

Bayesian Reasoning Using 3D Relations for Lane Marker Detection

Bart Boesman, Lars Baunegaard With Jensen, Emre Başeski, Nicolas Pugeault and Norbert Krüger

Department of Industrial Engineering
KaHo St. Lieven University College Belgium
Email: bartboesman@hotmail.com

The Mærsk Mc-Kinney Møller Institute
University of Southern Denmark
Email: {lbwj, emre, nicolas, norbert}
@mmmi.sdu.dk

Abstract

We introduce a lane marker detection algorithm that integrates 3D attributes as well as 3D relations between local edges and semi-global contours in a Bayesian framework. The algorithm is parameter free and does not make use of any heuristic assumptions. The reasoning is based on the complete conditional probabilities of the different cues which are estimated from a training set. The importance of the individual visual cues can be computed using a standard measure and the cues can then be combined in an optimal way. In addition we show that when doing 3D reasoning, the uncertainties connected to the reconstruction process need to be taken into account to make the reasoning process more stable. The results are shown on a publicly available data set.

1 Introduction

Lane marker detection is a crucial part of autonomous vehicles and driver assistance systems. It usually combines four different stages (for a recent review, see [1, 2, 3]):

Feature extraction: Visual features, usually edge features, that potentially correspond to lane structure become extracted.

Post processing: The features extracted in the first step are filtered to eliminate non-lane structures.

Lane modeling: A global street model is used, based on the post-processed structure to further eliminate non-lane contours

Tracking: The global lane model is used to track the lane structure in the scene.

In this paper, we mainly address the second and third step. In the literature, most constraints that are used for the post processing step are in 2D as for example collinearity (see, e.g., [4]) or more global relations (e.g., connected to the vanishing line) coded

for example in a Hough transformation (see, e.g., [5]). Less often 3D features are used as in [6] and then these features are constrained to the 3D position. The application of the constraints is often done in a heuristic rule based way (see, e.g., [7]). However, there exist statistically grounded methods like [1] where probabilistic dependencies are used to find the collinear lane markings starting from a known lane point, given the fact that lane markings are aligned in the heading direction of the car.

In our approach, we start with the observation that lane structures are defined not only by their local appearance but also by their relations to each other: lane structures, besides usually being connected to collinear high contrast edges, they are usually found close to the ground plane, are mutually coplanar and are part of parallel structures with certain distance ranges connected to the lane width. Note that, such 3D relations are by definition invariant under perspective transformations.

None of these different relations on its own is sufficient to describe lane structures; however, as shown in this paper, their statistical combination results in a rather stable classification. For this, we combine the different cues in a Bayesian reasoning process. In our system, we do not make use of any prior assumption on the road structure except that it can be defined as a statistical combination of the mentioned relations. Hence we neither make use of any explicit road model nor make any assumption about the relative importance of the individual cues. Instead the road model is learned and the individual cues are combined in an optimal way. We can show that by this Bayesian combination, we can reach a performance of about 88% correctly classified lane entities from one stereo frame without using any temporal regularization (which usually further stabilizes the extraction process).

The novelty of our approach is twofold. First,

in addition to 2D information and simple local 3D point information that is used commonly, we make use of higher level 3D relations between groups of 3D line segments extracted by stereo. Second, for each cue we compute the complete conditional probability density given that an edge stems from a lane structure. This allows us to make a statement of the relevance of each individual cue.

In the context of our work, we reflect on two issues when dealing with 3D relations. First, limits of 3D reasoning stemming from the uncertainties involved in the reconstruction process need to be taken into account (in particular due to the large depth variation in street scenes). Second, since the space of relations grows exponentially with the number of entities being related to each other as well as the order of the relations as such, the reasoning process needs to be done on spatially extended groups instead of local entities to keep computational efficiency. In addition, we show that reasoning on semi-global contours reduces computational cost while increasing the performance since the relations can become more efficiently computed on non-local entities.

We have evaluated our approach on three publicly available sequences that are used in the context of benchmarking driver assistance systems¹. We used a set of hand labeled data for training and test phases, for a statistically plausible performance evaluation. The prior probabilities and conditional probability densities have been calculated by analyzing the frames that are included in the training set. In this manner, we learn a lane model that is trained on a certain set of frames (55 stereo image pairs). To enable a lane recognition that works on a broad range of situations, such as illumination changes, dashed or connected lane markings, curves, slopes, other traffic etc., a large variation has been used in the training data set.

The rest of the paper is structured as follows: In Section 2, we describe the visual features we use. In Section 3, the calculation of prior probabilities and conditional probability densities using the training set is discussed. This information is then used in a Bayesian framework to do reasoning on the test set, which is hand labeled as well. We evaluate the performance of our approach in Section 4 for local and semi-global visual features in the presence and

absence of uncertainty.

2 The Early Cognitive Vision System

In this work we make use of a visual representation based on local descriptors called primitives [8]. They are extracted sparsely along image contours and form a feature vector that contains visual modalities such as position, orientation, phase, color and optical flow ($\pi = (\mathbf{x}, \theta, \phi, (\mathbf{c}_l, \mathbf{c}_m, \mathbf{c}_r), \mathbf{f})$) where color of a patch is defined by left, right and middle color. 2D-primitives are matched across two stereo views and pairs of corresponding primitives afford the reconstruction of a 3-dimensional equivalent called 3D-primitive which is encoded by the vector $\Pi = (\mathbf{X}, \Theta, \Phi, (\mathbf{C}_l, \mathbf{C}_m, \mathbf{C}_r))$. Extracted 2D and 3D primitives for a sample stereo image pair are illustrated in Figure 1.

The local 2D and 3D primitives are grouped together by using the perceptual organization scheme described in [9] to create semi-global contour structures. Since contours are based on good continuation in terms of geometry and appearance, they also have modalities like color and orientation. This sparse and symbolic nature of the primitives and the contours allows for perceptual relations defined on them that express relevant spatial relations in 2D and 3D (e.g., coplanarity, co-colority) which can be applied in different contexts. In the rest of this section, a brief description of the relations that have been used as visual cues in this work is given.

Gradient: Since both primitives and contours have left and right colors, the gradient is defined as the color difference between the left and right side of an entity.

Ground Plane: A rough estimation of the ground plane can be done once yaw, pitch, roll angles and the height of the camera is known (see [10]). For any primitive or contour, this relation is defined as the Euclidean distance of the visual feature to the calculated ground plane.

Angle: While the angle between two primitives is defined as the angle between their orientation, the angle between two contours is defined as the angle between their principal components².

Normal Distance: For both primitives and contours, normal distance is the parallel distance between the entities. For primitives, it is defined as the

¹www.mi.auckland.ac.nz/EISATS, SET 3: Colour stereo sequences Drivsc

²The eigen-vector of the highest eigen-value in PCA.

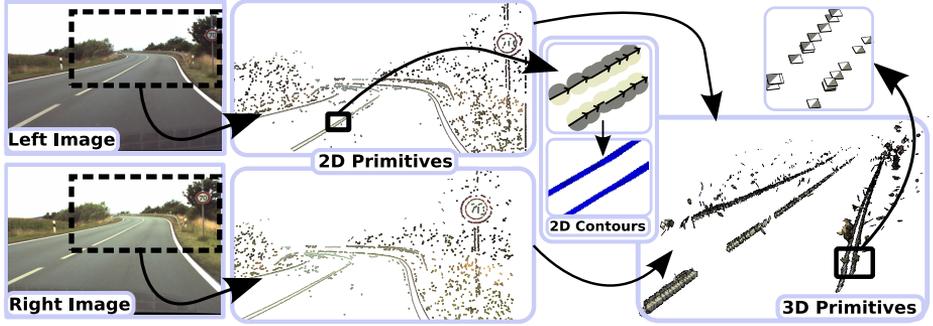


Figure 1: Extracted 2D and 3D primitives for a sample image pair. Note that, 2D primitives are used to reconstruct 3D primitives.

distance between one primitive and the line defined by the orientation and the position of the other. For contours, it is defined by the distance of the centroid of one contour to the line created by the orientation of the principal component of the other.

Coplanarity: Since a primitive has a position and an orientation, a common plane can be defined between two primitives. Similarly, a common plane can be defined for two non-collinear contours by fitting a plane. Coplanarity between two entities is defined as the distance between the entities and the common plane.

An important issue while reasoning in 3D is having an uncertainty model for the entities. For the 3D primitives discussed above, the uncertainty calculation has been shown by Pugeault et al. in [11]. In this work, the uncertainty of a contour is calculated as the mean of the matrix traces of the uncertainty matrices of the primitives that form the contour.

3 Bayesian Reasoning

To merge different cues as well as to deal with uncertainties, we make use of a Bayesian framework. The advantage of Bayesian reasoning is that it allows for: **a)** introduction of learning in terms of prior and conditional probabilities, **b)** assessing the relative importance of each cue for the detection of a given object, using the conditional probabilities.

Bayes' formula (e.g., see [12]) enables to infer the probability of an unknown cue or relation conditioned to other observable cues and to prior likelihoods. Let $P(c_i^e)$ be the prior probability of occurrence of the i^{th} cue c_i^e applied to an entity e (e.g.,

the probability that any primitive lies in the ground plane). Then, $P(c_i^e|e \in \mathcal{L})$ is the conditional probability of the visual cue c_i given an object \mathcal{L} .

Our aim is to compute the likelihood of an entity e being part of a lane \mathcal{L} given a number of visual cues relating to the entity:

$$P(e \in \mathcal{L} | c_i^e). \quad (1)$$

According to Bayes' formula, equation 1 can be expanded to:

$$\frac{P(c_i^e | e \in \mathcal{L}) P(e \in \mathcal{L})}{P(c_i^e | e \in \mathcal{L}) P(e \in \mathcal{L}) + P(c_i^e | e \notin \mathcal{L}) P(e \notin \mathcal{L})}. \quad (2)$$

In this work we assume independence between the cues c_1^e, \dots, c_n^e . If c_1^e, \dots, c_n^e are independent then $P(c_1^e, \dots, c_n^e | e \in \mathcal{L})$ can be written as:

$$P(c_1^e, \dots, c_n^e | e \in \mathcal{L}) = P(c_1^e | e \in \mathcal{L}) \cdot \dots \cdot P(c_n^e | e \in \mathcal{L}), \quad (3)$$

and

$$P(c_1^e, \dots, c_n^e | e \notin \mathcal{L}) = P(c_1^e | e \notin \mathcal{L}) \cdot \dots \cdot P(c_n^e | e \notin \mathcal{L}), \quad (4)$$

3.1 Prior Probabilities

The prior probability $P(e \in \mathcal{L})$ is necessary for calculating the posterior probabilities by following Bayes' theorem (see Equation (1)). We calculate the prior probabilities by using the hand labeled data and counting the entities that are on the lane and not on the lane for all the training images. The

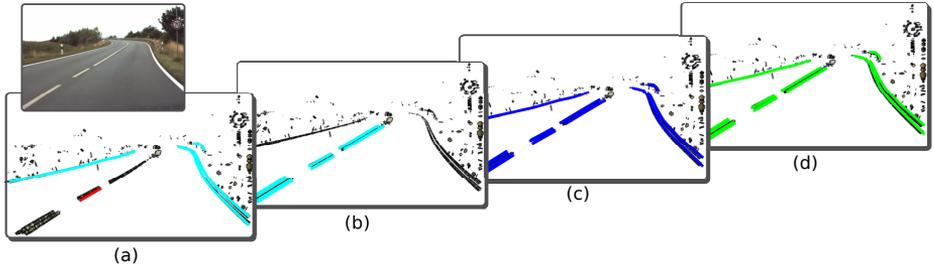


Figure 2: Illustration of 3D relations. (A selected 3D contour is marked in red in (a)). (a) All contours that have 1500-3500 mm normal distance to the selected contour. (b) All contours that have 0-200 mm normal distance to the selected contour. (c) All contours that are coplanar to the selected contour. (d) All contours that have 0-5° angle to the selected contour.

distribution of the prior probabilities for individual primitives and contours are displayed in Figure 3(a) and 4(a).

3.2 Conditional Probability Densities

For each relation, a conditional probability density is calculated from the training set. During the test phase, these probability densities are used to find the conditional probability of a calculated relation (e.g., $P(c_i^e | e \in \mathcal{L})$ and $P(c_i^e | e \notin \mathcal{L})$ can be calculated for a specific value of c_i^e , once the probability density is known.). The normalized conditional probability densities of the relations are shown in Figures 3(b-f) and 4(b-f). Note that, the x axis shows a possible value for a particular relation and the y axis shows the conditional probabilities. While green bars represent values for entities that are part of the lane ($P(c_i^e | e \in \mathcal{L})$), red bars represent values for entities that are not part of the lane ($P(c_i^e | e \notin \mathcal{L})$).

3.3 Explicit Relevance Measure

The relevance of each relation can be derived from the normalized densities by calculating the $L^1 - norm$ distance as the sum of absolute differences between the individual bins of the densities. Although this value is not used directly in the Bayesian reasoning process, it gives an idea about importance of an individual cue. For the densities X and Y with n bins, $L^1 - norm$ is defined as:

$$L^1 - norm(X, Y) = \sum_{i=0}^n (X_i - Y_i) \quad (5)$$

which corresponds to the difference between the areas of the two densities. Therefore, in this work, this value corresponds to the non-overlapping area between the part of lane (green) and not part of lane (red) densities for a given conditional probability density. Note that the highest possible value is 2 and a high value indicates a high importance for discriminating lane and non-lane structures. For example, the contrast relation applied on individual primitives in Figure 3(b) exhibits an $L^1 - norm$ value of 1.2. Here, a clear difference is visible between lane and non-lane entities. Lane entities exhibit a higher contrast gradient because of the white lane markings on a dark road. All the other non-lane entities generally display a low contrast gradient. Thus, the conditional probability density functions make already an a priori classification of lane and non-lane entities. To improve the classification, different relations or cues are merged into a Bayesian framework according to Equation 2.

3.4 The Effect of Uncertainty

As discussed earlier, one of the drawbacks of reasoning in 3D is the uncertainty of the data originating from 2D and reconstruction uncertainties. Although this uncertainty can be modeled, the use of data with very high uncertainty values may lead to a decrease in the performance of geometrical reasoning processes. A straightforward method to get rid of this problem is neglecting this kind of data by putting a threshold on the modeled uncertainty. An example is shown in Figure 5, where an uncertainty threshold is introduced for relations applied on contours. By eliminating uncertain entities, the

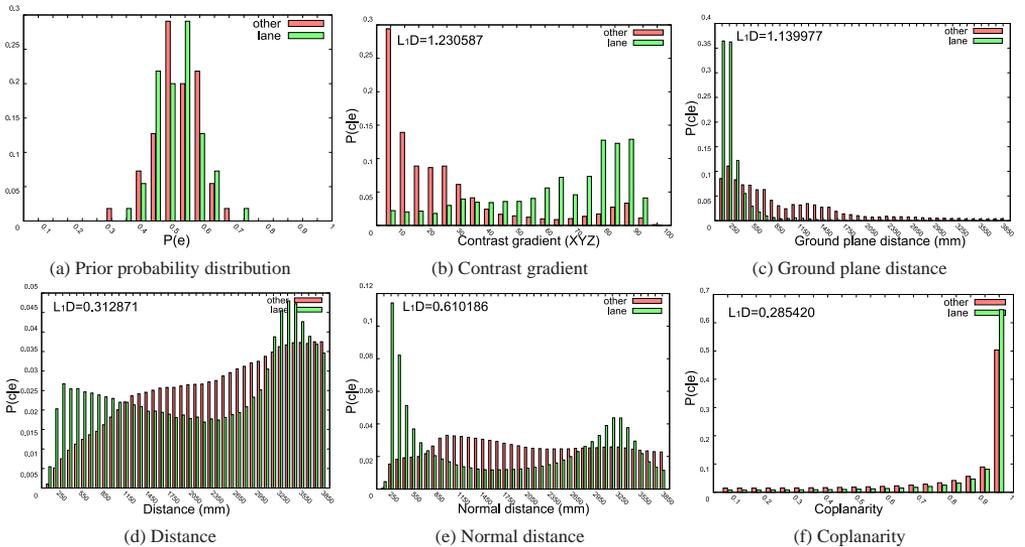


Figure 3: The prior probability distribution (a) and conditional probability densities (b-f) for relations between primitives.

L^1 – norm value of the different cues increases.

3.5 The Posterior Probabilities

Once the prior probabilities and the densities for the conditional probabilities are calculated from the training data, Equation 2 can be used to calculate the posterior probabilities which show the probability of primitives and contours of the test set being on the lane for the given cues. In Figure 6, the distribution of the posterior probabilities based on relations between primitives and between contours are shown. Note that the L^1 – norm value is highest for the case, where contours are used after an uncertainty thresholding. In Figure 7(a-c), samples of extracted lanes are shown. Once the reasoning is done in 3D, 2D data can be used to extend the information. For example in Figure 7(d-f), the 2D contours that contain the 3D lane contours are displayed.

4 Evaluation

Table 2 shows the classification of entities obtained from a test set of 30 hand labeled frames for three cases: In the first case, we make use of relations between primitives; in the second case, we use rela-

tions between contours that contain a maximum of 6 primitives and in the third case, we apply an uncertainty threshold to contours before we use their relations (1500 contours were eliminated with the thresholding). Since we use a hand labeled test set, each visual entity can be classified by using the calculated posterior probability. Table 1 shows the possible class labels, depending on the ground truth and the calculated posterior probabilities.

Table 1: Class labels of entities

		Ground Truth	
		lane	non-lane
Calculated	lane	true positive	false positive
	non-lane	false negative	true negative

The evaluation has been done by measuring two values for each case. We calculate the classification success rate as the percentage of true positives plus true negatives in the whole set. We also have a positive success rate, which is defined as the percentage of true positives in the set of true positives plus false positives. While the classification success rate mea-

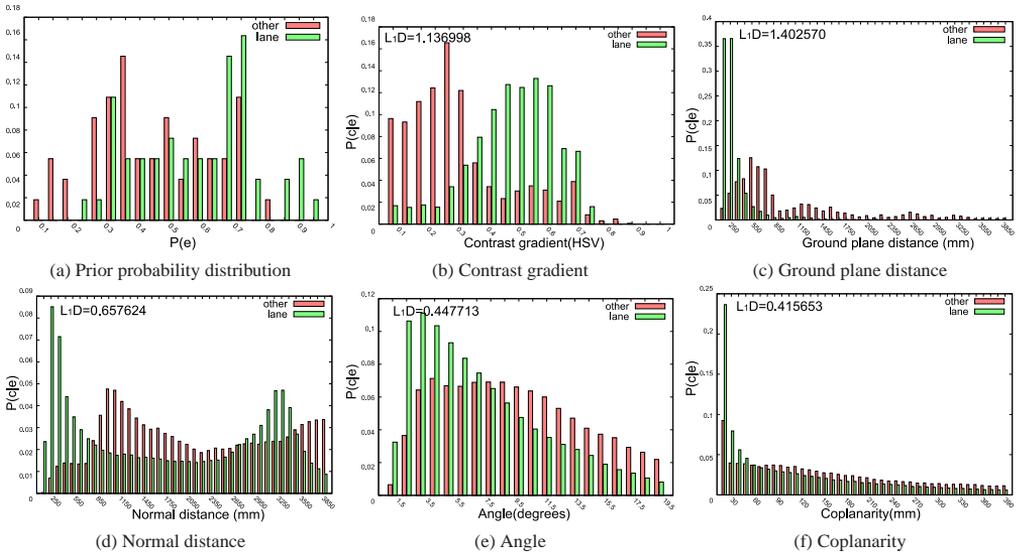


Figure 4: The prior probability distribution (a) and conditional probability densities (b-f) for relations between contours that contains maximum 6 primitives.

sures how successful the algorithm is for classifying entities in a scene as lane and non-lane, the positive success rate measures how successful the algorithm is for finding lane structures. When the relations between primitives are used, we obtain a classification success rate of 78.4% and a positive success rate of 58%. The usage of contour relations increases these ratios to a classification success rate of 79.6% and a positive success rate of 77%. Once the contour relations are used after an uncertainty threshold, we obtain a classification success rate and a positive success rate of 87.7%. Note that, the effect of loss of structure can be compensated by inferring from 2D, as discussed in Section 3.5.

5 Conclusion

We introduced a parameter free non-heuristic approach to characterize lane structures in a Bayesian reasoning process as a combination of 3D attributes and relations. We tested our algorithm in a publicly available data set. We investigated the relevance of the individual relations and we demonstrated the importance of 3D information as well as relational information for lane detection, which both is rarely used in current systems.

Acknowledgements

This work has been supported by the European Commission - FP6 Project DRIVSCO (IST-016276-2).

References

- [1] Joel C. McCall and Mohan M. Trivedi. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):20–37, 2006.
- [2] Y. Wang, E.K. Teoh, and D.G. Shen. Lane detection and tracking using b-snake. *Image and Vision Computing*, 22(4):269–280, April 2004.
- [3] Massimo Bertozzi, Alberto Broggi, and Alessandra Fascioli. Vision-based intelligent vehicles: state of the art and perspectives. *Robotics and Autonomous Systems*, 32(1):1–16, 2000.
- [4] E.D. Dickmanns and B.D. Mysliwetz. Recursive 3-d road and relative ego-state recognition. *IEEE Transactions on Pattern Analy-*

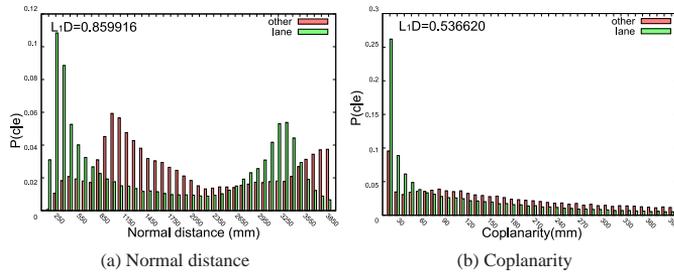


Figure 5: The conditional probability densities of two cues for contours after an uncertainty threshold

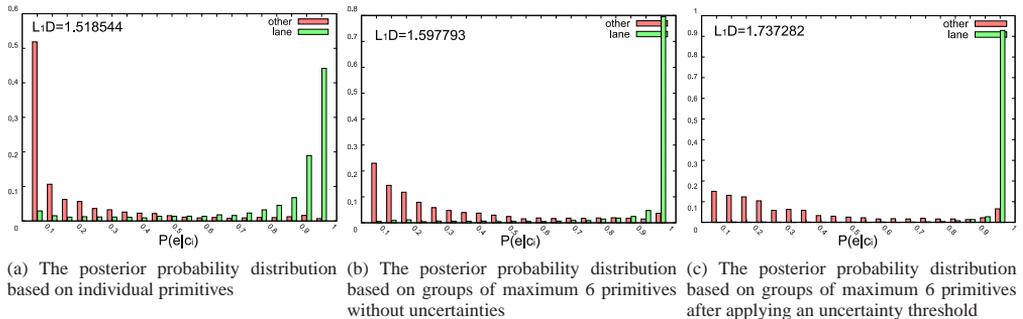


Figure 6: The posterior probability densities for individual and grouped entities, together with the influence of grouping on the $L^1 - norm$ value

sis and Machine Intelligence, 14(2):199–213, 1992.

[5] Joon Woong Lee. A machine vision system for lane-departure detection. *Computer Vision and Image Understanding*, 86(1):52–78, 2002.

[6] S. Nedeveschi, R. Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, and C. Pocol. 3D Lane Detection System based on Stereovision. In *IEEE Intelligent Transportation Systems Conference*, pages 161–166, Oct. 2004.

[7] Massimo Bertozzi and Alberto Broggi. Gold: A parallel real-time stereo vision system for generic obstacle and lane detection. *IEEE Transactions on Image Processing*, 7:62–81, 1998.

[8] N. Krüger, M. Lappe, and F. Wörgötter. Biologically motivated multi-modal processing of visual primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5):417–428, 2004.

[9] N. Pugeault, F. Wörgötter, and N. Krüger. Multi-modal Scene Reconstruction Using Perceptual Grouping Constraints. In *Proc. IEEE Workshop on Perceptual Organization in Computer Vision (in conjunction with CVPR’06)*, 2006.

[10] A. Broggi. A Massively Parallel Approach to Real-Time Vision-Based Road Markings Detection. In *Proceedings of the Intelligent Vehicles ’95 Symposium*, pages 84–89, 1995.

[11] N. Pugeault, S. Kalkan, E. Baseski, F. Wörgötter, and N. Krüger. Reconstruction Uncertainty and 3D Relations. In *Proceedings of Int. Conf. on Computer Vision Theory and Applications (VISAPP’08)*, 2008.

[12] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

Table 2: Classification of lane and non-lane entities

Class	Amount of selected entities		
	primitives	contours of 6 primitives	contours of 6 primitives with uncertainties
True Positive	12347	2666	2247
True Negative	21304	803	259
False Positive	8687	796	316
False Negative	572	91	34

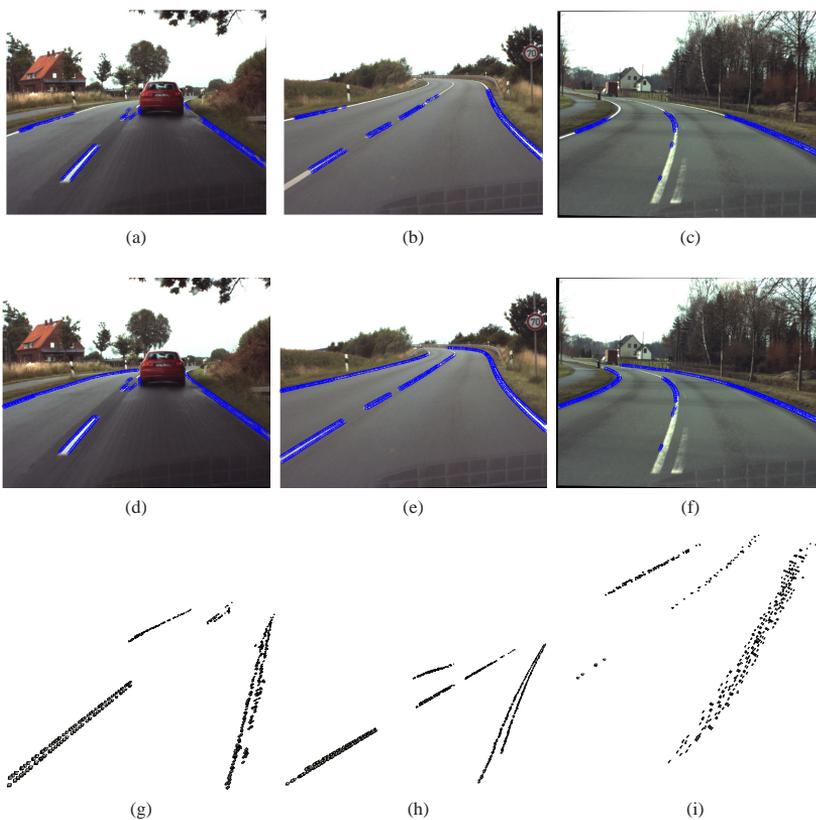


Figure 7: Results of the Bayesian framework applied on different frames. **(a-c)** Extracted lane structures by using only 3D reasoning. **(d-f)** 2D contours that contain the 3D lane contours. **(g-i)** 3D primitives corresponding to the 2D contours in (d-f).