Visual Confirmation of Mobile Objects Tracked by a Multi-layer Lidar

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

▶ **To cite this version:**

Sergio Alberto Rodriguez Florez, Vincent Fremont, Philippe Bonnifait, Véronique Cherfaoui. Visual Confirmation of Mobile Objects Tracked by a Multi-layer Lidar. IEEE Conference on Intelligent Transportation Systems (IEEE ITSC2010), Sep 2010, Portugal. pp.IEEE ITSC2010. hal-00521902

HAL Id: hal-00521902
https://hal.science/hal-00521902v1

Submitted on 28 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visual Confirmation of Mobile Objects Tracked by a Multi-layer Lidar

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*—Integrity of the information provided by a perception system is crucial for advanced driver assistance systems intended for safety applications, like obstacle avoidance systems. A method to ensure integrity is to use different kinds of perception sources. Lidars are key sensors for multiple objects detection and tracking. Stereo vision systems (SVS) can be used to improve the tracking but, in this paper, we use also SVS to confirm the real existence of potential obstacles thanks to 3D dense reconstruction in focused regions of interest. Synchronization issues between the different sensors are addressed using predictive filtering. The proposed approach is evaluated in real conditions thanks to five use cases relevant to urban situations. Results show that this visual confirmation strategy is efficient.**

***Index Terms*—Intelligent vehicles, perception, lidar, stereo vision, multi-sensor fusion, obstacle detection and localization**

I. Introduction

Nowadays, intelligent vehicles refer to vehicles able to drive autonomously or to provide pertinent information to the driver mainly for safety reasons. The problem addressed in this paper deals with this latter issue: we focus on Advanced Driver Assistance Systems (ADAS). In this context, the driving activity is in charge of the human driver who is helped by the machine. In case of danger, the driver can be informed through warnings. A Previous work [1] showed that False Alarms (FA) are potentially more harmful than non-detections, because if the system issues two many FA, the driver will switch it off. Let consider an obstacle detection system. Lidars are very efficient systems to detect and localize obstacles. Unfortunately, they can raise alarms on non-hazardous objects such as some located on the edge of the road. A first processing is therefore to retain objects that are actually on the road. This spatial filtering is not sufficient because there may be false alarms on the road due for instance to snowflakes. A stereo vision system can be used in this case to confirm the existence of a real object. For a human driver, it is not possible to transmit both outputs simultaneously, given the short time he has to analyze and react. In this case, the integrity of information provided by the warning system is essential for the driver confidence. One way to strengthen the integrity is to use different perception methods which check that the same causes produce the same effects. Integrity is one of the major attributes related to the performance of ADAS functions. To ensure this, information has to be sensed by at least two different sensors principles [2]. Then, if integrity is checked with respect to a chosen level of confidence, appropriate actions can be done by the ADAS system. We study in the following a method for visual confirmation of a target detected and localized by a lidar. This approach does not treat non-detections of the lidar. On the contrary, it filters false alarms and thus increases the integrity of the alert message.

Exteroceptive perception for automotive applications have been widely studied using multi-modality schemes which often involve radars, lidars and vision systems. Broggi et al. have shown the effectiveness of a multi source pedestrian detector in specific urban situations for an automatic braking system [3]. In [4], a car-following approach is presented using a radar-based vehicle detection function, then the radar results are confirmed using vision. In [5], a different multi-modal fusion system is proposed performing independently vehicle detection using lidar and monocular vision. Perrollaz at al. have proposed in [6], a laser scanner-based obstacle detection and tracking system which confirms objects using stereo vision at a long range using a digital zoomed Region Of Interest (ROI) strategy.

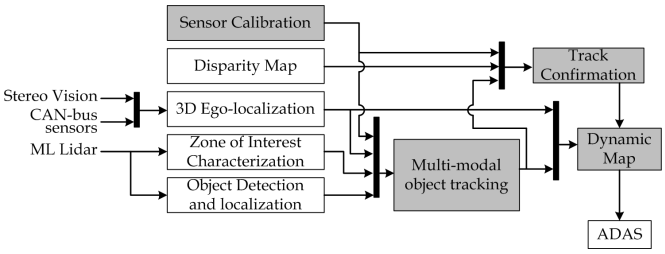

Figure 1. Multi-Modal Perception Block Diagram

The overall strategy is illustrated in Fig. 1. Firstly, surrounding objects are detected using a multi-layer (ML) lidar. The SVS aided by the proprioceptive vehicle sensors is then used to estimate the 3D ego-localization of the vehicle. Subsequently, the detected objects are localized and tracked w.r.t. a world reference frame increasing the tracking performance since the dynamics of the mobile objects are better modeled. Finally, tracked objects are reported in a vehicle-centered frame in order to be confirmed by the SVS, by taking into account the different sampling instants of the two modalities. Confirmed tracks are then declared safe.

In the sequel, section II describes the sensors set-up and their geometrical models. Section III introduces the proposed objects localization and tracking methodology based on lidar data and 3D visual ego-localization. Section IV presents the visual track confirmation technique. Synchronization issues are discussed in section V. Finally, section VI provides experimental results. Evaluating such a system in real conditions constitutes a complex task. This is why the system has been

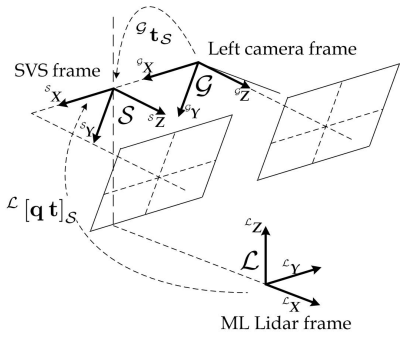

Figure 2. The multi-sensor frames

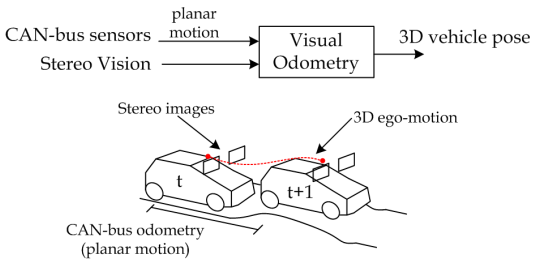

Figure 3. Multi-modal 3D ego-localization scheme

evaluated in specific use cases which are reported in the paper.

II. Multi-sensor system configuration

The considered multi-sensor system is geometrically defined by 3 local sensor frames as illustrated in Fig. 2.

The stereo vision camera system is represented by two classical pinhole camera models (i.e. considering the focal length f in pixels units and $[u_0\ v_0]^T$ the principal point coordinates, no distortion and zero skew [7]) rigidly linked and horizontally aligned on a baseline distance, b. As illustrated in Fig. 2, the reference frame of the SVS denoted $\mathcal{S}$, is located at the middle of the two cameras. Information referenced w.r.t. the left camera frame, $\mathcal{G}$, can be then expressed in the $\mathcal{S}$ frame (X-Right, Y-Down and Z-Front) by a translation ${}^{\mathcal{G}}\mathbf{t}_{\mathcal{S}} = [-b/2\ 0\ 0]^T$. The image pairs delivered by the vision system are rectified and the cameras parameters (i.e. intrinsic and extrinsic) are considered as known.

The ML-lidar provides a sparse perception of the 3D environment. Set up at the front of the vehicle, this sensor emits 4 crossed-scan-planes with a 3.2° field of view in the vertical direction, 140° 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). The ML lidar technology is well adapted for automotive applications since objects can still be observed even when pitch movements occur contrary to the single layer lidars. Additionally, the 4-layer configuration allows the extraction of 3D scene structure attributes e.g. the road plane and sidewalks borders.

In order to sense information in a common perception space, the relative pose of the sensors frames (i.e. SVS and ML-lidar frames) have to be estimated. This is the function done by the sensor calibration module in Fig. 1. Extrinsic parameters can be obtained using the left camera images and the ML-lidar measurements since the frame transformations between cameras composing the SVS are known (see Fig. 2). This process was carried out using the method detailed in [8]. The complete frame transformation from the lidar frame $\mathcal{L}$ into the vision frame $\mathcal{S}$ is noted ${}^{\mathcal{L}}[\mathbf{q}\ \mathbf{t}]_{\mathcal{S}}$ and composed of a unit quaternion and a translation vector.

III. Object Detection and Tracking

The object detection and tracking strategy is done in a fixed frame and is composed of three main stages. First, the vehicle localization is simultaneously estimated with the object detection process. Then, the detected objects are localized and tracked. However, tracking static and mobile objects may constitute a complex task in urban environments since they may be numerous. The third stage reduces the working space into a zone of interest which is detected thanks to the ML-lidar.

A. Vehicle localization

3D ego-localization is carried by the SVS aided by proprioceptive sensors. This task consists in estimating the 3D pose of the vehicle as a function of time with respect to a fixed initial frame. SVS can provide very precise 3D pose estimations based on multiple view geometrical relations (i.e. quadrifocal constraints). Here, 3D vehicle ego-localization is estimated using sparse visual odometry for high speed processing aided by the embedded proprioceptive sensors of the vehicle [9].

The ego-motion of the vehicle is defined as an elementary relative transformation (rotation-translation composition, 6 degrees-of-freedom) performed in a sampling time Δt. This estimate is represented by a rotation and a translation, ${}^{\mathcal{S}_{t-1}}[\Delta\boldsymbol{\omega}\ \Delta\mathbf{v}]^T_{\mathcal{S}_t}$ referenced in the SVS frame, $\mathcal{S}$. Firstly, an initial planar motion guess is computed using the proprioceptive sensors. Secondly, a 3D visual motion estimation algorithm is initialized with this motion guess and is iteratively refined (see Fig. 3). Since the vehicle localization can be required by other asynchronous functions, it has been implemented a predictive filter of the 6 ego-motion parameters. For this, a linear Kalman filter with a constant accelerated model has been implemented since the ego-vehicle can experience important speed changes in breaking maneuvers situations.

Let be $\mathcal{W}$, the world reference frame and $\mathcal{E}$, the ego frame (i.e body-frame) which is linked to the vehicle. The vehicle localization at time t is noted ${}^{\mathcal{W}}\mathcal{S}_t = {}^{\mathcal{W}}[\mathbf{q}_t\ \mathbf{p}_t]^T$ and is represented in the world frame by its attitude - ${}^{\mathcal{W}}\mathbf{q}_t$ a unit quaternion - and its position - ${}^{\mathcal{W}}\mathbf{p}_t$ - a vector in meters. It is obtained as follows:

$$ {}^{\mathcal{W}}\mathbf{q}_t = {}^{\mathcal{S}}\mathbf{q}_{\mathcal{E}} \star {}^{\mathcal{S}}\mathbf{q}_t \quad (1) $$

$$ {}^{\mathcal{W}}\mathbf{p}_t = {}^{\mathcal{S}}\mathbf{q}_{\mathcal{E}} \star \begin{bmatrix} 0 \\ {}^{\mathcal{S}}\mathbf{p}_t \end{bmatrix} \star {}^{\mathcal{S}}\bar{\mathbf{q}}_{\mathcal{E}} \quad (2) $$

where $\star$ denotes the quaternion operator and $\bar{\mathbf{q}}$ represents the corresponding quaternion conjugate. As illustrated in Eq. 1 and 2, the transformation ${}^{\mathcal{S}}\mathbf{q}_{\mathcal{E}}$ is used to compute the

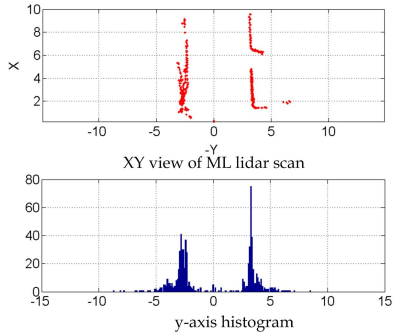

Figure 4. Zone of interest characterization using y-axis histogram

relative orientation of the world frame, $\mathcal{W}$, w.r.t the SVS frame, $\mathcal{S}$, since $\mathcal{W}$ has been chosen as the initial position of the ego frame, $\mathcal{E}_t$, at time $t = 0$ ($\mathcal{W}$ is not coplanar to the road plane). The rigid transformation ${}^{\mathcal{S}}[\mathbf{q}_t\ \mathbf{p}_t]$ corresponds to the visual odometry given by the following equations:

$$ {}^{\mathcal{S}}\mathbf{q}_t = {}^{\mathcal{S}}\mathbf{q}_{t-1} \star \mathbf{q}(\Delta\boldsymbol{\omega}) \quad (3) $$

$$ \begin{bmatrix} 0 \\ {}^{\mathcal{S}}\mathbf{p}_t \end{bmatrix} = {}^{\mathcal{S}}\mathbf{q}_{t-1} \star \begin{bmatrix} 0 \\ \Delta\mathbf{v} \end{bmatrix} \star {}^{\mathcal{S}}\bar{\mathbf{q}}_{t-1} + \begin{bmatrix} 0 \\ {}^{\mathcal{S}}\mathbf{p}_{t-1} \end{bmatrix} \quad (4) $$

where $\mathbf{q}(\Delta\boldsymbol{\omega})$ is the unit quaternion corresponding to the estimated ego-motion ${}^{\mathcal{S}_{t-1}}[\Delta\boldsymbol{\omega}\ \Delta\mathbf{v}]^T_{\mathcal{S}_t}$.

B. Zone Of Interest Characterization

The Zone Of Interest (ZOI) characterization function detects two lateral limits of the navigable space. This function has been proposed in [10] and is mainly based on lidar scan histogram maxima's detection which is represented by two local limits in the x-axis direction of the lidar frame. As illustrated in Figure 4, a 4-layer scan data projected onto the ${}^{\mathcal{L}}xy$ plane (see the upper subplot) provides an easy-to-exploit histogram into the ${}^{\mathcal{L}}y$ axis (see the lower subplot). Objects like security barriers, walls and parked vehicles reduce efficiently the ZOI. The detected limits are finally filtered using a fixed-gain Luenberger observer in order to reduce the oscillations produced by important pitch changes situations. Turns and roundabouts scenes may lead to the lost of histogram peaks. In such a case, no update of the ZOI limits is provided.

C. Objects detection, localization and tracking

Based on each ML-lidar scan, an object detection function delivers a set of surrounding objects obtained by 3D Euclidean distance clustering. These objects are characterized by their planar location (i.e. ${}^{\mathcal{L}}Z = 0$) in the lidar frame $\mathcal{L}$, their dimension (i.e. a bounding circle) and a detection confidence indicator [11] which is estimated using the following criteria:

- The ability of the ML-lidar to detect vertical objects,
- The beam divergence which worsens the measurement precision particularly in situations of non-perpendicular incidence angle,
- The theoretical maximum number of laser impacts (per layer) lying on a detected object. This factor can be computed as a function of the object dimension, the detection range and the laser scanner resolution.

Knowing the 3D localization of the vehicle, the detected objects can be localized with respect to the world frame, $\mathcal{W}$. For instance, let be ${}^{\mathcal{L}}\mathbf{o} = [x\ y\ 0]^T$ the coordinates of a detected object at time t. Its corresponding localization in $\mathcal{W}$ can be computed using two transformations:

$$ \begin{bmatrix} 0 \\ {}^{\mathcal{S}}\mathbf{o} \end{bmatrix} = {}^{\mathcal{L}}\mathbf{q}_{\mathcal{S}} \star \begin{bmatrix} 0 \\ {}^{\mathcal{L}}\mathbf{o} \end{bmatrix} \star {}^{\mathcal{L}}\bar{\mathbf{q}}_{\mathcal{S}} + \begin{bmatrix} 0 \\ {}^{\mathcal{L}}\mathbf{t}_{\mathcal{S}} \end{bmatrix} \quad (5) $$

$$ \begin{bmatrix} 0 \\ {}^{\mathcal{W}}\mathbf{o} \end{bmatrix} = {}^{\mathcal{W}}\mathbf{q}_t \star \begin{bmatrix} 0 \\ {}^{\mathcal{S}}\mathbf{o} \end{bmatrix} \star {}^{\mathcal{W}}\bar{\mathbf{q}}_t + \begin{bmatrix} 0 \\ {}^{\mathcal{W}}\mathbf{p}_t \end{bmatrix} \quad (6) $$

where ${}^{\mathcal{S}}\mathbf{o}$ and ${}^{\mathcal{W}}\mathbf{o}$ are the corresponding coordinates of the detected object in the SVS and the world frame respectively.

Only the detected objects lying in the ZOI are localized w.r.t. the world frame using Eq. 5 and 6. Then, they are tracked independently using Kalman filters. This tracking helps to robustify the perception scheme. Assuming that the motion of the objects is linear and uniform:

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

where ${}^{\mathcal{W}}\hat{\mathbf{x}}_t$ is the predicted state of the tracked object, $\mathbf{A}_t$ is the state transition matrix, Δt is the sampling time period (which is not considered constant) and ${}^{\mathcal{W}}\mathbf{x}_{t-1} = [x\ z\ v_x\ v_z]^T$ is the state vector consisting of the ${}^{\mathcal{W}}XZ$ plane coordinates $(x\ z)$ in meters and $(v_x\ v_z)$ the planar velocity in m/s. The object size is considered as an attribute of the track but is not included in the state.

At each ML-lidar sampling, the tracks are updated. For this, the new detected objects and tracks are associated using the nearest neighbor criterion [12] (i.e. $min(d)$) where the metric is computed as follows:

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

with $\mu_t = \mathbf{C} \cdot {}^{\mathcal{W}}\hat{\mathbf{x}}_t - {}^{\mathcal{W}}\mathbf{o}_{xz}$ and $\mathbf{C} = [\mathbf{I}_{2\times2}\ \mathbf{0}_{2\times2}]$

${}^{\mathcal{W}}\mathbf{o}_{xz}$ represents the XZ coordinates of the detected object in the $\mathcal{W}$ frame, $\mathbf{C}$ the observation matrix and $\hat{P}_t$ the covariance matrix of the predicted state, ${}^{\mathcal{W}}\hat{\mathbf{x}}_t$. The uncertainties of the lidar objects localization and the object motion model are taken into account through the covariance of the measurement noise, $\mathbf{R}$ and the covariance of the state transition model, $\mathbf{Q}$.

It is worth to recall that the object tracking stage increases the robustness of the system by allowing objects occlusion and since tracks contain information confirmed several times by the same source (the ML lidar here).

IV. Visual Track Confirmation

The multi-modal system presented up to this section provides a precise localization of the vehicle and the surrounding objects representing potential obstacles with respect to a fixed-reference frame. However, the object detection relies on a single source: the ML-lidar. Visual track confirmation is proposed here as a way to increase the integrity of the

information provided by the system.
It is performed using the following strategy. Firstly, each lidar-tracked object is transformed into the ego frame, $\mathcal{E}$, and its corresponding bounding cylinder (lidar bounding circle at an arbitrary height) is reprojected into the stereo images. In each image, this provides a Region Of Interest (ROI). Secondly, the pixels composing the ROI are reconstructed by stereo in the 3D space in order to provide a 3D points cloud. Afterward, this set of 3D points is segmented into 2 clusters assuming that the ROI is composed of two classes: the object and the background. Finally, the track is confirmed if one of the 3D points clusters is associated using a Mahalanobis distance thresholded at a given confidence level.

A. Region Of Interest in the images

A track is localized in the ego frame, $\mathcal{E}$ at time t by doing:

$$\begin{bmatrix} 0 \\ {}^{\mathcal{E}}\mathbf{t} \end{bmatrix} = {}^{\mathcal{S}}\mathbf{q}_{\mathcal{E}} \star \left({}^{\mathcal{W}}\bar{\mathbf{q}}_t \star \begin{bmatrix} 0 \\ {}^{\mathcal{W}}\mathbf{t} - {}^{\mathcal{W}}\mathbf{p}_t \end{bmatrix} \star {}^{\mathcal{W}}\mathbf{q}_t \right) \star {}^{\mathcal{S}}\bar{\mathbf{q}}_{\mathcal{E}} \quad (9)$$

where ${}^{\mathcal{E}}\mathbf{t}$ is the resulting position of the track in $\mathcal{E}$ and ${}^{\mathcal{W}}\mathbf{t}$ is the 3D position of the object in the world map.

The ROI is characterized by re-projecting the bounding box vertex of the track. These vertexes are estimated from the track size (the track height is known a priori) and its 3D centroid position (see Fig. 5).

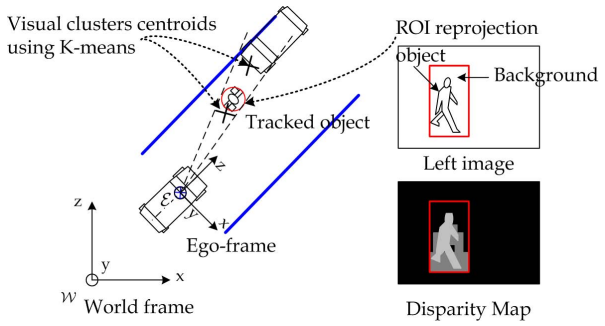

Figure 5. Visual 3D Track Confirmation

The track position is projected into the image plane by the following equation:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \mathbf{K} \cdot \left(\left({}^{\mathcal{S}}\bar{\mathbf{q}}_{\mathcal{E}} \star \begin{bmatrix} 0 \\ {}^{\mathcal{E}}\mathbf{t} \end{bmatrix} \star {}^{\mathcal{S}}\mathbf{q}_{\mathcal{E}} \right) - {}^{\mathcal{G}}\mathbf{t}_{\mathcal{S}} \right) \quad (10)$$

where $u\,v$ are the image coordinates and $\mathbf{K}$ is the intrinsic camera matrix . The operator $\sim$ represents up to a scale factor.

B. 3D dense reconstruction of the ROI

Each pair of images contains 3D dense information of the scene since the pixel images correspondence and the camera parameters are known. This information can be represented by a disparity map which is considered in this study referenced w.r.t. the left camera of the SVS.

The 3D dense reconstruction of the ROI consists in firstly overlapping the ROI and the disparity map as illustrated in Fig. 6. Then, the set of corresponding disparity values are extracted. Finally, the 3D coordinates of each pixel are estimated by performing a classical triangulation process [7].

Figure 6. 3D dense reconstruction of the ROI

C. Track confirmation

The existence of a tracked object is confirmed if its 3D position matches with the visual 3D points cloud. However, the ROI usually contains the observed track and also the scene background as illustrated in Fig. 6.

Assuming that the objects and the scene background in the ROI are distinguishable in the 3D space (see Fig. 5), the reconstructed 3D points cloud is clustered into two classes: track and background. For this, the K-means method [13] is very well adapted. It is based here on an Euclidean distance between 2 clusters characterized by theirs centroids, ${}^{\mathcal{E}}\mathbf{c}_{1,2}$, and their associated points. The 3D cluster corresponding to the tracked object is determined based on a Mahalanobis distance, ξ, with respect to the track position. This distance is formalized by the following expression:

$$\xi = \left[{}^{\mathcal{E}}\mathbf{c}_{xz} - {}^{\mathcal{E}}\mathbf{t}_{xz} \right] \cdot (P_{\mathbf{c}} + P)^{-1} \cdot \left[{}^{\mathcal{E}}\mathbf{c}_{xz} - {}^{\mathcal{E}}\mathbf{t}_{xz} \right]^T \quad (11)$$

where ${}^{\mathcal{E}}\mathbf{c}_{xz}$ and ${}^{\mathcal{E}}\mathbf{t}_{xz}$ are the ${}^{\mathcal{E}}XZ$ coordinates of the centroid cluster to be tested and the tracked object respectively. P is the covariance of the tracked object location and $P_{\mathbf{c}}$ is the covariance of the cluster centroid which is estimated based on its depth stereo reconstruction error and modeled by a score function [14].

$$P_{\mathbf{c}} = \begin{bmatrix} \frac{1}{k_1 \cdot \tau} & 0 \\ 0 & \frac{1}{k_2 \cdot \tau} \end{bmatrix} \quad (12)$$

$$with \quad \tau = \begin{cases} 1 - (1 - (\alpha \cdot \frac{b \cdot f}{{}^{\mathcal{E}}\mathbf{c}_z})^2) & , \alpha < \frac{{}^{\mathcal{E}}\mathbf{c}_z}{b \cdot f} \\ 1 & , \alpha \geq \frac{{}^{\mathcal{E}}\mathbf{c}_z}{b \cdot f} \end{cases} \quad (13)$$

where τ is a score which takes into account the confidence on the 3D reconstruction as a function of the depth observed w.r.t the SVS and a tolerance error factor, α in meters. The weighting parameters k_1 and k_2 have been chosen considering that reconstruction errors in depth have more impact in the ${}^{\mathcal{E}}Z$ axis direction (i.e. $k_2 > k_1$).

Finally, the 3D cluster centroid which satisfies the nearest neighbor criterion with respect to the track position is associated to the track, ${}^{\mathcal{W}}\mathbf{x}_t$. A gating of the Mahalanobis distance is then applied to confirm the real existence of the tracked object. Integrity is checked since two independent sources have been used.

V. Synchronization Issues

The multi-modal perception system is made of three main functions interacting together: disparity map computation, 3D-ego localization and object detection. These functions are asynchronous and run in different threads at different

frequencies (26, 16 and 15 Hz respectively). In order to solve this asynchronicity, predictive filters have been used and referenced on precisely stamped data. The treatments performed by the system over time are:

- New 3D-ego localization is available: the predictive filter of the ego-motion is time-updated and its state is corrected.
- New objects are detected by the ML-lidar: the last known vehicle localization is predicted up to this time. Then, these objects are localized in the world frame and the tracked objects are updated.
- A new disparity map is available: the vehicle localization and the tracked objects are extrapolated at this time. Predicted objects are localized in the ego-frame using the predicted vehicle pose. Tracks are proposed to confirmation using the disparity map.

Fig. 7 shows an example of possible measurements arrival. At t_0, the disparity map and the localization information are available but there is no object. Thus, only vehicle pose is updated. At t_1, objects have been detected. They are localized using the predicted vehicle pose. At t_2, tracks can be confirmed using the predictions of the objects and of the vehicle pose.

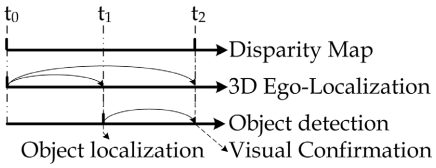

Figure 7. Example of possible data arrival

We have noticed experimentally that this mechanism deeply increases the performance particularly when the scene is composed of high dynamics.

VI. Experimental results

The results reported here were obtained in real conditions using an experimental vehicle of the Heudiasyc Laboratory. As illustrated in Fig. 8, the vehicle was equipped with a 47cm-baseline Videre SVS. This SVS is composed of two CMOS cameras with 4.5mm lens set-up to acquire 320x240 gray-scale images at 30 ftps. An IBEO Alasca XT lidar installed at the font of the vehicle provides a sparse perception of the 3D environments at 15 Hz. A CAN-bus gateway provides the speed of the rear-wheels (WSS) and the yaw rate of the vehicle (from the ESP).

The disparity map, 3D-ego localization and object detection functions have been processed in real time and their outputs were logged. These results have been provided as input to the object tracking and track confirmation functions that have been tested under Matlab by post-processing.

A. Use Cases and Evaluation Methodology

In order to evaluate the the performance of the system, we report experimental results of the visual confirmation function through five sequences. These use cases are relevant to common scenarios in urban environments. Fig. 9 gives a graphical

Figure 8. The experimental vehicle "Carmen"

description of the evaluated situations involving two kinds of mobile objects: pedestrians, wheel chair pedestrians and cars.

In the reported experiments, the ML-lidar didn't performed any miss-detection. The evaluation methodology aims at quantifying the percentage of time that the object tracking function is made unavailable because of a visual non confirmation.

The ground truth was referenced manually in the left image plane of the SVS: the center point coordinates of the observed objects of interest was selected, frame by frame. All objects considered in the ground truth were localized in a common perception region for the SVS and the ML-lidar. The confirmation track rate was checked by counting the times when the bounding box of the confirmed track contains the ground truth.

B. Real Data Results

The results obtained thanks to the ground truth are reported in Table I. A total of 650 frames in 5 different situations have shown that at least 81% of the time, the detected objects of interest were confirmed by the two modalities. If one can conclude that the visual track confirmation may some times decrease the true positive rate of the system, it should be noticed that the confirmed tracks ensures the integrity of the perception process. In use cases C and E, it was observed that important changes of vehicle pitch angle can influence the precision of the object tracking, since object motion is considered to be planar and the vehicle pitch angle is unknown.

Video Sequence	A	B	C	D	E
Duration (s)	5	4	5	6	10
Number of Analyzed Frames	110	90	90	125	235
Number of scans	78	51	62	94	137
Positioning updates	72	37	65	121	160
Visual Confirmation Rate (%)	100	100	81.8	98.5	83.5

Table I
RATE OF DETECTED OBJECTS CONFIRMED BY VISION

In Fig. 10, the left side illustrates the world map where the ego-vehicle and the detected objects are localized and tracked. The right side of the figure shows the reconstructed points of the ROI image. By observing the ego-map, one can notice that one of the centroids of the clustered points cloud has been associated to the track. This association confirms the detected object as illustrated in the upper image of Fig. 11.

Fig. 11 presents few examples of confirmed objects. Their bounding box (in red) and their speed vector projection (in green) show a quite good localization even for fast objects. This results validate the synchronization strategy.

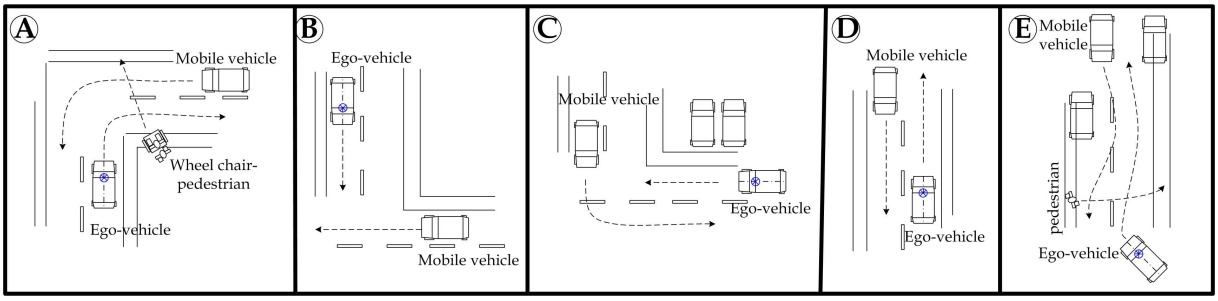

Figure 9. Use cases considered in the evaluation test

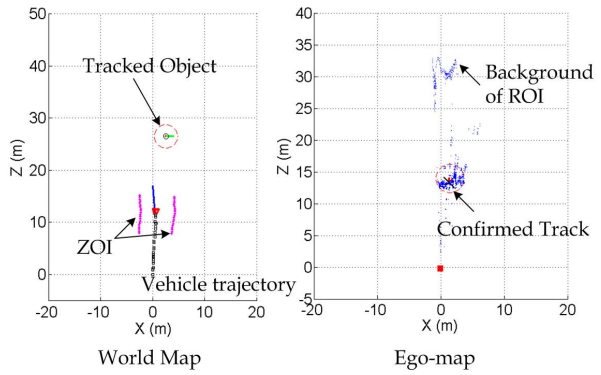

Figure 10. Example of a confirmed tracked object using the SVS

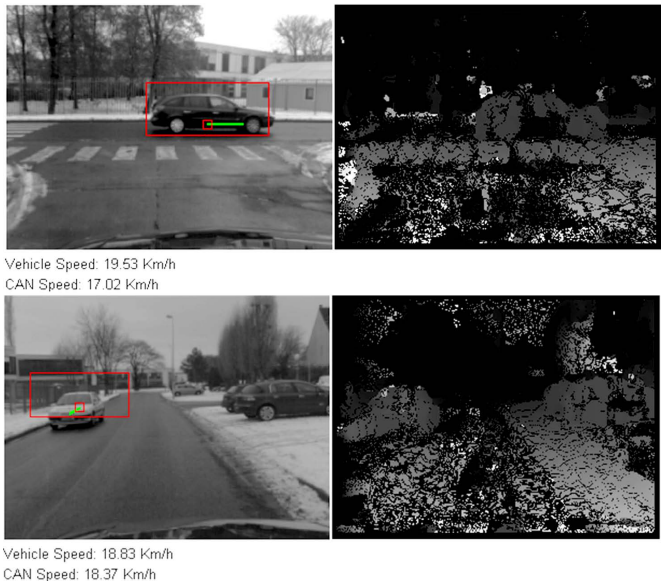

Figure 11. Confirmed objects

VII. Conclusion

An asynchronous multi-modal object localization and tracking system was presented and studied experimentally. This approach takes advantage of the broad functional spectrum of stereo vision systems. Synchronization issues were taken into account to ensure the temporal system consistency. A visual confirmation strategy was proposed to check the integrity of the information provided. The approach consists in focusing on ROI which are processed in a dense way. 3D points are reconstructed and compared to the lidar-tracked object. This method was tested in five different scenarios proving a good confirmation rate. Indeed, even if the visual confirmation reduces inevitably the availability of the detection function, the rate obtained seems compatible with the development of ADAS functions.

VIII. Acknowledgments

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

References

[1] J. P. Bliss and S. A. Acton, "Alarm mistrust in automobiles: how collision alarm reliability affects driving," *Applied Ergonomics*, vol. 34, no. 6, pp. 499 – 509, 2003.

[2] C. Stiller, *Intelligent Vehicle Technologies, Theory and Applications*. Butterworth Heinemann, 2001, ch. Towards Intelligent Automotive Vision Systems, pp. 113–128.

[3] 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.

[4] A. Haselhoff, A. Kummert, and G. Schneider, "Radar-vision fusion with an application to car-following using an improved adaboost detection algorithm," *IEEE Intelligent Transportation Systems Conference*, vol. 1, pp. 854 – 858, 2007.

[5] 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.

[6] M. Perrollaz, R. Labayarde, C. Royere, N. Hautiere, and D. Aubert, "Long range obstacle detection using laser scanner and stereo vision," *IEEE Intelligent Vehicles Symposium*, vol. 1, pp. 182–187, 2006.

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

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

[9] ——, "An experiment of a 3d real-time robust visual odometry for intelligent vehicles," in *IEEE International Conference on Intelligent Transportation Systems*, vol. 1, Saint Louis, USA, 2009, pp. 226 – 231.

[10] 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.

[11] ——, "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.

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

[13] J. MacQueen, "Some methods for classification and analisys multivariate observations," *Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

[14] S. D. Blostein and T. S. Huang, "Error analisys in stereo determination of 3d point positions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 752–765, 1987.