Ground Traversability Estimation using Computer Vision

Suzanne Ong
University of Adelaide
Australia
suzanne.ong@student.adelaide.edu.au

Dr Feras Dayoub
University of Adelaide
Australia
feras.dayoub@adelaide.edu.au

ABSTRACT

Ground traversability estimation is an important ability for robotic appliances, particularly mobile ground robots, to predict the path that is traversable. It involves the analysis of the area traversability on the ground, which can be learned from the images classified via computer vision. In this paper, we review the effect of computer vision approaches available in ground transversability estimation, and compare them based on the compute budget generated. It aims to evaluate the difference in runtime and average rate of performance within the selected range of newly-developed frameworks while implementing in a simulation of real-world elevation dataset.

1 INTRODUCTION

Ground traversability estimation is the ability of autonomous, mobile appliances to articulate a traversable path in order to reach a designated target position. Depending on the elevation, slope and obstacles on the ground, it relies on computer vision approaches, such as image classification or object detection, to partition areas that are either traversable and non-transversable (i.e. when faced obstacles like walls or static objects) for the appliances to move accordingly [1]. Traditionally, it is done by utilising sensors such as LIDAR to detect and identify the areas.

However, the complexity increases when it comes to uneven ground scenarios, because there are more constraints to consider other than the partitioning of areas based on images. For instance, given an offroad robot with a limited level of operational power, it may not be able to ascend upslope even though it is considered to be the shortest path based on the plan generated by algorithm used, hence rendering it infeasible to compute. This problem is also dependent on the centre of mass of the robots, as robots with a disproportionately great height but smaller surface area may result in the , despite having sufficient power to either ascend or descend a slope. External factors such as the ground clearance, robot orientation, zero moment point distance, force-angle stability measure, traction efficiency and distance stability measure may impact the robot optimality in traversing to an estimated path based on the image generated above ground level [11].

Recent development in computer vision techniques, such as object detection, semantic segmentation and image classification, has provided a research gap in exploring more optimised ways to better articulate the estimation of transervability based on the given imagery datasets. Through algorithms that categorized traversable and non-transversable areas based on the recorded footage and outliers, there have been numerous attempts to reduce the compute budget of the process in studies over the last five years [1,5,6].

In this project, the problem statement can be stated as follows: Given a generated image representing the heightmap of a ground patch, identify the best approach in computer vision that allows the robot to traverse from right to left with the least computational demand. This is suggested with consideration that low compute budget is a strong indication of an efficient algorithm, which is a favourable outcome when it is used in large datasets and limited computational resources like time, cost and power [3,5,9]. Moreover, because each approach is built differently in terms of its structure and approach towards each pixel of the images, results may vary in computational complexity depending on the robotic appliances used and diverse labeling methods on images from the dataset [6]. Hence, to observe the optimal approach from current findings, we reviewed three publicly retrievable, state-of-the-art computer vision frameworks to determine the technique that yields the lowest compute budget in ground transversability estimation. Based on the benchmark datasets from Sensefly, the configurations would be calibrated according to the implementation of images taken using the Robotic Operating System (ROS), particularly in off-road robotic applications, the wall clock time and operational power used at a constant set of imagery dataset to address the problems faced in ground traversability problems. Added with the lack of standardized metric and comparison study to analyse the compute budget for these frameworks, it offers a research gap for us to estimate the ground traversability for both simulations and real-world robotic appliances [1,4,17]. Hence, for this paper, we would be focusing on finalising two concepts, which are, identifying the computer vision approach that uses the least computational resources like the runtime, wall clock time, average rate of performance, and the storage space complexity of the computer vision approach, and secondly, obtain the difference of compute budget used for newly developed approaches.

2 MOTIVATION

The motivation for this paper is to extend the literature of available computer vision techniques on ground transversability estimation, inspire future research in addressing any identified challenges, and making a comparison study for the community as an additional reference for testing these approaches based on their compute budget generated.

Our research scope would be focusing primarily on testing up-to-date computer vision techniques that are preferably open-sourced, configurable in previous researches and can be implemented in our research environment. Although there is a

boost in research implementation for computer vision techniques in practical applications such as autonomous driving prediction [5, 10], up to date evaluation and analysis on traversable vegetation still holds theoretical significance in deciding the optimal technique in computer vision to use for their respective fields due to the introduction of improved algorithms that helps mitigate the effects of existing limitations like outliers and roughness of the terrain [4]. The need to benchmark these approaches is further emphasised with the recent development of image classification approach suggested in the 2018 conference [1] and scene recognition approach by Matsuzaki that utilised unsupervised domain adaptation for semantic segmentation to navigate through transversable plants to navigate the robot's path based on the geometric information of the environment [9]. The development of new frameworks and usage of enhanced methods in other fields like autonomous driving that are optimized in efficiency, simplicity, and accuracy would be an interesting research aspect that may generate optimal solutions for problems faced in ground traversability estimation.

Moreover, the idea of comparing different computer vision methods from published frameworks, especially in the overall computation demand, would provide us with a better insight into the runtime, quality and performance of the algorithms used in each framework. For instance, assuming that we have an imagery dataset of 2.6GB of images and environment footage, will there be any impact on the runtime performance and overall efficiency? Another extension of this idea would be the effect of qualitative parameters on the quality or prediction behaviour of the algorithm, which would be useful in identifying the constraints of the selected approach on the estimation problem.

3 LITERATURE REVIEW

A review of research papers was conducted to explore the recent development of computer vision techniques, the improvements and limitations of at least three used methods on estimating ground transversability and identifying any available datasets that allow us to test and benchmark the compute budget of different computer vision approaches.

Computer vision is the construction of systems that process, perceive and evaluate visual data artificially. It is widely used in various fields due to its ability to interpret visual information based on image and video data for further analysis through statistical modelling. Its purpose to translate visual data into insights readable for humans based on contextual inputs provides us with more detailed understanding to make informed decisions in business or solutions for complex real world problems. Through accumulating a training dataset of labelled images, we provide them as input to the computer for data processing. Some of computer vision applications included autonomous driving, allocation of vegetation and network security [20]. In this paper, we identified three computer vision techniques, which are image classification, semantic segmentation and object detection that provided well-documented literature with open-sourced software frameworks to implement, as well as alternative methods that have not been used, but has a feasible potential to be conducted into ground traversability estimation.

Image classification is a computer vision technique that involves the categorizing and labeling of land cover classes into groups of image pixels based on one or multiple desirable characteristics. By casting the problem of estimating ground traversability as a supervised image classification problem on the dataset, we can categorize the traversable areas for the robots to plan out a path with the lowest compute budget. This is backed by Chavez-Gracia's research, which compares the difference in accuracy and area under the Receiver Characteristics Operator (ROC) curve between two image classification techniques. The first option, which is the feature-based approach, extracted descriptive features such as the average terrain steepness for each heightmap patch based on the motion direction of the robot and the maximum height of any steps in the patch using the Histogram of Gradients (HOG). This is followed by the application of random forest classification with 10 trees onto the HOG computed. Alternatively, the second option involved using Convolutional Neural Networks (CNNs), another architecture that utilises the networks to categorize the data when images are inputted. In the CNN-based approach, it is expected that the network autonomously learns meaningful, problem-specific features; because the input shape is high-dimensional and no prior knowledge of the problem is provided to the model, this approach requires more training data. Using software like Keras to build the networks, Adadelta optimizer to reduce the cross-entropy loss through training for 50 epochs and Tensorflow to operate the frontend integration, the researchers were able to implement a connected layer with two output neurons [1]. Although it is being constrained the fact that it was only experimented in V-REP simulator using heightmaps, and its lack of consideration of other factors, such as the compactness, friction and robot dynamics, there is a research gap available for us to implement these approach in robot-centric perceptions to provide a detailed research literature over the implication of image classification approaches onto robotic appliances, and determine the validity of Chavez-Garcia's result in ground transversability estimation based on the underlying algorithm performance. As a result, through evaluating the performance metrics, the research suggested that CNN approach has a better performance than the feature-based approach, which allows additional room for subsequent research to compare both approaches based on the compute budget that is not specified in the paper and determine whether CNN approach is computationally feasible despite its better algorithm performance than the feature-based approach [1].

In addition, another well-known approach in computer vision would be semantic segmentation. Given that pixel-wise predictions from footage can be conducted from models, semantic segmentation is another main approach in distinguishing the images into pixel grouping, which not only detects the land cover classes like image classification, but also classify all image pixels to determine whether if there is an object detected that would hinder the traversability on the ground. Recent studies have exploited the use of both a semantic segmentation branch for object classification and a traversability estimation branch that operated onto the pixels. In plant-rich environments, researchers

managed to train the semantic segmentation branch using an unsupervised domain adaptation method and the traversability estimation branch using label images generated from the robot's traversal experience for the data acquisition phase. This resulted in the robot capability to recognise traversable plants with better accuracy than a conventional semantic segmentation with traversable and non-traversable plant classes [9]. A further extension of this approach would be the addition of Transformers with lightweight multilayer perception (MLP) decoders, which mitigates the complexity in decoders and effect of interpolation of positional codes, both that may lead to decreased algorithm performance while operating on training and testing sets. Experimented in a framework known as SegFormer in 2021 for a dataset compiled from urban street scenes, it yields a 84% increase at best in mean intersection of union (mIoU), thus suggesting that optimization using transformers improves the efficiency and compute budget of the approach [18].

Recent improvement in semi-supervised frameworks, such as GoNet and its updated version known as VuNet [6,7] , has been experimented and documented for ground transversability estimation. By enhancing deep generative model, known as Generative Adversarial Network (GAN) that assist in categorizing images taken from a fish-eye camera based on their transversability, GoNet developers provided images that depicted traversable areas as positive examples for robot navigation to train GAN, with a disproportionately low number of non-transversable areas compared to the positive examples. Such discrepancy in the size of both positive and negative examples in data allows GoNet's underlying learning algorithm to exceed both supervised and unsupervised baselines in terms of its practical robustness [6]. Demonstrated by capturing images from fisheye cameras attached in a robot, it significantly improves the accuracy of robot navigation through predicting risky paths, such as collisions with an immobile object like a wall based on the negative examples learned, and minimizes the probability of robots being damaged due to collisions [6]. Hence, it is suggested that GoNet framework offers the opportunity to be improvised and used in a communal setting for healthcare systems, network security and the establishment of a system to issue warnings when an obstacle is detected for the refinement in ground transversability estimation for robot navigation. The experimental method of GoNet's research also presents an opportunity to compare its overall accuracy in compute budget on outdoor door settings, as it was only limited to indoor campus environments where the impact of external obstacles like plants and vegetation on ground is negligible. This approach is further improvised with the update of VuNet framework, which handles scene view synthesis problems by making predictions on future images using RGB images for static and dynamic environments [7].

Besides these three known computer vision frameworks, there are two notable mentions in computer vision that may be useful in resolving the ground traversability estimation problem, which are object detection. Object detection algorithms, such as Faster R-CNN developed in 2015, provided a well-tested, open source improvement on CNN approach, which allows a separate network to predict the region proposals and followed by predicting the offset values through reshaping of the pooling layer. This significantly improves the traditional selective search approach, and unlike the YOLO, or You Only Look Once algorithm [9,13] , since it uses regions to localize the object, it is not constrained by small obstacles detected in the image. Through leveraging statistical techniques like cluster analysis [3] , it is able to estimate the ground transversability in accordance with the class probabilities generated with greater efficiency.

4 PROJECT GOAL AND CHALLENGES

For the project goal to be facilitated, this research aims to compare the selected frameworks in accordance to their compute budget when assigned onto an image dataset through ROS and LeggedRobotics for traversability estimation, either through simulations or actual robotic appliances like LIDAR sensors as suggested by Eriksson's findings [3], as well as holding a high level of relevance for the wider research community to encourage the use of computer vision techniques. The project should be deemed complete once the following procedures below are accomplished in the given timeframe:

1. Install the required software packages for framework compilation and download the required dataset;
2. Build a prototype of different computer vision approaches that are publicly available or established in frameworks and ensure that it is successfully compiled using PyTorch and assigned programming languages like Python and C++ ;
3. Measure and compare the computational demand for the selected methods using benchmark datasets used in previous experiments to check for its feasibility and accuracy;
4. Evaluate the compute budget based on the CPU runtime and the average rate of performance in Hz. Other performance metrics such as mean squared error and mIoU may be considered as a stretch ;
5. Record and present the overall finding, success or unsuccessful result in report, codebase and presentation.

The initial scope will be focused on implementing known computer vision approaches for ground and terrain transversability estimation, which are the feature-based and CNN approaches from Chavez-Garcia's image classification framework, semantic segmentation using SegFormer, and the VuNet approach. If time permits, additional approaches such as object detection algorithms like Fast R-CNN will be considered as well in the research as a stretch goal. Unless alternative datasets are suggested otherwise, a Sensefly case-study dataset from a drone excavation in outdoor Schwarzenbach, which is located approximately 60 kilometres south of Vienna may be tested for its validity of the selected computer vision approaches with other research. This is suggested in this project for documentation purposes as shown in Steps 3 and 4 due to its versatility in performance tracking, automated benchmarking on all runtimes

and satisfactory measurement aggregation. For contingency planning in dataset selection, in case the images do not resonate well with either the algorithms or research method, the retrieval of GO Stanford 3 indoor dataset may be applied for its previous usage in robotic appliances as a replacement to the original proposal. ROS is used to ease compatibility and the connection of the given robotic appliances and its specific sensors.

However, this project presents us with three challenges, which are the possibility of negative results, lack of on-hands experience in ROS or robotic appliances and the vast size of the dataset, that may hinder progress. As we have yet to conduct much experiment or usage on the proposed frameworks, there may be indications that a negative result may be obtained regardless of the techniques' performance. The lack of insights of the ROS software and large imagery dataset may also require additional learning time and storage. Therefore, it is necessary to implement contingency planning for all challenges faced, which is known to be a good practice for any project design. By implementation, in a scenario where a negative result is obtained, it should not be disregarded and should be presented as a valid conclusion that provides room for additional research in the future. The practice of cherry-picking results in the project should not be practised as well to maintain the level of accuracy and consistency of the project. A buffer of at least one week in the timeline for adequate time allocation, and an external hard-disk plug-in is used as an alternative source for storage to address the challenges.

5 TIMELINE

The expected completion of the project is defined by the successful construction of the final report and compilable algorithms stored in a GitHub repository by Week 12, added with a finalised presentation for the general audience in Week 13.

During the short weeks before the semester, and in both week 1 and 2, we structured the initial layout based on the conference and rubric requirements, with additional revision on the available computer vision techniques used and conducted thorough research on existing papers. A variety of benchmarked datasets available are reviewed, as well as a quick glance through over the implications of the software package Robotic Operating System (ROS) over off-road robotic applications. Added with the provision of time for the construction of the design document draft, Week 3 provides time to address additional feedback and evaluation over the feasibility of the project.

With these preparations onhand, week 4 and 5 primarily focused on constructing the computer vision approaches and training the models used in code using C++, Python, PyTorch and potentially the Jupyter Notebook framework from Anaconda. A buffer period in Week 6 will be scheduled to update the progress into the final report, address the lesson learned or arising problems throughout, and discuss the need to implement additional approaches or evaluation metric depending on the progress. We also aim to evaluate the results obtained, record any findings or additional resources, and finally review the feasibility of implementing any stretch goals identified between Week 7 and Week 11. Ideally, the project aims to be finalised with a few rounds of feedback on the report and presentation by the end of week 11 and 12.

REFERENCES

[1] Chavez-Garcia, RO, Guzzi, J, Gambardella, LM & Giusti, A 2017, 'Image classification for ground traversability estimation in robotics', in International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, pp. 325-336.

[2] Chavez-Garcia, RO, Guzzi, J, Gambardella, LM & Giusti, A 2018, 'Learning ground traversability from simulations', IEEE Robotics and Automation letters, vol. 3, no. 3, pp. 1695-1702.

[3] Eriksson, D & Harström, J 2019, 'Object detection by cluster analysis on 3D-points from a LiDAR sensor'.

[4] Guan, T, He, Z, Manocha, D & Zhang, L 2021, 'TTM: Terrain traversability mapping for autonomous excavator navigation in unstructured environments', arXiv preprint arXiv:2109.06250.

[5] Guastella, DC & Muscato, G 2020, 'Learning-based methods of perception and navigation for ground vehicles in unstructured environments: A review', Sensors, vol. 21, no. 1, p. 73.

[6] Hirose, N, Sadeghian, A, Vázquez, M, Goebel, P & Savarese, S 2018, 'Gonet: A semi-supervised deep learning approach for traversability estimation', in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, pp. 3044-3051.

[7] Hirose, N, Sadeghian, A, Xia, F, Martín-Martín, R & Savarese, S 2019, 'Vunet: Dynamic scene view synthesis for traversability estimation using an rgb camera', IEEE Robotics and Automation letters, vol. 4, no. 2, pp. 2062-2069.

[8] Kiran, PSR, Kumar, A & Mohan, R 2019, 'Aerial-Ground Robotic system for Terrain estimation and Navigation', in 2019 Fifth Indian Control Conference (ICC), IEEE, pp. 101-106.

[9] Matsuzaki, S, Masuzawa, H & Miura, J 2022, 'Image-based scene recognition for robot navigation considering traversable plants and its manual annotation-free training', IEEE Access.

[10] Onozuka, Y, Matsumi, R & Shino, M 2021, 'Weakly-supervised recommended traversable area segmentation using automatically labeled images for autonomous driving in a pedestrian environment with no edges', Sensors, vol. 21, no. 2, p. 437.

[11] Papadakis, P 2013, 'Terrain traversability analysis methods for unmanned ground vehicles: A survey', Engineering Applications of Artificial Intelligence, vol. 26, no. 4, pp. 1373-1385.

[12] Prágr, M, Čížek, P & Faigl, J 2018, 'Incremental learning of traversability cost for aerial reconnaissance support to ground units', in International Conference on Modelling and Simulation for Autonomous Systems, Springer, pp. 412-421.

[13] Ren, S, He, K, Girshick, R & Sun, J 2015, 'Faster r-cnn: Towards real-time object detection with region proposal networks', Advances in neural information processing systems, vol. 28.

[14] Ross, PJ 2016, 'Vision-based traversability estimation in field environments', Queensland University of Technology.

[15] Sevastopoulos, C & Konstantopoulos, S 2021, 'A Simulated Environment for Traversability Estimation Experiments in Field Robotics Applications', in The 14th PErvasive Technologies Related to Assistive Environments Conference, pp. 256-257.

[16] Shan, T, Wang, J, Englot, B & Doherty, K 2018, 'Bayesian generalized kernel inference for terrain traversability mapping', in Conference on Robot Learning, PMLR, pp. 829-838.

[17] Sun, P, Kretzschmar, H, Dotiwalla, X, Chouard, A, Patnaik, V, Tsui, P, Guo, J, Zhou, Y, Chai, Y & Caine, B 2020, 'Scalability in perception for autonomous driving: Waymo open dataset', in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2446-2454.

[18] Wu, Z, Shen, C & Van Den Hengel, A 2019, 'Wider or deeper: Revisiting the resnet model for visual recognition', Pattern Recognition, vol. 90, pp. 119-133.

[19] Xie, E, Wang, W, Yu, Z, Anandkumar, A, Alvarez, JM & Luo, P 2021, 'SegFormer: Simple and efficient design for semantic segmentation with transformers', Advances in Neural Information Processing Systems, vol. 34.

[20] Yin, W, Zhang, J, Wang, O, Niklaus, S, Mai, L, Chen, S & Shen, C 2021, 'Learning to recover 3d scene shape from a single image', in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 204-213.

[21] Zhao, J, Masood, R & Seneviratne, S 2021, 'A review of computer vision methods in network security', IEEE Communications Surveys & Tutorials.