Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Building a probabilistic grid-based road representation from direct and indirect visual cues

Edoardo Casapietra, Thomas H Weisswange, Jannik Fritsch, Franz Kummert, Christian Goerick

2015

Preprint:

This is an accepted article published in IEEE Intelligent Vehicles Symposium (IV 2015). The final authenticated version is available online at: https://doi.org/10.1109/IVS.2015.7225698

Building a Probabilistic Grid-based Road Representation from Direct and Indirect Visual Cues

E. Casapietra, T.H. Weisswange, C. Goerick, F. Kummert, J. Fritsch

***Abstract*—Detecting the road terrain ahead of the ego-vehicle is an important issue for modern driver assistance systems. In particular, vehicle motion planning in inner city environment requires the detection of road terrain up to 3 seconds in advance. State-of-the-art visual road terrain detection systems have a hard time fulfilling this task, due to their limited range and the presence of occlusions (other vehicles, buildings, etc.), which are expected to occur often in complex scenarios. However, those systems provide significant information where the conditions are favorable (proximity to the ego-vehicle, no occlusions). Therefore, a complementary approach is needed to enhance already existing and established detection systems.**

In this paper we propose a probabilistic grid-based approach based on the observation and interpretation of other vehicles' behavior in the scene. It exploits their movements in order to infer the presence and location of occluded road surface. We will show that this approach presents various advantages over current visual road terrain detection systems, especially in those situations that are the most challenging for them. We will illustrate how our approach is designed to work in concert also with other available resources, e.g. offline road maps.

Qualitative results on real-world scenes taken from the KITTI benchmark[11] demonstrate that the fusion of this method with visual road terrain detection can potentially extend our time horizon well over the 3 seconds mentioned above. Finally, we will show how our approach is planned to develop into a semantically enriched representation of the road, including road properties such as availability, lanes and directions.

I. Introduction

Modern day Advanced Driver Assistance Systems (ADAS) provide a very large range of potential uses, such as cruise-control in highways, parking assistance or emergency breaking. In order to provide warnings or control, these systems rely on sensors to get information and awareness about the surrounding of the vehicle. In particular, the detection of road is an important task for the majority of future ADAS. Knowing the position and geometry of the road is essential for safety systems: a detected object can be a dangerous obstacle if it lies in the middle of a road, it can be only a potential danger if it is on the sides of the road, and it can be harmless if it is outside. Similarly, spatial knowledge about the road is necessary for motion control systems, that need to know exactly where the ego-car can move on.

For systems that work in a highly dynamic environment, time is crucial. The earlier a road is detected, the more time the system has to plan a trajectory. Control systems have different levels of planning: the lowest one is basic reaction to detected obstacles, which plans control signals on a very short time scale. However, the aim of motion control is not to simply move without accidents, but it is also to do it in a comfortable way for human passengers. Therefore, a higher level of planning is necessary, in order to plan smooth and reasonable trajectories. This level needs a much longer time scale than its lower counterpart, usually planning motion control signals up to 3 seconds ahead[3] (at a speed of 50 km/h, this corresponds to a sensing range of about 40 meters), and therefore would strongly benefit from an early road detection. Already existing autonomous driving systems rely on navigational road maps for this level of planning[21][7][17]. However, navigational road maps present a problem: they focus on connectivity, sacrificing the spatial aspect. The planning of a smooth trajectory would ideally require a sub-meter spatial accuracy, that a navigational map cannot provide. Furthermore, maps cannot deal with unexpected events (e.g. the creation of a construction site). All these characteristics make navigational road maps more fitting for an even higher level of planning, which deals with a time scale of minutes or even hours.

Of the three levels of planning mentioned, the mid-level is the one with the most unsolved issues. In fact, while navigational maps can comfortably deal with route planning, the "reactive" level has a wide variety of already established methods that have been developed over the years, relying on a wide array of sensors, such as lasers, radars and cameras, which provides awareness about free space, obstacles and road terrain in the immediate surroundings of the ego-car. In order to cope with the requirements of the mid-level (with a particular attention to the problematic inner-city environment), our aim is to build a spatial representation, which can provide to the ego-car spatially accurate information about road terrain with a high range. The use of visual road detection systems, although it provides good results for the reactive level, is not sufficient. These systems have significant drawbacks which prevent them from being effective: their accuracy decreases with distance due to the limited camera resolution, they rely on good road surface, good imaging conditions (proper illumination, no shadows...) and they require a clear field of view (FOV). The road texture is fairly smooth and homogeneous on highways, but in inner city it can give significant issues to visual road terrain detection systems: different patches of asphalt, cobbles, or other unusual surfaces can pose challenges to those systems. Difficult weather conditions such as snow can also significantly limit their accuracy. Furthermore, illumination conditions can vary greatly in inner city, due to the abundance and proximity of tall buildings. However, the most important problem, also for systems based on other sensors, is presented by occlusions: in an inner city environment, a free FOV is a very unlikely condition. The main obstacles are other traffic participants, which can be at very close distance to the

ego car, hiding not only the areas of road directly under them, but also blocking large parts of the FOV. The same can be said about buildings and other static obstacles, that very often can hide entire roads, especially at the sides of the FOV.

In this paper we propose a novel approach, which exploits the behavior of other vehicles in the scene in order to infer road terrain probability in a 2D grid-based map. In an inner city environment it can happen frequently that a new vehicle appears in the field of view of the ego-car from behind an obstacle (like a building, or another vehicle). The appearance of this vehicle is an event that holds information: there must be road terrain behind the obstacle. The same reasoning holds for the disappearance of a vehicle behind an obstruction. The information is not limited to the mere presence of a road, but also its position and initial direction can be roughly estimated, provided that we can detect the position and direction of the vehicle with sufficient precision. This can be done at a far greater distance than the range of any visual road terrain detector[20], providing extra time for an ADAS to plan according to the new information. We will illustrate our approach through qualitative examples on real-world scenes taken from the KITTI benchmark[11], and show that the proposed approach can detect roads with a horizon of more than 3 seconds (or 40 meters).

Since it relies on other vehicles in the scene (the more there are, the more it can infer), our approach works best exactly in scenarios where visual road detection gives the poorest results. On the other hand, we still need to detect road where there are no vehicles, so visual road detection or other resources are necessary. Therefore we propose a system architecture where our approach is designed to work together with the SPRAY-based visual road terrain detection system developed by Fritsch et al.[10], and with the possible inclusion of navigational road maps.

Finally, we will show how we plan to exploit the further potential of this approach by adding different semantics to our road map, to create a multi-modal representation of the environment.

II. Related Work

Road detection is a key issue for modern ADAS. The most versatile sensor for that kind of detection is the camera. Road detection systems relying on cameras started detecting the road by observing lane markings ("marked" road detection) and exploiting prior knowledge about road geometry, and thus were effective only in highly structured environments, like highways. In order to detect road-like areas even in cases where lane markings are not sufficient ("unmarked" road detection), visual road detectors started analyzing low-level features, such as color and texture. One example is the SPRAY-based road detection system[10] used by our system. Ultimately, methods combining marked and unmarked detection together with other priors, including navigational maps, have been developed[2]

In order to make the detected properties of the environment accessible to the system, a form of representation of the environment itself is needed[4][6][22]. A very popular type of representation is the occupancy grid, which is a spatial representation describing the environment in terms of free or occupied area. Occupancy grids are very useful for short-term maneuvers. They can exploit different sensors, like lasers[22][6], radars[6] or stereo cameras[4], and they can eventually include prior knowledge[13] or stochastic dynamic models[4][8]. Besides occupancy grids, other types of representations have been implemented, such as probabilistic drivability maps [19][14]

Most of these representations are uni-modal, since direct sensory cues only provide information about one particular property of the environment. Aiming to achieve different kinds of information from the same measurements, research started investigating indirect cues. One indirect cue that has been studied is the behavior of other entities in the environment. Behavioral assumptions have been exploited to improve the tracking of vehicles[1], and to predict the motion of vehicles at intersections[16]. Information about context has been used to predict the behavior and classify the related danger of pedestrians[5] or vehicles[23]. The approach we present in this paper aims to do the opposite: using the behavior to infer the context. In this direction, the observation of pedestrians' behavior has been exploited to infer navigational maps for robots[18]. The motion of vehicles has been used, together with lane markings detection and context models, to infer road geometry at long range in highways[9]. This work used a Bayesian inference algorithm to obtain the parameters of a clothoid-based road model. Similarly, the behavior of vehicles has also been used together with other visual cues in order to infer the topological and geometrical layout of road intersections[12], out of a predefined set of possible topologies, using a probabilistic generative model. However, both of these last systems used only the current behavior of vehicles, while the approach we will present in the following will also exploit the predicted behavior.

III. Concepts

Our representation map is based on a regular 2D grid, where each patch has a fixed global position and holds a road probability. A global representation has been chosen because a key aspect of the system is to exploit multiple passages of the same area, increasing the information available at every passage. This poses the problem of the management of the representation map, which has to be dynamically extended whenever the ego-car enters a new area, and will have to compress and store distant areas in order to retrieve them when needed. The exact way of this compression/extension will be part of future improvements of the concept, therefore for now the system operates within a few hundred meters. The initial value of the road probability of each patch, in absence of any prior information, is set to 50% for simplicity. However, this value can be refined by using statistics (e.g. average percentage of road terrain area in the region) or other spatial priors.

In figure 1 the architecture of our system is presented. The necessary inputs to our approach are position, direction and

speed of all vehicles in the scene. The system also needs information about position and velocity of the ego-car, in order to be able to perform the coordinates transformation from the ego-centered reference frame to the global one (all these input are named "Vehicles state" in figure 1). In addition to those inputs, the system is designed to handle other resources: one that has been included is the camera-based road detection. Another possible input is a commercially available navigational road map. These inputs are processed into a road probability value for each patch, which is compared at each time step with the previous one and updated (see formula (2)).

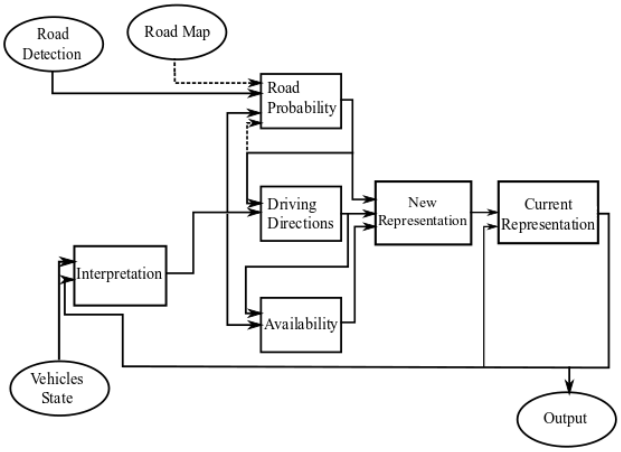


Fig. 1: System architecture of our proposal.

Our approach aims to exploit the very same obstacles which hamper visual road terrain detection: the traffic participants. The underlying concept is that, even if moving vehicles occlude parts of road, they can still provide potentially significant information. In fact, they all possess different information about the environment due to their different position and field of view, and behave according to their understanding of the environment. By observing their behavior, we can retrieve some of that understanding and use it to improve ours.

This behavior interpretation is based on the assumption that vehicles in the scene move on roads, from roads and towards roads. The first and most straightforward thing we need to consider is that the area below a moving vehicle must be road. Thus, all the observed positions of all moving vehicles are considered road with probability 100% by our system. Secondly, by observing the motion of the vehicles, it is possible to infer the road probability of a patch in the surroundings of the vehicles. This is achieved by estimating the probability of the patch being part of the future and past trajectory of a vehicle, taking into account the assumption that vehicles move on roads: if a vehicle is going to, or coming from, that patch, it must be road. Assuming we have an estimation of the vehicle future and past trajectory, the calculation of the road probability $p(R)$ of the patch i (we will use only one index for simplicity) can be done as follows:

$$p(R_i) = p(R_i|\bar{C}_i) + \left[p(R|C) - p(R_i|\bar{C}_i)\right] p(C_i) \quad (1)$$

Where $p(C_i)$ is the probability of the event C_i, that is the patch i being "hit" by a vehicle, $p(R|C)$ is the conditioned probability of a patch being road terrain if a vehicle hits it (which we consider a constant 100% for simplicity, but can be fine-tuned in case of necessity) and $p(R_i|\bar{C}_i)$ is the conditioned road probability in the event that no vehicles hit the patch: this last value is particularly interesting, and will be discussed later in this section.

At each time step, $p(R_i)$ of each patch is recalculated: the estimation of the future trajectories is performed again and this could change $p(R_i)$. If this happens, the new value substitutes the previous one, with a simple weighted average to minimize the impact of isolated errors. A possible formulation is

$$p_t(R_i) = \alpha p(R_i) + \beta p_{t-1}(R_i) \quad (2)$$

In a case with more than one vehicle in the scene, every patch will have to combine the probabilities of being hit by any of them. For simplicity, the combination is performed as the simple union of independent events: we only need one vehicle to hit the patch, in order to consider it road.

$$p(C_i) = p\left(\exists C_{i,j} = 1 \mid j \in [1, n]\right) \quad (3)$$

Where n is the total number of vehicles that can hit the patch i. In case of more than two vehicles, the calculation can be done iteratively, by combining the first two contributions and then incorporating the others one by one.

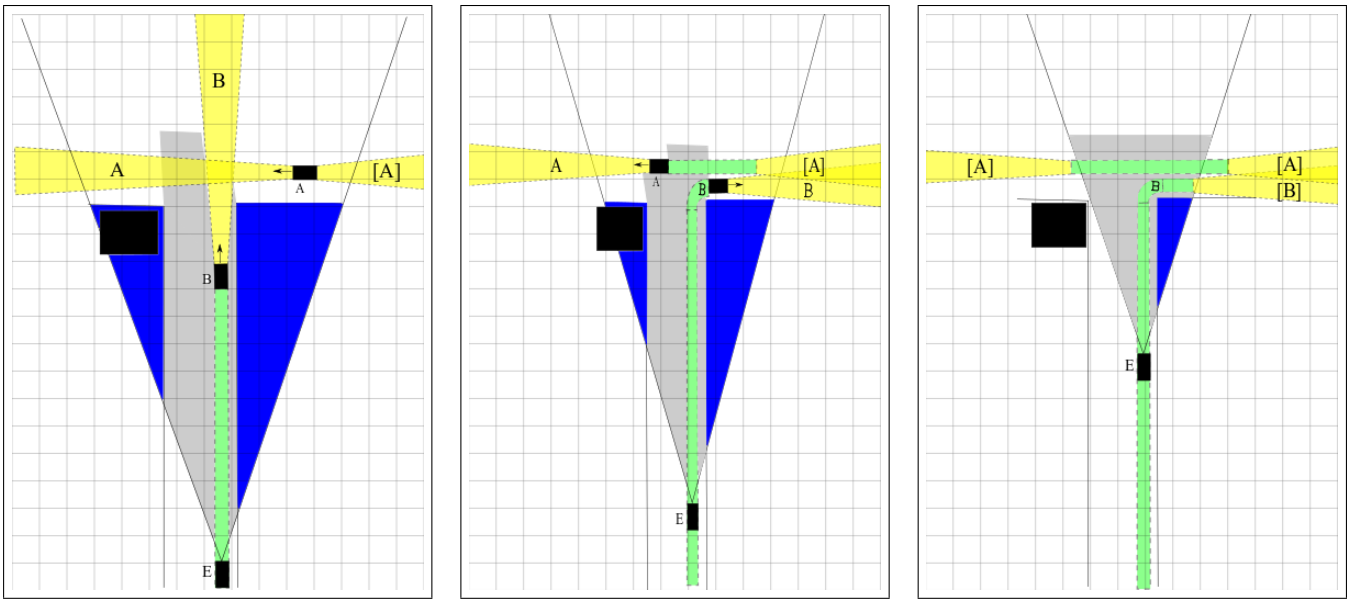
$$\begin{aligned} p_n(C_i) &= p(C_{i,1}) + p(C_{i,2}) - p(C_{i,1})p(C_{i,2}) \text{ if } n = 2 \\ p_n(C_i) &= p_{n-1}(C_i) + p(C_{i,n}) - p_{n-1}(C_i)p(C_{i,n}) \text{ if } n > 2 \end{aligned} \quad (4)$$

Subsequently, the question is how to estimate $p(C_{i,j})$, given the current position, direction and speed of the vehicle. The final result of this process will be to find a spatial 2D function, centered on the position of the vehicle, that assigns to each patch i a probability for the vehicle j to pass there, with direction and speed of j as parameters. Ideally, we need a function whose values decrease with distance (because the vehicle will have a bigger chance to change direction, spreading the probability) and that are higher along the current driving direction of the vehicle. In order to include speed, we decided to use the time to contact $TTC_{i,j}$ instead of the simple distance. The function we have chosen is a Gaussian:

$$p(C_{i,j}) = \frac{1}{\sqrt{2\pi}\sigma_{i,j}} exp\left(\frac{-\alpha_{i,j}^2}{2\sigma^2}\right) \quad (5)$$

Where $\alpha_{i,j}$ is the change in yaw angle needed for the vehicle j to hit patch i, and the variance $\sigma_{i,j}^2 \propto TTC_{i,j}$. This function is applied both for the past and the future trajectory estimation, but always only for the patches whose $\alpha_{i,j}$ is within a certain range.

At any time, each patch does not hold only its $p(R_i)$ value, but it actually contains a list of all vehicles in the



(a) (b) (c)

Fig. 2: Illustration of the concept of temporary and permanent contributions. (a) The ego-car E is following vehicle B, whose trace (in green) is considered known road, and its future trajectory estimation contributes temporarily to the area in front of it. Vehicle A is detected on the right side, going left. Its past trajectory estimation (on the right) is a permanent contribution (marked with squared brackets), while its future trajectory estimation (on the left, behind the building in black) is temporary. (b) Vehicle B has turned right, causing a different trajectory estimation. (c) Both vehicles have disappeared, and their last trajectory estimations have become permanent. Colored in grey is the road terrain visible by the ego-car.

scene which affect it, together with their relative contribution $p(C_{i,j})$. Having this list is useful because $p(R_i)$ does not depend only on the vehicles currently in the scene, but also potentially on vehicles that have disappeared already. The list divides the contributions into two different sets: *temporary* and *permanent*. While the future trajectory estimation is done at every time step (a *temporary* contribution), the estimation of the past trajectory of a vehicle is done only once, when it appears in the FOV (as soon as an estimation of its speed is available). Any further behavior of the vehicle does not affect this estimation, which is therefore considered a *permanent* contribution. When a vehicle disappears from the scene, future trajectory contributions would disappear, but the disappearance of a vehicle is an event that holds information just as its appearance. Hence, when the disappearance of a vehicle is detected, all its *temporary* contributions are saved as *permanent* (see figure 2). This guarantees to keep all the relevant information available. Hence, $p(C_i)$ can be written as follows:

$$p(C_i) = p\left(\exists C_{i,j} = 1 \mid j \in [1, N] \cup \exists \hat{C}_{i,k} = 1 \mid k \in [1, M]\right) \tag{6}$$

Where $C_{i,j}$ are the events related to all the *temporary* estimations (N in total), and $\hat{C}_{i,k}$ are the events related to all the *permanent* estimations (M in total).

Since keeping each *permanent* contribution forever would be a pointless burden for the system, after some time the old *permanent* contributions are condensed in a single "overall past contribution". Keeping the different contributions for a while is a security measure against potential detection fails: if a vehicle is not detected for just one time step, but then it is detected again in the next one, the system can retrieve its contribution and prevent inconsistencies.

Up until now, vehicles can only increase the value of $p(R_i)$. Since our system is designed to pass multiple times through the same area, and to observe many vehicles, there is a big risk of saturating the map. In order to avoid this, it is necessary to introduce a way to decrease the road probability based on a vehicle's behavior. In principle, it is reasonable to consider a patch unlikely to be road if vehicles appear to avoid it. This is achieved by considering the conditional probability $p(R_i|\bar{C}_i)$ mentioned in (1) as a variable instead of a constant. In particular, it should have values always lower than 50% and a shape similar (but opposite) to the trajectory estimation, with the lowest values along the direction of the vehicle, to model the reasoning that if the vehicle does not end up on the patch it was heading to, it is because it actively avoided it. The result, in case of only one vehicle, would be to obtain a difference of Gaussians, with an overall "negative" (i.e. lower than 50%) contribution on the sides. This will help sharpening the estimated road and preventing a saturation of the map.

However, we can also combine the behavior-based approach with a direct road detection system[10] by defining a function, the *visibility* p(V) of each patch, which expresses the

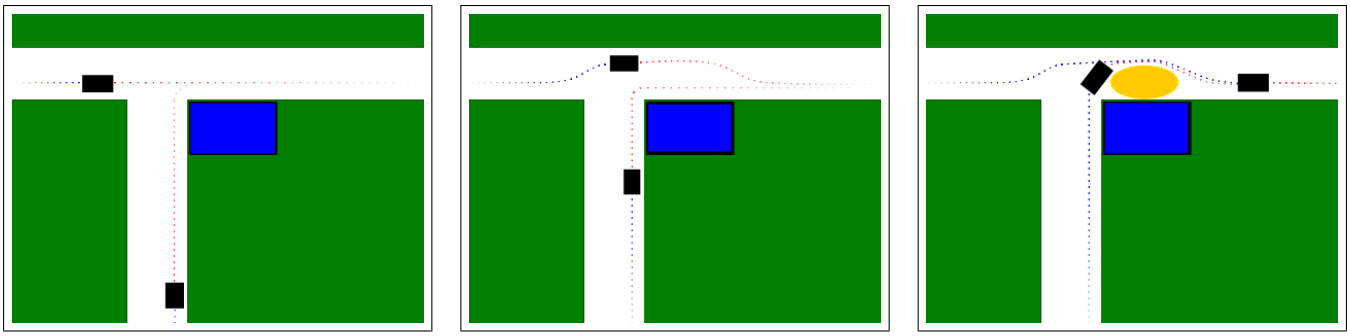


Fig. 3: A possible application of inferred driving directions: two vehicles deviate from the usual paths and seem to avoid the same area behind the obstacle (in blue), therefore the system marks that area as unavailable (yellow).

probability of the patch to be seen correctly by the sensors. The value of this function will be high for the patches lying inside the FOV and zero for occluded ones. The resulting road probability is computed as:

$$p(R_i) = p(R_i|\bar{C}_i) + \left[p(R|C) - p(R_i|\bar{C}_i)\right] p(C_i) \left[1 - p(V_i)\right] + \\ + \left[p(R_i|V_i) - p(R_i|\bar{C}_i)\right] p(V_i) \quad (7)$$

Where $p(V_i)$ is the visibility value, and $p(R_i|V_i)$ is the road terrain confidence provided by the detector. The formula is arranged in such a way that, if a patch is considered visible, the road detector will provide the most important contribution, while if it is considered not visible, the behavior interpretation will become the most important contributor.

An important resource that modern ADAS are exploiting is the availability of commercial navigational road maps. Given their increasing popularity, it is reasonable to include them in our system. There are a few downsides to this idea: the self-localization and the alignment between the navigational map and the representation grid, as well as the different spatial accuracy between them, are important issues. This particular data-fusion problem has been recently addressed in[15]. Furthermore, the format of navigational maps, which are focused on connectivity and are graph-based, differ greatly from our representation, that describes the spatial layout of the environment. Our aim is to assign to each of the patches of our representation a road probability value derived from the navigational map, using the location of the checkpoints and the width of the road segments as a reference. The conversion between two intrinsically different formats will generate road probabilities instead of the binary road/non-road labeling typical in navigational maps, caused by the uncertainty of the actual shape of the road. Due to their nature, though, these probabilities cannot be included in formula (7), like it has been done with the SPRAY-based road detection results. That is because navigational maps do not change with time. Therefore, our intention is to use them as an initialization of our road map: when a new area of the grid is created, the initial value of road probability will not be 0.5, but it will be the one provided by the annotated map.

It is worth noting that our system merges together information relevant for all levels of planning: it has a road terrain detector for short-range, sub-second planning, a behavior interpreter providing information useful for planning a few seconds ahead, and a coarse spatial navigation map for high level route planning.

Most of the representations used by modern systems are uni-modal: they represent only one particular property of the environment, e.g. road/non-road or occupied/free. The representation we presented in this paper belongs to this category too, as it only represents roads. However, this is just the first step. Our approach of using the behavior of road actors can potentially provide more information than just road. For example, a very useful property we plan to include in our representation is the driving direction of a road patch: every time the systems will observe a car on a patch, it will store the rough direction of the car in a histogram and consider the most ”popular” direction(s) of each patch in order to create virtual lanes. These lanes will not only replicate the ones explicitly shown by the road markings, but will also have the advantage to include paths to follow in less structured environments, such as intersections. This approach will be very useful for ego-car navigation as well as for the prediction of other vehicle’s motion, which will improve safety oriented ADAS and potentially refine our own system, since it will be possible to include these directions in formula (5).

Another advantage provided by the driving directions is that our system will be able to detect, and react to, sudden changes of vehicles’ behavior. In particular, it would be possible to label a road area as temporarily unavailable if we observe cars avoiding it by disregarding the driving directions. The unavailability of a road area will be another property included in our representation, and it will be useful in all those situations where the viability changes for any reason. An example is provided in figure 3: the vehicles are observed disregarding the previously stored driving direction, consistently avoiding an occluded area. The system then marks the area as unavailable and will be able to plan a new maneuver to avoid it.

IV. Experiments

In this section we apply our approach to real-world scenes taken from the KITTI benchmark[11]. Unfortunately, a quantitative evaluation of this approach is problematic for various

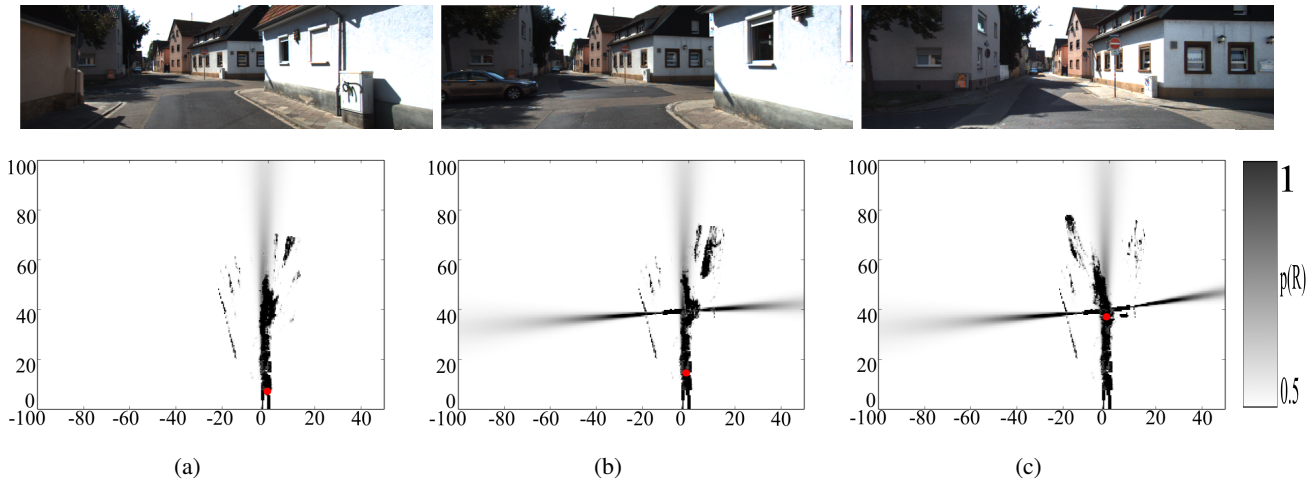


Fig. 4: On top, three frames taken from KITTI stream Nr 0046. On the bottom, the corresponding representation maps. All axis are in [m]. The red mark represents the ego-car position.

reasons. First of all, it is not trivial to get a ground truth for occluded road, although it can be done heuristically. Secondly, our approach is dependent on external events (appearance, behavior and abundance of other vehicles), so it is not trivial to come up with a proper evaluation method. We will deal with these issues iteratively. Here, we will illustrate the potential of this approach through qualitative examples. In particular, we will show that our approach can detect a road various seconds in advance, even roads that would not be detected at all by a visual road detector. Secondly, we will show how even in very simple scenes our approach can detect large portions of road at a much longer range ($> 40m$) than current visual road detectors. For clarity, we have not used the contribution of navigational maps, nor the negative effect of $p(R|C_i)$, that was unnecessary in this concrete examples. Hence, all the probability values lie between 0.5 and 1. For the detection of the vehicles, we used the data provided by the KITTI benchmark tracklets.

Figure 4 shows a very common inner city scene: buildings on both sides hiding an intersecting road. In this case the ego-car (even though traveling at very low speed) is unable to properly detect the intersecting road by using only the visual road detector (figure 4a). Eventually, just before entering the intersection, a car appears from the left road, directed towards the right. This event alone is sufficient for our system to infer the presence of an intersecting road, as shown by the representation map (figure 4b). The last frame shows the scene three seconds after the appearance of that vehicle: the ego-car is in the middle of the intersection, turning left. Only in that moment, when the maneuver has already begun, the SPRAY-based algorithm can start to detect the left road, which up to that point has always been occluded (figure 4c).

Figure 5 shows a different scene: a wide two way road, where the FOV of the ego-car is partially occluded by another car in front of it at close distance. That car renders the visual road terrain detection system useless beyond its position (15m from the ego-car). Here we want to show that our approach allows us to detect large portions of the road beyond the occluding vehicle. Consequently, the representation shown does not include the visual detection results, but only the effects of the behavior interpretation. In figure 5a, the ego-car comes from a smooth left bend, where it has briefly observed two cars in the opposing lane. The observation of these two cars (coupled with their speed) generates a permanent probability contribution which allows the system to be aware of the road ahead of it for about 100m. This contribution is also complemented by the estimated trajectory of the vehicle directly in front of us. In figure 5b two other cars have appeared in the opposing lane, extending our representation of the road up to about about 150m ahead. This extended horizon is due to the higher speed of these two cars with respect to the previous ones. In figure 5c, another car is observed, extending again the represented road.

V. Conclusions

We illustrated a spatial representation based on the interpretation of the behavior of traffic participants, which can also incorporate information from two other sources, visual road detectors and navigational maps. We showed that the behavior interpretation approach fits together with these two sources, serving as a middle level of spatial description between them, and how the system is planned to develop into a semantic representation of the environment. Finally, we presented qualitative examples taken from real-world data that show the advantages of this spatial representation, that can provide information for motion planning.

Future developments will focus on further testing and evaluation of our system, as well as on increasing its complexity. Our ultimate target is to create a semantic representation of the environment, by adding more and more properties above the layer road/non-road presented in this paper.

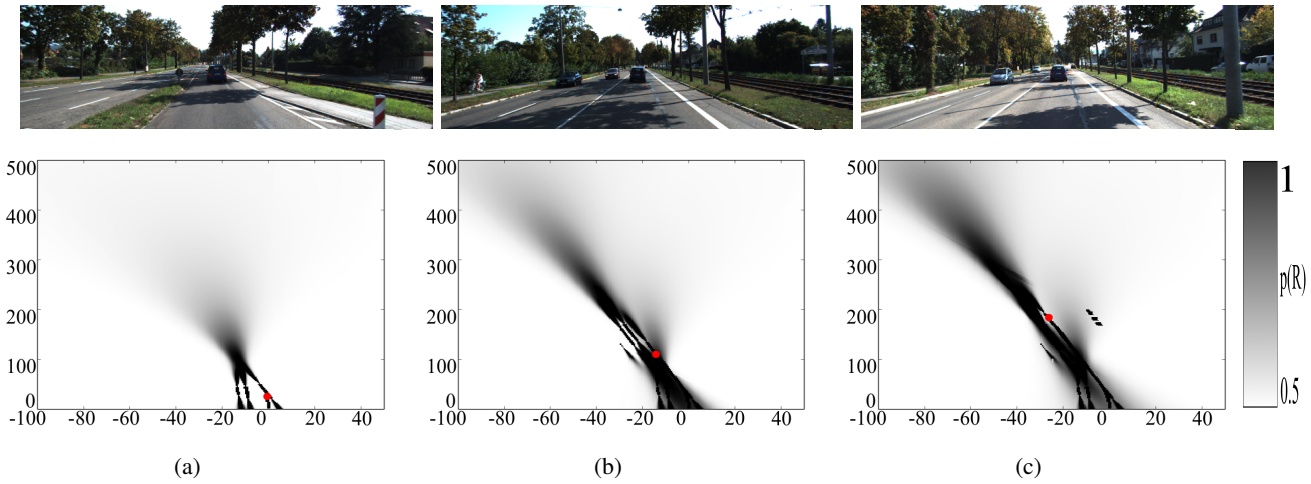


Fig. 5: On top, three frames taken from KITTI stream Nr 0056. On the bottom, the corresponding representation maps. All axis are in [m]. The red mark represents the ego-car position. This representation is composed without input from the visual road detection.

REFERENCES

[1] Andreas Alin, Martin V. Butz, and Jannik Fritsch. Tracking moving vehicles using an advanced grid-based Bayesian filter approach. *2011 IEEE Intell. Veh. Symp.*, (Iv):466–472, June 2011.

[2] Jose M. Alvarez, Antonio M. Lopez, Theo Gevers, and Felipe Lumbreras. Combining Priors, Appearance, and Context for Road Detection. *IEEE Trans. Intell. Transp. Syst.*, 15(3):1168–1178, June 2014.

[3] Andrew Bacha, Cheryl Bauman, Ruel Faruque, Michael Fleming, Chris Terwelp, Charles Reinholtz, Dennis Hong, Al Wicks, Thomas Alberi, David Anderson, Stephen Cacciola, Patrick Currier, Aaron Dalton, Jesse Farmer, Jesse Hurdus, Shawn Kimmel, Peter King, Andrew Taylor, David Van Covern, and Mike Webster. Odin: Team VictorTango's entry in the DARPA Urban Challenge. *J. F. Robot.*, 25(8):467–492, August 2008.

[4] Hernan Badino, Uwe Franke, and Rudolf Mester. Free Space Computation Using Stochastic Occupancy Grids and Dynamic Programming. In *ICCV 2007 Work. Dyn. Vis.*, Rio de Janeiro, Brazil, 2007.

[5] Sarah Bonnin, Thomas Weisswange, Franz Kummert, and Jens Schmüdderich. Pedestrian Crossing Prediction using Multiple Context-based Models. In *17th Int. IEEE Conf. Intell. Transp. Syst.* IEEE, October 2014.

[6] Mohamed Essayed Bouzouraa and Ulrich Hofmann. Fusion of occupancy grid mapping and model based object tracking for driver assistance systems using laser and radar sensors. In *IEEE Intell. Veh. Symp. (IV 2010)*, pages 294–300, San Diego, USA, June 2010. IEEE.

[7] Mark Campbell, Magnus Egerstedt, Jonathan P How, and Richard M Murray. Autonomous driving in urban environments: approaches, lessons and challenges. *Philos. Trans. A. Math. Phys. Eng. Sci.*, 368(1928):4649–72, October 2010.

[8] Radu Danescu, Florin Oniga, and Sergiu Nedevschi. Modeling and Tracking the Driving Environment With a Particle-Based Occupancy Grid. *IEEE Trans. Intell. Transp. Syst.*, 12(4):1331–1342, December 2011.

[9] Maryam Fatemi, Lars Hammarstrand, Lennart Svensson, and Ángel F García-Fernández. Road Geometry Estimation Using a Precise Clothoid Road Model and Observations of Moving Vehicles. In *17th IEEE Int. Conf. Intell. Transp. Syst. (ITSC 2014)*, pages 238–244, Qingdao, China, 2014. IEEE.

[10] Jannik Fritsch, Tobias Kühnl, and Franz Kummert. Monocular Road Terrain Detection by Combining Visual and Spatial Information. *IEEE Trans. Intell. Transp. Syst.*, 2014.

[11] A Geiger, P Lenz, and R Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR 2012)*, pages 3354–3361. IEEE, June 2012.

[12] Andreas Geiger, Martin Lauer, Christian Wojek, Christoph Stiller, and Raquel Urtasun. 3D Traffic Scene Understanding from Movable Platforms. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–14, September 2013.

[13] Tobias Gindele, Sebastian Brechtel, Joachim Schröder, and Rüdiger Dillmann. Bayesian Occupancy Grid Filter for Dynamic Environments Using Prior Map Knowledge. pages 669–676, 2009.

[14] Jungwon Kang and Myung Jin Chung. Stereo-vision based free space and obstacle detection with structural and traversability analysis using probabilistic volume polar grid map. *2011 IEEE 5th Int. Conf. Robot. Autom. Mechatronics*, pages 245–251, 2011.

[15] Marek Kurdej, Julien Moras, Véronique Cherfaoui, and Philippe Bonnifait. Map-Aided Evidential Grids for Driving Scene Understanding. *IEEE Intelligent Transportation Systems Magazine*, 7(1):30–41, 2015.

[16] Martin Liebner, Michael Baumann, Felix Klanner, and Christoph Stiller. Driver intent inference at urban intersections using the intelligent driver model. In *IEEE Intell. Veh. Symp. (IV 2012)*, pages 1162–1167, Alcalá de Henares, Spain, June 2012. Ieee.

[17] Isaac Miller, Mark Campbell, Dan Huttenlocher, Frank-Robert Kline, Aaron Nathan, Sergei Lupashin, Jason Catlin, Brian Schimpf, Pete Moran, Noah Zych, Ephrahim Garcia, Mike Kurdziel, and Hikaru Fujishima. Team Cornell's Skynet: Robust perception and planning in an urban environment. *J. F. Robot.*, 25(8):493–527, August 2008.

[18] Simon T. O'Callaghan, Surya P. N. Singh, Alen Alempijevic, and Fabio T. Ramos. Learning navigational maps by observing human motion patterns. In *IEEE Int. Conf. Robot. Autom. (ICRA 2011)*, pages 4333–4340, Shanghai, China, May 2011. IEEE.

[19] Sayanan Sivaraman and Mohan Manubhai Trivedi. Dynamic Probabilistic Drivability Maps for Lane Change and Merge Driver Assistance. *IEEE Intell. Transp. Syst. Mag.*, pages 1–11, 2014.

[20] Marvin Struwe, Stephan Hasler, and Ute Bauer-Wersing. A two-stage classifier architecture for detecting objects under real-world occlusion patterns. In *International Conference on Artificial Neural Networks ICANN 2014*. Marvin Struwe, Stephan Hasler, and Ute Bauer-Wersing, 2014.

[21] Chris Urmson, Joshua Anhalt, Drew Bagnell, ..., Michael Darms, and Dave Ferguson. Autonomous driving in urban environments: Boss and the Urban Challenge. *J. F. Robot.*, 25(8):425–466, August 2008.

[22] Thorsten Weiss, Bruno Schiele, and Klaus Dietmayer. Robust Driving Path Detection in Urban and Highway Scenarios Using a Laser Scanner and Online Occupancy Grids. In *IEEE Intell. Veh. Symp. (IV 2007)*, pages 184–189, Istanbul, Turkey, June 2007. IEEE.

[23] Thomas H. Weisswange, Bram Bolder, Jannik Fritsch, Stephan Hasler, and Christian Goerick. An integrated ADAS for assessing risky situations in urban driving. In *Intell. Veh. Symp. (IV), 2013 IEEE*, pages 292–297, 2013.