Word count – 3700+15 figures and tables TRB Paper Number: 11-4162

Evaluation of a Computer-Vision Tracking System for Collecting Traffic Data

Neeraj K. Kanhere
Department of Electrical and Computer Engineering
207-A Riggs Hall
Clemson University, Clemson, SC 29634
Phone: (864) 650-4844, FAX: (864) 656-5910
E-mail: nkanher@clemson.edu

Stanley T. Birchfield
Department of Electrical and Computer Engineering
207-A Riggs Hall
Clemson University, Clemson, SC 29634
Phone: (864) 656-5912, FAX: (864) 656-5910
E-mail: stb@clemson.edu

Wayne A. Sarasua
Department of Civil Engineering
110 Lowry Hall, Box 340911
Clemson University, Clemson, SC 29634
Phone: (864) 656-3318, FAX: (864) 656-2670
E-mail: sarasua@clemson.edu

Sara Khoeini
Department of Civil Engineering
110 Lowry Hall, Box 340911
Clemson University, Clemson, SC 29634
Phone: (864) 656-3318, FAX: (864) 656-2670
E-mail: skhoein@clemson.edu

November 15, 2010

ABSTRACT

This paper evaluates a video-based traffic monitoring system developed at Clemson University that is capable of collecting vehicle count, speed, and classification data. In contrast with traditional video-based approaches based upon virtual detection zones, the system relies upon a combination of computer vision techniques to track vehicles through the series of image frames. Vehicles are classified using an algorithm that calculates a vehicle's length, width, and height. The system is evaluated based on its ability to collect volume, speed, and classification data using more than 90 hours of traffic data collected from installations along highways in New York, Maryland, and South Carolina. The evaluation is performed with respect to loop detector and/or piezo sensors installed at the locations, where available, as well as with respect to ground truth obtained by manually viewing the video. The system's performance is measured under different environmental conditions, at different times of day/night, with and without shadows, and with cameras both low and high off the ground. Overall, the system is able to produce counts hourly per-lane with less than 5% error, with the error reducing to less than 0.5% when aggregated over a 24-hour period. Average speeds are estimated with less than 1 mph error, and vehicles are classified into three length-based bins with less than 3% error. The evaluation indicated that the Clemson sensor performed as well or better than the commercial sensors installed at the testbed sites.

INTRODUCTION

Traffic data needs vary with different traffic agencies and traffic applications. In 1997, Volpe National Transportation Systems Center *(1)* conducted a nationwide survey of public agencies in urbanized areas with populations of more than 200,000. The survey focused on traffic monitoring programs and their various applications. The responses showed that traffic volume, vehicle classification, and speed/travel time were the most sought-after data.

The FHWA Traffic Monitoring Guide provides suggestions and examples on how statewide traffic data collection programs should be structured. It does not recommend any particular traffic monitoring sensor. Thus, state agencies have a great deal of latitude to identify particular sensors that work best for them *(2,3)*.

In 2001, the Arizona Department of Transportation conducted a survey of 50 state DOTs regarding different detector technologies used to collect traffic data. They were asked about their level of satisfaction with the devices they used. The results of the survey showed that the most popular methods for collecting different data types at the time were: inductive loops, pneumatic rubber tubes, and piezoelectric sensors for collecting traffic volume and speeds; and pneumatic tubes and piezoelectric sensors for collecting vehicle classification data. The non-intrusive technologies including radar, video image detection, passive acoustic, and passive magnetic rated consistently lower than intrusive sensors and pneumatic tubes. In this survey, video image detection ranked amongst the lowest of the non-intrusive sensors with only passive acoustic ranking lower. Accuracy was one of the reasons why video image detection was not rated higher *(4)*.

Since the late 1980s, video detection systems have been marketed in the U.S. and elsewhere. Most popular video detection systems use high-angle cameras to count traffic by detecting vehicles passing digital sensors. As a pattern passes over the digital detector, the change is recognized and a vehicle is counted. The length of time that this change takes place can be translated into speed and vehicle classification. These systems use significant built-in heuristics to differentiate between shadows and to be able to detect vehicles in various weather conditions. The accuracy of such systems is compromised if the cameras are mounted too low or have poor perspective views of traffic. If the camera is not mounted high enough, a vehicle's image will "spill over" onto neighboring lanes, resulting in double counting (Figure 1a). In Figure 1b, because the Clemson video-based system correctly handles spillover, it is able to track the truck as a single vehicle.

FIGURE 1: (a) A traditional video-based sensor relies on virtual detectors (green lines), thus resulting in over counting of trucks due to the triggering of multiple virtual detectors. (b) The Clemson video-based system correctly handles spillover, allowing the truck to be tracked as a single vehicle.

A promising approach to video detection is to track vehicles over time which makes such a system less susceptible to occlusion and spillover errors. Several years ago, Clemson researchers introduced such a tracking approach to video detection in *(5)*. Since that time, the algorithm has been significantly improved through extensive field testing at testbed locations in New York, Maryland, and South Carolina.

In this paper, we evaluate the Clemson system's accuracy in collecting volume, speed, and classification data compared with commercial loop detectors, piezo sensors, and manual ground truth.

TRAFFIC DETECTOR TECHNOLOGIES

Axle, visual, and presence sensors are most frequently used for collecting vehicle class volume information and each use a different mechanism for classifying vehicles. A report produced by Martin *(6)* summarized the results of a number of traffic detector studies conducted prior to 2003 by a number of different organizations including Minnesota DOT (MNDOT), Georgia Tech, New Mexico State University, Oregon DOT, and the Texas Transportation Institute (TTI). A summary of the results of these studies is shown in Table 1.

Table 1: Detection Performance on Highways (Adapted from ref. 6)

Detector Technology	Count Accuracy		Speed Accuracy	Classification Accuracy (1)	Environmental Effect	Non-intrusive
	Low Volume	High Volume				
Inductive loop	■	■	◩	◩	■	
Magnetic	■	■	■	?	■	
Pneumatic road tube	◩	□	□	?	◩	
Active infrared	■	■	◩	■	◩	√
Passive infrared	◩	◩	□	□	■	√
Radar Doppler	■	◩	■	□	■	√
True presence	■	■	◩	◩	■	√
Passive acoustic	◩	◩	■/◩	□	□	√
Pulse ultrasonic	■	■	□	□	■	√
Video Detection	■	■	◩	□	□	√

Note: ■= Excellent (< 5%); ◩= Fair (< 10%); □= Poor (> 10%); ? = Unknown

Table 1 indicates that most of the detector technologies evaluated collect traffic count data with better than 5% error on average. This accuracy degrades slightly for speed and vehicle classification. Vehicle classification accuracy in this table is a little misleading because the evaluations were not standardized in terms of number of vehicle classes.

Vehicle Classification Data

The number of classes can vary by application. FHWA requires 13 classes for reporting Highway Performance Monitoring System (HPMS) data. However the Highway Capacity Manual has only 3 classes for doing capacity analysis *(7)*. Length based classification can be used to estimate 13 classes but accuracy would be suspect. The FHWA Traffic Monitoring Guide suggests that a broad 3 to 4 class scheme can meet most of the data demands in traffic analysis *(2)*. Factors from axle studies can be used to convert classification data to 13 classes.

More recent studies related to length-based vehicle classification were conducted by MNDOT, Pennsylvania DOT (PennDOT), and California PATH. MNDOT evaluated non-intrusive detectors (passive acoustic and 2 microwave radar) and concluded that all three provide three-class classification fairly reliably *(8)*. A PennDOT study found that classification even in two classes (short and long vehicles) can be problematic with overhead sensors. An axle counting infrared sensor installed at road level worked well in PennDOT tests but was labor-intensive and did not work well for multilane roads with a pronounced crown *(9)*. Research on vehicle classification with a side fire microwave radar sensor performed at California PATH was tested for 4 classes in non-congested and congested traffic flow. The researchers concluded that the sensor performed well in uncongested flow but only marginal in congested traffic *(10)*.

A simple study conducted by Auffray et al. indicated that AutoSCOPE had a classification average error of 4.4% versus manual ground truth counts for two classes. AutoSCOPE is a popular video image detection system that uses virtual detection to count and classify vehicles *(11)*.

Yu et al. (*12)* evaluated AutoSCOPE for vehicle classification using 5 vehicle classes and found that accuracy varied, with the average ranging from 65 to 90% depending on the type of facility and the number of lanes. The non-freeway sites performed better than the freeway sites. The study indicated that the intensity of shadows or the presence of rain degrades the average accuracy by an additional 10%.

CLEMSON TRAFFIC MONITORING SYSTEM

In contrast with the traditional approach of virtual detectors, the system evaluated in this paper builds upon recent developments in computer vision. The first version of the system, presented in Kanhere et al. *(5)*, relied upon tracking features points and isolating features that were low to the ground. Over the next several years, the research team at Clemson introduced a number of algorithmic extensions such as pattern detection and vehicle base fronts, leading to improvements such as real-time processing, classification of motorcycles, and automatic calibration *(13,14,15,16,17)*.

One of the difficulties of any video-based system is the extreme changes that occur in the visual data as the result of changing lighting and/or environmental conditions. In particular, it is not uncommon for algorithms that work well during the daytime to not work well at night. Moreover, some of the most difficult times occur during dawn and dusk when the sun in low on the horizon, causing long shadows as well as severe glare when the camera is facing the direction of the sun. In addition, many cameras automatically switch from color mode during the daytime to low-lux black-and-white mode as the ambient light is reduced below a threshold due to the setting sun, and vice versa when the sun rises again in the morning. Although the low-lux mode can enable a human viewer to see more of the scene than would be otherwise visible, the increased sensitivity of the camera often results in significant blooming around the vehicle headlights and their reflections on the pavement, wreaking havoc for computer vision-based algorithms. Finally, environmental conditions such as rain, ice, and snow cause significant changes in the visual appearance, thus producing further challenges for such algorithms.

The system evaluated in this paper is the latest version of the system which incorporates the extensions mentioned above, as well as a number of developments to handle the lighting and environmental conditions just mentioned. By using a combination of edge detection, feature tracking, and vehicle base-fronts to detect and track vehicles, the system overcomes several limitations of the legacy commercial video image detection systems. In particular, the system is versatile, capable of monitoring both directions of traffic in real time and able to work even with cameras mounted as low as 26 feet above the ground because it properly handles the perspective effects that cause occlusion and spillover (Figure 1b). With the recent developments of the system, it is also capable of exhibiting robust behavior in the presence of changing lighting and environmental conditions.

FIELD EVALUATION AND EXPERIMENTAL RESULTS

The Clemson system was evaluated using data from four locations. The parameters evaluated matched those that were most commonly evaluated in the literature review: traffic volumes, speed, and vehicle classification. Vehicle occupancy and density are additional parameters that can be collected but were not included in the evaluation. However, the Clemson system uses detection zones to determine speeds by tracking vehicles over extensive distances depending on vantage point and field of view. This will likely give a better indication of space mean speed than what is possible from point sensors such as short inductive loop detectors, piezos and commercial vision sensors that use virtual detectors. Accurate density estimated can be calculated for from volume and space mean speed data. The system was installed at test bed locations in New York and Maryland with cooperation of personnel from the respective DOTs. This involved mounting the cameras, setting up power and communications, and configuring the software. Both installations were designed to allow for remote access from our research lab in Clemson. The cameras were installed at heights much lower than the minimum heights recommended for most of today's commercially available video image detectors. In addition, the system was tested using a temporary roadside setup and an existing TMC camera, both in South Carolina.

Maryland Installation

The first system was installed in Columbia, MD, on December 16, 2008 along Hwy-29, Northbound at Gates Lane (+39° 12' 37.60", -76° 51' 14.86"), as shown in Figure 2. Hwy-29 is a 4-lane divided highway, and the system monitors the two northbound lanes. The loop detector at this site collects continuous volume data. The team installed a pan-tilt-zoom (PTZ) dome camera, a frame grabber to convert the analog video signal from the camera into a digital data stream, a processing board to run the algorithm, and a wireless modem for communications and for monitoring the system remotely. The camera was mounted at the top of a wooden utility pole 28'5" above the pavement surface and 20'6" from the nearest traffic edge line.

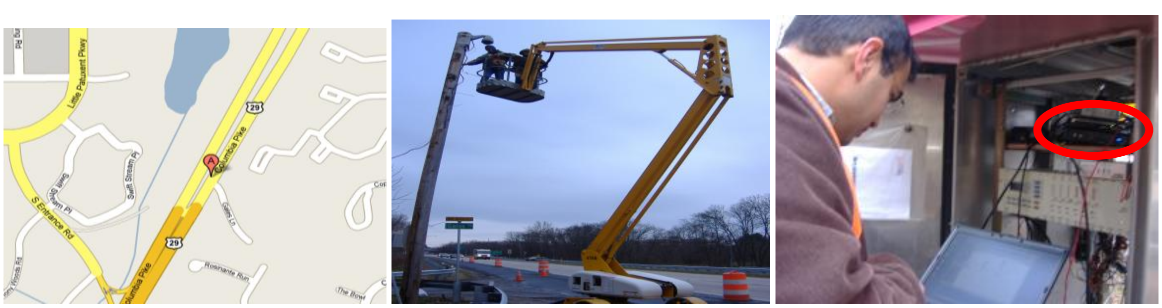


FIGURE 2: Installation at Maryland. The red oval indicates the video processor.

New York installation

The second system was installed in Long Island, NY on December 17, 2008 along I-495 East, near exit 23 (+40° 44'25.96", -73° 49' 13.13"), as shown in Figure 3. This is a busy 6-lane freeway. The location has a loop detector for collecting volume data and a piezo axle sensor for collecting speed and classification data. The camera was mounted at 26’4” above the pavement and 30’ from the edge line. The only source of power available at this location is a pair of solar panels mounted on the pole. The installation in New York was more time-consuming than the one in Maryland, due to the need to manually aim and focus the fixed camera. To conserve power, a USB-controlled relay was used to programmatically control the heater/fan in the camera housing.

FIGURE 3: Installation in New York. The red oval indicates the video processor.

Testbed Evaluation

Several days’ worth of video data was collected at both sites. Data provided by the DOTs allow the computer-vision system to be compared with the existing sensor, although discrepancies between the two sensors are difficult to interpret since it is not known whether the difference is due to error in one sensor or the other (or both). To resolve these discrepancies, graduate students at Clemson were assigned the tedious task of manually counting and classifying the vehicles in the video. To date, more than 90 hours of data have been manually annotated. This ground truth is indispensable for objectively calculating the accuracy of the vision system, as well as that of the existing sensors (such as loop detectors).

Maryland Testbed Volumes

Figures 4 and 5 show the output from the Maryland installation for two typical days – one weekday and one weekend day. The plots on the left display the hourly counts reported by the loop detector and our vision system, while the plots on the right show the percentage difference between the two outputs, along with a 10% dashed line for reference. For most hours, the difference in per-lane volumes is well below 10%.

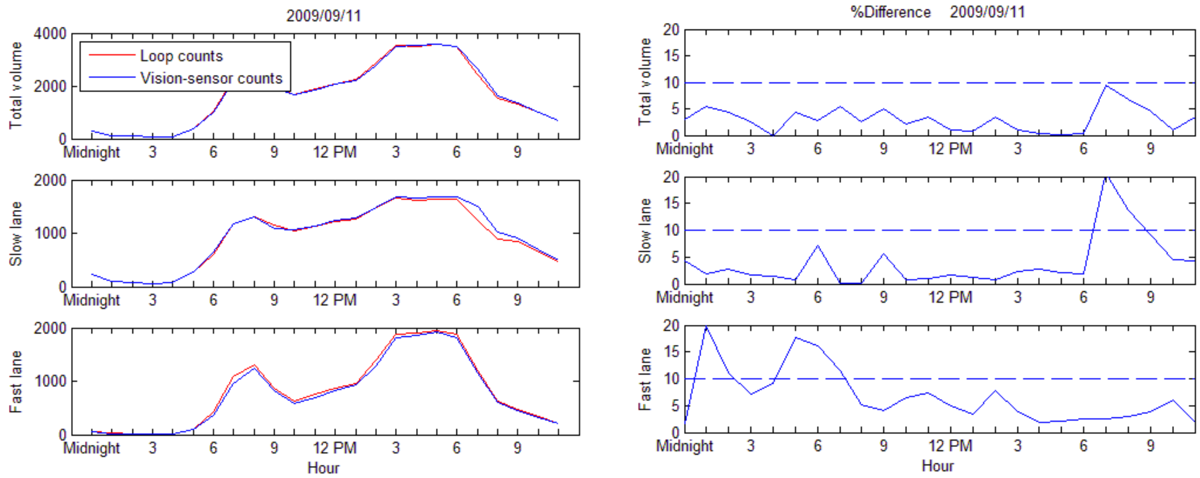


FIGURE 4: Comparison of vision-system counts and loop detector counts in Maryland shown for Friday, Sept 11, 2009.

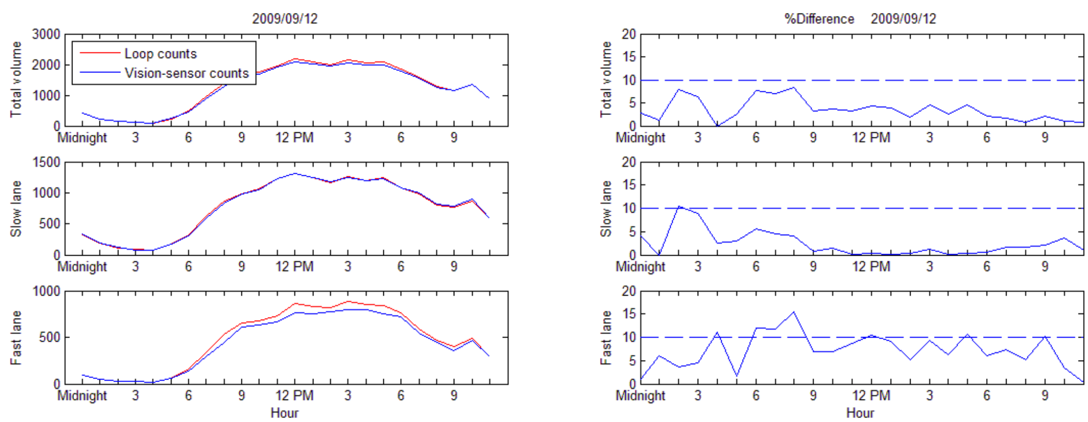


FIGURE 5: Comparison of vision-system counts and loop detector counts in Maryland shown for Saturday, Sept 12, 2009.

Note that low volumes during the early morning hours result in large percentage differences. To analyze these differences in detail, we recorded the output video from the Clemson system and manually annotated several hours of video data. Table 2 shows the result of this analysis. From the plots in Figure 4 we see that at 1:00 AM on Sept 11, the difference between the two sensor counts is almost 20% in the fast lane. However when compared to the ground truth counts it is revealed that the vision system actually performs better than the loop detector for this lane. Similarly at 8:00 AM on Sept 12 the difference in the fast lane is almost 15% whereas in reality the error is only 10%, just 4% worse than the loop detector (Figure 5).

Date				Manual counts (Ground truth)			Loop detector			Vision-sensor			% Error (Loop detector)			% Error (Vision-sensor)		
Year	Month	Day	Hour	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total
2009	9	11	1	104	22	126	102	25	127	100	20	120	-1.92	13.64	0.79	-3.85	-9.09	-4.76
2009	9	11	4	74	10	84	74	11	85	75	10	85	0.00	10.00	1.19	1.35	0.00	1.19
2009	9	11	6	618	391	1009	600	438	1038	643	367	1010	-2.91	12.02	2.87	4.05	-6.14	0.10
2009	9	11	7	1139	1027	2166	1176	1090	2266	1178	965	2143	3.25	6.13	4.62	3.42	-6.04	-1.06
2009	9	12	2	127	24	151	114	27	141	126	26	152	-10.24	12.50	-6.62	-0.79	8.33	0.66
2009	9	12	8	835	498	1333	870	528	1398	834	446	1280	4.19	6.02	4.88	-0.12	-10.44	-3.98

TABLE 2: Comparison of loop-detector and vision-system counts with ground-truth counts. Numbers in red indicate hours during which the loop detector error exceeds that of the vision-system. Negative signs indicate undercounting.

Figure 6 shows the accuracy of the two sensors compared with the ground truth counts for another 24-hour period. This video was captured on Thursday, February 19, 2009, stored, and manually annotated. From the per-lane count results (left plots in the figure), both the vision-system and the loop detector produce values that are almost indistinguishable from the ground truth. When we compare the actual accuracy of the sensors (right plots), we find that the accuracy of the vision-system is slightly better than the loop detector for a majority of the hours. Figure 6 shows that the Clemson system exceeded 5% error only one time in each lane during the 24 hour period, and the total 24-hour 2-lane volume is within 0.23% of ground truth. The 24-hour results for this day are summarized in Table 3.

	Manual counts (Ground truth)			Loop detector			Vision-system		
	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total	Slow lane	Fast lane	Total
Counts	22109	18662	40771	22024	18750	40774	22256	18609	40865
% Error				-0.38	0.47	0.01	0.66	-0.28	0..23

TABLE 3: 24-hour comparison of loop-detector and vision-system counts with ground-truth counts in Maryland, Thursday, February 19, 2009.

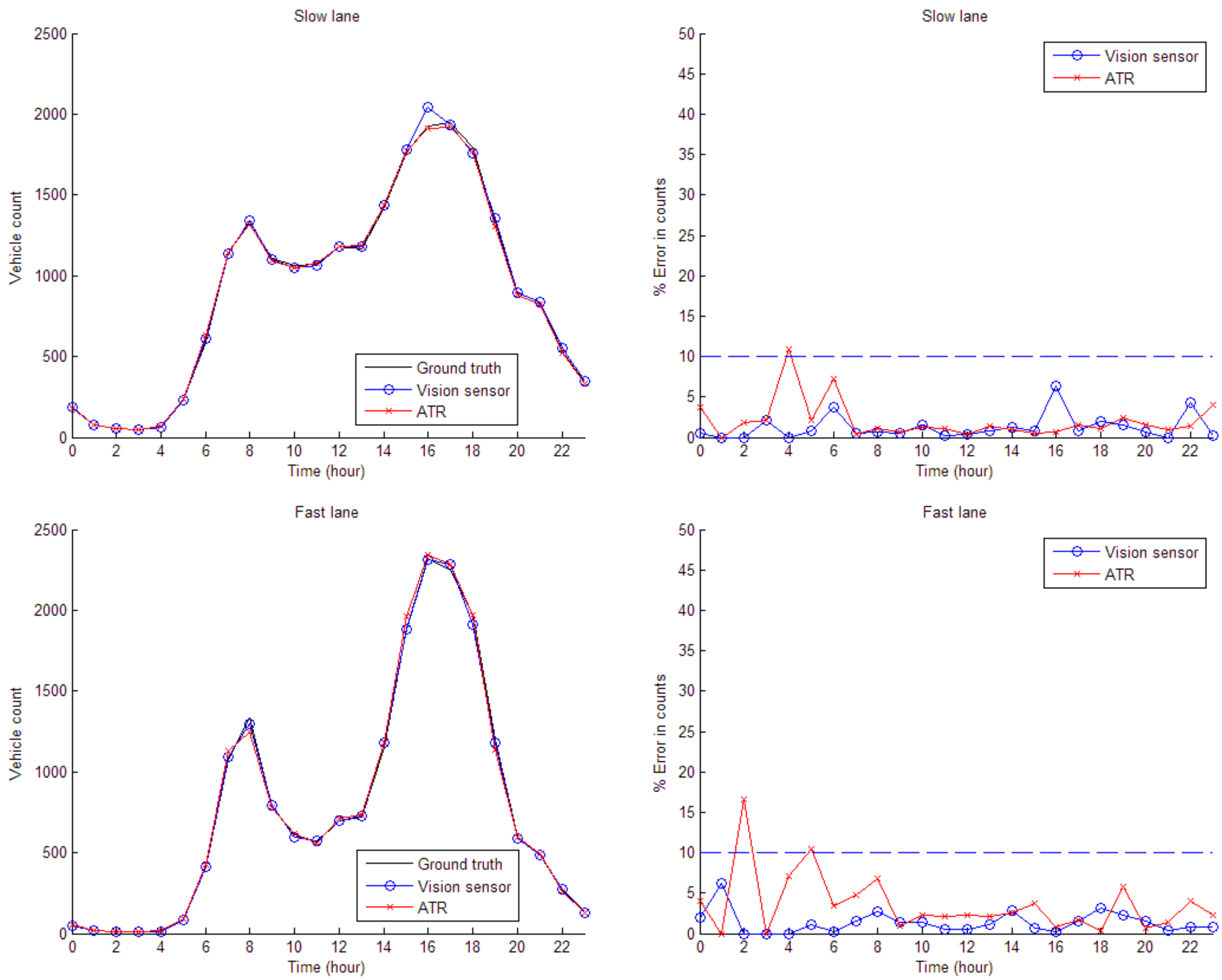
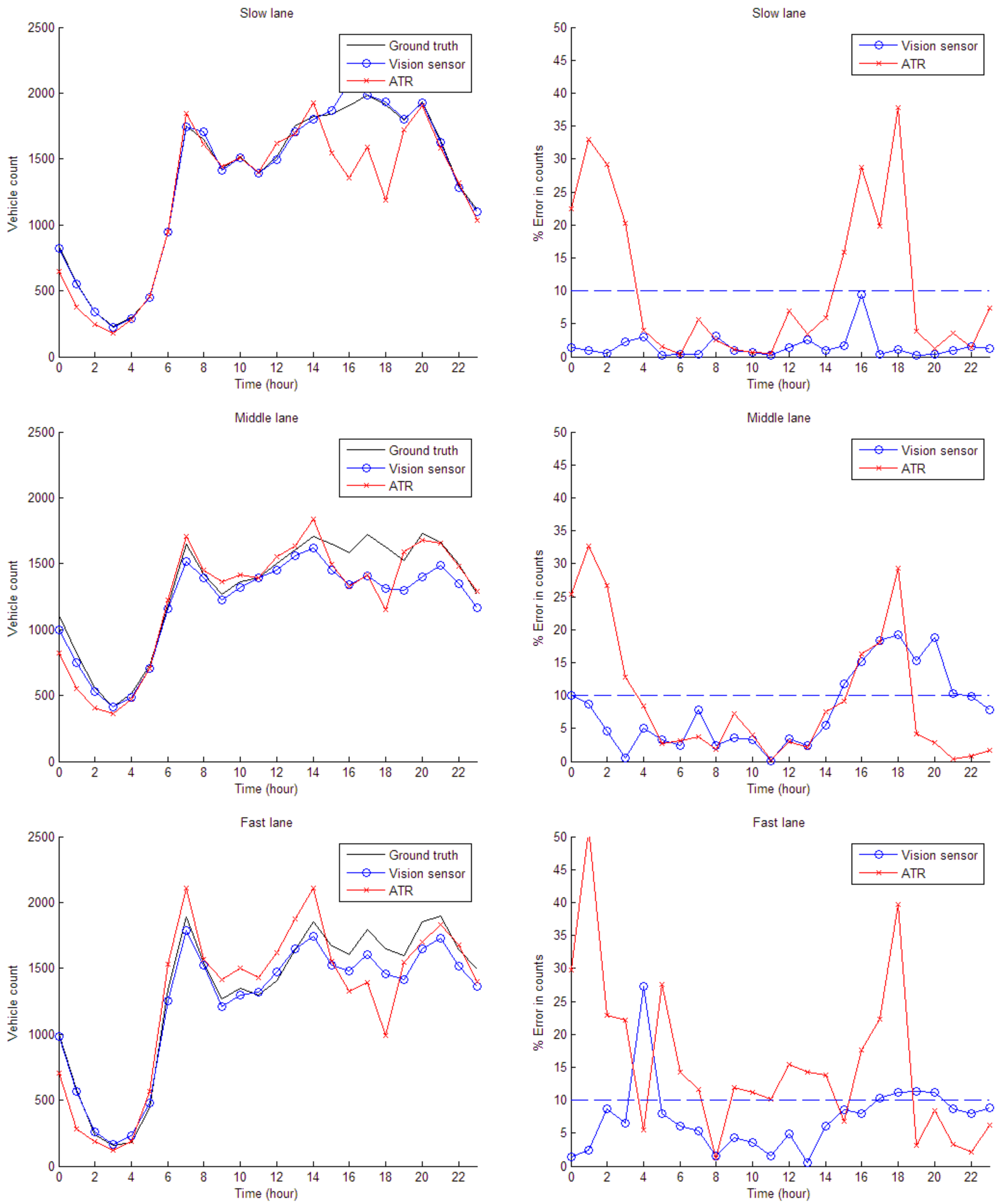


FIGURE 6: Accuracy of our vision system and the loop detector on another 24-hour period in Maryland, Thursday, February 19, 2009.

New York Testbed Volumes

Due to the dense traffic, frequent occlusions, and additional lane at the site in New York, manual annotation of the video data at that location was much more time consuming. Nevertheless, shown below in Figure 7 is the lane-by-lane comparison of the vision-sensor and loop-detector counts, compared with ground truth, for a 24-hour video collected on Thursday, November 5, 2009. Overall, the Clemson system generates less error than the loop detector, especially in the slow lane where the loop detector generates errors as high as 40% in one hour of counting while the error from the vision system never exceeds 10%. In the other lanes the Clemson system performs worse than the loop detector for a few hours just before midnight, but otherwise performs better. Notice that in all three lanes the maximum error of the loop detector is higher than that of the Clemson system. The error for 24 hours total for all 3 lanes is just under 5% for both the loop detector and the Clemson System.



1 **FIGURE 7: Results for the system in New York for a 24-hour period (midnight to midnight) on**
2 **Thursday, Nov. 5, 2009.**

24-Hour Speed and Classification Evaluation

The Maryland site included a single inductive loop that does not collect speed data. Thus, an evaluation of continuous speed data was not performed for this location. The New York site included a piezoelectric sensor for collecting speed and classification data. In this section, we compare the piezo sensor with the Clemson sensor using 24 hours of speed data and 8 hours of classification data from Thursday, Nov 5, 2009. Ground truth speed data and classification data was not available, thus it is unclear from the comparisons which device is more accurate. However, sampled time periods that compare the Clemson sensor with ground truth for both speed and classification data are included in the next section.

Figure 8 shows the difference in speeds between the piezo sensor and the Clemson system. The data is aggregated hourly. The hourly average speed difference never exceeds 4.5 MPH. The average difference weighted by traffic volume over the 24 hour period is 2.1 MPH.

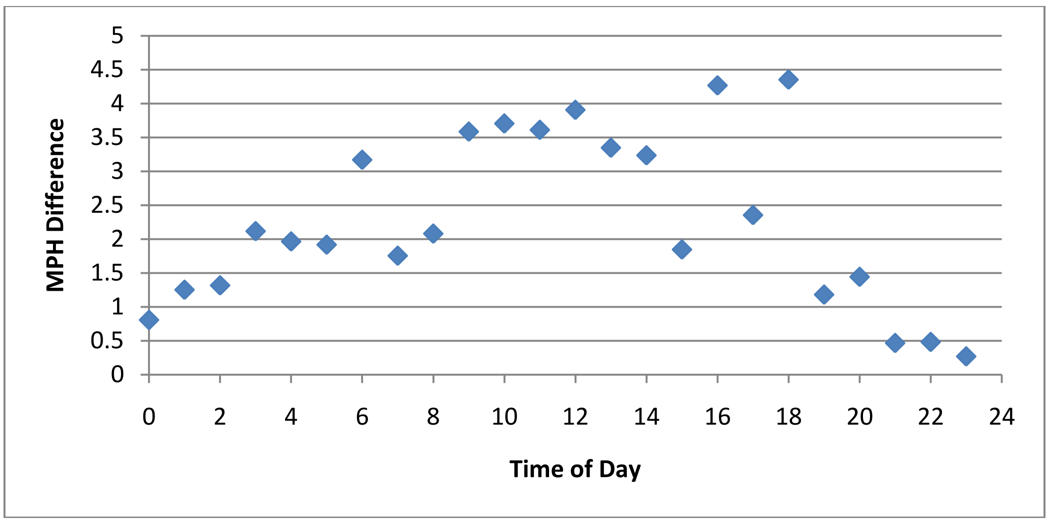


FIGURE 8: Hourly difference in speeds between the piezoelectric sensor and the Clemson system in New York on Thursday, Nov. 5, 2009.

The classification comparison was done for video data collected during daylight hours on November 5, 2009. The reason that only daylight hours were compared was because there was insufficient ambient light for the Clemson system to classify vehicles at night. Because it is an axle counter, the Piezo electric sensor is capable of collecting 13 classifications. The Clemson system uses dimensions and patterns to classify vehicles, thus accurately classifying 13 classes is not possible unless the axles are clearly visible from above. For this comparison, the FHWA classes were aggregated into a 3-class scheme as follows:

- Class A: FHWA classes 1,2,3
- Class B: FHWA classes 4-7
- Class C: FHWA classes 8-13

Note that the Clemson system was shown to accurately classify motorcycles in a previous work *(17)*. Table 4 shows the results of the vehicle classification comparison between the piezo sensor and the Clemson system. The results for the two devices are almost identical in terms of the proportions in each class. The differences aggregated over the 8 hour period are usually less than 1% for each travel lane. Table 5 provides a summary of the number of unclassified vehicles for each sensor during the 8-hour period, showing that the piezo sensor produces far fewer unclassified vehicles than the vision-based system.

TABLE 4: Summary of Classified Vehicles from 8:00 AM to 4:00 PM 11/5/2009

Class	Piezoelectric Sensor				Clemson System			
	Slow	Middle	Fast	Total	Slow	Middle	Fast	Total
A	92.07	84.68	99.68	92.44	92.11	83.31	99.29	91.51
B	5.63	8.76	0.17	4.71	5.22	10.37	0.24	5.33
C	2.30	6.55	0.02	2.85	2.67	6.32	0.47	3.16

TABLE 5: Summary of Unclassified Vehicles from 8:00 AM to 4:00 PM 11/5/2009

Lane	Piezoelectric Sensor			Clemson System		
	Unclassified	Total	% Unclassified	Unclassified	Total	% Unclassified
Slow	353	14030	2.5%	1163	14928	7.8%
Medium	403	13522	3.0%	1695	13755	12.3%
Fast	150	14459	1.0%	905	12562	7.2%
Total	906	42011	2.2%	3763	41245	9.1%

ADDITIONAL RESULTS IN A VARIETY OF CHALLENGING SCENARIOS

Manually annotating video data for speed and classification results is much more time-consuming than doing so for counts. As a result, it is prohibitive to obtain manual ground truth for such results over many consecutive hours as was done for counts. Instead, nine video clips – each several minutes long – were selected from among the four camera locations and manually annotated for counts, speeds, and classification. These clips were chosen to represent a variety of camera locations, lighting conditions, and weather conditions. Figure 9 shows a camera image from each of the video sequences that were evaluated. The number appended to the name of a video clip shows the hour in which the video was collected, so that “-09” means the video was extracted in the morning, between the hours of 9:00am and 10:00am; “-21” means between 9:00pm and 10:00pm, and so forth. This evaluation includes sequences from the Maryland and New York sites just described, including one Maryland sequence captured when icicles on the camera housing obstructed the field of view, and one Maryland sequence in which it was snowing. Other sequences were captured by a temporary camera along I-85 in South Carolina mounted 28’ above the road and over 45’ from the white edge line. The sequence from this camera was included to evaluate the Clemson system when the perspective view is not ideal. The system was also tested on a video sequence for a section of a 6-lane freeway in Greenville, SC. The camera at this section is an existing ITS camera operated by the SCDOT Traffic

Management Center. The camera is mounted over 60' above the pavement surface, which provides a good vantage to see across all 6 lanes of traffic (note that entrance and exit ramps were not included in this study). The system was tested for its ability to determine volume, speed, and classification by lane for both directions simultaneously for this video. The Greenville TMC clip was collected in the afternoon, while the MD-icicle and MD-snow clips were captured during the middle of the day. Note that it was actually snowing during the latter video clip.

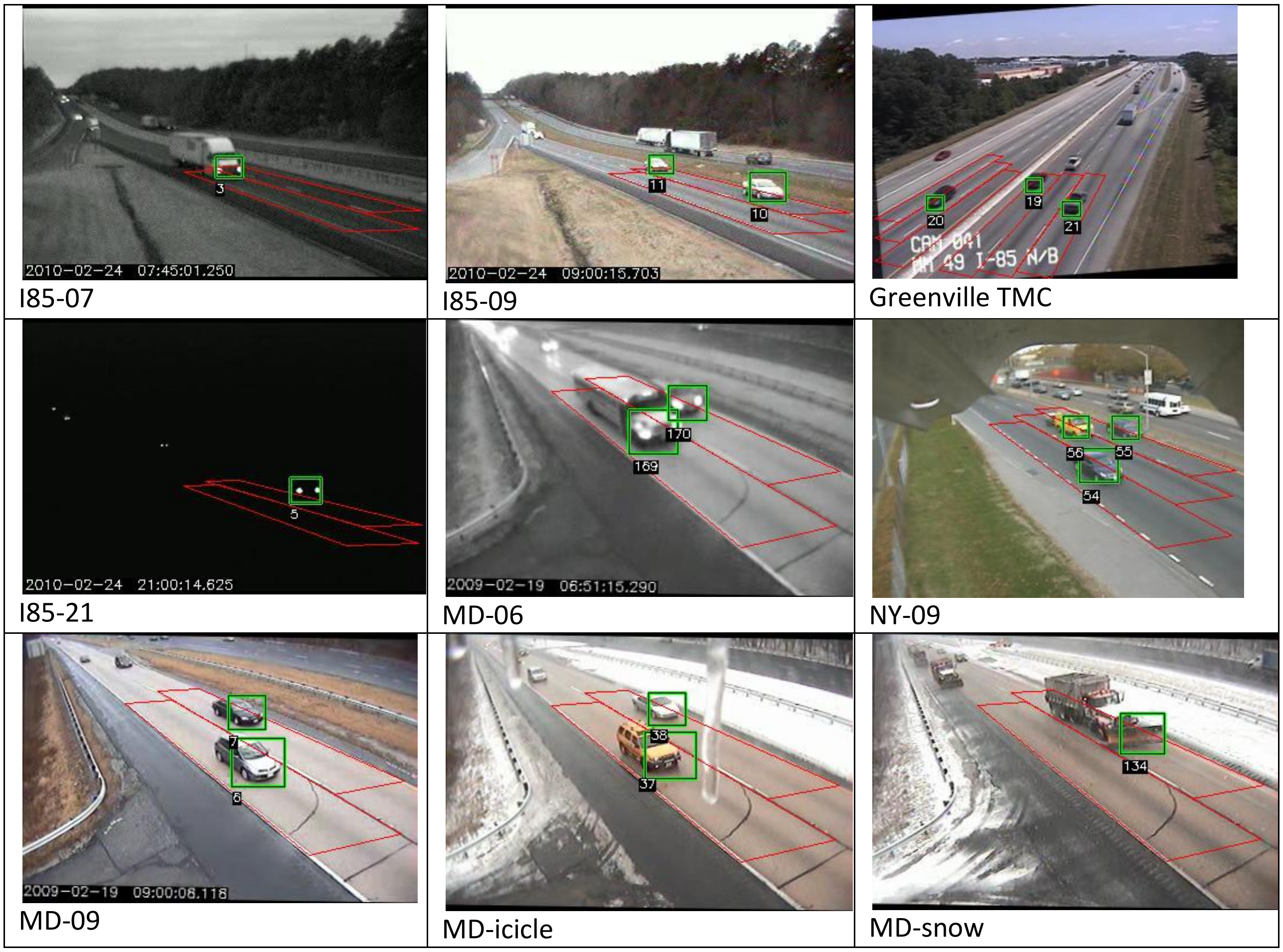


Figure 9: Images of the nine video sequences used for evaluating speeds and classification results.

Volume and Vehicle Classification Ground Truth

Ground truth for these additional sections was obtained by manually observing the video. The system assigns each vehicle a numerical ID. During manual observation, each ID is checked to determine whether the vehicle was counted and classified correctly by the system. False positives (an ID was assigned which should not have been) and missed vehicles (no assignment was made) are noted. This method, while tedious, gives a much better indication on the accuracy of the system because aggregating data can skew results. For example, if the number of false positives and the number of missed vehicles are roughly the same, these errors will cancel out if the data is aggregated.

Speed Ground Truth

Manual collection of ground truth speeds is done by determining the amount of time it takes for a vehicle to travel over a known distance in the video. Time is estimated by counting the number of frames it takes for a stable feature on the vehicle to travel the distance, then dividing by the constant frame rate of the camera, namely, 15 frames per second. Thus, if a vehicle in the video takes 30 frames to travel 160 feet, the speed of the vehicle is calculated as follows:

Speed = Distance/time=160ft/[(1sec/15 frames)*30 frames] =80 ft/sec=54.4 MPH

The stable feature that is tracked must be low to the ground for greatest accuracy, thus a tire is normally used. During the evening, vehicle headlights are used.

There are some random errors that occur when using this method to calculate speeds. First, there is error caused by pixel resolution. The pixel resolution will distort vehicle features. Non-constant frame rate can also contribute to the error. Ideally, a vehicle's tire will appear in the start and end frame at the exact moment when it touches the start and end lines in the video. This is rarely the case. In the example above, where the vehicle is traveling at 80 ft/sec, the vehicle is traveling roughly 5' during that frame. It can be shown that this will induce a maximum error of 2.5' in the distance. Thus, the actual vehicle distance will range between 157.5 to 162.5, which corresponds to a speed error of +/-0.85 MPH. As the sample size increases, the average error due to the frame rate will approach 0. Errors caused by the observer counting frames is relatively small because stop action is used. Further, each frame is numbered so the difference in the frame numbers at the start and end lines is the number of frames that have passed.

Results

Table 6 shows a summary of the results of the video samples under different circumstances. The travel lengths for measuring speeds are as follows: 80 feet for MD, 100 feet for NY, 160 feet for I-85, and 120 feet for TMC. The detection error ranges from 0.5% to 6.6%, while the classification error ranges from 0.8% to 2.6%. The only videos in which vehicles are unclassified are those captured along I-85, due to the vantage point of the camera. The maximum speed error is computed on a per-vehicle basis and provides the worst-case value for all vehicles in the video, while the average speed provides a more realistic picture of the accuracy of the system by averaging over all vehicles in the video. The maximum speed error remains under 10 MPH, while the average speed error never exceeds 1.4 MPH.

TABLE 6: Results of the video-based system computing counts, speeds, and classification on a variety of video clips.

	Video duration (min:sec)	# vehicles	# detected	Detect. error	Class. error	# unclass.	Max speed error (mph)	Avg. speed error (mph)
TMC	02:53	228	237	3.9%	2.2%	0	3.6	0.6
I85-07	15:04	235	231	1.7%	1.7%	10	7.7	0.6

I85-09	09:59	155	148	4.5%	2.6%	4	3.6	0.2
I85-21	15:03	150	140	6.6%	N/A	N/A	3.0	0.1
NY-09	07:29	378	393	4.0%	1.6%	0	6.6	0.3
MD-06	22:20	370	381	3.0%	1.9%	0	9.0	0.1
MD-09	14:59	356	350	1.7%	0.8%	0	6.0	1.4
MD-snow	15:03	199	198	0.5%	1.5%	0	3.0	0.9
MD-icicle	15:00	169	172	1.8%	1.8%	0	5.0	0.1

CONCLUSION

This paper has evaluated a video-based traffic monitoring system developed at Clemson University that is capable of collecting volume, speed, and classification data. More than 90 hours of video data was manually annotated to generate ground truth, to enable the system to be compared not only with existing sensors but also with respect to a fixed reference. Over a diverse set of data including four different locations and a variety of environmental conditions in three states, the system exhibited fairly robust behavior. Per-lane hourly counts were generally within 5% of ground truth, while aggregate counts were less than 1%. Classification results using three length-based bins never exceeded 3%, while average speed estimates were accurate within 1 MPH. A particularly interesting property of the system is that its accuracy did not vary much even when a single camera operated across a 6-lane freeway monitoring both directions simultaneously. The ability to collect accurate speed data from existing PTZ surveillance cameras mounted that has view across all lanes of a highway will allow the system to identify incidents in real-time as well as the lane in which it occurs. It should be emphasized that the system operates in real time, making it applicable to a variety of transportation applications.

According to the literature, commercially available video image sensors that use virtual detectors usually provide accurate volumes with the cameras are mounted at heights greater than 50 feet but provide mixed results for speed and, for the most part, do not do a good job classifying vehicles except when only a 2 class scheme was used. Because of the tracking approach combined with sophisticated image processing algorithms, the Clemson system produces accurate results for volume, speed, and vehicle classification whether the camera is mounted high above the ground (60') or as low as just 26 feet above the ground. This is the first video-based system of which we are aware that is capable of achieving such results.

One of the drawbacks of the system is its inability to classify vehicles at night. Therefore, a possible research activity would be to develop a classification algorithm for night time detection where there is little or no ambient light. In this case, the only stable features that can be seen by the camera are the running lights. Using color and placement standards for running lights, it may be possible to extend the classification results to situations in which ambient light is not available. The researchers plan to pursue this in future research.

REFERENCES

1. Mergel, J., An Overview of Traffic Monitoring Programs in Large Urban Areas, Center for Transportation Information of Volpe National Transportation Systems Center, Cambridge, MA, July 1997.
2. Office of Highway Policy Information. *Traffic Monitoring Guide*, FHWA, USDOT, May 2001.

3. Office of Highway Policy Information. *Traffic Monitoring Guide Supplement*, Section 4S FHWA, USDOT, April, 2008.
4. Skszek, S. L., "State-of-the-Art" Report on Non-Traditional Traffic Counting Methods, FHWAAZ-01-503, October 2001.
5. Neeraj K. Kanhere, Shrinivas J. Pundlik, and Stanley T. Birchfield. Vehicle segmentation and tracking from a low-angle off-axis camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1152–1157, June 2005.
6. Martin, P. T., Y. Feng, X. Wang. *Detector Technology Evaluation*. (November, 2003. http://www.mountain-plains.org/pubs/html/mpc-03-154/index.php. Accessed July 27, 2009.
7. *Highway Capacity Manual*. Transportation Research Board, Washington, D.C.. 2000.
8. Kotzenmacher, Jerry, Erik Minge, and Bingwen Hao, *Evaluation of Portable Non-Intrusive Traffic Detection System.* Final Report: MN-RC-2005-37, Minnesota Department of Transportation, September 2005.
9. French Engineering, LLC, *Traffic Data Collection Methodologies, Final report for Pennsylvania Department of Transportation*, Bureau of Planning and Research, April 2006.
10. James H. Banks, *Evaluation of Portable Automated Data Collection Technologies: Final Report*, California PATH Research Report, UCB-ITS-PRR-2008-15, August 2008.
11. Auffray, B., K. A. Tufte, Z. Horowitz, S. Matthews, and R. Bertini, *Evaluation of Single-Loop Detector Vehicle-Classification Algorithms using an Archived Data User Service System*, presented at the ITE District 6 Annual Meeting, Honolulu, June 2006.
12. Yu, X., P. Prevedouros, G. Sulijoadikusumo. Evaluation of Autoscope, SmartSensor HD, and TIRTL for Vehicle Classification Detectors. Presented at 89th Annual Meeting of the Transportation Board, paper #10-1846, Transportation Research Board, Washington, D.C., 2004.
13. Kanhere, N., S. Birchfield, and W. Sarasua. Vehicle Segmentation and Tracking in the Presence of Occlusions, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1944, pp. 89-97, 2006.
14. Kanhere, N., S. Birchfield, W. Sarasua, and T. Whitney. Real-Time Detection and Tracking of Vehicle Base Fronts for Measuring Traffic Counts and Speeds on Highways, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1993, pp. 155-164, 2007.
15. Kanhere, N., and S. Birchfield. Real-Time Incremental Segmentation and Tracking of Vehicles at Low Camera Angles Using Stable Features. *IEEE Transactions on Intelligent Transportation Systems*, 9(1):148-160, March 2008.
16. Kanhere, N., S. Birchfield, and W. Sarasua. Automatic Camera Calibration Using Pattern Detection for Vision-Based Speed Sensing, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2086, pp. 30-39, 2008.
17. Kanhere, N., S. Birchfield, W. Sarasua, and S. Khoeini. Traffic Monitoring of Motorcycles During Special Events Using Video Detection, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2160, pp. 69-76, 2010.