Robotics and Autonomous Systems 62 (2014) 696–706

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Visual ground segmentation by radar supervision

Annalisa Milella [a,*], Giulio Reina [b], James Underwood [c], Bertrand Douillard [c]

[a] Institute of Intelligent Systems for Automation, National Research Council, via G. Amendola 122/D-O, 70126 Bari, Italy

[b] University of Salento, Department of Engineering for Innovation, via Arnesano, 73100 Lecce, Italy

[c] Australian Centre for Field Robotics, University of Sydney, Rose Street Building (J04), 2006 Sydney, Australia

ARTICLE INFO

Article history:
Available online 8 November 2012

Keywords:
Perception
Self-supervised visual classifier
Radar–vision combination
Ground segmentation

ABSTRACT

Imaging sensors are being increasingly used in autonomous vehicle applications for scene understanding. This paper presents a method that combines radar and monocular vision for ground modeling and scene segmentation by a mobile robot operating in outdoor environments. The proposed system features two main phases: a radar-supervised training phase and a visual classification phase. The training stage relies on radar measurements to drive the selection of ground patches in the camera images, and learn online the visual appearance of the ground. In the classification stage, the visual model of the ground can be used to perform high level tasks such as image segmentation and terrain classification, as well as to solve radar ambiguities. This method leads to the following main advantages: (a) self-supervised training of the visual classifier across the portion of the environment where radar overlaps with the camera field of view. This avoids time-consuming manual labeling and enables on-line implementation; (b) the ground model can be continuously updated during the operation of the vehicle, thus making feasible the use of the system in long range and long duration applications. This paper details the algorithms and presents experimental tests conducted in the field using an unmanned vehicle.

1. Introduction

Optical sensors, both active and passive, have proved to be effective to provide a mobile robot with the ability to understand its surroundings and successfully accomplish its tasks. Among active imaging sensors, millimeter-wave radars have been employed in autonomous vehicle systems, since they provide relatively accurate measurements of obstacles in low visibility conditions, including dust, fog, and rain [1]. Radar also provides a rich source of information allowing for multiple object detection within a single beam, whereas other range sensors are generally limited to one target return per emission. However, radar has shortcomings as well, including large footprint, specularity effects, and limited range resolution, all of which may result in poor environment survey or make it difficult to extract object features for classification and scene interpretation tasks [2,3].

Consequently, to support an expanded range of applications, radar can be used in conjunction with other sensor modalities. Video sensors lend themselves very well to this purpose. Being passive devices, cameras are affected by environmental factors, such as lighting conditions. Nevertheless, they generally supply high resolution in a suitable range of distances and provide several useful features for classification of different objects present in the scene [4]. Due to the complementary characteristics of the two sensors, it is reasonable to combine them in order to get improved performance.

In this paper, we propose a novel radar–vision combination for accurate ground modeling and scene segmentation by an autonomous vehicle operating in natural terrain. A self-supervised learning framework is developed, whereby radar guides the selection of ground patches in the camera images to build a visual model of the ground online. This model can be used to perform high level tasks, such as image segmentation and terrain characterization, or to supplement the radar sensor by solving radar ambiguities deriving from reflections and occlusions. Radar supervision allows one to eliminate time consuming manual labeling. In addition, since the ground model can be continuously updated based on the most recent radar scans, the proposed approach is suited to long range navigation applications with differing ground appearance, or long duration applications, where the appearance of the ground is likely to be affected by changing lighting conditions.

The proposed system was integrated with the CAS Outdoor Research Demonstrator (CORD), an 8 wheel skid-steering all-terrain unmanned vehicle (Fig. 1(a)), and validated in the field. The robot sensor suite is shown in Fig. 1(b) including a Prosilica Mono-CCD megapixel Gigabit Ethernet camera, pointing down (a few degrees of pitch) and a 95 GHz Frequency Modulated Continuous

* Corresponding author.
E-mail address: milella@ba.issia.cnr.it (A. Milella).

0921-8890/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2012.10.001

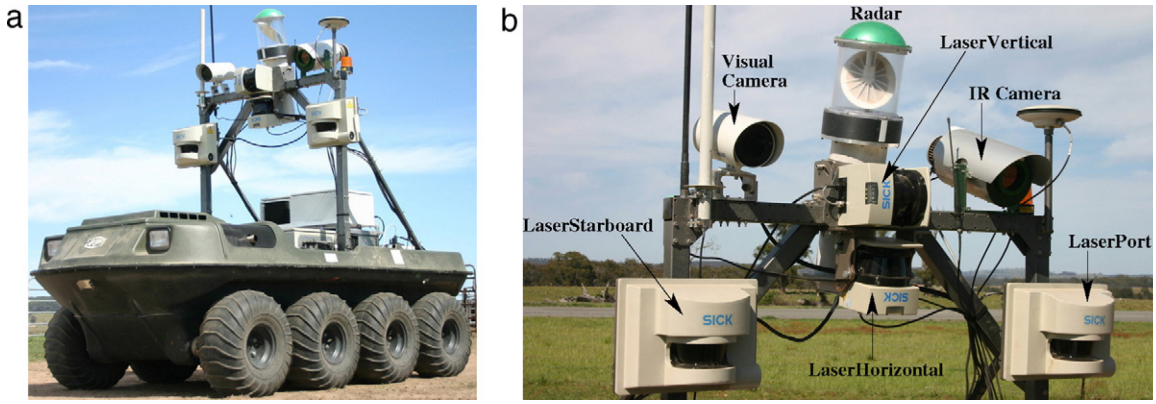

Fig. 1. The CORD UGV used in this research (a), and its sensor suite (b).

Table 1
Specifications of the custom-built radar system.

	Max. range	Raw range resolution	Horizontal FOV	Instantaneous FOV	Scan angle rate
Radar	120 m	0.25 m	360 deg	3×3 deg	3.0 Hz

Table 2
Specifications of the camera.

	Image pixel dimensions	Resolution	Frame rate
Camera	1360 × 1024	72 × 72 ppi	10 fps

Wave (FMCW) radar, custom built at the Australian Centre for Field Robotics (ACFR) for environment imaging [1]. The main technical properties of the two sensors are collected in Tables 1 and 2, for the radar and the camera, respectively. The robot was also equipped with other sensors, including four 2D SICK laser range scanners, a thermal infrared camera, and a RTK DGPS/INS unit providing accurate position and tilt estimation of the vehicle during the experiments. The remainder of the paper proceeds as follows. Related works are presented in Section 2. Section 3 details the radar-supervised visual classifier that is experimentally validated in Section 4. Conclusions are drawn in Section 5.

2. Related work

Radar and vision fusion has been discussed in the context of driver assistance systems, mainly featuring object detection and classification modules [2,3,5,6]. For instance, in [2] radar and vision independently detect targets of interest, then a high level fusion approach is adopted to validate radar targets based on visual data. A radar–vision fusion method for object classification into the category of vehicle or non-vehicle is developed in [3]. It uses radar data to select visual attention windows, which are then assigned a label and processed to extract features to train a Multi-layer In-place Learning Network (MILN). In [5], a vehicle detection system fusing radar and vision data is proposed. First, radar data are used to locate areas of interest on images. Then, a search is performed in these areas mainly based on vertical symmetry. A guard rail detection approach and a method to manage overlapping areas are also developed to speed up and improve the performance of the system.

In this work, we propose a novel radar–vision combination that aims at building a model of the ground online, to perform visual scene segmentation into ground and non-ground regions.

Persistent ground segmentation is critical for a robot to improve perception under all conditions, with many important applications, including environment classification and scene interpretation. While in structured environments, such as in urban contexts, the task of ground identification can be effectively performed by exploiting some distinctive roadway markings, in natural terrain, no *a priori* information about the ground surface is usually available. Furthermore, ground structure and appearance may significantly change during the vehicle operation; therefore, road detection algorithms based on specific cues are not suitable, unless retuning of road markers or re-training of classifiers is performed, generally with human supervision.

To overcome the limitations of these methods, self-supervised terrain classification approaches have been developed, whereby training examples are automatically labeled by another classification algorithm instead of being produced manually. For instance, in [7], data from a stereo camera are used to train a monocular image classifier that segments the scene into obstacles and ground patches, in the submodular Markov Random Field (MRF) framework. A self-learning framework is also proposed in [8] to automatically train a ground classifier for autonomous agricultural vehicles based on multi-baseline stereovision. In [9], self-supervised terrain classification is performed using a previously trained vibration-based classifier, which provides labels to train online a visual classifier. Another notable example of self-supervised ground segmentation can be found in [10], using a laser scanner and a monocular camera. Specifically, the laser is employed to scan for a flat, drivable surface area in the vicinity of the vehicle. Then, this area is projected in the camera image and is used as training data for a computer vision algorithm to learn online a visual model of the road. In [11], the authors propose a self-learning framework to train a classifier for radar image interpretation.

In this work, we propose a self-supervised ground segmentation method, using a radar–vision system. The main contribution of the proposed approach relies on the combination of a radar-based segmentation method with a vision-based classification system to incrementally construct online a visual model of the ground and perform scene segmentation. Specifically, the visual classifier is continuously retrained using radar ground labeled samples. Visual segmentation of the entire scene is then performed using a one-class classification system with a multivariate Gaussian model of the terrain and a Mahalanobis distance-based outlier detection approach.

3. Radar-supervised visual ground segmentation

In this work, a self-supervised learning framework for visual ground segmentation is discussed. "Self-supervised learning" refers to automatic training of a classification system by another classifier. Specifically, while in a conventional supervised classifier a human user provides manually labeled training examples for a given class of interest, a self-supervised framework uses the output