

Reference Machine Vision for ADAS Functions

May 2021

Final Report



VIRGINIA TECH
TRANSPORTATION INSTITUTE
VIRGINIA TECH.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 04-115	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Reference Machine Vision for ADAS Functions		5. Report Date May 2021	
		6. Performing Organization Code:	
7. Author(s) Abhishek Nayak (TAMU) Sivakumar Rathinam (TAMU) Adam Pike		8. Performing Organization Report No. Report 04-115	
9. Performing Organization Name and Address: Safe-D National UTC Texas A&M University Texas A&M Transportation Institute 3135 TAMU College Station, Texas 77843-3135 USA		10. Work Unit No.	
		11. Contract or Grant No. 69A3551747115/ Project 04-115	
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report	
		14. Sponsoring Agency Code	
15. Supplementary Notes This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.			
16. Abstract Studies have shown that fatalities due to unintentional roadway departures can be significantly reduced if Lane Departure Warning and Lane Keep Assist systems are used effectively. However, these systems have not been widely adopted due, in part, to the lack of suitable standards for pavement markings that enable reliable functionality of sensor systems. The objective of this project is to develop a reference lane detection system that will provide a benchmark for evaluating different lane markings and perception algorithms. The project will also validate the effectiveness of lane markings' material characteristics as well as the vision algorithms through a systematic testing of lane detection algorithms in a robust test/vehicle environment.			
17. Key Words Lane detection, Lane marking materials, Computer vision		18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website , as well as the following repositories: VTechWorks , The National Transportation Library , The Transportation Library , Volpe National Transportation Systems Center , Federal Highway Administration Research Library , and the National Technical Reports Library .	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 40	22. Price \$0

Form DOT F 1700.7 (8-72)

Reproduction of completed page

authorized

Abstract

Studies have shown that fatalities due to unintentional roadway departures can be significantly reduced if Lane Departure Warning and Lane Keep Assist systems are used effectively. However, these systems have not been widely adopted due, in part, to the lack of suitable standards for pavement markings that enable reliable functionality of sensor systems. The objective of this project is to develop a reference lane detection system that will provide a benchmark for evaluating different lane markings and perception algorithms. The project will also validate the effectiveness of lane markings' material characteristics as well as the vision algorithms through a systematic testing of lane detection algorithms in a robust test/vehicle environment.

Acknowledgments

Special thanks to Ken Smith from The 3M Company for being an advisor on this project and serving as the subject matter expert in compiling this report, and to Adam Pike from TTI for guiding us at every step. We also thank Kyle Kingsbury and Davis Dobrovolsky from TTI for their assistance in the lane marking data collection, and Dr. Eun Sug Park for her support in carrying out the statistical analysis for this project.

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

INTRODUCTION	1
BACKGROUND	1
METHOD	4
Video Data	4
Video Data Collection	4
Stage 1: On-road LD evaluation in Central Texas	4
Stage 2: Closed Course 3M Panel Dataset	6
Video Data Evaluation Metric	9
Material Characteristics Data	11
Retroreflectivity and Other Material Properties	11
Material Data Collection	12
RESULTS	12
College Station Dataset	12
Closed Course 3M Panel Dataset	15
Discussion	16
Conclusions and Recommendations	24
ADDITIONAL PRODUCTS	26
Education and Workforce Development Products	26
Technology Transfer Products	26
Data Products	26
REFERENCES	27
APPENDIX 1. US-290 DATASET	29
APPENDIX 2. RELLIS CLOSED COURSE EVALUATION	32

APPENDIX 3. ADDITIONAL TABLES AND PLOTS FROM 3M PANEL DATASET EVALUATION.....	34
--	-----------

List of Figures

Figure 1: A driving map for the College Station dataset. Map Credits: https://www.google.com/maps	5
Figure 2: Images of different roads in the College Station dataset collected during daytime.	5
Figure 3. Panel identification scheme.....	7
Figure 4: Google Earth view of runway 35L at RELLIS campus and locations of 3M panel placement.	8
Figure 5: Images of 40 ft gap pavement markings on 35L (Set 2: 06W-6 on left and 02W-6 on right) during day (above) and night (below).	8
Figure 6: Definition of intersection over union. Source: https://www.pyimagesearch.com/	10
Figure 7: LD outputs of 40 ft gap lane markings on 35L (Set 2: 06W-6 on left - 02W-6 on right) during day (above) and night (below).	11
Figure 8: Illustration of types of light reflections from a surface. Source: https://madebydelta.com/	12
Figure 9: Image showing the area-split for near FOV vs far FOV evaluations.	16
Figure 10: Plot of least square mean F1-score capturing the effect of panel spacing on LD performance.	18
Figure 11: Plot of least square mean F1-score capturing the effect of 2-way interaction of panel spacing on panel material for daytime LD performance.	19
Figure 12: Plot of least square mean F1-score capturing the effect of two-way interaction of panel spacing on panel material for nighttime LD performance.	19
Figure 13: Plot of least square mean F1-score capturing the effect of nighttime illumination on LD performance.	20
Figure 14: Plot of least square mean F1-score capturing the effect of two-way interaction of nighttime illumination on panel material for LD performance.	20
Figure 15: Plot of least square mean F1-score capturing the effect of 2-way interaction of panel spacing and driving direction on LD performance.	21
Figure 16: Plot of Least square mean F1-score capturing the effect of evaluation area on LD performance.	22
Figure 17: Plot of least square mean F1-score capturing the effect of panel material for LD performance.	24

Figure 18: Image of a road segment with white pavement marking followed by black (driving westwards).	30
Figure 19: Image of a road segment with white pavement marking followed by black (driving eastwards).	30
Figure 20: Image of a road segment with white pavement markings with a black border (driving eastwards).	30
Figure 21: Image of a road segment with white pavement markings with a black border (driving westwards).	31
Figure 22: Image of a road segment with white pavement markings, and significant sun glare (driving eastwards).....	31
Figure 23: Plot of lane detection performance of SCNN & LaneNet algorithms (measured as F1 score) vs continuous R_L data for pavement markings on RELLIS 35L.....	32
Figure 24: Plot of LD performance vs diffused reflectance (Q_d) for the 3M pavement panels....	35
Figure 25: Plot of LD performance vs retroreflectivity (R_L) for the 3M pavement panels.	35
Figure 26: Panel-wise LD performance summary for 3M panels.....	36
Figure 27: LD performance summary for Set 1 of 3M panels based on evaluation area.	37
Figure 28: LD performance summary for Set 2 of 3M panels based on evaluation area.	38
Figure 29: LD performance summary for Set 3 of 3M panels based on evaluation area.	39

List of Tables

Table 1: Combinations of Driving Conditions for 3M Pavement Panel Data Collection	7
Table 2: Performance of Different Algorithms (IoU Threshold = 0.5) on College Station Dataset	13
Table 3: Lane Marking Characteristics Data of College Station Dataset Based on ASTM Standards	14
Table 4: 3M Pavement Panel Material Characteristic Data – 05/21/2020 – RELLIS 35L – Retroreflectivity and Color	16
Table 5: List of Factors Considered in the ANOVA Model.....	17
Table 6: List of Statistically Significant Factors During Nighttime Based on ANOVA Model for LD Performance (F1-Scores).....	17
Table 7: List of Statistically Significant Factors During Daytime Based on ANOVA Model for LD Performance (F1-Scores)	17

Table 8: Least Square Means Table Comparing the Effect of Panel Spacing on LD Performance	18
Table 9: Least Square Means Table Comparing the Effect of Nighttime Illumination on LD Performance	20
Table 10: Least Square Means Table Comparing the Effect of Driving Direction on LD Performance	21
Table 11: Least Square Means Table Comparing the Effect of Evaluation Area on LD Performance	22
Table 13: Least Square Means Table Comparing the Effect of Panel Material and Width on LD Performance	23
Table 14: LD Performance of Different Pavement Marking on US 290	31
Table 15: Material Characteristics of Lane Markings on RELLIS 35 L	33
Table 16: Tukey HSD Test ($\alpha = 0.050$) of LSMean Differences for Two-way Interaction Between Panel Spacing and Panel Material During Day (Top) and Night (Bottom). (Levels not Connected by the Same Letter are Significantly Different).....	34
Table 16: Tukey HSD Test ($\alpha = 0.050$) of LSMean Differences for 2-way Interaction Between Panel Spacing and Driving Direction During Day (Left) and Night (Right) – Levels Not Connected by the Same Letter Are Significantly Different.....	39
Table 17: LSMeans Differences Tukey HSD test ($\alpha = 0.050$) for Two-way Interaction Between Panel Spacing and Nighttime-Illumination – Levels Not Connected by the Same Letter Are Significantly Different	40
Table 18: Tukey HSD Test ($\alpha = 0.050$) of LSMeans Differences for Two-way Interaction Between Nighttime Illumination and Panel Material – Levels Not Connected by the Same Letter Are Significantly Different.	40

Introduction

Lane Detection (LD) systems are an integral part of most commercial Advanced Driver Assistance Systems (ADAS), which are designed to improve the safety of automobiles. The systems can include features such as lane departure warning (LDW), lane keep assist (LKA), adaptive cruise control (ACC), lane centering, lane change assist, and autonomous driving modes. LDW and LKA systems alone have the potential to prevent or mitigate 483,000 crashes—87,000 nonfatal injury and 10,345 fatal—in the United States every year [1]. While LDW and LKA technologies are widely available, customer acceptance and market penetration of these technologies remain low. These deficiencies can be traced to the inability [1] [2] of many perception systems to consistently recognize lane markings and localize the vehicle with respect to the lane marking in a real-world environment of variable markings, changing weather, and other vehicles. These challenges translate to (i) inconsistent detection of lane markings, (ii) misidentification of lane markings, and (iii) the inability of the systems to locate lane markings in some conditions. These challenges can be addressed both by improving the consistency and detectability of the lane markings and by improving the perception algorithms currently employed in the sensors. Currently, there is no available standard or benchmark to evaluate the quality of either lane markings or perception algorithms or how they relate to ADAS functions [2].

The key functional feature for a reliable LDW or LKA system is road and lane perception. The main perceptual cues for driving used by both human drivers and autonomous vehicles include road color and texture, road boundaries, and lane markings [2]. While different modes of sensing, including the use of LIDAR and radar, exist for road vehicles, due to recent advances in image processing techniques and the low-cost of cameras, vision-based LD has become the most prominent mode of sensing employed in modern LD systems. The prominence of vision systems in LD can also be attributed to the fact that road markings are primarily developed for human vision [2]. This goal of this project was to conduct a detailed study on lane infrastructure and its effect on the vision systems used in LD. An in-depth review of factors and components (e.g., lane marking characteristics, marking quality, etc.) that affect LD performance was carried out by incorporating pavement marking material characteristics into the LD system evaluation framework. The framework was tested extensively using data collected from College Station, TX and nearby cities, as well as using lane marking materials provided by 3M at the Texas A & M University System RELLIS Campus.

Background

LD systems can be broadly divided into three functional components: (i) hardware, (ii) software, and (iii) infrastructure. The hardware component corresponds to the equipment used for the sensing modality of the LD system. Although vision-based LD systems traditionally suffer from

functional limitations due to changing environmental conditions like illumination variation (direct sun on camera, glare, oncoming vehicle lights), shadows, and bad weather (rain, fog, snow), vision-based detection is still a widely adopted ADAS sensing mode and is expected to continue dominating in the future. Monocular camera and stereo vision cameras are the main sensors that have been used in vision-based LD systems. The variations in the hardware components include the type of sensor, type of lens (wide angle, fisheye), lens properties (field of view, focal length), and camera specification (pixel size, megapixels, resolution, frame rate). The software component refers to the algorithms that are employed on the LD system to detect lanes and help navigate the vehicle.

Vision-based LD systems generally consist of three main subprocesses: (i) image preprocessing, (ii) LD, and (iii) lane tracking. According to Xing et.al [3], conventional vision-based LD algorithms can be roughly classified into two categories: feature-based and model-based. Feature-based methods rely on the detection of lane marking features such as colors, textures, and edges. Model-based methods usually assume that lanes can be described with a specific model, such as a linear, parabolic, or various kinds of spline models. Since the advent of machine learning and deep learning techniques, several new algorithms that leverage the power of deep networks, parallel computing, and large data approaches for LD have been developed. Many deep learning algorithms have consistently produced significantly better results as compared to conventional approaches. Bei et al. reported that by using a convolutional neural network (CNN), the LD accuracy increased from 80% to 90% compared with traditional image processing methods. Several review papers in the literature [2], [3], [4] give a detailed account of the various works that have been carried out towards the development of LD algorithms. However, most conclude that the challenges and limitations for future research extend to the scope of developing better road understanding techniques and methods to increase detection reliability [2].

The infrastructure component corresponds to the lane markings and pavement surfaces that will be used by machine vision systems for sensing lanes. The variations in lane markings could include color, geometry (continuous/intermittent lanes, width, length), lane marking performance characteristics (luminance, retroreflectivity, color), and other pavement variables (asphalt, concrete). Several independent studies have investigated the effects of various lane marking properties towards an effective machine vision system [5], [6], [7], [8]. Studies ([6], [7], and [8]) evaluating the effect of wet retroreflective properties of lane markings and their effect on machine vision indicate that considering the lane markings into the detection framework helps improve LD performance.

The essential requirement for a safe LD system is to provide accurate and robust LD. While most of the recent machine learning algorithms strive to provide accurate results, robustness remains an issue, since lane measurements are affected by heavy traffic and adverse weather conditions. According to Xing et al. [3], the factors that limit progress towards achieving robustness are the lack of public benchmarks and data sets due to the difficulty of labeling lanes as the ground truth. Hillel et al. [2] also make a similar observation, stating that the challenge of current research is the

inability to compare the performance of different LD methods due to the lack of public annotated benchmarks.

Developing a large public video benchmark has the potential to reduce evaluation costs. To address this issue, recently, large public datasets focusing on LD have been developed. The most prominent among those are the CULane Dataset (developed by the Multimedia Laboratory at The Chinese University of Hong Kong) [9], TuSimple Benchmark (developed by the San Diego-based tech startup TuSimple) [10], and the lane marking dataset in BDD100K (Developed by the Berkley Artificial Intelligence Research (BAIR) Lab at UC Berkley) [11]. CULane consists of more than 55 hours of video data collected by cameras mounted on six different vehicles driven on the roads of Beijing. The test set is divided into one normal and eight challenging categories, including crowded scenes, shadows, night, dazzle light, curved roads, crossroads, arrows, and no lanes. The TuSimple dataset is made up of clips with 20 frames collected at 1 Hz in good and medium weather conditions during different times of the day and in different traffic conditions.

However, none of these datasets have provisions to evaluate the effects of hardware and lane markings towards comparing the performance of LD systems. As noted in studies [5], [6], [7], and [8], different types of lane markings do affect the performance of an LD system and thus should be an integral part of the evaluation methodology for LD performance. According to [3], another challenging task in LD systems is to design an evaluation system that can verify system performance. Due to the lack of standard evaluation metrics that can comprehensively assess the system performance for both accuracy and robustness, no objective assessment of the lane estimation process currently exists. An American Traffic Safety Services Association report in 2019 [12] noted a few key recommendations (based on a summary of road-marking challenges by Mobileye) to improve infrastructure to aid LD performance. These recommendations were also included in the key-summaries developed by the National Committee on Uniform Traffic Control Devices, as described by Council Member Paul Carlson in MTC-CAV TF.Criteria.v08 [13]. A few notable recommendations include the following:

- Robust markings that are visible in a variety of lighting and weather conditions.
- Minimum pavement marking retroreflectivity levels.
- Longitudinal markings (edge lines, centerlines, and lane lines) shall be six inches wide on roads with a posted speed of 40 mph (64 km/h) or greater.
- Lane line markings shall be 15 feet long (about 4.5 meters) with a gap of 25 feet (about 7.6 meters).
- Dotted edge line extensions shall be marked along the exit and entrance ramps on roads with a posted speed of 40 mph (64 km/h) or greater.

- Crosshatch (i.e., chevron) markings shall be included in gore areas on roads with a posted speed ≥ 40 mph.
- Non-reflective Botts' Dots should be eliminated or only used when supplementing pavement markings.
- Contrast striping should be required in Portland cement concrete roadways with a posted speed of 40 mph (65 km/h) or greater.

This work plans to address some of these limitations through a detailed study on lane infrastructure. An in-depth review of factors and components (e.g., lane marking characteristics, marking quality) that affect LD performance was carried out by incorporating pavement marking material characteristics into the LD system evaluation framework.

Method

Datasets were developed to comprehensively evaluate the different material and hardware properties and their effect on LD vision systems. The datasets were collected in two stages. Stage 1 included data from pavement markings on roads in Central Texas under different environmental conditions. Stage 2 included data from custom lane markings (provided by 3M) on a closed course runway at the Texas A&M University System RELLIS campus. Stage 2 data collection was carried out in a controlled environment in the sense that different lane marking materials of known material properties were evaluated on the same pavement under similar environmental conditions, a scenario that is difficult to simulate on typical roads.

Video Data

Video Data Collection

The video data were collected by using a 5 MP camera mounted on a standard test vehicle owned by Texas A&M University. The camera setup used was a Blackfly BFS-U3-51S5C-C camera sensor (Sony IMX250 CMOS – 5 MP – USB 3.1 camera) attached with a Kowa LM8HC Manual Iris C-Mount f=8 mm/F 1.4 Lens. To maintain homogeneity throughout the dataset, the videos were collected at 25 frames per second.

Stage 1: On-road LD evaluation in Central Texas

College Station Dataset

Stage 1 data collection included the College Station Dataset and US 290 Dataset (discussed in Appendix 1. US-290 Dataset). For the College Station dataset, video data were collected by driving on roads with varied lane markings in and around Central Texas. These roads included asphalt and concrete road surfaces, multilane and two-lane two-way roadways, and sections with tangents and curves. The pavement markings on these roadways varied from new to several years old to represent a range of marking conditions. To study the effect of natural light vs incident light on

LD performance, video data were collected at three different times of day during summer: (1) morning (11 a.m. CST, sunny day with clear sky), (2) evening (6 p.m. CST, sunny day with clear sky) and (3) nighttime (10 p.m. CST with clear sky). Figure 1 shows the driving map for the College Station dataset. Figure 2 shows the images of different roads that were considered in this dataset.

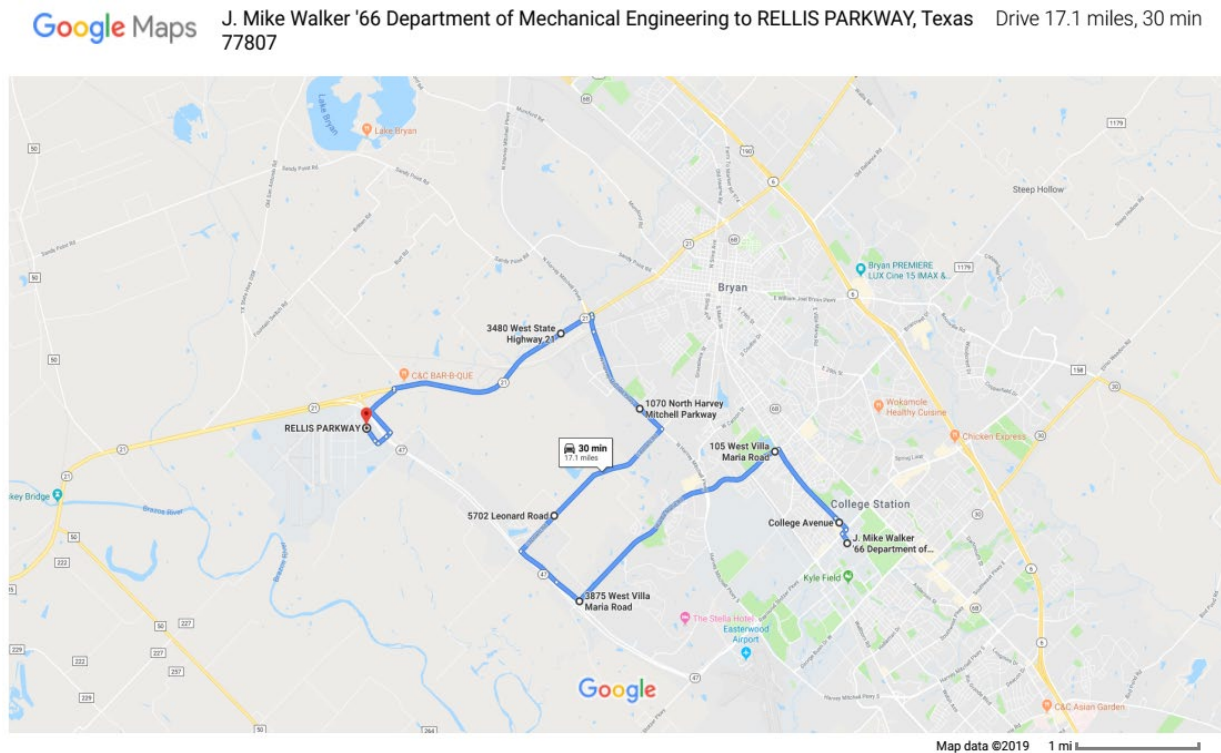


Figure 1: A driving map for the College Station dataset. Map Credits: <https://www.google.com/maps>



Figure 2: Images of different roads in the College Station dataset collected during daytime.

Stage 2: Closed Course 3M Panel Dataset

Stage 2 consisted of the closed course 3M panel dataset. For both the College Station dataset and the 3M dataset, video data were collected along with the pavement marking material characteristic data. We evaluated LD performance from state-of-the-art LD algorithms on the video data and the different factors that affect it. The data collection procedure and LD evaluation methods are discussed below.

RELLIS Closed Course Pavement Marking Evaluation

Environmental conditions like sun glare, shadows, and road illumination change depending on the road and driving direction. In order to selectively evaluate the effect that material characteristics of lane markings have on LD performance, it is beneficial to eliminate external environmental factors that may affect the study. Therefore, the team chose to evaluate different lane marking materials in a controlled environment by conducting the data collection of different materials on the same road (pavement). Doing so reduces impacts from other factors, such as changing sun-glare, lane marking deteriorations, and other factors that may affect LD performance.

One area of the closed course (Runway 35L) on the RELLIS campus was identified to carry out the experiments for this study. To check the acceptability of the runway, lane marking data were collected for the existing lane markings on runway 35L. Details of the study can be found in Appendix 2. RELLIS Closed Course Evaluation. There were limited interactions from the ghost markings, concrete block edges, or faded lane markings that affected the measurement and LD performance on 35L. Therefore, the runway area was deemed acceptable to study the LD performance of the 3M panels.

3M Panel Data Collection

Video data were collected by laying out the different 3M panels on the same runway section where the other closed-course evaluations took place (Runway 35L). In this study, four different pavement marking materials (01, 02, 06, 08) were evaluated. Each panel was either 4-inches wide (01W-4, 02W-4, 06W-4, 08W-4) or 6-inches wide (02W-6, 06W-6). The first number in the name refers to the specific marking material, which has its inherent properties like daytime reflectivity (Q_d), nighttime reflectivity (R_L), and Luminance Factor Y. “W” in the panel name refers to the white color of the panel. The last number refers to the marking width. Figure 3 explains the panel identification scheme used. These panels were grouped into three sets—Set 1: [01W-08W], Set 2: [02-6-06W-6], and Set 3: [02W-06W]. A set refers to a pair of markings that were laid out on the road during data collection, one on each side of the data collection vehicle; i.e., 01W on the left and 08W on the right. Each set of the markings were evaluated for LD performance (F1 score) using the spatial convolutional neural network (SCNN) algorithm. The “combined” data corresponds to the analysis that involves both set markings in the image. Individual F1 scores corresponding to each material were extracted from the combined data. Figure 4 shows the layout configuration of the panels in the test area. Figure 5 shows images from the closed course 3M panel dataset.

These three sets of panels were placed on the runway in two different spacing configurations: (a) 30 ft gap between panels and (b) 40 ft gap between panels, to evaluate the effect of panel spacing on LD performance. LD data were collected for the two panel configurations during daytime (clear sky conditions, around 10 a.m., in the month of May) and nighttime (clear sky conditions, around 9:30 p.m. on the same date, using high beam light and low beam light separately), by driving in two different directions (northbound and southbound). Table 1 captures the different testing conditions under which the video data was collected.

Table 1: Combinations of Driving Conditions for 3M Pavement Panel Data Collection

Vehicle Driving Direction	Panel Gap	Driving Conditions
Northbound	30 ft	Daylight
Northbound	30 ft	Night High Beam
Northbound	30 ft	Night Low Beam
Northbound	40 ft	Daylight
Northbound	40 ft	Night High Beam
Northbound	40 ft	Night Low Beam
Southbound	30 ft	Daylight
Southbound	30 ft	Night High Beam
Southbound	30 ft	Night Low Beam
Southbound	40 ft	Daylight
Southbound	40 ft	Night High Beam
Southbound	40 ft	Night Low Beam

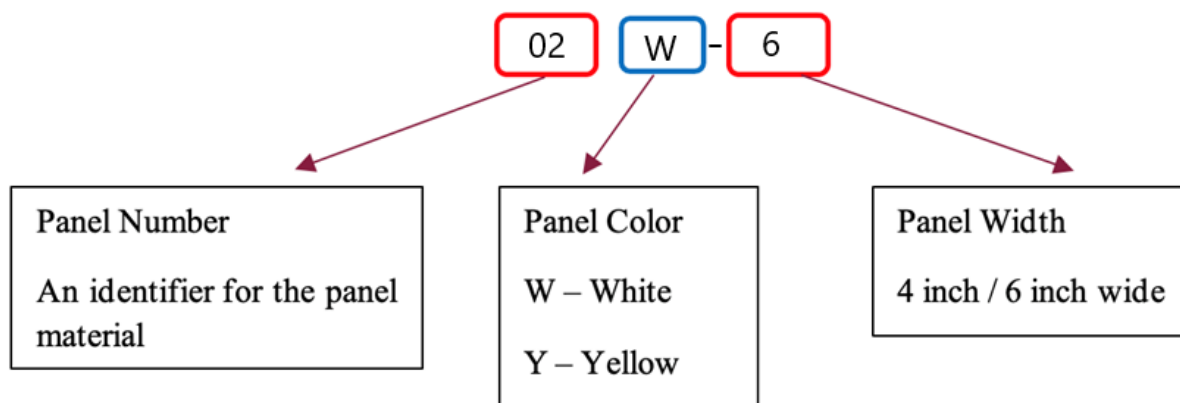


Figure 3. Panel identification scheme.

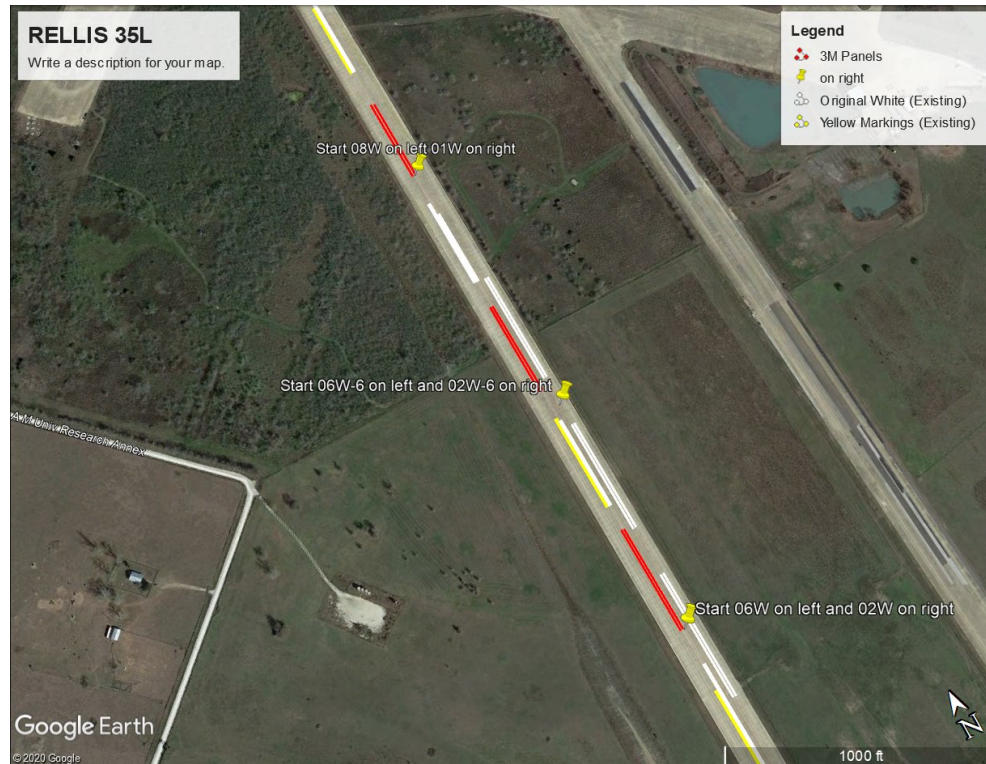


Figure 4: Google Earth view of runway 35L at RELLIS campus and locations of 3M panel placement.



Figure 5: Images of 40 ft gap pavement markings on 35L (Set 2: 06W-6 on left and 02W-6 on right) during day (above) and night (below).

Video Data Evaluation Metric

To enable evaluation of LD performance, images were extracted from the videos and annotated using the Scalabel Annotation tool [14]. Scalabel was developed by the Berkeley Deep Drive group for labeling their BDD100K [11] dataset. The Scalabel tool supports different types of annotations needed for training computer vision models, especially for a driving environment. For each image in the dataset, the research team manually annotated the lane markings using 2D polylines as supported in Scalabel. The annotations include three feature attributes for the lane markings—(i) lane categories (road curb, white, yellow, crosswalk), (ii) lane type (single, double) (iii) lane continuity (continuous/solid, dashed/skip)—and the lanes are annotated accordingly.

LD performance was evaluated using LD algorithms powered by three different state-of-the-art neural networks. (i) SCNN [9] (ii) LaneNet [15] (iii) ENet [16]. SCNN was developed to address prevalent LD issues, including processing speeds and complexity, and to efficiently learn the spatial relationship of structured objects like lane markings in driving scenarios. SCNN generalizes the traditional deep layer-by-layer convolutions to slice-by-slice convolutions within feature maps and enables message passing between pixels across rows and columns in each layer. Thus, SCNN type algorithms are particularly suitable for long continuous shape structures or large objects, with a strong spatial relationship but minimal appearance features, such as poles, walls, traffic lanes, etc. An optimized implementation of SCNN won first place in the TuSimple Benchmark Lane Detection Challenge, achieving an accuracy score of 96.53%. To evaluate the performance of the SCNN algorithm on the College Station dataset, the annotations in BDD100K format were converted to annotations to suit the SCNN format. SCNN supports the detection of three lanes, which corresponds to a maximum of four pavement markings in an image. The ground truth information from the annotation files is used to evaluate LD performance.

Neven et al. [17] developed an end-to-end algorithm that approaches LD as an instance segmentation problem. LaneNet is used as the backbone CNN, which combines the benefits of binary lane segmentation by forming an instance of each lane that can be trained end-to-end. Additionally, a network referred to as H-Net estimates the parameters for an “ideal” perspective transformation customized for each input image in contrast to the typical “bird’s eye view” transformation, thus ensuring a robust lane fitting model. LaneNet supports a maximum of five lane markings in lanes, where four lane markings correspond to the current lane and left/right lanes. The extra lane is in place for cases during lane changes in order to reduce confusion in identifying the current lane. Annotations contain polyline data of the lane markings constructed using pixel data organized by the same distance gap (“h_sample” in each label data) from the recording car/bottom of the image frame.

ENet, short for Efficient neural network, is a deep neural network architecture specifically created for tasks like semantic segmentation, which requires low latency in its execution, that can operate in real-time on low-power mobile devices. ENet-label [16] is a light-weight LD model based on ENet and adopts self-attention distillation. It has 20 times fewer parameters and runs 10 times faster compared to SCNN. ENet-label achieves an F1-measure of 72.0 on the CULane testing set

(better than SCNN, which achieves 71.6). It also achieves 96.64% accuracy in the TuSimple testing set. The researchers choose to evaluate this algorithm because of its claims of high performance and low latency.

The LD algorithm performance is measured in terms of the conventional pixel-accuracy-based performance metrics, such as True Positive (TP), False Positive (FP), False Negative (FN), F-Measure, etc. Both SCNN and ENet-label follow the same performance evaluation method. To evaluate if a lane marking is successfully detected, the lane markings are detected as lines with widths equal to 30 pixels. The intersection-over-union (IoU) metric (Figure 6) is calculated using the ground truth annotation and the lane prediction from the algorithm. The lane predictions where IoUs are larger than a certain threshold (in our evaluation we consider 0.5) are viewed as TPs. Based on the predictions the F-Measure is calculated as

$$F - measure = \frac{(1 + \beta^2)(P * R)}{\beta^2(P + R)}$$

where,

$$Precision (P) = \frac{TP}{(TP + FP)} = \frac{True Positives}{Total detections marked Positive}$$

$$Recall (R) = \frac{TP}{(TP + FN)} = \frac{True Positive}{Total Positives in Ground truth}$$

$\beta = 1$, which gives the harmonic mean (F1-measure).

Algorithm performance is related to Precision (P), Recall (R), and F1 scores on a linear scale. Higher scores represent better algorithm LD performance. The LD outputs from the SCNN algorithm are illustrated in Figure 7. Algorithm outputs in “Green” are the ground truth markings. The algorithm outputs in “Red” are the FP lane predictions. Algorithm outputs in “Blue” are the TP lane predictions that match the ground truth with an IoU > 0.5. IoU is the intersection area over the union measure between the ground truth and the prediction.

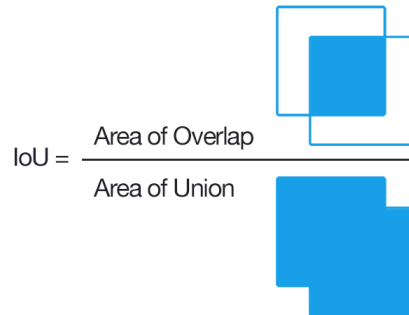


Figure 6: Definition of intersection over union. Source: <https://www.pyimagesearch.com/>

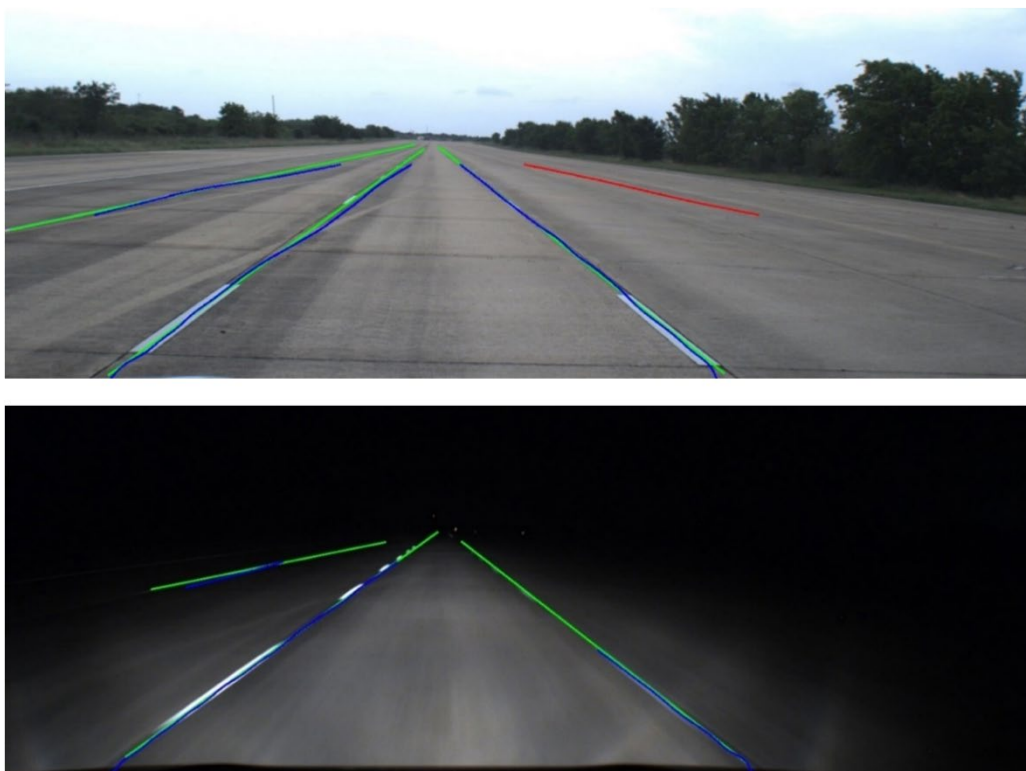


Figure 7: LD outputs of 40 ft gap lane markings on 35L (Set 2: 06W-6 on left - 02W-6 on right) during day (above) and night (below).

Material Characteristics Data

Retroreflectivity and Other Material Properties

Most surfaces have mixed reflection, made up of diffuse reflection, retroreflection, and mirror reflection. Figure 8 shows an illustration of different types of light reflections from a surface. Retroreflection is when an object returns light back toward the direction of the light source. The physical measure of the brightness of a surface is luminance L , which is the luminous intensity of light towards the driver's eyes in proportion to the apparent area of the surface. As the luminous intensity is measured in units of candela (cd) and surface area is measured in square meters (m^2), the unit of luminance is $cd \cdot m^{-2}$. Daylight reflection of pavement markings is described by the luminance factor β (or Y) measured in $45^\circ/0^\circ$ geometry and/or by the luminance coefficient under diffuse illumination (Q_d) with observation as in standard 30 m geometry [18]. Retroreflection is a very useful marking property for nighttime driving. A driver of a vehicle sees the retroreflection from the marking that is brighter than the pavement surface, thus making the marking stand out. The luminance coefficient is applicable for a retroreflecting surface but is referred to as “coefficient of retroreflected luminance” or “retroreflectivity (R_L).” R_L is a surrogate measure for how visible the marking will be at night. A higher retroreflectivity value indicates a marking that is more efficient at returning light from the vehicle headlamps back toward the vehicle's driver, making the marking appear bright. Each marking was evaluated for Q_d and R_L along the entire length of the test area.

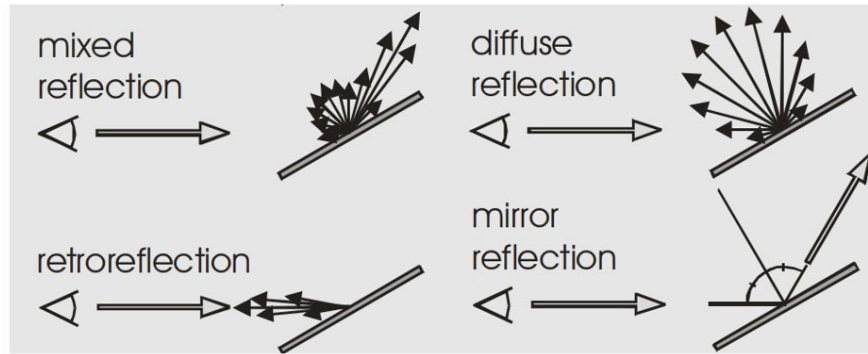


Figure 8: Illustration of types of light reflections from a surface. Source: <https://madebydelta.com/>

Material Data Collection

The color of the markings was evaluated using a spectrophotometer in the CIE 1931 x, y, Y, color space using illuminant D65 (representing daylight; ISO/CIE 10526) and a 2-degree standard observer. The x and y values are the color coordinate locations on the CIE 1931 chromaticity diagram. The CIE Y value is the brightness, with 0 representing perfect black and 100 representing perfect white. Color measurements were conducted at multiple locations in each test area at locations that were representative of the test area. A combination of mobile and portable retroreflectometers was used to capture the pavement marking reflectivity values. Lane marking performance characteristic data (including color, Q_d , and R_L) were collected around the same time that the video data was collected.

Results

College Station Dataset

The performance of the LD algorithms for different test scenarios in the College Station dataset was evaluated in terms of Precision, Recall, and F1 scores; results are tabulated in Table 2. The evaluation was carried out without explicitly training the algorithms on the College Station dataset. The reasoning behind this was that we aimed to evaluate the performance of an LD algorithm when it encounters a completely new test scenario and check how the LDs could co-relate to the lane marking characteristics.

Table 3 lists the lane marking performance characteristic data collected along the same roads. ENet-Label performed best on the College Station dataset among the three LD algorithms evaluated. The LD algorithms had the lowest performance on 03 Jones Road, mainly due to the absence of pavement markings for most of the road. Lane markings on 04 Leonard Road had the best overall performance in terms of detectability. This can be attributed to several factors, including high contrast between the roadway and the pavement markings, and viewing conditions, which include observation direction and time of day. LD algorithms performed better detections overall during morning times as compared to other times of the day.

Since the sun was more overhead at 11 a.m., the glare-inducing light sources were absent and the luminance of the roadway surface was relatively constant, resulting in improved LD performance. Roads with higher CapY and R_L values of pavement markings produced better LD performance scores. Overall, LD performance was significantly lower on roads (01, 02, 03, 05, 06) during evening times than other times of day when driving westwards. This can be attributed to the direct sun glare on the camera sensor and the low angle solar illumination of the road. In these cases, since the source of light (i.e., the sun) was over the horizon, emitting light towards the camera at a low angle, the specular reflections on the roads were high. These specular reflections tend to reduce the contrast between pavement marking and road, affecting the ability of LD algorithms to detect pavement markings.

However, we observed the exact opposite trend on roads where the data were collected when driving eastwards (04 Leonard Road) during the evening. Since the source of light was illuminating the road along with the camera's field-of-view (FOV), these lighting conditions resulted in high road illumination, improving the detectability of the lane markings, which resulted in better LD performance during the evening. During the nighttime, since the source of light (vehicle headlights) illuminated the road along the camera's FOV, the pavement markings light up because of retroreflection, whereas the roadway appears darker because it is not retroreflective. Thus, the perceived levels of contrast are expected to be much higher on roads with higher pavement marking retroreflectivity and CapY, resulting in better LDs.

Similarly, as seen in Table 2, roads with pavement markings having higher retroreflectivity (R_L) values (01, 02, and 04) showed better performance during nighttime detection as compared to roads with lower retroreflectivity values (03 and 05). The retroreflectivity values and CapY appeared to have a positive effect on LD performance during nighttime conditions. However, roads 06 and 07 exhibited unanticipated behavior. Even though they had the highest retroreflectivity values among all the roads, their nighttime detections were found to be poor, which indicates that additional parameters also need to be considered to predict road detectability behavior. Parameters like road surface roughness, road surface cracks, ghost markings and speed of vehicle can also influence the performance of LD algorithms, which we did not account for in this analysis.

Table 2: Performance of Different Algorithms (IoU Threshold = 0.5) on College Station Dataset

Road (Driving Speed)	Road Conditions*	Driving Direction	Time of Day	Image Count	SCNN Precision (P)	SCNN Recall (R)	SCNN F1-Score (F1)	LaneNet Precision (P)	LaneNet Recall (R)	LaneNet F1-Score (F1)	ENet- label Precision (P)	ENet- label Recall (R)	ENet- label F1-Score (F1)
01. S. College Ave (50 mph)	2-Lane Road, M Quality, M Contrast	NW	Morning	84	0.6	0.547	0.572	0.587	0.603	0.595	0.734	0.650	0.689
01. S. College Ave (50 mph)	2-Lane Road, M Quality, M Contrast	NW	Evening	25	0.306	0.339	0.322	0.424	0.485	0.452	0.435	0.357	0.392
01. S. College Ave (50 mph)	2-Lane Road, M Quality, M Contrast	NW	Night	94	0.494	0.391	0.437	0.325	0.468	0.384	0.506	0.382	0.435
02. W Villa Maria (45 mph)	2-Lane Road, M Quality, M Contrast	SW	Morning	225	0.445	0.444	0.445	0.318	0.357	0.336	0.551	0.377	0.448

Road (Driving Speed)	Road Conditions*	Driving Direction	Time of Day	Image Count	SCNN Precision (P)	SCNN Recall (R)	SCNN F1-Score (F1)	LaneNet Precision (P)	LaneNet Recall (R)	LaneNet F1-Score (F1)	ENet- label Precision (P)	ENet- label Recall (R)	ENet- label F1-Score (F1)
02. W Villa Maria (45 mph)	2-Lane Road, M Quality, M Contrast	SW	Evening	76	0.246	0.358	0.292	0.224	0.325	0.265	0.321	0.386	0.351
02. W Villa Maria (45 mph)	2-Lane Road, M Quality, M Contrast	SW	Night	166	0.470	0.407	0.437	0.412	0.384	0.397	0.462	0.450	0.456
03. Jones Road (30 mph)	1-Lane Road, Yellow Centre Line / No lane Markings	NW	Morning	116	0.100	0.491	0.167	0.102	0.152	0.122	0.089	0.403	0.146
03. Jones Road (30mph)	1-Lane Road, Yellow Centre Line / No lane Markings	NW	Evening	58	0.081	0.474	0.128	0.076	0.228	0.114	0.107	0.632	0.183
03. Jones Road (30 mph)	1-Lane Road, Yellow Centre Line / No lane Markings	NW	Night	97	0.107	0.684	0.186	0.085	0.156	0.110	0.098	0.449	0.161
04. Leonard Road (55 mph)	1-Lane Road, G Quality, H Contrast	NE	Morning	116	0.799	0.794	0.796	0.751	0.698	0.723	0.619	0.651	0.635
04. Leonard Road (55 mph)	1-Lane Road, G Quality, H Contrast	NE	Evening	71	0.811	0.771	0.791	0.824	0.726	0.772	0.791	0.749	0.769
04. Leonard Road (55 mph)	1-Lane Road, G Quality, H Contrast	NE	Night	192	0.786	0.765	0.775	0.691	0.712	0.701	0.659	0.721	0.688
05. Harvey Mitchell Pway (60 mph)	2-Lane Road, G Quality, M Contrast	NW	Morning	87	0.729	0.866	0.791	0.735	0.811	0.771	0.755	0.931	0.834
05. Harvey Mitchell Pway (60 mph)	2-Lane Road, G Quality, M Contrast	NW	Evening	48	0.547	0.761	0.637	0.673	0.587	0.627	0.589	0.787	0.674
05. Harvey Mitchell Pway (60 mph)	2-Lane Road, G Quality, M Contrast	NW	Night	95	0.494	0.391	0.437	0.356	0.483	0.410	0.506	0.381	0.435
06. TX – 21 (65 mph)	2-Lane Highway, G Quality, H Contrast	SW	Morning	130	0.676	0.792	0.729	0.659	0.735	0.695	0.682	0.759	0.718
06. TX – 21 (65 mph)	2-Lane Highway, G Quality, H Contrast	SW	Evening	106	0.577	0.687	0.627	0.563	0.628	0.594	0.528	0.683	0.595
06. TX – 21 (65 mph)	2-Lane Highway, G Quality, H Contrast	SW	Night	242	0.470	0.407	0.437	0.441	0.412	0.426	0.462	0.450	0.456
07. Hway 47 (75 mph)	2-Lane Highway, G Quality, H Contrast	SE	Morning	35	0.712	0.743	0.727	0.699	0.724	0.711	0.701	0.744	0.722
07. Hway 47 (75 mph)	2-Lane Highway, G Quality, H Contrast	SE	Evening	15	0.611	0.628	0.619	0.712	0.602	0.652	0.536	0.628	0.579
07. Hway 47 (75 mph)	2-Lane Highway, G Quality, H Contrast	SE	Night	28	0.103	0.450	0.167	0.105	0.126	0.114	0.098	0.449	0.162

*Good = G, Medium = M, High = H

Table 3: Lane Marking Characteristics Data of College Station Dataset Based on ASTM Standards

Road	Left Markings Color	Left Markings	Left Markings	Left Markings	Left Markings	Right Markings Color	Right Markings	Right Markings	Right Markings	Right Markings	Right Markings
------	---------------------------	------------------	------------------	------------------	---------------	----------------------------	-------------------	-------------------	-------------------	-------------------	-------------------

	Lane Type	x	y	Cap Y	Average R_L (med/m ² /lx)	Lane Type	x	y	Cap Y	Average R_L (med/m ² /lx)
01. S. College Ave	White Skip	0.334	0.35	24.758	190.34	White Edge	0.3375	0.3530	28.7339	200
01. S. College Ave	White Skip	0.3361	0.3522	34.7373	212	No Lane Marking				
01. S. College Ave	Yellow Centre	0.4237	0.3958	25.2798	73.33	White Edge	0.3316	0.3493	44.87	149
02. W Villa Maria	White Skip	0.333	0.3502	35.8026	114.33	No Lane Marking				
03. Jones Road	Yellow Centre	0.4057	0.3894	24.124	88	No Lane Marking				
04. Leonard Road	Yellow Centre	0.4111	0.3900	24.3590	107.67	White Edge	0.3406	0.3559	34.6	171
05. Harvey Mitchell Pway	White Skip	0.3303	0.3469	36.4182	56.67	White Edge	0.3329	0.3495	33.324	88
06. TX – 21	White Skip	0.3375	0.3534	38.09	254.33	White Edge	0.3356	0.3518	33.67	259.34
07. Hwy 47	White Skip	0.3406	0.3570	37.61	191	White Edge	0.3344	0.3519	46.21	199

Closed Course 3M Panel Dataset

Based on the different testing conditions employed, or aim was to evaluate different factors' effects on LD performance (F1 scores). The following factors were chosen based on which had a major influence on LD performance of markings in previous trials and the literature.

- Factor 1: Spacing between panels: 30 ft vs 40 ft.
- Factor 2: Time (illumination): daytime vs nighttime (high beam headlamps [HB] vs low beam headlamps [LB]),
- Factor 3: Driving Direction: northbound (NB) vs southbound (SB)
- Factor 4: Evaluation Area: full length vs near FOV vs far FOV.
 - Near FOV corresponds to the visible area up to 30 ft ahead of where the camera sees the road. Far FOV corresponds to the visible area 30 ft away from where the camera sees the road up to the horizon of visibility. Figure 9 illustrates the area-split for near FOV vs far FOV evaluations.
- Factor 5: Panel Material [01W-4, 02W-4, 02W-6, 06W-4, 06W-6, 08W-4].
 - Each panel is associated with its material property: a) Q_d for daytime data, b) R_L for nighttime data, which are evaluated simultaneously.

We expected there to be some overlap in interactions that affect LD performance, depending on the specific factor being studied. We studied these interactions to form a benchmark reference system on factors that affect LD. Table 4 lists the pavement material characteristic data collected under different test conditions.

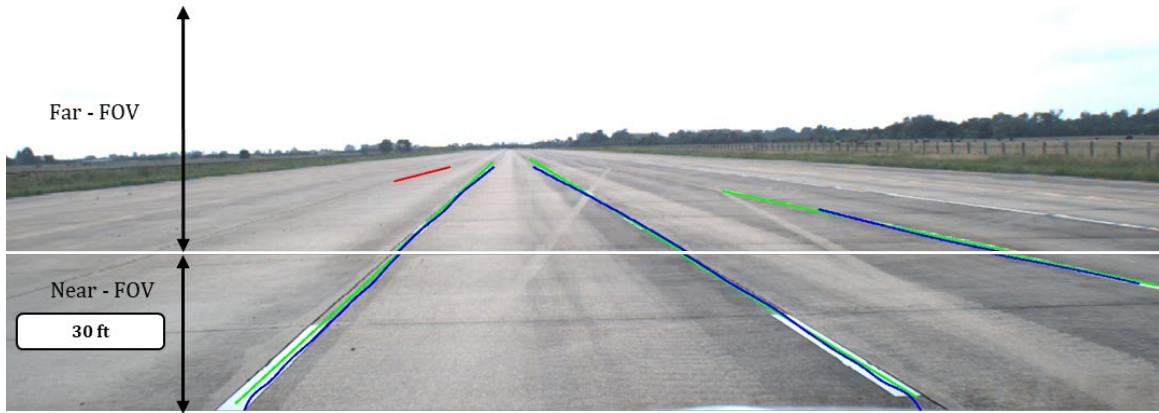


Figure 9: Image showing the area-split for near FOV vs far FOV evaluations.

Table 4: 3M Pavement Panel Material Characteristic Data – 05/21/2020 – RELLIS 35L – Retroreflectivity and Color

Group	Marking Type	Measurement Direction	R _L	Q _D	x	y	Y
Set 1 – L	01W	NB	94	203	0.3297	0.3504	51.99
Set 1 – L	01W	SB	84	199	0.3297	0.3504	51.99
Set 1 – L	01W	Pavement	18	86	0.3680	0.3678	18.59
Set 1 – R	08W	NB	1404	184	0.3256	0.3447	58.15
Set 1 – R	08W	SB	1045	192	0.3256	0.3447	58.15
Set 1 – R	08W	Pavement	18	78	0.3690	0.37	20.08
Set 2 – L	02W-6	NB	136	179	0.3227	0.3419	43.48
Set 2 – L	02W-6	SB	127	181	0.3227	0.3419	43.48
Set 2 – L	02W-6	Pavement	20	90	0.3687	0.3681	21.17
Set 2 – R	06W-6	NB	312	171	0.3227	0.3419	43.48
Set 2 – R	06W-6	SB	279	171	0.3227	0.3419	43.48
Set 2 – R	06W-6	Pavement	18	77	0.3649	0.37	18.46
Set 3 – L	02W	NB	141	178	0.3221	0.3413	43.06
Set 3 – L	02W	SB	126	181	0.3221	0.3413	43.06
Set 3 – L	02W	Pavement	19	86	0.3686	0.3693	22.20
Set 3 – R	06W	NB	313	173	0.3237	0.3428	38.27
Set 3 – R	06W	SB	288	172	0.3237	0.3428	38.27
Set 3 – R	06W	Pavement	17	71	0.3738	0.3714	21.41

Discussion

A statistical approach was employed to analyze the 3M panel data and investigate the effect of different marking and evaluation condition factors on LD performance. An analysis of variance (ANOVA) model was employed to analyze the individual factors as well as the effect of two-way interactions between factors to identify those with statistical significance. Table 5 lists the factors that were considered in the ANOVA model with F1 scores as the response variable. The analysis

was carried out using the JMP software suite. ANOVA was conducted separately for daytime and nighttime data. Two-way interaction factors and individual factors without a statistically significant two-way interaction were removed from consideration in the final ANOVA model. Table 6 and Table 7 show the statistically significant factors as predicted by the ANOVA model for the nighttime and daytime testing.

Table 5: List of Factors Considered in the ANOVA Model

Factor	Level
Spacing	30 ft gap, 40 ft gap
Light	Day, Night-LB, Night-HB
Driving Direction	SB, NB
Evaluation Area	Near, Full Length, Far
Panel-Width	01W-4, 02W-4, 02W-6, 06W-4, 06W-6, 08W-4

Table 6: List of Statistically Significant Factors During Nighttime Based on ANOVA Model for LD Performance (F1-Scores)

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Spacing	1	1	0.74899927	682.2828	< 0.0001*
Light	1	1	0.00107002	0.9747	0.3255
Driving Direction	1	1	0.02126749	19.3731	< 0.0001*
Eval. Area	2	2	0.04663060	21.2385	< 0.0001*
Panel	5	5	0.02830087	5.1560	0.0003*
Spacing*Light	1	1	0.00490784	4.4707	0.0365*
Spacing*Driving Direction	1	1	0.02145441	19.5434	< 0.0001*
Spacing*Panel	5	5	0.13685526	24.9330	< 0.0001*
Light*Panel	5	5	0.01803733	3.2861	0.0081*

Table 7: List of Statistically Significant Factors During Daytime Based on ANOVA Model for LD Performance (F1-Scores)

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Spacing	1	1	0.08099427	62.8570	< 0.0001*
Driving Direction	1	1	0.03735097	28.9868	< 0.0001*
Eval. Area	2	2	0.14112119	54.7597	< 0.0001*
Panel	5	5	0.03871050	6.0084	0.0002*
Spacing*Driving Direction	1	1	0.04192916	32.5398	< 0.0001*
Spacing*Panel	5	5	0.02143691	3.3273	0.0106*

Factor 1: Spacing Between Marking Panels: 30 ft vs 40 ft:

- As seen in Figure 10, LD performance was observed to be higher for 30 ft gap markings as compared to 40 ft gap markings during all times of day. This trend can be attributed to the fact that closer lane markings essentially mean more visual features for the LD algorithms to detect,

leading to higher LD performance. Table 8 lists the least square mean value of F1 scores output for panels with different spacing configurations.

- Figure 11 and Figure 12 shows the 2-way interaction of panel spacing on panel material for LD performance during daytime and nighttime respectively.
- Figure 24, Figure 25, and Figure 26 (Appendix 3) give a comparison of panel-wise LD performance for 30 ft gap vs 40 ft gap panel configurations. In Figure 24 (Appendix 3), panel F1 scores are compared against Q_d values. In Figure 25 (Appendix 3), panel F1 scores are compared against R_L values. Similar to the observations made on combined performance (pairwise) of panels, individual panels (panel wise) were observed to perform better when spaced 30 ft apart than when spaced 40 ft apart.
- Table 15 (Appendix 3) contains results from the Tukey Test for two-way interaction between panel spacing and panel materials. In a Tukey table, the levels that are not connected by the same letter are significantly different from each other statistically. We can see that there exists a significant difference between the 30 ft gap (Level A) and the 40 ft gap (Level D) during daytime, as noted in our previous discussion of the results.

Table 8: Least Square Means Table Comparing the Effect of Panel Spacing on LD Performance

Level	Least Sq Mean	Std Error	Mean
30 ft – Daytime	0.72230940	0.00598273	0.722309
40 ft – Daytime	0.65522973	0.00598273	0.655230
30 ft – Nighttime	0.86972840	0.00406634	0.869728
40 ft – Nighttime	0.72548716	0.00406634	0.725487

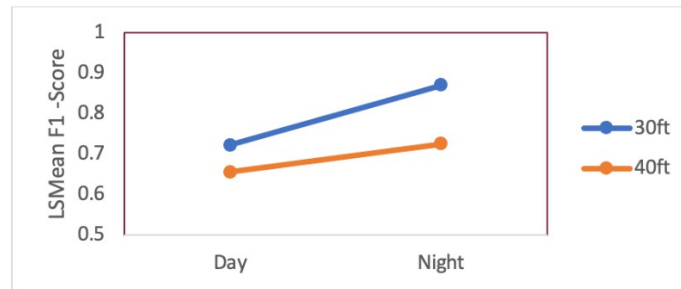


Figure 10: Plot of least square mean F1-score capturing the effect of panel spacing on LD performance.

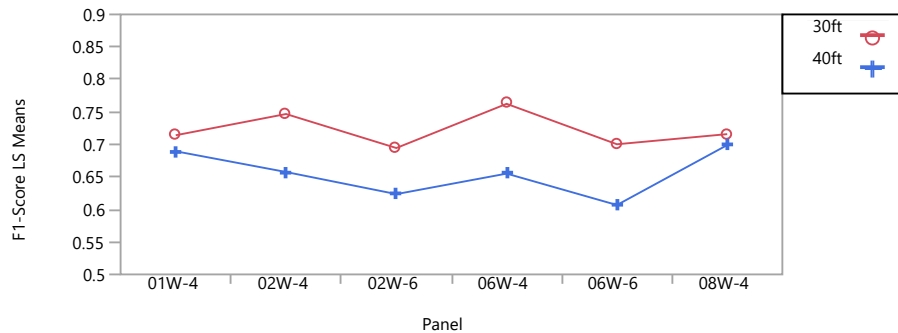


Figure 11: Plot of least square mean F1-score capturing the effect of 2-way interaction of panel spacing on panel material for daytime LD performance.

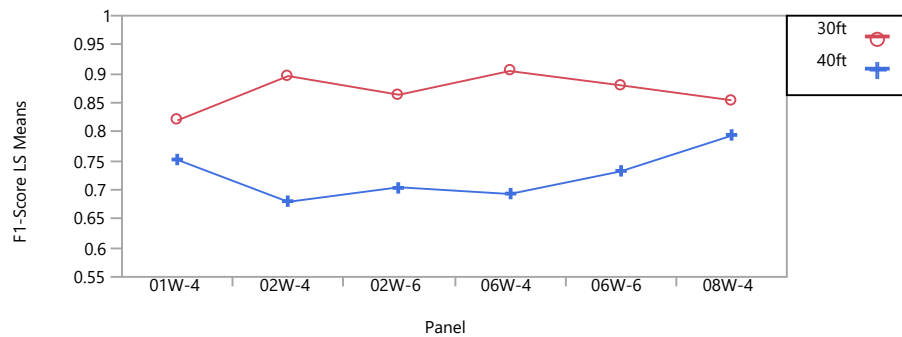


Figure 12: Plot of least square mean F1-score capturing the effect of two-way interaction of panel spacing on panel material for nighttime LD performance.

Factor 2: Time (Illumination) – Day vs Night (High Beam Headlamps vs Low Beam Headlamps):

- The panels' nighttime LD performance was observed to be overall better as compared to the summer daytime performance. This trend was also observed for both the combined and panel-wise LD performance shown in Figure 26 (Appendix 3.) and for area-based LD evaluation, as shown in Figure 27, Figure 28, and Figure 29 (Appendix 3.).
- Figure 14 captures the effect of two-way interactions of nighttime illumination on panel material for LD performance. Panel 08W-4 was expected to have the highest LD performance, whereas 01W-4 was expected to have the lowest LD performance. The results show that even though there is a general increasing trend for both high beam and low beam data, the 08W-4 LMS F1 score for low beam illumination is comparatively lower. A clear conclusion could not be drawn for this observation. This counter-intuitive observation can be attributed to the fact that the calculation of LMS scores for the two-way interactions of nighttime illumination on panel material inherently includes other interactions, such as panel spacing (30-ft gap vs 40-ft gap) and evaluation-area (near FOV vs far FOV). Another explanation for this observation could be the fact that since 08W (highest R_L) is paired with 01W (lowest R_L), the panel material with a lower F1 score could adversely affect the performance of the panel it was paired with, decreasing its detection. This reduced performance is higher in far FOV measurements (refer Figure 27, Figure 28, and Figure 29 in Appendix 3.), which might result in lower LD scores for 08W.
- In each pair of lane markings, markings with higher R_L values generally exhibited better LD performance (had higher F1 scores as compared to panels with lower R_L values) for panel-wise evaluations during nighttime (08W in Set 1, 06W-6 in Set 2, 06W in Set 3; Figure 14). A similar trend was observed in evaluation-area based LD performance scores (Figure 27–Figure 29 in Appendix 3.).
- The difference in LD performances between panels (of different materials) in each Set was observed to be greater during daytime than nighttime (Figure 26, Figure 27–Figure 29 in Appendix 3.). A possible explanation for this trend is that LD performance is more sensitive

to the reflectivity of the panel during daytime (Q_d values) than during nighttime (R_L values). This observation can also be corroborated from the LD scores in Figure 14, which do not vary significantly and lie in the small band even though retroreflectivity values are observed to vary significantly.

- Table 9 shows the least square means table comparing the effect of nighttime illumination on LD performance. Table 17 and Table 18 in Appendix 3. list the least-square mean differences Tukey HSD test results ($\alpha = 0.050$) for two-way interaction between panel spacing and nighttime-illumination, and between nighttime illumination and panel material, respectively.

Table 9: Least Square Means Table Comparing the Effect of Nighttime Illumination on LD Performance

Level	F1 Score (LS Mean)	Std Error	Mean
High Beam (HB)	0.79488186	0.00390474	0.794882
Low Beam (LB)	0.80033371	0.00390474	0.800334

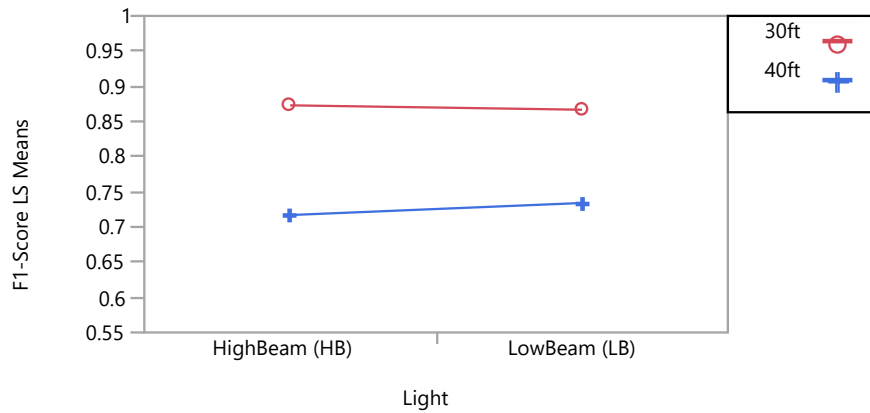


Figure 13: Plot of least square mean F1-score capturing the effect of nighttime illumination on LD performance.

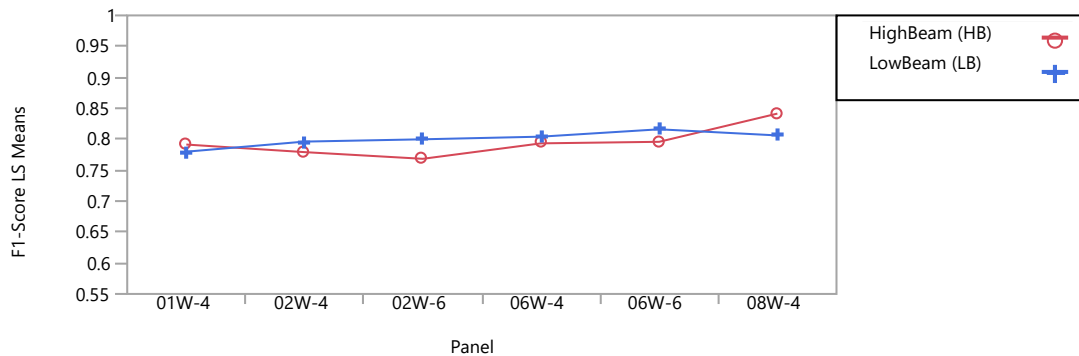


Figure 14: Plot of least square mean F1-score capturing the effect of two-way interaction of nighttime illumination on panel material for LD performance.

Factor 3: Driving Direction: Northbound (NB) vs Southbound (SB)

- The LD performance of markings was observed to be generally higher when driving southwards as compared to driving northwards during the daytime, as seen in Figure 21 (in Appendix 3.).
- The driving direction did not impact the nighttime data collection for the 30 ft spacing, but there was some impact for the 40 ft spacing. We generally would not expect an impact at night because illumination is relatively consistent from the vehicle headlamps.
- Table 10 is the least square means table comparing the effect of driving direction on LD performance. Table 16 (in Appendix 3.) lists the Tukey HSD test ($\alpha = 0.050$) of LSMeans differences for two-way interaction between panel spacing and driving direction during daytime (Left) and nighttime (Right).

Table 10: Least Square Means Table Comparing the Effect of Driving Direction on LD Performance

Level	Least Sq Mean	Std Error	Mean
Northbound (NB) – Day	0.66599319	0.00598273	0.665993
Southbound (SB) – Day	0.71154594	0.00598273	0.711546
Northbound (NB) – Night	0.80976059	0.00390474	0.809761
Southbound (SB) – Night	0.78545497	0.00390474	0.785455

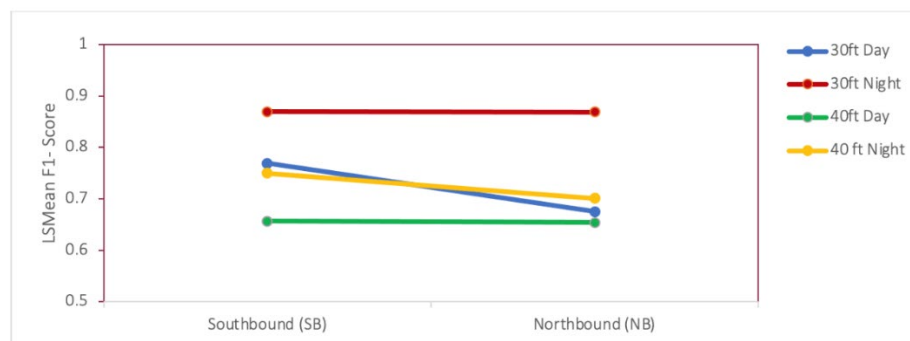


Figure 15: Plot of least square mean F1-score capturing the effect of 2-way interaction of panel spacing and driving direction on LD performance.

Factor 4: Evaluation Area [Full Length vs Near FOV vs Far FOV]

- LD performance for near FOV was observed to be generally higher than that of far FOV (Figure 27, Figure 28, and Figure 29 in Appendix 3.). This trend was consistently observed across all times of day for all driving directions. Due to the perspective transformation of the image captured by the camera, it is expected that the objects near the camera appear with greater detail than the objects that are farther away. Thus, the LD algorithms have more image features available to detect lane markings in near FOV images as compared to far FOV images.
- As observed in Figure 27, Figure 28, and Figure 29 (Appendix 3.), the panels LD performance was better in the 30 ft gap configuration as compared to 40ft gap in both Near FOV and Far FOV evaluations.

- LD performance for Near FOV was observed to be generally higher than Combined LD performance whereas LD performance for Far FOV was observed to be generally lower (Figure 27, Figure 28, and Figure 29 in Appendix 3.) across all times of the day. Since the combined LD scores were obtained from an image that combines the features from Near FOV and Far FOV images, it is expected to have an F1 score somewhere in-between.
- In each pair of lane markings, panel materials with higher R_L values generally exhibited better LD performance (had higher F1 scores as compared to panels with lower R_L values) for panel-wise evaluations during Night-time for both Near FOV and Far FOV (08W in Set1, 06W-6 in Set 2, 06W in Set 3).
- The difference in performance between Near vs Far FOV evaluations was generally higher in the 30ft gap configuration as compared to 40 ft gap configuration (Figure 27, Figure 28, and Figure 29 in Appendix 3.). Since the near FOV evaluation area for LD was 40 ft from where the camera saw the road, it was expected that a few image frames for the 40 ft gap configuration would not contain any lane markings. Conversely, the 30 ft near FOV videos always contained some lane marking characteristics, which aided LD, thus reducing the difference in LD scores between near vs far FOV evaluations as compared to the 40 ft gap.
- Another limitation in the right vs left evaluation is that the LD algorithm also predicts the adjacent lane markings, which affects the LD performance scores.

Table 11: Least Square Means Table Comparing the Effect of Evaluation Area on LD Performance

Level	Least Sq Mean	Std Error	Mean
Day – Far FOV	0.63541809	0.00732731	0.635418
Day – Full Length	0.68706855	0.00732731	0.687069
Day – Near FOV	0.74382205	0.00732731	0.743822
Night – Far FOV	0.77465624	0.00478231	0.774656
Night – Full Length	0.79956223	0.00478231	0.799562
Night – Near FOV	0.81860487	0.00478231	0.818605

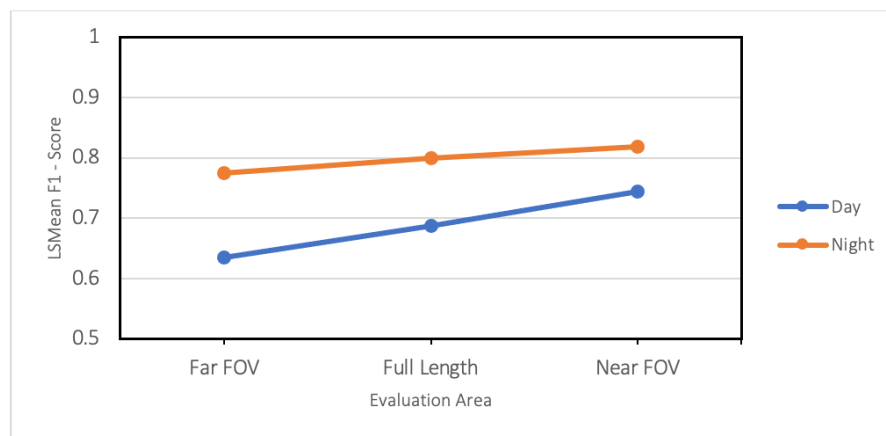


Figure 16: Plot of Least square mean F1-score capturing the effect of evaluation area on LD performance.

Factor 5: Panel Material [01W, 02W, 06W, 08W]

- Lane markings with higher R_L values generally exhibited better LD performance during the nighttime (higher F1 scores). Panels 01W, 02W, 06W, and 08W had R_L values that increased in the order listed here (Table 4), and this trend can be observed in Figure 17, where nighttime F1 scores appear to be increasing with increasing retroreflectivity (R_L) values from 01W to 08W. (Figure 17). During the daytime, all 4-inch-wide panels had similar F1 scores.
- This trend can be correlated to the fact that the Q_d value, which is an important property that affects LD performance during the daytime, is also similar ($Q_d \sim 170$ to 200 as seen in Table 4) for these panels. A limitation of this evaluation is that the Q_d values of the samples chosen do not vary much between samples. However, the existing pavement markings on the closed section had a wide range of values. As seen in Figure 23 (Appendix 3.), LD performance had an increasing trend for Q_d values.
- Set 1 contains 08W material panels, which have the highest R_L value (~ 1200) and 01W material panels, which have the lowest R_L value (~ 80). However, the LD performance of 08W was observed to be just 4% better than that of 01W (see Table 12). There could be several factors that may have influenced this behavior.
 - The increase in R_L values may not necessarily convert to improved LD performance by similar magnitudes.
 - F1 scores for individual performances of 01W and 08W were extracted by processing an image that included both 01W and 08W (Set 1). There may exist some co-operative interactions between 01W and 08W that improved the individual F1 score of 01W panel, or conversely, 08W's performance may have been reduced due to 01W, leading to the two panels having similar F1 scores. Further investigations are required to better understand this observation.
- LD performance of the 6-inch-wide panels was observed to be generally lower than that of the 4-inch-wide panels during daytime (Figure 17). It was expected that the performance of the LD algorithms would be higher with the wider 6-inch panels since the algorithm has more features to detect lanes. However, this trend was not observed in this study.

Table 12: Least Square Means Table Comparing the Effect of Panel Material and Width on LD Performance

Level (Day)	F1 Score (LS Mean)	Std Error	Mean
01W-4	0.70177019	0.01036239	0.701770
02W-4	0.70216415	0.01036239	0.702164
02W-6	0.65900532	0.01036239	0.659005
06W-4	0.70877024	0.01036239	0.708770
06W-6	0.65349822	0.01036239	0.653498
08W-4	0.70740926	0.01036239	0.707409

Level (Night)	F1-Score (LS Mean)	Std Error	Mean
01W-4	0.78559844	0.00676321	0.785598
02W-4	0.78762791	0.00676321	0.787628
02W-6	0.78408181	0.00676321	0.784082
06W-4	0.79870553	0.00676321	0.798706
06W-6	0.80590180	0.00676321	0.805902
08W-4	0.82373120	0.00676321	0.823731

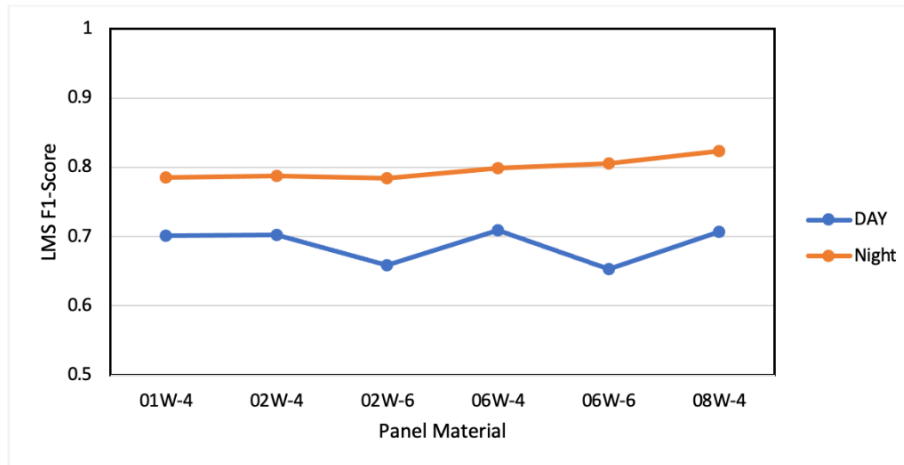


Figure 17: Plot of least square mean F1-score capturing the effect of panel material for LD performance.

Factor 6: Material Property a) Q_d for Daytime Data, b) R_L for Nighttime Data

- In each pair of lane markings, markings with higher R_L values generally contributed more towards better LD performance (had higher F1 scores) during nighttime (both low beam and high beam conditions) for panel-wise evaluations.
- In each pair of lane markings, markings with higher Q_d values generally contributed more towards better LD performance (had higher F1 scores) during daytime for panel-wise evaluations (Figure 26 in Appendix 3). Similar trends were observed in evaluation area-based LD performance (Figure 27, Figure 28, and Figure 29 in Appendix 3).

Conclusions and Recommendations

Future Work:

- Effect of parameters like road surface roughness, road surface cracks, ghost markings and speed of vehicle influences the performance of LD algorithms. This study did not quantitatively account for these effects in both on-road and closed course analysis.
- In the closed course evaluations of 3M panels, the ANOVA Model for LD Performance (F1-Scores) did not list the panel width (4in vs 6in) as statistically significant. This observation is puzzling since presumably more pixels are associated with the wider markings for similar distances and fields of view. Hence it is expected that the 6in panel must perform better than a 4in panel of the same material. As observed in Figure 26, 6in panels perform marginally better than 4in panels during nighttime, however the trend appeared to be reversed during daytime. A further study is needed to investigate this aspect better.
- Cross-interaction was observed in LD performance of Set 1 which contained 08W material panels (having high R_L value ~ 1200) and 01W material panels (having low R_L value ~ 80). The LD performance of 08W was observed to be just 4% better than that of 01W, Further study is proposed to study the cross interactions observed by pairing lane markings of varied Q_d/R_L values.

Conclusion:

The broad objective of this project was to carry out an in-depth study into the different factors affecting the performance of state-of-the-art LD algorithms by incorporating pavement marking material characteristics into the evaluation framework. We studied the effect of environmental factors (daytime vs nighttime, driving direction), lane marking material inherent characteristics (reflective properties like Q_d/R_L , marking quality), lane making layouts (30 ft gap vs 40 ft gap, 4-inch-wide vs 6-inch wide) and LD algorithm characteristics (type of algorithm, near FOV vs far FOV). Observations were made on how these different factors interact with each other. We noted that, at some level, each of these factors had an impact on the performance of machine-vision LD. Different annotated image datasets were also generated: 1) College Station Dataset, 2) 3M panel dataset, and 3) US 290 Dataset. Researchers can use these datasets as a reference/benchmark system to evaluate their LD algorithms and determine how their performance relates to the different material characteristics of the lane markings in these datasets. We hope this work will lead to cooperative infrastructure development of pavement markings, making them better suited for modern ADAS and automated vehicles.

Additional Products

The Education and Workforce Development (EWD) and Technology Transfer (T2) products created as part of this project can be downloaded from the Safe-D website [here](#). The final project dataset is available [Safe-D Dataverse](#).

Education and Workforce Development Products

Graduate student Abhishek Nayak's Ph.D. dissertation work on "Planning and vision-based systems for Autonomous vehicles" was funded by this Safe-D project. The following students also worked part-time or for a short span conducting annotations for the datasets in this project: Jay Shah, Andrew Lewis, and Ruicong Xie.

Technology Transfer Products

The following paper was published by graduate student Abhishek Nayak as part of his Ph.D. work. His Ph.D. work was partly supported by this project.

Nayak, A., Rathinam, S., Pike, A., & Gopalswamy, S. (2020). Reference Test System for Machine Vision Used for ADAS Functions (No. 2020-01-0096). SAE Technical Paper.

The PI, Dr. S. Rathinam, presented a summary of the work performed in this project and participated in a panel session at the Autonomous cars conference, Brookings Institution, Washington D.C. on July 25, 2019.

Dr. S. Rathinam presented an invited talk on this project titled "Reference Machine Vision for Advanced Driver Assist Systems (ADAS)" in the breakout session on *Reading the Road Ahead: Preparing Highway Infrastructure for ADAS and High Automation*, Automated Vehicles Symposium, July 16, 2019. He also participated in a panel discussion in this breakout session.

Dr. S. Rathinam participated in a panel and presented an invited talk on this project titled "Understanding the Correlation between the Quality of Markings and Lane Detection/Following Systems" at the IEEE-ITSS and ITE joint effort on *Development of Needs and Scope for Cooperative Infrastructure-Vehicle Detection and Localization for Automated Vehicles*, 3rd IEEE-ITS symposium, Sept 21, 2020.

Abhishek Nayak volunteered as a judge in the 2021 Virtual Edition of the Virginia State Science and Engineering Fair (VSSEF) conducted on April 10, 2021 and participated in knowledge sharing discussions.

Data Products

The data sets used for testing the lane detection performance of lane marking materials on different LD algorithms have been uploaded to the Safe-D Dataverse [DOI: 10.15787/VTI/AT0RHF](https://doi.org/10.15787/VTI/AT0RHF).

References

- [1] D. Pape and F. Habtemichael, "Infrastructure Initiatives to Apply Connected-and Automated-Vehicle Technology to Roadway Departures (No. FHWA-HRT-18-035)," in *United States. Federal Highway Administration*, Office of Safety Research and Development, 2018.
- [2] A. B. Hillel, R. Lerner, D. Levi and G. Raz, "Recent progress in road and lane detection: a survey.," *Machin Vision and Applications*, 25(3), pp. 727-745, 2014.
- [3] Y. Xing, C. Lv, L. Chen, H. Wang, H. Wang, D. Cao, E. Velenis and F.-Y. Wang, "Advances in vision-based lane detection: algorithms, integration, assessment, and perspectives on ACP-based parallel vision.," *IEEE/CAA Journal of Automatica Sinica*, pp. 5(3), 645-661, 2018.
- [4] S. P. Narote, P. N. Bhujbal, A. S. Narote and D. M. Dhane, "A review of recent advances in lane detection and departure warning system.," in *Pattern Recognition*, Elsevier, 2018, pp. 216-234.
- [5] C. Davies, "Effects of Pavement Marking Characteristics on Machine Vision Technology (No. 17-03724)," 2017.
- [6] P. J. Carlson, J. D. Miles, M. P. Pratt and A. M. Pike, "Evaluation of wet-weather pavement markings: First year report.," No. FHWA/TX-06/0-5008-1, 2005.
- [7] A. Pike, S. Clear, T. Barrette and P. J. Carlson, "Machine Vision Detection of Pavement Markings (No. 19-01495R1)," Transportation Research Board, 2018.
- [8] A. M. Pike, T. P. Barrette and P. J. & Carlson, " Evaluation of the Effects of Pavement Marking Characteristics on Detectability by ADAS Machine Vision," in *National Cooperative Highway Research Program (NCHRP)*, Washington, DC, 2018.
- [9] X. Pan, J. Shi, P. Luo, X. Wang and X. Tang, "Spatial as Deep: Spatial Cnn for Traffic Scene Understanding," *AAAI Conference on Artificial Intelligence (AAAI)*, February 2018.
- [10] TuSimple, "TuSimple Benchmark," [Online]. Available: <https://github.com/TuSimple/tusimple-benchmark>.
- [11] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning.," in

Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

- [12] A. T. S. S. Association, "ATSSA," 14 03 2019. [Online]. Available: <https://www.reflective-systems.com/wp-content/uploads/2019/04/Policy-re-Road-Markings-for-Machine-Vision-Systems.pdf>. [Accessed 01 02 2021].
- [13] P. J. Carlson, "NHCRP work on line marking / Austroads CAV6119, s.l.: s.n.," 2018.
- [14] "Scalabel," [Online]. Available: <https://www.scalabel.ai/>. [Accessed 01 02 2021].
- [15] Z. Wang, W. Ren and Q. Qiu, "Lanenet: Real-time lane detection networks for autonomous driving.," *arXiv preprint arXiv:1807.01726*, 2018.
- [16] A. Paszke, A. Chaurasia, S. Kim and E. Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation.," *arXiv preprint arXiv:1606.02147*, 2016.
- [17] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans and L. Van Gool, "Towards end-to-end lane detection: an instance segmentation approach.," *2018 IEEE intelligent vehicles symposium (IV)*, pp. 286-291, 2018.
- [18] A. E1710-18, "Standard Test Method for Measurement of Retroreflective Pavement Marking Materials with CEN-Prescribed Geometry Using a Portable Retroreflectometer," *ASTM International*, 2018.
- [19] H. Bei, R. Ai, Y. Yan and X. Lang, "Lane marking detection based on convolution neural network from point clouds," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016.

Appendices

Appendix 1. US-290 Dataset

US-290 consists of several pavement marking types which could provide some key insights towards developing a reference vision system. To add more variables into consideration the video data were collected early morning driving eastwards and around noon driving westwards. Driving east during the morning hours resulted in the sun being directly over the horizon adding glare on the camera sensor. This glare effect is absent during noon when the road is illuminated from overhead. Evaluation of these conditions is expected to give some key insights into the effect of glare on the lane detection (LD) performance in addition to the effects of the different pavement marking patterns.

Some of the unique pavement marking patterns encountered on US290 include:

1. White pavement marking followed by black (WB) [Figure 18, Figure 19]
2. White pavement markings with a black border (BWB) [Figure 20, Figure 21]
3. Combination of 4-inch and 6-inch wide white markings (White) [Figure 22]

Experimental Conditions:

Date of data collection: 2020/04/16

Time of day: 8:30 am to 12:30 pm

Camera Configuration used:

Sensor: BFS-U3-51S5C-C (Sony IMX250 CMOS - 5MP – USB3.1 camera)

Lens: Kowa LM8HC Manual Iris C-Mount Lens f=8mm/F1.4

Combined video clip length:

Driving Eastwards: 15min 45sec

Driving Westwards: 14min 24sec

We collected the data on these roads and present it as an annotated dataset for researchers to study the LD characteristics of these unique roads. The dataset was evaluated for LD performance on SCNN [9] lane detection algorithm and the results are presented in Table 13.



Figure 18: Image of a road segment with white pavement marking followed by black (driving westwards).



Figure 19: Image of a road segment with white pavement marking followed by black (driving eastwards).



Figure 20: Image of a road segment with white pavement markings with a black border (driving eastwards).



Figure 21: Image of a road segment with white pavement markings with a black border (driving westwards).



Figure 22: Image of a road segment with white pavement markings, and significant sun glare (driving eastwards).

Table 13: LD Performance of Different Pavement Marking on US 290

Marking	Driving Direction	Precision (P)	Recall (R)	F1 - Measure
BWB	East	0.56824	0.63833	0.601249226
BWB	West	0.59065	0.648249	0.618110551
WB	East	0.576458	0.785637	0.6649855310
WB	West	0.849475	0.917485	0.882171153
White	West	0.506709	0.521368	0.513933991

Appendix 2. RELLIS Closed Course Evaluation

To check the acceptability of a closed course (Runway 35L) to conduct tests on 3M panels, lane marking data were collected for the existing lane markings. Runway 35L consists of 4 sets of paint-based pavement markings and 2 sets of tape-based pavement markings. Video data were collected by driving northbound (NB) and southbound (SB) on a Lincoln MKZ vehicle fitted with the same camera equipment as discussed in the video data section. The video data were evaluated for LD performance on SCNN [9] and LaneNet [15] lane detection algorithm. Lane marking material characteristic data were collected for these pavement markings around the same time as the video data collection. Table 14 tabulates the lane marking material characteristic data and lane detection (LD) performance data in terms of the Precision (P), Recall (R), and F1 Measure (F1) on SCNN and LaneNet. Figure 23 shows the plot of F1 scores vs Q_d values for the pavement materials on RELLIS 35L. There were limited interactions from the Ghost Markings, Concrete block edges, or faded lane markings that affected the measurement and LD performance on 35L. Therefore, the runway area was deemed acceptable to study LD performance of the 3M panels.

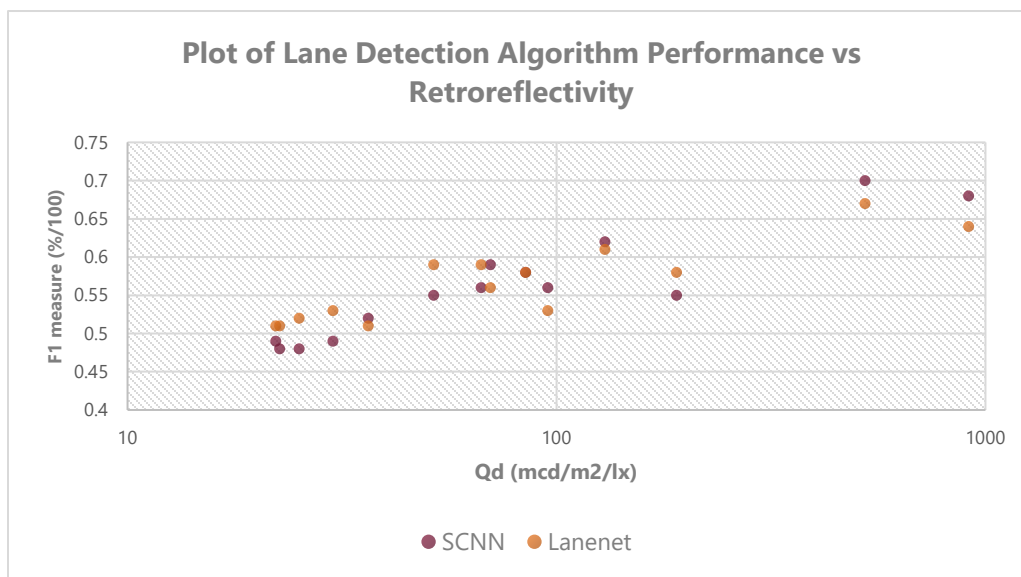


Figure 23: Plot of lane detection performance of SCNN & LaneNet algorithms (measured as F1 score) vs continuous R_L data for pavement markings on RELLIS 35L.

Table 14: Material Characteristics of Lane Markings on RELLIS 35 L

Section			Color	Width (in)	Qd (mcd/m²/lx)	x	y	Y	SCNN			LaneNet		
									Precision	Recall	F1 Score	Precision	Recall	F1 Score
S1	NB	L	Yellow	4	523.89	0.32	0.34	51.954	0.70	0.63	0.67	0.74	0.61	0.67
		R	White	4	1000.76	0.40	0.38	37.845						
	SB	L	White	4	913.52	0.32	0.34	51.954	0.68	0.59	0.63	0.66	0.63	0.64
		R	Yellow	4	470.37	0.40	0.38	37.845						
S2	NB	L	Yellow	4	190.47	0.32	0.33	22.892	0.55	0.59	0.57	0.57	0.59	0.58
		R	White	6	65	0.37	0.37	30.161						
	SB	L	White	6	66.72	0.32	0.33	22.892	0.56	0.61	0.59	0.59	0.6	0.59
		R	Yellow	4	202.05	0.38	0.37	30.161						
S3	NB	L	Yellow	4	95.43	0.32	0.34	45.234	0.56	0.58	0.57	0.55	0.51	0.53
		R	White	4	133.85	0.38	0.37	35.631						
	SB	L	White	4	129.67	0.32	0.34	45.234	0.62	0.59	0.60	0.68	0.56	0.61
		R	Yellow	4	102.14	0.38	0.37	35.632						
S4	NB	L	Yellow	6	22.62	0.32	0.33	44.494	0.48	0.53	0.50	0.52	0.51	0.51
		R	White	6	54.809	0.32	0.34	10.966						
	SB	L	White	6	51.714	0.32	0.33	44.493	0.55	0.51	0.53	0.60	0.58	0.59
		R	Yellow	6	21.94	0.32	0.34	10.966						
S5	NB	L	Yellow	4	36.4	0.34	0.35	28.821	0.52	0.58	0.55	0.51	0.52	0.51
		R	White	4	28.88	0.35	0.36	26.900						
	SB	L	White	4	30.12	0.34	0.35	28.821	0.49	0.52	0.50	0.56	0.51	0.53
		R	Yellow	4	33.48	0.35	0.36	26.900						
S6	NB	L	Yellow	4	84.72	0.33	0.35	27.137	0.58	0.6	0.59	0.58	0.59	0.58
		R	White	4	66.08	0.35	0.36	29.950						
	SB	L	White	4	70.12	0.33	0.35	27.137	0.59	0.55	0.57	0.62	0.52	0.56
		R	Yellow	4	67.08	0.35	0.36	29.950						
S7	NB	L	White	4	25.12	0.36	0.36	22.222	0.48	0.56	0.52	0.49	0.56	0.52
		R	White	4	22.76	0.36	0.36	21.681						
	SB	L	White	4	22.16	0.36	0.36	22.222	0.49	0.52	0.50	0.48	0.55	0.51
		R	White	4	24.96	0.36	0.36	21.681						

Appendix 3. Additional Tables and Plots from 3M Panel Dataset Evaluation

Table 15: Tukey HSD Test ($\alpha = 0.050$) of LS Mean Differences for Two-way Interaction Between Panel Spacing and Panel Material During Day (Top) and Night (Bottom). (Levels not Connected by the Same Letter are Significantly Different).

Level (Day)						F1 Score (LS Mean)
30ft, 06W-4	A					0.76244328
30ft, 02W-4	A	B				0.74676662
30ft, 08W-4	A	B	C			0.71583223
30ft, 01W-4	A	B	C			0.71410252
30ft, 06W-6	A	B	C			0.70024279
40ft, 08W-4	A	B	C			0.69898629
30ft, 02W-6	A	B	C			0.69446894
40ft, 01W-4		B	C	D		0.68943787
40ft, 02W-4			C	D	E	0.65756167
40ft, 06W-4			C	D	E	0.65509721
40ft, 02W-6				D	E	0.62354169
40ft, 06W-6					E	0.60675365

Level (Night)								F1 Score (LS Mean)
30ft, 06W-4	A							0.90463257
30ft, 02W-4	A	B						0.89601716
30ft, 06W-6	A	B						0.88009106
30ft, 02W-6	A	B	C					0.86376944
30ft, 08W-4		B	C					0.85432671
30ft, 01W-4			C	D				0.81953346
40ft, 08W-4				D	E			0.79313570
40ft, 01W-4					E	F		0.75166342
40ft, 06W-6						F	G	0.73171254
40ft, 02W-6							G	0.70439418
40ft, 06W-4							G	0.69277848
40ft, n 02W-4							H	0.67923865

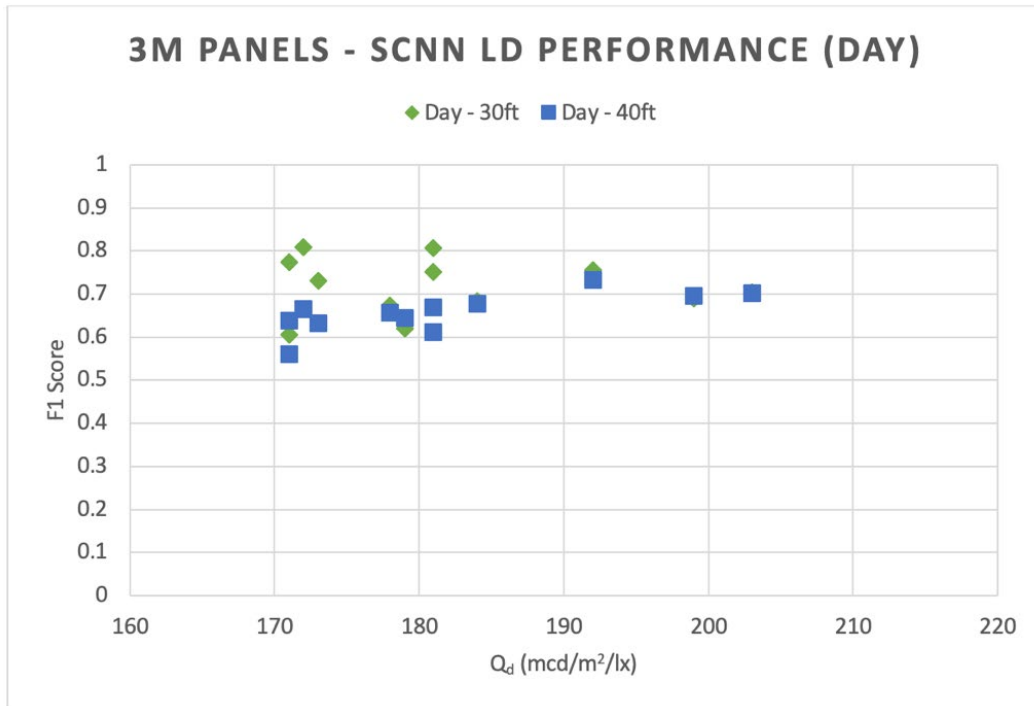


Figure 24: Plot of LD performance vs diffused reflectance (Q_d) for the 3M pavement panels.

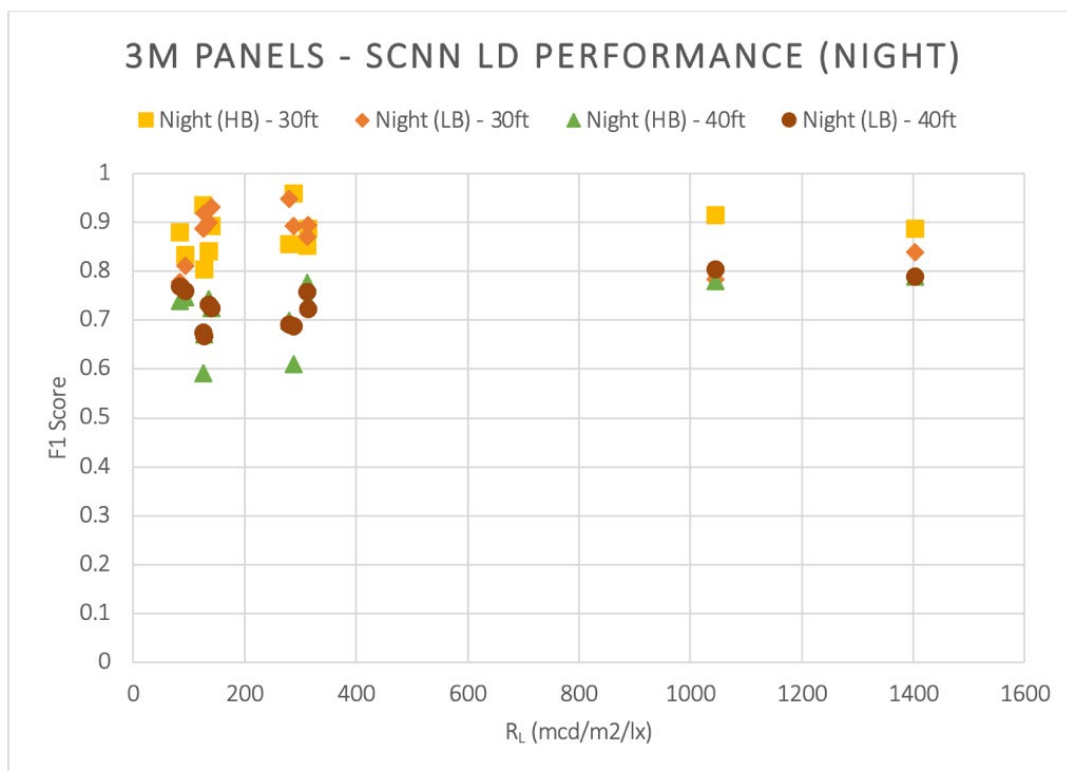


Figure 25: Plot of LD performance vs retroreflectivity (R_L) for the 3M pavement panels.

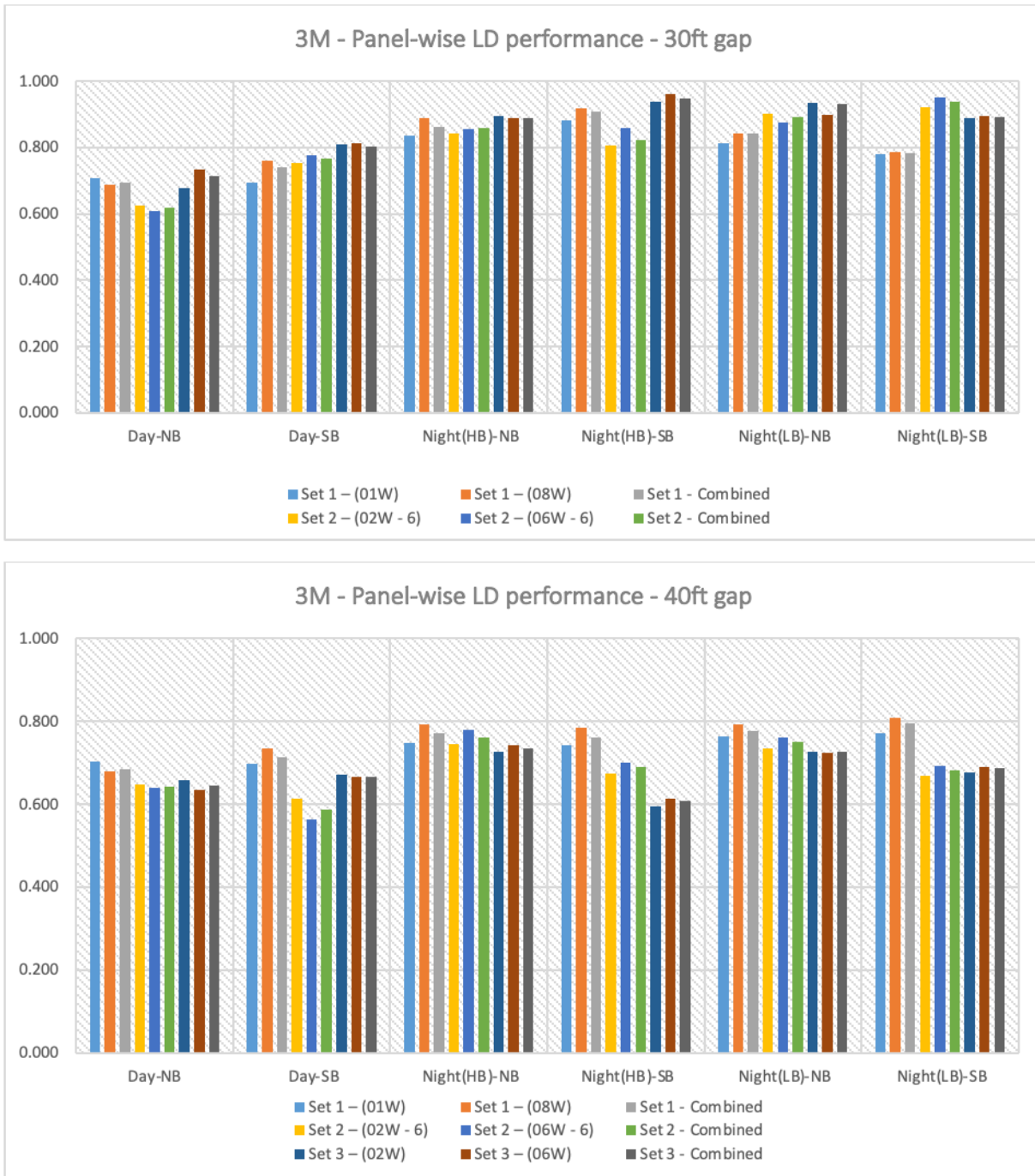


Figure 26: Panel-wise LD performance summary for 3M panels.

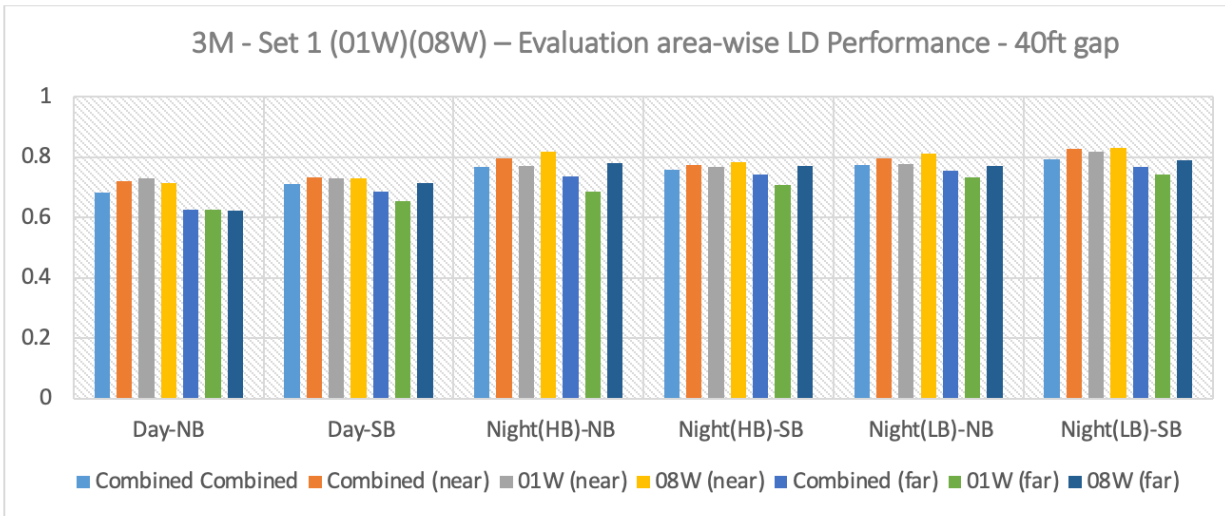
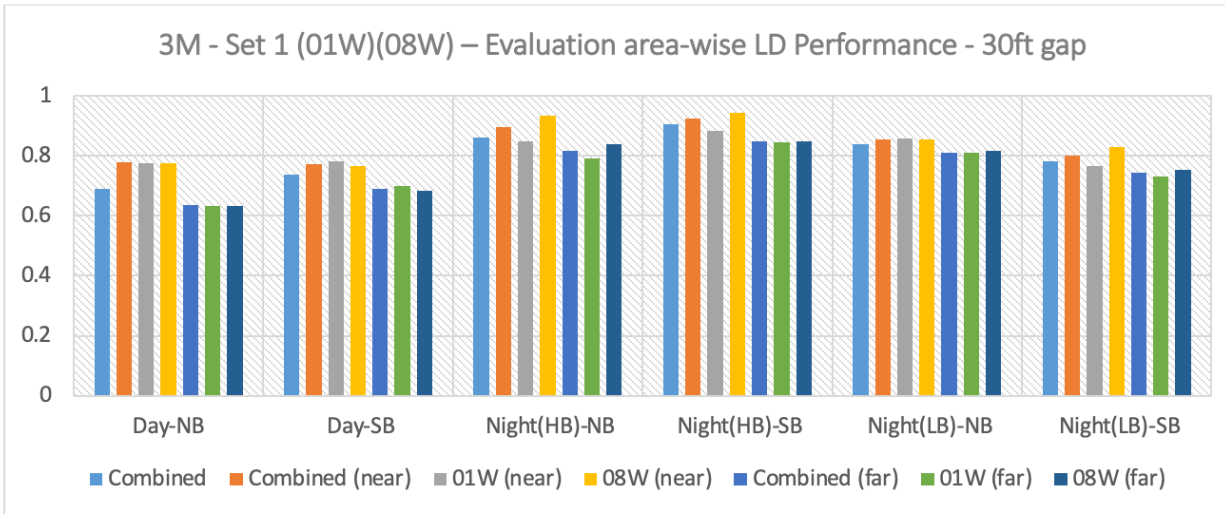


Figure 27: LD performance summary for Set 1 of 3M panels based on evaluation area.

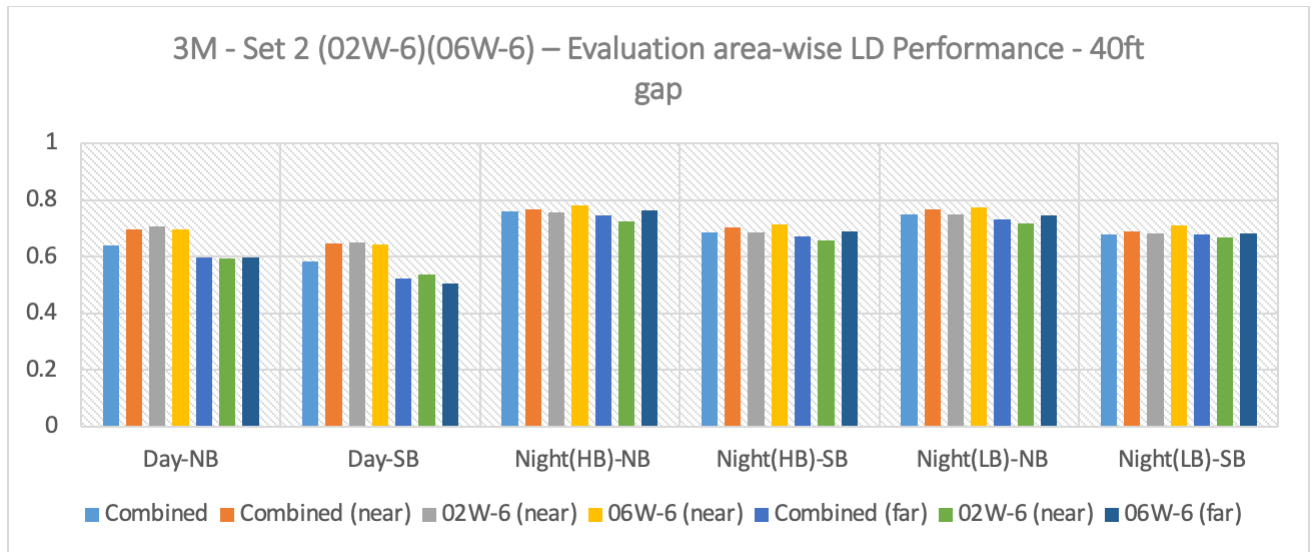
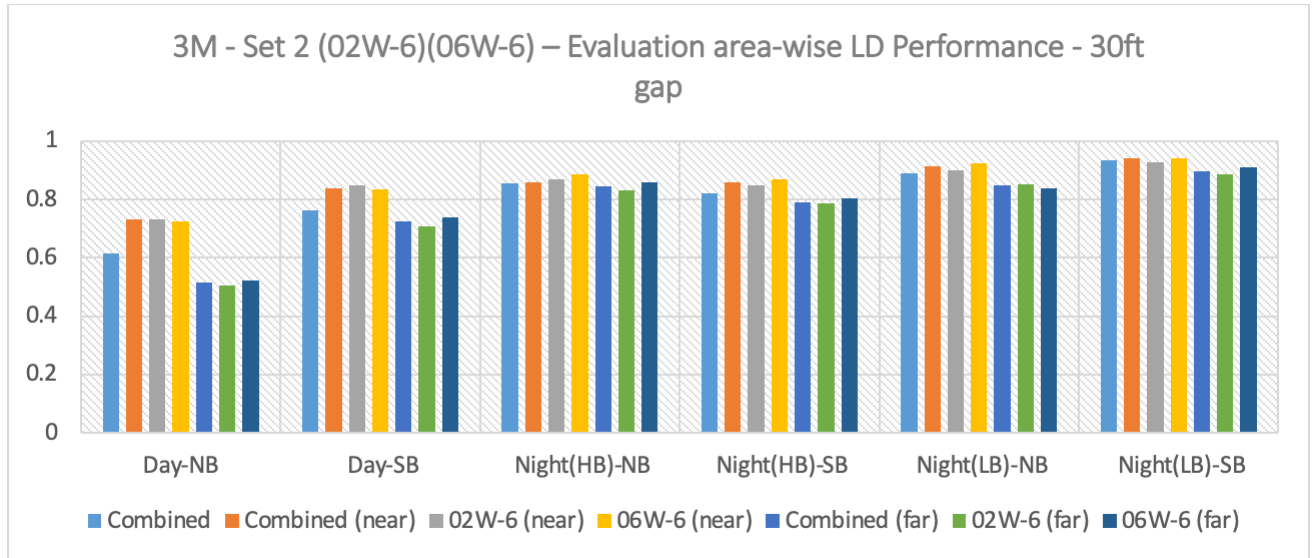


Figure 28: LD performance summary for Set 2 of 3M panels based on evaluation area.

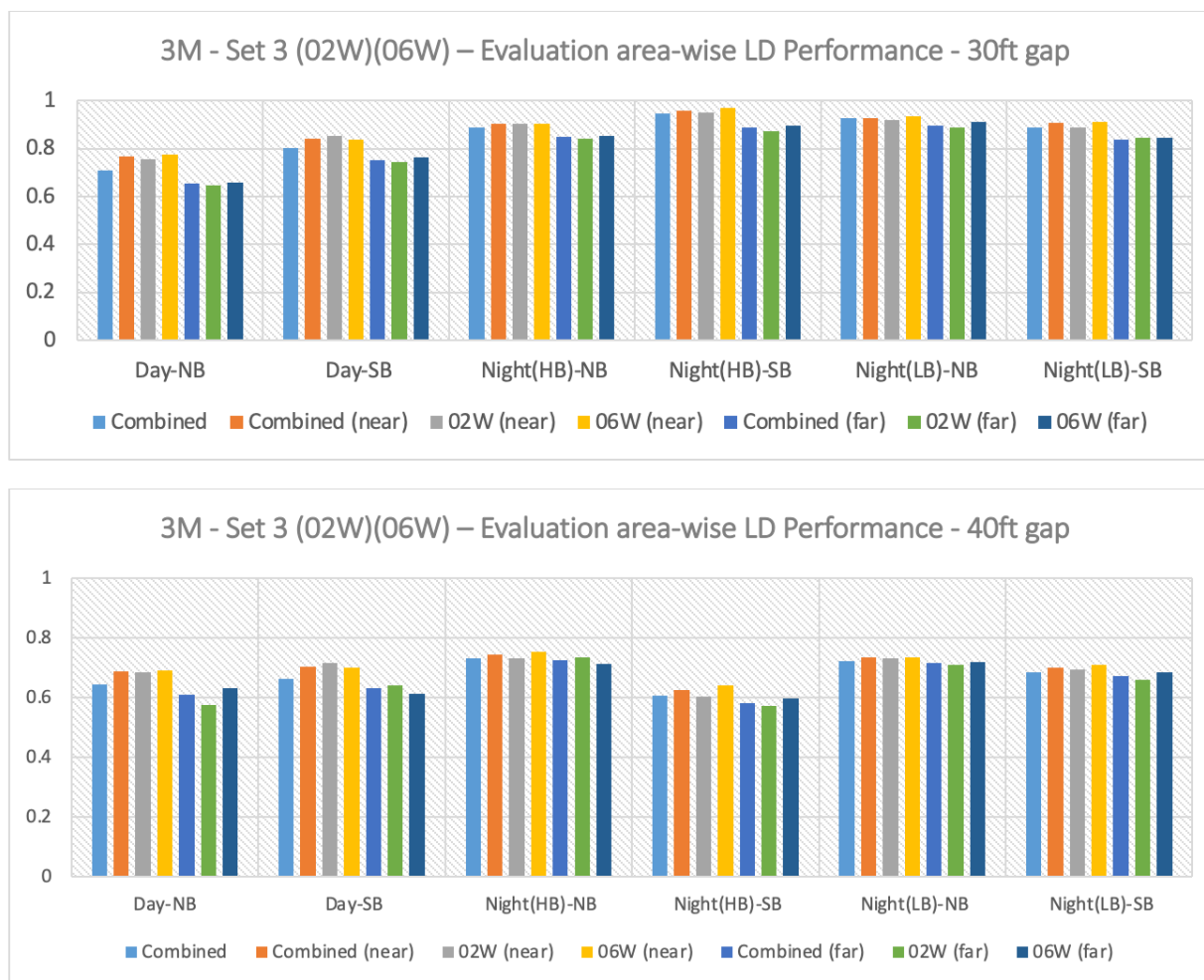


Figure 29: LD performance summary for Set 3 of 3M panels based on evaluation area.

Table 16: Tukey HSD Test ($\alpha = 0.050$) of LSMean Differences for 2-way Interaction Between Panel Spacing and Driving Direction During Day (Left) and Night (Right) – Levels Not Connected by the Same Letter Are Significantly Different

Level (Day)			Least Sq Mean
30ft, Southbound (SB)	A		0.76921770
30ft, Northbound (NB)		B	0.67540110
40ft, Northbound (NB)		B	0.65658527
40ft, Southbound (SB)		B	0.65387419

Level (Night)			Least Sq Mean
30ft, Southbound (SB)	A		0.86978169
30ft, Northbound (NB)	A		0.86967511
40ft, Northbound (NB)		B	0.74984608
40ft, Southbound (SB)		C	0.70112824

Table 17: LSMeans Differences Tukey HSD test ($\alpha = 0.050$) for Two-way Interaction Between Panel Spacing and Nighttime-Illumination – Levels Not Connected by the Same Letter Are Significantly Different

Level			F1 Score (LS Mean)	Std Error
30 ft - High Beam (HB)	A		0.87284047	0.00552214
30 ft - Low Beam (LB)	A		0.86661633	0.00552214
40 ft - Low Beam (LB)		B	0.73405108	0.00552214
40 ft - High Beam (HB)		B	0.71692324	0.00552214

Table 18: Tukey HSD Test ($\alpha = 0.050$) of LSMeans Differences for Two-way Interaction Between Nighttime Illumination and Panel Material – Levels Not Connected by the Same Letter Are Significantly Different.

Level				Least Sq Mean
High Beam (HB), 08W-4	A			0.84136985
Low Beam (LB), 06W-6	A	B		0.81599750
Low Beam (LB), 08W-4	A	B	C	0.80609255
Low Beam (LB), 06W-4	A	B	C	0.80411763
Low Beam (LB), 02W-6	A	B	C	0.79983597
Low Beam (LB), 02W-4		B	C	0.79613115
High Beam (HB), 06W-6		B	C	0.79580610
High Beam (HB), 06W-4		B	C	0.79329342
High Beam (HB), 01W-4		B	C	0.79136946
Low Beam (LB), 01W-4		B	C	0.77982743
High Beam (HB), 02W-4		B	C	0.77912466
High Beam (HB), 02W-6			C	0.76832765