The Ohio Department of Transportation
Office of Statewide Planning and Research
Research Section
1980 West Broad Street
Columbus, OH 43207
614-644-8135
Research@dot.state.oh.us
www.dot.state.oh.us/Research

Executive Summary Report

Validating the Performance of Vehicle Classification Stations

FHWA Report Number:	FHWA/OH-2012/7
Report Publication Date:	May 2012
ODOT State Job Number:	134516
Project Duration:	22 months
Start Date:	July 29, 2010
Completion Date:	May 29, 2012
Total Project Funding:	$89,949
Research Agency:	The Ohio State University
Researchers:	B. Coifman, H. Lee, S. Kim
ODOT Project Manager:	
ODOT Subject Matter Experts:	David Gardner, Lindsey Pflum

For copies of this final report go to http://www.dot.state.oh.us/research.

Project Background

Vehicle classification data are used in many transportation applications, including: pavement design, environmental impact studies, traffic control, and traffic safety. 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. Each sensor technology has its own strengths and weaknesses regarding costs, accuracy, 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.

To address these challenges, 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 (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.

Study Objectives

1) Use manual data reduction from concurrent video to evaluate how well ODOT classification stations perform.
2) Identify any chronic problems in the classification performance so that ODOT can ensure accurate vehicle classification.
3) Evaluate the axle based classification tree for the studied locations to see if any improvements

can be realized.

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

Description of Work

In the first part of the study we use concurrent video based ground truth to manually evaluate the performance of three freeway permanent vehicle classification stations providing length and axle based classification, one freeway traffic monitoring station providing length based classification, and two arterial temporary pneumatic tube deployments providing axle 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. The primary focus was the classification stations, and we examined 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 ODOT axle classifier. 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 revised 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 the second part of the study 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, and then cluster 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. 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.

In the third part, we seek to automate the process of evaluating the classification stations, i.e., addressing the problem in the first part with the tools from the second part. 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, 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 just one of them, and missed five altogether.

Research Findings & Conclusions

1) At the existing permanent classification stations overall performance was good, with only 3%-4% of the vehicles being 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 classification decision tree.
2) Diagnosing the axle classification errors, it was found that all of the observed errors could be attributed to one of six causes. About two thirds of the errors were unavoidable, due to overlapping axle spacing ranges or length ranges between neighboring classes. Upon reviewing the literature with this problem in mind, the figures and tables show evidence that the disproportionate relative errors appear to impact most of the length and axle based vehicle classification sensors in those studies, though the relative errors are rarely mentioned in the various reports.
3) The remaining 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, revising the tests, and adding an additional outcome from the tree to indicate a vehicle is unclassifiable. Our revised decision tree is shown in the report (Table 2-6). After making these changes our new axle-based classification decision tree was able to catch almost all of these errors, leading to an additional 10% of the SUT being correctly classified, with smaller improvements in almost every other metric.
3) 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 (akin to the SUT problem above). With increased interest in correctly classifying motorcycles combined with more commuter cars on the road there is a need to devise a means to separate the two types of vehicles.
4) The LIDAR based PNVCS worked well and was able to catch all of the chronic errors exhibited by the classification stations under review, even across four lanes of traffic.
5) The two permanent classification stations evaluated with the LIDAR PNVCS exhibited chronic problems detecting motorcycles. While the LIDAR system detected 15 passing motorcycles, the classification stations correctly classified just one of them, and missed five altogether.

Implementation Recommendations

1) This work developed a new axle-based classification decision tree that caught about a third of the errors arising from the existing decision tree, with the greatest impact on SUT (roughly a 10% improvement in the success rate). 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.
2) The overlapping range of axle spacings and vehicle lengths across different classes 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.
3) Similarly, all subsequent uses of the classification data (e.g., planning and measuring freight flows) must accommodate this unavoidable blurring of SUT with PV.
4) 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.
5) 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) 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 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) The LIDAR based PNVCS offers a means to rapidly evaluate refinements in the conventional classification scheme, e.g., evaluating solutions to address the large percentage of motorcycles that were misclassified or passed completely undetected in this study.

References

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