Vehicle self-localization using off-the-shelf sensors and a detailed map

IEEE Intelligent Vehicles Symposium, Proceedings

Vehicle Self-Localization Using Off-the-Shelf Sensors and a Detailed Map

Malin Lundgren, Erik Stenborg, Lennart Svensson, Lars Hammarstrand

***Abstract*— In the research on autonomous vehicles, self-localization is an important problem to solve. In this paper we present a localization algorithm based on a map and a set of off-the-shelf sensors, with the purpose of evaluating this low-cost solution with respect to localization performance. The used test vehicle is equipped with a Global Positioning System receiver, a gyroscope, wheel speed sensors, a camera providing information about lane markings, and a radar detecting landmarks along the road. Evaluation shows that the localization result is within or close to the requirements for autonomous driving when lane markers and good radar landmarks are present. However, it also indicates that the solution is not robust enough to handle situations when one of these information sources is absent.**

I. Introduction

Self-driving cars have been envisioned for almost as long as the car has been around. The self-driving car can bring many benefits to society such as increased traffic safety, comfort and productivity for drivers, as well as better use of roads, parking facilities and fuel.

A self-localization system is an essential component of an autonomous car. 20 years ago, in the Prometheus project [1], the approach was to use only on-board sensors for both localization and obstacle detection. Prometheus showed that this was possible, albeit for a limited set of scenarios, e.g., highway driving.

More recent localization approaches often utilize a detailed digital map in addition to the on-board sensors, in order to cover more complex scenarios. The map typically encodes the position of a number of landmarks that are visible using various types of sensors. By measuring the distance and/or the angle to the landmarks, it is possible to deduce ones location. Two examples of methods for generating a map, and possible representations of map features, can be found in [2] and [3].

The localization problem has been solved to some extent also in urban scenarios, and at moderate speeds, in the DARPA Urban Challenge [4]. However, all top contenders in the Urban Challenge relied on a rotating multi-beam laser scanner. Such a device is currently unrealistic, both for cost and packaging reasons, in a commercial product.

Camera sensing, in combination with a high resolution map of lane markings and other visual landmarks, can be considered a low cost set-up eligible for commercialization. Typical information provided by a camera system is shape and position of lane markings [5], traffic signs or, as in [6], arrows, speed limits and pedestrian crossings painted on the road. The information from the camera is often fused with measurements from a Global Positioning System (GPS) receiver and internal vehicle sensors [7]–[9]. A potential problem with this approach is that the localization becomes heavily dependant on a single sensor. If the camera stops to function due to glare, blockage, fog or worn out lane markers, the whole localization algorithm will break. Some of these issues are possible to solve by using redundant cameras, but fog and worn out lane markers, for example, would benefit more from a second sensing technology. One example of this is the work presented in [10], where the vehicle, in addition to a camera, is equipped with a laser scanner.

In this paper we look specifically at the self-localization problem using one combination of relatively low-cost on-board sensors. The aim of this work is to evaluate if this low cost solution can solve the localization for the purpose of autonomous cars at acceptable levels of accuracy. We use a sensor set-up containing a camera and a 76 GHz radar, in order not to be limited to the visible light spectrum. The focus of this paper is more on sensor fusion and the addition of radar as a sensor in self-localization rather than developing visual lane detection with the camera. In addition to the radar and camera, standard sensors available in most cars like wheel speed sensors, gyroscope and GPS receiver are used. During the data collection, the vehicle was equipped with an RTK-GPS which is only used as a reference in the generation of a map and as ground truth in the evaluation of the self-localization algorithm.

II. Problem formulation

The aim in this work is to evaluate the potential in an available set of sensors, regarding the self-localization performance. The problem of localization is to estimate the position of the ego vehicle in global coordinates, or, as we do here, relative to a road. The two problems are equivalent when the road is accurately described in global coordinates. In order to be safely implemented, an autonomous vehicle requires very accurate position and heading estimates. How accurate the localization must be is highly dependent on the scenario, the control algorithms, etc. In this paper, we consider position accuracy expressed in lateral and longitudinal errors. The longitudinal position is defined as the distance along a reference route on the road while the lateral position is the normal distance to the route. The requirements are set

This work was supported by the Strategic Vehicle Research and Innovation Program (FFI), which is funded by the Swedish Agency for Innovation Systems (VINNOVA). M. Lundgren, E. Stenborg, L. Svensson and L. Hammarstrand are with the Department of Signals and Systems, Chalmers University of Technology, SE-412 96, Gothenburg, Sweden. E. Stenborg is also with Volvo Car Corporation, Gothenburg Sweden. (E-mail: {malin.lundgren, erik.stenborg, lennart.svensson, lars.hammarstrand}@chalmers.se).

to lateral errors below 0.2 m and longitudinal errors below 1 m, which we believe would be sufficient for autonomous driving in many scenarios. The remainder of this section describes the used sensors and the limitations under which the localization is performed.

The data available for the localization is collected using commercially available sensors mounted in a test vehicle. There is a GPS receiver, a gyroscope and wheel speed sensors, providing information about position, heading, speed and turn rate. To sense the surrounding environment, there is a camera that observes the lane markings, and a forward-looking radar detecting objects on and along the road. In addition to the standard sensors, the vehicle is equipped with an RTK-GPS system which is an accurate reference system that can be used as ground truth in the evaluation of the localization algorithm and as a reference for generating a map. A summary of the sensor properties can be found in Table I. Information about the reference route and the road

	Information	Rate	Field of view
GPS	Position, heading, speed	1 Hz	-
Gyroscope	Turn rate	50 Hz	-
Speedometer	Speed	50 Hz	-
Camera (0.3 megapixel monovision)	Lane markers	10 Hz	Range: 60 m Angle: $\pm 30°$
Radar (76 GHz)	Guard rails, delineator posts, traffic signs	40 Hz	Range: 60 m Angle: $\pm 30°$
RTK-GPS	Reference	100 Hz	-

TABLE I: Summary of the sensors

is collected on the test track depicted in Fig. 1. The track is a 5 km long rural road where the number of lanes varies. The main part of the road segments have lane markers and many segments have guard rails, delineator posts or traffic signs along the road. During the data collection, there is only one vehicle on the track, hence we do not gain any information about the road shape from detected vehicles. On the other hand, there are no vehicles obscuring the surrounding environment from the sensors. From the test track we have data collected during two laps using both the standard sensors and the reference system.

III. Generating a map

The access to RTK-GPS reference data, together with the information from the other sensors, allow us to create a map of the road and its closest surroundings. Having such a map provides prior knowledge about the road and stationary objects and enables a matching of the received measurements to features in the global coordinate frame. Considering the test track in Fig. 1, we want to describe lane markings, guard

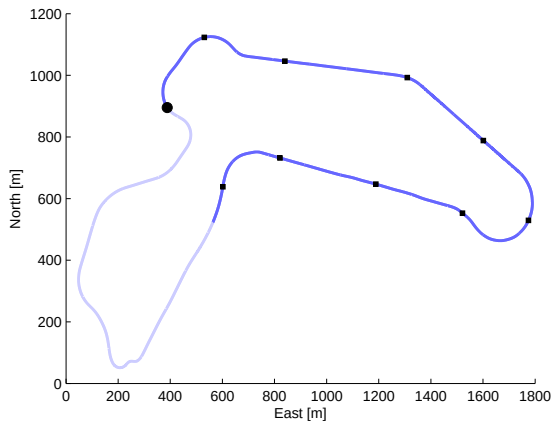


Fig. 1: The test track where the evaluation scenario is marked with darker blue. The starting point of the evaluation is marked by a large black circle and the true vehicle position every 20 s is marked by black squares.

rails, delineator posts and traffic signs in the map. In addition to these, we include a reference route in the map, which is used to define the road aligned coordinate system.

A. Lane markings and the reference route

To efficiently describe the lane markings and the reference route in the map, we want a smooth line representation that we can fit to the sensor data. There are many different descriptions that can be used for this purpose, e.g. b-splines, Beziér splines or clothoids. In this paper we use third-order Beziér splines to describe lines in the map. We use this representation because it is easily fitted to the collected data, and provides a natural segmentation that allows us to consider sub parts of the map. The spline consists of connected third order Beziér curves expressed as:

$$\mathbf{B}(t) = (1-t)^3\mathbf{P}_0 + 3(1-t)^2 t\mathbf{P}_1 + 3(1-t)t^2\mathbf{P}_2 + t^3\mathbf{P}_3, \quad (1)$$

where $t \in [0,1]$ and $(\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$ are the so-called control points. Each curve segment is described by its control points, which are found by solving a least-squares problem under the constraints that the starting point of a segment is the same as the end point of the previous segment, and that the tangent vector should have the same direction at both sides of the common control point. By imposing these constraints, we ensure that the resulting spline is continuous and sufficiently smooth.

The route is measured by the reference system, which provides more accurate position measurements compared to the standard GPS receiver mounted in the vehicle. The RTK-GPS delivers a position measurement every 10 ms and to these measurements we fit a Beziér spline that describes the route in the map, see Fig. 2. The lane marking measurements are defined relative the vehicle but, using the reference data, they are transformed into global coordinates. Then, similar as for the route, each line can be described by a spline.

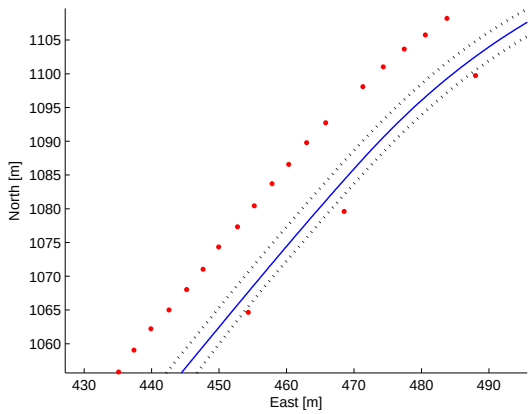


Fig. 2: A part of the map where the blue solid line is the reference route, the black dashed lines are the lane markings and the red dots show the radar landmarks. Here, the landmarks consist of a guard rail on one side of the road and three delineators on the other side.

B. Radar landmarks

The available radar measurements consist of pre-processed estimates of stationary objects, described in a local vehicle coordinate frame. By transforming the complete set of radar data into global coordinates, it is possible to cluster estimates into groups representing potential landmarks. The position of an identified landmark in the map is then given by the mean position of the estimates associated with that landmark. Both guard rails and more obvious point objects, such as traffic signs, are described using point sources in the map, see Fig. 2. Objects further away from the road, that generate a sufficient amount of detections, are included in the map as clutter generating features. Examples of clutter sources can be bankings or vegetation along the road. The reason for keeping the point sources and the clutter sources separated, is that the measurements from them behave differently. For example, the point sources are more likely to generate measurements while the measurements from the clutter sources spread more. It is worth noting that it is possible to characterize the measurement distribution from each landmark individually and store the information in the map. However, we have chosen to only distinguish between two types of landmarks in the map and incorporate the differences in the measurement model.

IV. Proposed Solution

In this section, we present the proposed self-localization algorithm. At each discrete time instant, k, we solve the localization problem by estimating the state vector, $\mathbf{x}_k$, containing the sought properties of the ego vehicle, such as position and speed. To handle non-linear models, and at the same time describe potentially multimodal densities, the algorithm is implemented using a particle filter [11]. In each step, the filter approximates the posterior distribution by a set of N particles, $\mathbf{x}_k^{(i)}$, and their corresponding weights, $w_k^{(i)}$, according to

$$p(\mathbf{x}_k|\mathbf{Z}_{1:k}, \mathbf{M}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}). \tag{2}$$

In (2), $\mathbf{Z}_{1:k}$ is the set of all received measurements up to time k and $\mathbf{M}$ is the map described in Section III. We use a bootstrap particle filter which requires a process model for propagating the particles in the prediction step, and a likelihood for each set of measurements in order to update the particle weights

$$w_k^{(i)} \propto p(\mathbf{z}_k|\mathbf{x}_k^{(i)}) w_{k-1}^{(i)}. \tag{3}$$

This section proceeds with a motivation of the choice of state vector describing the ego vehicle and a description of the process model. Thereafter, given the state and the map, the measurement models required in the particle filter are presented.

A. The state vector

When predicting the position of the vehicle, it is desirable to incorporate the knowledge that the vehicle is travelling on the road. To facilitate this we describe the vehicle state in local road coordinates which are defined relative the reference route in the map. That is, the state of the vehicle is described by the state vector

$$\mathbf{x}_k = [l_k, n_k, \varphi_k, v_k, \omega_k]^T, \tag{4}$$

where l_k is the longitudinal distance along the route and n_k is the normal distance to the route. Furthermore, in (4), φ_k is the heading angle relative the route, v_k is the speed in the heading direction and ω_k is the yaw rate relative the turn rate of the road.

To describe the measurements from all sensors except the speedometer, given the state, we use a transformation from road coordinates to the global map. This transformation is denoted $T(\mathbf{x}_k, \mathbf{M})$ and results in the global state vector

$$[E_k, N_k, \tilde{\varphi}_k, v_k, \tilde{\omega}_k]^T = T(\mathbf{x}_k, \mathbf{M}) \tag{5}$$

where (E_k, N_k) is the two dimensional north-east position on the map, $\tilde{\varphi}_k$ and $\tilde{\omega}_k$ are the global heading angle and turn rate, respectively, while v_k is the speed as given in road coordinates.

B. Process model

The process model, that describes how the state vector evolves over time, is assumed to be a constant turn and velocity model, which in continuous time is given by:

$$\begin{aligned} \dot{l}(t) &= v(t)\cos(\varphi(t)) \\ \dot{n}(t) &= v(t)\sin(\varphi(t)) \\ \dot{\varphi}(t) &= \omega(t) \\ \dot{v}(t) &= e_v(t) \\ \dot{\omega}(t) &= e_\omega(t) \end{aligned} \tag{6}$$

where $e_v(t)$ and $e_\omega(t)$ are zero-mean Gaussian noise representing changes in speed and yaw rate. Depending on the

choice of discretization method, there are different discrete versions of this model, all on the form

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k) + \mathbf{v}_k, \tag{7}$$

where f is a nonlinear function of the state and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ is the process noise. Discrete models based on an Euler approximation and a second order Taylor expansion are presented and evaluated in [12], where we have used the latter one.

C. GPS measurement model

From the GPS we receive measurements on the vehicle position, heading and speed. The measurements are given in longitude-latitude coordinates, but are mapped into the north-east coordinate frame before being used in the localization algorithm. The employed measurement model is somewhat simplistic and ignores both multi-path signals and time-correlated noise

$$\mathbf{z}_k^{\text{gps}} = \begin{bmatrix} E_k^{\text{gps}} \\ N_k^{\text{gps}} \\ \varphi_k^{\text{gps}} \\ v_k^{\text{gps}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} T(\mathbf{x}_k, \mathbf{M}) + \mathbf{w}_k^{\text{gps}} \tag{8}$$

where $\mathbf{w}_k^{\text{gps}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^{\text{gps}})$ is the measurement noise. Note that the model is nonlinear due to the transformation $T(\mathbf{x}_k, \mathbf{M})$.

D. Measurement models for speedometer and gyroscope

The speedometer and the gyroscope provide information about speed and yaw rate, respectively. The speed measurements provided by the speedometer are described by the linear model

$$z_k^{\text{speed}} = v_k^{\text{speed}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + w_k^{\text{speed}} \tag{9}$$

where $w_k^{\text{speed}} \sim \mathcal{N}(0, \sigma_{\text{speed}}^2)$ is the measurement noise. Ignoring the drift in the gyroscope, the model describing the yaw rate measurements can be stated as:

$$z_k^{\text{gyro}} = \omega_k^{\text{gyro}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix} T(\mathbf{x}_k, \mathbf{M}) + w_k^{\text{gyro}} \tag{10}$$

where $w_k^{\text{gyro}} \sim \mathcal{N}(0, \sigma_{\text{gyro}}^2)$ is the noise.

E. Camera measurement model

The car is equipped with a camera that observes the lane markers in front of the car. The camera reports at most two lines at each time k, one to the right and one to the left.

The measurements from the camera are coefficients to a third degree polynomial that represents the shape of the detected lane markers. The polynomial is defined in a local 2d Cartesian coordinate frame that has its origin fixed to the center of the rear axle of the vehicle, and where x and y are straight ahead and to the left, respectively. Besides the coefficients, $a_{k,i}$ for $i = 0...3$, there is also a parameter, x_k^l, that defines how far along the x-axis the polynomial is valid. In total, the measurement vector contains the following parameters,

$$\mathbf{z}_k^{\text{cam}} = [a_{k,0}, a_{k,1}, a_{k,2}, a_{k,3}, x_k^l]^T. \tag{11}$$

For $0 \leq x \leq x_k^l$, this measurement specifies the polynomial

$$\begin{aligned} y_k^{\text{cam}}(x) &= a_{k,0} + a_{k,1}x + a_{k,2}x^2 + a_{k,3}x^3 && (12) \\ &= L(x) + w_k^{\text{cam}}(x), && (13) \end{aligned}$$

where $L(x)$ is the detected line in the map and $w_k^{\text{cam}}(x) \sim \mathcal{N}(0, \sigma_{\text{cam}}^2(x))$ is the measurement noise. Due to the projection of the road plane onto the camera sensor, lane markers are measured with greater accuracy near the car. Assuming a linearly increasing standard deviation of the noise, the variance is given as $\sigma_{\text{cam}}^2(x) = (r + sx)^2$, where r and s are positive constants.

1) The measurement likelihood: At each time instant, the camera can detect at most one line on each side of the vehicle. Assuming that these measurements are independent, they can be treated separately. Hence, in the following we only consider a single measurement. The situation when the sensor does not deliver a measurement can be informative, however, in this paper we do not consider this negative information and the likelihood in (3) is set to constant in this case.

When a lane marking measurement is received, there are normally multiple lines in the map that can be considered as the detected line. Starting with the i:th line, denoted L_i, the measurement-to-line-matching is performed by comparing samples from the line in the map to the corresponding samples of the polynomial in (12). More specifically, we chose two points on the Beziér curve that describes L_i in the map. The points are denoted $P_{i,j} = [\tilde{E}_{i,j}, \tilde{N}_{i,j}]^T$ for $j = 1, 2$, and are chosen such that $P_{i,1}$ is near the position given by $T(\mathbf{x}_k, \mathbf{M})$ and $P_{i,2}$ is located at a significant distance ahead, but not beyond x_k^l. To enable the matching to the polynomial, the points are transformed to the camera coordinate frame, resulting in the points $Q_{i,j} = [\tilde{x}_{i,j}, \tilde{y}_{i,j}]^T$, $j = 1, 2$. Given that the measurement $\mathbf{z}_k^{\text{cam}}$ originates from the i:th line, the likelihood can then be stated as

$$p(\mathbf{z}_k^{\text{cam}} | \mathbf{x}_k, \mathbf{M}, L_i) = \prod_{j=1}^{2} \mathcal{N}(y_k^{\text{cam}}(\tilde{x}_{i,j}); \tilde{y}_{i,j}, \sigma_{\text{cam}}^2(\tilde{x}_{i,j})). \tag{14}$$

Considering all possible lines, the total likelihood for a camera measurement is

$$p(\mathbf{z}_k^{\text{cam}} | \mathbf{x}_k, \mathbf{M}) = P_{\text{FA}} + (1 - P_{\text{FA}}) \sum_i w_{k,i} p(\mathbf{z}_k^{\text{cam}} | \mathbf{x}_k, \mathbf{M}, L_i) \tag{15}$$

where $w_{k,i}$ are the normalized weight depending on the distance between the vehicle and L_i, and P_{FA} is the probability of a false alarm. In this work we approximate (15) by only considering the most likely measurement to line association according to

$$p(\mathbf{z}_k^{\text{cam}} | \mathbf{x}_k, \mathbf{M}) = P_{\text{FA}} + (1 - P_{\text{FA}}) P_{\max}, \tag{16}$$

where $P_{\max}$ is given as

$$P_{\max} = \max_i p(\mathbf{z}_k^{\text{cam}} | \mathbf{x}_k, \mathbf{M}, L_i). \tag{17}$$

F. Radar measurement model

From the radar processing unit, we receive a set of m_k measurements, each a pre-processed estimate of a stationary object. The measurement set is denoted $\mathbf{Z}_k^{\text{radar}} = \{\mathbf{z}_{k,1}^{\text{radar}}, ..., \mathbf{z}_{k,m_k}^{\text{radar}}\}$ where $\mathbf{z}_{k,i}^{\text{radar}}$ is described in the same local (x, y) coordinates as the camera measurements. While the radar measurements are given in this local coordinate frame the landmarks in $\mathbf{M}$, to which we want to match the radar measurements, are described in global north-east coordinates. Hence, by transforming the vehicle state to global coordinates we can easily match the measurements to landmarks in the map. Given the global vehicle state, $[E_k, N_k, \tilde{\varphi}_k, v_k, \tilde{\omega}_k]^T = T(\mathbf{x}_k, \mathbf{M})$, and conditioned on that the i:th landmark at position $(E_{\text{map}}^i, N_{\text{map}}^i)$ is detected by the radar, the produced measurement is described by

$$\mathbf{z}_k^{\text{radar}} = \begin{bmatrix} x_k^{\text{radar}} \\ y_k^{\text{radar}} \end{bmatrix} = h\big(T(\mathbf{x}_k, \mathbf{M}), (E_{\text{map}}^i, N_{\text{map}}^i)\big) + \mathbf{w}_{k,i}^{\text{radar}}$$
$$= \begin{bmatrix} (E_{\text{map}}^i - E_k)\cos\tilde{\varphi}_k + (N_{\text{map}}^i - N_k)\sin\tilde{\varphi}_k \\ (N_{\text{map}}^i - N_k)\cos\tilde{\varphi}_k - (E_{\text{map}}^i - E_k)\sin\tilde{\varphi}_k \end{bmatrix} + \mathbf{w}_{k,i}^{\text{radar}} \tag{18}$$

where $\mathbf{w}_{k,i}^{\text{radar}} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i^{\text{radar}})$ is the measurement uncertainties. As discussed earlier, depending on if the measurement originate from a point source or a clutter source the uncertainties differ. More specifically, the noise covariance is described by

$$\mathbf{R}_i^{\text{radar}} = \begin{cases} \mathbf{R}_m & \text{if landmark } i \text{ is a point source} \\ \mathbf{R}_m + \mathbf{R}_c & \text{if landmark } i \text{ is a clutter source} \end{cases} \tag{19}$$

where $\mathbf{R}_c$ is accounting for the scattering properties of the clutter sources.

1) Likelihood based on a Poisson assumption: For a given set of of measurements, at time k, we need to calculate the likelihood used for updating the particle weights according to (3). The origin of each measurement is unknown which leads to data association uncertainties. Summation over the association hypotheses is time consuming when there are many landmarks and measurements. A method for calculating the needed likelihood without considering all hypotheses explicitly is presented in [13], where the measurement set is modelled as an inhomogeneous Poisson process.

The assumption that the measurements are described by Poisson processes suits us reasonably well since many of the landmarks, although being points in the map, might generate multiple detections. In the used map, there are both point sources and clutter generators which are treated similarly as the targets in [13]. However, the two types of landmarks are treated separately since they are modelled with different expected number of measurements and measurement uncertainties. In addition to the landmarks, we have a uniform clutter density with intensity β describing radar data originating from objects not included in the map. Under these assumptions, the likelihood can be written:

$$p(\mathbf{Z}_k^{\text{radar}}|\mathbf{x}_k, \mathbf{M})$$
$$= \frac{e^{-\mu(V(\mathbf{x}_k))}}{m_k!} \prod_{j=1}^{m_k} \Bigg\{ \beta + \sum_{n=1}^{N_P} \mu_n(V(\mathbf{x}_k)) p(\mathbf{z}_k^j|\mathbf{x}_k, (E_{\text{map}}^n, N_{\text{map}}^n))$$
$$+ \sum_{m=1}^{N_C} \mu_m(V(\mathbf{x}_k)) p(\mathbf{z}_k^j|\mathbf{x}_k, (E_{\text{map}}^m, N_{\text{map}}^m)) \Bigg\}, \tag{20}$$

where $V(\mathbf{x}_k)$ is the field of view and N_P and N_C denote the number of point sources and clutter generators, respectively, in the map. Considering the i:th landmark, $p(\mathbf{z}_k^j|\mathbf{x}_k, (E_{\text{map}}^i, N_{\text{map}}^i))$ is given by the models in (18) and (19). The expected number of measurements in the field of view from a landmark is denoted $\mu_i(V(\mathbf{x}_k))$, i.e. if the i:th landmark is located outside the field of view $\mu_i(V(\mathbf{x}_k)) = 0$. The total expected number of measurements at time k is denoted $\mu(V(\mathbf{x}_k)) = \sum_{i=1}^{N_P+N_C} \mu_i(V(\mathbf{x}_k))$.

V. Evaluation

In this section we evaluate the presented self-localization filter using data collected on the test track described by the map. The result is compared to the requirements stated in Section II, i.e., we aim for a lateral error less than 0.2 m while a longitudinal error up to 1 m is allowed. To get a better understanding of the robustness of the system, the evaluation is divided into two parts, one where all available sensors are used, and one where either the radar or the camera is deactivated.

The evaluation is performed on a part of the test track and the number of detectable lane markings and visible point landmarks during the scenario are depicted in Fig. 3.

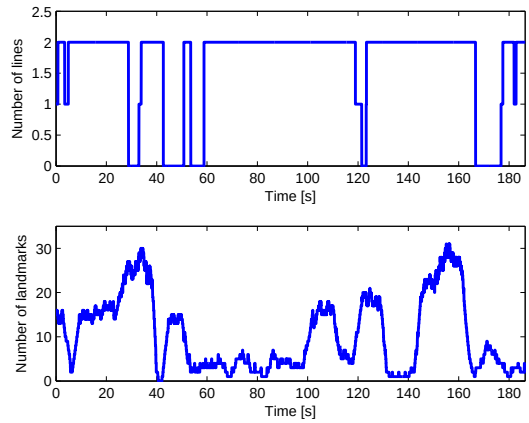


Fig. 3: The number of visible lane markings and the number of point sources within the radar field of view as a function of the reference system time during the data collection.

At each time instant k when the particle weights are updated using new information, an approximation of the maximum a posteriori estimate is computed as the weighted

sum of the particles

$$\hat{\mathbf{x}}_{k|k} = \sum_{i=1}^{N} w_k^{(i)} \mathbf{x}_k^{(i)}. \tag{21}$$

This estimate is compared to the ground truth provided by the reference system to get the errors in the computed estimates.

A. Implementation details

To initiate the position, heading and speed in the particle filter, $N = 10^4$ particles are drawn from a density given by the first GPS measurement and its uncertainties. The yaw rate is initiated in a similar way using the gyroscope. The initial particle weights are set to $w_0^{(i)} = 1/N$, $i = 1...N$.

The process and measurement noise covariances used in the implementation are set to: $\mathbf{Q} = \text{diag}\{\sigma_v^2, \sigma_\omega^2\} = \text{diag}\{5^2, 0.5^2\}$, $\mathbf{R}^{\text{gps}} = \text{diag}\{10^2, 10^2, 1^2, 0.2^2\}$, $\sigma_{\text{speed}}^2 = 2^2$, $\sigma_{\text{imu}}^2 = 0.2^2$, $\sigma_{\text{cam}}^2(x) = (0.25 + 0.05x)^2$ and for the radar measurements $\mathbf{R}_{\text{m}} = \text{diag}\{2^2, 2^2\}$, $\mathbf{R}_{\text{c}} = \text{diag}\{5^2, 2^2\}$.

To calculate the likelihood in (15), the probability of false alarm is set to $P_{\text{FA}} = 0.01$.

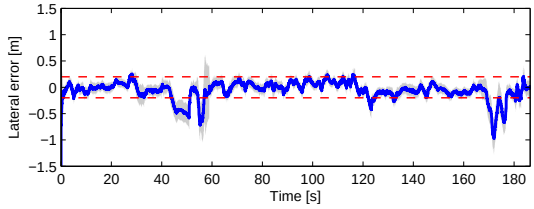
For the radar measurements, and the likelihood in (20), we use $V(\mathbf{x}_k)$ according to the field of view in Table I. For point sources $\mu_i(V(\mathbf{x}_k)) = 0.7$ and for clutter sources $\mu_i(V(\mathbf{x}_k)) = 0.2$, if the landmark is located within $V(\mathbf{x}_k)$. The number of landmarks in the map is $N_P + N_C = 751 + 431 = 1182$. The clutter intensity is set to $\beta = 5 \cdot 10^{-4}$ which results in 1.5 expected clutter measurements, in addition to those generated by the clutter sources, at each scan.

B. Performance using all sensors

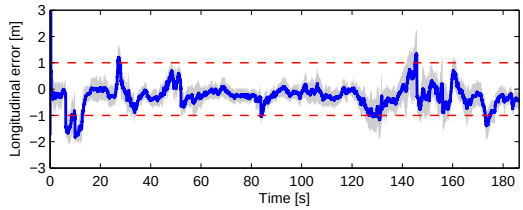
Using all available sensor data, except for the reference system, we run the localization filter and compute the estimated position, heading angle and speed according to (21). The errors are illustrated in Fig. 4 together with the requirements on the accuracy for the lateral and the longitudinal errors. Comparing the error with the requirements, we see that during most parts of the scenario, the performance is within or close to the set limits. More specifically, the lateral error is below the set limit 87 % of the time and the corresponding number for the longitudinal error is 94 %. Taking a closer look at the intervals where the error exceeds the limits, we see that it is often during road segments where there are no (or at least less than 2) lane markings, such as $t \in [40, 60]$. It is worth noting that during $t \in [165, 175]$, there is a curve with neither lane markings nor guard rails.

C. Robustness

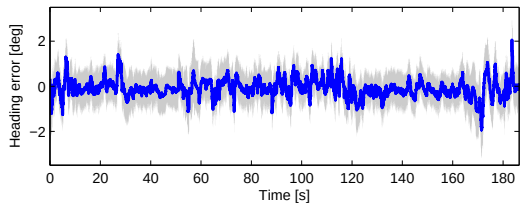
To study the robustness of the localization algorithm, we repeat the evaluation on the same scenario and with the same data set, but once without using the lane marking measurements and once without the radar data. This illustrates the expected performance on a road without lane markings or radar landmarks. The results for the position and the heading angle for the two cases are shown in Fig. 5 and 6. The speed estimates are omitted here since no interesting changes were observed. As expected, the overall result becomes worse when one of the sensors is left out. For example, the



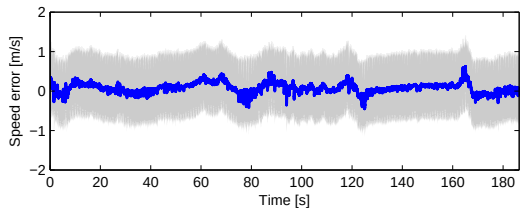
(a) The lateral error. The requirement on a lateral error below 20 cm is illustrated by the red dashed lines.



(b) The longitudinal error. The requirement on a longitudinal error below 1 m is illustrated by the red dashed lines.



(c) The error in heading angle



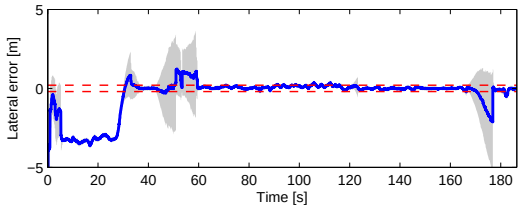
(d) The error in the speed estimates

Fig. 4: The estimation error when localization is performed using all sensors. The grey area illustrates the filter uncertainties (standard deviation) based on the particles.

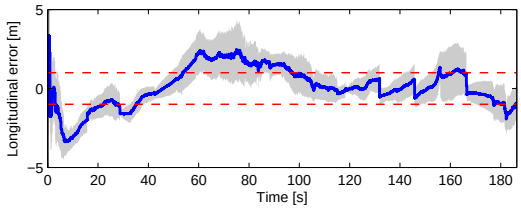
longitudinal error grows when no radar data is available. The problem becomes very clear at $t \in [60, 80]$ in Fig. 5(b), since there is less longitudinal information in the lane markings when driving on a straight road. The lateral error on the other hand is small during segments with lane markings, except in the beginning where the filter is initiated in the wrong lane due to poor GPS measurements. Not using the camera, the lateral error becomes bigger while the longitudinal error stays close to the requirements. This is expected as the lane markings provide the system with lateral information. To facilitate a comparison of the different cases, the results are summarized in Table II.

VI. Conclusions

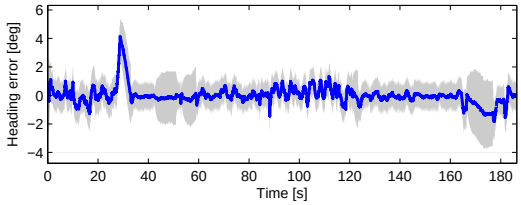
This paper presents a localization algorithm based on a detailed map and a set of available off-the-shelf sensors. The



(a) The error in lateral position.

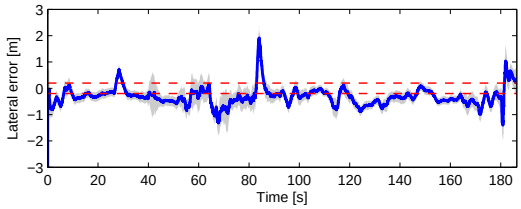


(b) The error in longitudinal position.

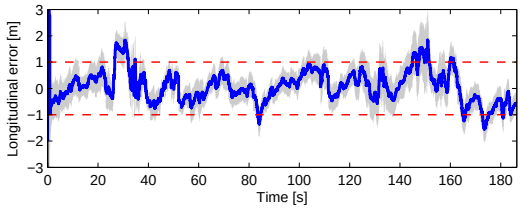


(c) The error in heading angle.

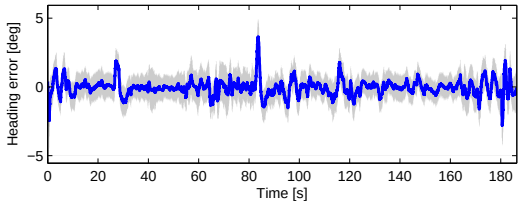
Fig. 5: The estimation error when the radar is deactivated.



(a) The error in lateral position.



(b) The error in longitudinal position



(c) The error in heading angle

Fig. 6: The estimation error when the camera is deactivated.

	Lat. error < 0.2 m	Long. error < 1 m
Using all sensors	87 %	94 %
Without radar	69 %	56 %
Without camera	23 %	90 %

TABLE II: Percent of the time that the position errors are within the requirements, computed from the results in Fig 4, 5 and 6.

map contains the position of lane markings and radar landmarks and was generated using a reference system together with camera and radar measurements from one data set. Evaluation on a second data set shows that the localization performance is within or close to the requirements set on position accuracy on the parts of the road where there are both lane markings and guard rails. However, the tests where either the camera or the radar is deactivated, indicate that the proposed algorithm is not robust enough to produce accurate position estimates in all situations. Possible improvements for these scenarios can be to use more sophisticated measurement models, e.g. incorporate low-level GPS information, or extract more information from the used sensors, such as information about additional landmarks from the camera.

References

[1] E. Dickmanns, R. Behringer, C. Brudigam, D. Dickmanns, F. Thomanek, and V. van Holt, "An all-transputer visual autobahn-autopilot/copilot," in *Fourth International Conference on Computer Vision*, 1993, pp. 608–615.

[2] K. Jo and M. Sunwoo, "Generation of a precise roadway map for autonomous cars," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–13, 2013.

[3] A. Schindler, G. Maier, and F. Janda, "Generation of high precision digital maps using circular arc splines," in *IEEE Intelligent Vehicles Symposium (IV)*, 2012, pp. 246–251.

[4] "Special Issue on the 2007 DARPA Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 423–860, 2008.

[5] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine Vision and Applications*, pp. 1–19, Feb. 2012.

[6] T. Wu and A. Ranganathan, "Vehicle localization using road markings," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2013, pp. 1185–1190.

[7] F. Chausse, J. Laneurit, and R. Chapuis, "Vehicle localization on a digital map using particles filtering," in *IEEE Intelligent Vehicles Symposium*, 2005, pp. 243–248.

[8] Z. Tao, P. Bonnifait, V. Frémont, and J. Ibañez Guzman, "Lane marking aided vehicle localization," in *IEEE Intelligent Vehicles Symposium (ITS)*, 2013.

[9] M. Schreiber, C. Knoppel, and U. Franke, "Laneloc: Lane marking based localization using highly accurate maps," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 449–454.

[10] A. Schindler, "Vehicle self-localization with high-precision digital maps," in *IEEE Intelligent Vehicles Symposium (IV)*, 2013, pp. 141–146.

[11] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb 2002.

[12] M. Morelande and N. Gordon, "Target tracking through a coordinated turn," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, 2005, pp. iv/21–iv/24 Vol. 4.

[13] K. Gilholm, S. J. Godsill, S. Maskell, and D. J. Salmond, "Poisson models for extended target and group tracking," *SPIE*, October 2005.