Multi-sensor-based online positive learning for drivable region detection

H. Lee, K. Kwak and S. Jo

A new method for detecting drivable regions in an unrehearsed and unstructured outdoor environment using multi-sensor information is presented. To achieve this goal, two key methods are developed: (i) robust and effective feature definition using colour and geometry and (ii) online learning algorithm using positive samples for detecting drivable regions. With real data sets, the effect of sensor modality is evaluated and is compared the performance of the algorithm to a cluster-based approach.

Introduction: Navigation in a complex outdoor environment that includes foliage, surfaces covered with leaves and dense vegetation is a difficult and challenging problem for autonomous robots. The primary concern of much of the previous research has been the identification of the kind of paved roads such as country roads and urban roads. Such roads are typically characterised by their road colour, lane markings or end boundary features. However, if a robot is placed in an unknown and unstructured outdoor environment without any prior information and goes to a goal location without human intervention, the robot should be able to identify drivable regions online.

In this Letter, we propose a new drivable region detection method in an unrehearsed and unstructured outdoor environment using a set of features derived from three-dimensional (3D) light detection and ranging (LIDAR) and a camera. The approach uses online positive learning to identify whether unknown regions in front of the vehicle are drivable or not while in motion and without prior manual labelling. Our contribution in this Letter is twofold. First, we define the robust and effective visual and geometric features used for drivable region detection in an unstructured environment. Secondly, we develop an online positive learning algorithm to identify a drivable region in an unknown environment without a pre-computed model. Our approach achieves robustness by combining multi-sensor information from a 3D-LIDAR and a camera. The extrinsic parameters between both sensors were estimated by Kwak *et al.* [1].

Features selection: The feature set used for our drivable region detection consists of an image and the corresponding geometry information derived from a camera and a 3D-LIDAR. We obtain superpixels from the over-segmentation of the image and the corresponding 3D-LIDAR points. Each feature set is defined as a histogram $H = [h_1, \ldots, h_B]$ with B bins. The histogram represents a normalised distribution of colour and geometric information in a superpixel.

For an image feature, we use CIE-LAB colour space [2] which is known to be more perceptually uniform and robust against outdoor lighting condition than the RGB (red, blue, green) and HSV (hue, saturation, value) colour spaces. To compensate for errors from the sensor calibration and to correctly estimate the boundaries of complex objects, we segment the image into superpixels. Superpixels are the result of a perceptual grouping of pixels and align better with image edges than rectangular image patches [3]. The image feature is measured by the colour histogram of a superpixel. Let h_b^c be the b index of the colour histogram value in superpixel I. It is defined as

$$h_b^c = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(y_i^c \in \mathrm{bin}(b)\right) \tag{1}$$

where y_i^c is a pixel in I, N is the total number of pixels and and $\mathbb{I}$ is the indicator function. This normalised colour histogram H^c gives the probability that the colour of the superpixel is in quantised colour bin(b).

The LIDAR features are defined as four normalised geometric distributions considering the shape, size and distance of LIDAR points in a superpixel. Among the four distributions, three of them represent the 3D geometrical shape of the region such as eigenvectors [4]. However, we focus on the distribution of points relative to the eigenvectors of the covariance matrix of 3D points. Thus it is invariant to translation and rotation. Let f_j^k denote a shape feature for the geometry histogram where $k = \{1, 2, 3\}$. It is defined as

$$f_j^k = 1 - \frac{1}{\pi}\arccos \frac{\boldsymbol{e}_k\left(y_j^g - \mu_Y^g\right)}{\|\boldsymbol{e}_k\| \left\| y_j^g - \mu_Y^g\right\|}, \quad k = \{1, 2, 3\} \tag{2}$$

where μ_Y^g is the centre of point sets $\boldsymbol{Y}^g = \{y_1^g, \ldots, y_M^g\}$, $y_j^g = \{x_j, y_j, z_j\}$ and $\boldsymbol{e}_k$ is the eigenvector. From this equation, we characterise the distribution of the 3D points in each superpixel with regard to shape. The feature f_j^k,where $k = \{4\}$, represents the size and scaling characteristics. The size and scaling factors are captured by the product of eigenvalues and distance from the sensor. The size and scaling feature is defined as

$$f_j^k = \left(\frac{\lambda_1\lambda_2\lambda_3 \times \|\mu_Y^g\|}{V \times D}\right)^2, \quad k = \{4\} \tag{3}$$

where λ_1, λ_2 and λ_3 are eigenvalues corresponding, respectively, to eigenvectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$, D is the maximum distance in the scene and V is the volume of the largest superpixel.

Since the four features are normalised with their value between [0, 1], we use it as a variable for a histogram. Similar to the colour histogram, h_b^g is the geometry histogram value and it is defined as

$$h_b^g = \frac{1}{4M}\sum_{j=1}^{M}\sum_{k=1}^{4} \mathbb{I}\left(f_j^k \in \mathrm{bin}(b)\right) \tag{4}$$

where M is the total number of 3D points in the superpixel. Given a normalised geometry histogram H^g for any superpixel I, it represents the statistical characteristics of 3D points with respect to shape, size and scale of the superpixel region from the LIDAR sensor.

Online positive learning: We use only positive samples to detect a drivable region in an unrehearsed and unstructured environment. Much of the standard research poses the drivable region detection as a binary classification problem. However, it is difficult to define negative samples in an unknown and unstructured outdoor environment. For example, if we trained dense vegetation samples as a negative, the vehicle may not detect the drivable region in an environment consisting of dense vegetation. We develop a drivable region detection approach using only positive samples. Our sample labelling system is inspired by Wellington and Stentz [5].

The classifier decides that a new region is drivable by comparing the region to previous drivable regions; if the similarity factor is greater than a parameter θ, then it is a drivable region. The threshold-based approach is a computationally efficient and simple prediction method for online region classification. The classifier is defined as

$$g(H_i, H') = \mathbb{I}\left[\frac{\sum_{s=1}^{S} I\left(D(H_i, H'_s) \geq \theta'\right)}{S} \geq \theta\right] \tag{5}$$

where H_i is the target feature histogram to predict, $H' = \{H'_1, \ldots, H'_S\}$ is the set of labelled feature histograms, $D(H, H')$ is the distance function to compare the similarity, θ' is the threshold for distance function and S is the number of histograms in H'. We define the distance function as χ^2 which is powerful for comparing the similarity of the histogram:

$$D(H, H') = \sum_{b=1}^{B}\left[w^c \frac{(H^c(b) - H'^c(b))^2}{H^c(b) + H'^c(b)} + w^g \frac{(H^g(b) - H'^g(b))^2}{H^g(b) + H'^g(b)}\right] \tag{6}$$

where w^c and w^g are weighted factors of each histogram distance.

Experimental results: Using the framework described above, we performed a series of experiments to evaluate the effectiveness of our proposed algorithm. The experiments include: (i) a comparison of our algorithm using LIDAR features, image features and a combination of LIDAR and image features and (ii) a comparison of our algorithm with a state-of-the-art method. We divide the data set used in our experiments into two classes: (i) an easy data set consisting of a surface covered with leaves and vegetation under 15 cm high and (ii) a challenging data set including foliage and dense vegetation of more than 40 cm high.

Fig. 1 provides the threshold averaging receiver operator characteristic (ROC) curves with respect to the two different data sets. As can be seen in both cases, the colour feature is more discriminative than the LIDAR-based features. In the challenging case, the LIDAR-based features are more affected by the shape variation of the dense vegetation than the case of the light terrain surface set. However, the combination of LIDAR and image-based features performs better than features from either modality individually. This result suggests that the information in the LIDAR and image features produced by our method is not entirely redundant.

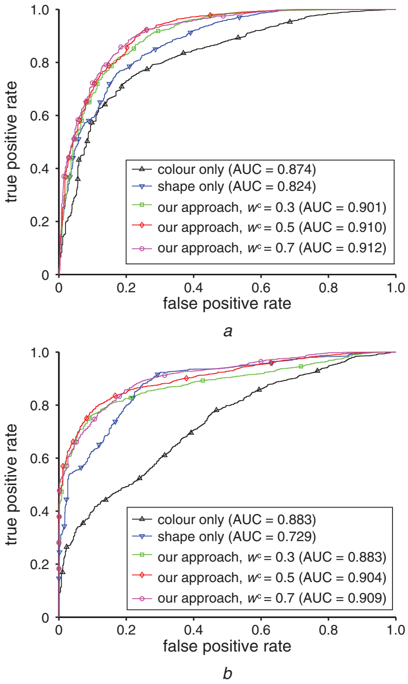

Fig. 1 *Effect of sensor modality*

a Easy sample case
b Challenging sample case

Fig. 2 shows the comparison results with the state-of-the-art method. Each image is selected from the easy and challenging data sets.

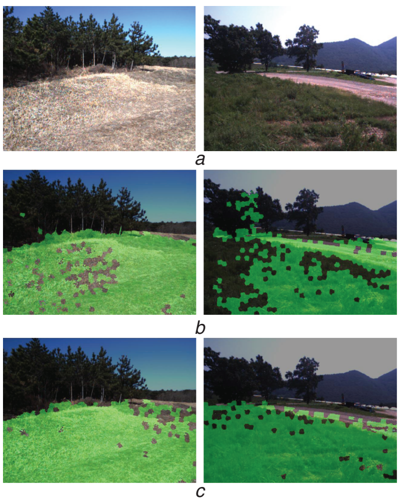

Fig. 2 *Estimated drivable regions (left: easy case, and right: challenging case)*

a Original image
b Result of cluster-based approach [6] ($k = 5$)
c Result of our approach ($w^c = 0.7$)

From our experiment, the algorithm of [6] is very strong with respect to assigning similar colour image patches into the same cluster. However, it is less accurate than our approach for geometry features.

Fig. 3 describes the performance comparison of the existing method and our approach. The results, in both ROC curves, show that our algorithm outperforms the existing approach for all cluster values. It has an improvement of nearly 5–7% in accuracy. In this experiment, the performance of the existing algorithm is affected by the data set significantly, but our algorithm shows a consistent pattern.

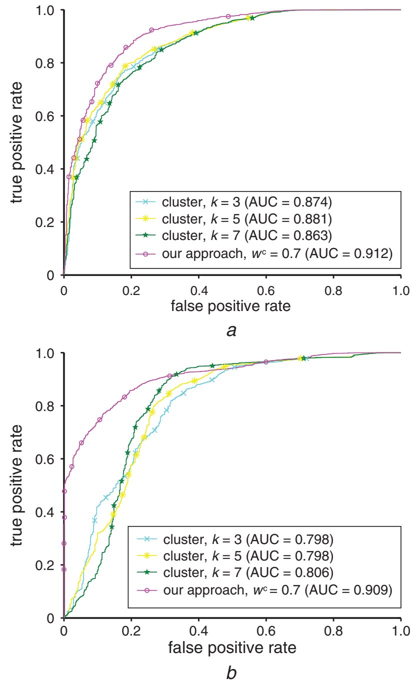

Fig. 3 *Performance comparison with existing approach*

a Easy sample case
b Challenging sample case

Conclusion: We have presented a novel method for detecting drivable regions in an unrehearsed and unstructured outdoor environment. To achieve this goal, we first defined robust and effective features as histograms that describe the colour and geometric characteristics of each superpixel region. Secondly, we developed an online positive learning algorithm. The approach is achieved by tightly combining information from a 3D-LIDAR and a camera. Through experimentation, we proved the performance of our method.

Acknowledgment: This project was funded by the Agency for Defense Development (ADD) under grant UD110111ID.

29 May 2014
doi: 10.1049/el.2014.1302
One or more of the Figures in this Letter are available in colour online.

H. Lee and S. Jo (*Department of Computer Science, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea*)

E-mail: shjo@kaist.ac.kr

K. Kwak (*Agency for Defense Development, Bukyuseongdaero 488gil, Yuseong, Daejeon 305-152, Republic of Korea*)

References

1 Kwak, K., Huber, D., Badino, H., and Kanade, T.: 'Extrinsic calibration of a single line scanning LIDAR and a camera'. IEEE Conf. on Intelligent Robots and Systems, San Francisco, CA, USA, September 2011, pp. 3283–3289, doi: 10.1109/IROS.2011.6094490
2 Commision Internationale de l'Eclairage (CIE): 'Recommendations on uniform color spaces, color difference equations and psychometric color terms', *CIE Publ.*, 1978, (15), pp. 9–12
3 Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S.: 'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2274–2282, doi: 10.1109/TPAMI.2012.120
4 Vandapel, N., Huber, D., Kapuria, A., and Hebert, M.: 'Natural terrain classification using 3-d LIDAR data'. IEEE Conf. on Robotics and Automation, New Orleans, LA, USA, April 2004, 5, pp. 5117–5122, doi: 10.1109/ROBOT.2004.1302529
5 Wellington, C., and Stentz, A.: 'Online adaptive rough-terrain navigation vegetation'. IEEE Conf. on Robotics and Automation, May 2004, Vol. 1, pp. 96–101
6 Ott, L., and Ramos, F.: 'Unsupervised incremental learning for long-term autonomy'. IEEE Conf. on Robotics and Automation, Saint Paul, MN, May 2012, pp. 4022–4029, doi: 10.1109/ICRA.2012.6224605