

3D LIDAR- and Camera-Based Terrain Classification Under Different Lighting Conditions

Stefan Laible, Yasir Niaz Khan, Karsten Bohlmann and Andreas Zell

Chair of Cognitive Systems, Computer Science Department, University of Tübingen, Sand 1,
72076 Tübingen, Germany

Abstract. Terrain classification is a fundamental task in outdoor robot navigation to detect and avoid impassable terrain. Camera-based approaches are well-studied and provide good results. A drawback of these approaches, however, is that the quality of the classification varies with the prevailing lighting conditions. 3D laser scanners, on the other hand, are largely illumination-invariant. In this work we present easy to compute features for 3D point clouds using range and intensity values. We compare the classification results obtained using only the laser-based features with the results of camera-based classification and study the influence of different lighting conditions.

1 Introduction

One of the most frequently mentioned applications for terrain classification in robotics is the recognition and avoidance of impassable terrain. The robot must be able to decide whether the terrain ahead is passable easily, passable with caution, or whether it is better to avoid this terrain and to plan another path. Apart from this, terrain classification can also improve the self-localization of a robot. GPS often is not very accurate and odometry is error-prone, especially in rough outdoor environments. Then, the knowledge of the terrain can be useful if the robot should follow a dirt road or a field boundary, for example. In our experiments we consider five types of terrain that are often encountered (see Fig. 1).

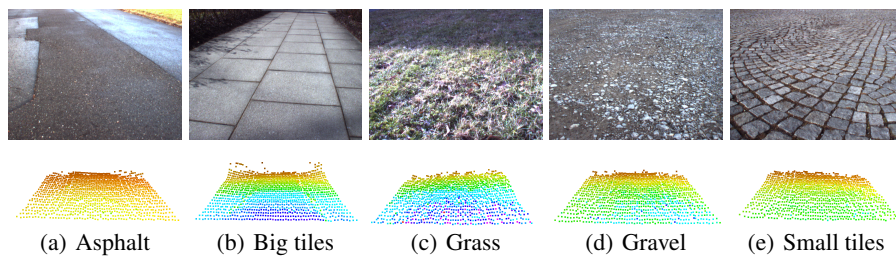


Fig. 1. Images and 3D scans of the considered five terrain types

A major challenge in the field of outdoor robotics is the changing lighting conditions. The texture of the ground may look very different depending on time of day and

the currently prevailing weather conditions. These external factors influence the results of camera-based terrain classification. Laser scanners, however, are largely independent of these outdoor conditions and 3D LIDARs, which scan an entire area at a high frame rate, provide enough information for terrain analysis. In this paper we show how to extract features from these 3D scans that are fast to compute and yet are capable of distinguishing different terrain types. We consider four very different lighting conditions to test the independence assumption and compare the results with a previously developed camera-based method [KKZ11].

2 Related Work

Camera-based approaches for terrain classification are well-studied. There, the problem is to find efficient and discriminative representations of texture information. This has been done, for instance, in terms of co-occurrence matrices, Local Binary Patterns, and texton-based approaches. For the experiments in this paper we use the methodology of an earlier work where we achieved good results using local image descriptors with a grid-based approach [KKZ11] (see Sec. 4.2).

There exist several approaches for terrain classification that use range data in addition. In [Ras02] color and texture features are combined with geometric features obtained from laser data for the purpose of road detection. A method for classifying the traversability of terrain is proposed in [HOJ06]. A stereo camera provides the data to learn geometric features for traversability, and color information is then used to enhance the geometric information. For a quick adaptation to different lighting conditions color models are learned in an unsupervised fashion. In [HLAP11] they use the 3D laser Velodyne HDL-64E S2 in addition to color cameras. The high-resolution data that this laser delivers is used together with color and texture information to classify the terrain in three classes: road, rough and obstacle. They further apply a Markov random field in order to take into account the context-sensitivity of the individual terrain grid cells.

In the just-mentioned approaches only the geometric information from range data is used. But most lasers also provide intensity values (also called remission or reflectance values), which indicates the proportion of the emitted light that arrives back at the laser. In [WSKB09] these intensity values are used to detect grass-like vegetation. They are able to distinguish between street and grass with an accuracy of over 99%. As an explanation of why this works so well, they state an effect well known from satellite images analysis [MHSM95], namely that chlorophyll, a green pigment found in almost all plants, strongly reflects near-IR light, such as that of a laser. Apart from the material the intensity values also depend on the distance and the angle of incidence of the laser beams.

We will show that with these values and the features presented in Sec. 4.1 not only grass and non-grass, but several terrain classes can be distinguished from each other. In a previous work where we presented a method for classifying plant species using a 3D LIDAR sensor and supervised learning [WBL⁺10] we also experienced the discriminative power of the intensity values as the features based on these values were the most important. There, the same 3D laser scanner was used, which we use here and which is described in the next section.

3 Hardware

All experiments were performed with our outdoor robot *Thorin*, which can be seen in Fig. 2. The robot is equipped with a Mini-ITX computer featuring a 2.26 GHz Core 2 Duo Mobile CPU and has, among other sensors, a Marlin F-046 C Color Camera and a FX6 3D Laser Scanner by Nippon Signal.



Marlin F-046 C Color Camera	
Vendor	Allied Vision Technologies GmbH
Resolution	780 x 582 pixels
Frame Rate	Max. 53 Hz
FX6 3D Laser Scanner	
Vendor	Nippon Signal Co., Ltd.
Resolution	29 x 59 data points
Frame Rate	8 or 16 Hz
Range	16 m
Scan Area	50° (hor.) and 60° (vert.)

Fig. 2. Outdoor robot *Thorin* with a Nippon Signal FX6 3D laser scanner and an AVT Marlin F-046 C color camera

The color camera is able to take pictures at a frame rate of up to 53 Hz. It has both manual and automatic white balance as well as an auto shutter and auto gain function. All three auto functions were enabled for the experiments, as this is necessary especially in changing lighting conditions.

The FX6 sensor uses a pulse laser in the near-IR range. It is lightweight and robust and largely illumination-independent, so that it works with ambient light of up to 100,000 Lux. Thus it is best suited to be used in different light and weather conditions. A drawback of the sensor is its low resolution with only 29 x 59 data points. In addition to the distance an intensity value is returned for each point, which indicates the proportion of the emitted light which arrives back at the sensor.

4 Terrain Classification

4.1 3D LIDAR-Based Feature Extraction

Filtering Since the laser scanner yields a lot of erroneous measurements, filtering of the point cloud is a crucial preprocessing step. Some mismeasurements can be easily identified by looking at the intensity values. Namely, extremely low or high values indicate that it is almost certainly not a valid measurement.

To reduce the noise in the data, every point of the cloud that does not have a certain number of neighbors within a fixed radius, is deleted. Filtering often removes up to 20-25% of the data.

Ground-plane detection After filtering the point cloud the next step is to detect the ground plane, which is an easy task if it includes most of the points (see Fig. 3(a)). Then one can simply apply sample consensus methods like RANSAC to find the largest plane in the scene. These iterative methods estimate the parameters of a mathematical model that best fits the observed data. We use a variant of the RANSAC algorithm called MSAC [TZ00] which uses an M-estimator-based error function and provides a more robust estimate without additional computational effort.

Scenes with multiple planes, especially when the ground plane is not the largest, are more difficult to treat, e.g. when the robot is driving near a house wall (see Fig. 3(b)). Then it is no longer enough to consider each frame individually. A simple yet robust method is to only consider points near a reference plane and to update the parameters of this plane every frame:

1. From the set P of all points consider only those points $P' = \{(x, y, z) | (x, y, z) \in P \wedge \hat{a} \cdot x + \hat{b} \cdot y + \hat{c} \cdot z + \hat{d} \leq d_{min}\}$ near the reference plane $R: \hat{a} \cdot x + \hat{b} \cdot y + \hat{c} \cdot z + \hat{d} = 0$ for a fixed threshold d_{min}
2. Estimate the parameters of the largest plane in P' : $(a, b, c, d) = \text{MSAC}(P')$
3. Update the reference plane:
 $\hat{a} := \hat{a} + (a - \hat{a})/n, \quad \hat{b}, \hat{c}, \hat{d}$ resp.
 $n := n + 1$

The implicit assumption made here is that the terrain can be described by a planar surface. As the robot drove only on relatively flat terrain in our experiments and the scanning range of the laser in front of the robot is very limited, this assumption is justified. For uneven terrain this model has to be extended appropriately, e.g. by hierarchically refining the plane.

With the known parameters of the ground-level plane the data points can be transformed so that the points belonging to the ground lie in the x - y -plane with the z -axis pointing upwards and the origin projected onto the plane. Thereafter, a grid is laid on the plane and only those cells in which there is a minimum number of data points are considered for classification. A pre-classification can be made by looking at the maximum height of the points of a cell, which now is simply the maximum z value. According to a chosen threshold the cells can be classified as drivable and non-drivable. The drivable cells can then be further divided into the different terrain classes. This is of course a much too simple approach to detect obstacles, however, the focus of this work lies not in obstacle avoidance but in distinguishing different terrain types.

Features Now, for every drivable cell, characteristic features can be calculated. Due to the low resolution of the laser scanner only very few points per cell are present. Despite this limitation the features must be discriminative enough to distinguish between different terrain types. Below we present the eleven features used.

The height of the points above the surface not only is suitable for distinguishing between drivable and non-drivable cells, but also within the drivable-terrain class height features can be used to distinguish between smooth and rough terrains:

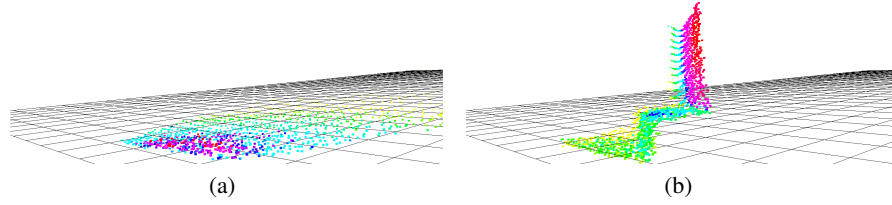


Fig. 3. Ground-plane detection: The ground plane is detected even when there are larger planes in the scene.

1. *Maximum height*: The maximum height of all points of a terrain grid cell regarding the detected ground plane. After the transformation of the point cloud this is simply the maximum z value.
2. *Standard deviation of height*: The standard deviation of all heights. Again, this is just the standard deviation of all z values.

As stated in Sec. 2 the intensity values provide good features for terrain classification:

3. - 4. *Minimum and maximum intensity*: The minimum and maximum intensity value of all points of a cell.
5. *Range of intensity*: The difference between the minimum and maximum intensity value.
6. - 8. *Mean, median and standard deviation of intensity*: The mean, median and standard deviation of the intensity values of all points of a cell.

The intensity values not only depend on the material but also on the distance and the angle of incidence. And since the number of points of a cell tends to decrease with the distance and thus also is characteristic, the remaining features are:

9. *Distance*: The distance of the grid cell to the laser origin.
10. *Angle of incidence*: The angle between the ground plane and the vector from the grid cell to the laser origin.
11. *Number of points*: The number of data points belonging to a grid cell.

4.2 Camera-Based Feature Extraction

For camera-based classification, a virtual grid is drawn on the image and local features are calculated across this grid. For this application we take three local image descriptors: Local Binary Patterns (LBP), Local Ternary Patterns (LTP) and TSURF. Two of these, LBP and LTP, are texture-based image descriptors, whereas TSURF is an interest point image descriptor. These image descriptors have shown their strength in terrain classification on mobile robots [KKZ11].

LBP is a very simple image descriptor which is calculated for each pixel considering a neighbourhood of 3×3 . All the neighbours of this pixel are thresholded based on the center pixel and resulted values are concatenated to form an 8-bit binary pattern

for this pixel. These binary patterns are calculated for all pixels in a patch and then a histogram is calculated from all of these patterns. This histogram then gives a 256-bit texture descriptor for this patch. LTP is an extension to LBP, in which a threshold value is introduced while thresholding the neighbourhood pixels. This gives a ternary pattern which is split into two binary patterns. These patterns are then concatenated and a histogram is generated which gives a 512-bit texture descriptor for the patch.

TSURF is an extension of the SURF image descriptor. In SURF, interest points are detected in an image which can be distinctively identified. These interest points are then described using box filters. They enable to track features across images. In TSURF, interest points are assigned to the intersections of the grid drawn on the image and then descriptors are calculated to describe these points. Details can be found in [KKZ11].

4.3 Classification with Random Forests

It has been found that for both the laser and the camera-based classification Random Forests are very well suited [Bre01]. A random forest is a collection of multiple decision trees. In the training phase each tree is constructed by not considering all of the N training instances but by randomly choosing N instances with replacement; so some are chosen more than once, while others are omitted. Then, for each node of each tree a different, also randomly chosen subset of all the features is considered to find the best split. In the classification phase a new sample is pushed down each tree so that it ends up in a leaf node with a corresponding class label. The final classification result then is the majority of labels assigned to the sample considering all trees of the forest.

5 Experiments and Results

In order to assess the capabilities of the classification methods described in Sec. 4 under varying lighting conditions, we tested them in four different settings:

1. *Cloudy morning*: no direct sunlight, soft shadows
2. *Sunny midday*: lots of sun, harsh shadows
3. *Dusk*: rapidly changing illumination
4. *Night*: diffuse light from street lamps

First, the laser- and camera-based approach were tested for each setting separately to see how well the two of them deal with each in particular. Then, for each of the two approaches a general model for all scenarios was built by taking the same number of training data from each setting. Both methods are grid-based and have been tested with different resolutions. In the laser-based method the grid is placed on the detected ground-plane in three-dimensional space and the height and width of a square grid cell is given in meters. In the camera-based approach a grid is drawn on the two-dimensional image and the size is measured in pixels. Only for cells that are located relatively close to the front of the robot ($\lesssim 2\text{m}$) enough laser measurements are available to classify the cell. To enable a fair comparison, only image patches belonging to this area were considered. In the camera-based approach each of the three descriptors was tested separately and then, for every setting and resolution, the best one was chosen. Fig. 4 shows

the true positive rates for the different settings and grid resolutions after a five-fold cross-validation using random forests with 100 trees each.

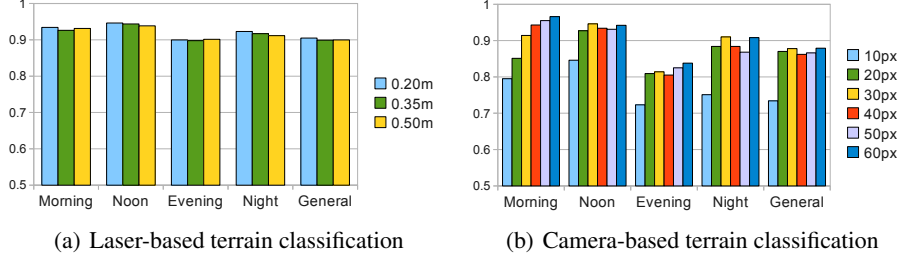


Fig. 4. Classification results (true positive rates) for different lighting conditions and grid resolutions. In (b), for every setting and resolution, the best result of the three descriptors is shown.

The laser-based classification provides consistently good results regardless of external conditions and the chosen grid resolution, with a true positive rate of up to 90.5% for the general model (see Fig. 4(a)). It performs best in the noon setting with a true positive rate of up to 94.6%. Only considering the two classes of grass and asphalt for classification demonstrates the laser scanner’s ability to detect vegetation as mentioned in Sec. 2. Then, the true positive rates for all settings and all grid resolutions are above 99.9% (see Fig. 5(b) for an example of a classified point cloud).

In the camera-based classification the results depend on the chosen grid resolution. LBP and LTP perform better with larger grid cells, while LTP outperforms LBP in all respects. LTP achieves the best results in the morning (96.6%) and at noon (94.2%) and the worst results in the evening (83.8%) and at night (78.3%), considering a grid resolution of 60 pixels each. TSURF performs better than LTP at small grid sizes and worse at larger ones, since the results of TSURF may get worse again with larger grid sizes. In contrast to LTP it copes very well with the night setting and achieves 91.0% for a grid size of 30 pixels, but the best result in the evening setting is only 81.4%, again for a resolution of 30 pixels. By selecting the most appropriate descriptor in each case, good results are obtained for all scenarios, only the true positive rates in the evening setting are somewhat lower (see Fig. 4(b)). That the results in the night setting are so good seems counterintuitive, but can be explained by the fact that there is little but constant light without harsh shadows.

6 Conclusions

In this work we studied 3D LIDAR- and camera-based terrain classification under different lighting conditions. We presented easy to compute 3D LIDAR features, and it turned out that the intensity values provide the most characteristic features and that the classification results are largely illumination-invariant and independent of the chosen grid resolution. In the general case, regarding all scenarios and five terrain classes, true positive rates of up to 90.5% were achieved. When only considering the two classes of grass and asphalt, the rates for all settings were above 99.9%. A disadvantage of the 3D

LIDAR is the limited range due to its low resolution. In the camera-based classification the results were somewhat more dependent on the lighting conditions and the resolution but are remarkably good in the night setting with a true positive rate of up to 91.0%. In the general case, rates of up to 87.9% were achieved, and the best results were made in the morning setting with rates up to 96.6%.

Since the random forest not only returns a class label, but also the proportion of trees that have voted for this class, one gets a measure that indicates how confident the corresponding classifier is; this can be used for sensor fusion in future work.

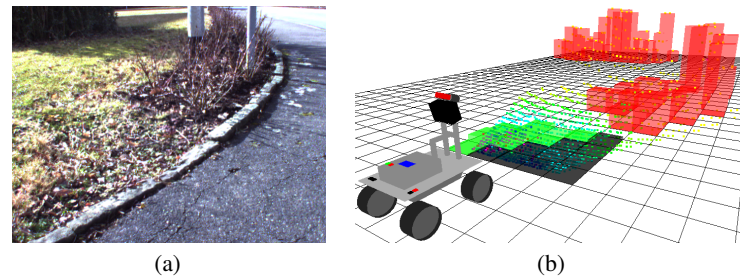


Fig. 5. Camera image and elevation map with laser-based terrain classification. Green: grass, black: asphalt, red: high elevations

References

- [Bre01] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [HLAP11] M. Häselich, D. Lang, M. Arends, and D. Paulus. Terrain classification with markov random fields on fused camera and 3d laser range data. In *Proceedings of the 5th European Conference on Mobile Robotics (ECMR)*, pages 153–158, 2011.
- [HOJ06] M. Happold, M. Ollis, and N. Johnson. Enhancing supervised terrain classification with predictive unsupervised learning. In *Robotics: Science and Systems*, 2006.
- [KKZ11] Y. N. Khan, P. Komma, and A. Zell. High resolution visual terrain classification for outdoor robots. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1014–1021, Barcelona, Spain, nov. 2011.
- [MHSM95] R. B. Myneni, F. G. Hall, P. J. Sellers, and A. L. Marshak. The interpretation of spectral vegetation indexes. *Geoscience and Remote Sensing, IEEE Transactions on*, 33(2):481–486+, 1995.
- [Ras02] C. Rasmussen. Combining laser range, color, and texture cues for autonomous road following. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 4320–4325, 2002.
- [TZ00] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.
- [WBL⁺10] U. Weiss, P. Biber, S. Laible, K. Bohlmann, and A. Zell. Plant species classification using a 3d lidar sensor and machine learning. In *Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on*, pages 339–345, dec. 2010.
- [WSKB09] K. M. Wurm, C. Stachniss, R. Kümmerle, and W. Burgard. Improving robot navigation in structured outdoor environments by identifying vegetation from laser data. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, St. Louis, MO, USA, oct. 2009.