Efficient Road Scene Understanding for Intelligent Vehicles Using Compositional Hierarchical Models

Daniel Töpfer, Jens Spehr, Jan Effertz, and Christoph Stiller, *Senior Member, IEEE,*

***Abstract*—In this article, we present a novel compositional hierarchical framework for road scene understanding that allows for reliable estimation of scene topologies, such as the number, location and width of lanes and the lane topology, i.e., parallel, splitting or merging. In our approach lanes and roads are represented in a hierarchical compositional model in which nodes represent parts of roads and edges represent probabilistic constraints between pairs of parts. A key benefit of our approach is the representation of lanes and roads as a set of common parts. This makes our approach applicable to scenes with rich topological diversity, while bringing along the much desired computational efficiency. To cope with the high-dimensional and continuous parameter space of our model and the non-Gaussian image evidence, we perform inference using nonparametric belief propagation. Based on this approximate inference algorithm, we introduce depth-first message passing for lane detection, that performs inference in several sweeps. Empirical results show that depth-first message passing requires significantly lower computation for performance comparable to classical belief propagation.**

***Index Terms*—Multi-lane recognition, Multi-feature fusion, Hierarchical graphical models, Nonparametric belief propagation.**

I. Introduction

THE ability of sensing and understanding the vehicles environment is a key technology for autonomous driving and Advanced Driver Assistance Systems (ADAS). Many of these applications require a robust estimation of geometrical road scene properties, such as the number, location and width of lanes as well as the recognition of the lane topology i.e., parallel, splitting or merging.

Lane perception, at least in its basic setting seams to be an easy task, since it only involves the recognition of the host lane. In fact, this basic tasks has been successfully treated in a large body of literature, e.g., [1]–[3]. While similar perception approaches are still used in commercial ADAS applications, such as lane keeping assist, these systems make strong assumptions on the structure of roads, e.g., a smooth and continuous curvature, parallel lanes and well-defined lane markings. Even so, these methods are sufficient for the recognition of lanes on well structured highways and highway-like roads, lane perception in less structured urban environments is still an open problem.

Still an open challenge is the development of multi-lane road recognition approaches for urban roads, as addressed by our approach. This is mainly due to complex lane topologies and a large amount of clutter and partial occlusion in urban scenes. Further, in contrast to highways lane markings are often not reliable as an individual cue. In order to allows for reliable results in complex real-world cluttered scenes, recent approaches in scene understanding use a combination of different sensory cues [4]–[6].

Similarly, our approach allows for the incorporation of different low-level visual cues. Thereby it is based on probabilistic graphical models allowing us to explicitly take into account that our knowledge about the environment is imperfect. A novelty of our models is that they do not impose any hard constraints on the lane geometry as imposed by the common clothoid or spline models. Instead, we assemble lanes as a composition of parts whose dependencies are encoded by weak probabilistic dependencies. A key benefit of these dependencies is that they incorporate prior scene knowledge which allows us to cope with clutter and partial occlusions.

This article extends our previous work [7] by means of an extended formal description that allows joint consideration of multiple visual cues, details on our real-time inference algorithms and extended experimental analysis covering complex urban scenarios.

II. Related Work

Lane detection and tracking has been successfully treated in a large body of literature. A Kalman filter for tracking the parameters of a clothoid model has been proposed in [1] and extended in many following approaches. The robust estimation of the lane course in [8] uses lane markings and a special particle filter. Fusing multiple visual cues including lane markings, edges, and road color has been proposed in [3]. In [9] the particle filter is used to fuse lane marking and curbstone cues for urban lane detection. A 3D spline model has been used as a lane model for rural roads whose parameters have been tracked in a Kalman particle filter in [2].

All of the above approaches have in common that they aim to detect the host vehicle lane, while constraining the features to be part of a specific lane geometry (i.e., clothoid or spline). Further, they assume that lanes and their visual cues are parallel. In contrast, our approach aims to detect multiple lanes with more complex lane toplogies, e.g., splitting and merging lanes. Further, we do not impose any hard constraints on the lane geometry. Instead, we express our prior expectations in the lane geometry using weak probabilistic constraints. These constraints allow us to not only account for spatial uncertainties, but to cope with clutter and partial occlusions. Issues that are not directly addressed by the above approaches.

Daniel Töpfer, Jens Spehr. and Jan Effertz are with the Driver Assistance and Integrated Safety Department, Group Research, Volkswagen AG, D-38436 Wolfsburg, Germany.

Christoph Stiller is with the Department of Measurement and Control, Karlsruhe Institute of Technology, 76128 Karlsruhe, Germany.

Manuscript received January 15, 2014; revised April 27, 2014.

Lane and road perception is also investigated in the field of scene understanding, where it is often a sub-task of a more holistic environment perception task. Many approaches treat scene understanding as a segmentation problem. A conditional random field is used by Wojek et al. [10] to jointly perform object detection and scene labeling. Sturgess et al. [11] developed a segmentation of road scenes based on appearance cues and structure-from-motion features. In contrast to segmentation based approaches, high-level scene understanding approaches aim to infer a more holistic scene knowledge, often using generative graphical models. Wang et al. [12] propose a hierarchical Bayesian network to perform activity detection in traffic scenes from a static platform. Interdependent Dirichlet processes are used in [13] to understand the behavior of moving objects in the scene. A generative model for 3d scene understanding was proposed in [6]. Their model jointly performs multi-class object detection, object tracking, scene labeling and 3d geometry estimation using a reversible-jump Markov Chain Monte Carlo (MCMC) scheme. Geiger et al. [4] also proposed to use reversible-jump MCMC to infer geometrical and topological scenes properties as well as semantic activities from movable platforms. Further, Spehr et al. [5] propose a high-level scene understanding approach for real-time parking lot interpretation. Towards this goal Spehr et al. propose the application of compositional hierarchical models and approximate inference algorithms.

Following [5], we propose to use compositional hierarchical models for real-time multi-lane road recognition. In our approach, however, we are not limited to parking lot scenes. Instead, we present a novel hierarchical model for heterogeneous road topologies, including multi-lane roads with parallel and non-parallel lanes. Furthermore, we propose the application of depth-first message passing and part sharing for real-time inference and the fusion of multiple lane cues for reliable lane recognition.

III. Multi-Cue Sensor Evidence

In our approach, we rely on a monocular vision sensor which provides the visual input frames for our approach. Given this visual input, we employ two different feature detection approaches. First, lane marking features are extracted from the visual input using the symmetrical local threshold method. We choose this method since according to the evaluation conducted in [14] it gives the best result in the general case, i.e., it is relatively robust against clutter and local illumination changes. Further, a Sobel edge detector is used to gain road edge features in image regions where markings are missing, but road edges, such as curbstones are present. We choose this edge detector mainly due to its high sensitivity, which allows us to reduce the amount of false negatives. However, this also means that the edge detector is highly affected by outliers due to clutter and shadow, as shown in Fig. 1. In our approach, we directly address this issues using the weak spatial constraints, as detailed in Sec. IV. These constraints reduce the influence of outliers on the overall recognition process by assessing their spatial plausibility. Therefore, our framework enables us to use very sensitive detector and thus to cope with challenging scenarios with sparse visual cues.

(a) (b)

Fig. 1. Results of feature extraction in an urban scenario. (a) Lane marking detections (green). (b) Road edge detections (red). Road edge cues are used to detect roads in scenarios where lane markings are not reliable.

The two feature extraction approaches obtain a set of lane marking features $\boldsymbol{m} = \{\boldsymbol{m}_1, \ldots, \boldsymbol{m}_{N_m}\}$ and a set of road edge features $\boldsymbol{r} = \{\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{N_r}\}$. A lane marking feature $\boldsymbol{m}_i = (x_i, y_i, \vartheta_i)$ is defined by its location $(x_i, y_i) \in \mathbb{R}^2$ and orientation $\vartheta_i \in [0, 2\pi)$. Similarly, a road edge feature is defined as $\boldsymbol{r}_i = (x_i, y_i, \vartheta_i)$. These features constitute the observable random variables $\boldsymbol{r}$ and $\boldsymbol{m}$ of our CHM, which have corresponding hidden random variables $\boldsymbol{x}_i^f$ defined on the same 3d state space. As Fig. 2 shows, these hidden feature variables constitute the first level $\mathcal{L}_1$ of our CHM.

The dependency of lane marking observations and hidden feature variables is modeled by observation potentials

$$\phi_i(\boldsymbol{x}_i^f, \boldsymbol{m}) = \epsilon^0 \mathcal{N}_0(\boldsymbol{x}_i^f; 0, \boldsymbol{\Sigma}_0) + (1 - \epsilon^0) \sum_{k=1}^{N_m} \pi_{i,k} \mathcal{N}(\boldsymbol{x}_i^f; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_{i,k}), \quad (1)$$

and analogously for observation potentials $\phi_i(\boldsymbol{x}_i^f, \boldsymbol{r})$ accounting for road edge features. Here, $\boldsymbol{\Sigma}_{i,k} \in \mathbb{R}^{3\times3}$ is the covariance matrix of the k^{th} mixture component. Formally, this observation potential is a kernel density estimation of the true likelihood, which allows to model multi-modal density distributions by assigning a Gaussian kernel to each feature $\boldsymbol{\mu}_k$. $\pi_{i,k}$ denotes the probability of the association of the k-th feature with the i-th hidden variable. Further, we augment the observation potential by a zero mean, high-variance Gaussian outlier process $\mathcal{N}_0(\boldsymbol{x}_i; 0, \boldsymbol{\Sigma}_0)$ that allows us to account for clutter and occlusions.

IV. Compositional Hierarchical Model of a Multi-Lane Road

Our approach to road scene understanding is based on a tree-structured graphical model, which captures the way the joint distributions over random variables can be decomposed into a product of factors. Each of these factors only depends on a subset of the variables and thus allows for the development of efficient inference algorithms.In particular, we represent multi-lane roads in a Compositional Hierarchical Model (CHM) [5], [15]–[17] that is encoded by a pairwise Markov random field. In this CHM the root represents a full instantiated model of a multi-lane road, with all its properties (see Fig. 2), and the nodes on the lower levels represent a recursive decomposition of the root object into parts and sub-parts. This decomposition leads to a layered object representation with decreasing part complexity in direction to the leaves. Thus, our CHM not only encodes the dependencies between low-level evidence and high-level scene topologies, but divides the perception

problem into sub-problems that are easier to solve. In our approach, we assume a flat road surface and model the scene in the vehicle coordinate system. The vehicle coordinate system is located at the center of the rear axis and follows the common axis definition ($x=$forward, $y=$left, $\theta=$yaw angle).

More formally, our CHM is encoded by an undirected tree structured graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\mathcal{V}$ and edges $\mathcal{E}$. The nodes $\mathcal{V}$ correspond to three disjoint sets of variables $\mathcal{V} = \boldsymbol{x} \cup \boldsymbol{m} \cup \boldsymbol{r}$, where $\boldsymbol{x}$ denotes the set of hidden random variables $\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. Each hidden variable $\boldsymbol{x}_i \subseteq \boldsymbol{x}$ represents a part or a sub-part of the multi-lane road, which is represented by the root node of our graphical model (see Fig. 2) and is defined on a multidimensional continuous state space. Further, the observable variables $\boldsymbol{m}$ and $\boldsymbol{r}$ correspond to the lane marking features and the road edge features, respectively. The edges $\mathcal{E}$ between pairs of hidden variables define spatial constraints $\psi_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which encode the dependencies between neighboring hidden variables $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ by means of conditional spatial distributions. Edges between hidden and observable random variables encode observation potentials $\phi_i(\boldsymbol{x}_i, \boldsymbol{m})$ and $\phi_i(\boldsymbol{x}_i, \boldsymbol{r})$, as in Eq. 1.

Given the above definition and assuming that the observations of lane markings $\boldsymbol{m}$ and road edges $\boldsymbol{r}$ are independent given $\boldsymbol{x}$, the joint probability distribution factorizes as

$$-\log(p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N | \boldsymbol{m}, \boldsymbol{r})) = \log(Z) + \sum_{(i) \in \mathcal{I}_m} \Phi_i(\boldsymbol{x}_i, \boldsymbol{m}) + \sum_{(i) \in \mathcal{I}_r} \Phi_i(\boldsymbol{x}_i, \boldsymbol{r}) + \sum_{(i,j) \in \mathcal{E}} \Psi_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j), \quad (2)$$

where $\Phi_i(\cdot) = -\log(\phi_i(\cdot))$, $\Psi_{ij}(\cdot) = -\log(\psi_{ij}(\cdot))$ and $Z \in \mathbb{R}$ denotes the partition function that normalizes the probability distribution. $\mathcal{I}_m$ denotes the indexes of the set of cliques that are contained in $\boldsymbol{x} \cup \boldsymbol{m}$ and $\mathcal{I}_r$ the indexes of the cliques in $\boldsymbol{x} \cup \boldsymbol{r}$. The above factorization is also shown in Fig. 2.

In the following, we detail the different levels of our CHM as well as their their spatial dependencies.

A. Features and local driveable Areas

Recall that the leaves of our CHM comprise feature variables $\boldsymbol{x}_i^f = (x_i, y_i, \vartheta_i)$ with associated observations. Two of the observed features define a local driveable area referred to as patch, which form the second level $\mathcal{L}_2$ of our CHM, as shown in Fig. 3a. This figure illustrates that each patch is defined by a left and a right lane boundary feature.

Formally, patches are defined by a five-dimensional state vector $\boldsymbol{x}_j^p = (x_j, y_j, \vartheta_j, w_j, \nu_{l_p})$. Here, $(x_j, y_j) \in \mathbb{R}$ is the patch location, $\vartheta_j \in [0, 2\pi)$ its orientation angle, $w_j \in \mathbb{R}^+$ its width and $\nu_{l_p} \in \mathbb{R}^+$ its length (see Fig. 4a). Note that in our experiments the patch length ν_{l_p} was chosen as a constant design parameter.

The dependencies between feature variables $\boldsymbol{x}_i^f$ and patch variables $\boldsymbol{x}_i^p$ are modeled using weak spatial constraints

$$\psi_{ij}(\boldsymbol{x}_i^f, \boldsymbol{x}_j^p) = \mathcal{N}(\boldsymbol{x}_j^p; S_{ij}(\boldsymbol{x}_i^f), \boldsymbol{\Sigma}_{ij}), \quad (3)$$

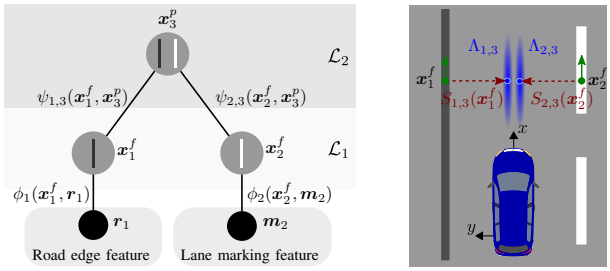


Fig. 3. CHM of a patch and illustration of the modeled spatial constraints. (a) A patch is decomposed into a left and a right lane boundary which are directly observable. (b) Illustration of the modeled spatial constraints, where spatial uncertainties are illustrated by showing 2D Gaussian distributions, where dark colors correspond to more likely locations.

where for a left feature the transformation function

$$S_{ij}(\boldsymbol{x}_i^f) = \begin{pmatrix} x_i \\ y_i \\ \vartheta_i \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{2}\sin(\vartheta_i) & 0 & 0\,0\,0 \\ 0 & -\frac{1}{2}\cos(\vartheta_i) & 0\,0\,0 \\ 0 & 0 & 1\,0\,0 \\ 0 & 0 & 0\,1\,0 \\ 0 & 0 & 0\,0\,1 \end{pmatrix} \begin{pmatrix} \nu_{w_p} \\ \nu_{w_p} \\ 0 \\ \nu_{w_p} \\ \nu_{l_p} \end{pmatrix} \quad (4)$$

returns the predicted location of variable $\boldsymbol{x}_j^p$ based on the expected lane width $\nu_{w_p} \in \mathbb{R}^+$ and the state of variable $\boldsymbol{x}_i^f$. The covariance matrix $\boldsymbol{\Sigma}_{ij} \in \mathbb{R}^{4\times4}$ is a design parameter that express uncertainties regarding the spatial dependencies, as illustrated in Fig. 3b.

We choose this patch definition due to its generality. In fact, patches can be composed from any lane and road cue that allows to predict the lane center (e.g., guardrails, delineators or bots'dots). This scalability is a key advantage of our framework, since it not only enables us to increase the reliability of our approach but to handle the enormous appearance diversity of lanes and roads [18], [19].

B. Local driveable Areas and Lanes

The next higher levels of our CHM comprise lanes of increasing length (see Fig. 2). As Fig. 4 illustrates, lanes are defined as a composition of N_P individual patches $\boldsymbol{x}_i^l = \{\boldsymbol{x}_1^p, \boldsymbol{x}_2^p, \ldots, \boldsymbol{x}_{N_P}^p, l_i^l\}$, i.e., lanes are represented by a polygonal path with piecewise constant orientation and width. Consequently, the length of a lane $l_i^l \in \mathbb{R}^+$ is defined as the sum of Euclidean distances between subsequent lane elements. A key development goal of the lane representation is to maintain a low complexity of the spatial constraints, which is crucial for a low computational complexity during inference. Particularly convenient are spatial constraints between lane and patch variables $\psi_{i,j}(\boldsymbol{x}_i^p, \boldsymbol{x}_j^l)$ since patches are expected to have the same spatial configuration as the adjacent lane element. Spatial constraints between pairs of lane variables $\psi_{i,j}(\boldsymbol{x}_i^l, \boldsymbol{x}_j^l)$, on the other hand, are used to predict the expected location of the subsequent lane element, using the transformation function

$$S_{ij}(\boldsymbol{x}_i^l) = \begin{pmatrix} x_i \\ y_i \\ \vartheta_i \\ w_i \end{pmatrix} + \begin{pmatrix} \cos(\vartheta_i) & 0 & 0\,0 \\ 0 & \sin(\vartheta_i) & 0\,0 \\ 0 & 0 & 1\,0 \\ 0 & 0 & 0\,1 \end{pmatrix} \begin{pmatrix} \nu_{l_p} \\ \nu_{l_p} \\ 0 \\ 0 \end{pmatrix}. \quad (5)$$

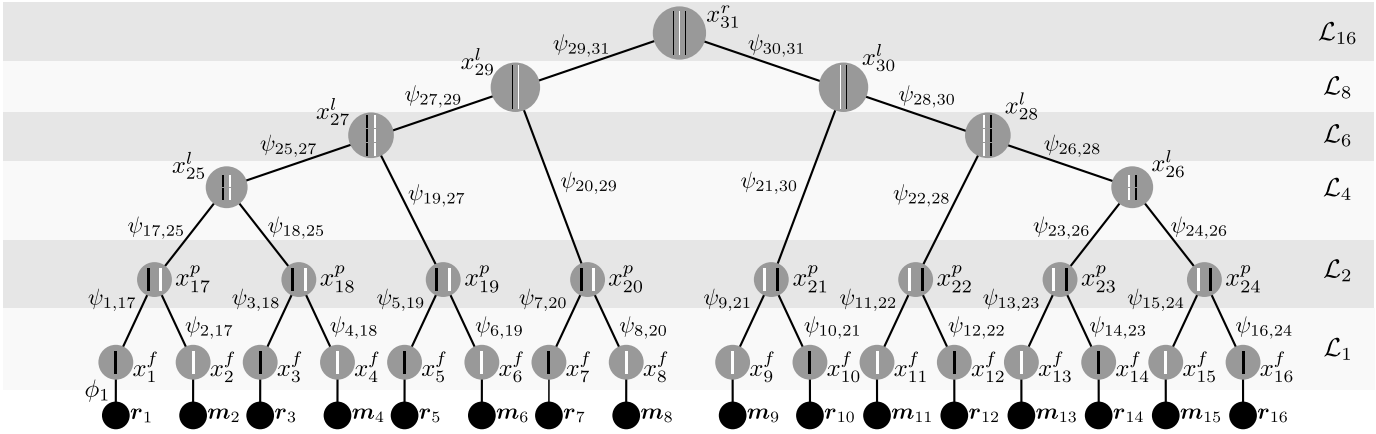
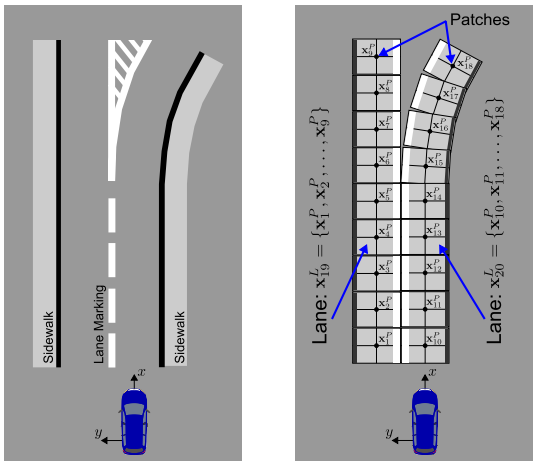


Fig. 2. CHM of a two lane road. This figure shows the factorization of the joint probability distribution in Eq. 2 using an undirected graphical model. Hidden random variables are depicted in grey and symbols illustrate their type, i.e., features, patches, lanes and multi-lane roads. Observable variables are shown in black and dependencies between random variables are highlighted using edges.



(a) (b)

Fig. 4. Patch-based lane representation. (a) Two lane road with splitting lanes. (b) Patches comprised to lanes.

Here, ν_{l_p} is the constant patch length defining the segmentation of the lane center line which allows us to control the complexity of inference.

Note that, in contrast to many state-of-the-art approaches, we do not impose hard constraints on the longitudinal lane geometry (e.g., clothoid or spline). Further, lanes are not restricted to a specific lateral model (e.g., parallel lanes or constant lane width). This great flexibility is a key benefit of our approach, since it makes our framework applicable to scenarios beyond highways and highway-like roads.

Yet, most ADAS applications involving vehicle control require a smooth lane representation and a small number of false detections. This commonly leads to the introduction of model assumptions on the lateral and longitudinal road topology [18], [19]. While in principal our lane representation could easily be extend to e.g., a clothoid model by introducing additional model parameters, it would limit its field of applicability. Therefore, instead of introducing specific lateral or longitudinal lane models, we consider such assumptions as the property of a specific road type, as detailed next.

C. Lanes and multi-lane Roads

As can be seen in Fig. 2, lanes are once again used as parts of more complex objects representing multi-lane roads. Since roads comprise all available information they are the output level of our approach.

A road $\boldsymbol{x}^r_i = \{\boldsymbol{x}^l_1, \boldsymbol{x}^l_2, \ldots, \boldsymbol{x}^l_{N_l}\}$ is composed of a finite number of N_l lanes and has a longitudinal model (e.g., clothoid, polyline or spline), defining its curvature and a lateral model defining its topology. The road topology, includes the number, position and width of lanes and the lane structure, i.e., parallel, splitting or merging.

An example of a CHM of a road is depicted in Fig. 2, showing a two-lane road with parallel lanes. In this case, the spatial constraints $\psi_{29,31}(\boldsymbol{x}^l_{29}, \boldsymbol{x}^r_{31})$ and $\psi_{28,31}(\boldsymbol{x}^l_{28}, \boldsymbol{x}^r_{31})$ model a parallel lane configuration. In a more complex model these spatial constraints can be used to encode dependencies between e.g., merging or splitting lanes. This example clarifies a key aspect of our approach. Namely, that our CHM can easily be adopted to new road topologies without altering the lower levels of our model. This flexibility not only allows to meet the heterogeneous demands of ADAS applications but to generalize our framework to scenarios with diverse topologies, as detailed in Sec. VI.

D. Periodic Variables

One of the challenges in modeling the spatial constraints is that the random variables representing angles $\vartheta \in [0, 2\pi)$ do not possess a natural origin. To overcome issues regarding the choice of origin, we adapt von Mises Fisher distributions [20]. The von Mises Fisher distribution is a convenient choice, since it can be derived from a bivariate Euclidean Gaussian distribution with mean $(\cos\vartheta, \sin\vartheta)$ [21], [22] and thus can easily be added to the Gaussian model of our spatial constraints.

V. Inference of a single Road Topology

In our framework the task of lane and road perception is equivalent to computing the marginal posterior distribution or

belief $b_i(\boldsymbol{x}_i)$ of all or a subset of the hidden variables in our graphical model. A task which can efficiently be performed using Belief Propagation (BP). In BP the belief over a variable $\boldsymbol{x}_i$ is computed by combining all incoming messages at node i with the local observation potential as

$$b_i(\boldsymbol{x}_i) \propto \phi_i(\boldsymbol{x}_i, \boldsymbol{m})\phi_i(\boldsymbol{x}_i, \boldsymbol{r}) \prod_{c\in\Xi(i)} m_{c,i}(\boldsymbol{x}_i) \prod_{p\in\Gamma(i)} m_{p,i}(\boldsymbol{x}_i)$$
$$\propto b_i^-(\boldsymbol{x}_i) \prod_{p\in\Gamma(i)} m_{p,i}(\boldsymbol{x}_i), \quad (6)$$

where the two products contain messages from the children $\Xi(i)$ and the parents $\Gamma(i)$ of node i, respectively. In BP, messages $m_{i,j}(\boldsymbol{x}_j)$ passed from node i to j predict which state node j should be in and can be computed recursively. Further, Eq. 6 defines the bottom-up belief state $b_i^-(\boldsymbol{x}_i)$ that is an intermediate processing results of the inference algorithms detailed in the following.

If the belief state of all variables is Gaussian the belief can be computed exactly using Eq. 6 [23]. However, in our case this is not the case due to the noisy and multi-modal sensory evidence and thus standard BP is not applicable. To overcome this issues, we perform inference using Nonparametric Belief Propagation (NBP) [24], [25], which is a generalization of the particle filter [26] for approximate inference in arbitrary graphs.

In NBP the belief $b_i(\boldsymbol{x}_i)$ is approximated by a set of L importance weighted samples $\{(\boldsymbol{s}_i^{(k)}, \pi_i^{(k)})\}_{k=1}^L$ as

$$b_i(\boldsymbol{x}_i) = \sum_{k=1}^{L} \pi_i^{(k)} \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{s}_i^{(k)}, \boldsymbol{\Lambda}_i). \quad (7)$$

Each of these samples $\boldsymbol{s}_i^{(k)}$ represents a hypothesis for the spatial configuration of variable $\boldsymbol{x}_i$ and is drawn from the product distribution

$$\boldsymbol{s}_i^{(k)} \sim \prod_{c\in\Xi(i)} m_{c,i}(\boldsymbol{x}_i) \prod_{p\in\Gamma(i)} m_{p,i}(\boldsymbol{x}_i) \quad (8)$$

using the efficient nearest neighbor product sampling method proposed in [27]. The corresponding importance weight $\pi_i^{(k)}$ of a sample $\boldsymbol{s}_i^{(k)}$ is then defined as

$$\pi_i^{(k)} \propto \phi_i(\boldsymbol{s}_i^{(k)}, \boldsymbol{m})\phi_i(\boldsymbol{s}_i^{(k)}, \boldsymbol{r}) \quad (9)$$

and represents the spatial plausibility of the hypothesis $\boldsymbol{s}_i^{(k)}$. Finally, the computational efficient rule of thumb [28] is used to construct a kernel density estimation from the raw sample set by assigning a Gaussian smoothing kernel with bandwidth $\boldsymbol{\Lambda}_i$ to each sample (see Eq. 7).

A convenient property of our framework is that evidence is exclusively injected into our model via the leave nodes, which allows us to perform the belief update of Eq. 7 in two stages. First, we compute the bottom-up belief state $b_i^-(\boldsymbol{x}_i)$ by passing messages from the observable leaves to the root. Second, we pass messages down from the root to the leaves to compute the belief $b_i(\boldsymbol{x}_i)$. These two phases have different goals. While the aim of the bottom-up phase is the fast generation of high-level hypotheses (e.g., lanes and roads), the top-down phase ensures the overall consistency of parts and their subparts. We generally follow the bottom-up/top-down message passing schedule with some adaptions to facilitate this standard message passing approach.

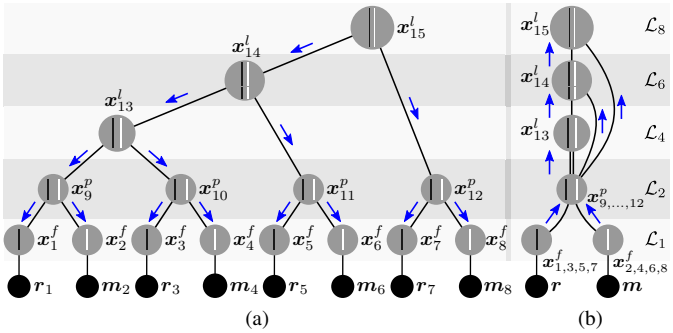


Fig. 5. Message passing using part-sharing. (a) During the bottom-up phase messages (blue) are passed from the leaves to the root using the sharing structure. During this phase the belief is shared between nodes on $\mathcal{L}_1$ and $\mathcal{L}_2$ to avoid redundant computations. (b) During the top-down phase, message passing is based on the structure of our CHM, since each node receives different contextual information from its parent.

As detailed above bottom-up/top-down message passing begins with the computation of the bottom-up belief of all feature variables $b_i^-(\boldsymbol{x}_i^f)$. During this step feature variables are conditioned on either the feature set $\boldsymbol{m}$ or the feature set $\boldsymbol{r}$. Consequently, all feature variables in $\boldsymbol{x}^f \cup \boldsymbol{m}$ and in $\boldsymbol{x}^f \cup \boldsymbol{r}$ comprise the same bottom-up belief and thus many messages send from the feature variables on $\mathcal{L}_1$ to the patch variables on $\mathcal{L}_2$ contain the same belief estimate. This means, we have to compute the same message product for each patch node in the CHM, leading to unnecessary computational complexity. To avoid such redundant computations, we adapt the part-sharing technique, which has been proposed in the field of vision based multi-view, multi-object detection [15], [27].

A. Part-Sharing

The fundamental idea of part-sharing is to merge those nodes during the bottom-up phase, which receive the same messages from their children. In our case, we can combine the patch nodes as well as the nodes in their sub-trees, as depicted in Fig. 5b. This figure shows the resulting sharing structure that includes three sharing-nodes $\boldsymbol{x}^f_{1,3,5,7} = \{\boldsymbol{x}^f_1, \boldsymbol{x}^f_3, \boldsymbol{x}^f_5, \boldsymbol{x}^f_7\}$, $\boldsymbol{x}^f_{2,4,6,8} = \{\boldsymbol{x}^f_2, \boldsymbol{x}^f_4, \boldsymbol{x}^f_6, \boldsymbol{x}^f_8\}$ and $\boldsymbol{x}^p_{9,\dots,12} = \{\boldsymbol{x}^p_9, \dots, \boldsymbol{x}^p_{12}\}$ clarifying that the bottom-up belief is shared between the combined hidden variables.

Using part-sharing, bottom-up inference is based on the sharing structure, where the bottom-up belief state of each node is only calculated once and then shared between its parents (see Fig. 5b). This procedure not only avoids redundant computations but ensures that only one object instance has to be memorized. After the belief of the root node $b_{15}(\boldsymbol{x}^l_{15})$ is computed, we begin with top-down message passing. Since, during this phase the patch nodes receive different messages from their parents, top-down message passing is performed in the graphical model illustrated in Fig. 5a. Before starting the top-down phase we decompose the sharing nodes into the comprised nodes by assigning the bottom-up belief $b_i^-(\boldsymbol{x}_i)$ of

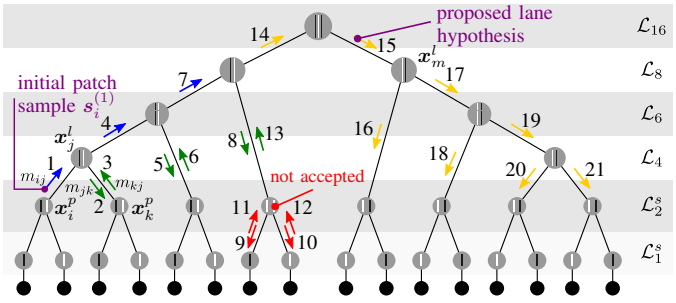


Fig. 6. Multi-lane road detection using depth-first message passing. Based on an initially selected patch sample lane hypotheses are generated (blue). An expectation-based step ensure the consistency of high-level hypotheses and local evidence (green). Further, road hypotheses can be evaluated without explicit bottom-up message passing (orange). Note that the bottom-up belief state of nodes on level $\mathcal{L}_1^s$ and $\mathcal{L}_2^s$ is computed using part-sharing.

the sharing-nodes to all aggregated nodes. Finally, the belief $b_i(\boldsymbol{x}_i)$ over each variable $\boldsymbol{x}_i$ is computed by combining its bottom-up belief with the incoming messages from its parents $m_{j,i}(\boldsymbol{x}_i)$ with $j \in \Gamma(i)$, as defined in Eq. 6.

B. Depth-First Message Passing

Bottom-up message passing in the sharing structure amounts in processing each level of our model one by one (see Fig. 5b). Intuitively, this can be thought of as a breadth-first search in the hypotheses space, since on each level all possible hypotheses are computed. The advantage of this approach is that it leads to a good approximation of the density distributions on all levels. The drawback, however, is that with each level the hypotheses space growth exponential, and therefore computations are only tractable for a small number of levels. Further, in our application we are interested in the fast detection of high-level hypotheses (i.e., lanes and roads) not in an exhaustive search for low-level hypotheses.

Therefore, we proposed a sequential message passing schedule [7], [27], which is inspired by the depth-first traversal for arbitrary tree structured graphs. The fundamental idea of depth-first message passing is to perform bottom-up message passing in several sequential sweeps. At the beginning of each sweeps we select a single patch sample that is likely to be part of a valid high-level hypotheses and propagate it through the graph. Since in this approach a single sample is propagated through the graph, we only compute a subset of the possible hypotheses on each level, resulting in a computational efficient depth-first search for valid hypotheses in the hypotheses space.

In the case of lane and road detection it is convenient to start depth-first message passing from patches close to the vehicle, since they are located in areas of low uncertainty. After selecting a patch sample, we can then extend the lane hypotheses from the vehicle into areas of higher uncertainty. Towards this goal, in each sweep a single patch samples $(\boldsymbol{s}_i^{(k)}, \pi_i^{(k)})$ is selected from the nonparametric density $b_i^-(\boldsymbol{x}_i^p)$ according to its weight $\boldsymbol{s}_i^{(k)} \sim \pi_i^{(k)}$. Subsequently, the selected sample is propagated through the graphical model as depicted in Fig. 6. In this example, message passing is initiated at variable $\boldsymbol{x}_i^p$ by selecting a single sample $\boldsymbol{s}_{ij}^{(1)}$ according to its weight. Using this sample the message $m_{i,j}(\boldsymbol{x}_j^l)$ is constructed, predicting the configuration of $\boldsymbol{x}_j^l$. This message can now be used to update the belief $b_j^-(\boldsymbol{x}_j^l)$ by computing the product of all incoming messages from the children of $\boldsymbol{x}_j^l$ (see Eq. 6). Towards this goal, nearest neighbor product sampling [27] searches for samples $\boldsymbol{s}_{kj}^{(q_{nn})}$ in the incoming messages at $\boldsymbol{x}_j^l$, which are similar to $\boldsymbol{s}_{ij}^{(1)}$. These samples are accepted according to the acceptance rate [27]

$$A(\boldsymbol{s}_{kj}^{(q_{nn})}) = \exp(-\frac{1}{2}(\boldsymbol{s}_{ij}^{(1)} - \boldsymbol{s}_{kj}^{(q_{nn})})^T \boldsymbol{\Sigma}_{ij}^{-1}(\boldsymbol{s}_{ij}^{(1)} - \boldsymbol{s}_{kj}^{(q_{nn})})), \tag{10}$$

where q_{nn} is the index of the nearest neighbor of the sample $\boldsymbol{s}_{ij}^{(1)}$ and $\boldsymbol{\Sigma}_{ij}$ is the covariance matrix of the spatial constraints $\psi_{ij}(\boldsymbol{x}_i^p, \boldsymbol{x}_j^l)$.

However, in depth-first message passing it is likely that no sample is accepted, since it is not guaranteed that variable $\boldsymbol{x}_j^l$ received messages from all its children. Therefore, if no sample is accepted a top-down bottom-up sweep is initiated that searches for evidence supporting the predicted configuration of $\boldsymbol{x}_j^l$. Intuitively, this sweep can be thought of as an aligning process that ensures the consistency of the lane hypotheses and the low-level observations.

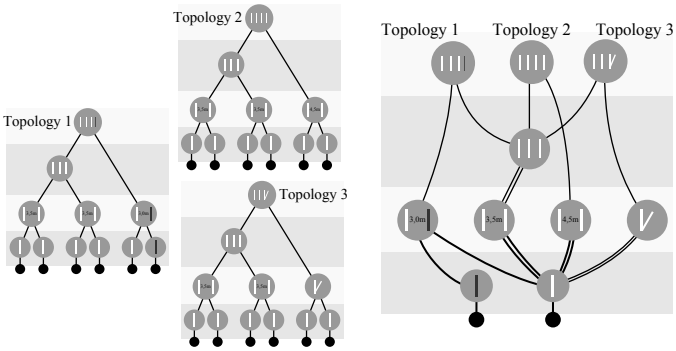
This alignment process is illustrated in Fig. 6 by showing green messages. Here, we begin by sending a message $m_{j,k}(\boldsymbol{x}_k^p)$ from node $\boldsymbol{x}_j^l$ to node $\boldsymbol{x}_k^p$, again containing a single sample $\boldsymbol{s}_{jk}^{(1)}$. This single sample is used in the nearest neighbor product sampling to search for samples in the messages $m_{h,k}(\boldsymbol{x}_k^p)$ with $h \in \Xi(k)$ send to $\boldsymbol{x}_k^p$ from its children. As before, samples $\boldsymbol{s}_{hk}^{(q_{nn})}$ are accepted according to Eq. 10, to decide whether a sample is supported by an observation or if it corresponds to the outlier process (see Eq. 1).

If for all messages $m_{h,k}(\boldsymbol{x}_k^p)$ a sample $\boldsymbol{s}_{hk}^{(q_{nn})}$ is accepted, we calculate the product of the incoming and the accepted sample, which amounts in calculating the product of two Gaussian distributions. If for some of the incoming messages no sample is accepted, we multiply the contained samples with the high-variance Gaussian outlier process. In the following, the product result is used to send a single sample message $m_{k,j}(\boldsymbol{x}_j^l)$ back to $\boldsymbol{x}_j^l$. Finally, the belief update at node $\boldsymbol{x}_j^l$ is performed by computing the product of $m_{i,j}(\boldsymbol{x}_j^l)$ and $m_{k,j}(\boldsymbol{x}_j^l)$. The above procedure is repeated for each level until the root node is reached. Then the next sweep is started.

An attractive property of depth-first message passing is that based on a partly estimated road hypothesis we can propose lane hypotheses, as shown in Fig. 6. This figure shows, that given the road geometry prediction of the left lane, we can propose a right lane hypothesis and evaluate its plausibility. This ability is a key advantage over many nonparametric lane detection approaches [9], [29], since it enables us to distribute lane samples based on the a priori knowledge of the spatial constraints and thus to detect lanes in areas of low belief.

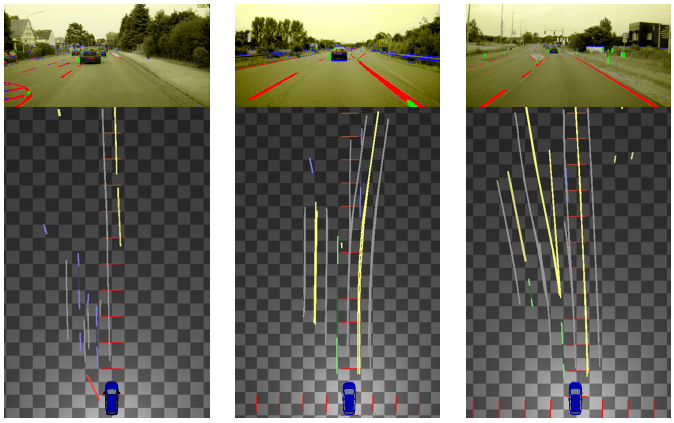
VI. Representing and Inferring Heterogeneous Road Topologies

So fare, we mainly focused on estimating a single road topology. However, a key challenge in lane and road perception is to handle the enormous topological diversity of traffic



(a) set of hierarchies (b) sharing structure

Fig. 7. Representation of heterogeneous road topologies using sets of CHMs and part-sharing. (a) Set of CHMs, where each CHM models a specific road topology. (b) Sharing structure representing the set of CHM by means of shared parts. Performing bottom-up inference in the sharing structure is not only computational efficient but allows to simultaneously estimate multiple road topologies. Note that the length of lanes and roads is not depicted.



(a) (b) (c)

Fig. 8. Results of lane recognition for (a) an urban road with sparse lane markings (b) a highway split and (c) an urban road with non-parallel lanes. The top shows the detected lane-marking features (red). The bottom shows the projected lane-marking detection into the vehicle reference frame (yellow, purple, green) and results of lane recognition (grey).

scenarios [18], [19]. While, in general, the spatial constraints of our CHM allow for a certain degree of spatial variation, as soon as the variations become too large or different road topologies have to be represented new separate CHMs have to be specified.

Therefore, a multi-scenario model is represented by a set of N_h CHMs $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_1, \ldots, \mathcal{G}_{N_h}\}$, where each CHM is a joint probability distribution defined over a hierarchical graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)$. An illustrative example of a set of CHM is depicted in Fig. 7, showing three CHMs each representing a different road topology. At the first glance, a separate CHM for each road topology seams unattractive, since it leads to an exponential growth of instances. However, it is reasonable to expect that many CHMs contain similar parts. In fact, the three CHMs in Fig. 7a comprise several common parts and thus we can represent them by means of shared parts, as depicted in Fig. 7b. The depicted sharing structure clarifies the similarities between the different road topologies, which e.g., share the two-lane road. Note that, sharing the two-lane road also requires us to share its sub-tree.

The key benefit of part-sharing is that during bottom-up message passing, we have to compute the bottom-up belief over all common parts only once and then share it between the associated parents. As before, bottom-up message passing is performed in the sharing structure, while top-down message passing is performed in the individual CHMs, corresponding to the root nodes of the sharing structure. Part-sharing is one of the key aspects in our framework, since it not only allows us to perform efficient bottom-up inference in a single CHM but to share complex high-level objects between different road topologies. In contrast to many recent lane and road perception approaches, which only aim to detect a single road topology this makes our approach applicable to various topologies.

VII. Experimental Results

In this section, we applying our framework to challenging real world scenarios and present a set of quantitative and qualitative results. These experiments aim to demonstrate that the proposed approach: (1) can benefit from the weak constraints incorporated on each level of the CHM to increase the reliability of the detection results, (2) can incorporate multiple low-level cues and thus increase recognition performance and (3) can perform lane and road perception in real-time using depth-first message passing [7].

To evaluate the performance of our approach, we tested it in highway, rural and urban scenarios (see Fig. 8) using our C++ implementation. The database used for our evaluation comprises several thousand individual images of urban, rural and highway scenarios and was captured during low traffic density, i.e., the road is completely visible. Clearly manual labeling of such a large database is impracticable. Therefore, we obtain ground-truth information from a high-accuracy map database [30], which contains an exact topological lane description. Given this map database, we perform the labeling process in three steps. First, we extract relevant data (e.g., patches, lanes or roads) from the map database and align them to our results using a high-accuracy DGPS and IMU system. Second, we compute an 2d overlap ratio between ground truth data and our results to judge if our result are positive or negative examples. In particular, we consider our results to be positive results if the overlap ratio exceeds 80%, i.e., our results have an geometrical consistency with the ground truth of over 80%. This soft labeling process allows to handle inaccuracies in the map database and the DGPS+IMU system as well as to access the accuracy and reliability of our approach. Finally, the importance weights given in Eq. 9 are used to draw precision recall curves.

The used CHM has a similar structure as depicted in Fig. 2. However, to cover the detection range of the used vision sensor, we extend the CHM by introducing random variables representing lanes composed of up to 40 patches. The root nodes of our graphical model represent roads with two or three lanes. For all tests we set the outlier probability ϵ^0 to 20% of the total likelihood [25], [26]. The a priori width and the length of the patches were set to $\nu_{w_p} = 3.5\,\text{m}$ and a length $\nu_{w_l} = 2.0\,\text{m}$.

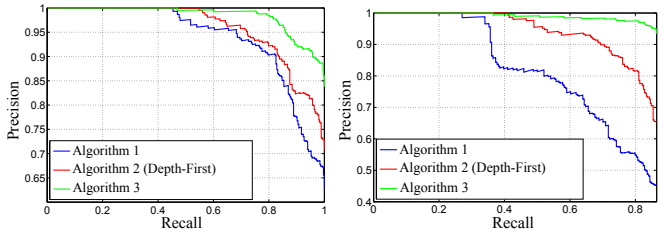


Fig. 9. Precision-Recall curves for highway scenarios (a) and for rural scenarios (b). Results are computed using different algorithms and evaluated against the ground truth. For details see text.

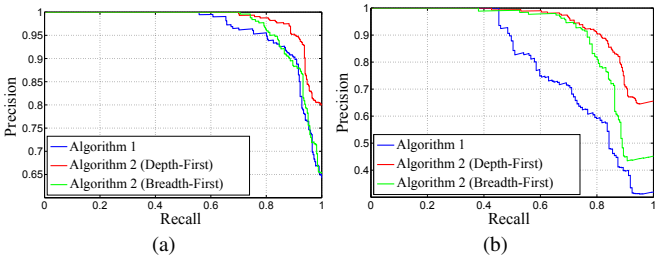


Fig. 10. Results of lane detection for highway (a) and for rural scenarios (b) using depth-first message passing (red) and breadth-first message passing (green). In both scenarios depth-first message passing (25 samples) shows more promising results than breadth-first message passing (150 samples).

A. Lane and Road Recognition

A key benefit of the proposed framework is that each level of the CHM incorporates geometrical and topological a priori knowledge comprised in the spatial constraints. We evaluate the importance of using such knowledge during road recognition by evaluating the belief $b_i(\boldsymbol{x}_i^p)$ over the patch locations at different stages of message passing. Towards this goal, we introduce three message passing algorithms.

- Alg. 1: We compute the bottom-up belief $b_i^-(\boldsymbol{x}_i^p)$ over the patch nodes by fusing messages received from their children $\boldsymbol{x}_j^f$ with $j \in \Xi(i)$.
- Alg. 2: We compute the bottom-up belief of both the lanes $b_i^-(\boldsymbol{x}_i^l)$ and patches $b_i^-(\boldsymbol{x}_i^p)$. Then, we compute the belief of the patches $b_i^{\backslash r}(\boldsymbol{x}_i^p)$ by propagating messages down from the lanes to the patches.
- Alg. 3: We compute the belief $b_i(\boldsymbol{x}_i^p)$ by performing a complete bottom-up/top-down sweep.

Here, Alg. 1 incorporates only the piecewise parallel lane assumption in the recognition process, while Alg. 2 and 3 introduce longitudinal and later lane models, respectively.

As can be seen in Fig. 9, the recognition performance of our model increases drastically, as we incorporate contextual information. This can be explained by the fact that patch nodes perform inference over a relatively small area. Accordingly, they strongly rely on the presence of local, visual evidence. This means e.g., missing, occluded or damaged lane cues have a significant impact on the recognition performance. Lanes, on the other hand, are based on a set of patches and thus combine sensory evidence from a larger area (e.g., about 200 patches for an average highway lane). Hence, the recognition performance is not as affected by missing local evidence as the one of patches. Furthermore, it can be seen in Fig. 9, that incorporating longitudinal and lateral model assumptions further improves the recognition performance of our approach, as it ensures the overall compatibility of objects and object parts.

Qualitative results for particular challenging scenarios are depicted in Fig. 8. It can be seen that our approach shows promising results in situations with sparse feature sets, lane splits and non-parallel lane structure, where conventional lane tracks are incapable of providing a full solution. Note that, hypotheses outside of the lane markings are supported by features on one side, and by the outlier process on the other side. As a result, weight is relatively low compared to hypotheses supported by two features. During the computation of road hypotheses it is less likely that these hypotheses are drawn from the message product (see Eq. 8) and thus they are often not present on the road level.

B. Depth-First Message Passing

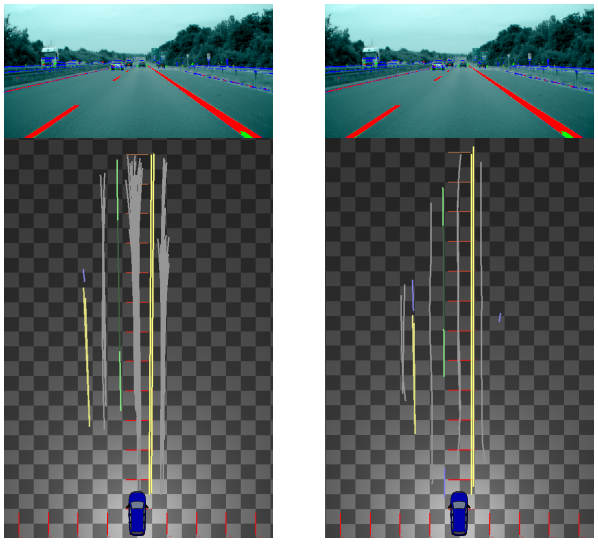
A key aspects of the proposed framework is depth-first message passing for lane detection [7]. We expect that depth-first message passing requires significantly lower running time as breadth-first message passing and thus to apply our approach in real-time. To test this hypothesis, we perform lane and road detection by applying both depth-first message passing and breadth-first message passing to the highway and rural scenarios of our dataset. To avoid an exponential growth of the lane-level hypotheses using breadth-first message passing, we introduce a resampling step after performing the belief update on the lane variables [24], [31]. This is used to limit the number of lane sample to 150.

It can be seen in Fig. 10 that depth-first message passing out-performs standard breadth-first message passing over the complete range of confidence, while using a significantly reduced sample set of only 25 samples. The reason for this major improvement is that by applying depth-first message passing, we first propagate those low-level hypotheses, which are likely to be part of valid high-level hypotheses. Consequently, during messages passing, we have to propagate less invalid hypotheses than using breadth-first message passing.

Qualitative results of applying both breadth-first and depth-first message passing are depicted in Fig. 11, showing the large amount of hypotheses computed during breadth-first message passing and the few likely hypotheses computed during depth-first message passing.

C. Runtime

In order to evaluate the runtime, we perform two experiments. First, we evaluate the runtime for highway scenarios using the lane marking detector, as they are the common target scenario for current ADAS. Furthermore, we access the runtime for complex multi-lane urban scenarios using both feature detectors. The results are summarized in Tab. I, showing that while in the more complex urban scenarios the runtime increases, patches, lanes and roads can still be computed in real-time (i.e., between two subsequent frames of the used vision sensors (25 fps)). Thereby, the results indicate that depth-first message passing is a key aspect in



(a) (b)

Fig. 11. Qualitative results of lane geometry estimation. (Top) Detected lane-marking features (red). (a) Results using breadth-first message passing. (b) Results using depth-first message passing.

Detection	Patches	Lanes (BF)	Lanes (DF)	Roads
Highway (ms)	1.47	75.29	4.41	5.38
Urban (m)	2.47	167.32	4.92	7.63

TABLE I
RUNTIME FOR HIGHWAYS USING LANE MARKING CUES AND FOR URBAN MULTI-LANE ROADS USING LANE MARKING AND ROAD EDGE DETECTORS.

archiving real-time performance, since it significantly reduces the computational complexity for lane detection.

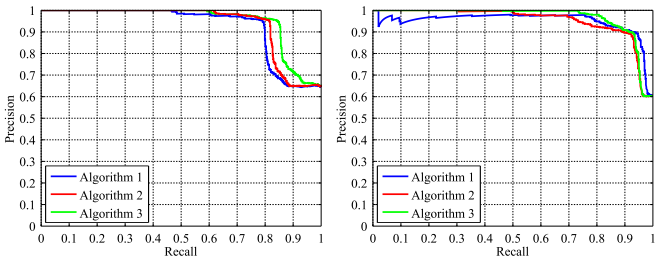
D. Multi-Cue Lane and Road Perception

A key benefit of our hierarchical framework is that it allows for the incorporation of multiple lane and road boundary cues and fuses them in an intelligent way. Particularly, in semi- or unstructured urban environments using multiple cues is expected to lead to an increased recognition performance. To verify this hypotheses we apply our hierarchical framework to the urban scenarios of our database. In this experiment, the recognition performance achieved using only lane marking cues is compared to the results obtained using both lane marking and the road edge cues.

The results of these experiments are depicted in Fig. 12, showing that as expected the additional usage of the road edge cues improves the recognition performance, since in many urban scenarios lane markings are not reliable. It can be seen that using both cues, we can obtain a precision of about 90% up to a recall of $90-95\%$, while precision drops drastically for a recall higher then $80-85\%$ using only lane marking cues. However, using multiple low-level cues also leads to additional computational complexity. In fact, the average computational time for the single cue setup is 21.0 ms, while for the multi-cue setup processing requires 23.3 ms.

VIII. CONCLUSION

We have presented a novel compositional hierarchical framework for multi-lane road recognition. Based on simple visual cues, our approach allows to reliably infer the topology of traffic scenes. Thereby, our road model is generic and compositional in the sense that we do not impose any hard constraints on the lane geometry as imposed by e.g., clothoids or splines. Instead, our prior expectations on the lane geometry are expressed through weak probabilistic constraints and we assemble lanes from a large number of lane patches. Furthermore, we introduced a new depth-first message passing algorithm for road recognition which in combination with part sharing allows to apply our approach in real-time. Finally, we proposed to use sets of hierarchies to represent heterogeneous road topologies, which allows to benefit from their similarities for real-time inference.



(a) (b)

Fig. 12. Results of multi-cue urban lane detection. (a) When relying on lane marking features the precision drops rapidly for a recall higher then $80-85\%$. (b) Using both lane marking and road edge features allows to obtain a precision higher then 90% up to a recall of $90-95\%$. In (b) the precision decreases slowly for a recall between $70-95\%$, since the second detector also causes additional false positives.

The proposed work is only one part of a next-generation environment understanding. In the future, we wish to extend our framework by including further modules, such as vehicle or pedestrian detections and navigation maps. These additional input sources can be used as observations on the different levels of our hierarchical model and as in [32] are expected to increase the performance in scenarios, where visual lane and road cues are occluded or not reliable.

ACKNOWLEDGMENT

This work was supported in part by the Bundesminesterium für Wirtschaft und Technologie, in the UR:BAN project. The authors thank the Institut für Robotik und Prozessinformatik at TU Braunschweig, headed by Prof. F. Wahl for supporting the development of the hierarchical models in an early stage.

REFERENCES

[1] E. D. Dickmanns and B. D. Mysliwetz, "Recursive 3-d road and relative ego-state recognition," *Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 199–213, 1992.
[2] H. Loose, U. Franke, and C. Stiller, "Kalman particle filter for lane recognition on rural roads," in *Intell. Veh. Symp.* IEEE, 2009, pp. 60–65.
[3] N. Apostoloff and A. Zelinsky, "Robust vision based lane tracking using multiple cues and particle filtering," in *Intell. Veh. Symp.* IEEE, 2003, pp. 558–563.
[4] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3d traffic scene understanding from movable platforms," *Trans. Pattern Anal. Mach. Intell.*, 2014.
[5] J. Spehr, D. Rosebrock, D. Mossau, R. Auer, S. Brosig, and F. Wahl, "Hierarchical scene understanding for intelligent vehicles," in *Intell. Veh. Symp.* IEEE, 2011, pp. 1142–1147.

[6] C. Wojek, S. Roth, K. Schindler, and B. Schiele, "Monocular 3d scene modeling and inference: Understanding multi-object traffic scenes," *European Conf. on Comput. Vision*, pp. 467–481, 2010.
[7] D. Töpfer, J. Spehr, J. Effertz, and C. Stiller, "Efficient scene understanding for intelligent vehicles using a part-based road representation," in *Internat. Conf. Intell. Transp. Syst.* IEEE, 2013, pp. 65–70.
[8] B. Southall and C. Taylor, "Stochastic road shape estimation," in *Internat. Conf. on Comput. Vision*, 2001, pp. 205–212.
[9] R. Danescu and S. Nedevschi, "Probabilistic lane tracking in difficult road scenarios using stereovision," *Intell. Veh. Symp.*, vol. 10, no. 2, pp. 272–282, 2009.
[10] C. Wojek and B. Schiele, "A dynamic conditional random field model for joint labeling of object and scene classes," *European Conf. on Comput. Vision*, pp. 733–747, 2008.
[11] P. Sturgess, K. Alahari, L. Ladicky, and P. Torr, "Combining appearance and structure from motion features for road scene understanding," 2009.
[12] X. Wang, X. Ma, and W. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 539–555, 2009.
[13] D. Kuettel, M. Breitenstein, L. Van Gool, and V. Ferrari, "What's going on? discovering spatio-temporal dependencies in dynamic scenes," in *Conf. on Comput. Vision and Pattern Recog.* IEEE, 2010, pp. 1951–1958.
[14] T. Veit, J. Tarel, P. Nicolle, and P. Charbonnier, "Evaluation of road marking feature extraction," in *Internat. Conf. Intell. Transp. Syst.* IEEE, 2008, pp. 174–181.
[15] L. L. Zhu, Y. Chen, A. Torralba, W. Freeman, and A. Yuille, "Part and appearance sharing: Recursive compositional models for multi-view multi-object detection," *Conf. on Comput. Vision and Pattern Recog.*, pp. 1919–1926, 2010.
[16] A. Torralba, K. P. Murphy, and W. Freeman, "Using the forest to see the trees: exploiting context for visual object detection and localization," *Communications of the ACM*, vol. 53, no. 3, pp. 107–114, 2010.
[17] K. Murphy, A. Torralba, and W. Freeman, "Using the forest to see the trees: a graphical model relating features, objects and scenes," *Adv. in Neural Inf. Process. Syst.*, vol. 16, 2003.
[18] A. B. Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine Vision and Applications*, pp. 1–19, 2012.
[19] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation," *Intell. Veh. Symp.*, vol. 7, no. 1, pp. 20–37, 2006.
[20] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the unit hypersphere using von mises-fisher distributions," in *Journal of Machine Learning Research*, 2005, pp. 1345–1382.
[21] L. Sigal, S. Bhatia, S. Roth, M. Black, and M. Isard, "Tracking loose-limbed people," in *Conf. on Comput. Vision and Pattern Recog.*, vol. 1. IEEE, june-2 july 2004, pp. I–421 – I–428 Vol.1.
[22] E. Sudderth, M. Mandel, W. Freeman, and A. Willsky, "Distributed occlusion reasoning for tracking with nonparametric belief propagation," *Adv. in Neural Inf. Process. Syst.*, vol. 17, pp. 1369–1376, 2004.
[23] Y. Weiss and W. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural Computation*, vol. 13, no. 10, pp. 2173–2200, 2001.
[24] M. Isard, "Pampas: Real-valued graphical models for computer vision," in *Conf. on Comput. Vision and Pattern Recog.*, vol. 1. IEEE, 2003, pp. I–613.
[25] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Conf. on Comput. Vision and Pattern Recog.*, 2003, pp. 605–612.
[26] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," *Internat. J. of Comput. Vision*, vol. 29, no. 1, pp. 5–28, 1998.
[27] J. Spehr, "On hierarchical models for visual recognition and learning of objects, scenes, and activities," Ph.D. dissertation, Technical University Braunschweig, 2013.
[28] B. W. Silverman, *Density estimation for statistics and data analysis*. Chapman & Hall, 1986, vol. 26.
[29] U. Franke, H. Loose, and C. Knoppel, "Lane recognition on country roads," in *Intell. Veh. Symp.* IEEE, 2007, pp. 99–104.
[30] K. Homeier and L. Wolf, "Roadgraph: High level sensor data fusion between objects and street network," in *Internat. Conf. Intell. Transp. Syst.* IEEE, 2011, pp. 1380–1385.
[31] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, MIT, 2006.
[32] S. Sivaraman and M. M. Trivedi, "Integrated lane and vehicle detection, localization, and tracking: A synergistic approach," *Trans. Intell. Transp. Syst.*, vol. 14, no. 2, 2013.

Daniel Töpfer received his diploma degree in mechatronics at TU-Braunschweig, Germany in 2010. He was a Ph.D. candidate at the Volkswagen Group Research under the supervision of Prof. Christoph Stiller from the Karlsruhe Institute of Technology from 2010 to 2014. Currently he is with the Volkswagen Group Research, where he works on safety applications for intelligent vehicles. His research is centered on machine learning, environment perception, control engineering and path planing.

Jens Spehr received the diploma degree in electrical engineering at TU-Braunschweig, Germany, in 2006. He was a research assistant at Institut für Robotik und Prozessinformatik at TU-Braunschweig, from which he obtained his Ph.D. degree in 2013. He is currently with the Volkswagen Group Research, where he works on scene understanding for intelligent vehicles. His research interests include computational models of vision, machine learning, and artificial intelligence.

Jan Effertz received the diploma degree in electrical engineering at TU-Braunschweig, Germany, in 2004. He was a research assistant at Institute of Control Engineering, TU-Braunschweig, from which he obtained his Ph.D. degree in 2009. He was responsible for the activities in sensor and sensor fusion systems for advanced driver assistance systems at Volkswagen Group Research from 2008 up to 2013. Currently he is responsible for the functional development of air conditioning control systems at Volkswagen passenger car development.

Christoph Stiller (S93-M95-SM99) studied Electrical Engineering in Aachen, Germany and Trondheim, Norway, and received the Diploma degree from TU Aachen in 1988. In 1988 he became a Scientific Assistant at TU Aachen. After completion of his Dr.-Ing. degree in 1994 he worked at INRS-Telecommunications in Montreal, Canada as a post-doctoral Scientist. In 1995 he joined the Corporate Research and Advanced Development of Robert Bosch GmbH, Hildesheim, Germany, where he was responsible for 'Computer Vision for Automotive Applications'. In 2001 he became chaired professor and director of the Institute for Measurement and Control Systems at Karlsruhe Institute of Technology. In 2010 he was appointed as Distinguished Visiting Scientist for three months at CSIRO in Brisbane, Australia. Dr. Stiller served as President of the IEEE Intelligent Transportation Systems (ITS) Society (2012-2013) and was Vice President for Publications (2009-2010) and for Member Activities (2006-2008). He served as Editor-in-Chief of the IEEE ITS Magazine (2009-2011) and as Associate Editor for the IEEE Transactions on Image processing (1999-2003), for the IEEE Transactions on ITS (2004-ongoing) and for the IEEE ITS Magazine (2012-ongoing). His Autonomous Vehicle AnnieWAY has been Finalist in the Urban Challenge 2007, in the USA and Winner of the Grand Cooperative Driving Challenge 2011 in Holland. In 2013, he collaborated with Daimler on the automated Bertha Benz memorial tour.