

# Visual Landmark based 3D Road Course Estimation with Black Box Variational Inference

Felix Trusheim<sup>1</sup>, Alexandru Condurache<sup>1</sup>, and Alfred Mertins<sup>2</sup>

<sup>1</sup>Robert-Bosch GmbH, Stuttgart, Germany

<sup>2</sup> ISIP, University of Luebeck, Germany

**Abstract.** In this paper we present an approach which estimates the course of a road over long distances based on static and dynamic scene cues detected by a video camera. The approach is based on a clothoid road model, a probabilistic fusion concept as well as a fast variational inference method. Our experimental results show that the approach outperforms a state-of-the-art road marking-based method in challenging real-world driving situations.

**Keywords:** Road Course Estimation, ADAS, Probabilistic Environment Model, Black Box Variational Inference, Clothoid Road-Course Model, 3D Scene Reconstruction

## 1 Introduction

Automated driving requires a robust and precise estimation of the road course. This is a challenging task, which for all conceivable driving situations and environments is still unsolved. State-of-the-art road-course detection systems fuse different complementary sensor signals to receive better estimation results. These are typically signals of radar, lidar and camera systems. Within these setups, camera systems are a valuable source of information. They usually contribute detected road markings to the road-course estimation. Unfortunately, this kind of information is often not sufficiently available in many driving situations like for example on some newly build roads or in dimly lit environments (e.g. at night). Therefore, we propose a camera-based estimation approach, which does not depend on information from road markings but instead on information of different road-course correlated scene cues. Thus, our approach estimates the course of a road within reach of 140 m on the base of static and dynamic scene cues like delineators or other traffic participants. We build our algorithm on a clothoid road model and fuse measured scene cues with the help of a probabilistic model and variational inference. We empirically evaluate our approach in challenging real-world driving situations with reduced light, and prove its performance.

### 1.1 Related Work

The importance of a robust road-course estimation for the realization of a vehicle-cruise-control system (ADAS) led to a high research interest in this subject early

on. The first commercially available systems were based primarily on radar sensors. Such systems detect the road course based on radar-signal-reflecting landmarks such as cars or guardrails. Since then the abilities of such systems progressed constantly. State-of-the-art radar systems depend heavily on increased and detailed environment models, probabilistic estimation approaches [1],[7] and advanced radar-hardware designs [2]. Of particular interest is the approach by Hammarstrand et al. [3]. They received a more robust estimation system by integrating a clothoid road model, which explicitly models the course of a road according to real-world design principles in road construction [23]. The first camera-based systems used road markings for a road-course estimation [4],[5]. Recent proposals for camera-based systems depend exclusively [11] or additionally [10] on semantic segmentation results computed with the help of a deep convolutional neuronal network. These kind of approaches produce remarkable results in day-light situations. But because these methods mainly exploit surface textures of objects for a scene segmentation, they often lack performance in dimly lit environments (e.g. at night) where surface textures are hardly visible. To overcome limitations of individual sensor systems and methods, fusion-based approaches, that integrate information of multiple methods as well as multiple sources, such as radar, lidar, cameras sensors and digital HDR maps, were proposed [6],[8],[7]. Popular representatives of these use an occupancy grid to fuse the information [9]. Another very promising fusion method has been proposed by Geiger et al. [12],[13] for a camera-only crossroad structure estimation. Based on graphical probabilistic modeling the approach showed remarkable estimation performances. However, a computational bottleneck of this approach is the used sampling-based inference method. Recent published variational-inference techniques [18],[19],[20] promise to solve this problem. Therefore, in this contribution we use a similar probabilistic fusion model combined with an efficient variational inference technique for an estimation of the road course based on camera-detected static and dynamic scene cues. Our versatile probabilistic fusion framework also allows the integration of information provided by other sensors, if these are available.

The remainder of the paper is structured as follows: Sec.2 describes our proposed estimation approach in detail, then in Sec.3 we evaluate and discuss our method in challenging driving situations. The Sec. 4 summarizes our approach and offers a brief outlook on future work.

## 2 Road Course Estimation using Variational Inference

### 2.1 Scene Cues

The foundation of our approach are scene cues or landmarks that reflect the course of a road. They are detected by a monoscopic camera system which is mounted behind the rear-view mirror in a car (see Fig. 1).

These scene cues (see Fig. 1) are:

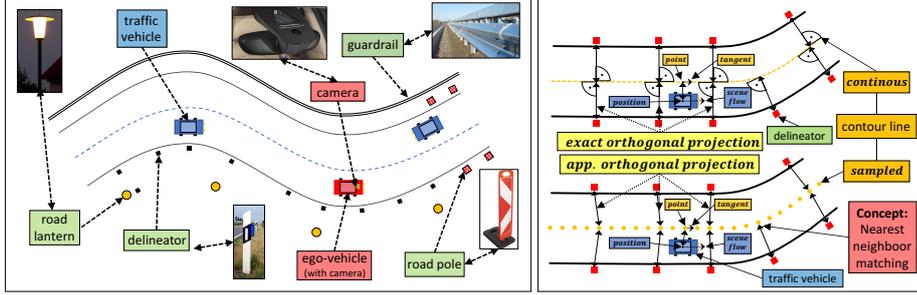


Fig. 1. Schematic road course with regarded evidence types (left) and different data-projection concepts (right)

- **Static Objects:** Guardrails, delineators, road poles, road lanterns and road-embedded reflectors.
- **Dynamic Objects:** Bicycle, cars, motorbikes and trucks.

The static and dynamic objects are detected with the help of different classifiers. Each classifier detection is rated by a confidence measure. The detected objects are tracked over time and thus generate tracklet information. The classifier and tracking [17] methods are not the subject of this publication, and therefore will not be discussed in detail.

In addition to the images, our road-course estimation approach makes also use of information from a 6-axis Inertial Measurement Unit (IMU).

## 2.2 Causality Model

The theoretical foundation of the proposed approach is a generative Bayesian network (BN) model which describes the causal relationship between the road-course defining parameters and the image projections of road-course correlated static and dynamic scene cues. A sketch of the context of our approach together with the proposed probabilistic model are shown in Fig. 2. To properly introduce the complete modeling, we begin with a description of a sub-part of the model. Hence, we start with the causal relationship between the road-course determining parameters and the 3D positions of the static landmarks. A look at the context-referencing Fig. 2, suggests that the course of a road can be described by a virtual contour line. This contour line can be quantified by a road model  $\mathbf{f}(\mathbf{C})$ , in which the  $\mathbf{C}$  represents the shape-defining parameters. The 3D positions  $\mathbf{y}_{Stat3DPos}$  of the road-course aligned static landmarks can then be described as objects lateral-shifted to the contour line along the contour-line normal  $\tilde{\mathbf{f}}(\mathbf{C})$ . This is exemplarily shown with delineators in Fig. 2. While driving, those 3D positions  $\mathbf{y}_{Stat3DPos}$  get projected on the image sensor according to the 3D pose of the user-vehicle camera  $\mathbf{y}_{UVCam3DPose}$  and thereby generate landmark-corresponding measurements  $\mathbf{y}_{Stat2DPos}$ . The same ideas can be transferred to dynamic objects as well. Analogous to static landmarks, the 2D

positions  $\mathbf{y}_{V2DPos}$  can be modeled as a causal superposition of the road-contour parameter  $\mathbf{C}$ , the vehicle-specific offsets  $O_D^{[p]}$  and the camera poses  $\mathbf{y}_{UVCam3Pose}$ . However, unlike static objects, dynamic objects, such as vehicles, change their positions over time and generate a 3D scene flow  $\mathbf{y}_{V3DFlow}$  as well as a corresponding sparse optical flow  $\mathbf{y}_{V2DFlow}$ . Because these generated flows are directly connected with the positions of the vehicles, they can be modeled equally by a causal superposition of the road-contour parameter  $\mathbf{C}$  and the vehicle lateral offsets  $\mathbf{O}_D$ . This model, in conjunction with the measured evidence  $\mathbf{y}_{Stat2DPos}$ ,  $\mathbf{y}_{V2DPos}$ ,  $\mathbf{y}_{V2DFlow}$  and the IMU data  $\mathbf{X}$  and  $\Theta$ , enables the estimation of the road-course parameters  $\mathbf{C}$ ,  $\mathbf{O}_S$  and  $\mathbf{O}_D$ .

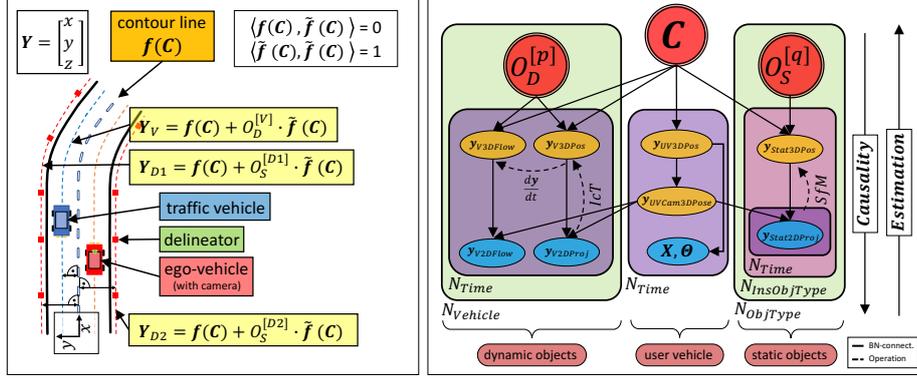
In order to increase the estimation robustness we build the proposed model on additional design principles:

- **Data Buffering:** To obtain a sufficient amount of evidence we accumulate ego-motion compensated data over a time period  $T$ .
- **Flat World Assumption:** To simplify the complexity of the contour model  $f(\mathbf{C})$  during the estimation process we make the assumption that the 3D road course is located on flat plane and therefore can be handled as a 2D road course.
- **Multi Stage Design:** We use a two stage signal processing pipeline (see Fig. 3) to reconstruct a 3D scene from measured 2D information and then estimate the course of the road in 3D.
  - In the **first stage** we reconstruct the 3D signals  $\mathbf{y}_{Stat3DPos}$ ,  $\mathbf{y}_{V3DPos}$  and  $\mathbf{y}_{V3DFlow}$  from the corresponding image measurements  $\mathbf{y}_{Stat2DPos}$ ,  $\mathbf{y}_{V2DPos}$  and  $\mathbf{y}_{V2DFlow}$ . The 3D positions of trackable static objects are obtained with the help of standard structure-from-motion (SfM) methods. In detail, we use a combination of an inverse-depth reconstruction [22] and a bundle adjustment calculation [21]. The necessary ego-motion information is obtained from the IMU. For the 3D reconstruction of vehicles we use an approach based on prior knowledge about the geometry of the vehicles, as well as the assumption of a distortion-free camera projection model. Hence, we calculate the 3D positions and the 3D scene flows with the help of the intercept theorem (IcT) [22] and standard tracklet-based differential methods.
  - In the **second stage** of the pipeline we estimate the road course based on the 3D evidence. This stage is presented in detail in Sec. 2.3.

### 2.3 Probabilistic Model

The purpose of our approach is the identification of the contour parameter  $\mathbf{C}$  as well as the lateral-offset parameters  $\mathbf{O}_S$  and  $\mathbf{O}_D$ , which explain the measured evidence data  $\mathbf{Y}$ . This directly corresponds to a regression problem. However, due to the chosen clothoid-contour model

$$f(\mathbf{C}) = \begin{bmatrix} x(l) \\ y(l) \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \int_0^l \begin{bmatrix} \cos(\phi(t)) \\ \sin(\phi(t)) \end{bmatrix} \Big|_{\phi(t)=\phi_0+\kappa_0 \cdot t + \frac{\kappa_1}{2} \cdot t^2} dt, \quad (1)$$



**Fig. 2.** Schematic road course (left) and proposed probabilistic model (right). The acronyms used in the figures are explained in the text.

this regression problem is not trivial. The integral terms complicates the mathematical handling of the road model within a regression problem. A solution to this problem can be derived by the clothoid-approximation framework of Bertolazzi [24]. This framework allows the definition of a clothoid based on the configuration of its start and end point. Therefore, the contour parameter  $\mathbf{C}$  is determined by

$$\mathbf{C} = [x_{Start}, y_{Start}, \alpha_{Start}, x_{End}, y_{End}, \alpha_{End}] .$$

In order to adapt the model to the measured evidence data we need an effective method to project this data orthogonally to road-contour model  $f(\mathbf{C})$  (see Fig. 1). However, an optimal orthogonal projection results in a computational heavy regression problem. To avoid that, we apply an approximative orthogonal projection concept similar to the procedure proposed by Geiger et al. [13]. In the first stage, we therefore sample the clothoid road model along its length. Based on the sampling, an evidence point  $\mathbf{y}$  is then assigned to the closest clothoid-sample point  $\mathbf{p}$ . Associated with that assignment, the scene flow of an evidence point  $\mathbf{y}$  (in case of a moving vehicle) is then connected to the corresponding clothoid-tangent at point  $\mathbf{p}$  (see Fig. 3). Therefore, the regression problem is no longer differentiable and hence can not be solved by an efficient gradient-based optimization method.

Alternatively, we reformulate the regression problem as a probabilistic maximum a-posteriori (MAP) estimation within a graphical model framework and solve that with the help of variational inference. In this reformulation, the regression parameters  $(\mathbf{C}, O_S, O_D)$  correspond to hidden random variables and the evidence data  $\mathbf{Y}$  correspond to observable random variables. The general

structure of the joint distribution of this MAP-problem follows the form

$$P(\mathbf{H}, \mathbf{Y}) \Big|_{\substack{\mathbf{H} = \{\mathbf{C}, \mathbf{O}_S, \mathbf{O}_D\} \\ \mathbf{Y} = \{\mathbf{Y}_S, \mathbf{Y}_D\}}} = P(\mathbf{C}) \cdot P(\mathbf{O}_S) \cdot P(\mathbf{O}_D) \quad (2) \\ \cdot P(\mathbf{Y}_S | \mathbf{C}, \mathbf{O}_S) \cdot P(\mathbf{Y}_D | \mathbf{C}, \mathbf{O}_D) .$$

Thereby, the evidence data  $\mathbf{Y}$  consists of static landmarks  $\mathbf{Y}_S$  and dynamic landmarks  $\mathbf{Y}_D$ . The prior terms of the regression parameters are assumed as Gaussian distributed. Thus, they are defined as

$$P(\mathbf{H}_i) = \frac{\exp\left(-\frac{1}{2}(\mathbf{H}_i - \boldsymbol{\mu}_{H_i})^T \boldsymbol{\Sigma}_{H_i}^{-1}(\mathbf{H}_i - \boldsymbol{\mu}_{H_i})\right)}{\left((2\pi)^{\dim(\mathbf{H}_i)} \cdot \det(\boldsymbol{\Sigma}_{H_i})\right)^{\frac{1}{2}}} . \quad (3)$$

Here, the parameters  $\boldsymbol{\mu}_{H_i}$  and  $\boldsymbol{\Sigma}_{H_i}$  represent manually defined hyperparameters. The likelihood terms of the joint distribution, which reflect the errors between the road model-predicted evidence data  $\hat{\mathbf{Y}}$  and the true evidence data  $\mathbf{Y}$ , are modeled similarly in our approach. Their structure is as follows:

$$P(\mathbf{Y}_S | \mathbf{C}, \mathbf{O}_S) = \prod_{n=1}^{N_S^{Pos}} P(\mathbf{Y}_S^{Pos[n]} | \mathbf{C}, \mathbf{O}_S) , \quad (4)$$

with

$$P(\mathbf{Y}_S^{Pos[n]} | \mathbf{C}, \mathbf{O}_S) = \frac{\beta_S^{Pos \frac{1}{2}} \cdot \exp\left(-\frac{\beta_S^{Pos}}{2} \cdot \frac{errorPos(\mathbf{C}, \mathbf{O}_S, \mathbf{Y}_S^{Pos[n]})^2}{\sigma_{\mathbf{Y}_S^{Pos[n]}^2}}\right)}{\left(2\pi \cdot \sigma_{\mathbf{Y}_S^{Pos[n]}^2}\right)^{\frac{1}{2}}}$$

and

$$P(\mathbf{Y}_D | \mathbf{C}, \mathbf{O}_D) = \prod_{n=1}^{N_D^{Pos}} P(\mathbf{Y}_D^{Pos[n]} | \mathbf{C}, \mathbf{O}_D) \cdot \prod_{n=1}^{N_D^{Flow}} P(\mathbf{Y}_D^{Flow[n]} | \mathbf{C}, \mathbf{O}_D) \quad (5)$$

with

$$P(\mathbf{Y}_D^{Pos[n]} | \mathbf{C}, \mathbf{O}_D) = \frac{\beta_D^{Pos \frac{1}{2}} \cdot \exp\left(-\frac{\beta_D^{Pos}}{2} \cdot \frac{errorPos(\mathbf{C}, \mathbf{O}_D, \mathbf{Y}_D^{Pos[n]})^2}{\sigma_{\mathbf{Y}_D^{Pos[n]}^2}}\right)}{\left(2\pi \cdot \sigma_{\mathbf{Y}_D^{Pos[n]}^2}\right)^{\frac{1}{2}}}$$

$$P(\mathbf{Y}_D^{Flow[n]} | \mathbf{C}, \mathbf{O}_D) = \frac{\beta_D^{Flow \frac{1}{2}} \cdot \exp\left(-\frac{\beta_D^{Flow}}{2} \cdot \frac{errorFlow(\mathbf{C}, \mathbf{O}_D, \mathbf{Y}_D^{Flow[n]})^2}{\sigma_{\mathbf{Y}_D^{Flow[n]}^2}}\right)}{\left(2\pi \cdot \sigma_{\mathbf{Y}_D^{Flow[n]}^2}\right)^{\frac{1}{2}}} .$$

Here, the proposed Gaussian character of those distributions shall reflect the typical lateral-fluctuations of landmark positions in the real-world, as well as inaccuracies in our 3D reconstruction and small sampling-caused projection errors (see Fig. 1). The value of  $errorPos$  corresponds to the euclidean distance between the road model-predicted position and the true position of a evidence data point. Similarly,  $errorFlow$  reflects the anti-correlation between the road model-predicted scene flow and the true scene flow of an evidence data point (see Fig. 1). The hyperparameters  $N_S^{Pos}$ ,  $N_D^{Pos}$ ,  $N_D^{Flow}$  depict the quantities of the different data types within the measured evidence data and the values of  $\sigma_{\mathbf{Y}_S^{Pos}}$ ,  $\sigma_{\mathbf{Y}_D^{Pos}}$ ,  $\sigma_{\mathbf{Y}_D^{Flow}}$  are related to the specific confidences of the measured evidence points. In contrast,  $\beta_S^{Pos}$ ,  $\beta_D^{Pos}$  and  $\beta_D^{Flow}$  represent design parameters which control the influence of the corresponding evidence data types in the MAP-problem.

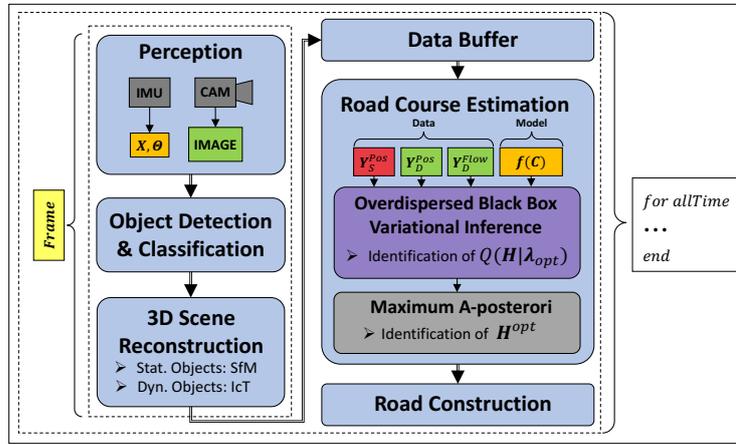


Fig. 3. Schematic representation of the estimation pipeline

## 2.4 Variational Inference

Solving the MAP-problem requires an analysis of the extreme value of the a-posteriori distribution  $P(\mathbf{H}|\mathbf{Y})$  which is defined by the joint distribution  $P(\mathbf{H}, \mathbf{Y})$ . However, this analysis is not trivial, because of the structure of the problem and the continuous random variables. Therefore, we propose to approximate the a-posteriori distribution initially and then infer the MAP-solution based on the generated approximation. In detail, we exploit a state-of-the-art variational inference technique, called Overdispersed Black-Box Variational Inference (O-BBVI) [18],[20]. This deterministic and fast-converging method lacks the typical high computational cost of popular sampling-based techniques [15],[16].

When applied to the MAP-problem, O-BBVI approximates the exact a-posteriori

distribution  $P(\mathbf{H}|\mathbf{Y})$  by a factorized distribution with the form

$$Q(\mathbf{H}|\boldsymbol{\lambda}) = \prod_{i=1}^N q_{H_i}(\mathbf{H}_i|\boldsymbol{\lambda}_i) ,$$

in which each component  $q_{H_i}$  is defined as:

$$q_{H_i}(\mathbf{H}_i|\boldsymbol{\lambda}_i) = \frac{\exp\left(-\frac{1}{2} \sum_{d=1}^{\dim(\mathbf{H}_i)} \left(H_i^{[d]} - \lambda_i^{[d]}\right)^2\right)}{(2\pi)^{\frac{1}{2} \dim(\mathbf{H}_i)}} . \quad (6)$$

Based on that, O-BBVI computes this approximation by an optimization over the parameters  $\boldsymbol{\lambda}_i$  of the factor terms  $q_{H_i}$ . The parameter-decoupling structure of the O-BBVI approximation modifies the solution of the MAP-problem to

$$\mathbf{H}^{opt} = \underset{\mathbf{H}}{\operatorname{argmax}} (P(\mathbf{H}|\mathbf{Y})) \Big|_{P(\mathbf{H}|\mathbf{Y}) \approx Q(\mathbf{H}|\boldsymbol{\lambda}_{opt})} \quad (7)$$

$$\approx \underset{\mathbf{H}}{\operatorname{argmax}} (Q(\mathbf{H}|\boldsymbol{\lambda}_{opt})) \Big|_{Q(\mathbf{H}|\boldsymbol{\lambda}_{opt}) = q(\mathbf{C}|\boldsymbol{\lambda}_{opt}^C) \cdot q(\mathbf{O}_S|\boldsymbol{\lambda}_{opt}^{O_S}) \cdot q(\mathbf{O}_D|\boldsymbol{\lambda}_{opt}^{O_D})} \quad (8)$$

$$\begin{aligned} \mathbf{C}^{opt} &= \underset{\mathbf{C}}{\operatorname{argmax}} (q_C(\mathbf{C}|\boldsymbol{\lambda}_C^{opt})) = \boldsymbol{\lambda}_C^{opt} \\ \implies \mathbf{O}_S^{opt} &= \underset{\mathbf{O}_S}{\operatorname{argmax}} (q_{O_S}(\mathbf{O}_S|\boldsymbol{\lambda}_{O_S}^{opt})) = \boldsymbol{\lambda}_{O_S}^{opt} \\ \mathbf{O}_D^{opt} &= \underset{\mathbf{O}_D}{\operatorname{argmax}} (q_{O_D}(\mathbf{O}_D|\boldsymbol{\lambda}_{O_D}^{opt})) = \boldsymbol{\lambda}_{O_D}^{opt} . \end{aligned}$$

Finally, this MAP-solution in conjunction with the applied road-contour model  $\mathbf{f}(\mathbf{C})$  returns the full 3D-description of an estimated road course.

### 3 Experiments and Discussion

In the following section, we will compare our scene cue based fusion approach on three real-world traffic examples with a corresponding state-of-the-art estimation method which is based on road markings [5]. The chosen examples cover various driving situations in which a road marking-based approach shows weaknesses in comparison with the landmark-based approach. The spectrum of the situations varies across different types of roads and different types, numbers and densities of available static and dynamic scene cues. All the situations are scenarios with reduced light. The course estimations for both methods are post-processed under real-time conditions on the base of recorded images and logged IMU data. However, for a better understanding of the scene structure, we visualize the IMU-recorded trajectory of the ego-vehicle as well as all the 3D-reconstructed static landmarks upfront. The results are shown in Fig. 4, 5, 6.

The first scenario in Fig. 4 represents a driving situation on a country road in low-light conditions. The road has roadside-markings and rows of delineators on both sides. The ego-vehicle follows an other vehicle through a curve. The results demonstrate that the range of the landmark-based estimation exceeds the corresponding estimation of road marking-based approach by more than 25 m or 33% because of the detected delineator on the right side. Making use of the delineators on the right side, our approach is even capable of seeing a short distance around the bending of the curve and hence identifies a part of the course which is not detected by the road-marking based system.

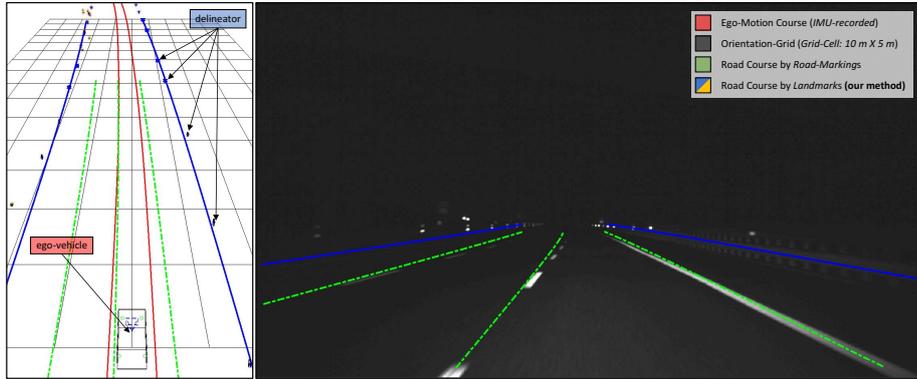
The second scenario in Fig. 5 reflects a ride on a highway at night. The scene is only illuminated by the high-beam head lamps of the ego-vehicle. On both sides of the road are reflectors mounted on the guardrails. Road markings are also available. The results illustrate that our approach achieves an estimation of the road course up to 140 m relative to the position of the ego-vehicle. This is accomplished with the help of the detected delineators. As a result, our estimation reaches 70 m or 100% further than the corresponding estimation based on road markings.

The scene in Fig. 6 shows a driving situation in an urban environment at night. The road has no delineators or roadside-markings. The scene is mainly illuminated by a few road-aligned road lanterns on the right and left side. Hence the road lanterns on the right side, our approach generates a virtual road boundary. This information can not be exploited by a road-marking based method. Next



Fig. 4. Rural Road: 3D Reconstruction (left), Camera Image (right)

to the comparison with the road-marking based method, we also evaluate our approach statically with labeled ground-truth road courses on a database of approximately 14,000 images of various night traffic situations. For this purpose, we labeled the regions in those images where we expect the boundaries of the estimated road courses with polygon-shaped tubes. Based on that labeling, we rate road-course estimations which are fully embedded within the labeled polygons



**Fig. 5.** Highway: 3D Reconstruction (left), Camera Image (right)



**Fig. 6.** Urban Road: 3D Reconstruction (left), Camera Image (right)

as true positives. Estimations which do not fulfill this criterion are rated as false positives. In this evaluation framework our approach achieves a true positive rate (TPR) of 92.18% and a false positive rate (FPR) of 4.68%. With the objective of using our approach on hardware platforms with limited computational resources, we further investigate the influence of the amount of the numerically expensive iterations within the O-BBVI inference method on the quality of the estimated road courses. In detail, we reduce the number of iterations to 75%, 50% and 25% of the amount of iteration which are needed for a complete convergence of the O-BBVI method during the inferences. The reduction to 75% results in a TPR of 81.8% and FPR of 15.4%. A further reduction to 50% causes a TPR of 74.5% and a FPR of 22.9% and a reduction to 25% lowers the TPR to 61.5% and FPR of 36.4%. These results implicate that the O-BBVI based inferences converge fast after a few iterations. Thus, this circumstance allows our approach to easily adapt to available hardware resources without entirely given up on estimation quality.

## 4 Conclusion and Future Work

In this publication we presented an approach that estimates the course of a road based on images of a monoscopic camera for ranges of 140 m, particularly in difficult situations with reduced light. The method therefore uses static and dynamic scene cues which are correlated to the course of a road. The underlying fusion concept is flexible and hence works with a variety of different landmark types and quantities. This makes the approach highly adaptive to varying evidence in a scene. In order to optimally respond to real-world road designs we proposed to use a clothoid road model. The associated complications with such a road model in a regression problem were addressed with a probabilistic model and a numerically efficient and adaptive variational inference. We demonstrated the performance of our algorithm in challenging low-light driving situations. Thereby, we proved that the approach can achieve larger estimation ranges than a comparable road marking-based method in the same situations. These results reflect that information fusion provides a framework to integrate expert knowledge over the problem setup with data-driven insights into the decision-making process.

In future work we plan to strengthen the presented approach in several areas. At first, we plan to exploit our current set of evidence data more effectively in the fusion process. Therefore, we would like to substitute manual-defined hyperparameters within the probabilistic model (see Eq. 3, 4, 5) by data-trained counterparts. In addition to that, we intend to develop a robust strategy to identify lane-changing vehicles. This would allow us to react better to lateral-shifts of vehicles during the estimation process. Beyond these improvements we would like to extend the current set of evidence data by integrating more camera signals, like road markings or semantic segmentation results, as well as signals from other sensors, such as lidar or radar into our probabilistic fusion framework. We expect that this will enhance the robustness and also will allow us to model even more complex driving scenarios, like splitting or reunifying roads. Furthermore, we plan to improve the variational inference procedure in order to achieve even faster and more precise estimates.

## References

1. Nieto, J., Hernandez-Gutierrez, A., Nebot, E.: Probabilistic Estimation Of Unmarked Roads Using Radar. In: *Journal of Physical Agents*, Vol. 4, 2010
2. Sarholz, F., Mehnert, J., Klappstein, J., Radig, B.: Evaluation Of Different Approaches For Road Course Estimation Using Imaging Radar. In: *Conference on Intelligent Robots and Systems*, 2011 IEEE
3. Hammarstrand, L., Fatemi, M., Angel F., Svensson, L.: Long-Range Road Geometry Estimation Using Moving Vehicles And Road-Side Observations. In: *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, pp. 2144–2158. 2016 IEEE
4. Goldbeck, J., Huertgen, B.: Lane detection and tracking by video sensors. In: *IEEE Intelligent Transportation Systems*, pp. 74-99, 1999

5. Liu, W., Zhang, H., Duan, B.: Vision-Based Real-Time Lane Marking Detection And Tracking. In: 11th International Conference on Intelligent Transportation Systems, 2008 IEEE
6. Konrad, M., Szczot, M., Schule F.: Generic Grid Mapping For Road Course Estimation. In: Intelligent Vehicles Symposium (IV), 2011 IEEE
7. Konrad, M., Nuss, D., Dietmayer, K.: Localization In Digital Maps For Road Course Estimation Using Grid Maps. In: Intelligent Vehicles Symposium (IV), 2012 IEEE
8. Schule F., Schweiger, R., Dietmayer, K.: Augmenting Night Vision Video Images With Longer Distance Road Course Information. In: Intelligent Vehicles Symposium (IV), 2013 IEEE
9. Schule F., Koch, C., Hartmann, O.: Probabilistic Fusion Of Rural Road Course Estimations. In: 16th International Conference on Intelligent Transportation Systems, 2013 IEEE
10. Limmer, M., Forster, J., Baudach, D., Schule F., Schweiger, R., Lensch, H. : Robust Deep-Learning-Based Road-Prediction for Augmented Reality Navigation Systems. In: Intelligent Vehicles Symposium (IV), 2012 IEEE
11. Teichmann, M., Weber, M., Zollner, M., Cipolla, R., Urtasun, R. : MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving. In: arXiv:1612.07695v1
12. Geiger, A., Wojek, C., Urtasun, R.: Joint 3D Estimation Of Objects And Scene Layout. In: In Advances in Neural Information Processing Systems (NIPS), 2011
13. Geiger, A., Lauer, M., Wojek, C., Stiller, C., Urtasun, R.: 3D Traffic Scene Understanding from Movable Platforms . In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 36, 2014
14. Bishop, C.: Pattern Recognition and Machine Learning. In: Springer, 2006
15. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distributions, And The Bayesian Restoration Of Images. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721-741, 1984
16. Goodman, J., Weare, J.: Ensemble Samplers With Affine Invariance. In: Communications in Applied Mathematics and Computational Science, Vol. 5, 2010
17. Trusheim, F., Condurache, A., Mertins, A.: Graphical Stochastic Models for Tracking Applications with Variational Message Passing Inference. In: 6th International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 2016
18. Ranganath, R., Gerrish, S., Blei, D.: Black Box Variational Inference. In: 17th International Conference on Artificial Intelligence and Statistics (AISTATS), Reykjavik, Iceland, 2014
19. Titsias, M., Lazaro-Gredilla, M.: Local Expectation Gradients For Black Box Variational Inference. In: 28th International Conference on Neural Information Processing Systems (NIPS), Montreal, Canada, 2015
20. Ruiz, F., Titsias, M., Blei, D.: Overdispersed Black-Box Variational Inference. In: The Conference on Uncertainty in Artificial Intelligence, New York, USA, 2016
21. Engels, C., Stewenius, H., Nister, D.: Bundle Adjustment Rules. In: Photogrammetric Computer Vision, 2006
22. Hartley, R., Zisserman, A.: Multiple View Geometry In Computer Vision. In: Cambridge University Press, 2004
23. Gackstatter, C., Thomas, S., Heinemann, P., Klinker, G.: Stable Road Lane Model Based On Clothoids. In: Advanced Microsystems for Automotive Applications, pp 133–143, 2010
24. Bertolazzi, E., Frego, M.:  $G^1$  Fitting With Clothoids. In: Wiley Online Library, 2013