An Embedded Multi-Modal System for Object Localization and Tracking

Sergio Alberto Rodriguez Florez, Vincent Fremont, Philippe Bonnifait, Véronique Cherfaoui

An Embedded Multi-Modal System for Object Localization and Tracking

Sergio A. Rodríguez F.^{1,2}, Vincent Frémont^{1,2}, Philippe Bonnifait^{1,2}, Véronique Cherfaoui^{1,2}
^{1}*Université de Technologie de Compiègne (UTC)*, ^{2}*CNRS Heudiasyc UMR 6599, France*

***Abstract*—Reliable obstacle detection and localization is a key issue for driving assistance systems particularly in urban environments. In this article, a multi-modal perception approach is studied in order to enhance vehicle localization and dynamic objects tracking, in a world-centric map. 3D ego-localization is done by merging a stereo vision system and proprioceptive information coming from vehicle sensors. Mobile objects are detected using a multi-layer lidar that is simultaneously used to characterize a zone of interest in order to reduce the complexity of the perception process. Objects localization and tracking is then performed in the fixed frame which simplifies the scene analysis and understanding. Real experimental results are reported to evaluate the performance of the multi-modal system.**

***Index Terms*—Multi-modal perception, visual odometry, object tracking, dynamic map, intelligent vehicles.**

I. Introduction

Advanced Driver Assistance Systems (ADAS) can improve road safety thanks to obstacle detection and avoidance functionalities. In this context, the knowledge of the location and the speed of the surrounding mobile objects constitutes a key information.

In the literature, different approaches address the object localization and tracking problem. Robotics approaches can be used to characterize the static part of the environment [1] and to detect simultaneously moving objects [2]. Leibe et al. have presented in [3] a stereo vision strategy to obtain a 3D dynamic map using a Structure-from-Motion technique and image object detectors. A lidar alone can be used to estimate the ego-motion and to detect mobile objects thanks to a dense 3D grid-map approach [4]. In [5] and [6] real time sensor-referenced approaches (i.e. ego-localization is not considered) are presented using multi-sensor systems showing the complementarity of lidar and vision systems in automotive applications.

A world-centric approach presents interesting properties once the ego-localization is estimated accurately (up to 1 cm per speed unit in Km/h). The tracking performance can be increased since the dynamics of the mobile objects are better modeled. Such an approach also facilitates scene understanding and ADAS implementation.

Ego-localization can be achieved using proprioceptive and exteroceptive sensors [7]. GPS is an affordable system that provides 3D positioning. Unfortunately, its performance can be significantly reduced in urban environments because of multi-paths and satellites outages. Dead-reckoning is a complementary solution. Stereo Vision Systems (SVS), often used for detection and recognition tasks, are also useful for dead-reckoning (also called 3D ego-motion estimation) [8].

Object tracking for ADAS is still an active research domain. Indeed, urban environments are characterized by complex conditions: moving and static objects, mobile perception, varied infrastructures. Object representation [9], [10], association methods [11], motion model and tracking strategies [12] are the key points which have to be considered with a particular attention.

In this work, we study a multi-modal system able to provide a 3D local perception of the environment of the vehicle in a world-centric frame (see Fig. 1). The environment is composed of static and moving objects and a zone of interest is defined in front of the vehicle . The contribution consists in estimating the dynamics of the surrounding objects (location and speed) based on different sensing modalities in order to build a *dynamic* map. Such a map is composed of a list of tracked objects states and the vehicle dynamics evolution in the 3D scene.

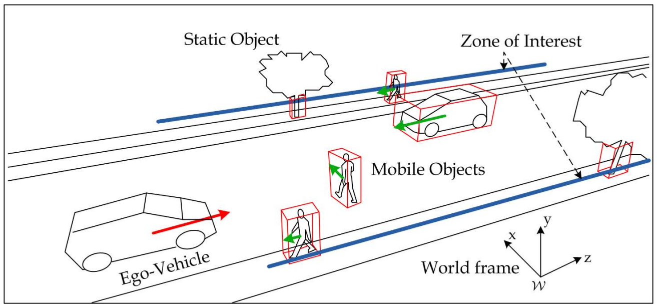

Figure 1. A Dynamic Map

The embedded multi-sensor system uses proprioceptive sensors (i.e. wheel speed sensors and a yaw rate gyro) and two exteroceptive sensors: a Multi-Layer lidar (denoted ML lidar) and a SVS.

The strategy is described on Fig. 2. Firstly, the vehicle ego-localization is done by merging CAN-bus (Controller-Area Network) information with visual odometry. Then, the ML lidar provides a 3D perception of the scene structure. Afterward, objects lying in the zone of interest are localized in the fixed-reference frame by compensating the motion of the ego-vehicle. Finally, objects are tracked in this world frame. Such an information can be exploited by an ADAS to estimate possible collisions.

This paper addresses 3D ego-localization, objects local-

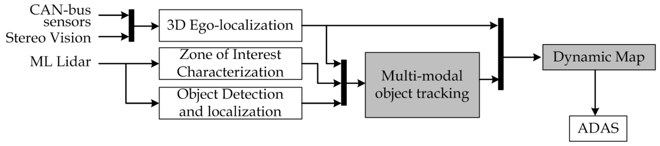
CAN-bus sensors
Stereo Vision
ML Lidar
3D Ego-localization
Zone of Interest Characterization
Object Detection and localization
Multi-modal object tracking
Dynamic Map
ADAS

Figure 2. Multi-Modal Strategy

ization and tracking. First, a detailed description of the embedded multi-sensor system setup is given in section II. In the sequel, each multi-modal function is presented and experimentally discussed. Section III is dedicated to the 3D ego-localization using vision and proprioceptive sensors. Object localization and tracking are studied in section IV.

II. Multi-Modal Perception System

Let us consider a vehicle with a ML lidar, a yaw rate gyro and wheel speed sensors (WSS) accessible through a CAN-bus gateway and a stereo vision system (SVS). These sensors constitute the asynchronous inputs of our perception system.

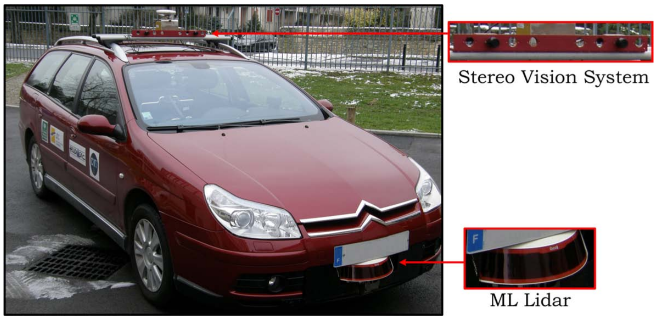

Figure 3. The experimental vehicle with the stereo vision system and the IBEO Alasca XT

A. Multi-Layer lidar

This vehicle is equipped with an IBEO Alasca XT lidar which provides a sparse perception of the 3D environment. Set up at the front of the vehicle (see Fig. 3), this sensor emits 4 crossed-scan-planes with a 3.2° field of view in the vertical direction, 150° in the horizontal direction with a 200m range. The ML lidar measurements (i.e. a 3D points cloud) are reported in a Cartesian frame, denoted $\mathcal{L}$ (X-Front, Y-Left and Z-Up).

B. CAN-bus sensors

A CAN-bus gateway allows accessing the speed of the rear-wheels provided by the WSS of the Anti-lock Braking System (ABS) and the yaw rate provided by a gyro of the Electronic Stability Program (ESP) of the vehicle. These measurements are referenced in a frame located at the middle of the rear-axis of the vehicle, denoted $\mathcal{C}$ (X-left, Y-Down and Z-Front). It has to be noticed that the measurements of the rear wheels are less sensitive to wheel slippage than the ones attached to the traction wheels.

C. Stereo Vision System

A 47cm-baseline Videre SVS has been installed at the top of the vehicle. This SVS is composed of CMOS cameras and 4.5mm lenses providing rectified 320x240 gray-scale images at 30 frames per second.

This system (see Fig. 4) is made as of two projective cameras rigidly joined, horizontally aligned and separated by a baseline distance, b. Both cameras are modeled using a classical pinhole projective model (i.e. the focal length f in pixels units and $[u_0\ v_0]^T$ the image coordinates of the principal point, assuming no distortion and zero skew [13]).

The SVS parameters (i.e. intrinsic and extrinsic) have been estimated using the camera's manufacturer toolbox. The extrinsic calibration of the exteroceptive sensors (i.e. the ML lidar and the two cameras of the SVS) were estimated using the calibration method detailed in [14]. These parameters are denoted by ${}^{\mathcal{L}}[\mathbf{q}\ \mathbf{t}]_{\mathcal{S}}$ which corresponds to the rigid-body transformation (quaternion-translation) from the lidar frame, $\mathcal{L}$, to the vision frame, $\mathcal{S}$. They are necessary to sense information in a common perception space.

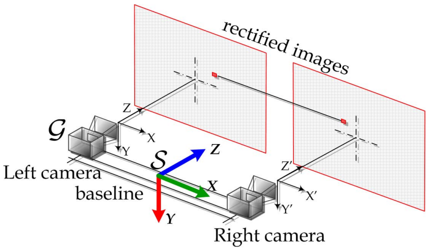

Figure 4. Stereo Vision System Model

SVS provides a 3D space perception (disparity map) computed using the x-image coordinate differences of the same observed feature in the stereo image pair. All measurements are initially expressed in the frame, $\mathcal{G}$, located in the left camera of the SVS. The measurements can be expressed in the stereo-vision-centered frame, $\mathcal{S}$, by the following expression:

$$ {}^{\mathcal{S}}\mathbf{p} = {}^{\mathcal{G}}\mathbf{p} + {}^{\mathcal{G}}\mathbf{t}_{\mathcal{S}} \tag{1} $$

where ${}^{\mathcal{G}}\mathbf{t}_{\mathcal{S}} = [-b/2\ 0\ 0]^T$ corresponds to the rigid-body transformation from $\mathcal{G}$ to $\mathcal{S}$, ${}^{\mathcal{G}}\mathbf{p}$ are the coordinates of a 3D point in the frame $\mathcal{G}$ and ${}^{\mathcal{S}}\mathbf{p}$, the corresponding coordinates in the SVS frame $\mathcal{S}$.

III. 3D Ego-Localization

3D ego-localization consists in estimating the 3D pose of the vehicle as a function of time with respect to a fixed initial frame. Odometry methods using stereo vision systems can provide very precise 3D pose estimations based on quadrifocal constraints as presented by Comport et al. [8]. However, visual odometry may require important computation time.

In order to achieve a good trade-off between precision and execution time, we estimate the 3D vehicle ego-localization using visual odometry [15] aided by the odometry using the CAN-bus sensors measurements.

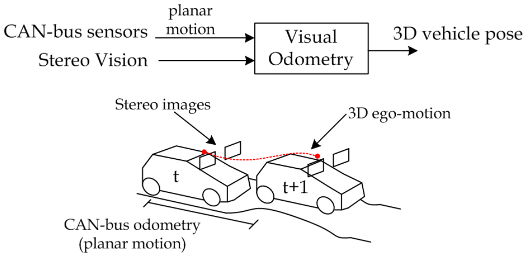

Figure 5. Multi-modal 3D ego-localization scheme

A. Visual odometry aided with CAN-bus sensors

The ego-motion of the vehicle is defined by an elementary transformation (rotation-translation composition, 6 degrees-of-freedom) performed in a interval of time. This estimate is represented by an axis-angle rotation and a translation vector, ${}^{\mathcal{S}_t}\left[\Delta\boldsymbol{\omega}\;\Delta\mathbf{v}\right]^T_{\mathcal{S}_{t+1}}$. First, we estimate an initial planar motion guess using the CAN-bus sensors in an interval of time Δt. Secondly, a 3D visual motion estimation algorithm is initialized with this motion guess and then iteratively refined (see Fig. 5).

Let $\mathcal{C}_t$ be the center of the body frame defined at time t. If the sampling frequency of the gyro and the WSS is high enough (about 40 Hz), the wheel speed is almost constant and the planar ego-motion can be approximated by a circle arc. As illustrated in Fig. 6, the planar ego-motion of the vehicle is modeled as follows [16]:

$$\Delta\boldsymbol{\omega}_0 = \begin{bmatrix} 0 \\ \Delta\theta \\ 0 \end{bmatrix} \quad \Delta\mathbf{v}_0 = \begin{bmatrix} \Delta s \cdot sin(\Delta\theta/2) \\ 0 \\ \Delta s \cdot cos(\Delta\theta/2) \end{bmatrix}$$

where $\Delta\theta$ is the angle obtained by integrating the yaw rate, Δs is the integrated rear-wheel odometry in meters. $\Delta\boldsymbol{\omega}_0$ is a vector representing the axis-angle rotation of the vehicle's motion and $\Delta\mathbf{v}_0$ is a vector representing the estimated displacement of the rear-wheel axis center.

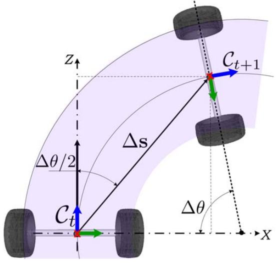

Figure 6. Yaw rate-WSS dead-reckoning for planar odometry estimation

Using successive stereo image pairs, we obtain a set of tracked stereo feature points, p, and their corresponding optical flow constituting the image motion. For this, a set of stereo feature points, p* , is extracted using Harris features associated with a ZNCC (Zero-mean Normal Cross Correlation) correlation criterion and image constraints (disparity and epipolar constraints). The stereo features, p*, are tracked over time using Lucas-Kanade method [17], thus defining the tracked stereo feature points set p.

A stereo feature can be predicted after a 3D motion of the vision system by using a warping function [8] based on geometrical constraints. These constraints are induced by the stereo configuration and the static scene assumption. The idea is to predict the set p̂ as a function of the set p* of stereo features at time t and the vehicle's motion incorporated in the trifocal tensors ${}^{l}\mathcal{T}_i^{jk}$ and ${}^{r}\mathcal{T}_i^{jk}$:

$$\begin{bmatrix} \hat{p} \\ \hat{p}^{'} \end{bmatrix} = \begin{bmatrix} p^{*} l_j^{'}\,{}^{l}\mathcal{T}_i^{jk} \\ p^{'*} l_j\,{}^{r}\mathcal{T}_i^{jk} \end{bmatrix} \tag{2}$$

where l_j and $l_j^{'}$ are the left and right image lines respectively passing through p^* and $p^{'}*$ and perpendicular to the epipolar line. ${}^{l}\mathcal{T}_i^{jk}$, is the trifocal tensor composed by the stereo image pair at time t and the left image at time $t+1$. The second tensor, ${}^{r}\mathcal{T}_i^{jk}$, is composed by the stereo image pair at time t and the right image at time $t+1$. The tensors ${}^{r}\mathcal{T}_i^{jk}$ and ${}^{l}\mathcal{T}_i^{jk}$ are non linear functions of the SVS parameters (i.e. intrinsic and extrinsic) and the vehicle motion ${}^{\mathcal{S}_t}\left[\Delta\boldsymbol{\omega}\;\Delta\mathbf{v}\right]^T_{\mathcal{S}_{t+1}}$.

However, since urban scenes are not only composed of static objects, the static scene assumption is violated. To address this issue, a robust iterative non-linear minimization is performed on the following criterion:

$$\min_{{}^{\mathcal{S}_t}\left[\Delta\boldsymbol{\omega}\;\Delta\mathbf{v}\right]^T_{\mathcal{S}_{t+1}}} (\epsilon) = \sum_{i=1}^{k} \mathbf{W}\left[||p_i - \hat{p}_i|| + ||p_i^{'} - \hat{p}_i^{'}||\right] \tag{3}$$

where k, is the number of tracked stereo feature pairs, p_i and $p_i^{'}$ are the left and right tracked stereo features at time $t{+}1$. $\hat{p}_i$ and $\hat{p}_i^{'}$ are the left and right stereo features at time t warped by the estimated motion (${}^{\mathcal{S}_t}\left[\Delta\boldsymbol{\omega}\;\Delta\mathbf{v}\right]^T_{\mathcal{S}_{t+1}}$) and the warping function stated in equation (2). $\mathbf{W}$ is the weighting matrix estimated by a M-estimator function [18] updated using the Iterative Re-weighted Least Squares algorithm (IRLS).
This robust minimization converges into a solution by rejecting the features points that are mainly generated by mobile objects. The convergence is guaranteed if at least 50% of stereo features points correspond to static objects (i.e. environment). The criterion of Eq. 3 is minimized by using the Levenberg-Marquard Algorithm (LM) on an IRLS loop [18]. The convergence speed of the LM algorithm is increased using the planar ego-motion ${}^{\mathcal{S}_t}\left[\Delta\boldsymbol{\omega}_0\;\Delta\mathbf{v}_0\right]^T_{\mathcal{S}_{t+1}}$, from the CAN-bus sensors. Indeed, this information provides a close initialization guess and then, helps to reduce the iteration cycles.

After convergence, the 3D localization of the vehicle with respect to the initial position $\mathcal{S}_0$ is estimated using the following state evolution equations.

Let $\mathcal{S}_t = \left[\mathbf{q}_t\;\mathbf{p}_t\right]^T$ be the 3D vehicle position at time t with $\mathbf{q}_t = \left[q_0\;q_1\;q_2\;q_3\right]^T$ and $\mathbf{p}_t = \left[p_0\;p_1\;p_2\right]^T$ representing the attitude as a unit quaternion and the vehicle position in meters. Thus, $\mathcal{S}_t$ can be computed as follows:

$$\mathbf{q}_{t+1} = \mathbf{q}_t \star \mathbf{q}(\Delta\boldsymbol{\omega})_{t+1} \tag{4}$$

$$\begin{bmatrix} 0 \\ \mathbf{p}_{t+1} \end{bmatrix} = \mathbf{q}_t \star \begin{bmatrix} 0 \\ \Delta\mathbf{v} \end{bmatrix} \star \bar{\mathbf{q}}_t + \begin{bmatrix} 0 \\ \mathbf{p}_t \end{bmatrix} \tag{5}$$

where $\mathbf{q}(\Delta\boldsymbol{\omega})$ is the unit quaternion corresponding to the estimated axis-angle rotation $\Delta\boldsymbol{\omega}$. The operator $(\star)$ means the quaternion multiplication and $\bar{\mathbf{q}} = [-q_0\ q_1\ q_2\ q_3]^T$ is the conjugated unit quaternion of $\mathbf{q}$.

B. Experimental Real time 3D Ego-Localization Results

A data set has been acquired in a urban environment composed of low-rise buildings, trees and moving objects (i.e. pedestrians and vehicles). During the experiment, the vehicle's speed was less than 30 Km/h. The vehicle trajectory describes a closed loop with pedestrians and vehicles. Low-textured scenes (e.g. rural environments and parking lots) were not considered in this study.

The 3D ego-localization function have been implemented in C/C++. The 3D trajectory is reconstructed in real time and is obtained by integrating the ego-motion estimations. Fig. 7 illustrates one of the performed tests. It consists of a 227m-clockwise loop (i.e. 90 seconds video sequence duration).

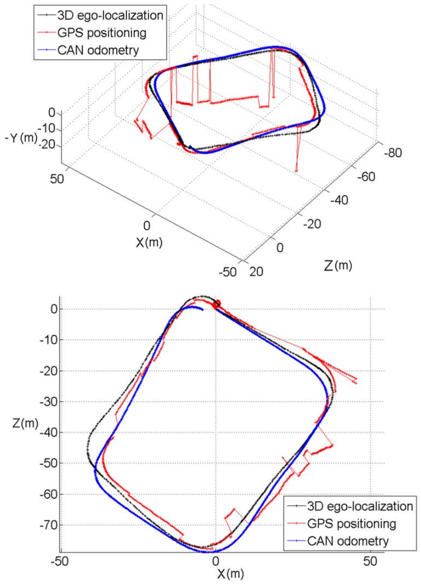

Figure 7. 3D Reconstructed Trajectory

These results show that the cooperative strategy helps to cope with errors of the CAN-sensor based odometry mainly due to wheel slippage. This technique has also improved the visual odometry performance in adverse conditions (e.g. high rotational speed in 90° turns and roundabouts). Those improvements were obtained thanks to the planar motion initialization which avoids any local minima ego-motion solution, improves outlier rejection and reduces the minimization iteration cycles. The 3D ego-localization system performs quite well in situations when GPS can not provide a precise position (see the GPS jumps shown in Fig. 7at the top).

IV. Object Localization and Tracking

The goal of this stage is to estimate the planar trajectory of a set of objects as they move in the 3D scene using a multi-modal approach (i.e. vision, lidar and WSS-Gyro sensing modalities). Object tracking also contributes to keep temporal coherence of the dynamic map and to provide information about the objects speed and size.

The proposed multi-modal strategy starts by detecting objects using a lidar-based technique. Then, objects are tracked using a Kalman filter algorithm [19] with motion constraints in a characterized zone of interest. These assumptions are done in order to simplify the object tracking problem.

The tracking strategy is presented in four parts: object detection, zone of interest characterization, track prediction, object-track association and updating. At the end of this section experimental results are reported to show the effectiveness of the approach.

A. Object Detection

A urban environment constitutes a very complex 3D scenery because of the presence of a large amount of static and mobile objects. Different approaches may be used to reduce the scene complexity (i.e. number of tracked objects) using for instance, the temporal track persistence (i.e. forgetting factor), the dynamic of the tracks and the uncertainty of the track localization. For this study, the 3D observation space for the object detection function is reduced using the detection of a Zone Of Interest (ZOI) based on the prior knowledge of the scene. This function was proposed and implemented in real time by Fayad et al. in [20]. The method is mainly based on lidar scan histogram maxima detection.

The object detection function provides a list of 2D objects at each scan cycle. The detection involves a 3D point clustering stage which can be efficiently implemented using maximal euclidean inter-distance. Geometric predefined features [10] can be an alternative but they require object prior knowledge . The output objects are characterized by their planar position in the lidar frame $\mathcal{L}$, their dimension (i.e. bounding circle) and detection confidence indicators [21].

The list of 2D object positions provided by the ML lidar at time t are transformed into the camera frame $\mathcal{S}$ and finally reported in a world frame (i.e. local dynamic map), $\mathcal{W}$, by compensating the vehicle's motion (see section III-A).

Let ${}^{\mathcal{L}}\mathbf{o} = [x\ y\ 0]^T$ be the coordinates of a 2D object in the lidar frame. Thus, using Eq. 5 their corresponding position in the world frame $\mathcal{W}$ is obtained as follows:

$$\begin{bmatrix} 0 \\ {}^{\mathcal{W}}\mathbf{o} \end{bmatrix} = \mathbf{q}_t \star \left({}^{\mathcal{L}}\mathbf{q}_{\mathcal{S}} \star \begin{bmatrix} 0 \\ {}^{\mathcal{L}}\mathbf{o} \end{bmatrix} \star {}^{\mathcal{L}}\bar{\mathbf{q}}_{\mathcal{S}} + \begin{bmatrix} 0 \\ {}^{\mathcal{L}}\mathbf{t}_{\mathcal{S}} \end{bmatrix} \right) \star \bar{\mathbf{q}}_t + \begin{bmatrix} 0 \\ \mathbf{p}_t \end{bmatrix} \tag{6}$$

Using the localized objects, new tracks are created considering only the objects lying in the zone of interest. The tracks and their state are referenced with respect to the fixed-frame $\mathcal{W}$, and they are managed independently within a Kalman filter. The track state is described by ${}^{\mathcal{W}}\mathbf{x}_t = [x\ z\ v_x\ v_z]^T$ consisting of the ${}^{W}XZ$ plane coordinates $(x_t\ z_t)$ in meters and $(v_x\ v_z)$ the planar velocity in m/s. Additionally, the object size is associated to the track but is not considered in the state.

B. Track Prediction

The track prediction, ${}^{\mathcal{W}}\hat{\mathbf{x}}_t$ at time t, can be calculated from the last track state, ${}^{\mathcal{W}}\mathbf{x}_{t-1}$, and an object motion model, $\mathbf{A}_t$. The motion model is assumed to be planar and linear at a constant speed. Therefore, ${}^{\mathcal{W}}\hat{\mathbf{x}}_t$ is given by:

$$ {}^{\mathcal{W}}\hat{\mathbf{x}}_t = \mathbf{A}_t \cdot {}^{\mathcal{W}}\mathbf{x}_{t-1}, \ \ with \ \mathbf{A}_t = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7) $$

where dt is the time period. The predicted covariance, $\hat{P}_t$, is computed using the covariance matrix of the model noise, $\mathbf{Q}_t$, which takes into account errors due to the motion model assumptions.

C. Track-Object Association and Updating

In this step, the predicted tracks are associated with the objects detected by the lidar under a mono-hypothesis assumption. The implemented association test relies on a nearest neighbor criterion based on the Mahalanobis metric [22]:

$$ \min(d) = \mu_t^T(\hat{P}_t + \mathbf{R})^{-1}\mu_t + \ln(\det(\hat{P}_t + \mathbf{R})) \quad (8) $$

with $\mu_t = \mathbf{C} \cdot {}^{\mathcal{W}}\hat{\mathbf{x}}_t - {}^{\mathcal{W}}\mathbf{o}_{xz}$ and $\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

${}^{\mathcal{W}}\mathbf{o}_{xz}$ represents the XZ coordinates of the detected object in the $\mathcal{W}$ frame and $\mathbf{C}$ is the observation matrix. The first term of Eq. 8, corresponds to the classical Mahalanobis metric and the second term, $\ln(\det(\hat{P}_t + \mathbf{R}))$, corresponds to a weighting factor computed from the track imprecision. The uncertainty of the lidar objects is taken into account by the covariance of the measurement noise, $\mathbf{R}$.

In order to cope with temporal object occlusions, the non associated tracks are kept for a fixed time interval (for example, 2 seconds). However, defining a long prediction time may lead to keep track artifacts. The non associated objects in the zone of interest generate new tracks until the algorithm reaches a predefined criterion of the maximum number of tracked objects. Here, a fixed number of tracks have been set sufficiently high, in order to track all detected objects in the ZOI.

The tracks states and their corresponding covariances are improved by combining the information provided by the associated lidar objects positions and the predicted tracks [22]. For that, we use the Kalman filter update equations.

D. Experimental Results

The 3D ego-localization, the zone of interest characterization and the object detection are real time functions and their results have been logged (see the function scheme in Fig. 2). Fig. 8 shows the mean output frequencies of the 3D localization function and the ML lidar-based functions (i.e. zone of interest characterization and object detection). One can observe that the convergence time of the 3D ego-localization function is not constant because it depends on the vehicle motion and the variability of the scene complexity.

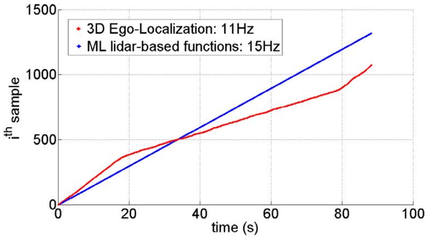

Figure 8. Real Time Output Frequency of 3D Ego-Localization and ML lidar based Functions

The object tracking function has been implemented in MATLAB. The reported results were obtained in an offline process. The input of the tracking algorithm were the logged results of the 3D ego-localization and the ML lidar-based functions.

Fig. 9 illustrates the XZ view of the reconstructed zone of interest in the local map. For this reconstruction, we use the 3D ego-localization of the vehicle and the ML lidar-vision extrinsic parameters results presented in the previous sections. It is important to highlight that at the beginning of the test sequence (i.e. initial position (0,0) on XZ view), the vehicle remains static which shows how the boundaries of the zone of interest converge. These results constitute a very interesting functionality which can be associated to GIS (Geographic Information System) for map-matching applications.

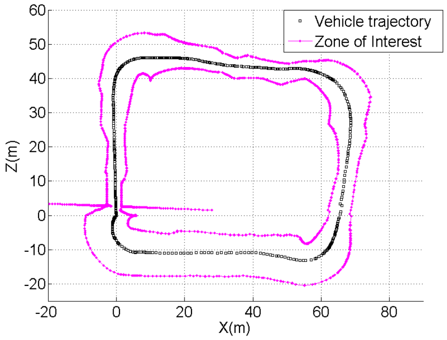

Figure 9. Zone of Interest reconstruction (XZ plane view)

Fig. 10 illustrates a zoomed area of the dynamic map. In this figure, we focus on a tracked vehicle. The size of the track is represented by its bounding circle in red and its center as a red triangle. The detected track size changes as the surface is impacted by the ML lidar. The corresponding image track projections (3D red boxes) and their speed vector (green line) are also illustrated in the upper side of the figure. By observing the image projection of the track speed vector, one can see that the multi-modal system performs quite well.

Fig. 11 shows another section of the dynamic map. No ground truth for the track localization was available during the experiment. However, the reconstructed trajectory corresponds to the observed trajectory followed by the pedestrian in the snapshot sequence.

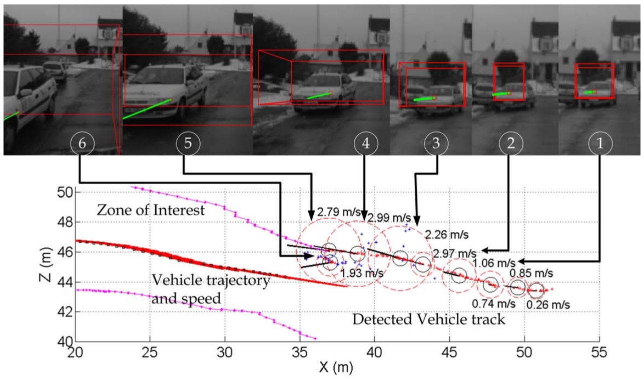

Figure 10. Trajectory of a tracked vehicle (XZ plane view)

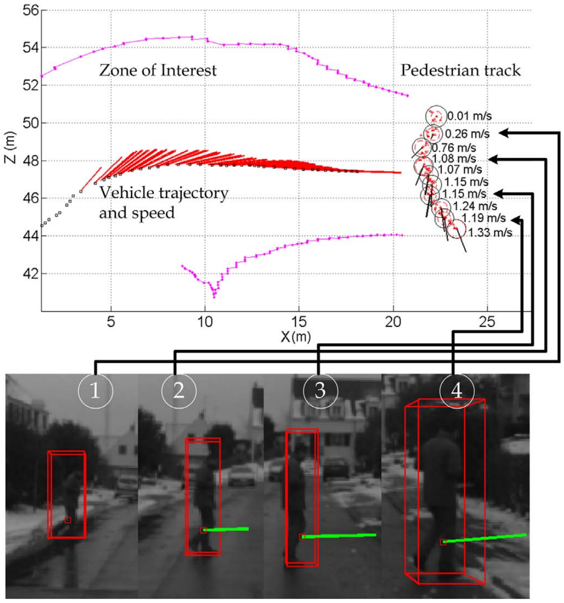

Figure 11. Trajectory of a pedestrian (XZ plane view)

V. Conclusion and Future Work

An embedded multi-modal system for object localization and tracking has been proposed and experimentally validated. The presented approach provides a 3D *dynamic* map of the vehicle surroundings. The method merges sensing information to achieve a robust and precise 3D ego-localization. This function is combined with a lidar-based object tracking focused in a zone of interest providing objects trajectories and speeds as they moves in the space. The obtained results make easy the scene analysis and understanding and can be used for ADAS (e.g. collision detection and avoidance).

One of the perspectives of this research is the improvement of object representation and recognition using multi-model motion track. The main perspective aims to enhance the object tracking taking advantage of a visual confirmation function.

VI. Acknowledgments

The authors want to thank Fadi Fayad for the implementation of the ML lidar functions and Gerald Dherbomez and Thierry Monglon for their experimental support.

References

[1] H. Durrant-Whyte and T. Bailey, “Simultaneous localisation and mapping (slam),” *IEEE Robotics & Automation Magazine*, vol. 13, pp. 99–110/108 – 117, 2006.

[2] C.-C. Wang, C. Thorpe, M. Herbert, S. Thrun, and H. Durrant-Whyte, “Simultaneous localization, mapping and moving object tracking,” *International Journal of Robotics Research*, vol. 26, pp. 889–916, 2007.

[3] B. Leibe, N. Cronelis, K. Cornelis, and L. V. Gool, “Dynamic 3d scene analysis from a moving vehicle,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*, vol. 1, 2007.

[4] T. Miyasaka, Y. Ohama, and Y. Ninomiya, “Ego-motion estimation and moving object tracking using multi-layer lidar,” *IEEE Intelligent Vehicles Symposium*, vol. 1, pp. 151–156, 2009.

[5] A. Broggi, P. Cerri, S. Ghidoni, P. Grisleri, and H. Jung, “A new approach to urban pedestrian detection for automatic braking,” *Journal of Intelligent Vehicles Systems*, vol. 10, no. 4, pp. 594–605, 2009.

[6] R. Labayrade, C. Royere, D. Gruyer, and D. Aubert, “Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner,” *Autonomous Robots*, vol. 19, pp. 117–140, 2005.

[7] C. Cappelle, M. E. E. Najjar, D. Pomorski, and F. Charpillet, “Multi-sensors data fusion using dynamic bayesian network for robotised vehicle geo-localisation,” *IEEE Fusion*, 2008.

[8] A. Comport, E. Malis, and P. Rives, “Accurate quadrifocal tracking for robust 3d visual odometry,” *IEEE International Conference on Robotics and Automation*, pp. 40–45, April 2007.

[9] A. Petrovskaya and S. Thrun, “Model based vehicle tracking in urban environments,” *IEEE International Conference on Robotics and Automation, Workshop on Safe Navigation*, vol. 1, pp. 1–8, 2009.

[10] F. Nashashibi, A. Khammari, and C. Laurgeau, “Vehicle recognition and tracking using a generic multisensor and multialgorithm fusion approach,” *International Journal of Vehicle Autonomous Systems*, vol. 6, pp. 134–154, 2008.

[11] Y. B. Shalom and W. D. Blair, *Multitarget/Multisensor Tracking: Applications and Advances*. Artech House Publishers, 2000.

[12] M. E. Liggins, D. L. Hall, and J. Llinas, *Handbook of Multi-Sensor Data Fusion*, M. E. Liggins, D. L. Hall, and J. Llinas, Eds. CRC Press, 2008.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision. Second Edition*, C. U. Press, Ed. Cambridge, 2003.

[14] S. A. Rodriguez, V. Fremont, and P. Bonnifait, “Influence of intrinsic parameters over extrinsic calibration between a multi-layer lidar and a camera,” *IEEE IROS 2nd Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, vol. 1, pp. 34–39, 2008.

[15] ——, “An experiment of a 3d real-time robust visual odometry for intelligent vehicles,” *IEEE International Conference on Intelligent Transportation Systems*, vol. 1, pp. 226 – 231, 2009.

[16] G. Dudek and M. Jenkin, *Springer Handbook of Robotics*. Springer Berlin Heidelberg, 2008, ch. Inertial Sensors, GPS, and Odometry, pp. 477–490.

[17] J.-Y. Bouguet, “Pyramidal implementation of the lucas kanade feature tracker description of the algorithm,” Intel Corporation Microprocessor Research Labs, Tech. Rep., 2002.

[18] C. V. Stewart, “Robust parameter estimation in computer vision,” *Society for Industrial and Applied Mathematics*, vol. 41, no. 3, pp. 513–537, 1999.

[19] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using Matlab*, Wiley, Ed. Wiley-Interscience Publication, 2001.

[20] F. Fayad and V. Cherfaoui, “Tracking objects using a laser scanner in driving situation based on modeling target shape,” *IEEE Intelligent Vehicles Symposium*, vol. 1, pp. 44–49, 2007.

[21] ——, “Object-level fusion and confidence management in a multi-sensor pedestrian tracking system,” *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Vehicles*, vol. 1, pp. 58–63, 2008.

[22] S. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, S. S. Blackman and R. Popoli, Eds. Artech House, Incorporated, 1999.