Validating the Performance of Vehicle Classification Stations

Benjamin Coifman
Ho Lee
Seoungbum Kim

for the
Ohio Department of Transportation
Office of Research and Development

State Job Number 134516

May 2012

1. Report No. **FHWA/OH-2012/7**	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and subtitle **Validating the Performance of Vehicle Classification Stations**		5. Report Date **May 2012**
		6. Performing Organization Code
7. Author(s) **Benjamin Coifman, Ho Lee, Seoungbum Kim**		8. Performing Organization Report No.
		10. Work Unit No. (TRAIS)
9. Performing Organization Name and Address **The Ohio State University Hitchcock Hall 470 2070 Neil Ave. Columbus, OH 43210**		11. Contract or Grant No. **State Job No. 134516**
12. Sponsoring Agency Name and Address **Ohio Department of Transportation 1980 West Broad Street Columbus, Ohio 43223**		13. Type of Report and Period Covered **Final Report**
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Vehicle classification is used in many transportation applications, e.g., infrastructure management and planning. Typical of most developed countries, every state in the US maintains a network of vehicle classification stations to explicitly sort vehicles into several classes based on observable features, e.g., length, number of axles, axle spacing, etc. Periodic performance monitoring is necessary to ensure the quality of collected data; however, such testing has been prohibitively labor intensive to do as thoroughly as needed. To address these challenges, this study examined three interrelated facets of vehicle classification performance monitoring. First, we manually evaluate the performance of vehicle classification station on a per-vehicle basis, second we develop a portable LIDAR (light detection and ranging) based vehicle classification system that can be rapidly deployed, and third we use the LIDAR based system to automate the manual validation done in the first part using the tools from the second part.

In the first part we examined over 18,000 vehicles, at several stations and found good performance overall, but performance for trucks was far worse than passenger vehicles. About a third of the errors were fixed by modifying the classification decision tree, the remaining two thirds of the errors are unavoidable because different classes have overlapping axle spacings or lengths (e.g., passenger vehicles and trucks, or commuter cars and motorcycles). All subsequent uses of the classification data must accommodate this unavoidable blurring. Next, we develop a side-fire LIDAR based classification system that does not require any calibration in the field. Finally, we develop a process to use the LIDAR system (or another temporary vehicle classification system) deployed concurrent to a permanent classification station to semi-automate the manual validation. The automated process does the bulk of the work, typically taking a user only a few minutes to validate all of the exceptions from all lanes over an hour of data. We found wide variance in performance from one station to the next. Since these errors are a function of the specific station, there would be benefit in the short term to leverage the LIDAR based system to evaluate the performance of many other classification stations to catch systematic errors that bias classification performance.

17. Key Words **Vehicle classification, Highway traffic monitoring, Portable non-intrusive vehicle detection, LIDAR, Performance evaluation**		18. Distribution Statement **No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161**	
19. Security Classif. (of this report) **Unclassified**	20. Security Classif. (of this page) **Unclassified**	21. No. of Pages **97**	22. Price

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

(This page intentionally left blank)

TITLE PAGE

Validating the Performance of Vehicle Classification Stations

Benjamin Coifman, PhD
Associate Professor
The Ohio State University
Joint appointment with the Department of Civil, Environmental, and Geodetic Engineering, and the Department of Electrical and Computer Engineering
Hitchcock Hall 470
2070 Neil Ave, Columbus, OH 43210
Phone: (614) 292-4282
E-mail: Coifman.1@OSU.edu

Ho Lee, PhD Candidate
Graduate Research Associate
Department of Civil, Environmental, and Geodetic Engineering
The Ohio State University
E-mail: lee.2406@osu.edu

Seoungbum Kim, PhD Candidate
Graduate Research Associate
Department of Civil, Environmental, and Geodetic Engineering
The Ohio State University
E-mail: kim.1936@osu.edu

Prepared in cooperation with the Ohio Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

We are particularly grateful for the assistance and input from David Gardner and Linsdey Pflum at the Ohio Department of Transportation, Gerald L. Wilcox at Traffic Engineering Services, and Zhuojun Jiang at the Mid-Ohio Regional Planning Commission (MORPC).

January, 2012

(This page intentionally left blank)

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

1 INTRODUCTION

This study examines three interrelated facets of vehicle classification and classification performance monitoring. The overall objectives of this study are:

1) Use manual data reduction from concurrent video to evaluate how well ODOT classification stations sort vehicles into the 13 FHWA classes and separately the three length classes used by ODOT.

2) Identify any chronic problems in the automated classification performance so that ODOT can ensure accurate vehicle classification and using the per vehicle ground truth evaluate the axle based classification decision tree for the studied locations to see if any improvements can be realized.

3) Investigate and develop non-labor intensive means to conduct these evaluations, to allow for on-going calibrations of classification stations.

To these ends, first, we manually evaluate the performance of vehicle classification station on a per-vehicle basis, second we develop a portable LIDAR (light detection and ranging) based vehicle classification system that can be rapidly deployed, and third we use the LIDAR based system to automate the manual validation done in the first part using the tools from the second part. Each component is discussed in a separate chapter, as follows.

In Chapter 2 we evaluate the performance of three freeway, permanent vehicle classification stations against concurrent video based ground truth (details of the various test sites can be found in Appendix A). All of the stations used in this chapter have dual loop detectors and a piezoelectric sensor in each lane, providing both axle-based and length-based classification. The performance evaluation is done at the "per-vehicle record" resolution, i.e., we compare every individual vehicle that passed during the study periods (over 18,000 vehicles, uncongested conditions). While the stations exhibited good performance overall (97% correct), across all three stations the performance for trucks was far worse, e.g., only 60% of the single unit truck/bus (SUT) - axle class 4-7- were correctly classified as SUT by the axle-based classification decision tree. We diagnosed all of the observed errors and some can be fixed quickly (e.g., gaps between bins) while others cannot. Using data from one site, we revise the axle-based classification decision tree to solve almost all of the fixable errors and then test the performance at another location. This new classification decision tree can be deployed immediately.

One chronic error found in this research is intrinsic to the vehicle fleet and may be impossible to correct with the existing sensors; namely, the shorter, SUT have a length range and axle spacing range that overlaps with passenger vehicles (PV) - axle class 1-3. Depending on the calibration, the error may be manifest as SUT counted as PV or vice versa. One should expect such errors at most classification stations. All subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of SUT with PV. The blurring also means that one cannot blindly use an axle classification station to calibrate the boundary between PV and SUT for length-based classification stations, otherwise, the unavoidable errors in the axle-based classification will be amplified in the length-based classification scheme.

In Chapter 3 we move out of the right-of-way and develop a LIDAR based classification system with the sensors mounted in a side-fire configuration next to the road. The first step is to distinguish between vehicle returns and non-vehicle returns. The algorithm then clusters the vehicle returns into individual vehicles. The algorithm examines each vehicle cluster to check if there is any evidence of partial occlusion from another vehicle. Several measurements are taken from each non-occluded cluster to classify the vehicle into one of six classes: motorcycle, passenger vehicle, passenger vehicle pulling a trailer, single-unit truck, single-unit truck pulling a trailer, and multi-unit truck. The algorithm was evaluated at six different locations under various traffic conditions (again, details of the various test sites can be found in Appendix A). Compared to concurrent video ground truth data for over 27,000 vehicles on a per-vehicle basis, 11% of the vehicles are suspected of being partially occluded. The algorithm correctly classified over 99.5% of the remaining, non-occluded vehicles. This research also uncovered emerging challenges that likely apply to most classification systems, e.g., differentiating commuter cars from motorcycles.

Occlusions are inevitable in this proof of concept study since the LIDAR sensors were mounted roughly 6 ft above the road, well below the tops of many vehicles. Ultimately we envision using a combination of a higher vantage point (in future work), and shape information (begun herein) to greatly reduce the impacts of occlusions.

Even with the impacts of occlusions, the LIDAR system is a valuable tool. In Chapter 4, we seek to automate the process of evaluating the classification stations, i.e., addressing the problem in Chapter 2 with the tools from Chapter 3. There are many classification technologies, each with its own strengths and weaknesses, but all of these systems depend on accurate calibration and validation to yield meaningful results. Such performance monitoring has been prohibitively labor intensive, prone to human error, and conventional aggregation periods are too coarse, allowing over counting errors to cancel undercounting errors. This work develops a classification performance monitoring system to allow operating agencies to monitor the health of their classification stations. We eliminate most of the labor demands and instead, deploy a portable non-intrusive vehicle classification system (PNVCS) to classify vehicles, concurrent with an existing classification station. Our system uses a LIDAR based PNVCS but our approach is compatible with many other portable vehicle classification systems. This pilot study used LIDAR sensors mounted on a van and our system does not require any calibration in the field. For longer-term deployments we envision a dedicated trailer that could be parked alongside the road.

To prevent classification errors from canceling one another in aggregate, we evaluate performance on a per-vehicle record basis. The approach requires several intermediate steps, developed herein, including synchronizing the independent clocks and matching observations of a given vehicle between the two classification systems. These algorithms automatically compare the vehicle classification between the existing classification station and the PNVCS for each vehicle. If the two systems agree, the given vehicle is automatically taken as a success. A human only looks at a given vehicle when the two systems disagree, and for this task we have developed tools to semi-automate the manual validation process, greatly increasing the efficiency and accuracy of the human user (typically on the order of 4 sec per vehicle- including seek time and loading time, translating to a few minutes to validate all of the exceptions from all lanes over an hour of data). The automated process does the bulk of the work, less than 8% of the vehicles required manual intervention. The methodology is applied to several permanent and temporary vehicle classification stations to evaluate axle and length-based classification. The evaluation datasets include over 21,000 vehicles. This evaluation also revealed a chronic problem detecting motorcycles at the two ODOT permanent classification stations studied. While the LIDAR system detected 15 passing motorcycles, the classification stations correctly classified only one of them, and missed five altogether.

2 AXLE AND LENGTH BASED VEHICLE CLASSIFICATION PERFORMANCE

2.1 Introduction

For many transportation applications it is important to know the mix of passing vehicles on the roadway. The volume of different vehicle classes are used for pavement design and management, modeling freight flows, and studying air quality since different vehicle classes make systematically different contributions [1]. The classification data are also important to ITS, e.g., for automated tolling the toll facility often classifies vehicles and charges different rates depending on the classification.

Typical of most developed countries, every state in the US maintains a network of vehicle classification stations to explicitly sort vehicles into several classes based on observable features, e.g., length, number of axles, axle spacing, etc.. Various technologies are used for this automated classification, the three most common approaches are: weigh in motion (WIM); axle-based classification from a combination of loop detectors, piezoelectric sensors or pneumatic sensors; and length-based classification from dual loop detectors. There are many more emerging technologies that also promise vehicle classification, e.g., video image processing and side-fire microwave radar.

This chapter examines the performance of three permanent vehicle classification stations operated by the Ohio Department of Transportation (ODOT) on different freeways around Columbus, OH (Figure 2-1(a)). Each lane at each of the stations has dual loop detectors to measure speed and vehicle length, and a piezoelectric sensor to detect the axle passages (Figure 2-1(b)), providing both the conventional 13 axle-based classes [1] and length-based classification. In the latter case, it is common to provide only three or four classes, which are intended to map to passenger vehicles (PV) - axle class 1-3, single unit truck/bus (SUT) - axle class 4-7, and multi-unit trucks (MUT) - axle class 8-13.

The performance evaluation is done at the "per-vehicle record" (pvr) resolution, i.e., we compare every individual vehicle that passed during the study periods (totaling over 18,000 vehicles, uncongested conditions). Evaluating the pvr data as we do in this work is uncommon; normally the pvr classifications are binned by fixed time periods, e.g., over 15 min or 1 hr, and the individual vehicle information is discarded. However, such conventional aggregation allows errors to cancel one another, which can obscure underlying problems.

While the stations exhibited good performance overall (97% correct), across all three stations the performance for trucks was far worse, e.g., only 60% of the SUT were correctly classified as SUT by the axle-based classification decision tree. We diagnosed all of the observed errors and some can be fixed quickly (e.g., gaps between bins) while others cannot. Using data from one site, at the end of this chapter we revise the axle-based classification decision tree to solve almost all of the fixable errors and then test the performance at another location.

One chronic error found in this research is intrinsic to the vehicle fleet and may be impossible to correct with the existing sensors; namely, the shorter, SUT have a length range and axle spacing range that overlaps with PV. Depending on the calibration, the error may be manifest as SUT counted as PV or vice versa. One should expect such errors at most classification stations. All subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of SUT with PV. The blurring also means that one cannot blindly use an axle classification station to calibrate the boundary between PV and SUT for length-based classification stations, otherwise, the unavoidable errors in the axle classification will be amplified in the length-based classification scheme.

The challenge from SUT blurring with PV is not unique to conventional detectors. Our group found similar problems between these two groups when using side-fire LIDAR to classify vehicle profiles (Chapter 3) and estimated vehicle length from single loop detectors [2]. Meanwhile, several non-invasive sensor manufacturers now offer length based vehicle classification as a feature of their sensors and their classification performance has been evaluated [3-8]. Most of these studies rely on manual counts for ground truth to quantify performance and typically found overall classification error rates between 5%-10%. Like the present study, however, most of the passing vehicles were PV. Many of the studies sampled counts over extended periods, e.g., 15 min or 1 hr [3-6], which as noted above, allows for over-counting errors to cancel under-counting errors. Even allowing the individual errors to cancel, the SmartSensor had an overall error rate for trucks (SUT and MUT combined) of 46% [3], 80% [4], 50%-400% [5], 20%-50% [6] and the RTMS had an error rate for trucks of 25% [3], 40%-97% [5]. Two studies used a small sample of pvr data, only a few hundred vehicles, and found the SmartSensor had an error rate for trucks of 13%-57% [7], 42% [8]. A few studies considered video systems, e.g., [6] found the length based classification from an Autoscope to be unacceptable, while [9] had an error rate for trucks of 73%. Although these studies reveal degraded non-invasive based classification for trucks, the authors do not explicitly investigate the causes.

2.1.1 Overview

The remainder of this chapter is as follows, in Section 2.2 we briefly review the details of the classification stations, collection of concurrent video based ground truth data, and data reduction processes for validation. Section 2.3 presents the overall performance of the stations and discusses the systematic errors that we observed. Using the ground truth data at one station, Section 2.4 develops a new classification decision tree to eliminate most of the preventable errors and then evaluates the performance at another station. Finally, Section 2.5 presents the conclusions and summarizes the results of this study.

2.2 Classification Stations and Concurrent Ground Truth Data

Table 2-1 enumerates summary statistics for the three classification stations used in this study and Figure 2-1(a) shows their locations. The observation periods ranged between 1 and 3.5 hours, during which time the per-vehicle record (pvr) data from the classifier were logged for the research and concurrent video was recorded for evaluation. The classifier uses the dual loop detectors and piezoelectric sensor to calculate vehicle speed, length, and axle spacing(s). The classifier uses fixed length-thresholds to assign length-class and a decision tree to assign axle-class based on the number of axles and their spacing(s). For each vehicle the pvr data include: time stamp, lane, speed, number of axles, axle spacing(s), axle-class, vehicle length, and length-class. Appendix B provides more details on the pvr and existing classification scheme.

After collecting the data in the field, we manually generate ground truth data from the video to evaluate the performance of the classification stations. Although both the video and pvr data are time stamped, the two clocks are independent, so the two datasets need to be time synchronized with one another. The time offset is a constant and in the absence of any detection errors, a sequence of observed headways in one dataset provides a unique pattern that can be found in the other dataset (similar to the vehicle reidentification in [10]). Or more formally, we manually extract 9 successive headways from the video. After only a few vehicles the sequence becomes distinct. We then look for this same sequence in the pvr data by finding the time off-set that minimizes the total relative error between the video and pvr data time stamps for the successive vehicles. Once the two datasets are time synchronized, we employ a semi-automated process to generate the ground truth data using a software tool to simultaneously view the pvr classification and the corresponding video frame. Figure 2-2 shows a screen shot from the tool as the user classifies a MUT.[1] Prior to selecting a class for the given vehicle, the user can step backward or forward in time if there is any uncertainty. The user then selects the axle classification for the vehicle (or in rare cases indicates either that the vehicle is unclassifiable or that it is a non-vehicle actuation). In any event, after the user has entered a class for the vehicle, the software immediately jumps to the next vehicle reported in the pvr data for the lane. Obviously this approach will not catch a vehicle that is completely missed by the classification station. The focus of the present work is on classification performance; however, one could use additional techniques to also catch missed vehicles (e.g., using a simple video image processing "trip wire", as in [11]; or an independent sensor, as in Chapter 4). In an ideal case, every single vehicle in the pvr would be assigned to its specific class, as was done in the I-70 dataset. The vast majority of the vehicles in our datasets are PV, axle-class 1-3. To greatly reduce the labor necessary to reduce the data, in the I-270 dataset we combined the PV into a single group, and in the SR-33 dataset we do a similar consolidation for SUT and MUT. In any event, all of the vehicles were manually classified at the resolution shown in the bottom row of Table 2-1.

2.3 Performance of the Classification Stations

Table 2-2 compares the pvr axle-class against the manual classifications over the 13 conventional axle-classes for the 8,079 vehicles in the I-270 dataset. A given vehicle is counted in a single cell, the row corresponding to its ground truth axle-class and the column corresponding to its pvr axle-class. Thus, each cell shows the total number of vehicles with the pairwise combination from the manual and pvr classifications. As noted above, for a vehicle with pvr axle-class 1-3 in I-270, the user only verifies that it is indeed a PV when generating the ground truth, and thus, the top left cells span three rows.

[1] In the event that this figure is hard to read, the key features are the integrated video view, detector data, and the fact that there are several buttons for user input. The figure may be clearer in the electronic version of the report available from ODOT.

Table 2-1, Summary statistics of ground truth datasets.

Location	Axle stations		
	Southbound I-270 at Rings Rd.	Eastbound I-70 at Brice Rd.	Northbound SR-33
Date	Nov 2, 2010	June 20, 2006	Aug 3, 2011
Traffic Conditions	Free flow	Free flow	Free flow
Time duration investigated	9:27~12:33	10:12~13:59	13:28~14:34
Average Speed	64 mph	65 mph	64 mph
Number of lanes	3	3	2
Average Flow (per lane)	873 vph	859 vph	627 vph
# of vehicles	8,079	9,746	1,255
# of occluded vehicles	30	377	0
Resolution of ground-truth	PV or axle class 4-13	axle class 1-13	PV, SUT, MUT

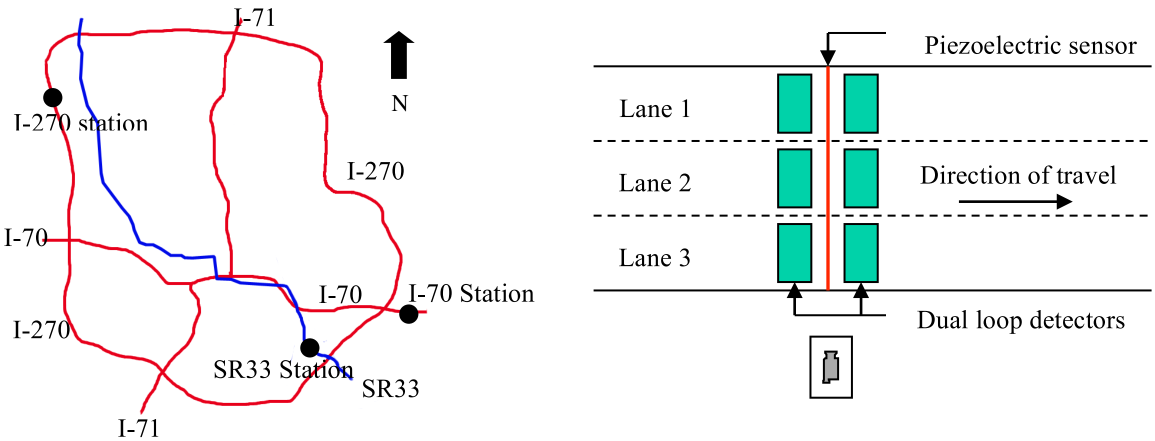


Figure 2-1, (a) Location of axle classification stations used in this study, around the Columbus, Ohio, metropolitan area, (b) Schematic of typical axle classification station, the camera icon shows the approximate location of our video recording.

Figure 2-2, Screen shot of the graphical user interface used to generate the ground-truth classifications- the primary video stream is in the bottom center, the classification options on the right, and several navigation buttons elsewhere on the display. The top two video streams are supplementary, showing concurrent views along the freeway to help when there is an occlusion of a far lane (note that the rear view, on the left, was flipped horizontally when filmed).

For classes 4-13, the cells on the diagonal tally the number of correct classifications, while all of the cells off of the diagonal tally classification errors. Overall 97% of vehicles are correctly classified (excluding any errors strictly among Classes 1-3), we reviewed all of the vehicles that fell in cells off of the diagonal in Table 2-2. The various sources of error are denoted with superscripts and will be discussed in the next section. These detailed results are similar to those from I-70 (see Appendix C for details). Although the resolution of the SR-33 ground truth precludes such a quantified comparison, qualitatively SR-33 performance seemed similar to the other stations.

2.3.1 Investigation of Axle-based Misclassifications

Table 2-2 shows that 2.2% (185 of 8,049) of the vehicles with ground truth were misclassified at the I-270 site. We reviewed the video and actuations from all 185 of the erroneous vehicles to diagnose the source of each error. We found six different sources of the misclassifications, denoted with superscripts in the table and described below.

<u>Case "a": Axle spacing falling in between two axle spacing bins</u>

In Table 2-2 there are 26 axle-class 2 vehicles misclassified as class 13, which is clearly an error since all of these vehicles had only two axles while class 13 is defined to have seven or more axles. Although Table 2-2 only uses 3 hrs of data, we have a total of 14.5 hrs of pvr data from the station. The remaining period does not have concurrent video, but is still useful for diagnosing this problem. Over the entire dataset we found 88 class 13 vehicles with only two axles. Looking at a distribution of their axle spacing measurements, we found all of these vehicles had one of five discrete axle spacing measurements, as shown in Table 2-3. The discretization is not in itself problematic or surprising, it merely reflects the sampling resolution of the classifier. However, reviewing the axle spacing criteria for the two-axle vehicle classes (axle-classes 1-5) in the classifier's decision tree (see Table B-2 in Appendix B), it became apparent that there were small gaps between upper-bound of one class and the lower-bound for the next. These 88 vehicles literally fell between the cracks between the classes (as shown in Figure 2-3). Obviously the bounds should be made continuous to avoid these errors. Since the decision tree does not differentiate between axle-class 13 and *unclassifiable*, the errors were compounded when the two-axle vehicles were assigned to axle-class 13. To prevent similar errors from going undetected, an operating agency should explicitly define axle-class 13 and then add a 14th class for the otherwise unclassifiable vehicles (e.g., as used in [3, 6]).

<u>Case "b": Two-axle SUT with short axle spacing</u>

In Table 2-2, there are 87 axle-class 5 vehicles misclassified as either axle-class 2 or axle-class 3 in the pvr data. Figure 2-4 shows typical examples of the seven different truck types that give rise to these 87 errors. Reviewing the decision tree in Appendix B, all of these vehicles had an axle spacing that falls into the pvr assigned axle-class (see Appendix D for a detailed review of these vehicles), i.e., this problem arises because these vehicles' axle-spacing falls below the boundary for SUT. The classifier correctly classified these vehicles given their measured axle spacing and there is no indication that the axle spacing measurements were inaccurate. This problem cannot be resolved by lowering the boundary, as shown in Figure 2-5(a), because many PV would then be misclassified as SUT. In fact if the boundary were moved any lower the number of PV misclassified as SUT would exceed the number of SUT errors that are eliminated. It is possible that some of these errors could be eliminated if the classifier considered both vehicle length and axle spacing in the decision tree. Although as discussed in Section 2.3.3, most of these errors would remain because the axle spacing and vehicle length of these vehicles are highly correlated. Marginal improvements could be made by also considering the distance between the last axle and rear bumper, but still many of these errors would persist.

With this mechanism in mind, returning to the decision tree, PV pulling trailers are classified as PV (based on the first axle spacing) even though they have more than two axles, while SUT pulling trailers are classified as MUT. This difference in handling vehicles with trailers caused some of the short axle spacing errors to impact vehicles classified as MUT in the ground truth data. Reviewing the SUT pulling trailers, we found nine axle-class 5 vehicles pulling trailers (thus, making them class 8 or 9, depending on the number of trailer axles) that were misclassified by the decision tree as class 3 in the pvr data because they had a short first axle spacing.

<u>Case "c": Errors from buses</u>

The decision tree assumes two-axle buses have a larger axle spacing than two-axle SUT. As shown in Figure 2-5(b), among the two-axle vehicle classes, the range of buses' axle spacing overlaps with that of class 5 SUT, giving rise to a problem similar to Case "b". In Table 2-2 there were six two-axle buses with axle spacing in the range for class 5 SUT, and four two-axle class 5 SUT with axle spacing in the range for class 4 buses. Thus, all 10 of these vehicles were misclassified. As with Case "b", the error is unavoidable and all subsequent analysis of the classification data must accommodate this disproportionately higher error rate.

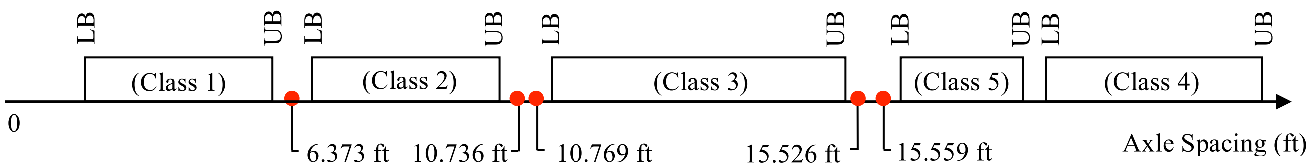
Table 2-2, Comparison between pvr and ground truth axle-class in the I-270 dataset.

Axle classification station in I270		ODOT Axle-based vehicle classification												
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13
Manual axle-based vehicle classification	class 1: Motorcycle				-	-	-	-	-	-	-	-	-	-
	class 2: Car	-	5881	1104	-	-	-	-	-	-	-	-	-	26[a]
	class 3: other 2axle, 4tire single-unit veh				-	-	-	1[f]	2[f]	-	-	-	-	-
	class 4: Bus	-	-	-	1	6[c]	-	-	3[c]	-	-	-	-	-
	class 5: 2 axle, 6tire, single-unit truck	-	6[b]	81[b]	4[c]	111	-	-	-	-	-	-	-	-
	class 6: 3 axle single-unit truck	-	-	-	-	-	71	-	-	-	-	-	-	-
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	1[d]	9	-	2[e]	21[e]	-	-	-
	class 8: 4 or fewer axle single-trailer truck	-	-	8[b]	1[c]	-	-	4[f]	29	-	-	-	-	1[a]
	class 9: 5 axle single-trailer truck	-	8[d]	1[b]	-	-	-	-	4[d]	616	-	-	-	-
	class 10: 6 or more axle single-trailer truck	-	2[d]	-	-	-	-	-	1[d]	-	13	-	-	1[a]
	class 11: 5 or fewer axle multi-trailer truck	-	1[d]	-	-	-	-	-	-	-	-	20	-	-
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	8	-
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1
	Unclassifiable vehicle or occluded vehicle	2	-	-	-	3	4	1	3	14	2	1	-	-
	Non-vehicle actuation	-	-	-	-	-	-	-	-	-	-	-	-	-

a: Axle spacing falling between two axle spacing bins
b: Two-axle SUT with short axle spacing
c: Errors from buses
d: Axle classification station reports incorrect number of axles
e: Errors from class 7
f: Errors from vehicles pulling trailers

Table 2-3, Axle spacing of axle-class 13 vehicles with only two axles.

Axle Spacing (ft)	# of Samples
6.373	1
10.736	41
10.769	43
15.526	2
15.559	1

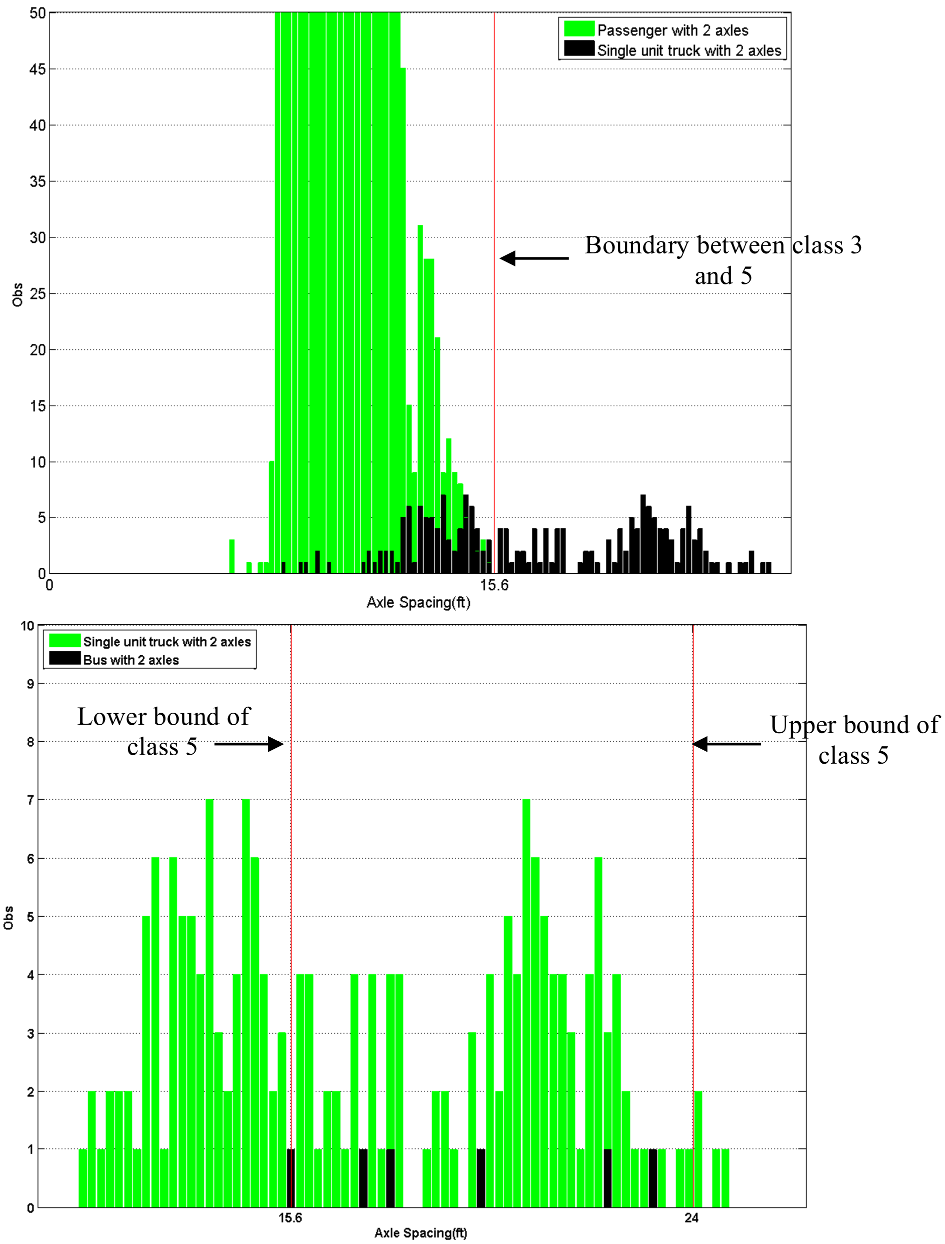


Where, “UB” and “LB” denote the upper and lower bounds respectively

Figure 2-3, Axle spacing of the axle-class 13 vehicles with only two axles against the bounds of the various two-axle vehicle classes.

Figure 2-4, Examples of the seven types of two-axle SUT that were misclassified as PV by the axle classification station because the axle spacing fell below the PV boundary.

(a)



(b)

Figure 2-5, (a) Histogram of axle spacing for PV and for SUT (excluding all vehicles pulling trailers), (b)Histogram of axle spacing for SUT and for buses (excluding all vehicles pulling trailers)

On the other hand, the decision tree does not explicitly consider buses pulling trailers. The existing decision tree only allows for up to 3 axle buses. There were three buses pulling a PV, yielding 4 axles and these vehicles were classified as if they were a SUT pulling a trailer (class 8). However, it is possible to catch some of these errors by explicitly looking for buses pulling trailers, as will be illustrated shortly.

Case "d": Axle classification station reports incorrect number of axles

There are 17 vehicles in Table 2-2 that had fewer axles in the pvr data than observed in the ground truth. It appears that the classification station missed one or more axles on each of these vehicles. Upon inspection, all of these vehicles were straddling the edge of the lane as they passed the station, either changing lanes or traveling partially on the shoulder. Fortunately, none of these misclassified vehicles were double counted in the adjacent lane. This type of error would not be easily identified from the axle classification station data, but the frequency is low. While the I-270 dataset in Table 2-2 only shows errors due to missing axles, the I-70 dataset also exhibits errors due to overcounting the axles (see Table C-1 in Appendix C), where we believe the piezoelectric sensors extended slightly into the adjacent lane and would occasionally detect axles from the wrong lane. In any event, the Case "d" errors are due to sensing faults, not the classifier.

Case "e": Errors from axle-class 7

Table 2-2 shows several axle-class 7 vehicles that were misclassified as MUT. The decision tree used at this station implicitly assumed axle-class 7 vehicles had *exactly* four axles, while the conventional definition is for SUT with four *or more* axles. So the 23 class 7 trucks with more than four axles were counted as MUT (class 9 or 10, depending on the number of axles). As will be illustrated, an added step in the decision tree can catch most of these vehicles, since an axle-class 7 vehicle with more than four axles will typically have much shorter axle spacings than a MUT with the same number of axles.

Case "f": Errors from vehicles pulling trailers

Like Cases "b" and "c", the axle spacings for vehicles pulling trailers overlap with axle spacings for MUT. There were three axle-class 3 vehicles with trailers long enough to look like trucks (class 7 and 8). While there were four MUT with small enough axle spacings that they looked like SUT. Like Cases "c" and "e", it is possible to prevent some of these errors, as will be discussed shortly.

2.3.2 Consolidating Classifications by Vehicle Type

At a more coarse level, the three shaded regions in Table 2-2 contain vehicles that were assigned the correct vehicle type: PV (class 1-3), SUT (class 4-7), or MUT (class 8-13). The off-diagonal cells within these shaded regions represent less severe errors, since the missclassified vehicles were still assigned the correct vehicle type. These intra-type errors represent 9.7% of the total misclassifications in the table. Using the three vehicle types, Table 2-4(a) reiterates the performance from the I-270 station at the coarser granularity and the off diagonal cells retain all 167 of the inter-type misclassifications. The bottom right cell shows the overall performance across all three vehicle types. Table 2-4(b) and 2-4(c) repeat this exercise for I-70 and SR-33. Note that although all three sets had over 97% success rate, the number of SUT that were correctly classified (i.e., by row) is on the order of 60%.

2.3.3 Length-Based Vehicle classification

Length-based vehicle classification uses less information than axle-based classification to sort vehicles into class. Due to the lower fidelity available from the length measurements, most length-based classification schemes only sort vehicles by type, e.g., length-class 1: PV, length-class 2: SUT (including buses), or length-class 3: MUT. As mentioned above, all three test-sites also report length-based classification in the pvr data. After looking at the distribution of vehicle lengths at the I-270 site, the classifier used 20.5 ft and 40.5 ft of physical vehicle length as the upper boundary for PV and SUT, respectively[2]. In the absence of detector errors, this resolution is comparable to the consolidated axle classes shown in Table 2-4. We use the ground truth vehicle types to indirectly evaluate the length-based classifications. To this end, using the 8,049 vehicle records with ground truth axle-classes at the I-270 site, the ground truth axle-classes are clustered into type (as was done in Table 2-4(a)) and then each length-based class is compared with the corresponding ground truth type in Table 2-5(a). Overall, the length-based classification is 97% accurate. Compared to Table 2-4(a), there are very few true SUT that are misclassified (91% of the SUT are correctly classified), but now there are many PV that are classified as SUT (only 69% of the vehicles classified as SUT are actually SUT). This result reflects the fact that the threshold between the two classes is lower in Table 2-

[2] [11] reports that ODOT typically uses 22 ft and 40 ft for the length thresholds.

5(a) than in Table 2-4(a). Table 2-5(b) and 2-5(c) repeat this exercise for I-70 and SR-33. Reviewing the vehicles with errors, the majority of the PV classified as SUT were pulling trailers (67 of 118). Ideally, the threshold would balance over-counting with under-counting, but that is impossible to do with a constant threshold since the optimal threshold also depends on the relative flow of SUT.

Figure 2-6 explicitly shows the trade-off from the axle spacing boundary and vehicle length boundary between PV and SUT for the two-axle vehicles. The two axle vehicles are sorted based on the ground truth classification, Figure 2-6(a) shows that all PV fell below the axle spacing boundary, but some were above the length boundary and (b) shows many SUT with two axles below one or both of the boundaries. Figure 2-6(c) shows PV and SUT together, a dark colored square highlights a SUT assigned to a PV axle-class in the pvr. It is impossible to choose a threshold on either dimension that would be error free, in each case the PV and SUT ranges overlap. This blurring means that one cannot blindly use an axle classification station to calibrate the boundary between PV and SUT for length-based classification stations, otherwise, the unavoidable errors in the axle classification will be amplified in the length-based classification scheme.

2.4 Improving the Axle-Based Classification Decision Tree

While axle classification shows over 97% correct, we observed systematic misclassification errors categorized by the six types discussed in Section 2.3.1, and summarized as follows for the I-270 dataset:

1: Unavoidable misclassifications (case “b”, case “c”, case “d”, case “f”)	68%
2: Misclassifications due to class 7 with 5 or more axles (case “e”)	12%
3: Misclassifications caused by decision tree (case “c”, case “f”)	5%
4: Misclassifications due to gaps between two classes (case “a”)	15%

While the majority of misclassifications (68%) are unavoidable due to an overlapping range in axle spacing among classes (case “b”, case “c”, case “f”) or the detector reporting the incorrect number of axles (case “d”), in this section we address the remaining misclassifications (32%) by recalibrating the ODOT decision tree and adding new steps to it. We recalibrate the decision tree using the I-70 dataset and then evaluate the performance using the I-270 dataset.

First, we address case "a" by closing the gaps between classes, add an explicit definition for class 13, and create a 14th bin for unclassifiable vehicles. Secondly, to address the case "e" misclassifications we add several new steps to classify class 7 vehicles with 5 or more axles. A typical class 7 with 5+ axles has 4+ closely spaced rear axles. This cluster of so many axles is unique among the observed vehicles can be used for identifying class 7 vehicles with 5+ axles. The five-axle, class 7 vehicles' axle spacing distributions show that S_2 and S_3 fall between 1 and 6 ft (where S_n denotes the n-th axle spacing), while S_4 has a longer upper bound of 13.1 ft. When there are more than five axles the final axle spacing is similar to S_4 in a five axle, class 7 vehicle, while the preceding spacings are similar to S_2 and S_3.

Next, we looked at all of the remaining case "c" and "f" misclassifications in the I-70 dataset, and then progressively updated the decision tree by adding steps, reordering steps, and changing boundaries to eliminate most of these errors at I-70. For example, the ODOT axle-based classification decision tree tends to misclassify class 7 vehicles with four axles as class 8 because the original decision tree checked for class 8 vehicles first and used boundaries that were too liberal. In the revised tree, we check for four axle class 7 vehicles before checking for class 8, and use more stringent criteria for both classes. If any change increased the number of misclassifications in the I-70 dataset, we kept the original conditions from the ODOT axle-based classification decision tree.

Although the range of length classes for two adjacent vehicle types tend to overlap (especially PV and SUT, e.g., Figure 2-6), we found that when combined with the axle spacings, vehicle length can help differentiate between SUT and MUT. We explicitly incorporate length class when segmenting three-axle class 6 vehicles from class 4 and 8 vehicles. We believe length class could also help segment axle-class 7 vehicles from MUT with the same number of axles; however, we did not observe enough vehicles in these classes to develop the threshold for vehicles with more than three axles.

Table 2-6 shows the resulting axle-based classification decision tree developed from on the I-70 dataset, after accounting for all of the above adjustments (compare to the original ODOT classification decision tree in Table B-2 in Appendix B). Note that this tree uses length class to select the three-axle class 6 vehicles, which catches five errors in the development dataset on I-70 that would occur if using axle spacing alone. The vehicle length test caught one more such error in the evaluation dataset after applying the decision tree to I-270. Aside from these six errors, the performance will not change if the length criterion is removed.

Table 2-4, Comparison between the pvr axle-class and ground truth vehicle type in (a) the I-270 dataset,(b) the I-70 dataset, and (c) the SR-33 dataset.

(a)

pvr axle versus ground-truth		pvr axle class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	6985	1	28	99.6%
	SUT	87	203	26	64.2%
	MUT	20	5	694	96.5%
% of column correct		98.5%	97.1%	92.8%	97.9%

(b)

pvr axle versus ground-truth		pvr axle class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	7494	3	21	99.7%
	SUT	107	255	56	61.0%
	MUT	13	18	1402	97.8%
% of column correct		98.4%	92.4%	94.8%	97.7%

(c)

pvr axle versus ground-truth		pvr axle class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	1127	2	4	99.5%
	SUT	14	44	16	59.5%
	MUT	0	1	47	97.9%
% of column correct		98.8%	93.6%	70.1%	97.1%

Table 2-5, Comparison between the pvr length-class and ground truth vehicle type in (a) the I-270 dataset, (b) the I-70 dataset, and (c) the SR-33 dataset.

(a)

pvr length versus ground-truth		pvr length class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	6867	118	29	97.9%
	SUT	26	286	4	90.5%
	MUT	0	11	708	98.5%
% of column correct		99.6%	68.9%	95.5%	97.7%

(b)

pvr length versus ground-truth		pvr length class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	7397	82	39	98.4%
	SUT	148	262	8	62.7%
	MUT	10	20	1403	97.9%
% of column correct		97.9%	72.0%	96.8%	96.7%

(c)

pvr length versus gound-truth		pvr length class			% of row correct
		PV	SUT	MUT	
Manual ground truth	PV	1117	14	2	98.6%
	SUT	19	54	1	73.0%
	MUT	1	3	44	91.7%
% of column correct		98.2%	76.1%	93.6%	96.8%

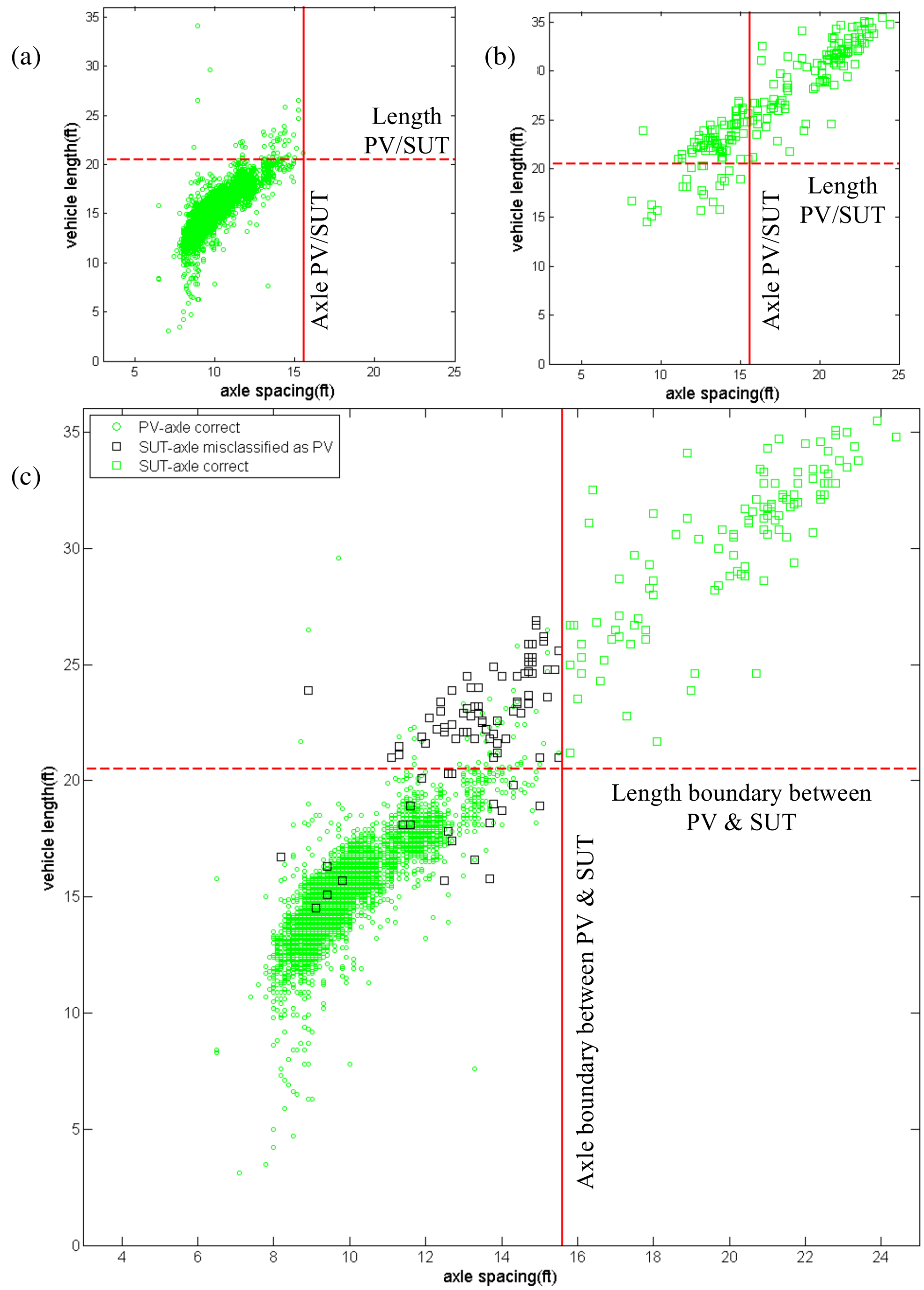


Figure 2-6, Length versus axle spacing of two-axle vehicles at the I-270 station (a) all PV, (b) all SUT, (c) all PV and SUT combined, highlighting the SUT-axle misclassified as PV by the axle boundary.

Due to the relatively small number of observations and larger variability in axle spacing, there are many SUT and MUT axle classes that would likely benefit from further data collection and decision tree refinement. Finally, note that this decision tree was tuned to the vehicles on central Ohio freeways. Obviously the decision tree would need further refinement if other axle configurations were present.

2.4.1 Evaluating the New Axle-Based Classification Decision Tree

To evaluate the performance of the new axle-based classification decision tree, we repeat the analysis from Table 2-2, comparing the axle-based classification results against the ground truth vehicle class. We apply the new decision tree first to the development dataset, Table 2-7(b), and then the evaluation dataset, Table 2-8(b). For reference, part (a) in both tables show the results from the original ODOT decision tree. In each table the numbers with a double strikethrough are unavoidable due to errors from overlapping ranges of axle spacing (case "b", case "c", case "f") and the numbers in parentheses are also unavoidable, due to the sensors reporting an incorrect number of axles (case "d"). Comparing Table 2-7(a) and (b), excluding the errors between PV classes, on I-70 there are 201 unavoidable errors due to case "b", "c", "d" or "f" from the ODOT classifier and 197 from the new axle-based classification decision tree. The small difference in unavoidable errors is simply noise, due to slight changes in the boundaries (e.g., if the boundary in Figure 2-5(a) moves slightly, the observed net error rate might change but there is little room for the expected net error rate to improve). The numbers in the black cells are potentially avoidable misclassifications that arise from the given classification decision tree. There are 59 such misclassifications in Table 2-7(a) due to the ODOT axle-based classification decision tree but only 1 in Table 2-7(b) from the new axle-based classification decision tree.

Repeating this comparison in the evaluation dataset on I-270, Table 2-8, we observe similar trends both in terms of unavoidable and avoidable misclassifications. The most noticeable difference from the development dataset is the larger number of two-axle vehicles that were assigned class 13 by the existing ODOT axle-based classification decision tree. In both the development dataset and evaluation dataset the new axle-based classification decision tree greatly reduced the number of avoidable misclassifications errors.

Table 2-9 and 2-10 reiterate the performance from the development and evaluation datasets at the coarser granularity of vehicle type. Part (a) in each table reiterates the results from Table 2-4. At this resolution on can see a roughly 10% improvement in the number of SUT that were correctly classified (row average). Since many of the now correctly classified SUT were erroneously classified as MUT by the ODOT axle-based classification decision tree, the percent of MUT classifications that are correct (column average) also improved by 4%-7%.

2.5 Conclusions

Vehicle classification stations are commonly used to sort vehicles into various classes based on observable features. Evaluating the pvr data as we do in this work is uncommon; both due to the inherent difficulty generating ground truth data, and the fact that normally the pvr classifications are binned by fixed time periods at which point the individual vehicle information is discarded. However, such conventional aggregation allows errors to cancel one another, which can obscure underlying problems. This study evaluated three permanent axle classification stations against concurrent video based ground truth in terms of axle-based and length-based classification. Only 3%-4% of the vehicles were misclassified, however, the relative impacts were much larger on the trucks, e.g., only 60% of the SUT were correctly classified as SUT by the existing axle-based classification decision tree.

Diagnosing the axle classification errors, it was found that all of them could be attributed to one of six causes. About a third of the errors among class 4-13 can be easily fixed by redefining the decision tree, e.g., ensuring that there are no gaps between successive classes and adding an additional outcome from the tree to indicate a vehicle is unclassifiable. Our revised decision tree is shown in Table 2-6. After making these changes, the new axle-based classification decision tree was able to correctly classify an additional 10% of the SUT, with smaller improvements in almost every other metric. Ideally the new decision tree should be deployed at a few new locations and the performance validated, then assuming no problems are found, be adopted as the new standard classification decision tree. In any event, most of the improvements of the decision tree should be incorporated in to standard practice (closing the gaps between bins, adding an "unclassifiable" class, and allowing for more than 4 axles in axle class 7).

One chronic error found in this research is intrinsic to the vehicle fleet and may be impossible to correct with the existing sensors; namely, the shorter, SUT have a length range and axle spacing range that overlaps with PV. Depending on the calibration, the error may be manifest as SUT counted as PV or vice versa. As discussed in the literature review, this PV/SUT blurring appears to impact other sensors as well. In any case, one should expect such errors at most classification stations. All subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of SUT with PV. The blurring also means that one cannot blindly use an axle classification station to calibrate the boundary between PV and SUT for length-based classification stations, otherwise, the unavoidable errors in the axle classification will be amplified in the length-based classification scheme.

Table 2-6, New axle-based classification decision tree, developed from the I-70 dataset ground truth. For each vehicle, the classifier will progress downward through the table until the vehicle first satisfies one condition, at which point the classifier stops and assigns that class to the vehicle.

# of axles	Class	Class name	Length	Axle Spacing (ft)	Index
2	1	Motorcycle		1~5.9	M
2	2	Car		5.9~**10.3**	M
2	3	other 2axle, 4tire, single-unit veh.		10.3~15	
2	5	2 axle, 6tire, single-unit truck		**15**~24	M
2	4	Bus		23.5~99.9	
3	6	3 axle single unit truck	0~40.5ft	any, 3.5~8	R
3	1	Motorcycle		1~**5.9**, any	M
3	2	Car		5.9~**10.3**, 10~18.8	M
3	3	other 2axle, 4tire, single-unit veh.		10.3~15, 10~18.8	
3	4	Bus		23.5~99.9, any	
3	8	4 or fewer axle single-trailer truck		any, **any**	M
4	7	4 or more axle single-unit truck		any, 1~6, 1~13.1	M, R
4	8	4 or fewer axle single-trailer truck		any, any, 3.5~8	M, R
4	8	4 or fewer axle single-trailer truck		any, 3.5~8, any	M, R
4	2	Class 2 pulling a trailer		1~10.3, **any**, any	M
4	3	Class 3 pulling a trailer		10.3~15, **any**, any	M
4	4	Bus pulling a trailer		23.5~99.9, any, any	A
4	4	Bus pulling a car		any, 17~99.9, 5.9~99.9	A
4	8	4 or fewer axle single-trailer truck		any, any, any	A
5	7	4 or more axle single-unit truck		any, 1~6, 1~6, 1~13.1	A
5	11	5 or fewer axle multi-trailer truck		any, 17~99.9, any, 6~99.9	M, R
5	9	5 axle single-trailer truck		any, 17~99.9, any, 3.5~11	A
5	9	5 axle single-trailer truck		any, **3.5~11**, any, 3.5~11	M, R
5	2	Class 2 pulling a trailer		1~**10.3**, any, 1~3.5, 1~3.5	M
5	3	Class 3 pulling a trailer		10.3~15, any, 1~3.5, 1~3.5	
5	9	5 axle single-trailer truck		any, any, any, any	A
6	7	4 or more axle single-unit truck		any, 1~6, 1~6, 1~6, 1~13.1	A
6	10	6 or more axle single-trailer truck		any, **1~8**, **1~8**, any, 8~99.9	M
6	12	6 axle multi-trailer truck		any, any, any, any, 8~99.9	
6	10	6 or more axle single-trailer truck		any, any, any, any, **1~8**	M
7	7	4 or more axle single-unit truck		any, any, 1~6, any, any	A
7	10	6 or more axle single-trailer truck		any, any, any, **1~8**, **1~8**	M
7	13	7 or more axle multi-trailer truck		any, any, any, any, any, any	A
8	10	6 or more axle single-trailer truck		any, 1~8, 1~8, any, 1~8, 1~8, 1~8	A
8	13	7 or more axle multi-trailer truck		any, any, any, any, any, any, any	A
9+	13	7 or more axle multi-trailer truck		any, any, any, any, any, any, any, any	A
any	14	Unclassified vehicle		others	A

M: Modified step from ODOT decision tree (changes are highlighted with bold text)
R: Reordered step from ODOT decision tree
A: Newly added step.

Table 2-6, continued- repeating the results in metric units.

# of axles	Class	Class name	Length	Axle Spacing (m)	Index
2	1	Motorcycle		0.3~1.8	M
2	2	Car		1.8~**3.1**	M
2	3	other 2axle, 4tire, single-unit veh.		3.1~4.6	
2	5	2 axle, 6tire, single-unit truck		**4.6**~7.3	M
2	4	Bus		7.2~30.4	
3	6	3 axle single unit truck	0~12.3 m	any, 1.1~2.4	R
3	1	Motorcycle		0.3~**1.8**, any	M
3	2	Car		1.8~**3.1**, 3~5.7	M
3	3	other 2axle, 4tire, single-unit veh.		3.1~4.6, 3~5.7	
3	4	Bus		7.2~30.4, any	
3	8	4 or fewer axle single-trailer truck		any, **any**	M
4	7	4 or more axle single-unit truck		any, 0.3~1.8, 0.3~4	M, R
4	8	4 or fewer axle single-trailer truck		any, any, 1.1~2.4	M, R
4	8	4 or fewer axle single-trailer truck		any, 1.1~2.4, any	M, R
4	2	Class 2 pulling a trailer		0.3~3.1, **any**, any	M
4	3	Class 3 pulling a trailer		3.1~4.6, **any**, any	M
4	4	Bus pulling a trailer		7.2~30.4, any, any	A
4	4	Bus pulling a car		any, 5.2~30.4, 1.8~30.4	A
4	8	4 or fewer axle single-trailer truck		any, any, any	A
5	7	4 or more axle single-unit truck		any, 0.3~1.8, 0.3~1.8, 0.3~4	A
5	11	5 or fewer axle multi-trailer truck		any, 5.2~30.4, any, 1.8~30.4	M, R
5	9	5 axle single-trailer truck		any, 5.2~30.4, any, 1.1~3.4	A
5	9	5 axle single-trailer truck		any, **1.1~3.4**, any, 1.1~3.4	M, R
5	2	Class 2 pulling a trailer		0.3~**3.1**, any, 0.3~1.1, 0.3~1.1	M
5	3	Class 3 pulling a trailer		3.1~4.6, any, 0.3~1.1, 0.3~1.1	
5	9	5 axle single-trailer truck		any, any, any, any	A
6	7	4 or more axle single-unit truck		any, 0.3~1.8, 0.3~1.8, 0.3~1.8, 00.3~4	A
6	10	6 or more axle single-trailer truck		any, 0.3**~2.4**, 0.3**~2.4**, any, 2.4~30.4	M
6	12	6 axle multi-trailer truck		any, any, any, any, 2.4~30.4	
6	10	6 or more axle single-trailer truck		any, any, any, any, 0.3**~2.4**	M
7	7	4 or more axle single-unit truck		any, any, 0.3~1.8, any, any	A
7	10	6 or more axle single-trailer truck		any, any, any, 0.3**~2.4**, 0.3**~2.4**	M
7	13	7 or more axle multi-trailer truck		any, any, any, any, any, any	A
2.4	10	6 or more axle single-trailer truck		any, 0.3~2.4, 0.3~2.4, any, 0.3~2.4,	A
8	13	7 or more axle multi-trailer truck		any, any, any, any, any, any, any	A
9+	13	7 or more axle multi-trailer truck		any, any, any, any, any, any, any, any	A
any	14	Unclassified vehicle		others	A

M: Modified step from ODOT decision tree (changes are highlighted with bold text)
R: Reordered step from ODOT decision tree
A: Newly added step.

Table 2-7, Comparison between the pvr axle-class and ground truth at I-70 (development set) using (a) the original ODOT classification decision tree, and (b) the new decision tree.

(a)

I-70 from ODOT classifier		Axle based vehicle classification												
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13
FHWA 13 vehicle classification	class 1: Motorcycle	31	-	-	-	-	-	-	-	-	-	-	-	-
	class 2: Car	-	3663	~~21~~	-	-	-	(2)	-	-	-	-	-	1
	class 3: other 2axle, 4tire single-unit veh	~~1~~	~~2278~~	1500	-	~~1~~	-	-	~~20~~	-	-	-	-	-
	class 4: Bus	-	-	~~1~~	2	~~11~~	-	-	5	-	-	-	-	-
	class 5: 2 axle, 6tire, single-unit truck	-	~~1~~	~~105~~	~~3~~	136	-	-	-	-	-	-	-	-
	class 6: 3 axle single-unit truck	-	-	-	~~1~~	-	95	(4)	-	-	-	-	-	1
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	(1)	1	2	9	39	-	-	-
	class 8: 4 or fewer axle single-trailer truck	-	-	~~12~~	-	-	(1)	1	55	-	-	-	-	-
	class 9: 5 axle single-trailer truck	-	(1)	-	-	-	(15)	-	(10)	1236	(1)	1	-	-
	class 10: 6 or more axle single-trailer truck	-	-	-	-	-	(1)	-	(1)	(6)	29	-	-	-
	class 11: 5 or fewer axle multi-trailer truck	-	-	-	-	(1)	-	-	(2)	-	-	46	-	-
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	14	-
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1

= Misclassifications from overlapping range of axle spacing (e.g., case “b”, “c”, “f”),

( ) Misclassifications from missing axles (e.g., case “d”),

Misclassifications from the given decision tree

Table 2-7, continued

(b)

I-70 from new axle-based classification decision tree		Axle based vehicle classification												
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13
FHWA 13 vehicle classification	class 1: Motorcycle	31	-	-	-	-	-	-	-	-	-	-	-	-
	class 2: Car	-	3663	~~21~~	-	-	-	(2)	(1)	-	-	-	-	-
	class 3: other 2axle, 4tire single-unit veh	~~1~~	~~2278~~	1499	-	~~1~~	-	-	~~21~~	-	-	-	-	-
	class 4: Bus	-	-	~~1~~	7	~~11~~	-	-	-	-	-	-	-	-
	class 5: 2 axle, 6tire, single-unit truck	-	~~1~~	~~102~~	~~2~~	140	-	-	-	-	-	-	-	-
	class 6: 3 axle single-unit truck	-	-	-	~~1~~	-	95	(4)	1	-	-	-	-	-
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	(1)	51	-	-	-	-	-	-
	class 8: 4 or fewer axle single-trailer truck	-	-	~~10~~	-	-	(1)	-	58	-	-	-	-	-
	class 9: 5 axle single-trailer truck	-	(1)	-	-	-	(12)	-	(13)	1237	(1)	-	-	-
	class 10: 6 or more axle single-trailer truck	-	-	-	-	-	(1)	-	(1)	(6)	29	-	-	-
	class 11: 5 or fewer axle multi-trailer truck	-	-	-	(2)	(1)	-	-	-	-	-	46	-	-
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	14	-
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1

Note: In Table 2-7(b), the one in the black cell turns out to be class 6 but the axle configuration is not typical of the other class 6 vehicles observed.

Table 2-8, Comparison between the pvr axle-class and ground truth at I-270 (evaluation set) using (a) the original ODOT classification decision tree, and (b) the new decision tree.

(a)

I-270 from ODOT classifier		Axle based vehicle classification												
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13
FHWA 13 vehicle classification	class 1: Motorcycle				-	-	-	-	-	-	-	-	-	-
	class 2: Car	-	5881	1104	-	-	-	-	-	-	-	-	-	26
	class 3: other 2axle, 4tire single-unit veh				-	-	-	1	~~2~~	-	-	-	-	-
	class 4: Bus	-	-	-	1	~~6~~	-	-	3	-	-	-	-	-
	class 5: 2 axle, 6tire, single-unit truck	-	~~6~~	~~81~~	~~4~~	111	-	-	-	-	-	-	-	-
	class 6: 3 axle single-unit truck	-	-	-	-	-	71	-	-	-	-	-	-	-
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	(1)	9	-	2	21	-	-	-
	class 8: 4 or fewer axle single-trailer truck	-	-	~~8~~	~~1~~	-	-	4	29	-	-	-	-	1
	class 9: 5 axle single-trailer truck	-	(8)	~~1~~	-	-	-	-	(4)	616	-	-	-	-
	class 10: 6 or more axle single-trailer truck	-	(2)	-	-	-	-	-	(1)	-	13	-	-	1
	class 11: 5 or fewer axle multi-trailer truck	-	(1)	-	-	-	-	-	-	-	-	20	-	-
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	8	-
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1

= Misclassifications from overlapping range of axle spacing (e.g., case "b", "c", "f"),

( ) Misclassifications from missing axles (e.g., case "d"),

Misclassifications from the given decision tree

Table 2-8, continued

(b)

I-270 from new axle-based classification decision tree		Axle based vehicle classification												
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13
FHWA 13 vehicle classification	class 1: Motorcycle				-	-	-	-	-	-	-	-	-	-
	class 2: Car	-	5907	1106	-	-	-	-	-	-	-	-	-	-
	class 3: other 2axle, 4tire single-unit veh				-	-	-	-	~~1~~	-	-	-	-	-
	class 4: Bus	-	-	-	4	~~6~~	-	-	-	-	-	-	-	-
	class 5: 2 axle, 6tire, single-unit truck	-	~~6~~	~~81~~	~~4~~	111	-	-	-	-	-	-	-	-
	class 6: 3 axle single-unit truck	-	-	-	-	-	71	-	-	-	-	-	-	-
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	(1)	32	-	-	-	-	-	-
	class 8: 4 or fewer axle single-trailer truck	-	-	~~8~~	~~1~~ 1	-	-	-	33	-	-	-	-	-
	class 9: 5 axle single-trailer truck	-	(8)	-	-	-	-	-	(4)	617	-	-	-	-
	class 10: 6 or more axle single-trailer truck	-	(2)	-	-	-	-	-	(1)	-	14	-	-	-
	class 11: 5 or fewer axle multi-trailer truck	-	(1)	-	-	-	-	-	-	-	-	20	-	-
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	8	-
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1

Note: In Table 2-8(b) the one in the black cell turns out to be class 5 pulling a car but the new axle-based classification decision tree classified it as a bus pulling a car.

Table 2-9, Comparison between the pvr axle-class and ground truth at I-70 (development set) using (a) the original ODOT classification decision tree, and (b) the new decision tree.

(a)

I-70 from ODOT classifier		pvr axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	7494	3	21	99.7%
	SUT	107	254	56	60.9%
	MUT	13	19	1402	97.8%
% of column correct		98.4%	92.0%	94.8%	97.7%

(b)

I-70 from new decision tree		pvr axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	7494	3	22	99.7%
	SUT	104	312	1	74.8%
	MUT	11	17	1406	98.0%
% of column correct		98.5%	94.0%	98.4%	98.3%

Table 2-10, Comparison between the pvr axle-class and ground truth at I-270 (evaluation set) using (a) the original ODOT classification decision tree, and (b) the new decision tree.

(a)

I-270 from ODOT classifier		pvr axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	6985	1	28	99.6%
	SUT	87	203	26	64.2%
	MUT	20	5	694	96.5%
% of column correct		98.5%	97.1%	92.8%	97.9%

(b)

I-270 from new decision tree		pvr axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	7013	0	1	100%
	SUT	87	229	0	72.5%
	MUT	19	2	698	97.1%
% of column correct		98.5%	99.1%	99.9%	98.6%

(This page intentionally left blank)

3 SIDE-FIRE LIDAR BASED VEHICLE CLASSIFICATION

3.1 Introduction

Vehicle classification data are used in many transportation applications, including: pavement design, environmental impact studies, traffic control, and traffic safety [1]. There are several classification methods, including: axle-based (e.g., pneumatic tube and piezoelectric detectors), vehicle length-based (e.g., dual loop and some wayside microwave detectors), as well as emerging machine vision based detection. As noted by the Traffic Monitoring Guide [1], each sensor technology has its own strengths and weaknesses regarding costs, accuracy, performance, and ease of use.

In the present study we add another technology to the mix and develop a vehicle classification algorithm for LIDAR (Light detection and ranging) sensors mounted in a side-fire configuration. Our prototype system consists of two LIDAR sensors mounted on the driver's side of a probe vehicle parked alongside the roadway. Each LIDAR scans a vertical plane across the roadway, providing a rich view of the passing vehicles. In practice, the LIDAR sensors could be mounted on a temporary deployment platform like this system, or permanently mounted on a pole adjacent to the roadway.

To classify vehicles, first we segment them from the background, next we look for possible occlusions using algorithms developed herein, and then we measure several features of size and shape for each vehicle. These features are subsequently used for classification into six categories. The classification algorithm is evaluated by comparing the individual vehicle results against concurrent video. Occlusions are inevitable in this proof of concept study since the LIDAR sensors were mounted roughly 6 ft above the road, well below the tops of many vehicles. The present work focuses primarily on the non-occluded vehicles. Ultimately we envision using a combination of a higher vantage point in future work (similar to wayside microwave detectors), and shape information (begun herein) to greatly reduce the impacts of occlusions.

LIDAR technology has been applied in various transportation applications, such as highway safety [12-13] and highway design [14-15]. There have been a few demonstrations of LIDAR or related optical range finding technologies to monitor traffic and sometimes classify the vehicles. The most notable example being the Schwartz Autosense [16], which consisted of a sensor mounted over the lane of travel; though this basic approach pre-dates the Autosense system [17]. While the overhead view eliminates occlusions, the need to mount the sensor over the roadway makes deployment more difficult. Others have contemplated using airborne LIDAR platforms for traffic monitoring [18-19]. For example, [19] collected LIDAR imagery data over transportation corridors, segmented individual vehicles from the road surface, and then extracted six parameters of vehicle shape and size for each vehicle. They classified vehicles in three categories (passenger vehicles, multi-purpose vehicles, and trucks) using principle component analysis. Finally, our group has also contemplated the use of LIDAR to classify vehicles from a moving platform [20-21].

The remainder of this chapter is organized as follows. First the process of collecting the LIDAR data and the procedure of segmenting the vehicles from the background are presented. Next, the LIDAR based vehicle classification algorithm is developed. Third, the algorithm is evaluated on a per-vehicle-basis against concurrent video ground truth from field data at six directional locations, exhibiting various traffic conditions, distance between LIDAR and target vehicles, and road type (freeway and arterial road). The evaluation dataset includes over 25,000 vehicles (23,000 non-occluded). Then, the chapter closes with conclusions.

3.2 LIDAR Measurements and Vehicle Detection

Figure 3-1(b) shows an overhead schematic of the prototype deployment. The two LIDAR sensors are each mounted at a height of about 6.7 ft above ground and they are 4.6 ft apart from one another. Each LIDAR sensor scans a vertical plane across the roadway at roughly 37 Hz. Each scan sweeps 180°, returning the distance to the nearest object (if any) at 0.5° increments with a ranging resolution of 0.1 inch and a maximum range of 262 ft. So each scan returns 361 samples in polar coordinates (range and angle) relative to the LIDAR sensor and these data are transformed into a Cartesian coordinate system (lateral distance and relative height) for analysis.

Using these LIDAR data, vehicle segmentation is split into two steps. First we distinguish between vehicle returns and non-vehicle returns (e.g., pavement, foliage, barriers, etc.). Then we cluster the vehicle returns into discrete vehicles.

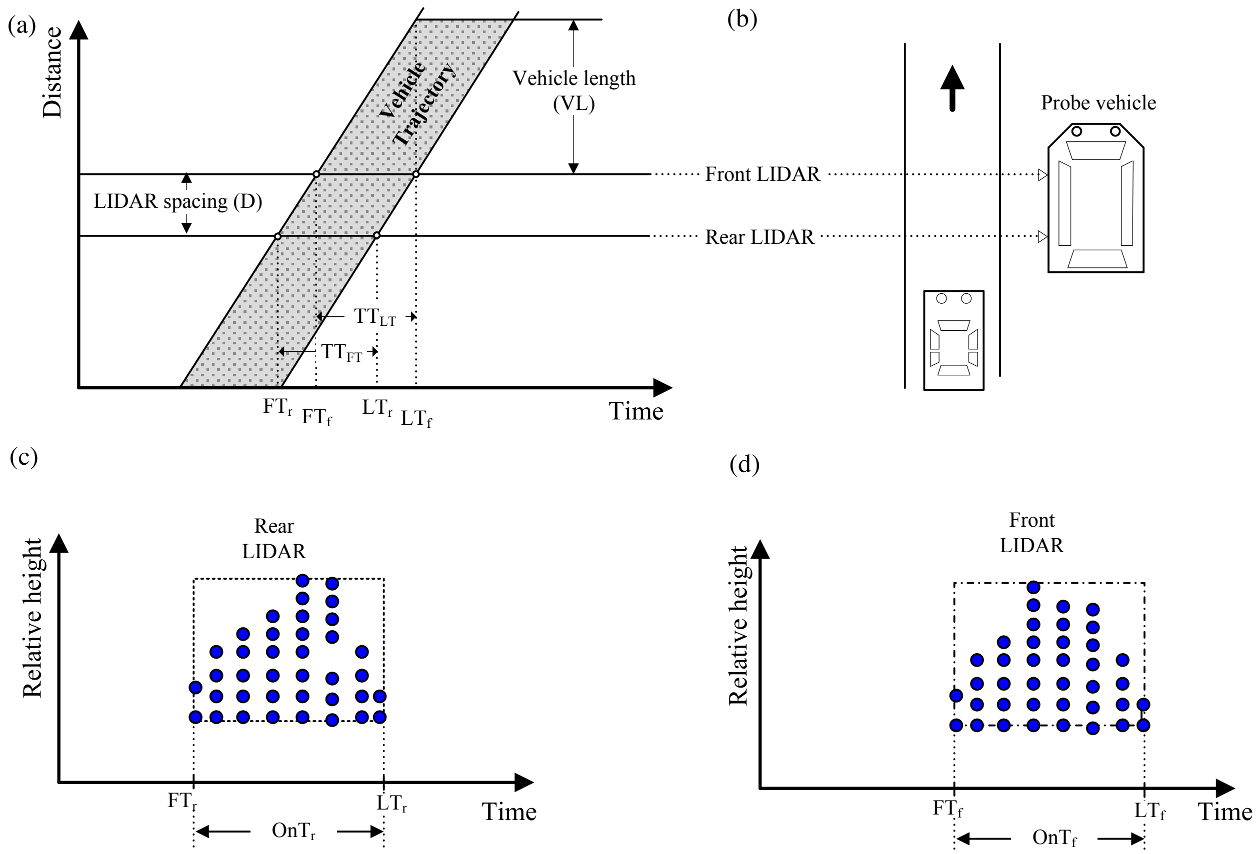


Figure 3-1, A hypothetical example of a vehicle passing by the two side-fire LIDAR sensors: (a) in time-space place; (b) a top-down schematic of the scene; and the corresponding returns from the vehicle from (c) the rear LIDAR sensor and (d) the front LIDAR sensor.

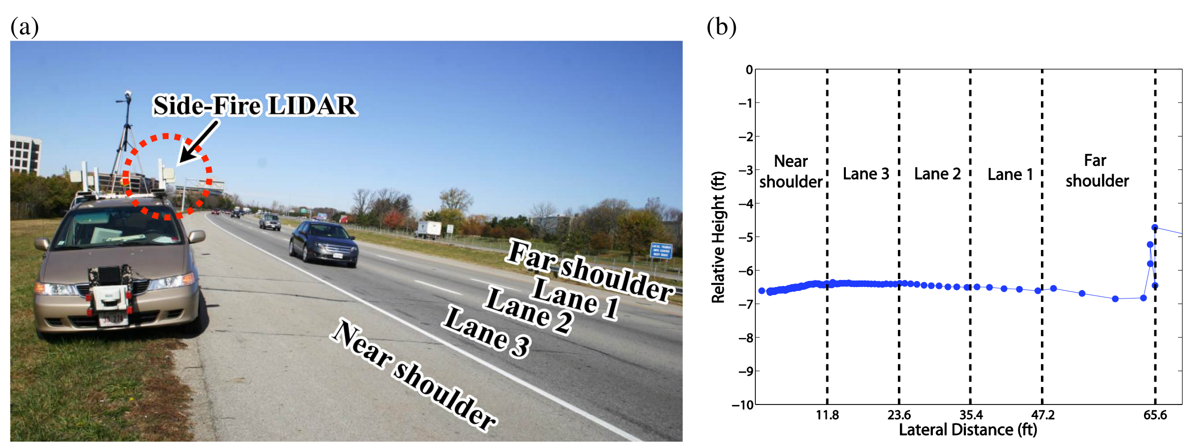


Figure 3-2, (a) the LIDAR data collection on I-270 southbound, on the west side of Columbus, Ohio; and (b) the corresponding background curve extracted from the data.

To segment the vehicle and non-vehicle returns we adapt background subtraction techniques from conventional image processing. The LIDAR are fixed, so when no vehicles are present they will return nearly identical scans of the *background*. Thus, the background's range at a given angle is (roughly) constant and over time the background returns are the dominant range reported at each angle. Formalizing this concept to extract the background from the LIDAR data, we set the background equal to the median range at each angle as observed over an extended time period with largely free flowing traffic. Whenever a vehicle is present, the vehicle's returns can only be at a range that is closer than the background range for the given angle. So in the absence of free flowing traffic, one could instead take the distribution of observed ranges at a given angle and set the background equal to the furthest mode of the distribution.

Figure 3-2(a) shows the data collection on I-270 southbound, on the west side of Columbus, Ohio. The probe vehicle was parked just off of the right hand shoulder to collect LIDAR data. Figure 3-2(b) shows the corresponding background that was extracted from the LIDAR data. Because the van would occasionally roll a small amount about its central axis as personnel entered or exited the vehicle during the data collection, returns from the background did not always fall on the measured background curve. All returns falling beyond the background curve as well as any returns that were within an inch above the background curve were considered non-vehicle returns and excluded from further analysis. However, if the low-lying returns prove critical to a subsequent application, one could estimate the LIDAR's instantaneous angle relative to the shoulders (0-11.8 ft, and 47.2-65.6 ft in Figure 3-2(b)) and normalize this angle across scans.

Only vehicle returns should remain after removing the background, however, these returns still need to be clustered into individual vehicles and we take the following steps to do so. First we establish the lane boundaries by looking at the distribution of the lateral distance across the vehicle returns. We expect to see one distinct mode per travel lane, corresponding to the near side of the vehicles when traveling in the given lane since the vertical edges on the vehicles will generally yield many returns at the same lateral distance; though, there will be other returns in the distribution from horizontal vehicle surfaces, vehicles changing lanes, and so forth. Provided the LIDAR sensors are not moved, this step only needs to be done once, using a few minutes of data.

Second, in each scan we segment the LIDAR returns by lane using the lane boundaries from the previous step. As long as a vehicle travels within a lane, all of the returns from that vehicle will fall between the respective lane boundaries in the given scan. In most cases even a single return in the lane will be taken as that lane being occupied in that scan. However, in the relatively rare cases when a vehicle changes lanes as it passes the LIDAR, that vehicle's returns may fall into two adjacent lanes (we saw this event occur 253 times out of 27,450 vehicles). To find the cases when a single vehicle is seen in adjacent lanes, we explicitly look for concurrent returns in neighboring lanes. When this occurs, we take the mode of lateral distance in the near lane and the far lane, respectively. Again, the nearside of a vehicle is characterized by a large number of returns at a given lateral distance, i.e., the mode lateral distance within the lane. If in the given scan the difference between the modes in successive lanes is less than the maximum feasible vehicle width (set to 8.5 ft, the maximum width of commercial motor vehicles [22]), the vehicle returns in the adjacent lanes are assumed to come from a single vehicle and are grouped together in the lane corresponding to the median lateral distance among the set of returns in question. Otherwise, the two modes are too far apart to come from a single vehicle and the groups are kept separate. Obviously this approach assumes that at most one vehicle can occupy a lane in a given scan; although we know that it is not always the case, e.g., when two motorcycles pass side by side within a lane, we have yet to observe any such exceptions in the LIDAR data so addressing these exceptions is left to future research.

Third, taking the temporal sequence by lane, the returns are clustered into vehicles. After each scan is processed, whenever a given lane is occupied, if there is not already an open vehicle cluster in that lane then a new vehicle cluster is begun with the corresponding returns; otherwise, the corresponding returns are added to the open vehicle cluster in that lane. On the other hand, if there is an open vehicle cluster and the lane has not been occupied for at least 1/4 sec (roughly 9 scans) then the open vehicle cluster is closed. To be retained, a closed vehicle cluster must span at least two scans and at least two of the scans must have different heights, otherwise, the vehicle cluster is discarded. Because the returns in a scan are grouped by lane before the clustering step and we make the above correction for vehicles changing lanes, it is theoretically possible for two neighboring vehicles to be erroneously clustered together. Though we have not seen this problem occur, to safeguard against it, if the net width of a closed cluster is greater than the maximum feasible vehicle width then the cluster is split in two, by lane. On the other hand, it is possible for a vehicle changing lanes to be assigned to different lanes at different time steps, resulting in separate clusters in each lane. To catch these breakups, when a cluster ends in one lane, we check the next scan to see if a new cluster begins in an adjacent lane a small distance away, i.e., if the difference between the mode lateral distance is less than 1.1 ft, the two clusters are merged together and assigned to the lane with the larger cluster. The segmentation and clustering steps are repeated for each lane across each successive LIDAR scan until all of the vehicle returns have been clustered into discrete vehicles.

3.2.1 Occlusion Reasoning

A key step in classifying a given vehicle is determining whether the entire vehicle was seen or if there was evidence of a partial occlusion. Table 3-1 shows that the latter case arose for about 12% of the vehicles observed on the multilane facilities. The frequency is so small because the spacing between vehicles is typically much larger than one might think, e.g., according to the HCM [23], LOS F on a freeway begins at 46 passenger cars per mile per lane or 117 ft per passenger car and passenger cars are generally on the order of 10-20 ft long. In any event, partially occluded vehicles are likely to be misclassified in our algorithm if the occlusion is not identified and handled separately from the non-occluded vehicles. Of course from the LIDAR data stream we cannot detect completely occluded vehicles, though we found these errors occurred between 3-6% in the three multilane datasets that had an independent detector to monitor occluded lanes (I-71 and I-270 in Table 3-1), and as one might expect, most of these occluded vehicles were passenger vehicles. A higher vantage point or using a second set of LIDAR to also monitor from the median of the roadway should reduce the frequency of completely occluded vehicles.

For any given vehicle cluster we suspect a partial occlusion occurred unless we see at least one non-vehicle return on all sides of the cluster (both temporally and spatially). To automatically detect partially occluded vehicles, first we check the vehicles seen in each scan of the LIDAR. If we cannot see the background curve between a given pair of vehicles the further vehicle is suspected of being partially occluded by the closer vehicle. Second, we check successive scans, if one vehicle is seen at a given angle in scan i, and a different vehicle is seen at the same angle in scan i+1, whichever vehicle cluster is further away is considered to be partially occluded.

3.3 LIDAR Based Vehicle Classification Algorithm

In this section we develop an algorithm to classify the vehicle clusters extracted from the LIDAR data in the previous section. The core algorithm focuses on the non-occluded vehicles and sorts them into six vehicle classes: motorcycle (MC) - axle class 1, passenger vehicle (PV) - axle class 2-3, PV pulling a trailer (PVPT) - axle class 2-3, single-unit truck/bus (SUT) - axle class 4-7, SUT pulling a trailer (SUTPT) - axle class 8-13, and multi-unit truck (MUT) - axle class 8-13. Note that the distribution of axle class to group differs slightly depending on whether we use three groups (Chapter 2), six groups (Chapter 3), or four groups (Chapter 3 & 4). These classes are a refinement of commonly used length-based classes (as noted in [1], a user might not need the full 13 axle-based classes and three or four simple categories may suffice). After classifying the non-occluded vehicles we separately handle the partially occluded vehicles, taking care to address the uncertainty about what went unobserved.

We derived the vehicle classification algorithm using a ground truth development dataset that consists of 24 min of free flow data collected across four lanes on I-71 southbound in Columbus, Ohio, between 11th Ave and 17th Ave on July 9, 2009. There were 1,502 non-occluded vehicles in this dataset and all of the vehicle classifications were manually verified from the video ground truth data. The two primary vehicle features used by the classification algorithm are length and height measured from the individual vehicle clusters, as shown in Figure 3-3(a). Compared to using length alone, as would be done from loop detectors (see, e.g., [2]), vehicle height helps separate different vehicle classes (e.g., SUT and PVPT). However, the boundaries of various classes still overlap in the length-height plane. To segregate vehicles pulling trailers we calculate up to six additional measurements of the vehicle's shape (for a total of eight shape measurements), as enumerated below and explained in the following subsections.

- Vehicle length (VL)
- Vehicle height (VH)
- Detection of middle drop (DMD)
- Vehicle height at middle drop (VHMD)
- Front vehicle height (FVH)
- Front vehicle length (FVL)
- Rear vehicle height (RVH)
- Rear vehicle length (RVL)

Table 3-1, Summary of LIDAR data collected to evaluate the algorithm and the performance of the algorithm by each dataset.

Data type	Road type	Location (direction)	Number of lanes	Date	Time period (Start time ~ End time)	Duration (hr: min)	Distance between LIDAR sensor and the nearest travelled lane (ft)	Average of the LIDAR speeds over the duration (mph)	Number of vehicles seen by LIDAR	Number of partially occluded vehicles	Number of totally occluded vehicles	Performance of the algorithm: Success	Performance of the algorithm: Errors	% errors
Development	Freeway	I-71 (SB)	4	July 9, 2009	18:09 ~ 18:33	0:24	58	63	1,813	311	65	1,494	8	0.5%
Evaluation	Freeway	I-71 (SB)	4	Nov 19, 2009	07:41 ~ 08:09	0:28	58	47	2,619	591	145	2,021	7	0.3%
		I-270 (SB)	3	Nov 2, 2010	09:29 ~ 14:29	5:00	15	65	13,397	1,376	422	11,934	87	0.7%
		SR-315 (NB)	2	Aug 12, 2010	14:57 ~ 17:57	3:00	2	41	6,900	660	n/a	6,230	10	0.2%
		Subtotal of Evaluation Freeway			-	8:28	-	-	22,916	2,627	567	20,185	104	0.5%
	Arterial Rd.	Dublin Rd (SB)	1	Oct 28, 2010	07:32 ~ 08:57 14:30 ~ 15:55	2:50	2	36	1,344	-	-	1,337	7	0.5%
		Wilson Rd (NB)	1	Oct 28, 2010	09:08 ~ 09:56 16:02 ~ 16:54	1:40	2	36	666	-	-	664	2	0.3%
		Wilson Rd (SB)	1	Oct 28, 2010	10:18 ~ 10:58 17:00 ~ 18:00	1:40	2	38	711	-	-	710	1	0.1%
		Subtotal of Arterial Rd.			-	6:10	-	-	2,721	-	-	2,711	10	0.4%
Evaluation data total					-	14:38	-	-	25,637	2,627	567	22,896	114	0.5%
Overall total					-	15:02	-	-	27,450	2,938	632	24,390	122	0.5%

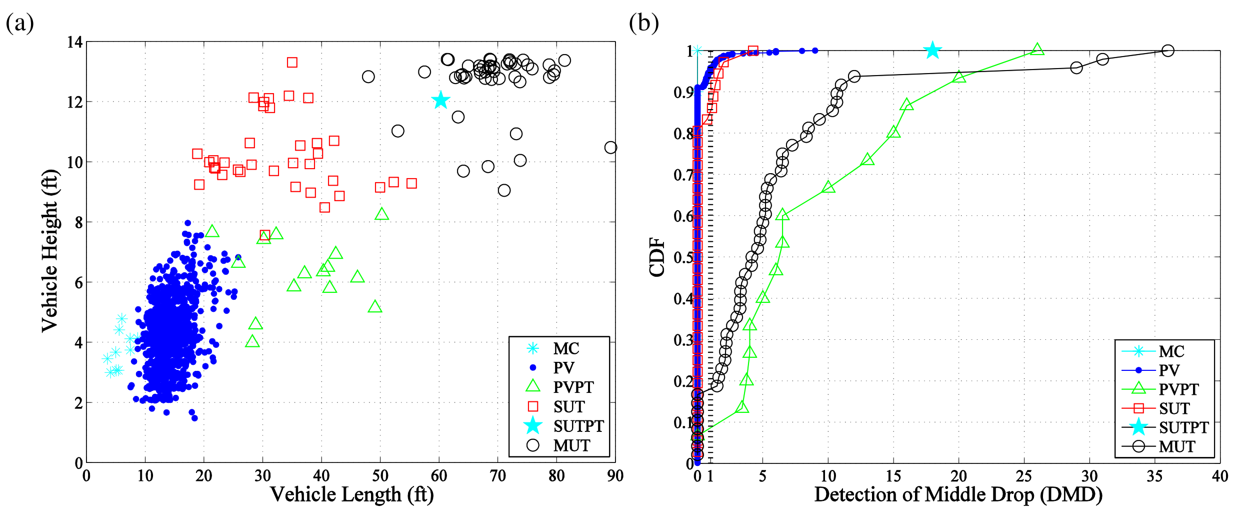


Figure 3-3, (a) A scatter plot of vehicle height and vehicle length of 1,502 non-occluded vehicles from the development dataset; and (b) the cumulative distribution of DMD for these vehicles.

3.3.1 Vehicle Length (VL) and Vehicle Height (VH)

The two side LIDAR sensors are mounted in a "speed-trap" configuration with 4.6 ft spacing. Any moving target will appear at different times in the two views, thereby allowing for speed measurement. Figure 3-1(a) shows a hypothetical example of the time-space diagram as a vehicle passes by the two LIDAR sensors and Figure 3-1(b) shows the corresponding schematic on the same distance scale. In this study a vehicle passes the rear LIDAR sensor first and then the front LIDAR sensor. Figure 3-1(c) and (d) show the vehicle returns from each of the two LIDAR sensors as the vehicle passes, where FT and LT respectively denote the first and last time samples in which the vehicle was scanned by the given LIDAR (subscript "r" for rear and "f" for front). OnT_f and OnT_r indicate the duration of time that a vehicle is scanned by the given LIDAR sensor, i.e., the on-time, where $OnT_x = LT_x - FT_x$, and x is either "r" or "f". Meanwhile, the traversal time is defined as the difference between the first scan time at the two sensors, i.e., $TT_{FT} = FT_f - FT_r$, or the last scan time, i.e., $TT_{LT} = LT_f - LT_r$. Speed is calculated via Equation (3.1) from the LIDAR spacing, D, and the traversal time. Vehicle length (VL) is calculated from the mean of V_{FT} and V_{LT}, multiplied by OnT_r (we arbitrarily select the rear LIDAR on-time in this study), yielding Equation (3.2). Finally, vehicle height (VH) is directly measured from the difference of the highest relative height and the lowest relative height across all of the returns in the given vehicle cluster from the rear LIDAR, yielding Equation (3.3). By using the difference in cluster heights, this step accounts for the fact that the road cross-section is not flat, each lane may be at a different height relative to the LIDAR sensor.

$$V_{FT} = \frac{D}{TT_{FT}}, \; V_{LT} = \frac{D}{TT_{LT}} \tag{3.1}$$

$$VL = \mathrm{mean}(V_{FT}, V_{LT}) \times OnT_r \tag{3.2}$$

$$VH = \max(h(t)) - \min(h(t)), \forall\, t \in [FT_r, LT_r], \forall\, h(t) \in \text{cluster} \tag{3.3}$$

where h(t) is height of a LIDAR return relative to a height of LIDAR sensor at time t.

Figure 3-3(a) shows a scatter plot of vehicle height versus vehicle length for the 1,502 non-occluded vehicles from the development dataset sorted by the six vehicle classes. The VH for almost all of the MC, PV and PVPT are below 8 ft, while VH for almost all of the SUT, SUTPT, and MUT are above 8 ft. As will be discussed shortly, the height of the trailer (or its load) is sometimes the tallest point on a PVPT or SUTPT and thus is reflected in VH for that vehicle. The observed VL are distributed between 5 ft and 89 ft, with a clear but overlapping progression from MC to PV to PVPT, and similarly from SUT to SUTPT to MUT. Based on this plot, we select VL = 7.5 ft as the dividing line between MC and PV. To segregate the remaining classes, we look for a characteristic "gap" before the start of a trailer (PVPT, SUTPT, and MUT) as follows.

3.3.2 Detection of a Middle Drop in a Vehicle (DMD)

The vertically scanning LIDAR captures the profile shape of the passing vehicles. This profile is useful to distinguish between vehicle classes with overlapping VL and VH ranges, e.g., SUT and MUT. For vehicles in these ranges, we look for the presence of a gap that is indicative of the start of a trailer, as manifest as one or more scans with a "drop" in the number of returns somewhere in the middle of the vehicle cluster. To determine whether a vehicle has such a *middle drop*, we first tally the number of LIDAR returns, nLR, as a function of each scan (i.e., time step) that the vehicle cluster was seen, yielding nLR(t). For example, Figure 3-4(a) shows the image of a pickup truck pulling a trailer (an example of PVPT) as it passes by the LIDAR sensors while Figure 3-4(b) shows the corresponding LIDAR returns from the vehicle cluster. Figure 3-4(c) shows the nLR(t) curve for the vehicle cluster. The curve does a good job highlighting the point where the trailer is connected to the pickup truck via the low nLR(t). Note that we deliberately use nLR(t) rather than the height of the vehicle because there are some trailers that have a return near the top of the gap even though most of the gap is open (e.g., tree trimming trucks).

Formalizing the process, once the nLR(t) curve is obtained, the set of local minimum points on the curve are considered as potential locations of a middle drop in the vehicle's shape, where $nLR(t^*_i)$ denotes the i-th minima. Since the middle drop should correspond to relatively few LIDAR returns in the given scan (but not necessarily zero due to the connecting link, e.g., the hitch in Figure 3-4(a)), we assume that nLR at a middle drop must be less than the average of nLR(t) for the cluster across all times, $\overline{nLR}$. So, we ignore i-th local minimum if it is greater than $\overline{nLR}$. Formalizing this process, a given scan is considered a possible middle drop if it satisfies all of the conditions in Equation (3.4).

(a)

(b)

(c)

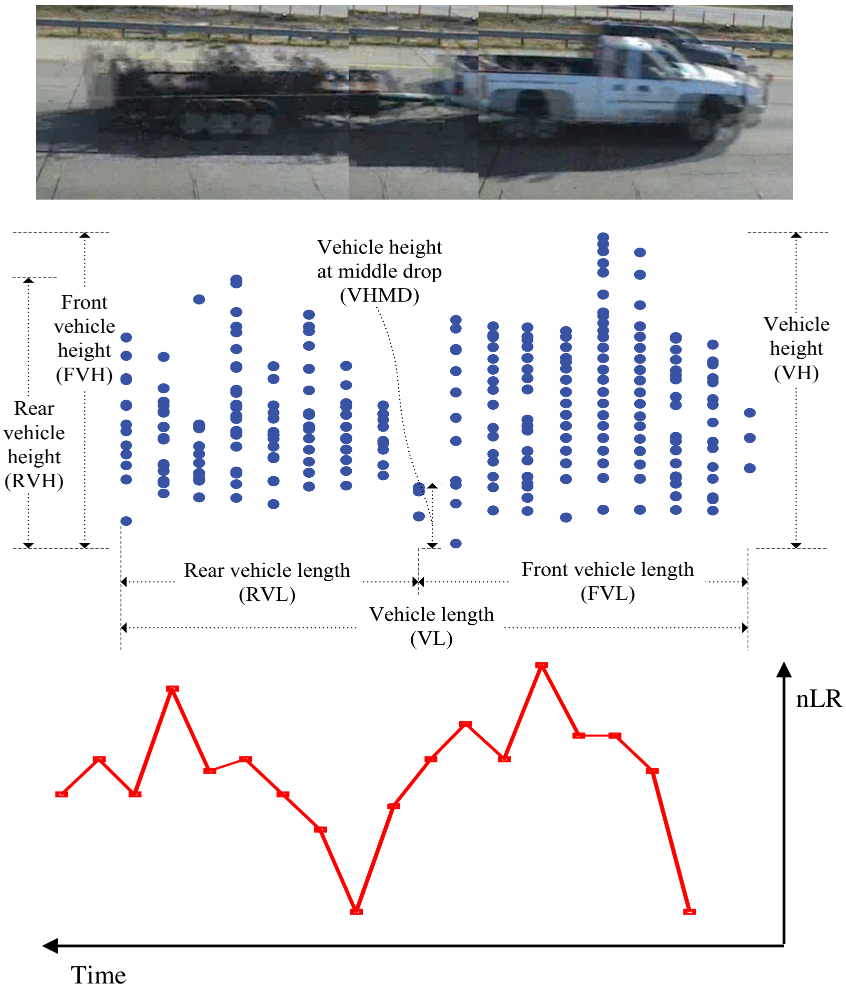


Figure 3-4, (a) A pickup truck pulling a trailer; (b) the corresponding vehicle cluster of returns and the various measurements used for vehicle classification; and (c) the number of returns by scan, capturing the vehicle shape. Note that time in Figure 3-4(c) is increasing to the left in this plot because the front of the vehicle is seen first and the vehicle orientation is presented consistent with the rest of the figure.

$$
\begin{aligned}
&nLR(t_i^*) < nLR(t_i^* - 1)\\
&nLR(t_i^*) < nLR(t_i^* + 1)\\
&nLR(t_i^*) < \overline{nLR}
\end{aligned}
\tag{3.4}
$$

For each minima at t^*_i, we take the difference of nLR(t) and $nLR(t^*_i)$ over all times, $t \in (FT_r, LT_r)$, denoted $\Delta n(t, t^*_i)$. We find $\max(\Delta n(t, t^*_i))$ over the α ft ahead of the scan at t^*_i (α = 4 ft in this study), add it to $\max(\Delta n(t, t^*_i))$ for α ft behind the scan and divide the sum by $nLR(t^*_i)$, yielding the *Sum of Relative Difference* (SRD) via Equation (3.5) at each t^*_i, i.e., $SRD(t^*_i)$. The use of distance rather than time is to make the algorithm robust to slow moving vehicles. Next we select the max $SRD(t^*_i)$ and call this value the *Detection of Middle Drop* (DMD) indicator, as expressed via Equation (3.6), and set t^* equal to the corresponding t^*_i. Figure 3-3(b) shows the cumulative distribution function of DMD for the 1,502 non-occluded vehicles by vehicle class in the development dataset. As expected PVPT, SUTPT and MUT have a wider range of DMD than MC, PV, and SUT. The latter three classes usually present zero DMD, indicative of a vehicle without a middle drop.

$$
\begin{aligned}
SRD(t_i^*) &= \frac{\Delta n(t_a, t_i^*) + \Delta n(t_b, t_i^*)}{nLR(t_i^*)}\\
&= \frac{nLR(t_a) + nLR(t_b) - 2 \times nLR(t_i^*)}{nLR(t_i^*)}
\end{aligned}
\tag{3.5}
$$

where,

$$
\begin{aligned}
t_a &= \max\left(t_i^* - \frac{\alpha}{mean(V_{FT}, V_{LT})}, FT_r\right)\\
t_b &= \min\left(t_i^* + \frac{\alpha}{mean(V_{FT}, V_{LT})}, LT_r\right)
\end{aligned}
$$

$$
DMD = \max(SRD(t_i^*)) \tag{3.6}
$$

Based on the distributions in Figure 3-3(b), if DMD < 1, the vehicle is presumed to be a single unit vehicle that is not pulling a trailer. Only 1 vehicle out of 16 vehicles pulling a trailer had DMD < 1 (a PVPT with zero DMD), or 6%. In addition, 40 out of 48 MUT (83%) had DMD > 1. Figure 3-3(b) also shows that 6% of PV and 15% of SUT had DMD > 1. From the development dataset, most of the vehicles with DMD < 1 can be correctly classified based on VL and VH. Correctly classifying the vehicles with DMD > 1 is the topic of the next section.

3.3.3 Additional Measurements of a Vehicle with Middle Drop

To correctly classify the vehicle clusters where DMD > 1, we segment a vehicle with middle drop into the front part of the vehicle (from the front bumper to the middle drop) and rear part of the vehicle (from the middle drop to the rear bumper). We then calculate the length of the front (FVL), height of the front (FVH), length of the rear (RVL), height of the rear (RVH), and the height of the vehicle at the middle drop (VHMD), as illustrated in Figure 3-4(b). Note that VH of a vehicle with middle drop corresponds to the maximum of FVH and RVH.

<u>Vehicle Height at Middle Drop (VHMD)</u>

The VHMD due to the hitch in PVPT or SUTPT should usually be lower than the VHMD due to the rear portion of a semi-trailer tractor in a MUT. We set a threshold height of the connection to be 2 ft. If VHMD is lower than this threshold the vehicle will be classified as either PVPT or SUTPT (depending on the FVL, discussed below). Otherwise, we need to check the other measurements to classify the vehicle. The VHMD is calculated via Equation (3.7) applied to the returns in the vehicle cluster.

$$
VHMD = \max(h(t^*)) - \min(h(t)), \forall t \in [FT_r, LT_r] \tag{3.7}
$$

Front Vehicle Height (FVH) and Front Vehicle Length (FVL)

The front part of a vehicle cluster with a true middle drop is either a PV, SUT, or the tractor of a MUT. As was shown in Figure 3-3(a), VH of SUT and MUT is usually higher than 8 ft, while VH of PV is usually lower than 8 ft. So we use FVH calculated via Equation (3.8) to capture height of the front portion of the cluster and if the height is below 8 ft, the vehicle is classified as PVPT. Otherwise, we need to check the other measurements to classify the vehicle.

$$FVH = \max(h(t)) - \min(h(t)),\ \forall t \in [FT_r, t^*] \tag{3.8}$$

In the case of PVPT or SUTPT the FVL calculated via Equation (3.9) is the VL of the PV or SUT portion of the cluster. From the development dataset we found the minimum length of the PV portion of the PVPT is above 15 ft. If the FVL is below 15 ft, we conclude that the middle drop is not due to a trailer and the vehicle is a single unit, PV or SUT.

$$FVL = V \times (t^* - FT_r) \tag{3.9}$$

Rear Vehicle Height (RVH) and Rear Vehicle Length (RVL)

The rear part of a vehicle with a true middle drop is trailer in a PVPT, SUTPT or MUT. If the RVH calculated via Equation (3.10) is sufficiently low (below 2.4 ft in the algorithm based on the development dataset), it is considered to be an empty flatbed trailer behind a PV or SUT and the complete cluster will be classified as either PVPT or SUTPT depending on the other measurements. If the RVH is sufficiently high (above 12 ft in the algorithm based on the development dataset), it is considered to be a semi-trailer and the complete cluster will be classified as a MUT. Otherwise, we need to check the other measurements to classify the vehicle. The trailer length is captured by RVL, Equation (3.11). If RVL is below 28 ft, we assume this trailer cannot come from a semi-trailer truck.

$$RVH = \max(h(t)) - \min(h(t)),\ \forall t \in [t^*, LT_r] \tag{3.10}$$

$$RVL = V \times (LT_r - t^*) \tag{3.11}$$

3.4 The LIDAR Based Vehicle Classification Algorithm

The eight shape measurements and various tests described above are combined into the LIDAR based classification decision tree shown in Figure 3-5. This figure shows our classification algorithm for non-occluded vehicles that we produced, based on the development dataset. As noted above, before applying this algorithm we automatically differentiate between non-occluded and partially occluded vehicles. For the latter group we cannot be as precise as Figure 3-5 for our classification, as follows.

3.4.1 Classifying Partially Occluded Vehicles

While some information is missing about the partially occluded vehicles, the intersection between the occluded and the occluder dimensions bound the size of the occluded vehicle, i.e., the size of the occluded part of the vehicle is no larger than the size of the occluder vehicle. Being careful not to double count scans where both the occluder and occluded are seen "overlapping", the length of the non-overlapping portion of the occluder vehicle is measured and the length of the occluded vehicle is bounded by Equation (3.12). Overlapping is not an issue for height, and the height of the occluded vehicle is bounded by Equation (3.13).

$$VL_o \le VL_o^{est} \le VL_o + VL_{c-NOL} \tag{3.12}$$

Where,

VL_o^{est} = estimation of unknown actual vehicle length of the occluded vehicle,

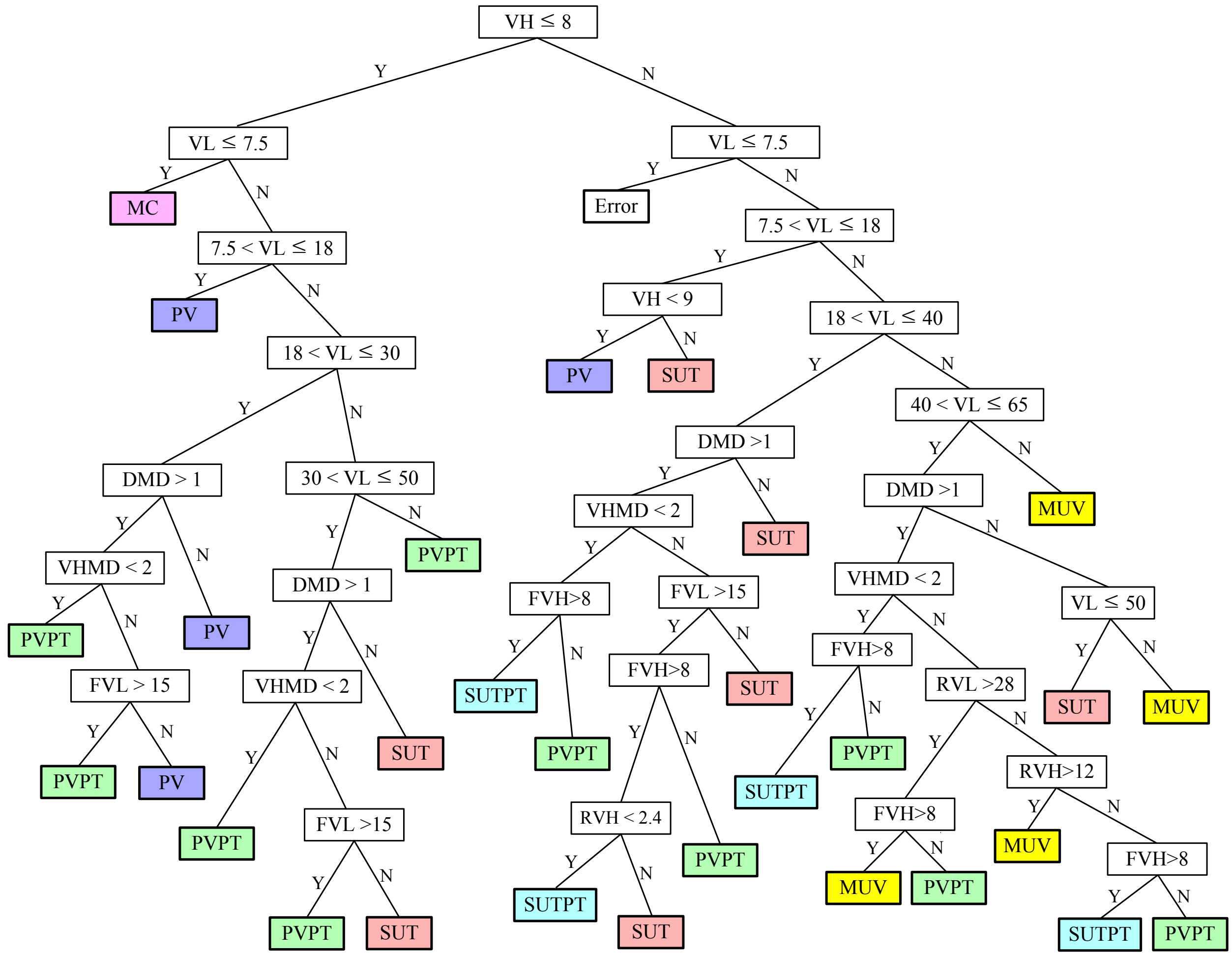
VL_o = vehicle length seen from the occluded vehicle,

$VL_{c\text{-}NOL}$ = vehicle length of non-overlapping portion of the occluder vehicle.

$$VH_o \le VH_o^{est} \le \mathrm{Max}(VH_o, VH_c) \tag{3.13}$$

Where,

VH_o^{est} = estimation of unknown actual vehicle height of the occluded vehicle,



Measurements	Type of vehicle
VH: LIDAR vehicle height (ft)	MC: Motorcycle
VL: LIDAR vehicle length (ft)	PV: Passenger vehicle
DMD: Detection of middle drop of a vehicle	SUT: Single-unit truck/bus
FVL: Front vehicle length (ft)	PVPT: PV pulling a trailer
FVH: Front vehicle height (ft)	SUTPT: SUT pulling a trailer
RVL: Rear vehicle length (ft)	MUV: Muti-unit truck
RVH: Rear vehicle height (ft)	
VHMD: Vehicle height at middle drop (ft)	

Figure 3-5, The decision tree underlying the non-occluded LIDAR based vehicle classification algorithm.

VH_o = vehicle height seen from the occluded vehicle,

VH_c = vehicle height from the occluder vehicle.

In this proof of concept study we only attempt to classify occlusions that involve two vehicles, though the principles could easily be extended to more complicated multi-vehicle occlusions. When classifying a partially occluded vehicle, the six classes are defined by static boundaries in the vehicle length and vehicle height plane, as shown in Figure 3-6 (compare to Figure 3-3). Because VL_o^{est} and VH_o^{est} each span a range, it is possible for a partially occluded vehicle to be associated with more than one class.

3.5 Evaluation of the LIDAR Based Vehicle Classification Algorithm

Thus far this research has used a single development dataset collected on July 9, 2009 to derive the classification algorithm. In addition to the development dataset, we collected three additional freeway datasets and three arterial datasets for evaluation. We used a total of just over 15 hrs of data: 24 min for development and the rest for evaluation. All of the datasets were collected in the Columbus metropolitan area. All locations were visited a single time in this study except for I-71, which we visited twice. The facility, number of lanes, date, time period, duration, and distance between the LIDAR sensors and travel lanes are shown in the first few columns of Table 3-1, while Appendix A provides further information about each site. The next four columns of Table 3-1 show the average speed over all vehicles seen in the data collection period, the number of vehicles seen, the number of vehicles that our algorithm labeled as partially occluded, and the number of totally occluded vehicles as counted by the detectors. Among the freeway datasets two come from free flow (5.4 hrs) and two from mild congestion (3.5 hrs). All of the data sets come from clear weather conditions.

Overall the algorithm suspected 2,938 out of 27,450 vehicles (11%) are partially occluded and these vehicles are excluded from the classification algorithm performance evaluation in Tables 3-1 and 3-2. Instead, we separately evaluate the classification performance on partially occluded vehicles at the end of this section. The highest rate of partially occluded vehicles occurred at the I-71 site on Nov 19, 2009 under mildly congested conditions (22.6%), while the lowest rate of partially occluded vehicles on the freeway segments occurred on SR 315 (9.6%). Not surprisingly, across the four freeway datasets the percentage of partially occluded vehicles increased as the number of lanes increased and at the I-71 location, as congestion increased (17.2% in free flow and 22.6% in mild congestion).

The vehicle class was manually reduced from the video ground truth data for all 27,450 vehicles in these datasets and the partial occlusions were verified at that time (see Chapter 4 for an example of the data reduction tool). We also ran the classification algorithm from Figure 3-5 on the datasets and the last three columns of Table 3-1 show the performance of the algorithm against the ground truth data. The errors are tallied on a per-vehicle basis, and thus, are not allowed to cancel one another across vehicles. Collectively, the algorithm correctly classifies 24,390 out of 24,512 non-occluded vehicles (99.5%) and misclassifies 122 vehicles (0.5%). The error rate was low across all seven datasets taken separately, the largest error rate was only 0.7%. The distance between the LIDAR and the roadway does not appear to have a large effect even though the further away a target vehicle is the smaller portion of the LIDAR field of view it occupies (and thus, the fewer angles in a LIDAR scan that provide vehicle returns). Among the freeway datasets the performance appears to degrade slightly as the average speed increases due to the 37 Hz sampling rate, but with only four datasets, the number is not large enough to draw any firm conclusions.

Table 3-2 shows the classification results by class against the ground truth data for all six evaluation datasets combined (see Appendix E for the results by station). The cells on the diagonal tally the number of vehicles where the LIDAR classification is the same as the ground truth classification, while the off-diagonal cells tally the incorrect vehicle classifications. The final row indicates the percentage correct among the vehicles assigned the given classification by the algorithm, while the second to the last column indicates percentage correct among the vehicles from the given class in the ground truth data. The last column tallies the number of partially occluded vehicles-by-class that are excluded from the non-occluded LIDAR based vehicle classification. Often an operating agency will group PVPT with PV and SUTPT with MUT, for reference, these supersets are shown in the table, denoted PV* and MUT*, respectively. If using the two supersets, 14% of the errors (16 vehicles) in Table 3-2 and 16% of the errors (20 vehicles) in Table 3-1 would be eliminated. Overall, the algorithm correctly classified a total of 22,896 out of 23,010 vehicles (99.5%) in the evaluation datasets.

Table 3-2, Comparison of LIDAR based vehicle classification and actual vehicle class from the six evaluation ground truth datasets.

From six evaluation datasets			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**31**	3	0	0	0	0	34	91.2%	4
	PV*	PV	10	**20,762**	3	15	0	0	20,790	99.9%	2,366
		PVPT	0	2	**192**	6	3	1	204	94.1%	25
	SUT		0	30	4	**688**	4	2	728	94.5%	61
	MUT*	SUPT	0	0	6	2	**31**	6	45	68.9%	2
		MUT	0	0	3	9	5	**1,192**	1,209	98.6%	169
Number of vehicles from LIDAR vehicle classification			41	20,797	208	720	43	1,201	23,010	99.5%	2,627
% correct			75.6%	99.8%	92.3%	95.6%	72.1%	99.3%	99.5%		

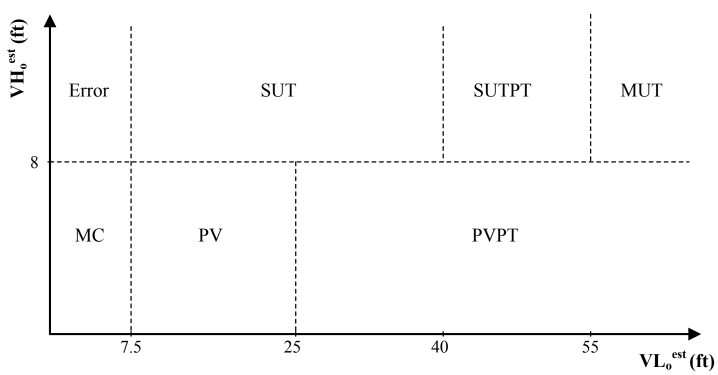


Figure 3-6, The classification space for partially occluded vehicles.

The most common errors are between PV and SUT because the length and height ranges of these vehicles overlap (30 SUT misclassified as PV and 15 PV misclassified as SUT), accounting for 39% of all errors. Also of note, we see 10 PV misclassified as MC. All of these PV were confirmed to have exceptionally short length, e.g., a 7.5 ft long commuter car (Smart Car). As with PV/SUT the MC/PV problem arises because the length and height ranges overlap between the two classes (3 MC were also misclassified as PV). This problem is not unique to LIDAR, the relatively new commuter cars will likely degrade the performance of most classification technologies when segregating MC. However, with the higher vantage point envisioned in our future research, the LIDAR should also be able to measure vehicle width, which should distinguish MC from commuter cars.

Finally, the algorithm for classifying partially occluded vehicles was applied to 1.5 hrs of the I-270 dataset. There were 465 partially occluded vehicles detected and of these, 219 are placed into a single feasible class (47% of partially occluded vehicles) and only six of these (3%) are incorrectly classified. The remaining 246 partially occluded vehicles are assigned two more feasible vehicle classes. Within this set, 34 (14%) were assigned all six classes. Out of the remaining 212 vehicles, 96% had the correct class among the two or more classes assigned to the given vehicle.

3.6 Conclusions

This chapter developed and tested a side-fire LIDAR based vehicle classification algorithm. The algorithm includes up to eight different measurements of vehicle shape to sort vehicles into six different classes. The algorithm was tested over seven datasets collected at various locations (including one development dataset). The results were compared against the concurrent video-recorded ground truth data on a per-vehicle basis. Overall, 2,938 out of 27,450 vehicles (11%) are suspected of being partially occluded and these vehicles are classified separately. Occlusions are inevitable given the low vantage point of the sensors in this proof of concept study. In future research we will investigate higher views (comparable to typical microwave radar detector deployments) to mitigate the impact of occlusions. These higher views should also provide additional features, e.g., vehicle width. Unlike video, a vehicle's width and height are easily separable in the LIDAR ranging data. The algorithm correctly classifies 24,390 of the 24,512 non-occluded vehicles (99.5%). While most side-fire detectors have challenges with occluded vehicles, the algorithms developed by this project are able to work around those problems. When a vehicle was partially occluded, we calculate the range of feasible length and height. These ranges are then used to assign one or more feasible vehicle classes to the given vehicle. Among these partially occluded vehicles, 47% were assigned a single class and 97% of these were correct.

Finally, this work also uncovered an emerging challenge facing most vehicle classification technologies: separating commuter cars from motorcycles. The two groups have similar lengths, axle spacing and height, though they differ in width and likely in weight. With increased interest in classifying motorcycles correctly, combined with more commuter cars on the road, there is a need to devise a means to separate the two types of vehicles.

Alternatively, recognizing the difficulty in distinguishing pairs of vehicle classes with the existing detector infrastructure (e.g., commuter cars and motorcycles, short SUT and PV), there may be a need to create buffer classes to impart greater confidence in the reported classifications, e.g., adding a new "class 3 or class 5" bin to the axle-based decision tree that takes the upper portion of axle class 3 and lower portion of axle class 5 axle spacings in Figure 2-3. Thus confining the uncertainty to a much smaller number of vehicles and ensuring much greater confidence that anything that is classified as "strictly class 5" is indeed axle class 5.

(This page intentionally left blank)

4 USING LIDAR TO VALIDATE THE PERFORMANCE OF VEHICLE CLASSIFICATION STATIONS

4.1 Introduction

Vehicle classification data are used in many transportation applications, including: pavement design, environmental impact studies, traffic control, and traffic safety [1]. There are several classification methods, including: axle-based (e.g., pneumatic tube and piezoelectric detectors), vehicle length-based (e.g., dual lop and some wayside microwave detectors), as well as emerging machine vision based detection. Each sensor technology has its own strengths and weaknesses regarding costs, performance, and ease of use. As noted in the Traffic Monitoring Guide [1], the quality of data collected depends on the operating agency to periodically calibrate, test, and validate the performance of classification sensors. However, such a periodic performance monitoring has been prohibitively labor intensive because the only option has been to manually validate the performance, e.g., classifying a sample by hand. Furthermore, the manual classifications are prone to human error and conventional aggregation periods allow classification errors to cancel one another.

In the present study we develop a classification performance monitoring system to allow operating agencies to monitor the health of their classification stations. We eliminate most of the labor demands and instead, deploy a portable non-intrusive vehicle classification system (PNVCS) to classify vehicles, concurrent with an existing classification station. For this study we use a side-fire LIDAR (light detection and ranging) based classifier for the PNVCS discussed in Chapter 3. Figure 4-1 shows a flowchart of our performance evaluation system, the existing classification station normally follows the three boxes within the dashed region when it is not under evaluation and the PNVCS is shown immediately to the right of the dashed region. To prevent classification errors from canceling one another in aggregate, we record per-vehicle record (pvr) data in the field from both systems. After the field collection the classification results are evaluated on a per-vehicle basis. Algorithms for time synchronization and for matching observations of a given vehicle between the two classification systems are developed in this study. These algorithms automatically compare the vehicle classification between the existing classification station and the PNVCS for each vehicle. The conventional 13 axle-based classes are consolidated into four classes to facilitate comparison with the LIDAR PNVCS in Chapter 3, i.e., motorcycle (MC) - axle class 1, passenger vehicle (PV) - axle class 2-3, single unit truck/bus (SUT) - axle class 4-7, and multiple unit truck (MUT) - axle class 8-13. If the two systems agree, the given vehicle is automatically taken as a success by the classification station (under the implicit assumption that few vehicles will be misclassified the same way by the two independent systems). The temporary deployment includes a video camera (right-most path in Figure 4-1) to allow a human to assess any discrepancies. A human only looks at a given vehicle when the two systems disagree, and for this task we have developed tools to semi-automate the manual validation process, greatly increasing the efficiency and accuracy of the human user. The datasets in this study take only a few minutes for the user to validate an hour of pvr data.

Although we use a LIDAR based system, the tools at the heart of the methodology are transferable to many PNVCS such as the TIRTL by Control Specialists, AxleLight by Quixote, and the prototype ORADS (more recently NTMS) by Spectra Research [24-27]. These systems were specifically developed to replace pneumatic tubes and use light beams just above the pavement to implement axle-based classification. The TIRTL performed very well at measuring axle spacing on two lane highways, typically above 95% accuracy [6], though some studies found an error rate of 24% among the truck classes due to the default decision tree [3, 8, 28]. While the AxleLight had an error rate for the truck classes up to 34% in high volume across four lanes [7, 8, 28], which was attributed to the sensor mistaking closely-following two-axle vehicles for multi-axle trucks. Most of the errors in [8, 28] were corrected by post-processing the pvr data from AxleLite and TIRTL using a new decision tree. Meanwhile, other studies found the TIRTL performance degrades on four lane roads [5]. Finally, as discussed in Section 2.1, side-fire microwave radar systems do not currently appear to offer sufficient classification accuracy to be used for this application.

This pilot study used LIDAR sensors mounted on a van (see, e.g., Figure 3-1(b)). This approach offers a distinct advantage over the other PNVCS since our system does not require any calibration in the field, in fact the van can be classifying vehicles as it pulls up to the site. For longer-term deployments we envision a dedicated trailer that could be parked alongside the road.

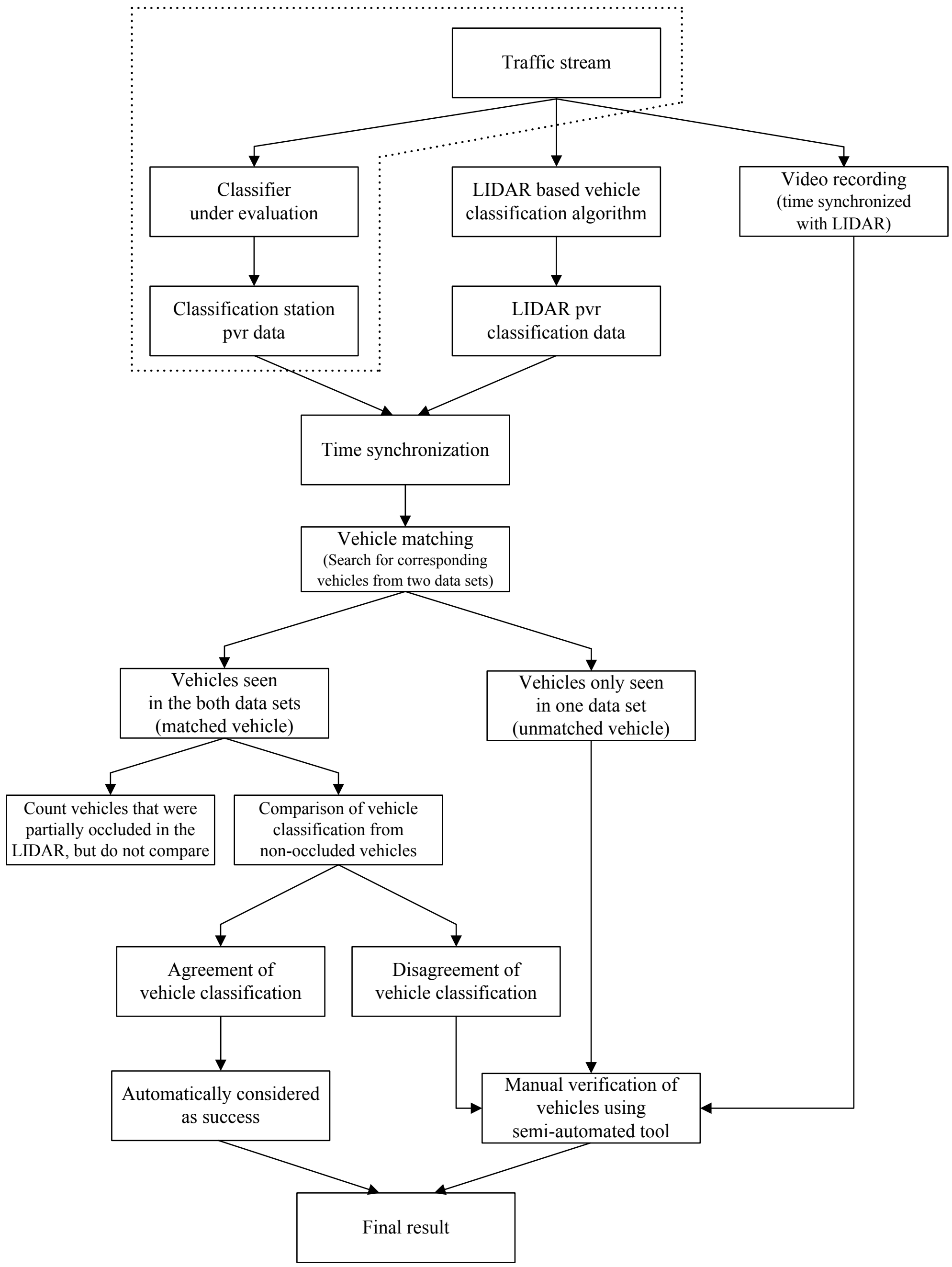


Figure 4-1, Flowchart of the evaluation of an existing vehicle classification station using LIDAR PNVCS vehicle classification. The existing station is shown in the dashed box at the top left. In normal operation most classifiers go one step further than shown in the dashed box and aggregate the pvr data by time period.

The remainder of this chapter is organized as follows. First the process of collecting the concurrent pvr vehicle classification data from the LIDAR and existing classification station is presented. Next the performance evaluation methodology is developed. Third, the methodology is applied to several permanent and temporary vehicle classification stations to evaluate axle and length-based classification. The evaluation datasets include over 21,000 vehicles, less than 8% of which required manual intervention. Finally, the chapter closes with conclusions.

4.2 Methodology of Using a PNVCS to Evaluate Classification Station Performance

This section develops the semi-automated performance evaluation methodology for an existing classification station using LIDAR PNVCS classification, as shown in Figure 4-1. There are four key steps discussed below, first the input classification data itself, then the time synchronization algorithm, next the vehicle matching algorithm to match observations of a given vehicle between the two classification systems, and finally the semi-automated tool to allow a human to rapidly review any discrepancies between the two classification systems. The discrepancies include both conflicting classifications and vehicles seen by just one of the systems. In the absence of a discrepancy, a vehicle is automatically recorded as a successful classification, without human intervention.

Given the low mounting location of the LIDAR sensors used in this study, vehicles in further lanes are susceptible to occlusions from vehicles in closer lanes. Totally occluded vehicles are a discrepancy handled in the above steps. Partial occlusions degrade the LIDAR classification performance, but the LIDAR classifier can automatically detect when a partial occlusion occurs (Chapter 3 found roughly 11% of the vehicles were partially occluded). These vehicles are counted to ensure both detectors saw a single vehicle pass, but for now the classifications are not used since a partial occlusion in the LIDAR should not be correlated with misclassifications by the existing station. In practice this approach would necessitate collecting a slightly larger dataset to accommodate the fact that some of the vehicles will not be used in the final comparison. Alternatively, if simply setting the partially occluded vehicles aside like this is unacceptable, then Section 3.4.1 presents a means to classify them to one or more classes. In the previous chapter roughly 50% of the partially occluded vehicles were assigned to a single class and could be processed automatically by the vehicle matching algorithm, the rest could be treated as a discrepancy and subjected to human evaluation with the semi-automated tool, thus, slightly increasing the number of vehicles sent for human assessment.

4.2.1 The Classification Data

Our prototype LIDAR based vehicle classification platform consists of two LIDAR sensors mounted at a height of about 6.7 ft above ground on the driver's side of a minivan parked alongside the roadway, as discussed in Chapter 3. The LIDAR sensors provide a rich view of the passing vehicles, each scan sweeps a 180° arc vertically across the road, returning the distance to the nearest object (if any) at 0.5° increments with a ranging resolution of 0.1 inch and a maximum range of 262 ft. To classify vehicles, first we segment them from the background, look for possible occlusions in further lanes, and then we measure several features of size and shape for each non-occluded vehicle. The algorithm uses these features to classify the vehicle clusters into six vehicle classes: MC, PV, PV pulling a trailer (PVPT), SUT, SUT pulling a trailer (SUTPT), and MUT. For this chapter PVPT are included with PV and SUTPT are included with MUT.

In the present study we evaluate both axle-based classification and length-based classification. We evaluate two permanent vehicle classification stations (total of three directional stations) with dual loop detectors and a piezoelectric sensor in each lane and two temporary vehicle classification deployments (total of four directional stations) with pneumatic tubes. Both systems provide the conventional 13 axle-based classes. The permanent vehicle classification stations also provide length-based vehicle classification with three length-classes that are intended to map to PV, SUT and MUT, respectively. Finally, we also tested the system at a single loop detector station using [2] for length-based classification. All of the datasets were collected in the Columbus, Ohio, metropolitan area (see Appendix A for more details).

4.2.2 Time Synchronization

The LIDAR PNVCS and the existing classification station clocks are independent, so before any comparisons are made it is necessary to first find the offset between the two systems. To automatically find this offset we borrow an approach from our earlier vehicle reidentification work, e.g., [29], only now the two locations are concurrent, so the vehicle headways become a unique signature and our algorithm looks for sequences of headways. The algorithm has to accommodate the fact that any given vehicle may be seen in just one dataset or the other due to detection errors and LIDAR occlusions, hence our use of the vehicle reidentification work.

The algorithm currently uses arrivals in one lane, over one minute.[3] We arbitrarily select one vehicle in the LIDAR data as the reference (0-th vehicle), all n vehicles that follow within a minute, and their arrival times, t_i^L. The only constraint is that there must be concurrent data from the classification station. We then successively step through the station's vehicles from the same lane, taking each one as the station's reference (K-th vehicle), all m vehicles that follow within a minute, and their arrival times t_j^C. The algorithm then tallies the number of times the n LIDAR vehicles arrive within one second of the m station vehicles, i.e., finds the rate of virtually matched vehicles (RVM_K) from Equation (4.1) for each value of K. Figure 4-2 shows an example of RVM_K versus the resulting offset time, $t_0^L - t_K^C$ from the K-th vehicle from SR 33 northbound in each lane. The algorithm selects the value of K with the largest RVM_K and uses this as the final offset, it then subtracts the corresponding offset time, $t_0^L - t_K^C$, from the entire LIDAR dataset. In Figure 4-2 the final offset time from lane 1 is -436.6 sec and from lane 2 is -436.5 sec. In this case the classification station clock is 436 sec later than the LIDAR.

$$RVM_K = \frac{1}{n}\sum_{i=0}^{n}\begin{cases}1, & \left|\left(t_i^L - t_0^L\right) - \left(t_j^C - t_K^C\right)\right| < 1\,\text{sec}, \forall j \in (K, K+m) \\ 0, & \text{otherwise}\end{cases} \tag{4.1}$$

4.2.3 Vehicle Matching

After time synchronization, most vehicles in one dataset have a unique match in the other. However, the pvr data from many of the classification stations used in this study only reported arrival times to the second. So there are many vehicles in either set that have two possible matches. With subsecond precision many of these ambiguities would be resolved, but some would likely remain. In any event, the vehicle matching algorithm seeks to find the best match for a vehicle that has two or more possible matches by accounting for the vehicles immediately before and after the ambiguity, as well as the vehicle classes assigned to these vehicles by the two sensor systems.

Formalizing the process, the i-th LIDAR PNVCS observation and j-th classification station observation are taken as a possible match if $\left|t_i^L - t_j^C\right| < 1\ sec$. The results can be summarized in a *feasible vehicle matrix*. The matrix is indexed by successive vehicle number in each dataset (LIDAR on the ordinate and classification station on the abscissa). Each element of the matrix is the outcome of the temporal comparison for the ij pair. Figure 4-3 shows an example of the feasible vehicle matrix using 11 successive vehicles from both datasets in lane 1 at SR-33 northbound. Most cells are empty, indicating there is no match, while "O" indicates a possible match for the ij pair of vehicles. The matrix shows that two classification station vehicles (379 and 383) and two LIDAR vehicles (380 and 381) have no matches in the other dataset. These unmatched vehicles will automatically be sent for manual review by the algorithm (see next section). Upon reviewing the concurrent video, the two unmatched classification station vehicles were totally occluded in the LIDAR while the two unmatched LIDAR vehicles were completely missed by the classification station.

A given vehicle can have at most one true match and indeed, most of the vehicles in Figure 4-3 have a single match. If a given possible match is the only match in the given row and column, that match is retained as a final match. Otherwise, the vehicle matching algorithm has to choose between the possible matches, e.g., classification station vehicle 374 and LIDAR vehicle 372 each have two possible matches. The algorithm assumes that vehicles maintain the same order in the two datasets, in which case, the true (but unknown) matches should fall into sequences in the feasible vehicle matrix (manifest as diagonal lines of possible matches at 45°). Whenever a vehicle has more than one possible match, the vehicle matching algorithm collects the group of all involved vehicles from each detector (classification station vehicles 373-374 and LIDAR vehicles 372-373 in Figure 4-3). Figure 4-4(a) shows an extreme hypothetical example, where almost every vehicle falls into one of three distinct groups of vehicles, as shown in Figure 4-4(b). If there is a single longest sequence in a group, the algorithm selects that sequence as final matches, Figure 4-4(c). Otherwise, if there are two or more sequences tied for the longest sequence, the algorithm considers the classifications assigned by the two sensor systems and chooses the sequence with the best classification agreement, e.g., as would be necessary for group 2 in Figure 4-4(c).

[3] Expanding to multiple lanes or longer duration would improve the precision in challenging conditions.

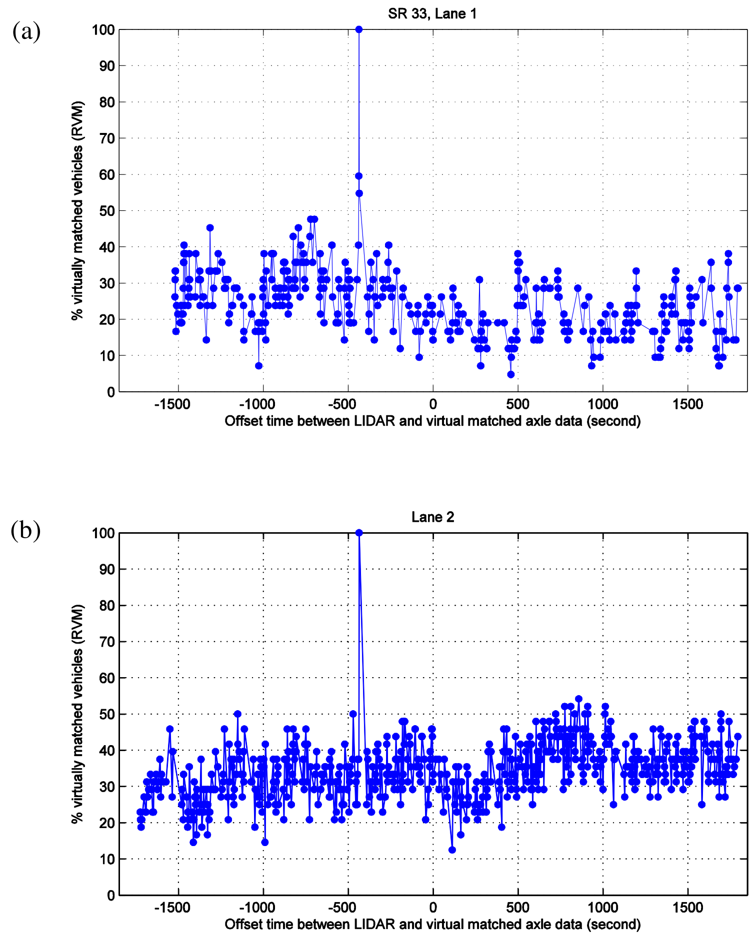


Figure 4-2, RVM$_K$ versus the resulting offset time as a function of K from SR 33 northbound, (a) Lane 1, the peak shows the final offset time is -436.6 sec, and (b) Lane 2, the peak shows the final offset time is -436.5 second.

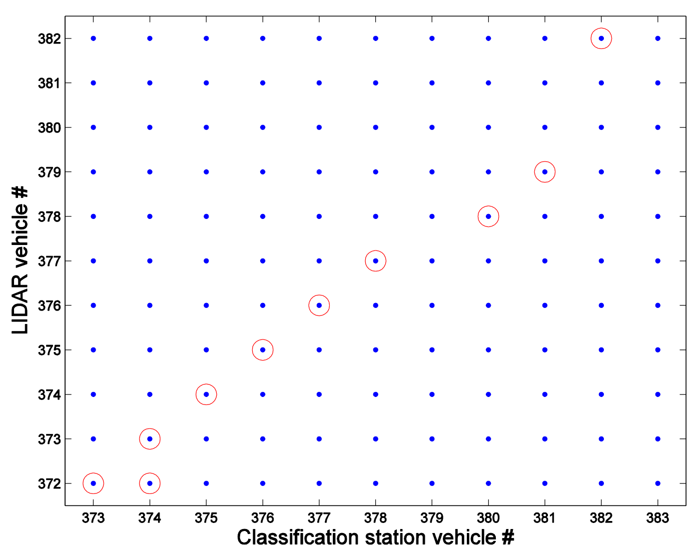


Figure 4-3, A feasible vehicle matrix, summarizing the outcome from the difference of arrival times between the LIDAR and classification station data in lane 1 at SR-33 northbound.

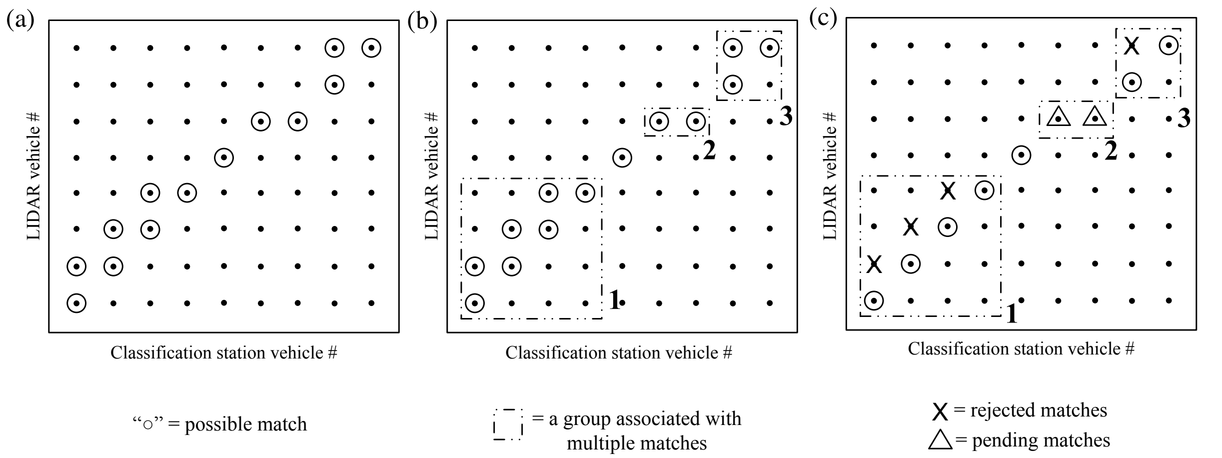


Figure 4-4, (a) hypothetical feasible vehicle matrix in which many rows and columns have multiple matches, (b) isolating the distinct groups of vehicles, the groups are numbered for reference, (c) selecting the longest sequence from the given group. Note that the two sequences in group 2 are equal length, so the algorithm would then compare the classification results from the two sensor systems and select the sequence with the strongest similarity between the two sensor systems.

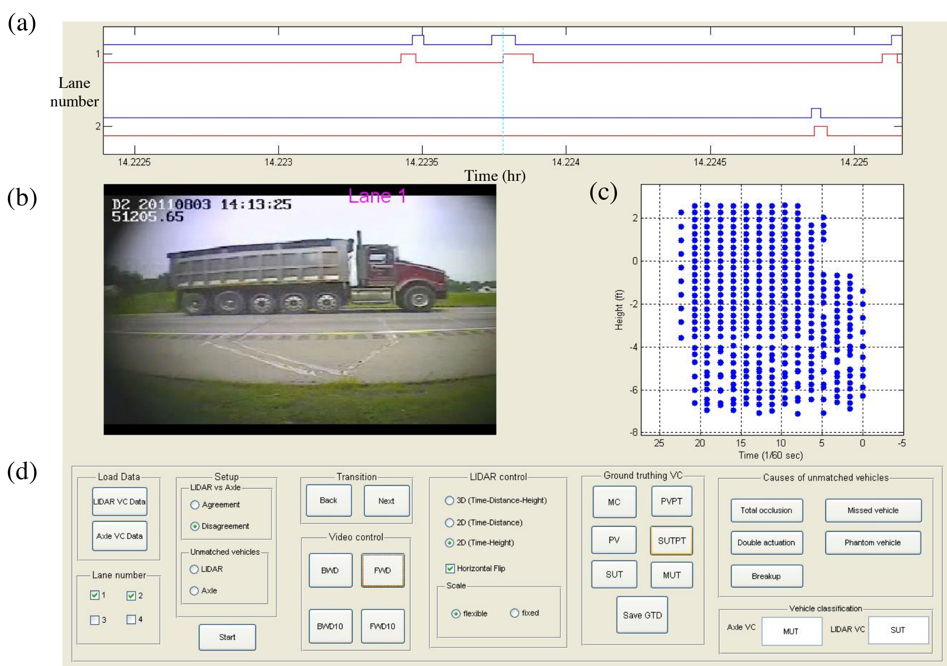


Figure 4-5, A snapshot of the semi-automated GUI verification tool processing a conflicting classification for a vehicle in lane 1 at SR 33 northbound. The GUI window consists of four interfaces: (a) plot of transition pulses, the plot shows for each lane the classification station data (top curve) and LIDAR data (bottom curve) and the current instant is shown with a vertical dashed line, (b) the current video frame, (c) the LIDAR returns from the vehicle in question, and (d) a panel for controlling the review and entering ground truth data. So in this case the GUI is at the second visible pulse in lane 1 (counted from the left hand side) and is ready for the user to assess the data using the buttons on the right of part (d).

4.2.4 Manual Verification Using a Semi-Automated Tool

Inspired by VideoSync [30], a purpose built software ground truthing tool with a graphical user interface (GUI) was developed in MATLAB to efficiently generate ground truth data and increase the accuracy of the human user. After the time synchronization and vehicle matching steps above, the GUI loads the pvr classifications from the classification station and the LIDAR PNVCS. The user can choose which set(s) of vehicles they wish to review: (i) seen only in LIDAR, (ii) seen only at the classification station, (iii) conflicting classifications between the two sources, and/or (iv) consistent classification between the two sources. Normally the user would select the three error conditions, i.e., sets i-iii. Next, the user chooses one or more lanes to review, then the GUI steps through all of the vehicles in the given set(s) and lane(s). Figure 4-5 shows an example of the GUI as a SUT passes.[4] For each vehicle the GUI displays the raw LIDAR data and the raw classification station data for a few seconds before and after the given vehicle detection (Figure 4-5(c) and (a), respectively). The GUI shows the video frame at the instant of the vehicle passage (Figure 4-5(b)), and allows the user to step forward or back in the video to see the evolution if necessary (Figure 4-5(d)). The bottom right corner of the GUI shows the user what vehicle class was assigned by the station and the LIDAR. After assessing the concurrent sensor and video data, the user records the observed vehicle class (or detection error) for the current actuation via the buttons in the two right-most boxes of Figure 4-5(d). As soon as the user enters a selection, the GUI jumps to the next actuation in the selected set(s) and lane(s) until all of the vehicles have been reviewed in the given set(s) from the entire time period with video data. In this study the user typically spent 3-5 sec per vehicle reviewed (including seek time and loading time), but only about 8% of the actuations required review. The automated process does the bulk of the work, in this study it typically took the human only a few minutes to process the exceptions from all lanes over one hour of data.

4.3 Results of Using a PNVCS to Evaluate Classification Station Performance

4.3.1 Axle-Based Classification Stations

As noted above, we collected concurrent LIDAR and classification station pvr data at two permanent axle classification stations (I-270 and SR-33) and two temporary axle classification deployments (Wilson Rd and Dublin Rd). Table 4-1 enumerates the location, date, duration, and number of lanes in the first few columns. All locations yielded data for the direction of travel adjacent to the minivan (top rows in the table). We parked the van on both sides of Wilson Rd, hence both NB and SB nearside data for this location. Almost all of the locations provided sufficient view of the far lanes in the opposing direction to allow LIDAR classification, shown in the lower portion of the table. The one exception was I-270, where the median barrier and superelevation precluded a view of the opposing lanes. In any event, all lanes are numbered successively from the LIDAR minivan, regardless of the direction of travel.

Columns (a) and (b) show the number of actuations reported by the LIDAR and classification data (including any non-vehicle actuations). Columns (c)-(e) show the number of matched and unmatched actuations after the vehicle matching algorithm. Column (f) sums columns (c), (d), and (e), yielding the number of actuations seen by one or both sensors. Column (g) tallies the number of partially occluded vehicles detected in the LIDAR (as per Section 3.2.1) and seen by the classification station. Since the partial occlusions do not reflect any error by the classification station, at present they are excluded from further analysis.[5] Column (h) shows the number of actuations for which the algorithm compared the respective classifications from the two systems and from this set (i) tallies the disagreement. The percentage of disagreement is below 8% for all lanes studied and below 4% for most of them. Columns (j) and (k) reiterate (d) and (e) as percentages of (f). Finally, column (l) tallies the number of vehicles subject to manual verification (sum of columns (d), (e) and (i), as a percent of (f)).

[4] In the event that this figure is hard to read, the key features are the integrated video view, detector data, and the fact that there are several buttons for user input. The figure may be clearer in the electronic version of the report available from ODOT.

[5] See Section 4.2 for a discussion on how the partially occluded vehicles can be handled if they are specifically of interest. Roughly half of these vehicles would require human review, slightly increasing the labor demands for the evaluation.

Table 4-1, Summary of the automated comparison of vehicle classification between LIDAR and axle data at seven directional classification stations,

LIDAR sensor direction relative to vehicle travel direction	Location (direction)	Date	Duration (hh:min)	Lane number from LIDAR	Number of vehicles seen in;		From the algorithm of vehicle matching			Number of vehicles passing the location (f)	Number of partially occluded vehicles (g)	Comparison of vehicle classification		% vehicles not detected by;		Number of vehicles manually confirmed (l)
					LIDAR (a)	Axle (b)	Number of vehicles seen in both LIDAR and Axle (c)	Number of vehicles only seen in; LIDAR (d)	Number of vehicles only seen in; Axle (e)			Number of compared vehicles (h)	Disagreement (i)	LIDAR (j)	Axle (k)	
Adjacent	I-270 (SB)	11/02/2010	5:00	1	5,415	5,452	5,389	26	63	5,478	n/a	5,389	188 (3.5%)	1.2%	0.5%	277 (5.1%)
				2	5,335	5,488	5,303	32	185	5,520	641	4,662	145 (3.1%)	3.4%	0.6%	362 (6.6%)
				3	2,647	2,789	2,615	32	174	2,821	713	1,902	24 (1.3%)	6.2%	1.1%	230 (8.2%)
	Dublin (SB)	10/28/2010	2:50	1	1,344	1,317	1,313	31	4	1,348	n/a	1,313	80 (6.1%)	0.3%	2.3%	115 (8.5%)
	Wilson (NB)	10/28/2010	1:40	1	666	664	658	8	6	672	n/a	658	24 (3.6%)	0.9%	1.2%	38 (5.7%)
	Wilson (SB)	10/28/2010	1:40	1	711	712	701	10	11	722	n/a	701	21 (3.0%)	1.5%	1.4%	42 (5.8%)
	SR 33 (NB)	08/03/2011	1:10	1	732	693	684	48	9	741	n/a	684	32 (4.7%)	1.2%	6.5%	89 (12.0%)
				2	569	562	547	22	15	584	65	482	6 (1.2%)	2.6%	3.8%	43 (7.4%)
Subtotal of adjacent			12:20	-	17,419	17,677	17,210	209	467	17,886	1,419	15,791	520 (3.3%)	2.6%	1.2%	1,196 (6.7%)
Opposite	Dublin (NB)	10/28/2010	2:50	2	940	943	933	7	10	950	75	858	52 (6.1%)	1.1%	0.7%	69 (7.3%)
	Wilson (NB)	10/28/2010	1:40	2	749	752	742	7	10	759	58	684	18 (2.6%)	1.3%	0.9%	35 (4.6%)
	Wilson (SB)	10/28/2010	1:40	2	741	735	723	18	12	753	47	676	24 (3.6%)	1.6%	2.4%	54 (7.2%)
	SR 33 (SB)	08/03/2011	1:10	3	592	587	548	44	39	631	53	495	9 (1.8%)	6.2%	7.0%	92 (14.6%)
				4	888	884	838	50	46	934	148	690	54 (7.8%)	4.9%	5.4%	150 (16.1%)
Subtotal of opposite			7:20	-	3,910	3,901	3,784	126	117	4,027	381	3,403	157 (4.6%)	2.9%	3.1%	400 (9.9%)
Overall			19:40	-	21,329	21,578	20,994	335	584	21,913	1,800	19,194	677 (3.5%)	2.7%	1.5%	1,596 (7.3%)

n/a: occlusions are infeasible in this lane because it is adjacent to the LIDAR sensor
(a) = (c + d)
(b) = (c + e)
(f) = (a + e) = (b + d)
(h) = (c) – (g)
(j) = (e) / (f)
(k) = (d) / (f)
(l) = (d + e + i), where the percentage is relative to (f)

Table 4-2, Manual verification of the vehicles with conflicting classifications or only seen by one sensor using the semi-automated tool,

LIDAR sensor direction relative to vehicle travel direction	Location (direction)	Lane number from LIDAR	Number of vehicles not detected by LIDAR (e)	Reason			Number of vehicles not detected by Axle (d)	Reason		Number of vehicles in disagreement (i)	Verification of disagreement			% axle miss-classified (r)	% total axle error (s)
				Totally occluded vehicle	LIDAR missed vehicle	Axle non-vehicle actuation (m)		Axle missed vehicle (n)	LIDAR non-vehicle actuation		LIDAR correct, Axle incorrect (p)	LIDAR incorrect, Axle correct	LIDAR incorrect Axle incorrect (q)		
Adjacent	I-270 (SB)	1	63	n/a	63	0	26	26	0	188	148	36	4	2.8%	3.2%
		2	185	116	69	0	32	32	0	145	113	30	2	2.5%	2.6%
		3	174	141	33	0	32	32	0	24	20	4	0	1.1%	1.7%
	Dublin (SB)	1	4	n/a	4	0	31	31	0	80	76	4	0	5.8%	7.9%
	Wilson (NB)	1	6	n/a	6	0	8	8	0	24	22	2	0	3.3%	4.5%
	Wilson (SB)	1	11	n/a	11	0	10	10	0	21	18	3	0	2.6%	3.9%
	SR-33 (NB)	1	9	n/a	9	0	48	48	0	32	26	4	2	4.1%	10.3%
		2	15	8	7	0	22	22	0	6	5	1	0	1.0%	4.5%
Subtotal of adjacent			467	265	202	0	209	209	0	520	428	84	8	2.8%	3.6%
Opposite	Dublin (NB)	2	10	5	1	4	7	7	0	52	48	3	1	5.7%	6.3%
	Wilson (NB)	2	10	5	5	0	7	7	0	18	15	3	0	2.2%	2.9%
	Wilson (SB)	2	12	10	2	0	18	18	0	24	22	2	0	3.3%	5.3%
	SR-33 (SB)	3	39	27	12	0	44	44	0	9	6	3	0	1.2%	7.9%
		4	46	41	4	1	50	50	0	54	42	11	1	6.2%	10.1%
Subtotal of opposite			117	88	24	5	126	126	0	157	133	22	2	4.0%	6.6%
Overall			584	353	226	5	335	335	0	677	561	106	10	3.0%	4.1%

n/a: occlusions are infeasible in this lane because it is adjacent to the LIDAR sensor

(r) = (p+q) / (h)
(s) = (p+q+m+n)/(f-m)

Note (f) and (h) are shown in Table 4-1.

Table 4-2 summarizes the results from manual verification for the vehicles with a discrepancy in Table 4-1 (columns (d), (e) and (i)). Of the vehicles only seen by the LIDAR, 60% (353 out of 584) are due to completely occluded vehicles, 39% (226 out of 584) are due to the LIDAR missing unoccluded vehicles, and 1% are due to non-vehicle actuations at the classification station. Upon review, it turns out that all 335 of the actuations that were not detected at the classification stations were due to those stations missing the vehicles. Of the vehicles with conflicting classification, the classification station was incorrect 84% of the time (571 out of 677). Assuming few vehicles are misclassified the same way by the two systems, all of the agreements are automatically tallied as a success by the classification station. As a result, the classification stations exhibited an overall misclassification rate of 3% (sum of columns (p) and (q) as a percent of (h)), and including the undetected vehicles, an overall error rate of 4.5% (sum of columns (m), (n), (p), and (q) divided by [(f)-(m)]). The highest error rate observed in a lane was 10.3%.

To ensure the validity of the assumption that no individual vehicles were misclassified the same way by both systems, (and thus, by extension, degrade the accuracy of the above results), we manually verified the class of 15,271 out of the 18,517 vehicles that the two systems gave the same class. As noted above, these vehicles would normally be assigned "success" automatically, without review by a person. Within this set, 99.8% (15,245 out of 15,271) were assigned the correct vehicle class and only 26 vehicles (0.2%) were incorrectly classified.

Table 4-3 compares the specific classification of the non-occluded vehicles detected by both sensors across all of the datasets. The columns show the axle classification and rows show the LIDAR classification. The bold numbers on the diagonal show the agreement between the two systems and all of the numbers off the axis reflect the disagreements. The third row from the bottom and the second column from the end tally the class of vehicles that were only seen by one of the detectors. The last column and second to the last row tally the row and column total, respectively. The final row presents the number of partially occluded vehicles that were excluded from the comparisons, sorted by axle class for reference. Collectively, 4.6% of the non-occluded vehicles (919 out of 20,113) are detected by only one sensor, of the remaining 19,194 non-occluded vehicles that were detected by both sensors, 96.5% (18,517 vehicles) were assigned the same classification from the two systems and 3.5% (677 vehicles) were not.

As noted above, all of the vehicles assigned the same class by both systems are automatically taken to be correct, while all of the conflicting classifications were manually validated (i.e., the off diagonal cells in Table 4-3). After conducting the manual validation we refer to the collection of the results as pseudo ground truth since the cells that were originally in agreement were not manually validated. The axle classification station performance across all of the datasets is compared against the pseudo ground truth in Table 4-4. There are a total of 19,760 vehicles in the pseudo ground truth data, including 19,194 non-occluded vehicles seen by both sensors, 335 vehicles not detected by the axle sensors, 226 vehicles not detected by the LIDAR sensors, and 5 non-vehicle actuations in the axle data. The remaining 353 vehicles from Table 4-3 were completely occluded in the video as well. The completely occluded vehicles are excluded from the comparison, but their assigned axle class is reported in the final row for reference. No vehicle changed columns from Table 4-3 since the axle classifications did not change, but many of the vehicles were reassigned to new rows as a result of the manual validation. The accuracy of pseudo ground truth data should be above 99% because most vehicles with the corresponding classification are correctly classified (as per above, we found that only 0.2% of the vehicles with the same classification from the two systems were incorrectly classified). The classification stations exhibited 95% accuracy overall, but dramatically different performance by class. The best performance was on PV and worst performance on MC. It is also important to take care reading the table, although 83% of the vehicles classified as SUT by the axle classification stations were indeed SUT (column total), only 66% of the SUT were correctly classified as such (row total). This pseudo ground truth analysis is repeated by individual station in Appendix F and Table 4-5 summarizes the performance by station. To help interpret these results, the final row of Table 4-5 summarizes Table 4-4. The first few columns report the number of vehicles seen in the pseudo ground truth for the given class (e.g., the second to the last column in Table 4-4), the next set of columns present the percentage of vehicles correctly classified in the given class (e.g., the last column in Table 4-4), and the last set of columns present the percentage of detector station classifications that were correct in the given class (e.g., the second to the last row in Table 4-4).

Table 4-4 shows the worst performance for motorcycles, with only 27% being correctly classified, but this table combines data from ODOT permanent classification stations and MORPC temporary pneumatic tube deployments. Unfortunately the pneumatic tubes were much better at detecting and classifying the motorcycles. Reviewing the data strictly from the two ODOT classification stations with concurrent LIDAR (Appendix F), the pseudo ground truth include 15 motorcycles, of which only 1 (7%) was correctly classified by the classification stations. Meanwhile, 9 (60%) of the motorcycles were misclassified as longer vehicles and 5 (33%) passed completely undetected. Given the fact that these data come from only two classification stations and the number of motorcycles is small, further study is warranted. For example, Table 2-2 shows two completely occluded vehicles being classified as motorcycles at I-270. Meanwhile, Table 2-7 shows that the I-70 data had 31 motorcycles correctly classified, but this location did not have an independent PNVCS and using the same data set, [11] found an additional 13 passing motorcycles that were missed by the detectors, resulting instead in "sensor miss" (SnMis) events in the pvr data.

4.3.2 Length-Based Classification Stations

As noted in the introduction, we also used this methodology to evaluate the performance of length-based classification. All of the permanent vehicle classification stations also provide length-based and we also tested the system at a single loop detector station using [2] for length-based classification. All vehicles below 28 ft are assigned to length class 1, all remaining vehicles below 47 ft are assigned to length class 2, and all vehicles above 47 ft are assigned length class 3; and these length classes are intended to roughly map to PV, SUT and MUT, respectively. So for our analysis we map LIDAR MC and PV to length class 1, LIDAR SUT to length class 2, and LIDAR MUT to length class 3. Tables 4-6 to 4-9 repeat the comparisons of the previous section, now applied to the length-based classification stations. The length-based performance and number of vehicle requiring manual validation are comparable to the axle-based classification. Appendix G show the length-based classification pseudo ground truth results by station.

4.4 Conclusions

Vehicle classification data are critical to many transportation applications, but the quality of data collected depends on the operating agency to periodically calibrate, test, and validate the performance of classification sensors. These studies are labor intensive and coarse, allowing over counting errors to cancel undercounting errors. To address these challenges, the present work develops a classification performance monitoring system to allow operating agencies to automatically monitor the health of their classification stations. We eliminate most of the labor demands and instead, deploy a LIDAR based PNVCS to classify vehicles, concurrent with existing classification stations. To prevent classification errors from canceling one another in aggregate, we record per-vehicle record (pvr) data in the field from both systems. After the field collection the classification results are evaluated on a per-vehicle basis. If the two systems agree, the given vehicle is automatically taken as a success by the classification station. The PNVCS includes a video camera to allow a human to assess the discrepancies. A human only looks at a given vehicle when the two systems disagree, and we developed tools to semi-automate the manual validation process, greatly increasing the efficiency and accuracy of the human user. The datasets in this study take only a few minutes for the user to validate an hour of pvr data. Although we use a LIDAR based system, the tools at the heart of the methodology are transferable to many PNVCS such as the TIRTL or AxleLight. This pilot study used LIDAR sensors mounted on a van. This approach offers a distinct advantage over the other PNVCS since our system does not require any calibration in the field, in fact the van can be classifying vehicles as it pulls up to the site. For longer-term deployments we envision a dedicated trailer that could be parked alongside the road.

The evaluation datasets come from several different classification stations, they include over 21,000 vehicles. We separately evaluated length-based classification stations and axle-based classification stations, each yielding similar results. In each case about 8% of the vehicles required manual intervention. In this study the user typically spent 3-5 sec per vehicle reviewed. The automated process does the bulk of the work, in this study it typically took the human only a few minutes to process the exceptions from all lanes over one hour of data.

This evaluation revealed a chronic problem detecting motorcycles at the two ODOT permanent classification stations studied. While the LIDAR system detected 15 passing motorcycles, the stations correctly classified one of them, and missed five altogether.

As this research has shown, there is wide variance in performance from one station to the next and these errors tend to have a higher frequency among the truck classes, particularly the SUT. Since these errors are a function of the specific station, there would be benefit in the short term if ODOT were to leverage the LIDAR based PNVCS system developed in this research to evaluate the performance of many other classification stations. Thereby catching systematic errors that bias classification performance at the given station.

Table 4-3, Comparison of LIDAR vehicle classification and axle vehicle classification across seven directional locations,

Overall		Axle vehicle classification: Motor-cycle	Passenger vehicle*	Single unit truck	Multiple unit truck*	Number of LIDAR vehicles not detected by axle sensor	Total number of LIDAR vehicles
LIDAR vehicle classification	Motorcycle	**6**	12	1	0	12	31
	Passenger vehicle*	2	**16,751**	127	159	283	17,322
	Single unit truck	1	212	**530**	96	28	867
	Multiple unit truck*	1	47	19	**1,230**	12	1,309
Number of axle vehicles not detected by LIDAR sensor		3	555	10	16	-	584
Total number of axle vehicles above		13	17,577	687	1,501	335	20,113
Number of partially occluded vehicles excluded in the comparison matrix		2	1,571	56	171	-	1,800

Passenger vehicle* includes passenger vehicle and passenger vehicle pulling a trailer.
Multiple unit truck* includes single unit truck pulling a trailer and multiple unit truck.

Table 4-4, Comparison of pseudo ground truth data and axle vehicle classification across seven directional locations,

Overall		Axle vehicle classification: Motor-cycle	Passenger vehicle*	Single unit truck	Multiple unit truck*	Number of LIDAR vehicles not detected by axle sensor	Row total	% correct
Pseudo ground truth data	Motorcycle	**6**	2	6	3	5	22	27%
	Passenger vehicle*	2	**17,001**	94	160	289	17,546	97%
	Single unit truck	1	196	**574**	79	25	875	66%
	Multiple unit truck*	1	30	9	**1,256**	16	1,312	96%
	Non-vehicle actuation in axle data	2	3	0	0	-	5	-
Column total above		12	17,232	683	1,498	335	19,760	-
% correct		50%	99%	84%	84%	-	-	95%
Totally occluded vehicle		1	345	4	3	-	353	-

Passenger vehicle* includes passenger vehicle and passenger vehicle pulling a trailer.
Multiple unit truck* includes single unit truck pulling a trailer and multiple unit truck.

Table 4-5, Summary of evaluation of axle vehicle classification station by a vehicle class. Note Wilson Rd northbound and southbound includes both Wilson Rd adjacent to and opposite from LIDAR sensor, respectively,

Location	Dire-ction	A number of vehicles from pseudo ground truth data				% of pseudo ground truth vehicle classified correctly				% of correct axle classification				% of correct classifi-cation over all vehicles
		MC	PV*	SUT	MUT*	MC	PV*	SUT	MUT*	MC	PV*	SUT	MUT*	
I-270	SB	7	10,561	500	1,140	14%	99%	61%	97%	50%	98%	95%	92%	97%
Dublin Rd	NB	2	795	63	6	50%	94%	87%	100%	20%	99%	72%	22%	93%
	SB	2	1,282	53	11	50%	92%	87%	100%	100%	99%	61%	21%	92%
Wilson Rd	NB	1	1,280	60	27	100%	96%	97%	100%	100%	100%	77%	60%	96%
	SB	2	1,360	29	27	100%	95%	90%	100%	100%	100%	63%	54%	95%
SR 33	NB	5	1,114	79	54	0%	95%	57%	87%	-	99%	94%	72%	92%
	SB	3	1,154	91	47	0%	93%	46%	79%	0%	98%	84%	62%	89%
Overall		22	17,546	875	1,312	27%	97%	66%	96%	50%	99%	84%	84%	95%

PV* includes passenger vehicle and passenger vehicle pulling a trailer. MUT* includes single unit truck pulling a trailer and multiple unit truck.

Table 4-6, Summary of the comparison of vehicle classification between LIDAR and loop detector data at four directional classification stations

Location (direction)	Date	Duration (hh:min)	Lane number from LIDAR	Number of vehicles seen in:		From the algorithm of vehicle matching			Number of vehicles passing the location (f)	Number of partially occluded vehicles (g)	Comparison of vehicle classification		% vehicles not detected by:		Number of vehicles manually confirmed (l)
				LIDAR (a)	Loop detector (b)	Number of vehicles seen in both LIDAR and loop detector (c)	Number of vehicles only seen in; LIDAR (d)	Loop detector (e)			Number of compared vehicles (h)	Disagreement (i)	LIDAR (j)	Loop detector (k)	
I-71 (SB): Free flow	07/09/2009	00:24	1	168	156	156	12	0	168	n/a	156	9	0.0%	7.1%	21 (12.5%)
			2	546	539	538	8	1	547	6	532	16	0.2%	1.5%	25 (4.6%)
			3	644	653	638	6	15	659	132	506	13	2.3%	0.9%	34 (5.2%)
			4	454	482	445	9	37	491	169	276	2	7.5%	1.8%	48 (9.8%)
I-71 (SB): Congestion	11/19/2009	00:28	1	191	182	181	10	1	192	n/a	181	20	0.5%	5.2%	31 (16.1%)
			2	859	848	848	11	0	859	18	830	12	0.0%	1.3%	23 (2.7%)
			3	772	798	771	1	27	799	228	543	10	3.4%	0.1%	38 (4.8%)
			4	797	912	795	2	117	914	343	452	0	12.8%	0.2%	119 (13.0%)
I-270 (SB): Free flow	11/02/2010	5:00	1	5,415	5,452	5,389	26	63	5,478	n/a	5,389	184	1.2%	0.5%	273 (5.0%)
			2	5,335	5,488	5,303	32	185	5,520	641	4,662	131	3.4%	0.6%	348 (6.3%)
			3	2,647	2,789	2,615	32	174	2,821	713	1,902	28	6.2%	1.1%	234 (8.3%)
SR-33 (NB): Free flow	08/03/2011	1:10	1	732	693	684	48	9	741	n/a	684	31	1.2%	6.5%	88 (11.9%)
			2	569	562	547	22	15	584	65	482	7	2.6%	3.8%	44 (7.5%)
SR-33 (SB): Free flow	08/03/2011	1:10	3	592	587	548	44	39	631	53	495	8	6.2%	7.0%	91 (14.4%)
			4	888	884	838	50	46	934	148	690	69	4.9%	5.4%	165 (17.7%)
Overall				20,609	21,025	20,296	313	729	21,338	2,516	17,780	540	3.4%	1.5%	1,582 (7.4%)

n/a: occlusions are infeasible in this lane because it is adjacent to the LIDAR sensor

(a) = (c + d)
(b) = (c + e)
(f) = (a + e) = (b + d)
(h) = (c) – (g)
(j) = (e) / (f)
(k) = (d) / (f)
(l) = (d + e + i), where the percentage is relative to (f)

Table 4-7, Manual verification using semi-automated tool of the vehicles with conflicting classifications or only seen by one sensor from the comparison of vehicle classification between LIDAR and loop detector data

Loca-tion (direc-tion)	Lane num-ber from LIDAR	Number of vehicles not detected by LIDAR (e)	Reason			Number of vehicles not detected by loop detector (d)	Reason		Number of vehicles in disagree-ment (i)	Verification of disagreement			% loop detector miss-classified (r)	% total loop detector error (s)
			Totally occluded vehicle	LIDAR missed vehicle	Loop detector non-vehicle actuation (m)		Loop detector missed vehicle (n)	LIDAR non-vehicle actuation		LIDAR correct, loop incorrect (p)	LIDAR incorrect, loop correct	LIDAR incorrect loop incorrect (q)		
I-71 (SB): Free flow	1	0	n/a	0	0	12	12	0	9	9	0	0	5.8%	12.5%
	2	1	1	0	0	8	8	0	16	15	1	0	2.8%	4.2%
	3	15	14	1	0	6	6	0	13	11	2	0	2.2%	2.6%
	4	37	32	5	0	9	9	0	2	2	0	0	0.7%	2.2%
I-71 (SB): Congestion	1	1	n/a	1	0	10	10	0	20	20	0	0	11.0%	15.6%
	2	0	0	0	0	11	11	0	12	12	0	0	1.4%	2.7%
	3	27	25	2	0	1	1	0	10	8	2	0	1.5%	1.1%
	4	117	104	13	0	2	2	0	0	0	0	0	0.0%	0.2%
I-270 (SB): Free flow	1	63	n/a	63	0	26	26	0	184	156	22	6	3.0%	3.4%
	2	185	116	69	0	32	32	0	131	112	15	4	2.5%	2.7%
	3	174	141	33	0	32	32	0	28	27	1	0	1.4%	2.1%
SR-33 (NB): Free flow	1	9	n/a	9	0	48	48	0	31	29	2	0	4.2%	10.4%
	2	15	8	7	0	22	22	0	7	6	1	0	1.2%	4.8%
SR-33 (SB): Free flow	3	39	27	12	0	44	44	0	8	8	0	0	1.6%	8.2%
	4	46	41	4	1	50	50	0	69	59	5	5	9.3%	12.3%
Overall		729	510	218	1	313	313	0	540	474	51	15	2.8%	3.8%

n/a: occlusions are infeasible in this lane because it is adjacent to the LIDAR sensor

(p) = (m+n) / (h)
(q) = (m+n+aa+bb)/(f-aa)

Note (f) and (h) are shown in Table 4-6.

Table 4-8, Comparison of pseudo ground truth data and length-based vehicle classification across four directional locations

Overall		Length class from loop detector: Class 1	Class 2	Class 3	Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
Pseudo ground truth data	Passenger vehicle**	15,623	256	66	271	16,216	96%	0
	Single unit truck	125	590	8	26	749	79%	0
	Multiple unit truck*	21	23	1,286	16	1,346	96%	0
	Non-vehicle actuation in loop detector data	1	0	0	-	1	-	-
Column total above		15,770	869	1,360	313	18,312	-	0
% correct		99%	68%	95%	-	-	96%	-
Totally occluded vehicles		498	7	5	-	510	-	-

Passenger vehicle** includes motorcycle, passenger vehicle, and passenger vehicle pulling a trailer.
Multiple unit truck* includes single unit truck pulling a trailer and multiple unit truck.

Table 4-9, Summary of evaluation of length-based vehicle classification station by a vehicle class.

Location (traffic condition)	Dire-ction	A number of vehicles from pseudo ground truth data			% of pseudo ground truth vehicle classified correctly			% of correct loop classification			% of correct classification over all vehicles
		Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3	
I-71 (free flow)	SB	1,428	32	51	96%	50%	94%	99%	42%	84%	95%
I-71 (Cong)	SB	1,967	38	40	97%	61%	98%	99%	47%	95%	97%
I-270 (free flow)	SB	10,546	509	1,153	97%	92%	97%	99%	70%	95%	97%
SR-33 (free flow)	NB	1,117	81	54	94%	69%	81%	98%	79%	96%	92%
	SB	1,158	89	48	92%	31%	73%	95%	60%	92%	87%

5 CONCLUSIONS AND RECOMMENDATIONS

This study examined three interrelated facets of vehicle classification and classification performance monitoring. First, we manually evaluate the performance of vehicle classification station on a per-vehicle basis, second we develop a portable LIDAR based vehicle classification system that can be rapidly deployed, and third we use the LIDAR based system to automate the manual validation done in the first part using the tools from the second part. Each component is discussed in a separate chapter, and the conclusions are presented at the end of each chapter. This section summarizes the conclusions from those chapters.

In Chapter 2 we used per-vehicle record (pvr) data to manually evaluate the performance of several classification stations. Evaluating the pvr data as we do in this work is uncommon; both due to the inherent difficulty generating ground truth data, and the fact that normally the pvr classifications are binned by fixed time periods at which point the individual vehicle information is discarded. However, such conventional aggregation allows errors to cancel one another, which can obscure underlying problems. This study evaluated three permanent axle classification stations against concurrent video based ground truth in terms of axle-based and length-based classification. Only 3%-4% of the vehicles were misclassified, however, the relative impacts were much larger on the trucks, e.g., only 60% of the single unit truck/bus (SUT) -axle class 4-7 - were correctly classified as SUT by the existing axle-based classifier.

Diagnosing the axle classification errors, it was found that all of them could be attributed to one of six causes. About a third of the errors among class 4-13 can be easily fixed by redefining the decision tree, e.g., ensuring that there are no gaps between successive classes and adding an additional outcome from the tree to indicate a vehicle is unclassifiable. Our revised decision tree is shown in Table 2-6. After making these changes, the axle-based classification decision tree was able to correctly classify an additional 10% of the SUT, with smaller improvements in almost every other metric. Ideally the new decision tree should be deployed at a few new locations and the performance validated, then assuming no problems are found, be adopted as the new standard classification decision tree. In any event, most of the improvements of the new decision tree should be incorporated in to standard practice (closing the gaps between bins, adding an "unclassifiable" class, and allowing for more than 4 axles in axle class 7).

One chronic error found in this research is intrinsic to the vehicle fleet and may be impossible to correct with the existing sensors; namely, the shorter, SUT have a length range and axle spacing range that overlaps with passenger vehicles (PV) - axle class 1-3. Depending on the calibration, the error may be manifest as SUT counted as PV or vice versa. As discussed in the literature review in Section 2.1, this PV/SUT blurring appears to impact other sensors as well. In any case, one should expect such errors at most classification stations. All subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of SUT with PV. The blurring also means that one cannot blindly use an axle classification station to calibrate the boundary between PV and SUT for length-based classification stations, otherwise, the unavoidable errors in the axle classification will be amplified in the length-based classification scheme.

In Chapter 3 we developed and tested a side-fire LIDAR based vehicle classification algorithm. The algorithm includes up to eight different measurements of vehicle shape to sort vehicles into six different classes. The algorithm was tested over seven datasets (including one development dataset) collected at various locations. The results were compared against the concurrent video-recorded ground truth data on a per-vehicle basis. Overall, 2,938 out of 27,450 vehicles (11%) are suspected of being partially occluded and these vehicles are classified separately. Occlusions are inevitable given the low vantage point of the sensors in this proof of concept study. In future research we will investigate higher views (comparable to typical microwave radar detector deployments) to mitigate the impact of occlusions. These higher views should also provide additional features, e.g., vehicle width. Unlike video, a vehicle's width and height are easily separable in the LIDAR ranging data. The algorithm correctly classifies 24,390 of the 24,512 non-occluded vehicles (99.5%). While most side-fire detectors have challenges with occluded vehicles, the algorithms developed by this project are able to work around those problems. When a vehicle was partially occluded, we calculate the range of feasible length and height. These ranges are then used to assign one or more feasible vehicle classes to the given vehicle. Among these partially occluded vehicles, 47% were assigned a single class and 97% of these were correct.

This work also uncovered an emerging challenge facing most vehicle classification technologies: separating commuter cars from motorcycles. The two groups have similar lengths, axle spacing and height, though they differ in width and likely in weight. With increased interest in classifying motorcycles correctly, combined with more commuter cars on the road, there is a need to devise a means to separate the two types of vehicles.

Alternatively, recognizing the difficulty in distinguishing pairs of vehicle classes with the existing detector infrastructure (e.g., commuter cars and motorcycles, short SUT and PV), there may be a need to create buffer classes to impart greater confidence in the reported classifications, e.g., adding a new "class 3 or class 5" bin to the axle-based decision tree that takes the upper portion of class 3 and lower portion of class 5 axle spacings in Figure 2-3. Thus confining the uncertainty to a much smaller number of vehicles and ensuring much greater confidence that anything that is classified as "strictly class 5" is indeed class 5.

In Chapter 4 we tackle the labor demands required to undertake the detailed study in Chapter 2. Vehicle classification data are critical to many transportation applications, but the quality of data collected depends on the operating agency to periodically calibrate, test, and validate the performance of classification sensors. These studies are labor intensive and coarse, allowing over counting errors to cancel undercounting errors. To address these challenges, this study develops a classification performance monitoring system to allow operating agencies to automatically monitor the health of their classification stations. We eliminate most of the labor demands and instead, deploy a LIDAR based portable non-intrusive vehicle classification system (PNVCS) to classify vehicles, concurrent with existing classification stations. To prevent classification errors from canceling one another in aggregate, we record pvr data in the field from both systems. After the field collection the classification results are evaluated on a per-vehicle basis. If the two systems agree, the given vehicle is automatically taken as a success by the classification station. The PNVCS includes a video camera to allow a human to assess the discrepancies. A human only looks at a given vehicle when the two systems disagree, and we developed tools to semi-automate the manual validation process, greatly increasing the efficiency and accuracy of the human user. The datasets in this study take only a few minutes for the user to validate an hour of pvr data. Although we use a LIDAR based system, the tools at the heart of the methodology are transferable to many PNVCS such as the TIRTL or AxleLight. This pilot study used LIDAR sensors mounted on a van. This approach offers a distinct advantage over the other PNVCS since our system does not require any calibration in the field, in fact the van can be classifying vehicles as it pulls up to the site. For longer-term deployments we envision a dedicated trailer that could be parked alongside the road.

The evaluation datasets come from several different classification stations, they include over 21,000 vehicles. We separately evaluated length-based classification stations and axle-based classification stations, each yielding similar results. In each case about 8% of the vehicles required manual intervention. In this study the user typically spent 3-5 sec per vehicle reviewed (including seek time and loading time). The automated process does the bulk of the work, in this study it typically took the human only a few minutes to process the exceptions from all lanes over one hour of data.

The Chapter 4 evaluation also revealed a chronic problem detecting motorcycles at the two ODOT permanent classification stations studied. While the LIDAR system detected 15 passing motorcycles, the classification stations correctly classified one of them, and missed five altogether.

As this research has shown, there is wide variance in performance from one station to the next and these errors tend to have a higher frequency among the truck classes, particularly the SUT. Since these errors are a function of the specific station, there would be benefit in the short term if ODOT were to leverage the LIDAR based PNVCS system developed in this research to evaluate the performance of many other classification stations. Thereby catching systematic errors that bias classification performance at the given station.

6 IMPLEMENTATION PLAN

First there are the classification errors that cannot be easily fixed with the existing detectors. We found that there are systematic errors arising from the vehicle fleet, they are most pronounced among the short single unit trucks and passenger vehicles. These two groups have an overlapping range of axle spacing and of vehicle length. So it is likely impossible to completely segregate the two groups with the existing detectors. Furthermore, upon reviewing the literature with this problem in mind, these errors appear to impact most of the length and axle based vehicle classification sensors, it is not limited to loop detector based systems. The problem can be accommodated in the accounting, however, by creating a buffer class that would catch most of the errors, ensuring that those vehicles far from the threshold (axle spacing or vehicle length) are classified with very high confidence. A similar problem was found between commuter cars (e.g., the Smart Car) and motorcycles and this problem will become more pronounced as the number of commuter cars increases.

Based on our investigation we feel the conventional solution is sub-optimal, i.e., attempting to find the threshold that yields an unbiased error rate, i.e., the number of over counting errors roughly cancel the number of undercounting errors. The conventional approach relies on two assumptions: (i) that the detectors are calibrated consistently across stations, when in fact sensitivity and responsiveness varies widely among detectors even after field calibration (see, e.g., [31]), and (ii) that the mix of vehicles is static when in fact the optimal threshold depends on the mix of vehicles. The approach offers no way of dynamically responding to changes in the composition of the passing fleet (see, e.g., [32] for examples of how the fleet changes dramatically over a single day). Recognizing the difficulty in distinguishing pairs of vehicle classes with the existing detector infrastructure, there may be a need to create buffer classes to impart greater confidence in the reported classifications, e.g., adding a new "class 3 or class 5" bin to the axle-based decision tree that takes the upper portion of axle class 3 and lower portion of axle class 5 axle spacings in Figure 2-3. Thus confining the uncertainty to a much smaller number of vehicles and ensuring much greater confidence that anything that is classified as "strictly class 5" is indeed axle class 5.

In any event, all subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of single unit trucks with passenger vehicles. The blurring also means that one cannot blindly use an axle classification station to calibrate the boundary between passenger vehicles and single unit trucks for length-based classification stations, otherwise, the unavoidable errors in the axle classification will be amplified in the length-based classification scheme.

Second are the classification errors that can be easily fixed. Upon the detailed review of the per-vehicle data, about a third of the classification errors could be eliminated by adjusting the classification decision tree. The original ODOT classification decision tree is shown in Table B-2 (Appendix B) and our modified axle-based classification decision tree is shown in Table 2-6. We used the data from one station to calibrate the new tree and evaluated the performance at another station. After making the changes, the axle-based classification decision tree was able to correctly classify an additional 10% of the SUT, with smaller improvements in almost every other metric. Ideally the new decision tree should be deployed at a few new locations and the performance validated, then assuming no problems are found, be adopted as the new standard classification decision tree. In any event, most of the improvements of the new decision tree should be incorporated in to standard practice (closing the gaps between bins, adding an "unclassifiable" class, and allowing for more than 4 axles in axle class 7). If communication costs are not a constraint, an even better solution would be to collect the pvr data, thereby allowing post-processing of the data, and thus, ODOT can apply new classification decision trees to historic data.

Third come the errors that have yet to be discovered. As already noted, each classification station is calibrated differently. The quality of data collected depends on the operating agency to periodically calibrate, test, and validate the performance of classification sensors. Such studies are labor intensive and coarse, allowing over counting errors to cancel undercounting errors. To address these challenges, we developed a classification performance monitoring system to automatically monitor the health of the classification stations. We eliminate most of the labor demands and instead, deploy a LIDAR based portable non-intrusive vehicle classification system (PNVCS) to classify vehicles, concurrent with existing classification stations. To prevent classification errors from canceling one another in aggregate, we evaluate the data on a per-vehicle record basis. A human only looks at a given vehicle when the two systems disagree, and we developed tools to semi-automate the manual validation process, greatly increasing the efficiency and accuracy of the human user. The datasets in this study take only a few minutes for the user to validate an hour of data. Although we use a LIDAR based system, the tools at the heart of the methodology are transferable to many PNVCS such as the TIRTL or AxleLight. This pilot study used LIDAR sensors mounted on a van. This approach offers a distinct advantage over the PNVCS since our system does not require any calibration in the field, in fact the van can be classifying vehicles as it pulls up to the site. For longer-term deployments we envision a dedicated trailer that could be parked alongside the road.

The Chapter 4 evaluation also revealed a chronic problem detecting motorcycles at the two ODOT permanent classification stations studied. While the LIDAR system detected 15 passing motorcycles, the classification stations correctly classified one of them, and missed five altogether. The LIDAR based PNVCS also offers a means to rapidly evaluate refinements in the conventional classification scheme, e.g., evaluating solutions to the large number of motorcycles that are misclassified or pass completely undetected.

As this research has shown, there is wide variance in performance from one station to the next and these errors tend to have a higher frequency among the truck classes, particularly the SUT. Since these errors are a function of the specific station, there would be benefit in the short term if ODOT were to leverage the system developed in this research to evaluate the performance of many other classification stations. Thereby catching systematic errors that bias classification performance at the given station.

This research and the outcomes have the promise to improve the accuracy of vehicle classification, which impacts operating agencies at many levels. The specific steps to implementation depend on the depth that ODOT wishes to pursue a given thrust. Some of the advances should be little or no cost, e.g., refining the classification decision tree. However, to ensure the changes are in the right direction ultimately someone would have to monitor progress, that task could either be handled by ODOT staff or be the subject of future research.

7 BIBLIOGRAPHY

[1] Federal Highway Administration. *Traffic Monitoring Guide*. USDOT, Office of Highway Policy Information, FHWA-PL-01-021, 2001.

[2] Coifman, B., Kim, S. Speed Estimation and Length Based Vehicle Classification from Freeway Single Loop Detectors. *Transportation Research Part-C*, Vol 17, No 4, 2009, pp 349-364.

[3] Kotzenmacher, J., Minge, E., Hao, B. *Evaluation of Portable Non-Intrusive Traffic Detection System*. Minnesota Department of Transportation, 2005, MN-RC-2005-37.

[4] Zwahlen, H. T., Russ, A., Oner, E. and M. Parthasarathy. Evaluation of Microwave Radar Trailers for Non-intrusive Traffic Measurements. *Transportation Research Record, no. 1917*, 2005, pp127-140.

[5] French, J., French, M. *Traffic Data Collection Methodologies*. Pennsylvania Department of Transportation, Contract 04-02 (C19), 2006.

[6] Yu, X., Prevedouros, P., Sulijoadikusumo, G. Evaluation of Autoscope, SmartSensor HD, and Infra-Red Traffic Logger for Vehicle Classification. *Transportation Research Record, no. 2160*, 2010, pp. 77–86.

[7] Banks, J. Evaluation of Portable Automated Data Collection Technologies: Final Report. California PATH Research Report, 2008, UCB-ITSPRR-2008-15.

[8] Minge, E. *Evaluation of Non-Intrusive Technologies for Traffic Detection*. Minnesota Department of Transportation, 2010, Final Report #2010-36.

[9] Schwach, J., Morris, T., Michalopoulos, P., *Rapidly. Deployable Low-Cost Traffic Data and Video Collection Device*. Center for Transportation Studies, University of Minnesota, 2009, CTS 09-21.

[10] Coifman, B. Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infrastructure. *Transportation Research Record, no. 1643*, Transportation Research Board, 1998, pp 181-191.

[11] Coifman, B. *Vehicle Classification from Single Loop Detectors*, Project 05-02, Midwest Regional University Transportation Center, University of Wisconsin, Madison, 2007

[12] Khattak, A. J., S. Hallmark, and R. Souleyrette. Application of Light Detection and Ranging Technology to Highway Safety. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1836, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 7-15

[13] Tsai, Y., Q. Yang, and Y. Wu. Identifying and Quantifying Intersection Obstruction and Its Severity Using LiDAR Technology and GIS Spatial Analysis. *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2011

[14] Veneziano, D., R. Souleyrette, and S. Hallmark. Integration of Light Detection and Ranging Technology with Photogrammetry in Highway Location and Design. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1836, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 1-6

[15] Souleyrette, R., S. Hallmark, S. Pattnaik, M. O'Brien, and D. Veneziano, *Grade and Cross Slope Estimation from LIDAR-Based Surface Models.* MTC Project 2001-02, FHWA, U.S. Department of Transportation, 2003

[16] Chenoweth, A. J., R. E. McConnell, and R. L. Gustavson. Overhead Optical Sensor for Vehicle Classification and Axle Counting. *Mobile robots XIV: proceedings of SPIE, vol. 3838*, Boston, MA, 1999

[17] Cunagin, W. D. and D. J. Vitello Jr. Development of an Overhead Vehicle Sensor System, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1200, Transportation Research Board of the National Academies, Washington, D.C., 1988, pp. 15-23

[18] Yao, W., S. Hinz, and U. Stilla. Traffic Monitoring from Airborne Full-Waveform LIDAR – Feasibility, Simulation and Analysis. *International Archives of the Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences*, Vol. 37(B3B), pp. 593-598, Beijing 2008

[19] Grejner-Brzezinska, D. A, C. Toth, and M. McCord. *Airborne LiDAR: A New Source of Traffic Flow Data.* FHWA/OH-2005/14, FHWA, U.S. Department of Transportation, 2005

[20] Yang, R., Vehicle Detection and Classification from a LIDAR Equipped Probe Vehicle, Masters Thesis, The Ohio State University, 2009

[21] Redmill, K., Coifman, B., McCord, M., Mishalani, R., Using Transit or Municipal Vehicles as Moving Observer Platforms for Large Scale Collection of Traffic and Transportation System Information, *Proc. of the 14th International IEEE Conference on Intelligent Transportation Systems*, Oct 5-7, 2011, Washington, DC.

[22] FHWA, Federal Size Regulations for Commercial Motor Vehicles, Federal Highway Administration, 2000

[23] TRB, *Highway Capacity Manual*. Transportation Research Board of the National Academies, Washington, DC, 2000

[24] http://controlspecialist2.com/TIRTL.aspx, accessed November 4, 2011.

[25] http://www.peektraffic.com/datasheets/axlelightdatasheet.pdf, accessed November 4, 2011.

[26] Little, G., M. Johnson, and P. Zidek. *Off-Road Axle Detection Sensor (ORADS) : Final Technical Report*, Ohio Department of Transportation, 2001, FHWA/HWY-04/2001

[27] http://www.spectra-research.com/inner/electro.htm, accessed November 4, 2011.

[28] Minge, E., S. Petersen, and J. Kotzenmacher. Evaluation of Non-Intrusive Technologies for Traffic Detection - Phase 3. *Proc. of the 90th Annual Meeting of the Transportation Research Board*, 2011.

[29] Coifman, B. and M. Cassidy. Vehicle Reidentification and Travel Time Measurement on Congested Freeways. *Transportation Research: Part A*, Vol 36, No 10, 2002, pp. 899-917.

[30] Caltrans, 2007, VideoSync, http://www.dot.ca.gov/research/operations/videosync, accessed on December 2, 2011.

[31] Lee, H., and Coifman, B. Identifying and Correcting Pulse-Breakup Errors from Freeway Loop Detectors, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2256, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 68-78.

[32] Coifman, B. Improved Velocity Estimation Using Single Loop Detectors, *Transportation Research: Part A,* vol 35, no 10, 2001, pp. 863-880.

8 APPENDIX A: DETAILS OF THE CLASSIFICATION STATIONS

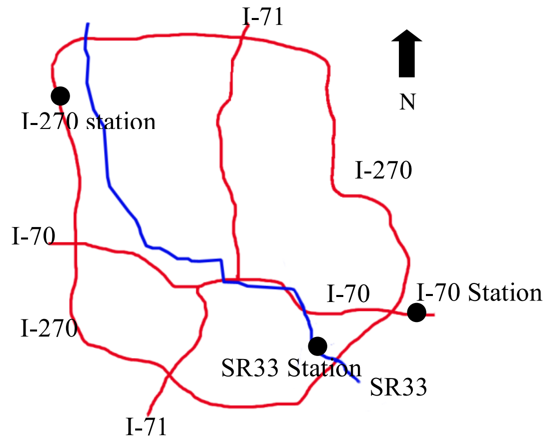


Figure A-1, Location of axle classification stations.

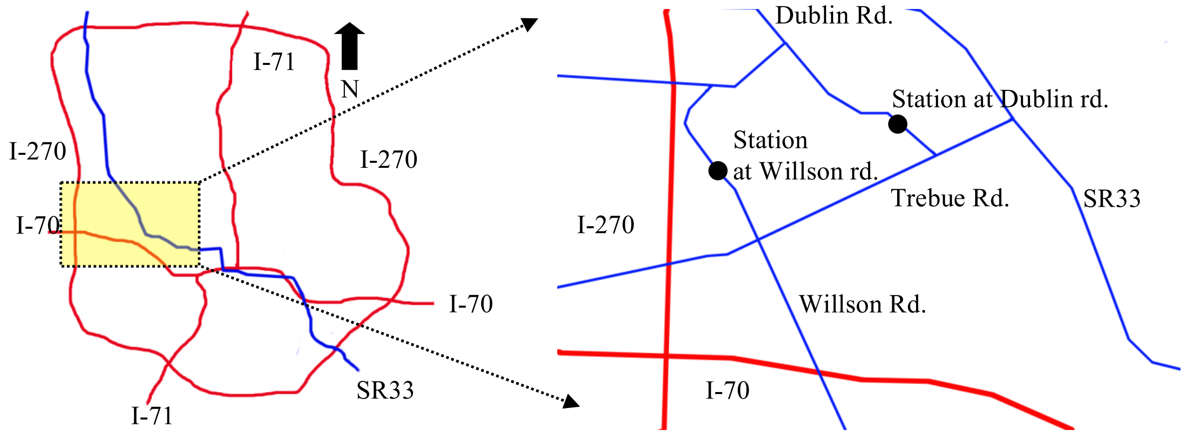


Figure A-2, Location of tube classification sites.

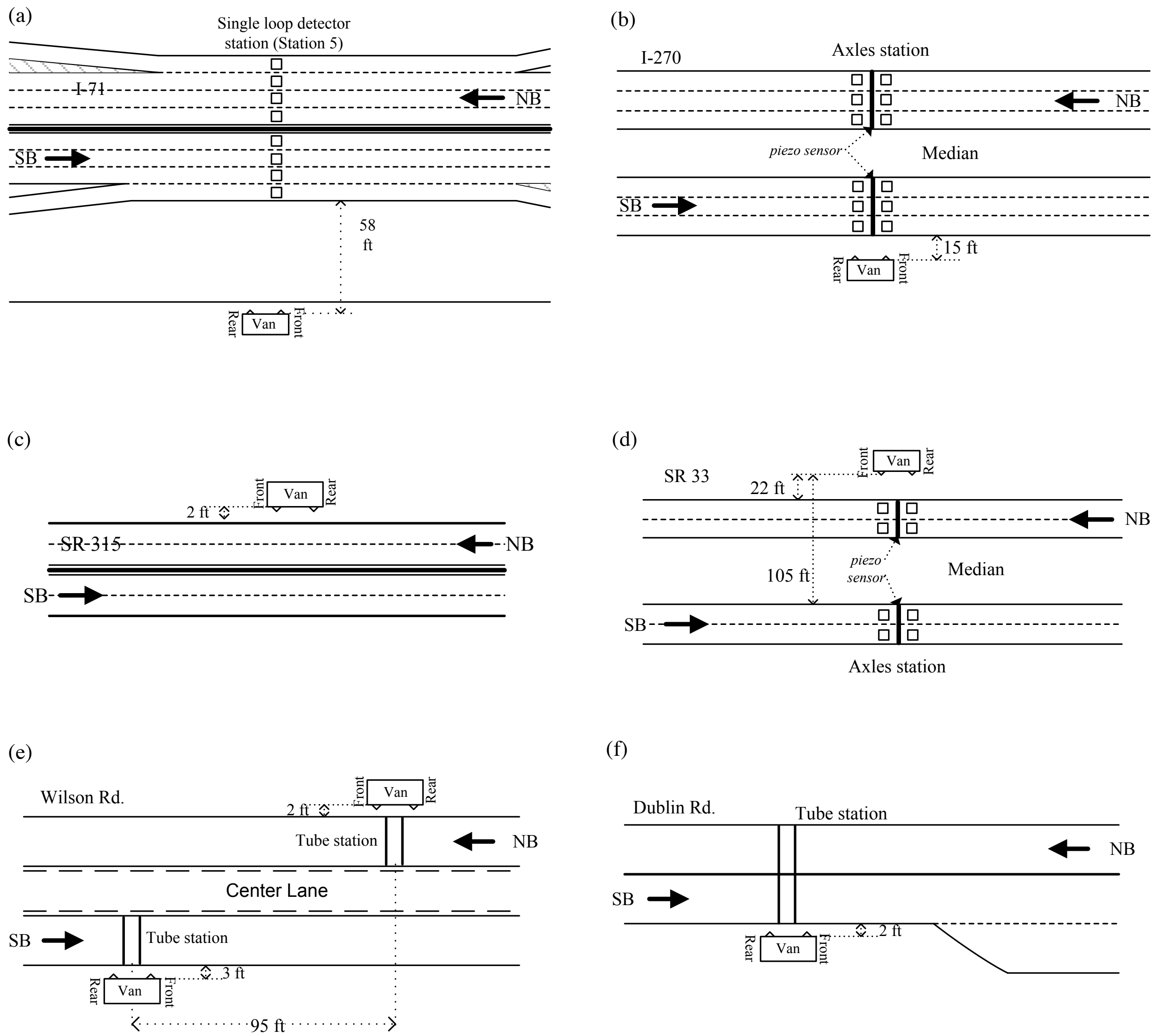
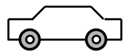
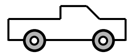
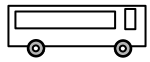
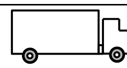
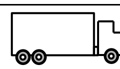
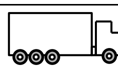
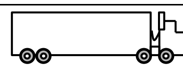
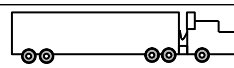
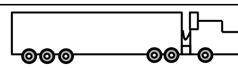
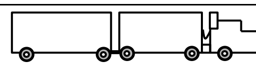
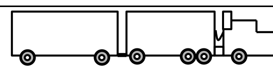
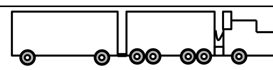


Figure A-3, Schematic of locations LIDAR data collected: (a) I-71 southbound, (b) I-270 southbound, (c) SR-315 northbound, (d) SR-33 northbound and southbound, (e) Wilson Rd northbound and southbound, (f) Dublin Rd northbound and southbound

Table A-1, 13 FHWA axle-based vehicle classes

Vehicle Class	Example (not to scale)	3 groups	4 groups	6 groups
Class 1: Motorcycle		PV	MC	MC
Class 2: Car		PV	PV	PV PVPT
Class 3: other 2 axle, 4 tire single-unit vehicle		PV	PV	PV PVPT
Class 4: Bus		SUT	SUT	SUT SUTPT
Class 5: 2 axle, 6 tire, single-unit truck		SUT	SUT	SUT SUTPT
Class 6: 3 axle single-unit truck		SUT	SUT	SUT SUTPT
Class 7: 4 or more axle single-unit truck		SUT	SUT	SUT SUTPT
Class 8: 4 or fewer axle single-trailer truck		MUT	MUT	MUT
Class 9: 5 axle single-trailer truck		MUT	MUT	MUT
Class 10: 6 or more axle single-trailer truck		MUT	MUT	MUT
Class 11: 5 or fewer axle multi-trailer truck		MUT	MUT	MUT
Class 12: 6 axle multi-trailer truck		MUT	MUT	MUT
Class 13: 7 or more axle multi-trailer truck		MUT	MUT	MUT

Note: any class 1 through class 3 vehicle pulling a trailer should be assigned the same class as if the vehicle were not pulling a trailer. However any single-unit truck pulling a trailer should be assigned to a multi-unit truck class (e.g., a class 6 pulling a trailer with two axles is treated as a class 9). The final three columns map the axle class to the three different groupings used in this study. So in the 3 groups and 4 groups PV pulling trailers (PVPT) are included with PV, and SUT pulling trailers (SUTPT) are included with MUT. While in the 6 groups PVPT and SUTPT are kept separate.

(This page intentionally left blank)

9 APPENDIX B: DETAILS OF PVR AND CLASSIFICATION SCHEMES ON I-270 AT RINGS RD.

Table B-1 shows samples of 9 per-vehicle records from I-270. Axle bin indicates the original axle class assigned by the classifier. Similarly, the length bin represents length class that is determined by vehicle length in the pvr. Table B-2 enumerates the existing ODOT decision tree used to assign axle class based on the number of axles and axle spacing in the pvr.

Table B-1, Details of pvr data in I270

HH:MM:SS	Lane *	Speed (mph)	# of axles	Length (ft)	Axle bin	Length bin	Axle Spacing(ft)**			
							S_1	S_2	S_3	S_4
9:27:56	5	63.2	3	28.4	6	2	16.7	5.1		
9:27:56	4	67	2	13	2	1	9.1			
9:27:58	6	59.9	5	73.6	9	3	17.3	4.7	33.8	4.4
9:27:58	5	62.4	2	15	2	1	10.2			
9:27:58	4	68.3	2	20.5	3	1	13.9			
9:28:00	5	64.7	2	15.8	2	1	10			
9:28:01	6	57.9	2	13.2	2	1	8.9			
9:28:02	5	64.6	2	12.9	2	1	8.9			
9:28:03	5	64.7	4	45.2	3	3	13.8	18.4	3.3	

* : Lane 4, 5, and 6 correspond to lane 1 (median), 2, and 3(shoulder) on southbound I-270
** : S_i indicates axle spacing (ft) between i^{th} axle and $i+1^{th}$ axle

Table B-2, ODOT axle based classification scheme

Class	# of axles	Class name	Spacing (ft)
1	2~3	Motorcycle	1~5.8, any
2	2~3	Car	5.9~10.2, 10~18.8
3	2~3	other 2axle, 4tire single-unit veh	10.3~15, 10~18.8
5	2	2 axle, 6tire, single-unit truck	15.1~24
4	2~3	Bus	23.5~99.9, any
8	3	4 or fewer axle single-trailer truck	any, 18.1~99.9
6	3	3 axle single-unit truck	any, 3.5~8
2	4~5	Class 2 pulling a trailer	1~10.2, any, 1~3.4, 1~3.4
3	4~5	Class 3 pulling a trailer	10.3~15, any, 1~3.4, 1~3.4
8	4	4 or fewer axle single-trailer truck	any, 5.1~99.9, 3.5~99.9
8	4	4 or fewer axle single-trailer truck	any, 1~5, 10~99.9
7	4	4 or more axle single-unit truck	any, any, any
11	5	5 or fewer axle multi-trailer truck	any, 6.1~99.9, any, any
9	5	5 axle single-trailer truck	any, 1~6, any, 3.5~11
3	5	other 2axle, 4tire single-unit veh w/ a trailer	9.9~14.9, any, any, 1~3.4
5	5	2 axle, 6tire, single-unit truck w/ a trailer	15.1~24, any, any, 1~3.4
9	5	5 axle single-trailer truck	any, any, any, any
10	6	6 or more axle single-trailer truck	any, 3.5~8, 3.5~8, any, 8.1~99.9
12	6	6 axle multi-trailer truck	any, any, any, any, 8.1~99.9
10	6~10	6 or more axle single-trailer truck	any, any, any, any, 3.5~8, 3.5~8, 3.5~8, 3.5~8, 3.5~8

Note that the classifier proceeds through the decision tree in the table a vehicle is assigned to the first test that it passes. So if no class is found for the vehicle, that vehicle is assigned to class 13. This table represents the default settings and the thresholds manually set in the field classifier may be different. For example, at the I-270 station it turns out that there is an 0.5 ft difference between the thresholds in this table and those that we empirically deduced. Thus we include this offset when we apply the classification decision tree in this table to the I-270 dataset (e.g., Figure 2-5 and 2-6, and Table D-1).

10 APPENDIX C: PERFORMANCE OF THE CLASSIFICATION STATIONS

Table C-1, Comparison between pvr and ground truth axle-class in (a) the I-70 dataset, (b) the SR-33 dataset

(a)

Axle classification station in I-70		Axle based vehicle classification													% of row correct
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13	
Manual FHWA 13 vehicle classification	class 1: Motorcycle	31	-	-	-	-	-	-	-	-	-	-	-	-	
	class 2: Car	-	3663	21	-	-	-	2	-	-	-	-	-	1	99.7%
	class 3: other 2axle, 4tire single-unit veh	1	2278	1500	-	1	-	-	20	-	-	-	-	-	
	class 4: Bus	-	-	1	2	11	-	-	5	-	-	-	-	-	
	class 5: 2 axle, 6tire, single-unit truck	-	1	105	3	136	-	-	-	-	-	-	-	-	60.9%
	class 6: 3 axle single-unit truck	-	-	-	1	-	95	4	-	-	-	-	-	1	
	class 7: 4 or more axle single-unit truck	-	-	-	-	-	1	1	2	9	39	-	-	-	
	class 8: 4 or fewer axle single-trailer truck	-	-	12	-	-	1	1	55	-	-	-	-	-	
	class 9: 5 axle single-trailer truck	-	1	-	-	-	15	-	10	1236	1	1	-	-	
	class 10: 6 or more axle single-trailer truck	-	-	-	-	-	1	-	1	6	29	-	-	-	97.8%
	class 11: 5 or fewer axle multi-trailer truck	-	-	-	-	1	-	-	2	-	-	46	-	-	
	class 12: 6 axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	14	-	
	class 13: 7 or more axle multi-trailer truck	-	-	-	-	-	-	-	-	-	-	-	-	1	
% of column correct		98.4%			92.0%				94.8%						

(b)

Axle classification station in SR-33		PVR axle class													% of row correct
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13	
Manual ground truth	PV	-	908	219	-	2	-	-	4	-	-	-	-	-	99.5%
	SUT	-	-	14	2	21	19	2	1	-	15	-	-	-	59.5%
	MUT	-	-	-	-	-	1	-	3	42	1	1	-	-	97.9%
% of column correct		98.8%			93.6%				70.2%						

10.1 Performance of the Pneumatic Tubes

Although not included in the results presented in Chapter 2, we also manually examined the performance of the pneumatic tube sites. The results are summarized in this section.

Table C-2, Summary statistics of ground truth data sets from pneumatic tubes.

Location	Tube stations		
	Southbound Dublin Rd.	Southbound Wilson Rd.	Northbound Wilson Rd.
Date	Oct 28,2010		
Road Type	Arterial		
Traffic Conditions	Moderate	Moderate	Moderate
Time duration investigated	7:42~8:56 14:30~15:54	10:18~10:57 17:00~18:00	9:08~9:54 16:02~16:53
Average Speed	36mph,38mph	41mph, 38mph	38mph,38mph
Average Flow (per lane)	689vph, 326vph	220vph, 561vph	325vph,480vph
# of vehicles	1317	712	664
# of obscured vehicles	0	0	0
Resolution of ground-truth	PV, SUT, MUT	PV, SUT, MUT	PV, SUT, MUT

Table C-3, Summary statistics of axle classification, for reference the top three rows come from Table 2-4.

Station	Location	% of ground-truth vehicle correctly classified			% correct of axle classification			Overall
		PV	SUT	MUT	PV	SUT	MUT	
Axle station	I270 (SB)	99.6%	64.2%	96.5%	98.5%	97.1%	92.8%	97.9%
	I70 (EB)	99.7%	61.0%	97.8%	98.4%	92.4%	94.8%	97.7%
	SR33 (NB)	99.5%	59.5%	97.1%	98.8%	93.6%	70.1%	97.1%
Tube station	Dublin Rd. (SB)	94.5%	83.6%	100%	99.3%	61.3%	21.2%	94.1%
	Wilson Rd. (SB)	97.5%	90.9%	100%	99.9%	62.5%	60.7%	97.5%
	Wilson Rd. (NB)	96.5%	93.3%	100%	99.5%	80.8%	59.3%	96.4%

Table C-4, Comparison between pvr and ground truth axle-class in (a) the Dublin Rd. dataset, (b) the southbound Wilson Rd. dataset, (c) the northbound Wilson Rd. dataset

(a)

Axle classification station in Dublin Rd.		PVR axle class													% of row correct
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13	
Manual ground truth	PV	1	975	206	4	25	-	-	40	-	-	-	-	-	94.5%
	SUT	-	3	5	18	23	4	1	1	-	-	-	-	-	83.6%
	MUT	-	-	-	-	-	-	-	10	1	-	-	-	-	100%
% of column correct		99.3%			61.3%				21.2%						

(b)

Axle classification station in southbound Wilson Rd.		PVR axle class													% of row correct
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13	
Manual ground truth	PV	1	588	78	-	6	-	-	10	-	-	1	-	-	97.5%
	SUT	-	-	1	2	5	3	-	-	-	-	-	-	-	90.9%
	MUT	-	-	-	-	-	-	-	13	4	-	-	-	-	100%
% of column correct		99.9%			62.5%				60.7%						

(c)

Axle classification station in northbound Wilson Rd.		PVR axle class													% of row correct
		class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class 10	class 11	class 12	class 13	
Manual ground truth	PV	1	472	109	1	9	-	-	11	-	-	-	-	-	96.5%
	SUT	-	-	3	4	37	1	-	-	-	-	-	-	-	93.3%
	MUT	-	-	-	-	-	-	-	4	11	-	1	-	-	100%
% of column correct		99.5%			80.8%				59.3%						

Table C-5, Comparison between pvr axle class and ground truth vehicle type in (a) the Dublin Rd. dataset, (b) the southbound Wilson Rd. dataset, (c) the northbound Wilson Rd. dataset

(a)

Dublin Rd. (SB)		PVR axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	1182	29	40	94.5%
	SUT	8	46	1	83.6%
	MUT	0	0	11	100%
% of column correct		99.3%	61.3%	21.2%	94.1%

(b)

Wilson Rd (SB)		PVR axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	667	6	11	97.5%
	SUT	1	10	0	90.9%
	MUT	0	0	17	100%
% of column correct		99.9%	62.5%	60.7%	97.5%

(c)

Wilson Rd (NB)		PVR axle class			% of row correct
		Passenger	Single-unit truck	Multi-unit truck	
Manual ground truth	PV	582	10	11	96.5%
	SUT	3	42	0	93.3%
	MUT	0	0	16	100%
% of column correct		99.5%	80.8%	59.3%	96.4%

11 APPENDIX D: TWO-AXLE SUT WITH SHORT AXLE SPACING

Figure D-1 shows a histogram of axle spacing for misclassified class 5 with the thresholds (after accounting for the offset discussed in Appendix B) of class 2 and class 3 in Table B-2.

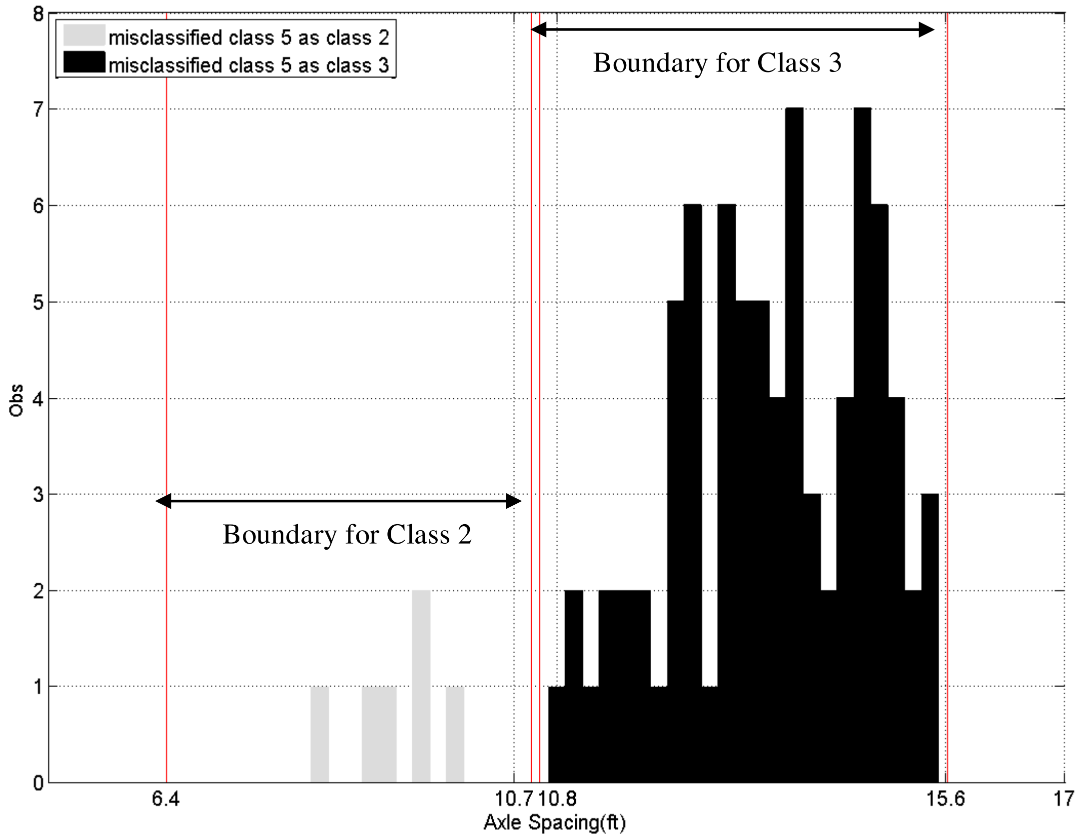


Figure D-1, Minimum and maximum axle spacing of misclassified class 5 as class 2 and class 3

(This page intentionally left blank)

12 APPENDIX E: LIDAR BASED VEHICLE CLASSIFICATION BY LOCATION

Table E-1, Comparison of LIDAR based vehicle classification and actual vehicle class from I-71 southbound free flow.

I-71 SB FF			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**9**	1	0	0	0	0	10	90.0%	1
	PV*	PV	0	**1,390**	2	0	0	0	1,392	99.9%	297
		PVPT	0	1	**14**	0	0	0	15	93.3%	3
	SUT		0	0	1	**33**	0	2	36	91.7%	1
	MUT*	SUPT	0	0	0	0	**1**	0	1	100%	0
		MUT	0	0	0	0	1	**47**	48	97.9%	9
Number of vehicles from LIDAR vehicle classification			9	1,392	17	33	2	49	1,502	99.5%	311
% correct			100%	99.9%	82.4%	100%	50%	95.9%	99.5%		

Table E-2, Comparison of LIDAR based vehicle classification and actual vehicle class from I-71 southbound mild-congested.

I-71 SB Mild-congested			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**0**	0	0	0	0	0	0	100%	0
	PV*	PV	1	**1,939**	1	4	0	0	1,945	99.7%	568
		PVPT	0	0	**9**	0	0	0	9	100%	1
	SUT		0	1	0	**34**	0	0	35	97.1%	5
	MUT*	SUPT	0	0	0	0	**1**	0	1	100%	0
		MUT	0	0	0	0	0	**38**	38	100%	17
Number of vehicles from LIDAR vehicle classification			1	1,940	10	38	1	38	2,028	99.7%	591
% correct			0%	99.9%	90.0%	89.5%	100%	100%	99.7%		

Table E-3, Comparison of LIDAR based vehicle classification and actual vehicle class from I-270 southbound free flow.

I-270 SB			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**3**	2	0	0	0	0	5	60.0%	0
	PV*	PV	6	**10,205**	2	11	0	0	10,224	99.8%	1,156
		PVPT	0	2	**138**	6	3	1	150	92.0%	20
	SUT		0	20	4	**479**	3	2	508	94.3%	49
	MUT*	SUPT	0	0	4	1	**21**	5	31	67.7%	2
		MUT	0	0	3	7	5	**1,088**	1,103	98.6%	149
Number of vehicles from LIDAR vehicle classification			9	10,229	151	504	32	1,096	12,021	99.3%	1,376
% correct			33.3%	99.8%	91.4%	95.0%	65.6%	99.3%	99.3%		

Table E-4, Comparison of LIDAR based vehicle classification and actual vehicle class from SR-315 northbound free flow.

SR-315 NB			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**24**	1	0	0	0	0	25	96.0%	4
	PV*	PV	3	**6,085**	0	0	0	0	6,088	100%	642
		PVPT	0	0	**15**	0	0	0	15	100%	4
	SUT		0	5	0	**70**	0	0	75	93.3%	7
	MUT*	SUPT	0	0	0	0	**2**	0	2	100%	0
		MUT	0	0	0	1	0	**34**	35	97.1%	3
Number of vehicles from LIDAR vehicle classification			27	6,091	15	71	2	34	6,240	99.8%	660
% correct			88.9%	99.9%	100%	98.6%	100%	100%	99.8%		

Table E-5, Comparison of LIDAR based vehicle classification and actual vehicle class from Dublin Rd southbound.

Dublin Rd SB			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**2**	0	0	0	0	0	2	100%	-
	PV*	PV	0	**1,258**	0	0	0	0	1,258	100%	-
		PVPT	0	0	**19**	0	0	0	19	100%	-
	SUT		0	2	0	**51**	1	0	54	94.4%	-
	MUT*	SUPT	0	0	2	1	**3**	1	7	42.9%	-
		MUT	0	0	0	0	0	**4**	4	100%	-
Number of vehicles from LIDAR vehicle classification			2	1,260	21	52	4	5	1,344	99.5%	-
% correct			100%	99.8%	90.5%	98.1%	75.0%	80.0%	99.5%		

Table E-6, Comparison of LIDAR based vehicle classification and actual vehicle class from Wilson Rd northbound.

Wilson Rd NB			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**1**	0	0	0	0	0	1	100%	-
	PV*	PV	0	**599**	0	0	0	0	599	100%	-
		PVPT	0	0	**5**	0	0	0	5	100%	-
	SUT		0	2	0	**43**	0	0	45	95.6%	-
	MUT*	SUPT	0	0	0	0	**1**	0	1	100%	-
		MUT	0	0	0	0	0	**15**	15	100%	-
Number of vehicles from LIDAR vehicle classification			1	601	5	43	1	15	666	99.7%	-
% correct			100%	99.7%	100%	100%	100%	100%	99.7%		

Table E-7, Comparison of LIDAR based vehicle classification and actual vehicle class from Wilson Rd southbound.

Wilson Rd SB			LIDAR vehicle classification						Number of vehicles from ground truth data	% correct	Number of partially occluded vehicles that are excluded from LIDAR based vehicle classification
			MC	PV*		SUT	MUT*				
				PV	PVPT		SUPT	MUT			
Ground truth data	MC		**1**	0	0	0	0	0	1	100%	-
	PV*	PV	0	**676**	0	0	0	0	676	100%	-
		PVPT	0	0	**6**	0	0	0	6	100%	-
	SUT		0	0	0	**11**	0	0	11	100%	-
	MUT*	SUPT	0	0	0	0	**3**	0	3	100%	-
		MUT	0	0	0	1	0	**13**	14	92.9%	-
Number of vehicles from LIDAR vehicle classification			1	676	6	12	3	13	711	99.9%	-
% correct			100%	100%	100%	91.7%	100%	100%	99.9%		

13 Appendix F: Comparison of Pseudo Ground Truth Data and Axle Vehicle Classification by Location

Throughout this appendix: PV* includes passenger vehicle and passenger vehicle pulling a trailer; and MUT* includes single unit truck pulling a trailer and multiple unit truck.

Table F-1, Comparison of pseudo ground truth data and axle vehicle classification at I-270 southbound adjacent to LIDAR sensor.

I-270 SB		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	2	4	0	0	7	14%	0
	PV*	1	**10,416**	8	56	80	10,561	99%	0
	SUT	0	153	**303**	41	3	500	61%	0
	MUT*	0	28	4	**1,101**	7	1,140	97%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		2	10,599	319	1,198	90	12,208	-	0
% correct		50%	98%	95%	92%	-	-	97%	-
Totally occluded vehicles		1	252	2	2	-	257	-	-

Table F-2, Comparison of pseudo ground truth data and axle vehicle classification at Dublin Rd southbound adjacent to LIDAR sensor.

Dublin Rd SB adjacent		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	0	1	0	0	2	50%	0
	PV*	0	**1,183**	28	41	30	1,282	92%	0
	SUT	0	6	**46**	0	1	53	87%	0
	MUT*	0	0	0	**11**	0	11	100%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total		1	1,189	75	52	31	1,348	-	0
% correct		100%	99%	61%	21%	-	-	92%	-
Totally occluded vehicles		0	0	0	0	-	0	-	-

Table F-3, Comparison of pseudo ground truth data and axle vehicle classification at Wilson Rd northbound adjacent to LIDAR sensor.

Wilson Rd NB adjacent		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	0	0	0	0	1	100%	0
	PV*	0	**583**	10	11	8	612	95%	0
	SUT	0	1	**42**	0	0	43	98%	0
	MUT*	0	0	0	**16**	0	16	100%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		1	584	52	27	8	672	-	0
% correct		100%	100%	81%	59%	-	-	96%	-
Totally occluded vehicles		0	0	0	0	-	0	-	-

Table F-4, Comparison of pseudo ground truth data and axle vehicle classification at Wilson Rd southbound adjacent to LIDAR sensor.

Wilson Rd SB adjacent		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	0	0	0	0	1	100%	0
	PV*	0	**666**	6	11	10	693	96%	0
	SUT	0	1	**10**	0	0	11	91%	0
	MUT*	0	0	0	**17**	0	17	100%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		1	667	16	28	10	722	-	0
% correct		100%	100%	63%	61%	-	-	96%	-
Totally occluded vehicles		0	0	0	0	-	0	-	-

Table F-5, Comparison of pseudo ground truth data and axle vehicle classification at SR-33 northbound adjacent to LIDAR sensor.

SR-33 NB adjacent		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**0**	0	0	2	3	5	0%	0
	PV*	0	**1,057**	2	2	53	1,114	95%	0
	SUT	0	12	**44**	15	8	79	56%	0
	MUT*	0	0	1	**47**	6	54	87%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		0	1,069	47	66	70	1,252	-	0
% correct		-	99%	94%	71%	-	-	92%	-
Totally occluded vehicles		0	8	0	0	-	8	-	-

Table F-6, Comparison of pseudo ground truth data and axle vehicle classification at Dublin Rd northbound on the opposite side of LIDAR sensor.

Dublin NB opposite		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	0	0	1	0	2	50%	0
	PV*	1	**748**	21	18	7	795	94%	0
	SUT	1	5	**55**	2	0	63	87%	0
	MUT*	0	0	0	**6**	0	6	100%	0
	Non-vehicle actuation in axle data	2	2	0	0	-	4	-	-
Column total above		5	755	76	27	7	870	-	0
% correct		20%	99%	72%	22%	-	-	93%	-
Totally occluded vehicles		0	5	0	0	-	5	-	-

Table F-7, Comparison of pseudo ground truth data and axle vehicle classification at Wilson Rd northbound on the opposite side of LIDAR sensor.

Wilson NB opposite		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**0**	0	0	0	0	0	-	0
	PV*	0	**647**	7	7	7	668	97%	0
	SUT	0	1	**16**	0	0	17	94%	0
	MUT*	0	0	0	**11**	0	11	100%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		0	648	23	18	7	696	-	0
% correct		-	100%	70%	61%	-	-	97%	-
Totally occluded vehicles		0	5	0	0	-	5	-	-

Table F-8, Comparison of pseudo ground truth data and axle vehicle classification at Wilson Rd southbound on the opposite side of LIDAR sensor.

Wilson SB opposite		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**1**	0	0	0	0	1	100%	0
	PV*	0	**628**	9	12	18	667	94%	0
	SUT	0	2	**16**	0	0	18	89%	0
	MUT*	0	0	0	**10**	0	10	100%	0
	Non-vehicle actuation in axle data	0	0	0	0	-	0	-	-
Column total above		1	630	25	22	18	696	-	0
% correct		100%	100%	64%	45%	-	-	94%	-
Totally occluded vehicles		0	10	0	0	-	10	-	-

Table F-9, Comparison of pseudo ground truth data and axle vehicle classification at SR-33 southbound on the opposite side of LIDAR sensor.

SR-33 SB opposite		Axle vehicle classification				Number of LIDAR vehicles not detected by axle sensor	Row total	% correct	Non-vehicle actuation in LIDAR data
		MC	PV*	SUT	MUT*				
Pseudo ground truth data	MC	**0**	0	1	0	2	3	0%	0
	PV*	0	**1,073**	3	2	76	1,154	93%	0
	SUT	0	15	**42**	21	13	91	46%	0
	MUT*	1	2	4	**37**	3	47	79%	0
	Non-vehicle actuation in axle data	0	1	0	0	-	1	-	-
Column total above		1	1,091	50	60	94	1,296	-	0
% correct		0%	98%	84%	62%	-	-	89%	-
Totally occluded vehicles		0	65	2	1	-	68	-	-

(This page intentionally left blank)

14 Appendix G: Comparison of Pseudo Ground Truth Data and Length Based Vehicle Classification by Location

Throughout this appendix: PV** includes motorcycle, passenger vehicle, and passenger vehicle pulling a trailer; and MUT* includes single unit truck pulling a trailer and multiple unit truck.

Table G-1, Comparison of pseudo ground truth data and length based vehicle classification from I-71 southbound free flow.

I-71 SB Free flow		Length class from loop detector: Class 1	Class 2	Class 3	Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
Pseudo ground truth data	PV**	**1,372**	19	4	33	1,428	96%	0
	SUT	9	**16**	5	2	32	50%	0
	MUT*	0	3	**48**	0	51	94%	0
	Non-vehicle actuation in loop detector data	0	0	0	-	0	-	-
Column total above		1,381	38	57	35	1,511	-	0
% correct		99%	42%	84%	-	-	95%	-
Totally occluded vehicles		46	0	1	-	47	-	-

Table G-2, Comparison of pseudo ground truth data and length based vehicle classification from I-71 southbound semi-congested.

I-71 SB Semi-congested		Length class from loop detector: Class 1	Class 2	Class 3	Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
Pseudo ground truth data	PV**	**1,917**	25	1	24	1,967	97%	0
	SUT	14	**23**	1	0	38	61%	0
	MUT*	0	1	**39**	0	40	98%	0
	Non-vehicle actuation in loop detector data	0	0	0	-	0	-	-
Column total above		1,931	49	41	24	2,045	-	0
% correct		99%	47%	95%	-	-	97%	-
Totally occluded vehicles		130	0	0	-	130	-	-

Table G-3, Comparison of pseudo ground truth data and length based vehicle classification from I-270 southbound.

I-270 SB		Length class from loop detector: Class 1	Class 2	Class 3	Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
Pseudo ground truth data	PV**	**10,221**	189	56	80	10,546	97%	0
	SUT	37	**467**	2	3	509	92%	0
	MUT*	18	8	**1,120**	7	1,153	97%	0
	Non-vehicle actuation in loop detector data	0	0	0	-	0	-	-
Column total above		10,276	664	1,178	90	12,208	-	0
% correct		99%	70%	95%	-	-	97%	-
Totally occluded vehicles		249	5	3	-	257	-	-

Table G-4, Comparison of pseudo ground truth data and length based vehicle classification from SR-33 northbound.

SR-33 NB		Length class from loop detector: Class 1	Class 2	Class 3	Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
Pseudo ground truth data	PV**	**1,047**	12	2	56	1,117	94%	0
	SUT	17	**56**	0	8	81	69%	0
	MUT*	1	3	**44**	6	54	81%	0
	Non-vehicle actuation in loop detector data	0	0	0	-	0	-	-
Column total above		1,065	71	46	70	1,252	-	0
% correct		98%	79%	96%	-	-	92%	-
Totally occluded vehicles		8	0	0	-	8	-	-

Table G-5, Comparison of pseudo ground truth data and length based vehicle classification from SR-33 southbound.

SR-33 SB		Length class from loop detector			Number of LIDAR vehicles not detected by loop detector	Row total	% correct	Non-vehicle actuation in LIDAR data
		Class 1	Class 2	Class 3				
Pseudo ground truth data	PV**	**1,066**	11	3	78	1,158	92%	0
	SUT	48	**28**	0	13	89	31%	0
	MUT*	2	8	**35**	3	48	73%	0
	Non-vehicle actuation in loop detector data	1	0	0	-	1	-	-
Column total above		1,117	47	38	94	1,296	-	0
% correct		95%	60%	92%	-	-	87%	-
Totally occluded vehicles		65	2	1	-	68	-	-