Robotica (2016) volume 34, pp. 2823–2841. © Cambridge University Press 2015
doi:10.1017/S0263574715000442

LIDAR and stereo combination for traversability assessment of off-road robotic vehicles

Giulio Reina†*, Annalisa Milella‡ and Rainer Worst§

†*Department of Engineering for Innovation, University of Salento, Via Arnesano, 73100 Lecce, Italy*
‡*Institute of Intelligent Systems for Automation, National Research Council, via G. Amendola 122 D/O, 70126, Bari, Italy. E-mail: milella@ba.issia.cnr.it*
§*Fraunhofer IAIS, Schloss Birlinghoven, 53757 Sankt Augustin, Germany. E-mail: rainer.worst@iais.fraunhofer.de*

(Accepted May 14, 2015. First published online: June 15, 2015)

SUMMARY
Reliable assessment of terrain traversability using multi-sensory input is a key issue for driving automation, particularly when the domain is unstructured or semi-structured, as in natural environments. In this paper, LIDAR-stereo combination is proposed to detect traversable ground in outdoor applications. The system integrates two self-learning classifiers, one based on LIDAR data and one based on stereo data, to detect the broad class of drivable ground. Each single-sensor classifier features two main stages: an adaptive training stage and a classification stage. During the training stage, the classifier automatically learns to associate geometric appearance of 3D data with class labels. Then, it makes predictions based on past observations. The output obtained from the single-sensor classifiers are statistically combined in order to exploit their individual strengths and reach an overall better performance than could be achieved by using each of them separately. Experimental results, obtained with a test bed platform operating in rural environments, are presented to validate and assess the performance of this approach, showing its effectiveness and potential applicability to autonomous navigation in outdoor contexts.

KEYWORDS: Robotic vehicles; Navigation systems; Sensor combination; Online learning strategy; Unmanned ground vehicles.

1. Introduction

To increase the level of driving automation, perception capability towards environment awareness is a fundamental requirement that presents many challenges.[20] Although autonomous navigation has inspired decades of research with many ground vehicles successfully demonstrated in a variety of application domains, such as on-road scene awareness and urban environments,[24,39] off-road terrain analysis for challenging vegetated areas,[36] deserts,[8] and planetary exploration,[3,27] it still remains an open and active field of investigation. One of the most critical issues is the ability to discriminate drivable terrain from obstacles, including man-made artifacts, ruts, cliffs, trees, bushes, shrubs, and other vegetation that can obstruct or endanger the robot's motion.[12] The last few years have seen an increasing interest towards solving this problem using LIght Detection And Ranging (LIDAR) sensors and stereo cameras. Both sensor modalities measure the range to objects within their field of view, providing as output a 3D point cloud. LIDAR and vision systems are complementary in many ways: for example, when performing a three-dimensional reconstruction of the world, LIDAR data offer a very accurate, yet sparse representation of the world, whereas vision provides dense, but less accurate measurements. LIDAR systems also perform better in low-lighting conditions than stereovision. However, scanning LIDARs typically provide low sampling rates (0.1–1 Hz) resulting in difficulties to capture dynamic obstacles. In addition, LIDAR may feature limited sensing range according to the particular sensor location onboard the vehicle, i.e. low height above the ground

* Corresponding author. E-mail: giulio.reina@unisalento.it

results in reduced look-ahead distance. This is the case for the vehicle used in this research, where the LIDAR features less range than the stereo system. Expensive LIDARs like the Velodyne HDL-64E are an exception (www.velodyne.com/lidar). On the other hand, images from a stereo camera can be processed at a fast frame rate allowing static and dynamic obstacles to be detected. This can be particularly useful for reactive planning responding to dynamic obstacles.

This paper develops a novel approach to ground surface detection in natural environments, which combines LIDAR with stereovision within a statistical framework, as preliminary introduced in ref. [29]. An adaptive self-learning scheme is proposed, whose basic principle is of general validity and it can be applied to any 3D sensor combination. In the context of this research, it is demonstrated for LIDAR and stereo data, although they differ in resolution, accuracy, and field of view. In detail, the proposed approach relies on an online learning strategy to build a statistical model of the ground using geometric features extracted from 3D data, and it implements a Mahalanobis distance (MhD) classification rule for traversable ground identification. Geometric features are computed as statistics obtained from the point coordinates associated with each patch of a grid-based representation of the scene. Since the characteristics of the ground may change geographically and over time, the ground model is continuously retrained: newest labeled data are added to the ground model replacing the oldest labels in order to incorporate changes in the ground geometric appearance. LIDAR and stereo data are first processed separately, thus featuring two independent self-taught ground classifiers. Afterwards, the output obtained from the two single-sensor classifiers are statistically combined to improve the overall perception capability.

This approach leads to the following main advantages: (a) improvement of the perception performance of the combined LIDAR/stereo system due to the complementarity of the sensor modalities, (b) self-learning training of the system, where the sensors allow a set of ground samples to be automatically acquired, eliminating the need for time-consuming manual labeling, (c) continuous updating of the system during the vehicle's operation, thus making it adaptive and feasible for long-range and long-duration navigation.

Experimental results, obtained with an off-road vehicle operating in rural environments, are presented to validate the proposed system. It should be noted that the work focuses on the geometric aspects of obstacle detection. Therefore, the problem of image appearance or LIDAR reflectance-based classification of obstacles and terrain types is not addressed.

The remainder of the paper is organized as follows. Section 2 surveys related research in the field. An overview of the proposed system and its architecture are presented in Sections 3 and 4, respectively. Section 5 describes 3D reconstruction using LIDAR and stereovision, whereas ground classification and classifier fusion are explained in Sections 6 and 7, respectively. In Section 8, the system is validated in field tests. Finally, Section 9 concludes the paper.

2. Related Work

In order to address the terrain estimation issue, three different strategies have been adopted in the past: vision-based, LIDAR-based, and vision/LIDAR-based. Many researchers rely on stereovision to generate a 3D point cloud at relatively high frequency.[6,15,17] By applying geometric and statistical heuristics, the terrain surface and obstacles can be modeled and classified. For example, in ref. [7] an approach for full 3D obstacle detection was proposed using a voxel-based representation of the environment, whereas 3D modeling based on octrees and probabilistic occupancy estimation was discussed in ref. [37]. However, stereo-generated maps may be affected by lighting conditions or poor reconstruction (due for instance to textureless areas or to the presence of repetitive structures). There are also many researchers using LIDAR sensors to detect terrain surface for robot navigation.[16,35,36] In general, LIDAR sensors can return accurate 3D point clouds, even though, scanning LIDAR sensors often operate at a relatively low frequency (1 Hz or less). Fixed (non-scanning) LIDAR sensors can operate at a high frequency; however, they require an additional algorithm to accumulate data as the robot moves. This algorithm relies on accurate robot pose estimation, which is difficult to achieve on rough terrains, and the presence of a tree canopy makes GPS signal reception problematic.

Due to the mentioned limitations, methods have been proposed to combine vision and LIDAR into one system and mitigate the drawbacks of each single approach. Two main methodologies have been proposed in the literature: *a priori* and *a posteriori* integration. The first approach combines ranges obtained from the two sensors at raw data or feature level. For instance, in ref. [2], LIDAR data are

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

integrated directly into the stereo algorithm to improve disparity computation. Ref. [23] propose an architecture to fuse stereo and LIDAR sensors with several modular algorithms to generate a single spatial representation of obstacles and free space around a mobile robot. As a further example, in ref. [1], the authors present a fusion method at object detection level, specifically developed for urban intersection safety.

In contrast, *a posteriori* methods aim to combine the output coming from high-level classifiers using single sensor data. As research on low-level fusion becomes well established and approaches maturity, research focus is currently shifting towards integration at higher levels.[9,10] New interest in this field has been generated by the application of probabilistic self-learning techniques.[13] The idea is that of using one sensor to automatically provide training examples to another sensor classifier. For example, a self-supervised approach is proposed by ref. [8], where monocular vision is supervised by a laser range finder. Specifically, the laser is used to scan for a flat drivable surface area in the vicinity of the vehicle. Once identified, this area is used as training data for the computer vision algorithm. A similar scheme was applied to the specific case of a forested environment in ref. [40]. Other sensors, including a radar,[19] and a stereo camera,[28] have also been used as the supervising sensor to automatically train a visual classifier.

In this work, a multi-sensor fusion approach for ground surface detection is proposed, where the results coming from single-sensor classifiers are statistically combined to produce a unique classification result, following *a posteriori* integration. This mitigates the drawbacks that would result in using each sensor modality separately. Indeed, stereovision and LIDAR produce compatible data (i.e. range measurements), which would make it feasible a data or feature level fusion. However, a high level fusion scheme is adopted in order to keep the system independent from data type and potentially applicable to any sensor combination, which is the case for autonomous vehicles requiring multi-sensory perception. One should also note that most of the algorithms proposed in the literature for terrain estimation work under the assumption of a "flat" world,[11,14,26] and thus the obstacle detection task amounts to identify objects that "stick out" of the ground. However, this assumption is of limited validity in outdoor unstructured environments. Some works have adopted more complex primitives to deal with non-planar ground surfaces,[22,34] at the cost of an increase in computational complexity.

In this research, for each single sensor classifier local geometric features are extracted using a grid-based representation of the environment and ground estimation is performed within a self-learning scheme; therefore, ground plane reasoning is not explicitly needed nor fitting of geometric primitives, and the system automatically adapts to the changing geometry of the terrain. This leads to an integrated LIDAR-stereo classifier for adaptive ground estimation.

Furthermore, the proposed approach aims to detect scene regions that are traversable-safe for the vehicle rather than attempting to explicitly identify obstacles.[17,32] This is a subtle, but significant difference; only those regions where there is evidence that it is safe are labeled as traversable, thereby avoiding both positive and negative obstacles without explicitly detecting them. An additional advantage of the proposed obstacle detection scheme is that the output traversability map can be directly employed by most grid-based planners.[31]

3. Overview of Proposed System

This paper presents a self-learning classification approach to identify drivable terrain surface, based on data provided by a single or multiple 3D sensors. The system is intended to be used onboard an autonomous vehicle operating in natural contexts.

The term "self-learning" refers to automatic training of the system. While in a traditional (i.e. manually) supervised classifier, a user provides labeled examples to train each class of interest, in a self-learning scheme these training instances are automatically produced. Self-learning systems eliminate the need for hand-labeled training data, thus gaining flexibility in unknown environments. Not only the burden of hand-labeling data is relieved, but the system can robustly adapt to changing environments on-the-fly.

In the context of this research, the use of a rolling training set is proposed. Initially, the robot has no knowledge of the ground class appearance. The training set is initialized at the beginning of the robot's operation and progressively updated. The underlying assumption is that the vehicle starts its operation from a clear (free of obstacles) area, so that each sensor initially "looks" at ground

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

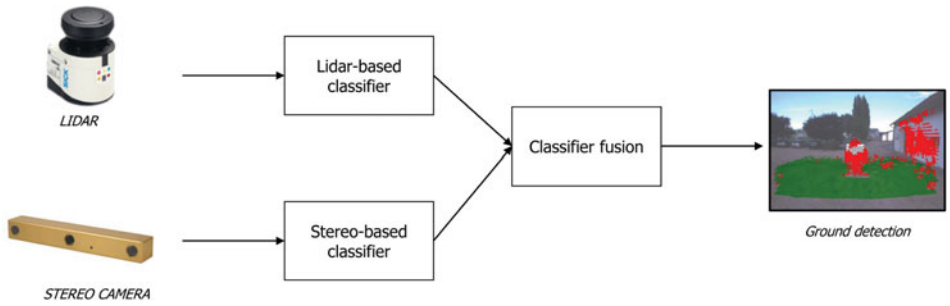

Fig. 1. Architecture of the proposed statistical framework for multisensory perception. Input data from LIDAR and stereo camera are used by different classifiers, whose results are fused to produce a unique classification output.

only. However, it should be said that this hypothesis may be not always feasible (i.e. bootstrapping in a forested environment). In these particular cases, alternative solutions may be thought of and implemented, including the use of a previously recorded database or looking at terrain patches that are directly in front of the robot. During the bootstrap stage, geometric features can be extracted from the sensor-generated 3D point cloud, and they can be reasonably associated with the ground class. When sufficient data is accumulated, the ground classifier can be trained, and the ground labels are related with 3D point cloud properties. The task is that of generalizing from training data to unseen situations to identify single new observations as ground or non-ground. This allows the system to predict the presence of ground in successive scenes based on past observations. Such a classification task is generally difficult as the geometric ground appearance is affected by a number of factors that are not easily measured and change over time, including the type of terrain surface, topology, etc. Therefore, an adaptive approach is necessary in which the image interpretation changes as the vehicle moves and conditions vary. To this aim, the model (i.e. the training set) is continuously updated using the most recent sensor readings. In every newly acquired scan, the latest training set is used to train the classifier.

It is important to note that the proposed classification scheme can be used to characterize the scene structure obtained from any 3D sensor. In this research, it is applied to develop two classifiers: one based on LIDAR data and one based on stereo data. Both classifiers adopt a one-class classification scheme, using a multivariate Gaussian model of the ground and a MhD-based classification rule. The single-sensor classifiers are then statistically combined to obtain a unique classification result, according to the scheme illustrated in Fig. 1.

4. System Architecture

4.1. Vehicle configuration

The proposed system was integrated with the off-road vehicle shown in Fig. 2. The test bed was equipped with a Point Grey Bumblebee XB3 camera and a scanning LIDAR. The vision unit features two stereo configurations: a narrow stereo pair with a baseline of 0.12 m using the right and middle cameras, and a wide stereo pair with a baseline of 0.24 m using the left and right cameras. Additional technical details of the stereo system are collected in Table I. The use of a trinocular configuration in place of a binocular one allows to combine the advantages of two different baselines.[5] A narrow baseline increases the shared field of view of the two cameras, while yielding to shorter maximum range. Conversely, a larger baseline decreases the common field of view, but leads to higher maximum range and accuracy at each visible distance. By employing the narrow baseline to reconstruct nearby points and the wide baseline for more distant points, the trinocular system takes the advantage of the small minimum range of the narrow baseline, while preserving, at the same time, the higher accuracy and maximum range of the wide baseline configuration. The interested readers are referred to ref. [18] for more details.

The LIDAR system consists of a SICK LMS111 that generates single line scans covering an angle of 270 deg with an angular resolution of 0.5 deg at a rate of 50 Hz. The LIDAR is mounted on a

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

Table. I. Specifications of the stereo vision system.

Camera	Model (baseline)	Image size (pixels)	Field of view	Optics	Range
Trinocular	Bumblebee XB3 (0.12/0.24 m)	1280 × 960	66 deg × 54 deg	focal length: 3.8 mm f2.0	2 to 30 m

Table II. Specifications of the LIDAR scanner.

LIDAR	Model	Number of points	Field of view	Range
Rotating rangefinder	3DSL with SICK LMS 11	∼ 80, 000	360 deg × 270 deg	0.5 to 17 m

Fig. 2. The test platform employed in this research. It was made available by IRSTEA within the research activity connected with the Ambient Awareness for Autonomous Agricultural Vehicles (QUAD-AV) project, funded by the 2011 ICT-AGRI FP7 European Research Program.

servo-controlled rotating stage that sweeps the scan plane through 360 deg around the longitudinal axis of the robot, generating a complete 3D point cloud with a size of 80,000 points in about 3 s. Multiple scans obtained during the robot's motion are aligned and integrated using ICP-based algorithms,[21] to generate a dense 3D point cloud in a fixed reference frame. The salient technical details of the LIDAR system are collected in Table II.

4.2. Sensor calibration

In order to integrate LIDAR and stereovision data, time-synchronization and calibration are required. As the test platform used in this work was equipped with different computers, the sensor data were coming from various sources with their own clock. For example, laser data were gathered as ROS (Robot Operating System) bags by a computer running Ubuntu Linux, whereas stereo data were processed using a C++ code running on a Windows 7 laptop. In order to handle time differences, sensory data were accurately time-stamped, and NTP (Network Time Protocol) was used to mitigate time synchronization errors between the two computers (i.e. less than 5 ms).

Calibration was performed using an offline procedure to estimate the position and orientation of each sensor with respect to the vehicle. Specifically, three reference frames were defined as shown in Fig. 2: a vehicle reference frame (VRF), a LIDAR reference frame (LRF), and a camera reference frame (CRF). The VRF is the frame linked to the body of the vehicle, its center being located approximately at the center of the vehicle. The axes are: X_v pointing forward from the vehicle, Y_v

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

pointing to the right side of the vehicle, and Z_v pointing up. The LRF is defined in a similar way as the vehicle frame (i.e. X_l forward, Y_l right side, Z_l up), but centered on the LIDAR sensor. The CRF is centered on the right camera of the trinocular sensor, with the Z_c axis aligned with the optical axis, the Y_c-axis pointing down, and the X_C-axis pointing right. Two categories of calibration have been made.

- *Camera-LIDAR calibration*: the XB3 sensor is pre-calibrated, i.e. intrinsic and extrinsic parameters are available in a calibration file provided by the manufacturer, therefore, only the transformation between the reference (right) camera of the trinocular sensor and the LIDAR sensor had to be estimated. Specifically, a method adapted from ref.[38] was employed. It is based on acquiring multiple views of a planar checkerboard pattern from the camera and the LIDAR simultaneously, and solving for constraints between the "views". The positions and orientations of the planes corresponding to the checker board visible in the images were estimated using the Matlab Camera Calibration Toolbox. Then, these positions were compared with the positions of the laser points hitting the board to estimate the pose of the CRF with respect to the LRF, based on an optimization process.
- *LIDAR-Vehicle calibration*: the estimation of the transformation between the frame associated to the LIDAR sensor and the vehicle frame was performed partially by construction, i.e. by assembling the sensor without any rotational offset and exactly in the middle between the front wheels, and partially by measuring the distances $X_l - X_v$ and $Z_l - Z_v$ with a folding ruler. The results of this calibration are the estimation of the three rotation angles and three translation offsets from the vehicle frame to the LIDAR frame.

5. 3D Scene Reconstruction

This section describes the algorithms adopted for 3D scene reconstruction using stereo and LIDAR data.

5.1. Stereo-based reconstruction

Three-dimensional scene reconstruction using stereo data is performed by processing separately the wide and narrow baselines of the trinocular sensor, and then fusing the two correspondent point clouds in a unique set. This approach ensures accurate information both in the near range (up to 15 m) and in the far range (up to 30 m). In detail, each stereo pair is processed according to the following steps:

- *Rectification*: transformation of each image plane so that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes. The use of rectified images allows the problem of stereo correspondence computation to be reduced from a 2D to a 1D search problem, which is typically performed along the horizontal raster lines of the rectified images. Rectification matrices are known from the calibration step, as described in Section 4.2.
- *Disparity map computation*: an algorithm based on a Sum of Absolute Differences (SAD) correlation approach is used to find corresponding points.[4] In order to improve the accuracy of the disparity map, a filtering scheme is applied to remove incorrect matches due to occlusions or lack of texture.[25]
- *3D point cloud generation in the reference camera frame*: being the stereo pair calibrated both intrinsically and extrinsically, disparity values can be converted in depth values, and 3D coordinates can be computed in the reference camera frame for all matched points.
- *Statistical filtering*: for each point, the mean distance from it to all its neighbors is computed. By assuming that the resulting distribution is Gaussian, all points whose mean distances are outside an interval given by the global distances mean and standard deviation can be removed as outliers from the data set.
- *Voxelization*: A 3D voxel grid with a leaf size of 10 cm is applied and all the points in each voxel are replaced with their centroid. This allows the computational burden of the algorithm to be reduced.
- *Transformation in the VRF*: 3D points are transformed from the camera to the vehicle frame.

Afterwards, the point clouds reconstructed separately by the narrow and wide stereo pair of the trinocular system are combined in a unique point cloud. It should be noted that if a point has been reconstructed by both the wide and narrow baseline, only information generated from the wide

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

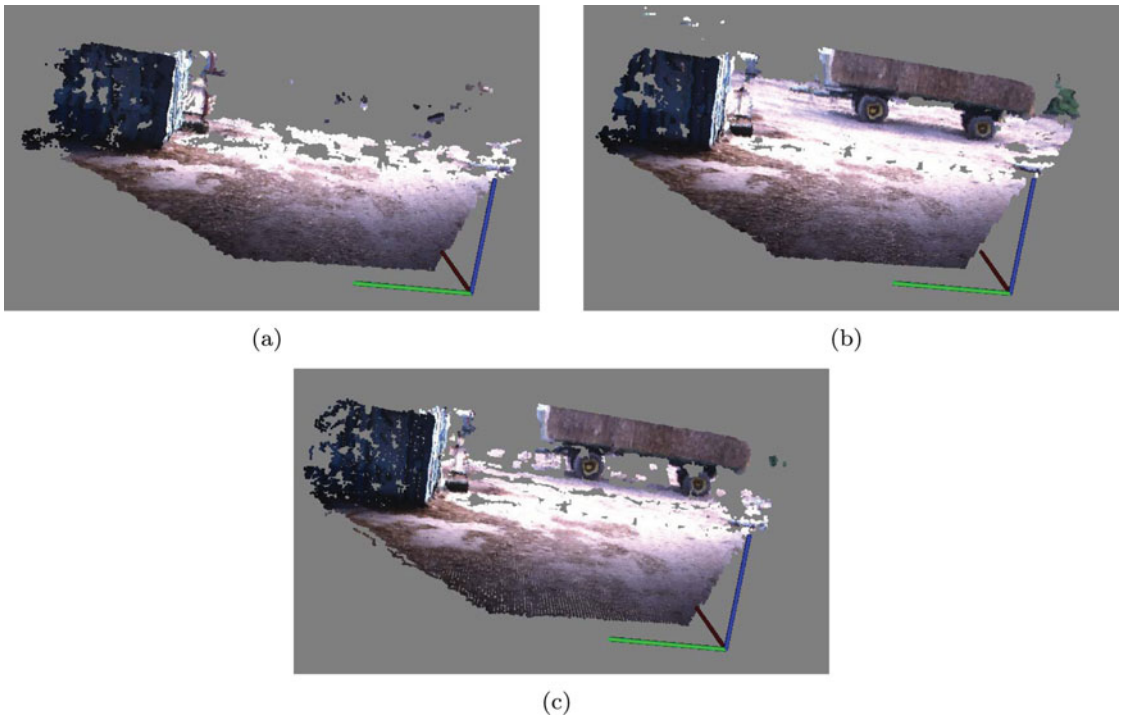

Fig. 3. (a) RGB point cloud obtained from the narrow baseline; (b) RGB point cloud obtained from the wide baseline; (c) Multiplexing of wide and narrow stereo pairs.

pair is retained, as it is known that wider baselines generally guarantee better accuracy at every distance. Figure 3 shows the results obtained from the stereo system in a sample scene. It is worth mentioning that fusion of the two stereo baselines was performed at data level (*a priori* integration), as data generated by the two subsystems were considered as two outputs of a unique sensor (i.e. the trinocular camera).

Finally, the accuracy of the 3D reconstruction can be quantified by comparing stereo-generated points with a set of points known as the ground truth. The resulting difference defines the 3D measurement error along the three axes, which resulted in an average value of 0.092 m with a standard deviation of 0.062 m.

5.2. LIDAR-based reconstruction

The scanning plane of SICK LMS111 is rotated around the X_l-axis, which points forward, thus delivering a 360 deg scan of the scene in front of the vehicle after each half round. The reconstruction of the point clouds contains the following stages:

- *2D scan*: there is one node running that reads the plain data from the laser scanner.
- *Rotary encoder*: simultaneously, there is one node running to sense the current position of the rotational unit, on which the laser scanner is mounted.
- *3D scan*: a third node merges the data from these two nodes and creates records containing the measured distances and the respective orientation of the scanning plane for one complete 2D scan.
- *Generation of point cloud*: the records of the previous stage are collected by another node and after a rotation of 180 deg, all these records are used to generate a 3D point cloud.

The further processing of the laser-generated point cloud follows the same rationale of the stereo-based reconstruction.

6. Ground Modeling and Classification

The objective is to classify a given terrain patch as being traversable or not. This problem can be formulated as a one-class classification.[33] In general, one-class classification methods are useful in

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

two-class classification problems, where one of the two classes (the outlier class), is relatively under-sampled or it is difficult to model. This is the case for most of outdoor applications, as the variation of all possible non-ground classes is "virtually" unlimited. In contrast, the ground class (or target class) is generally less variable, although it changes geographically and over time. To model the ground class, a multivariate Gaussian distribution is assumed. Then, a MhD-based classification approach is adopted to assess whether a new pattern is an instance of the ground class. In more detail, an outlier detection strategy is implemented by looking at the MhD and its distribution to predict if this reading has an extremely low probability of belonging to ground and may be suspected to be an outlier.

6.1. Geometric features

The appearance of ground is constructed upon a set of geometric features that can be extracted from the 3D scene reconstruction performed using either stereo data or LIDAR data, as previously described in Sections 5.1 and 5.2, respectively. The raw point cloud is first divided into a grid of 0.4×0.4 m terrain patches projected onto a horizontal plane. Geometric features are, then, statistics obtained from the point coordinates associated with each terrain patch. This geometric information is used as input to a self-learning classification scheme that labels all cells as traversable and non-traversable. As a result, positive and negative obstacles, as well as, unknown regions can be implicitly detected and avoided. The first element of the geometric feature vector is the average slope of the terrain patch, i.e. the angle θ between the least-squares-fit plane and the horizontal plane. The second component is the goodness of fit, E, measured as the mean-squared deviation of the points from the least-squares plane along its normal. This is the same as the minimum singular value of the points' covariance matrix. The third element is the variance in the height of the range data points with respect to the reference plane, σ_h^2. The fourth component is the mean of the height of the range data points, $\overline{h}$. Thus, the geometric properties of each patch is represented by a four-element vector $x = [\theta, E, \sigma_h^2, \overline{h}]$.

6.2. Ground class model

Let X_t be an $n \times m$ data table that represents at time t a set of x_i feature vectors with $i = 1, 2, \ldots n$, each characterized by m traits ($m = 4$, in our case): $X_t = \{x_1, \ldots, x_n\}$. This set constitutes the training set at this time. Then, the ground model can be defined at time t as $M_t(\mu_t, \Sigma_t)$, where μ_t is the mean and Σ_t the covariance of the data in X_t. A given new observation x_{new}, acquired in the next sensor scan can be classified by estimating its squared MhD d^2 from the ground model

$$d^2 = (x_{new} - \mu_t)\Sigma_t^{-1}(x_{new} - \mu_t)^T \tag{1}$$

It can be proved that d^2 is asymptotically distributed as the m degrees of freedom (DOF) chi-square distribution χ_m^2, under the assumption that vectors x_i are independent and have Gaussian distribution, as previously discussed by the authors in ref. [18]. Then, the quantile β of the m DOF chi-square distribution can be used as the delimiter (cutoff) for outlying observations.[30] Let β denote a constant probability level: $0 < \beta < 1$, and $\chi_{m;\beta}^2$ the appropriate quantile of the distribution, it holds

$$p(d^2 \geq \chi_{m;\beta}^2) = 1 - \beta \tag{2}$$

which means that values of d^2 greater than (or equal to) the value $\chi_{m;\beta}^2$ appear with a probability equal to $1 - \beta$. Now, the cutoff for the MhD can be defined as

$$L_\beta = \sqrt{\chi_{m;\beta}^2} \tag{3}$$

In summary, a given observation with MhD d satisfying the inequality $d \geq L_\beta$ may be suspected to be an outlier and, therefore classified as a non-ground observation.

6.3. Model update

One important aspect of the proposed classification scheme is that the model is recursively updated at each new sensor scan. At the start, the training set is initialized under the assumption that the vehicle is in an area free of obstacles, so that the sensor "looks" at ground only. Afterwards, the ground model is updated during the vehicle's operation by incorporating new observations labeled by

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

the system as ground and discarding an equal number of the oldest ground instances. Therefore, the ground model is assumed as a rolling window, whose size is kept constant, that always reflects the latest appearance of the ground encountered by the vehicle. By denoting with $Z_{t+1} = \{z_1, z_2, \ldots, z_l\}$ the set of l ground-labeled returns classified at time $t + 1$, then the training set used to classify the next sensor scan is obtained as

$$X_{t+1} = \{(x_{l+1}, \ldots, x_n), Z_{t+1}\} \tag{4}$$

7. Classifier Fusion

The single-sensor ground classifiers can be combined statistically. Thus, one can exploit their individual advantages in order to reach an overall better performance than could be achieved by using each of them separately. Combining classifiers aims to exploit the complementary information residing in the single classifiers. Assume that we are given a set of classifiers, which have already been trained to provide as output the class *a posteriori* likelihood in the form of the MhD from the class center. For a one-class classification task, given an unknown observation x, each classifier produces estimates of the *a posteriori* class likelihood, that is $M_i(x)$, $i = L, S$, where L stands for LIDAR-based classifier and S for stereo-based classifier. The goal is to devise a way to come up with an improved estimate of a final *a posteriori* likelihood $\widehat{M}(x)$ based on all the resulting estimates from the individual classifiers. One way is to weight the individual output obtained from the classifiers with their prior probabilities that can be statistically quantified using ground-truth data. This analysis would provide various statistical quantities including for instance a confusion matrix. By appropriately normalizing true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), obtained from the confusion matrix, we can construct (empirical) expected rates of TP detections or precision (P) and false alarms or rejection precision (RP),

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$RP = \frac{TN}{TN + FN} \tag{6}$$

Being normalized, these rates are also probabilities. So now, we have values of uncertainty associated with the LIDAR and stereo only ground detection in the form of prior probabilities over these detections. These (prior) probabilities can be used as weights to combine statistically the decision of each classifier through a weighted sum and obtain a unique classification result,

$$\widehat{M}(x) = \sum_{i=L}^{S} \frac{W_i M_i(x)}{W_L + W_S} \tag{7}$$

where the weight, W_i, is equal to P_i or RP_i if the observation x is labeled as ground or non-ground, respectively, by the classifier i. As a result, the uncertainty associated with each classifier is propagated throughout the final classification result.

8. Results

In this section, experimental results are presented to validate the proposed approach for terrain estimation using LIDAR-stereo combination. The system was integrated with the experimental platform of Fig. 2 and tested in rural settings. Various scenarios were analyzed including different types of terrains (asphalt, country trails, and agricultural terrain) and obstacles (trees, crops, metallic poles, buildings, agricultural equipment, humans and animals). During the experiments, the vehicle was driven by a user with a travel speed ranging between 10 and 20 km/h, as the onboard sensors acquired data from the surrounding environment. Then, the proposed classification framework was applied offline. Several experimental data sets were collected over the course of three days. Each

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

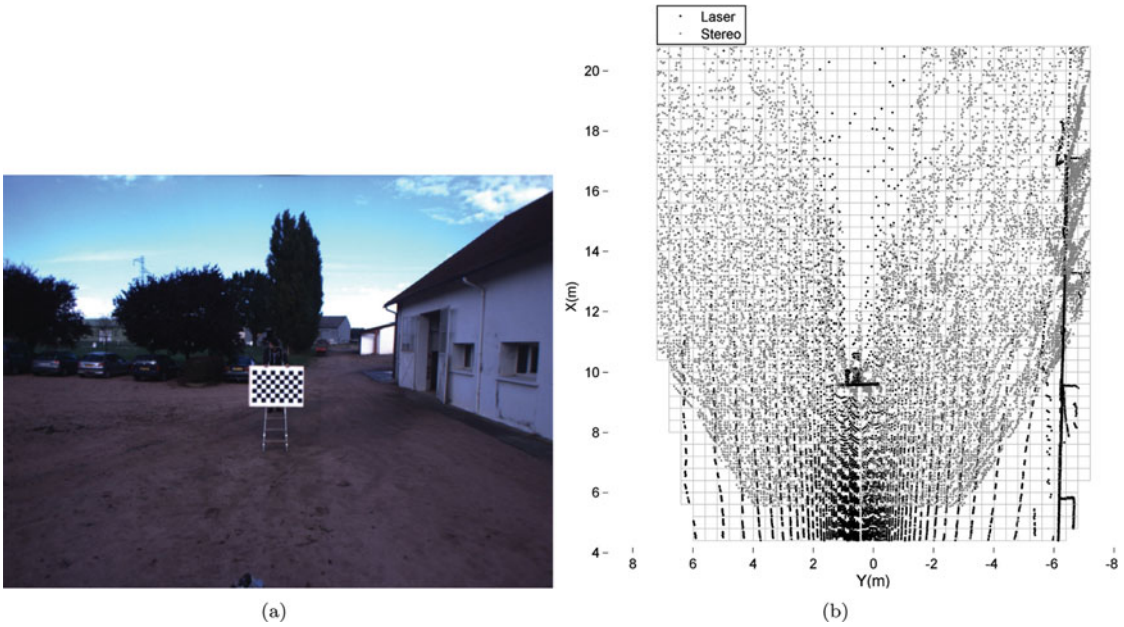

Fig. 4. (a) Original visual image. (b) Reference grid divided into 0.4 m × 0.4 m cells. Points reconstructed by laser scanner are denoted in black. Points reconstructed by stereovision are marked in grey.

data set consisted of a time series of LIDAR and stereo data recorded during traverses of at least 250 m (up to 3 km). For each data set, the vehicle started its operations from an area that was clear of obstacles in order to initialize the ground model for both LIDAR-based and stereo-based classifier. Few scans (s_0 = 3, in our case) were necessary to complete the bootstrap phase, requiring a short time interval (e.g., a 10 s window if a sampling rate of 0.3 Hz was used for the LIDAR scanner). After the starting training, both single-sensor classifiers were able to predict the presence of ground in successive acquisitions.

Note that, in all the experiments, the tilt experienced by the vehicle was relatively limited with roll and pitch angle up to 5 deg. For higher values of tilt, an onboard navigation system would also be necessary in order to estimate and compensate for the vehicle tilt (i.e. to correctly transform the sensor readings from the vehicle to the world reference frame).

One important aspect of the proposed methodology is the definition of the prior classification probabilities of the LIDAR and stereo-based classifiers. In this implementation, they were estimated off-line using ground-truth data for one data set. Automatic methods to update online prior probabilities would be desirable. However, this is not addressed in this work and it will be part of future research. Based on the knowledge of the prior classification performance, it was possible to fuse the two single-sensor classifiers using (7) for terrain estimation in different data sets. A significance level of 0.1% ($\beta = 0.999$) was assumed for the cutoff threshold expressed by (3), in both the LIDAR-based and stereo-based classifiers.

In the remainder of this section, first the results obtained for single test cases are presented. Then, a more comprehensive assessment of the system performance is evaluated over a subset of selected images and using a sequence of images.

8.1. Test cases in the field

In Fig. 4(a), a typical scene is shown acquired during the field calibration and testing of the system. Figure 4(b) shows the upper view of the 3D reconstruction, as obtained independently by laser scanner (marked by black points) and stereovision (denoted by grey points). By combining LIDAR with stereovision the overall field of view is increased with LIDAR and vision providing data mostly in the short and long range, respectively. In addition, the sparseness of LIDAR data is mitigated by dense stereo reconstruction. In Fig. 5, the results obtained from the LIDAR-based classifier are shown for the scene of Fig. 4(a) in terms of traversability map. Cells labeled as ground are marked in green, whereas cells that are labeled as non-ground are denoted in red. In Fig. 5(b), the same results

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

Table III. Classification results obtained from the single-sensor and combined classifiers over the common cell subset, i.e. the cells that are "seen" by both sensors in Fig. 4(a).

	LIDAR-based	Stereo-based	Combined
Precision	99.1%	96.5%	99.6%
Rejection Precision	87.5%	96.1%	98.3%

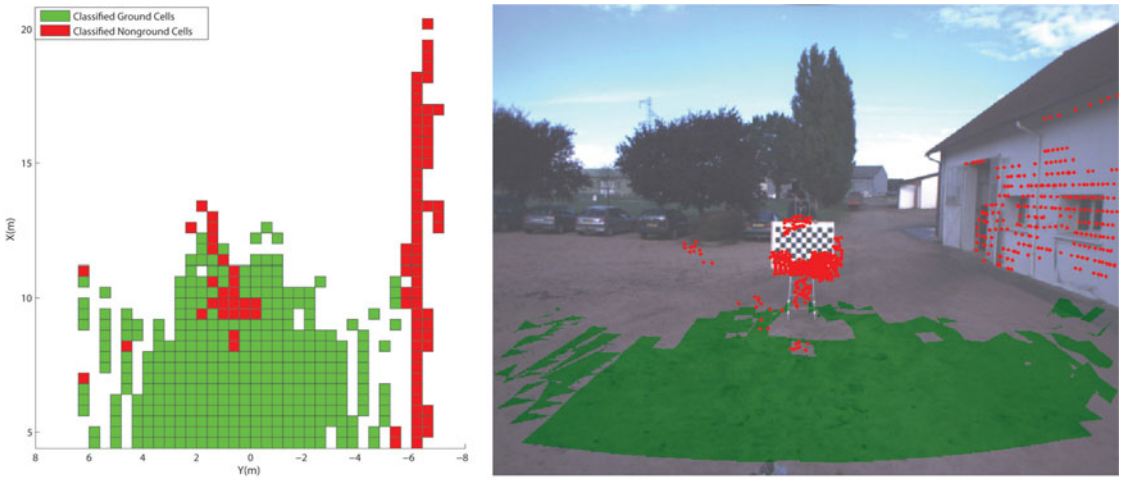

Fig. 5. (a) Traversability map obtained by the LIDAR-based ground classifier. (b) Results projected over the original image. Pixels associated with ground- (non-ground-) labeled cells are marked using green (red).

are projected over the co-located image for visualization and comparison purposes. Pixels associated with ground and non-ground cells are marked using green and red, respectively. The LIDAR-based classifier correctly detected both the obstacle in front of the vehicle and the building on the right side, as well as the traversable ground. For this scene, the precision and RP of the LIDAR-based classifier were 99.0% and 80.1%, respectively. Conversely, Fig. 6 shows the results obtained from the stereo-based classifier applied to the running example of Fig. 4(a). Again, cells labeled as ground are marked in green, whereas cells that are labeled as non-ground are denoted in red. Classification results are projected over the original image as shown in Fig. 6(b), where pixels associated with ground and non-ground cells are marked using green and red, respectively. The stereo-based classifier was also correct in labeling obstacles and traversable ground. The precision and RP of the stereo-based classifier resulted in 98.9% and 66.0%, respectively. When the two single-sensor classifiers are combined by weighting their results with the associated prior probability, the performance of the overall system resulted in a precision (P) and a RP of 99.3% and 76.5%, respectively. The traversability map obtained from the combined classifier is shown in Fig. 7. At a first glance, it may seem that the combined classifier performs worse than the LIDAR classifier in terms of RP. However, it should be noted that the single-sensor and the combined classifiers differ in field of view, thus a direct comparison of their classification performance is possible only when considering the cells that are "seen" simultaneously by both sensors. As a matter of fact, if a given cell is observed only by one sensor modality, no fusion will be possible and this cell will be labeled by the single-sensor classifier available. Table III collects the results obtained from each classifier over the common cell subset that amounts to about the 35% of the total number of labeled cells. These results clearly demonstrate that the statistical fusion of the two classifiers helps in increasing the information content and accuracy of the combined output.

Figure 8 shows a different scenario taken from a sequence where the vehicle drives on a country asphalt road. While the LIDAR-based module performs well, as shown in Figs. 8(a)–(b), the stereo-based classifier produces many FN due to the presence of heavy shadowing on the road, as shown in Fig. 8(c)–(d). The traversability map, obtained fusing the two classifiers, is shown in Fig. 8(e). For this scene, the LIDAR-based classifier provided a precision and RP of 99.0% and 98.3%, respectively,

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

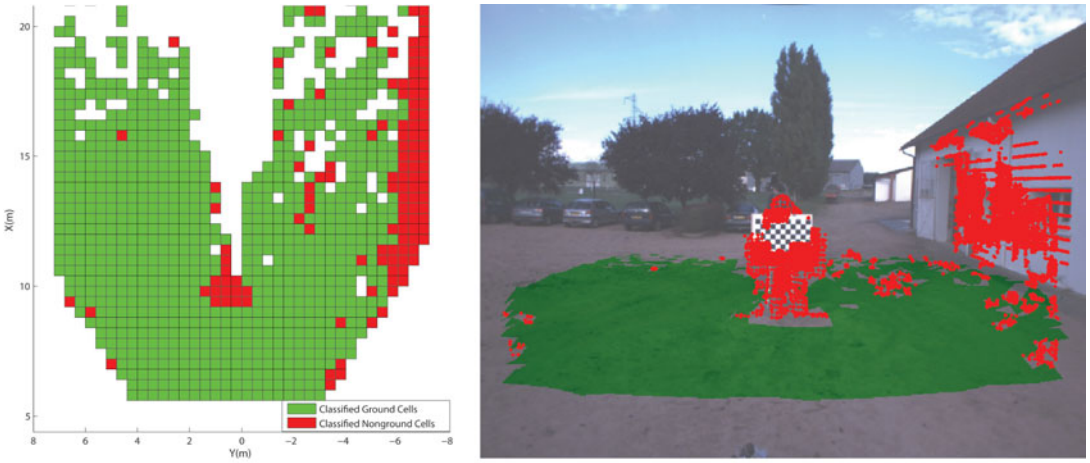

Fig. 6. (a) Traversability map obtained by the stereo-based ground classifier. (b) Results projected over the original image. Pixels associated with ground- (non-ground-) labeled cells are marked using green (red).

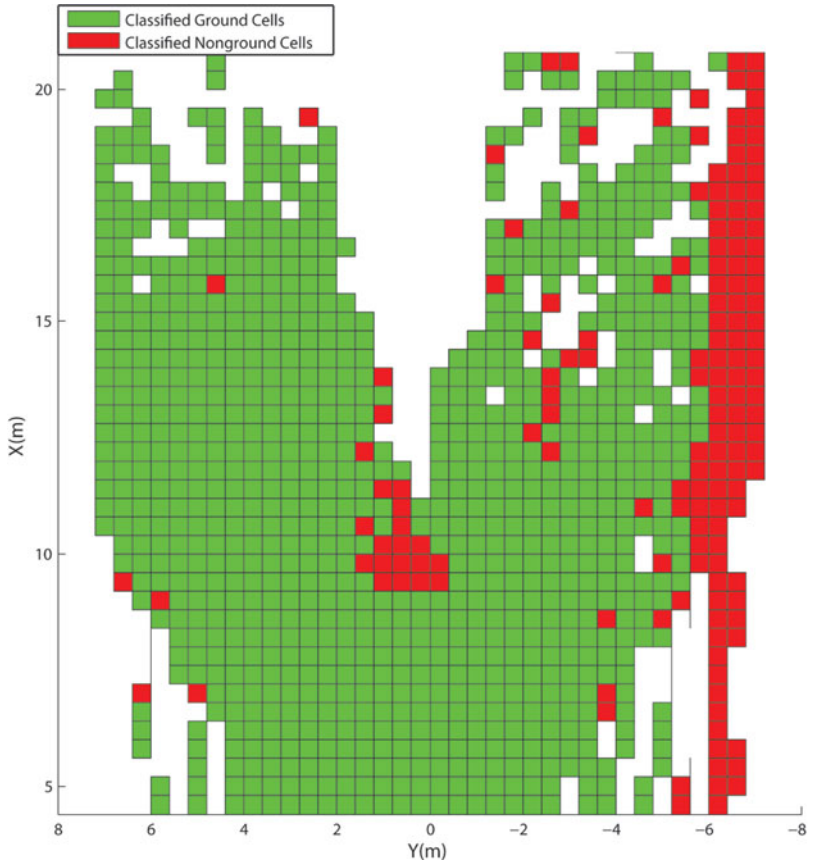

Fig. 7. Traversability map obtained by the combined LIDAR-stereo system for the scene of Fig. 4(a).

against values of 97.1% and 49.1%, respectively, for the stereo-based classifier. When the two systems are fused, P and RP resulted in 98.2% and 75.1%, respectively. Again, the combination of the two sensor modalities allowed the overall field of view to be increased, and the low RP provided by the vision to be compensated, while preserving, at the same time, high precision in detecting ground examples.

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

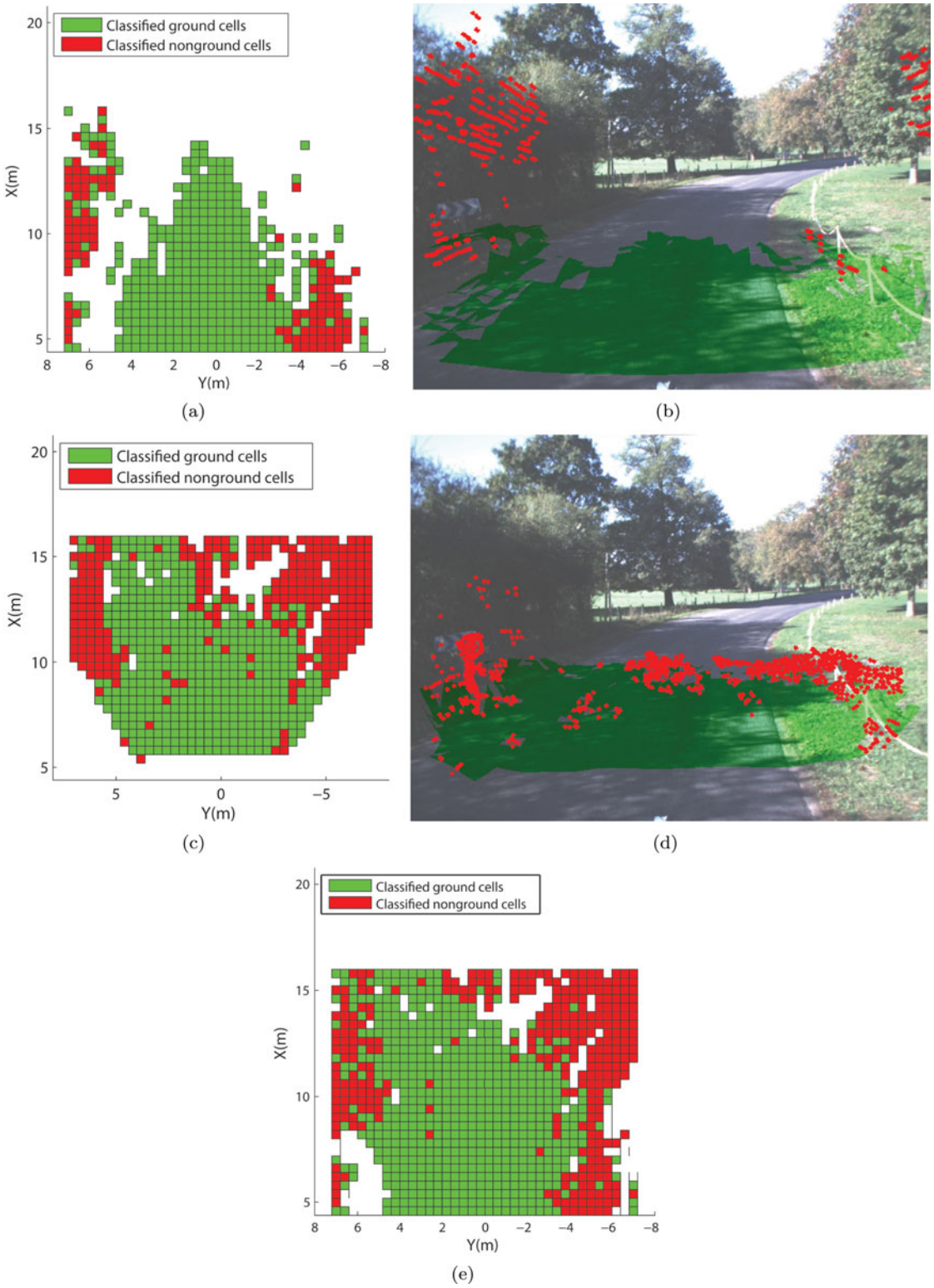

Fig. 8. (a) Traversability map obtained by the LIDAR-based ground classifier, (c) and the stereo-based ground classifier. (b), (d) Results projected over the original visual image. (e) Traversability map obtained by the combined LIDAR/stereo system. Ground-labeled cells and associated pixels are marked using green. Non-ground labeled cells and associated pixels are denoted in red.

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

Table IV. Classification results obtained from the single-sensor and combined classifiers for a subset of salient images.

	LIDAR-based	Stereo-based	Combined
Precision	97.3%	96.9%	97.4%
Rejection Precision	83.6%	82.6%	90.6%
Recall	97.5%	97.4%	98.7%
Specificity	82.7%	79.7%	82.8%
Accuracy	95.5%	95.1%	96.5%
F1-score	97.4%	97.2%	98.0%

Finally, another scenario is shown in Fig. 9, where the vehicle travels on agricultural terrain facing a small metallic pole on the right. The LIDAR-based system fails to detect the obstacle due to laser data sparseness in this area (see Figs. 9(a)–(b)), whereas the pole is correctly flagged by the stereo-based classifier (see Figs. 9(c)–(d)), and it is detected in the combined traversability map (see Fig. 9(e)). In this case, the vision module compensates for the LIDAR limitation by detecting thin obstacles, demonstrating once again the utility of a multi-sensor approach.

In summary, the combination of stereo and LIDAR data is useful in that (1) vision can help to overcome limitations of LIDAR, such as sparseness of data and low acquisition frequency, by producing dense maps at relatively high frequency, (2) being less affected by lighting conditions, LIDAR can help to overcome limitations of vision, such as reconstruction errors due to poor lighting conditions, shadows, and lack of texture. In addition, depending on the specific configuration of the sensors, an increment of the overall sensing area may be obtained.

8.2. Overall classification performance

A quantitative evaluation of system performance was obtained over a subset of salient images (i.e., s_b=113) taken from various data sets and referring to significant scenarios (see, for example, Figs. 10 and 11). This subset was manually labeled to gain reference ground-truth data. The combined LIDAR-stereo classifier was compared against the LIDAR and stereo classifier using only those cells that were observed by both sensors. For the selected frames, the "true" ground patches amounted to a total of 24,953, whereas the "true" non-ground patches were 3862. The results are collected in Table IV expressed in terms of the main classification metrics (i.e. precision, RP, recall, specificity, accuracy, F1-score). One can note that the combined LIDAR-vision system leads to a general improvement of all metrics with an accuracy of 96.5% and F1-score of 98.0%.

Some typical results obtained from the classifier are shown in Figs. 10 and 11. In each figure, the first row shows, from left to right, the original visual image overlaid with the points reconstructed by LIDAR, stereo and LIDAR-stereo fusion, respectively. Green dots denote points belonging to ground-labeled cells, whereas red dots are used to indicate pixels falling into non-ground-labeled cells. The second row shows, from left to right, the traversability map as obtained using only LIDAR data, only stereo data and the combination of both. Green cells denote ground patches, whereas red cells denote non-ground cells.

As a further demonstration, Fig. 12(a) shows the variation of F1-score for a sequence data set spanning a period of time of 180 s, during which the vehicle followed a path of about 500 m through a forested area. The two single-sensor classifiers are successfully fused with a general improvement of the single-sensor algorithms' performance. It is worth noting that the sensor fusion strategy ensures good results even when one of the two sensors performs poorly. This happens, for example, between the 135th and 145th second when the vehicle drives through a high-vegetated area and the stereo classifier's F1-score drops significantly due to the presence of heavy shadowing and saturation, as shown in the sample scene of Fig. 12(b). For the same sequence, it is also interesting to look at the adaptation of the ground model built by the stereo classifier, as shown in Fig. 13. For the reader's sake, only the average height, $\bar{h}$, of the 4D feature space (see previous Section 6.1) is plotted as a function of time and it is compared against the ground truth model obtained by manual inspection. As expected, the discrepancy of the stereo-generated model with true ground is large for the same time window.

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

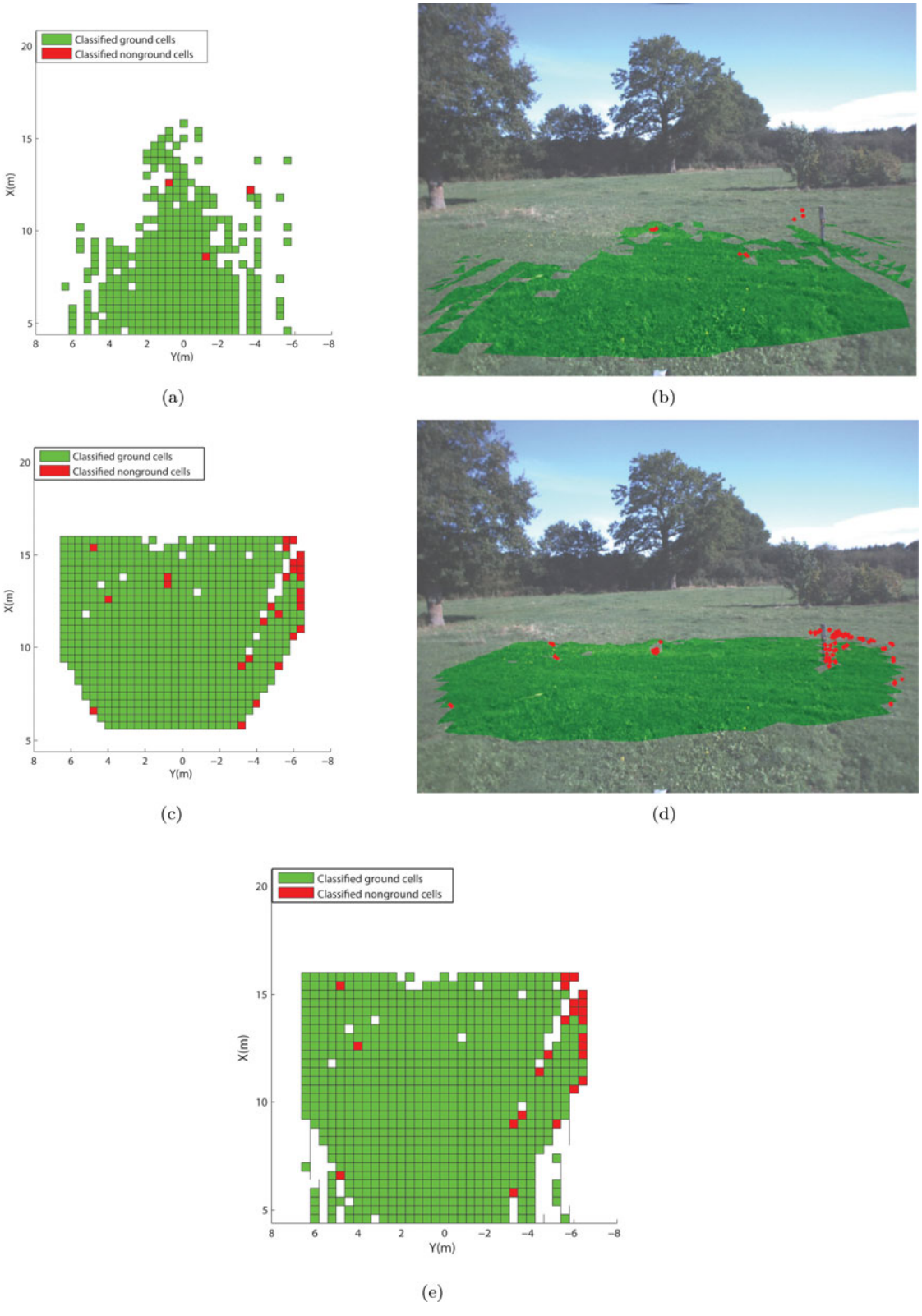

Fig. 9. (a) Traversability map obtained by the LIDAR-based ground classifier, (c) and by the stereo-based ground classifier. (b), (d) Results projected over the co-located visual image. (e) Traversability map obtained by the combined LIDAR/stereo system. Ground-labeled cells and associated pixels are marked using green. Non-ground labeled cells and associated pixels are denoted in red.

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

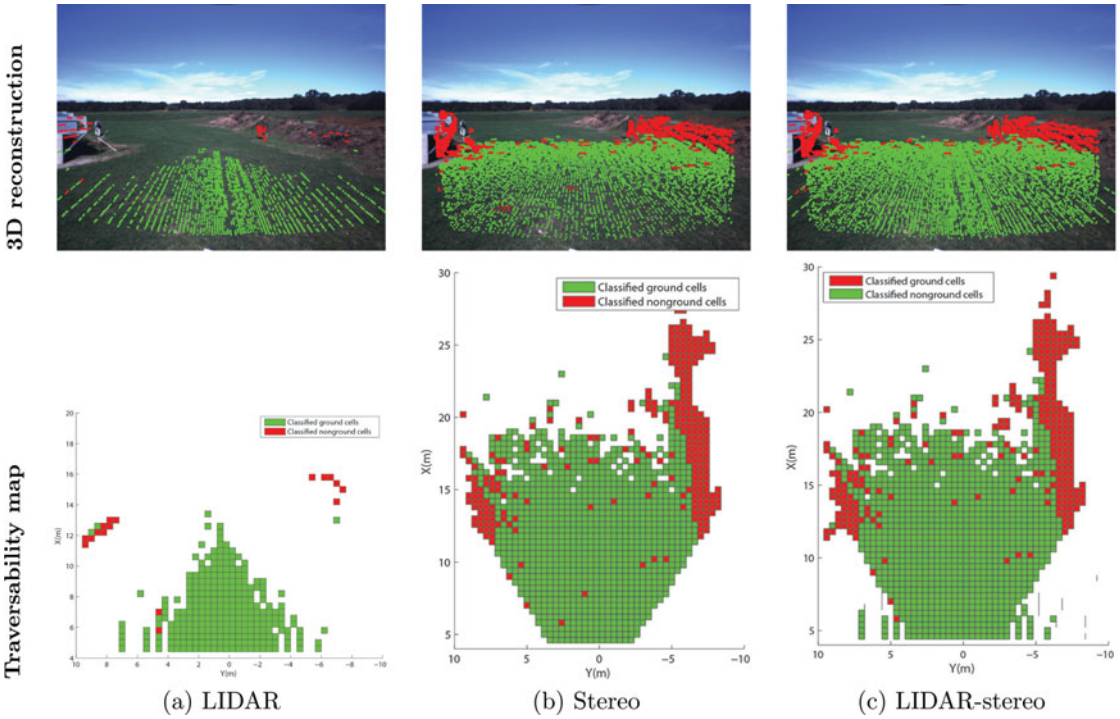

Fig. 10. Terrain estimation for a scene with agricultural terrain and mounds in the presence of people: (a) LIDAR classifier, (b) Stereo classifier, (c) LIDAR-stereo classifier.

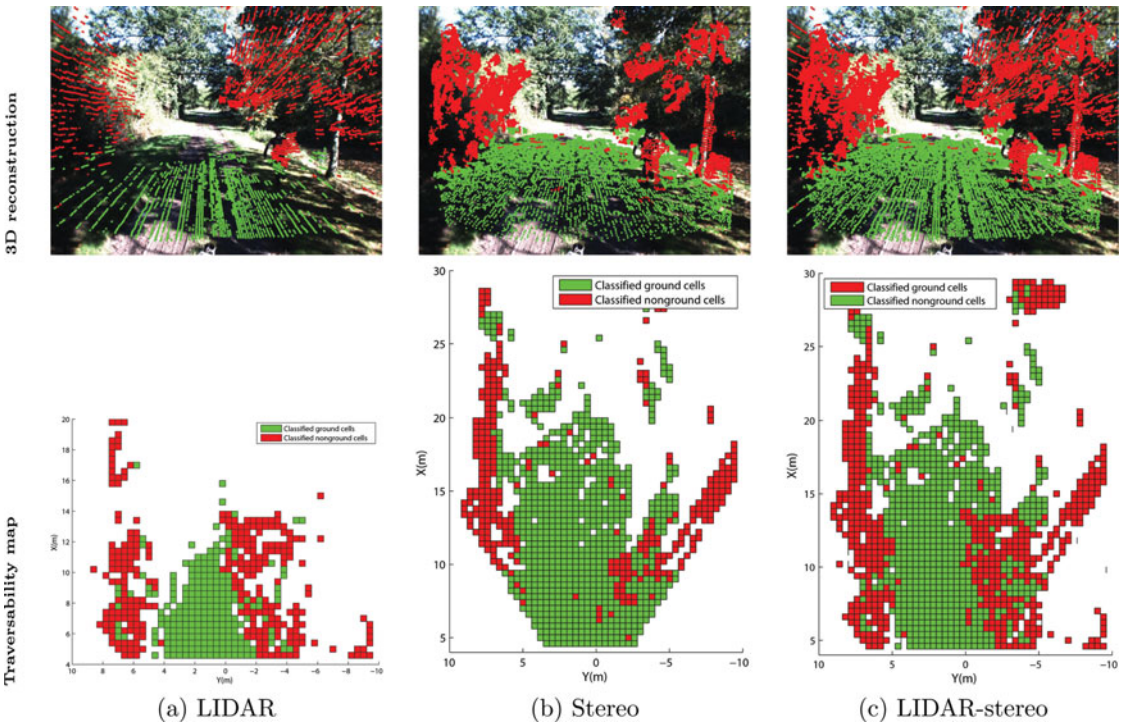

Fig. 11. Terrain estimation in a forested environment: (a) LIDAR classifier, (b) Stereo classifier, (c) LIDAR-stereo classifier.

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

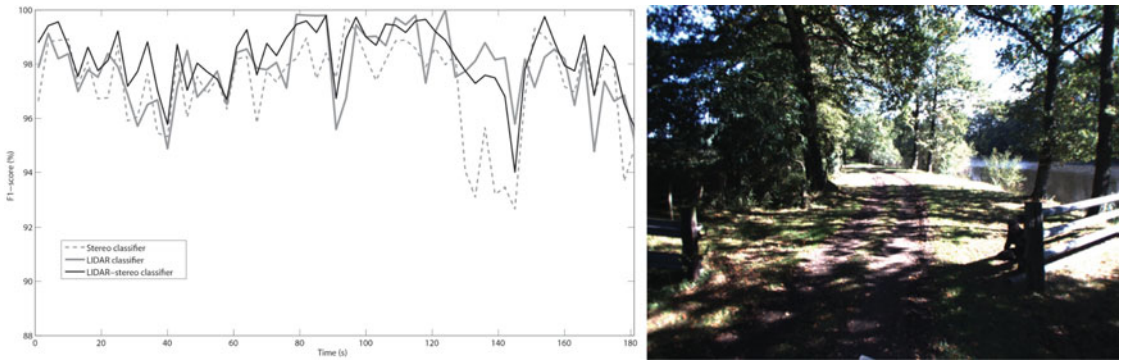

Fig. 12. (a) F1-score as a function of time for a sample sequence, (b) Visual image at the 140th second.

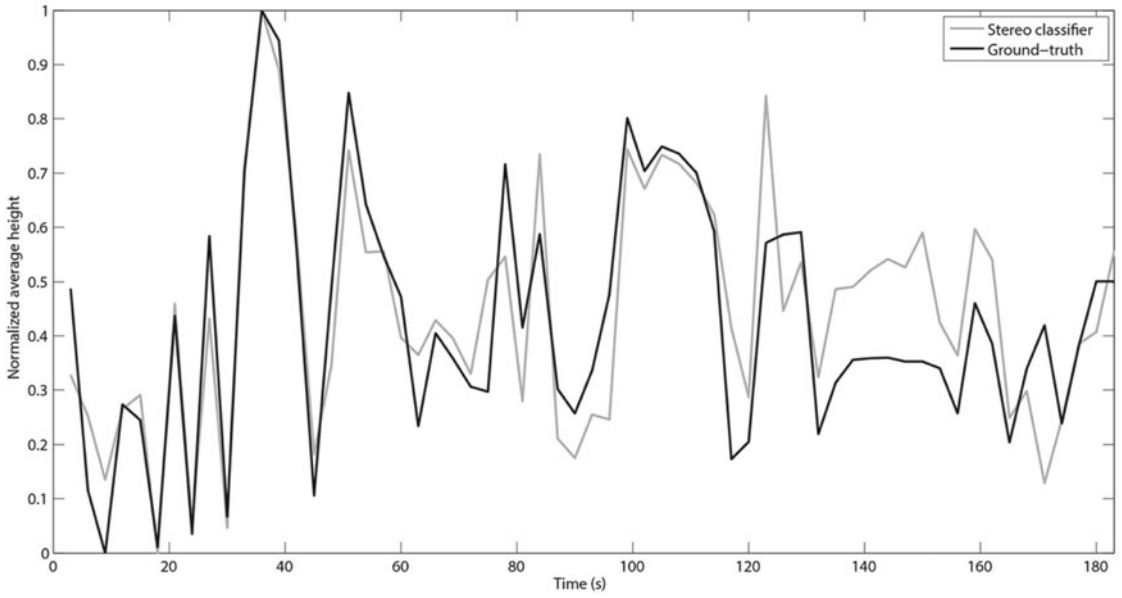

Fig. 13. Normalized average height $\bar{h}$, i.e., the fourth feature of the 4D feature ground model, for the stereo classifier as a function of time.

9. Conclusions

Integrating the data coming from different sensors is a challenging operation in robotic perception, but it can also significantly improve the overall system behavior by leveraging the strengths of both approaches and overcoming their limits. In this paper, a multi-sensor fusion method was presented for terrain analysis, where LIDAR and stereovision were combined within a statistical self-learning framework. Experimental results obtained using a test platform in natural scenarios validated this approach showing good classification performance. The classifier led to the following main advantages: (a) improvement of the perception performance of the combined LIDAR/stereo system due to complementary of the two sensor modalities, showing a classification precision and RP of 97.4% and 90.6%, respectively, (b) self-learning training of the system, where the sensors allows the vehicle to automatically acquire a set of ground samples, removing the need for time-consuming manual labeling, (c) continuous updating of the system during the vehicle's operation, thus making it adaptive and feasible for long-range and long-duration navigation applications.

In the current implementation, the main limitations of the system are: LIDAR-vision combination is based on fixed weights that were obtained offline using ground-truth data, no navigation system is employed to compensate for the tilt experienced by the vehicle during the experiments, and the system is started up under the assumption of absence of obstacles. Future research to address these issues will focus on the development of methods to update online the prior probabilities of the two single-sensor classifiers, inclusion of a full six-DOF navigation system, and implementation of

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press

alternative bootstrapping methods, e.g. looking at terrain patches that are directly in front of the robot.

Acknowledgments
The financial support of the ERA-NET ICT-AGRI, FP7 European Research Program, through the grant Ambient Awareness for Autonomous Agricultural Vehicles (QUAD-AV) is gratefully acknowledged.

References

1. O. Aycard, Q. Baig, S. Bota, F. Nashashibi, S. Nedevschi, C. D. Pantilie, M. Parent, P. Resende and Trung-Dung Vu, "Intersection Safety using Lidar and Stereo Vision Sensors," *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, (2011), pp. 863–869.
2. H. Badino, D. Huber and T. Kanade, "Integrating Lidar into Stereo for Fast and Improved Disparity Computation," *3D Imaging, Modeling, Processing, Visualization, Transmission (3DIMPVT)*, International Conference on, Hangzhou, China, (2011) pp. 1–8.
3. M. Bajracharya, M. W. Maimone and D. Helmick, "Autonomy for mars rovers: Past, present, and future," *Computer* **41**(12), 44–50 (2008).
4. G. Bradski and A. Kaehler, "*Learning OpenCV: Computer Vision with the OpenCV Library* (O'Reilly Media, 2008).
5. A. Broggi, A. Cappalunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P. Porta, M. Posterli and P. Zani, "Terramax vision at the urban challenge 2007," *IEEE Trans. Intell. Transp. Syst.* **11**(1), 194–205 (2010).
6. A. Broggi, E. Cardarelli, S. Cattani and M. Sabbatelli, "Terrain Mapping for Off-Road Autonomous Ground Vehicles using Rational B-Spline Surfaces and Stereo Vision," *IEEE Intelligent Vehicles Symposium* (2013), Gold Coast City, Australia, pp. 648–653.
7. A. Broggi, S. Cattani, M. Patander, M. Sabbatelli and P. Zani, A Full-3D Voxel-Based Dynamic Obstacle Detection for Urban Scenario using Stereo Vision," *International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)* (2013) pp. 71–76.
8. H. A. Dahlkamp, D. S. Kaehler, S. Thrun and G. Bradski, "Self-Supervised Monocular road Detection in Desert Terrain," *Robotics Science and Systems Conference* (2006), Philadelphia, USA, pp. 1–6.
9. C. Dima, N. Vandapel and M. Hebert, "Classifier fusion for outdoor obstacle detection," *IEEE International Conference on Robotics and Automation* (2004), New Orleans, LA, USA, pp. 665–671.
10. P. Foo and G. NG, "High-level information fusion: An overview," *J. Adv. Inform. Fusion* **8**(1), 33–72 (2013).
11. R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffie and K. Kavukcuoglu, "Learning long-range vision for autonomous off-road driving," *J. Field Robot.* **26**(2), 120–144 (2009).
12. T. Hague, J. Marchant and N. Tillett, "Ground-based sensing systems for autonomous agricultural vehicles," *Comput. Electron. Agric.* **25**(1–2), 11–28 (2000).
13. T. Hastie, R. Tibshirani and J. Friedman, "*The Elements of Statistical Learning* (Springer-Verlag, New York, 2003).
14. K. Konolige, M. Agrawal, M. R. Blas, R. C. Bolles, B. P. Gerkey, J. Solà and A. Sundaresan, "Mapping, navigation, and learning for off-road traversal," *J. Field Robot.* **26**(1), 88–113 (2009).
15. K. Konolige, J. Bowman, J. Chen, P. Mihelick, M. C. V. Lepetit and P. Fua, "View-based maps," *Int. J. Robot. Res.* **29**(8), 941–957 (2010).
16. J. Lalonde, N. Vandapel, D. Huber and M. Hebert, "Natural terrain classification using three-dimensional ladar data for ground robot mobility," *J. Field Robot.* **23**(10), 839–861 (2006).
17. R. Manduchi, A. Castano, A. Talukder and L.Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Auton. Robot* **18**, 81–102 (2004).
18. A. Milella and G. Reina, "3D reconstruction and classification of natural environments by an autonomous vehicle using multi-baseline stereo," *Intell. Serv. Robot.* **7**, 79–92 (2014).
19. A. Milella, G. Reina and J. Underwood, "A self-learning framework for statistical ground classification using radar and monocular vision," *J. Field Robot.* **32**(1), 20–41 (2015).
20. H. Mousazadeh, "A technical review on navigation systems of agricultural autonomous off-road vehicles," *J. Terramech.* **50**(3), 211–232 (2013).
21. A. Nüchter, K. Lingemann, J. Hertzberg and H. Surmann, "6D SLAM–3D mapping outdoor environments," *J. Field Robot.* **24**, 699–722 (2007).
22. S. Nedevschi, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, T. Graf and R. Schmidt, "High Accuracy Stereovision Approach for Obstacle Detection on Non-Planar Roads," *Proceedings of IEEE INES*, (2004) Cluji-Napoca, Romania (2004) pp. 211–216.
23. K. Nickels, A. Castano and C. M. Cianci, "Fusion of Lidar and Stereo Range for Mobile Robots," *International Conference on Advanced Robotics* (2003), Coimbra, Portugal, pp. 1–6.
24. M. Perrollaz, J. D. Yoder and C. Laugier, "Using Obstacles and Road Pixels in the Disparity-Space Computation of Stereo-Vision Based Occupancy Grids," *IEEE Conference on Intelligent Transportation Systems (ITSC)* (2010) pp. 1147–1152.
25. Point Grey, "Triclops software development kit," Available at: http://www.ptgrey.com/triclops. [accessed on June 1st, 2015].

26. J. Poppinga, A. Birk and K. Pathak, "Hough-based terrain classification for realtime detection of drivable ground," *J. Field Robot.* **25**(1–2), 67–88 (2008).
27. G. Reina, G. Ishigami, K. Nagatani and K. Yoshida, "Vision-Based Estimation of Slip Angle for Mobile Robots and Planetary Rovers," *Proceedings of IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA (2008) pp. 486–491.
28. G. Reina and A. Milella, "Toward autonomous agriculture: Automatic ground detection using trinocular stereovision," *Sensors* **60**(11), 12405–12423 (2012).
29. G. Reina, A. Milella, W. Halft and R. Worst, "LIDAR and Stereo Imagery Integration for Safe Navigation in Outdoor Settings," *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)* (2013) pp. 1–6.
30. G. Reina, A. Milella and J. Underwood, "Self-learning classification of radar features for scene understanding," *Robot. Auton. Syst.* **60**(11), 1377–1388 (2012).
31. E. Rohmer, G. Reina and K. Yoshida, "Dynamic simulation-based action planner for a reconfigurable hybrid legwheel planetary exploration rover," *Adv. Robot.* **24**(8–9), 1219–1238 (2010).
32. P. Santana, M. Guedes, L. Correia and J. Barata, "A Saliency-Based Solution for Robust Off-Road Obstacle Detection," *IEEE International Conference on Robotics and Automation*, Anchorage, Alaska, USA (2010), pp. 3096–3101.
33. D. Tax, One-Class Classification. Concept Learning in the Absence of Counter Examples *Ph.D. Thesis* (Delft University of Technology, Delft, Netherlands, 2001).
34. A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke and D. Cremers, "B-spline modeling of road surfaces with an application to free-space estimation," *IEEE Trans. Intell. Transp. Syst.* **10**(4), 572–583 (2009).
35. U. Weiss and P. Biber, "Plant detection and mapping for agricultural robots using a 3D LIDAR sensor," *Robot. Auton. Syst.* **59**(5), 265–273 (2011).
36. C. Wellington and A. Stentz, "Online Adaptive Rough-Terrain Navigation in Vegetation," *Proceedings of International Conference on Robotics and Automation* (2004), New Orleans, LA, USA, pp. 96–101.
37. K. M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss and W. Burgard, "Octomap: A Probabilistic, Flexible, and Compact 3D Map Representation for Robotic Systems," *ICRA Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation*, Anchorage, Alaska, USA (2010), pp. 1–8.
38. Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," *Proceedings of IEEE/RSJ International Conference onIntelligent Robots and Systems*, Sendai, Japan, vol. 3 (2004) pp. 2301–2306.
39. J. Zhao, J. Katupitiya and J. Ward, "Global Correlation Based Ground Plane Estimation using v-Disparity Image," *International Conference on Robotics and Automation*, Rome, Italy (2007), pp. 529–534.
40. S. Zhou, J. Xi, M. W. McDaniel, T. Nishihata, P. Salesses and K. Iagnemma, "Self-supervised learning to visually detect terrain surfaces for autonomous robots operating in forested terrain," *J. Field Robot.* **29**(2), 277–297 (2012).

https://doi.org/10.1017/S0263574715000442 Published online by Cambridge University Press