Map-Aided Localization with Lateral Perception

Dominique Gruyer, Rachid Belaroussi, and Marc Revilloud

Abstract—Accurate localization of a vehicle is a challenging task as GPS available on the market are not designed for lane-level accuracy application. Although dead reckoning helps, cumulative errors from inertial sensors result in a integration drift. This paper presents a new method of localization based on sensors data fusion. An accurate digital map of the lane marking is used as a powerful additional sensor. Road markings are detected by processing two lateral cameras to estimate their distance to the vehicle. Coupled with the map data in an EKF filter it improves the ego-localization obtained with inertial and GPS measurements. The result is a vehicle localization at an ego-lane level of accuracy, with a lateral error of less than 10 centimeters.

I. INTRODUCTION

Driver Assistance Systems developed over the last decade have required a precise and robust estimation of road scene major features. Those features include obstacles (vehicle, pedestrian), road (marking, lanes, traffic signs), and the ego-vehicle (localization and dynamics of the vehicle). Usually each feature is addressed separately, for instance obstacle detection in collision avoidance, road ego-vehicle in lane keeping assistance, ego-localization in navigation systems. Recently, automated driving systems have made it necessary to fuse the attributes of different features to obtain more precise, robust and complete information.

In the frame of french, european and international projects (respectively ABV, eFuture, CooPerCom) we have tackled the task of perception of the environment including the attributes of these three features. More specifically, we develop an application localizing the ego-vehicle in its lane allowing a positioning and a lateral control precise enough to be applicable.

There is a body of work in the field of robust localization by hybrid data fusion (proprio and exteroceptive): mono-model approaches (EKF, UKF, DD1, DD2) [1], [2], [3], multi-model [4], [5] and particular filter [6] achieve localization with a precision to the meter. For instance, in [5], the approach is only centered on the ego-vehicle but can compute the likelihood of each model to build the finale estimation. In [6] a map is used to filter out the particles out of a road. In this case, the localization process converges rapidly toward a solution inside a roadway. However it cannot specify which lane the vehicle is on.

It is clear that the use of a geographic map can be an advantage: commercial navigation systems routinely operate MapMatching by coupling maps and vehicle positioning. In [7] a map-matching algorithm using the visual odometry motion trajectory estimation (from a stereovision rig) as input and corrected using the digital map features, greatly improve the global localization performance. In [8], a standard navigation map is matched with a laser scanner occupancy grid map, a video based grid map and a lane marking grid map for road course prediction. Alternately, in [9] a radar local map is built by estimating the ego-motion from odometry measurements with an EKF. Matching the digital road map and the grid maps is optimized with a Levenberg-Marquardt or a particle filter.

Recent approaches assume the road network has been surveyed accurately beforehand. In [10] a loosely coupled GPS/INS system is used with a camera: the digital map is made of polylines and area information divided into three classes (line landmark, road surface and background) which are projected in the image plane. Appearance of these classes are modeled by the distribution of their coherency values, and used in a particle filter. In [11] the absolute position of corners of ground markings (such as arrows, speed limit, or texts) were precisely surveyed and are matched to corners detected by FAST with a camera in order to estimate its pose and location. In [12], a vehicle with a backward facing camera is used to create a 3D landmarks map of the environment. Landmarks are matched into the current image and back projection errors are minimized yielding a rough single shot pose estimate. IMU measurements are blended with past single shot estimates yielding the final ego-pose.

In order to enhance accuracy and potentially to determine the traffic lane being travelled, [13], [4] proposed to use a map of the lane marking with a centimeter level accuracy coupled with a differential GPS EGNOS and inertial datas. The combination is done by way of a particle filter. The map is used as a geometrical constraint in the ego-postioning of the vehicle and for map-matching purposes. However, the lack of GPS data or in case of multipath effect, this approach approach is not able to function properly and to guaranty the integrity of its results. A relevant solution is to add a local processing providing a lane marking detection. It can provide a more accurate localization of the vehicle inside its lane.
Recently, [1] proposed an EKF-based algorithm fusing GPS, IMU and lane marking information: they have shown that the use of visual features can improve the lateral localization up to a centimeter-level accuracy (less than 30 cm). Their experimental setup uses one camera directed frontward to detect lane marking and estimate their lateral distance to the vehicle. In [14] a similar setup (GPS, IMU, front camera) is used with the addition of a laser scanner, the digital map also containing location of landmarks such as traffic signs, tree or guide post. Extracted features (lane markings and landmarks) are associated with the elements in the digital map to correct the pose estimation which is roughly initialized with GPS. A particle filter is to implement the ego-localization algorithm.

In [15], a method to improve global localization in an intersection is based on the alignment of visual landmarks with the information from an extended digital map. A stereovision system provides a detailed 3D perception of road landmarks such as lateral lane delimiter, painted traffic signs, curbs and stop lines. Combination of visual and enriched map of the intersection is done with a Bayesian network, yielding to a global localization with a submeter level of accuracy.

In this paper, we propose an algorithm that fuses localization data, road marking detection and a digital map of the road (location of the edges of the left and right lanes) in order to obtain a centimeter lateral localization of our vehicle. Our approach uses two lateral cameras detecting left and right markings independently, with a focus on lane segment close to the vehicle leading. We believe this setup improves the lateral localization and is specific enough for the vehicle to be considered as moving on rails.

II. IN-VEHICLE APPARATUS

The experimental setup uses two cameras positioned on each side of the vehicle and directed toward the ground. In this configuration, cameras can be low cost and have a low resolution because the scope of information to be extracted is only a little over a meter. This configuration is shown in Figure 1. Several coordinate frames have to be considered: for each sensor, there is the image frame \((u, v)\) and the camera frame to be able to describe the characteristic of a road mark in respect of each camera.

The other sensors considered are proprioceptive sensors:
- a GPS-receiver with a slow refresh rate (1Hz): it provides vehicle absolute position in the world coordinate frame \(R_0\),
- an Inertial Navigation System INS (66Hz), including motion sensors (accelerometer for each axis) and rotation sensors (gyroscope providing roll, pitch and yaw rate),
- an odometer (20Hz) measuring the distance traveled,
- a shaft encoder to measure the wheel steering angle.

Figure 1 also illustrates the vehicle coordinate frame \(R_v = (X_v, Y_v, Z_v)\) where \(Z_v\) denotes the altitude, and \((X_v, Y_v)\) plane is parallel to the ground.

The rigid transformation between the camera frames and the vehicle frame is know by experimentally measuring the extrinsic parameters for each camera. As the camera are calibrated, the intrinsic parameters are also known.

---

(a) Source image.

(b) Detected primitives.

(c) Projection on the \(R_v\) frame.

Fig. 2. Lane marking primitive detection with a frontward camera.
Fig. 3. Overview of the Satory track map. The road lane is divided into
band portions bounded by two triplet of points (yellow circles). Triplets of
points are sampled every 4 to 20 meters depending on the curvature of the
road.

An other feature required for the proposed approach is
a map with the precise location of the lane marking. Our
experiments are run on the Satory (Versailles-France) test
track which is a one way, 3.5 km long asphalt carriageway,
with sharp bends, hairpin turns, as well as straight sections.
A top-view of its map is displayed in Figure 3. The total road
width is 7 meters: it is a two-lanes way with three painted
white lines defining the lane boundaries. Land surveys on the
track provide a centimeter level accuracy of its road marking.
The test track of Satory is a 2-lanes road precisely surveyed
by professionals: coordinates of the road edges and the center
line has been registered with a centimeter level accuracy.
The resulting digital map is made of 380 triplets of points
(1 sample every 5 meters on bends, 1 sample/20 meters
on straight portions) also contains road features such as its
curvature and type (straight, clothoid, arch). The sampling
of the map is reduce to 10 cm by interpolation.

III. ROAD MARKING DETECTION USING TWO LATERAL CAMERAS

The road marking detector we developed is an adaptation
of the work presented in [16], which will be summarized
in this section for the sake of clarity. It is a three stages
algorithm:

- road primitive extraction: this step determines which
  pixels belong to a lane marker. For each camera, this
  set of point is projected in the vehicle coordinate frame
  using the known intrinsic and extrinsic parameters.
- lane marking detection: the spatial distribution of ex-
  tracted point on the $Y_v$ axis is analysed to determine the
  center of the potential marking lanes.
- lane shape estimation using a polynomial fitting. From
  this last step, two parameters are estimated: the vehicle
  yaw and its distance to the lane.

In order to improve the primitive extraction step, we pro-
posed a cooperative combination strategy of two intensity-
based extractors: the Median Local Threshold MLT and
the Symmetric Local Threshold. Figure 2(a) illustrates the
detected road primitives (in blue) superimposed on the source
image.

Knowing the intrinsic and extrinsic camera parameters, it
is possible to match a point $P$ of the image with a real world
point under the assumption that it is a point of the road.
The set of pixels classified as road marking primitives of
Figure 2(b) projected on the $(X_v, Y_v)$ plane of the vehicle
frame is illustrated on Figure 2(c). A 1D projection on the
$Y_v$ axis is performed using dynamic templates: it results in
a 1D-histogram showing the spatial distribution of the road
primitive along the $Y_v$ axis (Figure 2(c)). Within this distri-
bution, clusters areidentified using a tierce derivative and
further selected based on mean belief, number of primitives,
cluster standard deviation and length.

A robust poly-fitting is finally applied on each cluster so
that the candidate lane is modeled by its equation on the
$(X_v, Y_v)$ plane:

$$ y = ax^2 + bx + c $$  \hspace{1cm} (1)

When the camera is directed frontward, the shape of the
road especially its curvature introduces a lot of possible
variations in the road marking appearance. Estimation of its
parameters (curvature, slope, bank) is limited by the accuracy
of the distance computation and can be sensitive to change
in lighting conditions. In our apparatus, only the part of the
marking the closest to the vehicle is being sensed as shown
in Figure 4(a). In this configuration, detected primitives are
less noisy and the effect of lighting variation is limited.
Moreover, lane marking can be assumed to be locally linear.
The distance $D_1$ (or $D_2$) from a camera to a lane as well as
its orientation $\phi$ can be extracted from parameters $b$ and $c$
of Equation 1: $b = \tan \phi$ and $c = D_1$. The lateral distance
ranges from 1 meter to 2.5 meters during our tests, while the vehicle traveled exclusively on the left lane.

IV. EGO-VEHICLE ABSOLUTE LOCALIZATION

The ego-vehicle localization is estimated using an Extended Kalman Filter defined by the following equations:

\[ X_{k+1} = f(X_k, U_k) + V_k \]  \hspace{1cm} (2)
\[ Y_{k+1} = h(X_{k+1}) + W_k \]  \hspace{1cm} (3)

\( X_k = (x_k, y_k, \theta_k)^T \) is the vehicle state at time k. \( Y_k \) is the measurement vector at time k. \( V_k \) and \( W_k \) are the process and observation noises, assumed to be zero mean multivariate Gaussian noises with covariance \( R_k \) and \( Q_k \) respectively.

Regarding the prediction step, dead reckoning is performed by merging the proprioceptive data coming from the steering wheel coder (from which the actual front wheels steering radius \( \theta_c \) is computed), the odometer measuring the distance \( d_k \) traveled at time \( t_k \), and the yaw rate \( \dot{\theta}_k \) from the INS. The EKF is based on the widely used bicycle non linear model. An ellipse \( E_k \) representing the level of confidence on the localization can be build from the covariance matrix.

As both speed and direction must be accurately measured at all time, dead reckoning is subject to significant errors, as illustrated by Figure 5(a): each estimate of position being relative to the previous one, errors are cumulative [17]. To overcome this drift in time, other source of information are used, the most common being the GPS-receiver. The correction introduced by exteroceptive measurement such as the GPS data significantly improves the accuracy of the positioning estimate as shown by Figure 5(b). However, GPS data are also subject to bias (see [18]) and a drift in vehicle localization can still be found especially in case of hairpin turn as illustrated in Figure 6(a) where the EKF filter results in a localization outside of the roadway. To reach a lane-level accuracy, we propose to integrate the lane map information into the Kalman filtering.

V. LANE MAP INTEGRATION

At this stage, the following capability are gathered:

- estimation of the distance from the vehicle to the left and right road marking of the lane,
- estimation of the vehicle localization using an extended Kalman filter.

The issue is now to adequately use these information to enhance the accuracy of the ego-localization. To build a bridge between these two approaches, we used the Satory cartography data.

Locally the track is approximated by a line segment. Using the estimated vehicle position, a point-to-segment based Map-Matching is performed to select a lane segment of the map. This segment is modeled by its polar equation featuring its polar coordinates \( (\rho, \theta) \):

\[ x \cos \theta_{seg} + y \sin \theta_{seg} - \rho_{seg} = 0 \]  \hspace{1cm} (4)

Image processing only provides local measurements. To obtain an absolute localization of the ego-vehicle, a cartography matching algorithm is applied. The distance \( D_l \) \((D_r) \) or \( D_2 \) see Figure 4(a)) between the camera and the marking is extracted from the lane marking equation \( y = ax^2 + \tan \phi x + D_i \) by image processing. The corresponding \( P_i \) point coordinates are computed in the vehicle frame \( R_v \):

\[ vP_i = \begin{bmatrix} vX_i \\ vY_i \end{bmatrix} = \begin{bmatrix} vX_c \\ vY_c \end{bmatrix} + D_i \begin{bmatrix} \cos \theta_c \\ \sin \theta_c \end{bmatrix} \]  \hspace{1cm} (5)

where \( (vX_c, vY_c) \) are the coordinates of the camera center in the vehicle frame, \( \theta_c \) the orientation of the camera frame in \( R_v \) \((\theta_c = -\pi/2 \text{ for the left camera and } \theta_c = \pi/2 \text{ for the right camera}) \) and \( D_i \) the distance of the closest point of the lane sensed by the camera.

The Jacobian of the measurement matrix from the polar measurement equation is:

\[ H_{k+1} = \begin{bmatrix} \frac{\partial h_{k+1}}{\partial x} & \frac{\partial h_{k+1}}{\partial y} & \frac{\partial h_{k+1}}{\partial \theta} \end{bmatrix} \]  \hspace{1cm} (6)

![Image](image_url)
Fig. 6. (a) Error made by the Kalman filter with the merger of Odometer / INS / GPS. (b) Results obtained with the correction using lateral distance and accurate mapping.

with

\[
\frac{\partial h_{k+1}}{\partial x} = \cos \theta_{\text{seg}}
\]

(7)

\[
\frac{\partial h_{k+1}}{\partial y} = \sin \theta_{\text{seg}}
\]

(8)

\[
\frac{\partial h_{k+1}}{\partial \theta} = (X_c + D_s \cos \theta_c) \sin (\theta_{\text{seg}} - \theta_{\text{seg}}) - (Y_c + D_l \sin \theta_c) \cos (\theta_{\text{seg}} - \theta_{\text{seg}})
\]

(9)

The coordinates of point \( P_1 \) and \( P_2 \) in the absolute frame \( R_0 \) are then computed:

\[
\begin{bmatrix}
X_{P_1} \\
Y_{P_1}
\end{bmatrix} = \begin{bmatrix}
\hat{x}_{kj} \\
\hat{y}_{kj}
\end{bmatrix} + \begin{bmatrix}
\cos \hat{\theta}_{kj} & \sin \hat{\theta}_{kj} \\
-\sin \hat{\theta}_{kj} & \cos \hat{\theta}_{kj}
\end{bmatrix} \begin{bmatrix}
X_c \\
Y_c
\end{bmatrix}
\]

\[
+ \begin{bmatrix}
\cos \theta_c & \sin \theta_c \\
-\sin \theta_c & \cos \theta_c
\end{bmatrix} \begin{bmatrix}
X_i \\
Y_i
\end{bmatrix}
\]

(10)

\( \hat{\theta}_{kj} \) is the estimated yaw at time \( k \). \((\hat{x}_{kj}, \hat{y}_{kj})\) are the estimated coordinates of the vehicle at time \( k \) (in the absolute frame).

If \( X_{P_1} \) and \( Y_{P_1} \) measurement are unnoisy measurement, \( P_i \) point belongs to matching segments of the map and the road marking. In this case, \( X_{P_1} \cos \theta_{\text{seg}} + Y_{P_1} \sin \theta_{\text{seg}} - \rho_{\text{seg}} = 0 \), else \( X_{P_1} \cos \theta_{\text{seg}} + Y_{P_1} \sin \theta_{\text{seg}} - \rho_{\text{seg}} \neq 0 \). This error is used to update the positioning; the measurement error is then computed as:

\[
h_{k+1} = \begin{bmatrix}
\cos \theta_{\text{seg}} & \sin \theta_{\text{seg}} \\
-\sin \theta_{\text{seg}} & \cos \theta_{\text{seg}}
\end{bmatrix} \begin{bmatrix}
X_{P_1} \\
Y_{P_1}
\end{bmatrix}
\]

(11)

The Kalman gain is computed with matrix \( H_k \) and enable to update the estimate:

\[
\hat{X}_{k+1|k+1} = \hat{X}_{k+1|k} + K_{k+1} h_{k+1}
\]

(12)

The covariance matrix is updated using \( H_{k+1} \) and \( R_k \).

VI. RESULTS AND COMMENTS

In order to validate this approach, we made a series of tests in simulated and real conditions. Results presented in this section come from a recording made in real conditions.

To properly see the benefits of the approach to an accurate location, we present 3 cases. First, we proposed an estimate location with only the use of an odometer and an inertial unit. Then we applied the updated estimates with GPS data. Finally we implemented the use of all data sources (Odometer, INS, GPS, and cameras).

With only the use of proprioceptive sensors, we see a drift of the filter due to the noise and bias of the inertial sensors. After traveling 30 meters, we already have 20 cm of error. After 100 meters, we get 1 meter of error. After 180 meters the error magnitude is 2 meters. This behavior is clearly observed in Figure 5(a). Once we implement the filter’s correction phase with the use of GPS data, we find that the results greatly improve as shown in Figure 5(b). However, a zoom in on the nonlinear part (tight turn), illustrated in Figure 6(a) shows that the estimated position of the ego-vehicle is off the track while the vehicle is actually traveling on the left lane.

Using additional information relating to the lateral gap between the ego-vehicle and the roadsides provided by a lane detection algorithm, we obtain significantly better results (see Figure 6(b)). These results clearly show that the vehicle is properly located on the left lane. The whole test results are demonstrated in two online videos:

- GPS-IMU ego-localization without visual features: youtube.com/watch?v=N-0ANY7zqJs
- map-aided localization with lateral vision: youtube.com/watch?v=nSfc1UM7qI4

Figure 7 illustrates the difference in accuracy at several time of the test run. The samples in column (a) are estimated with the GPS and the IMU alone, while the samples of column (b) are improved with the use of lateral vision.
Localization by integration of lateral distance and accurate mapping. The confidence level on the estimation (yellow ellipse) is higher with visual features. (b) also integrate the lane marking detection. During this test, the vehicle traveled on the left lane, and we can see that the GPS+dead reckoning approach is biased: it is mainly due to a high uncertainty on the GPS signal, which make the confidence ellipse $E_k$ spatial extent quite important (yellow ellipses). Map-aided localization makes a good estimate of the ego-lane, with a high confidence on the lateral position, the ellipse being stretched only in the longitudinal direction.

On this experiment, the lateral position estimation accuracy is less than 10 cm on average: Figure 8(a) shows an example of trajectory drawn inside the ego-lane (the dotted line is the middle of the left lane). Figure 8(b) illustrated the trajectory in the carriageway during an obstacle avoidance maneuver. However it is important to note that these good results are not only due to the quality of lane detection but also to the quality and accuracy of the mapping. It is obvious that mapping a large area with centimeter accuracy is heavy both in memory level resources than financial cost level.

The lane detection algorithm is also interesting because its accuracy is sufficient to identify the driver maneuvers. In addition, we do not actually use the runway heading information, which can be very useful to update the vehicle heading.

VII. Conclusion

Currently our approach relies on the use of width markings information. However this information is not constant and can therefore generate inaccuracies. In future work, we propose to use a multi-lane detection algorithm with multi-camera fusion, with a range limited to a maximum of 10 meter. Regarding the topology of the cameras, we will test several configurations (one front camera and one rear camera, 2 front cameras,...). What is interesting about our approach is that our traffic lane model is based on polynomial of degree $n$. Generally we limit the polynomials to order 2 to estimate the curvature, the direction and the lateral deviation of the marking. In this application dedicated to the location, only orientation information and lateral deviation are useful. In view of the results of our multi-lane detection algorithm, it is expected that the results will be really good.

Acknowledgement

The authors would like to thanks the french ANR Agence Nationale de la Recherche. The works presented in this paper was funded by the ANR as a part of the CooPerCom project.

References


