Robust Bayesian Detection and Tracking of Lane Boundary Markings for Autonomous Driving

by

Michael Smart

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Mechanical Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Lane detection is a fundamental and challenging task in autonomous driving and must be performed safely and robustly to avoid catastrophic failures. Current methods do not perform effectively in the challenging scenarios arising from degraded or worn lane markings and preclude the broader deployment of autonomous driving technologies. Additionally, many methods lack provisions for safe failures, and will return false positive detections as the strongest lane marking candidate instead of declaring that no lane marking was found.

This work proposes several changes to the current state of the art in robust lane detection and tracking and builds on existing methods using Dynamic Bayesian Networks with heuristic features. A new training approach is proposed for learning heuristic feature distributions from unlabelled data with greatly reduced sensitivity to initialization. The null hypothesis is then reformulated to provide a fail-safe so that in the absence of a successful detection, the lane detection system will be able to declare a detection failure instead of producing a high-risk false positive. The Bayesian Inference formulation used in the current state of the art is then generalized to support different lane marking configurations. Lastly, a stereo threshold filter is proposed as a method for reducing dangerous false positives caused by out-of-plane features.

The proposed methods were tested against several datasets, including the new WAterloo Representative Roads (WARR) dataset, covering a 40 km route around the Waterloo region captured at 3 Hz. When tested against the KITTI dataset, the proposed stereo filter has a negative predictive value of over 95% and provides a dramatic reduction in dangerous false alarms. The proposed detection method is effective in scenarios that match the expected single-lane road model and fails safely in 84% of the scenarios that do not adhere to the expected model. Of the dangerous failures, approximately 90% were model failures and may be corrected through use of a different detector within the proposed generalized form that is more compatible with the failure scenario. Such a reduction in model failures could dramatically reduce the rate of potentially dangerous failures and represents a significant improvement on the state of the art.

Acknowledgements

I would like to thank my supervisor, Professor Steven Waslander, for providing me with the opportunity to explore such an interesting field. His guidance, encouragement, and support have been pivotal in the success of this work. I would also like to thank him for allowing me to take the extra time to examine much more challenging real-world scenarios that would otherwise have been possible.

I would also like to thank all of the members of the Waterloo Autonomous Vehicle Laboratory: Mike, Arun, Sirui, Adeel, Kevin, James, Neil, Nima, Abdel, Stan, Devinder, Bek, Jungwook, Chris, and Yu. They have been wonderful to work with and have been great for bouncing around ideas.

I would like to specifically thank Dr. Michael Tribou for helping me figure out how planar homographies work in the context of projective camera models. I would like to thank Dr. Kris Pryke for the use of her vehicle in collecting the data for this thesis. I would like to thank Natalie Isotupa for driving during data collection and for her help in planning the collection route; her deep knowledge of the Waterloo region enabled the route to be planned in a few hours instead of several days. I would also like to thank Professor Sapna Isotupa for the many helpful and illuminating conversations about statistics and probability.

Lastly, I would like to thank my family and friends for their continuous support and encouragement.

Dedication

This thesis is dedicated to Nat for her support, patience, and encouragement over these past few years, and for generally putting up with it.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Lane detection and lane marking detection are challenging and fundamentally critical tasks required for autonomous driving. Without a reliable and accurate estimate of the vehicle's position and orientation relative to its currently occupied lane, it becomes impossible to maintain safe and continued control of the vehicle, and increases the risk of a catastrophic unintended lane departure. Since lane markings are often the only available information for identifying the location of the boundary between lanes, failures in lane marking detection whether due to faded markings or occlusions could have dangerous consequences. For this reason it is imperative for lane detection methods to not just be robust, but also support fail-safes so that inevitable detection errors do not result in loss of life, requirements that have yet to be fully satisfied. Lane marking detection demands are not restricted solely to the ego lane, but must detect all markings required for any vehicle maneuver, such as the lane markings for the adjacent lane in the case of a lane change.

Commercial lane detection systems have advanced considerably, with the Mobileye 560, used in Tesla's Autopilot autonomous highway driving system, being the most prominent example, but have not yet demonstrated full solution of the lane detection problem. For liability reasons, both Mobileye and Tesla are understandably very conservative in their claimed system limitations. Mobileye's 5-series system limitations indicate that they "are intended for paved roads with lanes that are clearly marked" [23], and that performance can be reduced by road, weather, or any other conditions that may occlude the camera's view [23]. Similarly, Tesla's Autopilot limitations claim reduced functionality due to factors such as: poor visibility, poor lane markings, unusual road width or curvature, excessive brightness or excessive shadows, or occluded views [28]. While many user experience examples indicate high performance beyond these claimed limitations, failures still occur regularly within the described system limitations and consequently both Tesla and Mobileye indicate

Figure 1.1: Lane Detection Focus on Drivability: Example ground truth from the [11]. The focus here is on detecting only the immediately drivable portions of the roadway (blue) and the ego-lane (green).

that their detection systems are for guidance purposes only with the driver still required to maintain continuous awareness and performance of the lane detection task [23] [28], proving that autonomous lane detection is still very much an open problem commercially.

In research, progress on the lane detection area has somewhat focused on the detection of the immediately drivable portion of the ego-lane, instead of the problem of detecting just the boundaries of the lane. For example, the KITTI Vision Benchmark Suite focuses on the detection of the immediately drivable portion of the ego-lane as shown in Figure 1.1 from [11]. Restricting the lane detection problem to the drivable portion of the ego-lane serves as a combination of the two separate tasks of lane boundary detection and drivability analysis, with mutual information improving performance and the top methods scoring F-1 metrics in excess of 92%.

Focusing on the detection of the immediately drivable portion of the ego-lane however introduces some severe limitations. In many cases, especially in traffic with a leading vehicle, lane markings can be visible at a distance well beyond the immediately drivable region, and limiting lane detection to the drivable region prevents the information provided by these markings from being used. Examples of these types of situations are shown in Figure 1.2.

While detection of the immediately drivable portion of the ego-lane is an important requirement for the overall task of fully autonomous driving, not all applications require such a restricted form of lane estimation. For example, a lane departure warning system is concerned with lateral positioning of the vehicle relative to the ego lane and depends on using as much lane marking information as can reliably be obtained, whether a leading

(a)

(b)

(c)

Figure 1.2: Markings Beyond Leading Vehicle: (a-c) Images of road scenes with leading vehicles. Significant portions of visible lane markings are not immediately drivable due to the presence of the leading vehicle. For dashed lines in particular, most of the visible segments would be ignored in the problem formulation of [11].

vehicle is present or not. Methods focused on detection of the immediately drivable portion of the ego-lane have their place, but they do not solve the distinct problem of detecting lane boundary markings themselves.

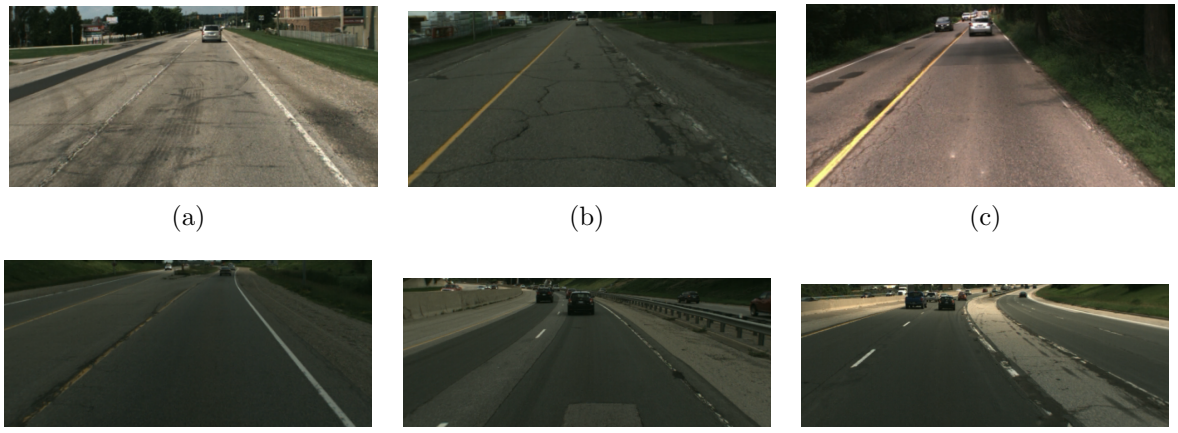
Lane detection methods that focus on the detection of lane boundary markings have also progressed significantly, however robustness remains a significant challenge, with very few works directly addressing the challenge posed by degraded lane markings. Most works are based on the unstated presumption that lane markings are clearly visible in the first place and not severely faded. While this assumption is often strongly true in many regions of the world, it also fails quite regularly in countries with severe winters where the deicing of road surfaces greatly accelerates lane marking degradation, such as in Canada, making for a much more challenging lane detection problem. The difference between the quality of lane markings encountered in many publications suffering from this limitation and the quality encountered on Canadian roads can be quite striking, as shown in Figures 1.3 through 1.5:

(a) (b) (c)

Figure 1.3: Typical Lane Marking Examples: (a) An example image from the KITTI Roads dataset [11]. (b) A lower quality lane marking in the KITTI Roads dataset. (c) An example image from the ROMA dataset [29]

(a) (b) (c)

Figure 1.4: Typical Canadian Lane Markings: These typical examples of Canadian lane markings are well within the scope of quality represented in datasets such as those shown in Figure 1.3



(a) (b) (c)

(d) (e) (f)

Figure 1.5: Challenging Canadian Lane Markings: (a-c) Rural roads and side roads. (d) A rural highway. (e, f) A controlled-access highway. For ground truth: (a) The barely visible left-hand lane marking is supposed to be a solid white marking. (b, c) The faded right-hand lane markings are supposed to be solid white markings. (d) The faded left-hand marking is supposed to be a solid yellow marking. (e) The right-hand marking is supposed to be a solid white marking. (f) The right-hand lane marking for the ego lane and the left-hand marking for the adjacent lane in the connecting off-ramp to the right are supposed to be solid white lines. These markings are clearly of lower quality than the scope covered in data sets shown in Figure 1.3

The stark difference between the quality of Canadian lane markings and the lane markings found in most published lane detection research has severe implications. First, the

Figure 1.6: Detection Failures from Worn Markings: The marking detection method of [14] fails to detect these markings and flags them for needing maintenance. The detection method of [14] is otherwise very effective for well maintained markings.

results achieved by methods that are trained and evaluated on datasets with high quality lane markings may not be representative of their performance on Canadian roads. Second, some of the best methods of lane detection demonstrated on such datasets may not even be feasible for Canadian roads since the problems of detecting well-maintained markings and detecting faded or low quality markings are so different. For example, [14] uses this exact principle as a method for detecting regions requiring maintenance - a marking detector using intensity segmentation and contour classification that performs very well with high quality markings is used and its results are compared to a known map of lane markings, with regions where the detector failed being designated as requiring maintenance. An important note is that the shown example of regions requiring maintenance, shown in Figure 1.6, is still of much higher quality than the Canadian examples presented in Figure 1.5.

Failure to robustly and reliably detect lane markings that are worn or faded precludes the wider deployment of autonomous vehicles to the affected regions and particularly to Canadian roads. While the markings shown in Figure 1.5 are only a small minority of Canadian lane markings, they are common enough that they are very regularly encountered by nearly all Canadian drivers. Autonomous vehicles must also be able to encounter such lane markings without failure if they are to be widely deployed on Canadian roads. Lastly, it is important to note that the need for detecting severely faded lane markings is not unique to Canada, and to some extent affects many countries that experience harsh winters, including the northern United States and much of northeastern Europe.

To date, the list of works addressing the specific problem of lane detection for weak or low quality lane markings is extremely limited as described in the survey papers [4, 22, 25] with [2, 3, 7, 8, 9, 15, 17, 16, 27, 30, 31] as examples that claim strong performance but do not all address the significant challenge posed by faded or worn markings and only demonstrate their results on scenes with bright and visible markings. Specifically, all of the preceding methods rely on slight variations of the 'Intensity Bump Method' - a method relying on the sharp dark-light-dark transition that lane markings show in a greyscale intensity image. These methods were specifically examined by [19] and found to significantly under-perform in lane marking pixel classification compared to neural networks or support vector machines - especially with worn or degraded markings.

Two methods that address degraded markings and robustness are [1] and [24]. The work of [1] achieves robust results using RANdom Sample And Consensus (RANSAC), while [24] uses tracking and a particle filter approach. Both [1] and [24] however use a variation of the 'Intensity Bump Method' for initial feature extraction and are consequently limited in the extent of challenging markings that they can address. For example, the method of [24] can address markings that temporarily disappear due to occlusion or other discontinuity, but only as long as the discontinuity does not contain other obfuscating image features that may resemble markings and spawn false positives through tracking. Although [1] should provide some improved performance with worn markings, the presented results ignore lower quality markings and instead use the improved robustness to detect markings in much more complicated scenarios, such as markings across three lanes.

In contrast to both [1] and [24], the method proposed by Kim [19] specifically focuses on detecting lower quality markings and uses both RANSAC and tracking to obtain robust results and position itself as the current best method for dealing with degraded lane markings. Kim's method uses a neural network patch classifier for image preprocessing followed by a combined RANSAC and dynamic Bayesian network for marking detection and classification with a specific focus on robustly detecting poor or faded lane markings. Kim's method is also resistant to partial occlusions and claims strong results. More importantly, the method also uniquely includes a null-hypothesis, allowing for a probabilistic evaluation of the possibility of a marking not being present in the image or being misdetected. Allowing for the null-hypothesis greatly improves safety as it allows for the detection system to declare a marking as not being detected instead of the detection of a non-marking as a dangerous false alarm.

There are some limitations to the results of [19]. First, the results have not been demonstrated on modern data sets with modern cameras; the videos used for testing were 352×240 in MPEG format with many of the resulting errors caused by MPEG compression effects and poor image quality [19]. Consequently, the results claimed in [19] are

not representative of the algorithm's potential - it could perform much better on higher quality images. Second, while the Bayesian Evaluation Framework is derived for a general deformable multi-part model, there is no discussion or guidance given towards the development of detectors addressing markings that do not fit the bounded single-lane model described. Lastly and most severely, the training method proposed for learning the required conditional distributions is extremely sensitive when applied to modern images and wider fields of view and fails to converge unless the initial estimate is extremely close to the optimal solution. This inability to train the system described in [19] limits the use of its overall method and precludes reasonable investigation of the method or its application to modern data sets.

In this work I make several contributions. First, I solve the challenge in training the method of [19] by reformulating the training problem as a mixture model problem and applying the well-known and robust Expectation Maximization (EM) algorithm. I also replace the tuned tracking parameters used by [19] with derived values that do not require any tuning and learned parameters obtained from EM. Second, I adjust the definition of misdetections used by [19] so that the null hypothesis becomes usable as a failsafe. If the detector does not find a suitably confident hypothesis, then the null hypothesis is returned - producing a system that either succeeds or is aware that it failed and can trigger the appropriate response. Third, I generalize the Bayesian Inference equations used in [19] to allow for arbitrary lane models so long as heuristic evidence metrics can be defined for them. Fourth, I propose a stereo filter that dramatically reduces the number of false alarms caused by out of plane elements. Lastly, these results are demonstrated on a dataset of 8644 images covering 40 km of driving around the region of Waterloo with a vehicle-mounted integrated stereo camera providing image and depth information. The results indicate a system that is effective in scenes where its road model is appropriate, and is also able to detect the majority of instances where it fails and fail safely as a result.

The remainder of this thesis proceeds as follows. Chapter 2 presents the necessary remaining background information upon which this work is built including a more detailed examination of the methods in [19] and the Expectation Maximization method for solving mixture model estimation problems. Chapter 3 presents the application of Expectation Maximization to training the method of [19] and provides a recommended training procedure before presenting a more general form of the lane marking model and several examples of alternative lane marking classes derived from this more general form. Chapter 4 presents the results of applying the developed method to 40 km of driving data around the region of Waterloo. Lastly, Chapter 5 provides a conclusion to this work as well as a discussion of future work that could further develop the contained method.

Chapter 2

Background

This chapter covers the background information required for the contributions made by this work. The Birds Eye View image space is discussed, followed by lane marking hypothesis generation, heuristic hypothesis features, and Bayesian hypothesis evaluation. Lastly, the Expectation Maximization algorithm is discussed as a method of solving mixture model problems, particularly with exponential family distributions.

2.1 Generating the Bird's Eye View

In order to perform detection and tracking of lane markings, a domain must be specified within which the detection and tracking is performed. In keeping with the KITTI-proposed evaluation framework [11], this work uses Bird's Eye View (BEV) images centered on a vehicle-fixed frame for evaluation of results. The reasons for performing evaluation in the BEV space are thoroughly covered in [19, 11] and the same arguments are used here, critically that the BEV space allows for the removal of perspective distortion, which allows the detection results to be evaluated independently of the specifics of the location and orientation of the camera; images from different cameras can be converted to a common frame, and so results can be compared across experiments and trials. These same arguments can be used to justify performing detection steps in the BEV space as well. Additionally, performing detection in the BEV space allows the entire detection pipeline to be evaluated within a common frame without requiring intermediate transforms or conversions when assessing individual components of the machine learning system.

There are multiple methods for obtaining a BEV of the ground plane from a given monocular image, two of which will be described in this work. Both methods rely on using

a planar homography to map the ground plane coordinates to corresponding image plane coordinates, but differ in their method of obtaining such a homography. The homography maps the ground plane coordinates (x, y) to the image points (u, v) in homogeneous coordinates through

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.1}$$

where $\mathbf{H}$ is a 3×3 homography matrix and λ is a scalar. For each desired point in the BEV image, equation (2.1) is used to obtain the corresponding coordinates in the source image. As the source image only contains discrete points, the desired pixel value for the BEV image is then determined using bilateral interpolation from the integer pixels surrounding the desired source pixel.

The first method used to obtain the homography required by equation (2.1) is to directly determine the homography by reducing the 3D projection equation for a calibrated pinhole camera to a 2D projection for points on a plane. The second method is used when full calibration information is not available, but instead some geometric features in the image are known and the homography is obtained by solving a series of equations.

2.1.1 Bird's Eye View Homography from Calibration

The 3D projection of point $[x, y, z]$ in world coordinates to the 2D image plane point $[u, v]$ for a fully calibrated pinhole camera is given by:

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2.2}$$

where the intrinsic camera matrix, $\mathbf{K} \in \mathbb{R}^{3\times3}$, is defined as

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

and consists of the x and y axis focal lengths f_x and f_y given in pixels and the principal image point (c_x, c_y) also given in pixels, and where the extrinsic camera matrix

$$[\mathbf{R}|\mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{2.4}$$

is a joint rotation-translation matrix that maps the coordinates of a point from world coordinates to a camera-fixed coordinate system with the camera center at $(0, 0, 0)$.

Noting that if the ground plane is defined in the world frame to be at $z = 0$, then the z element of the world-coordinate point and the corresponding column of the $[\mathbf{R}|\mathbf{t}]$ matrix can be deleted, reducing equation (2.2) to

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.5}$$

which yields

$$\mathbf{H} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix} \tag{2.6}$$

as the desired 3×3 homography matrix.

2.1.2 Bird's Eye View Homography from Image Features

The fully extrinsic camera matrix $[\mathbf{R}|\mathbf{t}]$ is not always readily obtainable for every image as the ground plane angle can change due to the interaction between vehicle suspension and the road surface and so a second method is used to obtain the desired homography in such cases. The desired homography is a projective transform with 8 degrees of freedom that can be constrained by a variety of image features [13]. The 8 degrees of freedom can be most readily constrained by 4 point correspondences with no three of the corresponding points being collinear in either image, allowing the desired homography to be solved directly [13]. Alternatively, $\mathbf{H}$ can be decomposed into three successive transforms as follows:

$$\mathbf{H} = \mathbf{H_S}\mathbf{H_A}\mathbf{H_P} \tag{2.7}$$

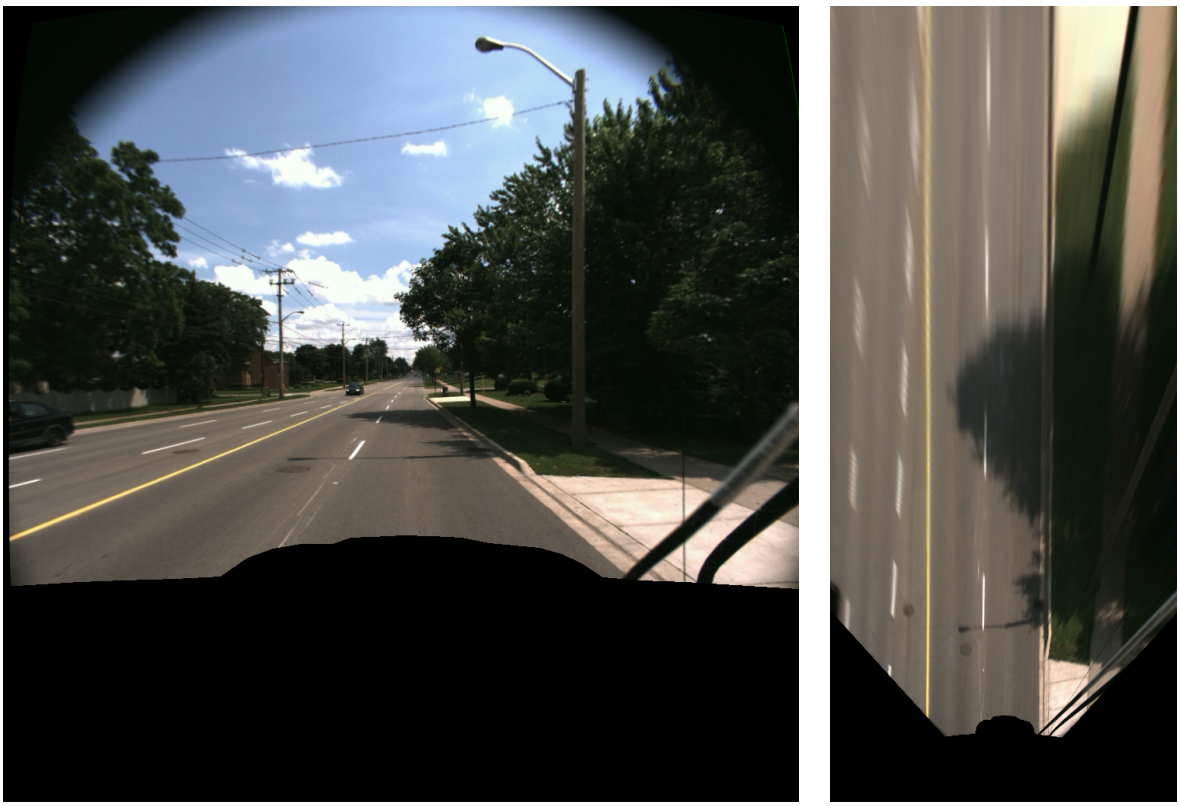
where $\mathbf{H_P}$ captures the perspective transform which has 2 degrees of freedom, $\mathbf{H_A}$ captures the affine properties of the transform which have another 2 degrees of freedom, and $\mathbf{H_S}$ is a similarity transform having the last 4 degrees of freedom [13]. The desired homography can then be obtained by using a variety of properties that may be more readily obtainable than a set of 4 point correspondences. The location of the vanishing line in the image directly provides $\mathbf{H_P}$, with the vanishing line obtainable from any two independent sets of lines parallel in the world frame. The transform $\mathbf{H_A}$ can be determined from 2 constraints given by either 2 pairs of orthogonal lines, a single imaged circle, or two known length ratios. The similarity transform can then be determined from 2 known points. In this work, the similarity transform is imposed by the desired orientation and location of the BEV. In order for all images to have a common center, the origin of the BEV is set to the BEV coordinates of the camera center and the x-axis is aligned to be horizontal in the BEV. A thorough discussion of this approach for obtaining $\mathbf{H}$ is contained in Chapter 2 of [13]. If the first method of obtaining $\mathbf{H}$ is used, then the resulting BEV is subsequently rotated and translated to satisfy the common center and alignment. An example BEV rectification is shown in Figure 2.1.

2.2 Lane Boundary Hypothesis Generation

Separate from the problem of evaluating a given set of lane marking hypotheses is the problem of generating such a set of lane marking hypotheses from an image in the first place. There are many methods used for identifying possible lane markings in images that are discussed in survey papers [25, 4, 22]. In this work, the hypothesis generation method of [19] is used in favour of other methods because it is robust enough to accommodate complicated situations arising from both worn or faded lane markings and from complicated scenes which may confuse other methods. Specifically:

1. A neural network patch classifier is used to identify potential lane marking pixels in the BEV image, as opposed to the faster but less robust alternatives of gradient-based or intensity-bump feature extraction, allowing detection of faded lane markings missed by some intensity-bump detectors without overly favouring noisy regions as in some gradient-based methods [19].

2. A RANSAC-based approach is used for generating hypotheses, allowing suitable hypotheses to be generated even in complicated urban scenarios where there may be many line segments identified as lane marking pixels but only a small minority of which represent suitable hypotheses for the lane of interest [19].



(a) (b)

Figure 2.1: Example of the Bird's Eye View: (a) A monocular image of a road scene. (b) The corresponding Bird's Eye View image obtained using the known intrinsic calibration parameters of the camera and estimated extrinsic parameters relative to the ground plane.

The hypothesis generation pipeline used in [19] generally divides into three major processes: the patch classifier, line segment extraction, and RANSAC hypothesis generation. The pipeline is illustrated in Figure 2.2

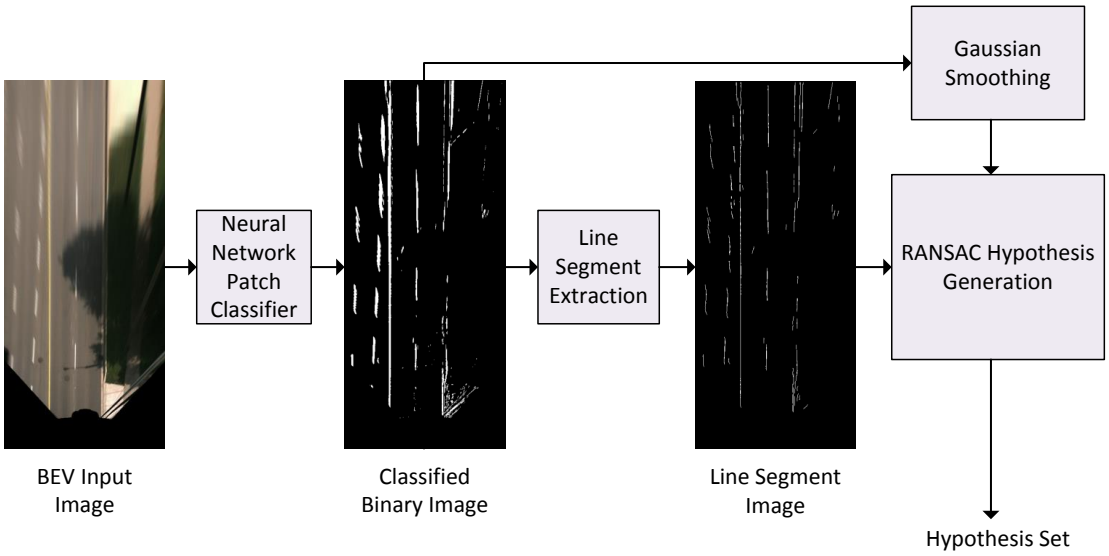


Figure 2.2: Hypothesis Generation Pipeline

2.2.1 Patch Classifier

The patch classifier first processes the BEV image at a local level, assessing 15×3 pixel patches to decide if the patch's center pixel represents a lane marking or not. The decision of whether or not a pixel represents a lane marking is made using a neural network classifier consisting of a single hidden layer and with the pixel values of the patch forming the input vector of either 45 elements for a grayscale BEV image or 135 elements for an RGB BEV image. The decision is illustrated below in Figure 2.3.

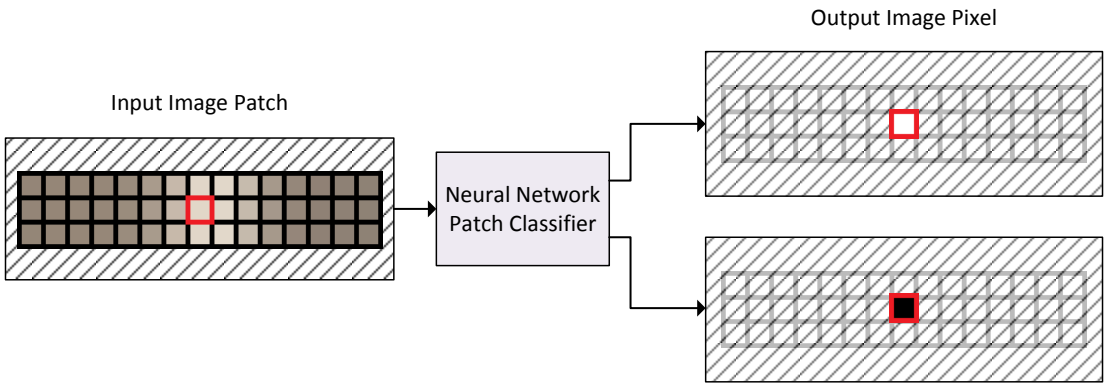
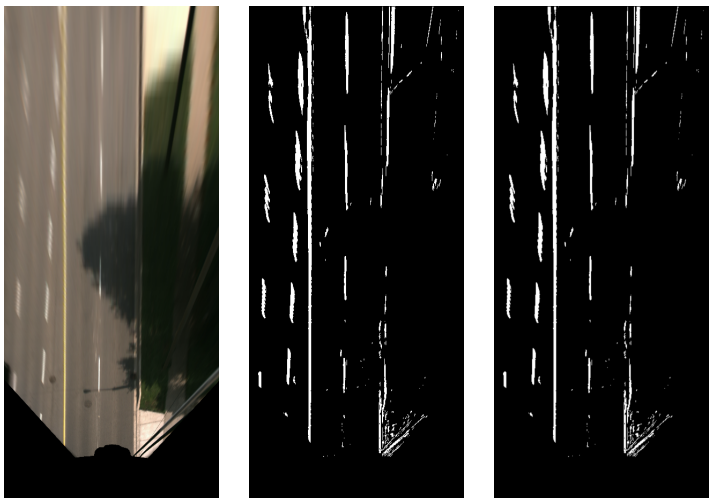


Figure 2.3: The Neural Network Patch Classifier: The neural network takes the pixels in a patch surrounding the pixel of interest (bordered in red) to determine if the corresponding pixel in the output image is classified as a lane marking pixel (top right) or not (bottom right).

The neural network classifier requires a training set of labelled examples of pixel patches from BEV images. Care must be taken when determining whether to use colour or monochromatic images, with the choice being determined by the available training data. If the training data sufficiently includes both yellow and white markings, then the classifier can be trained to detect both types. If the training data does not include enough of either colour, then the resulting classifier will overfit and incorrectly classify the absent marking colour. Consequently, if the labelled training data includes sufficient representation of both white and yellow markings, then the colour form of the patch classifier can be used to take advantage of the additional available information. Otherwise, using the monochromatic form of the patch classifier is more appropriate. In this work, the labelled ROMA dataset was used to train the neural network classifier and the monochromatic form of the classifier was used since the dataset does not include yellow markings[29]. A blurred image, called the score image, is created by applying standard Gaussian smoothing to the binary image created by the patch classifier [19]. The score image is retained for later use in RANSAC hypothesis generation as well as for hypothesis evaluation. An example of a classified image is illustrated in Figure 2.4



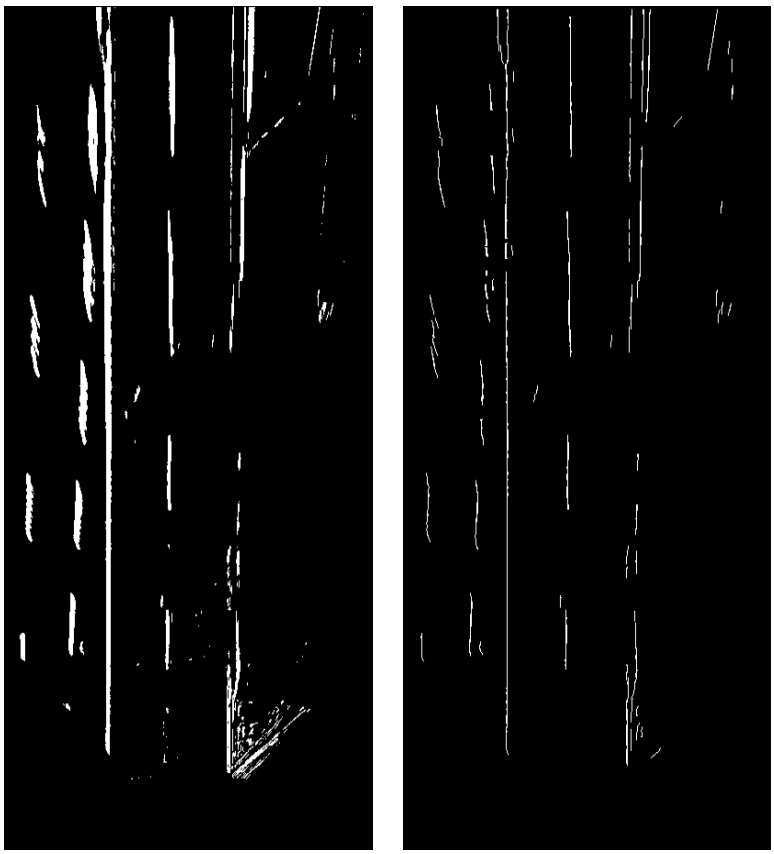
(a) (b) (c)

Figure 2.4: Patch Classifier Results: (a) A source BEV image. (b) The resulting binary image produced by the neural network patch classifier. (c) The score image produced by slightly blurring the image in (b).

2.2.2 Line Segment Extraction

The next stage of image processing extracts line segments from the classified binary image produced by the neural network patch classifier. First, the image is subjected to standard Gaussian smoothing [18]. Next, the image is subjected to non-maxima suppression where all pixels which have an adjacent pixel with a higher value are suppressed [18], followed by a thresholding where all remaining pixels below a specified value are suppressed. The resulting filtered image is then converted back into a binary image having significantly less noise than the original binary image provided by the patch classifier. Connected component analysis, also known as blob extraction, is then performed using eight-element adjacency to identify connected groups of pixels within the filtered binary image [18]. Groups with a population below a specified threshold are discarded. The remaining groups are then evaluated using singular value decomposition to determine the principal components of the group's pixel distribution. The variance in the group's second principal component is compared to the variance in its first principal component, and groups where this ratio is below a specified threshold are retained as the ultimately extracted line segments [18]. All of the above specified thresholds and parameters used in the extraction of line segments are determined by inspection and vary slightly based on the camera and BEV configuration.

An example of the line segments extracted from a binary image produced by the neural network patch classifier is shown in Figure 2.5.



(a) (b)

Figure 2.5: Line Segment Extraction: (a) A binary patch classified image. (b) The extracted line segments.

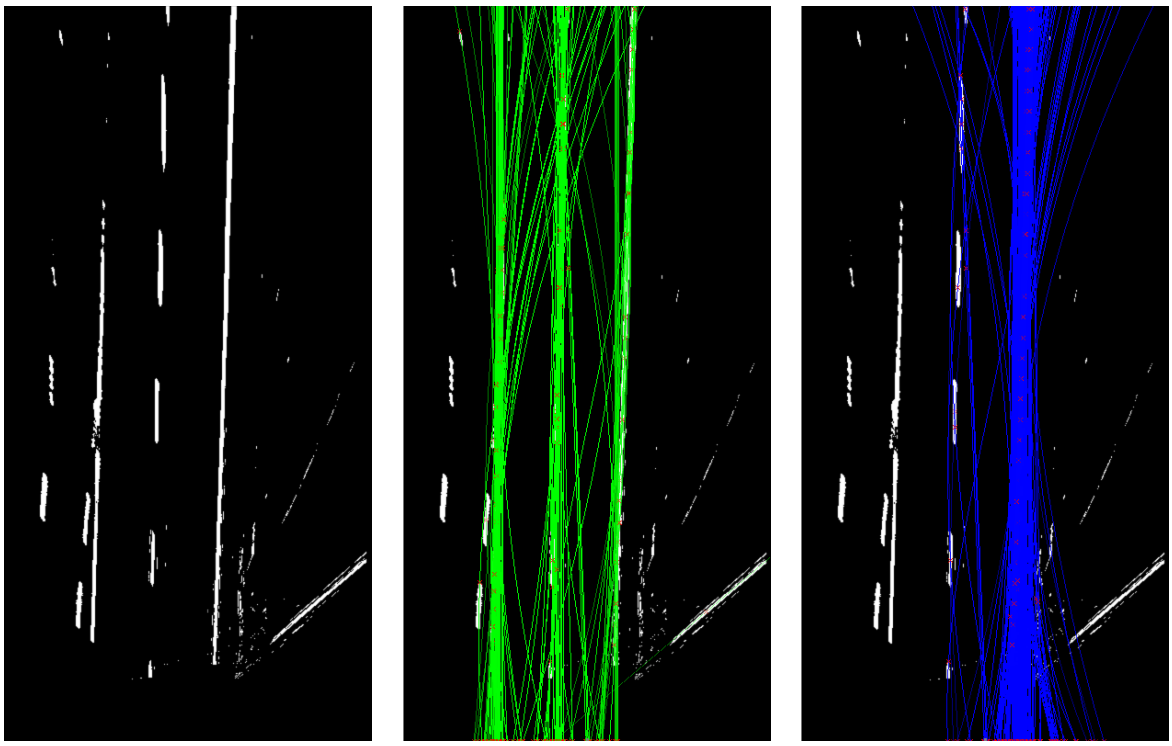
2.2.3 Hypothesis Generation using RANSAC

Hypotheses for each marking in the lane model are generated individually from the line segment image using the RANdom Sample And Consensus(RANSAC) approach, and are in the form of uniform cubic splines [19], where here a uniform cubic spline is a cubic spline whose control points all lie on the spline itself and whose control points are spaced as uniformly as possible. Uniform cubic splines do not typically produce fittings as well

as B-splines, but the RANSAC approach allows the poor fitting spline candidates to be ignored [19]. The hypotheses are generated from one of 4 methods chosen at random:

- A hypothesis is generated from two control points using a single randomly chosen line segment. The first control point is set to where the extrapolated line segment meets the bottom of the BEV image, while the second control point is chosen to be the farther end point of the segment.
- A hypothesis is generated from two control points using two randomly chosen line segments. The first control point is set to where the extrapolated line of best-fit constructed from the two line segments meets the bottom of the BEV image, while the second control point is chosen as the farthest end point of the line segments.
- A hypothesis is generated from three control points drawn from two line segments. The first control point is set to be where an extrapolated spline fit of both line segments meets the bottom of the BEV image, while the third control point is chosen as the farthest end point of the line segments. The second control point is then chosen as whichever of the three remaining line segment end points is closest to the midpoint between the first and third control points.
- A hypothesis is generated from four control points drawn from three line segments. The first control point is set to be where an extrapolated spline fit of just the first two line segments meets the bottom of the BEV image, while the fourth control point is chosen as the farthest end point of the line segments. The second and third control points are then chosen from the remaining end points to provide the most uniform spacing between each of the four control points

Once a hypothesis is randomly generated, it is scored for consensus with the observed lane marking pixels by summing over the score image values for all pixels along the spline. Many part candidates are required to produce valid hypotheses for the lane model as a whole, and so the top K RANSAC-generated candidates for each part (left or right markings) are kept for generating hypotheses of the lane model as a whole, where the value of K is determined by the memory and speed requirements of the application with $(K+1)^2$ hypothesis pairs being produced from K hypotheses for each of the two parts. Figure 2.6 illustrates the part hypothesis populations produced by the hypothesis generation process.



(a) (b) (c)

Figure 2.6: Hypothesis Generation and RANSAC Populations: (a) A source score image. (b) The population of generated part hypotheses for the left-hand marking. (c) The population of generated part hypotheses for the right-hand marking.

The lane model hypotheses are generated by considering each possible pairing of left part hypotheses and right part hypotheses, including the possibility that either part or both could be misdetected or missing in the image. These whole-lane hypotheses are then evaluated for different heuristic metrics that form the features used to evaluate hypothesis validity.

2.3 Heuristic Hypothesis Features

Several heuristic metrics are evaluated for each hypothesis and are used as features to evaluate hypothesis validity. In this work, we select 8 metrics associated with the hypothesis' evidence support in the current frame, encapsulated in the vector $\mathbf{e_c}$, and 6 metrics

associated with the tracking of part hypotheses from frame to frame, encapsulated in the vector $\mathbf{e_t}$. In Chapter 3.3.2 these metrics will be adjusted and expanded to provide a list of heuristics useful for evaluating hypotheses for a variety of lane marking configurations, not just the configuration described in [19].

2.3.1 Current Frame Metrics

The 8 heuristic metrics in the current frame feature vector, $\mathbf{e_c}$, used for evaluating hypothesis support in the current frame are divided into 3 sub-vectors $\mathbf{e_c} = [\mathbf{e_L}, \mathbf{e_R}, \mathbf{e_{LR}}]$ with $\mathbf{e_L}$ for the metrics that are specific to the left part hypothesis, $\mathbf{e_R}$ for the metrics that are specific to the right part hypothesis, and the joint metrics $\mathbf{e_{LR}}$ that describe the compatibility of the two part-hypotheses.

The part-specific evidence vectors $\mathbf{e_L}$ and $\mathbf{e_R}$ each consist of 2 common evidence metrics with $\mathbf{e_L} = [e_{L,1}, e_{L,2}]$ and $\mathbf{e_R} = [e_{R,1}, e_{R,2}]$ for a total of 4 part-specific evidence metrics for the current frame:

1. $e_{L,1}$ and $e_{R,1}$: These part-specific metrics are the part's lane marking support score. A part's lane marking support score is computed by summing over the score image values for all pixels along the part's spline.

2. $e_{L,2}$ and $e_{R,2}$: These part-specific metrics are a curvature penalty determined by summing all of the direction changes in the spline that occur without lane marking support, penalizing over-fitting hypotheses that maximize their support scores.

There are 4 joint metrics $\mathbf{e_{LR}}$ that describe the strength of the hypothesis as a whole with $\mathbf{e_{LR}} = [e_{LR,1}, e_{LR,2}, e_{LR,3}, e_{LR,4}]$. The first 3 are obtained by sampling the lane width along the distance of the BEV image and fitting the samples to a linear model using least-squares linear regression.

1. $e_{LR,1}$: This metric is taken as the absolute difference between the average lane width and the nominal lane width expected by the model.

2. $e_{LR,2}$: This metric is taken as the absolute rate of change of the lane width with respect to distance.

3. $e_{LR,3}$: This metric is taken as the maximum absolute residual of the width samples against the linear fit.

4. $e_{LR,4}$: This metric is taken as the distance from the center of the estimated lane at the bottom of the BEV image to the origin of the BEV image, penalizing otherwise excellent hypotheses that are too far away to belong to the lane enclosing the ego vehicle, as may occur if the adjacent lane has lane markings that are much more clearly visible than the markings for the ego lane. This metric is not present in the original work of [19], but specifically addresses misdetections that can otherwise occur when adjacent lanes are more visibly marked than the ego-lane [26].

2.3.2 Part-Tracking Metrics

The 6 heuristic metrics used for evaluating the tracking of part hypotheses from frame to frame are divided into 2 subvectors $\mathbf{e_t} = [\mathbf{e_{t_L}}, \mathbf{e_{t_R}}]$ with $\mathbf{e_{t_L}} = [e_{t_L,1}, e_{t_L,2}, e_{t_L,3}]$ containing the tracking metrics specific to the left part and with $\mathbf{e_{t_R}} = [e_{t_R,1}, e_{t_R,2}, e_{t_R,3}]$ containing the tracking metrics specific to the right part.

1. $e_{t_L,1}$ and $e_{t_R,1}$: These metrics are taken as the absolute difference in direction between the current part-hypothesis and the previous part-hypothesis, with the directions being obtained as the average rate of change of the part's lateral position along the length of the BEV image.

2. $e_{t_L,2}$ and $e_{t_R,2}$: These metrics are taken as the absolute difference between the current and past part hypotheses at the bottom of the BEV image.

3. $e_{t_L,3}$ and $e_{t_R,3}$: A least-squares linear regression of the lateral difference between the previous and current part along the longitudinal axis of the BEV image. The maximum residual is then taken as this tracking metric.

2.4 Bayesian Hypothesis Evaluation

Each lane hypothesis consists of a pair of part-hypotheses - one each for the left and right lane markings. The validity of each hypothesis can be represented as a 2-vector of binary random values capturing the validity of each part-hypothesis as either True or False. The available evidence metrics associated with the lane hypothesis can be used to assess its likelihood, i.e. the probability that it is correct:

$$\Pr(\mathbf{x}_i|\text{Evidence}) \tag{2.8}$$

where $\mathbf{x}_i = \{L_i = \text{True}, R_i = \text{True}\}$ is a shorthand representing the validity of lane hypothesis i which consists of part-hypotheses L_i and R_i for the left and right-hand lane markings respectively. As part hypotheses can also be missing or misdetected, a shorthand of $L = \phi$ or $R = \phi$ is used for $L_\phi = \text{True}$ and $R_\phi = \text{True}$ respectively for missing part hypotheses L_ϕ and R_ϕ or $x_j = \phi$ for $x_{j,\phi} = \text{True}$ for a missing part hypothesis $x_{j,\phi}$ for a general part j. A shorthand of $x_j \neq \phi$ is similarly used for $x_{j,\phi} = \text{False}$.

The evidence metrics associated with a given hypothesis in equation (2.8) are divided into three categories represented as vectors: $\mathbf{e_c}$ represents the evidence in the current frame, $\mathbf{e_p}$ represents the evidence for the previous image frame, and $\mathbf{e_t}$ represents the transitional evidence between the two frames. As was also assumed in [19], it is assumed here that these three categories of evidence are independent.

Bayes' rule is then used to re-write equation (2.8) more specifically:

$$\Pr(\mathbf{x}_i|\text{Evidence}) = \Pr(\mathbf{x}_i|\mathbf{e_c}, \mathbf{e_t}, \mathbf{e_p}) \tag{2.9}$$

$$= \frac{\Pr(\mathbf{e_c}|\mathbf{x}_i, \mathbf{e_t}, \mathbf{e_p}) \Pr(\mathbf{x}_i, \mathbf{e_t}, \mathbf{e_p})}{\Pr(\mathbf{e_c}, \mathbf{e_t}, \mathbf{e_p})} \tag{2.10}$$

It is then assumed that the evidence vectors are conditionally independent given the hypothesis validity, i.e. if the hypothesis validity is known, then any of $\mathbf{e_c}$, $\mathbf{e_t}$, or $\mathbf{e_p}$ give no additional information about the other evidence vectors:

$$\Pr(\mathbf{x}_i|\text{Evidence}) = \frac{\Pr(\mathbf{e_c}|\mathbf{x}_i) \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p})}{\Pr(\mathbf{e_c})} \tag{2.11}$$

The derivation then divides into two sections: first, evaluation of the current frame components, and second, the evaluation of the tracking components.

2.4.1 Evidence Probabilities From the Current Frame

The terms associated with the current frame consist of the conditional probability of the current-frame evidence $\Pr(\mathbf{e_c}|\mathbf{x}_i)$ and the prior probability of the current-frame evidence $\Pr(\mathbf{e_c})$.

Conditional Evidence Probability

In order to assess $\Pr(\mathbf{e_c}|\mathbf{x}_i)$ it is again assumed that the components of $\mathbf{e_c}$ are conditionally independent given the hypothesis validity, i.e. each evidence metric gives us no additional information about the other metrics if the hypothesis validity is already known.

$$\Pr(\mathbf{e_c}|\mathbf{x}_i) = \prod_{e \in \mathbf{e_c}} \Pr(e|\mathbf{x}_i) \tag{2.12}$$

The current frame evidence metrics are then separated according to the relevant part-hypotheses, noting that evidence metrics for the left part-hypothesis are independent of the right part-hypothesis and vice versa, with the joint evidence metrics depending on both parts:

$$\Pr(\mathbf{e_c}|\mathbf{x}_i) = \prod_{e_L \in \mathbf{e_c}} \Pr(e_L|L_i = \mathrm{T}) \prod_{e_R \in \mathbf{e_c}} \Pr(e_R|R_i = \mathrm{T}) \prod_{e_{LR} \in \mathbf{e_c}} \Pr(e_{LR}|L_i = \mathrm{T}, R_i = \mathrm{T}) \tag{2.13}$$

There are two part-specific evidence metrics for each of the left and right part-hypotheses and four joint evidence metrics for the hypothesis as a whole as defined in Section 2.3, resulting in 8 conditional probabilities. These 8 conditional probability distributions can be determined through training, and so equation (2.13) can be evaluated for $\Pr(\mathbf{e_c}|\mathbf{x}_i)$.

Prior Evidence Probability

The value of $\Pr(\mathbf{e_c})$ is evaluated by marginalizing the conditional probabilities evaluated in equation (2.13) over all possible values of $\mathbf{x}_i$ with our prior estimates of $\mathbf{x}_i$ coming from transitional and past evidence:

$$\Pr(\mathbf{e_c}) = \sum_{\mathbf{x}_i} \Pr(\mathbf{e_c}|\mathbf{x}_i) \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}) \tag{2.14}$$

As $\mathbf{x}_i$ is a binary random 2-vector, the summation in equation (2.14) consists of 4 summands: $\{(L_i = \mathrm{T}, R_i = \mathrm{T}), (L_i = \mathrm{T}, R_i = \mathrm{F}), (L_i = \mathrm{F}, R_i = \mathrm{T}), (L_i = \mathrm{F}, R_i = \mathrm{F})\}$. As with equation (2.13), the conditional probability distributions required to evaluate equation (2.14) can be determined through training, and so provided the prior estimates are available for each of the 4 summands, equation (2.14) can be evaluated for $\Pr(\mathbf{e_c})$.

Missing Parts

It is impossible to compute evidence metrics for a part-hypothesis that doesn't exist, and so hypotheses consisting of missing parts require slightly different evaluation, with the

conditional evidence distributions $\Pr(\mathbf{e_c}|\mathbf{x}_i)$ reduced from their full form in equation (2.13) to only consist of the terms associated exclusively with the non-missing part. For example, for a hypothesis where the right part is missing, equation (2.13) becomes

$$\Pr(\mathbf{e_c}|\mathbf{x}_i) = \prod_{e_L \in \mathbf{e_c}} \Pr(e_L|L_i = \mathrm{T})$$

and equation (2.14) is unchanged.

2.4.2 Evidence Probabilities From Tracking Components

Past evidence and transitional evidence are incorporated into hypothesis assessment through the term $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p})$. Mapping each past hypothesis likelihood to the current hypothesis' likelihood is difficult however as any summation of past likelihoods over the past hypotheses will not be a strict probability as it will not necessarily sum to one. Instead, mapping from past hypotheses to the current hypothesis is done by the inclusion of an additional random variable, $\mathbf{H} = \mathbf{h}_k$, where $\mathbf{H}$ represents the previous frame's true lane marking pair, and $\mathbf{h}_k$ represents the k^{th} hypothesis for that lane marking pair in the previous frame. It is important to note that $\Pr(\mathbf{H} = \mathbf{h}_k)$ has a very different meaning from $\Pr(\mathbf{x}_i)$. While $\mathbf{x}_i$ stands for $\{L_i = \mathrm{T}, R_i = \mathrm{T}\}$, $\mathbf{H} = \mathbf{h}_k$ instead stands for $\{L = L_k, R = R_k\}$ in the previous frame. Marginalizing and conditioning $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p})$ over $\mathbf{H} = \mathbf{h}_k$ yields:

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}) = \sum_k \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}, \mathbf{H} = \mathbf{h}_k) \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_t}, \mathbf{e_p}) \tag{2.15}$$

Next, it is assumed that $\mathbf{e_t}$ gives no information about $\mathbf{H} = \mathbf{h}_k$ if the past evidence $\mathbf{e_p}$ is already given unless it also has information from the current frame to link the present to the past, implying:

$$\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_t}, \mathbf{e_p}) = \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p}) \tag{2.16}$$

The Markov assumption, that the history is conditionally independent given a previous state, is then applied which means that $\mathbf{e_p}$ gives no information about $\mathbf{x}_i$ if $\mathbf{H} = \mathbf{h}_k$ is already given as any past information that $\mathbf{e_p}$ would contribute has already been incorporated in evaluating $\mathbf{H} = \mathbf{h}_k$, implying

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}, \mathbf{H} = \mathbf{h}_k) = \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k) \tag{2.17}$$

Applying equations (2.16) and (2.17) to (2.15) then yields:

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}) = \sum_k \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k) \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p}) \tag{2.18}$$

Equation (2.18) again separates into two components. The transitional evidence linking hypotheses $\mathbf{x}_i$ to the past hypothesis $\mathbf{h}_k$ is captured in $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k)$, while the past evidence supporting the past hypothesis $\mathbf{h}_k$ is captured in $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$.

Equation (2.18) can be interpreted as propagating each evaluated past probability, $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$, to the current hypothesis likelihood prior $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p})$ using the transition probabilities $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k)$. Note that since the set of $\mathbf{h}_k$ past hypotheses include missing part hypotheses, the summation spans $\mathbf{H}$ and $\sum_k \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p}) = 1$. This interpretation is illustrated in Figure 2.7.

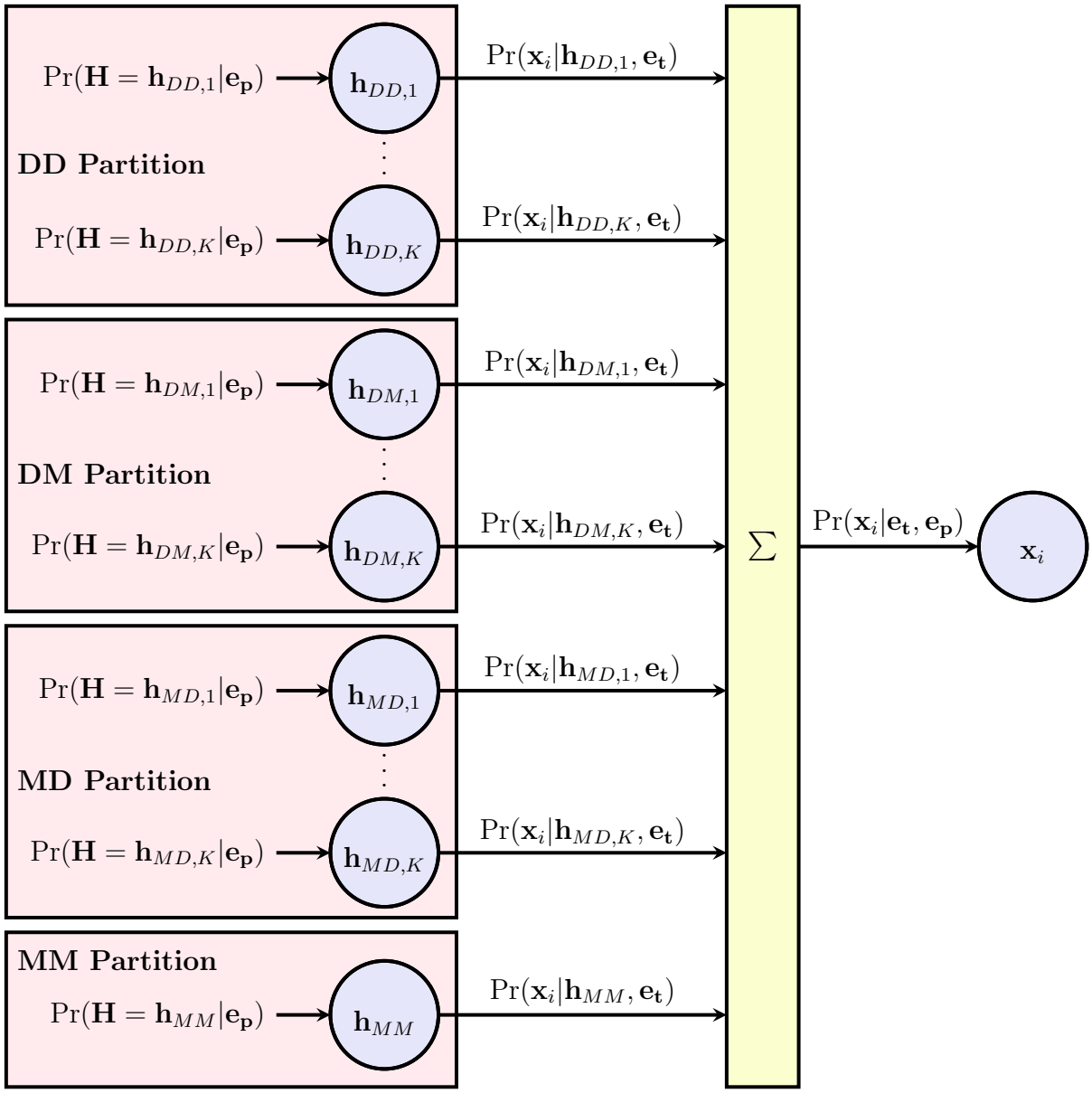


Figure 2.7: Hypothesis Tracking: Hypothesis i in the current frame is compared against all of the retained hypotheses from the previous frame. The past possibilities for $\mathbf{H}$, shown in the circles on the left, are separated into partitions corresponding to each part being either missing (M) or non-missing (D). The union of these partitions span $\mathbf{H}$ and therefore the sum of their probabilities given $\mathbf{e_p}$ is one. Here, it is assumed that K past hypotheses from each partition other than MM are retained for tracking.

Transitional Evidence

The transitional evidence $\mathbf{e_t}$ represents how compatible a current frame hypothesis is with a specific hypothesis from the past frame independently of their respective frame's evidence. This evidence aims to answer the question "How likely is it that these hypotheses represent a transition between frames of a single object, marking, or feature?"

The transitional evidence is incorporated through $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k)$, which is evaluated by first changing the variable $\mathbf{H} = \mathbf{h}_k$ to match the likelihood approach used by the overall system. Here, the conditioned variable can be directly changed from $\mathbf{H} = \mathbf{h}_k$ to $\mathbf{h}_k$ because if $\mathbf{H} = \mathbf{h}_k$ is already given, then it is also known that $\mathbf{h}_k$ is true. The part-hypotheses are then separated by assuming that the individual j parts of a hypothesis have independent tracking histories:

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k) = \prod_{j \in L,R} \Pr(x_{i,j}|h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j}) \tag{2.19}$$

where $x_{i,j}$ represents the validity of the j^{th} part-hypothesis of hypothesis $\mathbf{x}_i$, and, $h_{k,j}$ represents the previous part-hypothesis that hypothesis k associated with part j, and $\mathbf{e}_{\mathbf{t}_j}$ represents the transitional evidence vector linking the two part-hypotheses $x_{i,j}$ and $h_{k,j}$.

The conditional distribution $\Pr(x_{i,j}|h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j})$ is then reorganized using Bayes' rule:

$$\Pr(x_{i,j}|h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j}) = \frac{\Pr(\mathbf{e}_{\mathbf{t}_j}|x_{i,j}, h_{k,j} = \mathrm{T}) \Pr(x_{i,j}|h_{k,j} = \mathrm{T})}{\Pr(\mathbf{e}_{\mathbf{t}_j}|h_{k,j} = \mathrm{T})} \tag{2.20}$$

where the conditional distribution $\Pr(\mathbf{e}_{\mathbf{t}_j}|x_{i,j}, h_{k,j} = \mathrm{T})$ is evaluated by assuming that the components of $\mathbf{e}_{\mathbf{t}_j}$, the tracking evidence metrics, give no information about each other if the validity of the associated part hypotheses are already known:

$$\Pr(\mathbf{e}_{\mathbf{t}_j}|x_{i,j}, h_{k,j} = \mathrm{T}) = \prod_{e^j \in \mathbf{e}_{\mathbf{t}_j}} \Pr(e^j|x_{i,j}, h_{k,j} = \mathrm{T}) \tag{2.21}$$

There are 3 evidence metrics associated with each potential track, resulting in 6 conditional distributions to determine through training.

The term $\Pr(\mathbf{e}_{\mathbf{t}_j}|h_{k,j} = \mathrm{T})$ is determined by marginalizing over the two possible values that the binary random variable $x_{i,j}$ can take:

$$\begin{aligned} \Pr(\mathbf{e}_{\mathbf{t}_j}|h_{k,j} = \mathrm{T}) = {} & \Pr(\mathbf{e}_{\mathbf{t}_j}|x_{i,j} = \mathrm{T}, h_{k,j} = \mathrm{T}) \Pr(x_{i,j} = \mathrm{T}|h_{k,j} = \mathrm{T}) \\ & + \Pr(\mathbf{e}_{\mathbf{t}_j}|x_{i,j} = \mathrm{F}, h_{k,j} = \mathrm{T}) \Pr(x_{i,j} = \mathrm{F}|h_{k,j} = \mathrm{T}) \end{aligned} \tag{2.22}$$

The last component of equation (2.20) to evaluate is $\Pr(x_{i,j}|h_{k,j} = \mathrm{T})$. As there is no evidence taken into account, this term is purely a prior probability. Depending on the presence of missing parts, $\Pr(x_{i,j}|h_{k,j} = \mathrm{T})$ is assumed to be a constant for all tracked hypotheses belonging to part j. Note here that the case of a missing part includes the possibility of the part being misdetected. Since either the current part hypothesis or past part hypothesis could be missing or non-missing, there are four cases each with their own constant value for $\Pr(x_{i,j}|h_{k,j} = \mathrm{T})$. There are dependencies between these four constants, and they can be reduced to three parameters by marginalizing over another binary variable, X_j, with $X_j = \phi$ or $X_j \neq \phi$ representing the case where part j is truly missing or truly not missing respectively. Conditioning on X_j allows the cases where a non-missing part hypothesis $x_{i,j}$ is false to be separated into cases where part hypothesis $x_{i,j}$ is false because it is a poor hypothesis and cases where part hypothesis $x_{i,j}$ is false because part j is missing or misdetected.

For the case where both the current part hypothesis is non-missing ($x_{i,j} \neq \phi$) and the past part-hypotheses is non-missing ($h_{k,j} \neq \phi$):

$$\begin{aligned}\Pr(x_{i,j}|h_{k,j} = \mathrm{T}) \;=\; & \Pr(x_{i,j}|X_j \neq \phi, h_{k,j} = \mathrm{T}) \Pr(X_j \neq \phi|h_{k,j} = \mathrm{T}) \\ & + \Pr(x_{i,j}|X_j = \phi, h_{k,j} = \mathrm{T}) \Pr(X_j = \phi|h_{k,j} = \mathrm{T}) \end{aligned} \tag{2.23}$$

Note that $\Pr(x_{i,j} = \mathrm{T}|X_j = \phi) = 0$ for $x_{i,j} \neq \phi$ since a part can't simultaneously have a true non-missing hypothesis and be truly missing, which yields:

$$\Pr(x_{i,j} = \mathrm{T}|h_{k,j} = \mathrm{T}) = \Pr(x_{i,j} = \mathrm{T}|X_j \neq \phi, h_{k,j} = \mathrm{T}) \Pr(X_j \neq \phi|h_{k,j} = \mathrm{T}) \tag{2.24}$$

For the case where the past part-hypothesis is missing ($h_{k,j} = \phi$) and the current part hypothesis is non-missing ($x_{i,j} \neq \phi$), one can similarly obtain:

$$\Pr(x_{i,j} = \mathrm{T}|h_{k,j} = \phi) = \Pr(x_{i,j} = \mathrm{T}|X_j \neq \phi, h_{k,j} = \phi) \Pr(X_j \neq \phi|h_{k,j} = \phi) \tag{2.25}$$

Since a missing past part hypothesis does not generate any transitional evidence metrics, $h_{k,j} = \phi$ does not contribute any information about a non-missing current part hypothesis if it is already known that the part is not truly missing in the current frame, which yields:

$$\Pr(x_{i,j} = \mathrm{T}|h_{k,j} = \phi) = \Pr(x_{i,j} = \mathrm{T}|X_j \neq \phi) \Pr(X_j \neq \phi|h_{k,j} = \phi) \tag{2.26}$$

For the case where the current part-hypothesis is missing ($x_{i,j} = \phi$) and the past part hypothesis is non-missing ($h_{k,j} \neq \phi$) , instead note that $\Pr(x_{i,j} = \phi|X_j \neq \phi, h_{k,j} = \mathrm{T}) = 0$ since a part can't be simultaneously missing and non-missing, and also that $\Pr(x_{i,j} =$

$\phi|X_j = \phi) = 1$ since if a part is known to truly be missing, then the missing part hypothesis must be true to similarly arrive at:

$$\Pr(x_{i,j} = \phi|h_{k,j} = \mathrm{T}) = \Pr(X_j = \phi|h_{k,j} = \mathrm{T}) \tag{2.27}$$

Lastly, for the case where both the current part hypothesis is missing ($x_{i,j} = \phi$) and that past part hypothesis is missing ($h_{k,j} = \phi$), again $\Pr(x_{i,j} = \phi|X_j \neq \phi, h_{k,j} = \mathrm{T}) = 0$ and $\Pr(x_{i,j} = \phi|X_j = \phi) = 1$ which yield:

$$\Pr(x_{i,j} = \phi|h_{k,j} = \phi) = \Pr(X_j = \phi|h_{k,j} = \phi) \tag{2.28}$$

Equations (2.24), (2.26), (2.27), and (2.28) contain six terms. In [19] the assumption that $h_{k,j}$ gives no information about $x_{i,j}$ if $X_j \neq \phi$ is given allows $\Pr(x_{i,j}|X_j \neq \phi, h_{k,j} = \mathrm{T})$ to be replaced by $\Pr(x_{i,j}|X_j \neq \phi)$ in (2.24) reducing the list to five terms. Since X_j is a binary variable, the five remaining terms are then reduced to three independent terms by the law of total probability:

$$\Pr(X_j = \phi|h_{k,j} = \phi) + \Pr(X_j \neq \phi|h_{k,j} = \phi) = 1 \tag{2.29}$$

$$\Pr(X_j \neq \phi|h_{k,j} = \mathrm{T}, h_{k,j} \neq \phi) + \Pr(X_j = \phi|h_{k,j} = \mathrm{T}, h_{k,j} \neq \phi) = 1 \tag{2.30}$$

The terms $\Pr(x_{i,j}|X_j \neq \phi)$, $\Pr(X_j \neq \phi|h_{k,j} = \phi)$, and $\Pr(X_j = \phi|h_{k,j} = \mathrm{T}, h_{k,j} \neq \phi)$ are then used as tracking parameters in [19].

The term $\Pr(x_{i,j}|X_j \neq \phi)$ represents the prior probability of any non-missing part-hypothesis given that the part is truly non-missing. This parameter is set to a high value 0.9999 in [19].

The term $\Pr(X_j \neq \phi|h_{k,j} = \phi)$ represents the prior probability of a non-missing part-hypothesis to emerge in the current frame from a missing part-hypothesis in the previous frame. This parameter determines how quickly the system can respond to emerging part-hypotheses, or conversely how sensitive it is to noise and is set to 0.1 in [19].

The term $\Pr(X_j = \phi|h_{k,j} = \mathrm{T}, h_{k,j} \neq \phi)$ represents the prior probability of a disappearing part-hypothesis, i.e. a part that has a true current part-hypothesis of missing, but a non-missing part-hypothesis in the previous frame and is set to $1e^{-8}$ in [19].

In Chapter 3, an alternative discussion of the three parameters $\Pr(x_{i,j}|X_j \neq \phi)$, $\Pr(X_j \neq \phi|h_{k,j} = \phi)$, and $\Pr(X_j = \phi|h_{k,j} = \mathrm{T}, h_{k,j} \neq \phi)$ will be introduced. Instead of selecting the parameter values by tuning or inspection, their values will be derived from desired functionality or learned directly from training.

Past Evidence

The past evidence $\mathbf{e_p}$ represents all evidence received prior to the current frame, and so includes the previous frame's $\mathbf{e_c}$, $\mathbf{e_t}$ and $\mathbf{e_p}$, with the previous frame's $\mathbf{e_p}$ itself capturing all evidence prior to the previous frame. The past evidence $\mathbf{e_p}$ is incorporated through $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$. Additional assumptions are required however, since for any given hypothesis in a specific frame, it is the likelihood estimate $\Pr(\mathbf{h}_k = \mathrm{T}|\mathbf{e_p})$ that is evaluated, not the required $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$. This difference does not hold for the missing part hypotheses however, because the cases of part j being truly missing and a true missing part hypothesis for part j imply each other.

The desired $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$ can be approximated from the likelihood estimates by first partitioning $\mathbf{H}$ based on the missing and detected parts of the hypothesis of interest, $\mathbf{H}_k = \{\mathbf{D}_k, \mathbf{M}_k\}$, where $\mathbf{D}_k$ and $\mathbf{M}_k$ represent the set of detected and missing parts for the k^{th} past hypothesis respectively. Under this partition the desired term is written:

$$\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p}) = \Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{2.31}$$

Equation (2.31) can be evaluated using an assumption that the ratio between the probability that the past frame's true lane marking pair was hypothesis k and the probability that the true lane marking pair's partition of missing and non-missing parts matched that of hypothesis k is the same as the ratio between the likelihood of hypothesis k and the sum of the likelihoods of all hypotheses sharing the partition of missing and non-missing parts of hypothesis k:

$$\frac{\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p})}{\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})} \approx \frac{\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})}{\sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})} \tag{2.32}$$

where $\mathbf{d}_i$ are all selected hypotheses from the previous frame that share hypothesis k's set of detected and missing parts. In words, it is assumed that the ratio between the chance that the detected part-hypotheses are the correct detections and the chance that the detected parts are indeed detected is equal to the ratio between the likelihood of the hypothesis and the total likelihood of all hypotheses sharing that same partition of detected and missing parts.

The desired past probability $\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p})$ is then evaluated through the resulting approximation

$$\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \frac{\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})}{\sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})} \tag{2.33}$$

where the terms $\Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$ and $\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$ are taken directly from their corresponding $\Pr(\mathbf{x}_i|\text{Evidence})$ values in the previous frame. The $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ term is evaluated through its complement and then by marginalizing over any available hypothesis:

$$\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) = \Pr(\mathbf{M}_k = \phi|\mathbf{e_p}) - \Pr(\mathbf{D}_k = \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{2.34}$$

with

$$\Pr(\mathbf{M}_k = \phi|\mathbf{e_p}) = \Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) + \Pr(\mathbf{d}_k = \mathrm{F}, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{2.35}$$

where the terms $\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$ and $\Pr(\mathbf{d}_k = \mathrm{F}, \mathbf{M}_k = \phi|\mathbf{e_p})$ are again taken directly from their corresponding $\Pr(\mathbf{x}_i|\text{Evidence})$ values in the previous frame.

2.4.3 Training Procedure and Limitations

Sections 2.4.1 and 2.4.2 provide a method for evaluating the likelihood of a lane hypothesis based on its associated evidence metrics and the conditional distributions of the heuristic metrics. As [19] notes, the validity of a hypothesis can be directly judged visually by a human observer, and so there are no hidden variables preventing the conditional distributions from being learned from a labelled training set. Labelling can be very labour intensive however, and so to reduce the required labelling effort, [19] provides a method for training the conditional distributions. The training method assumes that the hypotheses for an image i having the highest and lowest likelihoods of being correct can be taken as ground truth example hypotheses, $h_{GT,i}$ and $h_{GF,i}$ respectively. The full algorithm is defined in Algorithm 1. The key requirement of the algorithm is that the initial parameter estimate $\boldsymbol{\theta}_0$ must produce classifications where hypotheses with the maximum likelihood are truly correct and hypotheses with the minimum likelihood are truly false, i.e. the initial parameter estimate must classify extreme examples correctly.

The limitation of the training approach of [19] is that it is often impossible to provide an initial parameter estimate that can consistently classify extreme examples correctly. Even if the complication is merely due to a larger view of a simple scenario it can be impossible to provide an initial distribution solely from intuition. For example, consider the common situation of multiple lanes shown in Fig. 2.8 with an image from the KITTI Roads Data Set [11]. The blue rectangle represents the approximate BEV area evaluated in [19]'s original work and contains only markings relevant to the ego lane. In contrast, the image as a whole represents the larger BEV area evaluated more recently in [11]. The larger BEV area contains many more markings for not just other vehicle lanes but also for

Algorithm 1 Training Process of [19]: Assuming that a set of lane marking hypotheses h from a training set of images are divisible into True (T) and False (F) classes with evidence metrics $\mathbf{X}$, learn the parameters $\boldsymbol{\theta}$ of their respective conditional evidence distributions $\Pr(\mathbf{X}|\boldsymbol{\theta}, \mathrm{T})$, and $\Pr(\mathbf{X}|\boldsymbol{\theta}, \mathrm{F})$ as well as their prior probabilities $\Pr(\mathrm{T}|\boldsymbol{\theta})$ and $\Pr(\mathrm{F}|\boldsymbol{\theta})$

$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}_0$
repeat
  for all hypotheses in training set **do**
    Update $\Pr(\mathrm{T}|\mathbf{X}, \boldsymbol{\theta})$ and $\Pr(\mathrm{F}|\mathbf{X}, \boldsymbol{\theta})$ using 2.11
  end for
  for all images i in training set **do**
    $h_{GT,i} = \arg\max_{h_i} \Pr(\mathrm{T}|\mathbf{X}, \boldsymbol{\theta})$
    $h_{GF,i} = \arg\max_{h_i} \Pr(\mathrm{F}|\mathbf{X}, \boldsymbol{\theta})$
  end for
  $\boldsymbol{\theta} \leftarrow \arg\max_{\boldsymbol{\theta}} \{\sum_{h_{GT,i}} \ln \Pr(\mathbf{X}|\boldsymbol{\theta}, \mathrm{T}) + \sum_{h_{GF,i}} \ln \Pr(\mathbf{X}|\boldsymbol{\theta}, \mathrm{F})\}$
until convergence of $\boldsymbol{\theta}$

the separated bicycle lane that runs alongside the roadway to the right. The correct lane markings for the ego lane are contained within the green boxes and lie within the smaller BEV area, while a competing false hypothesis introduced in the larger BEV area consists of the markings in the red boxes.

Providing an initial distribution estimate that correctly classifies the green hypothesis as true without classifying the red hypothesis as true in 2.8 requires the solution of a challenging trade-off. As the left marking of the red hypothesis is solid, it has much stronger marking support than the dotted left marking of the green hypothesis. Conversely, as the red hypothesis represents two lanes instead of one, the red hypothesis has a much more severe lane width penalty. Tightening the lane width distributions too much has the result of favouring hypotheses with poorer lane marking support but ideal lane widths, while not tightening the lane width distributions enough favours the red hypothesis in this example over the green. Balancing these competing metrics is not trivial, even for this relatively simple situation. The challenge of providing a valid initialization for the training method provided in [19] makes training very challenging if not impossible, and significantly limits the overall method as a whole, restricting its use to simpler situations and smaller BEV ranges and rendering it unusable for most real-world driving scenarios.

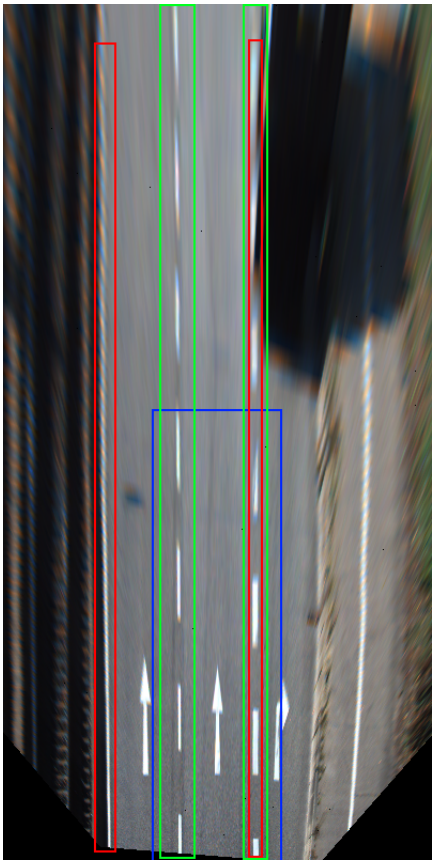


Figure 2.8: Competing Metrics: The blue rectangle indicates the approximate field of view used in [19] while the image as a whole indicates the field of view used in [11]. The true hypothesis is enclosed in green and has a superior lane width score. The false hypothesis is enclosed in red and has a superior lane marking score as its left marking is solid instead of dotted.

2.5 Expectation Maximization

When seeking maximum likelihood estimates of model parameters, a common complication is the presence of unobserved latent variables. Consider the log-likelihood function for a set of observed data $\mathbf{X}$ given model parameters $\boldsymbol{\theta}$

$$L(\mathbf{X}|\boldsymbol{\theta}) = \ln \Pr(\mathbf{X}|\boldsymbol{\theta}) \tag{2.36}$$

and now involving unobserved latent variables $\mathbf{Z}$ which require marginalization

$$L(\mathbf{X}|\boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} \Pr(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \right\} \tag{2.37}$$

The summation within the logarithm in equation (2.37) often renders the maximization of the log-likelihood function (2.36) with respect to $\boldsymbol{\theta}$ intractable. The Expectation Maximization (EM) algorithm provides an iterative solution to such a problem by iteratively maximizing the expected likelihood $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ [6], which is given by:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}} \Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}}) \ln[\Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) \Pr(\mathbf{Z}|\boldsymbol{\theta})] \tag{2.38}$$

Note that the discretization widths used to evaluate the probability density functions for continuous random variables are dropped because they are constant relative to the maximization of the maximum likelihood estimate. The EM algorithm is detailed in Algorithm 2.

Algorithm 2 Expectation Maximization: Given joint distributions $\Pr(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ over observed data $\mathbf{X}$ and unobserved latent variables $\mathbf{Z}$, maximize the likelihood function with respect to the model parameters $\boldsymbol{\theta}$.

1: $\boldsymbol{\theta}_{\text{old}} \leftarrow \boldsymbol{\theta}_{\text{initial}}$
2: **repeat**
3: E Step: Compute $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}})$
4: M Step: $\boldsymbol{\theta}_{\text{new}} \leftarrow \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$
5: $\boldsymbol{\theta}_{\text{old}} \leftarrow \boldsymbol{\theta}_{\text{new}}$
6: **until** convergence of $\boldsymbol{\theta}$

where the expectation, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$, is given by:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}} \Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}}) \ln[\Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) \Pr(\mathbf{Z}|\boldsymbol{\theta})] \tag{2.39}$$

The EM algorithm has two significant advantages. First, EM is guaranteed to converge to a local optimum [6]. Second, if a prior $\Pr(\boldsymbol{\theta})$ is defined over the model parameters, then augmenting the maximization in step 4 as

$$\boldsymbol{\theta}_{\text{new}} \leftarrow \arg\max_{\boldsymbol{\theta}} \{Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) + \ln(\boldsymbol{\theta})\} \tag{2.40}$$

produces a local optimum that is also a maximum posterior (MAP) solution [6].

The EM algorithm can be particularly effective for mixture model problems, where multiple models are used to represent the observed data, with the assignment of each data point to its generating model treated as a latent variable. For mixture model problems equation (2.39) can be reformulated as:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \phi_{i,k} \{\ln \Pr(x_i|\mathbf{Z}, \boldsymbol{\theta}) + \ln \Pr(\mathbf{Z}|\boldsymbol{\theta})\} \tag{2.41}$$

where $\phi_{i,k}$ is the value of $\Pr(\mathbf{Z}|x_i, \boldsymbol{\theta}_{\text{old}})$ for data point x_i with model $\mathbf{Z} = \mathbf{z}_k$. For mixture model problems, $\phi_{i,k}$ is the ownership of data point x_i by model k and is obtained in the E-step of the algorithm. Distributing $\phi_{i,k}$ and noting that $\Pr(\mathbf{Z}|\boldsymbol{\theta})$ is a blind prior independent of the observed data and ownership:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}=\mathbf{z}_k} \ln \Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) \sum_i \phi_{i,k} + \sum_{\mathbf{Z}=\mathbf{z}_k} \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) \sum_i \phi_{i,k} \tag{2.42}$$

EM can be particularly convenient when the conditional distributions $\Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta})$ are members of the exponential family which have the convenient property that the log probability can be written in a form where the data are separable from the distribution parameters.

For an exponential family distribution of the form

$$f(x|\boldsymbol{\theta}) = h(x)g(\boldsymbol{\theta}) \exp(\boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(x)) \tag{2.43}$$

where x is the random variable, $\boldsymbol{\theta}$ are the distribution parameters, and $h, g, \boldsymbol{\eta}$ and $\mathbf{T}$ are functions then

$$\ln f(x|\boldsymbol{\theta}) = \ln h(x) + \ln g(\boldsymbol{\theta}) + \boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(x) \tag{2.44}$$

and therefore the total data log-likelihood of all of the independent identically distributed

x_i variables becomes

$$\ln \prod_i f(x_i|\boldsymbol{\theta}) = \sum_i \ln f(x_i|\boldsymbol{\theta}) \tag{2.45}$$

$$= \sum_i (\ln h(x_i) + \ln g(\boldsymbol{\theta}) + \boldsymbol{\eta}(\boldsymbol{\theta}) \cdot \mathbf{T}(x_i)) \tag{2.46}$$

$$= \sum_i \ln h(x_i) + \sum_i \ln g(\boldsymbol{\theta}) + \boldsymbol{\eta}(\boldsymbol{\theta}) \sum_i \mathbf{T}(x_i) \tag{2.47}$$

Equation (2.47) decouples the total data from the parameters, requiring only the sufficient statistics of the data, $\sum_i \ln h(x_i)$ and $\sum_i \mathbf{T}(x_i)$, and not the data itself when evaluating the log likelihood of the data for different values of the distribution parameters $\boldsymbol{\theta}$. Applying equation (2.47) to equation (2.42) produces:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \left\{ \phi_{i,k} \ln h(x_i) + \ln g(\boldsymbol{\theta}) \phi_{i,k} + \boldsymbol{\eta}(\boldsymbol{\theta}) \phi_{i,k} \mathbf{T}(x_i) + \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) \right\} \tag{2.48}$$

With equation (2.48), both the ownership values from the E-step and the total data are separated from the parameters $\boldsymbol{\theta}$ used to maximize $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ in the M-Step. Such a separation allows an arbitrary number of data points to be compressed into the terms $\sum_i \phi_{i,k} \ln h(x_i)$, $\sum_i \phi_{i,k}$, and $\sum_i \phi_{i,k} \mathbf{T}(x_i)$, with the maximization becoming independent of the size of the data set and therefore dramatically reducing the time required by the M-step.

For example, consider a mixture model problem that is a fitting of two Gamma Distributions with parameters $\mathbf{z}_1 = \{\alpha_1, \beta_1\}$ and $\mathbf{z}_2 = \{\alpha_2, \beta_2\}$ where the gamma distribution

$$f(x|\boldsymbol{\theta}) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \tag{2.49}$$

can be rewritten in the form of

$$f(x|\boldsymbol{\theta}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp((\alpha - 1) \ln x - \beta x) \tag{2.50}$$

The gamma distribution then has the following $h, g, \boldsymbol{\eta}$ and $\mathbf{T}$ functions

$$\begin{aligned} h(x) &= 1 && (2.51)\\ g(\boldsymbol{\theta}) &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} && (2.52)\\ \boldsymbol{\eta}(\boldsymbol{\theta}) &= \begin{bmatrix} \alpha - 1 \\ -\beta \end{bmatrix} && (2.53)\\ \mathbf{T}(x) &= \begin{bmatrix} \ln(x) \\ x \end{bmatrix} && (2.54) \end{aligned}$$

Then $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ for the M-step then becomes:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) &= C_1 \ln(\frac{{\beta_1}^{\alpha_1}}{\Gamma(\alpha_1)}) + \begin{bmatrix} \alpha_1 - 1 \\ -\beta_1 \end{bmatrix} \cdot \mathbf{S}_1 + C_1 \ln \Pr(\mathbf{Z} = \mathbf{z}_1 | \boldsymbol{\theta}) \\ &\quad + C_2 \ln(\frac{{\beta_2}^{\alpha_2}}{\Gamma(\alpha_2)}) + \begin{bmatrix} \alpha_2 - 1 \\ -\beta_2 \end{bmatrix} \cdot \mathbf{S}_2 + C_2 \ln \Pr(\mathbf{Z} = \mathbf{z}_2 | \boldsymbol{\theta}) \qquad (2.55) \end{aligned}$$

with C_k being the total estimated membership of model $\mathbf{z}_k$:

$$C_k = \sum_i \phi_{i,k} \qquad (2.56)$$

and $\mathbf{S}_k$ being the membership-weighted sufficient statistics used by model $\mathbf{z}_k$:

$$\mathbf{S}_k = \begin{bmatrix} \sum_i \phi_{i,k} \ln x_i \\ \sum_i \phi_{i,k} x_i \end{bmatrix} \qquad (2.57)$$

where both C_k and $\mathbf{S}_k$ are constant in the M-Step.

Similarly, consider another example fitting two Exponential Distributions with parameters $\mathbf{z}_1 = \{\lambda_1\}$ and $\mathbf{z}_2 = \{\lambda_2\}$ where the exponential distribution

$$f(x|\boldsymbol{\theta}) = \lambda \exp(-\lambda x) \qquad (2.58)$$

has the following $h, g, \boldsymbol{\eta}$ and $\mathbf{T}$ functions

$$\begin{aligned} h(x) &= 1 && (2.59)\\ g(\boldsymbol{\theta}) &= \lambda && (2.60)\\ \boldsymbol{\eta}(\boldsymbol{\theta}) &= -\lambda && (2.61)\\ \mathbf{T}(x) &= x && (2.62) \end{aligned}$$

Then $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ for the M-step becomes:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) &= C_1 \ln \lambda_1 + -\lambda_1 \cdot \mathbf{S}_1 + C_1 \ln \Pr(\mathbf{Z} = \mathbf{z}_1 | \boldsymbol{\theta}) \\ &\quad + C_2 \ln \lambda_2 + -\lambda_2 \cdot \mathbf{S}_2 + C_2 \ln \Pr(\mathbf{Z} = \mathbf{z}_1 | \boldsymbol{\theta}) \end{aligned} \tag{2.63}$$

with C_k being the total estimated membership of model $\mathbf{z}_k$:

$$C_k = \sum_i \phi_{i,k} \tag{2.64}$$

and $\mathbf{S}_k$ being the membership-weighted sufficient statistics used by model $\mathbf{z}_k$:

$$\mathbf{S}_k = \sum_i \phi_{i,k} x_i \tag{2.65}$$

where both C_k and $\mathbf{S}_k$ are constant in the M-Step.

This chapter covered the background information required for the contributions made by this work including the generation of BEV images, lane marking hypothesis generation, heuristic hypothesis features, Bayesian hypothesis evaluation and Expectation Maximization with exponential family distributions. These concepts provide the foundation for the contributions that follow in Chapter 3.

Chapter 3

Generalized Bayesian Detection and Tracking

Building on the efforts of [19], this work makes several contributions to the detection and tracking of lane markings. First, the training problem in [19] is completely reformulated as a mixture-model problem and solved using the Expectation Maximization (EM) algorithm. As a consequence of the reformulation, the tracking parameters that required tuning in [19] can be replaced by values determined by EM or derived from desired functionality. The resulting training process is robust with respect to initial parameter estimates, and requires very little human labelling effort; the lane model used in [19] converges without requiring any human labelling when using the proposed training process. Second, the Bayesian Lane Detection approach proposed in [19] is generalized and extended for the detection of multiple different lane marking configurations. Third, a stereo filter is proposed to reduce the impact of false alarms caused by out-of-plane features such as other vehicles.

3.1 Revisiting Missing and Misdetected Parts

There are limitations to the past evidence derivation as it concerns a pair of edge cases. The first provides the possibility for the most likely hypothesis returned by the system to be more likely false than true while simultaneously more likely than the misdetection case, while the second produces erroneous results when there are very few hypotheses available from the previous frame.

The first limitation is that the formulation in [19] allows for the most likely hypothesis returned by the system to be more likely false than true while simultaneously more likely

than the misdetection case. The probability that a part is missing is allowed to fluctuate over time based on the observed evidence and the parameters governing the changes of disappearing or emerging parts. This fluctuation intuitively makes sense, as lane markings don't just emerge and disappear from frame to frame at random. However, the null hypotheses aren't defined to solely include the possibility that a part is truly missing - they also include the possibility that a part is present but was misdetected. Consider then a hypothetical hypothesis $\mathbf{x}_F$ that isn't very strong and has likelihoods of:

$$\Pr(\mathbf{x}_F = \mathrm{T}|\mathbf{e}) = 0.2 \tag{3.1}$$
$$\Pr(\mathbf{x}_F = \mathrm{F}|\mathbf{e}) = 0.8 \tag{3.2}$$

Since $\Pr(\mathbf{x} = \phi)$ is free to fluctuate over time, $\Pr(\mathbf{x} = \phi)$ may very well have a value below 0.2. If so, a peculiar contradiction is arrived at when considering that the null hypothesis $\mathbf{x} = \phi$ is supposed to cover both misdetections and missing parts:

$$\Pr(\mathbf{x}_F = \mathrm{T}|\mathbf{e}) > \Pr(\mathbf{x} = \phi|\mathbf{e}) \tag{3.3}$$
$$\Pr(\mathbf{x}_F = \mathrm{F}|\mathbf{e}) > \Pr(\mathbf{x}_F = \mathrm{T}|\mathbf{e}) \tag{3.4}$$

In words, a hypothesis that is more likely false than true could be more likely than the estimated probability of a misdetection. This contradiction is not a problem so long as there is a stronger hypothesis more likely to be true than false available for the system to return as the most likely hypothesis. In some cases however, particularly in very cluttered scenes, if RANSAC is unable to pick a correct set of inliers to produce a valid hypothesis, a very poor hypothesis may be returned instead of a misdetection. Alternatively, when a part is emerging or disappearing, the most likely hypothesis may be more likely false than true during the time that the tracking components take to respond to the change in the probability of a missing part. Both sources for such a contradiction were observed during initial implementations.

A second limitation lies in the fact that the approximation used in 2.33 appears to break down when the set of detected hypotheses in the previous frame contain a hypothesis k that satisfies the approximation

$$\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{3.5}$$

such as may occur if either the population of past hypotheses in the summation is decreased to one or if there is a past hypothesis in the summation that is dramatically more likely than all of the others:

$$\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \gg \Pr(\mathbf{d}_{i\neq k} = \mathrm{T}, \mathbf{M}_{i\neq k} = \phi|\mathbf{e_p}) \tag{3.6}$$

If the approximation in equation (3.5) is satisfied, then the estimated past probability for hypothesis k from equation 2.33 becomes:

$$\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{3.7}$$

In a similar vein as the first limitation, depending on the likelihood of hypothesis k and value of $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$, equation (3.7) may not be an appropriate approximation, particularly if hypothesis k's previous likelihood $\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$ is significantly lower than $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$. This can result in poor hypotheses generated from features present in the image, such as cracks in the road, producing unreasonably high past probabilities and tracking strongly from frame to frame despite poor frame-specific evidence. This phenomenon was also observed during implementation of the method of [19], specifically in image sequences without any true lane markings present and in conjunction with the previous problem produced most likely lane marking estimates that were both much more likely to be false than true as well as very persistent from frame to frame.

3.1.1 Past Probability Approximation

The assumption made in [19], that

$$\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \frac{\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})}{\sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})} \tag{3.8}$$

breaks down when $\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$ for some past hypothesis k. As the resulting estimate of $\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p})$ would be approximately $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$, the estimate could be inappropriately high. If the hypothesis has a high likelihood, then it makes intuitive sense that it should inherit nearly all of $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ if there are no other hypotheses. If the hypothesis is very poor however, it may inherit a higher probability than its likelihood if $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ is higher than the likelihood of the hypothesis, thus producing the contradictory result: If some hypothesis k satisfies

$$\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{3.9}$$

$$\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) < \Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{3.10}$$

then

$$\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) > \Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p}) \tag{3.11}$$

In words, this contradiction implies that it is more likely for the lane markings to truly take the form of hypothesis k than for hypothesis k to be correct.

A solution to the contradiction in (3.11) is to adjust the approximation of $\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p})$ by enforcing a relationship between $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ and $\sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})$, noting that misdetections are free to be defined as necessary, and so their probability can also be defined as necessary. If the estimated value of the $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ partition of $\mathbf{H}$ is otherwise given by $\Pr(\mathbf{D}_k, \mathbf{M}_k)$, then updating $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ according to

$$\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) \leftarrow \min\left(\Pr(\mathbf{D}_k, \mathbf{M}_k), \sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})\right) \tag{3.12}$$

will prevent any hypothesis from triggering the contradiction in (3.11).

Such an assignment will cause the partition probabilities to become improper as they will no longer sum to one. The partition probabilities can be normalized by proportionally distributing the probability removed from partitions affected by (3.12) among the partitions that are not affected by (3.12).

3.1.2 Minimum Misdetection Probabilities

The tracking formulation provided in [19] allows the probability of misdetections to fluctuate without an explicit lower bound. Poor hypotheses that are more likely to be true than false can then possibly have a higher likelihood than the misdetection case, particularly in cases where the tracking components have not yet responded to a rapidly disappearing part. Such poor hypotheses are then returned as the most likely hypothesis if no better hypotheses are available - an inappropriate result. If the system evaluates its most plausible hypothesis as more likely to be wrong than correct, it should result in a misdetection.

The misdetection case can be redefined such that its probabilities have a minimum bound so that if the system cannot detect a hypothesis that is more likely than not, then a misdetection will be returned. As will be shown, asserting a number of required relations allow the missing part probabilities to be derived from desired functionality, as opposed to being parameters that must be tuned.

The following four assertions prevent a hypothesis that is more likely to be false than true from being returned in favour of a misdetection.

1. If the left part of a hypothesis is more likely to be true than false, then the probability of the left part being missing or misdetected should be greater than the probability of that left part hypothesis being true, i.e. if $\Pr(L_i = \mathrm{F}|\mathbf{e}) \geq \Pr(L_i = \mathrm{T}|\mathbf{e})$, then $\Pr(L = \phi|\mathbf{e}) \geq \Pr(L_i = \mathrm{T}|\mathbf{e})$ must hold.

2. Similarly for the right part if $\Pr(R_i = \mathrm{F}|\mathbf{e}) \geq \Pr(R_i = \mathrm{T}|\mathbf{e})$, then $\Pr(R = \phi|\mathbf{e}) \geq \Pr(R_i = \mathrm{T}|\mathbf{e})$ must hold.

3. If a hypothesis with the left part as missing and a non-missing right part is less likely to be true than a hypothesis with the left part missing and the same non-missing right part being false, then the probability of both parts missing should be greater than the probability of the former, i.e. if $\Pr(L = \phi, R_i = \mathrm{F}|\mathbf{e}) \geq \Pr(L = \phi, R_i = \mathrm{T}|\mathbf{e})$, then $\Pr(L = \phi, R = \phi|\mathbf{e}) \geq \Pr(L = \phi, R_i = \mathrm{T}|\mathbf{e})$ must hold.

4. Similarly for the right part being missing, if $\Pr(L_i = \mathrm{F}, R = \phi|\mathbf{e}) \geq \Pr(\Pr(L_i = \mathrm{T}, R = \phi|\mathbf{e})$, then $\Pr(L = \phi, R = \phi) \geq \Pr(L_i = \mathrm{T}, R = \phi|\mathbf{e})$ must hold.

Assertions 1 and 2 directly give minimum values for $\Pr(L = \phi|\mathbf{e})$ and $\Pr(R = \phi|\mathbf{e})$. Since $\Pr(L = \mathrm{T}|\mathbf{e}) + \Pr(L = \mathrm{F}|\mathbf{e}) = 1$, then the condition of $\Pr(L = \mathrm{F}|\mathbf{e}) \geq \Pr(L = \mathrm{T}|\mathbf{e})$ bounds $\Pr(L = \mathrm{T}|\mathbf{e})$ to be within $[0, \frac{1}{2}]$ which requires $\Pr(L = \phi|\mathbf{e})$ to be at least $\frac{1}{2}$. Similarly, $\Pr(R = \phi|\mathbf{e})$ must be at least $\frac{1}{2}$.

Assertions 3 and 4 provide minimum values for $\Pr(L = \phi, R = \phi|\mathbf{e})$. Since $\Pr(L = \phi, R = \mathrm{T}|\mathbf{e}) + \Pr(L = \phi, R = \mathrm{F}|\mathbf{e}) = \Pr(L = \phi|\mathbf{e})$, then a condition of $\Pr(L = \phi, R = \mathrm{F}|\mathbf{e}) \geq \Pr(L = \phi, R = \mathrm{T}|\mathbf{e})$ bounds $\Pr(L = \phi, R = \mathrm{T}|\mathbf{e})$ to be within $[0, \frac{1}{2}\Pr(L = \phi|\mathbf{e})]$. If the condition $\Pr(L = \phi, R = \phi|\mathbf{e}) \geq \Pr(L = \phi, R = \mathrm{T}|\mathbf{e})$ is to then be held, $\Pr(L = \phi, R = \phi|\mathbf{e})$ must be at least $\frac{1}{2}\Pr(L = \phi|\mathbf{e})$. Similarly, $\Pr(L = \phi, R = \phi|\mathbf{e})$ must also be at least $\frac{1}{2}\Pr(R = \phi|\mathbf{e})$. Thus, the minimum misdetection probabilities required in order to prevent hypotheses that are more likely to be false than true from being returned by the system are given by:

$$\Pr(L = \phi|\mathbf{e})_{min} = \frac{1}{2} \tag{3.13}$$

$$\Pr(R = \phi|\mathbf{e})_{min} = \frac{1}{2} \tag{3.14}$$

$$\begin{aligned} \Pr(L = \phi, R = \phi|\mathbf{e})_{min} &= \frac{1}{2}\max\{\Pr(L = \phi|\mathbf{e}), \Pr(R = \phi|\mathbf{e})\} \\ &= \frac{1}{4} \end{aligned} \tag{3.15}$$

Note that the minimum requirement for (3.15) is redundant as parts are assumed to track independently and missing parts have no current frame evidence metrics, so $\Pr(L = \phi, R = \phi|\mathbf{e}) = \Pr(L = \phi|\mathbf{e}) \Pr(R = \phi|\mathbf{e})$ causing $\Pr(L = \phi|\mathbf{e}) \geq \Pr(L = \phi|\mathbf{e})_{min}$ and $\Pr(R = \phi|\mathbf{e}) \geq \Pr(R = \phi|\mathbf{e})_{min}$ to imply $\Pr(L = \phi, R = \phi|\mathbf{e}) \geq \Pr(L = \phi, R = \phi|\mathbf{e})_{min}$.

Ensuring that the values in equations (3.13) and (3.14) are satisfied is done by asserting that the blind priors used in initialization satisfy these values and then that the tracking parameters are chosen to also satisfy these values. The blind prior probabilities, the probabilities before any evidence is obtained, are $\Pr(L = \phi)$, $\Pr(R = \phi)$, $\Pr(L = \phi, R = \phi)$ and can be directly set to $\frac{1}{2}, \frac{1}{2}$ and $\frac{1}{4}$ respectively. As missing parts do not generate evidence metrics and so have neither $\mathbf{e_c}$ nor $\mathbf{e_t}$ vectors, the likelihood of a missing part is solely given by its prior probability through the propagation parameters $\Pr(X = \phi|H \neq \phi)$ and $\Pr(X = \phi|H = \phi)$:

$$\begin{aligned} \Pr(X = \phi|\mathbf{e_t}, \mathbf{e_p}) &= \sum_k \Pr(X = \phi|H = h_k) \Pr(H = h_k|\mathbf{e_p}) && (3.16) \\ &= \sum_{k \in h_k \neq \phi} \Pr(X = \phi|H \neq \phi) \Pr(H = h_k|\mathbf{e_p}) \\ &\quad + \Pr(X = \phi|H = \phi) \Pr(H = \phi|\mathbf{e_p}) && (3.17) \end{aligned}$$

where $\Pr(X = \phi|H \neq \phi)$ is the probability of a part disappearing and $\Pr(X = \phi|H = \phi)$ is the probability of a missing part remaining missing and is the complement of the probability of an emerging part. Noting that $\Pr(H = \phi|\mathbf{e_p}) + \sum_{k \in h_k \neq \phi} \Pr(H = h_k|\mathbf{e_p}) = 1$ and denoting M_{prev} as $\Pr(H = \phi|\mathbf{e_p})$ and M_{next} as $\Pr(X = \phi|\mathbf{e_t}, \mathbf{e_p})$ then yields

$$M_{next} = (1 - M_{prev}) \Pr(X = \phi|H \neq \phi) + M_{prev} \Pr(X = \phi|H = \phi) \tag{3.18}$$

which relates the past frame's misdetection probability, M_{prev}, to the current frame's misdetection probability, M_{next}, through the tracking parameters, $\Pr(X = \phi|H \neq \phi)$ and $\Pr(X = \phi|H = \phi)$.

Equation (3.18) can be re-arranged as a difference equation in the form:

$$\begin{aligned} M(n) &= c_1 M(n-1) + c_0 && (3.19a) \\ c_1 &= \Pr(X = \phi|H = \phi) - \Pr(X = \phi|H \neq \phi) && (3.19b) \\ c_0 &= \Pr(X = \phi|H \neq \phi) && (3.19c) \end{aligned}$$

which has the solution

$$M(n) = \left[M_0 - \frac{c_0}{1 - c_1} \right] c_1^n + \frac{c_0}{1 - c_1} \tag{3.20}$$

with $c_0 \in (0,1)$, $c_1 \in (-c_0, 1-c_0)$, and M_0 as the initial value $M(0)$. The dynamics of the relationship of equation Equation (3.18) can then be discussed as a specific example of a simple linear system, such as described in [21].

Since $|c_1| < 1$, the system in equation (3.20) has a steady state value M_∞ given by:

$$M_\infty = \frac{c_0}{1 - c_1} \tag{3.21}$$

and the system can be re-written as

$$M(n) = (M_0 - M_\infty)c_1^n + M_\infty \tag{3.22}$$

which provides a more direct representation of the system's dynamics. The memory of the system is determined by c_1 and defines how long a disturbance in M, such as through application of equation (3.12), will persist in its effect on the partitions of H. For example, if a frame has very few hypotheses and has its M value increased, the parameter c_1 determines how much of an effect that will have on the next frame's likelihood, with $c_1 = 0$ producing a memoryless system that always predicts the same liklelihood of its corresponding part being missing independent of the past frame's detection. By contrast, if the previous frame had a very high likelihood of the part not being detected, then a system with $|c_1| > 0$ will favour the misdetection case in the next frame. Also, negative values of c_1 introduce oscillatory behaviour as equation (3.22) can be rewritten for negative values of c_1 as:

$$M(n) = (M_0 - M_\infty)|c_1|^n cos(n\pi) + M_\infty \tag{3.23}$$

Returning to the parameters $\Pr(X = \phi | H \neq \phi)$ and $\Pr(X = \phi | H = \phi)$, c_1 and M_∞ can be written as:

$$c_1 = \Pr(X = \phi | H = \phi) - \Pr(X = \phi | H \neq \phi) \tag{3.24}$$

$$M_\infty = \frac{\Pr(X = \phi | H \neq \phi)}{1 - [\Pr(X = \phi | H = \phi) - \Pr(X = \phi | H \neq \phi)]} \tag{3.25}$$

Requiring the steady state value M_∞ to be greater than or equal to $\Pr(X = \phi | \mathbf{e})_{min} = \frac{1}{2}$ from equations (3.13) and (3.14) yields:

$$\Pr(X = \phi | H \neq \phi) + \Pr(X = \phi | H = \phi) \geq 1 \tag{3.26}$$

Also noting that equation (3.23) implies that $M(n)$ could be below M_∞ for negative values of c_1, requiring c_1 to be positive ensures that $M(n)$ will always be above M_∞ provided

that M is never disturbed to be below M_∞. As the only disturbance to M is from the application of (3.12) which only ever increases the value of M, M will never fall below M_∞ so long as the following equations are satisfied by the parameters $\Pr(X = \phi | H \neq \phi)$, $\Pr(X = \phi | H = \phi)$, and the initial blind prior of $\Pr(X = \phi)$:

$$\Pr(X = \phi) \geq \frac{1}{2} \tag{3.27a}$$

$$\Pr(X = \phi | H = \phi) - \Pr(X = \phi | H = \phi) \geq 0 \tag{3.27b}$$

$$\Pr(X = \phi | H \neq \phi) + \Pr(X = \phi | H = \phi) \geq 1 \tag{3.27c}$$

The parameters of $\Pr(X = \phi | H \neq \phi)$, $\Pr(X = \phi | H = \phi)$ and the initial blind prior of $\Pr(X = \phi)$ are then chosen to satisfy desired functionality. First, it will be asserted that the misdetection case will be less likely than a hypothesis that is more likely to be true than false, requiring inequalities (3.27b) and (3.27c) to be satisfied at equality:

$$\Pr(X = \phi) = \tfrac{1}{2} \tag{3.28}$$

$$\Pr(X = \phi | H \neq \phi) + \Pr(X = \phi | H = \phi) = 1 \tag{3.29}$$

Second, it will be asserted that a non-disappearing part hypothesis that is just as likely to satisfy a true non-disappearing track as a false non-disappearing track will be just as likely as a disappearing track, i.e. if

$$\Pr(x_{i,j} = \mathrm{T}, X_j \neq \phi | H_j = h_{k,j}, \mathbf{e_t}) = \Pr(x_{i,j} = \mathrm{F}, X_j \neq \phi | H_j = h_{k,j}, \mathbf{e_t}) \tag{3.30}$$

for non-missing $h_{k,j}$ then

$$\Pr(X_j = \phi | H_j = h_{k,j}, \mathbf{e_t}) = \Pr(x_{i,j} = \mathrm{T}, X_j \neq \phi | H_j = h_{k,j}, \mathbf{e_t}) \tag{3.31}$$

With the terms in equations (3.30) and (3.31) being mutually independent and summing to one, the solution of $\Pr(X = \phi | H \neq \phi) = \frac{1}{3}$ is obtained. Equation (3.29) then gives the solution of $\Pr(X = \phi | H = \phi) = \frac{2}{3}$. Thus the parameters and blind prior that satisfy the desired assertions are

$$\Pr(X = \phi) = \tfrac{1}{2} \tag{3.32a}$$

$$\Pr(X = \phi | H = \phi) = \tfrac{2}{3} \tag{3.32b}$$

$$\Pr(X = \phi | H \neq \phi) = \tfrac{1}{3} \tag{3.32c}$$

It may also be possible to learn the system parameters from training data, or to derive them from a different set of desired functionality assertions. Such parameter selection is the subject of future work; for the remainder of this work, the values in (3.32) will be used.

3.2 Hypothesis Evaluation as a Mixture Model Problem

The method proposed in [19] requires the conditional distributions of the hypothesis evidence metrics, $\mathbf{e}$, given the hypothesis validity $\mathbf{x}_i$. It is easy to generate a very large number of hypotheses from a training set of images, but very laborious to manually label even a small proportion of them. Additionally, even if labour weren't a limitation, hard labelling might not be appropriate because $\mathbf{x}_i$ is a binary vector, and each element corresponds to a part hypothesis being either True or False. A hypothesis that isn't strongly False should not be treated the same as a hypothesis that is utterly terrible. Instead, a better approach would use soft labelling, where the label assigned to each example's validity could be anywhere in the range $[0, 1]$ instead of just either 0 or 1, allowing for varying confidence in the labels assigned to each hypothesis.

Training the Bayesian Lane Detection approach proposed by [19], require a training method that:

- Allows for gradations in hypothesis validity during training through soft labels, so that iterations where a hypothesis may lie on the incorrect side of the decision boundary do not permanently and severely inhibit training.
- Is either semi-supervised or unsupervised in order to reduce labelling labour.
- Has lenient requirements for initialization. While the training method in [19] requires very little labelling effort, it is extremely difficult to initialize.

3.2.1 Current Frame Training with Expectation Maximization

The training problem from [19] turns out to be suitable for solution using Expectation Maximization (EM) if viewed as a mixture model problem. If the validity of a hypothesis is treated as an unobserved latent variable $\mathbf{Z}$, then the probability of observing a specific hypothesis having m current frame evidence metrics in vector $\mathbf{e_c}$ with distribution parameters $\boldsymbol{\theta}$ is given by

$$\Pr(\mathbf{e_c}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} \Pr(\mathbf{e_c}, \mathbf{Z}|\boldsymbol{\theta}) \tag{3.33}$$

where $\mathbf{Z} \in \{\mathrm{TT}, \mathrm{TF}, \mathrm{FT}, \mathrm{FF}\}$ is a class variable capturing the possible validity values, the models, for the two part hypotheses in [19]. While the actual values of $\mathbf{Z}$ may be

unobservable, their probabilities can still be calculated from the evidence and a given set of parameters using Bayes' Rule:

$$\phi_{\mathbf{Z}} = \Pr(\mathbf{Z}|\mathbf{e_c}, \boldsymbol{\theta}) \tag{3.34}$$

$$= \frac{\Pr(\mathbf{e_c}|\mathbf{Z}, \boldsymbol{\theta}) \Pr(\mathbf{Z}|\boldsymbol{\theta})}{\sum_{\mathbf{Z}} \Pr(\mathbf{e_c}|\mathbf{Z}, \boldsymbol{\theta}) \Pr(\mathbf{Z}|\boldsymbol{\theta})} \tag{3.35}$$

where the soft label $\phi_{\mathbf{Z}}$ denotes the membership of the hypothesis of interest in class $\mathbf{Z}$, or alternatively, the responsibility of class $\mathbf{Z}$ for the hypothesis of interest, where $\Pr(\mathbf{e}|\mathbf{Z}, \boldsymbol{\theta})$ is the evidence model of class $\mathbf{Z}$ comprising the conditional distributions that are sought in training, and where $\Pr(\mathbf{Z})$ represents the prior probability of any hypothesis having a validity of $\mathbf{Z}$. Collectively, all of the ϕ values in the hypothesis set are the mixture of the models $\mathbf{Z}$.

Consider now a set of n hypotheses, each with their own evidence m-vector $\mathbf{e_c}$ and latent membership vectors, producing an $n \times m$ data matrix $\mathbf{X}$ containing all of the evidence vectors. Recall the objective function of EM from 2.39:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}} \Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}}) \ln[\Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) \Pr(\mathbf{Z}|\boldsymbol{\theta})] \tag{3.36}$$

Maximizing $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ also maximizes the likelihood of the observed hypothesis evidence $\mathbf{X}$, and directly includes the conditional distributions $\Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta})$ needed from training and the soft hypothesis labels $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$ that are useful for verifying results. The only term not yet included in adapting EM to training for Bayesian Lane Detection is the presence of $\Pr(\mathbf{Z}|\boldsymbol{\theta})$, the blind prior probability of a hypothesis's validity, instead of the prior used in hypothesis evaluation which is obtained through tracking. The presence of the blind prior is easily remedied by simply adding the prior probabilities as a parameter to the models.

It is important to note that missing parts do not generate metrics and so do not contribute to the mixture model. The conditional distributions obtained from EM are therefore additionally implicitly conditioned on the part hypotheses not being missing. The resulting blind prior obtained from training is then $\Pr(x_{i,j} = \text{T}|X^j \neq \phi)$ which is critically important for evaluating emerging parts. Where in [19], $\Pr(x_{i,j} = \text{T}|X^j \neq \phi)$ is one of the tracking parameters that must be tuned, in this EM formulation the probability of any particular hypothesis being true without evaluating evidence can be learned through training. The desired functionality asserted in Section 3.1.2 provides a blind prior probability for a part being missing or non-missing and for a part either emerging or disappearing, while the prior obtained in $\Pr(x_{i,j} = \text{T}|X^j \neq \phi)$ provides the link between a non-missing part and a specific part hypothesis.

Continuing with the mixture model and EM formulation, recall the mixture model expectation from equation 2.42:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) = \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \phi_{i,k} \ln \Pr(\mathbf{X}|\mathbf{Z}, \boldsymbol{\theta}) + \sum_{\mathbf{Z}=\mathbf{z}_k} \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) \sum_i \phi_{i,k} \tag{3.37}$$

where $\phi_{i,k}$ is the value of $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}})$ for hypothesis i with $\mathbf{Z} = \mathbf{z}_k$, i.e. $\phi_{i,k}$ is the ownership of hypothesis i by the k^{th} hypothesis validity model in the mixture (one of $\{\text{TT}, \text{TF}, \text{FT}, \text{FF}\}$) as evaluated in the E-Step.

We additionally note that each of the m evidence metrics contained in $\mathbf{X}$ are conditionally independent given $\mathbf{Z}$, which gives

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) &= \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \phi_{i,k} \{\ln \prod_m \Pr(X_{i,m}|\mathbf{Z}, \boldsymbol{\theta}) + \ln \Pr(\mathbf{Z}|\boldsymbol{\theta})\} & (3.38) \\ &= \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \sum_m \phi_{i,k} \ln \Pr(X_{i,m}|\mathbf{Z}, \boldsymbol{\theta}) + \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \phi_{i,k} \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) & (3.39) \end{aligned}$$

To evaluate the terms in 3.39 that depend on $\boldsymbol{\theta}$ the assumption of [19] is kept that the conditional distributions can be drawn from the exponential family allowing the usage of sufficient statistics in the M-step instead of the whole data matrix $\mathbf{X}$. The metrics are treated here as either exponentially distributed ($x_m \sim \text{Exp}(\beta)$) or gamma distributed ($x_m \sim \Gamma(\alpha, \beta)$), which allows 3.39 to ultimately be evaluated as:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) &= \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_{x_m \sim \text{Exp}} \left\{ \sum_i \phi_{i,k} \ln \lambda_{m,k} + \lambda_{m,k} \sum_i \phi_{i,k} x_i \right\} \\ &\quad + \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_{x_m \sim \Gamma} \left\{ \sum_i \phi_{i,k} \ln(\frac{\beta_{m,k}^{\alpha_{m,k}}}{\Gamma(\alpha_{m,k})}) + \begin{bmatrix} \alpha_{m,k} - 1 \\ -\beta_{m,k} \end{bmatrix} \cdot \begin{bmatrix} \sum_i \phi_{i,k} \ln x_i \\ \sum_i \phi_{i,k} x_i \end{bmatrix} \right\} \\ &\quad + \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_i \phi_{i,k} \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) \end{aligned} \tag{3.40}$$

where the parameter subscripts m, k on $\lambda_{m,k}$, $\alpha_{m,k}$, and $\beta_{m,k}$ represent the parameter within $\boldsymbol{\theta}$ belonging to the distribution of metric m required by class k. For example class $\mathbf{z}_k = \text{TF}$ requires the conditional distributions for the left part being true and the right part being false, whereas the reverse is true for $\mathbf{z}_k = \text{FT}$.

Since missing parts do not generate evidence metrics, missing part hypotheses are absent from the data matrix $\mathbf{X}$ and so the distributions and priors obtained from EM will be implicitly conditioned on the hypotheses being non-missing. As the missing part hypotheses are covered exclusively through their prior probabilities obtained through tracking, the only impact this has on the current frame training is that the obtained class prior $\Pr(\mathbf{Z}|\boldsymbol{\theta})$ will represent $\Pr(\mathbf{Z}|L \neq \phi, R \neq \phi, \boldsymbol{\theta})$ instead of $\Pr(\mathbf{Z}|\boldsymbol{\theta})$. These blind priors are not used for the current frame, however they are used for tracking when evaluating emerging hypotheses.

There are still some critical limitations of EM that need to be dealt with. First, EM guarantees a solution that is a local optimum, but this does not at all guarantee a global optimum. Second, a common problem of EM in mixture models is that distributions can often "collapse" onto single data points as a local optimum, where the distribution's mean value is that point and the distribution's variance is zero [6]. A third problem occurs in mixture models where the classes are severely skewed with one class being much more populous than another, where a local optimum can be achieved by simply ignoring the smaller class as noise and using both models to better fit the larger class. A recommended solution to these problems is to use the MAP form of EM by augmenting the optimization of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ with a prior distribution over the parameters $\boldsymbol{\theta}$ as in 2.40 [6]:

$$\boldsymbol{\theta}_{\text{new}} \leftarrow \arg\max_{\boldsymbol{\theta}} \{Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) + \ln(\boldsymbol{\theta})\}$$

The resulting estimate will then be a maximum posterior estimate based on both the observed data and the prior domain knowledge encapsulated in $\Pr(\boldsymbol{\theta})$. As will be shown in Section 3.2.3, the prior domain knowledge does not need to be very specific, and in this application even vague prior distributions fix the challenges EM can frequently face in mixture model problems.

3.2.2 Tracking Training with Expectation Maximization

Training the conditional tracking distributions proceeds as in the current frame training, but the results of training are used differently than in [19] and so further discussion is required. First, recall from Section 2.4.2 that the tracking for part j requires the distributions:

- $\Pr(\mathbf{e}_{\mathbf{t}j}|x_{i,j} = \text{T}, h_{k,j} = \text{T})$
- $\Pr(\mathbf{e}_{\mathbf{t}j}|x_{i,j} = \text{F}, h_{k,j} = \text{T})$

and the parameters:

- $\Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T})$
- $\Pr(x_{i,j} = \mathrm{F} | h_{k,j} = \mathrm{T})$

These terms are required in order to evaluate equation (2.20) for the transition probability from past part hypothesis $h_{k,j}$ to current part hypothesis $x_{i,j}$:

$$\Pr(x_{i,j} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j}) = \frac{\Pr(\mathbf{e}_{\mathbf{t}_j} | x_{i,j}, h_{k,j} = \mathrm{T}) \Pr(x_{i,j}, h_{k,j} = \mathrm{T})}{\Pr(h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j})}$$

As in the current frame training, missing part hypotheses do not generate evidence metrics, and so will not contribute to the mixture model. Consequently, the distributions and parameters learned from EM will be implicitly conditioned on part j not being missing, written again as $X^j \neq \phi$. Thus, the distributions and parameters that will be obtained by EM will be:

- $\Pr(\mathbf{e}_{\mathbf{t}j} | x_{i,j} = \mathrm{T}, h_{k,j} = \mathrm{T}, X^j \neq \phi)$
- $\Pr(\mathbf{e}_{\mathbf{t}j} | x_{i,j} = \mathrm{F}, h_{k,j} = \mathrm{T}, X^j \neq \phi)$

and the parameters:

- $\Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, X^j \neq \phi)$
- $\Pr(x_{i,j} = \mathrm{F} | h_{k,j} = \mathrm{T}, X^j \neq \phi)$

which differ from the terms required by the previous derivations in [19].

A slightly different derivation however allows $\Pr(x_{i,j} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j})$ to be evaluated more directly by separating the evaluation of $\Pr(x_{i,j} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j})$ based on missing or non-missing parts before the application of Bayes' rule in (2.20) instead of after the application of Bayes' rule as in the derivation in [19].

$x_{i,j} \neq \phi, h_{k,j} \neq \phi$: For the case where neither part hypothesis is missing, the desired probability $\Pr(x_{i,j} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j})$ can be viewed as being implicitly conditioned on part j not being missing, i.e on $X^j \neq \phi$, because $\Pr(x_{i,j} = \mathrm{T} | X^j = \phi) = 0$ for a non-missing $x_{i,j}$. The desired transition probability can then be written as:

$$\Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}j}) = \Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, X^j \neq \phi, \mathbf{e}_{\mathbf{t}j}) \tag{3.41}$$

Applying Bayes' rule then yields:

$$\Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, \mathbf{e}_{\mathbf{t}_j}) = \frac{\Pr(\mathbf{e}_{\mathbf{t}_j} | x_{i,j} = \mathrm{T}, h_{k,j} = \mathrm{T}, X^j \neq \phi) \Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, X^j \neq \phi)}{\Pr(\mathbf{e}_{\mathbf{t}_j} | h_{k,j} = \mathrm{T}, X^j \neq \phi)} \tag{3.42}$$

with

$$\begin{aligned} \Pr(\mathbf{e}_{\mathbf{t}_j} | h_{k,j} = \mathrm{T}, X^j \neq \phi) \quad = \quad & \Pr(\mathbf{e}_{\mathbf{t}_j} | x_{i,j} = \mathrm{T}, h_{k,j} = \mathrm{T}, X^j \neq \phi) \Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \mathrm{T}, X^j \neq \phi) \\ & + \Pr(\mathbf{e}_{\mathbf{t}_j} | x_{i,j} = \mathrm{F}, h_{k,j} = \mathrm{T}, X^j \neq \phi) \Pr(x_{i,j} = \mathrm{F} | h_{k,j} = \mathrm{T}, X^j \neq \phi) \end{aligned} \tag{3.43}$$

where all of terms in (3.42) and (3.43) are directly given by EM, and the prior conditional probability $\Pr(X^j \neq \phi | h_{k,j} = \mathrm{T})$ required to link the non-missing past part hypothesis $h_{k,j}$ to current non-missing part hypothesis $x_{i,j}$ alone is simply the complement of the parameter $\Pr(X^j = \phi | H \neq \phi)$ and whose value is determined by derivation in Section 3.1.2.

$x_{i,j} \neq \phi, h_{k,j} = \phi$: For the case where the current part hypothesis $x_{i,j}$ is not missing but the past hypothesis $h_{k,j} = \phi$ is missing, the desired transition probability is assumed to be constant as there is no transitional evidence to evaluate. As in [19], $h_{k,j} = \phi$ provides no information about $x_{i,j}$ if $X^j \neq \phi$ is given. The transition probability is then:

$$\Pr(x_{i,j} = \mathrm{T} | h_{k,j} = \phi) = \Pr(x_{i,j} = \mathrm{T} | X^j \neq \phi) \Pr(X^j \neq \phi | h_{k,j} = \phi) \tag{3.44}$$

where $\Pr(X^j \neq \phi | h_{k,j} = \phi)$ is simply the complement of the parameter $\Pr(X^j = \phi | H = \phi)$ and whose value is determined by derivation in Section 3.1.2, and where $\Pr(x_{i,j} = \mathrm{T} | X^j \neq \phi)$ is simply the prior probability of a part hypothesis given that it is non-missing and is obtained current frame training by marginalizing the blind class priors over the other parts in the model.

$x_{i,j} = \phi, h_{k,j} \neq \phi$: For the case where the current frame hypothesis is missing but the previous part hypothesis is not, the transition probability is simply the parameter $\Pr(X^j = \phi | H^j \neq \phi)$ as in [19] and whose value is determined through derivation in Section 3.1.2.

$x_{i,j} \neq \phi, h_{k,j} \neq \phi$: For the case where both part hypotheses are missing, the transition probability is again simply given by the parameter $\Pr(X^j = \phi | H^j = \phi)$ whose value is determined through derivation in Section 3.1.2.

Since parts are assumed to track independently, treating the transition validity of each track as a latent variable allows the same EM process to be used but with $\mathbf{Z} \in \{x_{i,j} =$

$\mathrm{T}, x_{i,j} = \mathrm{F}\}$, and instead of n hypotheses being considered it is the n tracks generated by the different $x_{i,j}$ and $h_{k,j}$ pairs, with each track having its own evidence m-vector $\mathbf{e_t}$ producing the $n \times m$ data matrix $\mathbf{X}$ containing all of the tracking evidence vectors. Using exponentially distributed or gamma distributed evidence metrics for tracking, the tracking equivalent to equation (3.40) thus becomes:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}}) \;=\; & \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_{x_m \sim \text{Exp}} \left\{ \sum_t \phi_{t,k} \ln \lambda_{m,k} + \lambda_{m,k} \sum_t \phi_{t,k} x_t \right\} \\ & + \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_{x_m \sim \Gamma} \left\{ \sum_t \phi_{t,k} \ln\left(\frac{\beta_{m,k}^{\alpha_{m,k}}}{\Gamma(\alpha_{m,k})}\right) + \begin{bmatrix} \alpha_{m,k} - 1 \\ -\beta_{m,k} \end{bmatrix} \cdot \begin{bmatrix} \sum_i \phi_{t,k} \ln x_t \\ \sum_i \phi_{t,k} x_t \end{bmatrix} \right\} \\ & + \sum_{\mathbf{Z}=\mathbf{z}_k} \sum_t \phi_{t,k} \ln \Pr(\mathbf{Z}|\boldsymbol{\theta}) \end{aligned} \tag{3.45}$$

where $\phi_{t,k}$ is the value of $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{old})$ for track t with $\mathbf{Z} = \mathbf{z}_k$, i.e. $\phi_{t,k}$ is the ownership of track t by the temporal consistency value $\mathbf{z}_k$ evaluated in the E-Step, and where the parameter subscripts m and k on $\lambda_{m,k}$, $\alpha_{m,k}$, and $\beta_{m,k}$ represent the parameters within $\boldsymbol{\theta}$ belong to the distribution of metric m required by class k.

3.2.3 Prior Parameter Distributions

In order to avoid many of the challenges described at the end of Section 3.2.1, prior distributions over the parameters $\boldsymbol{\theta}$ are used [6]. It turns out that the prior knowledge of the parameters does not need to be specific and can in fact be extremely vague and still produce valid results. In this work both extreme values and mass boundaries are used as types of weak prior knowledge.

Extreme Values

Instead of directly assigning a distribution for $\Pr(\boldsymbol{\theta})$, it is possible to infer the likelihood of a set of parameters from their performance in classifying some synthetic examples with extreme values that should produce very high confidence likelihoods. For example, an example hypothesis, $\mathbf{x}_{\text{best}}$, with ideal current frame evidence metrics, $\mathbf{e}_{\mathbf{c}\text{best}}$ should have $\Pr(\mathbf{x}_{\text{best}}|\mathbf{e}_{\mathbf{c}\text{best}}, \boldsymbol{\theta}) \approx 1$. Similar examples can also be created for the likelihoods of individual metrics. For example the likelihood of a hypothesis, $\mathbf{x}_{\text{poor}}$, with a very poor value for metric m can be calculated using just the parameters relevant to the conditional distributions for metric m, the prior $\Pr(\mathbf{x} = \mathrm{T}|\boldsymbol{\theta})$ and Bayes' Rule as:

$$\Pr(\mathbf{x}_{\text{poor}}|e_m, \boldsymbol{\theta}) = \frac{\Pr(e_m|\mathbf{x}_{\text{poor}}, \boldsymbol{\theta}_m)\Pr(\mathbf{x} = \text{T}|\boldsymbol{\theta})}{\Pr(e_m|\mathbf{x}_{\text{poor}}, \boldsymbol{\theta}_m)\Pr(\mathbf{x} = \text{T}|\boldsymbol{\theta}) + \Pr(e_m|\mathbf{x}_{\text{poor}}, \boldsymbol{\theta}_m)\Pr(\mathbf{x} = \text{F}|\boldsymbol{\theta})} \quad (3.46)$$

where e_m is the extremely poor value for metric m. For every evidence metric, at least one extreme example can be created from one of the ends of the metric's range. For penalty metrics, only one extreme example should be taken from their maximum values, while for the support score metrics both ends of the range can be taken. The weak prior knowledge obtained from these extreme examples can be described as penalizing any parameter sets that fail to satisfy the simple assumptions "extremely high penalties are bad", "extremely high scores are good", and "extremely low scores are bad". The parameter likelihood can then be evaluated as the product of the likelihoods of each extreme example having the correct classification:

$$\Pr(\boldsymbol{\theta}|\text{Extreme Examples}) = \prod_g \Pr(\mathbf{x} = X_g|\mathbf{e}_g, \boldsymbol{\theta}) \quad (3.47)$$

where $\mathbf{e}_g$ are the evidence values corresponding to the g^{th} extreme example, and X_g is the classification that the extreme example should have.

Mass Boundaries

The second type of weak prior knowledge used to evaluate $\Pr(\boldsymbol{\theta})$ is an estimate of where the majority of examples of a True or False class should lie or not lie for a specific metric. For example, the vast majority of true examples should be in the lower range for a penalty metric. Consider a value of $E_{m,\text{T}}$ for metric m where the extreme majority of True examples should have a corresponding value of metric m below $E_{m,\text{T}}$. The cumulative distribution of metric m's corresponding conditional distribution for the true class should be approximately equal to one:

$$\mathbf{CDF}_m(E_{m,\text{T}}|\mathbf{x} = \text{T}, \boldsymbol{\theta}) = \int_{-\infty}^{E_{m,\text{T}}} \Pr(e_m|\mathbf{x} = \text{T}, \boldsymbol{\theta}) \approx 1 \quad (3.48)$$

For each metric and distribution where such mass boundaries can easily be obtained by inspection, the parameter likelihood can be evaluated as the product of corresponding probability masses within the desired bounds:

$$\Pr(\boldsymbol{\theta}|\text{Mass Boundaries}) = \prod_b \mathbf{CDF}_{m_b}(E_{m_b,X_b}|\mathbf{x} = X_b, \boldsymbol{\theta}) \quad (3.49)$$

for a group of b mass boundaries. In this work mass boundaries for the false class were not used for prior knowledge of the parameters because the False class did not have any boundaries that were visible by inspection. In contrast, boundaries for the True class can easily be determined towards the upper bound of penalty metrics. The benefit of the parameter prior is demonstrated in Chapter 4.

3.2.4 Training Procedure

With Sections 3.2.1 through 3.2.3 providing solutions to the training problems posed by [19], it is now possible to provide a training procedure for the Bayesian Lane Detection method that does not suffer from the problems identified in Section 2.4.3. The proposed training method divides into two separate phases, one for training the current frame parameters $\boldsymbol{\theta}_c$ and one for training the tracking parameters $\boldsymbol{\theta}_t$ detailed in Algorithms 3 and 4 respectively.

In the first phase, a data matrix is created to capture a large number of example hypotheses' current frame evidence metrics from each image in the training set. EM is then used as in Section 3.2.1 to provide estimates of the conditional distributions $\Pr(\mathbf{e_c}|\mathbf{x}_i)$ as well as the priors for $\Pr(\mathbf{x}_i)$.

In the second phase, new data matrices are created for each part by evaluating each pair of consecutive images in the training set to create a large number of tracks and their associated metrics. Creating tracks requires an evaluation of the first frame's hypotheses in order to select suitable candidates for tracking, with the current frame $\boldsymbol{\theta}_c$ parameters estimated in the first phase being used for evaluation. EM is then used as in Section 3.2.2 to estimate the conditional distributions $\Pr(\mathbf{e_t}|x_{i,j}, h_{k,j})$ and priors $\Pr(x_{i,j}|h_{k,j})$ for each part. The tracking training is thus broken up into a pair of independent EM training problems because it has been assumed in equation (2.19) that parts track independently.

In both phases, the prior parameter distributions use coarse knowledge as described in Section 3.2.3. If for some reason prior knowledge is either unavailable or insufficient to avoid unwanted local optima, the method reverts to a semi-supervised learning system using the following labelling procedure:

1. Select a small sample of hypotheses from the class with significant misclassification, typically the True class.
2. Review the sample hypotheses. For each sample either:
   (a) Assign a hard label to its membership if it is obvious.

(b) Leave the example's membership as variable if it is not an obvious member of a single class

(c) For current frame metrics that have validity vectors instead of single variables, assign a negative label to a class to identify it as not a member of that class, such as: $\mathbf{x}_i$ is not $\{L_i = \mathrm{T}, R_i = \mathrm{T}\}$.

It may often be necessary after executing either of the labelling steps to reset the parameter values to initial estimates. While the additional labelling information may help avoid the unwanted local optimum in the first place, it might not necessarily be sufficient to allow the algorithm to escape the local optimum if it has already become stuck there.

Algorithm 3 Proposed Current Frame Training Method: Assuming that a set of lane marking hypotheses from a training set of images are divisible into True (T) and False (F) classes with current-frame evidence metrics $\mathbf{X}$, learn the parameters $\boldsymbol{\theta}_c$ of their respective conditional current-frame evidence distributions $\Pr(\mathbf{X}|\boldsymbol{\theta}_c, \mathrm{T})$, and $\Pr(\mathbf{X}|\boldsymbol{\theta}_c, \mathrm{F})$ as well as their prior probabilities $\Pr(\mathrm{T}|\boldsymbol{\theta}_c)$ and $\Pr(\mathrm{F}|\boldsymbol{\theta}_c)$, with the parameters' prior distribution $\Pr(\boldsymbol{\theta}_c)$ defined as in Section 3.2.3.

$\boldsymbol{\theta}_c \leftarrow \boldsymbol{\theta}_{c,0}$
repeat
  for all hypotheses in training set **do**
    Compute $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}})$ from $\Pr(\mathrm{T}|\mathbf{X}, \boldsymbol{\theta}_c)$ and $\Pr(\mathrm{F}|\mathbf{X}, \boldsymbol{\theta}_c)$.
  end for
  $\boldsymbol{\theta}_c \leftarrow \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ with $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ from (3.39)
  Perform labelling procedure if necessary.
until convergence of $\boldsymbol{\theta}_c$

The proposed training procedure has many advantages over the training method described in [19]. First, it is very robust with respect to initialization; so long as the initialization values provides very basic trends correctly, such as the True class having higher scores and lower penalties than the False class, it appears to converge to a suitable result. Second, the labelling procedure is often not required if the prior parameter distribution is able to eliminate unwanted local optima. For example, in applying this training procedure to the model in [19], no labelling was required even with extremely vague prior knowledge, as will be shown in Chapter 4. As such, beyond the information contained in the prior parameter distribution, the system can effectively function as an unsupervised learning system and obtain acceptable results without any additional human labour.

Algorithm 4 Proposed Tracking Training Method: Assuming that a set of potential part hypothesis tracks consisting of both past and current part hypotheses are divisible into True (T) and False (F) classes with tracking evidence metrics $\mathbf{X}$, learn the parameters $\boldsymbol{\theta}_t$ of their respective conditional tracking evidence distributions $\Pr(\mathbf{X}|\boldsymbol{\theta}_t, \mathrm{T})$, and $\Pr(\mathbf{X}|\boldsymbol{\theta}_t, \mathrm{F})$ as well as their prior probabilities $\Pr(\mathrm{T}|\boldsymbol{\theta}_t)$ and $\Pr(\mathrm{F}|\boldsymbol{\theta}_t)$, with the parameters' prior distribution $\Pr(\boldsymbol{\theta}_t)$ defined as in Section 3.2.3.

$\boldsymbol{\theta}_t \leftarrow \boldsymbol{\theta}_{t,0}$
repeat
  for all hypotheses in training set **do**
    Compute $\Pr(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}_{\text{old}})$ from $\Pr(\mathrm{T}|\mathbf{X}, \boldsymbol{\theta}_t)$ and $\Pr(\mathrm{F}|\mathbf{X}, \boldsymbol{\theta}_t)$.
  end for
  $\boldsymbol{\theta}_t \leftarrow \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ with $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ from (3.45)
  Perform labelling procedure if necessary.
until convergence of $\boldsymbol{\theta}_t$

3.3 Generalized Bayesian Detection and Tracking

With Section 3.2 providing a training method that requires almost no human labelling and only requires very coarse prior knowledge of the distribution parameters, it is now possible to not just apply the Bayesian Lane Detection method introduced by [19] and described in Section 2.4, but to extend it. Critically, neither the EM training approach proposed in Section 3.2 nor the Bayesian Lane Detection method itself are strictly dependent on either the evidence metrics or the lane model defined in [19]. In fact, the derivation in [19] specifically addressed the general case of a deformable multi-part model, but provided no adaptable method for training or defining arbitrary models or suitable heuristics. This Section will provide a definition for arbitrary lane marking models adapted from the form of [19], a list of useful general heuristic metrics that can serve as hypothesis features, and some example models.

Definition 1 *An arbitrary lane marking model ξ can be defined as consisting of a set X of N parts $X = \{X_1, X_2, ..., X_N\}$ and a set L of M links $L = \{L_1, L_2, ...L_M\}$ between parts where each link L_i is defined as the subset of parts in X that the link, L_i, connects: $L_i = \{X_j | j \in L_i\}$. Evaluating hypotheses for model Θ require three types of heuristic metrics:*

- *$h_{\mathbf{c}X_j}$: Each part X_j requires a set of heuristic metrics $h_{\mathbf{c}X_j}$ that can be used to evaluate the current frame evidence supporting part X_j of a hypothesis.*

- $h_{\mathbf{c}L_j}$: *Each link L_j requires a set of heuristic metrics $h_{\mathbf{c}L_j}$ that can be used to evaluate the current frame compatibility of the parts in link L_j of a hypothesis.*
- $h_{\mathbf{t}X_j}$: *Each part X_j requires a set of heuristic metrics $h_{\mathbf{t}X_j}$ that can be used to evaluate the likelihood of a past hypothesis for part X_j tracking to a current part hypothesis for part X_j. Parts are assumed to track independently.*

Training the conditional distributions of the metrics is then performed using EM as in Algorithm 3 to train the current frame conditional distributions for $h_{\mathbf{c}X_j}$ and $h_{\mathbf{c}L_j}$ as in Section 3.2.1 followed by using Algorithm 4 to train the tracking conditional distributions for $h_{\mathbf{t}X_j}$ as in Section 3.2.2. Tracking is again assumed to be independent across different parts, which is especially important for models with several parts as the combinations of parts increases exponentially with the number of parts in a model.

3.3.1 Generalized Hypothesis Evaluation

Evaluating a hypothesis for model ξ proceeds as in Section 2.4, but with some minor changes to take advantage of the additional information that the EM training method gives about the evidence distributions. The desired likelihood remains as in equation (2.11)

$$\Pr(\mathbf{x}_i|\text{Evidence}) = \frac{\Pr(\mathbf{e_c}|\mathbf{x}_i)\Pr(\mathbf{x}_i|\mathbf{e_t},\mathbf{e_p})}{\Pr(\mathbf{e_c})} \tag{3.50}$$

but with the validity of hypothesis i, $\mathbf{x}_i \in \{0,1\}^N$, now as a binary N-vector capturing the validity of each of the N part-hypotheses comprising hypothesis i, and $\mathbf{e_c}$ comprising the values of all of the current frame metrics $h_{\mathbf{c}X_j}$ and $h_{\mathbf{c}L_j}$ associated with hypothesis i. The evaluation of $\Pr(\mathbf{e_c}|\mathbf{x}_i)$ is now given by

$$\Pr(\mathbf{e_c}|\mathbf{x}_i) = \prod_{m \in h_{\mathbf{c}X_j}} \Pr(e_m|X_j = x_{i,j}) \prod_{m \in h_{\mathbf{c}L_j}} \Pr(e_m|X_v = x_{i,v}, \forall X_v \in L_j) \tag{3.51}$$

where $x_{i,j}$ and $x_{i,v}$ are the validity values of part j and v of hypothesis i respectively and where L_j represents link j. $\Pr(\mathbf{e_c})$ is then calculated as before:

$$\Pr(\mathbf{e_c}) = \sum_{\mathbf{x}_i} \Pr(\mathbf{e_c}|\mathbf{x}_i)\Pr(\mathbf{x}_i|\mathbf{e_t},\mathbf{e_p}) \tag{3.52}$$

where the summation in equation (3.52) is performed over all 2^N permutations of $\mathbf{x}_i$ values.

The priors $\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p})$ for each permutation of $\mathbf{x}_i$ are determined from

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}) = \sum_k \Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{H} = \mathbf{h}_k) \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p}) \tag{3.53}$$

As it is again assumed as in [19] that parts track independently, equation (3.53) can be written as

$$\Pr(\mathbf{x}_i|\mathbf{e_t}, \mathbf{e_p}) = \sum_k \left[\left(\prod_{j=1}^{N} \Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}) \right) \Pr(\mathbf{H} = \mathbf{h_k}|\mathbf{e_p}) \right] \tag{3.54}$$

for a model with N parts.

Transitional Components

There are four possible forms for the transitional component $\Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j})$ in equation (3.54) corresponding to each of $x_{i,j}$ and $h_{k,j}$ being missing or non-missing. The probabilities of the matching false tracks $\Pr(x_{i,j} = \mathrm{F}|\mathbf{e_t}_j, h_{k,j})$ are given by the complement: $\Pr(x_{i,j} = \mathrm{F}|\mathbf{e_t}_j, h_{k,j}) = 1 - \Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j})$. The evaluations of these transitional components are nearly identical to [19], with the only major difference being that the tracking parameters here are derived from desired functionality in Section 3.1 as opposed to being treated as tunable parameters as in [19].

For neither of $x_{i,j}$ or $h_{k,j}$ being missing, the transitional component is evaluated as

$$\Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}) = \Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}, X_j \neq \phi) \Pr(X_j \neq \phi|H \neq \phi) \tag{3.55}$$

where $\Pr(X_j \neq \phi|H \neq \phi) = 1 - \Pr(X_j = \phi|H \neq \phi)$ is given by the tracking parameter $\Pr(X_j = \phi|H \neq \phi)$ for part j derived as in Section 3.1 and where $\Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}, X_j \neq \phi)$ is given by Bayes rule:

$$\Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}, X_j \neq \phi) = \frac{\Pr(\mathbf{e_t}_j|x_{i,j}h_{k,j}, X_j \neq \phi) \Pr(x_{i,j}|h_{k,j}, X_j \neq \phi)}{\sum_{x_{i,j}} \Pr(\mathbf{e_t}_j|x_{i,j}h_{k,j}, X_j \neq \phi) \Pr(x_{i,j}|h_{k,j}, X_j \neq \phi)} \tag{3.56}$$

using the conditional distributions and track priors obtained from the EM tracking training in Section 3.2.2.

For only the current part $x_{i,j}$ being missing, the transitional component is evaluated as simply

$$\Pr(x_{i,j} = \phi|\mathbf{e_t}_j, h_{k,j}) = \Pr(X_j = \phi|H \neq \phi) \tag{3.57}$$

where $\Pr(X_j = \phi|H \neq \phi)$ is the tracking parameter derived as in Section 3.1.

For only the previous part $h_{k,j}$ being missing, the transitional component is evaluated as

$$\Pr(x_{i,j}|\mathbf{e_t}_j, h_{k,j}) = \Pr(x_{i,j}|X_j \neq \phi)\Pr(X_j \neq \phi|H = \phi) \tag{3.58}$$

where $\Pr(X_j \neq \phi|H = \phi)$ is given by the complement $\Pr(X_j \neq \phi|H = \phi) = 1 - \Pr(X_j = \phi|H = \phi)$ with the tracking parameter $\Pr(X_j = \phi|H = \phi)$ derived as in Section 3.1 and where $\Pr(x_{i,j}|X_j \neq \phi)$ is the blind prior probability of a hypothesis for part j being true without any past evidence obtained by marginalization:

$$\Pr(x_{i,j}|X_j \neq \phi) = \sum_{\mathbf{Z}|X_j=\mathrm{T}} \Pr(\mathbf{Z}|\boldsymbol{\theta}) \tag{3.59}$$

over the blind priors $\Pr(\mathbf{Z}|\boldsymbol{\theta})$ obtained from the EM tracking training in Section 3.2.2 that have part j as true.

For both parts in the track being missing, the transitional component is simply given by

$$\Pr(x_{i,j} = \phi|\mathbf{e_t}_j, h_{k,j}) = \Pr(X_j = \phi|H = \phi) \tag{3.60}$$

where the tracking parameter $\Pr(X_j = \phi|H = \phi)$ is derived as in Section 3.1.

Past Component

The past probability of past hypothesis k, $\Pr(\mathbf{H} = \mathbf{h_k}|\mathbf{e_p})$, is evaluated by separating the set of retained past hypotheses into $(\mathbf{D}, \mathbf{M})$ partitions, where all of the hypotheses in a partition have the same sets of missing parts $\mathbf{M}$ and detected parts $\mathbf{D}$. With N parts, there are 2^N ways for parts to be detected or missing, resulting in 2^N different sets of $\mathbf{D}$ and $\mathbf{M}$. Since parts track independently, the predicted past probability of a $(\mathbf{D}, \mathbf{M})$ partition, $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})_{pred}$ can be predicted from the past partition probabilities $\Pr(\mathbf{D}_q \neq \phi, \mathbf{M}_q = \phi|\mathbf{e_p})_{prev}$ of the frame preceding the previous frame, i.e. the frame two prior to the current frame, and the tracking parameters $\Pr(X_j = \phi|H_j = \phi)$ and $\Pr(X_j = \phi|H_j \neq \phi)$:

$$\begin{aligned}\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})_{pred} \;=\; & \prod_{j\in\mathbf{D}_k}\Big[\Pr(H_j \neq \phi)_{prev}(1-\Pr(X_j=\phi|H\neq\phi)) \\ & +\Pr(H_j=\phi)_{prev})(1-\Pr(X_j=\phi|H=\phi))\Big] \\ & \prod_{j\in\mathbf{M}_k}\Big[\Pr(H_j\neq\phi)_{prev}\Pr(X_j=\phi|H\neq\phi) \\ & +\Pr(H_j=\phi)_{prev})\Pr(X_j=\phi|H=\phi)\Big] \end{aligned} \tag{3.61}$$

where the products over j covering the j parts that are missing and non-missing respectively and with

$$\Pr(H_j \neq \phi)_{prev} = \sum_{q|X_j\in\mathbf{D}_q} \Pr(\mathbf{D}_q \neq \phi, \mathbf{M}_q = \phi|\mathbf{e_p})_{prev} \tag{3.62}$$

and

$$\Pr(H_j = \phi)_{prev} = \sum_{q|X_j\in\mathbf{M}_q} \Pr(\mathbf{D}_q \neq \phi, \mathbf{M}_q = \phi|\mathbf{e_p})_{prev} \tag{3.63}$$

where the summations over q cover all past hypotheses q where part X_j is included in the detected parts' partition $\mathbf{D}_q$ and where X_j is included in the missing parts' partition $\mathbf{M}_q$ respectively.

The predicted value $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})_{pred}$ is then compared to the sum likelihood of the past hypotheses in partition $(\mathbf{D}_k, \mathbf{M}_k)$ and the value of $\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})$ to be used in estimating $\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p})$ is chosen as the lesser of value of the two:

$$\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p}) \leftarrow \min\big(\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})_{pred}, \sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})\big) \tag{3.64}$$

Lastly, as in [19], the approximation

$$\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) \approx \frac{\Pr(\mathbf{d}_k = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})\Pr(\mathbf{D}_k \neq \phi, \mathbf{M}_k = \phi|\mathbf{e_p})}{\sum_i \Pr(\mathbf{d}_i = \mathrm{T}, \mathbf{M}_k = \phi|\mathbf{e_p})} \tag{3.65}$$

is used to estimate $\Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$, where $\Pr(\mathbf{D}_k = \mathbf{d}_k, \mathbf{M}_k = \phi|\mathbf{e_p}) = \Pr(\mathbf{H} = \mathbf{h}_k|\mathbf{e_p})$ with the detected and missing parts of $\mathbf{h}_k$ separated into $\mathbf{h}_k$'s respective $(\mathbf{D}, \mathbf{M})$ partition.

3.3.2 General Heuristic Hypothesis Features

The lane marking model introduced in [19] and described in Chapter 2 uses eight current frame metrics and six tracking metrics as features to evaluate hypothesis validity. These metrics can be adjusted and expanded to provide a list of heuristics that may be useful for other models within the generalized model framework. The key restriction on selecting heuristic metrics for a lane marking model is that the metrics must be conditionally independent given the hypothesis validity. In order to streamline the training process, metrics that provide at least some crude prior knowledge as in Section 3.2.3 are favourable as prior parameter distributions reduce or eliminate the required labelling. Metrics do not necessarily need to be human defined heuristics and if metrics obtained from machine learning with labelled data are available, then they could be used as an alternative, so long as they can provide some crude prior knowledge as in Section 3.2.3.

There are three categories of heuristic metrics: metrics $h_{\mathbf{c}X_j}$ for the current frame that are specific to part X_j, metrics $h_{\mathbf{c}L_j}$ for the current frame that measure the compatibility of the parts linked by link L_j, and the tracking metrics $h_{\mathbf{t}X_j}$ that capture the temporal correlation between past and present part hypotheses for part X_j. Since parts are assumed to track independently, there are no tracking metrics for links.

Part-Specific Current Frame Metrics $h_{\mathbf{c}X_j}$

Part-specific metrics typically rely on a balance of scores capturing local pixel support and geometrically inspired penalties to penalize over-fitting.

An obvious part-specific metric is the support score created by summing the pixel-level support provided for a part over every pixel in the part hypothesis as discussed in Section 2.3.1 for metrics $e_{L,1}$ and $e_{R,1}$. The neural network patch classifier used in [19] is specific for detecting longitudinal lane markings, but support scores could be similarly generated for other marking orientations by training a patch classifier to detect straight line lane markings at any angle.

To avoid over-fitting, the curvature penalty metric used in [19] can be used to penalize part hypotheses that undergo dramatic shape changes in image areas without local pixel support for the change such as with metrics $e_{L,2}$ and $e_{R,2}$ in Section 2.3.1.

The system proposed in [19] is intended to accept both straight and curved lane markings, however a system supporting multiple models could treat these as separate types of lane markings. A model specifically for straight lane markings could penalize overall

part-hypothesis curvature while a model specifically designed for curve lane markings could instead penalize part hypotheses that lack the appropriate curvature.

The system proposed in [19] is biased against dotted lane markings as hypotheses with lower pixel support scores are weighted equally whether or not the support is periodically spaced out. Another part-specific metric could be added so that hypotheses with periodically spaced pixel support are penalized less than hypotheses with the same pixel support score but no periodic behaviour and is the subject of future work.

Link-Specific Current Frame Metrics $h_{\mathbf{c}L_j}$

Link-specific metrics typically rely on the general geometric relationship between the parts comprising the link in order to represent the compatibility of the part hypotheses. For example, the metrics $e_{LR,1}$, $e_{LR,2}$, $e_{LR,3}$ described in Section 2.3.1 attempt to represent the common geometric relationship that left and right-hand lane markings are parallel.

Another type of simple link-specific metric can relate the location of the marking model as a whole relative to the ego vehicle. For example, the metric $e_{LR,4}$ described in Section 2.3.1 captures the location of the center of the estimated lane in an attempt to differentiate between ego lane markings and adjacent lane markings. Colour relationships could also be used, such as yellow left-hand markings linking to white right-hand markings.

Tracking Metrics $h_{\mathbf{t}X_j}$

Tracking metrics aim to capture the temporal compatibility of part hypotheses from frame to frame based on the model's assumption about the motion of the ego vehicle. The model in [19] specifically addresses the ego-lane, with the ego vehicle expected to be travelling along the ego lane and the lane markings maintaining an approximately constant location relative to the ego vehicle as the driver stays within them. Consequently, the tracking metrics described in Section 2.3.2 aim to represent a constant relative location for the lane markings by penalizing frame to frame changes in lateral position, direction, or shape. If information regarding vehicle motion is available from an IMU or other sensors, the tracking metrics used in the lane marking model could be adjusted to account for changes in relative position such as by tracking lane markings through a lane-change maneuver. For the present work, only camera images are used.

3.3.3 Example Classes

Two example classes of lane markings will be considered within the general form presented in Definition 1. First, the lane marking model presented in [19] will be described in the context of Definition 1 as ξ_1. Second, an extension to ξ_1 that also considers the right-hand lane marking of the adjacent right-hand lane will be described in the context of Definition 1 as ξ_2.

Single Lane

The lane marking model in [19] that was described in Chapter 2 can be expressed in the general form of Definition 1 as follows:

ξ_1 consists of $N = 2$ parts, $X = \{X_1, X_2\}$, with X_1 being the left-hand lane marking and X_2 being the right-hand lane marking, and $M = 1$ link, $L = \{L_1\}$, with $L_1 = \{X_1, X_2\}$ and with the following heuristics:

- $h_{\mathbf{c}X_1}$: The left-hand lane marking uses metrics $h_{\mathbf{c}X_1} = \{e_{L,1}, e_{L,2}\}$ with $e_{L,1}$ and $e_{L,2}$ defined as in Section 2.3.1.
- $h_{\mathbf{c}X_2}$: The right-hand lane marking uses metrics $h_{\mathbf{c}X_2} = \{e_{R,1}, e_{R,2}\}$ with $e_{R,1}$ and $e_{R,2}$ defined as in Section 2.3.1.
- $h_{\mathbf{c}l_1}$: The link between the left-hand and right-hand markings use compatibility metrics $h_{\mathbf{c}l_1} = \{e_{LR,1}, e_{LR,2}, e_{LR,3}, e_{LR,4}\}$ with $e_{LR,1}$, $e_{LR,2}$, $e_{LR,3}$, and $e_{LR,4}$ defined as in Section 2.3.1.
- $h_{\mathbf{t}X_1}$: The left-hand lane marking uses tracking metrics $h_{t_{X_1}} = \{e_{t_{L_1}}, e_{t_{L_2}}, e_{t_{L_3}}\}$ with $e_{t_{L_1}}$, $e_{t_{L_2}}$, and $e_{t_{L_3}}$ defined as in Section 2.3.1.
- $h_{\mathbf{t}X_2}$: The right-hand lane marking uses tracking metrics $h_{t_{X_2}} = \{e_{t_{R_1}}, e_{t_{R_2}}, e_{t_{R_3}}\}$ with $e_{t_{R_1}}$, $e_{t_{R_2}}$, and $e_{t_{R_3}}$ defined as in Section 2.3.1.

Ego Lane and Right Side Lane

A multiple lane model can be constructed to consider both the ego lane and the right-hand adjacent lane as an extension of the single lane model.

ξ_2 consists of $N = 3$ parts, $X = \{X_1, X_2, X_3\}$, with X_1 being the left-hand lane marking of the ego lane, X_2 being the right-hand lane marking for the ego lane and the left-hand

marking for the adjacent right-hand lane, and X_3 being the right-hand lane marking for the adjacent right-hand lane, and $M = 4$ links, $L = \{L_1, L_2, L_3, L_4\}$, with $L_1 = \{X_1, X_2\}$, $L_2 = \{X_2, X_3\}$, $L_3 = \{X_1, X_3\}$, $L_4 = \{X_1, X_2, X_3\}$, and with the following heuristics:

- $h_{\mathbf{c}X_1}$: The left-hand lane marking uses metrics $h_{\mathbf{c}X_1} = \{e_{L,1}, e_{L,2}\}$ with $e_{L,1}$ and $e_{L,2}$ defined as in Section 2.3.1.
- $h_{\mathbf{c}X_2}$: The right-hand lane marking uses metrics $h_{\mathbf{c}X_2} = \{e_{R,1}, e_{R,2}\}$ with $e_{R,1}$ and $e_{R,2}$ defined the same as $e_{L,1}$ and $e_{L,2}$.
- $h_{\mathbf{c}X_3}$: The far right-hand lane marking uses metrics $h_{\mathbf{c}X_2} = \{e_{F,1}, e_{F,2}\}$ with $e_{F,1}$ and $e_{F,2}$ defined the same as $e_{L,1}$ and $e_{L,2}$.
- $h_{\mathbf{c}l_1}$: The link between the left-hand and right-hand markings use compatibility metrics $h_{\mathbf{c}l_1} = \{e_{LR,1}, e_{LR,2}, e_{LR,3}\}$ with $e_{LR,1}$, $e_{LR,2}$, and $e_{LR,3}$ defined as in Section 2.3.1.
- $h_{\mathbf{c}l_2}$: The link between the right-hand and far right-hand markings use compatibility metrics $h_{\mathbf{c}l_2} = \{e_{RF,1}, e_{RF,2}, e_{RF,3}\}$ with $e_{RF,1}$, $e_{RF,2}$, and $e_{RF,3}$ defined the same as $e_{LR,1}$, $e_{LR,2}$, and $e_{LR,3}$.
- $h_{\mathbf{c}l_3}$: The link between the left-hand and far right-hand markings use compatibility metrics $h_{\mathbf{c}l_3} = \{e_{LF,1}, e_{LF,2}, e_{LF,3}\}$ with $e_{LF,2}$ and $e_{LF,3}$ defined the same as $e_{LR,2}$ and $e_{LR,3}$, and with $e_{LF,1}$ defined as being the absolute difference between the average width between the left-hand and far right lane markings and double the nominal expected lane width.
- $h_{\mathbf{c}l_4}$: The link between all three parts use the compatibility metric $h_{\mathbf{c}l_4} = \{e_{LRF,1}\}$ with $e_{LRF,1}$ defined as the distance from the center of the estimated ego-lane to the origin of the BEV image.
- $h_{\mathbf{t}X_1}$: The left-hand lane marking uses tracking metrics $h_{t_{X_1}} = \{e_{t_{L_1}}, e_{t_{L_2}}, e_{t_{L_3}}\}$ with $e_{t_{L_1}}$, $e_{t_{L_2}}$, and $e_{t_{L_3}}$ defined as in Section 2.3.1.
- $h_{\mathbf{t}X_2}$: The right-hand lane marking uses tracking metrics $h_{t_{X_2}} = \{e_{t_{R_1}}, e_{t_{R_2}}, e_{t_{R_3}}\}$ with $e_{t_{R_1}}$, $e_{t_{R_2}}$, and $e_{t_{R_3}}$ defined the same as $e_{t_{L_1}}$, $e_{t_{L_2}}$, and $e_{t_{L_3}}$.
- $h_{\mathbf{t}X_3}$: The far right lane marking uses tracking metrics $h_{t_{X_3}} = \{e_{t_{F_1}}, e_{t_{F_2}}, e_{t_{F_3}}\}$ with $e_{t_{F_1}}$, $e_{t_{F_2}}$, and $e_{t_{F_3}}$ defined the same as $e_{t_{L_1}}$, $e_{t_{L_2}}$, and $e_{t_{L_3}}$.

The key difference between ξ_1 and ξ_2 is that ξ_2 includes a third part for the far right lane marking that is individually treated the same as the other two markings, but is different in that its links treat the whole model as covering two adjacent lanes instead of a single lane.

A limitation of the metrics used in models ξ_1 and ξ_2 is that none of the metrics account for the location of image features in the image. Features located close to the vehicle are treated the same as features at the far end of the BEV, where the BEV may be less accurate. Metrics that provide for decreasing feature confidence based on distance are the subject of future work.

3.4 Stereo Filtering

The probabilistic lane detection method discussed thus far in Chapter 3 provides detection with a potentially very low false alarm rate with results discussed in greater detail in Chapter 4. However, there remains a category of false lane marking candidates that the method has severe difficulty correctly classifying. These high-confidence misclassifications stem from the application of a planar homography transform to an image containing significant out-of-plane elements.

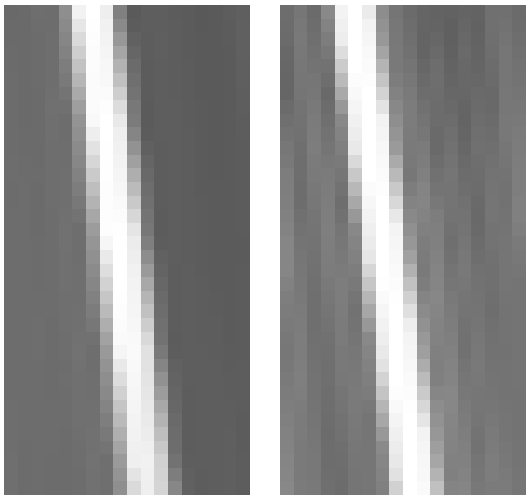
The planar homography correction for perspective distortion transforms the image such that the imaged line of infinity maps to infinity in the BEV image [13]. As a result, any vertical edges in the image are stretched out in the BEV image, with the stretching becoming increasingly severe the farther a point is from the estimated ground plane, and growing without bound with image points lying on the imaged line of infinity being located infinitely far away in the BEV image. Figure 3.1 illustrates the streaking distortion created by vertical out-of-plane edges and features

(a) (b)

Figure 3.1: BEV Distortions: (a) An image of a road scene containing obstacles and out-of-plane features. (b) The resulting Bird's Eye View image. Even relatively small features in the source image produce large streaks if they occur far enough above the ground plane. The rear window of the leading vehicle is not prominent in the source image but takes up roughly a third of the BEV image.

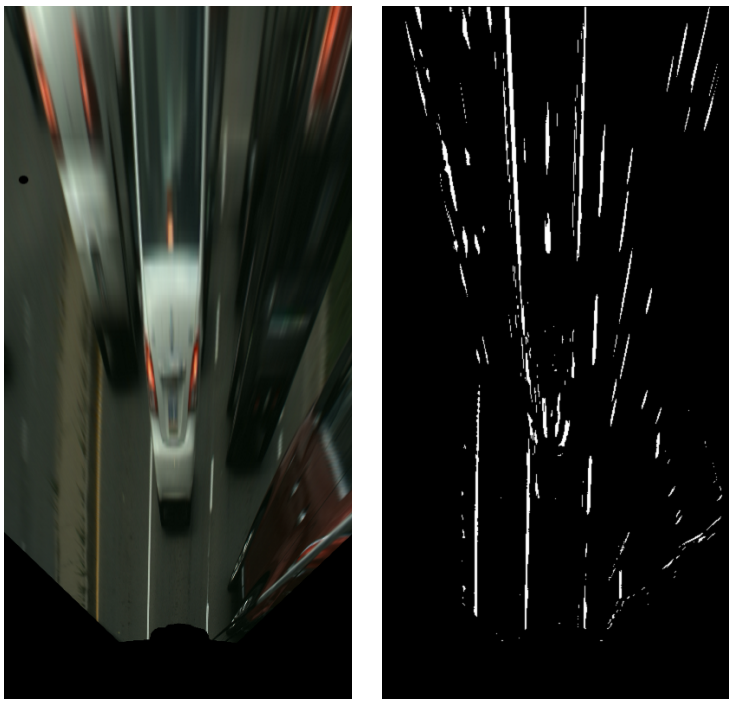
If the distorted vertical features also happen to contain a dark-light-dark edge transitions, then the resulting distortion can look identical to true lane markings at a local pixel level. Consequently, the local neural network pixel patch classifier identifies the distorted vertical features as lane markings. With even a trained human observer being unable to correctly classify the distorted pixel patch as not representing a lane marking pixel in many cases, it is difficult if not impossible for the neural network patch classifier or indeed any local patch classifier to correctly classify these distorted features. Figure 3.2 illustrates the similarity of these distorted pixel patches to true lane markings and Figure 3.3 illustrates

the consequence of this similarity on the resulting support score image.



(a) (b)

Figure 3.2: Pixel Patch Classification Errors from Distortion: (a) A pixel patch representing a true lane marking that is classified as a lane marking pixel. (b) A pixel patch containing a distorted signpost against a dark background that is incorrectly classified as a lane marking pixel. The similarity of (a) and (b) make correct classification of many distorted vertical features effectively impossible for the local patch classifier. Pixel patch examples (a) and (b) obtained from the KITTI Roads data set [11].



(a) (b)

Figure 3.3: Distortion in Patch Classified Image: (a) A BEV image with distorted out-of-plane features. (b) The resulting binary image produced after applying the neural network classifier.

The distorted vertical features also have characteristics in common with true lane markings beyond local pixel patch classification. The perspective distortion often causes these vertical features to stretch out as lines along the BEV image producing false hypotheses that are geometrically similar to true lane markings in that they have extremely low curvature and do not have curvature changes without lane marking support. Since heuristic metrics used for part-specific lane marking evidence are solely comprised of local pixel support scores and the geometrically inspired curvature penalty, the distorted vertical features can produce extremely high confidence lane marking candidates that are indistinguishable from true lane markings within the framework outlined in Chapters 2 and 3. An example of an image producing a high confidence false positive is shown in Figure 3.4. If the features maintain a consistent position relative to the ego vehicle, such as in the case of a leading vehicle as in Figure 3.4, then even the tracking components of the system are unable to remedy the situation.

Figure 3.4: High Confidence False Positive from Distortion: This distorted rear window of the leading vehicle produces a very high confidence false positive with an assessed likelihood of 91 percent even after tracking.

An additional problem posed by the vertical features is the crowding imposed upon RANSAC. The vertical features incorrectly identified as lane markings by the patch classifier are often prominent both in terms of their individual pixel populations and in their sheer number, as is the case in Figure 3.3(b). These features greatly increase the proportion of outliers in the population of line segments, which either results in a dramatic decrease in the probability of RANSAC generating a correct hypothesis in a given number of iterations, or results in a dramatic increase in the number of RANSAC iterations required to achieve a given probability of correct hypothesis generation [10].

A simple solution to the problem posed by out-of-plane features that will be shown to be effective in Chapter 4 is to utilize a stereo camera to filter out vertical features above the roadway. Since the main problem for the filter to address here is to screen out the types of vertical features that create high confidence false positives that the local patch classifier cannot detect, a simple threshold is used where points more than 25cm above the estimated roadway are not used in constructing the BEV images. Other approaches to detect objects

can also be employed, such as visual vehicle detection, laser depth measurement or even radar detection. In each case, pixels associated with non-ground objects can be removed from the lane marking detection process. In this specific application, a 25cm threshold catches the most severe out-of-plane features without being overly sensitive to the natural variations observed on real roadways that are not truly flat. Figure 3.5 illustrates the functionality of the stereo filter and demonstrates the significant removal of out-of-plane line segments. Detailed results are discussed in Chapter 4.

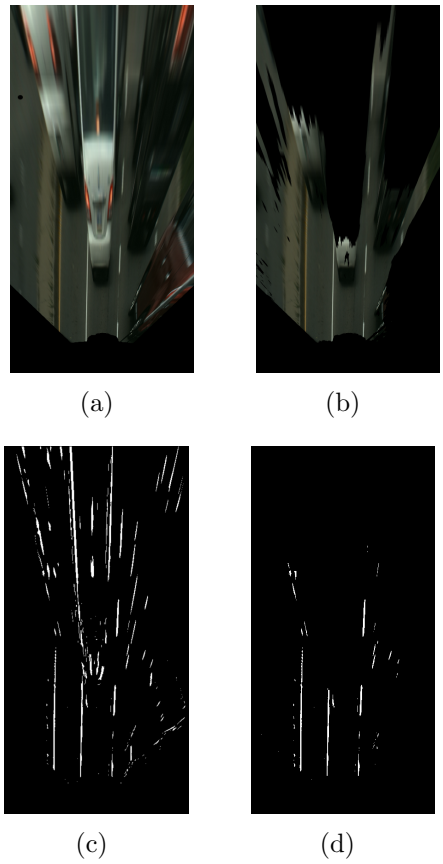


Figure 3.5: Result of Stereo Filter: (a) A BEV image produced without the stereo filter. (b) The corresponding BEV image produced with the stereo filter. (c) The resulting patch classified image without the stereo filter. (d) The resulting patch classified image with the stereo filter.

Chapter 4

Experiment and Results

This chapter describes validation and results of the lane detection method described in Chapters 2 and 3 as well as the experimental configuration used for implementation and testing. The contributions presented in Sections 3.2, 3.3, and 3.4, are each validated individually in addition to the testing of the lane detection system as a whole, with each test having its own data requirements.

There are a few limiting factors that determine the data requirements for testing. Implementation and training of the Neural Network patch classifier requires labelled pixel-level ground truth for a small set of images and is required for the input to all contributions in this work. Implementing and training the tracking components of the system, necessary for testing the system as a whole, require image sequences with sufficient temporal resolution for tracking to be possible and to provide significant overlap of image features between subsequent frames. Validation of the training method proposed in Section 3.2 requires a large number of images for the unsupervised learning method to converge to an appropriate optimum that does not overfit the training data. Validation of detectors for multiple models proposed in Section 3.3 requires that the data for both training and testing have sufficient representation of each of the included models if the models are to be trained and evaluated with representative results. Lastly, validation of the stereo filter proposed in Section 3.4 simply requires that the datasets include stereo image pairs, instead of only monocular images.

4.1 Datasets Used for Validation

No single publicly available dataset has been found that satisfies all of the testing requirements in this work. Consequently, validation is instead performed using a combination of three different data sets, including a new dataset collected in Waterloo region for this work - the WAterloo Representative Roads Dataset (WARR). The ROad MArkings (ROMA) dataset provides pixel-level ground truth for monocular lane marking classification in varying illumination conditions, but is severely limited in size. The KITTI Roads stereo dataset is widely used, but only provides ground truth for the drivable portion of the ego-lane and not for the markings themselves [11]. The KITTI Roads dataset also lacks sufficient temporal resolution to implement tracking and lacks the combination of breadth and depth to allow for training lane detectors with different marking models. The new WARR dataset contains a much larger amount of images with the depth and breadth to accommodate different marking models and with sufficient temporal resolution to allow for tracking, but is completely unlabelled. These datasets and their use in this work are discussed further in Sections 4.1.1 through 4.1.3.

4.1.1 The ROMA Dataset

The ROMA dataset [29] consists of 116 monocular images captured in a variety of illumination conditions with no overlapping features and with pixel-level ground truth of the lane markings, but lacks the detailed calibration information necessary to directly re-construct the Bird's Eye View Homography. An example image from the ROMA dataset and the corresponding ground truth image are shown in Figure 4.1:

In order to convert the images to the Bird's Eye View without the camera extrinsics, the method described in Section 2.1.2 was used to obtain the BEV Homography using image features that were easily obtained by inspection. Once converted to the BEV space, pixel patches were used as ground truth training targets for the neural network patch classifer, with the training completed using the MATLAB neural network toolbox [5]. The resulting trained patch classifier is then used to classify BEV images as the first phase of the hypothesis generation pipeline discussed in Section 2.2 used by this work.

4.1.2 The KITTI Roads Dataset

The KITTI Roads dataset is a road surface and lane estimation stereo image benchmark within the popular KITTI Vision Benchmark Suite [11]. It contains 289 training and

(a) (b)

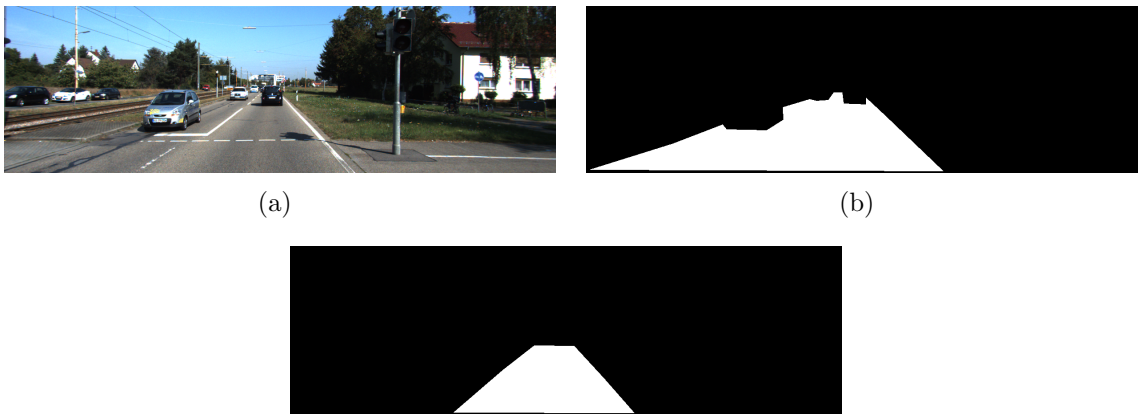
Figure 4.1: Example Images from the ROMA Dataset: (a) An image of a road scene. (b) The corresponding pixel-level ground truth image.

290 test image pairs divided approximately into even thirds in the categories Urban Unmarked(UU), Urban Marked (UM), and Urban Multiple Marked (UMM). Calibration matrices are provided for converting the camera frames to the road surface frame as well. Ground truth is provided for pixels belonging to the roadway as a whole and also for pixels belonging to the immediately drivable portion of the ego lane. An example image and corresponding ground truth images are shown in Figure 4.2:

The image sequences in the KITTI Roads dataset are spaced so as to minimize overlap of image features between subsequent frames and consequently precludes the use of the dataset for the validation of tracking-dependent systems. As the stereo filter proposed in Section 3.4 only requires stereo image pairs however, the KITTI Roads dataset is used for its validation, with testing performed against the UM and UMM training folders [11]. The existing libELAS stereo algorithm was used perform stereo matching and obtain depth maps from the stereo image pairs within the dataset [12].

4.1.3 The Waterloo Representative Roads Dataset

As no public dataset was found that provided a combination of the temporal resolution to allow for tracking and the depth and breadth required to train and test the lane detection method presented in Sections 3.2 and 3.3, a new dataset was collected. The WAterloo



(a) (b)

(c)

Figure 4.2: Example Images from the KITTI Dataset: (a) An image of a road scene. (b) The corresponding pixel-level ground truth image of the roadway. (c) The corresponding pixel-level ground truth image of the ego-lane.

Representative Roads Dataset (WARR) consists of 8644 stereo image pairs captured at approximately 3 Hz along a 40 km route around the Waterloo region shown in Figure 4.3.

Data capture was completed using a Carnegie Robotics MultiSense S7S stereo camera by mounting it above the roof-rack of a 2007 Saturn Vue Hybrid vehicle, as shown in Figure 4.4.

The dataset includes a wide variety of paved roads including single lane rural roads, multiple lane roads, labelled and unlabelled intersections, controlled access highways and more, with many independent physical instances captured of each type. The breadth provided by many independent physical examples of different lane marking models combined with the depth offered by the many images produced by a 3 Hz capture rate support the testing of both the training method proposed in Section 3.2 and the different lane detection configurations proposed in Section 3.3.

Critically, the WARR dataset also includes significant amounts of degraded and faded lane markings that are more representative of Canadian roads than other publicly available datasets. The 40km route covers the wide range of marking quality that Canadian drivers encounter regularly on paved roads. Consequentially, results obtained from testing against the WARR dataset should be much more representative of system performance on Canadian roads than other publicly available datasets, such as ROMA or KITTI. The WARR dataset was calibrated relative to the nominal road surface using a least-squares estimate

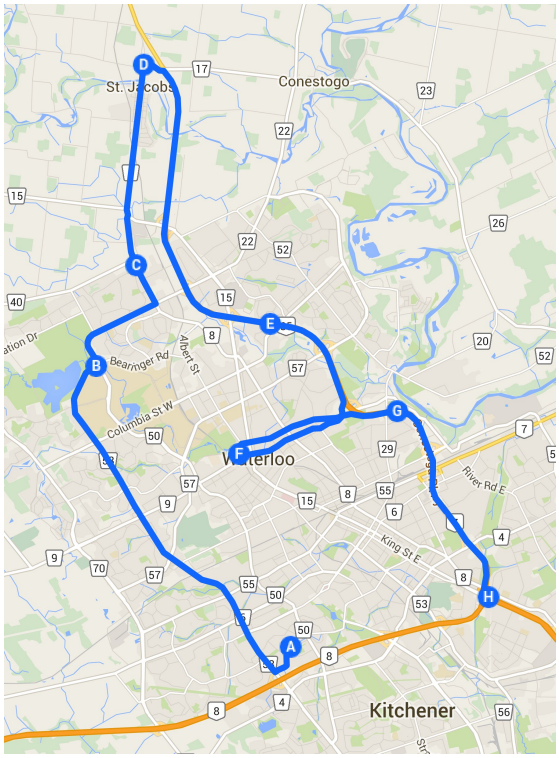


Figure 4.3: WARR Route Map: The 40km route traversed in the dataset.

and the relative positions and orientations of known features on the road surface.

4.2 Training Validation

As discussed in Section 2.4.3, the training method proposed in [19] fails to converge meaningfully on more complicated data sets such as the KITTI Roads data set or the Waterloo data set. Specifically, the method proposed in [19] either eventually converges to one of two terrible local optima depending on the initial parameter estimate. The first classifies solid lines as ego-lane markings and mostly ignores lane width, while the second classifies any features with ideal lane widths even if they lack any significant marking support. Either of these optima generate terrible classifications with most classifications being incorrect. It is therefore not possible to quantitatively compare the training method described in Section 3.2 against the training method originally proposed in [19] in any representative fashion.

(a) (b)

Figure 4.4: Data Capture Configuration for WARR Dataset: (a) The platform used for capturing the dataset. (b) A closer view of the camera.

The proposed EM-based training method is instead validated by testing the training method against different disjoint 80-image subsets of WARR training data and checking for indicators of severe bias, over-fitting, or erroneous local optima are not present. As the parameter estimate obtained using the full set of data and parameter prior uses all of the available knowledge, it is the best parameter estimate available and is denoted as $\boldsymbol{\mu}_*$, with the parameter prior distribution as described in Section 3.2.3. The parameter estimates obtained using three mutually disjointed subsets of the data in addition to the parameter prior are denoted as $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, and $\boldsymbol{\mu}_C$, while the parameter estimates obtained using these same disjoint subsets but without the parameter prior are denoted as $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$. Lastly, the parameter estimate obtained solely from using the parameter prior described in Section 3.2.3 and ignoring the data entirely during EM optimization is denoted as $\boldsymbol{\mu}_P$. Comparisons between these eight parameter estimates allow for the following indicators to be used for validation:

1. If the proposed training method does not suffer from excessive over-fitting, then the estimates $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, and $\boldsymbol{\mu}_C$ should be consistent with $\boldsymbol{\mu}_*$.

2. If $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_*$ differ significantly, then the parameter prior distribution is not dominant over the training data in the EM optimization.

3. If the parameter prior described in Section 3.2.3 is providing a significant contribution, then $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$ should differ from $\boldsymbol{\mu}_*$ by more than $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, and

$\boldsymbol{\mu}_C$.

4. If the estimates $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$ are inconsistent with each other, then the parameter prior may be preventing unfavourable local optima that would otherwise be representative of over-fitting.

The parameter estimates $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ were compared using the Kullback-Leibler divergence [20] of the probability distributions defined by their values, $\Pr(X|\boldsymbol{\mu}_1)$, with $\Pr(X|\boldsymbol{\mu}_2)$ defined as:

$$\mathrm{KL}(\boldsymbol{\mu}_1 \| \boldsymbol{\mu}_2) = \int_{-\infty}^{\infty} \Pr(X|\boldsymbol{\mu}_1) \log \frac{\Pr(X|\boldsymbol{\mu}_1)}{\Pr(X|\boldsymbol{\mu}_2)} dX \tag{4.1}$$

Comparing the parameter estimates $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, $\boldsymbol{\mu}_C$, $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, $\boldsymbol{\mu}_{C,X}$ and $\boldsymbol{\mu}_P$ to the best estimate $\boldsymbol{\mu}_*$ using (4.1) produces the differences shown in Table 4.1, while comparing the parameter estimates $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$ to each other using (4.1) produces the differences shown in Table 4.2.

Table 4.1: Kullback-Leibler Divergence of Disjoint Parameter Estimates to Best Estimate

$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_A$	$\boldsymbol{\mu}_B$	$\boldsymbol{\mu}_C$	$\boldsymbol{\mu}_{A,X}$	$\boldsymbol{\mu}_{B,X}$	$\boldsymbol{\mu}_{C,X}$	$\boldsymbol{\mu}_P$
$\mathrm{KL}(\boldsymbol{\mu}_1 \| \boldsymbol{\mu}_*)$	0.2314	0.4949	0.0533	4.6064	2.0229	1.5565	4.4111

Table 4.2: Kullback-Leibler Divergences of Disjoint Parameter Estimates without Parameter Prior

$\mathrm{KL}(\boldsymbol{\mu}_1 \| \boldsymbol{\mu}_2)$	$\boldsymbol{\mu}_2$		
$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_{A,X}$	$\boldsymbol{\mu}_{B,X}$	$\boldsymbol{\mu}_{C,X}$
$\boldsymbol{\mu}_{A,X}$	0	1.1289	1.1651
$\boldsymbol{\mu}_{B,X}$	1.0816	0	0.9562
$\boldsymbol{\mu}_{C,X}$	0.8699	0.7072	0

The training results shown in Tables 4.1 and 4.2 validate the training method proposed in Chapter 3 by satisfying each of validations 1 through 4:

1. Although each of the parameters $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, and $\boldsymbol{\mu}_C$ are obtained from mutually disjoint subsets of the training data used to estimate $\boldsymbol{\mu}_*$, all three are consistent with $\boldsymbol{\mu}_*$ and have lower divergences than any of the other estimates.

2. The parameter $\boldsymbol{\mu}_P$ differs significantly from $\boldsymbol{\mu}_*$, which indicates that the training data is not dominated by the parameter prior during optimization.

3. Each of $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$ diverge from $\boldsymbol{\mu}_*$ by 4.6064, 2.0229, and 1.5565, respectively, while $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, and $\boldsymbol{\mu}_C$ only differ from $\boldsymbol{\mu}_*$ by at most 0.4949, indicating that the parameter prior is providing a significant contribution to the estimate.

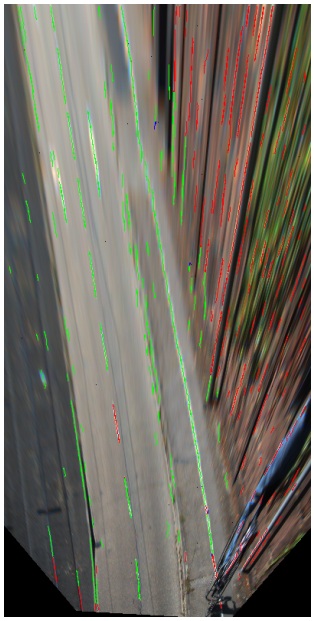
4. The local optima $\boldsymbol{\mu}_{A,X}$, $\boldsymbol{\mu}_{B,X}$, and $\boldsymbol{\mu}_{C,X}$ are inconsistent with each other with Kullback-Leibler divergences ranging from 0.7072 to 1.1651 indicating that these three estimates are local optima suffering from some degree of over-fitting and that the over-fitting is inhibited by the inclusion of the parameter prior. The inclusion of the parameter prior is at least justified in this sense.

Since there is no available ground truth, it is difficult to obtain further validation of the training method. While it is theoretically possible that the inclusion of the parameter prior distribution is causing $\boldsymbol{\mu}_*$ to converge to an undesirable local optimum, the fact that the natural local optimum of the prior-only estimate $\boldsymbol{\mu}_P$ is so much farther from $\boldsymbol{\mu}_*$ than any of $\boldsymbol{\mu}_A$, $\boldsymbol{\mu}_B$, or $\boldsymbol{\mu}_C$ indicates such a problem is unlikely. Additionally, the strong performance demonstrated in Section 4.4 provide further indication against $\boldsymbol{\mu}_*$ being an undesirable local optimum.

4.3 Stereo Filter Performance

The stereo filter proposed in Section 3.4 is evaluated based on its removal of line segments from the hypothesis generation pipeline that do not belong to lane markings when tested against the UM and UMM training folders of the KITTI Roads Dataset[11]. Additionally, the segments removed by the filter are then separately provided to the hypothesis generation pipeline and evaluated to assess the threats posed by false hypotheses arising from out-of-plane segments when they are not removed.

Line segments removed by the stereo filter that are not part of the lane markings are assessed as true negatives (TN), while line segments removed that are part of the lane markings are assessed as false negatives (FN). Conversely, line segments retained after filtering that are part of the lane markings are assessed as true positives (TP), while line segments retained that are not part of the lane markings are assessed as false positives (FP). An example of the filter's performance on a sample image from the KITTI dataset is shown in Figure 4.5 with removed segments shown in red and retained segments shown in green.



(a)

Figure 4.5: Stereo Filtering of Out-of-Plane Features: Highlighted red segments are removed by the stereo filter (Negatives), while green segments are retained (Positives).

Without fully labelled ground truth data, it is difficult to obtain full statistics for the performance of the proposed stereo filter. However, as is apparent in Figure 4.5, the number of false negatives (FN) is sufficiently low as to be easily counted manually, and since the total number of negatives are known by the algorithm the number of true negatives (TN) can be determined. Therefore, the Negative Predictive Value (NPV) of the stereo filtering process can be determined as well as the proportion of the population affected by the filter. These values are given in Table 4.3.

While an NPV of greater than 95% is promising, NPV alone does not prove strong filter performance as it does not account for relative proportions of true positives and false positives for segments not removed by our filter. Unlike the false negatives, the true positives are much more numerous and too laborious to count as is apparent in Figure 4.5. However, it was observed that in all of the examined images, the removed line segments belonging to lanes were a small minority of the total line segments belonging to the lanes, and also that the number of remaining line segments belonging to the lane were far more than sufficient for a suitable reconstruction of the lane marking as a whole.

Also notable is that the proportion of line segments removed by the filter is significant at

Table 4.3: Results of Stereo Filtering on Lane Marking Line Segments

Property	**Value**
Total TN	17334
Total FN	531
Total NPV	97.0%
Average Number of Segments per Image	183.15
Average Flagged as Negative per Image	93.54
Total Number of Segments (TP+TN+FP+FN)	34982
Total Flagged as Negative (TN+FN)	17865
Proportion Flagged as Negative (TN+FN)	51.1%

approximately 50% of the total population, which greatly reduces the size of the RANSAC phase of the hypothesis generation sub-problem. In conjunction with the very high NPV, the large population reduction greatly increases the chance that RANSAC can select an appropriate group of segments to create a valid hypothesis resulting in much better lane detection performance as will be demonstrated in Section 4.4.1.

Next, the stereo filter's impact on potentially dangerous false positive lane marking candidates is assessed by exclusively providing the segments removed by the filter to the hypothesis generation pipeline and evaluating the resulting lane marking candidates in the Bayesian framework of Section 3.3 using solely the current frame detections. This assessment is performed on the UMM training folder of the KITTI road dataset with the number of part hypothesis generations set to 50. This produces a total of 973 part hypotheses. Table 4.4 lists the relative likelihoods of this population of part hypotheses.

Table 4.4: Part Hypotheses Generated by Removed Line Segments

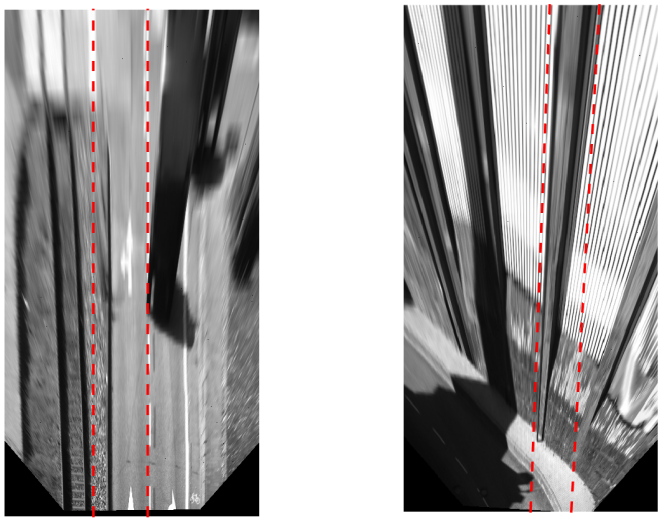
Property	**Value**	**Percentage**
Total Number	973	100
Total Less Likely Than Missing	438	45.0
Total More Likely Than Missing	535	55.0
Total More Likely Than False	92	9.5
Total More Than 95% Likely	42	4.3

The data in Table 4.4 confirms that while nearly half of the part hypotheses generated from the removed data are less likely than the missing part hypothesis and will be safely ignored, the majority of part hypotheses are more likely than the missing part hypothesis, meaning that if paired with a sufficiently strong matching part then they would form falsely viable hypotheses. Additionally, some of these false part hypotheses are extremely viable individually, with 4.3% of them having an individual likelihood above 95%. These hypotheses can easily generate dangerous results if combined with an even partially compatible hypothesis, such as occurs along the train tracks in Figure 4.6(a) or in front of the building in figure 4.6(b).

4.4 Lane Detection Performance

The lane detection system defined in Chapter 3 is evaluated against its performance on the WARR dataset. Comparative results will first be presented between current frame detection results for individual images both with and without the use of the stereo filter. Comparative results will then be presented comparing detection results for stereo filtered image sequences between current frame detections only and detections that also use the tracking components of the system. The detection results obtained by using the full system, combining the stereo filter and tracking, are then presented in more detail and with an examination of the detection failures. Lastly, the multiple lane model is demonstrated with current frame detection results to show that the training and detection methods can be extended to other categories of lane markings.

As the WARR dataset is completely unlabelled, detection results are assessed qualitatively as in [19]. For each image, the detection results are manually assessed as belonging to the following categories:



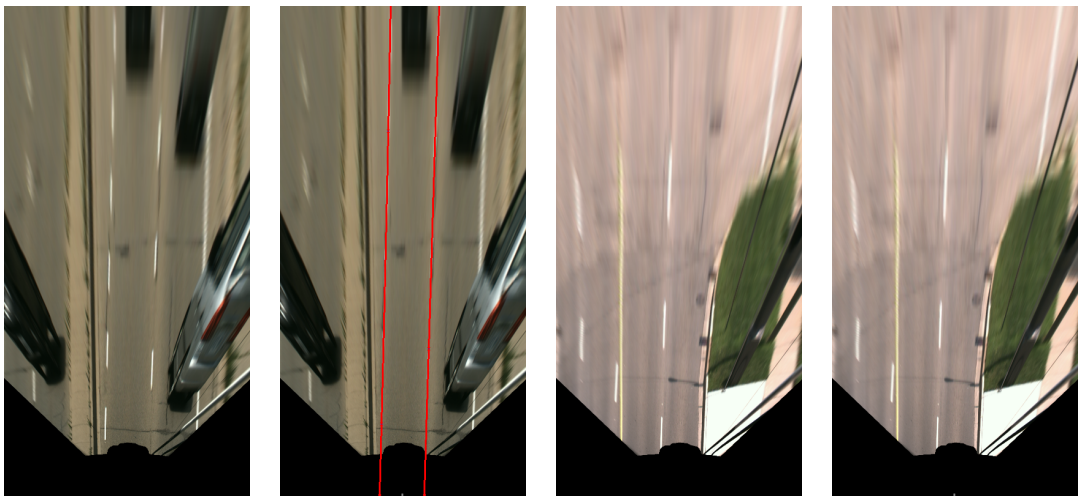
(a) (b)

Figure 4.6: Example Dangerous False Positives: (a) A false positive generated from train tracks. (b) A false positive generated from vertical features of a building.

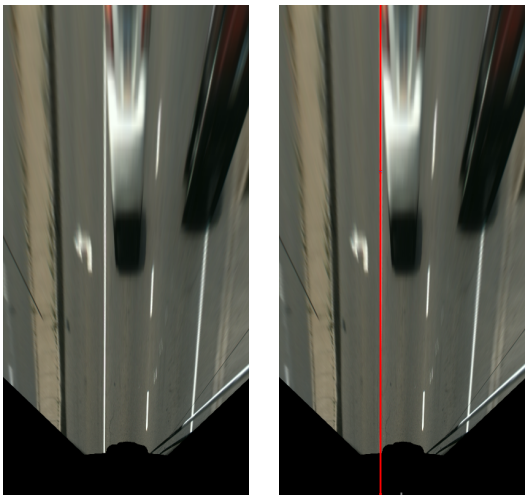
- Correct Detection (CD): The lane marking estimates were correctly detected or were correctly identified as missing or not conforming to the trained model.
- Slight Misalignment (SM): The lane marking estimates were slightly misaligned with at least one line segment being missed by the estimates but not more than one third of the line segments being missed for either marking.
- Major Misalignment (MM): The lane marking estimates contain a major misalignment and miss more than one third of the line segments belonging to either marking being missed.
- False Alarm (FA): One or more image features that are not truly lane markings were incorrectly identified as a lane marking instead of that marking being declared missing.
- False Failure (FF): Both lane markings were returned as missing when there truly was a lane marking present.

An additional group of categories exists for hypotheses where a single marking was returned as missing when there truly was a lane marking present. Since hypotheses with missing part hypotheses lack compatibility metrics, their individual part hypotheses are evaluated independently and they do not effectively fit into the above categorization. Instead they are categorized as a combination of the FF outcome with whatever categorization is appropriate for the non-missing part: CD/FF, SM/FF, MM/FF, or FA/FF. These detection categories are illustrated in Figures 4.7, 4.8, and 4.9.

Of the detection categories, Correct Detections (CD) are the most favourable, with False Failures (FF) being less favourable but still a much better outcome then any of the other failure categories as it is not a dangerous failure. The FF outcome is unique in that it is both a positive and negative result. The system evaluates that it did not find a lane marking, which is correct, however there truly is an undetected lane marking in the image that the system should have detected, making the outcome also a failure. The FF outcome is not a dangerous failure in comparison to the other failure categories because it does not introduce any errors that are not known to the system; the algorithm knows that it did not find the markings and can trigger an appropriate fail-safe response such as alerting the driver, whereas the other failure outcomes return erroneous results as though they were true and introduce unknown failure modes to the system. Of the dangerous failure modes, the Major Misalignments (MM) and False Alarms (FA) are the most dangerous as they can produce lasting errors in estimates for lateral lane position or lane heading or both. Slight Misalignments (SM) are less dangerous as they produce smaller errors for lane position and heading, and are typically corrected and refined through tracking.

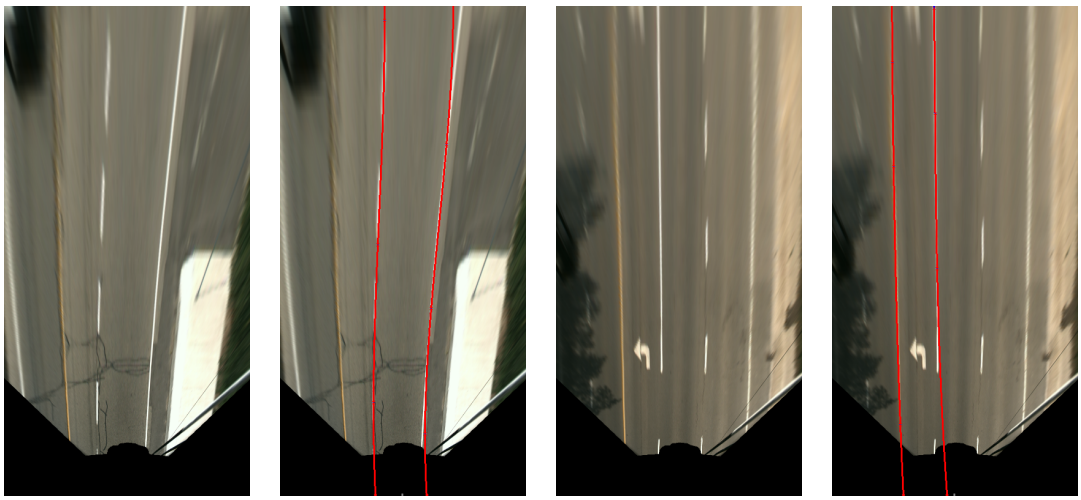


(a) (b) (c) (d)

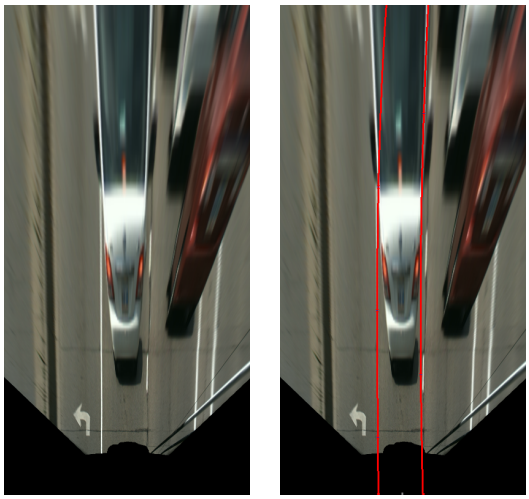


(e) (f)

Figure 4.7: Non-Dangerous Qualitative Assessment Categories: (a) & (b): Source and example of CD outcome. (c) & (d): Source and example of FF outcome showing no detection. (e) & (f): Source and example of CD/FF outcome.

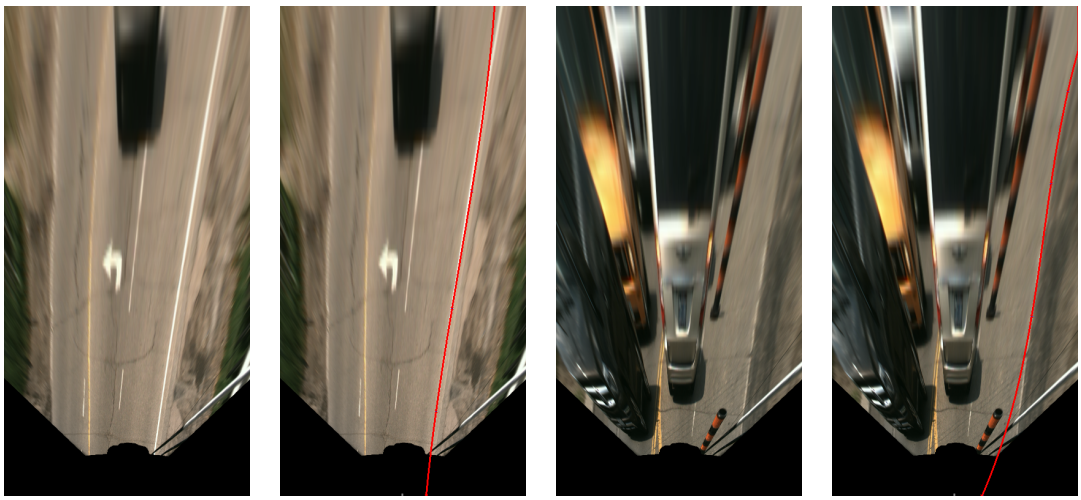


(a) (b) (c) (d)

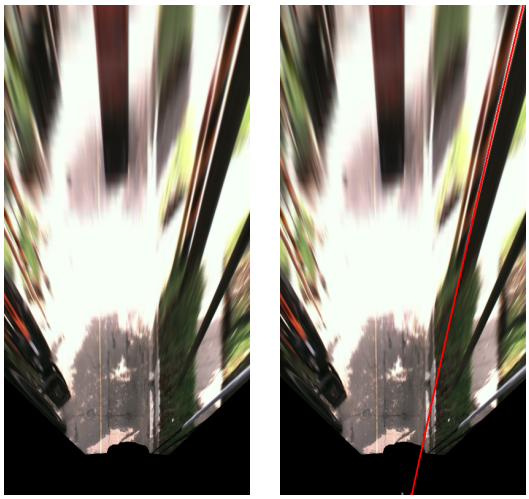


(e) (f)

Figure 4.8: Potentially Dangerous Qualitative Assessment Categories: (a) & (b): Source and example of SM with right-hand hypothesis missing either ends of the true marking. (c) & (d): Source and example of MM with detection of the adjacent lane instead of the ego-lane. (e) & (f): Source and example of FA with the right-hand hypothesis tracking the rear window of the leading vehicle instead of the true markings.



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



(e) (f)

Figure 4.9: Potentially Dangerous Qualitative Assessment Categories with Partial False Failures: (a) & (b): Source and example of SM/FF with right-hand hypothesis missing far end of the true marking. (c) &(d): Source and example of MM/FF with marking for adjacent lane being detected instead. (e) & (f): Source and example of FA/FF with out-of-plane post being detected as a lane marking.

4.4.1 Monocular vs Stereo

A randomly selected set of 250 test images were selected from the WARR dataset and evaluated without tracking components for detection results both with and without the stereo filter proposed in Section 3.4. The required current frame detection parameters were obtained by using the proposed training method on a random set of 250 images selected from a disjoint training partition of the WARR dataset, with training being performed once with the stereo filter and once without. The current-frame detection test results are detailed in Table 4.5 in order of increasing category danger.

Table 4.5: Current Frame Detection Results on 250 Images: Mono vs Stereo

Category	**Mono**	**Stereo**
Non-Dangerous Categories	193 (77.2%)	224 (89.6%)
CD	123 (49.2%)	135 (54%)
CD/FF	56 (22.4%)	43 (17.2%)
FF	14 (5.6%)	46 (18.4%)
Potentially Dangerous Categories	57 (22.8%)	26 (10.4%)
SM/FF	6 (2.4%)	4 (1.6%)
SM	(1.2%)3	4 (1.6%)
MM/FF	12 (4.8%)	13 (5.2%)
FA/FF	14 (5.6%)	0 (0.0%)
MM	10 (4.0%)	5 (2.0%)
FA	12 (4.8%)	0 (0.0%)

Table 4.5 shows that the stereo filter provides a significant improvement to the current frame detection results. Specifically, the stereo filter is intended to reduce the risk posed by FA outcomes created by out of plane segments, and appears to have been very successful in this set of test images; it has completely eliminated the FA/FF and FA outcomes which made up a total of 26 images, 10%, of the outcomes in the monocular case, with these 26 images instead generating CD, CD/FF, or FF outcomes. The stereo filter does not appear to have had a significant impact on the other categories as the other failure categories are typically due to image features that are unaffected by the stereo filter, or due to model failures and will be discussed in greater detail in Section 4.4.3.

4.4.2 Current Frame Detections vs Tracking

A randomly selected set of 50 sequences of 20 images in length were evaluated using two variants of the proposed detection method with the stereo filter present. The first uses only the current frame detection results and treats each image in a sequence in isolation. The second uses the tracking components of the system and so represents the full system proposed in Chapter 3. Both use the current frame parameters obtained from training the stereo detector as in Section 4.4.1, while the tracking system parameters are obtained by using the proposed training method on 250 randomly selected sequential image pairs selected from the disjoint training partition of the WARR dataset. The detection results are detailed in Table 4.7 in order of increasing category danger.

Table 4.6: Image Sequence Detection Results: Current Frame Only vs Tracking

Category	**Current Frame Only**	**Tracking**
Non-Dangerous Categories	917 (91.7%)	947 (94.7%)
CD	515 (51.5%)	460 (46.0%)
CD/FF	190 (19.0%)	292 (29.2%)
FF	212 (21.2%)	195 (19.5%)
Potentially Dangerous Categories	83 (8.3%)	53 (5.3%)
SM/FF	4 (0.4%)	1 (0.1%)
SM	17 (1.7%)	1 (0.1%)
MM/FF	38 (3.8%)	43 (4.3%)
FA/FF	0 (0.0%)	0 (0.0%)
MM	14 (1.4%)	6 (0.6%)
FA	10 (1.0%)	2 (0.2%)

The most notable results provided by including the tracking components is a significant reduction in the number of potentially dangerous outcomes, but also of CD outcomes in lieu of CD/FF outcomes. The conversion from CD to CD/FF outcomes is largely due to the impact tracking has on emerging parts, with current-frame-only detection able to respond to changes immediately, and tracking detection taking a few frames to respond if the emerging part has very weak part-specific evidence. This was observed when passing through intersections with a dashed lines taking longer to be re-detected after disappearing,

as shown in Figure 4.10.

The increase in CD/FF outcomes relative to CD outcomes from emerging parts was somewhat offset by cases where the tracking components generated CD outcomes instead of FF outcomes. Many of the FF and CD/FF outcomes in the current frame case occur when hypotheses are detected, but do not have enough support to overcome the null-hypothesis provided by the missing-part case. The inclusion of temporal and past evidence through tracking allows these hypotheses to build support from frame to frame and eventually overtake the null hypothesis, most notably for dotted lane markings as shown in Figure 4.11 but also for poor quality or faded lane markings.

Additionally, the dangerous failure categories, especially SM and FA, are also reduced by the inclusion of tracking. As the correct hypotheses build temporal support over time, they overtake the false positive hypotheses producing SM or FA outcomes, resulting in the near elimination of the SM failure mode. A significant number of MM detection errors remain after the inclusion of tracking. A more detailed evaluation of the full system performance and its failure modes follow in Section 4.4.3.

4.4.3 Full System Performance

A randomly selected set of 250 sequences of 20 images in length were evaluated using the full detection system proposed in Chapter 3 in order to obtain a larger sample of the system's remaining failure modes, with the detection parameters from Section 4.4.2 being used. The detection results are detailed in Table 4.7 in order of increasing category danger. A more detailed assessment of the failure modes present in the full detection system follows.

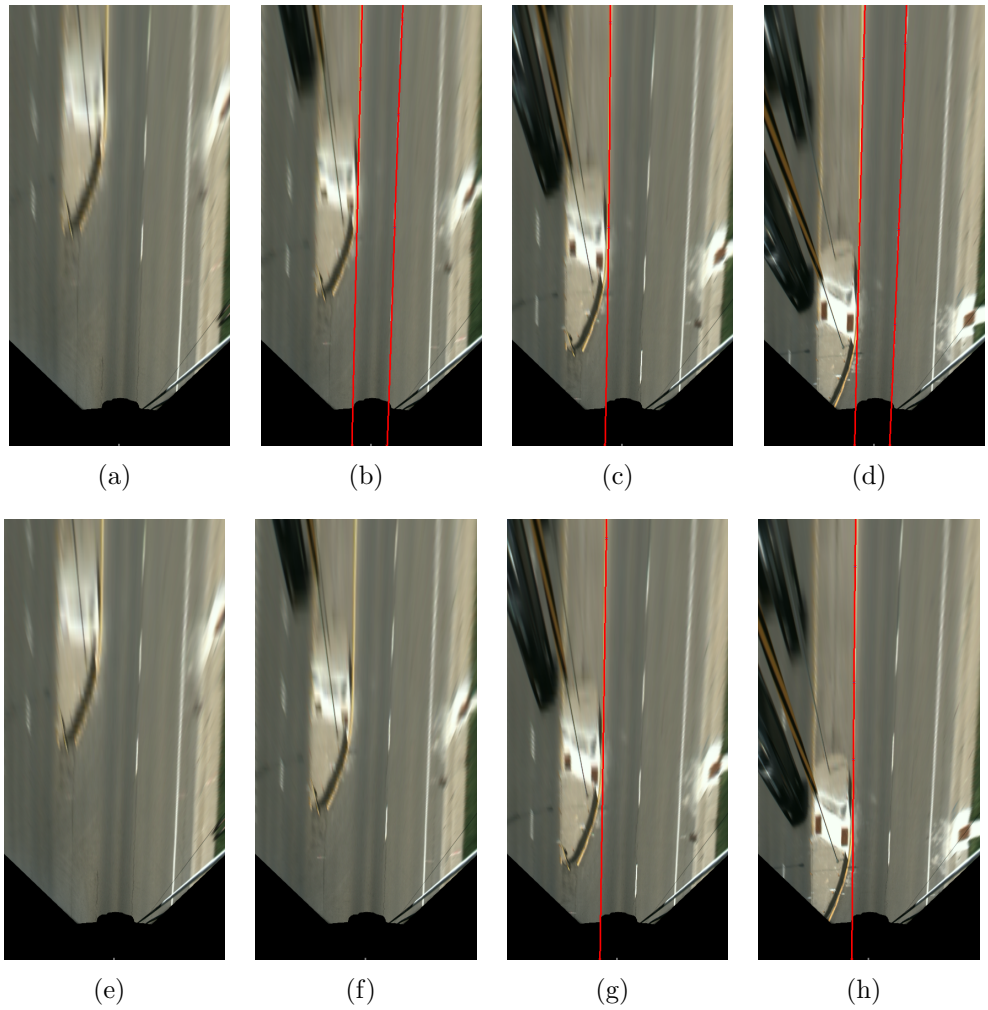


Figure 4.10: Emerging Part Example: (a-d): Four frame sequence of current-frame-only detection results. (e-h): Four frame sequence of detection results with tracking. Tracking slows down the response to emerging parts.

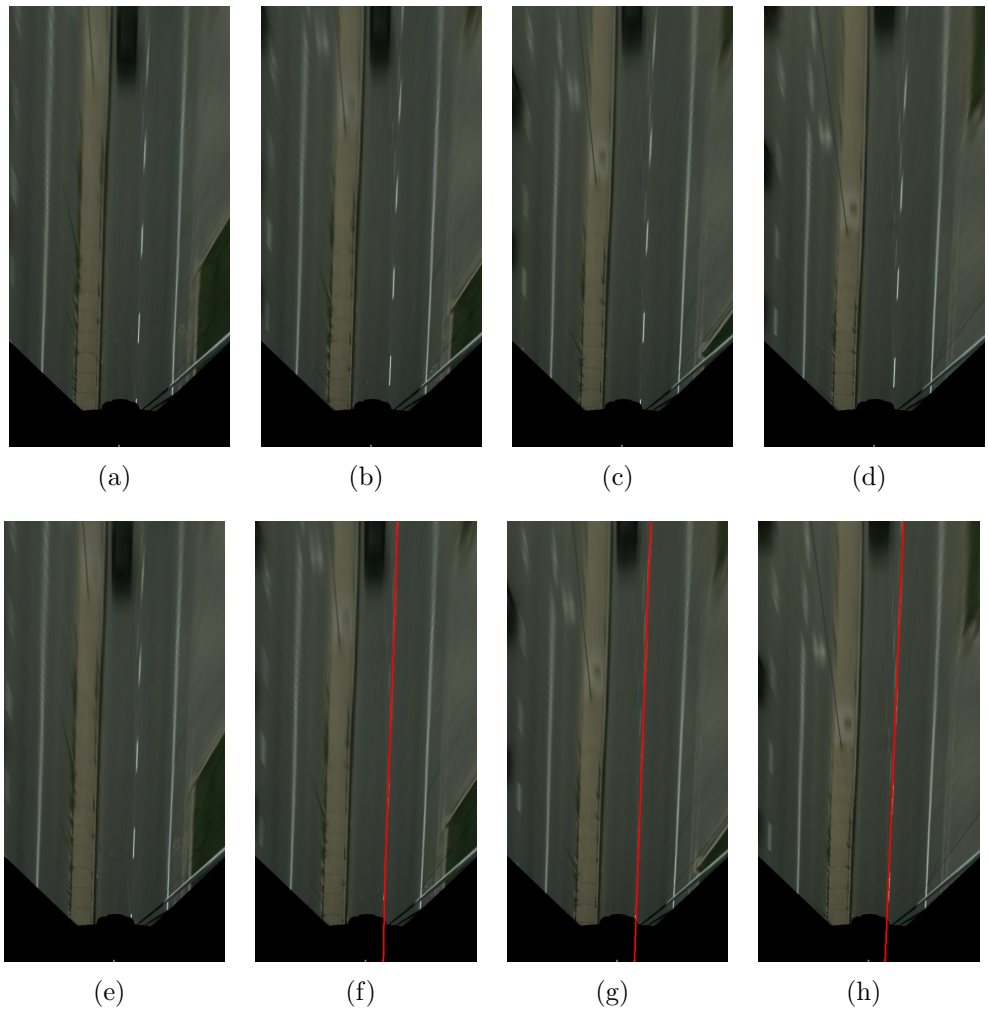


Figure 4.11: Four Frame Tracking Example: (a-d): FF detection results occurring without tracking. (e-h): CD detection results occurring with Tracking after initial FF frame.

Table 4.7: Detection Results: Full System Performance

Category	**Number of Frames**	**Percentage (%)**
Non-Dangerous Categories	4685	93.70
CD	2495	49.90
CD/FF	1228	24.56
FF	962	19.24
Potentially Dangerous Categories	315	6.30
SM/FF	33	0.66
SM	6	0.12
MM/FF	203	4.06
FA/FF	12	0.24
MM	34	0.68
FA	27	0.54

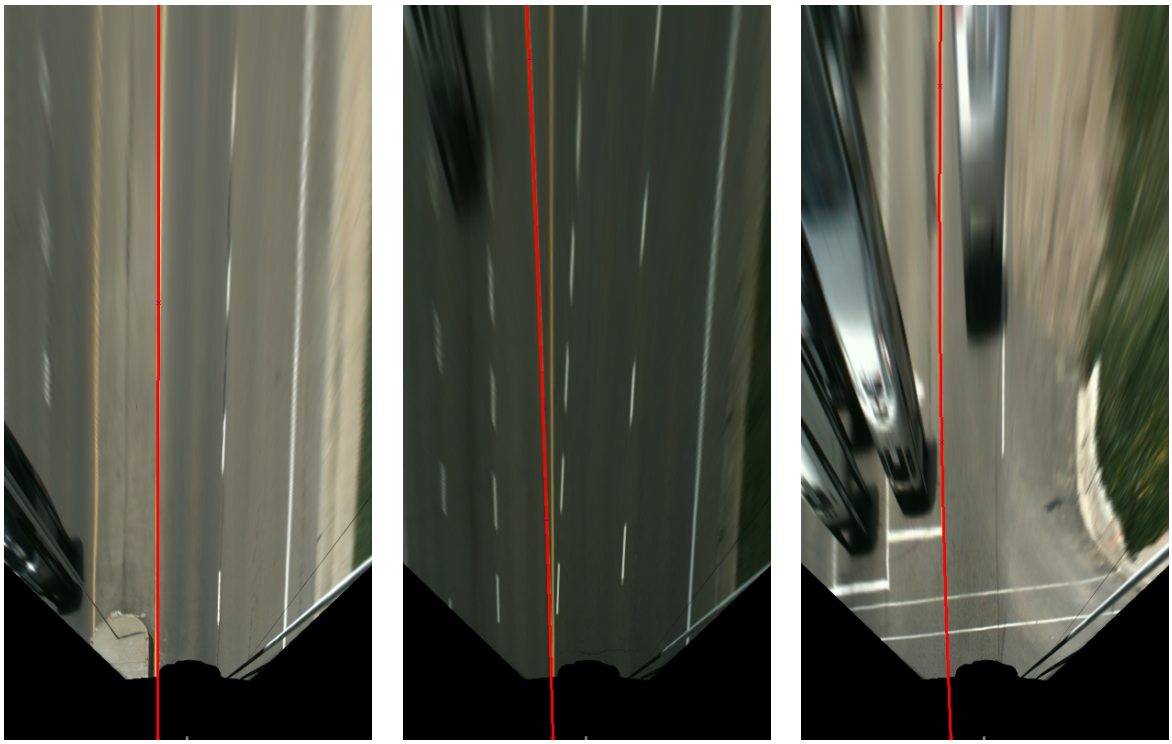
As Table 4.7 shows, the majority of outcomes are not dangerous, with the system either detecting the markings perfectly, or correctly declaring that it has not found them. Reviewing both the dangerous and non-dangerous failures, i.e. all outcome categories except for CD, it is possible to categorize the failure modes of the method so that potential future improvements can be identified. Table 4.8 lists the various failure modes as well as both the number of non-CD outcomes as well as the potentially dangerous outcomes that they generate. Note that some of the failure modes co-occur, for example a scene that does not conform to the single-lane model may make a tracking error more likely.

Table 4.8: Detection Results: Failure Modes

Failure Mode	**Safe Failures**	**Potentially Dangerous Failures**
Model Failures	1539	286
Multi-lane Roads	1524	281
Dashed Markings	1412	270
Intersections	207	23
Merging or Splitting Lanes	138	162
External Disturbances	182	12
Turns or Lane Changes	81	0
Construction Pylons	20	7
Speed Bumps or Potholes	13	1
Excessively Challenging	68	4
Other Identified Failure Modes	59	30
Transient Tracking Failure	59	27
Stereo Filter Failures	0	3
Total - Known and Unknown Causes	2190	315

As Table 4.8 indicates, the vast majority of failures are due to model failures, especially for the potentially dangerous failures where more than 90% are due to model failures. These failures constitute situations where the single-lane detection model described in Chapter 3 fails to describe the observed scene. The most commonly encountered model failures were due to multiple-lane scenarios, dashed markings, splitting or merging lanes, and intersections with examples shown in Figure 4.12. All of the dashed marking examples and splitting or merging examples co-occur within multi-lane examples.

Dashed lane markings are a challenge for this method because the metrics defined in [19] do not account for the spatial distribution of the lane marking pixels, only the total number along the hypothesis. In this sense, hypotheses with randomly distributed pixel support can be treated the same as hypotheses with structure to their pixel distribution. Adapting the lane marking support metric to account for the periodic distribution of lane marking pixels in dashed lines could mitigate this failure mode. Since dashed markings also occur in most of the multi-lane failure examples, stronger support for dashed markings could greatly improve the multi-lane failure mode as well, addressing the vast majority of



(a) (b) (c)

Figure 4.12: Model Failures: (a): Multiple lanes with a dashed right-hand marking. (b): Splitting lanes with multiple viable candidates for the left-hand marking. (c): Intersection example. The system is slow to respond to the emerging right-hand part hypothesis.

failures. Adaptations to support both solid markings and dashed lane markings are the subject of future work.

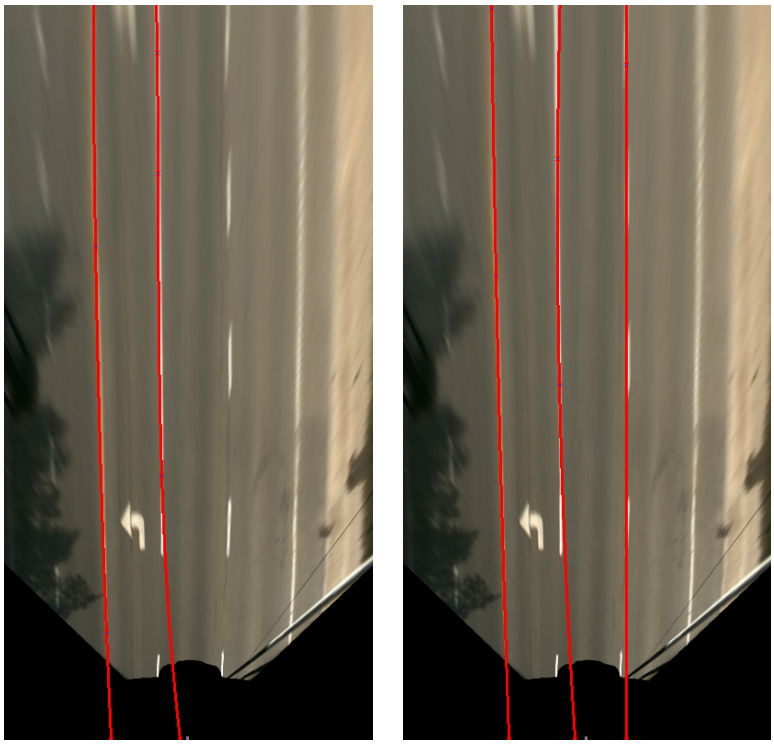
Multiple lane scenarios can be difficult because not only do they frequently have dashed markings, but they also often have additional lane markings for adjacent lanes that compete with the ego-lane markings for support. The metric for lateral lane position mitigates the challenge of multiple lanes for situations with solid but lower quality markings, but was not sufficient to overcome the combined challenge of multiple lane scenarios with dashed lane markings.

Scenarios with merging or splitting lanes are especially challenging for this method. They typically include both multiple-lane and dashed marking challenges in addition to the presence of additional overlapping hypotheses introduced from the split or merge. As illustrated in Figure 4.12(c), splitting or merging lanes often have far too many viable lane marking candidates for the single-lane model to be effective.

All of the model failures, which account for the vast majority of the dangerous outcomes and the overall failures, can be mitigated or corrected by using different models as proposed in Section 3.3. For example, the ξ_2 model for detecting both the ego-lane and adjacent right lane produces the following current frame detections where the ξ_1 model failed as shown in Figure 4.13.

If multiple models of the form described in Section 3.3 are trained and available it may be possible to run their detection methods in parallel with the common model failures of one detector addressed by a second. If any navigation information or crude map information is available, then the task of selecting the appropriate detector can be simplified. For example, if the vehicle is known to be on a multi-lane highway then the single lane detector can be ignored. Effectively utilizing multiple detectors remains the subject of future work.

A final caveat to these results is that all of the detection methods and results presented in this work use only camera images. No other sensor data, such as GPS/IMU, laser intensity, or wheel odometry are used. If other sensor data is available to provide estimates of the vehicle's position, velocity and orientation, then the temporal heuristics could be adapted to provide much stronger tracking results. Estimates of the vehicle's position and orientation could also allow for a dynamically calculated BEV image instead of relying on static calibration, which would improve the systems response in the presence of external vehicle disturbances such as speed bumps or potholes. Incorporation of a motion model into this system is the subject of future work.



(a) (b)

Figure 4.13: Model Comparison: (a): Dangerous misalignment occurring with the ξ_1 model detector. (b): Correct detection using the ξ_2 model. The ξ_1 model misses the right hand marking and instead detects the adjacent lane, while the ξ_2 instead detects both lanes and avoids a dangerous misalignment.

Chapter 5

Conclusion

Lane detection is a challenging problem that is central to the field of autonomous driving. Systems aiming for broad deployment, particularly in Canada, must have reliable and robust methods for dealing with degraded or worn lane markings. These systems must also be able to accommodate a wide range of lane marking configurations without failing in a way that poses a risk to passenger or driver safety.

This work presents several improvements to the current state of the art for robust detection of lane markings. The training method proposed in previous work for training Dynamic Bayesian Networks (DBN) with heuristic features is reformulated to use an augmented form of the Expectation Maximization algorithm. The new training method is not sensitive to initialization, is robust against unwanted local optima, and requires no labelling provided several easily satisfied criteria are met. The current Bayesian formulation was then reformulated so that the null-hypothesis is returned instead of a false positive whenever the system's confidence is below a desired threshold, creating a safe failure mode that can trigger appropriate corrective action if necessary. The Bayesian formulation is also generalized to support different lane marking configurations. A stereo filter is then proposed as a method for reducing dangerous false positives caused by out-of-plane features.

The proposed methods are demonstrated against several datasets, including the new WAterloo Representative Roads (WARR) dataset. The WARR data set has more challenging scenarios that are representative of the degraded or worn markings regularly encountered on Canadian roads than any publicly available dataset to date. The proposed methods are effective in scenarios matching the expected single-lane model and to fail safely in most non-matching scenarios. The vast majority of demonstrated failures were model failures stemming from scenarios that cannot be suitably described by the simple single-

lane model, and can be mitigated in future work by implementing additional detectors as cases of the generalized formulation proposed. As model failures account for more than 90% of the observed potentially dangerous failures, if these future detectors can achieve comparable performance to the single-lane detector for scenarios matching the expected model, then the overall dangerous failure rate could be reduced dramatically.

The framework presented in this work and the corresponding detection and training methods have significant potential for extension. All of the currently used evidence metrics and distributions stem from human generated heuristics evaluated as distributions in a Bayesian fashion. Improvements such as sensor or navigation integration could greatly improve the system's performance. Any additionally available information can be used directly in the system so long as some simple criteria are satisfied that permit the use of the unlabelled training method. Additional future work includes real-time implementation, multiple model support, dynamic BEV homography calculation, additional sensor integration, and adverse weather extensions. With an extremely low dangerous failure rate for trained scenarios and extendability to other scenarios, this work makes a significant contribution towards the performance and reliability required for broader deployment of autonomous vehicles.

References

[1] M. Aly. Real time detection of lane markers in urban streets. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 7–12, Eindhoven, Netherlands, June 2008.

[2] N. Apostoloff and A. Zelinsky. Robust vision based lane tracking using multiple cues and particle filtering. In *IEEE Intelligent Vehicles Symposium*, pages 558–563, June 2003.

[3] Ji. H. Bae and J. B. Song. Monocular vision-based lane detection using segmented regions from edge information. In *2011 8th International Conference on Ubiquitous Robots and Ambient Intelligence, URAI*, Incheon, Korea, Dec 2011.

[4] Aharon Bar Hillel, Ronen Lerner, Dan Levi, and Guy Raz. Recent progress in road and lane detection: a survey. *Machine Vision and Applications*, 25(3):727–745, 2014.

[5] M. H. Beale, M. T. Hagan, and H. B. Demuth. *Matlab Neural Network Toolbox*. The Mathworks.

[6] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[7] J. Byun, K. I. Na, M. Noh, and S. Kim. ESTRO: design and development of intelligent autonomous vehicle for shuttle service in the ETRI. In *Workshop on Planning, Perception, and Navigation of Intelligent Vehicles at the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Algarve, Portugal, Oct 2012.

[8] J. Choi, J. Lee, D. Kim, G. Soprani, P. Cerri, A. Broggi, and K. Yi. Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment. *IEEE Transactions on Intelligent Transportation Systems*, 13(2):974–982, 2012.

[9] H. Deusch, J. Wiest, S. Reuter, M. Szczot, M. Konrad, and K. Dietmayer. A random finite set approach to multiple lane detection. In *2012 IEEE 15th International Conference on Intelligent Transportation Systems (ITSC)*, pages 270–275, Anchorage, AK, USA, Sept 2012.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[11] J. Fritsch, T. Kuhnl, and A. Geiger. A new performance measure and evaluation benchmark for road detection algorithms. In *2013 16th International IEEE Conference on Intelligent Transportation Systems - (ITSC)*, pages 1693–1700, The Hague, Netherlands, Oct 2013.

[12] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *Asian Conference Computer Vision (ACCV)*, Queenstown, New Zealand, Nov 2010.

[13] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

[14] L. Hazelhoff, I. Creusen, T. Woudsma, X. Bao, and P. H. N. de With. Combined generation of road marking and road sign databases applied to consistency checking of pedestrian crossings. In *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pages 439–442, Tokyo, Japan, May 2015.

[15] P. Jeong and S. Nedevschi. Efficient and robust classification method using combined feature vector for lane detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):528–537, 2005.

[16] R. Jiang, R. Klette, T. Vaudrey, and S. Wang. New lane model and distance transform for lane detection and tracking. In *Computer Analysis of Images and Patterns*, pages 1044–1052. Springer, 2009.

[17] R. Jiang, R. Klette, T. Vaudrey, and S. Wang. Lane detection and tracking using a new lane model and distance transform. *Machine vision and applications*, 22(4):721–737, 2011.

[18] Z Kim. Geometry of vanishing points and its application to external calibration and realtime pose estimation. Technical Report Paper UCB-ITS-RR-2006-5, Institute of Transportation Studies, 2006.

[19] Z. Kim. Robust lane detection and tracking in challenging scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 9(1):16–26, March 2008.

[20] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[21] B. P. Lathi. *Linear Systems and Signals*. Oxford University Press, Oxford, UK, 2nd edition, 2009.

[22] J.C. McCall and M.M. Trivedi. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *Intelligent Transportation Systems, IEEE Transactions on*, 7(1):20–37, March 2006.

[23] Mobileye Technologies Limited. *Mobileye 5-Series User Manual*, 2011.

[24] S. Sehestedt, S. Kodagoda, A. Alempijevic, and G. Dissanayake. Robust lane detection in urban environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 123–128, San Diego, CA , USA, Oct 2007.

[25] B. S. Shin, Z. Xu, and R. Klette. Visual lane analysis and higher-order tasks: a concise review. *Machine Vision and Applications*, 25(6):1519–1547, 2014.

[26] M. Smart and S. L. Waslander. Stereo augmented detection of lane marking boundaries. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2491–2496, Las Palmas, Spain, Sept 2015.

[27] J. Tao, B.S. Shin, and R. Klette. Wrong roadway detection for multi-lane roads. In *Computer Analysis of Images and Patterns*, pages 50–58. Springer, 2013.

[28] Tesla Motors, Inc. *Model S Owner's Manual*, 2016.

[29] T. Veit, J. P. Tarel, P. Nicolle, and P. Charbonnier. Evaluation of road marking feature extraction. In *Proceedings of 11th IEEE Conference on Intelligent Transportation Systems (ITSC'08)*, pages 174–181, Beijing, China, 2008.

[30] Y. Wang, E. K. Teoh, and D. Shen. Lane detection and tracking using b-snake. *Image and Vision computing*, 22(4):269–280, 2004.

[31] B. F. Wu, C. T. Lin, and Y. L. Chen. Dynamic calibration and occlusion handling algorithms for lane tracking. *IEEE Transactions on Industrial Electronics*, 56(5):1757–1773, 2009.