Ground segmentation from outdoor environments in rural areas

Gabriel da Silva Vieira[*§1], Fabrizzio A.A.M.N. Soares[*2], Samuel A. Santos[§3], Gustavo T. Laureano[*4], Junio Cesar de Lima[§5], Ronaldo M. Costa[*6], Juliana Paula Félix[*7], Thamer H. Nascimento[§8]

[*]Federal University of Goiás, *Pixelab Laboratory*. Goiânia - GO, Brazil.

{fabrizzio[2], gustavo[4], ronaldocosta[6], julianafelix[7]}@inf.ufg.br

[§]Federal Institute Goiano, *Computer Vision Laboratory*. Urutaí - GO, Brazil.

{gabriel.vieira[1], junio.lima[5], thamer.nascimento[8]}@ifgoiano.edu.br, samuelalvesv4[3]@gmail.com

***Abstract*—The understanding of the complexity of outdoor environments is an essential issue for the development of efficient processes of autonomous mobility, especially in areas with uneven illumination and without a well-defined road. In this context, the detection of ground and obstacles plays a relevant role in giving the first impressions of the external surroundings to a machine. Furthermore, it can guide independent movements and decisions. In this study, we introduce a segmentation method that detects ground and non-ground points of complex scenes under different exposures to illumination, textures, and shading. We prepared a dataset with images collected from some environments in which trees are prominent obstacles. The proposed method uses contrast templates, statistical measures, and morphological operators to reach the ground segmentation. Experiments showed satisfactory results in which trees were well detected and the ground was efficiently segmented with the maintenance of the structure of the image.**

Keywords: image segmentation; tree trunk detection; ground detection; feature engineering; computer vision.

I. Introduction

In this study, we present an approach to segment ground areas from unstructured and dynamic environments to deal with the perception of the scenes for autonomous navigation. It is an essential task for outdoor mobile robots because the delineation of the ground can be used to estimate the traversability of the terrain in order to navigate safely, as avoiding obstacles.

Traversability estimation is quite challenging due to a variety of adversities that make it an ill-posed problem. Terrain characteristics, such as relief, slopes, depressions, and obstructions, as well as natural light and overlapping textures, increase the challenge a bit more. Thus, the scope of practical applications is defined by making some assumptions to simplify the problem, i.e., it is necessary to know what the vision module is supposed to do and what it is not [1].

In urban areas, typical properties of roads or highways can be considered to model a suitable solution. For example, Soquet et al. [2] proposed an approach to find a path in the free space of urban roads by estimating the geometric content of the road scenes. Cubber et al. [3] used a similar approach to distinguish traversable from non-traversable areas. In other studies, the segmentation of road regions was made considering the shape of the roads on which line segments and local orientations were used to estimate vanishing points [4], [5].

In rural environments, the detection of road regions is also a difficult or even more complicated task due to the absence of common properties such as those found on urban roads [6]. The properties of these roads can change along the way, making it difficult to design a model capable of handling different scenarios. Besides, terrain irregularities, unclear road boundaries, and different surrounding obstacles as vegetation, rocks, and trees, may appear in different formats which are hard to predict.

However, solutions are being developed, primarily to comply with agricultural demands. Mendes et al. [7] developed a vision-based detector to carry-out crop monitoring tasks in steep slope vineyards. Cabo et al. [8] proposed an automatic dendrometry method to estimate the diameter of trees using laser scanning. Liu et al. [9] dealt with the detection of citrus fruits and tree trunks in the context of intelligent fruit-harvesting.

In these cases, the detection of potential obstacles and the estimation of the free space are also important for autonomous driving. The problem is that in such cases there are no friendly road properties, or even roads may not exist. In this way, these proposals often deal with the collection of key features and the development of training models, thus being treated as a machine learning problem.

Considering that the terrain delineation of outdoor environments is a relevant task of the auto-determination of traversable areas, we present in this paper a segmentation method that is responsible for the detection of ground and non-ground points of complex scenes under different exposures to illumination, textures, and shading.

The major contributions of this paper can be summarized as follows:

- It provides a trunk detection method.
- It provides a safe path design based on ground segmentation.
- It uses the Signal Detection Theory (SDT) to measure the signal distributions of ground and non-ground.

The paper is organized as follows. The following section, Section II presents related work. Section III, introduces the proposed methods with details of its implementation. This is followed by the methodology, dataset description, and testing measures used for evaluation in Section IV. This leads onto a discussion of the experimental results, which is then finalized by a conclusion and future work in Section V.

II. Related Work

Ali et al. [10] used a fusion of color and texture to segment images into tree trunks and background objects. They evaluated methods of classification, feature descriptors, and prepared a three-step pipeline: training, classification, and

978-1-7281-0319-8/19/$31.00 ©2019 IEEE

segmentation. Yildiz [11] followed a similar approach in which suitable stationary landmarks, i.e. trees, were detected and extracted using training and classification models. In both studies, the authors selected samples of objects from training images and labeled them manually as tree or non-tree.

In a different manner, Mendes et al. [7] used a key-point extractor based on the Sobel operator to define a descriptor in the region around each detected feature. Othmani et al. [12] used a morphological segmentation algorithm to extract the bark of tree trunks to feed a classifier. Likewise, Kim et al. [13] developed a self-supervised learning framework which aimed to improve the traversability estimation capabilities in unknown terrain.

Other studies dealt with the tree detection task without using learning algorithms. Shao et al. [14] presented a method of recognizing tree trunks based on Hough transform for the detecting vertical edges. Lu and Rasmussen [15] proposed a contrast-based method for detecting trees in which bar filters were modeled to emphasize the opposite contrast between trunks and the background. These two works are the closest to our method. The first one uses morphological operators and a segmentation approach that is based on thresholds. The second one assumes that trunks are nearly vertical with high contrast on both sides. In contrast to these two works, our method is prepared to detect more than one trunk per image and to align the tree bottoms without camera calibration.

III. Proposed Method

It first searches for points that contrast with the background and then the initial information on areas that are not terrain are provided; thus, it is used to label non-ground areas and the remaining points are marked as ground areas. Thereby, we look at points that differ from regions of ground and, based on this detection we segment the ground by induction.

The proposed method takes an image, I, in the RGB color space as input in such a way that the channel *green* is used to appraise the pixels in respect to their contrast differences. In this channel, the contrast between dark areas and the other components of a scene is more evident and, because of that, the contrast models may perform better on it than on other channels.

In our study, we considered complex scenes where tree trunks are prominent obstacles. The proposal begins with the detection of trees to thus delineate the ground. The input image, I, is divided into small pieces and each of them goes through contrast models that distinguish objects based on their color and brightness. In the tests, the size of the slices was defined as 20 rows $\times$ d columns, where d was the width of I.

Considering that natural light produces a relevant contrast between the background and foreground of a scene, we used the kernel function proposed by [15] to create models that extract vertical features of trees. Eq. 1 reproduces this contrast model function.

$$x' = y\sin(\theta), y' = y\cos(\theta)$$
$$B(x,y) = \exp(-0.5*(\frac{x'^2}{s_x}+\frac{y'^2}{s_y}))\cos(2\pi f x'), \quad (1)$$

where s_x and s_y are half width and height of the kernel, $x \in [-s_x, s_x], y \in [-s_y, s_y]$. The frequency and orientation of the model filter are set to be $f = 1/(2*s_x)$ and $\theta = 90°$, respectively.

These models are convolved with the input image I, $F = (B*I)$, to produce curves of contrasting areas. Each model filter produces its own curve through the summed-up value of each column of F, Eq. 2.

$$curve(1,j) = \sum_{j=1}^{n}\sum_{i=1}^{m} F(i,j) \quad (2)$$

$curve$ is a vector containing the sum of the values of all rows i of each column j of the F, n is the number of columns and m is the number of rows of F.

The valleys of the curves indicate dark areas that contrast with a bright background. The width of these valleys is measured and based on them some bounding boxes, or patches, are placed in the slice of the image I to select only those areas. Besides, a threshold, $0 < \alpha \leq 1$, is applied to accept only valleys at a specific depth. If the α is set to a higher value, then it accepts more curves. In the tests, α was set with values $0,1$, 0.3, 0.4, or 0.6.

As the best model filter B is unknown, it is necessary to explore different filter sizes. Consequently, each one of them forms their own bounding boxes and to select the most appropriate for an area we used the Greedy Non-Maximum Suppression (GreedyNMS) [16]. In the tests, we used models with size 8, 14, 22, 44.

To reduce the influence of brighter points inside the patches, a local evaluation is applied to discard these points. These patches are converted to the HSV (Hue, Saturation, Value) color space and the channel *value* v is used to detect points with high values, (Eq. 3).

$$\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i \qquad \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(v_i - \overline{v})^2} \quad (3)$$

where n is the number of pixels in a patch and i is the index of each one of them. Hence, v passes through a threshold evaluation such that

$$V(i) = \begin{cases} 1 & \text{if } v_i < (\overline{v} + 3\sigma), \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Eq. 4 is a logical binarization that is based on a condition which detect points further from the mean, $\overline{v}$, in 3 times the standard deviation, σ. Bright points inside a patch are labeled with 0 and other points are labeled with 1. Based on this result, points that were labeled with 0 are removed from the original patch (RGB image), and the others are maintained. After these steps, a partial segmented image, I_s, is obtained.

Next, we calculate the average value for each segmented region in I_s. Points which are larger than the mean value, as well as points which are not connected with the top of the image, are discarded, thus the remaining regions go through some morphological operations to close small holes and then they are used to define the bounding boxes around the trees. Based on these steps, the points inside the bounding boxes in I are parts of the detected trees, and we can conjecture that the ground is the other points of I, i.e., the points which are outside the bounding boxes.

IV. Experimental Results and Discussion

We prepared a dataset to evaluate the proposed method, 20 images were taken from outdoor environments in areas with the presence of trees. The camera was positioned near the ground to simulate a robot that was "walking" through the forest, and also, some photos were obtained from the Internet to evaluate the generalization of the proposed method. This dataset was manually segmented to produce ground truth images; then the results could be compared with the reference images in the dataset, which is available on-line[1].

Fig. 1 shows a sample of the dataset. The first line displays the original image and the second line displays its respective segmented image (in binary format). The third line shows the tree regions, and the fourth line presents the ground area.

Fig. 1: A sample of the dataset. From top-down: the input image I; followed by its respective segmented image; the tree regions; and the ground area.

That binary image was used to measure the amount of data that was correctly identified concerning tree and ground segmentation. Thus, the problem of calculating hits and fails of the proposed method could be treated as collections of objects. Firstly, we computed the number of trees in each scene; and after that, a binary classification test was performed in which the samples were labeled as ground or non-ground.

The proposed method detected 113 trees out of a total of 133 in the dataset (as shown in Table I). It means that 84.96% of possible tree regions were detected. We looked at these results and we noted that only 7 trees were wrongly labeled as trees, which shows that 93.80% of the results were correctly applied. The errors arose due to shadows or non-tree objects that created dark artifacts in the scenes, and some trees were not detected due to the influence of light in the tree trunk or because they were away from the camera.

[1]The dataset is available on https://github.com/vicom-ifgo-urutai/datasets/blob/master/dataset_002.rar

TABLE I: Tree detection.

Number of Trees in the dataset	Detected trees	Hits	Failures
133	113	106	7

The classification assigned as correct or incorrect was defined as:

- True Positive Rate (TPR): the rate of ground pixels correctly identified as ground,
- True Negative Rate (TNR): the rate of non-ground pixels correctly identified as non-ground,
- False Positive Rate (FPR): the rate of non-ground pixels incorrectly identified as ground,
- False Negative Rate (FNR): the rate of ground pixels incorrectly identified as non-ground,

TABLE II: Confusion Matrix.

		Actual	
		Ground	Non-Ground
Predicted	Ground	92	23
	Non-Ground	7	76

Table II presents the results obtained when these two classes are considered in each statistical rate. TPR achieved an assertiveness result equal to 92% and TNR reported 76% in the detection of non-ground points. FPR presented a considerable number of false positives, 23%, and the FNR reported few errors in data classification, only 7%.

Based on these statistical measures, we referred to Signal Detection Theory (SDT) to calculate d' and its correlated metrics. d' is useful to measure how the signal distributions are distinguishable, in our case, the signals: ground and non-ground. Values larger than 0 for d' indicate a greater ability to distinguish both signals [17]. In addition, we calculated $A_{d'}$ and A' to estimate the receiver operating characteristic (ROC) area; β and c to measure the response bias; and β'', a nonparametric measure of response bias.

Table III presents the SDT measures. The function ϕ^{-1} converts probabilities, TPR and FNR, into z scores. d' shows that the distance between the means is more than twice as large as the standard deviations of the two distributions. $A_{d'}$ and A' with results close to 1 indicate an impressive performance. The other metrics, $\beta = 1$, $c \approx 0$, and $\beta'' \approx 0$, show that neither response was favored.

TABLE III: SDT Measures.

$\phi^{-1}(TPR)$	$\phi^{-1}(FNR)$	d'	$A_{d'}$	A'	β	c	β''
1.48	-1.48	2.96	0.98	0.96	1.00	≈ 0.00	≈ 0.00

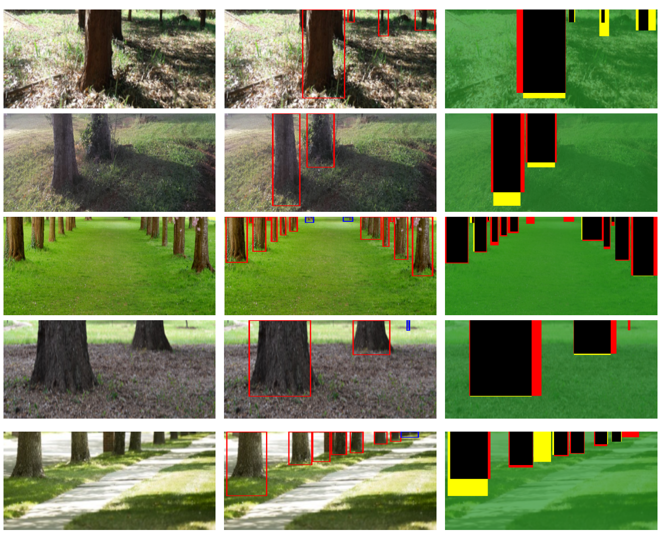
We also calculated the overlap score between the detected ground regions and the ground truth, $overlap(A, B) = n(A \cap B)^2/(n(A)n(B))$, [15]. Considering the average of the dataset the $overlap$ was 0.84 which is a fine result.

Table IV presents the results of the proposed method by taking into account some quality measures: Structural Similarity Index (SSIM), Complex-Wavelet Structural SIMilarity (CW-SSIM), Normalized Cross-Correlation (NCC), and Peak Signal to Noise Ratio (PSNR). The results show that the proposed method can effectively maintain the structure of the image by adding only a few noises in the segmented ground.

TABLE IV: Quality Measures.

SSIM	CW-SSIM	NCC	PSNR
0.77	0.80	0.93	18.6

Fig. 2 shows the visual results of the proposed method. The input images are presented in the first column. In the second column, the detected trees (which are shown in red boxes), as well as the undetected trees (in blue bounding boxes), are displayed. Finally, the third column shows the segmented ground area. Looking at the outputs, we can notice that the ground is highly preserved and tree trunks have been substantially removed to keep only the ground points. Therefore, this method can be used as an unsupervised method to select features and also to define safe paths.



(a) (b) (c)

Fig. 2: Results from the proposed method. The input images I are shown in column (a). The segments of trees and ground are shown in columns (b) and (c). Besides, column (c) shows the detected ground areas (green), the false positive (red), and false negative (yellow).

V. Conclusion and Future Work

Ground detection is a critical task in mobile robotics, as well as the detection of salient obstacles, such as trees. In this study, we proposed a method to identify tree trunks and detach them from the ground using contrast templates, statistical measures, and morphological operator. We applied this proposed method in outdoor scenes which are often complex due to illumination issues as photometric distortion and noise. Experiments showed satisfactory results in which trees were well detected and the ground was segmented with robustness to present safe paths of navigation. In addition, we referred to Signal Detection Theory to discriminate and interpret the classification probabilities. As a next step, we want to use this segmentation method to extract features from outdoor images and use them in machine learning algorithms to increase the robustness of the proposal.

References

[1] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 65–65.

[2] N. Soquet, D. Aubert, and N. Hautiere, "Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation," in *IEEE Intelligent Vehicles Symposium, Istanbul, Turkey*, 2007, pp. 160–165.

[3] G. De Cubber, D. Doroftei, L. Nalpantidis, G. C. Sirakoulis, and A. Gasteratos, "Stereo-based terrain traversability analysis for robot navigation," in *IARP/EURON Workshop on Robotics for Risky Interventions and Environmental Surveillance, Brussels, Belgium*, 2009.

[4] M. C. Le, S. L. Phung, and A. Bouzerdoum, "Lane detection in unstructured environments for autonomous navigation systems," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 414–429.

[5] M. Valente and B. Stanciulescu, "Real-time method for general road segmentation," in *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE, 2017, pp. 443–447.

[6] K. B. Ozutemiz, A. Hacinecipoglu, A. B. Koku, and E. I. Konukseven, "Adaptive unstructured road detection using close range stereo vision," in *Control Conference (ASCC), 2013 9th Asian*. IEEE, 2013, pp. 1–6.

[7] J. Mendes, F. N. dos Santos, N. Ferraz, P. Couto, and R. Morais, "Vine trunk detector for a reliable robot localization system," in *2016 International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 2016, pp. 1–6.

[8] C. Cabo, C. Ordóñez, C. A. López-Sánchez, and J. Armesto, "Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 69, pp. 164–174, 2018.

[9] T.-H. Liu, R. Ehsani, A. Toudeshki, X.-J. Zou, and H.-J. Wang, "Detection of citrus fruit and tree trunks in natural environments using a multi-elliptical boundary model," *Computers in Industry*, vol. 99, pp. 9–16, 2018.

[10] W. Ali, F. Georgsson, and T. Hellstrom, "Visual tree detection for autonomous navigation in forest environment," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 560–565.

[11] T. Yıldız, "Detection of tree trunks as visual landmarks in outdoor environments," Ph.D. dissertation, bilkent university, 2010.

[12] A. A. Othmani, C. Jiang, N. Lomenie, J.-M. Favreau, A. Piboule, and L. F. C. L. Y. Voon, "A novel computer-aided tree species identification method based on burst wind segmentation of 3d bark textures," *Machine Vision and Applications*, vol. 27, no. 5, pp. 751–766, Jul 2016. [Online]. Available: https://doi.org/10.1007/s00138-015-0738-2

[13] D. Kim, S. M. Oh, and J. M. Rehg, "Traversability classification for ugv navigation: a comparison of patch and superpixel representations," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2007, pp. 3166–3173.

[14] L. Shao, X. Chen, B. Milne, and P. Guo, "A novel tree trunk recognition approach for forestry harvesting robot," in *Industrial Electronics and Applications (ICIEA), 2014 IEEE 9th Conference on*. IEEE, 2014, pp. 862–866.

[15] Y. Lu and C. Rasmussen, "Tree trunk detection using contrast templates," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1253–1256.

[16] J. H. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression." in *CVPR*, 2017, pp. 6469–6477.

[17] H. Stanislaw and N. Todorov, "Calculation of signal detection theory measures," *Behavior research methods, instruments, & computers*, vol. 31, no. 1, pp. 137–149, 1999.