

# Vehicle Detection and Tracking by Collaborative Fusion Between Laser Scanner and Camera

Dominique Gruyer, Aurélien Cord and Rachid Belaroussi

**Abstract**—This paper presents a new approach to fuse 3D and 2D information in a driver assistance setup, in particular to perform obstacle detection and tracking. We propose a new cooperative fusion method between two exteroceptive sensors: it is able to address highly non linear dynamic configuration without any assumption on the driving maneuver. Information are provided by a mono-layer laser scanner and a monocular camera which are unsynchronized. The initial detection stage is performed using the 1D laser data, which computes clusters of points which might correspond to vehicles present on the road. These clusters are projected to the image to define targets, which will be tracked using image registration techniques. This multi-object association and tracking scheme is implemented using belief theory integrating temporal and spatial information, which allows the estimation of the dynamic state of the tracks and to monitor appearance and disappearance of obstacles. Accuracy of the method is evaluated on a database made publicly available, focus is cast on the relative localization of the vehicle ahead: estimations of its longitudinal and lateral distances are analysed.

## I. INTRODUCTION

For many on-board automotive driver assistance systems DAS (such as collision avoidance, blind spot monitoring, adaptive cruise control, or parking assistant), robust and reliable vehicle detection is a critical step. On-road vehicle detection concerns systems where sensors are mounted on the vehicle rather than being fixed on the infrastructure such as cameras for traffic monitoring systems [1].

The most common vehicle detection systems are using active sensors: laser, radar or sonar. Such sensors detect the distance of objects by measuring the travel time of a signal they emitted after its reflection by the object. Laser scanners are popular sensors for such a purpose [2], [3]: they are usually mounted on the front bumper and perform a horizontal scanning; objects are detected on a given horizontal plane (mono-layer). The data coming from laser scanner are easier to cluster than radar and they are more accurate. Moreover, it is easier to quantify the reliability and to model the uncertainties of such data. However, laser sensors fail to overcome some situations such as non-planar road configuration, or a varying pitch angle due to the ego-vehicle maneuver depending on an acceleration or road shape variations (turns, road bumps . . . ). Radar are less subject to such issues, but their radio waves energy are reverberated by walls in a tunnel (wave guide effect); they can also be reflected by objects that can be safely overridden (metal plate, a guardrail or a Botts' dot).

Authors are with IFSTTAR, COSYS, LIVIC, 77 rue des chantiers, F-78000, Versailles, France, e-mail: dominique.gruyet@ifsttar.fr

Passive sensors such as cameras provide a refined and more complete view of the environment at a lower cost. Visual information is also interesting as recognition of different kind of shapes can be performed on videos (lane detection, traffic sign recognition, visual odometry, pedestrian detection), so an increasing number of DAS systems already include one or several on-board cameras. An extensive survey on visual-based approaches for on-road vehicle detection and tracking can be found in [4]. Detection methods are classified into three categories: knowledge-based [5] (edges, corners, colors, texture), stereo-based [6], [7] (disparity, inverse perspective mapping) and motion-based [8] (optical flow).

Systems based solely on computer vision are not powerful enough to handle complex traffic situations: multiple sensors, active and passive, are required. They can be used in a collaborative way as in [7]: a stereoscopic camera rig is used to validate the targets provided by a laser scanner; the outputs of the two filtered sensors are then merged by checking redundancy. In [9], a Lidar and a camera datas are processed providing a set of targets: the sum rule is used to combine the classifiers outputs. A more elaborated way of combining a laser rangefinder and a camera is proposed in [1] for a traffic surveillance application (sensors are fixed on the infrastructure). The telemetric data are incorporated in the likelihood function of a particle filter tracking vehicles motion in the image. In track-to-track fusion systems [10], each local sensor data is filtered to provide a list of objects sent to a central fusion module that fuses all the local sensors objects lists into a single global objects list. Local sensor-level tracks are fused asynchronously using the information matrix fusion algorithm. In these works, the issue of data association (identifying which object of two sensors correspond to the same target) is not raised.

In this paper, we present a new approach to efficiently detect and track on-road vehicles using multiple sensors, namely a laser scanner and a camera: the focus is made on the issue of data association of simultaneous measurements from multiple sensors. In our approach, detection and tracking are addressed in a unified framework: targets coming from laser data processing are used in order to build and to manage tracks (tracking stage). This tracking step allows to improve target knowledge by use of temporal and spatial information. With a propagation module, a confidence index is computed for each track. This index quantifies the accumulation of temporal evidence about target existence.

Another issue in the field of vehicle detection and tracking is the lack of representative benchmarks and evaluation

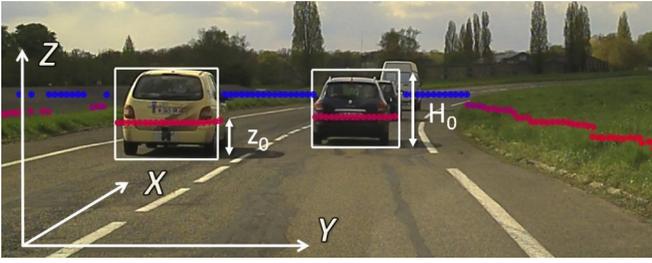


Fig. 1. Laser points are on the same layer  $Z = z_0$ . A point's color is related to its distance: from red (closest) to blue (out of range). After clustering, targets are filtered and plotted with a white rectangle.

procedures. For example, the method proposed in [6] is not evaluated quantitatively, although it is illustrated by some images. Performances of [5] are presented by ROC curves on a test set made of positive and negative examples extracted from 420 road images, when a more relevant test would process each image with a sliding window at different resolution. The approach proposed in [11] is evaluated on situations with only 2 objects by counting the number of targets detected at each time. Accuracy of their localization is not assessed. In [10], longitudinal and lateral localization errors are estimated; the evaluation is carried out of an overtaking scenario with one target vehicle for 25 seconds. In this paper, we present a new dataset made available to the broader scientific community. To facilitate comparison, we present our results in terms of localization accuracy of the vehicle ahead. Qualitative illustrations are also available in the form of online videos.

## II. METHOD OVERVIEW

The experimental setup is made of a laser scanner and a mono-camera mounted on a vehicle. The camera is calibrated: its focal and pixels size are known, as well as its position relative to the laser scanner. The laser telemeter is a mono-layer scanner with a given field of view (in degree) and maximal range (in meter): its measurement noise characteristics are known.

Obstacles detection is done with the laser scanner in its XYZ frame: tracking of these objects is implemented in the image plane. A set  $\mathcal{P} = \{P_j, j \in [1, N_p]\}$  of  $N_p$  regions of interest (ROI) define the tracks previously pursued: at a given time, either laser datas or an image are processed. The laser sensor generates a set of points in a horizontal XY plane: they are clustered according to their relative distance and the sensor noise model. A set of targets is thus defined and projected in the image plane giving a set  $\mathcal{C}$  of  $N_c$  ROIs:  $\mathcal{C} = \{C_i, i \in [1, N_c]\}$ . When a target of  $\mathcal{C}$  can be associated to a track of  $\mathcal{P}$ , the corresponding ROIs are merged and the confidence in this track is incremented. If no track matches a target, a new track is created and added to the set  $\mathcal{P}$  with an initially small confidence value.

When the camera provides a new image, a registration technic is used to track the ROIs of  $\mathcal{P}$ . Local features are matched between images at time  $t$  and  $t-1$  using a pyramid decomposition: the position that yields the best match is

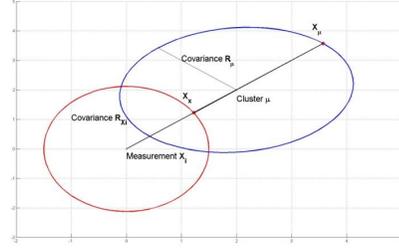


Fig. 2. Clustering of a measurement.

used to update the ROI. Confidence assigned to each track is decreased depending on the last time it was matched with a laser target.

## III. VEHICLE DETECTION USING A LASER SCANNER

### A. Similarity Distance Between Laser Impacts and Clusters

The laser scanner outputs a set of points in a (X,Y) plane illustrated in Fig. 1. To infer the set of objects in the observed scene, an iterative clustering process is implemented. A first cluster is defined by the first laser point. Whether a new laser point 'i' belongs to an existing cluster 'mu', or a new one, depends on a distance function. The chosen distance  $D_{i,\mu}$ , must comply with the following criteria [12]:

- $D_{i,\mu}$  ranges from 0 (i is part of cluster  $\mu$  with a full confidence) to 1 (small confidence on the membership) if measurement i is inside cluster  $\mu$ .
- $D_{i,\mu} > 1$  if point i is out of cluster  $\mu$ .
- $D_{i,\mu}$  has the properties of distance functions.

$D_{i,\mu}$  estimation uses cluster and measurement covariance matrices  $R_\mu$  and  $R_X$ .

$$D_{i,\mu} = \frac{\|X_i - \mu\|}{\|X_\mu - \mu\| + \|X_X - X_i\|} \quad (1)$$

$\mu$  is the center of the cluster, center  $X_i$  is the point measured by the laser, as shown in Fig.2 with an illustration of  $R_\mu$  and  $R_X$ .  $X_\mu$  and  $X_X$  are computed from the covariance matrices  $R_\mu$  and  $R_X$ .

Distance  $\rho_0$  and angle  $\theta_0$  being the polar representation of the measurement covariance matrix [13]:

$$R_X = \begin{bmatrix} \sigma_{\rho_0}^2 & \sigma_{\theta_0}^2 \rho_0^2 \\ \sigma_{\rho_0}^2 \rho_0 \cos \theta_0 & \sigma_{\theta_0}^2 \rho_0^2 \sin \theta_0 \end{bmatrix} \quad (2)$$

with, using a first order approximation:

$$\sigma_{x_0}^2 = \sigma_{\rho_0}^2 \cos^2 \theta_0 + \sigma_{\theta_0}^2 \rho_0^2 \sin^2 \theta_0 \quad (3)$$

$$\sigma_{y_0}^2 = \sigma_{\rho_0}^2 \sin^2 \theta_0 + \sigma_{\theta_0}^2 \rho_0^2 \cos^2 \theta_0 \quad (4)$$

$$\sigma_{x_0 y_0}^2 = \frac{1}{2} \sin^2 \theta_0 (\sigma_{\rho_0}^2 - \sigma_{\theta_0}^2 \rho_0^2) \quad (5)$$

$\sigma_{\rho_0}^2$  and  $\sigma_{\theta_0}^2$  are the variances on the distance and angle of a measurement.

From each covariance matrix, the eigenvalues  $\{\sigma_1, \sigma_2\}$  and the eigenvectors  $\{\vec{V}_1, \vec{V}_2\}$  are extracted. The parametric equation of the corresponding ellipse is then:

$$x = V_{11} |\sigma_1| \cos \Phi + V_{12} |\sigma_2| \sin \Phi \quad (6)$$

$$y = V_{21} |\sigma_1| \cos \Phi + V_{22} |\sigma_2| \sin \Phi \quad (7)$$

$x$  and  $y$  give the position of any point on the ellipse.

The equation of the line joining  $X_i$  and  $\mu$  is linear:  $y = ax + b$ . The intersection between the ellipse and the line are points with the following parameter:

$$\Phi = -\arctan \frac{|\sigma_1|(V_{21} - aV_{11})}{|\sigma_2|(V_{22} - aV_{12})} \text{ with } \Phi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \quad (8)$$

Two solutions are possible using  $R_\mu$ 's eigenvectors  $P_\mu = [\vec{V}_1 \vec{V}_2]$  and its eigenvalues  $(\sigma_1, \sigma_2)$ :

$$X_\mu = P_\mu \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} \begin{bmatrix} \cos \Phi \\ \sin \Phi \end{bmatrix} \text{ and } X_\mu = P_\mu \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} \begin{bmatrix} \cos(\Phi + \pi) \\ \sin(\Phi + \pi) \end{bmatrix} \quad (9)$$

$X_X$  is estimated the same way: point  $X_i$  belong to cluster  $\mu$  if  $D_{i,\mu} \leq 1$ .

### B. Obstacle Detection by Dynamic Programming

Dissimilarity between a laser scanning impact and a given cluster can be computed using Eq. 1. The association of a point with a cluster can go two ways:

- if the impact does not correspond to any clusters then a new one is created: it is centered on the point and is attributed the noise model of laser data to represent its imprecision,
- else, the impact is associated to an existing class: the point is added to the set of points of this cluster; its center and covariance matrix are updated.

The classification process outputs are centimetric 2D positions of each class in the horizontal plane XY.

Once the initial set of clusters is defined, the classification process based on the dissimilarity is re-run on the classes in order to merge the connected ones. This procedure is recursively called until convergence: a final set of targets is reached.

Fig. 3 shows the clustering result of batch of scanner data. Each cluster is represented with its position and size along the two axes. In the following stage, cluster are filtered according to their spatial extent in order to focus on vehicles.

### C. Target Projection in the Image

Once a cluster is defined in the XY plane it is checked for its width: this stage selects target whose size is compliant to the rear view of a vehicle. Objects' width should be equal to 1.80 meters  $\pm 30$  cm, else they are discarded: this threshold is common to the one used in [6]. This stage enables to reduce the number of false positives. Under this configuration, motorbikes and persons are not considered: detectable vehicles ranges from small cars to trucks. Fig. 1 shows the targets remaining after filtering according to their width (white rectangles). Actually an extension of this filter module allows to identify several classes of object from geometrical attributes and some intrinsic criteria. As the scanner used has only one layer, the set of points belongs to the same plane at a height equal to  $z_0$  from the ground. The spatial extent (along the Z-axis) of a detected object is defined by a rectangle. This  $\mathbf{r}$  rectangle is with a height  $H_0 = 1.60$  meters.

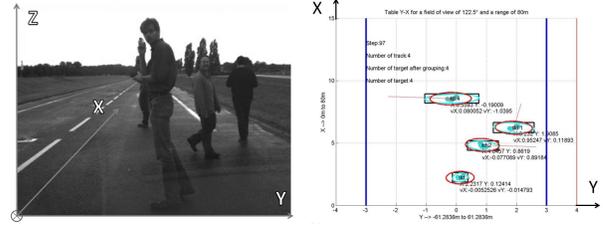


Fig. 3. Left: the source scene scanned contains 4 obstacles. Right: laser points and the 4 filtered clusters in the XY plane.

Projecting a point  $P$  of the real world onto the image plane requires two reference frames: the laser frame  $\mathcal{R}_l$ , and the camera frame  $\mathcal{R}_c$ . The rigid transformation between  $\mathcal{R}_l$  and  $\mathcal{R}_c$  is measured during the experimental setup. Camera's *intrinsic parameters* are also known beforehand, so that the projection of rectangle  $\mathbf{r}$  onto the image plane is only a matter of product of transform matrix.

This step of the system enables data association and tracking in the image plane  $(u, v)$ .

### IV. TARGET PROPAGATION BASED ON IMAGE MOTION ANALYSIS

Once a new image is produced by the camera, the motion of each ROI  $P_j$  of  $\mathcal{P}$  is evaluated. The affine transformation of the considered ROI from the previous image to the new one is calculated using the Motion2D software [14]. This software is an object-oriented library to estimate 2D parametric motion models in an images sequence. It exploits only the spatio-temporal derivatives of the image intensity function. It is an extended and optimized implementation of the robust, multi-resolution and incremental estimation method. For each  $P_j$ , the position is updated by the best match of the algorithm. Finally, the confidence  $\gamma_p$  granted to  $P_j$  is decreased.

This steps allows us to ensure a good tracking of objects, even when the laser scanner temporally loses the target. This could happen if there are slope variation on the roads inducing a strong variation of vehicle's pitch.

### V. TARGET ASSOCIATION AND TRACKING BASED ON BELIEF THEORY

To monitor appearance or disappearance of tracks and estimate the dynamic state of targets generated by the laser scanner, a multi-objects association is implemented. Locations of known objects are predicted at time  $t$  using the image registration outputs eventually coupled with a Kalman filter. These objects are already pursued tracks and will be denoted  $P_j$  in what follows:  $P_j \in \mathcal{P}$ ,  $j = \{1, \dots, N_p\}$ . Detected objects at time  $t$  are targets designated by  $C_i$ :  $C_i \in \mathcal{C}$ ,  $i = \{1, \dots, N_c\}$ . The multi-objects association uses the the belief theory [15].

The current problem is to associate a perceived ROI  $C_i$  to known tracks  $\{P_j\}$ . It can also be stated as the identification of object  $C_i$  among a set of  $N_p$  propositions  $\{P_j\}$ , one of them supposedly being true.

A basic belief (mass  $m_\theta(\cdot)$ ) characterizes a proposition: it is defined in a  $[0, 1]$  interval. In order to achieve the

generalisation of the Dempster combination rule and thus reducing its combinatorial complexity, the reference frame of definition is limited with the constraint that a perceived object (target) can be connected with one and only one known object (track). For example, for a perceived object, in order to associate among it to three known objects, the frame of discernment is  $\Omega = \{P_1, P_2, P_3, P_\otimes\}$ , where  $P_i$  means that "target C and track  $P_i$  are the same object". To complete the frame of discernment another hypothesis noted  $P_\otimes$  is added ([16]). It represents the proposition "target is not associated with any of the existing tracks". Each  $P_j$  corresponds a local view of the world (an already existing track) while  $P_\otimes$  represents the rest of the world (a new track to pursue).

The definition of the *basic belief assignment* is directly related to the data association applications. The basic belief assignment on the association between  $C_i$  and  $P_j$  is the function  $m_{ij}()$  defined on the following frame of discernment  $\Omega$ :

$$\Omega = \{P_1, P_2, \dots, P_{N_p}, P_\otimes\} \quad (10)$$

and more precisely on focal elements  $\{P, \bar{P}, \Omega\}$  where  $\bar{P}$  means *not P*.

Each one has the following meaning:

- $m_i\{C_i\}(P_j)$  is the degree of belief on the proposition *target  $C_i$  is associated with track  $P_j$* .
- $m_i\{C_i\}(\bar{P}_j)$  mass on the proposition *target  $C_i$  is not associated with track  $P_j$* .
- $m_i\{C_i\}(\Omega)$ : degree of *ignorance about the association between  $C_i$  and  $P_j$* .
- $m_i\{C_i\}(P_\otimes)$ : mass allocated to rejection:  $C_i$  is in relation with nothing ( $C_i$  constitutes a brand new track).

In the mass distribution  $m_i\{X\}()$ , X denotes the perceived objects and the index i alone indicates that the mass is applied to all known objects (tracks).

To compute these mass distribution, an initial mass set must be defined. A first matrix  $\mathfrak{D} = (d_{ij})$  stores the rate of overlap of the ROI of C and P. Let us call  $A_R$  the area of a rectangle R. The rate of overlap  $d_{ij}$  between target  $C_i$  and track  $P_j$  - is computed as follow:

$$d_{ij} = \frac{A_{C_i \cap P_j}}{A_{C_i \cup P_j}} \quad (11)$$

Therefore,  $\mathfrak{D}$  is a  $N_c \times N_p$  matrix containing quantities related to the similarity between targets and tracks: measurements related to target  $C_i$  are stored in the  $i^{th}$  line, those related to track  $P_j$  can be found in the  $j^{th}$  column.

From  $\mathfrak{D}$  three matrices  $N_c \times N_p$  are computed to represent the initial set of mass. Elements  $m_{ij}(P_j)$ ,  $m_{ij}(\bar{P}_j)$  and  $m_{ij}(\Omega)$  are respectively the belief on the association of target  $C_i$  with track  $P_j$ , not with track  $P_j$  and the ignorance on this association, without taking the other tracks into account. These specialized belief assignment are defined given a local view of the association of  $C_i$  with  $P_j$ . The closer  $d_{ij}$  is to 1, the more similar the two rectangles are; in this case, there is a high likelihood that  $C_i$  and  $P_j$  correspond to the same object. These initial masses are plotted with respect to  $d_{ij}$  in Fig. 4:  $m_{ij}(P_j)$  is represented by the solid red line,  $m_{ij}(\bar{P}_j)$

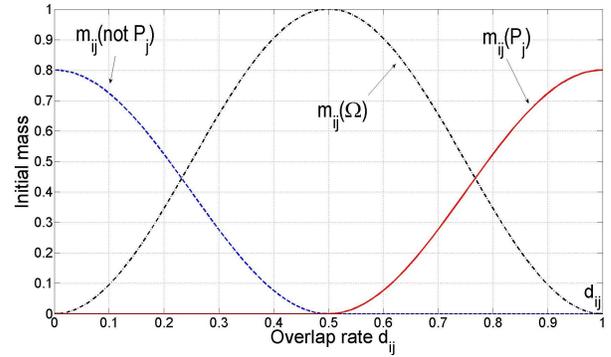


Fig. 4. Elementary belief functions with  $\alpha = 0.8$ .

by the dotted blue line and  $m_{ij}(\Omega)$  in dotted thin black line.  $\alpha$  is a parameter linked to the confidence granted to the laser sensor.

The generalized combination rules are computed without any recurrent stage based on the formalism given in [17]. A strong exclusive hypothesis is made : an object cannot be at the same time associated and not associated to another object. It results in a reduction of the influence of conflicts and a decrease in complexity [18]. The rules are :

$$m_i\{C_i\}(P_j) = m_{ij}(P_j) \prod_{k \neq j} (1 - m_{ik}(P_k)) \quad (12)$$

$$m_i\{C_i\}(P_\otimes) = \prod_{j=1}^{N_p} m_{ij}(\bar{P}_j) \quad (13)$$

$$m_i\{C_i\}(\Omega) = \prod_{j=1}^{N_p} m_{ij}(\Omega) \quad (14)$$

By repeating these operations for each  $C_i$  a set of  $N_c$  basic belief assignment are obtained:

$$m_1\{C_1\}(), m_2\{C_2\}(), \dots, m_{N_c}\{C_{N_c}\}() \quad (15)$$

To reach a decision, a pignistic transformation is applied for each  $m_i\{C_i\}()$  with  $i \in [1..N_c]$ .  $(m_i\{C_i\}(P_j))_{ij}$  is a  $N_c$ -by- $N_p$  matrix, whilst  $(m_i\{C_i\}(\emptyset))_i$  dimension is  $N_c$ -by-1. By concatenating these two matrices, the pignistic probabilities  $BetP_{\{C_i\}(P_j)}$  of each  $P_j$  hypothesis are summarized in a matrix corresponding to the target point of view:

$$BetP_{\{C_i\}(P_j)} = \begin{bmatrix} (m_i\{C_i\}(P_j))_{1j} & (m_i\{C_i\}(P_\otimes))_1 \\ \dots & \dots \\ (m_i\{C_i\}(P_j))_{N_c j} & (m_i\{C_i\}(P_\otimes))_{N_c} \end{bmatrix} \quad (16)$$

$\underbrace{\hspace{10em}}_{N_p \text{ columns}}$

This matrix is the pignistic probabilities of each target not taking into account the other targets : columns are independent. A dual approach should assess the association of a track with the targets in order to have the tracks' point of view.

From the tracks perspective, the frame of discernment is  $\Theta = \{C_1, C_2, \dots, C_{N_c}, C_\otimes\}$ . For one track  $P_j$ , the specialized basic belief assignments are then:

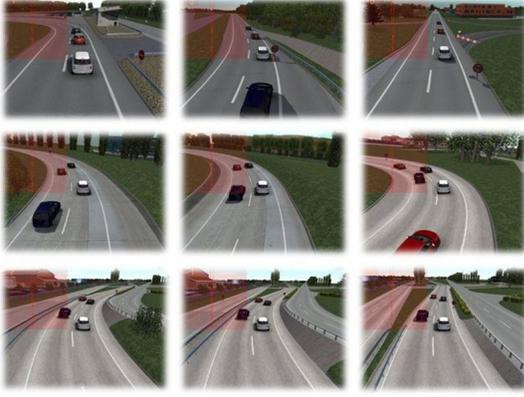


Fig. 5. SiVIC's visual rendering: multiple captures from the Versailles-Satory's test tracks scenery.

- $m_{ij}^\ominus(C_i) = m_{ij}(P_j)$ : degree of belief on the proposition  $P_j$  is associated with  $C_i$ .
- $m_{ij}^\ominus(\bar{C}_i) = m_{ij}(\bar{P}_j)$ : belief on the proposition  $P_j$  is not associated with  $C_i$ .
- $m_{ij}^\ominus(\Theta) = m_{ij}(\Omega)$ : degree of ignorance on the association between  $P_j$  and  $C_i$ .

By combining them using Eq. 12 and 13, the basic belief assignment  $m_j^\ominus\{P_j\}()$  is estimated. Repeating these operations for all tracks  $P_j$ , a set of  $N_p$  basic belief assignments is defined:

$$m_1^\ominus\{P_1\}(), m_2^\ominus\{P_2\}(), \dots, m_{N_p}^\ominus\{P_{N_p}\}() \quad (17)$$

A second matrix is obtained involving the pignistic probabilities  $BetP_{\{P_j\}(C_i)}^\ominus$  about the tracks by concatenating the  $N_p$ -by- $N_c$  matrix  $(m_j\{P_j\}(C_i))_{ji}$  and the  $N_c$ -by-1 matrix  $(m_j\{P_j\}(\emptyset))_j$ :

$$BetP_{\{P_j\}(C)}^\ominus = \begin{bmatrix} (m_1\{P_1\}(C_i))_{1i} & (m_1\{P_1\}(P_\ominus))_1 \\ \vdots & \vdots \\ (m_{N_p}\{P_{N_p}\}(C_i))_{N_p i} & (m_{N_p}\{P_{N_p}\}(P_\ominus))_{N_p} \end{bmatrix} \quad (18)$$

$N_c$  columns:  $i=1\dots N_c$

The last stage of this algorithm consists in establishing the best decision from the previously computed associations using the two pignistic probabilities matrices  $BetP_{\{C_i\}(P_j)}$  and  $BetP_{\{P_j\}(C_i)}^\ominus$ . The decision stage is done with the maximum pignistic probability rule. This rule is applied on each row of both pignistic probabilities matrices. With the first matrix, this rule answers to the question "which track  $P_{k_i}^*$  is associated with target  $C_i$ ?"

$$P_i^* = \arg \max_{j=1\dots N_p+1} [BetP_{\{C_i\}(P_j)}] \quad (19)$$

With the second matrix, this rule answers to the question "which target  $C_{k_j}^*$  is associated to the track  $P_j$ ?"

$$C_j^* = \arg \max_{i=1\dots N_c+1} [BetP_{\{P_j\}(C_i)}^\ominus] \quad (20)$$

Let us suppose that target  $C_i$  is associated with track  $P_{i_0} = P_i^*$  according to Eq. 19: in the best case scenario, the target  $C_{i_0}^*$

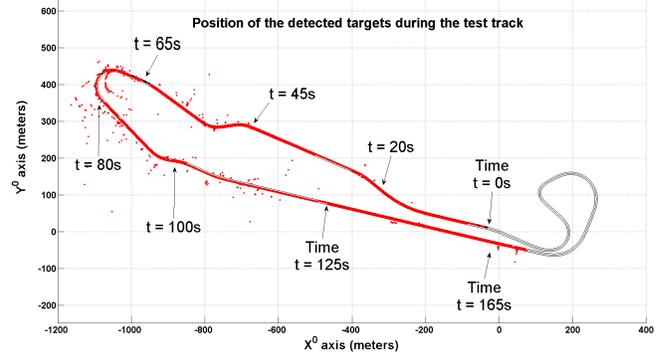


Fig. 6. Overview of all the target detected along the VeLaSCa track.

associated with track  $P_{i_0}$  using the second pignistic matrix (Eq. 20) is target  $C_i$  ( $C_{i_0}^* = C_i$ ). Then it can be safely inferred that target  $C_i$  is the object already tracked  $P_{i_0}$ , or a new object if  $P_i^* = \emptyset$ .

Sometimes the decision extracted from one pignistic matrix is equivocal or there can be a conflict between the decisions reached by the two pignistic matrices. Conflicts and ambiguities are both solved by an assignment method called the Hungarian algorithm detailed in [19].

## VI. EXPERIMENTAL RESULTS

### A. Presentation

In this section, the results are reported according to two viewpoints:

- detection and tracking of all the vehicles in front of the ego-vehicle are illustrated in two scenarios: on a highway with a dense traffic flow, and on simulated images of Satory's test facility.
- precision of the relative longitudinal and lateral localization of the vehicle ahead is evaluated on Satory's test track.

Real world tests were performed on a highway. They illustrate qualitatively the proposed approach, but no ground truth was available. To finely evaluate the performances of the system, a realistic simulation was used, providing a data logger of all the vehicles present during the test (especially their absolute position in the world frame).

### B. The VeLaSCa database

To produce a quantitative evaluation of the data sensors fusion, a driving environment including several vehicles and their accurate location is required. For this purpose, a software named SiVIC<sup>TM</sup> was used [20], [21]. This virtual platform was developed at LIVIC since 2003 in order to model and to prototype exteroceptive and proprioceptive sensors embedded on a vehicle. It is a platform used to prototype virtual sensors (cameras, radar, laser, GPS...). Its objective is to reproduce, in the most truthful manner, the realistic aspect of a situation, the behavior of a vehicle and the functioning of the sensors that could be embedded on such vehicle. The main advantages of SiVIC are to simulate

situations that are difficult to reproduce in real life, to allow the use of several sensors such as laser scanner, camera, GPS receiver and inertial measurement unit IMU. The images simulated are photo-realistic as shown on Fig. 5.

The database *VeLaSCa* simulates the journey of an ego-vehicle moving on a two-lane road with 5 vehicles used as targets, on the the Satory (Versailles, France) test track. It is a test road of 3.5 km with a 3D topography accurately measured by professional land surveyors and used in order to build the Satory's 3D realistic virtual environment. The *VeLaSCa* test is run during 165 seconds on a subpart of the Satory's track: this 2.5 km long portion is timestamped on Fig. 6. The following data are logged:

- camera data with grey level  $640 \times 480$  images at 25 img/s: a total of 4100 images are recorded, some examples are illustrated on Fig. 8 at different times.
- laser scanner data from the ego-vehicle: this sensor scans a  $100^\circ$  angle aperture with a 80 meters of range, a  $0.25^\circ$  resolution, and a 25Hz frequency. This sensor provides at each time a frame of 401 laser impacts. The detected and filtered objects coming from the laser data processing are displayed in Fig. 6 (red points).
- current positioning  $(X_0, Y_0, Z_0)$  and heading of the ego-vehicle in the world frame.
- current positioning for each obstacle in both the world frame and the relative ego-vehicle referential. Datas  $(X_k - X_0, Y_k - Y_0)$  for  $k \in [1..5]$  are the record of the longitudinal (depth) and lateral relative position of the obstacles.

The database is publicly available at: <http://www.inrets.fr/linstitut/unites-de-recherche-unites-de-service/livic/logiciels/databases/velasca.html>

### C. Vehicles detection and tracking results

Figure 8 shows examples of detection and tracking using the presented approach. Laser impacts are drawn as colored circles with a hot colormap: from red for a close point to blue if it is out of range. When a target has been detected, a white rectangle is drawn at its location.

Tracks' ROI are drawn using a varying color depending on their confidence value  $\gamma_p$ : red while  $\gamma_p < 0.6$ , orange if  $0.6 < \gamma_p < 0.8$  and green if  $\gamma_p > 0.8$ . Typically when a target has just been detected the corresponding track is red (Fig. 8(g)). If it is confirmed by the laser scanner in the following frames, its  $\gamma_p$  value increase and its color turns to orange (Fig. 8(h)), then green (Fig. 8(i) and (j)) meaning the track is indeed (with a great confidence) a vehicle. When a track is not confirmed during some successive frames (for example when the vehicle is out of range or occluded by another obstacle), its  $\gamma_p$  value is decreased, and its color turns to orange first then fade to red before it disappears from the  $\mathcal{P}$  set.

Fig. 7(a) illustrates the confidence value of the car ahead of the ego-vehicle (track  $P_1$ ). In the first second  $\gamma_p$  reaches its maximal value because the  $P_1$  is very close (10 meters) to the ego-vehicle in the beginning of the test. At time  $t=65s$ , the  $\gamma_p$  slowly decreases because the laser impacts do not hit

the vehicle ahead for a few seconds:  $P_1$  is at 50 m and not at the same altitude as the ego-vehicle (the road is non planar), as illustrated on Fig. 8(e) and (f). At time  $t=125s$ , the same  $P_1$  outage phenomena occurs at 60 m long enough to induce a short lost of  $P_1$ : a laser target retrigger its tracking 800 ms later.

We posted some examples of videos:

- [www.youtube.com/watch?v=MmejKj-aA5M](http://www.youtube.com/watch?v=MmejKj-aA5M) on a highway (A86 in France) at a high speed (70 mph) with a dense traffic flow,
- [www.youtube.com/watch?v=g2mLq4yrySM](http://www.youtube.com/watch?v=g2mLq4yrySM): on *VeLaSCa* database (test track of Versailles Satory).

Our approach handles correctly multiple moving vehicles with very few false positives.

### D. Performances on relative localization

The first track  $P_1$  is the car ahead of the ego-vehicle. This first track is correctly followed during 125 seconds, then its confidence  $\gamma_p$  drops below 0.6 so the tracking is cancelled (for 800 ms, before a laser detection retrigger it, as shown in Fig. 7(a)). Fig. 7 illustrates the relative localization of the first vehicle ahead during this test.

In the world frame, the ego-vehicle is localized at point  $(X_0, Y_0, Z_0)$ . Fig. 7(b) gives the longitudinal positioning of  $P_1$  relatively to the ego-vehicle frame of reference:  $X_1 - X_0$ . It is the depth of the vehicle ahead, its distance to the ego-vehicle: it is an important index of the current state of risk of the ego-vehicle. Knowing the speed of the ego-vehicle, this inter-vehicle distance can be used to assess the local risk level of a current situation. Moreover with use of the ego-vehicle speed, it is possible to determine if the inter vehicular safety time is respected. The red curve is the measurement obtained with the proposed approach, the green line is the ground truth. The precision obtained is less than one meter, with an error of 80 cm on average on the *VeLaSCa* database. Most of the dispersion on this error occurs at time  $t=65s$  when track  $P_1$  is almost lost: the distance is overestimated at this particular time, due to the lack of track evolution model. A risk assessment system should take into account the low confidence value  $\gamma_p$ . In [10] the accuracy reported on an overtaking scenario is about 30 cm during the 10 seconds where the other vehicle is in front of the ego-vehicle. Such an accuracy is obtained at great cost, by fusing a four layer laser scanner, a radar and a camera.

Fig. 7(c) is the relative lateral position of the car immediately ahead  $Y_1 - Y_0$ . The red line is the measurement obtained, it is very close to the ground truth (green line). The lateral positioning can reach a centimetric accuracy with an average absolute error of only 15 cm, close to the one reported in [10] but with much less sensors.

## VII. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a generic solution for road environment perception applicable to low cost sensors. The proposed method is based on a collaborative fusion approach with two sensor types: one telemetric sensor and one optical sensor. In this configuration, the telemetric sensor allows to

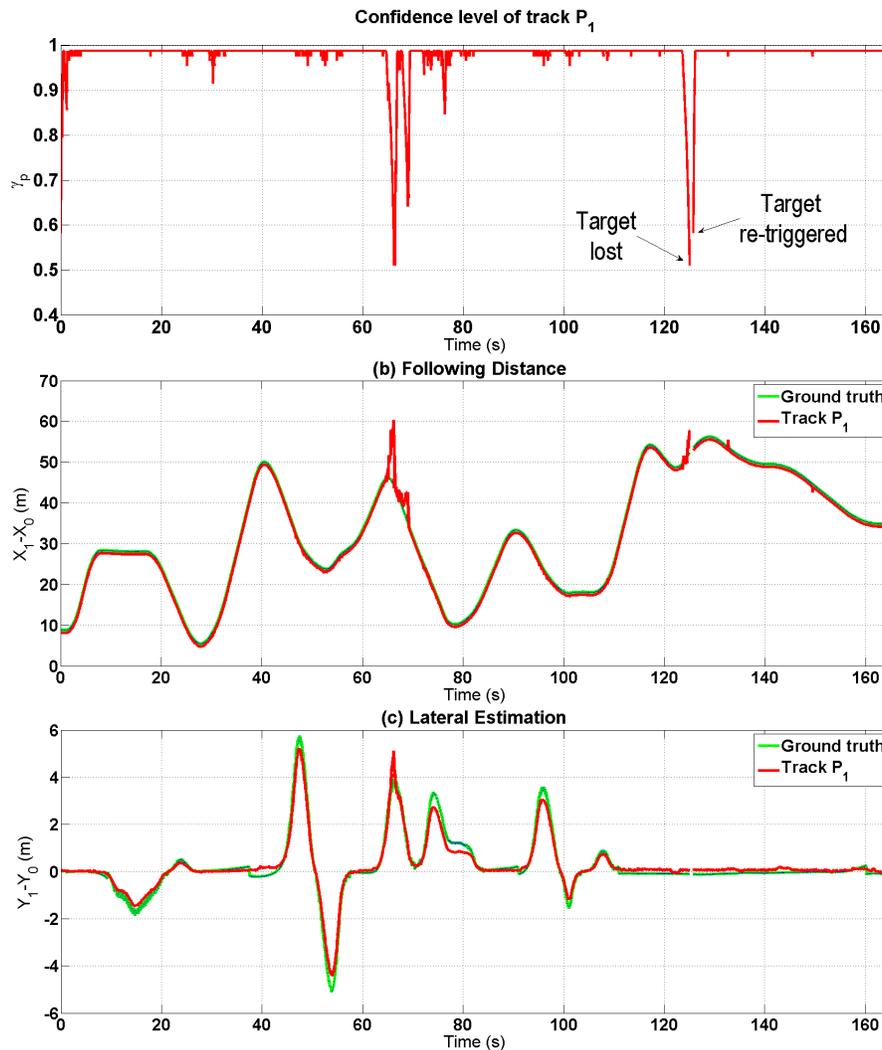


Fig. 7. Confidence value  $\gamma_p$ : evolution during VeLaSCa test run (a). Relative position of the car ahead.  $X_1 - X_0$  is the longitudinal position of track  $P_1$  relative to the ego-vehicle (b).  $Y_1 - Y_0$  is the lateral relative position (c).

provide targets in the forward view of the host vehicle with a characterization of uncertainty on each target. This telemetric sensor can be replaced by any other type of telemetric sensor like RADAR. Nevertheless, the laser scanner appears to perfectly match with perception task specifications. In the presented approach, the main goal of this telemetric sensor and attached algorithmic processing is to provide to the fusion and image processing modules an accurate and reliable description of targets used to initialize the regions of interest (ROI).

In the fusion loop, the regions of interest initiates the track creation in the video processing. Once a track is created then the image processing loops on tracks propagation, association and update steps in the image referential.

The association stage allowing to manage the object tracking is based on belief theory. This approach has been adapted in order to track objects in the image space and to provide a generic way to manage the appearance, disappearance and propagation of tracks. In addition, this approach offers the ability to manage conflicts and ambiguities. Unlike many

approaches of collaborative fusion, this approach is not affected by temporary and potential troubles in the target generation (laser process). In fact, this step is only required for the generation of the zones of interest. Once these ROI are created, the image processing module (Motion2D) coupled to the tracking module ensures a vehicle tracking in good condition. However, if the laser processing provides continuous detection of targets then the fusion stage will be better and the track attributes assessment will be more accurate and certain.

Moreover, it is important to mention that this approach can work without the use of vehicle evolution model unlike other tracking approaches. In our case, no a priori is done on the track dynamics. This allows taking into account the nonlinearities in obstacle maneuvers. The spatial and temporal track alignment, in order to perform an association stage between targets and tracks, is carried out by the expansion of the tracks ROI.

In future developments, several types of telemetric sensors with varied performances will be implemented in order to

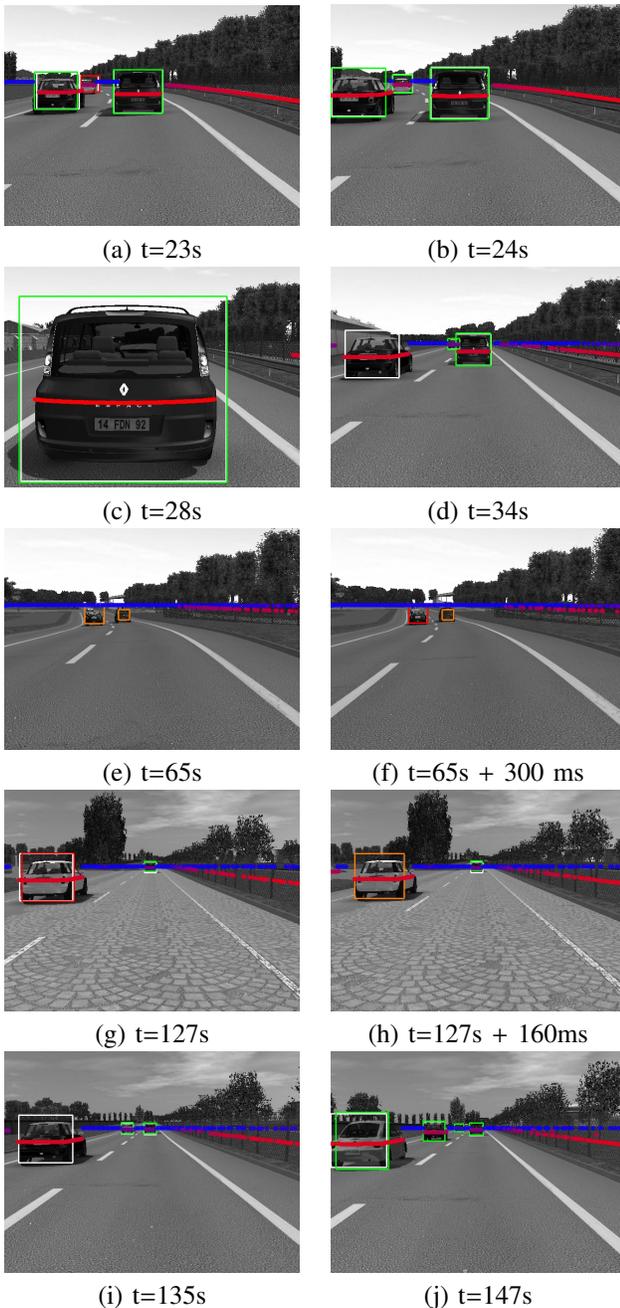


Fig. 8. White rectangles are targets of  $C$ . Tracks of  $\mathcal{P}$  are colored according to their confidence level: from red (appearance or disappearance) to green (high value).

estimate the robustness of our perception method. In the association module, an extension will be integrated to manage multi-hypothesis scenarios. Finally, several improvement with single and multiple evolution models will be tested to take into account the dynamic behavior of the ROI and to assess their impact compared to the current method.

#### Acknowledgments:

This work is part of the french-canadian *CooPerCom* project (Cooperative Perception and Communication in vehicular technologies) funded by the ANR. The authors thank

the Vista research team at Irisa/Inria Rennes for the use of the Motion2D software

#### REFERENCES

- [1] Y. Goyat, T. Chateau, and L. Trassoudaine, "Tracking of vehicle trajectory by combining a camera and a laser rangefinder," *Machine Vision and Applications*, vol. 21, pp. 275–286, 2010.
- [2] H. Zhao, C. Wang, W. Yao, F. Davoine, J. Cui, and H. Zha, "Omni-directional detection and tracking of on-road vehicles using multiple horizontal laser scanners," in *IEEE Intelligent Vehicles Symposium (IV'12)*, 2012, pp. 57–62.
- [3] C. Wang, C. Thorpe, and A. Suppe, "Ladar-based detection and tracking of moving objects from a ground vehicle at high speeds," in *IEEE Intelligent Vehicles Symposium (IV'03)*, 2003.
- [4] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 694–711, 2006.
- [5] M. Cheon, W. Lee, C. Yoon, and M. Park, "Vision-based vehicle detection system with consideration of the detecting location," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1243–1252, 2012.
- [6] S. Lefebvre and S. Ambellouis, "Vehicle detection and tracking using mean shift segmentation on semi-dense disparity maps," in *IEEE Intelligent Vehicles Symposium (IV'12)*, 2012, pp. 855–860.
- [7] 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.
- [8] A. Jazayeri, H. Cai, J. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 583–595, 2011.
- [9] C. Premebida, M. Goncalo, U. Nunes, and P. Peixoto, "A lidar and vision-based approach for pedestrian and vehicle detection and tracking," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, 2007, pp. 1044–1049.
- [10] M. Aeberhard, A. Rauch, M. Rabiaga, N. Kaempchen, and T. Bertram, "Track-to-track fusion with asynchronous sensors and out-of-sequence tracks using information matrix fusion for advanced driver assistance systems," in *IEEE Intelligent Vehicles Symposium (IV'12)*, 2012, pp. 1–6.
- [11] L. Lamard, R. Chapuis, and J.-P. Boyer, "Dealing with occlusions with multi targets tracking algorithms for the real road context," in *IEEE Intelligent Vehicles Symposium (IV'12)*, 2012, pp. 371–376.
- [12] D. Gruyer, C. Royere, R. Labayrade, and D. Aubert, "Credibilistic multi-sensor fusion for real time application. application to obstacle detection and tracking," in *IEEE International Conference on Advanced Robotics (ICAR'03)*, 2003.
- [13] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [14] J. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of visual communication and image representation*, vol. 6, no. 4, pp. 348–365, 1995.
- [15] Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
- [16] C. Royere, D. Gruyer, and V. Cherfaoui, "Data association with believe theory," in *International Conference on Information Fusion (FUSION'00)*, 2000.
- [17] M. Rombaut, "Decision in multi-obstacle matching process using theory of belief," in *Advances in Vehicle Control and Safety (AVCS'98)*, 1998.
- [18] D. Gruyer and V. Berge-Cherfaoui, "Matching and decision for vehicle tracking in road situation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'99)*, 1999.
- [19] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows, theory, algorithms, and applications*. Editions Prentice-Hall, 1993.
- [20] D. Gruyer, C. Royere, N. du Lac, G. Michel, and J.-M. Blosseville, "Sivic and rtmads, interconnected platforms for the conception and the evaluation of driving assistance systems," in *Intelligent Transportation Systems Conference, 2006. ITSC 2006. IEEE*, London, October 2006.
- [21] D. Gruyer, S. Glaser, and B. Monnier, "Sivic, a virtual platform for adas and padas prototyping, test and evaluation," in *FISITA10*, 2010.