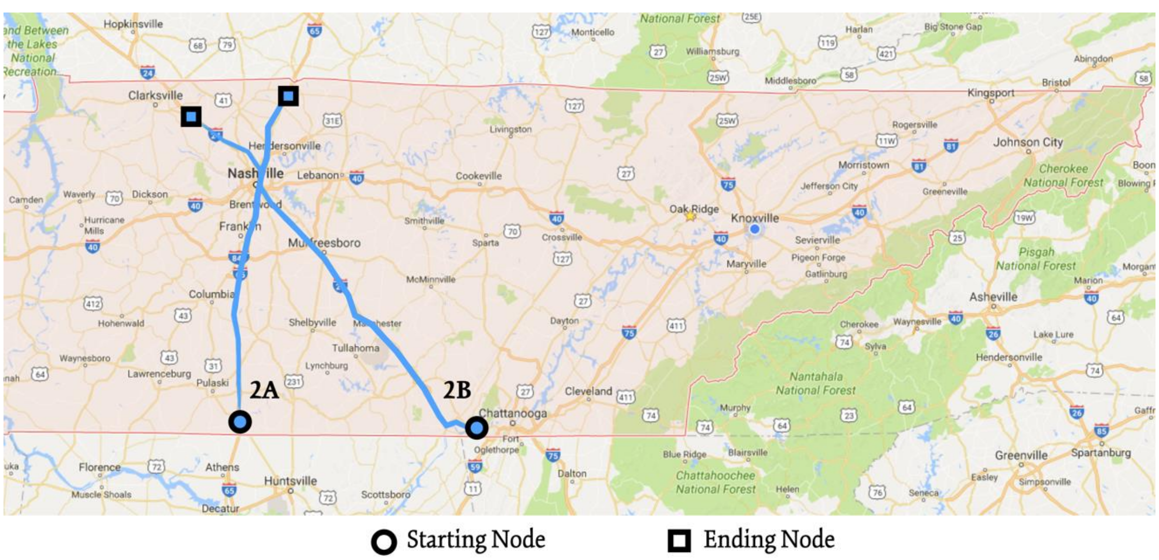


Figure 3-3 The location map of the second field study

For Case 2A, the distance from the TN/AL border on I-65 to TN/KY border near Clarksville on I-24 is about 85.7 miles, and the distance from the TN/AL border on I-65 to TN/KY border on I-65 is about 89.5 miles.

For Case 2B, the distance from the TN/GA border on I-24 to TN/KY border near Clarksville on I-24 is about 136 miles, and the distance from the TN/GA border on I-65 to TN/KY border on I-65 is about 141 miles.

The ALPR devices were deployed at three locations each day from 6 AM to 6 PM. The field study team was limited to three locations because of the number of ALPR units available.

3.3.2 Data Assessment

For Case 2A, the ALPR units collected 1,290, 1,588, and 1,214 license plates at the I-65 South, I-65 North, and I-24 West locations, respectively. Figure 3-4 shows the trucks captured by ALPR at these locations per time of day.

Similarly, for Case 2B, the ALPR units collected 1,293, 2,886, and 1,841 license plates at the I-24 East, I-65 North and I-24 West locations, respectively. Figure 3-5 shows the trucks captured by ALPR at these locations per time of day. Both sets of data showed similar patterns with more trucks during peak hours and slightly less during the middle of the day, as expected.

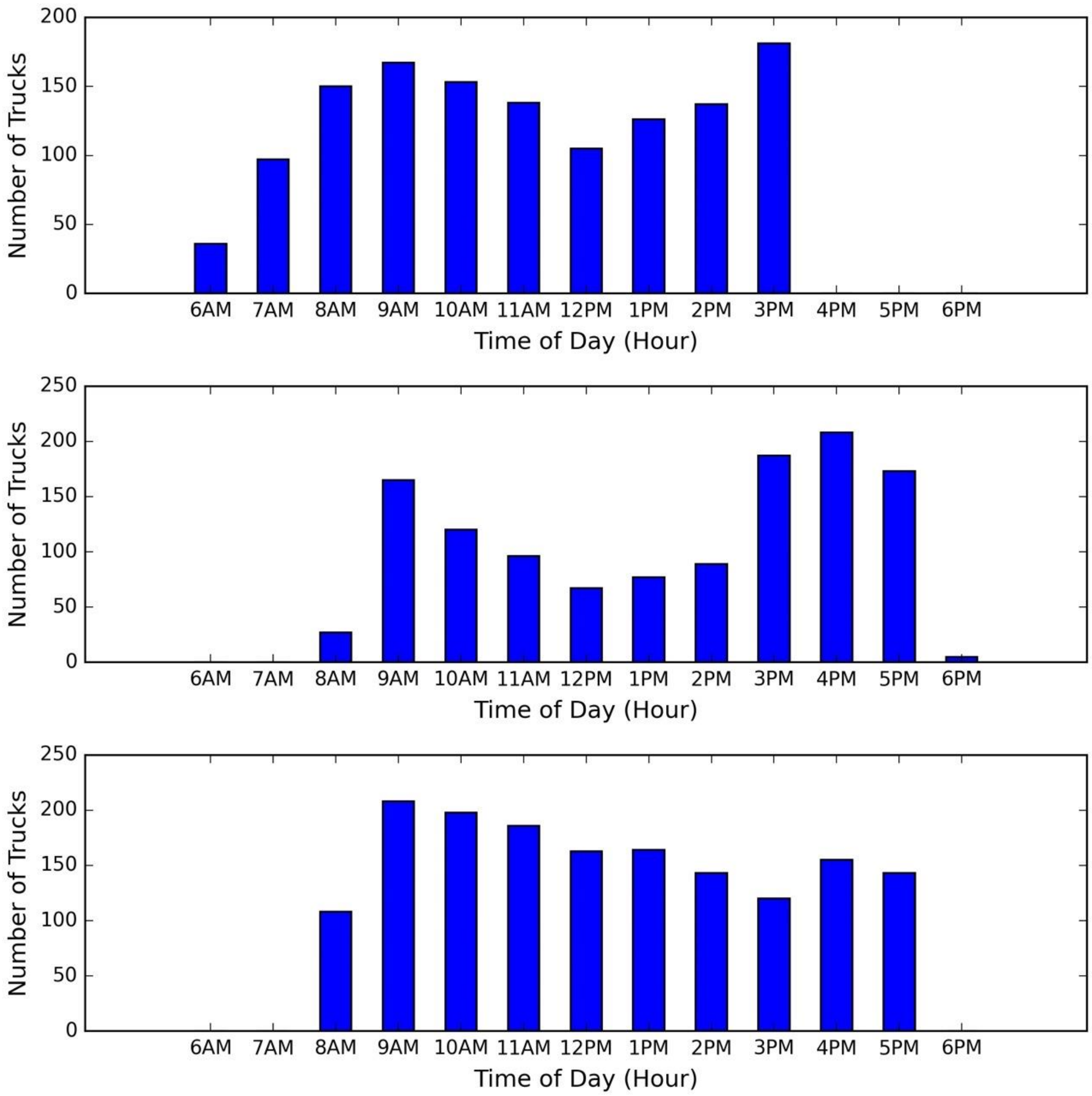


Figure 3-4 Trucks captured per day of time at three Case 2A locations: I-65 South (top), I-65 North (middle) and I-24 West (bottom), on April 8, 2017.

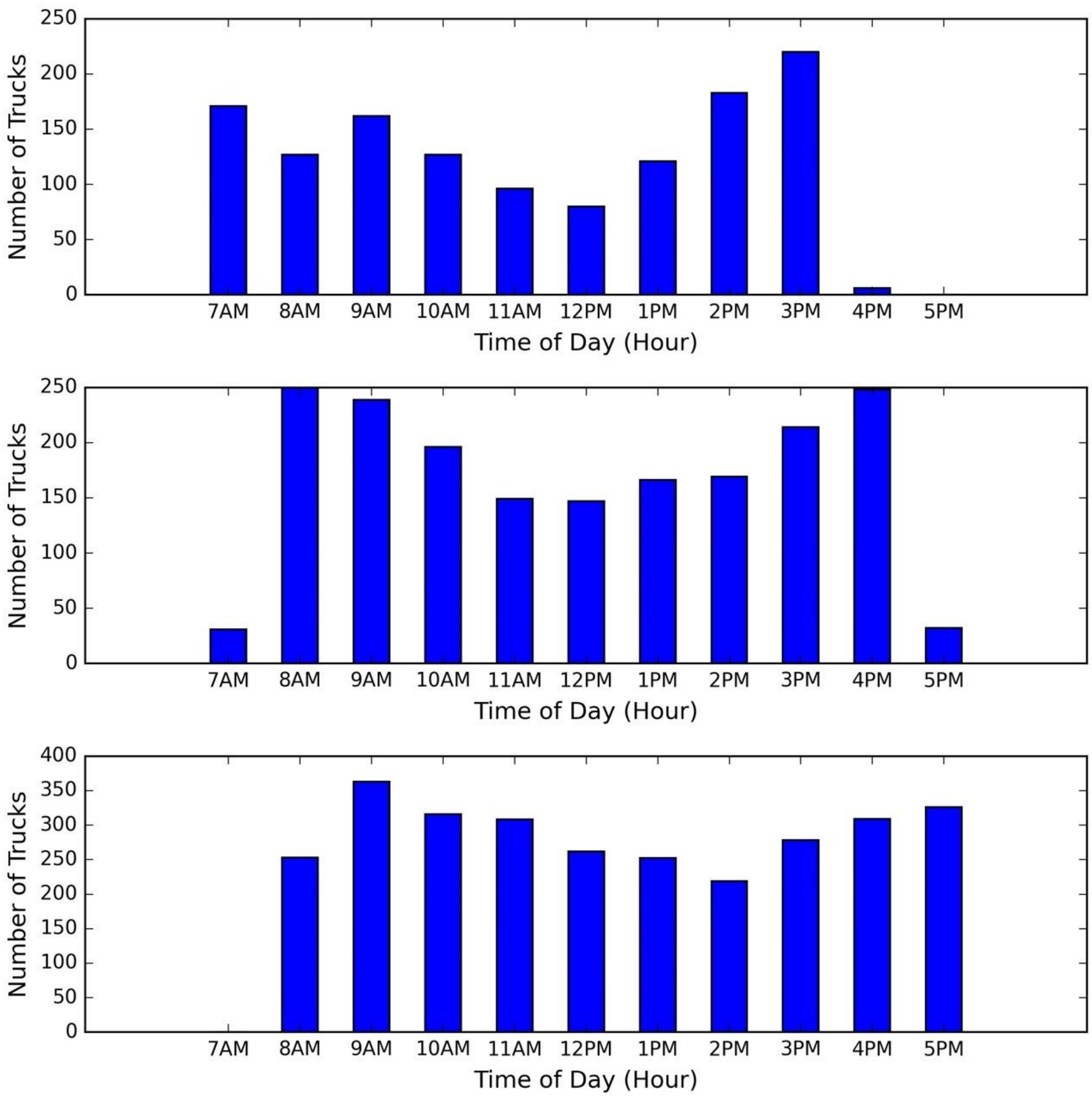


Figure 3-5 Trucks captured per day of time at three Case 2B locations: I-24 East (top), I-65 North (middle) and I-24 West (bottom), on April 9, 2017.

3.3.3 Plate Matching Results

For Case 2A, which starts on I-65 near the TN/AL border, we found 24.4% of the truck license plates were matched at I-65 North and 23.5 were matched at I-24 West, see Table 3-2. This may indicate that the northbound trucks from I-65 South are evenly split to I-65 North and I-24 West.

For Case 2B, which starts on I-24 near the TN/GA border, we found 12.2% and 27.4% of the truck plates were matched at I-65 North and I-24 West, respectively. Thus, for all the northbound trucks from I-24, about two-third stayed on I-24 while one-third switched to I-65.

Table 3-2 License plate matching results of the second field study

April 8th Case 2A	Truck Path	Distance (miles)	# of Matches	Starting Node	Ending Node	Matching Percentage
	I-65 → I-65	89.5	315	1,290	1,588	24.4%
	I-65 → I-24	85.7	303	1,290	1,214	23.5%
April 9th Case 2B	Truck Path	Distance (miles)	#of Matches	Starting Node	Ending Node	Matching Percentage
	I-24 → I-65	141	158	1,293	2,886	12.2%
	I-24 → I-24	136	354	1,293	1,841	27.4%

Again, Section 3.2.3 explains some of the factors affecting the matching percentage values. The shorter distances in this second field study likely contributed to the decidedly higher matching percentages than those in the first study.

3.3.4 Data Analysis

For Case 2A, the license plate matching results showed that the mean journey time from I-65 South to I-65 North is about 2 hours 42 minutes (162 minutes) with an average journey speed of 33 mph and the mean journey time from I-65 South to I-24 West is about 3 hours and 20 minutes (200 minutes) with an average journey speed of 26 mph. The journey speed figures on both truck paths are significantly lower than the speed limit and the typical average travel time for passenger cars for a mere 90-mile distance. While some of the trucks did travel through the corridor without stopping, with an average travel time of just under an hour, it appears that a significant portion of the trucks must have stopped for whatever reasons in the vicinity of Nashville, for as long as half a day, before continuing on the same route. It should be noted that the extra time, in hours, that took these trucks to traverse the corridor were not resultant from local traffic congestions but stops.

Figure 3-6 shows the distribution of individual truck journey time from I-65 South to I-65 North and to I-24 West for Case 2A. We can observe that both routes show a similar pattern with a long tail, many occurrences tapering away far from the central portion of the distribution, which is consistent with the typical long-tailed journey time distribution. Figures 3-7 and 3-8 show the mean truck journey time based on time of day and the number of matched trucks, respectively, at the TN/AL location (the starting node). Figure 3-7 may, at a first glance, suggest trucks passing the first station in the morning are more likely to have a longer journey time. The reality though is that's the artifact of a decreasing number of trucks captured at both stations as time goes on. This effect becomes pronounced when truck journey times are comparable or, for some cases, longer than the remaining data collection time at the second station. While all trucks with short journey time, e.g., an hour, can be captured at both stations for most of the day, only the early trucks can be captured at the second station if they had very long journey time. Since the time window for capturing a truck again at the second station continues to shrink as time goes on, the portion of trucks with longer journey time would not make it during the study time window would continue to shrink also. This eventually led to a reduce number of trucks captured and reduced average journey time throughout the day.

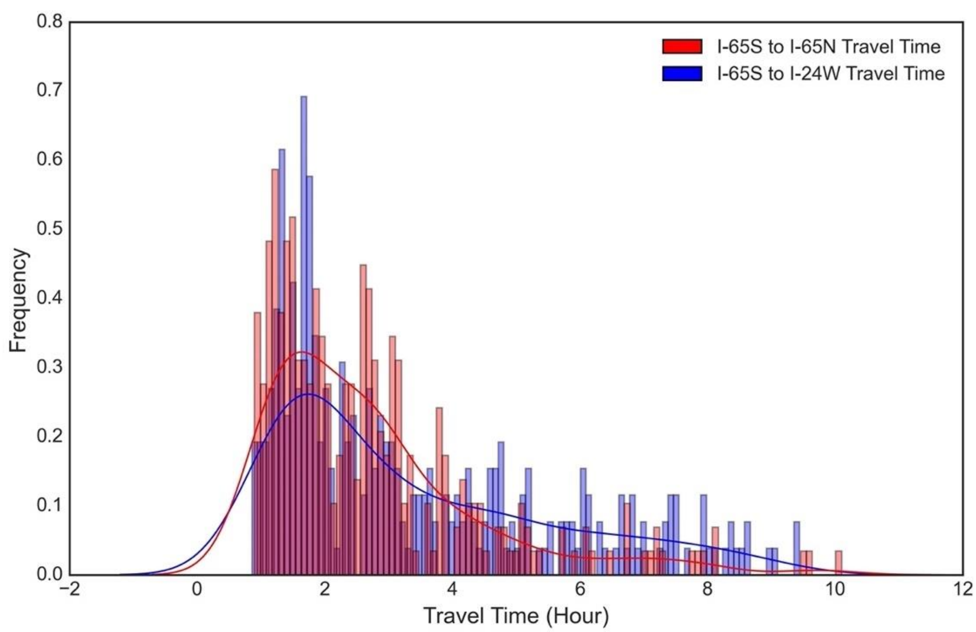


Figure 3-6 Distribution of truck journey time for Case 2A

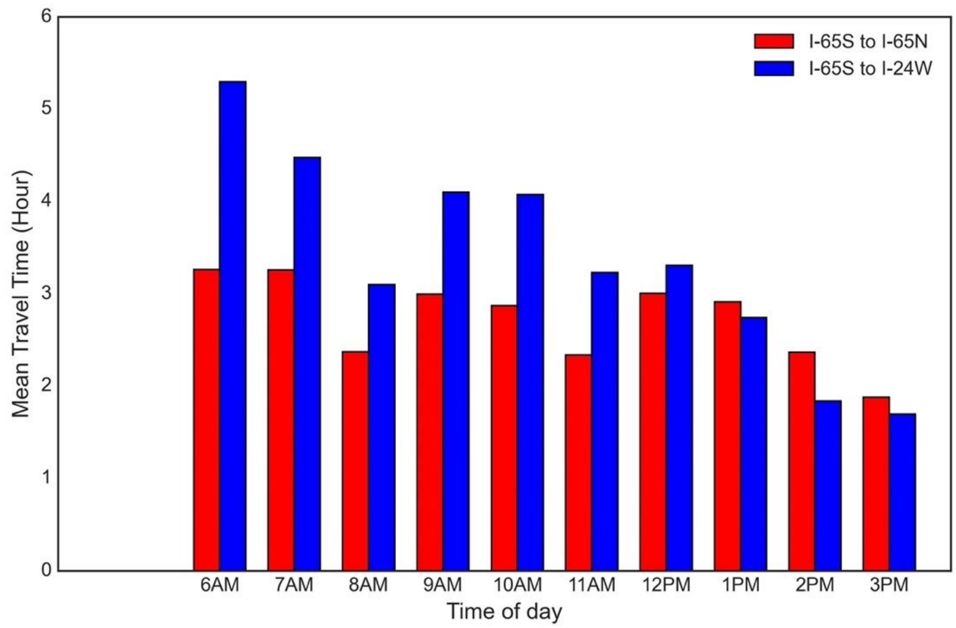


Figure 3-7 Mean journey time based on time at starting node for Case 2A

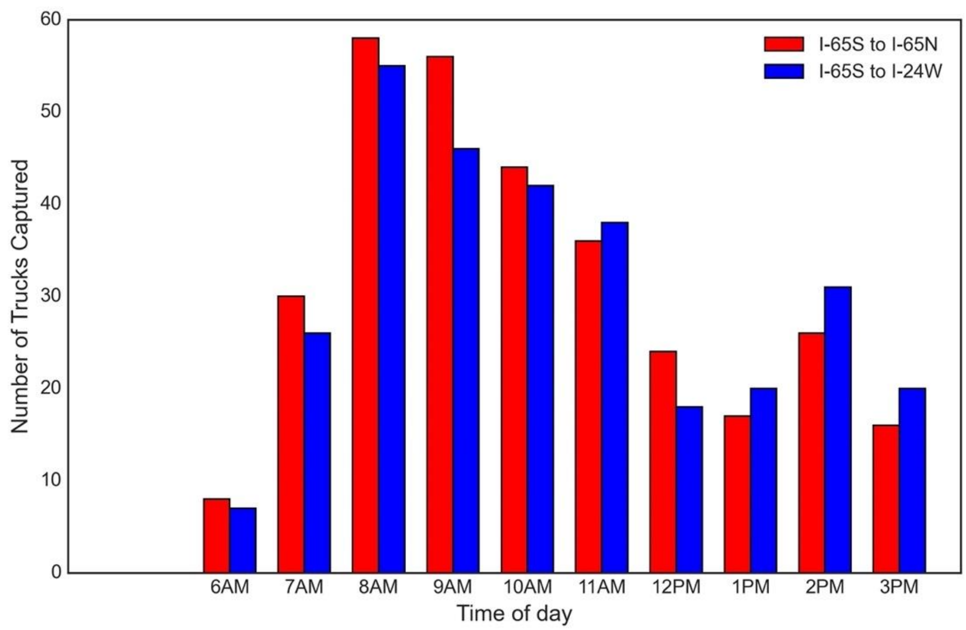


Figure 3-8 Number of matched trucks based on time at starting node for Case 2A

For Case 2B, the license plate matching results showed that the mean journey time from I-24 East to I-65 North is about 5 hours 12 minutes (312 minutes) with an average journey speed of 27 mph and the mean journey time from I-24 East to I-24 West is about 3 hours and 4 minutes (184 minutes) with a slightly more reasonable average journey speed of 44 mph. Figure 3-9 shows the distribution of travel time for I-24 East to I-65 North I-24 West. We can observe that the two directions showed different patterns. Few of the trucks staying on I-24 all the way through stopped and took longer time than needed. But a much larger portion of trucks switching over to I-65 stopped along the way.

Figures 3-10 and 3-11 show the mean truck journey time per hour of the day and the number of matched trucks passing the TN/GA border station each hour of the day for Case 2B. From Figure 3-10, for the direction from I-24E to I-65N, the mean journey time continuously and steadily declined, while for the direction from I-24E to I-24W, the mean travel time also declined steadily but of a much less magnitude throughout the study duration. Also, more trucks are captured in the morning for direction from I-24E to I-65N, whilst for direction I-24E to I-24W, more trucks are captured in the early afternoon.

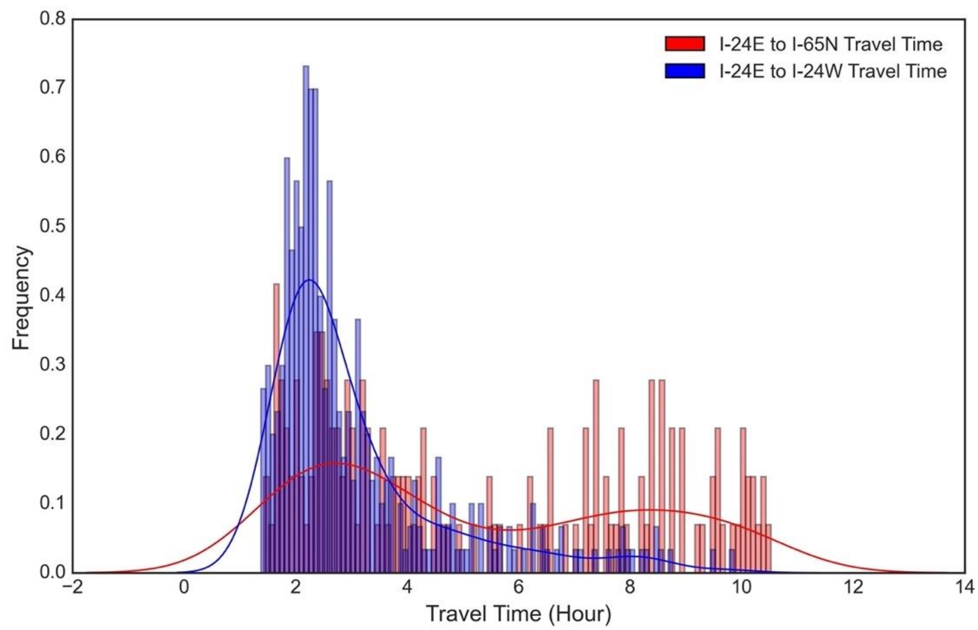


Figure 3-9 Distribution of truck journey time for Case 2B

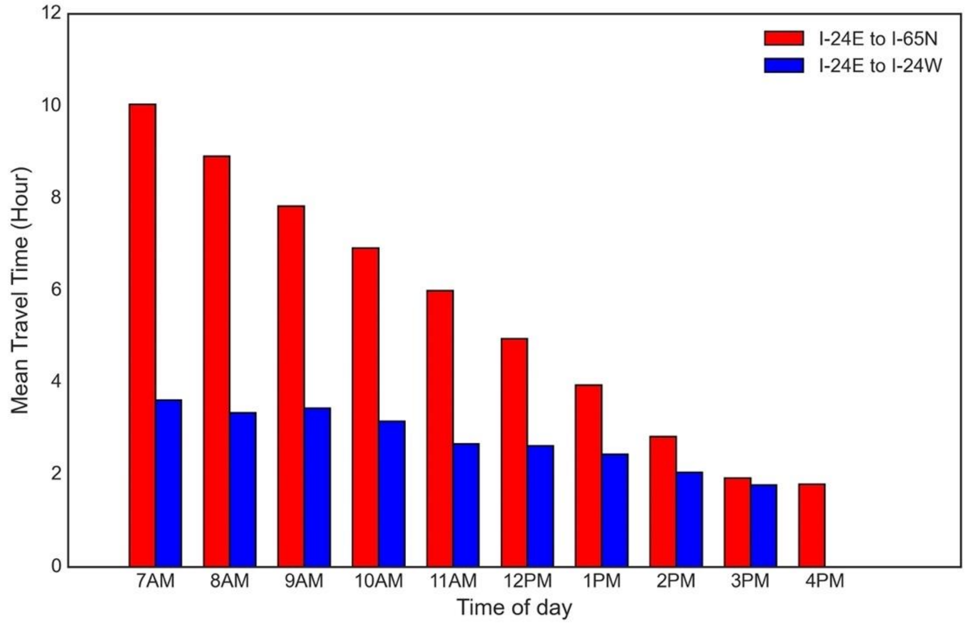


Figure 3-10 Mean journey time based on time at starting node for Case 2B

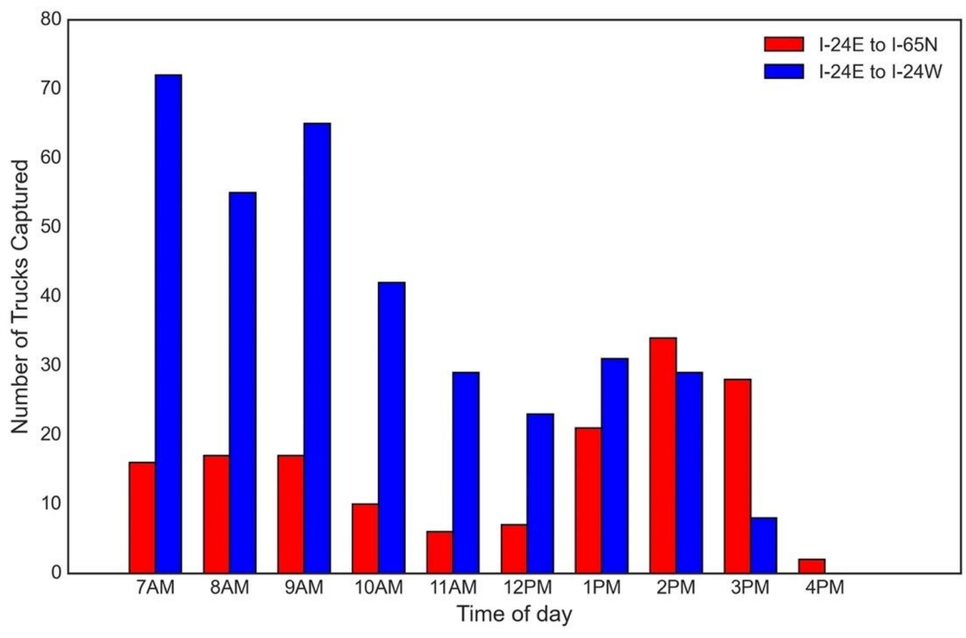


Figure 3-11 Number of matched trucks based on time at starting node for Case 2B

3.3.5 Findings

Two data collection outings with different configurations were conducted for the second field task, tracking northbound trucks in all lanes on I-65 South/I-24 East to I-65 North and I-24 West. Overall, the ALPR matching rates were good, maxing out at about 25%.

The results for the field Case 2A show that the northbound trucks from I-65 South are evenly split onto I-65 North and I-24 West. For both directions, more trucks are captured in the morning and tend to have a long journey time, which would be due to stops in the middle of the long day.

For field Case 2B, the matching/tracking results show that for all northbound trucks on I-24 from Georgia through Nashville and to Kentucky, about two-third stayed on I-24 while one-third switched to I-65.

3.4 Case 3 – I-40/840 to I-40

3.4.1 Field Study Setup

The third field study was conducted on May 8, 2017, tracking Eastbound trucks from the west side of Nashville on I-40E and I-840E to the east side of the city after the two roads merge, see Figure 3-12. The distance of the path along I-40E is about 59.3 miles and the distance of the path along I-840 is 77.28 miles. This study attempts to understand the truck route choice when two routes are presented.

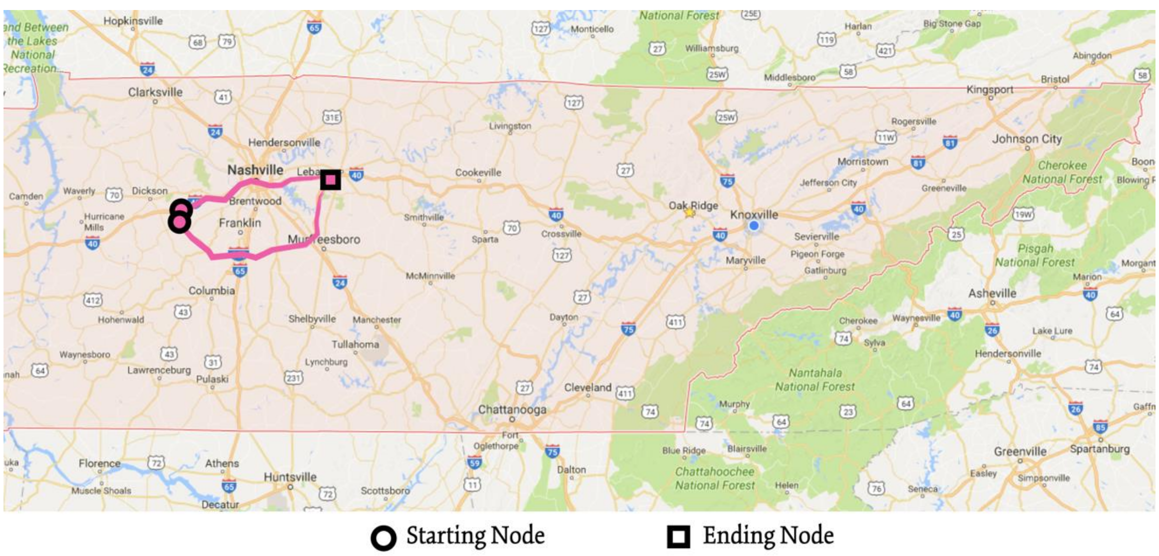


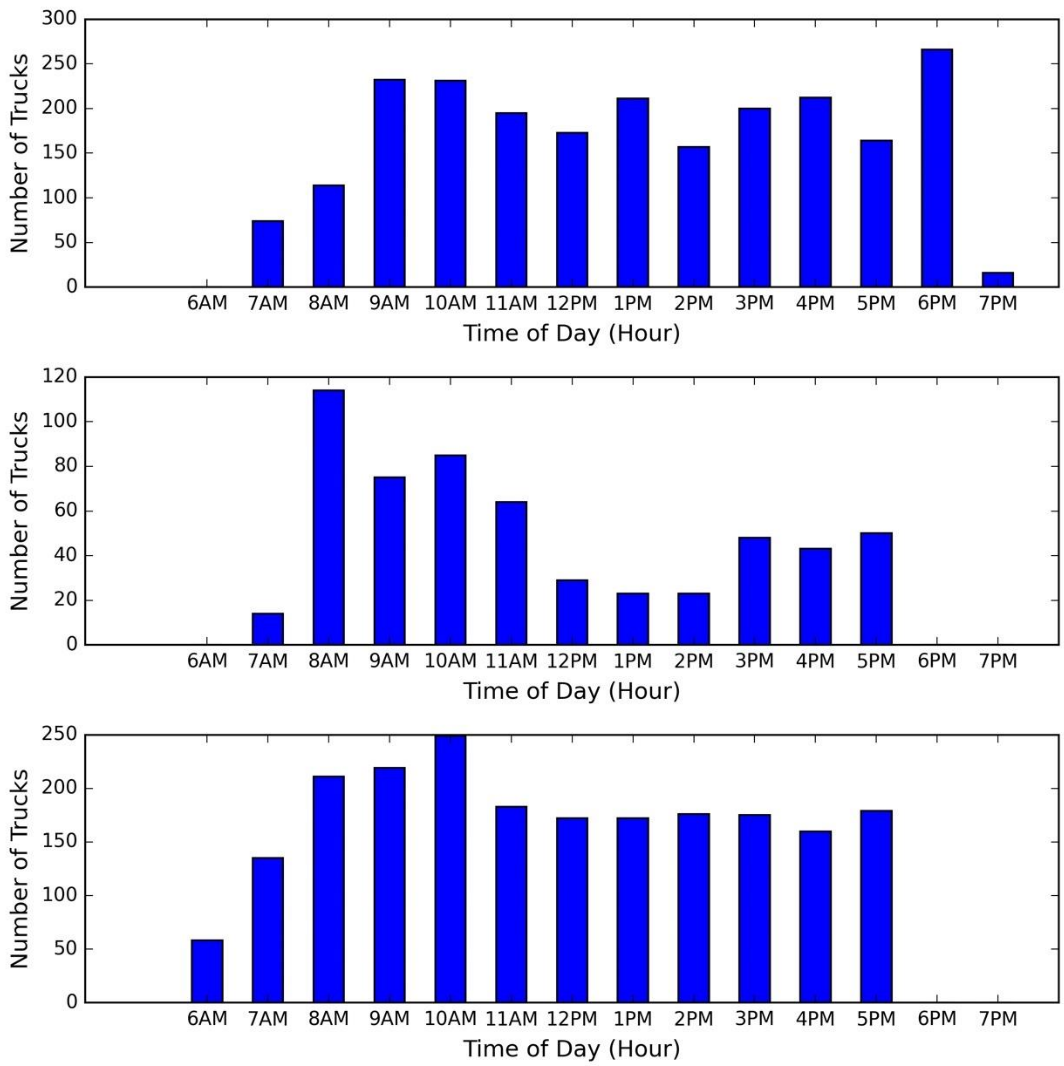
Figure 3-12 Location map of the third field study

The ALPR devices were set up by well-trained graduate and undergraduate students at the University of Tennessee. Three groups of students set up the ALPR devices at the three strategic locations covering the two ends of each truck path from 6AM to 6PM.

It should be mentioned that during the study, we also collaborated with TDOT's Region 3 Traffic Management Center (TMC), which put out messages on the Dynamic Message Signs (DMS) to encourage trucks take I-840 to avoid congestion during peak hours.

3.4.2 Data Assessment

Again, the field data collection effort yielded 2,089; 568; and 2,298 truck license plates at I-40E west of Nashville, I-840 west of Nashville, and I-40E east of Nashville, respectively. Figure 3-13 shows the trucks captured by ALPR at three locations based on time of day for Case 3. From the figure, constantly high truck volumes were observed from 9 AM to 6 PM on I-40E on both ends of Nashville. whileI-840E carries a relatively lower truck volume.



;

Figure 3-13 Trucks captured based on time of day at I-40 West of Nashville (top), I-840 West of Nashville (middle) and I-40 East of Nashville (bottom), on May 8, 2017.

3.4.3 Plate Matching Results

After running the license plate matching algorithm, the results are tabulated in Table 3-3. Some 602 out of 2,298 (26.2%) license plates were from I-40E west of Nashville while 287 out of 2,298 (12.5%) were from I-840. Thus, for eastbound trucks on I-40 E, about two-third passed through the city via I-40E while one-third used I-840E.

Table 3-3 License plate matching results of the third field study

May 8th	Truck Path	Distance (miles)	# of Matches	Starting Node	Ending Node	Matching Percentage
	I-40E → I-40E	59.3	602	2,089	2,298	26.2%
	I-840 → I-40E	78.5	287	568	2,298	12.5%

3.4.4 Data Analysis

For field Case 3, the license plate matching results showed that the mean journey time along I-40E was about 3 hours 16 minutes (196 minutes) with an average journey speed of 18 mph, and the mean journey time on I-840 is about 2 hours and 48 minutes (168 minutes) with an average journey speed of about 28 mph. Similar to the discussion in 3.3.4, the shrinking time window for capturing the eastbound trucks on I-40 or I-840 underestimated the average journey time on both routes. The journey times for trucks making no stops through these routes were short, around an hour, and well represented in the collected data. But the journey times of trucks making long stops or taking long detours then resuming the trip east, either on I-40 or I-840, were not all captured and decreasingly so as time went on during the study. Figure 3-14 shows the distribution of journey time on I-40 and on I-840. We observe that both directions showed similar patterns with a long tail, a slowly declining number of occurrences away from the main part of the distribution, which is consistent with the typical journey time distribution on the road.

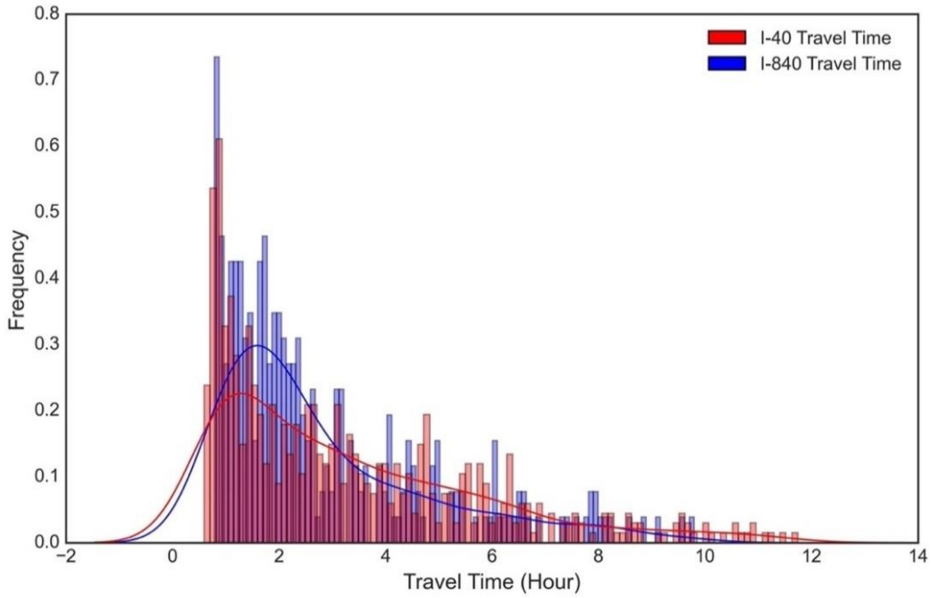


Figure 3-14 Distribution of truck journey time for field Case 3

Figures 3-15 and 3-16 show the mean truck journey time per time of the day and number of trucks captured each hour of the day for Case 3, respectively. From Figure 3-15, we can observe that, for both routes (I-40E and I-840E), trucks are more likely to have a longer journey time in the morning. A plausible explanation is while all trucks with short journey time can be captured throughout most of the day, only the earlier trucks can be captured at the ending node if they had very long journey time, due to stops or detours. Since the time window for capturing a truck again at the second station continues to shrink as time goes on, the portion of trucks with longer journey time would not make it to the east side of Nashville during the study time window would continue to shrink also. This eventually led to a reduced number of trucks captured and a declining average journey time throughout the day.

3.4.5 Findings

Case 3 attempted to understand the truck route choice when two alternate routes were present. The ALPR matching algorithm was performed, and the matching rates are within the acceptable range, as high as 26%.

The results for Case 3 indicate about two-thirds of truck drivers opted for I-40 through Nashville while the other one-third opted for an 18-mile longer but less congested route bypassing Nashville. The mean journey time on I-840 is almost half an hour shorter than that on I-40. This could have something to do with the nature of the type and length of stop made on each route.

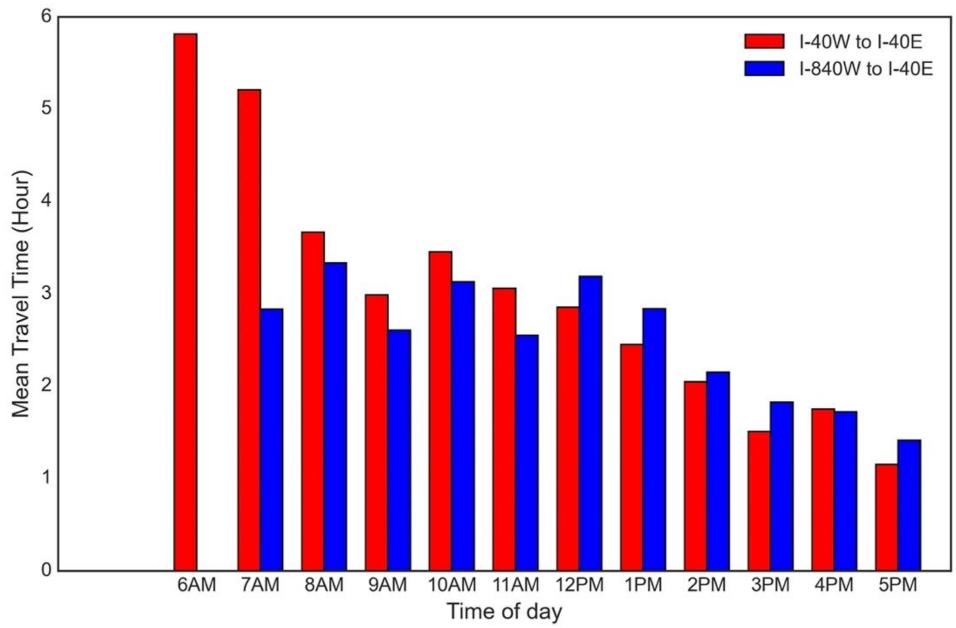


Figure 3-15 Mean journey time based on time at starting node for Case 3

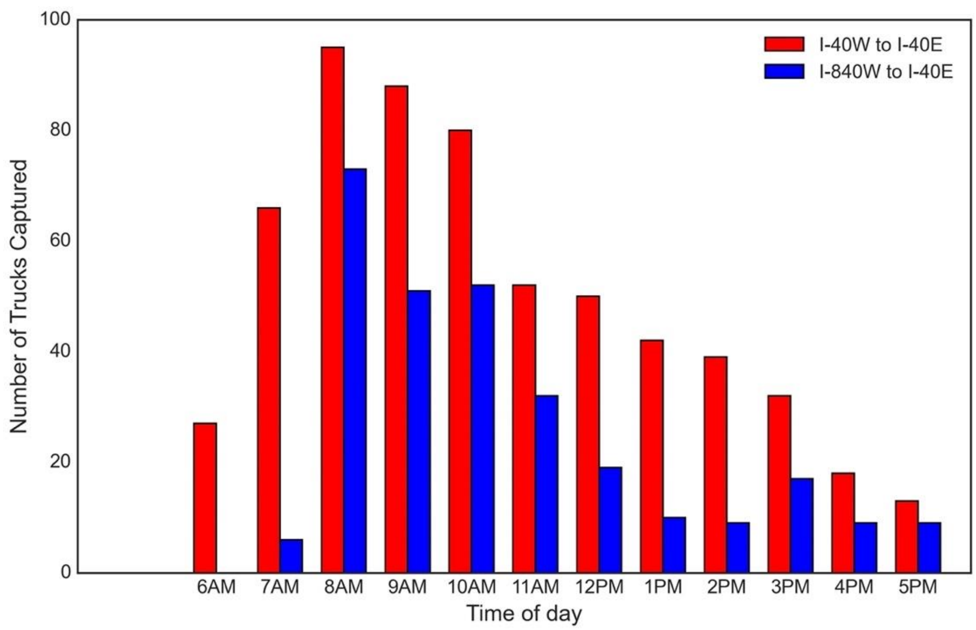


Figure 3-16 Number of matched trucks based on time at starting node for Case 3

3.5 Summary

This chapter consists of three field case studies performed by UTK research team. The cases were conducted for evaluating the LPR matching algorithm in long-distance scenarios which is challenging for vehicle tracking and unsupervised learning algorithms. For each study, three locations were chosen as starting node or ending node; then ALPR devices were deployed to capture truck license plates at these locations. Truck license plates captured at the starting node and ending node were matched using ALPR matching algorithm developed by the UTK research team. Truck travel patterns and route choices were observed.

The results demonstrated the potential of using ALPR matching algorithm in long-distance truck tracking scenarios. For the first case, the results showed that for all the northbound trucks from I-75 South, about three-quarters stayed on I-75 while one-quarter switched to I-81. For the second case, the results Case 2A show that northbound trucks from Alabama (I-65N) were evenly split onto I-65N and I-24 West trough Tennessee. Case 2B, in the meantime, showed that northbound trucks on I-24 from Georgia through Nashville and to Kentucky split with about two-thirds staying on I-24 while one-third headed to I-65N. For the third filed study, the results showed that about two-thirds stayed on I-40E while one-third opted for I-840E.

CHAPTER 4 WEIGH STATION DATA ANALYSIS

4.1 Data Source and Data Assessment

The team was eventually provided some ALPR data at a selected number of weigh stations by Tennessee Department of Safety and Homeland Security (TDSHS). The data was from a period of about three months (November 2016 to February 2017) with license plates strings, but no images, collected at nine Tennessee weigh stations, including:

- Coffee Co Manchester I-24E and I-24W @ MM:115
- Giles Co I-65N @ MM:5
- Greene Co I-81S @ MM:21*
- Haywood Co I-40E and I-40W @ MM:50
- Knox Co I-40W*
- Knox Co I-40E
- Unicoi Co I-26W

* Two stations had no data for the 3-month study period but will have data and can be used for future studies

A total of 1,071,295 license plates captured at these stations were obtained in the TDSHS database. Unfortunately, there were no entries for Knox I-40W or Greene I-81S. Figure 4-1 illustrates the locations of the weigh stations and the number plates captured at these stations.

Haywood	Giles	Coffee	Knox	Greene	Unicoi
I-40E: 134,319	I-65N: 123,084	I-24E: 268,361	I-40E: 215,689	I-81S: 0	I-26W: 31,449
I-40W: 132,832		I-24W: 165,561	I-40W: 0		

Figure 4-1 Location and obtained number of license plates for each weigh station

Given the available data, we chose to track eastbound (EB) trucks from Haywood EB to Coffee EB and Knox EB to explore truck travel patterns with the LPR matching algorithms developed by Dr. Han's research team.

4.2 Case 4 – Eastbound Trucks from Haywood EB to Coffee EB and Knox EB

4.2.1 Field Study Setup

In Case 4, the Eastbound trucks from Haywood EB weigh station were tracked to Coffee EB weigh station and Knox EB weigh station (Figure 4-1). The distance from Haywood EB to Coffee EB is about 229 miles, and the distance from Haywood EB to Knox EB is about 322 miles.

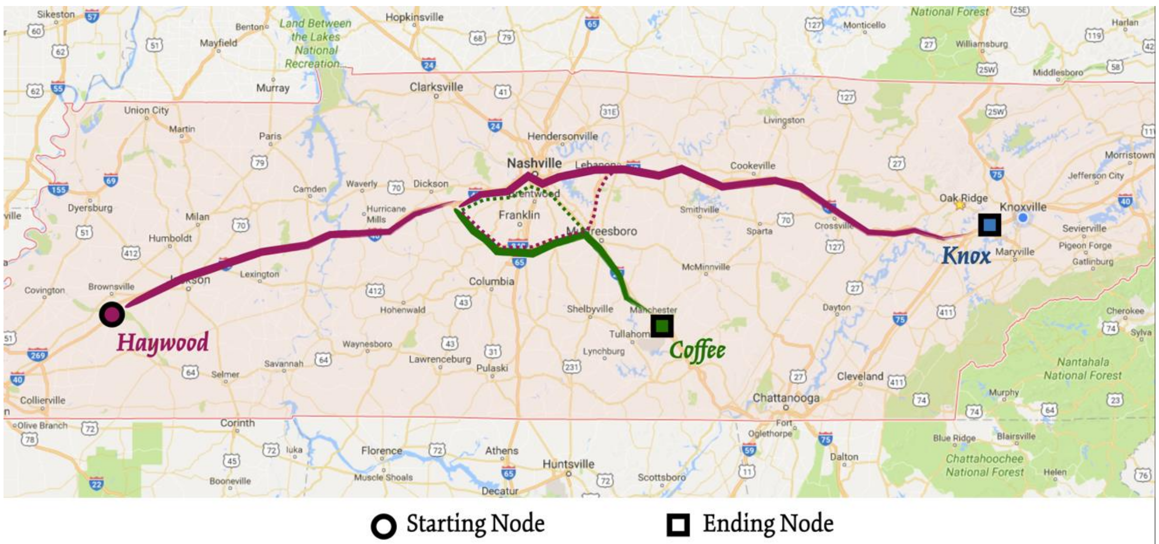


Figure 4-2 Map of route case 4

4.2.2 Data Assessment

We obtained 134,319; 215,896; and 268,361 license plates for Haywood EB, Knox EB, and Coffee EB, respectively. The distribution of the number of license plates captured by the time of the day at these three weigh stations are shown in Figures 4-3 through 4-5. A few notable points:

- For Haywood EB, most of the trucks arrive after noon. While the station was capable of operating 24/7, it is unclear if there were equipment failures or closures due to other reasons during the study period.
- For Knox EB, the pattern for weekdays is significantly different from that for weekends. Most of the trucks arrive in the afternoon and evening on weekdays. On Saturdays, more trucks arrive in the early morning. Still more trucks arrive in the afternoon (around 6 PM) on Sundays.
- For Coffee EB, it seems that the pattern is constant for each day of the week. More trucks arrive at the weigh station during the PM peak.
- All these weigh stations are designed to function 24/7, but closures due to various reasons could have occurred. Also, trucks equipped with PrePass transponder can bypass the weigh stations and would not be captured if they did so.

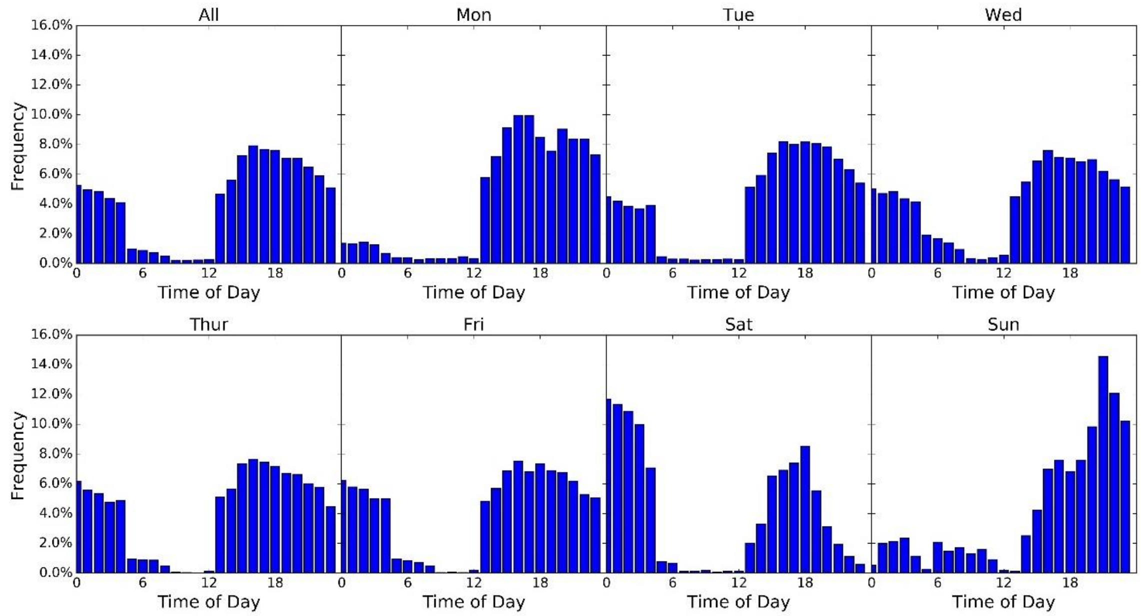


Figure 4-3 The weekly pattern of license plates by the hour at Haywood EB weigh station

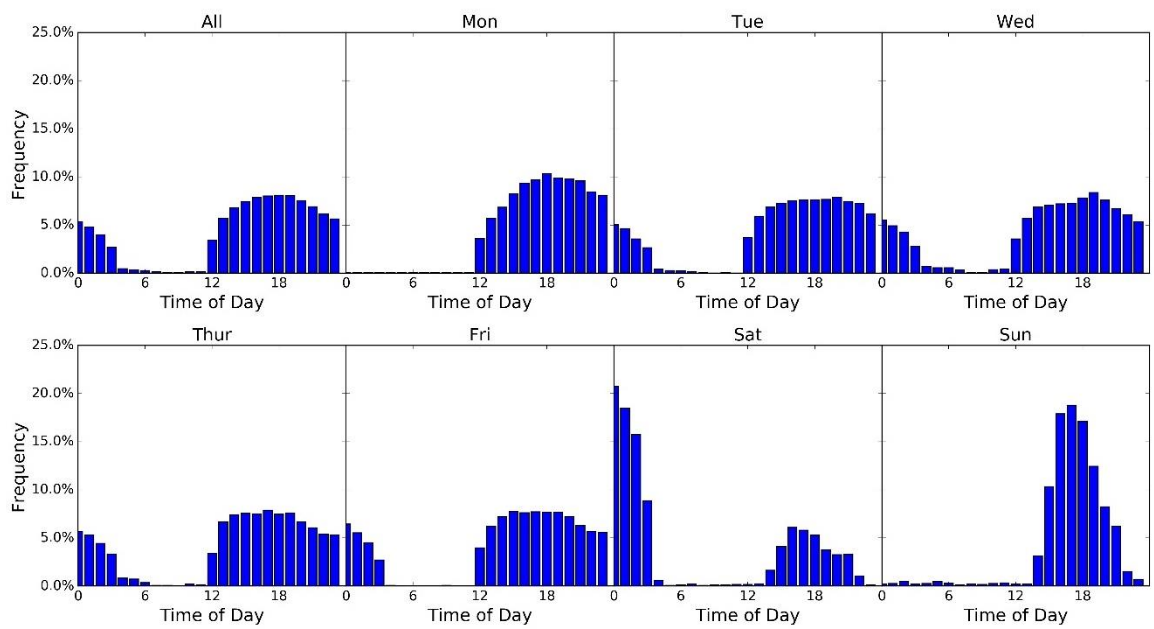


Figure 4-4 The weekly pattern of license plates by the hour at Knox EB weigh station

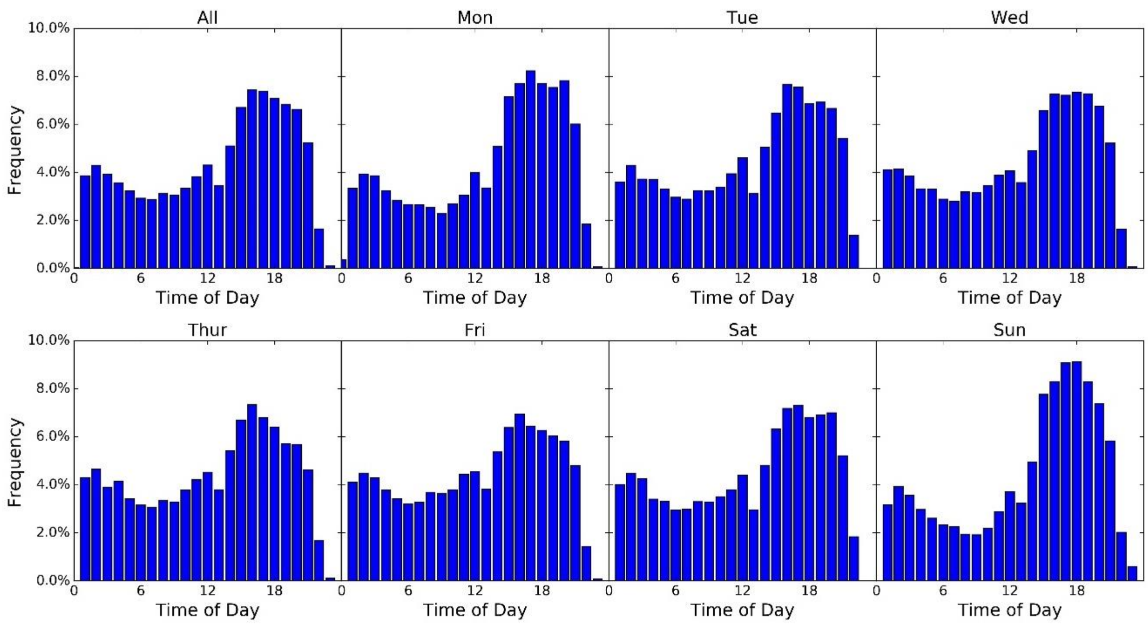


Figure 4-5 The weekly pattern of license plates by the hour at Coffee EB weigh station

4.2.3 Plate Matching Results

Truck plates captured from Haywood EB to Knox EB or to Coffee EB were matched up with the ALPR matching algorithm. Results show 13.2% of Haywood EB trucks trekked their ways to Knox EB while only 4.0% from Haywood EB went to Coffee EB. Considering more than 20,000 plates were matched among millions of possible matching permutations, the matching algorithm seemed to have performed well.

Table 4-1 License plate matching results of case 4

Locations	Distance (miles)	Number of Matches	Starting Node	Ending Node	Matching Percentage
Haywood EB → Knox EB	322	17,692	134,319	215,896	13.2%
Haywood EB → Coffee EB	229	4,872	134,319	268,361	4.0%

4.2.4 Data Analysis

Studying the patterns of matched license plates for Haywood EB and Knox EB, we observe that:

- The travel time distribution is different between weekdays and weekends. On weekdays, there is a bimodal travel time distribution with a significant number of the trucks having a reasonably short travel time around 6 hours. On weekends, many trucks have much longer travel times making stops, perhaps around Nashville area, along the way (Figure 4-6).

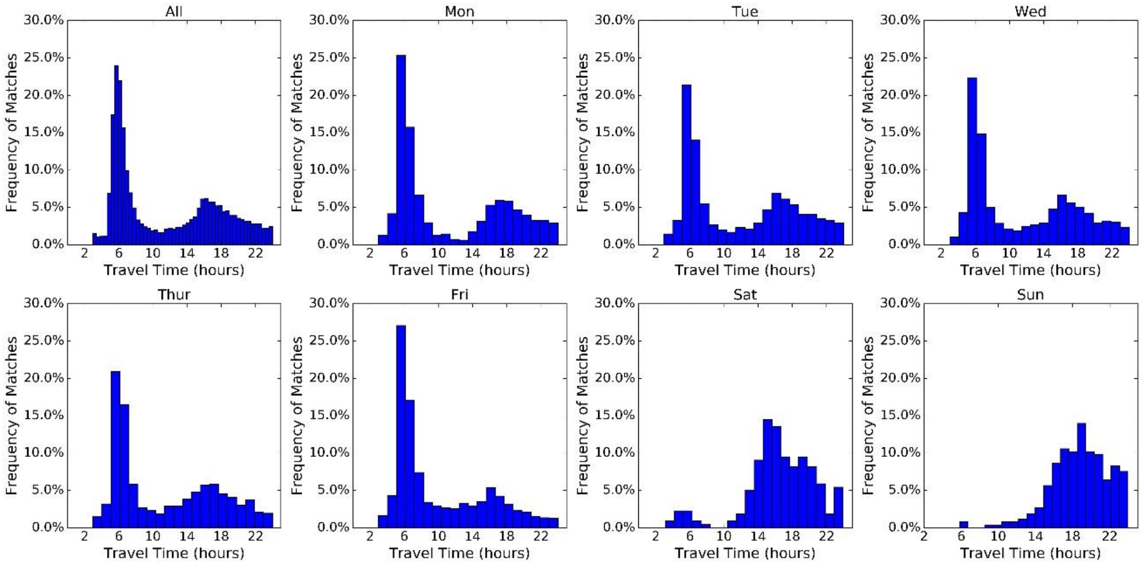


Figure 4-6 Journey time distribution for all matched trucks from Haywood to Knox

- Most matched trucks passed through the Haywood weigh station after noon time and very few passed Haywood in the late morning. This may have something to do with the truck traffic demand pattern. There appears to be a flood of trucks passing through Haywood in early Saturday hours for some reason. On the other hand, Sunday is quiet with most trucks passing through Haywood weigh station after dinner hours. (Figure 4-7).

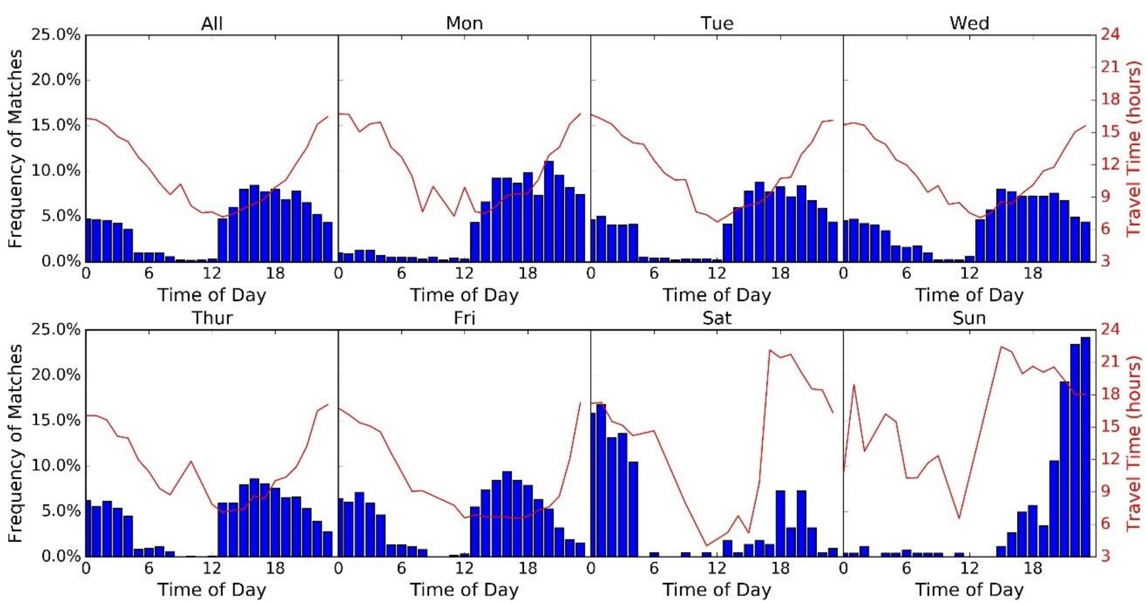


Figure 4-7 Number of trucks matched from Haywood EB to Knox EB per time of day and associated mean journey time

- Trucks passing through Haywood weigh station during daytime have a lower overall journey time. Trucks passing by during early morning or late afternoon tend to take much longer to arrive at Knoxville weigh station. (Figure 4-7).

For Haywood EB to Coffee EB, we found that,

- The travel time distribution seems to be relatively consistent for all days of the week (Figure 4-8). The journey of about 220 miles, which should take a 3 to 4 hours, saw many trucks taking a far longer time to complete for some reason.

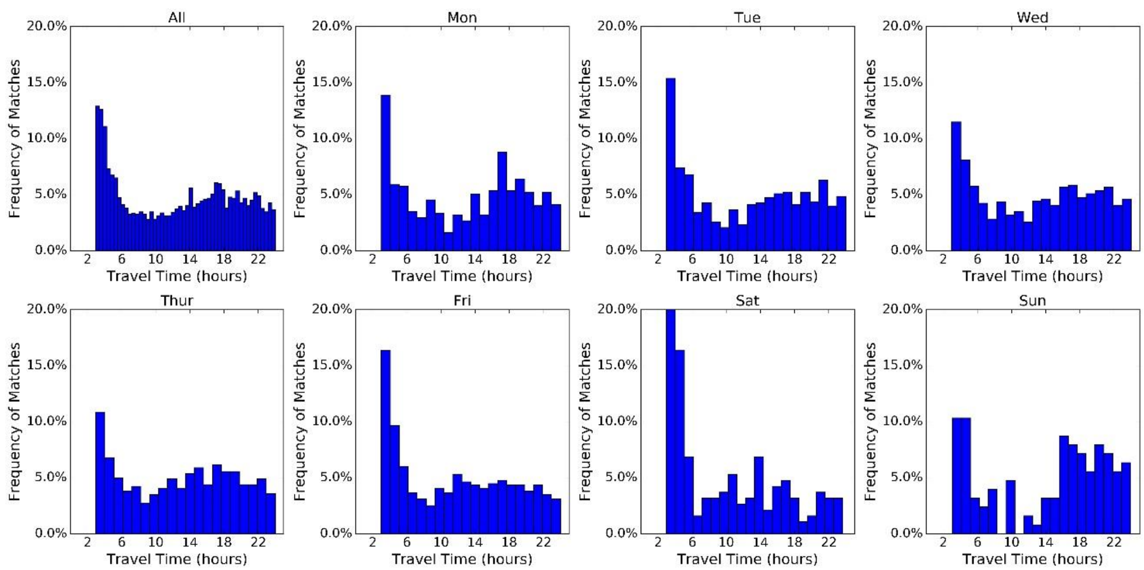


Figure 4-8 Travel time distribution for all matched trucks from Haywood EB to Coffee EB

- Weekday travel pattern appears to be rather stable and is quite different from those on weekends. Most matched trucks passed through the Haywood weigh station from noon time to midnight from Sunday through Friday, but for Saturday, most matches are found in the early morning. (Figure 4-9).
- Trucks passing through Haywood weigh station in the Coffee weigh station direction have a wide range of journey times. It is less stable or predictable like the travel time for trucks from Haywood toward Knoxville (Figure 4-9).

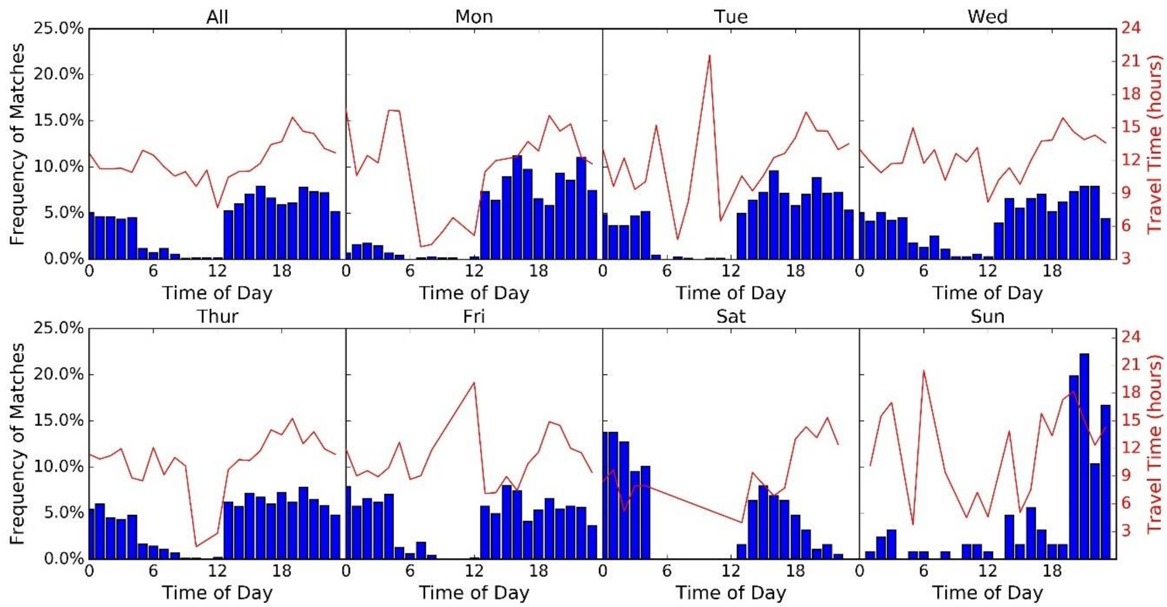


Figure 4-9 Number of trucks matched from Haywood EB to Coffee EB based on time of day and associated mean travel time

4.2.5 Summary of Findings

Case 4 study explored the travel pattern of trucks from Haywood EB to Coffee EB and Knox EB, which can provide insights into the freight distribution pattern. After running the LPR matching algorithm, 13.2% and 4.0% Eastbound trucks from Haywood EB were tracked at Knox EB and Coffee EB, respectively.

The results for Case 4 study show that for trucks from Haywood EB weigh station, about three-fourths were matched at Knox EB weight station, while one-fourth was tracked at Coffee EB weigh station. In addition, the distribution of journey time differs substantially between weekdays and weekends for both routes (Haywood EB to Knox EB and Haywood EB to Coffee EB).

4.3 Thoughts

The case study in this chapter demonstrates the potential usefulness of TDSHS weigh station license plate data in understanding freight truck route choice and freight mobility pattern. Some of the pros and cons are identified below.

The Pros:

- Temporal Coverage – Unlike the mobile ALPR units used in the first three field studies, as presented in Chapter 3 of this report, TDSHS use permanently mounted ALPR units, which can operate 24/7/365 as long as the stations are open. This would effectively address the shrinking capture time window issue discussed in Chapter 3 that could result in underestimated average journey times.

- ALPR Accuracy – Permanently mounted ALPR units usually have higher accuracy than the mobile units.
- Already Deployed – These weigh station ALPR units are already deployed by TDSHS and have collected tens of millions of license plates in the past. Using the data would not incur additional costs to TDOT. They could also be incorporated into a larger ALPR truck monitoring network that TDOT could develop and deploy throughout the state.

The Cons:

- Spatial Coverage – Only slightly more than a handful of the weigh stations are equipped with operational ALPR units. To effectively cover all the major truck routes, significantly more ALPR stations are needed to obtain more complete truck travel statistics.
- Station Outages – It was evident that there were recurring scenarios when these weigh stations were not open during certain time of day or under certain conditions. For the purpose of truck tracking and license plate matching, the outage of a single station could render data collected at many other stations less useful.
- Bypassing Rate – For as little as $18/month/truck, truckers equipped with PrePass are allowed to bypass weight stations. Since there are about 600,000 users of the PrePass system, a portion of the trucks may not be effectively tracked by weight station ALPR units.

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

This study accomplished all tasks it set out to complete except the analysis with ATRI individual truck GPS trajectory data, which was not available. Instead, TDSHS weigh station license plate data were obtained to explore the freight mobility patterns. By using the ALPR matching algorithm developed by UTK, we were able to evaluate the travel patterns and the proportion of on different routes among the matched trucks. For example, Case 3 of this study was conducted to track eastbound (EB) trucks using I-40E or I-840E pass through the Nashville metropolitan area. The results show that about two-thirds opted for I-40E while the other one-third opted for the much longer route that is I-840.

With the TDSHS weigh station ALPR data, we also attempted to explore the route choice of trucks with a long travel distance. For example, for route Haywood EB to Coffee EB and Knox EB, 13.2% and 4.0% of eastbound trucks from Haywood EB were tracked at Knox EB and Coffee EB, respectively. The analysis results demonstrated the potential and challenges of using the existing (already collected) weigh station ALPR data for acquiring better understanding of freight mobility patterns.

The project did not yield high matching percentages while attempting the long-distance truck tracking objective with mobile ALPR units. This was an expected result because long distance tracking using passive ALPR technology has the challenge of the traffic flow being spatially diluted and temporally dispersed. A network of permanent ALPR installations would overcome many of the challenges we identified below. The TDSHS ALPR data employed in this study is an existing and paid-for source of information to TDOT and could be of use in the future.

Based on the finding from this study, here are several conclusions and recommendations:

- Passive tracking using portable ALPR devices (as demoed by UT in this project) can be labor-intensive and, hence, expensive. The lower percentages of trucks successfully tracked in this study can be attributed to several reasons:
    1) The technical challenges associated with long-distance tracking were known to be difficult. The self-learning algorithm that worked well with short-distance tracking scenarios did not have enough training samples, a high number of trucks traversing both ALPR stations within the study time window, to be proficiently "trained."
    2) The field studies were performed over a few hours, from sunrise to sunset for most cases, in the interest of the safety/security of the students in the field. With hundreds of miles to travel for the subject trucks, only a portion of them were able to complete the journey within the study time window and be tracked at both the start and the

end stations. Many of the trucks stopped along the way for various reasons. This is evident from the fact that a significant number of trucks had journey times much longer than the non-stop travel time between the two stations otherwise. As such, many of the trucks captured at one station would not show up at the other station, or vice versa, during our studies.

3) A significant number of trucks that happened to pass one station would not pass the other station hundreds of miles away because of the wide ranges of origins, destinations, and route choices. For two ALPR stations a short distance apart, say 5 miles, on a stretch of Interstate highway, it is highly likely a truck would be captured at both stations within a few minutes. But when that distance is extended to 50 miles with some stops and alternate routes in between, the likelihood that the same truck would be captured at both stations declines quickly and requires a longer "time window" to allow for the variability in travel time. When that distance is extended to 250 miles with many Interstate junctions and even more potential stops in between, the likelihood a truck would pass though the two stations within a few hours drops down to single percentage points.
4) A potential measure to handle the leakage problem of trucks and address the constraint of the study's short time window is to add additional ALPR stations between the two end stations hundreds of miles apart. Such would break down the long-distance tracking challenge into an array of more handleably shorter-distance tracking exercises. This approach was cost-prohibitive for this pilot study but is likely feasible for larger scale implementations by TDOT.
5) An alternative or a supplement approach to additional ALPR stations mentioned above is to use combined and complementary tracking mechanisms. For example, the Bluetooth technology could be used at ALPR stations to collect unique identification ID that could then be tracked subsequently more economically. We have used ALPR and Bluetooth technologies in a TDOT project previously for vehicle speed studies on I-40 in Nashville.

- Permanent stations deployed at strategic locations, would be more reliable than the mobile research units used in this study, if TDOT wants to track trucks on a continuous basis. This could significantly open the narrow "time window" our studies were subject to and would track trucks even if the journey time were over a day. Costs would be a consideration depending on the associated investment in the ALPR infrastructure and extensiveness of deployment and connectivity. Other considerations include:

1) Active GPS tracking is expensive, not in real-time, and TDOT will depend on an external for-profit entity, such as ATRI, to provide the aggregated statistical information.
2) Weigh station ALPR data are already being collected and managed by TDSHS and should be free to TDOT through state-level interagency data-sharing or data-exchange agreements. UT could help maximize the utility of these data sources and curate them for TDOT's freight mobility planning and operation purposes. In this study, we only acquired a limited amount of data from some TDSHS weight stations. A more thorough year-round study on the recurring travel patterns of all recorded trucks would be desirable. Currently many of these stations do not operate 24/7; a discussion with TDSHS on that could make the database more complete and useful. A continuous analysis report or an annual report based on the data could be insightful. If and when TDOT deploys its own WIM or roadside weigh stations, data from TDSHS stations could still be very valuable.
3) Using dynamic message signs and other means, such as the WAZE app, to entice and shape driver behaviors could be reviewed in the future as well. ALPR stations can be used to monitor and assess the performance of such "behavior modification" attempts.
4) TDOT already has video camera footages from their Smart Way Plus system. A potential use of video imaging-based technology could help identify trucks and provide truck counts at strategic locations. This will not actually track the trucks, per se, but will help TDOT gain a better understanding of the classifications of vehicles on the State's major Interstate locations.
5) TDOT's RDS database has about 800 stations on the state's Interstate highways. Of these, about 100 stations have a "long count" feature estimating "longer" vehicles, such as trucks. While this feature is not yet fully verified, UT has had some initial success in developing, in parallel to the long counts, some algorithms to estimate truck percentage from the high-resolution RDS raw data. Further development and field verification could lead to a very useful tool for the state to obtain much better appreciation of truck traffic, at lane-by-lane and minute-by-minute level, at all 800 stations across the state's major urban areas.

ACKNOWLEDGMENTS

At TDOT, Dr. Casey Langford provided guidance and feedback from inception through the stages of this study. Brad Freeze provided insightful suggestions for study sites and ALPR locations. Aayush Thakur provided guidance to the study for several months.

Significant help came from Tennessee Department of Safety and Homeland Security (TDSHS) where Captain Brandon Douglas, Captain Doug Taylor, and Lieutenant Allen England made available a month worth of weigh station ALPR truck plate data, which enabled us to expand the study beyond its original scope with promising results.

Dr. Hyeonsup Lim modified the code to implement the unsupervised learning algorithm for long-distance tracking scenarios. Dr. Yuandong Liu helped with the analyses of the massive real-world field datasets. She also participated in the field data collection effort in remote locations with other students including Bruce Applegate, Dr. Bumjoon Bae, Dr. Ali Boggs, Jason Chai, Brandon Whetsel, and Brandon Worley.

When one of the ALPR unit was fried early on after the first field study, Tad Aarant came to our assistance. Even though we were not able to recover the data on that unit, we were able to get back to the full number of units to keep the study going.

We are grateful for all their help.

REFERENCES

- Bureau of Transportation Statistics (2017) Freight Facts and Figures, U. S. Department of Transportaiton, 111 pages.
- Federal Motor Carrier Safety Administration (2015) Interstate Driver's Guide to Hours of Service, U.S. Department of Transportaiton, 27 pages.
- Hargrove, S.R., Lim, H., & Han, L.D. (2016). Expanding license plate matching capabilities with secondary self-learning algorithm. *Transportation Research Record*, 2594(1), 51-60.
- Hargrove, S.R., Lim, H., Han, L.D., & Freeze, P.B. (2016). Empirical evaluation of the accuracy of technologies for measuring average speed in real time. *Transportation Research Record*, 2594(1), 73-82.
- Oliveira-Neto, F. M., Han, L. D., & Jeong, M. K. (2012). Online license plate matching procedures using license-plate recognition machines and new weighted edit distance. *Transportation Research Part C: Emerging Technologies, 21*(1), 306-320. doi:10.1016/j.trc.2011.11.003
- Oliveira-Neto, F. M., Han, L. D., & Jeong, M. K. (2013). An online self-learning algorithm for license plate matching. *IEEE Transactions on Intelligent Transportation Systems, 14*(4), 1806-1816.
- Turner, S. M., Eisele, W. L., Benz, R. J., & Holdener, D. J. (1998). *Travel time data collection handbook*. Retrieved from
- Wagner, R. A., & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the ACM (JACM), 21*(1), 168-173.
- Whetsel, B. C. (2017). Enhancing License Plate Matching Algorithm: A New Evaluation Measurement and Threshold Determination.

APPENDIX

Appendix A. A Short Description of the License Plate Matching Algorithm

License plate matching techniques are increasingly being used in data collection and traffic studies. In general, license plate matching techniques consist of collecting vehicle license plate numbers and arrival times at various checkpoints, matching the license plates between consecutive checkpoints, and computing travel times from the difference in arrival times. There are several techniques developed to conduct license plate matching, and considerable work has been done to improve the matching process (e.g. matching rate). In 2009, Dr. Han's team, developed algorithms to improve the matching process by incorporating the use of edit distance (ED) and travel time thresholds. Edit distance is a technique aiming to measure how close two strings (sequences of characters) are from each other based on weight functions (which can be unitary values or based on statistical data) to compare each individual pair of characters. It was first developed by Wagner and Fischer (1974), the edit distance $d(x \rightarrow y)$ between two strings x and y, can be calculated based on the following recurrent equation:

$$d(x \rightarrow y) = \min \begin{cases} d(i-1, j-1) \; + \; \gamma(x_i \rightarrow y_j) & \textit{substitution} \\ d(i-1, j) + \gamma(x_i \rightarrow \varepsilon) & \textit{deletion} \\ d(i, j-1) + \gamma(\varepsilon \rightarrow y_j) & \textit{insertion} \end{cases}$$

Where $d(i, j)$ is the edit distance between substrings $x[1, \dots, i]$ and $y[1, \dots, j]$, of x and y, respectively, and $d(0,0) = 0$. The $\boldsymbol{\gamma}$ are the weight functions. For example, $\gamma(x_i \rightarrow y_j)$ is the cost for the change (substitution) from x_i to y_j. The cost of $\gamma(x_i \rightarrow \varepsilon)$, where ε represents the empty character, is incurred by a deletion of x_i, and the cost of $\gamma(\varepsilon \rightarrow y_j)$ is incurred from an insertion of y_j.

Later in 2012, Oliveira-Neto, Han, and Jeong (2012) increased the matching rate to 97% by incorporating the use of a newly proposed weight function in edit distance (ED) and travel time thresholds. Basically, they devised a new weight function (association matrix) where each cell represents the likelihood of certain pair-wise character symbol occurrence. For example, there is a relatively high chance of certain characters (e.g. "1," "0," and "B") being misread (e.g. "I," "O," and "8," respectively) by the LPR machine. More details on this were presented in chapter 4.

In addition, previous matching techniques always required ground truth data, which requires much more work. To overcome this challenge, Dr. Han's team proposed a self-learning algorithm to avoid manually extracting ground truth from images. The self-learning algorithm is simple and straight-forward. Initially, it started with a blank association matrix. The algorithm analyzes each new plate string to continuously improve the matrix. Each new iteration of the matrix is compared to the previous one to see if there would be an increase in performance. Once the matching performance cannot be improved, the algorithm stops, and the final association matrix is ready. The pseudocode of the algorithm is shown below.

1. Initialize a blank matrix $\boldsymbol{C_0}$.
2. Match the license plates with $\boldsymbol{C_k}$, and the result is $\boldsymbol{M_k}$.
3. Get the association matrix $\boldsymbol{C_{k+1}}$ based on $\boldsymbol{M_k}$.
4. Stop if the association matrix converges ($\|\boldsymbol{C_k} - \boldsymbol{C_{k+1}}\| < \varepsilon$).

A.1 Association Matrix

The Edit Distance and probability method are both based on the association matrix. Previously, the association matrix has been derived from the LPR data and ground truth data. For an association matrix $\boldsymbol{C}$ in two LPR stations (g, h), there is a square matrix whose elements are the conditional probabilities $p(b|a)$ of observing a character reading b in station h for a given character reading a, in station g. The conditional probability is calculated from the two confusion matrices in any two LPR stations. The element of the confusion matrix is the odds of the LPR machines reading (diagonal elements) or misreading characters (off-diagonal elements). However, the estimation of the truth matrix requires extensive manual extraction by visual inspection of the ground truths of many plate images. To overcome this, a self-learning algorithm was developed to get the association matrix. It starts with a blank association matrix, and iteratively learns from the matched license plates in two LPR stations. By the end, when the association matrix converges, then the association matrix is ready.

A.2 Edit Distance

The probabilities for each character are acquired from the association matrix and are then used in the equation below to calculate the weighted ED for each plate string. The most likely match or sequence of editing operations is given by the sequence with the highest probability, which implies the minimization of the negative natural logarithm, as follows (Oliveira-Neto et al., 2013):

$$d(x \rightarrow y) = \min\{\sum_{k=0}^{n} \log(\frac{1}{p(i_k, j_k)})\}$$

Where n is the total number of characters in the plate string, the $p(i_k, j_k)$ is the condition probability of the character pair (i_k, j_k). To find the match, we calculated every possible match of two LPR stations within a travel time window. Then, the plate match with the smallest ED is selected as a potential match and is checked to see if it falls below the ED threshold.

A.3 Probability Method

The probability method is used to calculate the likelihood that two license plates are a match, which is expressed as a percentage, is more easily understood than ED and can be explained in a matter of minutes to almost anyone. To calculate the likelihood, we also need to use the association matrix. First, we compute the probability that two plate strings are a match. The equation is shown below:

$$\mathrm{P}_{E|Match} = \prod_{k=1}^{n} p(i_k, j_k)$$

Where $p(i_k, j_k)$ is the probability of having a character i_k at station g and j_k at station h, can be obtained from the association matrix, n is the number of characters on the license plate.

For a case of non-match, the probability of having a character i_k at station g will not be correlated with the character j_k at station h. Then the probability of having i_k at station g is affected by the distribution of characters in population. Since there are 37 possible characters (A to Z, 0 to 9, and a null character), the probability of having i_k at station g is calculated simply by $1/37$. Therefore, the probability of the event with the condition of a non-match, $\mathrm{P}_{E|Non-Match}$, can be calculated as follows:

$$\mathrm{P}_{E|Non-Match} = \prod_{k=1}^{n} p(i_k, j_k) = \frac{1}{37}^{n}$$

Now, using Bayes' theorem, we can reformulate the equation to calculate the probability of matching with the given pair of conditions (Whetsel, 2017).

$$\mathrm{P}_{Match|E} = \frac{\mathrm{P}_{E|Match} * P_{Match}}{\mathrm{P}_{E|Match} * P_{Match} + \mathrm{P}_{E|Non-Match} * P_{Non-Match}}$$

Here, the P_{Match} is the probability of matching without considering the given pair or event, which is equivalent to the probability of capturing the same vehicle at station g, when the vehicle is captured at station h.

$$P_{Match} = \frac{Number\ of\ vehicels\ captured\ at\ Station\ g\ and\ h}{Number\ of\ vehicles\ captured\ at\ station\ h}$$

Last, the values are ranked: the match with the highest probability is selected as the potential match.

Appendix B. Field Study ALPR Equipment Setup

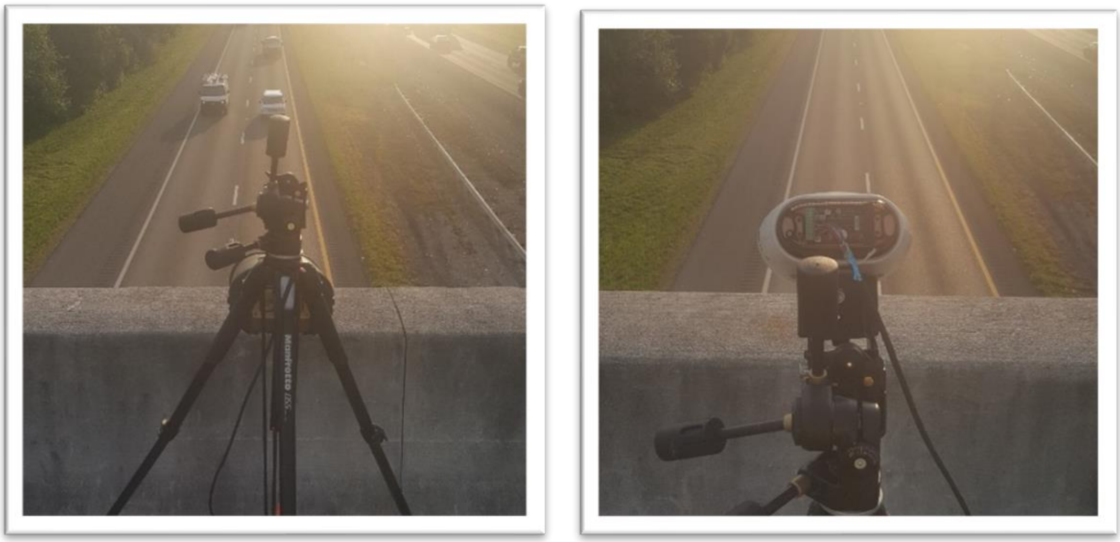


Figure B-1 LPR device setup

Figure B-2 On-site LPR data collection in April 8-9, 2017

Figure B-3 On-site LPR data collection in May 8, 2017

Figure B-4 Sample photos of captured trucks

Appendix C. Sample Weigh Station ALPR Data

We obtained the weigh station ALPR data from TDSHS. The data is a nearly 3-month (November 2016 to February 2017) license plates data for nine weigh stations, which are:

- Coffee Co Manchester I-24 EW MM:115 (x2)
- Giles Co I-65 N MM:5
- Greene Co I-81 S MM:21
- Haywood Co I-40 EW MM:50 (x2)
- Knox Co I-40 W no data
- Knox Co I-40 E
- Unicoi Co I-26

The data contains a lot of information, like the license plate, time, license jurisdiction, lane ID, vehicle ID, and location. This information can be used to track trucks and explore freight mobility patterns. While several stations do not have the associated license plates data, which may due to the malfunction or other reasons. The table below is the number of license plates captured for each weigh station. Also, a sample weigh station ALPR data is given.

Table C-1 Number of license plates captured at each weigh station

Weigh Station	License Plates Number	Weigh Station	License Plates Number
Coffee EB	268, 361	Coffee WB	165, 561
Haywood EB	134, 319	Haywood WB	132, 832
Knox EB	215, 689	Knox WB	0
Giles NB	123, 084	Unicoi	31, 449
Greene	0		

Location	VehicleId	laneId	DateTime	LicenseJui	LicensePlate
Coffee EB	1163867	12	2016-11-23 20:28:40		1069844
Coffee EB	1163868	12	2016-11-23 20:28:45	USIN	2140462
Coffee EB	1163869	12	2016-11-23 20:28:49	USMN	PAN3095
Coffee EB	1163882	12	2016-11-23 20:29:34	USMT	PA11672
Coffee EB	1163885	12	2016-11-23 20:29:42	USIN	2247633
Coffee EB	1163891	12	2016-11-23 20:29:53		No Plate Detected
Coffee EB	1163895	12	2016-11-23 20:30:00	USTX	R25062
Coffee EB	1163899	12	2016-11-23 20:30:08	USMO	10AR7H
Coffee EB	1163902	12	2016-11-23 20:30:14	USIN	2285277
Coffee EB	1163905	12	2016-11-23 20:30:23	USIN	2417212
Coffee EB	1163926	12	2016-11-23 20:31:18	USTN	D0736HY
Coffee EB	1163929	12	2016-11-23 20:31:28		4184L14
Coffee EB	1163931	12	2016-11-23 20:31:36	USTN	D0407HY
Coffee EB	1163932	12	2016-11-23 20:31:39	USGA	C325AE
Coffee EB	1163935	12	2016-11-23 20:31:56		I1075
Coffee EB	1163938	12	2016-11-23 20:32:09	USGA	C445AH
Coffee EB	1163943	12	2016-11-23 20:32:43	USTN	E0294HY
Coffee EB	1163950	12	2016-11-23 20:33:22		1072504
Coffee EB	1163954	12	2016-11-23 20:33:38	USNE	C13806
Coffee EB	1163958	12	2016-11-23 20:33:58		FR1B77T
Coffee EB	1163969	12	2016-11-23 20:34:26	USIN	2292882
Coffee EB	1163978	12	2016-11-23 20:34:54		147523
Coffee EB	1163986	12	2016-11-23 20:35:31		No Plate Detected
Coffee EB	1163992	12	2016-11-23 20:35:44	USIN	951879
Coffee EB	1163996	12	2016-11-23 20:35:56	USTN	D0415HY
Coffee EB	1164007	12	2016-11-23 20:36:30	USIN	2333649
Coffee EB	1164013	12	2016-11-23 20:36:47	USTN	D4241HY
Coffee EB	1164020	12	2016-11-23 20:37:05	USMO	39AR6J
Coffee EB	1164023	12	2016-11-23 20:37:12	USIN	1065928
Coffee EB	1164030	12	2016-11-23 20:37:35	USIN	996680
Coffee EB	1164032	12	2016-11-23 20:37:40	USIN	995231
Coffee EB	1164033	12	2016-11-23 20:37:43	USIN	994236
Coffee EB	1164045	12	2016-11-23 20:38:19	USTN	C6673HY
Coffee EB	1164048	12	2016-11-23 20:38:39	USIN	2273641
Coffee EB	1164049	12	2016-11-23 20:38:53	USTN	G0116HY
Coffee EB	1164050	12	2016-11-23 20:38:54	USIN	1171128

Figure C-1 Sample of weigh station license plate data

Appendix D. Sample ALPR Matching Results

	A	B	C	D	E	F	G	H	I	J	K	L
1	Index_g	Index_h	Time_g	Time_h	GED	Probabilit	TRUE	Exact	Travel Time	Mean	STDV	
2	71	257	45.1	194.9167	44.06599	0.942015	0	0	149.8166667	151.755	11.49289	
3	43	259	35.15	195.85	46.69404	0.974284	0	0	160.7	152.9883	10.4011	
4	76	261	46.31667	197.3667	33.83028	0.933302	0	0	151.05	154.76	10.00903	
5	50	262	37.28333	197.45	31.10371	0.99748	0	0	160.1666667	156.0983	8.200931	
6	17	263	24.53333	198.9833	14.76001	0.996874	0	0	174.45	157.6983	7.101519	
7	116	264	57.6	199.25	36.17999	0.985618	0	0	141.65	158.705	8.690221	
8	99	265	52.56667	199.75	30.77284	0.998757	0	0	147.1833333	157.375	10.21308	
9	33	266	31.05	199.9833	51.64554	0.923384	0	0	168.9333333	156.245	10.69079	
10	18	267	24.65	200.0167	39.79699	0.994432	0	0	175.3666667	158.5217	10.73015	
11	5	270	17.45	200.8167	12.70813	0.999936	0	0	183.3666667	159.8167	11.96308	
12	139	272	64.06667	203.5167	30.39824	0.99936	0	0	139.45	161.2683	13.9043	
13	123	273	60.16667	203.9167	28.20895	0.995228	0	0	143.75	160.2317	15.18083	
14	83	277	48.28333	208.6833	26.04327	0.998627	0	0	160.4	158.5367	16.04443	
15	54	278	38.2	209.5167	38.3041	0.936412	0	0	171.3166667	159.4717	15.83068	
16	88	280	49.36667	210.1667	45.73939	0.993186	0	0	160.8	160.5867	16.27159	
17	126	285	61.38333	211.95	22.24704	0.945391	0	0	150.5666667	159.2217	15.53528	
18	295	286	119.65	311.5833	16.22991	0.993532	0	0	191.9333333	160.1133	14.64505	
19	206	287	84.05	312.9167	30.25409	0.99893	0	0	228.8666667	164.5883	16.91603	
20	183	288	76.23333	314.1333	20.35266	0.996487	0	0	237.9	170.5817	26.51833	
21	405	289	172.75	315.2333	29.46749	0.995871	0	0	142.4833333	176.835	34.06989	
22	395	290	167.7	315.3167	14.38908	0.004631	0	0	147.6166667	172.7467	35.61686	
23	169	291	71.36667	315.4333	42.37497	0.997138	0	0	244.0666667	173.5633	34.85402	
24	102	292	53.01667	315.5667	37.96117	0.999513	0	0	262.55	183.595	39.45288	
25	6	293	19.11667	315.8833	21.82415	0.999587	0	0	296.7666667	193.81	45.53532	
26	13	294	23.1	316.2667	17.20488	0.999443	0	0	293.1666667	206.355	54.95609	
27	145	299	65.56667	318.1167	21.7806	0.930841	0	0	252.55	219.5917	58.5856	
28	195	300	80.41667	318.4833	44.88258	0.914389	0	0	238.0666667	229.79	53.92608	
29	34	304	32.7	318.9	37.86603	0.999564	0	0	286.2	234.4033	52.27571	
30	362	308	146.8167	321.5833	14.81672	0.999359	0	0	174.7666667	240.1367	54.6893	
31	191	309	79.68333	321.6667	29.59436	0.997444	0	0	241.9833333	233.8233	58.48829	
32	338	310	136.4167	322.1	54.19746	0.939545	0	0	185.6833333	243.7733	48.90068	
33	289	314	117.5333	322.4667	50.39097	0.948784	0	0	204.9333333	247.58	41.50626	
34	233	315	94.28333	322.9333	10.43142	0.997402	0	0	228.65	243.6667	43.66308	
35	166	316	70.66667	323.4333	7.285429	0.999516	0	0	252.7666667	240.2767	43.34894	
36	100	317	52.85	323.5833	9.942994	0.999324	0	0	270.7333333	235.8767	38.99211	
37	345	318	138	323.6333	26.92503	0.989138	0	0	185.6333333	233.6333	35.84841	

Figure D-1 Sample of results of LPR matching

Appendix E. Literature Review of Truck Tracking

A number of technologies and practices for vehicle/truck tracking were identified in a literature review conducted by the University of Tennessee.

Inductive Loop Detector (ILD)

Inductive Loop detector (ILD) is the most prevalent surveillance system for traffic systems. ILD can not only generate inductive signature data, but it can also estimate vehicle speed and vehicle length. Other useful information might be the time difference between two proposed matching signatures at two distinct ILD stations. Sun et al. (1999) provides one of the first vehicle reidentification study using ILD data. It first extracts a multi-dimensional vector, concatenating all features, then trying to find closest upstream feature given any downstream features. A multi-objective optimization problem is formulated, and a Pareto optimal solution is given by narrowing down searching space to consider one constraint at a time by lexicographic order (objective goals indexed by importance). The target of that study aimed to find the traffic performance between two ILD stations and get sectional traffic measures like sectional travel speed. Data from two loop stations on SR-24 freeway in Lafayette, 1.2 miles apart, were collected. One moderate traffic dataset and one congested traffic dataset were collected with video recordings as ground truth data. The results showed that the sectional traffic estimates were close to ground truth data, but the method of interpolating point measures of speed and densities could not represent sectional traffic data.

Another pioneering study by Abdulhai et al. (2003) further discussed machine learning techniques' ability in improving this vehicle re-identification problem at successive ILD stations. Distance measure that is insensitive to affine distortion/normalization of spatiotemporal patterns were defined. A (back-propagated) neural network was trained using a genetic algorithm to measure the distances, thus making better matches compared to traditional distance measures. For each ILD waveform, the distance between that waveform and a group of 10, 20 or 30 upstream vehicles were calculated using neural networks, conventional distance measures and spatiotemporal measures for both the moderate traffic case and congested traffic case.

1. Sun C, Ritchie SG, Tsai K, Jayakrishnan R. Use of vehicle signature analysis and lexicographic optimization for vehicle reidentification on freeways. Transportation Research Part C: Emerging Technologies. 1999 Aug 1;7(4):167-85.
2. Abdulhai B, Tabib SM. Spatio-temporal inductance-pattern recognition for vehicle re-identification. Transportation Research Part C: Emerging Technologies. 2003 Jun 1;11(3-4):223-39.

WIM & Loop Detectors

Weigh-in-motion (WIM) stations are data collection points to collect heavy vehicle information (vehicle's presence, the weight of axles, axles spacing, speed, etc.) using a set of different sensors (inductive loop detector, piezoelectric sensor, bending plate sensor, axle sensor, etc.). The sensors are embedded on the highway, so the data is collected while one is moving along. There are only around 800 WIM station in the United States, while there are much more inductive loop detector infrastructures. Jeng et al (2014) provided a method to track heavy vehicles using both WIM data and inductive loop signature data, in case the WIM data is not present or the sensor is a different model. A framework was developed and consisted of two different models: a classification model and a reidentification model. For each WIM station, the classification model tries to classify the truck classes using information like axle number, spacing, weigh, etc. The reidentification model tries to reidentify by matching the record of the same truck from the upstream station, using a set of algorithms like decision tree, wavelet-k nearest neighbor, etc. Based on the availability of data sources (availability of vehicle detection station (VDS) or WIM), a different comparison algorithm was designed to match vehicles. "Travel time window" was also used in the algorithm to keep reasonable matchings. Notice that the study only matches the vehicles reaching the downstream station in reasonable time, only considering if they never left the highway between stations. In other words, for those vehicles go off for loading/unloading or other activities, they would not be paired in the case. The "travel time window" was estimated using Caltrans performance measuring system (PeMS). WIM data and loop data can also be useful in distinguishing those vehicles as these attributes may change when vehicle's loading status or tractor is changed. A case study using the data from two WIM stations 19 miles apart was also studied. The ground truth matching rate was estimated by manually checking vehicle plate information. Compared with using only vehicle signature data, when using both vehicle signature data and WIM data, the matching performance improved significantly based on the defined reidentification performance metrics. Four indexes were proposed to measure the performance of the matching algorithm in matching tasks: 1) Ideal Match Rate; 2) Correct Match Rate; 3) Over Match Rate; 4) Error Rate. Notice that only a small portion of trucks upstream could travel to downstream or vice versa.

The same idea of using WIM and inductive loop detectors was further investigated by Hyun et al. (2017). With the aim of estimating flows of long-haul trucks across a region, a selective weighted Bayesian model was proposed. The model tried to explore the marginal distribution of each data feature in discerning ability between matched pairs and non-matched pairs (the labels are paired by manually identifying plate information). The features with a higher information gain (i.e., discerning abilities) were assigned with higher weights. The distribution traits were classified in terms of distinguished vehicles and a joint probability model under the (weighted) naïve Bayesian model. This was used to predict if two observations were a match or not. Two WIM stations along I-5, 26 miles apart, were used for developing the model, with vehicles manually labeled using plate information. The dataset was separated into a training set and a testing set to develop and evaluate the model. Two WIM stations along SR-99, 65 miles apart, were used for the case study applying the trained model to see its usability for long distance traveling.

1. Jeng ST, Chu L. Tracking heavy vehicles based on weigh-in-motion and inductive loop signature technologies. IEEE Transactions on Intelligent Transportation Systems. 2014 Jul 17;16(2):632-41.
2. Hyun KK, Tok A, Ritchie SG. Long distance truck tracking from advanced point detectors using a selective weighted Bayesian model. Transportation Research Part C: Emerging Technologies. 2017 Sep 1;82:24-42.

GPS Data

The above papers mainly focus on studying the vehicle re-identification between successive point measurements at data collecting stations. GPS data, on the other hand, can track a vehicle constantly, which provides the opportunity to look into the behavior and operation details of trucks. Thakur et al. (2015) developed a procedural algorithm to identify the trip segments and stop points. Furthermore, the algorithm tried to distinguish the stops for logistics operations from other purposes including refueling, congestion, resting, etc. GIS polygons of major truck stops were used in the algorithm. 145 million GPS records from the American Transportation Research Institute (ATRI) with origin/destination (OD) in Florida over 4 months were used in the study. The procedures and implementation claimed to be useful for converting raw GPS records to useful OD truck information, considering the data scale and the state-wide geometry scale. Zanjani et al. (2015) further estimated the OD matrix of traffic within, into, and out of Florida state.

1. Thakur A, Pinjari AR, Zanjani AB, Short J, Mysore V, Tabatabaee SF. Development of algorithms to convert large streams of truck GPS data into truck trips. Transportation Research Record. 2015 Jan;2529(1):66-73.
2. Zanjani AB, Pinjari AR, Kamali M, Thakur A, Short J, Mysore V, Tabatabaee SF. Estimation of statewide origin–destination truck flows from large streams of GPS data: Application for florida statewide model. Transportation Research Record. 2015 Jan;2494(1):87-96.

RFID

Radio-frequency identification (RFID) is an automatic identification technology, which is widely applied to many industries. RFID is widely used in developing countries for transportation management. In highway systems, open road tolling (ORT) systems are also widely implemented using RFID technologies. Wang et al. (2010) proposed a way to use RFID systems to monitor highway running vehicles and identify speeding vehicles. Ren et al. (2009) discussed the design details of the tollway usage of RFID system. According to Chinese state news media Xinhua News, by December 23rd in 2019, there were 197 million electronic tollway collection system (ETC) users . The number had increased by over 102 million in just one year.

Digital license plates with RFID imbedded is another practice in transportation. Abd Rahman et al. (2013) and Liang et al. (2017) designed a license plate with an RFID tag integrated. According to media, in the city of Shenzhen, China, 8 types of vehicles, (including school buses, heavy duty vehicles, etc.) were required to install RFID tags for better management.

1. Hongjian W, Yuelin T, Zhi L. RFID technology applied in highway traffic management. In2010 International Conference on Optoelectronics and Image Processing 2010 Nov 11 (Vol. 2, pp. 348-351). IEEE.
2. Zhengang R, Yingbo G. Design of electronic toll collection system in expressway based on RFID. In2009 International conference on environmental science and information application technology 2009 Jul 4 (Vol. 3, pp. 779-782). IEEE.
3. Abd Rahman T, Rahim SK. RFID vehicle plate number (e-plate) for tracking and management system. In2013 International Conference on Parallel and Distributed Systems 2013 Dec 15 (pp. 611-616). IEEE.
4. Liang Z, Ouyang J, Yang F, Zhou L. Design of license plate RFID tag antenna using characteristic mode pattern synthesis. IEEE Transactions on Antennas and Propagation. 2017 Jul 31;65(10):4964-70.

Image and License Plate Matching

Surveillance video information is very useful in tracking vehicles. Jelaca et al. (2013) designed a method to track vehicles. To quickly matching vehicles, a set of signatures were extracted from the images by project profiles vertically, horizontally, diagonally, etc. If one image was of low quality, a tuple of images was projected together to form one signature. Tunnel surveillance video recordings were used to match vehicles. The project was done before the explosion of deep learning technologies, although many ideas can perform comparably to deep learning. The algorithm is claimed to work online.

Oliveira-Neto et al. (2012) proposed an online license plate matching procedure using license plate recognition technology. Under a two-point LPR survey, license plate information might be misread at one or both stations. A character-transition matrix was used to measure the probability of a true character given predicted ones. With both stations having their own transition matrix, the probability of an upstream station's identified character given its downstream character can be estimated. Furthermore, a generalized edit-distance (GED) was used for describing the information gain/loss in a better way. Time passage information was was considered between stations to speed up the searching process and make the algorithm more efficient.

1. Oliveira-Neto FM, Han LD, Jeong MK. An online self-learning algorithm for license plate matching. IEEE Transactions on Intelligent Transportation Systems. 2013 Aug 2;14(4):1806-16.
2. Hyun KK, Tok A, Ritchie SG. Long distance truck tracking from advanced point detectors using a selective weighted Bayesian model. Transportation Research Part C: Emerging Technologies. 2017 Sep 1;82:24-42.
3. Jelača V, Pižurica A, Niño-Castañeda JO, FríAs-VeláZquez A, Philips W. Vehicle matching in smart camera networks using image projection profiles at multiple instances. Image and Vision Computing. 2013 Sep 1;31(9):673-85.

Deep Learning (Vehicle Re-Identification)

With the increase of video surveillance data and the fast development of computer vision technology using deep learning technology, a large number of researchers are proposing new deep learning methods to the vehicle reidentification problem. Liu et al. (2013) proposed a convolutional neural network (CNN). Unlike vehicle detection, tracking and classification problem, vehicle re-identification (Re-Id) can be found as near duplicate image retrial (NDIR) problem. Two contributions are made for this project: 1) a new dataset called VehicleReID (VeRi) "containing 40,000 images and 619 vehicles captured by 20 cameras covering a 1.0 km^2 area in 24 hours". The data was labeled manually including drawing boundary boxes (also abbreviated BBoxes), adding text information of color, vehicle type and brands. Moreover, all 619 vehicles in the dataset were captured by 2-18 cameras, which gave many true matchings samples for the Re-Id task. 2) The research group compared the performance of a set of methods, including texture based feature (BOW-SIFT), Color based feature (BOW-CN), semantic feature extracted by deep neural network (AlexNet, GoogleLeNet), feature fusion (AlexNet+BOW-CN), Fusion of Attributes and Color features (FACT)). The results showed that while BOW-SIFT performs badly, all other methods have a similarly good performance. FACT, summing the weights of rank scores of each feature, unsurprisingly, performs the best. The team proposed using FACT for coarse vehicle identification and then using vehicle plate recognition to further improve the results.

Liu et al. (2016) further developed their own specified neural network for vehicle Re-Id problem named PROgressive Vehicle re-ID (PROVID), using on the same VeRi dataset proposed before. There were three major steps for the developed method:

1) an appearance-based searching step; 2) a license plate fine match; and 3) a spatiotemporal re-rank step. The model borrows two useful principles in the searching process: coarser-to-finer principle and near-to-distance search. The first step is to identify vehicles sharing similar features (e.g., color, textile, etc.) to narrow down the searching scope down to a few potential candidates. After narrowing down the searching scope, the second step is to look into license plate recognition to compare plates for matching. For the license plate recognition, the group exploited the Siamese Neural Network for license plate verification. Thirdly, based on the spatiotemporal distribution of vehicles, a spatiotemporal proximity index was designed to describe similarity between any proposed matching images. To combine the spatiotemporal information, either a post-fusion strategy or re-ranking strategy can be employed.

Liu et al. (2018) proposed a method of identifying local regions by identifying unique decorations, stickers in the front windshield. The Region-Aware deep Model (RAM) method achieves a significantly better performance compared to many other methods.

More developments in deep learning have been made in recent years. Huang et al. (2019) proposed a vehicle reidentification model using a temporal attention model and matadata-reranking techniques. The group collected some video clips for training the re-ID task. A frame-level feature (one frame from a video clip) was extracted by using the key point identification method and identifying 36 key points on the vehicle. Then, these points were used to estimate the orientation of vehicles by relationship of these points in a 2D image, and this information was mapped into 18-dimensional feature outputs. This information can help us to tell whether the frame's viewpoint is back view, front view, side view, etc. With this information, the group developed a viewpoint-aware temporal attention model. Each video clips contained a set of video frames. For each video frame in a video clip, both viewpoint features and CNN features (for appearance) were generated and fed to the temporal attention-based method for frames in the video clip. Thus, the model learns more about one vehicle observed at different angles. The re-rank strategy was applied by using metadata information. The metadata included type, brand and color. It ranked 2nd in the AI city challenge 2019 dataset achieving mean average precision (mAP) of 79.17%.

1. Liu X, Liu W, Ma H, Fu H. Large-scale vehicle re-identification in urban surveillance videos. In2016 IEEE International Conference on Multimedia and Expo (ICME) 2016 Jul 11 (pp. 1-6). IEEE.
2. Liu X, Liu W, Mei T, Ma H. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In European Conference on Computer Vision 2016 Oct 8 (pp. 869-884). Springer, Cham.

3. Liu X, Zhang S, Huang Q, Gao W. Ram: a region-aware deep model for vehicle re-identification. In2018 IEEE International Conference on Multimedia and Expo (ICME) 2018 Jul 23 (pp. 1-6). IEEE.
4. Zhou Y, Shao L. Aware attentive multi-view inference for vehicle re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 6489-6498).
5. Huang TW, Cai J, Yang H, Hsu HM, Hwang JN. Multi-View Vehicle Re-Identification using Temporal Attention Model and Metadata Re-ranking. In CVPR Workshops 2019 Jun 16 (Vol. 2).

Appendix F. Scholarly Papers by UTK on LPR Matching Methodology

Transportation Research Part C 21 (2012) 306–320

Contents lists available at SciVerse ScienceDirect

Transportation Research Part C

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

Online license plate matching procedures using license-plate recognition machines and new weighted edit distance

Francisco Moraes Oliveira-Neto [a], Lee D. Han [a,*], Myong K. Jeong [b]

[a] Department of Civil and Environmental Engineering, The University of Tennessee, Perkins Hall 223, Knoxville, TN 37996, USA
[b] Department of Industrial and Systems Engineering and RUTCOR, Rutgers University, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:
Received 15 November 2010
Received in revised form 14 November 2011
Accepted 14 November 2011

Keywords:
Vehicle tracking
License-plate recognition
Edit distance
Text mining
Online plate matching
Two-point plate survey

ABSTRACT

License-plate recognition (LPR) technology has been widely applied in many different transportation applications such as enforcement, vehicle monitoring, and access control. Recently, there has been effort to exploit an LPR database for vehicle tracking using popular template matching procedures. Existing template matching procedures assume that the true reference string is always available. However, under a two-point LPR survey, a vehicle could have its plate misread at both locations generating a pair of misread strings (or templates) with no reference for matching. To compensate for LPR misreading problem, we propose a new weight function based on a probability model to match the observed outcomes of a dual LPR setup. Also, considering that reversal errors are never made in LPR machines, new editing constraints as a function of the string lengths are proposed to avoid compensation for reversal errors. These editing constraints are incorporated into the constraint edit distance formulation to improve the performance of the matching procedure. Finally, considering that previous template matching procedures do not take advantage of passage time information available in LPR databases, we present an online tracking procedure that considers the properties of probability distribution of vehicle journey times in order to increase the probability of correct matches. Experimental results show that our proposed procedure can improve the accuracy of LPR systems and achieve up to 97% of positive matches with no false matches. Further research is needed to extend the ideas proposed herein to plate-matching with multiple, i.e., more than two, LPR units.

Published by Elsevier Ltd.

1. Introduction

Among intelligent transportation systems, automated vehicle identification (AVI) is a powerful tool for electronic toll, traffic management, commercial vehicle operations, motor vehicle law enforcement, origin–destination (O–D) survey, and access control, among other applications (Nelson, 2000a). All these applications require a unique identification of a vehicle at a checkpoint; some of them require the vehicle to be tracked at multiple points, e.g., O–D survey. To this end, it would be desirable if each vehicle is required to carry a radio-frequency identification (RFID) tag, or some hardware of similar mechanism, to provide uniquely identifiable and useful vehicle and, perhaps even, driver information. This idea, however, has experienced much resistance and is unlikely to be implemented in the short term due to a number of reasons with privacy concerns being the most prohibitive.

A more traditional and feasible approach is to identify vehicles by their license plates, which all vehicles are required to have, using automated license-plate recognition (LPR) systems. These systems were developed with the main objective of

* Corresponding author.
E-mail address: lhan@utk.edu (L.D. Han).

0968-090X/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.trc.2011.11.003

An Online Self-Learning Algorithm for License Plate Matching

Francisco Moraes Oliveira-Neto, Lee D. Han, and Myong Kee Jeong, *Senior Member, IEEE*

***Abstract*—License plate recognition (LPR) technology is a mature yet imperfect technology used for automated toll collection and speed enforcement. The portion of license plates that can be correctly recognized and matched at two separate stations is typically in the range of 35% or less. Existing methods for improving the matching of plates recognized by LPR units rely on intensive manual data reduction, such that the misread plates are manually entered into the system. Recently, an advanced matching technique that combines Bayesian probability and Levenshtein text-mining techniques was proposed to increase the accuracy of automated license plate matching. The key component of this method is what we called *the association matrix*, which contains the conditional probabilities of observing one character at one station for a given observed character at another station. However, the estimation of the association matrix relies on the manually extracted ground truth of a large number of plates, which is a cumbersome and tedious process. To overcome this drawback, in this study, we propose an ingenious novel self-learning algorithm that eliminates the need for extracting ground truth manually. These automatically learned association matrices are found to perform well in the correctness in plate matching, in comparison with those generated from the painstaking manual method. Furthermore, these automatically learned association matrices outperform their manual counterparts in reducing false matching rates. The automatic self-learning method is also cheaper and easier to implement and continues to improve and correct itself over time.**

***Index Terms*—Edit distance (ED), license plate recognition (LPR), text mining, vehicle tracking.**

I. Introduction

LICENSE plate recognition (LPR) technology is a mature yet imperfect technology used for automated toll collection and speed enforcement. Because of LPR's limited accuracy, around or less than 60%, depending on the model, installation, variation of the license plates in the traffic stream, and other factors, the portion of license plates that was correctly recognized and matched at two separate stations was typically in the range of 35% or less. This number further deteriorates if one tries to match the same plate through more than two sequential stations. Existing techniques for improving plate matching accuracy consist of intensive manual data reduction and posterior training [1], [2].

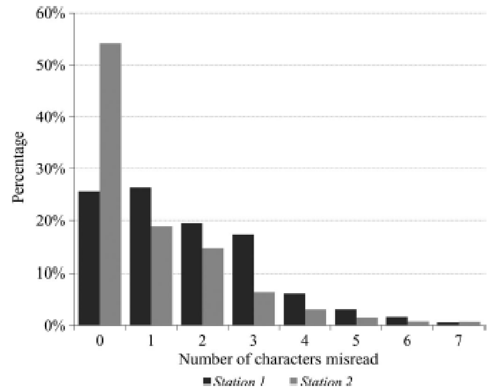


Fig. 1. Sample LPR character-reading error rates per plate for two units installed at two different locations.

Through the study of the characteristics of the errors made by LPR hardware, we can find out that, while significant portion of plates were recognized incorrectly, many of these misread plates only had one, two, or three misread characters out of the entire string of six or so total characters, as shown in Fig. 1. In other words, the correct recognition rate of individual characters is much higher than the correct recognition rate of entire plates. This is a simple yet powerful fact unexplored by hardware manufacturers and researchers in the area of LPR technology and video image processing in general. Our idea takes advantage of this simple fact and explores the likelihood that two seemingly different license plate strings (sequence of alphanumeric characters) resultant from two LPR stations are actually a match.

It should be pointed out that the license plate matching is far more challenging than a traditional template matching problem because license plate strings typically 1) do not have a readily available dictionary to compare to, 2) do not have a context to help "guess" the meaning of the plate, and 3) include both

Manuscript received September 28, 2012; revised March 23, 2013; accepted June 3, 2013. Date of publication August 1, 2013; date of current version November 26, 2013. This work was supported in part by the National Transportation Research Center, Inc. through the U.S. Department of Transportation (USDOT) Research and Innovative Technology Administration under Grant DTRT-06-0043-U09 and in part by the USDOT Federal Highway Administration through the Dwight David Eisenhower Graduate Scholarship Program under Grant DDEGRD-09-X-00407. The Associate Editor for this paper was Prof. S. Sun.

F. M. Oliveira-Neto is with the Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN 37996 USA, and also with the Center for Transportation Analysis, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA.

L. D. Han is with the Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN 37996 USA, and also with the School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410004, China.

M. K. Jeong is with the Department of Industrial and Systems Engineering and the Rutgers Center for Operations Research (RUTCOR), Rutgers University, Piscataway, NJ 08854 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2013.2270107

Expanding License Plate Matching Capabilities with Secondary Self-Learning Algorithm

Stephanie R. Hargrove, Hyeonsup Lim, and Lee D. Han

To perform the postprocessing matching of license plates between two license plate recognition (LPR) stations, a self-learning matching algorithm was employed. The key component of this algorithm is an association matrix that is a unique translator, associating two LPR units in relation to how they may see or recognize the same characters differently, for a host of reasons. This association matrix consists primarily of high-confidence matches between two LPR stations estimated directly from a set of matched character pairs. The matching algorithm's performance decreases as the distance between the two LPR stations increases because of vehicles no longer traveling within an average travel time window, a low sample of vehicles traveling between the two LPR stations, or both. This paper proposes using a third LPR station to generate additional information to derive a better association matrix for an existing pair of LPR stations and thus replaces the existing learned-association matrix. In other words, the added LPR unit facilitates secondary or transferred learning to improve the matching performance of the first two units, even after the third LPR unit is subsequently removed. To evaluate this derived association matrix, the authors employed two simulations. They were to determine when the newly derived matrix should be used and to evaluate the overall performance of license plate matching.

License plate recognition (LPR) technology has been widely applied to numerous transportation applications, including automated speed and law enforcement, vehicle tracking, and automated highway tolling. These applications require LPR technology to match a license plate at two or more locations. To do so without additional postprocessing, manual labor particularly, each license plate string (sequence of characters) must be identified correctly to declare a match. If just one character is misread, then a match cannot be declared without further efforts and delay, which typically involve human intervention.

LPR technology uses optical character recognition (OCR) engines to identify the text strings of license plates. The matching of OCR-recognized license plates is far more complicated than the matching of traditional OCR text, such as text from books. The matching of traditional OCR text has the benefit of readily available dictionaries containing a finite number of vocabulary or strings, context to help determine the likely meaning of a word, and a standard syntax for all characters. In contrast, license plate strings almost never have meanings and multiple potential syntaxes; one cannot even be certain a plate string was recognized correctly by the LPR algorithm without manual verification. However, not all is lost, because the process of matching two license plate strings can go beyond looking at the string as a whole and instead using the sequence of comparisons of individual characters.

Consider, for example, that license plate strings X and Y are read at two LPR stations, with the results being X = ABC123 and Y = A8C1Z3. By comparing these two strings, one can guess whether X and Y are a match. To perform the matching of these strings, Oliveira-Neto et al. proposed using the Levenshtein edit distance technique (*1, 2*). To apply this technique to the example, two fundamental operations (the substitutions B to 8 and 2 to Z) are required to convert X to Y; hence, the total edit distance is two. If the edit distance between the two strings falls below an assigned threshold value, a match is declared with some level of certainty.

By examining the predictability of OCR recognition patterns (1/I, 0/O, 2/Z, 8/B, and 5/S), one could make a more educated guess whether X and Y are a match. Oliveira-Neto el al., consequently, proposed a self-learning license plate–matching algorithm that included a generalized edit distance technique with a weight function (*3, 4*). This generalized technique assigns different weights to the edit operations as a function of the character to allow a measure of similarity between the two plate strings. This technique relies on an association matrix to provide the measure of similarity for whether a character (e.g., B) recognized by an upstream LPR station is recognized as B or 8 (or any other characters) at the downstream station. These measures of similarity are based on the performance of an associated pair of LPR stations.

LPR technology can achieve different levels of accuracy that depend on the hardware and software of the camera, setup (mobile or stationary), location (side of the road or overpass), and on-site calibration. Outside factors also play a role in the level of accuracy (e.g., traffic and weather conditions). The two facets of LPR technology that have the largest effect on the overall performance are the capturing and reading of license plates (*1*). First, the "capture rate" is the rate of successful plate recognitions in the field of view. This rate is commonly affected by uncontrollable outside factors and parameters pertaining specifically to the camera's hardware, installation, or on-site calibration. Second, the "read rate" is the rate of correct interpretation of an entire license plate and is based solely on the performance of the OCR engine. The accuracy of these facets is commonly uncontrollable by the data output user and dependent on the performance of the LPR technology.

Experience shows that the typical read rate of LPR cameras rarely surpasses 80% and more commonly performs at 60% or less (*1, 3–5*). Consequently, the portion of license plate strings that are correctly

S. R. Hargrove and H. Lim, Room 311, and L. D. Han, Room 319, John Tickle Engineering Building, Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Knoxville, TN 37996-2313. Corresponding author: L. D. Han, lhan@utk.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2594, Transportation Research Board, Washington, D.C., 2016, pp. 51–60.
DOI: 10.3141/2594-09