

Situational awareness and road prediction for trajectory control applications

Christian Lundquist, Thomas B. Schön, Fredrik Gustafsson

Abstract This chapter is concerned with the problem of estimating a map of the immediate surroundings of a vehicle as it moves. In order to compute these map estimates sensor measurements from radars, lasers and/or cameras are used together with the standard proprioceptive sensors present in a car. Four different types of maps are discussed. Feature based maps are represented by a set of salient features. Road maps make use of the fact that roads are highly structured, since they are built according to clearly specified road construction standards, which allows relatively simple and powerful models of the road to be employed. Location based maps, where occupancy grid maps belong and finally intensity based maps, which can be viewed as a continuous version of the location based maps. The aim is to provide a self-contained presentation of how these maps can be built from measurements. Real data from Swedish roads are used throughout the chapter to illustrate the methods introduced.

1 Introduction

Most automotive original equipment manufacturers today offer longitudinal control systems, such as adaptive cruise control (ACC) or collision mitigation systems. Lateral control systems, such as lane keeping assistance (LKA), emergency lane assist

Christian Lundquist

Department of Electrical Engineering, Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden e-mail: lundquist@isy.liu.se

Thomas B. Schön

Department of Electrical Engineering, Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden e-mail: schon@isy.liu.se

Fredrik Gustafsson

Department of Electrical Engineering, Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden e-mail: fredrik@isy.liu.se

(ELA) (Eidehall et al. 2007) and curve speed warning, are currently developed and released. These systems can roughly be split into safety applications, which aim to mitigate vehicular collisions such as rear end or blind spot detection; and comfort applications such as ACC and LKA, which aim at reducing the driver's work load. The overview article by Caveney (2010) describes the current development of trajectory control systems. The requirements on the position accuracy of the ego vehicle in relation to other vehicles, the road and the surrounding environment increases with those control applications that are currently under development and expected to be introduced to the market.

The systems available or in development today are based on two basic tracking and decision principles: longitudinal systems use a radar, possibly supported by a camera, to track leading vehicles and they decide on braking warnings or interventions. On the other hand, lateral systems use a camera to track the lane markers and they decide on steering warnings or interventions. Future safety and comfort functions need more sophisticated situational awareness and decision functionality:

- A combination of lateral and longitudinal awareness will be needed, where all lanes are monitored, all of their vehicles are tracked, and the road-side conditions are modeled to allow for emergency maneuvers. The result is a situational awareness map, which is the topic for this chapter.
- This will allow for more sophisticated decision functionality. First, the possible evasive driver maneuvers are computed, and only if the driver has no or very little time for evasive actions, the system will intervene. Second, more complex automatic evasive maneuvers can be planned using the situational awareness map, including consecutive lateral and braking actions.

It should be remarked that the accuracy of the navigation systems today and in the near future, see Chapters "In-car Navigation Basics" and "The evolution of in-car navigation systems", are not of much assistance for situational awareness. The reason is that satellite based navigation gives an accuracy of 10-20 meters, which is not sufficient for lateral awareness. Even in future systems, including reference base stations, enabling meter accuracy, the standard road maps will limit the performance since they are not of sufficient accuracy. Thus, two leaps in development are needed before positioning information and standard maps can be used to improve situational awareness maps. Another technical enabler is car to car communication (C2C), which may improve tracking of other vehicles and in the end change the transportation system as has already been done with the transponder systems for aircraft and commercial surface ships. Still, there will always be vehicles and obstacles without functioning communication systems. The need for accurate situation awareness and road prediction to be able to automatically position the car in a lane and derive drivable trajectories will evolve and remain important.

The different types of situation awareness maps used to represent the environment are introduced in Section 2. Details of these maps are presented in Sections 3 to 6. The chapter is concluded in Section 7.

2 Modeling the Environment with a Map

The transportation system may be described and represented by a number of variables. These variables include state variables describing the position, orientation, velocity and size of the vehicles. Here one can distinguish between the own vehicle, the so called ego vehicle, and the other vehicles, referred to as the targets.

The state variable of the ego vehicle at time k is denoted $\mathbf{x}_{E,k}$. The trajectory of the ego vehicle is recorded in $\mathbf{x}_{E,1:k} = \{\mathbf{x}_{E,1}, \dots, \mathbf{x}_{E,k}\}$, and it is assumed to be a priori known in this work. This is a feasible assumption, since the absolute trajectory in world coordinates and the relative position in the road network are separable problems.

The state variable of the targets at time k is denoted $\mathbf{x}_{T,k}$. The road and the environment may be modeled by a map, which is represented by a set of variables describing N_m landmarks in the environment according to

$$\mathbf{M}_k = \left\{ \mathbf{m}_k^{(1)}, \mathbf{m}_k^{(2)}, \dots, \mathbf{m}_k^{(N_m)} \right\}. \quad (1)$$

According to Thrun et al. (2005), there exists primarily two types of indexing for probabilistic maps. In a *feature based map* each $\mathbf{m}^{(n)}$ specifies the properties and location of one object, whereas in a *location based map* the index n corresponds to a location and $\mathbf{m}^{(n)}$ is the property of that specific coordinate. The *occupancy grid map* is a classical location based representation of a map, where each cell of the grid is assigned a binary occupancy value that specifies if the location n is occupied ($\mathbf{m}^{(n)} = 1$) or not ($\mathbf{m}^{(n)} = 0$), see e.g., Elfes (1987), Moravec (1988).

The ego vehicle perceives information about the other vehicles and the environment through its sensors. The sensors provide a set of noisy measurements

$$\mathbf{Z}_k = \left\{ \mathbf{z}_k^{(1)}, \mathbf{z}_k^{(2)}, \dots, \mathbf{z}_k^{(N_{z,k})} \right\} \quad (2)$$

at each discrete time instant $k = 1, \dots, K$. Common sensors used for automotive navigation and mapping measure either range and bearing angle, as for example radar and laser, or bearing and elevation angles, as for the case of a camera. A signal pre-processing is always included in automotive radar sensors and the sensor provides a list of detected features, defined by the range r , range rate \dot{r} and bearing ψ . The pre-processing, the waveform design and the detection algorithms of the radar is well described by e.g., Rohling & Möller (2008), Rohling & Meinecke (2001). Laser sensors typically obtain one range measurement per beam, and there exists both sensors which emit several beams at different angles and those which have a rotating beam deflection system. They all have in common that the angles at which they measure range are quantized, thus providing a list of range and bearings of which only the ones which are less than the maximum range shall be considered. Another commonly used automotive sensor is the camera. The camera measurements are quantized and the data is represented in a pixel matrix as an image.

Note that the indexing of sensor data is analogous to the representation of maps. Each range and bearing measurement $\mathbf{z}^{(n)}$ from a radar or laser specifies the properties and location of one observation, i.e., it is a feature based measurement. However, the indexing of camera measurement is location based, since the index n corresponds to a pixel in the image and $\mathbf{z}^{(n)}$ is the property of that specific coordinate.

The aim of all stochastic mapping algorithms, independent of indexing, is to estimate the posterior density of the map

$$p(\mathbf{M}_k | \mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k}), \quad (3)$$

given all the measurements $\mathbf{Z}_{1:k}$ from time 1 to k and the trajectory of the ego vehicle $\mathbf{x}_{E,1:k}$. The conditioning on $\mathbf{x}_{E,1:k}$ is implicit in this chapter, since it is assumed to be a priori known. To be able to estimate the map, a relation between the map and the measurements must first be found and modeled. This model h is referred to as the measurement equation and for one combination of a measurement $\mathbf{z}_k^{(i)}$ and a map variable $\mathbf{m}_k^{(j)}$ it may be written according to

$$\mathbf{z}_k^{(i)} = h(\mathbf{m}_k^{(j)}) + \mathbf{e}_k, \quad (4)$$

where \mathbf{e}_k is the measurement noise. The primary aim in this chapter is to create a momentary map of the environment currently surrounding the ego vehicle. Hence, it is just the present map data that is recorded in the vehicle. As soon as a part of the environment is sufficiently far from the ego vehicle the corresponding map entries are deleted. Environmental models must be compact, so that they can be transmitted to and used efficiently by other automotive systems, such as path planners. The maps must be adapted to the type of environment they aim to model. For this reason four different map representations, which are relevant for modeling the environment surrounding a vehicle, are described in the following sections.

| | |
|--------------------|---|
| Feature based map | The map is represented by a number of salient features in the scene. Feature representation and tracking as part of a map is described in Section 3. |
| Road map | This is a special case of the feature based map, where the map variables model the geometry of the road. Roads are highly structured; they are built according to road construction standards and contain primarily, straight lines, curves and clothoids. Maps of road lanes and edges are described in Section 4. |
| Location based map | One of the most well established location based map is the occupancy grid map, which is described in Section 5. The map is defined over a continuous space, but discretized with a grid approximation. |

Intensity based map The intensity density may be interpreted as the probability that one object is located in an infinitesimal region of the state space. The intensity based map is a continuous approximation of a location based map and it is described in Section 6.

The estimated maps can be used to increase the localization accuracy of the ego vehicle with respect to its local environment. Furthermore, the maps may be used to derive a trajectory, which enables a collision free path of the vehicle.

3 Feature Based Map

Features corresponding to distinct objects in the physical world, such as tree trunks, corners of buildings, lampposts and traffic signs are commonly denoted *landmarks*. The procedure of extracting features reduces the computational complexity of the system, as the features are on a more compact format than the original measurement. The form of the measurement equations (4) depends on the type of sensor used, and the signals measured by the sensor. In this section we will briefly describe the use of features and the corresponding measurement equations in both radar and laser sensors (Section 3.1) as well as cameras (Section 3.2).

The feature based approach may together with existing road maps be used to supplement the GPS-based position of the vehicle. This approach is also commonly referred to as visual odometry, see e.g., Nistér et al. (2006).

3.1 Radar and Laser

As mentioned in the introduction, radar and laser sensors measure at least range and bearing of the landmark relative to the vehicles local coordinate, i.e., the measurement vector is composed of

$$\mathbf{z}^{(i)} = [r^{(i)} \ \psi^{(i)}]^T. \quad (5)$$

Notice that, for the sake of simplicity, the subscripts k specifying the time stamps of the quantities is dropped throughout this chapter. The assumption will be made that the measurements of the features are independent, i.e., the noise in each individual measurement $\mathbf{z}^{(i)}$ is independent of the noise in the other measurements $\mathbf{z}^{(j)}$, for $i \neq j$. This assumption makes it possible to process one feature at a time in the algorithms. Assume that the ego vehicle pose is defined by

$$\mathbf{x}_E = [x_E \ y_E \ \psi_E]^T, \quad (6)$$

where x_E, y_E denote the horizontal position of the vehicle and ψ_E denotes the heading angle of the vehicle. Furthermore, let us assume that one feature j in the map is defined by its Cartesian coordinate,

$$\mathbf{m}^{(j)} = \begin{bmatrix} x_m^{(j)} & y_m^{(j)} \end{bmatrix}^T. \quad (7)$$

The measurement model (4) is then written as

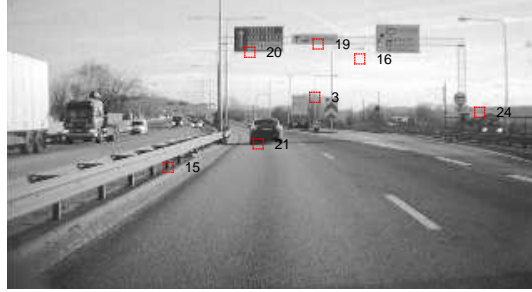
$$\begin{bmatrix} r^{(i)} \\ \psi^{(i)} \end{bmatrix} = \begin{bmatrix} \sqrt{(x_m^{(j)} + x_E)^2 + (y_m^{(j)} + y_E)^2} \\ \arctan \frac{y_m^{(j)} - y_E}{x_m^{(j)} + x_E} - \psi_E \end{bmatrix} + \begin{bmatrix} e_r \\ e_\psi \end{bmatrix}, \quad (8)$$

where e_r and e_ψ are noise terms. The i^{th} measurement feature corresponds to the j^{th} map feature. The data association problem arises when the correspondence between the measurement feature and the map feature cannot be uniquely identified. A correspondence variable, which describes the relation between measurements and map features is introduced. This variable is also estimated at each time step k and there exists a number of different algorithms to do this, see e.g., Blackman & Popoli (1999). There are quite a few different possibilities on how to define features when radars and lasers are used.

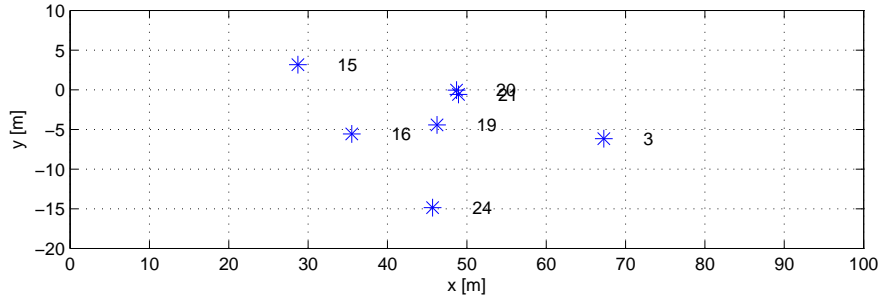
3.2 Cameras and Computer Vision

Features form the bases in many computer vision algorithms, especially when it comes to building maps. There exists a myriad of feature detectors, which extract edges, corners or other distinct patterns. Some of the most well know are the Harris corner detector (Harris & Stephens 1988), SIFT (Lowe 2004) and MSER (Matas et al. 2004), see Fig. 1 for an example, where the Harris corner detector is used. For a more complete account of various features used, see e.g., Szeliski (2010). Using features to build maps from camera images has been studied for a long time and a good account of this is provided by Davison et al. (2007).

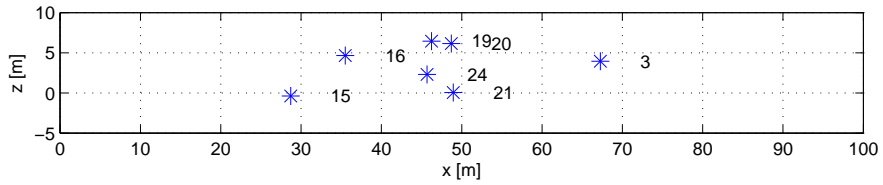
A key component in building maps using features is a good mathematical description of how the features detected in the image plane are related to the corresponding positions in world coordinates. The distance (commonly referred to as the depth) to a landmark cannot be determined from a single image and this fact should be encoded by the mathematical parameterization of the landmark. The so called *inverse depth parameterization* by Civera et al. (2008) provides an elegant uncertainty description of the fact that the depth (i.e., distance) to the landmark is unknown. Here, the landmark (lm) state vector is given by



(a)



(b)



(c)

Fig. 1 Features detected using the Harris corner detector are shown in Fig. (a). Fig. (b) and (c) shows the estimated position of the landmarks in the $x - y$ and $x - z$ plane, respectively.

$$\mathbf{m}^{(j)} = \begin{bmatrix} \mathbf{c}^{(j)} \\ \psi^{(j)} \\ \phi^{(j)} \\ \rho^{(j)} \end{bmatrix} = \begin{bmatrix} \text{camera position first time lm was seen} \\ \text{azimuth angle of lm seen from } \mathbf{c}^{(j)} \\ \text{elevation angle of lm seen from } \mathbf{c}^{(j)} \\ \text{inverse distance (depth) from } \mathbf{c}^{(j)} \text{ to lm} \end{bmatrix} \quad (9a)$$

$$\mathbf{c}^{(j)} = [x^{(j)} \ y^{(j)} \ z^{(j)}]^T, \quad (9b)$$

where $\mathbf{c}^{(j)}$ is the position of the camera expressed in world coordinates at the time when landmark j was first seen, $\psi^{(j)}$ is the azimuth angle of the landmark as seen from $\mathbf{c}^{(j)}$, relative to the world coordinate frame. The elevation angle of the landmark as seen from $\mathbf{c}^{(j)}$, relative to world coordinate frame directions is denoted $\phi^{(j)}$,

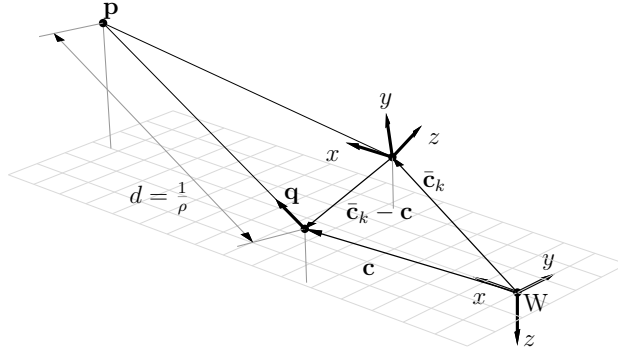


Fig. 2 The inverse depth parameterization used for the landmarks. The position of the landmark \mathbf{p} is parameterized using the position \mathbf{c} of the camera the first time the feature was seen, the direction $\mathbf{q}(\phi, \psi)$ and the inverse depth ρ . The position of the camera at time step k is denoted $\bar{\mathbf{c}}_k$.

and the inverse depth, which is the inverse of the distance, from $\mathbf{c}^{(j)}$ to the landmark is denoted $\rho^{(j)}$.

The landmark state $\mathbf{m}^{(j)}$ is a parametrization of the Cartesian position $\mathbf{p}^{(j)}$ of landmark j , see Fig. 2. The relationship between the position of the landmark and the inverse depth state representation is given by

$$\mathbf{p}^{(j)} = \mathbf{c}_k^{(j)} + \underbrace{\frac{1}{\rho^{(j)}} \begin{bmatrix} \cos \phi^{(j)} \cos \psi^{(j)} \\ \cos \phi^{(j)} \sin \psi^{(j)} \\ \sin \phi^{(j)} \end{bmatrix}}_{\mathbf{q}^{(j)}}. \quad (10)$$

The measurement model (4) for the landmarks is given by

$$h(\mathbf{m}^{(j)}) = \mathcal{P}_n(\mathbf{p}^{C,(j)}) = \frac{1}{x_p^{C,(j)}} \begin{bmatrix} y_p^{C,(j)} \\ z_p^{C,(j)} \end{bmatrix}, \quad (11)$$

where $\mathcal{P}_n(\mathbf{p}^{C,(j)})$ is used to denote the normalized pinhole projection and $\mathbf{p}^{C,(j)} = \begin{bmatrix} x_p^{C,(j)} & y_p^{C,(j)} & z_p^{C,(j)} \end{bmatrix}^T$ denotes the position of feature j at time k in the camera coordinate frame C . Note that before an image position can be used as a measurement together with the measurement equation (11), the image position is adjusted according to the camera specific parameters, such as focal length, pixel sizes, etc. The transformation between pixel coordinates $\begin{bmatrix} u & v \end{bmatrix}^T$ and normalized camera coordinates $\begin{bmatrix} y & z \end{bmatrix}^T$, which is the kind of coordinates landmark measurements $\mathbf{z}^{(i)}$ (see (11)) are given in, is

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \frac{u - u^{ic}}{f_u} \\ \frac{v - v^{ic}}{f_v} \end{bmatrix}, \quad (12)$$

where $[\mathbf{u}^{\text{ic}} \ \mathbf{v}^{\text{ic}}]^T$ denotes the image center, and f_u and f_v are the focal lengths (given in pixels) in the lateral u -direction and the vertical v -direction, respectively. The transformation between world and camera coordinates is given by

$$\mathbf{p}^{C,(j)} = R^{CE} \left(R^{EW} \left(\left(\mathbf{c}^{(j)} - \begin{bmatrix} x_E \\ y_E \end{bmatrix} \right) + \frac{1}{\rho^{(j)}} \mathbf{q}^{(j)} \right) - \mathbf{c}^E \right), \quad (13a)$$

$$R^{CE} = R(\psi^C, \phi^C, \gamma^C), \quad (13b)$$

$$R^{EW} = R(\psi_E, \phi_E, \gamma_E), \quad (13c)$$

where \mathbf{c}^E denotes the position of the camera expressed in the ego vehicle body coordinate frame. The rotation matrix $R(\alpha, \beta, \gamma)$ transforms coordinates from coordinate frame B to coordinate frame A , where the orientation of B relative to A is ψ (yaw), ϕ (pitch) and γ (roll). Furthermore, ψ^C , ϕ^C and γ^C are the constant yaw, pitch and roll angles of the camera, relative to the vehicle body coordinate frame.

The landmarks are estimated recursively using e.g., a Kalman filter. An example of estimated landmarks is shown in Fig. 1. The estimated position $\mathbf{p}^{(j)}$ for seven landmarks is shown in the image plane, as well as in the world $x-y$ and $x-z$ plane, where the ego vehicle is in origin.

4 Road Map

A road map describe the shape of the road. The roads are mainly modeled using polynomial functions which describe the lane and the road edges. The advantages of road models are that they require sparse memory and are still very accurate, since they do not suffer from discretization problems. General road models are presented in Section 4.1. Lane estimation using camera measurements is described in Section 4.2 and finally road edge estimation based on feature measurements is described in Section 4.3.

4.1 Road Model

The road, as a construction created by humans, possesses no dynamics; it is a static time invariant object in the world coordinate frame. The building of roads is subject to road construction standards, hence, the modeling of roads is geared to these specifications. However, if the road is described in the ego vehicle's coordinate frame and the vehicle is moving along the road it is possible and indeed useful to describe the characteristics of the road using time varying state variables.

A road consists of straight and curved segments with constant radius and of varying length. The sections are connected through transition curves, so that the driver can use smooth and constant steering wheel movements instead of stepwise changes

when passing through road segments. More specifically, this means that a transition curve is formed as a clothoid, whose curvature c changes linearly with its curve length x_c according to

$$c(x_c) = c_0 + c_1 \cdot x_c. \quad (14)$$

Note that the curvature c is the inverse of the radius. Now, suppose x_c is fixed to the ego vehicle, i.e., $x_c = 0$ at the position of the ego vehicle. When driving along the road and passing through different road segments c_0 and c_1 will not be constant, but rather time varying state variables

$$\mathbf{m} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} \text{curvature at the ego vehicle} \\ \text{curvature derivative} \end{bmatrix}. \quad (15)$$

In section 3 the map features were expressed in a fixed world coordinate frame. However, note that in this section the road map is expressed as seen from the moving ego vehicle. Using (14), a change in curvature at the position of the vehicle is given by

$$\left. \frac{dc(x_c)}{dt} \right|_{x_c=0} = \dot{c}_0 = \frac{dc_0}{dx_c} \cdot \frac{dx_c}{dt} = c_1 \cdot v, \quad (16)$$

where v is the ego vehicle's velocity. Furthermore, the process model is given by

$$\begin{bmatrix} \dot{c}_0 \\ \dot{c}_1 \end{bmatrix} = \begin{bmatrix} 0 & v_x \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} + \begin{bmatrix} 0 \\ w_{c_1} \end{bmatrix}. \quad (17)$$

This model is referred to as the *simple clothoid model* and it is driven by the process noise w_{c_1} . Note that the road is modeled in a road aligned coordinate frame, with the components (x_c, y_c) , and the origin at the position of the ego vehicle. There are several advantages of using road aligned coordinate frames, especially when it comes to the process models of the other vehicles on the same road, these models are greatly simplified in road aligned coordinates. However, the flexibility of the process model is reduced and basic dynamic relations such as Newton's and Euler's laws cannot be directly applied. The road model (14) is transformed into Cartesian coordinates (x, y) using

$$x(x_c) = \int_0^{x_c} \cos(\chi(x)) dx \approx x_c, \quad (18a)$$

$$y(x_c) = \int_0^{x_c} \sin(\chi(x)) dx \approx \frac{c_0}{2} x_c^2 + \frac{c_1}{6} x_c^3, \quad (18b)$$

where the heading angle χ is defined as

$$\chi(x) = \int_0^x c(\lambda) d\lambda = c_0 x + \frac{c_1}{2} x^2. \quad (18c)$$

The origin of the two frames is fixed to the ego vehicle, hence, integration constants (x_0, y_0) are omitted.

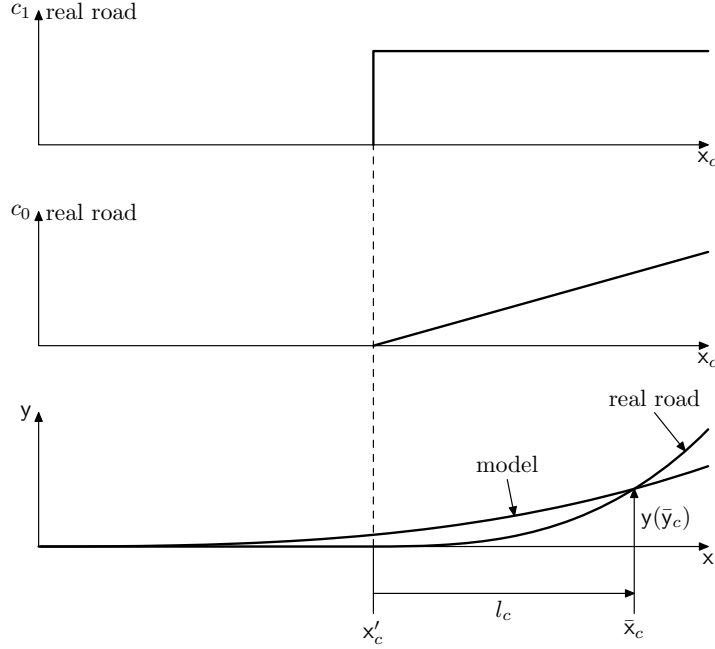


Fig. 3 A straight and a curved road segment are modeled with the averaging road model. The two upper plots show the parameters c_1 and c_0 of the real road, the bottom plot shows the real and the modeled roads in a Cartesian coordinate frame.

A problem appears when two or more clothoid segments, with different parameters c_0 and c_1 , are observed in the same camera view. The parameter c_0 will change continuously during driving, whereas c_1 will be constant in each segment and change stepwise at the segment transition. This leads to a dirac impulse in \dot{c}_1 at the transition. The problem can be solved by assuming a high process noise w_{c_1} in (17), but this leads to less precise estimates of the state variables when no segment transitions occur in the camera view. To solve this problem an averaging curvature model was proposed by Dickmanns (1988), which is perhaps best described with an example. Assume that the ego vehicle is driving on a straight road (i.e., $c_0 = c_1 = 0$) and that the look ahead distance of the camera is \bar{x}_c . A new segment begins at the position $x'_c < \bar{x}_c$, which means that there is a step in c_1 , and c_0 is ramped up, see Fig. 3. The penetration into the next segment is $l_c = \bar{x}_c - x'_c$. The idea of this model, referred to as *averaging* or *spread-out dynamic curvature model*, with the new state variables c_{0m} and c_{1m} , is that it generates the true lateral offset $y(\bar{x}_c)$ at the look ahead distance \bar{x}_c , i.e.,

$$y_{\text{real}}(\bar{x}_c) = y_{\text{model}}(\bar{x}_c), \quad (19)$$

but it is continuously spread out in the range $(0, \bar{x}_c)$. The lateral offset of the real road as a function of the penetration l_c , for $0 \leq l_c \leq \bar{x}_c$, is

$$y_{\text{real}}(l_c) = \frac{c_1}{6} l_c^3, \quad (20)$$

since the first segment is straight. The lateral offset of the averaging model as a function of the penetration l_c is

$$y_{\text{model}}(l_c) = \frac{c_{0m}(l_c)}{2} \bar{x}_c^2 + \frac{c_{1m}(l_c)}{6} \bar{x}_c^3, \quad (21)$$

at the look ahead distance \bar{x}_c . The equation

$$c_1 \frac{l_c^3}{\bar{x}_c^2} = 3c_{0m}(l_c) + c_{1m}(l_c) \bar{x}_c, \quad (22)$$

is obtained by inserting (20) and (21) into (19). By differentiating (22) with respect to l_c and using the relations $\frac{dc_1}{dl_c} = 0$, $\frac{dc_{0m}(l_c)}{dl_c} = c_{1m}(l_c)$ and $\frac{d(\cdot)}{dl_c} = \frac{d(\cdot)}{dt} \cdot \frac{dt}{dl_c}$ the following equation is obtained

$$\dot{c}_{1m} = 3 \frac{v}{\bar{x}_c} (c_1 (l_c / \bar{x}_c)^2 - c_{1m}), \quad (23)$$

for $l_c < \bar{x}_c$. Since $(l_c / \bar{x}_c)^2$ is unknown it is usually set to 1 (Dickmanns 2007), which finally yields

$$\dot{c}_{1m} = 3 \frac{v}{\bar{x}_c} (c_1 - c_{1m}). \quad (24)$$

The state variable vector of the averaging model is defined as

$$\mathbf{m} = \begin{bmatrix} c_{0m} \\ c_{1m} \\ c_1 \end{bmatrix} = \begin{bmatrix} \text{curvature at the ego vehicle} \\ \text{averaged curvature derivative} \\ c \text{ derivative of the foremost segment} \end{bmatrix}, \quad (25)$$

and the process model is given by augmenting the simple clothoid model (17) with (24) according to

$$\begin{bmatrix} \dot{c}_{0m} \\ \dot{c}_{1m} \\ \dot{c}_1 \end{bmatrix} = \begin{bmatrix} 0 & v & 0 \\ 0 & -3 \frac{v}{\bar{x}_c} & 3 \frac{v}{\bar{x}_c} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_{0m} \\ c_{1m} \\ c_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ w_{c_1} \end{bmatrix}. \quad (26)$$

The model is driven by the process noise w_{c_1} , which also influences the other states. The averaging model is well described in the recent textbook Dickmanns (2007) and some early results using the model are presented by e.g., Dickmanns & Mysliwetz (1992).

4.2 Mapping of the Road Lanes

The problem of mapping road lanes or lane estimation as it is often called, is a *curve estimation* problem. The task is to obtain the best possible estimate of the curve describing the lane by exploiting the measurements provided by the onboard sensors. The most important sensor type here is exteroceptive sensors, such as for example cameras and lasers. Currently the camera is the most commonly used sensor for the lane estimation problem and in this section we will show how camera images can be used to obtain lane estimates.

The lane estimation problem is by no means a new problem, it has been studied for more than 25 years, see e.g., Waxman et al. (1987), Dickmanns & Mysliwetz (1992) for some early work and Wang et al. (2008) for more recent contributions. A complete overview of what has been done on this problem is available in the survey paper McCall & Trivedi (2006). In this section the problem is broken down into its constituents and one way of solving the lane estimation problem when a camera is used as a sensor is shown.

The lane estimation problem can be separated into two subproblems, commonly referred to as the *lane detection* problem and the *lane tracking* problem. As the name reveals, the lane detection problem deals with detecting lanes in an image. The lane tracking problem then makes use of the detected lanes together with information about the temporal and the spatial dependencies over time in order to compute lane estimates. These dependencies are mathematically described using a road model. Traditionally, lane tracking is done using an extended Kalman filter, see e.g., Dickmanns & Mysliwetz (1992), Guiducci (1999). There are also interesting approaches based on the particle filter by Gordon et al. (1993) available, see e.g., Zhou et al. (2006), Wang et al. (2008), Kim (2008).

The lane is here modeled in the image plane (cf. Section 3.2) as a linear function close to the ego vehicle and as a quadratic function far away, i.e.,

$$\mathbf{l}_\theta(v) = \begin{cases} a + b(v - v_s) & v > v_s \\ a + b(v - v_s) + c(v - v_s)^2 & v \leq v_s, \end{cases} \quad (27a)$$

where v_s denotes the (known) vertical separation in pixels between the linear and the quadratic model (illustrated by the horizontal line in Fig. 4) and subindex θ is used to denote the dependence on the parameters

$$\theta = [a \ b \ c]^T, \quad (27b)$$

which are to be estimates. These parameters all have geometrical interpretations in terms of offset (a), local orientation (b) and curvature (c) of the lane in the image plane. The lane estimates (here, the estimates of the parameters θ in (27)) carry important information about the states in the road model introduced in Section 4.1, which are expressed in world coordinates (in contrast to pixel coordinates u , v). These road model states are typically what we are interested in and we will return to this important connection at the end of this section.

Given the fact that the problem of lane detection has been studied for more than 25 year there are many ideas on how to solve this problem available. Rather than trying to give a complete account of all the different methods available, we will here be very specific and explain one way in which lanes can be detected and show some results on real sequences. The solution presented here is very much along the lines of Jung & Kelber (2005), Lee (2002).

The initial lane detection is performed using a linear lane model, which is found from the image using a combination of the edge distribution function (EDF) (Lee 2002) and the Hough transform (Hough 1962, Duda & Hart 1972). The EDF is defined as the gradient orientation

$$\varphi(u, v) = \arctan\left(\frac{D_u}{D_v}\right), \quad (28)$$

where D_u and D_v are approximations of the gradient function

$$\nabla I(u, v) = \left[\frac{\partial I}{\partial u} \quad \frac{\partial I}{\partial v}\right]^T \approx [D_u \quad D_v]^T \quad (29)$$

for the gray scale image $I(u, v)$. The two largest peaks (α^l, α^r) of $\varphi(u, v)$ provide the most probable orientations of the lanes in the image. This is used to form an edge image $g(u, v)$ as

$$g(u, v) = \begin{cases} |\nabla I(u, v)| \approx |D_u| + |D_v|, & |\varphi(u, v) - \alpha^l| < T_\alpha \text{ or } |\varphi(u, v) - \alpha^r| < T_\alpha \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

where T_α is a threshold, here $T_\alpha = 2^\circ$. Applying the Hough transform to the edge image $g(u, v)$ provides two initial linear models $l(v) = a + bv$, one for the left lane markings and one for the right lane markings. These models are used to form a region which will serve as the search region in the subsequent image. This region, which we refer to as the lane boundary region of interest (LBROI) is simply defined by extending the linear model w pixels to the right and w pixels to the left (here $w = 10$).

Given that an initial LBROI is found, the task is now to make use of this information in order to compute estimates of the parameters in the lane model (27),

$$\theta = [(\theta^l)^T \quad (\theta^r)^T]^T, \quad \text{where} \quad \theta^l = [a^l \quad b^l \quad c^l]^T, \quad \theta^r = [a^r \quad b^r \quad c^r]^T, \quad (31)$$

where superscript l and r have been used to indicate the left lane marking and the right lane marking, respectively. Estimates of these parameters θ are obtained by solving a constrained weighted least squares problem for each image. The cost function is given by

$$V(\theta) = \sum_{i=1}^N \left(M_i^l(u_i^l - \mathbf{I}_{\theta^l}(v_i^l))^2 + (M_i^r(u_i^r - \mathbf{I}_{\theta^r}(v_i^r))^2 \right), \quad (32)$$

where N denotes the number of relevant pixels in the lateral u direction, \mathbf{l} denotes the lane model given in (27) and M_i denotes the magnitude in the thresholded edge-image $g_2(u, v)$, $M_i = g_2(u_i, v_i)$, defined as

$$g_2(u, v) = \begin{cases} g(u, v), & g(u, v) \geq \frac{1}{2}M_{\text{mean}} \\ 0, & \text{otherwise} \end{cases} \quad (33)$$

where M_{mean} denotes the mean magnitude of the $g(u, v)$.

Constraints are introduced in order to account for the fact that the right lane and the left lane are related to each other. This is modelled according to the following linear (in θ) inequality constraint

$$a^r - a^l + (b^r - b^l)(v_1 - v_m) + (c^r - c^l)(v_1 - v_m)^2 \leq \delta, \quad (34)$$

which states that the left and the right lanes cannot be more than δ pixels apart furthest away (at v_1) from the host vehicle. In other words, (34) encodes the fact that the left and the right lanes must have similar geometry in the sense that the quadratic parts in (27) are strongly related.

From (32)–(34) it is now clear that lane estimation boils down to a curve estimation problem, which here is quadratic in the unknown parameters θ . More specifically, inserting the lane model (27) into the cost function (32) and writing the problem on matrix form results in a constrained weighted least squares problem on the form

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2} \theta^T H \theta + f^T \theta \\ \text{s.t.} \quad & L \theta \leq \delta. \end{aligned} \quad (35)$$

This is a quadratic program (QP) implying that a global minimum will be found and there are very efficient solvers available for this type of problems. Here, we have made use of a dual active set method¹ according to Gill et al. (1991). An illustration of the lane estimation results is provided in Fig. 4. The estimate of θ is then used to form the LBROI for the new image, simply as region defined by $\mathbf{l}_{\theta}(v) \pm w$ for each lane.

The lane estimates that are now obtained as the solution to (35) can be expressed in world coordinates, seen from the ego vehicle, using geometrical transformation along the lines of what has already been described in Section 3. These transformations are discussed in detail in Guiducci (2000). Once the lane estimates are available in the world coordinates they can be used as camera measurements in a *sensor fusion* framework to make a very important contribution to the estimate of map variables \mathbf{m} (i.e., (15) or (24)) in the road model (perhaps most importantly the curvature c_0 and the curvature derivative c_1) as it is derived in Section 4.1.

¹ The QP code was provided by Dr. Adrian Wills at the University of Newcastle, Australia, see <http://sigpromu.org/quadprog>. This code implements the method described by Goldfarb & Idnani (1983), Powell (1985).



(a)



(b)

Fig. 4 Lane estimation results (in red) overlaid onto the camera image. From this figure the lane model (27) is clearly illustrated, the model is linear for $v > v_s$ and quadratic for $v \leq v_s$. The method assumes that the road surface is flat and when this assumption is true the results are good, see Fig. (a). However, when this assumption does not hold the estimates are not that good on a longer horizon, see Fig. (b).

4.3 Mapping of the Road Edges

Feature based measurements of landmarks along the road may be used to map the road edges. This section describes a method to track line shaped objects, such as guardrails using point measurements from radar, laser or extracted camera features. Tracking point objects, was covered in Section 3 and is not repeated here. The line shaped and curved guardrails are described using the polynomial road model (18) and tracked as extended targets in a Cartesian frame. However, to allow a more general treatment of the problem in this section the extended targets are modeled using n^{th} order polynomials given as

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n, \quad (36)$$

in the range $[x_{\text{start}}, x_{\text{end}}]$ where $\mathbf{m}_a \triangleq [a_0 \ a_1 \ \dots \ a_n]^T$ are the polynomial coefficients and $[x \ y]^T$ are planar Cartesian coordinates. Note that the coordinate y is a function of x and that the direction of the coordinate frame is chosen dependent on the application in mind. The state vector of a map object j is defined as

$$\mathbf{m}^{(j)} \triangleq \left[(\mathbf{m}_a^j)^T \ x_{\text{start}}^j \ x_{\text{end}}^j \right]^T. \quad (37)$$

The map \mathbf{M} is modeled by a set of such polynomial shapes according to (1).

Suppose the 2-dimensional noisy feature based sensor measurements are given in batches of Cartesian x and y coordinates as follows

$$\left\{ \mathbf{z}_k^{(i)} \triangleq \begin{bmatrix} x^{(i)} & y^{(i)} \end{bmatrix}_k^T \right\}_{i=1}^{N_{z,k}}, \quad (38)$$

for discrete time instants $k = 1, \dots, K$. In many cases in reality (e.g., radar, laser and stereo vision cf. (5)) and in the practical application considered in this section, the sensor provides range r and azimuth angle ψ given as,

$$\left\{ \bar{\mathbf{z}}_k^{(i)} \triangleq \begin{bmatrix} r^{(i)} & \psi^{(i)} \end{bmatrix}_k^T \right\}_{i=1}^{N_{z,k}}. \quad (39)$$

In such a case some suitable standard polar to Cartesian conversion algorithm is used to convert these measurements into the form (38).

The state model considered in this section is described, in general, by the state space equations

$$\mathbf{m}_{k+1} = f(\mathbf{m}_k, \mathbf{u}_k) + \mathbf{w}_k, \quad (40a)$$

$$\mathbf{y}_k = h(\mathbf{m}_k, \mathbf{u}_k) + \mathbf{e}_k, \quad (40b)$$

where \mathbf{m} , \mathbf{u} and \mathbf{y} denote the state, the input signal, and the output signal, while $\mathbf{w} \sim \mathcal{N}(0, \mathcal{Q})$ and $\mathbf{e} \sim \mathcal{N}(0, \mathcal{R})$ are the process and measurement noise, respectively. The use of an input signal \mathbf{u} is important in this framework. For the sake of simplicity, the tracked objects are assumed stationary, resulting in very simple motion models (40a).

A polynomial is generally difficult to handle in a filter, since the noisy measurements are distributed arbitrarily along the polynomial. In this respect, the measurement models considered contain parts of the actual measurement vector as parameters. The methodology takes into account the errors caused by using the actual noisy measurements as model parameters. This scheme is an example of the so called ‘‘errors-in-variables’’ (EIV) framework, see e.g., Söderström (2007), Diversi et al. (2005), Björck (1996).

The general convention in modeling is to make the definitions

$$\mathbf{y} \triangleq \mathbf{z}, \quad \mathbf{u} \triangleq \emptyset, \quad (41)$$

where \emptyset denotes the empty set meaning that there is no input. Notice that the subscripts k , specifying the time stamps of the quantities, is omitted for the sake of simplicity. In this setting, it is extremely difficult, if not impossible, to find a measurement model connecting the outputs \mathbf{y} to the states \mathbf{m}_a in the form of (40b). Therefore, other selections for \mathbf{y} and \mathbf{u} , need to be used. Here, the selection

$$\mathbf{y} \triangleq y, \quad \mathbf{u} \triangleq x. \quad (42)$$

is made. Although being quite a simple selection, this choice results in a rather convenient linear measurement model in the state partition \mathbf{m}_a ,

$$\mathbf{y} = H_a(\mathbf{u})\mathbf{m}_a + \mathbf{e}, \quad (43)$$

where $H_a(\mathbf{u}) = [1 \times x^2 \cdots x^n]^T$. It is the selection in (42) rather than (41) that allows to use the standard methods in target tracking with clever modifications. Such a selection as (42) is also in accordance with the EIV representations where measurement noise are present in both the outputs and inputs, i.e., the observation \mathbf{z} can be partitioned according to

$$\mathbf{z} = \begin{bmatrix} \mathbf{u} \\ \mathbf{y} \end{bmatrix}. \quad (44)$$

The measurement vector given in (38) is expressed in terms of a noise free variable \mathbf{z}_0 which is corrupted by additive measurement noise $\tilde{\mathbf{z}}$ according to

$$\mathbf{z} = \mathbf{z}_0 + \tilde{\mathbf{z}}, \quad \tilde{\mathbf{z}} \sim \mathcal{N}(0, \Sigma_c), \quad (45)$$

where the covariance Σ_c can be decomposed as

$$\Sigma_c = \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_y \end{bmatrix}. \quad (46)$$

Note that, in the case the sensor provides measurements only in polar coordinates (39), one has to convert both the measurement $\tilde{\mathbf{z}}$ and the measurement noise covariance

$$\Sigma_p = \text{diag}(\sigma_d^2, \sigma_\delta^2) \quad (47)$$

into Cartesian coordinates. This is a rather standard procedure. Note that, in such a case, the resulting Cartesian measurement covariance Σ_c is, in general, not necessarily diagonal and hence Σ_{xy} of (46) might be non-zero.

Since the model (43) is linear, the Kalman filter measurement update formulas can be used to incorporate the information in \mathbf{z} into the extended source state \mathbf{m}_a . An important question in this regard is what would be the measurement covariance of the measurement noise term \mathbf{e} in (43).

Neglecting the errors in the model parameters $H_a(\mathbf{u})$ can cause overconfidence in the estimates of recursive filters and can actually make data association difficult in tracking applications (by causing too small gates). A simple methodology is used to take the uncertainties in $H_a(\mathbf{u})$ into account in line with the EIV framework. Assuming that the elements of the noise free quantity \mathbf{z}_0 satisfy the polynomial equation exactly according to

$$\mathbf{y} - \tilde{\mathbf{y}} = H_a(\mathbf{u} - \tilde{\mathbf{u}})\mathbf{m}_a, \quad (48a)$$

$$y - \tilde{y} = [1 \times x - \tilde{x} (x - \tilde{x})^2 \cdots (x - \tilde{x})^n] \mathbf{m}_a, \quad (48b)$$

which can be approximated using a first order Taylor expansion resulting in

$$\mathbf{y} \approx H_a(\mathbf{u})\mathbf{m}_a - \tilde{H}_a(\mathbf{u})\tilde{\mathbf{x}}\mathbf{m}_a + \tilde{\mathbf{y}} \quad (49a)$$

$$= H_a(\mathbf{u})\mathbf{m}_a + \tilde{h}_a(\mathbf{m}_a, \mathbf{u}) \begin{bmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{y}} \end{bmatrix}, \quad (49b)$$

with

$$H_a(\mathbf{u}) = [1 \times x^2 \cdots x^n], \quad (49c)$$

$$\tilde{H}_a(\mathbf{u}) = [0 \ 1 \ 2x \cdots nx^{n-1}], \quad (49d)$$

$$\tilde{h}_a(\mathbf{m}_a, \mathbf{u}) = [-a_1 - 2a_2x - \cdots - na_nx^{n-1} \ 1]. \quad (49e)$$

Hence, the noise term \mathbf{e} of (43) is given by

$$\mathbf{e} = \tilde{\mathbf{y}} - \tilde{H}_a\tilde{\mathbf{x}}\mathbf{m}_a = \tilde{h}(\mathbf{m}_a, \mathbf{u}) \begin{bmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{y}} \end{bmatrix} \quad (50)$$

and its covariance is given by

$$\begin{aligned} \Sigma_a = E(\mathbf{e}\mathbf{e}^T) &= \Sigma_y + \mathbf{m}_a\tilde{H}_a\Sigma_x\tilde{H}_a^T\mathbf{m}_a^T - 2\tilde{H}_a\Sigma_{xy} \\ &= \tilde{h}(\mathbf{m}_a, \mathbf{u})\Sigma_c\tilde{h}^T(\mathbf{m}_a, \mathbf{u}). \end{aligned} \quad (51)$$

Note that the EIV covariance Σ_a depends on the state variable \mathbf{m}_a , which is substituted by its last estimate in recursive estimation.

Up to this point, only the relation of the observation \mathbf{z} to the state component \mathbf{m}_a has been considered. It remains to discuss the relation between the observation and the start x_{start} and the end points x_{end} of the polynomial. The measurement information must only be used to update these components of the state if the new observations of the extended source lie outside the range of the polynomial. The following (measurement dependent) measurement matrix can be defined for this purpose:

$$H_{se} = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{if } x \leq x_{\text{start},k|k-1} \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \text{if } x \geq x_{\text{end},k|k-1} \\ \begin{bmatrix} 0 & 0 \end{bmatrix} & \text{otherwise.} \end{cases} \quad (52)$$

The complete measurement model of an extended object can now be summarized by

$$\mathbf{z} = H\mathbf{m} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(0, R(\mathbf{m})), \quad (53a)$$

with

$$H = \begin{bmatrix} \mathbf{0}^{1 \times n} & H_{se} \\ H_a & \mathbf{0}^{1 \times 2} \end{bmatrix}, \quad (53b)$$

$$R(\mathbf{m}) = \text{blkdiag}(\Sigma_x, \Sigma_a(\mathbf{m})). \quad (53c)$$

Put in words, if the x -component of a new measurement is closer to the sensor than the start point of the line x_{start} it is considered in the measurement equation (52) and can be used to update this state variable. Analogously, if a new measurement is more distant than the end point of the line x_{end} it is considered in (52). Further, if a measurement is in between the start and end point of the line, the measurement model is zero in (52) and there is no relation between this measurement and the state variables x_{start} or x_{end} .

Any model as e.g., the standard constant velocity or the coordinated turn model may be used for the targets. For simplicity it is assumed that the targets are stationary in this contribution, thus the process model on the form (40a) is linear and may be written

$$\mathbf{m}_{k+1} = F\mathbf{m}_k + \mathbf{w}_k. \quad (54)$$

To increase the flexibility of the extended object an assumption about the dynamic behavior of its size is made. The size of the extended object is modeled to shrink with a factor $0.9 < \lambda < 1$ according to

$$x_{\text{start},k+1} = x_{\text{start},k} + \lambda(x_{\text{end},k} - x_{\text{start},k}), \quad (55a)$$

$$x_{\text{end},k+1} = x_{\text{end},k} - \lambda(x_{\text{end},k} - x_{\text{start},k}), \quad (55b)$$

leading to the following process model for the polynomial

$$F = \begin{bmatrix} \mathbf{I}^{n \times n} & \mathbf{0}^{n \times 2} \\ \mathbf{0}^{2 \times n} & \begin{bmatrix} 1 - \lambda & \lambda \\ \lambda & 1 - \lambda \end{bmatrix} \end{bmatrix}. \quad (56)$$

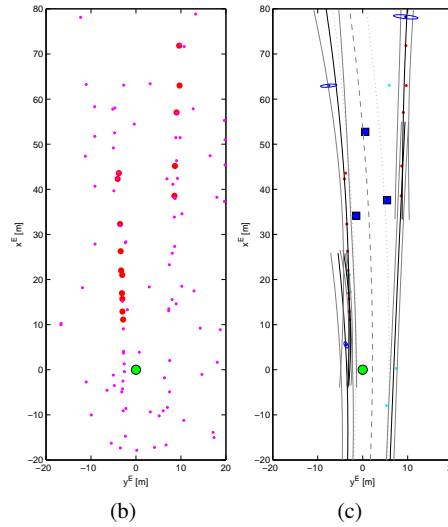
This shrinking behavior for the polynomials allows for automatic adjustment of the start and end points of the polynomials according to the incoming measurements.

The association of measurements to state estimates is treated in Lundquist, Orguner & Gustafsson (2011), where a generalized nearest neighbor method is applied.

The section is concluded with some results based on the information given by an ordinary automotive ACC radar, for the traffic situation shown in Fig. 5a. The ego vehicle, indicated by a green circle, is situated at the (0,0)-position in Fig. 5b, and the red dots are the radar reflections, or stationary observations, at one time sample. The smaller magenta colored dots are former radar reflections, obtained at earlier time samples. Fig. 5c shows the estimated points and lines for the same scenario using the KF EIV method presented in this contribution. The mean values of the states are indicated by solid black lines or blue points. Furthermore, the state variance, by means of the 90% confidence interval, is illustrated by gray lines or cyan colored ellipses, respectively. The estimate of the lane markings (18), illustrated by the gray dashed lines and derived according to the method presented in Lundquist & Schön (2010), is shown here as a comparison. Tracked vehicle in front of the ego vehicle are illustrated by blue squares.



(a)



(b)

(c)

Fig. 5 A traffic situation is shown in Fig. (a). Fig. (b) shows the radar measurements, and Fig. (c) the resulting tracked points and lines. The circle in the origin is the ego vehicle, the square is the tracked vehicle in front and the dashed gray lines illustrate the tracked road curvature.

5 Occupancy Grid Map

An occupancy grid map is defined over a continuous space and it can be discretized with, e.g., a grid approximation. The size of the map can be reduced to a certain area surrounding the ego vehicle. In order to keep a constant map size while the vehicle is moving, some parts of the map are thrown away and new parts are initiated. Occupancy grid mapping (OGM) is one method for tackling the problem of generating consistent maps from noisy and uncertain data under the assumption that the ego vehicle pose, i.e., position and heading, is known. These maps are very popular in the robotics community, especially for all sorts of autonomous vehicles equipped with laser scanners. Indeed several of the DARPA urban challenge vehicles used OGM's, see Buehler et al. (2008). This is because they are easy to acquire, and they capture important information for navigation. The OGM was introduced by Elfes (1987) and an early introduction is given by Moravec (1988). To the best of the

author’s knowledge Borenstein & Koren (1991) were the first to utilize OGM for collision avoidance. Examples of OGM in automotive applications are given in Vu et al. (2007). A solid treatment can be found in the recent textbook by Thrun et al. (2005).

This section begins with a brief introduction to occupancy grid maps, according to Thrun et al. (2005). Using this theory and a sensor with high resolution usually gives a nice looking bird eye’s view map. However, since a standard automotive radar is used, producing only a few range and bearing measurements at every time sample, some modifications are introduced as described in the following sections.

5.1 Background

The planar map \mathbf{M} is defined in the world coordinate frame W and is represented by a matrix. An occupancy grid map is partitioned into finitely many grid cells

$$\mathbf{M} = \{\mathbf{m}^{(j)}\}_{j=1}^{N_m}. \quad (57)$$

The probability of a cell being occupied $p(\mathbf{m}^{(j)})$ is specified by a number ranging from 1 for occupied to 0 for free. The notation $p(\mathbf{m}^{(j)})$ will be used to refer to the probability that a grid cell is occupied. A disadvantage with this design is that it does not allow for dependencies between neighboring cells.

The occupancy grid map was originally developed to primarily be used with measurements from a laser scanner. A laser is often mounted on a rotating shaft and it generates a range measurement for every angular step of the mechanical shaft, i.e. a bearing angle. This means that the continuously rotating shaft produces many range and bearing measurements during every cycle. The OGM algorithms transform the polar coordinates of the measurements into Cartesian coordinates in a fixed world or map frame. After completing one mechanical measurement cycle the sensor provides the measurements for use.

The algorithm loops through all cells and increases the occupancy probability $p(\mathbf{m}^{(j)})$ if the cell was occupied according to the measurement $\mathbf{z}_k^{(i)}$. Otherwise the occupancy value either remains unchanged or is decreased, depending on if the range to the cell is greater or less than the measured range. The latter implies that the laser beam did pass this cell without observing any obstacles. If the measured range is too large or the cell size is too small, it might be necessary to consider the angular spread of the laser beam and increase or decrease the occupancy probability of several cells with respect to the beam width.

The map is assumed to be static, i.e., it does not change during sensing. In this section the map estimation problems is solved with a binary Bayes filter, of which OGM is one example. In this case the estimation problem is solved with the binary Bayes filter, of which OGM is one example. the state can either be free $\mathbf{m}^{(j)} = 0$ or occupied $\mathbf{m}^{(j)} = 1$. A standard technique to avoid numerical instabilities for probabilities close to 0 and to avoid truncation problems close to 0 and 1 is to use

the log odds representation of occupancy

$$\ell_{j,k} = \log \frac{p(\mathbf{m}^{(j)}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k})}{1 - p(\mathbf{m}^{(j)}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k})}, \quad (58)$$

or put in words, the odds of a state is defined as the ratio of the probability of this event $p(\mathbf{m}^{(j)}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k})$ divided by the probability of its complement $1 - p(\mathbf{m}^{(j)}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k})$. The probabilities are easily recovered using

$$p(\mathbf{m}^{(j)}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k}) = 1 - \frac{1}{1 + \exp \ell_{j,k}}. \quad (59)$$

Note that the filter uses the inverse measurement model $p(\mathbf{m}|\mathbf{z}, \mathbf{x})$. Using Bayes' rule it can be shown that the binary Bayes filter in log odds form is

$$\ell_{j,k} = \ell_{j,k-1} + \log \frac{p(\mathbf{m}^{(j)}|\mathbf{Z}_k, \mathbf{x}_{E,k})}{1 - p(\mathbf{m}^{(j)}|\mathbf{Z}_k, \mathbf{x}_{E,k})} - \log \frac{p(\mathbf{m}^{(j)})}{1 - p(\mathbf{m}^{(j)})}, \quad (60)$$

where $p(\mathbf{m}^{(j)})$ represents the prior probability. The log odds ratio of the prior before processing any measurements is defined as

$$\ell_{j,0} = \log \frac{p(\mathbf{m}^{(j)})}{1 - p(\mathbf{m}^{(j)})}. \quad (61)$$

Typically $p(\mathbf{m}^{(j)}) = 0.5$ is assumed, since before having measurements nothing is known about the surrounding environment. This value yields $\ell_0 = 0$.

5.2 OGM with Radar Measurements

The radar system provides range and bearing measurements for observed targets at every measurement cycle. The main difference to a laser is that there is not one range measurement for every angular position of the moving sensor. The number of observations depends on the environment. In general there are much fewer observations compared to a laser sensor. There is also a limit (usually around 32 – 64) on the number of objects transmitted by the radar equipment on the CAN-bus, and a proprietary selection is performed in the radar. Moving objects, which are distinguished by measurements of the Doppler shift, are prioritized and more likely to be transmitted than stationary objects. Furthermore, it is assumed that the opening angle of the radar beam is small compared to the grid cell size. With these the OGM algorithm is changed to loop through the measurements instead of the cells, in order to decrease the computational load. A radar's angular uncertainty is usually larger than its range uncertainty. When transforming the polar coordinates of the radar measurements into the Cartesian coordinates of the map, the uncertainties can either be

transformed in the same manner or it can simply be assumed that the uncertainty increases with the range.

5.3 Experiments and Results

Fig. 6a shows an OGM example of a highway situation. The ego vehicle's camera view is shown in Fig. 6c. The size of the OGM is 401×401 m, with the ego vehicle in the middle cell. Each cell represents a 1×1 m square. The gray-level in the occupancy map indicates the probability of occupancy $p(\mathbf{M}|\mathbf{Z}_{1:k}, \mathbf{x}_{E,1:k})$, the darker the grid cell, the more likely it is to be occupied. The map shows all major structural elements as they are visible at the height of the radar. This is a problem if the road is undulated and especially if the radar observes obstacles over and behind the guardrail. In this case the occupancy probability of a cell might be decreased even though it was previously believed to be occupied, since the cell is between the ego vehicle and the new observation. The impact of this problem can be reduced by tuning the filter well.

It is clearly visible in Fig. 6a that the left border is sharper than the right. The only obstacle on the left side is the guardrail, which gives rise to the sharp edge, whereas on the right side there are several obstacles behind the guardrail, which also cause reflections, e.g., noise barrier and vegetation. A closer look in Fig. 6b reveals that there is no black line of occupied cells representing the guardrail as expected. Instead there is a region with mixed probability of occupancy and after about 5 m the gray region with initial valued cells tell us that nothing is known about these cells.

6 Intensity Based Map

The bin-occupancy filter, which is described in Erdinc et al. (2009), aims at estimating the probability of a target being in a given point. The approach is derived via a discretized state-space model of the surveillance region, where each grid cell (denoted bin in this approach) can or may not contain a target. One of the important assumptions is that the bins are sufficiently small so that each bin is occupied by maximum one target. In the limiting case, when the volume of the bins $|\mathbf{v}|$ tends to zero, it is possible to define the bin-occupancy density

$$D_{k|k} \triangleq \lim_{|\mathbf{v}| \rightarrow 0} \frac{\Pr(\mathbf{m}_k^{(j)} = 1 | \mathbf{Z}_{1:k})}{|\mathbf{v}|}, \quad (62)$$

where $\Pr(\mathbf{m}_k^{(j)} = 1 | \mathbf{Z}_{1:k})$ is the probability that bin j is occupied by one target. The continuous form of the bin-occupancy filter prediction and update equations are the same as the probability hypothesis density (PHD) filter equations (Erdinc et al.

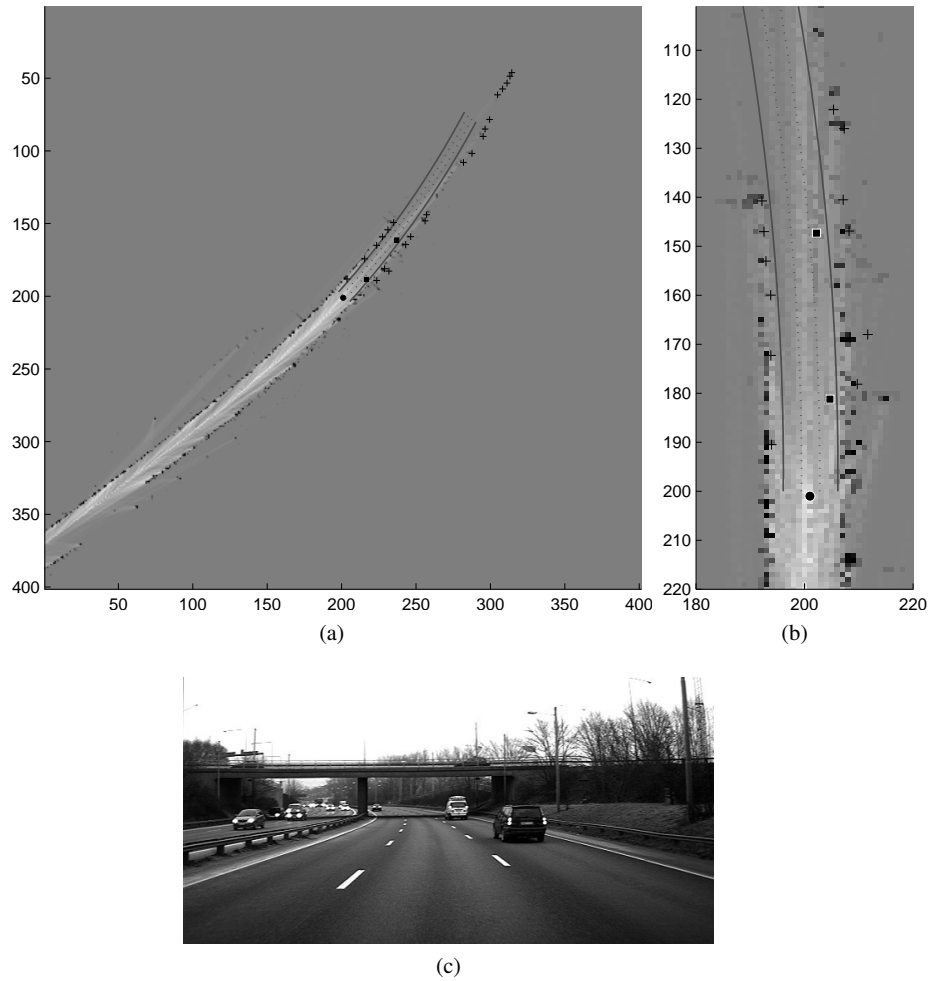


Fig. 6 The filled circle at position (201,201) in the occupancy grid map in Fig. (a) is the ego vehicle, the + are the radar observations obtained at this time sample, the black squares are the two leading vehicles that are currently tracked. Fig. (b) shows a zoom of the OGM in front of the ego vehicle. The gray-level in the figure indicates the probability of occupancy, the darker the grid cell, the more likely it is to be occupied. The shape of the road is given as solid and dashed lines, calculated as described in Section 4. The camera view from the ego vehicle is shown in Fig. (c), the concrete walls, the guardrail and the pillar of the bridge are interesting landmarks. Furthermore, the two tracked leading vehicles are clearly visible in the right lane.

2009). Furthermore, the PHD is the first moment density or *intensity density* in point process theory, see e.g., Mahler (2007), and a physical interpretation is given in Daley & Vere-Jones (2003) as the probability that one target is located in the infinitesimal region $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$ of the state space, divided by $d\mathbf{x}$. The continuous form of the physical bin model leads us to a continuous location based map which we denote *intensity-based map*, and intend to estimate with the PHD filter.

The bin occupancy filter or the PHD filter was developed for target tracking of point sources, however the aim in this section is to create a probabilistic location based map of the surroundings of a moving vehicle. One of the main differences between standard target tracking problems and the building of a location based map, is that many objects such as, guardrails or walls, are typically not point targets, but extended targets (Mahler 2007, Gilholm & Salmond 2005). Furthermore, there is no interest in estimating the number of objects in the map, and there is also no interest in keeping track of specific objects. Nevertheless, the bin-occupancy filter attempts to answer the important question: "Is there an object (target) at a given point?". Erdinc et al. (2009) poses the following assumptions for the bin occupancy filter:

1. The bins are sufficiently small so that each bin is occupied by at most one target.
2. One target gives rise to only one measurement.
3. Each target generates measurements independently.
4. False alarms are independent of target originated measurements.
5. False alarms are Poisson distributed.

Here, only point 2 needs some extra treatment if the aim of the algorithm is mapping and not target tracking. It can be argued that the measurements of point sources belongs to extended objects and that the aim is to create a map of those point sources. Also for mapping purposes, the assumption that there will not be two measurements from the same point at the same time is justified. The relation described is modeled by a likelihood function $p(\mathbf{Z}_k | \mathbf{M}_{k|k})$, which maps the Cartesian map to polar point measurements.

So far in this section the discussion has been quite general and the PHD or the intensity has only been considered as a surface over the surveillance region. The first practical algorithms to realize the PHD filter prediction and measurement update equations were based on the particle filter, see e.g., Vo et al. (2003), Sidenbladh (2003), where the PHD is approximated by a large set of random samples (particles). A Gaussian mixture approximation of the PHD (GM-PHD) was proposed by Vo & Ma (2006). The mixture is represented by a sum of weighted Gaussian components and in particular the mean and covariance of those components are propagated by the Kalman filter. In this work we represent the intensity by a Gaussian mixture, since the parametrization and derivation is simpler than for a particle filter based solution. The modeling of the intensity through a number of Gaussian components also makes it simpler to account for structures in the map. We will return to these structures in the next two sections.

The GM-PHD filter estimates the posterior intensity, denoted $D_{k|k}$, as a mixture of Gaussian densities as,

$$D_{k|k} = \sum_{i=1}^{J_{k|k}} w_{k|k}^{(i)} \mathcal{N} \left(m_{k|k}^{(i)}, P_{k|k}^{(i)} \right), \quad (63)$$

where $J_{k|k}$ is the number of Gaussian components and $w_{k|k}^{(i)}$ is the expected number of point sources covered by the density $\mathcal{N} \left(m_{k|k}^{(i)}, P_{k|k}^{(i)} \right)$. In Lundquist, Hammarstrand & Gustafsson (2011) it is shown how the intensity is estimated with the GM-PHD filter. The Gaussian components are parametrized by a mean $m_{k|k}^{(i)}$ and a covariance $P_{k|k}^{(i)}$, which are expressed in a planar Cartesian coordinate frame, according to

$$m_k^{(i)} = \begin{bmatrix} x_k^{(i)} & y_k^{(i)} \end{bmatrix}^T. \quad (64)$$

The aim of the mapping algorithm is to estimate the posterior density (3). The considered intensity based map is continuous over the surveillance region, thus, for the number of elements in (1) it holds that $N_m \rightarrow \infty$. Furthermore, the intensity is a summary statistic of the map according to

$$p(\mathbf{M}_k | \mathbf{Z}_{1:k}) \sim p(\mathbf{M}_k; D_{k|k}), \quad (65)$$

see e.g., Mahler (2003), and the estimated intensity $D_{k|k}$ is parametrized by

$$\mu_k^{(i)} \triangleq \left\{ w_k^{(i)}, m_k^{(i)}, P_k^{(i)} \right\} \quad (66)$$

of the Gaussian sum (63). The intensity based map is a multimodal surface with peaks around areas with many sensor reflections or point sources. It is worth observing that the map \mathbf{M} is described by a location based function (63), with feature based parametrization (66).

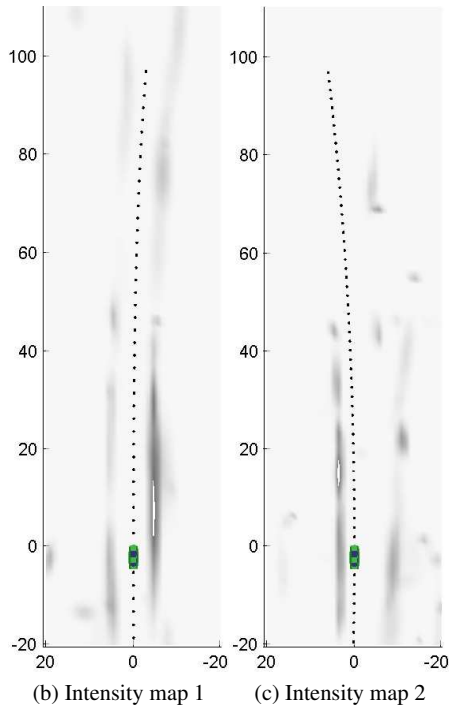
Experiments were conducted with a prototype passenger car. One example of the estimated intensity at a freeway traffic scenario is shown as a bird's eye view in Fig. 7b. Darker regions illustrate higher concentrations of point sources, which in this figure stem from the guardrails to the left and the right of the road. As expected, the path of the ego vehicle, indicated by the black dots, is in between the two regions of higher object concentration. The driver's view is shown in Fig. 7a.

A second example is shown in Fig. 7c and 7d. Here, the freeway exit is clearly visible in the intensity map, which shows that the proposed method to create maps is very conformable.

The Gaussian components are generally removed from the filter when the vehicle passed those parts of the map. However, to give a more comprehensive overview, these components are stored and the resulting intensity based map is shown together with an occupancy grid map (OGM) and a flight photo in Fig. 8. The top figure is the map produced as described in this section. The OGM, described in the previous Section 5, is based on the same data set and used as a comparison. The gray-level of the OGM indicates the probability of occupancy, the darker the grid cell the more likely it is to be occupied. As seen in the figure the road edges are not modeled as



(a) Camera view 1



(b) Intensity map 1

(c) Intensity map 2



(d) Camera view 2

Fig. 7 The image in (a) shows the driver's view of the intensity map in (b), and the image in (d) is the driver's view of the intensity map in (c). The darker the areas in the intensity map, the higher the concentration of objects. The driver's path is illustrated with black dots and may be used as a reference. Note that snap shot in (c) and (d) is obtained only some meters after the situation shown in Fig. 1.

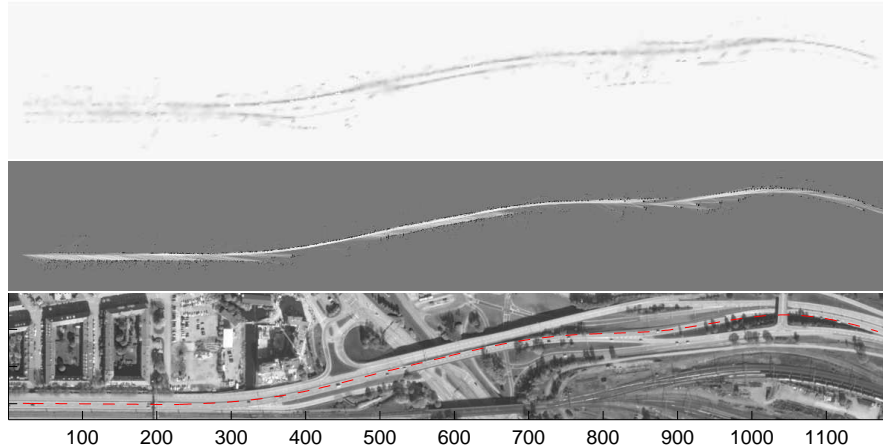


Fig. 8 The top figure shows the intensity based map obtained from radar measurements collected on a freeway. The OGM in the middle figure serves as a comparison of an existing algorithm. The bottom figure is a flight photo used as ground truth, where the driven trajectory is illustrated with a dashed line (©Lantmäteriet Gävle 2010. Medgivande I 2011/0100. Reprinted with permission). Note that the drivers view at 295 meter is shown in Fig. 7d, and about the same position is also shown in Fig. 1 and Fig. 5.

distinct with the OGM. The OGM representation of the map is not very efficient, since huge parts of the map are gray indicating that nothing is known about these areas. An OGM matrix with often more than 10000 elements must be updated and communicated to other safety functions of a car at each time step. The compact representation is an advantage of the intensity based map. Each Gaussian components is parametrized with 7 scalar values according to (66). Since most maps are modeled with 10 – 30 components it summarizes to around 70 – 210 scalar values, which easily can be sent on the vehicles CAN bus to other safety functions. Finally, the bottom photo is a very accurate flight photo (obtained from the Swedish mapping, cadastral and land registration authority), which can be used as ground truth to visualize the quality of the intensity based map.

7 Conclusion

The use of radar, laser and camera for situation awareness is gaining popularity in automotive safety applications. In this chapter it has been shown how sensor data perceived from the ego vehicle is used to estimate a map describing the local surroundings of a vehicle. The map may be modeled in various different ways, of which four major approaches have been described. In a feature based map each element of the map specifies the properties and location of one object. This can either be a point source in the space; or it can be an extended object such as the position and shape of

the lane or the road edges. Furthermore, in a location based map the index of each element corresponds to a location and the value of the map element describes the property in that position. One example is the occupancy grid map, which is defined over a continuous space but discretized with a grid approximation. Another example is the intensity based map, which is a continuous approximation, describing the density of objects in the map. The four approaches presented in this chapter have all been evaluated on real data from both freeways and rural roads in Sweden.

The current accuracy of GPS receivers is acceptable only for route guidance, where the provided global position is sufficient. For automotive active safety systems, the local position of the ego vehicle with respect to its surroundings is more important. The estimated maps, described in this chapter, can be used to increase the localization accuracy of the ego vehicle. Furthermore, the maps may be used to derive a collision free trajectory for the vehicle.

Acknowledgment

The authors would like to thank the Sensor Fusion for Safety (SEFS) project within the Intelligent Vehicle Safety Systems (IVSS) program, the strategic research center MOVIII, funded by the Swedish Foundation for Strategic Research (SSF) and CADICS, a Linneaus Center funded by the Swedish Research Council for financial support.

References

- Björck, Å. (1996), *Numerical methods for least squares problems*, SIAM, Philadelphia, PA, USA.
- Blackman, S. S. & Popoli, R. (1999), *Design and Analysis of Modern Tracking Systems*, Artech House, Norwood, MA, USA.
- Borenstein, J. & Koren, Y. (1991), 'The vector field histogram-fast obstacle avoidance for mobile robots', *IEEE Transactions on Robotics and Automation* **7**(3), 278–288.
- Buehler, M., Iagnemma, K. & Singh, S., eds (2008), *Special Issue on the 2007 DARPA Urban Challenge, Part I-III*, Vol. 25 (8–10), Journal of Field Robotics.
- Caveney, D. (2010), 'Cooperative vehicular safety applications', *IEEE Control Systems Magazine* **30**(4), 38–53.
- Civera, J., Davison, A. & Montiel, J. (2008), 'Inverse depth parametrization for monocular SLAM', *IEEE Transactions on Robotics* **24**(5), 932–945.
- Daley, D. J. & Vere-Jones, D. (2003), *An introduction to the theory of point processes. Vol. 1, Elementary theory and method*, 2 edn, Springer, New York, NY, USA.

- Davison, A. J., Reid, I., Molton, N. & Strasse, O. (2007), 'MonoSLAM: Real-time single camera SLAM', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(6), 1052–1067.
- Dickmanns, E. (1988), Dynamic computer vision for mobile robot control, in 'Proceedings of the International Symposium on Industrial Robots', Sydney, Australia.
- Dickmanns, E. D. (2007), *Dynamic Vision for Perception and Control of Motion*, Springer, London, UK.
- Dickmanns, E. D. & Mysliwetz, B. D. (1992), 'Recursive 3-D road and relative ego-state recognition', *IEEE Transactions on pattern analysis and machine intelligence* **14**(2), 199–213.
- Diversi, R., Guidorzi, R. & Soverini, U. (2005), 'Kalman filtering in extended noise environments', *IEEE Transactions on Automatic Control* **50**(9), 1396–1402.
- Duda, R. O. & Hart, P. E. (1972), 'Use of the hough transformation to detect lines and curves in pictures', *Communications of the ACM* **15**(1), 11–15.
- Eidehall, A., Pohl, J., Gustafsson, F. & Ekmark, J. (2007), 'Toward autonomous collision avoidance by steering', *IEEE Transactions on Intelligent Transportation Systems* **8**(1), 84–94.
- Elfes, A. (1987), 'Sonar-based real-world mapping and navigation', *IEEE Journal of Robotics and Automation* **3**(3), 249–265.
- Erdinc, O., Willett, P. & Bar-Shalom, Y. (2009), 'The bin-occupancy filter and its connection to the PHD filters', *IEEE Transactions on Signal Processing* **57**(11), 4232–4246.
- Gilholm, K. & Salmond, D. (2005), 'Spatial distribution model for tracking extended objects', *IEE Proceedings of Radar, Sonar and Navigation* **152**(5), 364–371.
- Gill, P. E., Murray, W., Saunders, M. A. & Wright, M. H. (1991), 'Inertia-controlling methods for general quadratic programming', *SIAM Review* **33**(1), 1–36.
- Goldfarb, D. & Idnani, A. (1983), 'A numerically stable dual method for solving strictly convex quadratic programs', *Mathematical Programming* **27**(1), 1–33.
- Gordon, N. J., Salmond, D. J. & Smith, A. F. M. (1993), 'Novel approach to nonlinear/non-Gaussian Bayesian state estimation', *IEE Proceedings on Radar and Signal Processing* **140**(5), 107–113.
- Guiducci, A. (1999), 'Parametric model of the perspective projection of a road with applications to lane keeping and 3D road reconstruction', *Computer Vision and Image Understanding* **73**(3), 414–427.
- Guiducci, A. (2000), 'Camera calibration for road applications', *Computer Vision and Image Understanding* **79**(2), 250–266.
- Harris, C. & Stephens, M. (1988), A combined corner and edge detector, in 'Proceedings of the Alvey Vision Conference', Manchester, UK, pp. 147–151.
- Hough, P. V. C. (1962), 'A method and means for recognizing complex patterns', U.S. Patent No. 3069654.
- Jung, C. R. & Kelber, C. R. (2005), 'Lane following and lane departure using a linear-parabolic model', *Image and Vision Computing* **23**(13), 1192–1202.

- Kim, Z. W. (2008), ‘Robust lane detection and tracking in challenging scenarios’, *IEEE Transactions on Intelligent Transportation Systems* **9**(1), 16–26.
- Lee, J. W. (2002), ‘A machine vision system for lane-departure detection’, *Computer vision and image understanding* **86**(1), 52–78.
- Lowe, D. G. (2004), ‘Distinctive image features from scale-invariant keypoints’, *International Journal of Computer Vision* **60**(2), 91–110.
- Lundquist, C., Hammarstrand, L. & Gustafsson, F. (2011), ‘Road intensity based mapping using radar measurements with a probability hypothesis density filter’, *IEEE Transactions on Signal Processing*. Accepted for publication.
- Lundquist, C., Orguner, U. & Gustafsson, F. (2011), ‘Extended target tracking using polynomials with applications to road-map estimation’, *IEEE Transactions on Signal Processing* **59**(1), 1–12.
- Lundquist, C. & Schön, T. B. (2010), ‘Joint ego-motion and road geometry estimation’, *Information Fusion*. DOI: 10.1016/j.inffus.2010.06.007.
- Mahler, R. P. S. (2003), ‘Multitarget Bayes filtering via first-order multitarget moments’, *IEEE Transactions on Aerospace and Electronic Systems* **39**(4), 1152–1178.
- Mahler, R. P. S. (2007), *Statistical Multisource-Multitarget Information Fusion*, Artech House, Boston, MA, USA.
- Matas, J., Chum, O., Urban, M. & Pajdla, T. (2004), ‘Robust wide baseline stereo from maximally stable extremal regions’, *Image and Vision Computing* **22**(10), 731–767.
- McCall, J. C. & Trivedi, M. M. (2006), ‘Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation’, *IEEE Transactions on Intelligent Transportation Systems* **7**(1), 20–37.
- Moravec, H. (1988), ‘Sensor fusion in certainty grids for mobile robots’, *AI Magazine* **9**(2), 61–74.
- Nistér, D., Naroditsky, O. & Bergen, J. (2006), ‘Visual odometry for ground vehicle applications’, *Journal of Field Robotics* **23**(1), 3–20.
- Powell, M. (1985), ‘On the quadratic programming algorithm of Goldfarb and idnani’, *Mathematical Programming Study* **25**(1), 46–61.
- Rohling, H. & Meinecke, M.-M. (2001), Waveform design principles for automotive radar systems, in ‘Proceedings on CIE International Conference on Radar’, Beijing, China, pp. 1–4.
- Rohling, H. & Möller, C. (2008), Radar waveform for automotive radar systems and applications, in ‘IEEE Radar Conference’, Rome, Italy, pp. 1–4.
- Sidenbladh, H. (2003), Multi-target particle filtering for the probability hypothesis density, in ‘Proceedings of the International Conference on Information Fusion’, Vol. 2, Cairns, Australia, pp. 800–806.
- Söderström, T. (2007), ‘Survey paper: Errors-in-variables methods in system identification’, *Automatica* **43**(6), 939–958.
- Szeliski, R. (2010), *Computer vision : algorithms and applications*, Springer, New York, NY, USA.
- Thrun, S., Burgard, W. & Fox, D. (2005), *Probabilistic Robotics*, The MIT Press, Cambridge, MA, USA.

- Vo, B.-N. & Ma, W.-K. (2006), 'The Gaussian mixture probability hypothesis density filter', *IEEE Transactions on Signal Processing* **54**(11), 4091–4104.
- Vo, B.-N., Singh, S. & Doucet, A. (2003), Random finite sets and sequential monte carlo methods in multi-target tracking, in 'Proceedings of the International Radar Conference', Adelaide, Australia, pp. 486–491.
- Vu, T. D., Aycard, O. & Appenrodt, N. (2007), Online localization and mapping with moving object tracking in dynamic outdoor environments, in 'Proceedings of the IEEE Intelligent Vehicles Symposium', Istanbul, Turkey, pp. 190–195.
- Wang, Y., Bai, L. & Fairhurst, M. (2008), 'Robust road modeling and tracking using condensation', *IEEE Transactions on Intelligent Transportation Systems* **9**(4), 570–579.
- Waxman, A., LeMoigne, J., Davis, L., Srinivasan, B., Kushner, T., Liang, E. & Sidalgaiah, T. (1987), 'A visual navigation system for autonomous land vehicles', *IEEE Journal of Robotics and Automation* **3**(2), 124–141.
- Zhou, Y., Xu, R., Hu, X. & Ye, Q. (2006), 'A robust lane detection and tracking method based on computer vision', *Measurement science and technology* **17**(4), 736–745.