AN EFFECTIVE FRAMEWORK OF AUTONOMOUS DRIVING BY SENSING ROAD/MOTION PROFILES

by

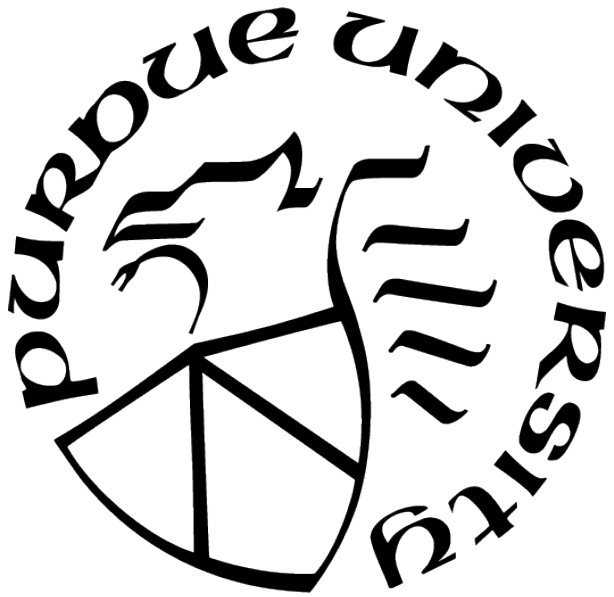
Zheyuan Wang

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Computer Science

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Jiang Yu Zheng, Chair

Department of Computer and Information Science

Dr. Voicu Popescu

Department of Computer Science

Dr. Xavier Tricoche

Department of Computer Science

Dr. Yuni Xia

Department of Computer and Information Science

Approved by:

Dr. Kihong Park

To my parents, Zhiqiang Wang, Yuxiang Li

ACKNOWLEDGMENTS

My greatest gratitude goes to my advisor, Dr. Jiang Yu Zheng, for guiding me towards one of the major achievements in my life. Dr. Zheng introduced me into the world of image video processing, intelligent vehicles, and advanced driver assistance systems. Dr. Zheng taught me how to think critically and independently about a research challenge, how to design tests to validate research hypotheses, and how to publish high-quality articles that effectively express research contributions. Without his commitment and zeal, this endeavor would be impossible to complete. His prompt and intellectual guidance, rigorous investigation, and scientific approach have all aided me in completing my assignment to a significant extent. They have been crucial during my whole PhD journey. Learning from him during this time has been an invaluable and unforgettable experience that will follow me the rest of my life. Thank you. Also, I am very thankful to Dr. Voicu Popescu, Dr. Xavier Tricoche and Dr. Yuni Xia for serving in my doctoral committee and offering me new perspectives in my study.

I feel really lucky to have met numerous friends during my six years in Indiana. This dissertation would be endless if I mentioned all of them here and I firmly believe that the time spent with them is among the most enriching experiences during this journey. I would like to thank all of my lab mates, specially Brook Cheng, for their valuable collaborations and assistance on my research, as well as for the great friendship. Dr. Chris Cross, who is also known as 'bad grandpa', coaches me not only in Jiujitsu, but also gives me help in my thesis.

Last, but not least, I would want to thank my father Zhiqiang Wang and mother Yuxiang Li for their unwavering support. When I have any problem or need some advisory, my parents are always there and willing to help, even though sometimes the result is the same, their consolation puts me at ease. Their unconditional encouragement to pursue and accomplish all of the life goals I've set has been crucial in my writing my dissertation. Additionally, I want to express my gratitude to Mu Wei, a friend and advisor. Without Wei's assistance and advice, I might not be able to pursue a Ph.D. or even study abroad.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

With more and more videos taken from dash cams on thousands of cars, retrieving these videos and searching for important information is a daunting task. The purpose of this work is to mine some key road and vehicle motion attributes in a large-scale driving video data set for traffic analysis, sensing algorithm development and autonomous driving test benchmarks. Current sensing and control of autonomous cars based on full-view identification makes it difficult to maintain a high-frequency with a fast-moving vehicle, since computation is increasingly used to cope with driving environment changes.

A big challenge in video data mining is how to deal with huge amounts of data. We use a compact representation called the road profile system to visualize the road environment in long 2D images. It reduces the data from each frame of image to one line, thereby compressing the video clip to the image. This data dimensionality reduction method has several advantages: First, the data size is greatly compressed. The data is compressed from a video to an image, and each frame in the video is compressed into a line. The data size is compressed hundreds of times. While the size and dimensionality of the data has been compressed greatly, the useful information in the driving video is still completely preserved, and motion information is even better represent more intuitively. Because of the data and dimensionality reduction, the identification algorithm computational efficiency is higher than the full-view identification method, and it makes the real-time identification on road is possible. Second, the data is easier to be visualized, because the data is reduced in dimensionality, and the three-dimensional video data is compressed into two-dimensional data, which is more conducive to the visualization and mutual comparison of the data. Third, continuously changing attributes are easier to show and be captured. Due to the more convenient visualization of two-dimensional data, the position, color and size of the same object within a few frames will be easier to compare and capture. At the same time, in many cases, the trouble caused by tracking and matching can be eliminated. Based on the road profile system, there are three tasks in autonomous driving are achieved using the road profile images.

The first application is road edge detection under different weather and appearance for road following in autonomous driving, based on the road profile image and linearity profile image in the road profile system. This work uses naturalistic driving video data mining to study the appearance of roads, which covers large-scale road data and changes. This work excavated a large number of naturalistic driving video sets to sample the light-sensitive area for color feature distribution. The effective road contour image is extracted from the long-time driving video, thereby greatly reducing the amount of video data. Then, the weather and lighting type can be identified. For each weather and lighting condition, select obvious features at the edge of the road to distinguish the edge of the road.

The second application is detecting vehicle interactions in driving videos via motion profile images, based on the motion profile image in the road profile system. This work uses visual actions recorded in driving videos taken by a dashboard camera to identify this interaction. The motion profile images of the video are filtered at key locations, thereby reducing the complexity of object detection, depth sensing, target tracking and motion estimation. The purpose is for decision making of vehicle actions such as lane changing, vehicle following, and cut-in handling.

The third application is motion planning based on vehicle interactions and driving video. Taking use of the fact that a car travels in a straight line, we simply identify a few sample lines in the view to constantly scan the road, vehicles, and environment, generating a portion of the entire video data. Without using redundant data processing, we performed semantic segmentation to streaming road profile images. We plan the vehicle's path/motion using the smallest data set possible that contains all necessary information for driving.

The results are obtained efficiently, and the accuracy is acceptable. The results can be used for driving video mining, traffic analysis, driver behavior understanding, etc.

1. INTRODUCTION

1.1 Background in Autonomous Driving And ADAS Research

In last decade, the automobile industry and information science have fused to advance autonomous driving in manufacturing areas. As one of the major areas of growth for AI, autonomous driving will ensure safety, reduce the workload, and boost the economy. The United States is a country that relies more on automobiles than of the any countries in Europe or Asia— where a large population uses public transportation like trains and subways. Replacing drivers with autonomous vehicles will not only have scientific significance in exploring human intelligence, but also create a workforce in many related industries such as automobile, retail, civil engineering, information systems, communication, manufacturing, travel, etc. The Society of Automotive Engineers (SAE) created a standard industry scale from zero to five to depict this continuum [1], but there are several gray regions where characteristics may overlap as shown in Figure 1.1. For self-driving automobiles, the most important tasks are autonomous lateral and longitudinal control to drive the vehicle in lane and avoid collision with other vehicles and road users. A comprehensive software stack for autonomous driving appears to be highly promising with an end-to-end concept. The automobile industry's current trend, along with active research by key technology firms, has shown that self-driving vehicles are the future.

The core of the self-driving system, consisting of three components: perception, planning, and control. Self-driving system technology entails the following: a self-driving car can recognize its surrounding environment and vehicle status using a variety of on-board sensors [2], [3](camera, Lidar, radar, GPS, inertial sensors, etc.), perform analysis and judgment, then control vehicle movement autonomously, and finally achieve self-driving based on the acquired environmental information (including road information, traffic information, vehicle location and Obstacle information, etc.) Given the usefulness of on-board sensors, this article provides techniques for self-driving systems' perception, planning, and control.

Autonomous vehicles have been equipped with many computer vision components and engineering devices such as driving video, stereo 3D camera[4], LiDAR depth measure[5], road detection[6], object detection, target tracking, lane marker awareness[7], traffic signal

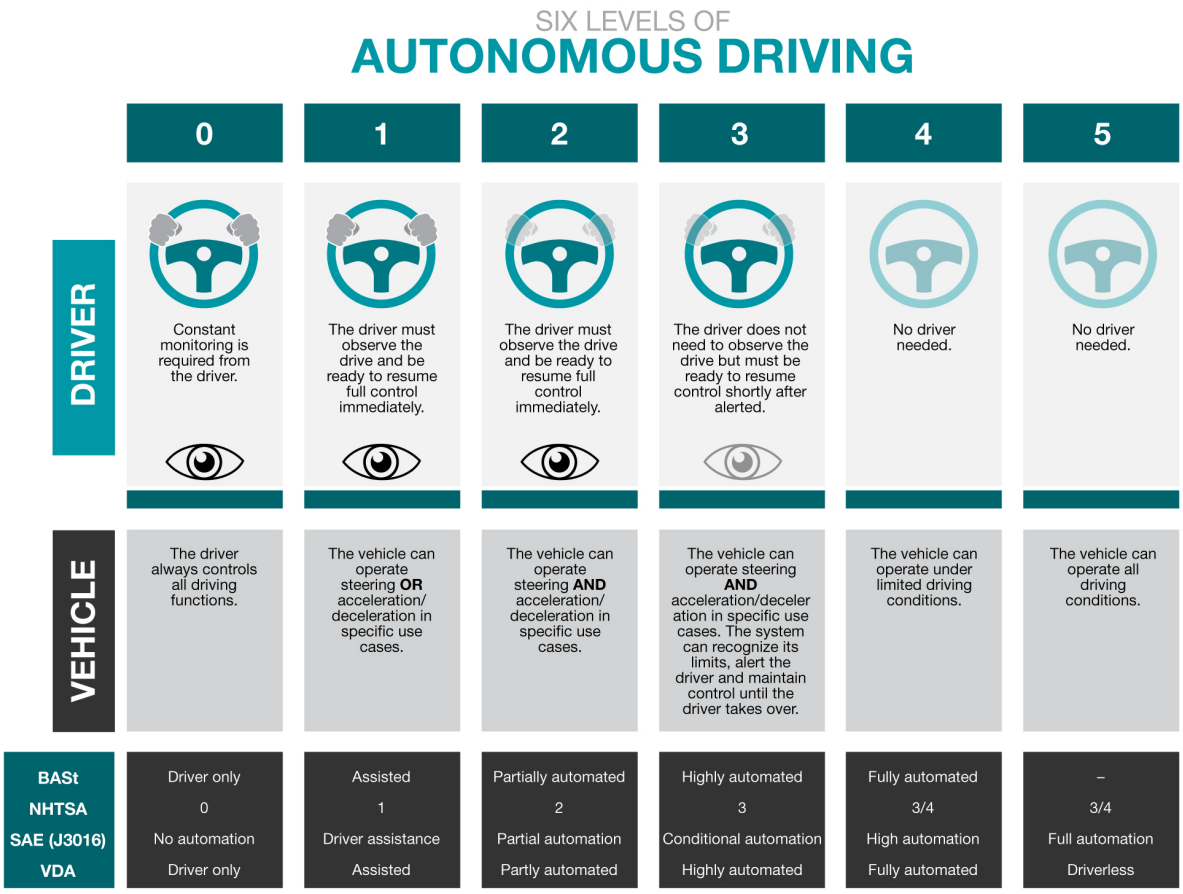


Figure 1.1. SAE International recently unveiled a new visual chart (below, and in gallery) that is designed to clarify and simplify its J3016 "Levels of Driving Automation" standard for consumers. The J3016 standard defines six levels of driving automation, from SAE Level Zero (no automation) to SAE Level 5 (full vehicle autonomy). It serves as the industry's most-cited reference for automated-vehicle (AV) capabilities.

recognition[8], wide area awareness, vulnerable road user detection, vehicle path planning, HD road mapping, GPS localization, traffic flow broadcasting, etc. However, many of these technologies cannot guarantee 100% accuracy in their output. Because of the large variations in road, traffic, and environmental conditions as well as traffic laws in various regions, autonomous vehicles are complex systems that prove difficult for scientists and engineers to model explicitly and precisely. Any failure in this decision-making chain can break down the driving process and cause an accident. In addition, the delay of each component using so-

phisticated algorithms and communications can make vehicle control less frequent and more fragile because the environment is highly varied when a vehicle moves at high speeds.

Successful technologies and breakthrough functions for autonomous vehicles include lane tracking functions to keep a vehicle in the proper lane on the road and radar to alert fast-approaching vehicles (LEVEL2 autonomous driving). Short-distance radar [9] or sonar in sideways directions prevent dangerous lane changes, and LiDAR [10], [11] keeps track of distance to static objects in all directions (a Tesla car occasionally hits static objects because it does not have the depth measurement capabilities of LiDAR). Researchers and developers have made efforts to raise autonomous driving to LEVEL 3. With the monitoring of a human driver, the vehicle can handle longitudinal speed and latitudinal steering on most roads and traffic under normal weather and illumination conditions. However, this is merely the first level of a driving school curriculum and engineers are developing and experimenting more advanced functions such as parking, interacting with other vehicles, following traffic rules, signs and flow, handling adverse weather and illumination, etc. On the other hand, higher levels of fully autonomous driving, e.g., LEVEL 4 and 5, are being tested in special areas for standard events and circumstances. With the new generation of mobile communication (5G for example) and infrastructure improvement in civil engineering, autonomous and connect vehicles have been evolved to behave like a connectable train moving in some designated road environments.

As for human driving intelligence, we are not checking every object in our surroundings. Our human perception cannot measure distance and speed as precisely as those sensors. However, our ability to handle complex events is much stronger and more robust. A skillful driver can ignore some procedural observation but focus on some critical spots briefly in information acquisition, e.g., skip the observation of roadside scenes but employ peripheral perception to sense an approaching vehicle and avoid a collision. Additionally, from the expansion rate of object size, mature drivers can predict the time to collision (TTC) [12] without measuring exact 3D depth and vehicle velocity precisely in miles per hour. These abilities are enhanced by previous experiences as a driver as well as location/space memorization. The more the experience is accumulated, the more stable and reliable the driver (and vehicle) acts. Overall, autonomous driving is to teach a machine the skills needed to pass a

route at a safe position and a proper speed with large variation and uncertainty regarding the road, traffic, environment, sensor, and system conditions.

1.2 Learning from Data in Naturalistic Driving for Algorithm Development

Recently, autonomous driving study has taken a data orientated method based on naturalistic driving data. There are significant scene changes in the gathered naturalistic driving footage owing to variances in roads, cars, traffic, pedestrian interactions, weather, and illumination, among other factors. For example, road illumination varies significantly depending on the weather (Fig.1.2), and no distinct lane markings are visible at junctions or on worn roads, but many lane markers make determining the proper way difficult. While certain significant and easy road and traffic scenarios have been detailed in detail, many additional cases have not been fully modeled using depth measures, kinematics and dynamics, and rule systems. To enable autonomous cars to operate on public highways alongside human drivers, more real-world scenarios and combinations of scenarios must be considered. Machine learning techniques based on large data will summarize typical practices and scenarios to serve as the neural network's ground truth. The convolutional neural network (CNN) is far more effective in sorting and storing spatial information in the deep layered network than previous approaches used by human researchers. Multilayer neural nodes and their connections through links and coefficients further simulate the three-dimensional motions of nearby vehicles and their variation within the network. With suitably deep layers and node coefficients, network training on large-scale datasets may solve difficult classes.

1.3 Objectives of This Work

1.3.1 A temporally-dense sensing framework with the minimum data set

The purpose of the road profile system is to reduce the size and dimensionality of driving videos to detect road edges, vehicle interactions, and motion planning. The profile image is a two-dimensional spatial-temporal image with the time dimension on one axis and the spatial dimension in the video frame on the other, as an example shown in Fig. 1.3 [12]. The motion profile images are generated from driving video according to the optical flow

Figure 1.2. Various road appearances in driving videos under different weather and illumination conditions. A brief categorization in terms of road type, off-road type, and weather/illumination is given by human annotation.

of scenes for detecting surrounding vehicles, self movement, other objects, and road profile images are used to detect road edge.

Spatial-temporal profile image on driving video has first proposed by Mehmet et al for predict vehicle collision and pedestrian detection [12], [13]. Mehmet's work greatly inspired my research about the way to use NDV. Instead of the vertical scanning line, this work used horizontal line in the frames which provide road and vehicle interaction information in the profile images.

Road edge detection is based on road profile image in variable weather as briefly shown in Fig. 1.4. Different from other road edge detection method [14]–[17], the approach in this paper is using road profile images to detect road edge, which require less data and

(a) Road profile image (b) Motion profile image

Figure 1.3. Spatial-temporal road profile image and motion profile image from the same video clip of 5min. They capture the road layout and motion of other vehicles at a far range, respectively. The time axes are upward and the temporal resolution is 9k frames at the sampling rate of 30 fps. The horizontal axes are the same as the image x axis in the video frame.

computation power. This is the first study to propose for the road detection based on weather and illumination.

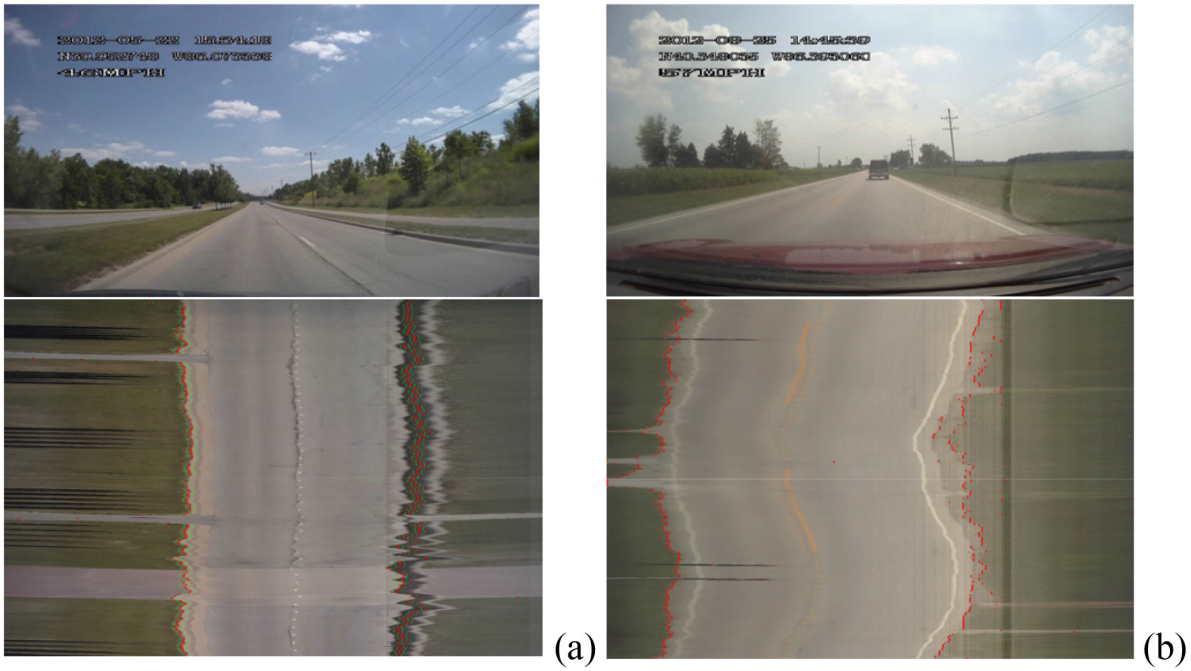


Figure 1.4. Road edge detection results at a middle range. Upper: video frames. Lower: sections of road profile images obtained from middle range and detected road edges marked in red dots.

1.3.2 Direct use of motion without heavy object recognition

Vehicle interaction detection is based on a motion profile image. This part proposed a novel approach for capturing vehicle interactions in driving footage for the purpose of analyzing driving behavior. Our method detects the direction of motion of cars in concentrated zones in driving films without relying on identification, tracking, or classification techniques [18]–[22]. We reduce the complexity of the problem by filtering the motion profile image, which is a compressed representation of driving video.

We propose a novel approach for capturing vehicle interactions in driving footage for the purpose of analyzing driving behavior. Our method detects the direction of motion of cars in concentrated zones in driving films without relying on identification, tracking, or

classification techniques. We reduce the complexity of the problem by filtering the motion profile image, which is a compressed representation of driving video.

1.3.3 Planning driving for fast moving vehicles

We implemented driving planning based on interaction detection and motion profile images. The task explores further details on route and motion planning in driving using this framework. We will provide a direct connection between temporal sensing and driving tasks, avoiding duplicate data and recognition. The road profile images are used to determine lateral path in lane and the detected headway distance adjust ego-vehicle speed. High-speed sensing and planning at video rate will result in smooth vehicle control and a rapid response to various events encountered when driving at high speeds.

1.4 Originality, Contribution and Impact

The three major contributions of this dissertation are the exploration, modeling, and implementation of following problem solving are as follow:

1. Generating road profile image, motion profile image, and other types of profiles with reduced data size and how the profiles still keep important information for further computation will facilitate modeling. These investigations establish a new temporally tense sensing framework of autonomous driving with much smaller data than normal driving videos. This can facilitate future autonomous driving devices on small machines.

2. Introducing a method of road edge detection robust in different weather conditions using road profile image will provide more information. The compact road profile images make data mining of a large naturalistic driving data set possible, which includes many critical scenarios of road, traffic, weather and illumination conditions. The developed road recognition algorithms are able to cope with the scene variations robustly for safety driving.

3. Introducing a method of vehicle interaction and motion planing using motion profile images. This avoids full range shape recognition and tracking in traditional computer vision methods in acquiring relative motion for driving. The complexity of sensing for autonomous

driving is also reduced significantly for the safe negotiation of autonomous vehicles on the road.

1.4.1 Compact data for real-time sensing and driving

Data indexing and retrieval are critical for further mining and learning with a huge dataset of driving video footage. The video's huge data size must be reduced in order to create a new motion-oriented architecture for vehicle interactions and motion planning. Consecutive video frames include overlapping view coverage for vehicle translation along a road. The processing of this duplicate data adds time to the vehicles' response to dynamic events.

Driving video profile images adopted in this work mainly includes two basic types, road profile image and motion profile image. Compared with the original driving video, both the road profile image and the motion profile image reduce the dimensionality of the data and greatly reduce the size of the data. The road profile image will record more detailed information, pixel-level color contrast, but at the same time it will be more easily affected by unstable factors during driving, road bumps, etc. Although the motion profile image does not have pixel-level color contrast, it can better record the motion trajectory of the object in the driving video, and is less susceptible to bumps. So these two types can play their respective roles in different scenarios.

The data used in this work is Natural Driving Video (NDV). These videos are used to generate road profile images and motion profile images. The resolution of each frame in the video is 1280 * 720 pixels in width and height, 30 frames per second, and the video length is about 5 minutes. Corresponding to each segment of NDV, the width of the road profile image and the motion profile image will be 1280 pixels, and the length will be close to 9000 pixels. The data reduction rate will be 2k/720, where k is the number of lines or zones selected for monitoring the road and traffic ahead. This will shorten the computation time drastically as well during the fast vehicle driving.

1.4.2 Motion orientated driving as human driver

To avoid collision, relative motion between ego-vehicle and targets have to be obtained to plan ego-vehicle action early, as the mechanical movement or stopping point of vehicles due to the vehicle inertia must be considered. The motion profile image directly provides target motion as traces in the spatial-temporal space. This avoids heavy load object detection and tracking process in video frames, and obtain the vehicle interaction from the position and motion of visual trajectories in the streaming input of a motion profile image.

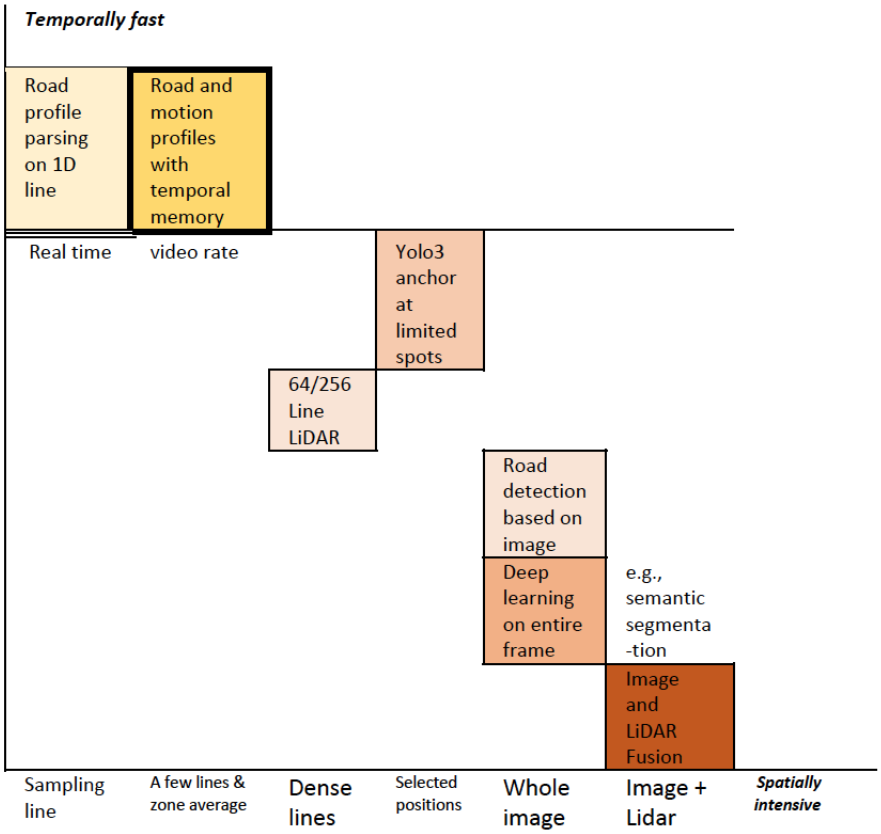


Figure 1.5. Comparison of spatial resolution and temporal efficiency of different methods in sensing and motion planning in terms of spatial resolution and response speed in sensing. The darker color intends to indicate a higher accuracy.

Although, object recognition in recent years has advanced significantly with deep learning, and numerous object identification techniques are capable of detecting the majority of objects in each frame of driving footage, they still need tracking of detected objects across frames in

order to generate motion information. For safety driving, the identify which types of vehicles are less important than the relative motion. The over-killing methodologies along with 3D data acquired by LiDAR may produce precise 3D models but are passed quickly.

A human driver, on the other hand, disregards the majority of this shape information, while extracting the motion information of surrounding objects from his/her peripheral vision. Such selectively captured motion is important and succinct in the subsequent phase of vehicle contact detection and motion planning. Because of this proposed scheme, our autonomous driving can respond to the real road environment with a small latency based on sufficient and effect data in the road/motion profile images. Figure 1.5 indicates the relation of our framework with other methods in the space of spatial resolution and sensing speed.

1.5 Related Works

1.5.1 Data mining of road environments in Naturalistic Driving Videos

Autonomous vehicle safety driving requires many vision tasks, such as road segmentation, lane mark detection, and vehicle recognition by frontal cameras. However, all these tasks can suffer due to drastic changes of weather and illumination. Data mining approach has been adopted in recent years to model the sensing and recognition system for autonomous driving. Large naturalistic driving video data sets such as KITTI [23] and Cityscape [24]are examples. TASI in IUPUI has also collected a driving dataset through one-year driving by 110 cars to cover various roads in the great Indianapolis Area [25]. To make vision a more robust function in driving, as it is for human drivers, this study models a spectrum of weather and illuminations visible in road environments. In weather study, a few papers have devoted to image weather recognition based on human tagged samples [26], [27]. Most of them provide a sky view rather than the road in driving views. Other works measure physical parameters and the results have not been directly used in road sensing [28]. For driving view, the study in has clustering illumination classes for a stable weather recognition (classification) at video frame.

We have implemented big-data mining on naturalistic driving videos through four seasons to understand the influence of weather and illumination [29], [30]. Weather sensitive

regions are sampled as image features to describe the illumination models qualitatively and quantitatively [31]. To understand how many distinct weather and illumination types exist for vision tasks, clustering is performed by unsupervised learning on all video samples [29], [32]. Typical views of a spectrum of weather and illumination conditions are generated using K-means clustering of feature distributions; we also find a stable number of clusters. The learned data is used to classify a driving view into one illumination type for guiding the road perception modules in autonomous driving. We further explored the sparse coding of vehicle views under various weather and illuminations [33].

1.5.2 Sensing and control in intelligent vehicles

Range and imagery sensors, such as cameras, are used in ADAS and automation to sense the vehicle's environment. The Radio Detection and Ranging (RADAR) sensor measures the range or velocity of target objects using radio waves. Radar is excellent at determining the target's distance and relative speed, which may be utilized to make direct judgments. Light Imaging, Detection, and Ranging (LiDAR) creates a three-dimensional picture of the surroundings by utilizing ultraviolet light. In recent years, autonomous driving has more applications of using LiDAR for depth in collision avoidance and partially road detection. It overcomes the problem of cameras output that is influenced by different weather and illumination conditions. However, LiDAR cannot provide lane mark, and is less accurate on road edges. In addition, it does not include motion information of dynamic vehicles; obtaining motion from LiDAR requires matching objects, which further involves object recognition and matching. The primary drawbacks of LiDAR is its high cost. Although the fusion of LiDAR depth and video images is being studied recently [34], there are no commercial products available on the market.

Cameras, on the other hand, are mostly utilized for object detection and motion estimation. Current automobiles can be fitted with in-car video cameras that enable the recording of driving scenes [35]. The cameras capture dynamic changes in road scenery, collisions, and the ego-motion of the car. The data has been extensively used in accident research, traffic pattern recognition, environmental surveying, driver behavior analysis, vehicle safety design, and object detection. Cameras have demonstrated a high degree of effectiveness in identi-

fying things including as cars, pedestrians, and lane markings. Motion is calculated using either optical flow between frames on monocular cameras or disparity pictures on stereo cameras. Each frame is analyzed individually in object identification, resulting in a significant computational cost. Only two consecutive frames are included in motion estimation, which leads to noisy results.

However, a critical issue is raised regarding the effective access and processing of large video volumes[36]. While individual frame processing is still possible for object detection and scene comprehension, it is computationally expensive, temporally sparse if some frames are skipped, and makes it difficult to visualize the continuous development of an event.

To avoid vehicle collision, Mehmet et al [37] calculated the instantaneous Time-to-Collision (TTC) for possible collisions using just motion data recorded by a vehicle-mounted camera. The contribution is the detection of risky occurrences and their severity immediately from motion divergence in a driving video, which is also a hint utilized by human drivers, without the need for previous vehicle recognition and depth measurement. In many collision-prone zones, both horizontal and vertical motion divergence are evaluated concurrently. Through filtering in the motion profile images, stable motion traces of linear feature components are produced. As a result, object identification and advanced depth sensing are omitted. The fine velocity computation produces a respectable TTC accuracy, which enables the video camera to avoid collisions only based on size changes in visual patterns.

Their technique computes TTC only on the basis of motion from a cluster of linear features, which is theoretically applicable to any backdrop and removes the need for complex vehicle scanning and recognition in the video. Selective zones for spatial temporal motion profiling have resulted in the detection of harmful collisions and an increase in computing efficiency for real-time processing. The approach is unique in that it relies entirely on motion, and the test was conducted on a variety of movies and situations. However, the TTC is a passive alarming but not an active path and speed planning. It can only used for alert in ADAS.

For dealing with the big data size and high data dimensionality problem of video, the motion profile image is also used for detecting potential collision objects including pedestrians as noted by Kilicarslan et al. [38]. In his work, motion profile image is used to show both

ego-motion and long term dynamic objects motion. The advantage of motion profile image comparing to video is that (1)much more compact data size than video, (2) continuous trajectories for exposing vehicle motion and activities, and (3) rapid item counting based on a global perspective of long term motion.

1.5.3 Machine learning applied to driving videos

In the driving videos, there are numerous scene variations due to the differences in roads, vehicles, traffic, pedestrian interactions, weather and illumination, etc. For example, no clear lane markers are visible at intersections or on worn roads and multiple lane markers make it difficult to determine a correct path. Current road and traffic scenarios have been described systematically in some major and straightforward circumstances, while many other cases have not been modeled explicitly according to depth measure, kinematics and dynamics, and rule systems. To make autonomous vehicles drive on normal roads with human drivers, more individual cases and the combinations of scenarios have to be taken into consideration from real-life circumstances. Machine learning based on big data will summarize common practices and scenes as ground truth in the neural network. The convolution neural network (CNN) can sort and store spatial information in the deep layered network in a much better way than traditional methods handcrafted by human researchers. Multilayer neural nodes and their connections with links and coefficients further model 3D movements of surrounding vehicles and their variation within the network. Network training using large-scale datasets can solve the complex classes with sufficiently deep layers and node coefficients.

Recent semantic segmentation uses deep learning framework. Road and roadside segments are detected mostly in normal weather such as sunny and cloudy. Structural and position information in the scenes are also considered through convolution, maximum pooling, etc. The results magically remove shadow and lane marks on road with training samples. Nevertheless, there is no evidence showing that the method is inherently good at diverse weather and illumination out of current training sets. The correctness of segmentation come from numerous sample annotation; new training must be performed if the method will be extended to diverse weather.

It is accomplished by detecting the presence of a road, a car, or a person in pictures [13], [39]. The spatial locations of distinct patterns are further monitored in driving footage to determine their mobility. They sample a single pixel line at each frame of driving footage, and the temporal confluence of lines from consecutive frames creates a picture of the road profile image. They demonstrate in this study how to learn the road profile image using Semantic Segmentation. Semantic Segmentation is used to efficiently comprehend both individual areas and their spatial relationships on the road by utilizing RGB-F pictures of the road profile image. They evaluated their technique using realistic driving video and found encouraging results.

1.6 Organization of This Thesis

The next chapter introduces the generation of road profile image and motion profile image and the reason why they are efficient. Chapter 3 introduces the method of road edge detection based on different weather condition using road profile image. Chapter 4 introduces the method of vehicle detection and interaction based on motion profile image. Based on the vehicle interactions, the motion planning and vehicle control decision making is described in Chapter 5. Chapter 6 outlines the experiment, gives evaluation results and followed by a conclusion in Chapter 7.

2. DRIVING VIDEO PROFILES AS COMPACT VISUAL DATA

2.1 Road Profile Image Generated from Driving Video

2.1.1 Line scanning on moving vehicle

A significant problem in video data mining is dealing with massive amounts of data. To display the road surroundings in a lengthy 2D picture, we utilize a concise representation called a road profile image. It condenses data from each frame (1280 * 720 pixels) to a single line, converting a video clip to a picture (data reduced to 1/720). We focus on a depth of 10-15m ahead of cars, which is the minimum distance necessary to avoid road departure incidents in the first or second following road edge detection in 30Hz video.

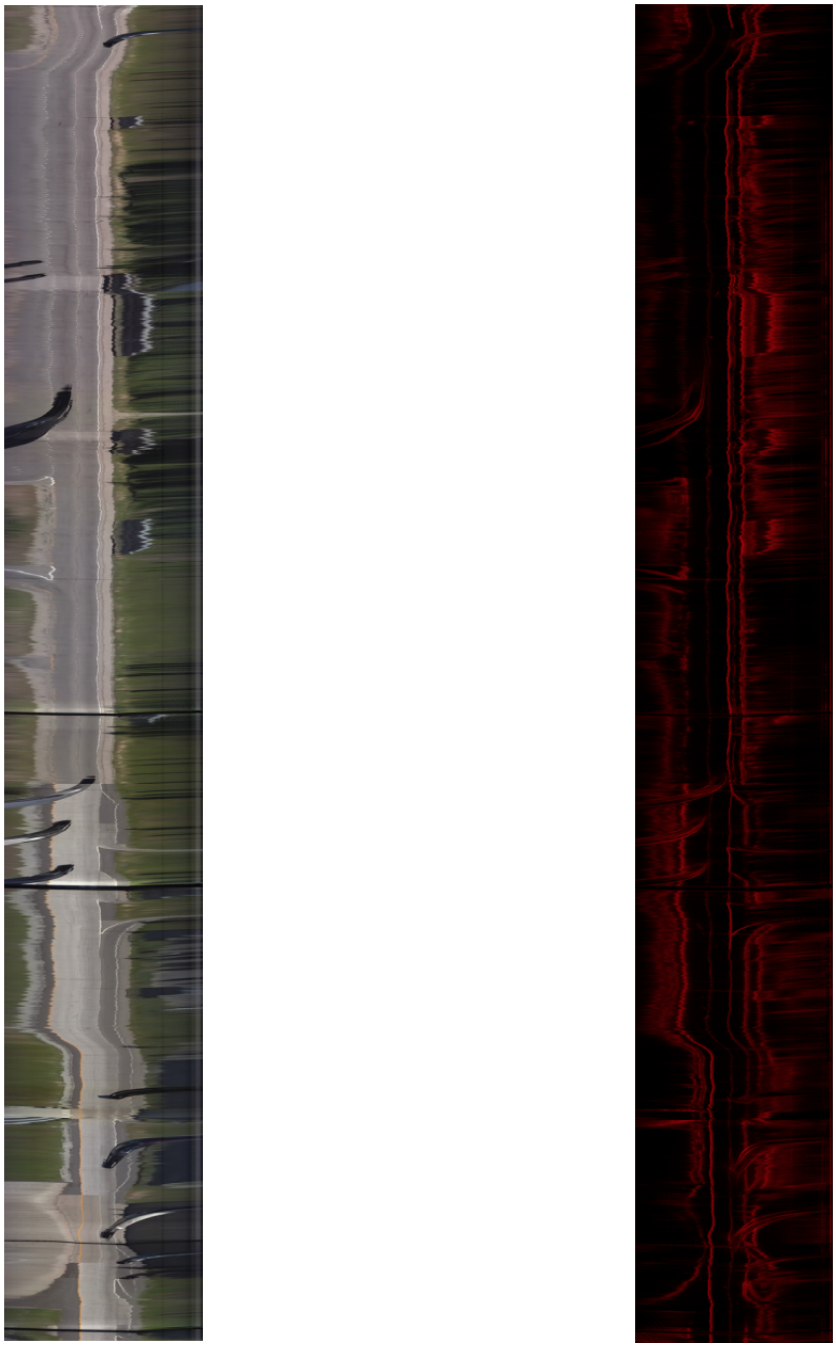
Naturalistic driving video is costly in storage, analysis, and test due to its large data volume. To facilitate data mining and efficient road detection, we convert video clips to temporal images called road profiles for data condensing. After the horizon is calibrated in HD video at a fixed height for a vehicle, we set a horizontal pixel line at h pixels below the horizon to sample all the video frames taken from that vehicle. The sampling line scans a road at about 10 15m ahead, as we consider a necessary time to respond to road departure. The line should not be set too far to capture other front vehicles. The sampled pixel lines from consecutive frames are connected in their frame numbers to form a temporal road profile, which shows road and off-road regions about four lanes wide. The road profile reduced video data to a small fraction (1/720-th of video) but contains major information on road and off-road scenes. A five-minute video generates 9,000pixel lines in the road profile image (video sampled at 30 fr/sec).

For each fixed camera pose on a vehicle, the horizon in the video frame is stable except on a rolling road. A horizontal sampling line, L, can be set on road to scan the road surface and its projection to the video frame, $I(x, y, t)$, is denoted as l, where t is the frame number and x, y are the image coordinates (Figure 2.1). It can cover the driving lane and a part of off-road scenes such as lawn, sidewalk, forests, field, etc. The color pixels on the line are sampled at consecutive t frames and copied to the profile image $P(x, t)$ as shown in Figure 2.2a. We sample the video at 30Hz and the resulting $P(x, t)$ has the height T as the video

Figure 2.1. A frame from driving video. The relevant image area around the sampling line is also located to obtain the spatial attributes of the road, such as the linearity of the road edge.

length in frame number. The sampling with the pixel line skips some interval on the road depending on the vehicle speed.

If a car keeps driving in a lane, the lane in the video and profile are approximately maintained at the same positions. Because the vehicle shakes in its fast motion mainly in roll and yaw, the lane marks and road edges are waved in the profile. This waving is more pronounced at the peripheral of the image and profile. Nevertheless, we still can observe the lanes, road surface, shadow, and the passing vehicles at the side lanes. The road profile covers an entire route. Even if the vehicle moves on a curved road, the road profile shows a straight lane in the image. This avoids complex geometry computation of road structure like road curvature, which is an important factor that causes road departure, and allows us to focus on color appearance study. In Figure 2.2, we can observe gray road surface and green lawn at side. Guardrails are visible from place to place in the green area. Some dark patterns initiated on the road surface are the traces of vehicles passing by. The dark horizontal regions spanning over the entire road are the shadow under bridges across the road, and such a dark region is followed by a bright region immediately due to the auto-exposure function of the camera when the vehicle moves out of the shadow area.



(a) (b)

Figure 2.2. The road profile image obtained from the sampling line 10-15 meters in front of the vehicle. (a) The resulting road outline is shown in the right column. (b) Linearity is also displayed in red intensity. The horizontal axis of (a) and (b) is x, and the vertical axis is the number of upward frames t.

2.1.2 Linearity computed from sampling belt in driving video

As most of other methods, we consider road shape shortly near sampling location 15 m ahead. The linearity at position x on the sampling line, denoted as $l(x)$, is an accumulated number of points on lines through x, roughly proportional to the length of line. To estimate the linearity, a belt of 30-pixel high is set around the sampling line. Within the belt, edge points are voted towards the sampling lines by using their gradient orientation.

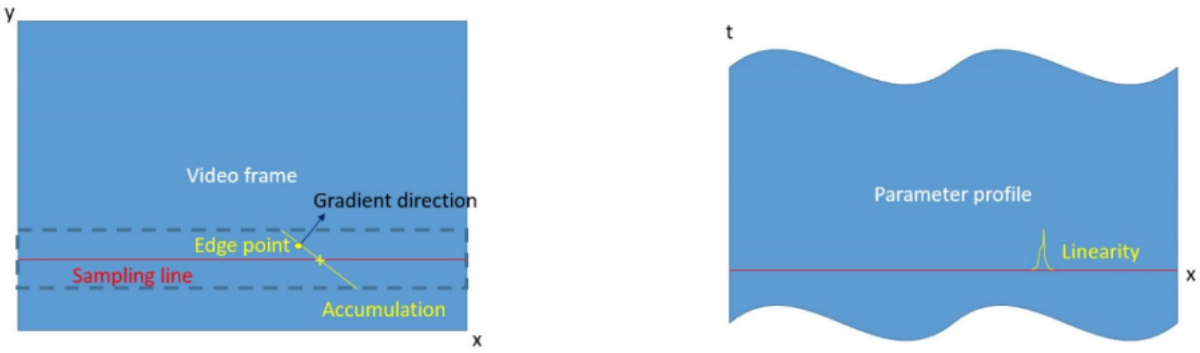
The linearity has strong response at a road edge, lane mark, curb, guide rails, lines on vehicles, as well as high-contrast texture on roadside. The linearity has low responses at local damage and noise spots on road surface because of their short lengths. We consider the road edges mainly as slanted line in video frames and exclude vertical lines in the frames that are mainly from surrounding vehicle rims, poles and buildings, and the reflection of objects and lights on wet roads. On a rainy day, the reflection of scenes and vehicle lights on wet surface affects road edge detection significantly.

In more details as in Figure 2.3, an edge point with its gradient stronger than a threshold is detected at (x, y) in the belt. The orientation $T(x, y)$ is from its gradient direction G_x/G_y, where G_x and G_y are the horizontal and vertical differential values. Extending the edge along $T(x, y)$ to the sampling line, the intersection is computed as k at

$$k = x + (y - h)G_y/G_x \text{ subject to } |G_y| + |G_x| > \delta \tag{2.1}$$

where h is the height of sampling line in the frame, and δ is a threshold for strong edges. We accumulate the vote at k by $l(k) = l(k) + 1$ for linearity. In addition, we store G_x and G_y as the homogeneity on road. Figure 2.4 displays the linearity in the road profile image. We have also found that tiny texture, e.g. roadside grass, has insufficient resolution for analysis. Linearity is incorrectly high in the high texture area due to its accumulation from high frequency points. Further, the linearity is classified as vertical and slanted lines in the view by voting at k_1 and k_2

$$l(k_1) = l(k_1) + 1, \text{ for vertical line if } |G_y|/|G_x| > \delta_1 \tag{2.2}$$



(a) (b)

Figure 2.3. Linearity estimation from an image frame to include shape information of road. (a) road edge and image belt in a frame, (b) stored linearity with the road profile image in the same coordinate system.

$$l(k_2) = l(k_2) + 1, \text{ for vertical line if } |G_y|/|G_x| > \delta_1 \tag{2.3}$$

where $l(k) = l(k1) + l(k2)$. Vertical linearity facilitates the separation of road edges from mirror reflection on wet ground.

Lane markings are frequently used on highways and are one of the most easily recognized and utilized forms of road border information. White and yellow are frequently employed as lane marking colors, which contrast significantly with the color of the road (usually gray). As a result, the linearity value will be very high, typically the highest line in the whole road profile, allowing for easy identification of the lane mark's location in the road profile.

2.2 Motion Profile Image Condensed from Driving Video for Vehicle Tracking

2.2.1 Motion Profile Image Generation

We implemented the scanning of road scenes from mid and closer range to avoid repetitive visual data computing over redundant areas. The scanning is at video rate such that we have a higher frequency sampling rate than those that rely on a spatial approach with individual video frames, and this difference can facilitate smooth vehicle control and earliest incident response. Moreover, horizontal belts are aligned at the image projection of horizon to obtain the motion trajectories of surrounding vehicles. Such motion profiles directly provide the relative motion without carrying out object recognition and tracking. The motion profiles are

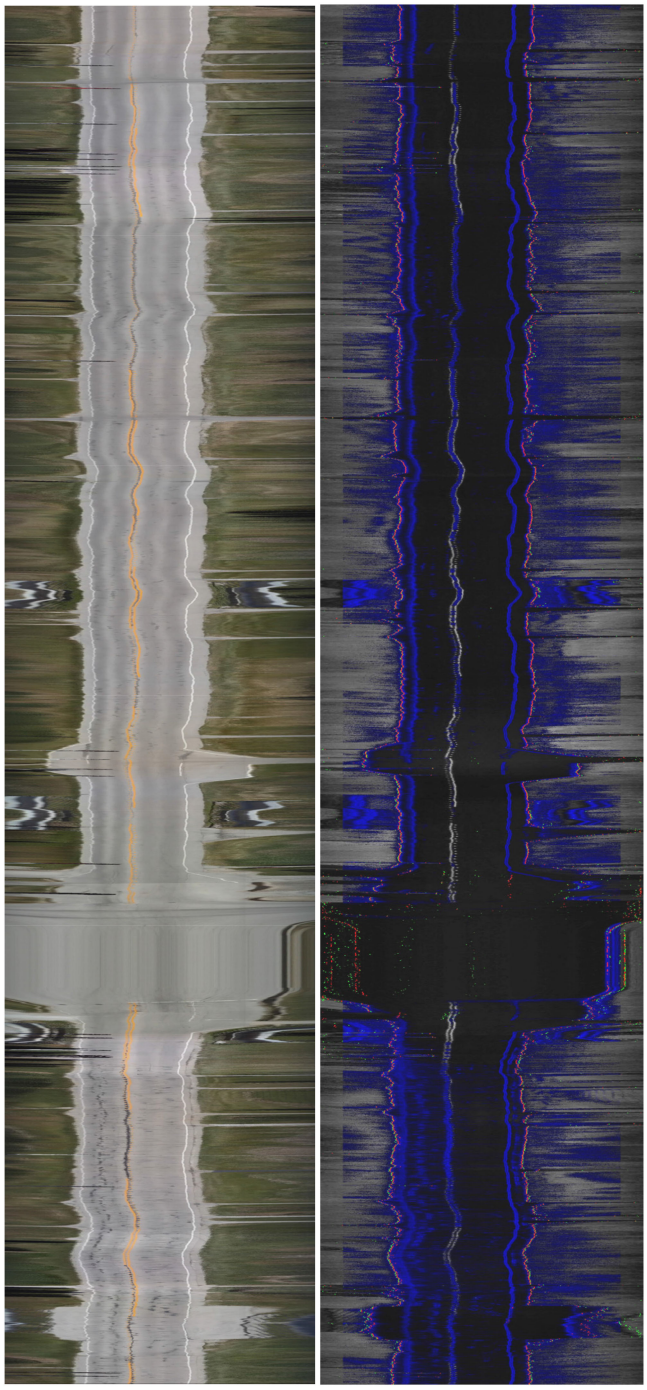


Figure 2.4. Five minute road profile image on a sunny day (left) and computed features (right) of chroma (in gray), overwriten by the mixture of linearity (in blue), the the highest chroma changes (green points), and the highest intensity changes (red points) in order on two sides of road.

one dimension lower than the video volume and are thus compact and efficient for learning and verification. In particular, the datasets can provide temporal information of dynamic surroundings and identify interactions with other vehicles on the road such as cutting in, changing lanes, vehicle following, merging, braking, etc.

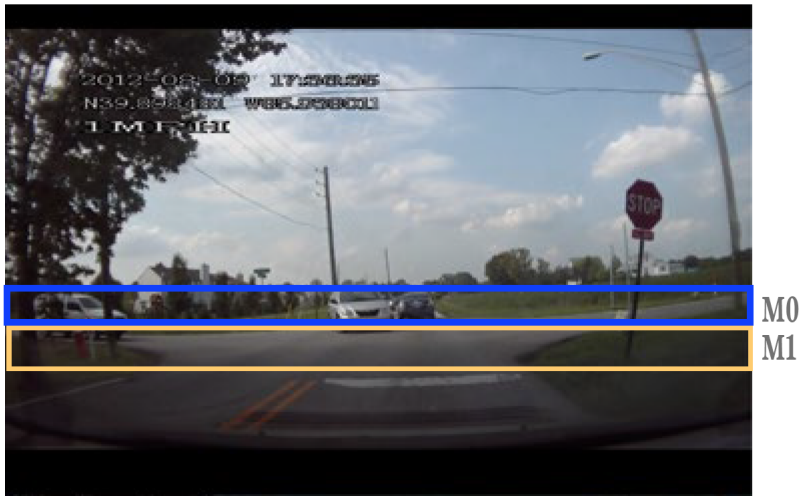


Figure 2.5. Example frame where zones or belts are set to average color there for motion profile images. M_0 is over the horizon in the frame to obtain traffic up to infinity, and M_1 is set for a closer range to finding surrounding vehicle motion.

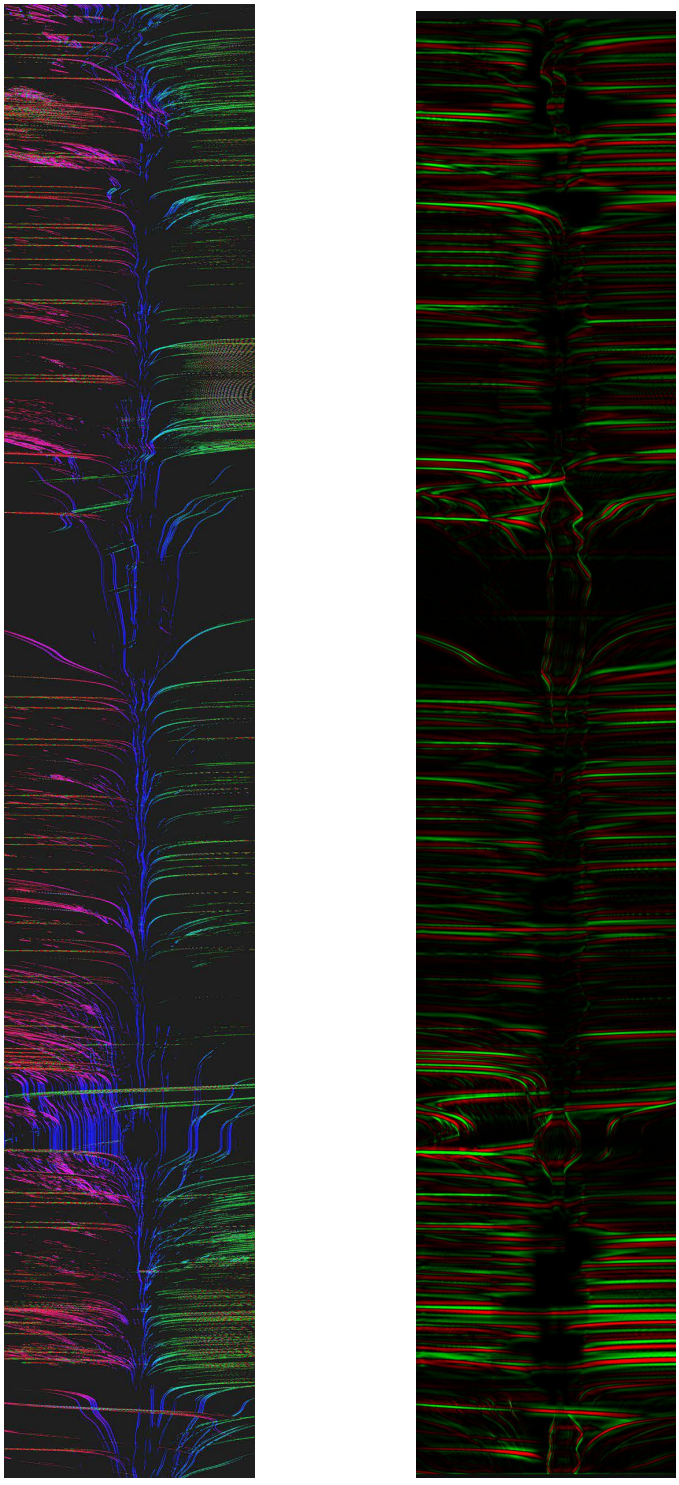
Due to the fact that we average the pixels vertically in the zone, sloping road borders and lane markings in the image are blurred into a wide belt in the motion profile image as in in Figure 2.6. Additionally, details in the road area and grass along the wayside are obscured. After pixel averaging, only vertical lines on vehicles retain their high contrast, forming unique trajectories in the motion profile image. As a result, we may concentrate on vehicle trajectories and disregard road borders and surface markings.

2.2.2 Image motion detection in Motion Profile Image

In the motion profile image M, differential filters $D(x)$ and $D(t)$ are applied horizontally and vertically (x, t). The outputs of the filters are combined to generate the gradient $G(x, t)$ of the motion profile image's strong traces. Figure 2.7(a) illustrates such a trace direction picture, where the pixel contrast is shown in terms of intensity and the tangent direction of trajectories is expressed in terms of colors. The orientation of the trace is transformed to an

(a) R_1 (b) M_1

Figure 2.6. road profile image (R_1) and motion profile image M_1 from middle range in the video of 5-min driving. The horizontal axis is the same as the image while the vertical axis upward is the time in pixel (frame number).



(a) (b)

Figure 2.7. (a) Motion direction of distinct trajectories in colors in $A(x,t)$, and (b) trace stripe image $L(x,t)$. (a) Vertical traces are marked in blue and horizontal traces are close to red or green according to their negative or positive direction. The intensity of color is the gradient value at that point. (b) Positive and negative values in red and green after filtering L_T.

angle in the range (-90, +90) degrees, with the vertical (ahead direction) set to 0. A vehicle trace is a collection of edge traces that all point in the same direction. Due to digitization problems and poor temporal resolution of the motion profile image for rapid vehicles, near to horizontal traces may have angle values ranging from 90 to -90 degrees after local pixel filtering. The trace angle is recorded in an image $A(x,t)$ for the purpose of determining the vehicle's direction of travel. We visualize the value of motion direction in color according to the RGB values set next.

$$A(x,t) = \begin{cases} [0, (G_t)^2 + (G_x)^2, 0], & G_x > 0, \\ [(G_t)^2 + (G_x)^2, 0, 0], & G_x < 0, \\ [0, 0, (G_t)^2 + (G_x)^2], & G_x \approx 0, \end{cases} \tag{2.4}$$

where G_x and G_t is the gradient value of the motion profile based on horizontal and vertical axis. If the motion trace direction of the point is going to leftward, the value is in red channel. IF the motion is close to zero (vertical trace in the motion profile image), the trace direction is displayed in blue. Finally, the value is in green channel, If the motion trace is rightward.

Along with the trace direction applied to the motion profile, another 1D Laplacian filter $L_T(t)$ is applied vertically to detect horizontal trace stripes bounded with two edges. The length of filter is T obtained from the average time of passing and passed vehicles in the driving video. They are set to sense the relative motion of vehicles in next lanes and check cut-in motion into driving lane.

$$L(x,t) = \frac{\partial^2 M(x,t)}{\partial t^2} \tag{2.5}$$

where M is the intensity of motion profile, x is the horizontal axis of the profile, and t is the time axis(vertical axis) of the profile. Figure 2.7(b) shows the entire motion profile filtered by this vertical filter resulting horizontal stripes in $L(x,t)$. If the value of $L(x,t)$ is positive, it shows in red channel in Figure 2.7(b) and if the value is negative, it shows in green channel.

2.3 Video Profile Dataset Creation for Mining and Algorithm Development

2.3.1 Profile data generation and attributes computation

In order to study the appearance of roads and road edges, we analyzed more than 7,000 video clips (2TB) of a large Naturalistic Driving Video (NDV) collection obtained from 110 cars for a period of one year. The four seasons and all weather are captured at different times. These roads include highways, rural roads, urban roads, and roads in a residential area of a large city in the States. The cameras that get NDV are all of the same type, and automatically expose according to the overall intensity of the picture. The exposure is stable during the movement of the vehicle, except when crossing a bridge or entering a tunnel, as the light of the entire picture will change instantaneously. Our approach is to understand the appearance of the road in relation to the scene, weather and lighting; we look for statistical attributes and qualitative signs in the video. We built a web interface on the Internet to visualize videos, key frames, road profile images and processing results. We can browse the road appearance and road overview in the video. The video is manually marked in its properties.

The NDV is collected by TASI in IUPUI which creates a driving dataset over a year driving by 110 cars to cover various roads in the great Indianapolis Area [25]. We used the driving videos from the TASI dataset and manually marked properties in these video by ourselves in this research.

To condense driving video for understanding interactions, the height of horizon in the video is first located for each camera. The frame is located with sampling regions or zones as well as sampling lines below the horizon. Top zone, M_0, is on the horizon to sense horizontal motion of all vehicles up to infinity. A middle range zone, M_1, covers range 10-20m for understanding actions of surrounding vehicles and path planning. A close-range zone, M_2, located even lower in the image captures the sudden invasion of side vehicle and approaching front vehicle for urgent braking.

In Figure 2.8, each sample zone produces a line of data by averaging pixels vertically. The lines from consecutive frames are further concatenated to form a long spatial-temporal image, $M_{\mathrm{i}}(x,t), \mathrm{i} = 0, 1, 2$, showing motion trajectories of scenes. A motion profile from zone

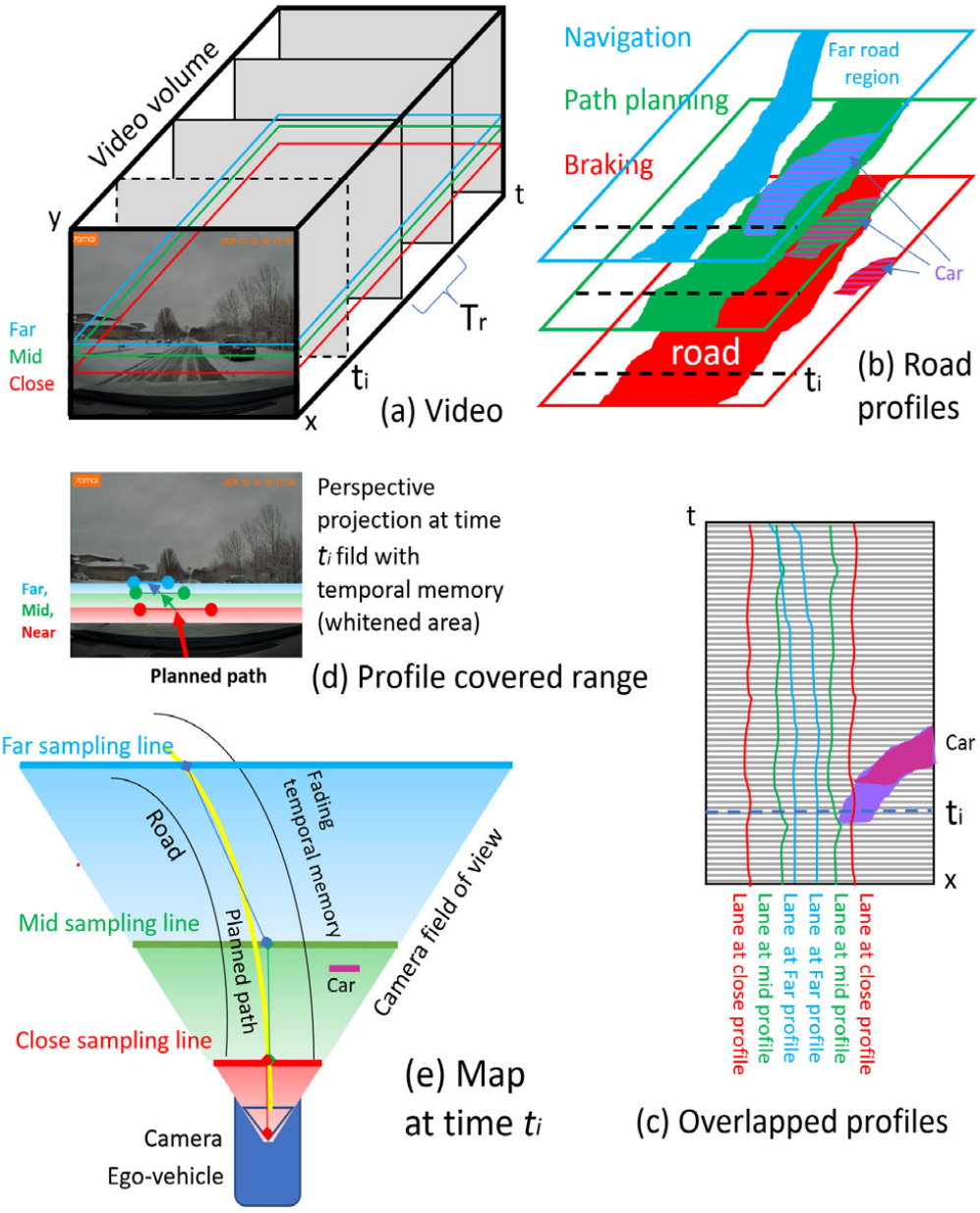


Figure 2.8. Image data structure of three slices of road profiles in the video volume, which are much smaller than the video volume.

M_1 and a road profile from a line in the zone show vehicles in both driving lane and next lane moving in parallel. The motion profile from closer range M_2 does not cover vehicle interaction but can be used for determining urgent braking if a vehicle trace is visible in it. On the other hand, the motion profile from far away, i.e., M_0, has dense traces of background, which

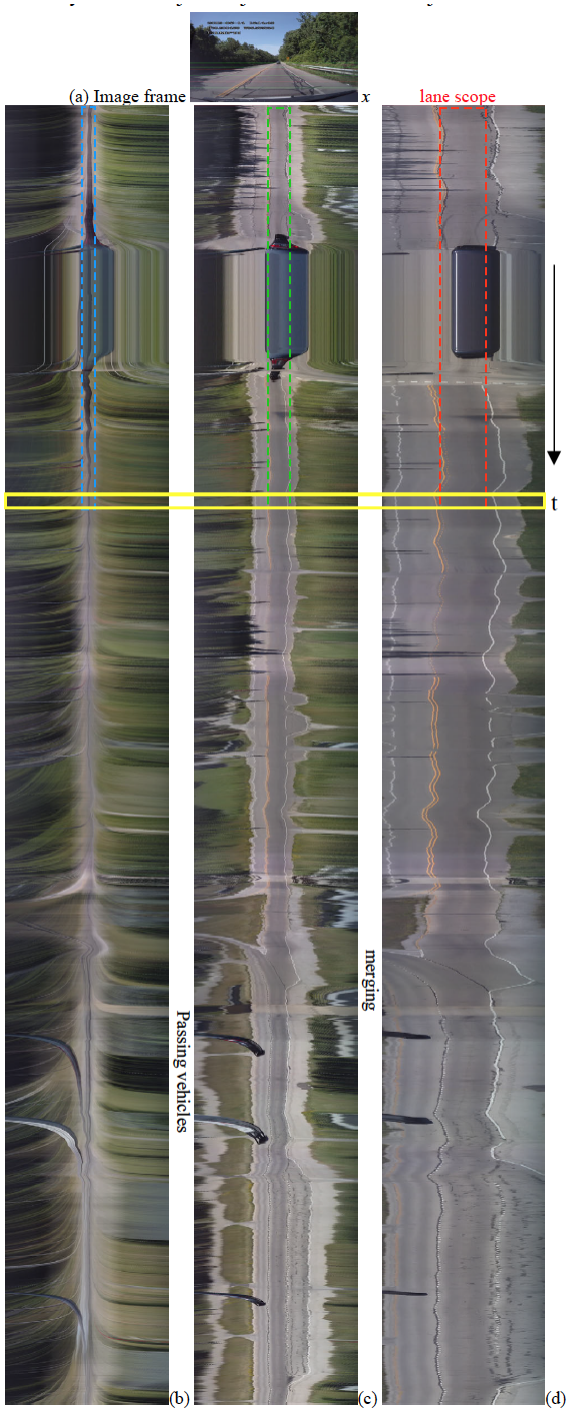


Figure 2.9. Five-minute driving yields motion profile image M_0 in (b) and road profile images R_2 and R_3 in (c), and (d) at far, mid, and close ranges ahead respectively. They are obtained continuously from the top zone on horizon and two lines below in the video shown in (a) when the vehicle moves forward. Color widths in blue, green, red above show the vehicle passing positions at different depths deviated from lane marks. A yellow box is the time window to detect scenes.

makes it hard to identify vehicle traces against background. We will use them occasionally as Figure 2.9 demonstrates.

If the sampling of for a road profile image is blocked by an object such as dynamic vehicle or static object, it will capture pixel line segments on the object as illustrated in 3D layout of scenes in Figure 2.10. The relation of the sampling line with the road (zero height) and objects (with certain heights) are described below.

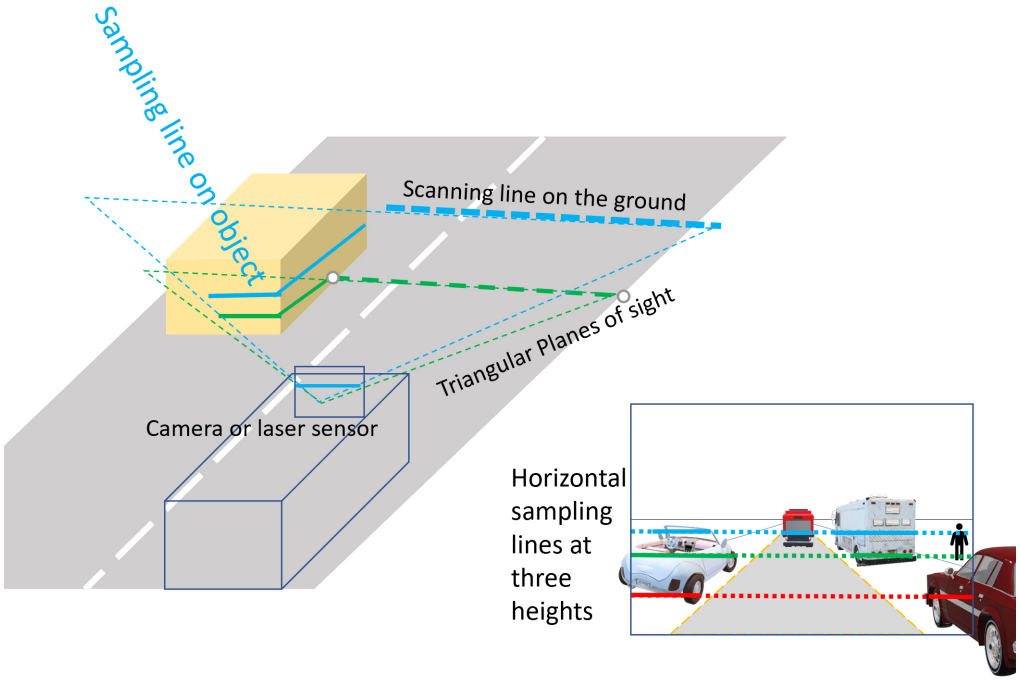


Figure 2.10. Multiple lines to cut road and objects at different depths. Its 3D layout including solid lines on objects and dashed lines on free road are depicted.

Objects: an object with a reasonable height (e.g., above bumper) will be cut by at least one plane of sight when it is closer than the farthest line. The object is sampled constantly at every moment rather than once. This is also true for a static object of the same height (stopped car, roadside pole apply).

Road: object lower than bumper height can be treated as static road fixture such as lane mark, curb, etc. These scenes are scanned once by a sampling line, and three times by three lines at different depths.

2.3.2 Semantic Segmentation of Road and Motion Profile Image

There have been endeavors made to segment road profiles to different regions of symbolic classes [39] such as road, off-road, vehicle, lane mark, etc. The lateral position and width can be gotten from shading and color. The longitudinal shapes are from the transient progression in the road profile scaled by the vehicle speed. Figure 2.11 is the semantic segmentation applied to road profile along with the motion profile as what has been applied on ordinary pictures so far. Cheng et. al [39] have categorized and classified pixels on the latest lines in the road profile into six types of semantic areas based on the surface materials and vehicle motion styles for path planning and autonomous driving on typical roads. The RGB values of the pixels are labeled as follows:

• Road (128, 64, 128): the road surface in temporal space.

• Roadsides (128, 128, 128): adjacent to the road on two sides, including the sidewalk, grass, buildings area, etc.

• Vehicles (64, 0, 128): moving or stopped vehicle seen from the driving view.

• Lane marks (255, 255, 255): include either yellow solid line or white dashed line on the road.

• Vertical obstacles (0, 128, 64): vertical objects on road side, including buildings, telegraph poles and so forth.

• Stopping period (192, 128, 64): the whole period in road profile while ego-vehicle is stopping temporally.

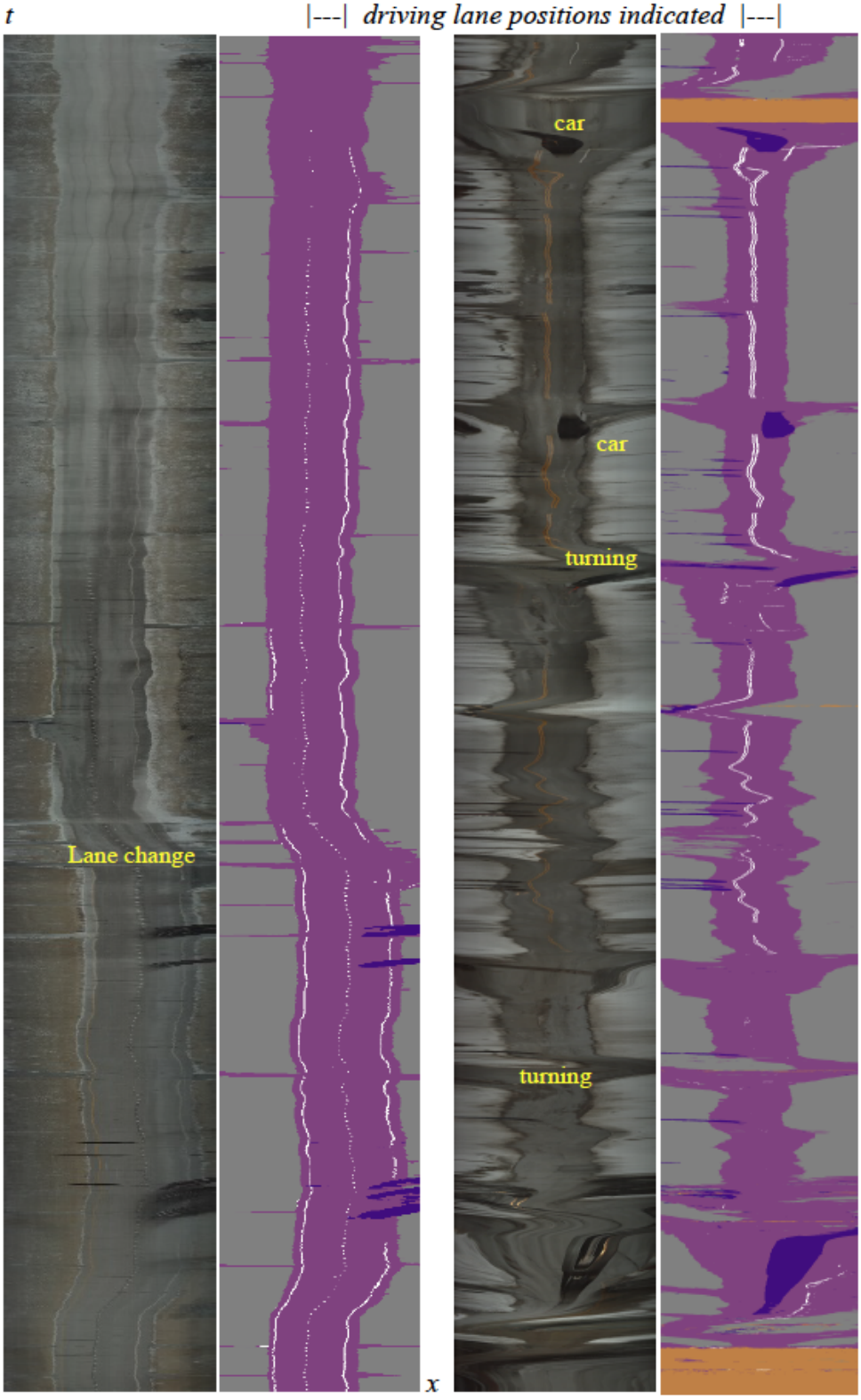


Figure 2.11. Examples of spatial-temporal road profile image $R_2(x,t)$ in cloudy and snow days and their segmented results to pixels of road (magenta), off-road (gray), vehicle (purple-blue), lane mark (white), and stopping (orange).

3. ROAD DETECTION UNDER VARIOUS ILLUMINATIONS

3.1 Various Appearances of Road Edges Based on Physics Conditions

Roads exhibit boundaries that differentiate the road surface from off-road areas. The cameras mounted on the vehicle collect visual data on on- and off-road appearance based on the following physical attributes and lighting conditions.

3.1.1 Material, reflectance and seasonal factor

The material of road surface ranges from asphalt, concrete, to gravel. Roadside can be concrete, gravel, brick, soil, grass, vegetation, water, snow, guardrail, construction cones, etc. Their combinations contain more variations. Even for an asphalt surface, it looks dark for new pavement or on wet road, while becomes gray and white after years or in a sunny day with specular reflection of certain degree. A uniformed color on road surface will not occur at repaired spots and shadow areas. In different seasons, off-road scenes can change largely from green to yellow, or even white due to snow coverage. These material differences are described by surface reflectance. The subdivision of materials are further listed in 3.1. We denote materials as $m_1, m_2, ...m_{\mathrm{i}}$, and their global frequency or probability, $Prob(m_{\mathrm{i}})$, in a country like US can be obtained from separate data sources such as GIS database.

Table 3.1. Subdivision of materials affacted by seasons

Material	Season and status
Asphalt	New, old, wet, dry
Grass	Green, yellow
Vegetation	Green, yellow, gray
Snow	Fully covered, whitened
Water	Wet, mirror reflection

The physical road edge is defined as the first position where material or elevation changes from road surface. The material changes such as asphalt-grass, asphalt-curb, asphalt-dirt, and so on cause differences in reflectance captured under sufficient illumination. An elevation change happens on curb, barrier, etc. where shading may change under sunshine or vehicle headlight. Road surface materials are mostly asphalt with old surface in gray and new

Table 3.2. Road and road side materials

Roadside materials	Season	In road profile image
grass	green	regions
grass	yellow/gray	
vege	green	with texture
vege	yellow/brown	with texture
tree	green	Vertical lines become hyperbolas
tree	brown	Vertical lines become hyperbolas
gravel		white region
soil/dirt		Yellow region
snow-covered		whiter than road
ditch/cliff		large variation
concrete barrier/curb		multiple line due to shaping and shadow
guardrail, construction cones		multiple line, original hyperbolas
vehicle on side lane		dark region next to road surface

surface in black. In addition, we treat wet road surface separately from dry road because it causes strong specular (mirror) reflection. Table 3.3 summarizes road side materials and objects on normal highway, local and rural road, urban street, etc. We integrate seasons with surface materials, because seasonal changes such as snow-covered ground, yellow grass, and wet road alter the reflectance on roadsides.

Table 3.3. Road, off-road, season, weather, and illumination conditions

Features	Classes
Road Materials	Asphalt new, Asphalt old, Asphalt repaired, Contrete, Gravel, Wet, Snow covered
Off-road Materials	Grass green, Grass yellow/gray, Vegetation green, Vegetation yellow/brown, Tree/forest green, Tree/forest brown Soil/dirt, Gravel, Concrete/Curb, Cliff/ditch, Cones, Guardrail/barrier, Vehicles
Weathers/Illuminations	Sunny face sun(SF), Sunny back to sun(BS), Snowy, Rainy, Shadow, Cloudy, Foggy, Dark lit, Night, Direct light, Dirty windshield

3.1.2 Weather, lighting condition and direction

The road color captured by a camera is also determined from illumination. The light can be directional sunlight in sunny day, diffused light in cloudy day, ambient light in shadow area, dark lit at dusk, and at night illuminated by vehicle headlights and street lights.

In driving, video cameras may capture direct illuminants during sun rise or sun set. Mostly cameras observe reflected light from scenes, which further include diffused reflection and specular reflection. Specular reflection occurs slightly on road surface when a car faces the sun, or the road surface has water in raining day; the surface will be brightened by reflecting sky or darkened due to tall objects like mountains, buildings and trees.

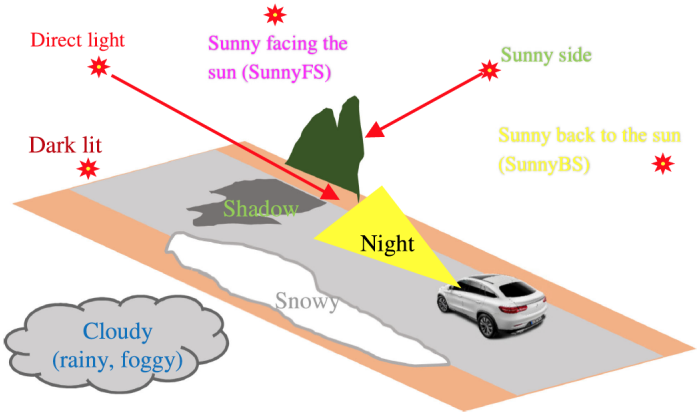


Figure 3.1. Directions of light source with respect to the forward vehicle camera.

In video, image intensity is determined not just by the camera aperture, but also by the camera's sensitivity to fog, rain, and a filthy windshield. During video capture, the auto-exposure function is utilized to balance color over the full field of vision. Due to the broad field of view of the cameras, the intensity variations in the collected driving films remain steady.

We start our weather analysis from the perspective of computer vision. In order to understand the weather on the road, we classify the driving vision affected by the weather, rather than daily weather observations of the sky or weather from forecasts. In the principles of physics and optics, the appearance of the road in the image is determined by the surface reflectance of road and non-road materials, environmental illuminations, and camera

exposure. The general form describing the light intensity from ambient light, reflected light, and direct light of illuminating objects to the camera can be written as

$$I_c = \alpha_1 L_\alpha + \alpha_2 R_d \bullet L + \alpha_3 (R_s \bullet V)\alpha_4 |L| + \alpha_5 |L| \tag{3.1}$$

Table 3.4. Integrated categories for road visual appearances

Features	**Classes**
Road Materials	Asphalt new, old, repaired, Contrete, Gravel, Soil/dirt
Off-road Materials	Grass, Soil/dirt, Gravel, Vegetable/field Concrete/Curb, Cliff/ditch, Tree/forest, Construction cone, Guardrail/barrier, Parked Vehicles
Seasons	Spring, Summer, Fall, Winter
Weathers	Rainy/Wet, Heavy rain, Snowing/snow-covered, Sunny, Cloudy, Foggy
Illuminations	Specular reflection, Dark lit, Direct light, Shadow, Night
Camera sensitivity	Dirty windshield, Normal, auto-exposure

According to Phong's reflection model, on a three-dimensional point. Here $\bullet$ represents the inner product of the vector. I_c is the camera input at that point. L is the vector toward the light source. If the point is an illuminating object, such as the sun or a car headlight, the brightness is the scale. V is the vector along the viewing direction. L_a is the ambient light intensity. R_d is the diffuse reflectance vector aligned with the surface normal of the point. R_s is a vector scaled by the specular reflectance of the point along the intercept direction of the incident ray, and $\alpha_\mathrm{i}, \mathrm{i} = 1, 2, ..5$ are the coefficients of different components, which not only reflect their size, but also due to snow and rain. The intervention of, fog, etc. reduces them. Although this model has been used for graphical simulation of road environments, it is difficult to accurately recover these real physical parameters from low-quality driving videos and graphics. These real physical parameters are accurately restored in quality driving videos. However, if our goal is to accurately estimate these parameters, it is not necessary if our goal is to guide the vehicle in real time. Therefore, we adopt data mining methods to study the road appearance model.

In an image with a wide field of view, the appearance of the road in Figure 1.2 is much more complicated than the point described by formula (1). Under normal circumstances,

because the surrounding objects may block the light, and the scene is constantly changing during driving, there is no light source that can continuously illuminate all areas. For example, in the shadows cast by trees and buildings, ambient light other than sunlight is lighting. On rainy days, wet roads will produce specular reflections, which may be much stronger than diffuse reflections, which interferes with road edge detection. On partly cloudy days, the sky may be full of clouds, but the road ahead may still be illuminated by sunlight. At sunset, when the vehicle is facing the sun, direct light enters the camera. All these changes related to the area or location in the image are determined by the factors listed in Table 3.4 that span the road material, season, weather, time, and camera direction. Such a large change makes the current road edge detection algorithms successful in good weather, but unstable in the case of insufficient light, because they rely on the reflectivity of the material and heuristic cues. The color captured by the camera is a combination of multiple factors, as shown in Table 3.4. However, the combination of all these factors (columns in Table 3.4) produces a large number of cases, even if not every case is concurrent.

The weather forecast categories used daily are not sufficient in describing the visual characteristics of roads and traffic. They are not sufficient in describing the visual characteristics of roads and traffic. Based on the reflection principle in computer vision, as shown in Figure 3.2. These categories are defined based on optical principles. As defined by the principle, there is still a lot of ambiguity when the video is marked. However, this classification method can be used for sample selection in data mining to cover the weather and light conditions as comprehensively as possible.

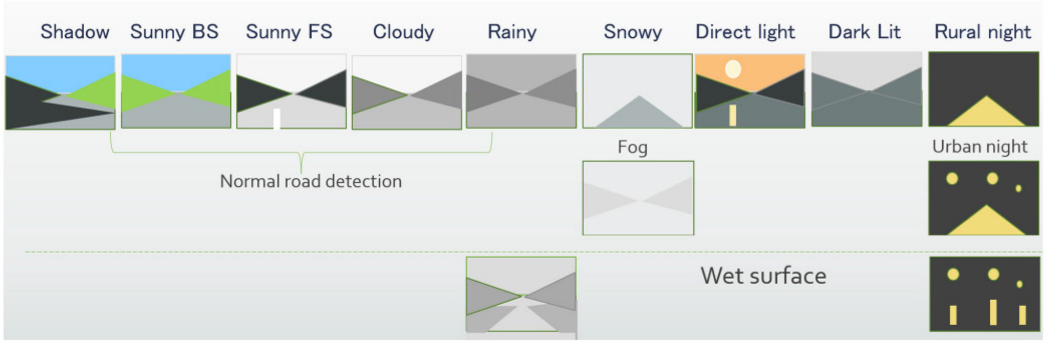


Figure 3.2. Iconic vehicle views of weather and illumination spectrum summarized from a large naturalistic driving dataset.

`Sunny back-to-the-sun` (sunnyBS): includes sunny `side-to-the-sun` . The camera receives diffused reflection in saturated color. It also includes partly cloudy, as a road is sunny while the sky far away has cloud.

`Sunny facing-the-sun` (sunnyFS): involves highlight or specular reflection on road surface, and shadow on most standing objects/scenes. SunnyFS generally has lower saturation than SunnyBS in the image.

`Shadow` : indicates sparse tree shadow and a whole section of shadow in forest or by buildings. It is like dark lit if a vehicle is in a forest. The intensity drops significantly in shadow.

`Cloudy` : is mainly overcast. It varies from bright sky to dark rainy.

`Raining` : is the situation when water drop reduces visibility, classified either when wiper is on or the windshield has raindrops on it. It is hard to distinguish from cloudy.

`Dark lit` : refer to low illumination on road where road edges are barely visible, while the sky is still bright as ambient light. This usually happens in dusk after sun set or bad weather.

`Foggy` : blurs entire scenes on their sharpness and whitens the field of view by adding a layer of water particles.

`Snowy` : is similar as raining while the visibility is reduced more. Roadside is probably being snow-covered more frequently than road surface.

`Direct light` : the sun is in the field of view. The glare suppresses the intensity of other areas due to narrow dynamic range of camera. This is difficult case even for human drivers.

`Dirty windshield` : includes dirty spots on glass and object reflection from dashboard to the camera, when the camera face the sun.

`Night` : video mostly have a dark sky but with street lights and vehicle headlights on. Wet ground at night has drastic reflection of lights.

3.2 Weather and Illumination Categorization and Classification

Our goal in this section is to try to learn and simulate all possible weather and light through big data mining, and to that end, we apply a clustering method to a huge driving

video data set. Knowing the weather and lighting can also improve other visual tasks during driving and provide test scenarios for the vehicle. We can also provide test scenes for vehicles and even design roads with obvious visual features.

Using the principles of computer vision, weather and time are combined to describe the intensity and direction of lighting. This resulted in 10 categories of weather and lighting, as shown in below. Dark lit, direct light and sunlight facing the sun are not terms used in daily life. The direction of sunlight may cause specular reflections on the road surface. We further combined the weather and the camera's sensitivity, because snow, rain and fog may block a certain percentage of the ground reflected light. Heavy rain and dirty windshields have the same effect on the sensitivity of the camera. Please note that a general description of daily weather is not sufficient to distinguish the nuances of the vehicle's field of view. In addition, the qualitative category here is related to dry roads. Some weather, such as rain, night and snow, completely changed the appearance of roads on wet and snow-covered roads. Below is our comment on weather and light from the perspective of sensing.

3.2.1 Video data condensing and feature extraction for weather mining

Observing a large Naturalistic Driving Video database, we conduct a qualitative analysis of the physical environment. Instead of using weather forecasts, we classify the visual performance of roads under "typical" weather, which has relatively few weather and light conditions.

For road edge detection, there are several types of visual features that are particularly important. These features include the linearity that describes the shape or structure of the road extending forward, the surface uniformity from the road surface paving, and the color determined by the road and non-road materials, that is, the surface reflectivity. These characteristics are closely related to the weather and the illumination in the driving field of vision. Therefore, the weather category must be sampled from light-sensitive areas. We do not consider the entire image full of dynamic objects and background scenes, but sampling the image areas that are less affected by the colors of other vehicles on the road ensure that the result is less dependent on the changeable scene. As shown in Figure 3.3, we have selected three regions and a line segment in the video to study the impact of weather. The

one sky area, one road area and two roadside areas are sensitive to light, but are less affected by dynamic traffic. The sky area is sampled at the highest position of the frame to avoid tall buildings and trees.

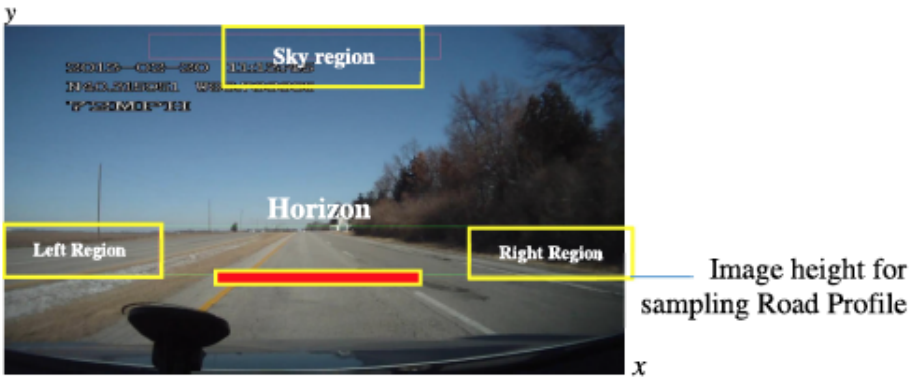


Figure 3.3. Four sampling regions in video frames for weather clustering. The averaged color in each region is recorded from every frame in the video.

After calibrating the horizon in the video frame, we locate a horizontal sampling line l at a fixed distance from the horizon to capture the road surface about 10 to 15 meters in front of the vehicle. The lane width is calibrated according to this setting so that the total width of the frame covers approximately four lanes at the sampling line. Side lanes or off-roads are covered in curbside areas. The road surface of the line sampling is slightly wider than the width of the lane.

The two roadside areas between the horizon and the sampling line avoid capturing tall objects/buildings at different locations, which reduces the color confusion of special objects or signs near the road. If the vehicle is not on a narrow city street, these areas will cover a boarder area. There is also a gap between the two roadside areas and the driving lane to avoid sampling inaccurate road edges in data mining due to the position of the road edge in the image often shifts with the swing of the vehicle. Due to the waving of the vehicle, the edge of the road often deviates. The area on the left has more possibilities to cover the oncoming road (for driving on the right).

Such a setting provides more reliable information for understanding the weather than a single sky area in the video, and avoids the preceding vehicles, trees and buildings in the

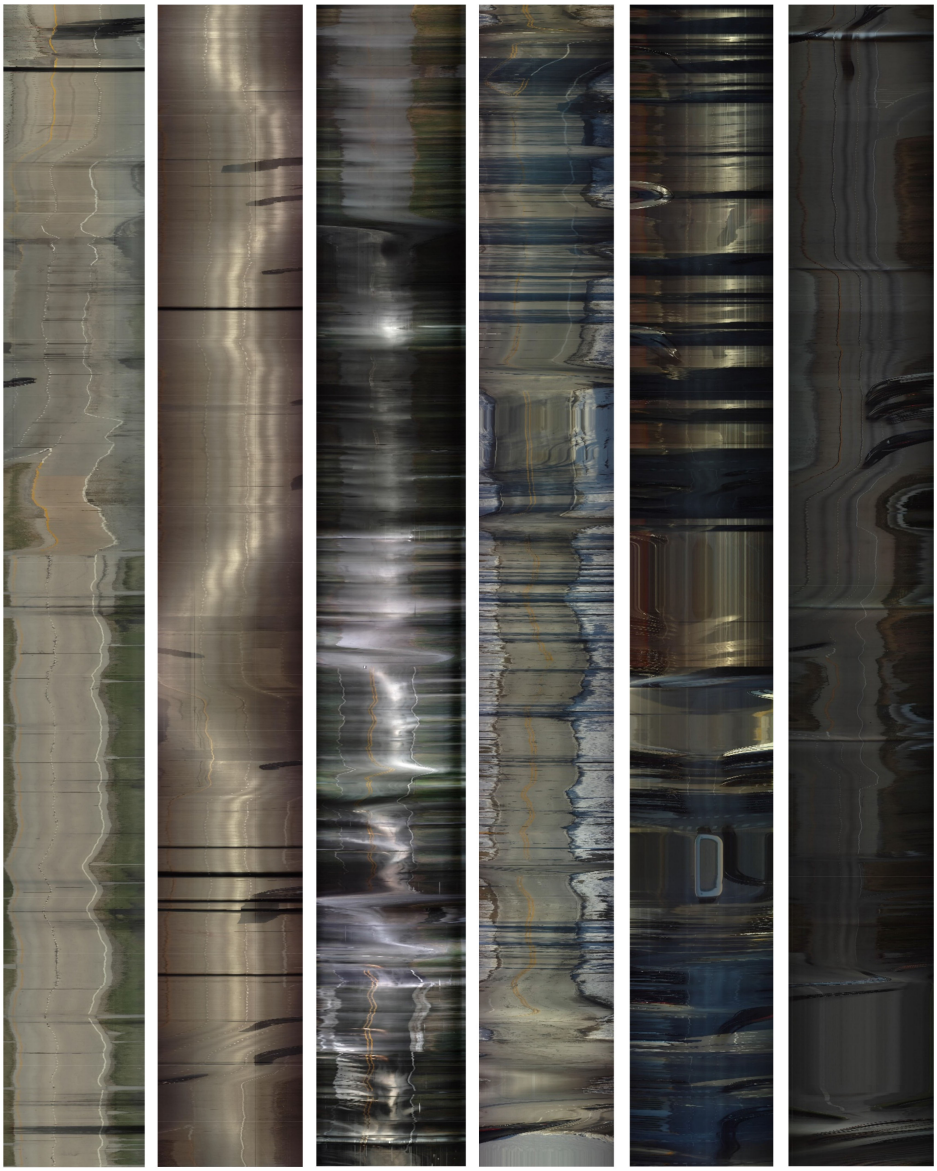


Figure 3.4. Road profile images captured in different weather and illuminations with time axes upward. (a) Sunny back to the sun with color roadside, (b) Sunny facing the sun with highlight on road, (c) Rainy on wet road with specular reflection of sky as highlight. Specular reflection is more dominant than road edges. (d) Snow-covered roadside with partial shadow, (e) Heavy shadow on urban road with road partly visible, and (f) dark lit. Road edge is barely visible.

weather assessment. We calculated the average color intensity, chromaticity and variation of these areas. We do not use hue as a feature because it is more related to roadside materials other than weather and light, such as green or yellow grass. When the intensity or chroma is low, the hue value becomes less reliable. For example, a gray road with almost no color (low saturation) will also produce a hue value. In dark light, the color of the scene is less, and its tonal value related to reflectivity becomes unstable when measured in the image. Chroma, that is, max(r, g, b)-min(r, g, b), rather than saturation, is the first choice, because saturation has a singular value when the intensity is zero, which is random and discontinuous for a dark night image.

Although the lane markings are the most obvious identification sign, we deliberately ignored the lane markings, filtered the pixels of this width, and looked at the resulting global area. It is obvious that the largely changing weather and lighting make the appearance of the road more change than the road material. Insufficient lighting can also make the edges of the road (direct light and rain as shown in Figure 3.4(c) and (e)) and colors lose visibility. For videos taken on rainy days, the movement of the wiper is either captured as a horizontal line, or raindrops form a unique texture on the outline of the road. We examined these phenomena from the perspective of data mining.

3.2.2 Robust clustering of weather and illuminations

In terms of quantity, we use NVD to cluster weather conditions starting from the categories of human annotations, instead of using optical equivalence to simulate. The video used was obtained from a hundred cars on various roads through different seasons, weather and lighting conditions. This work uses data mining methods to explore the visual appearance of roads under different lighting. We use natural driving videos taken throughout the year to study the effects of various weather and light conditions. 300 videos with less traffic jam views (roads and edges blocked by traffic) manually marked 10 weather categories C_{i} at the segment level through the video interface; each frame is a sample, and is labeled in a category as in Figure 3.2.

We mined video data and found the distribution of visual attributes. Through the unsupervised learning of the K-means algorithm, clusters are generated in the feature space,

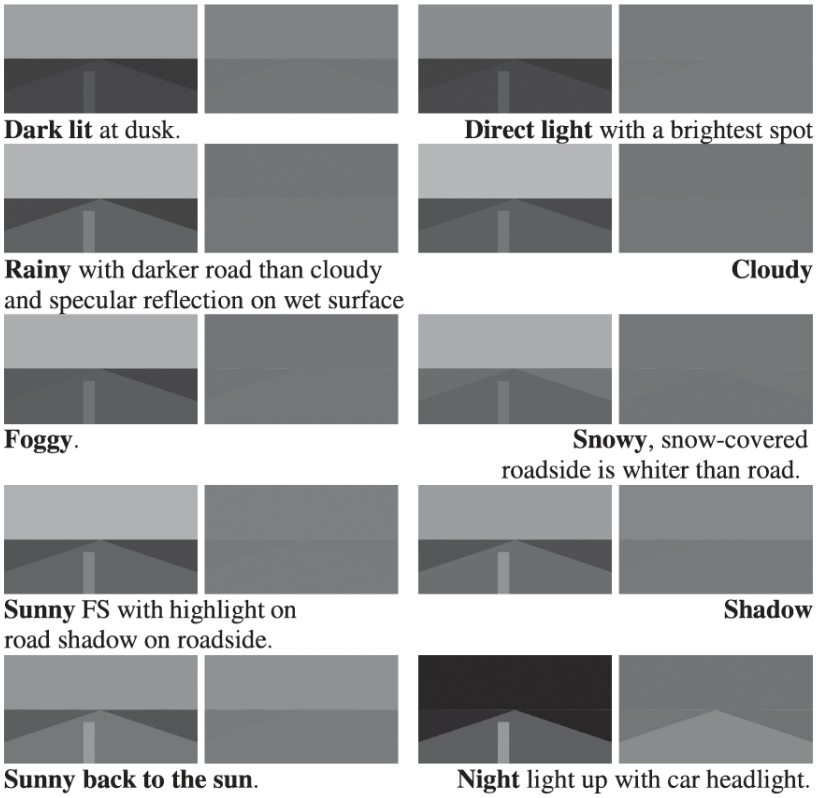


Figure 3.5. Typical intensity and chroma pairs averaged from 300 sample videos with human tagged weather categories. The vertical bars at centers of road areas indicate the color variation against road surface due to specular reflection, shadow, road repairing, etc. Compare to the intensity, chroma averages are much smaller (displayed with 10% scale up in value).

and the spectrum of weather and light is generated in the hierarchical structure. In order to model the weather and lighting, the video is annotated as shown in above. However, the categories of human markers have ambiguity and overlap in the feature space. In order to classify weather and lighting for road edge detection, the K-means clustering algorithm divides all samples digitally in the feature space. Because we think that weather and lighting are continuous in the physical world, they should be used in the feature space. To describe the simple clustering, the 7 classes are a good choice for classifying the weather and lighting in the driving view, with high accuracy. More clusters cannot guarantee a robust classification, and fewer clusters cannot provide detailed guidance for road detection. This kind of evenly divided clusters in the feature space will also give their center points. In this article,

we intuitively add a close category name marked by humans to each K-mean cluster for easy understanding.

We visualize the average image intensity and chromaticity in Figure 3.5. We can notice the difference in the parameters of different weather categories. Night is the easiest to separate from other categories. As an environmental light source, the sky is always better than the ground (road and off-road) is brighter except at night. The intensity of the sky and the intensity and chromaticity of the sky are more important than the left and right side of the road. At night, the left side of the road is brighter than the right side Because it involves the opposite lane, there are bright vehicle headlights from time to time. Similarly, the opposite lane is usually made of asphalt and brighter than off-highway materials, so that the strength on the left is in most triangles, and the strength on the left Slightly higher than the right one. Unless it is snowing (the road on the opposite side melts earlier than the road on the right, the right side of the road melts earlier). The triangles of cloudy, rainy and foggy days are very similar. It is very similar, so it is difficult to compare the appearance of these weather conditions on the appearance of the road.

By drawing C_{i} categories, i = 1..10, representative images with average characteristics in four areas (including single line segments on the road), we can generate typical views of the ten weather categories in Figure 3.5. We can intuitively observe the difference between highway and off-road, and we can also find that it is difficult to detect the weather at the edge of the road for a short time. In all weather, the blue sky of sunny-BS has the highest chromaticity. Compared with the daytime, the road has the largest chromaticity at different gray levels at night, because vehicles with yellow headlights were used when natural driving video (NDV) was acquired a few years ago. Because the manual marking of the shadow is performed on a set of frames of the road section (not detailed to each frame). Therefore, the average value of shadows is like the average value of sunny days-BS. Rainy and cloudy days are more similar in nature, because the movement of the wiper is ignored in the road profile image. After ignoring the movement of the wiper in the road profile image, the nature of rainy and cloudy days is more similar. Human markings of the weather Human-made markings of the weather may also miss some small collections, such as city nights or even

wet ground. Finally, the human-labeled categories may be ambiguous, and there is a large overlap between the categories in the feature space.

3.2.3 Classification of weather and illumination

We utilize the learning results to classify new videos into these types of weather and illumination to guide road edge detection. The decision tree, K-NN and sparse coding methods are compared, this achieves an accuracy rate of up to 90% and makes the weather/light recognition model more powerful.

The K-NN and decision tree use the intensity of sky, left and right roadside regions to classify weather and illumination conditions. To get the nearest neighborhood, we compute Euclidian distances of normalized intesenty value of sky, left and right roadside regions. Classification using Euclidian distances to each cluster takes no time in detecting weather and illuminations at the frame level, which is adequate for a real-time system as a pre-stage for subsequent processing. Additionally, the decision tree does not require much time for an input to reach the tree leaves via a series of feature checks starting at the root. Matlab's computation time for K-NN is likewise rather small. This means that the approaches discussed in this section are suitable for real-time vehicle on-board applications. The overall accuracy of K-NN is 88.3% when K = 5 and 87.1% when K = 1, and fine decision tree has 62%.

3.3 Road Edge Detection Based on Weather and Illumination Data Mining

The illuminance has a great influence on the appearance of the road. Through our observations in NDV, we can simply divide the light source into sunny sunlight, cloudy or shaded ambient light, and overhead/street light in night reflection mode. In addition, at sunrise and sunset, direct exposure to the camera from the sun is another mode. The reflection mode also includes diffuse reflection and specular reflection (even highlight), which is caused by the material at the edge of the road. Specular reflections appear on the asphalt facing the sun, while specular reflections appear on wet roads. Not all light sources exist at the same time, and some light sources illuminate different parts of the field of view. There

is no classification technique for detecting all types of road edges; the majority of algorithms are focused on a certain type of road.

3.3.1 Mining visual property of road around edges

Road segmentation has been accomplished by the use of color, intensity, hue, edge continuity, and uniformity of road surface, among other factors. However, categorization thresholds are difficult to predict. Due to the fact that the mix of materials and lighting may result in hundreds of possible combinations, not all of them are visually unique. There should be clusters of roadways with comparable aesthetic features. These traits may be sorted via big data mining of driving footage. This is to examine the complete range of visual appearances on highways in comparison to off-road situations. Several visual characteristics that we evaluate are useful for real-time road boundary identification, which are listed as $V = \{$on-road and off-road colors, edge contrast, linearity of road edges, homogeneity on road surface$\}$. $Prob(V)$ may be determined by sampling millions of places from driving footage.

The observing cars in the driving video can change lanes and make abrupt turns at street crossings that should be kept away from the road margins. Periods of halting exist when waiting for signals and during traffic bottlenecks. While driving, other cars are present. These should be kept out of roadside sceneries. It is well established that white and yellow lane markings offer a far stronger indication of road than any road boundary does in terms of road identification. Thus, we pick rural roads that lack obvious lane markings or even eliminate lane markings entirely for the purpose of collecting visual property at road borders. Roads in residential neighborhoods are also chosen owing to their absence of lane markings. If both the land mark and the road edge are visible, the road edge identification algorithm will use the maximum win strategy in favor of the result from lane mark recognition.

Road edge detection is more difficult than lane mark detection due to variations in road and roadside materials, and scenes are more heterogeneous when a road is influenced by weather. Road edge is easily confused in poor illumination conditions. The primary goal of road detection is to separate road and off-road regions from their notable image difference. There are not so many works targeting road edge detection, most using color difference of materials. However, those works have not exhaustively examined all types of materials,

weather and illuminations. The weather and time dependent illumination appear to be the biggest factor influencing road edge appearance because of the qualitative changes in the illumination and reflection model such as specular reflection on ground. Although a difference is observable in normal weather when materials differ across road edges, a road edge may be invisible or located incorrectly when shadow and highlight exist, or illumination is dark.

3.3.2 Qualitative feature selection for road edge candidates

In a broad spectrum of road environments influenced by weather and illumination, we seek effective features for detecting road edges from a large data set of driving videos. Our road edge detection is carried out in the road profile. After features of color, linearity, and homogeneity are computed on road and off road, road edges are searched from the center, i.e., the vehicle heading direction, towards both sides. The road surface color is captured from there.

For input frames $F(x, y, t)$, we compute linearity map $L(x, t)$ and the road profile in intensity map $I(x, t)$ and chroma map $s(x, t)$. Both road profile maps are filtered with a horizontal median filter in a size much wider than the lane mark in the frame, such that no lane marks will give any evidence to road edge identification. The large regions obtained then gives more trustable on-road and off-road separation. This operation also removes the small portholes captured by edges. Above that, edge detection with a large differential filter is applied horizontally on the two maps to find global edges. By examining videos and road profiles of each weather and illumination clusters, we visualize the intensity and saturation, and linearity in road profiles as in Figure 2.4. We can observe that the method can locate nearly perfect road edges under good illumination conditions. For good weather and illuminations such as sunnyBS and sunnyFS, the maximum contrast points computed from intensity and chroma are consistent in position. which confirms that the algorithms developed so far for good weather are successful.

In general, roadside rough materials like grass, tree, gravel, are not as smooth as paved road surface. The reflectance there is thus smaller than that on road. The exceptions appear on two cases: new asphalt paved on surface has color close to black, and snow covered

roadside. Some roadside section like a road crossing is another exception. The roadside has the same intensity as road surface (though not necessary to be identified as roadside). Therefore, a simple sampling of road area ahead based on a general road width is good to separate this exception. Based on our observations, the value of maximum intensity change can serve the road edge detection reliably on most of roads. The exception are the cases of shadow, mirror reflection on wet road as follows.

Shadow has been investigated in many works. Intensity invariant features (e.g., color between green and red) has been proposed to ignore intensity changes around shadow. However, this assumes that roadside has different color regions such as tree, grass, and vegetation. It is less effective if roadside materials such as rock, gravel, concrete, curb, and dirt are in the shadow. The roadside chroma is not always strong in wild area, so as to the chroma difference in shadowed off-road (seeing Figure 2.4). Therefore, road edge detection in shadow has to rely on linearity in road structure. Because shadow is casted from roadside objects, it is horizontal on a road and thus in the road profile. The vertical differential value $|G_y|$ is mostly non-zero on shadow edges, which removes large shadow in the road profile. Partial shadow with boundary on road is more problematic than fully covered by shadow. Even in this scenario, we observed from data that it can be classified as hard and soft shadows depending on the density and size of objects that cast shadow. Soft shadow in winter from defoliated trees on roadside does not affect road edges with strong color difference. Hard shadow in summer casted by buildings and forest reduces intensity, which causes chroma and hue less reliable.

Across different weathers, passing and parked cars on side lane have dark intensity at lower part due to tires and shadow. Their traces in the road profile are hyperbolas. The intensity difference can catch such locations. Generally, concrete curb has shading or even shadow to be detected with a filter. If a curb is facing the sun, it is still possible to be identified using the maximum intensity contrast according to its material difference from an asphalt road and lawn next to it.

SunnyFS has a high contrast between bright sky and road, and dark shadows on most objects with an elevation (Figure 3.6). The view is less colorful (low chroma) because of shadow and white sky as compared to sunnyBS. Specular reflection is visible on road surface

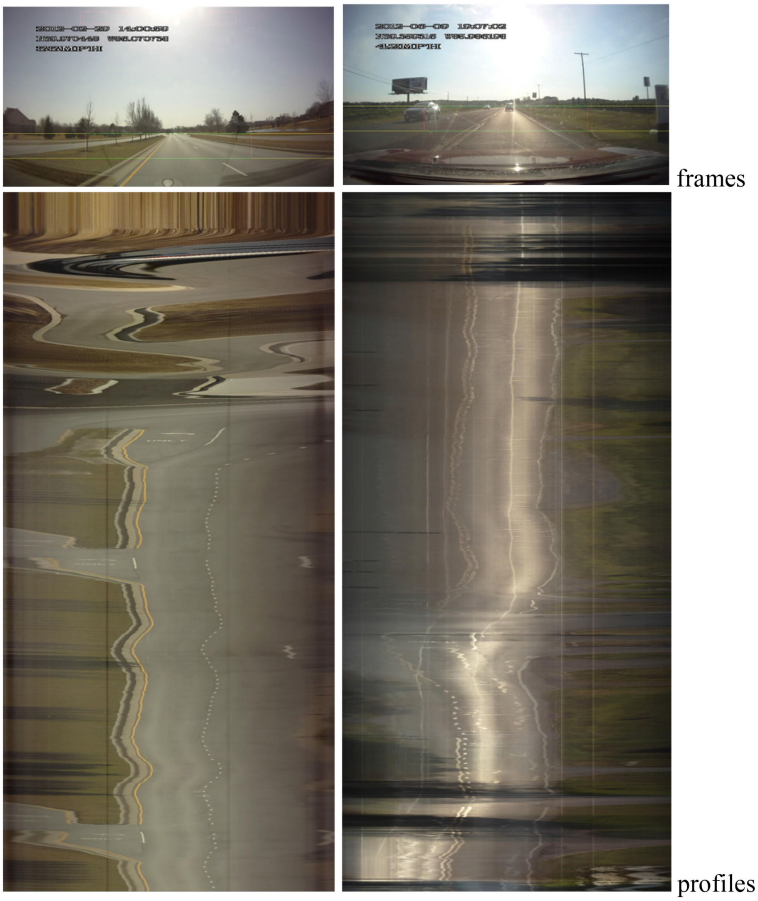


Figure 3.6. SunnyFS. (a) Clear curb visible in road profile. (b) Road surface may have bright specular reflection of sunlight.

sometime, though the boundary of specular reflection is not sharp. It can be omitted in candidate selection for the maximum intensity difference. Off-road area has less specular reflection because of hazard materials. Overall, sunnyFS is not a difficult case in finding road edges.

Gravel and concrete have less difference from old asphalt (close white color) than grass further outside a road. It is hard to detect the boundary between gravel and road than outside grass. The maximum intensity difference may be correctly located if no grass and vegetation are visible. Even if the edge is located at grass boundary, we consider it is correct, as the gravel regions are buffering zones drivable in emergency.

Dark lit has weak light on the ground despite the sky may still be bright. Road edge is still the strongest separator since other details are also suppressed in contrast (Figure 3.7).

Figure 3.7. Road profiles with detected maximum intensity change for dark lit after sun set, and direct light condition before sun set.

For direct light, however, the road edge contrast is suppressed to none by the automatic exposure to a very strong illuminant (the sun). This situation is impossible to handle even for human drivers.

Night is examined for several scenes: night curb on urban road, night urban road under street lights, rural road without light (Figure 3.8), and finally night scenes with a wet road surface. If a road is narrow, road edge is still visible within the vehicle headlight scope. Whether the road edge is at curb or green grass, it can be located with the maximum intensity changes. However, if an off-road region is yellow grass or dirt on rural road without any lane mark, road edge is invisible in yellow headlight. On the other hand, if a road is wide and its edge is outside headlight lit scope, the detectable edge mistakenly falls at the

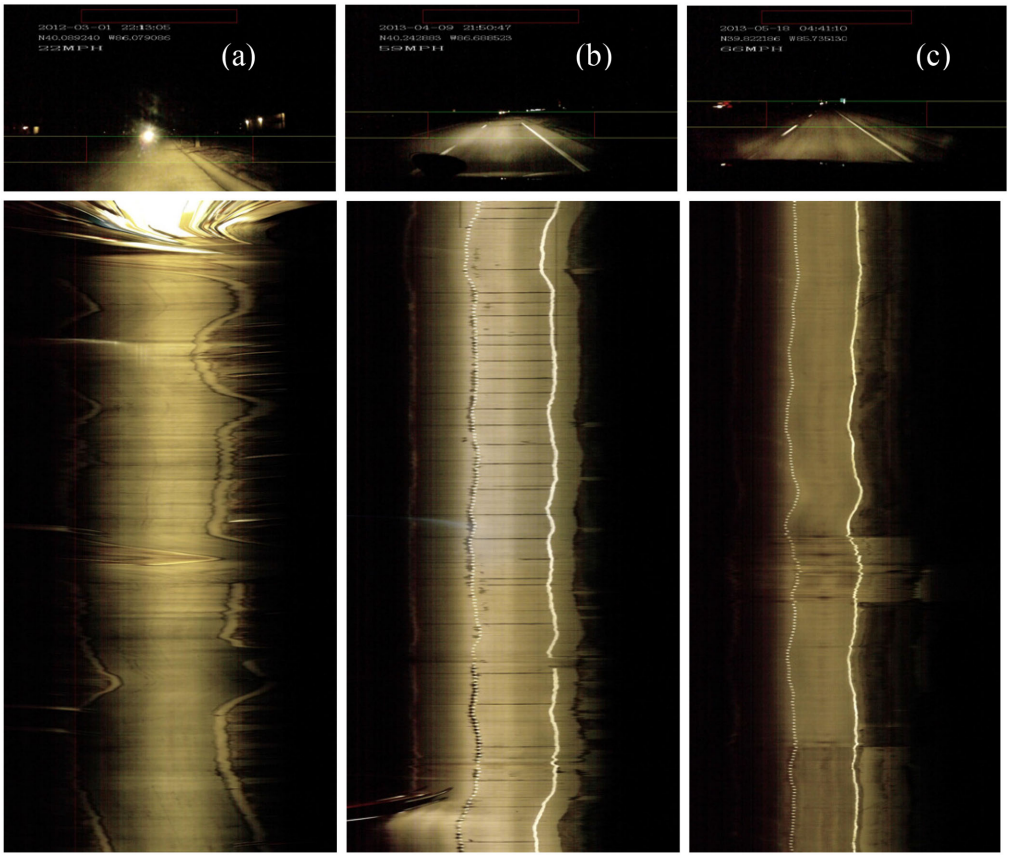


Figure 3.8. Night road profiles. (a) Night curb, (b) edge at green grass and (c) yellow grass out of lit scope of headlight. Lane mark is much distinct than road curb and grass edges in night.

margin of lighting scope. In such a circumstance, the only clue for a human driver to keep a vehicle on road is lane mark. Overall, street light illuminated roads (Figure 3.9) are easier in road edge detection.

Rainy: cannot be separated from cloudy according to the averaged color properties unless a wiper is on. More seriously, road surface can be wet on a raining day and at night, both having strong specular reflection from roadside objects.

In daytime, road surface may contain the mirror reflection of dark trees and buildings in daytime, and various highlights from street lights, traffic signals and vehicle lights at night (Figure 3.10). Mirror-reflected objects and lights are all dragged long vertically on wet road in the video. As a result, they leave hyperbola traces in the road profile image like vertical objects, as a vehicle moves straight at a constant speed. Strong edges in the

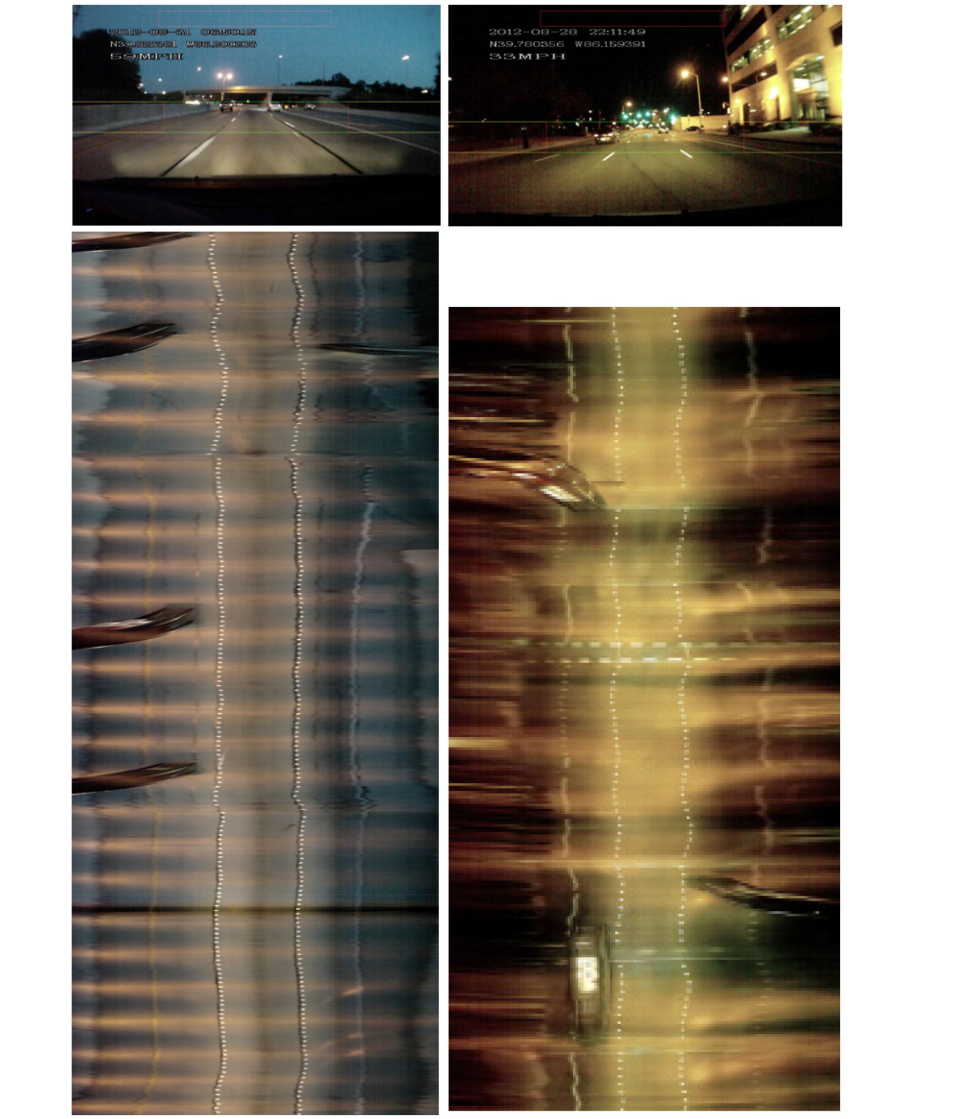


Figure 3.9. Highway and urban road with street lights. Periodic lit areas appear in the road profiles. The roadsides are much brighter than those in Figure 3.8

road profile images do not imply road edges. The non-vertical linearity is a clue to avoid a miss-classification of road edge for mirror reflection. At night, if highlights appear on a wet road, real road edges are mostly invisible along with the bright light reflection. Human drivers have the same difficulty in identifying a road edge in this condition; the driver relies on front vehicles to infer the road direction. Inversely, because off-road regions have less

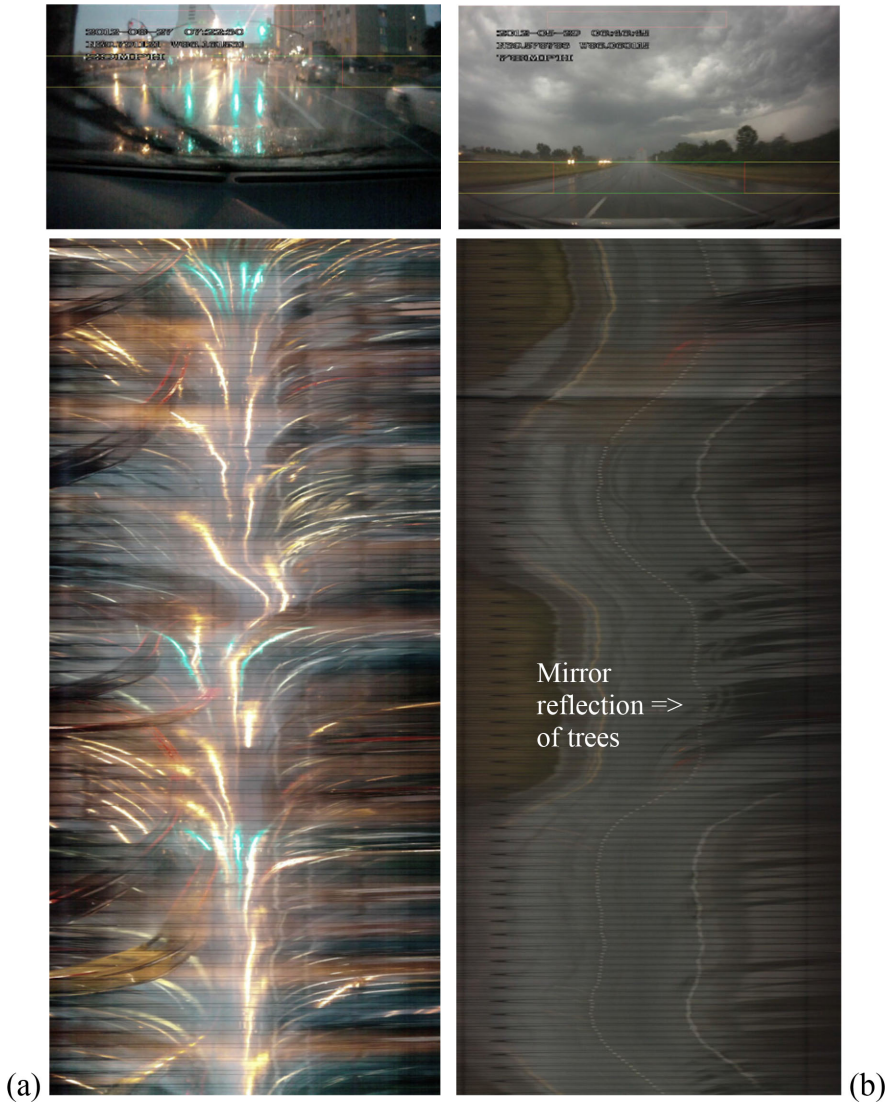


Figure 3.10. Mirror reflection on wet road. (a) Night time reflection of street lights and signals. (b) Tree mirror reflection in dark color. Left road edge is still visible with the maximum intensity changes, while the right mirror reflection of trees in the road area has to be ignored.

mirror reflection due to their non-smooth surface, we can inversely predict that the highlight scope is likely a road area.

Up to this stage, our method is similar as other methods in feature extraction. However, our road detection performs differently when edges have many uncertain candidates under weak illuminations. The methods proposed so far for road edges have focused on optimization to overcome the uncertainty in an adhoc approach; no solid evidence from real data

supports the optimization in many circumstances. Through the data survey, we examine discriminative features in different weather and use the learned data to confirm road edge decision. Qualitatively, the feature responses are visualized in the road profile images for all clusters, and their effectiveness in locating road edges is examined.

3.3.3 Quantitative road edge detection

Our road edge detector further uses Bayesian methods to benefit from data clustering. On each line of the road contour, a minimum threshold is set to accept non-horizontal edges as candidates for road edges. These edges may divide the road and non-road areas. For multiple edges of $x_n, n = 1, 2, 3, ...$, the edge of the road is searched from the center to both sides, as shown in Figure 3.11. For each identified weather and light condition, we select the final road edge, the difference value of which is the most appropriate in the data mining stage. The strongest difference in intensity or chorma is not necessarily the correct edge of the road. For example, on a wet road, the reflection of the rear lights of the vehicle is much stronger than the real road edge, but it should be avoided.

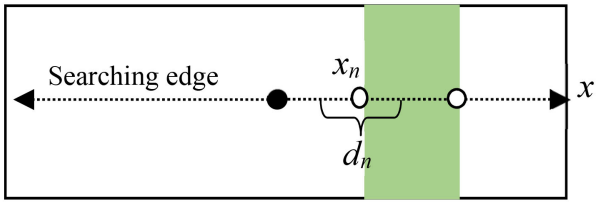


Figure 3.11. Road profile for road edge selection from multiple edge candidates. Feature properties on two sides of edge candidates are considered as evidence.

The key to using big data is to calculate the possibility of features under different weather and lighting conditions. In the weather/lighting cluster, we study the visual characteristics of the edge of the road. Since the road surface is mostly asphalt, the color ranges from black to gray, so the chromaticity value is low, so the color difference d (road-off-road) of the color and intensity space is calculated from each frame of the video set. In the distribution of each cluster is obtained, where the d_{i} of weather cluster i is visualized in the color display with R and G channels, which represent the difference between left and right, respectively.

This is taken as the likelihood probability $P(d_{\mathrm{i}} \mid E, S_{\mathrm{i}}), \mathrm{i} = 1..10$, where E is the edge of the road and S_{i} is a cluster. In this distribution, we can observe several points with high values, which are related to the long duration of the special material on the side of the road. Due to the low of the road, except for snow and dark lights, the chromaticity difference of all clusters is almost negative. With the exception of snow-covered roadsides and new asphalt roads, the strength differences of almost all clusters are positive because the road surface is usually brighter than the non-pavement.

In each frame, the difference of x_n across each edge candidate is the contrast of intensity (ΔI) and chromaticity (Δs), and linearity (L), which is calculated by the horizontal difference filter in $d_n = \Delta I, \Delta s, L..$ After the weather/light level S_{i} is identified, the difference probability $P(d_n \mid E, S_{\mathrm{i}})$ of the likelihood is used to determine the road edge in the frame. For multiple candidates with $n = 1, 2, 3$ in x_n, each side has a difference d_n, we find a roadside to find the roadside with the highest probability among all n, where roadside E only one. Edge E is unique on one side of the road, that is

$$x_n = \arg max_n P(E \mid d_n) \tag{3.2}$$

Using the Bayesian rule, $P(E|d_n)$ is calculated as

$$P(E \mid d_n) = P(E, d_n)/P(d_n) = P(d_n \mid E)P(E)/P(d_n) \tag{3.3}$$

where prior probability $P(E)$ is constant 0 for invisible road edge on a wide road, and 1 for a visible road edge in the frame. The likelihood probability $P(d_n \mid E)$ has been obtained for all possible d. Probability $P(d_n)$ is the edge contrast distribution accumulated for the entire video clips since it is related to the camera exposure for long driving.

4. MOTION DETECTION AND VEHICLE INTERACTIONS

Finding interactions of vehicles on the road will be valuable for accident analysis and driving behavior assessment, because half of the accidents are collision with other vehicles, as well as road departure. Vehicle interactions include cut-in, merging, crossing, frontal approaching, etc. Vehicle interaction does not imply an accident, but a crash or near crash always starts from an interaction. In driving behavior analysis, aggressive drivers have more lane changes passing many vehicles and bumper to bumper chasing. A tired driver may drive slowly and be passed by most vehicles on multi-lanes. Major accidents are bumping into a frontal vehicle or with a cut-in vehicle to the same lane if ego-driver does not brake or avoid promptly. These interactions have been recorded in Naturalistic Driving Videos (NDV) that yield clear visual motion footage.

Driving video provides rich information to understand events between vehicles. We limit our sensing depth up to middle range in examining vehicle interactions. A horizontal zone reaching middle range in the frames can capture image velocity of surrounding vehicles. The range is determined from the relative speed and distance of vehicles with the camera. A motion profile is sampled from such a zone such that surrounding scenes have their trajectories in the generated spatial-temporal image. We discovered that different events around cameras correspond to unique trajectories. By filtering traces at special locations over time, we can classify trace types reflecting different events with surroundings.

Related works on vehicle interactions can be found in [18], [21], [22]. Traditional methods have spent much more efforts to achieve the goal. It starts from object recognition using detector, which requires more computing resources and time. Tracking bounding boxes is followed and motion trajectories have to be estimated and classified to interactions and events. Our method skips object recognition of objects by focuses on vehicle motion observed at special locations in video. This reduced the complexity of the problems significantly because the shape recognition of vehicles suffers from the variations in vehicle shape, orientation, type, color, and illumination conditions. As preparation for motion understanding, the temporal profile has been proposed to directly reflect the relative motion of surrounding vehicles. The

TTC computation has been carried out in the motion profiles. Multiple motion profiles have been employed to sort the near crashing based on the motion trajectories in the field of view.

4.1 Vehicle Interactions on the Road

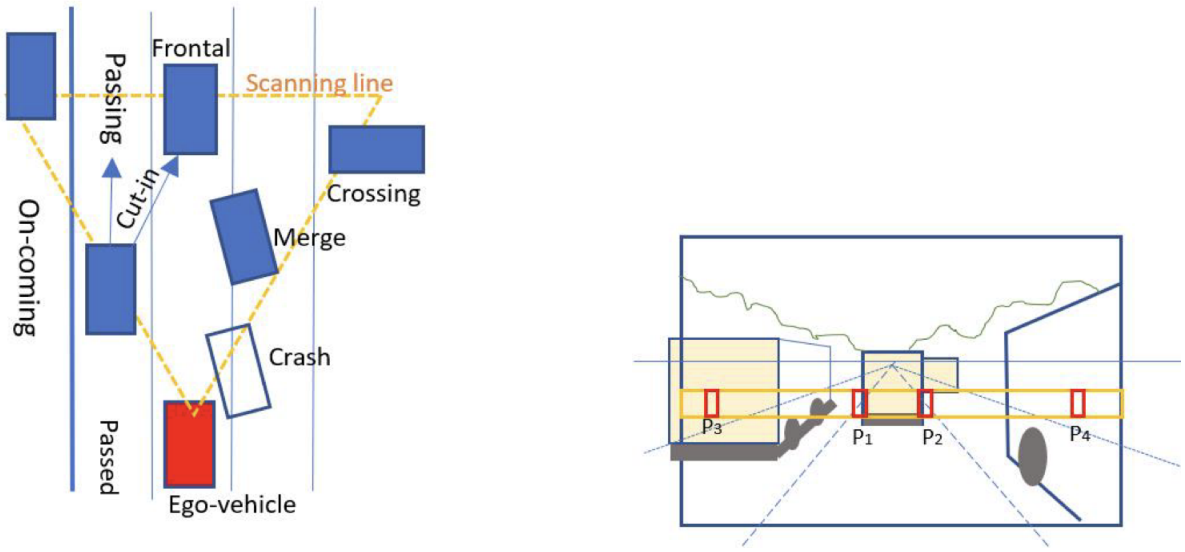
The vehicle interactions discussed in this section are summarized in Table 4.1, and many examples are illustrated in Figure 4.1. Figure 4.2 illustrates several interactions visible in the frames.

In the ego-vehicle-centered area, a vehicle contact occurs at a depth of up to around 20m. Beyond that, vehicle actions are disregarded if they do not immediately result in a crash. Vehicle interactions impair safe driving and may result in a near-crash if not addressed appropriately. We detect vehicle interactions by referring to the full process of vehicle motion. The following criteria are applied in this work: (1) Vehicles are almost same in width due to the road system's design, and hence lane widths are nearly identical. (2) Vehicle mobility is accomplished by the use of four-wheeled vehicles traveling along a nicely curved lane on the road plane. (3) Locally, the ego-vehicle is on a straight road; otherwise, it is depicted as turning or changing lanes in a brief duration.

We discovered through NDV mining that the bottom portions of cars include tire and shadow, which are rather dark in virtually all weather situations except for night driving and heavy rain on a wet road. The road is significantly more illuminated than the lower vehicles. This condition or phenomenon is universally applicable to all vehicles regardless of their color, shape, or kind. The horizontal motion of the car on the road is projected onto the image, while the vertical motion represents depth change. If a vehicle's depth is near, it shows in the lower frame. Monitoring can be removed in the sky and high area in the images. As a result, we develop a method for detecting vehicle interactions in driving videos using motion detection.

Table 4.1. Major vehicle interactions happening on the road and their motion observable in video

	Interaction Types	**Observable Positions in frames**	**Relative Motion in the Field of View**
Frontal vehicle	Following same lane	Front, middle depth range	Width keeps, enlarge and reduce
	Crossing vehicle: at intersection side from turn from opposite road	Side or center, across horizontally at least half field of view	Rightward or leftward across driving lane to one margin
Side vehicle	Passing (overtaking) vehicle goes ahead faster in the next lane	From either side, toward middle depth range	Centered (inward) motion ending at the next lane at middle depth
	Passed (overtaken) by ego-vehicle next lane	Toward either side, from middle depth range	Motion starting from the next lane at middle depth range outward to margin
	Cut-in vehicle from the next lane into the driving lane quickly	Either side to center, at middle depth. Only partially visible in crash	Centered motion faster than Passing ending at the same lane of ego-vehicle
	Merging or ramp slowly merging into the driving lane	From either side, toward center	Faster motion than cut-in from margin
	On-coming in the opposite lane	Starting from opposite lane at center, middle depth, fast speed leftward	Fast leftward motion due to fast relative speed between opposite vehicles and ego-vehicle.
Ego-vehicle	Stopping for signal or sign	Most part of field of view are static except crossing vehicles	Scenes are static except crossing vehicles having motion, left-turning vehicle from opposite road moving rightward.
	Turning at interaction	Scenes shift in entire field of view	Opposite to the turning direction
	Changing lane	Scenes shift in entire field of view	Short deviation of all scenes
	Speeding	Roadside scenes on two sides move outward	Roadside scene expending fast to margins from center
Traffic	High density vehicle flow	Many vehicles at middle depth	Stable motion near center
	Leaving vehicles: turning away without interaction	Two sides	Motion outward to a margin



(a) (b)

Figure 4.1. Different interactions between ego-vehicle and surrounding vehicles. (a) Interactions in a top view where ego-vehicle is moving upward. Many interactions are symmetric on left and right sides except for on-coming vehicles on left side. (b) Front view with a horizontal zone (M_1 in orange color) covering scenes up to a middle range of road for vehicle motion identification. In addition, M_0 and M_2 zones can also be obtained in the same way above and below M_1 for farther and closer vehicles, respectively.

4.2 Motion Profile Images Capturing Horizontal Motion of Vehicles

4.2.1 Focusing on specific positions at Motion Profile Images

To condense driving video for understanding interactions, the height of horizon in the video is first located for each camera. The frame is located with sampling regions or zones as well as sampling lines below the horizon. Top zone, M_0, is on the horizon to sense horizontal motion of all vehicles up to infinity (Figure 4.3). A middle range zone, M_1, covers range 10-20m for understanding actions of surrounding vehicles and path planning. A close-range zone, M_2, located even lower in the image captures the sudden invasion of side vehicle and approaching front vehicle for urgent braking.

In each sampling zone, pixels are averaged vertically to produce a line of data. The lines from consecutive frames are further concatenated to form a long spatial-temporal image, $M_{\mathrm{i}}(x,t), \mathrm{i} = 0, 1, 2$, showing motion trajectories of scenes. Because we average the pixels

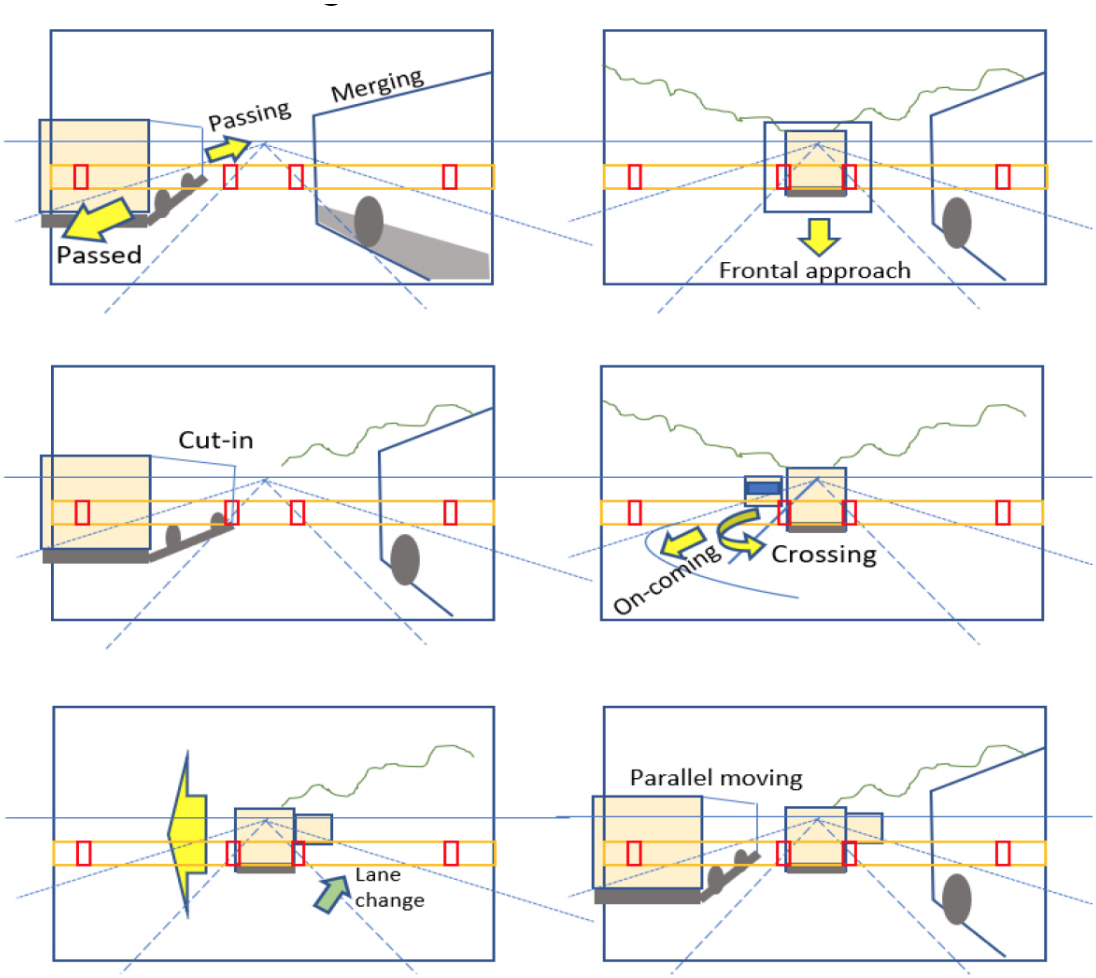


Figure 4.2. Front views of driving video in different events. Yellow arrows indicate observable vehicle motion. Green arrow is for ego-vehicle.

vertically in the zone, slanted road edges and lane marks in the image are blurred to a wide belt in the motion profile. Details in road area and grass at roadside are also blurred. Only vertical lines on vehicles remain high contrast after the pixel averaging, which further form distinct trajectories in the motion profile. Therefore, we can focus on trajectories for vehicles and ignore road edges and surface marks. As displayed in Figure 4.4, a motion profile from zone M_1 and a road profile from a line in the zone show vehicles in both driving lane and next lane moving in parallel. The motion profile from closer range M_2 does not cover vehicle interaction but can be used for determining urgent braking if a vehicle trace is visible in it. On the other hand, the motion profile from far away, i.e., M_0, has dense traces

of background, which makes it hard to identify vehicle traces against background. We will use them occasionally.

Figure 4.3. Example frames of right turn of an opposite vehicle at an intersection.

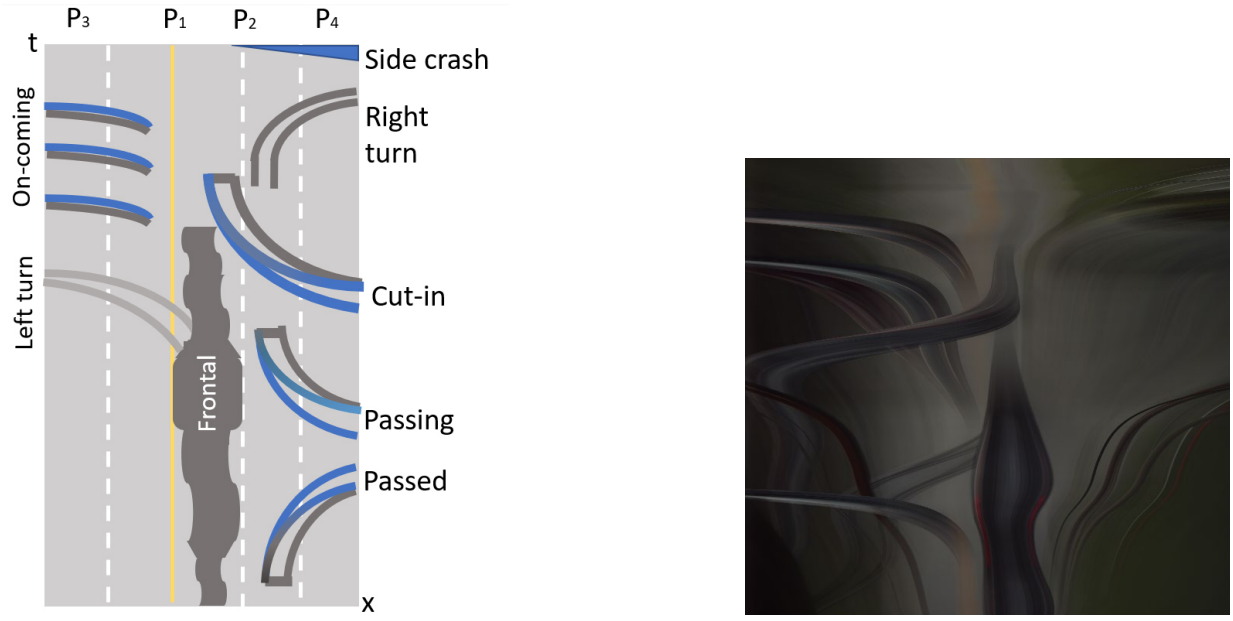
We discovered vehicle traces as illustrated in Figure 4.5 by monitoring vehicle motion in Naturalistic Driving Videos as well as the traces in the produced motion profile images.

- A cut-in car follows an inward path toward the center and eventually merges into the same lane as the ego vehicle.
- A passing vehicle traveling faster than the ego-vehicle on either side leaves an inward trail that vanishes towards the picture center at medium range.
- A vehicle that is passed faster than the ego-vehicle on each side has a trace toward the margin.

(a) Road profile(R_1)

(b) Motion profile image(M_1)

Figure 4.4. Road profile (R_1) and motion profile image M_1 from middle range in the video of 5-min driving. The horizontal axis is the same as the image while the vertical axis upward is the time in pixel (frame number).



(a) motion profile image(M_1)

(b) An example

Figure 4.5. Characters of motion traces in different events. Blue and gray traces indicate side and back of vehicles, respectively. The traces can be flipped to the other side horizontally except for on-coming vehicle traces.

- A frontal and side vehicle may travel in tandem with the ego-vehicle. Their traces remain vertical and about the same breadth.
- A merging vehicle that is successful may have a quick trace toward the center while its width decreases; alternatively, if its trace expands at the same place, it may cause a collision and need braking.
- When a car departs, its breadth decreases. Inversely, whether it is approaching the camera because it is slowing down or the ego-vehicle is traveling faster, the vehicle's breadth is expanding. The time required to reach a collision is calculable.
- Due to the high relative speed, approaching cars in the other lane appear in the image center and move outward to the left margin in a highly horizontal way. They are more visible from a road's left lane.
- When the ego-vehicle turns or changes lanes, all of the traces in the profiles diverge in the opposite direction of the turning direction within a brief interval.
- Except for a few crossing vehicles, a halt period of the ego-vehicle results in pure vertical traces in the motion profile image over the whole field.

- Lane markings and road borders are visible in the road profiles; as a vehicle changes lanes, the lane markings bend and pass through the profile's center. This is used to identify lane change events involving the ego-vehicle.

4.2.2 Detecting Interaction in the Motion Profile Image

Along with the trace direction applied to the motion profile image, a vertically oriented 1D Laplacian filter $LT(t)$ is applied to multiple horizontal positions p_1, p_2, p_3, and p_4 in the motion profile image (Figure 4.5) to identify horizontal trace stripes with two edges. The filter's length is T, which is determined by the average time between passing and passed cars in the driving video. The four horizontal monitoring points are denoted by tiny red boxes in Figure 4.3. Two outer places, p_3 and p_4, are near to the picture boundaries, while two inner spots, p_1 and p_2, are at the midway depth of the driving lane. They are programmed to detect relative motion between cars in adjacent lanes and to prevent cut-in motion into the driving lane. Figure 2.7(b) illustrates the whole motion profile image after it has been filtered vertically, resulting in horizontal stripes in $L.(x, t)$.

$$A(p,t) = \begin{cases} \ll 90 & \text{passing vehicle} \\ \gg 90 & \text{passed vehicle} \\ \approx 90 & \text{oncoming vehicle} \end{cases} \tag{4.1}$$

A cut-in might begin with a passing vehicle from the rear or with a parallel car moving into the next lane. To differentiate a cut-in with a similar trace to a passing vehicle, a further check is made at location p_1 or p_2 to determine if the vehicle enters the driving lane of the ego-vehicle at middle range. If a cut-in occurs at a shallow depth, as in a side collision, the front edge of the cut-in vehicle creates an inward horizontal trace over a broad span, but the back-side edge may not have formed as depicted in Figure 4.5. Additionally, the Laplacian filter $LT(t)$ detects such a strong edge in M_1 from a close cut-in vehicle. There are some example of cut-in detection results in motion profile in Figure 4.6.

We detect such interaction or motion of other vehicles in the vehicle regions. Particular lane scopes and specific directions are examined by filtering a motion profile with differential

operators and computing the gradient direction. The filters are applied in small lane scope or positions in mid-range $M_1(x,t)$, as a light processing for real-time output. The background moves outward in M_0, and are less visible in other motion profiles. For passing, passed, and parallel driving vehicles, only a side lane scope in M_1 is monitored to find their inward, outward, and vertical trace direction respectively. For cut-in, merging, and crossing events, the driving lane scope has different periods of motion profiles show the traces of surrounding vehicles and their moving directions.

For instance, passing and passed vehicles in next line leave inward and outward directions separately. Parallel moving vehicles in next line have moderately vertical follows in $M_1(x,t)$. Cutting-in vehicles into the driving lane moves internal yet end inside the lane scope at mid-profundity. In the event that a vehicle moves into the lane in close $M_2(x,t)$, it's anything but an expected crash to perform slowing down. Cutting-in vehicles have long even directions over street scope. Approaching vehicles in the contrary lane have leftward follows with a lot quicker relative speed than passed vehicles; their follows in $M_2(x,t)$ are more level. To acquire dynamic occasions from movement, we find spatial-worldly channels at unique situations after vehicle pixels are removed by semantic division in street profiles. These channels are applied in little degree or positions in $M_1(x,t)$ as a light handling for constant yield. As displayed in Fig. 4.6, channels are situated at the edge places of driving lane where cutting-in and getting vehicles may have their follows through.

To track the motion of a front vehicle, the traces of the vehicle's edges are tracked over time to determine the vehicle's size change. This tracking begins with the appearance of a front vehicle in M_1's middle depth range. We designate two horizontal windows around positions p_1 and p_2, with their spacing suited to the width of the vehicle. The windows correspond to the major edge traces in $A(x,t)$ that are near to vertical (blue). If two main traces are taken from such windows and their x locations are denoted by $x_1(t)$ and $x_2(t)$, the frontal vehicle's divergence/convergence rate is

$$S(t) = [tanA(x_2(t),t) – tanA(x_1(t),t)]/[x_2(t) - x_1(t)] \tag{4.2}$$

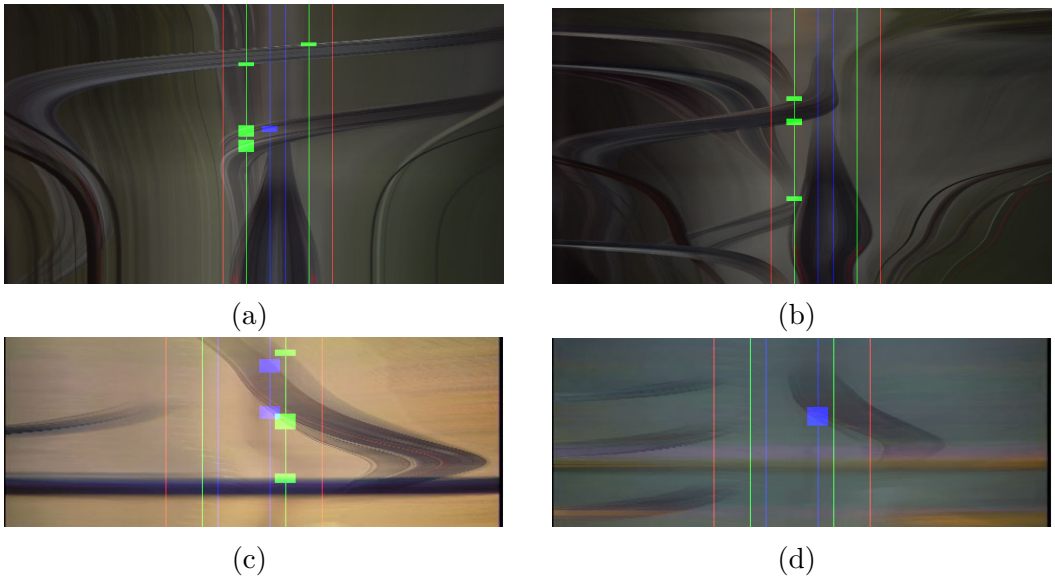


Figure 4.6. Cutting-in moments detected at driving lane in M_1 marked with small green boxes. Lane widths are colored in green lines for M_1 and in red for M_2. (a) Left-turn of an opposite vehicle and a crossing vehicle at crossing road are included. (b) Left-turn vehicles from crossing road into the driving lane of ego-vehicle are visible. (c) Right turn and cut-in of a vehicle from crossing road. (d) Cutting-in from right lane is visible. Time axes are upward.

If $S(t) > 0$, the front vehicle is approaching; if S(t)=0, the front vehicle is departing. $S(t) = 1/TTC$ has been deduced, as shown in Figure 4.7.

Although we may collect ego-vehicle motion from vehicle control data through the CAN bus, we can also detect exceptional times of halting, turning, and rapid acceleration by observing the movement of surrounding sceneries in the video and motion profile images. During a standstill at an intersection or in a traffic congestion, the surrounding environment remains largely unchanged, with the exception of some crossing traffic or cars moving on side lanes. This results in the formation of many pure vertical traces in the motion profile images M_0 and M_1. Vertical traces of this type can be identified at places $S(x,t)$ by restricting their angles $|A(x,t)|$ to a small value. By counting the number of such points at frame t, $\|S(x,t)\|$, we may assert that the ego-vehicle is coming to a halt at time t if $\|S(x,t)\|$ reaches a certain value. From Figure 4.5, at points p_1, p_2, p_3, and p_4 in L, a trace of a crossing vehicle is inspected (x,t). For cars making a left turn in the opposite lane, p_1, p_2, and p_3 are checked.

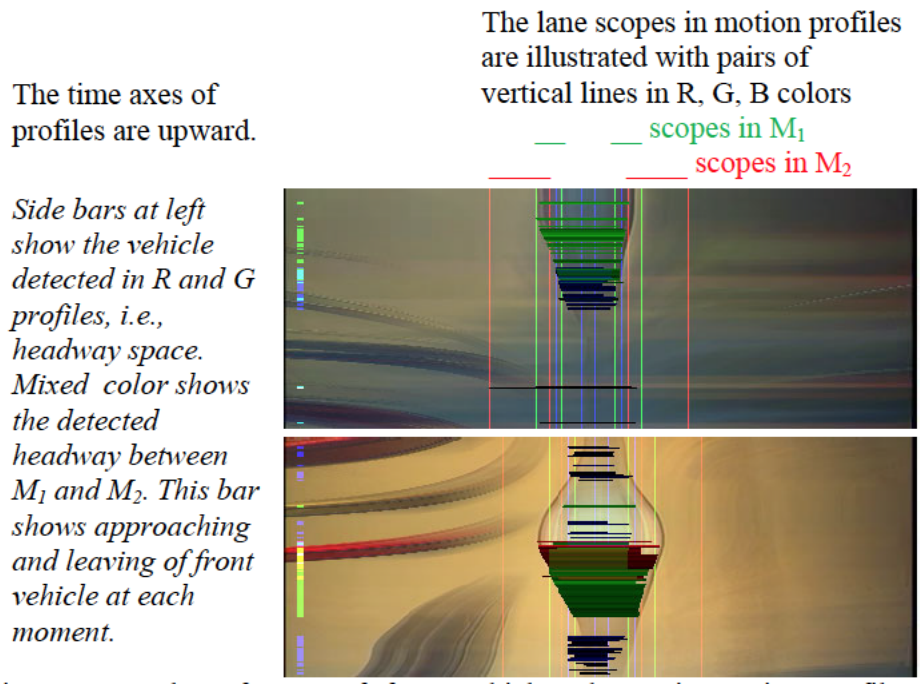


Figure 4.7. Examples of TTC of front vehicles shown in motion profiles. Filter-searching scopes are bounded by pairs of vertical lines. Color in R and G indicate two scopes in M_1 and M_2. The expansion rates of vehicle trace are computed overtime from pairs of edge filters in the scopes. A detected pair is visualized with a horizontal line. $1/ttc = w\prime/w$ is calculated and visualized by the line intensity and color indicates the detected profiles. Dark means safe TTC.

4.3 Detect Ego-vehicle Action and Other Trajectories in Motion Profile Images

The motion profile image of roadside sceneries, such as tall trees and houses, shows traces fanning out from the center. This is especially true when the ego-vehicle is in the rightmost lane. The scenes on the opposite side of the road are blurred owing to the greater distance; their traces are less visible. The quicker the ego-vehicle travels, the more horizontal the motion profile image's traces.

Other traces left in the motion profile images include sudden light changes, such as when the ego-vehicle enters a shadow area or passes beneath a bridge. The profiles of such dark and instantaneous trajectories are almost horizontal over the whole field of vision. Because of the sample zone's average of pixels, sloping road edges and patterns in the pictures are blurred. Nevertheless, certain white and yellow lane markers around the image center may

leave short but broad traces in the motion profile image. They are weaker than vehicle traces from vertical borders after blurring and so may be filtered out.

In addition to interactions, we also detect ego-vehicle actions in $M_1(x,t), R_2(x,t)$, or $M_2(x,t)$ precisely for monitoring vehicle control and movement. Such qualitative actions include stopping, turning, and lane changing. As basic parameters, we compute the flow direction $ang(x,t)$ at solid angle focuses in $M_1(x,t)$ with 5*5 level and vertical differential administrators. The bearing is gotten from the digression of inclination, where $ang(x,t) = 0$ for a follow along the time hub. Then, at that point, isolated at focus point x_0, the significant stream course on the two sides of $M_1(x,t)$ are registered at each line t as

$$mfd_l(t) = average_{x<x_0} ang(x,t) \tag{4.3}$$

$$mfd_r(t) = average_{x>x_0} ang(x,t) \tag{4.4}$$

Based on the flow directions of traces, we calculate following.

1. A stopping moment/frame contains many absolutely vertical edge points in $M_1(x,t)$ due to static traces along the time axis. We count the number of zero-flow points, i.e., $|ang(x,t)| \approx 0$, at each frame. Stopping moments are found at frames with large numbers of such points (Fig. 6.9).

2. A turning at intersection causes a major flow in the opposite direction. In $M_1(x,t)$, long traces are nearly parallel over the entire field of view in the turning period. Comparing the major flow directions on both sides at each line, if they are highly consistent and are non-horizontal as $|mfd_l(t) - mfd_r(t)| \approx 0, |mfd_l(t)| \ll 90°, |mfd_r(t)| \ll 90°$, we obtain a turning frame at t as shown in Fig. 6.9.

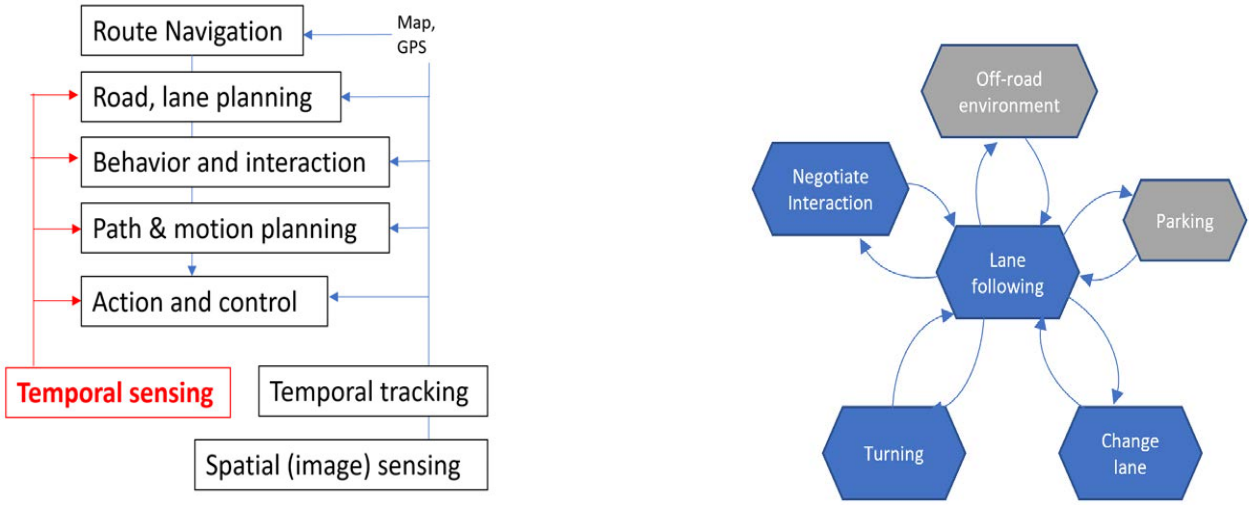
3. As a side effect in $M_1(x,t)$, instantaneous light changes generate close-to-horizontal stripes over entire field of view, when ego-vehicle goes under bridges or strong shadow. They are not danger and are ignored if $|mfd_l(t) - mfd_r(t)| \approx 0, |mfd_l(t)| \approx 90°, |mfd_r(t)| \approx 90°$.

4. For understanding a lane changing action of ego-vehicle, the central position x_0 in $R_2(x,t)$ and $R_3(x,t)$ are monitored overtime to see if a curved lane mark (pixel) goes across x_0.

5. DRIVING PLANNING FOR AUTONOMOUS VEHICLES

5.1 Driving Behaviors and Task Planning Hierarchy

One of the goals in this work targets motion planning using the temporally dense road profiles from several depths. Given the driving direction from navigation map at a qualitative level, ego-vehicle will plan the driving path and speed at middle range and hard brake at close range accordingly, to perform behaviors of lane changing, merging, turning, road following, and vehicle following. The quantitative level sensor input such as vehicle interaction and headway depths are used in the path planning and action control. Our new scheme makes short sensing and control circle with a prompt response to incoming events. This will facilitate fast vehicle motion and smoothen interaction in traffic flow. The hierarchy of motion planning and decision making are summarized in Figure 5.1.



(a) (b)

Figure 5.1. Temporal sensing supports autonomous driving. (a) Hierarchy of autonomous vehicle control and our temporal sensing architecture. (b) Behavior layer to perform different actions via state transition in automaton.

We adopt a different approach to obtain motion at surrounding directly in the temporal domain called video profile, which bypasses the current route that is to obtain shape first in individual frames and then track them for motion for driving planning. We use several horizontal lines to sample driving video at looking ahead depths to omit large view redundancy in video. Taking the key advantage of vehicle motion along a smooth path, road scenes move

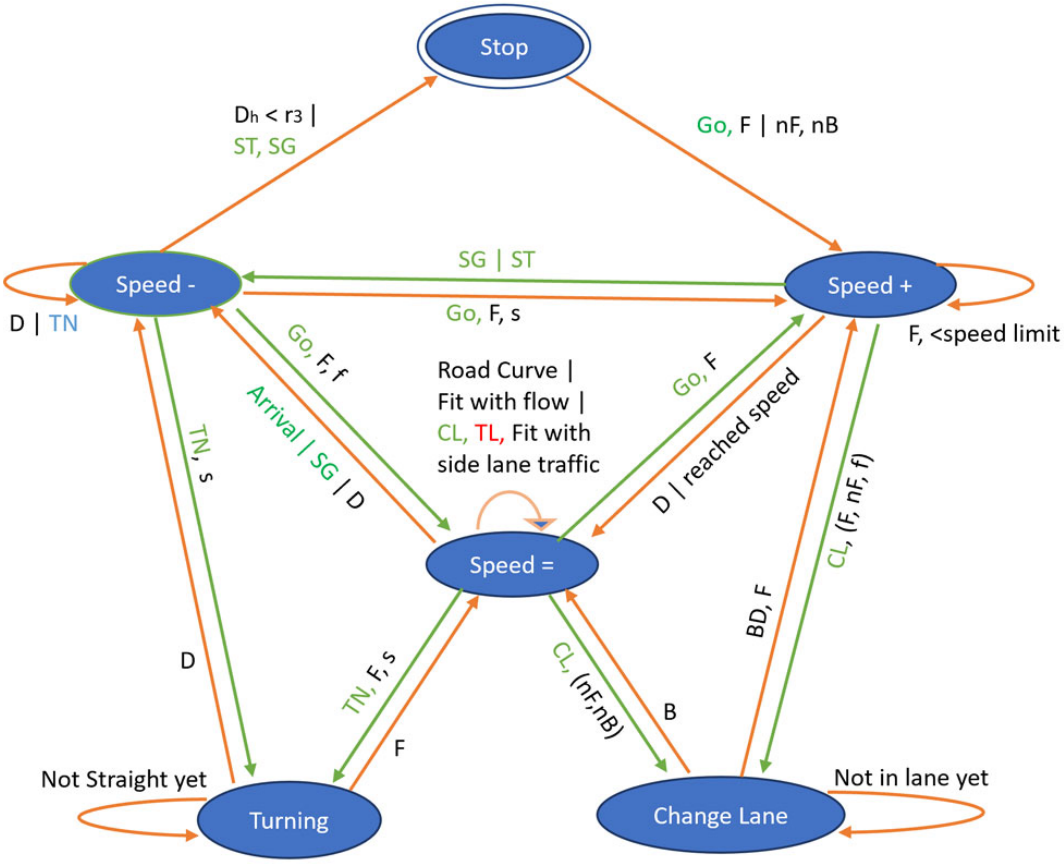


Figure 5.2. Action state transition in performing behavior tasks guided by upper level route navigation and feedback from lower level sensing. +, -, = indicate acceleration, deceleration, and keep speed, respectively.
Green transition→: from map navigation and communication "—": or. ",": and. CL: change lane. TN: turning. Go: move ST: stop sign; SG: red signal. D_h: headway space; D_b: braking distance F: front safe, i.e., $D_h(r) > D_b(V)$. D: front danger, i.e., $D_h(r) \leq D_b(V)$. s: slow ego-vehicle speed; f: fast speed nF: front safe on next lane targeted nB: back safe on next lane targeted TL: turn on turning light Red transition→: by sensor feedback

closer gradually and a line can scan road once without scene losses. Although the spatial resolution is dropped, the space between lines can be recalled from their temporal memory if lines are sampled at the maximum rate (30fps). The road scanned by a far line will be scanned again by a closer line when the vehicle approaches. The road layout is projected to a road profile in a deformed parallel-perspective projection, which provides both latitudinal shape (direction and size) and longitudinal motion in a continuous form.

Although the temporal projection is not similar to normal perspective images, it is equally difficult/easy as normal images for deep learning algorithms such as semantic segmentation to interpret based on large training sets, because deep learning deals with shape and their association, and profiles have converted spatial and temporal information in video to spatial layout. In classifying incoming lines pixel-wise to road, vehicle, obstacle, pedestrians etc. semantic segmentation has taken a history of previous moments into account to guarantee the latest output, which maintains the classification accuracy but in the minimum latency (1 frame) at video rate. Vehicle interaction such as front vehicle tracking, cutting in and merging, passing and passed in next lanes, crossing and turning, etc. have also been identified from trajectories in motion profiles without recognition in every frame. Pedestrian detection has been reached to a high accuracy purely based on the motion patterns and human width at several depths. Road area detection in road profiles under different weather and illumination conditions has been studied based on the appearances and their clustering.

5.2 Road and Lane Following with Road Profile Images

Different from the traditional path and motion planning in an infrequent rate at discrete positions, our temporally dense detection of road and headway space can respond to dynamic events promptly with updated path direction and speed. Given classified pixels belonging to different regions, it is straight forward to perform road following within the area wider than mapped ego-vehicle widths. As shown in Figure 5.3, (a) the lane scope at close line is mostly centered if ego-vehicle has been safe driving so far on road. If any object enters this scope in close depth, urgent braking or steering are performed for collision avoidance. (b) The mid depth is suitable for path planning to determine steering and speed. (c) The far-depth along a curved road may deviate from the center in $R_1(x,t)$; the heading affirms the course route from the guide. Between two depths, road regions can be gotten from the transient memory checked minutes prior in the farther road profile. A relative change is applied by the vehicle direction. At each frame, two trapezoidal regions at the three depths are detected from road pixels in road profiles along with the image widths of the ego-vehicle for path finding. A path with minimum curvature is generated through three key points of trapezoidal regions.

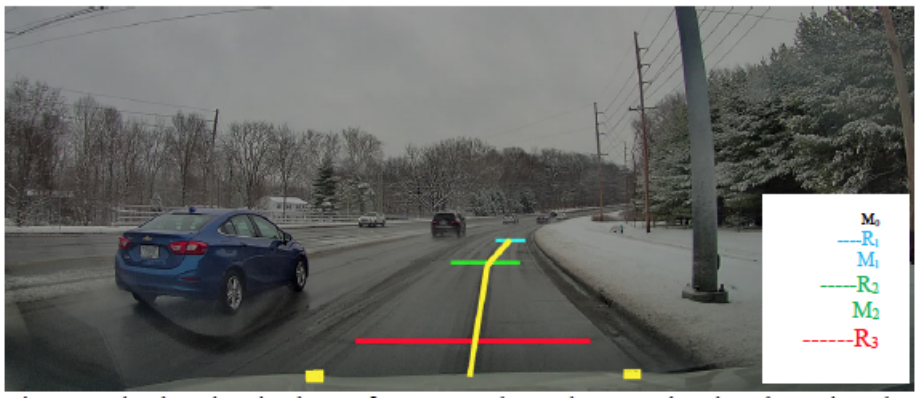


Figure 5.3. Path planning in dense frequency through control points by using the calibrated vehicle widths (red, green, blue) in the road profiles. This yields a smooth path up to three depths at each moment.

Frequently, a free road region is not able be found from the far road profile due to street end, stop sign, halting line, signal, turning intersection, ego-vehicle is ready for a full stop. To stop from current speed V, ego-vehicle continues to slow down as depth is detected in order at r_{i}, i=1,2,3, i.e., for headway length $D_b(V)$, V has to satisfy

$$D_b(V) \leq r_{\mathrm{i}}, \mathrm{i} = 1, 2, 3 \tag{5.1}$$

Figure 5.4 shows the semantic segmentation results where road, off-road (road edge), lane marks and vehicles are labeled at pixel level in different colors. For path planning, each road profile obtained from the semantic segmentation has one red line and two black line in the ego vehicle lane generated in the path planning. The red line shows the center of ego-vehicle in the path planning, and black lines are ego-vehicle margins representing the width of ego-vehicle at that depth. The margins and widths at three depths have been calibrated once in pixels for flat roads. To ensure safety of vehicle passing, the algorithm makes a interval (safe gap) between black lines and lane marks in the road area, which will let ego-vehicle drive in middle of the lane and avoid obstacles.

If the left lane mark of the ego-vehicle lane is clearly observed, the algorithm use the left lane mark as reference point. Left vehicle margin is on the right of left lane mark with safe gap, while keep the vehicle width within the free road area.

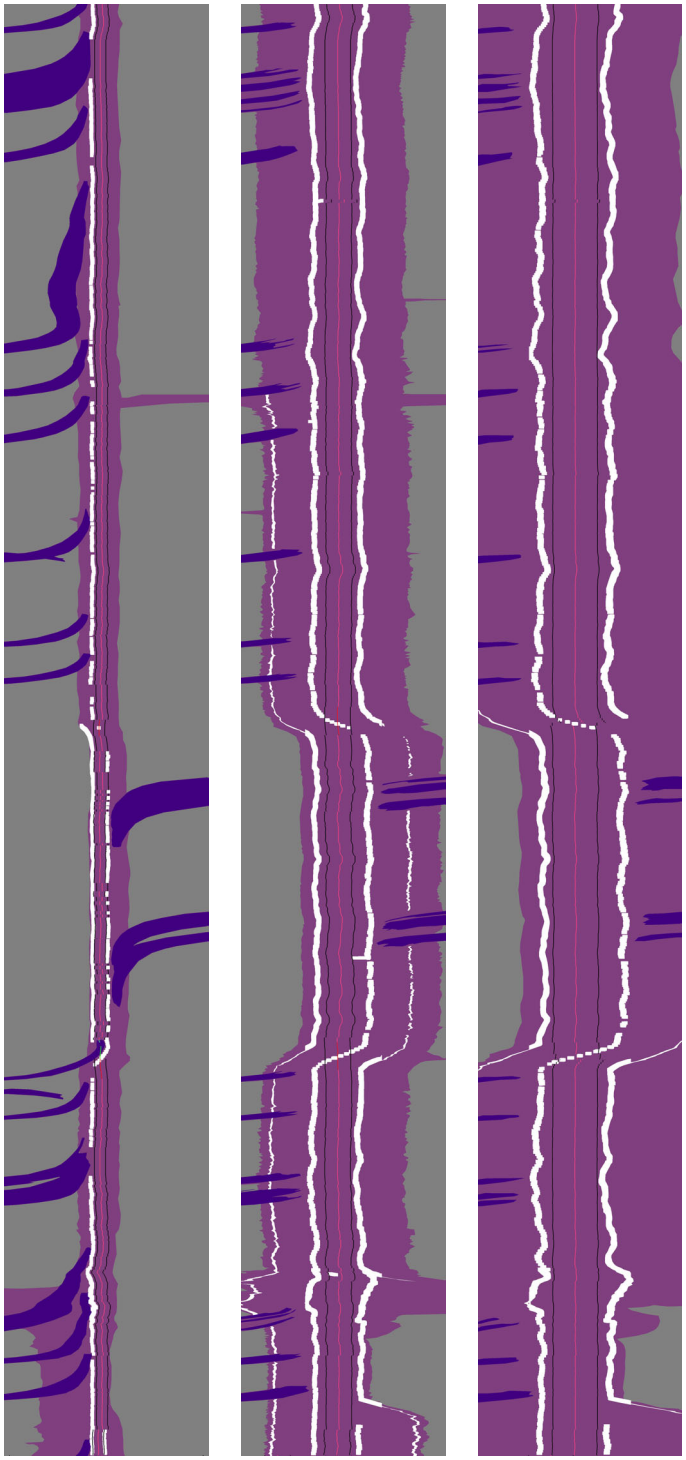
If the left lane mark is not clear due to a low visibility, occlusion of obstacles, or an invading vehicle, while right lane is clear in the road area, the right vehicle margin is planned on the left of right lane mark with a safe gap, in order to let ego-vehicle passing on the road area. If no right lane is visible, right road edge is used for the path planning of that depth.

If neither lane marks or road edge are visible in road area around the given direction provided by upper level navigation module, which means the road is wide at an intersection or parking lot, the center line and margin line follow the last frames.

If a road is not visible at a road end, T-junction, sharply bending road, or occluded by a front vehicle, ego-vehicle will slow down and the path planning shrinks to a shorter depth, as the vehicle following state discussed in next section. In such a case, the vehicle occupied width is not possible to be captured within the road area in the road profile, given the direction of navigation from high-level module.

For a turning at road crossing, mid and close depths are used for path alignment, even if far road is visible (ignored). For a lane changing, mid depth path is already located in the scope of next lane. Figure 5.4 also shows the planned path points at three depths for every moment, based on the semantic segmentation results of road and motion profiles provided from Guo Cheng [39]. Lane changes are observed and many vehicle-passing happened when ego-vehicle stayed on the right lane. Trucks were passed when ego-vehicle took over from left-lane. The interrupted path in far road profile is caused by plan change due to a low visibility and dynamic traffic. The mid-range path follows planned direction at far range, but keeps the vehicle width in lane. The path at close range is mostly in lane because the plan will keep driving continuous and smooth.

On the other hand, stop signal and stop sign are provided via other channels from digital map, GPS, V2I, or vision method. For a turning action at an intersection, the mid and close lines are used to track road margin on the turning side, after far and car turning car lane change turning mid lines exceed the intersection first. The turning path will keep the vehicle width and tire positions (yellow boxes in Figure 5.3) inside the street region.



(a) (b) (c)

Figure 5.4. Path planning results on a highway for 5min visualized on the road profiles from far, mid, and close depths. The time axes are upward. At each moment, the planned vehicle center and margin positions are displayed in red and black curves at three depths. White lane marks are widened for increasing connectivity. The green line is the period when a front vehicle is detected or road becomes invisible.

5.3 Vehicle Speed Planning Based on Headway Distance

5.3.1 Braking distance of vehicle speed for safe stopping

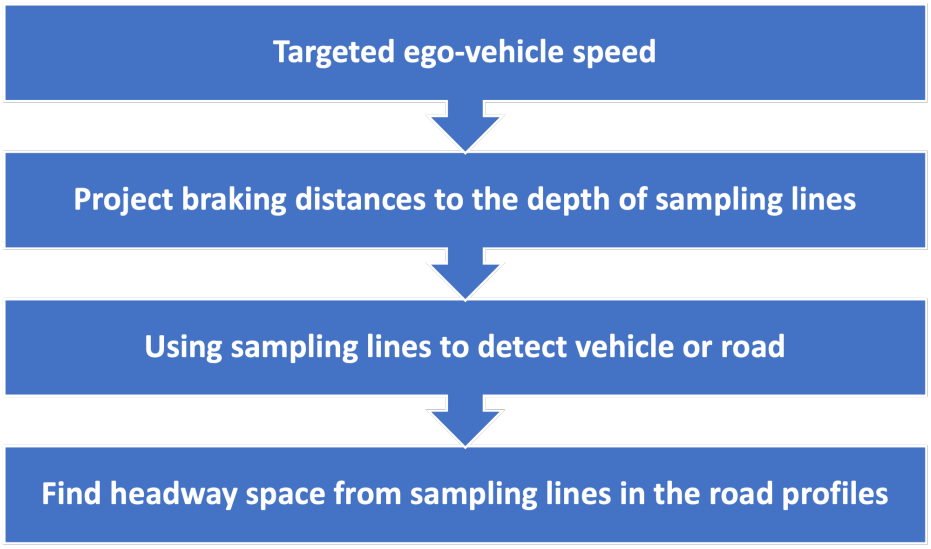


Figure 5.5. There is the workflow for speed planning control architecture.

Figure 5.5 shows the diagram of this process. For each vehicle, its braking distance is fixed as a function $V^2/\mu g$ for a particular speed V, where μ is the coefficient of friction and g is the gravity. If a constant deceleration is put on, V is reduced evenly. Assume front vehicle speed is S_{slow}, S_{slow}=0 if it stops like other obstacles, pedestrians, etc. The safe distance for ego-vehicle to fully brake is

$$D_h > B_v(V) \tag{5.2}$$

where D_h is the free detected headway distance from a series of road profiles. If a front vehicle is also moving forward, i.e., $S_{slow} > 0$, headway D_h can be shorter than $B_v(V)$,

$$D_h > B_v(V) – S_{slow}\tau(V) \tag{5.3}$$

where τ is the time for ego-vehicle to fully stop and is also a function of V. Because S_{slow} of front vehicle has a large variation from steady speed, sharp braking (braking light on),

to sudden accident, we will ignore the capacity of the second term. Thus a safe speed of ego-vehicle is

$$S < B_v^{-1}(D_h)\ (D_h)^{1/2} \tag{5.4}$$

which can be illustrated in Figure 5.6 with three road profiles and the motion profile from the horizon in the image frame. More profiles up to 10 can increase the granularity of headway distance to refine speed levels for smooth traffic flow.

5.3.2 Headway distance detection from Road Profile Images

The camera is set at windshield with a normal height. In the camera view, the horizon is specified first. Below the horizon, multiple (three in our illustration) horizontal sampling lines are set at different heights in the frame to scan the ground at far, mid, and close depths (e.g., 8m, 18m, 35m). Through these lines, slanted planes of sight cut the space ahead the camera. An object with a reasonable height (e.g., above bumper) will be cut by at least one plane of sight when it is closer than the farthest line. The object is sampled constantly at every moment rather than once. This is also true for a static object of the same height (stopped car, roadside pole apply). Object lower than bumper height can be treated as static road fixture such as lane mark, curb, etc. These scenes are scanned once by a sampling line, and three times by three lines at different depths.

As shown in Figure 5.6, we preset several targeting speed of vehicle as S_{slow}, S_{mid}, S_{fast} and $S_stop = 0$ is vehicle stopped status. For these speeds, their braking distances l_1, l_2, l_3 are calculated and we set the sampling depths of road profile images R_1, R_2, R_3 there, respectively.

At every moment, a sampling plane of sight intersects with an object closer than its sampling depth (dotted line) in Figure 5.6. This will be captured in the recognition of road profile image corresponding to that depth. In Figure 5.6, solid lines are depicted on the objects when they are intersected by the slanted planes of sight. A close object may be cut by multiple sampling lines on its surface (displayed in solid lines around vehicles). The closest plane of sight determines the headway space. As seen in Figure 5.6, free headway

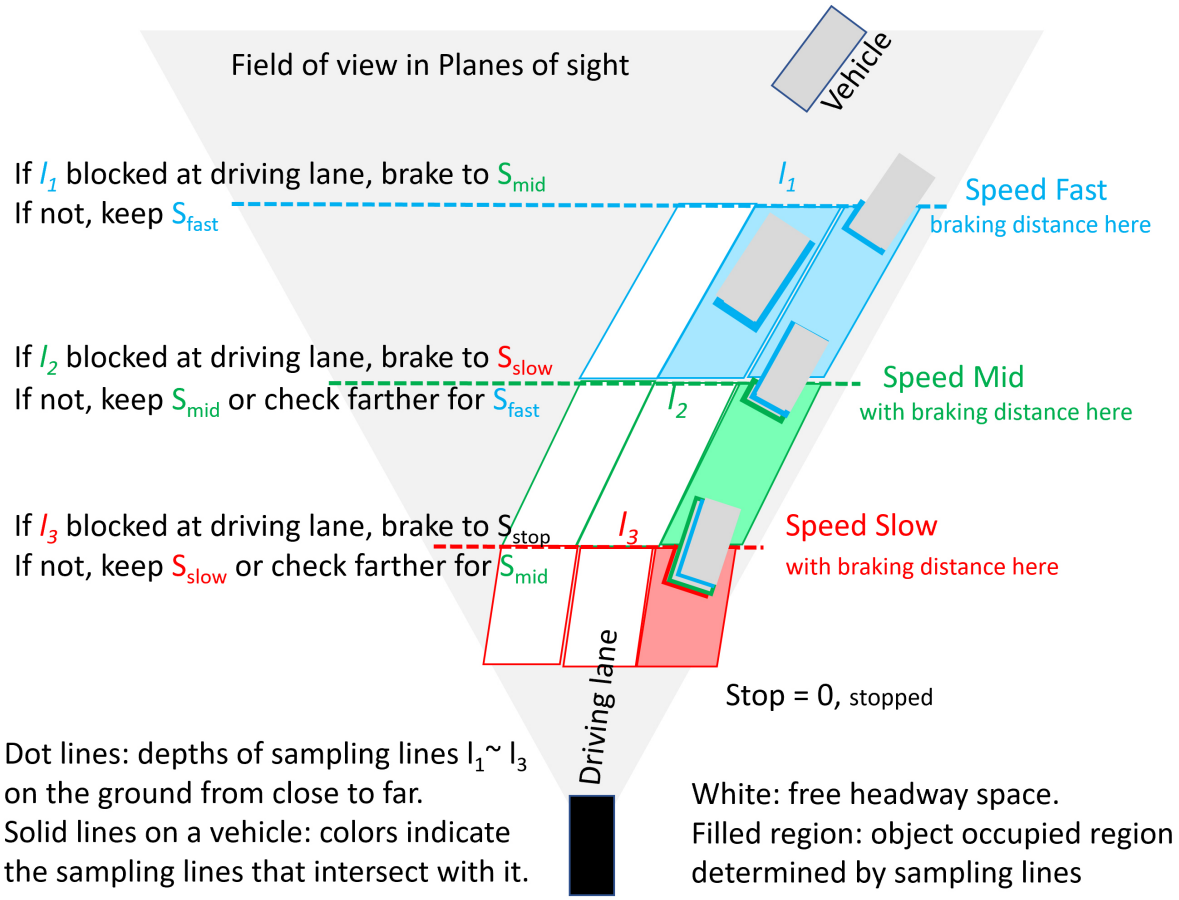


Figure 5.6. Block-wise headway space for speed planning at each moment. The vertical axis is the distance in the camera covered field of view. The road can be curved to some extent. Braking distance for each speed is pre-calculated for several speeds $S_{stop} \sim S_{fast}$ from low to high. Sampling lines $l_1 \sim l_3$ are set at the braking distances $r_1 \sim r_3$ of these speeds from close to far for detecting objects. We adjust vehicle speed toward one of $S_{stop} \sim S_{fast}$ consecutively according to the detected headway space. More levels of speed can be set with additional sampling lines (e.g., 10 lines) for finer speed control.

blocks are displayed in white and vehicle occupied blocks are colored by their ranges (red, green, blue, gray etc.)

With the latest lines captured on the road at far, mid, and close depths, vehicle path planning obtains free area at these depths, as already described in previous section. They are scanned at denser video rate (30 fps) than discrete images (e.g., 10 fps) due to the small data size. For objects on the road including dynamic vehicles and static obstacles, they are

sampled at every moment by at least one sampling line if they are closer than the farthest line. Their dynamic headway $D_h(t)$ at every moment are captured by sampling lines.

5.3.3 Dynamic speed planning responding to detected headway space

Let us focus on a front vehicle following by ego-vehicle to avoid collision. Vehicles further ahead may be occluded so that the farther profiles captured has the same vehicle but at a higher position. We flag the lowest line (closest road profile) where a front vehicle is detected. The headway is

$$D_h = min(d_{\mathrm{i}} | R_{\mathrm{i}} \in vehicle), \mathrm{i} \in [1, 10] or [close, mid, far] \tag{5.5}$$

where road profile i detects a vehicle region. If ego-vehicle speed V is known from CAN, the TTC to front vehicle can be obtained as $D_h/(S - S_{slow})$. A vehicle-following action is to follow the target in a lower or equal speed with a longer headway than braking distance. As the target speed is unknown, ego-vehicle will adjust speed according to equation 5.4. To avoid frequent braking and accelerating, an additional length as green trace (Figure 5.7) can be added to the braking distance in the vehicle following.

Table 5.1 is an example of step-wise speed transition between discrete levels based on headway distance in following traffic flow. Four typical speeds, S, M, F, and H, are selected evenly at 20, 35, 50, and 65 mph and their braking distances are referred to (0.7 as friction rate) in setting the sampling line in video frames for road profiles. The monitoring headway distances for frontal vehicle are approximately at 5m, 16m, 35m and 60m. The ego-vehicle can maintain a safe speed at S, M, F, or H, if such a headway space is kept. The farthest motion profile M_0 is from the horizontal zone covering the horizon in the image for identifying vehicles up to 60m; narrow traces of far vehicles show up in $M_0(x, t)$. Farther than this distance, vehicles may be blurred and invisible in $M_0(x, t)$ after the pixel averaging to obtain $M_0(x, t)$. We detect narrow traces of far vehicles by semantic segmentation on $M_0(x, t)$.

Different from a full braking process to stop on road eventually, a slow-down action may prevent further approaching of a front vehicle, if the front vehicle keeps moving forward. The ego-vehicle speed is then maintained properly and adapted frequently for front vehicle staying

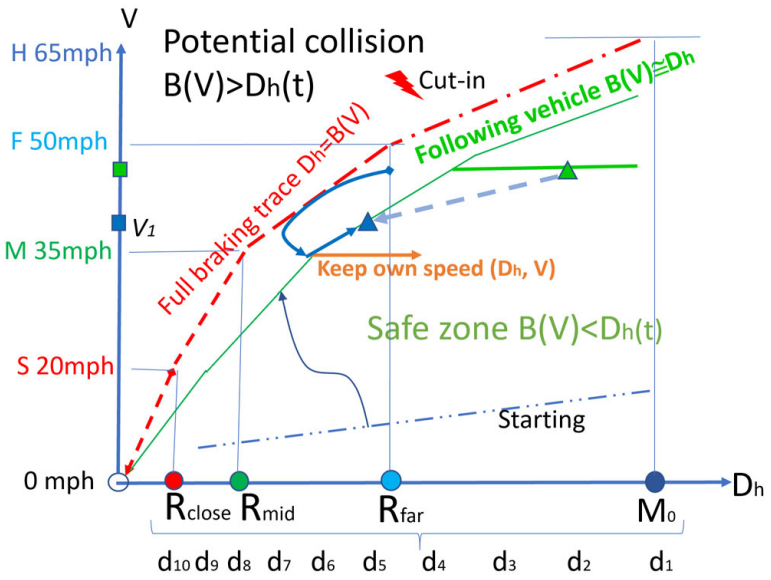


Figure 5.7. Speed transition based on headway space. The red path is the speed change of full braking to stop from all speeds. The vehicle in green area has safe headway. The vehicle will brake speed to a lower level if its state shifts on to the red path. If headway rebound again due to a leaving front vehicle, ego-vehicle speed tends stable as the example depicted in dark-blue trace.

away from safe depths. In Table 5.1, the speeds in left-upper corner are ideal speeds for sensed headways. Anti-diagonal cells can speed up as a headway distance extends. Certainly, a strategy not to speed up aggressively can be taken to avoid frequent speed changes. For a stopping before hitting an obstacle, traffic jam, or before stop line, the right-down cells without urgent braking and collision perform step-wise deceleration down to stop before collision. Urgent braking causes a collision if the front vehicle does not move away. On a highway, headway can be kept shorter and more tolerant (italic in Table 5.1) in busy traffic flow because no sudden stop is allowed; otherwise, the status is informed by front vehicles via braking lights.

Ego-vehicle will prepare a full stop at road end, stop sign, red signal, turning crossing, or traffic jam, required from navigation task or sensor output, unless free road area reveals again in the vehicle-following mode. Stop sign position and signal can be received from a digital map, GPS, V2I, or a vision method. For turning action at an intersection, the mid

Table 5.1. Speed transition depending on headway distances
d_h: headway, =: Keep a speed, ↑: Speed up to a speed, ↓: Slow down to a speed, U.B.: urgent braking; Italic: highway

	Slow S 20 mph	Median M 35 mph	Fast F 50 mph	Highway H 65 mph
Braking Distance	4.6m	16m	35m	57m
No objects detected	↑ M	↑ F	↑ H	=H
$M_0 d_h \leq 60m$	↑ M	↑ F	= F	=H or ↓ F
R_1, at far line $r_1 d_h \leq 35m$	↑ M	= M	↓ M	↓ F
R_2, at mid line $r_2 d_h \leq 16m$	= S	↓ S	U.B.	↓ M
R_3, at close line $r_3 d_h \leq 5m$	Stop	U.B.	Collision	Collision

and close lines are used to track road margin, after far or mid sampling lines have exceed the intersection. The turning path will keep the vehicle width within road or lane as well.

6. EXPERIMENTS AND DISCUSSION

6.1 Naturalistic Driving Data Processing and Online Database Construction

The experiment used a subset of Naturalistic Driving Video (NDV), which was taken from various types of roads (highways, urban, rural, residential roads, etc.) on 110 cars in different seasons, weather and time for more than a year. The cameras on all vehicles are of the same type, so they respond the same to external brightness. The video with a resolution of 1280×720 pixels is sampled at a rate of 30 frames per second, and the high-definition camera has an automatic exposure function. In this study, a 2TB video subset was uniformly selected for each date and time. The horizon in all segments is manually determined, and road profiles are generated from 5800 segments; each segment has a length of 5 minutes (9000 frames or position). The profiling is performed offline, so it will not affect the real-time classification of weather and lighting using K-means, decision trees, and sparse coding.

In order to retrieve and annotate driving videos, an interface was developed based on PHP, SQL database and distributed storage system. Videos can be searched by the following attributes: road and off-road materials, weather and lighting conditions, and video ID number. Visualization of each individual video provides key frames of the video, MPEG format video, road overview, the videos attributes, the intermediate and final results of the calculated features, and classification result of weather and road edge. By mapping a video clip to the road profile, we have achieved effective video inspection and annotation, which has been impossible to achieve so far. The assessment of weather recognition is also displayed directly on the time axis, together with the road overview, to reduce the number of frames of work. The profiles are generated by Python program using NDV. The vertical axis of the profiles in every frame are depicted in chart 6.1.

In our data mining of weather and illumination, the K-means algorithm on Matlab starts with the average number of categories marked by humans. The human marking of the weather produced 360 relatively uniform segments, and the vehicle moved smoothly. The 9 features obtained in 27k frames are input into K-means clustering, and the clusters are obtained in about an hour. It turns out that although the K-means algorithm tends to

Table 6.1. Vertical-axis of road profile and motion profile in frame

profile name	vertical-axis
potion profile 0(M_0)	from horizon to horizon + 35 pixels
motion profile 1(M_1)	from horizon + 35 to horizon + 100 pixels
motion profile 2(M_2)	from horizon + 100 to horizon + 200 pixels
road profile 1(R_1)	horizon + 50
road profile 2(R_2)	horizon + 150

give different clustering values, because our sample data becomes such a large set, they are close to similar and stable clusters. Therefore, we manually selected 360 video clips with low traffic volume, and there are road edges at appropriate positions in the video. The curb on the left is brighter than the curb on the right, because the area on the left may include the opposite road, separation zone, central separation zone, etc. Therefore, the difference on the right is more reliable for road edge detection. At night, the road surface illuminated by headlights or street lamps has high chromaticity from these light sources.

In our data clustering and classification, we have identified several factors. As the number of clusters increases, on the one hand, the classification rate becomes very low. On the other hand, classification with a small number of clusters (such as three clusters) has less impact on the road detection algorithm, although its accuracy is higher than seven clusters.

By calculating the Euclidean distance of each cluster for classification, there is no need to spend time to identify the weather and light at the frame level, which is sufficient for the real-time system as a pre-stage of other processing. Decision trees also don't need to spend time for the input to pass through several feature checks at the root to reach the leaves. For K-NN, the calculation time on Matlab is also very small. This means that the method we study here is sufficient for real-time automotive in-vehicle applications.

6.2 Road Edge Detection Under Various Road and Illumination Conditions

We apply data mining approach to driving videos captured over a year. Various types of roads including urban, rural, residential roads and highway are scanned through all seasons . To test road edge detection results, we use the same HD driving video database but different clips with all weather/illumination. The experiments have been carried out on 150 randomly

selected road profiles (each 9k frames) along with the extracted sequences of weather, linearity, and homogeneity. Their weather/illumination types are classified first so that we can select features in road edge detection accordingly. The weather classification has reached a good accuracy to apply the maximum likelihood to the limited number of candidates by for a final road edge. For the remaining error in the classification of illumination, if the class is adjacent to the correct class, e.g., the class often uses the same features in identifying road edges, we continue the process and evaluate the final position of detected road edges.

Road edges are detected correctly under the bright illumination (Figure 6.1), which is not much different from what other works have achieved. Figure 1.4 shows the curb detection based on shadow/shading on concrete surfaces. Sunny facing the sun: the illumination makes bright road surface and background scenes in shadow(Figure1.4b). The chroma is small because of the gray-to-white road and shadow on objects. The intensity change is relatively large.

Figure 6.1. Road edges detected in road profile without using lane marks. Red bounded by two side green points are marked at road edges. High accuracy results in output outside road shoulder. Both passing vehicles (with the close-to-horizontal traces) and the road edges are located almost perfectly.

Dark road surfaces: for raining, shadow, and dark lit, the accuracy of road edges are much lower. Road edges under dark lit can still be detected correctly because there is less interference from bright sky as shown in Figure 6.2. On the other hand, direct light has brighter edges at highlight than road edges. For example, sunlight causes unstable intensity, dirty windshield patterns, and highlight on road, which largely reduce the accuracy.

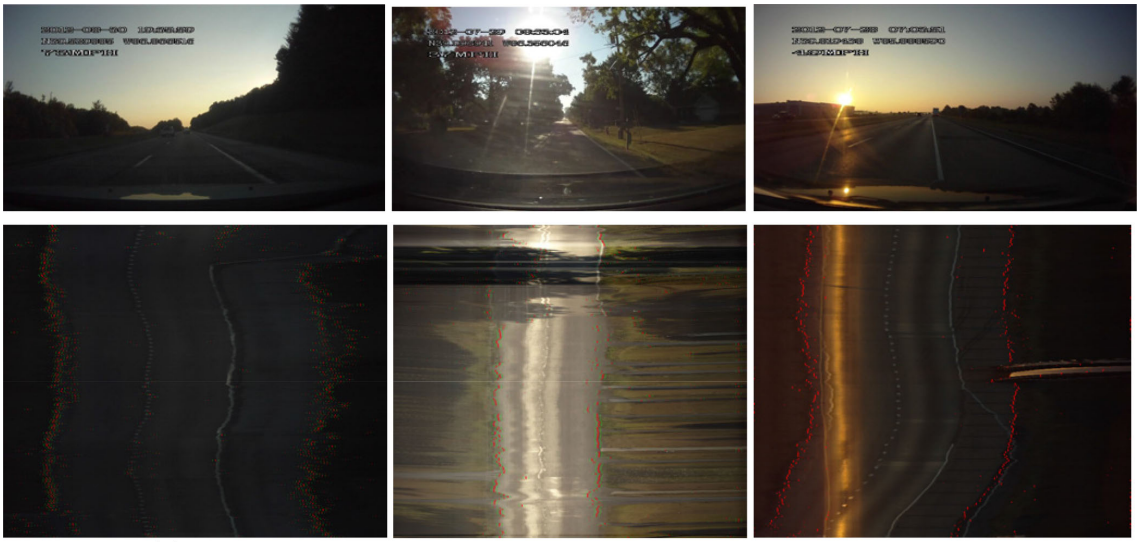


Figure 6.2. Detect dark road stably by our algorithm. (a) Dark lit roadside barely visible from intensity changes. The detected positions out of three lanes are marked correctly in red. (b, c) Direct light and detected road edge marked in red. (b) is closer to SunnyFS with shadow of tree. (c) is closer to dark lit case.

Night and invisible roadside: Because of the limited scope lit by vehicle headlight, road edge in dark and rural area is unable to be located if road is wide (Figure 6.3). We count such an edge safe since it does not reach the real edge at wide end. Wet urban roads at night are a problem since the traffic and street lights have a long reflection on the road, which causes much confusion.

Raining and wet road: rain has a lower illumination than cloudy. If road is less wet, the algorithm performs correctly as cloudy case (Figure 6.4(a)). When a road accumulates much water, road area reflections on manifold vertical edges from roadside buildings, poles, street lights, and vehicle lights. Highway with fewer high buildings in background is less affected. Our linearity computation from non-vertical lines reduces such influences. However, the most annoying case is a wet urban road at night (Figure 6.4(b)), where road edges have weak intensity and linearity.

Snowing and snow-covered roadside: we found the intensity on-road lower than off-road because roadsides always get snow covered earlier than road surface (Figure 6.5). The

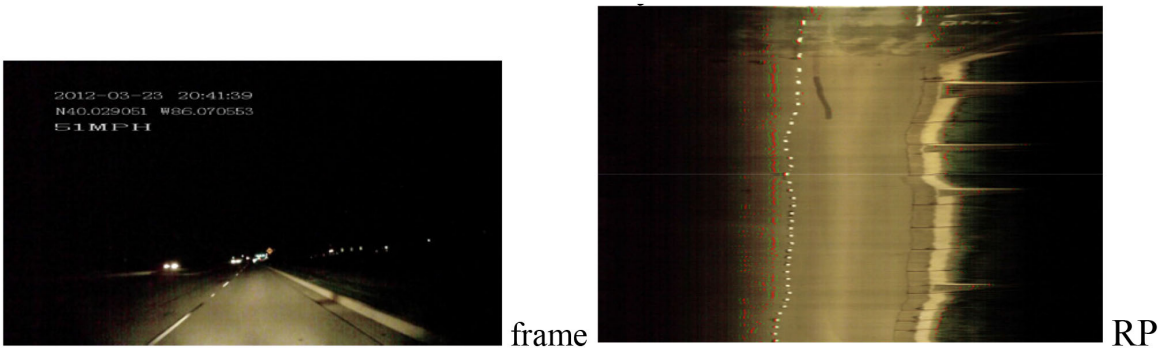


Figure 6.3. If road edge on left side is outside the field of view, the margin of headlight scope is marked as road edge. If a road edge is visible in a close position as the right side, it can be detected correctly. The algorithm only detects close road edge within the lit region.

accuracy in snow is lower than sunny, even if it has bright environment, because some snow remains on road. Linearity is not always available if the snow has not melted yet.

The data mining takes time off-line to investigate videos from the extraction of road profile to the feature detection. This also include extracting nine features of intensity and chroma in the selected regions in a frame for weather identification, median filtering to remove lane marks, and computing distributions of clusters for different weather and illumination.

The second stage is to detect road edges in near real-time, which includes (1) road contour sampling, extracting nine features together with gradient and contrast, (2) weather and illumination classification and recognition, and then (3) through one-dimensional filtering and selection Carry out road edge detection. The time to use Matlab is longer than the time to play the video. A 5-minute video has 9000 frames, it takes 369.766 seconds, and each frame takes 4ms (the advancing time of the video frame is 3.3ms). Even if more modules are triggered to deal with the ambiguity in light recognition (for example, the first three weather categories), the time will not increase much due to the common filtering.

Table 6.2 shows the road edge detection accuracy under different lighting conditions. Each lighting cluster was tested for 27k frames in the configuration file, which is much longer than the previous method. We calculated the number of frames where the curb was correctly positioned to determine the accuracy of the road profile length. This boundary-based standard requires that the detected edge differs from its true position by 5 pixels

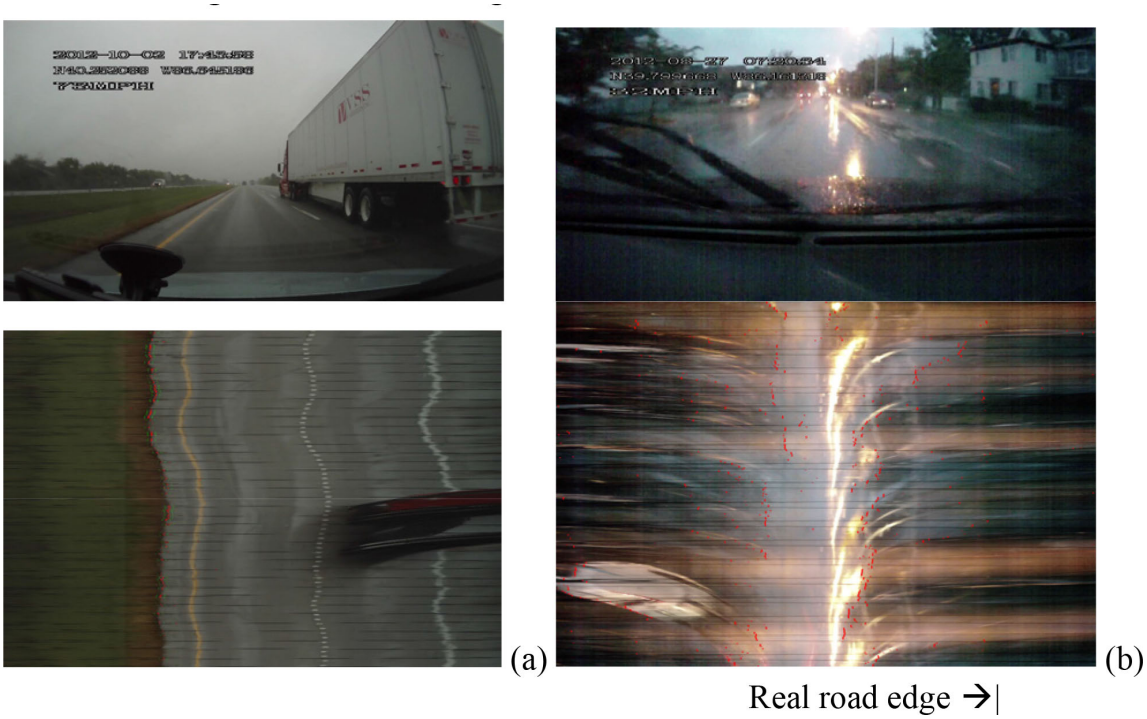


Figure 6.4. Raining weather. (a) Highway not very wet yet.Wiper motion visible as black horizontal lines in the road profiles but has little influence. (b) Raining on wet road with light reflection at night. Detected edge are wrong frequently.

(0.4% in high-definition video), which is more stringent than the area-based measures used in semantic segmentation.

For good light conditions, such as sunny and cloudy days, our results have about 90% accuracy. For lighting that cannot show clear roads, such as wet roads at night, direct light, and road edges that are not illuminated by car headlights, there is currently no work report experiment and accuracy. The overall accuracy rate is 78%, and we have broken down the lighting categories to understand these difficult situations. We report that in the case of insufficient lighting, the detection rate is very low; a false edge may cause the vehicle to run off the road or stop the vehicle.

Our method can also consider the temporal and spatial continuity along a longer path instead of local linearity to produce a more reliable road edge. This has been achieved by median filtering the X positions of the edge points along the road contour, where most X positions will win in each time-space window. As shown in Figure 6.6, the local outliers in

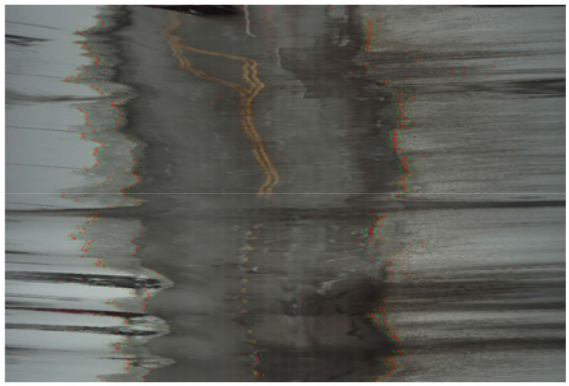

frame

RP

Figure 6.5. Snow-covered edges have poor linerity, but are brighter than drive way.

the road contour are ignored. The wiper movement in the rain is completely removed. As shown in Table 6.2, the accuracy rate of road edge has increased by 79% on average to 87%, although it may miss a short section of imperfect road structure. In addition, if we further understand the width of the road or the lane in which the vehicle is traveling, we can further improve the stability of road detection.

Table 6.2. Accuracy of road edge detection and improved accuracy using temporal continuity (ATC).
We count the frame numbers when the road edges are located correctly among the entire road profiles. Each weather cluster is tested at 27,000 frames in the profiles.

Weather	Night	Darklit	Direct light	Rainy	Cloudy
Accuracy	90.58%	81%	63.63%	54.36%	88.04%
ATC	93.56%	91.64%	75.5%	65.58%	92.22%
Weather	Foggy	Snowy	SunnyFS	Shadow	SunnyBS
Accuracy	84.53%	79.42%	91.08%	70%	87.03%
ATC	88.82%	89.87%	95.04%	84.39%	92.78%

6.3 Semantic Segmentation of Road and Vehicle Motion

This work has used semantic segmentation of road profiles resulted from driving video [39]. An open-loop path and speed planning are carried out then and the result is overlapped with with driving video to check the planned ego-vehicle activities along the road and the interaction with other vehicles. Although the planned actions have not down to mechanical

Table 6.3. Accuracy of semantic segmentation on different weather

Weather	#images	Boundary Acc.	Area Based IoU
rainy	6	0.84	0.79
shadow	5	0.81	0.895
sunny	5	0.83	0.892
night	4	0.80	0.81
cloudy	5	0.84	0.882
foggy	3	0.86	0.86
snow	3	0.81	0.81
average		0.83	0.87

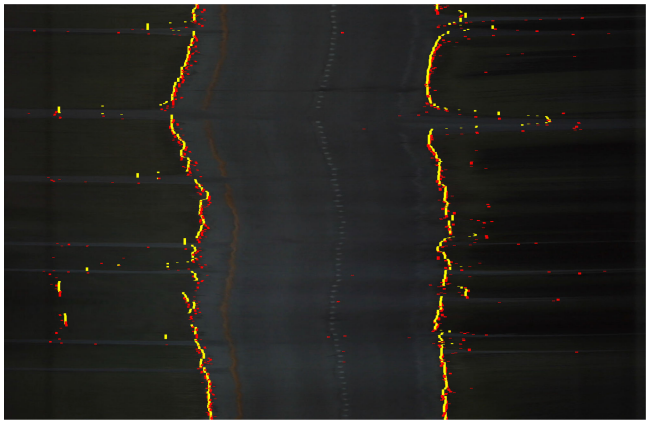


Figure 6.6. Improve road edge accuracy with temporal continuity. Red: detected road edge. Yellow: improved road edge. Many isolated red points are removed.

levels of steer/choke/brake to control an intelligent vehicle, we affirm the right choices in the planned path and speed. It is nothing more than a driving instructor observing new driver's activity through his own view, which is the video in our case. The detecting and planning yield in 30 fps output with zero-latency. According to the result[39], the accuracy of semantic segmentation is sufficient for path planning.

The semantic segmentation is the key sensing component has reached the accuracy shown in Table 6.4 by referring to previous 32 frames (about one second video). It is considered as equivalent accuracy as using spatial images. Learning images and temporal profiles have the same degree of difficulty as long as sufficiently large training sets are provided, because the video profiles have converted motion and position in video to a spatial layout without missing critical driving information.

Table 6.4. Accuracy of semantic segmentation on different object provided from Cheng, et. al [39]

True+ \ False+	road	roadside	vehicle	lane mark	stopping	vehicle objects	pedest -rian
road	0.96	0.03	0.005	0.001	0.004	0	0
roadside	0.02	0.97	0.002	0.001	0.007	0	0
vehicle	0.04	0.05	0.87	0.001	0.03	0.007	0.002
lane mark	0.1	0.03	0.02	0.83	0.018	0	0.002
stopping	0.038	0.027	0.011	0.002	0.92	0.001	0.001
vehicle objects	0.01	0.009	0.03	0.001	0.01	0.93	0.01
pedestrian	0.018	0.005	0.032	0.003	0.03	0.002	0.91

6.4 Detecting Vehicle Interactions in Driving Videos from Motion Profiles

The Naturalist Driving Videos of 1280×720 pixels have been used and a subset generates motion profiles. The field of view covers 120 degree horizontally, and this covers about four lane width in the middle range depth. Three motion profiles are sampled at [0, 35], [35, 100], [100, 200] pixels below the horizon after it is picked in each video clips. Each five-minute of driving video yields 9000-pixel long motion and road profiles, notated as $M_{\mathrm{i}}(x,t), \mathrm{i} = 0, 1, 2$, and $R_k(x,t), k = 1, 2$. The video set contains various types of road including rural, urban, highway, and local roads. Output events extracted are saved in labels as shown in Figure 6.7. The data reduction from driving video to profiles make the computation much faster than the traditional methods using vehicle detection, tracking and motion estimation. The understanding of temporal process is turned to the identification of spatial relation of trajectories in profiles.

Ego-vehicle is driving in its own lane. The approaches of frontal vehicles are detected correctly in ego-vehicle's lane. If a pair of inward traces on left and right sides of ego-vehicle's driving lane, it means a front vehicle leaving away from to ego-vehicle. Oppositely, a pair of outward traces on both sides means a front vehicle approaching closer. The TTC can be computed from the traces in the colored areas by using (2). Crossing vehicles at intersections are detected by the traces through margins, i.e., p_3 and p_4. If a vehicle trace from left lane

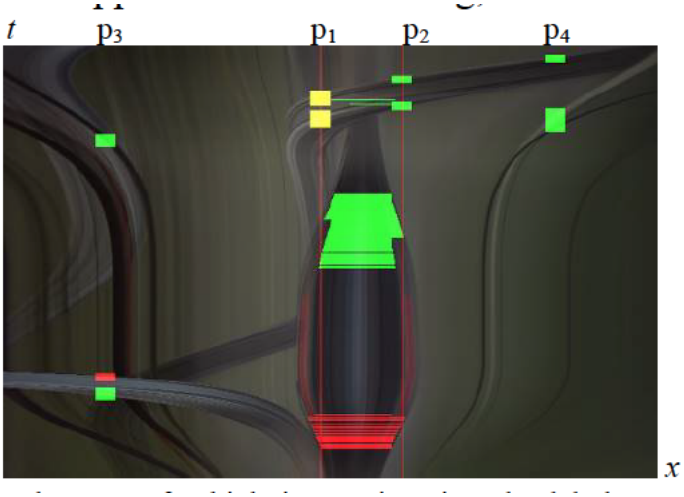


Figure 6.7. Detected events of vehicle interactions in color labels. Vertical orange lines at p_1 and p_2 are the ego-vehicle driving lane. The red area at center means front vehicle approaching to ego-vehicle, and the green area means leaving of front vehicle. At p_1 and p_2 as well as p_3 and p_4, yellow and green boxes are inward are outward traces respectively. The crossing trace from left to right is a right turning vehicle from opposite lane.

through center to right, the vehicle is turning (its) left from opposite lane at crossing, as shown in Figure 6.7.

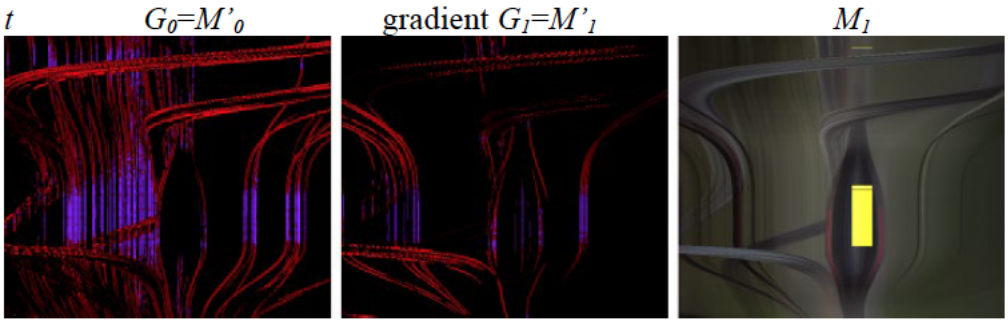
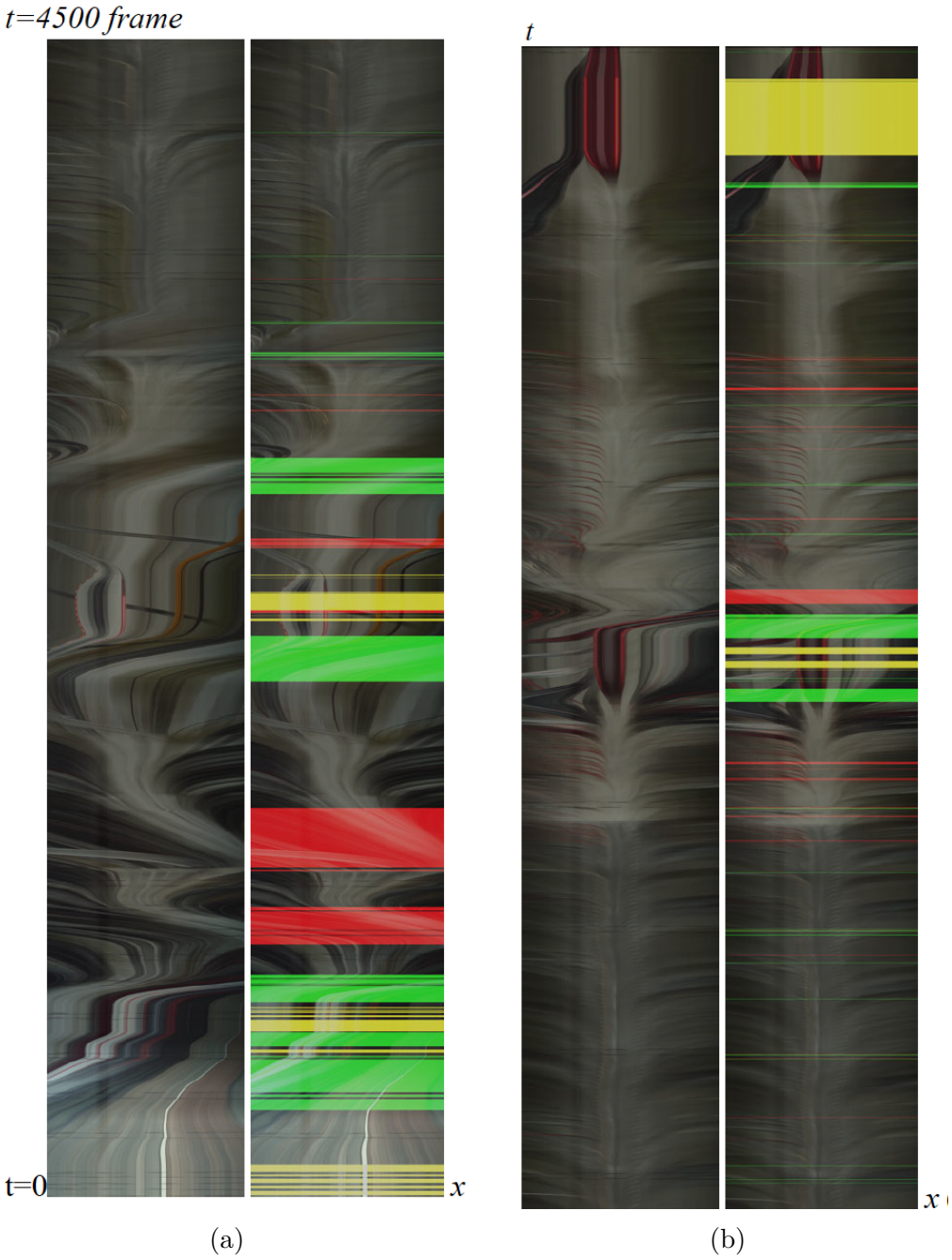


Figure 6.8. Motion traces with vertical edge points are marked in blue and rest of traces in red from motion profiles M_0 and M_1. The resulting stopping period is marked with yellow lines at center.

For stopping and turning period detection of ego-vehicle, we refer to static scenes relative to the camera. Occasional cases of parallel driving of a nearby vehicle may generate vertical traces locally, but such traces only remain in short segments. Figure 6.8 shows a period of stopping with many vertical traces marked in the motion profiles, and the period is marked



(a) (b)

Figure 6.9. Effective detection of stopping and turning frames in motion profiles of a 5 min video. Left: motion profiles, Right: detect moments in color visualization. Yellow: stopping; Red: right turn; Green: left turn. A temporal median filter can further remove isolate frames (false positives).

as stopping framewise when such traces exceed a given number. Figure 6.9 shows a period of stopping with many traces with an angle to vertical axis in the motion profiles. In motion

profile, if the object in the background are with a left angle to vertical axis, ego-vehicle turns to right. Similarly, if the trace goes to right side of vertical axis, ego-vehicle turns to left.

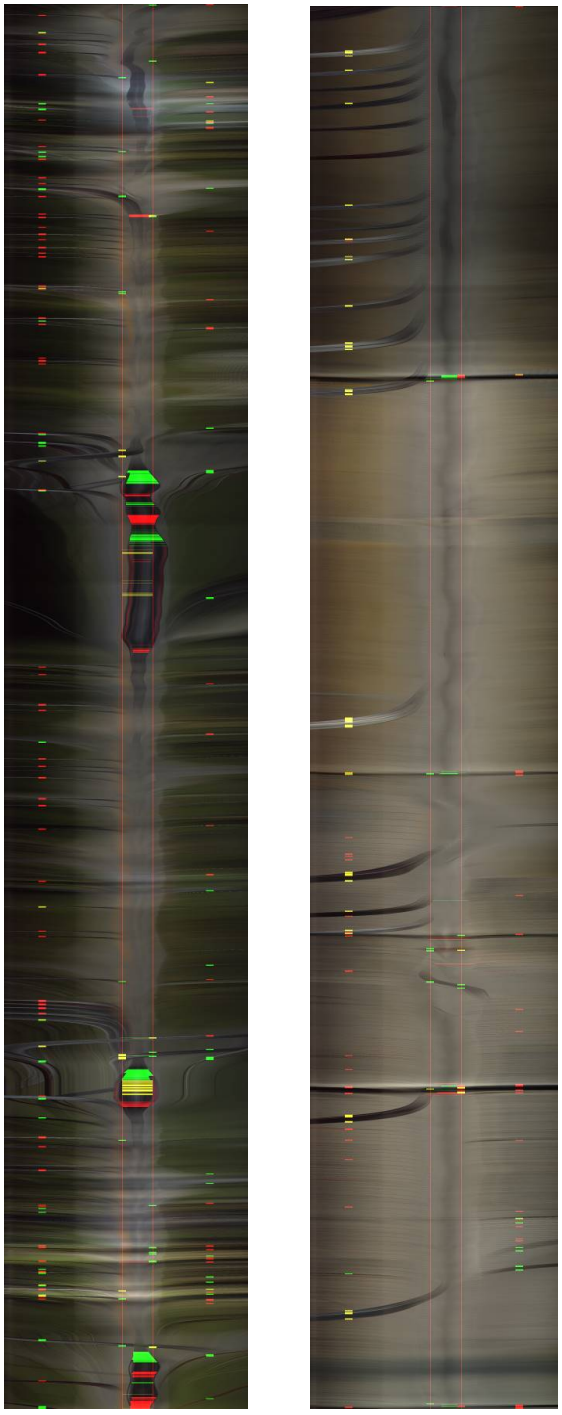
As motion events detected by color in Figure 6.10, green and red at center indicate a frontal vehicle at middle range moving away and getting closer respectively, computed from trace shrinking and expansion. Frequent shrinking and expansion mean a bumper-to-bumper traffic scenario. On the other hand, red color onside indicates on-coming vehicles, and some objects near driving lane due to their fast image velocity (traces are close to horizontal). Green and yellow further indicate outward and inward traces that are passed and passing vehicles on side lane, respectively.

For the overall evaluation, we select 47 clips of 5-min driving with rich vehicle traces in M_1. The interactions are labeled by humans and detected marks are compared with these labels. By counting events and periods, the accuracy is evaluated in Table 6.5. Our method reduces the complexity of problem based on the camera setting and vehicle motion constrained by traffic rules and roads. The filtering at focused positions in a motion profile needs much lower computation cost than current deep learning methods in the recognition of vehicle and their motion. The motion profiles are obtained in video rate and all filtering takes at most 0.07s for 1s video on MAC with 2.2GHz Intel Core i7.

Table 6.5. Accuracy of interaction detection based on motion traces

Events	TP	FP	FN	Precision	Recall	F1
Passing(fast)	113	11	4	0.91	0.97	0.94
On-coming	260	27	51	0.90	0.84	0.87
Passed(slower)	104	13	7	0.89	0.94	0.91
Cut-in	38	7	2	0.84	0.95	0.89
Front approach	28	4	1	0.88	0.97	0.92
Front leaving	36	2	2	0.95	0.95	0.95

Our interaction classification is based on motion traces at critical locations. Some event detection has to wait for the entire process happened and is thus more suitable for batch processing of NDV than predicting vehicle behavior ahead. This work has not considered turning/braking lights on vehicles, traffic lights ahead of interactions, or future V2V com-



(a) (b)

Figure 6.10. Detected vehicle interactions in color labels. 5-min drivings with time axes upward. Orange lines at center indicate drive-able width. On both sides, inward traces are detected in yellow, outward traces in green, and horizontal traces in red at sides. At center, green, red, and yellow indicate leaving, approach, and fixed distance of frontal vehicle. Yellow at center also indicates stop of ego-vehicle. (a) Rural driving. (b) A highway with passing vehicles.

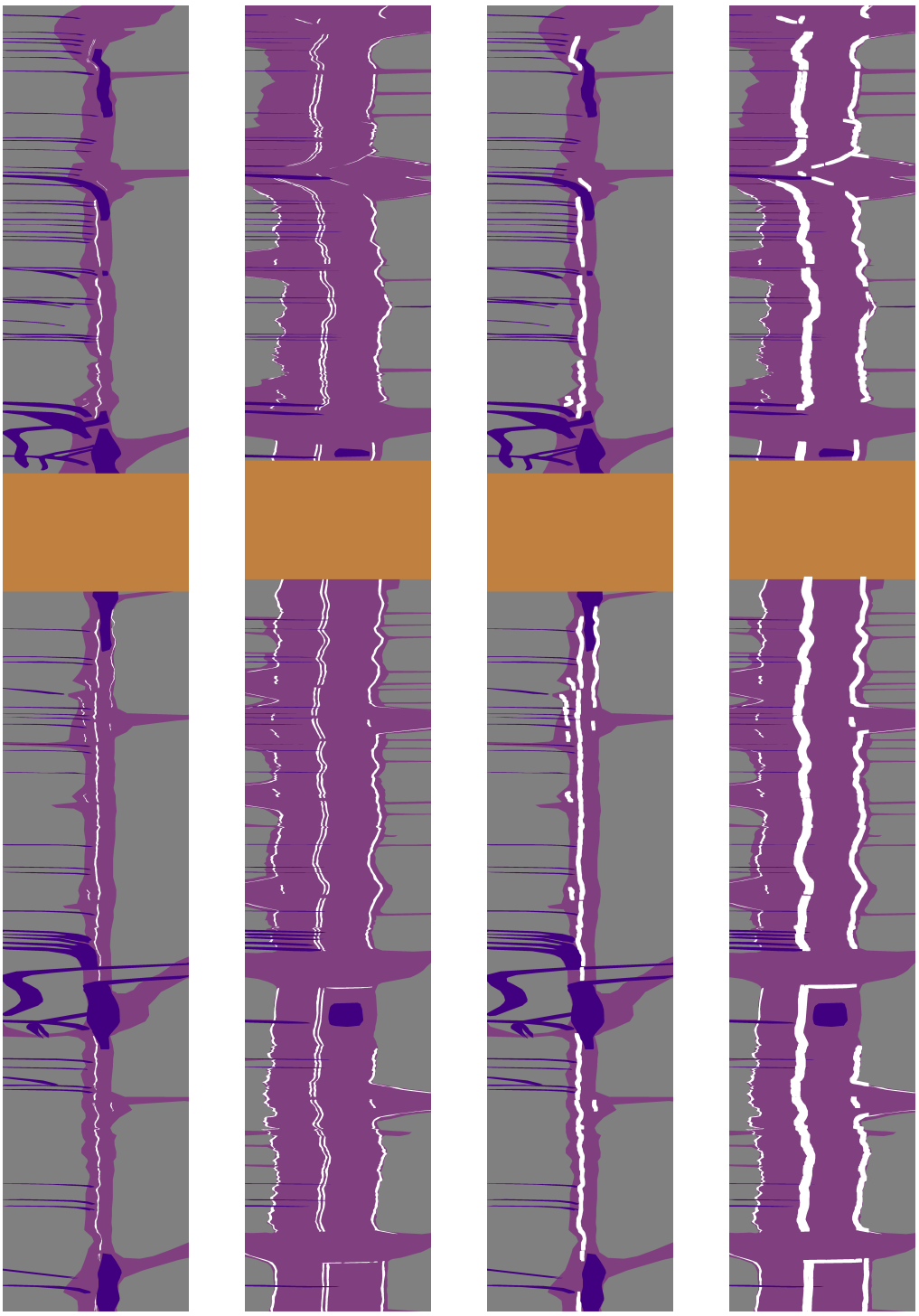
munication during interactions. Recognizing different lights requires locating vehicles and then light blinking.

6.5 Vehicle Motion Planning with Road Profile Images

The vehicle motion planning is based on segmentation results on road profile and vehicle interaction detection using motion profile. Based on these results, we could project our motion planing on driving video as shown in Figure 6.12. The headway space is visualized dynamically in red, green and blue color to show its scope extending to close, mid, and far range. These are results from the semantic segmentation of three road profiles at the depths [39]. As shown in Figure 6.11, the purple regions are detected as on-road surfaces, yellow regions are detected as stop moment of ego-vehicle, blue regions are detected as surrounding vehicles, and white parts are lane marks. Because there are some lane marks not detected or dotted line itself, we enhance the lane marks with wider and connected pixels, as shown in Figure 6.11.

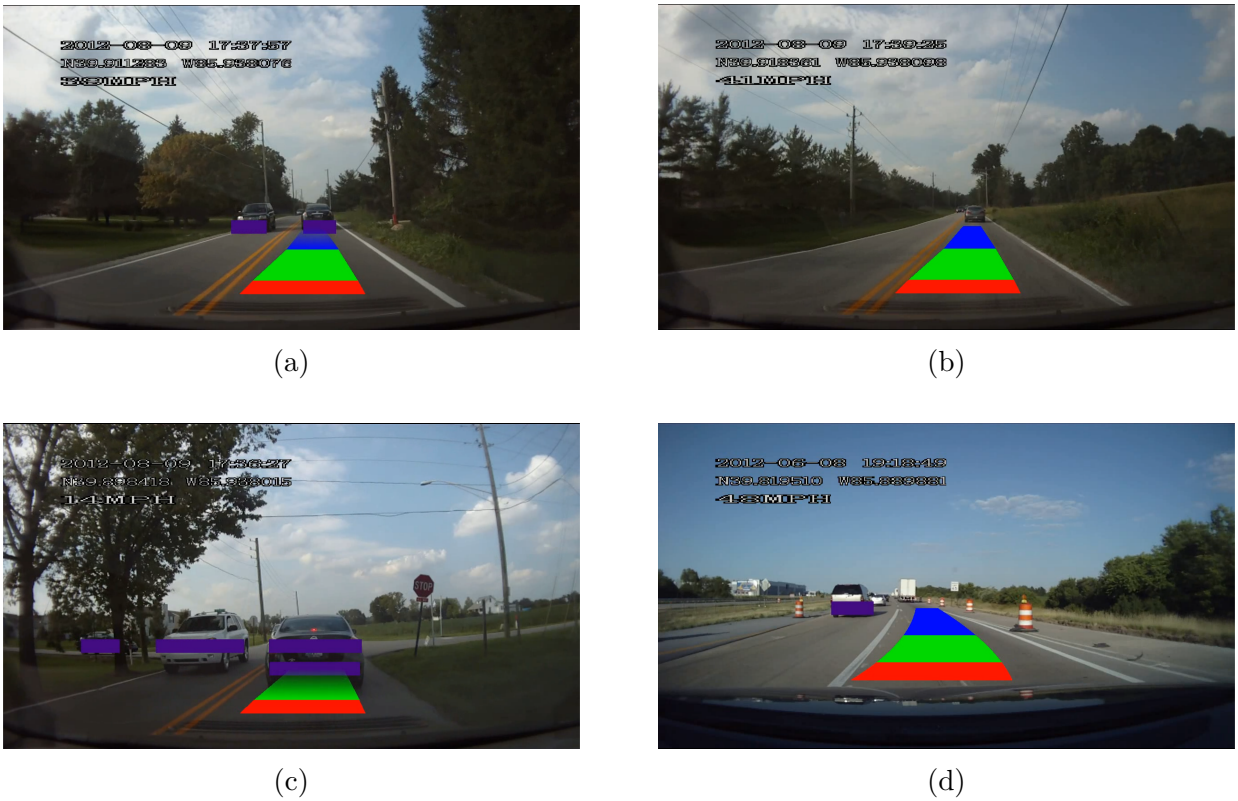
Figure 6.12 is a frame from our motion planning video, and planning path is in the middle. The red, green, and blue region represent the headway space of ego-vehicle from motion profile M_2, M_1, and M_0, which are also the belt locations where the motion profiles are acquired. If there are no vehicle shown in the motion profile, the color of mapped region is solid, otherwise, the region has gradation from front vehicle location, and is disappeared if the belt is totally blocked by front vehicle. The colored headway space is safe for ego-vehicle to drive at a slow, middle, and fast speed accordingly.

For ego-vehicle path control at turning crossing, two turning sections from motion profiles and road profiles are displayed in Figure 6.13 and Figure 6.14. In motion profiles, turning could be detected based on the traces of background objects are all turning to the other side. In road profiles, the traces of road edges are not obvious, which make the detection more difficult for vehicle turning control. Among different moving actions, the turning at crossing is performed at short period. The observed scenes are wide and fast transited in the field of view. The complexity of scenes include complex road structure and vehicles passing and waiting nearby. This causes recognition of road edges more difficult. At such section, the switch from road profiles to normal image for key components localization is necessary.



(a) (b) (c) (d)

Figure 6.11. Enhanced road segmentation using deep learning semantic segmentation. (a),(b) original segmentation result of R_0 and R_1 provided from Cheng et. al [39]. (c), (d) The enhanced lane marks of R_0 and R_1.



(a) (b)

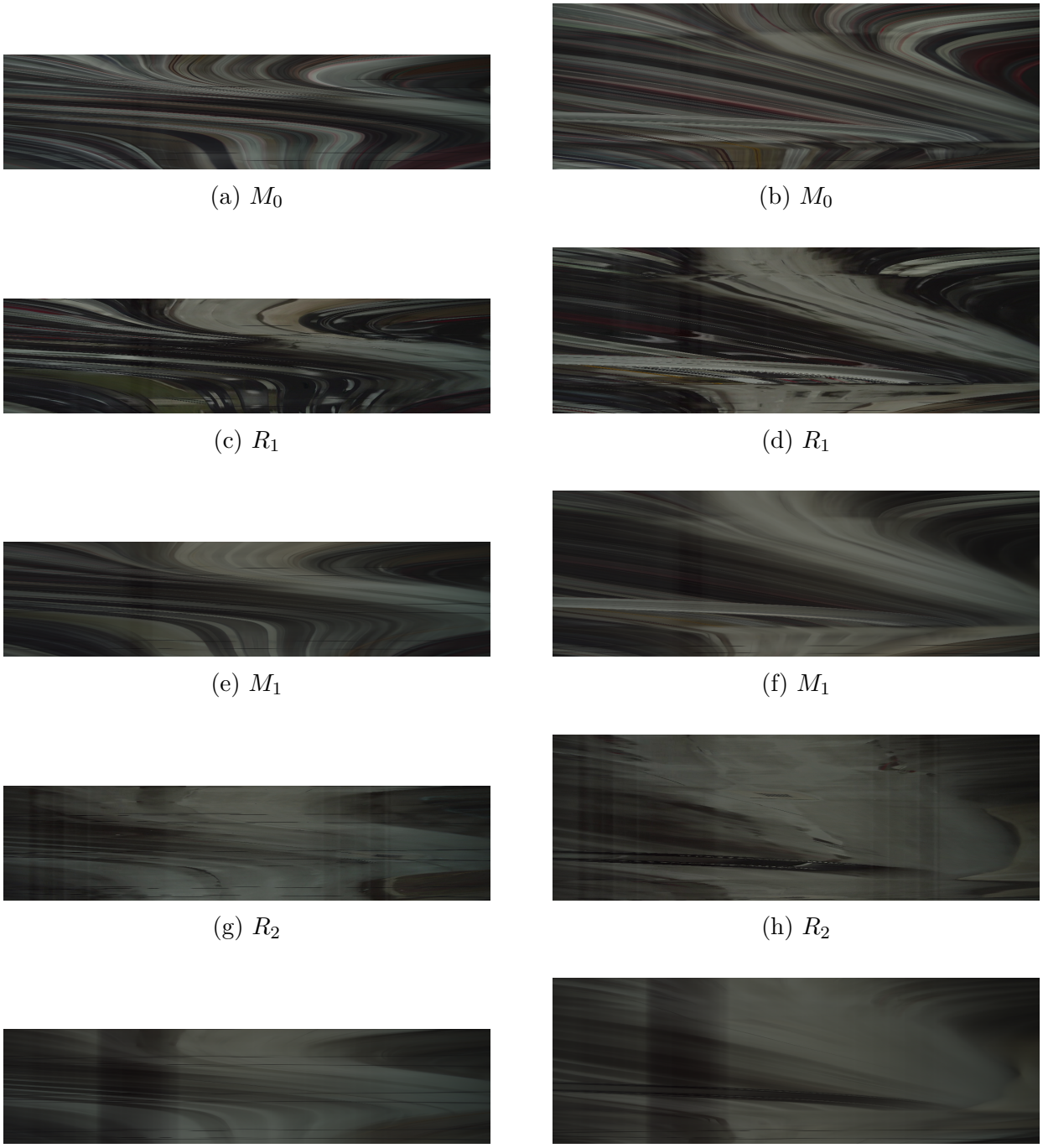
(c) (d)

Figure 6.12. Planned speed on frames with the vehicle path in color and speed in the stacked trapezoidal shapes. (a), (b), (c) are recorded in local road video. There is a vehicle from opposite lane labeled (purple), and a front vehicle labeled. The blue region is half transparent when a front vehicle is closer than the far sampling line and appear in motion profile M_0. (d) is from a speed way video. There is a curve following by the land mark.

6.6 Discussions

In general, the road profiles and motion profiles have been successful in recognizing normal road layout, regular traffic motion under smooth ego-motion. When the road is occupied with other vehicles during ego-vehicle turning, the road structure understanding and vehicle recognition are difficult because of the occlusion of road and changing shapes of vehicles. At these moments, the ego-vehicle can always open "full-eye" to analyze entire video frames with complicated algorithms or use LiDAR data to acquire 3D layout at road intersections.

In the recognition of road environment using semantic segmentation, it is observed that the vehicle identification is reliable on successful road recognition ,which is further relying



(a) M_0 (b) M_0

(c) R_1 (d) R_1

(e) M_1 (f) M_1

(g) R_2 (h) R_2

(i) M_2 (j) M_2

Figure 6.13. Two right-turning sections from motion profile and road profile. The small number profiles capture scene motion faraway including buildings across the street, vehicles at opposite lanes, etc. The larger number profiles record road surface and turning side curb.

(a) M_0 (b) M_0

(c) R_1 (d) R_1

(e) M_1 (f) M_1

(g) R_2 (h) R_2

(i) M_2 (j) M_2

Figure 6.14. Left and right columns show two left-turning sections from motion profile and road profile. The far profiles such as R_1 contains lane separator, which is a landmark for ego-vehicle to turn into the right lane of intersecting road. It also shows right road edge sometime.

on the smooth camera motion driven by the four-wheel vehicle along curved paths on flat surface.

Vehicle motion: For a four-wheeled vehicle, the vehicle motion on the road has limited roll, mild yaw (except turning at street corners), and no pitch influence on local road up to mid distance. Even on hill roads, the pitch against far road only change the scanning depth of lines and belts at a far range, because road construction requires local slope to be invariant.

Lane changing has been smoothly recorded in the road profile and motion profiles, Which has no difficulty in the road environment recognition. More random motion in off-road driving are not included in our scope of investigation.

Road: the lanes are supposed to be visible for precise path planning and vehicle control. For wider area than a lane, upper level navigation should provide road guidance on global position and moving direction of road. Slow driving in parking lots and pedestrian walking zones (market, construction site, etc.) are not included in our study. For traffic jam or parking with very little vehicle space, the road may not be visible for reference and control. Such cases can use LiDAR and sonar to guide parking and flow merging. This work aims at straightforward driving with a higher speed.

Traffic: this work consider traffic on the road as vehicle interaction. It has not included other road users such as pedestrians, bikes, bicyclists, scooter, etc. For pedestrian recognition in the road profile, some other work has provided successful results of detecting pedestrian trajectories. Such results can be used in path and speed planning using the same headway space constraint. Other object recognition can be carried out through deep learning by adding more classes so that there trajectories can be located in multiple motion profiles for headway space in motion planning.

Fortunately, road vehicles have to follow traffic rules, which generate regular motion patterns in the filed of views. These data are learnable through machine learning tools for each driving actions. The outliers of such regular patterns are considered as abnormal situations and may cause accidents such as collision and run off road.

The compact data of road and motion profiles provides a fast detection of road environment and dense sampling rate. This is particularly important in responding to sudden events such as other vehicle's cut-in, merging, and allows the autonomous driving to move fast with less computing cost and light-weight devices on board vehicles.

7. CONCLUSION

This dissertation introduced the road profile image representation, which reduces the size and dimensionality of driving videos, and reveal the applications of the road profiles along with motion profiles in road edge detection, vehicle interaction monitoring, and vehicle motion planning. The profile image is a two-dimensional spatial-temporal image with the time dimension on one axis and the spatial dimension in the video frame on the other. The motion profile images are generated from driving video according to the optical flow of scenes for detecting surrounding vehicles, self movement, and other background objects. Road profile images are used to detect road edge, and finding headway spaces in driving. Both road profile and motion profile compress a video to a temporal image. This representation allows direct extraction of motion as trajectories, which simplified traditional approaches applied to full view from object recognition, view matching, motion tracking, to path planning. Compared to other method, our method detects the direction of motion of vehicles in concentrated zones in driving video without relying on identification, tracking, or classification techniques. It has three significant advantages to impact autonomous driving. First, we reduce the complexity of the problem by filtering the motion profile image. It avoids the errors and computation costs in complex scene recognition, scene matching including occlusion and partial view, and path/speed planning without latency. Second, our method has fast computing speed and thus dense in sensing rate. This allows the vehicle to response promptly to the sudden events on the road and make driving smoothly in busy traffic flow. There are one line information from the frame in road profile and one line compressed from frame in motion profile for each frame. Our approach easily processes results in real time using such compressed data. In other methods detecting on each frame, it is impossible to process all frames object detection, matching and generate traces in one second. Third, our method is not an involved of complicated neural network training and detect. The computing resource in our method are mostly used on filtering and profiling. It means that our detection method requires less computing resource and could use on smaller machine like smart phone.

REFERENCES

[1] *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems*, 2014.

[2] S. Shalev-Shwartz, S. Shammah, and A. Shashua, *On a formal model of safe and scalable self-driving cars*, 2017. eprint: arXiv:1708.06374.

[3] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, *End to end learning for self-driving cars*, 2016. eprint: arXiv:1604.07316.

[4] P. Li, T. Qin, *et al.*, "Stereo vision-based semantic 3d object and ego-motion tracking for autonomous driving," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 646–661.

[5] W. Saeys, B. Lenaerts, G. Craessaerts, and J. De Baerdemaeker, "Estimation of the crop density of small grains using lidar sensors," *Biosystems Engineering*, vol. 102, no. 1, pp. 22–30, 2009.

[6] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, 2013, pp. 1693–1700.

[7] S.-S. Huang, C.-J. Chen, P.-Y. Hsiao, and L.-C. Fu, "On-board vision system for lane recognition and front-vehicle detection to enhance driver's awareness," in *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, IEEE, vol. 3, 2004, pp. 2456–2461.

[8] F. Lindner, U. Kressel, and S. Kaelberer, "Robust recognition of traffic signals," in *IEEE Intelligent Vehicles Symposium, 2004*, IEEE, 2004, pp. 49–53.

[9] S. M. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 22–35, 2017. DOI: 10.1109/MSP.2016.2628914.

[10] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019.

[11] J. Hecht, "Lidar for self-driving cars," *Optics and Photonics News*, vol. 29, no. 1, pp. 26–33, 2018.

[12] M. Kilicarslan and J. Y. Zheng, "Predict vehicle collision by ttc from motion using a single video camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 2, pp. 522–533, 2019. DOI: 10.1109/TITS.2018.2819827.

[13] G. Cheng, J. Y. Zheng, and M. Kilicarslan, "Semantic segmentation of road profiles for efficient sensing in autonomous driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 564–569. DOI: 10.1109/IVS.2019.8814259.

[14] H. Kong, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 96–103. DOI: 10.1109/CVPR.2009.5206787.

[15] T. Veit, J.-P. Tarel, P. Nicolle, and P. Charbonnier, "Evaluation of road marking feature extraction," in *2008 11th International IEEE Conference on Intelligent Transportation Systems*, 2008, pp. 174–181. DOI: 10.1109/ITSC.2008.4732564.

[16] C. Fernández, D. F. Llorca, C. Stiller, and M. A. Sotelo, "Curvature-based curb detection method in urban environments using stereo and laser," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 579–584. DOI: 10.1109/IVS.2015.7225747.

[17] A. Gern, R. Moebus, and U. Franke, "Vision-based lane recognition under adverse weather conditions using optical flow," in *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, 2002, 652–657 vol.2. DOI: 10.1109/IVS.2002.1188025.

[18] Z. Wei, C. Wang, P. Hao, and M. J. Barth, "Vision-based lane-changing behavior detection using deep residual neural network," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 3108–3113. DOI: 10.1109/ITSC.2019.8917158.

[19] O. Bagdadi, "Assessing safety critical braking events in naturalistic driving studies," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 16, pp. 117–126, 2013, ISSN: 1369-8478. DOI: https://doi.org/10.1016/j.trf.2012.08.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847812000770.

[20] K. Choi and H. G. Jung, "Cut-in vehicle warning system exploiting multiple rotational images of svm cameras," *Expert Systems with Applications*, vol. 125, pp. 81–99, 2019, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2019.01.081. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417419300971.

[21] A. Khosroshahi, E. Ohn-Bar, and M. M. Trivedi, "Surround vehicles trajectory analysis with recurrent neural networks," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 2016, pp. 2267–2272. DOI: 10.1109/ITSC.2016.7795922.

[22] J. Nilsson, J. Silvlin, M. Brannstrom, E. Coelingh, and J. Fredriksson, "If, when, and how to perform lane change maneuvers on highways," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 4, pp. 68–78, 2016. DOI: 10.1109/MITS.2016.2565718.

[23] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013. DOI: 10.1177/0278364913491297.

[24] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[25] R. Tian, L. Li, K. Yang, S. Chien, Y. Chen, and R. Sherony, "Estimation of the vehicle-pedestrian encounter/conflict risk on the road based on tasi 110-car naturalistic driving data collection," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 623–629. DOI: 10.1109/IVS.2014.6856599.

[26] M. Elhoseiny, S. Huang, and A. Elgammal, "Weather classification with deep convolutional neural networks," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3349–3353. DOI: 10.1109/ICIP.2015.7351424.

[27] C. Lu, D. Lin, J. Jia, and C.-K. Tang, "Two-class weather classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014.

[28] X. Yan, Y. Luo, and X. Zheng, "Weather recognition based on images captured by vision system in vehicle," in *Advances in Neural Networks – ISNN 2009*, W. Yu, H. He, and N. Zhang, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 390–398, ISBN: 978-3-642-01513-7.

[29] G. Cheng, Z. Wang, and J. Y. Zheng, "Modeling weather and illuminations in driving views based on big-video mining," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 522–533, 2018. DOI: 10.1109/TIV.2018.2873920.

[30] G. Cheng, J. Y. Zheng, and H. Murase, "Sparse coding of weather and illuminations for adas and autonomous driving," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 2030–2035. DOI: 10.1109/IVS.2018.8500385.

[31] Z. Wang, G. Cheng, and J. Zheng, "Road edge detection in all weather and illumination via driving video mining," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 232–243, 2019. DOI: 10.1109/TIV.2019.2904382.

[32] M. Roser and F. Moosmann, "Classification of weather situations on single color images," in *2008 IEEE Intelligent Vehicles Symposium*, 2008, pp. 798–803. DOI: 10.1109/IVS.2008.4621205.

[33] G. Cheng, J. Y. Zheng, and M. Kilicarslan, "Semantic segmentation of road profiles for efficient sensing in autonomous driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 564–569. DOI: 10.1109/IVS.2019.8814259.

[34] F. Wulff, B. Schäufele, O. Sawade, D. Becker, B. Henke, and I. Radusch, "Early fusion of camera and lidar for robust road detection based on u-net fcn," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1426–1431. DOI: 10.1109/IVS.2018.8500549.

[35] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 33–55, 2016. DOI: 10.1109/TIV.2016.2578706.

[36] C. E. Akbaş and M. Kozubek, "Condensed u-net (cu-net): An improved u-net architecture for cell segmentation powered by 4x4 max-pooling layers," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 446–450. DOI: 10.1109/ISBI45749.2020.9098351.

[37] M. Kilicarslan and J. Y. Zheng, "Direct vehicle collision detection from motion in driving video," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1558–1564. DOI: 10.1109/IVS.2017.7995931.

[38] M. Kilicarslan, J. Y. Zheng, and K. Raptis, "Pedestrain detection from motion," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 1857–1863. DOI: 10.1109/ICPR.2016.7899907.

[39] G. Cheng, J. Y. Zheng, and M. Kilicarslan, "Semantic segmentation of road profiles for efficient sensing in autonomous driving," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 564–569. DOI: 10.1109/IVS.2019.8814259.