Abstract Much work is currently devoted to increasing the reliability, completeness and precision of the data used by driving assistance systems, particularly in urban environments. Urban environments represent a particular challenge for the task of perception, since they are complex, dynamic and completely variable. This article examines a multi-modal perception approach for enhancing vehicle localization and the tracking of dynamic objects in a world-centric map. 3D ego-localization is achieved by merging stereo vision perception data and proprioceptive information from vehicle sensors. Mobile objects are detected using a multi-layer lidar that is simultaneously used to identify a zone of interest to reduce the complexity of the perception process. Object localization and tracking is then performed in a fixed frame which simplifies analysis and understanding of the scene. Finally, tracked objects are confirmed by vision using 3D dense reconstruction in focused regions of interest. Only confirmed objects can generate an alarm or an action on the vehicle. This is crucial to reduce false alarms that affect the trust that the driver places in the driving assistance system. Synchronization issues between the sensing modalities are solved using predictive filtering. Real experimental results are reported so that the performance of the multi-modal system may be evaluated.

Keywords Multi-modal perception · Visual odometry · Object tracking · Dynamic map · Intelligent vehicles

Abbreviations
ADAS Advanced driver assistance system
GPS Global positioning system
SVS Stereo vision system
FA False alarms
FOV Field-of-view
WSS Wheel speed sensors
ESP Electronic stability programme
ML Multi-layer
CAN Controller area network
ROI Region of interest
ZOI Zone of interest

1 Introduction

Advanced driver assistance systems (ADAS) can improve road safety through their obstacle detection and avoidance functions. For these functions, the location and the speed of nearby mobile objects are key pieces of information. In the literature, a number of different approaches have been applied to problems of object localization and tracking. Robotics approaches have been used to distinguish the static part of the environment [9] and to simultaneously detect moving objects [32]. Leibe et al. have presented in [18], a stereo vision strategy for obtaining a 3D dynamic map using a structure-from-motion technique and image object detectors. Using lidar information only, it is possible to estimate
the ego-motion and to detect mobile objects through a dense 3D grid-map approach [21]. In contrast, for [4] and [17], 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/km/h). The tracking performance can be improved, since the dynamics of the mobile objects are better modeled. This sort of approach simplifies the understanding of the scene and the ADAS implementation, and is also well-adapted to cooperative applications [26] (e.g. vehicle-to-vehicle communication). The present study addresses the problem of ego-localization and object tracking in urban environments.

Ego-localization can be achieved using proprioceptive and exteroceptive sensors [5]. GPS is an affordable system that provides 3D positioning. Unfortunately, GPS performance can decrease significantly in urban environments because of multi-paths and satellites outages. Dead-reckoning is a complementary solution which can be used when GPS information is unreliable. Stereo vision systems (SVS), often used for 3D reconstruction, detection and recognition tasks, are also useful for dead-reckoning (also called 3D ego-motion estimation) [6].

Object tracking for ADAS is still an active research domain. Urban environments are characterized by complex conditions: moving and static objects, mobile perception, varied infrastructures. Object representation [22,23], association methods [29], motion model and tracking strategies [19] are key points requiring particular attention.

In this work, we study a tightly coupled multi-modal system able to provide a 3D local perception of the vehicle’s surrounding environment in a world-centric frame, as depicted in Fig. 1. The modeled environment is composed of static and moving objects and a zone of interest localized in front of the vehicle. Our contribution lies in the estimation of the dynamics (location and speed) of the surrounding objects to build a dynamic map, and in ensuring the map’s integrity by using different, independent sensing principles. A dynamic map is composed of a list of the states of tracked objects together with the changing vehicle dynamics in 3D scene. A particular difficulty faced in this study is the use of asynchronous modalities which are sampled independently. To address this issue, we propose a multi-sampled strategy.

The multi-sensor system, we have designed, makes use of essential information provided by an SVS coupled with a Multi-Layer lidar (denoted ML lidar) and proprioceptive sensors (i.e. wheel speed sensors and a yaw rate gyro).

The overall system design and architecture of the proposed system are illustrated in Fig. 2. First, nearby objects are detected based on ML lidar data. The SVS and the proprioceptive vehicle sensors are used simultaneously to estimate 3D ego-localization of the vehicle. Subsequently, the detected objects are localized and tracked w.r.t. a world reference frame. Finally, tracked objects are transferred to a vehicle-centered frame in order to be confirmed by the SVS, by taking into account the different sampling times of the two sensing modalities. Confirmed tracks that are declared as verified then become the input into an ADAS for the detection of possible collisions.

In this article, four main topics are addressed: 3D ego-localization, exteroceptive sensor calibration (i.e. data alignment), object localization and tracking, and visual track confirmation. First, a detailed description of the embedded multi-sensor system set-up is given in Sect. 2. Next, the perception function is presented and experimental results are discussed. Sect. 3 is devoted to 3D ego-localization using vision and proprioceptive sensors. The extrinsic calibration procedure is described in detail in Sect. 4. Object localization and tracking are examined in Sect. 5. Visual track confirmation and synchronization issues are discussed in Sect. 6. Finally, the conclusion and perspectives of this work are presented.

2 Multi-modal perception system

2.1 Experimental set-up

An experimental vehicle belonging to the Heudiasyc Laboratory was used for the implementation and validation of the global perception system in real-life conditions. As illustrated in Fig. 3, the vehicle is equipped with a 47-cm-baseline Videre SVS. This SVS is composed of two CMOS cameras, with 4.5 mm lenses configured to acquire 320 × 240 grayscale images at 30 fps. The vision system provides essential information for ego-motion estimation and visual track confirmation. This system covers 45° with an effective range up to 50 m in front of the vehicle.

A large surrounding region is covered by an IBEO Alasca XT lidar which transmits a sparse perception of the 3D environment at 15 Hz. Installed at the front of the vehicle, this