A System Proposal for Vegetation Traversal Cost Estimation in Off-Road Autonomous Mobile Robots

John Gettings
UC Riverside Master of Science
June 2019

Abstract

This paper proposes a system architecture to calculate energy costs and resistive forces of vegetation traversal for autonomous off-road robots. An advanced semantic segmentation algorithm and dataset are discussed to first classify vegetation at a lower level than ever before for autonomous mobile robots. For each class, a regression model would be built to calculate costs and forces through experimentation using lidar and stereo vision, and then updated continuously in-field with unsupervised online learning.

1 Introduction

The purpose of this paper is to suggest a more robust method for autonomous robots to handle vegetation in outdoor environments. Currently off-road autonomous systems are very limited in their ability to assign traverability costs to vegetation. The overly-simplistic evaluations currently in place greatly hinders a robot's capabilities and efficiency by restricting it to avoid viable paths with traversable vegetation.

For most current autonomous off-road systems in research, all vegetation is labeled in a simple binary fashion as either 'traversable' or 'non-traversable' [21], [19], [18], [22]. The robot then uses a conservative, brute-force method to plow through a vegetation type that is determined to be traversable or it simply avoids all vegetation. In some instances the robot will identify a few groups such as tall grass, sparse vegetation, dense vegetation, and trees [23], [29], [15]. With each of these groups a single cost will be blanketed over many types of vegetation, or it will be labeled as traversable or non-traversable.

This type of evaluation for vegetation costs not only constricts the autonomous vehicle to paths mostly free of vegetation, but also doesn't allow for evaluation between two paths of different vegetation. For example, if there is a section of semi-tall grass that can be easily traversed through but recognized as non-traversable, a robot may find itself taking a much longer route wasting time and energy. If there is one viable path of semi-tall grass and another path of short shrubbery of the same height, current robots would not be able to give preference to one or the other thus giving the opportunity to get stuck or waste much more energy in the shrubbery.

With the advancement of robotics, artificial intelligence, and sensors, robots are being used for more and more applications in every industry. Currently they do quite well in organized, man made environments and are frequently designed for many autonomous tasks. They are also being designed for outdoor environments such as farming [33], [34], [35], but crops and orchards have a high level of man-made predictability as well. In the future, off-road robots can be used for exploration, search and rescue, surveying, landscaping, plant and animal monitoring, firefighting, border patrol, and military assignments to name only the most obvious applications.

For off-road robots that only require to follow a pre-made path, the current classification and evaluation scheme can suffice. For robots that will be used for true off-road traveling and exploring, binary classification and cost assignment would not be sufficient. In reality, robots in populated urban and rural environments will also have a difficult time in yards, parks, farms, and overgrown areas with the current vegetation modeling technology.

The development of robots that can travel through unstructured outdoor environments is an inevitable progression. Our dependency on autonomous systems will only increase with time in both industry and home life. As autonomous travel on clearly defined surfaces and geometric surroundings is coming close to mainstream use, and travel on outdoor but defined roads is making a fair amount headway, efficient off-road travel in vegetated environments is an early area of research in terms of practical applications. As more options in traveling, transporting, exploring, and surveying wilderness becomes available, more uses for the technology will be discovered.

This paper proposes a method for creating such a system which is able to automatically segment an outdoor scene and assign costs to each vegetation instance grouping. First, the scene is segmented into hierarchical classes from a large dataset which has yet to be created. Then a regression algorithm will be assigned to each class which will receive input features such as height, width, density, time of year, texture features, etc. to be verified through experimentation. The vegetation cost and class will then be mapped to a 3D voxel structure using Lidar and stereo imaging data. Once vegetation is traversed through, actual costs will be measured through proprioceptive feedback and current measuring sensors. The regression algorithm will update in real time through online learning.

The caveat to all this, of course, is that more accurate models require more testing, training, and data collection. Due to the stochastic nature of vegetation, models will likely require even more training data than that of man made structures and objects for the same accuracy of classification. On the other hand, off-road vehicles will not require as high of an accuracy or as quick of a response time due to slower speeds and less risk of harming pedestrians. Although the process will take time, it is an inevitable step in advancing autonomous mobile robots for outdoor environments.

2 Related Work

2.1 Traversability metrics

As mentioned above, most off-road robotics research papers only mention the ability to detect traversable or non-traversable locations. Some are only concerned with detecting man-made paths in outdoor environments and labels all or most vegetation as “non-traversable” [24], [25], [26], [27]. In others, there are only two groups which are “traversable” for roads, sparse vegetation, and short vegetation, or “non-traversable” for everything else [21], [28], [22], [19], [18], [17]. In some instances vegetation is segmented into a limited set of approximately three to five groups, which only give an overarching cost to vegetation type, or a simple binary classification for traversability [12], [13], [15], [23], [29]. To my knowledge there is very limited research on estimating actual forces or energy consumption from different types of vegetation.

Currently, the military uses the NATO Reference Mobility Model (NRMM) [30] to predict the capabilities of vehicles in various outdoor terrains. This was written mainly for large vehicles that are comfortable with plowing through various types of vegetation with no cost calculation. The only vegetation calculations that it describes are those for overriding small trees. These empirical equations were developed through experimentation in the 1960's and have not changed very drastically since.

The NRMM contains years of empirical data collected from UGVs and uses simple look-up tables from past data to predict mechanical mobility [31]. In order to stray away from mechanical and driver-centric factors, the U.S. Army Engineer Research and Development Center (ERDC) is currently developing a Reference Autonomous Mobility Model (RAMM) [32] which is symbiosis of empirical data and modeling the sensor-environment interactions. Though the paper did not mention modeling vegetation specifically, RAMM may handle vegetation in a more robust manner than the NRMM.

Ordonez et. al. [38] wrote one of the only papers I found which attempts to model vegetation. The robot uses a torque-force sensor to train a regression model based on the equation for thin rods anchored at the ground. As the name of the paper suggests, this method could be useful for modeling pliable, single stemmed vegetation. I believe the various complex structures of other vegetation types will not be as easily estimated by physical models.

2.2 Scene Understanding

When it comes to autonomous classification of an environment, it can be argued that there are two major methods; object detection and semantic segmentation. Object detection algorithms evaluates an image and places a bounding box over any objects belonging to one of the predefined classes. Semantic segmentation labels every pixel in the image (including 'unknown') which helps with scene understanding and setting exact boundary locations.

Object detection is useful for finding and tracking specific objects in an image and is commonly seen in self-driving car demonstrations to locate pedestrians, cyclists, and other cars. A couple popular image detection algorithms are Faster R-CNN and You Only Look Once (YOLO). Impressively, YOLO v2 was able to train on 9,000 classes and classify images at 40 fps with an accuracy of 78.6% mAP on the VOC 2007 dataset [4]. Faster R-CNN also has impressive metrics with accuracy of 73.2% mAP at 5fps on the same dataset [5].

With that being said, semantic segmentation should be used for off-road travels instead of object detection. Object detection is useful for finding specific objects in a scene, but ignores the majority of the background image. It is important to understand the entire scene of an outdoor landscape, including the specific boundaries of roads, vegetation, and obstacles.

One possible algorithm for semantic segmentation is ICNet [6], an impressive image cascade network that can fully segment high resolution images with 19 classes at an accuracy of 70% mIoU. In the paper, a Nvidia TitanX was used to segment 1024 x 2048 images at 30 fps. It was also tested on the COCO-stuff dataset which has 181 classes, and achieved an accuracy of 29.1% mIoU at 35.7 fps.

Alternatively, there are other algorithms which have higher accuracy but runs at around 1 fps. These algorithms include PSPNet[1], Resnet38[2], and DUC[3]. Each of these three methods for semantic segmentation have approximately the same accuracy and classification speed as each other, but their methods may be more or less difficult when adjusting the balance between speed and accuracy.

The Microsoft COCO dataset is an important reference for semantic segmentation. In 2017, the COCO-Stuff challenge was presented to focus on segmentation of "stuff" objects [7]. Stuff objects are amorphous background objects of no distinct shape. This includes grass, sky, soil, and many types of vegetation that an offroad robot would need to analyze. A second challenge, the 2018 Panoptic challenge [8] focuses on both stuff and thing objects, which may be necessary for the system described in this paper.

Some of the less sophisticated off-road robots simply pull in RGB features to a train a classifier to segment the image into traversable/ not traversable or ground/ obstacle [21], [22], [17], [18]. Other papers mention the benefits of deep neural networks but choose to use a simpler model to save on processing, such as Texton Boost [29], random forest [12], or a shallow network classifier [24]. [13] Shows a more sophisticated and robust segmentation method using a CNN to segment the scene into a few classes, which will be more appropriate for our system.

2.3 Sensors

Many off-road autonomous robots have been built using stereo alone with successful models [14], [17], [18], [21], [23], [24], [29]. RGB images are the best data source for classifying paths, obstacles, vegetation, and pedestrians. Stereo cameras allows robots to build disparity maps and give 3D locations to each pixel in the image. For many applications, stereo vision is all that is required to build a 3D global map of viable paths and obstacles.

Lidar is also very useful in autonomous robotics but is rarely seen being used by itself. Advantages to lidar are being able to see in the dark and collect a rich 3D point cloud of the environment, which are more reliable than disparity maps from stereo. Disadvantages to lidar is that there are gaps in the point cloud due to scan line limitations and they are inable to classify objects at the level that images can.

More advanced and robust autonomous systems will be seen using both stereo and lidar [12], [13], [15], [22], [36]. The two sensors are used to their strengths and makes up for the inadequacies of the other. This usually comes in the form of using RGB features to classify items in an environment and using lidar data to build the environment map, or using features from both to build a traversability map. Although it can become computationally and memory expensive, it is usually worth using both sensors for more robust autonomous systems.

3 System Outline

Overview

The robot will record a continuous video feed with a stereo camera. Using a semantic image segmentation algorithm the environment will be segmented automatically. Paths with obstacles that are assumed to give the highest resistive forces (rocks, trees, dense vegetation) will be excluded from vegetation cost calculations. Only vegetation with the most viable paths, based on their respective classes, will then be evaluated using their physical characteristics as features fed into individual regression models. The segmented image will be projected onto the lidar point cloud to build a 3D voxel cost map.

Each class will have its own regression model that will take inputs such as height, lidar density, speed of robot, then output estimated energy consumption and possibly maximum forces in the x and z direction. Using the 3D stereo and lidar, the energy consumption estimates will be calculated for each path and sent to the motion planning algorithm. Data for each segment of vegetation in consideration will be saved temporarily. After a certain vegetation is traversed, the true values will be recorded and the respective regression model will be updated online and unsupervised through stochastic gradient descent.

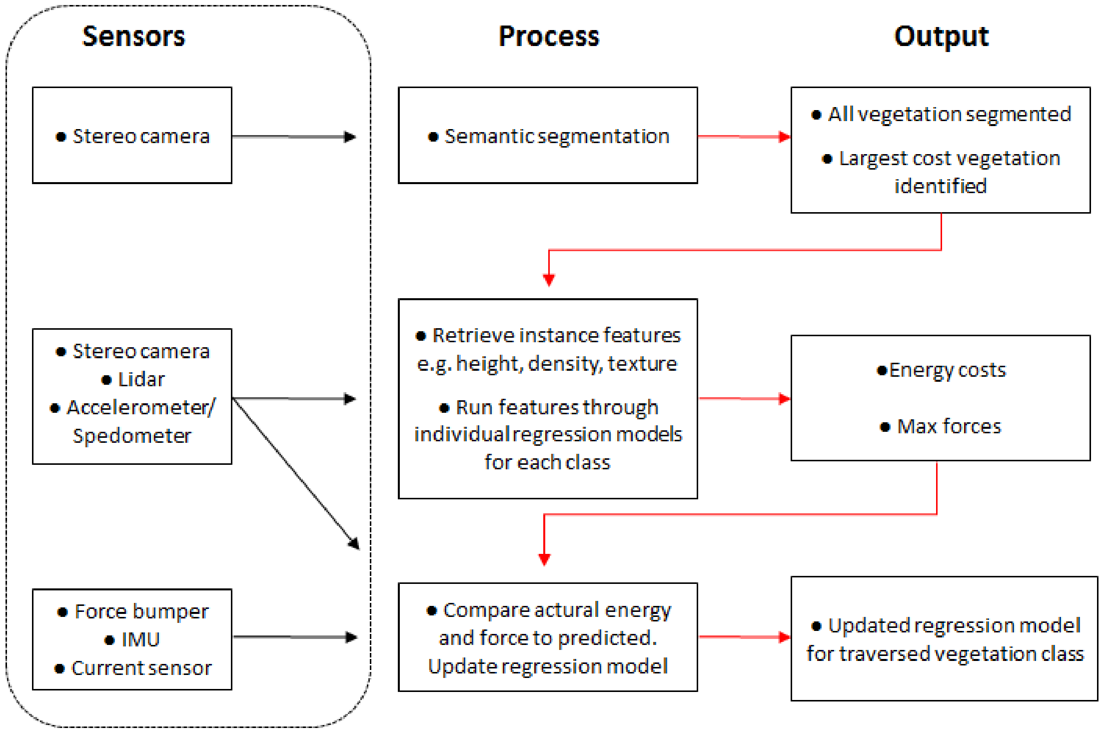

Figure 1: System Diagram

3.1 Avoiding an Individual Regression Model

A major component to my system proposal is building a large dataset to construct multiple hierarchical vegetation classes, with a regression algorithm attached to each class. It can be argued that decent results of vegetation forces can be obtained by pulling features from various sensors and putting them into one regression algorithm. For example, the regression algorithm could be a convolutional neural network with external lidar features fed in towards the end after convolution and output cost.

This would simplify the system, cut down on in-field processing, and reduce the effort on training the classification model. The pros to a single regression neural network, random forest, support vector machine, etc. are tempting, but we should take into account where machine learning and computer vision are headed.

Machine learning, automation, artificial intelligence and computer vision are extremely popular and rapidly advancing fields. Although I believe that the system described in this paper can be created at our current technology level, the implementation of it will quickly become easier and easier. Processor capabilities will only keep increasing at lower costs, large real-world object classification datasets will continue to grow, and algorithms will continue to improve.

If a system to calculate vegetation costs was to be created with a single regression algorithm, I believe it would be obsolete fairly quickly. The future of artificial intelligence is headed in a direction of scene understanding, logical deductions, and relational learning. These ideas work best when there are classes to connect various objects and ideas. Scene understanding, especially, will be an important function of future autonomous robots, and will require many classes to understand relationships of an environment. It will be much more difficult to understand an environment of continuous values than defined classes.

Aside from all this, I also believe that the accuracy will be much higher when a regression model is assigned to individual classes. As I have mentioned already, and will elaborate in further sections, individual vegetation species have various hidden nuances embedded into their structure. It may be difficult to capture these inherent properties in one all-encompassing network that relies on pattern features or lidar point cloud data. Each vegetation type can have these properties built into the regression algorithm, and then use sensor features as supplemental data.

3.2 Semantic Segmentation

3.2.1 Expanding the number of classes

Semantic segmentation is commonly used for autonomous off-road robots in many path planning algorithms. If a robot is to leave the confines of a man-made road, the problem of current segmentation algorithms is how few classes the models present. Quite often there are only a few classes such as sky, ground, vegetation, trees, grass. This limits all types of vegetation into one or two classes and greatly hinders the ability of the robot to analyze its surroundings if traveling off man-made roads.

One importance of including more classes is to capture the underlying and universal properties of each vegetation type. When analyzing visual data for size or texture patterns, there are many aspects to the features that may be difficult to build into an all-encompassing neural network for something as stochastic as vegetation. This could become increasingly difficult in future advancements of this system when factoring in properties of the robot and environment such as speed, acceleration, time of year, or ecosystem type.

"Underlying properties" could also be described as the properties important to building a theoretical physical model through machine learning. For example, a tree will act similar to a rod pinned at the bottom and will have a high impact force on the robot. Tall grass will require bending of the blades at the wheel or track level and cause some wheel slip when traversing over. Vines will be based on tension forces and will have a high risk of getting tangled. Short bushes or hedges will require the robot 'climb' over top of them and will likely have a physical model different than anything else listed above. This is why one single neural network may have a difficult time building an all-encompassing model.

An important feature in one type of vegetation may not be so important in another when it comes to resistive forces. Two types of vegetation may give similar feature values, but the DNN could have trouble determining rigidity and strength of branches, if the plant has harmful barbs or thorns, if the vegetation has vines that can tangle the robot, whether a plant has a rigid center base hidden behind foliage or sprouts up as a fountain grass, etc.

The idea is to have the regression model for each vegetation class built on a foundation of these inherent properties that sensors may not be able to detect. The proposed individual regression model will know that a fountain grass plant with higher density and larger size will still be a more efficient option than the small hedge of thick, rigid branches. These inherent properties will allow the robot to automatically filter the image into a hierarchy of traversability likelihood (i.e. eliminating trees, hedges, or cacti patch opposed to grass or sparse bushes).

Simple features such as height, density, and texture patterns pulled from various sensors will then be applied to the remaining vegetation segments. In theory, this should provide a more accurate output of energy and maximum forces than a large deep neural network due to each regression model being tailored to a specific vegetation type. With the knowledge of stiffness, saturation, vines, rigid branches, hidden trunks, and other properties hidden from sensors, the model will be able to compute more accurate resistive forces based solely on the vegetation class.

A second advantage to creating more segmentation classes is the ability of real-time online learning. With one single neural network to calculate all forces, it would be extremely difficult to do any real-time adjustments to the cost model. This concept may be important when exposing the robot to new environments where the vegetation looks similar but regional factors cause it to have varying properties. The idea also holds for the same vegetation in different seasons.

For example, tall grass in Ireland during spring will have different properties of tall grass in Northern Africa during summer. Although it is an extreme example, it can be understood that depending on area and time of year, vegetation will have different variations and inherent properties. With a stochastic gradient descent algorithm, the robot can use the smaller base regression models and modify them automatically as it moves through the vegetation of a new area. Additionally, the algorithm can use a weighted coefficient algorithm which puts more importance on recent data coming in from new locations to update the regression models.

Another advantage of creating more vegetation classes is to give the robot the ability of visual relationships. Visual relationships have shown to give much higher scene understanding in computer vision. It is the process of using the position of certain objects to predict the possibilities of other objects. This can help categorize or predict different classes that would be grouped together or usually next to, above, or around another object.

For instance, if a robot is having trouble determining what is sky and what is water, it can assume anything below ground level will be water and anything above will be sky. Another simple example is determining what are treetops and what are larger bushes based on their relative location. This idea can be expanded to possibly find patterns in nature which reveal predicted locations of certain items like rivers, cliff drop offs, weeds, obstacles, or specific plants due to natural patterns of other classes that surround them. It can also be applied to help robots classify the type of environment they are in based on the plants around such as garden, desert, forest, farm, jungle, etc.

Aside from traversing more safely and being energy efficient when traveling off-road, a more comprehensive vegetation classification dataset will help robots of the future identify specific plant life for agriculture, landscaping, or surveying purposes. If a robot is to be surveying and vegetation mapping, then a greater classification set of vegetation is necessary. As robots become more versatile and robust in the farming industry, classification of various crops and weeds will be essential. As robots become more common in the service industry, landscaping robots will need classes of commercial and residential vegetation for various maintenance routines.

3.2.2 Hierarchical Classification

The proposed system will classify vegetation into a hierarchy of subclasses (See Figure 2). The COCO-Stuff dataset [7] mentions using subclasses in their paper to help annotators classify more quickly when creating the dataset. They state that it took the annotators much less time to filter down through each subclass instead of having one large list of classes. This will prove to be useful in the creation of large dataset that will be necessary for classifying and cost estimation.

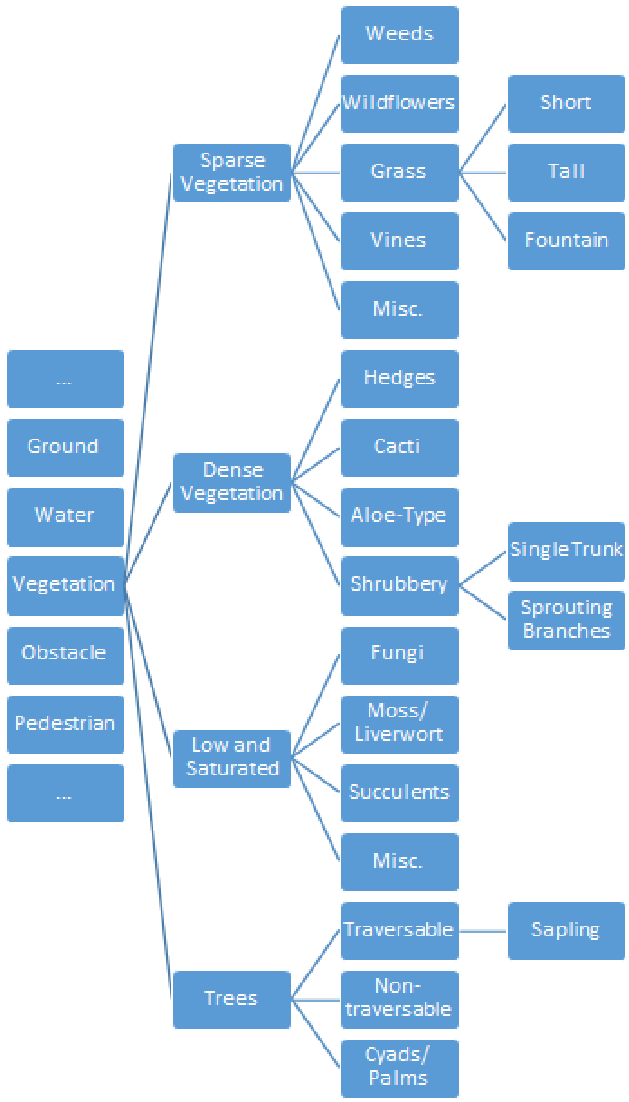

Figure 2: Class Hierarchy Example

Aside from helping with the initial creation of the dataset and keeping the data organized, a hierarchical classification is predicted to be beneficial when analyzing data as well. It could provide levels of classification for different scales of granularity in various implementations and purposes. As an example, if a user wants to design a system that doesn't need to calculate the costs for every type of vegetation in view, then it can automatically eliminate the vegetation of the highest costs such as trees and hedges.

In addition, multiple levels of classification could be useful for scene understanding. If a robot is being used for simple off-road mobility, the highest level of classes may be used for scene understanding but the lowest level for calculating costs. If a robot is being used for more specific tasks that require interaction with, or analyzation of specific vegetation, then the lowest levels of classification would be more useful for scene understanding.

If a new piece of vegetation is detected in-field, the robot would have more options on how to handle traversability costs than to simply bucket it into the closest subclass. One option would be to complete an k nearest neighbor approach and base costs off the superclass characteristics. Simply, building a hierarchical class structure will give more levels of granularity, more options to optimize evaluation, and simplifies the annotation process.

3.2.3 Segmentation Overview

A network such as ICNet [6] can be built to segment high resolution images at rates of around 2-5 fps. The only input to the model will be an image from the robot's stereo vision. The output will be a fully segmented image which sets boundaries around every group or instance of thing and stuff objects [7]. After segmentation, the robot will do a high level path elimination based on vegetation class and height.

The number of classes will need to be determined through experimentation and further investigation of plant class, family, and genus. There is a huge availability of granularity in the classification, but expanding the amount of subclasses will be unnecessary after a certain point. The optimal number of classes would be an amount that would be able to separate the main characteristics of each vegetation type that would affect resistive forces. Further classification would be necessary for landscaping, agriculture, and plant monitoring but a base dataset can be built for simple off-road autonomous travel.

Transfer learning is a powerful tool used to increase the speed of training neural networks [21], [22], [23]. Vegetation is predicted to be much more difficult to classify than man-made objects due to its stochastic nature and the similarities across all vegetation types. Training the network for segmentation and classification is expected to be a timely process. There are networks available which classify between a handful of vegetation types in outdoor environments, and using these as a base to start the training will cut down on convergence time.

3.3 Regression Algorithm

The purpose of the regression algorithm is to output costs for traversing through vegetation. These costs can come in the form of maximum vertical and horizontal force for impact analysis, predicted energy spent per distance, or a general cost measure which aligns with other environmental costs already set in place for the robot. These vegetation costs will then be combined with other costs such as slip, slope, and obstacle traversal to plan the optimal path.

It was determined that the robot could either have a simple classification system of a few classes and a robust regression model or a robust classification model with simple regression models. The first option of a robust regression model would have likely resulted in a convolutional neural network with a continuous output for vegetation forces. This would have required much more data collection for actual vegetation traversal and less data for labeled vegetation types.

The second option of a strong classification model and weak regression model was chosen because it is predicted to be more robust, more accurate, and more scalable as robotics hardware and software advances as mentioned above. As long as the inherent properties of vegetation are separated correctly into appropriate classes then the regression algorithm will need substantially fewer features to accurately predict vegetation costs, and less training data to build the model.

There are many options for regression algorithms which will need to be tested once actual data is collected. A simple polynomial or nonlinear regression may suffice since the heavy lifting for finding vegetation properties will be done previously by the segmentation network. If maximum impact forces are desired along with the energy costs, a multivariate regression model will be used.

Another option would be to use an Extreme Learning Machine (ELM) [10] which is a type of simple feedforward neural network with only one iteration of training. It has very fast training time which is necessary to update the regression models in real time, but requires the full training data in memory. It was used by Liu et al. [11] for terrain classification with promising results but can be used for regression tasks as well. Though an ELM is not practical for data of high dimensionality and has received some criticism for its simplicity, it may prove useful for our data.

The inputs to the regression algorithm will include image features, lidar data, and possibly speed or acceleration values in the future. The image features will include simple metrics such as size and texture patterns, to be later finalized through experimentation. The lidar data will provide density information of leaves, branches, and possibly hidden trunks. Speed and acceleration is an important factor when determining forces and energy to overcome vegetation, but it will also require a lot more data to properly implement. The first implementation of this system will likely evaluate a robot moving at constant speed.

3.4 Environment Mapping

Ideally, a 3D voxel map as in [12], [14] would be constructed with information containing height, class, and vegetation cost. A 2D traversability map or occupancy grid [16] [17], [18], [19], [20], [21] will not be applicable because it usually only contains a traversability cost, if not just a binary traversability label. Height values are necessary in eliminating vegetation early on by evaluating height and class. Although it contains height data, a 2.5D heightmap [13], [15] needs to be dismissed as well due to it not being able to recognize overhanging vegetation or two types of vegetation in the same 2D cell.

3D voxel maps can be created using a dense stereo correlation algorithm to create disparity maps as in [14] or 2D maps as in [17] and [15]. Although for many applications stereo alone is sufficient, stereo based distance estimation can be unreliable after 30 or 40 feet [18] and images have a hard time detecting objects hidden by foliage. Lidar is very good at returning accurate 3D locations for objects and finding hidden solid objects, but it can not classify the vegetation as we need it to. Due to these reasons, it is recommended that both Lidar and stereo data be used when building the 3D voxel maps, possibly by using a simple mapping feature as in [13] and [22].

3.5 Online Learning

The regression model structure as described above allows for the use of online learning. If the regression model is smaller and easier to modify, it can be updated in real time using stochastic gradient descent to adjust to new environments. This will be much more important to update regularly than the classification of vegetation because the robot will continuously be collecting new data on vegetation costs. If a robot enters a new environment with different weather or new vegetation types, the regression model can be structured so that it is updated with more weight given to new data coming in and adjust costs accordingly.

Stochastic gradient descent will be used to train the regression models in real time. It differs from standard batch gradient descent in that it trains a model using one sample or minibatch of samples at a time [9]. In this way we can continuously update the regression model without having to completely retrain it every time a new sample of data comes in. A downside to this method is that the path to minimum error is more noisy due to the variance of individual data samples, opposed to a combined sample set average.

Once the scene is segmented and possible paths are determined, the features for vegetation in each path will be computed and temporarily stored. A patch of vegetation will then be traversed and measurements of force and energy will be measured using and IMU, current sensors, and/ or a force bumper. The stored features and measured output will be fed into the appropriate regression model for an individual or mini-batch update with stochastic gradient descent.

The classification neural network will need to be updated periodically offline. In reality, new annotated vegetation data will not be created often. When it is decided that new data should be included to either add new vegetation types or increase accuracy, the current network will work as a base and retrained with the new data.

3.6 Sensors

The autonomous vehicle using this system for calculating vegetation costs should have both a lidar sensor and stereo vision. This dual setup is fairly standard in most research papers for off-road autonomous travel as mentioned in the related work section. Each have their strengths and weaknesses and would likely result in poor output if used on their own.

Stereo vision will be essential in the segmentation and classification process because texture and pattern features found from a convolutional neural network is the most powerful scene segmentation method available at this time. If the quality is high enough, camera images can detect sparse objects, thin objects, color characteristics, and slight pattern differences which will help greatly in understanding the robot's environment.

Lidar is very useful for highly geometric man-made objects but has trouble detecting sparse and thin objects properly due to scan angle and resolution limitations. This of course will be a problem by itself when trying to determine vegetation characteristics. Lidar will be useful for general environment mapping, which can be combined with the stereo vision distance measurements if needed. Another great property of lidar is that it can calculate density characteristics and in some cases "see" behind foliage to detect solid objects occluded by sparse vegetation [37], [39]. Lastly, it is also able to see with little to no light just as well as it can during the day.

Both [13] and [36] explicitly state that lidar geometric features are not sufficient on its own for off-road travel. I expect that stereo vision will be used by itself to segment the scene and lidar will be used to help create an environment map. Once the scene is evaluated and the robot calculates vegetation cost from the regression algorithm, it will include some simple image features and lidar density features.

For initial data collecting and online learning, the robot will need to measure the actual energy spent and impact force from traversing over vegetation. An IMU and force bumper will be able to calculate deceleration and impact from vegetation. A current sensor can be placed on the motors to detect power consumption just before vegetation traversal and during vegetation traversal. The difference between the two can be recorded as the power/ energy required to surpass that specific type of vegetation.

3.7 Efficiency Considerations

The system described thus far requires processing of both lidar and RGB data, segmentation, regression, adjusting models, and environment mapping on top of all of the many other logical functions of the robot. Making the algorithms as efficient as possible will be an important factor when implementing the ideas described above. To reduce some computation time and energy I have listed a few possible efficiency considerations below that will be developed further at a later time.

In order to prevent calculating costs for all vegetation in a given scene, a very simple algorithm will partition the segmented scene into vegetation that should have costs calculated and vegetation that should not. The algorithm will likely be based solely on height and superclass, and will eliminate vegetation that is clearly going to give much more resistance than others. Vegetation in the paths of solid impassable obstacles will also be removed from calculations because they will not be traversed through regardless.

Vegetation at a further distance should be analyzed differently than vegetation within a shorter distance of the robot. Although vegetation in the distance can be important for a more complete and robust motion planning, the calculations for cost do not need to be as precise as the ones requiring an immediate decision. If the robot calculates vegetation at all distances visible by the sensors, the data will likely be unreliable, costly, and not nearly as important as the vegetation closer to the robot. I suggest a much simpler algorithm to feed into the motion planning for vegetation after a certain distance.

Due to the simpler metrics provided for vegetation at a greater distance, it would be advantageous to transition from a 3D voxel map to a 2.5D height map past a certain distance which can still hold data for motion planning but does not take up unnecessary memory. Only a rough estimate of occupancy, class, and height should be sufficient for planning past the immediate radius useful for calculating vegetation costs.

4 System Simulation

Consider the following scene in Figure 3, which is a simple hill with various obstacles obstructing the path between the robot and its objective. If the vegetation is taller than a conservative pre-set height, many robots will mark any paths with vegetation as non traversable. The red line represents the theoretical path for smaller robots with current vegetation assessment capabilities. The purple line is the path with a more in-depth vegetation analysis. Although it is a simple example, it can be understood how multiple instances of this situation can cause the robot to waste a lot of time and energy avoiding vegetation, or possibly never even being able to reach its objective.

Figure 3

First and foremost the robot will use a neural network to semantically segment the scene and output data as visualized in Figure 4. The robot will segment the scene around 5 fps and label a classification for every pixel as an output. At the same time, lidar will be collected to gain a better understanding of spatial and density information (Figure 5). The lidar will useful in building a 3D map of the environment and finding hidden attributes while the stereo vision will be necessary for classifying the different types of vegetation.

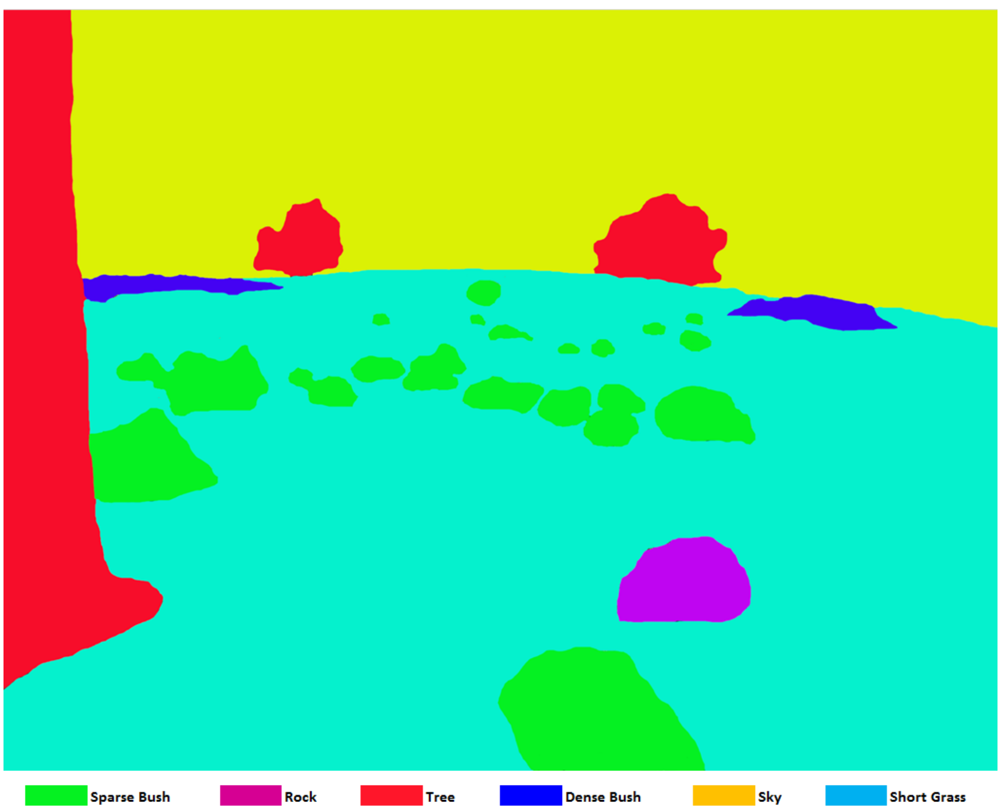

Figure 4

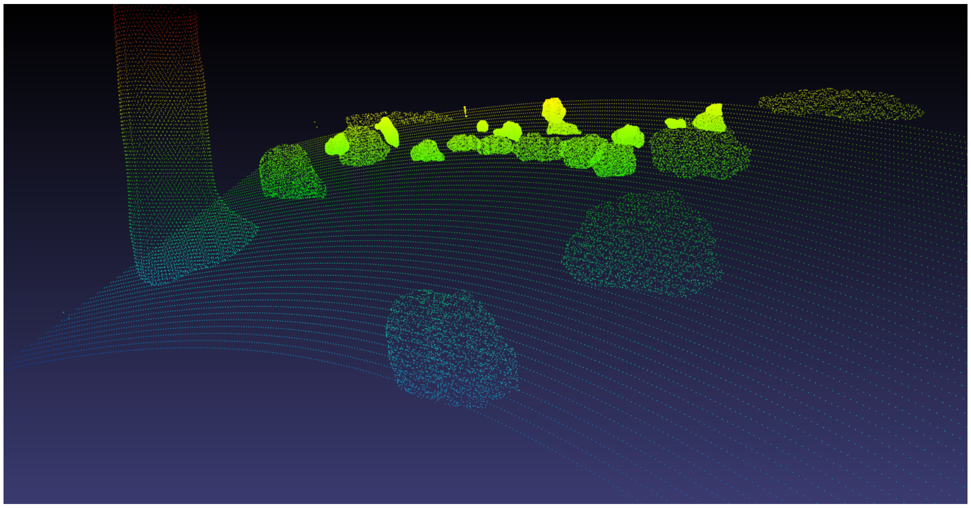

Figure 5

The two sets of data can be merged by using a mapping function as illustrated in Figure 6. Once the two coordinate systems are carefully calibrated, the classification data can be assigned lidar point cloud attributes and vise versa. As mentioned previously, vegetation such as trees and large hedges will be eliminated from evaluation as well as vegetation obstructed by obstacles. In the case of our example, the trees and any vegetation around the rock will not be evaluated. The rest will be run through a regression algorithm to determine estimated cost of traversal.

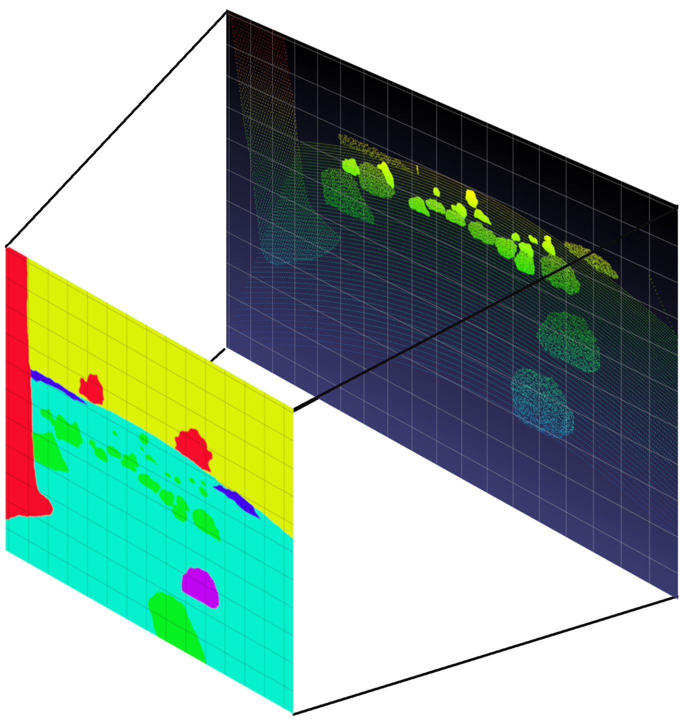

Figure 6

Once the attributes of class, height, and cost are collected, a 3D voxel map will be created to gain a comprehensive understanding of the environment (Figure 7). A motion planning system can then collect all other costs outside of vegetation and plan the most efficient path accordingly. Proprioceptive feedback sensors will collect data of 'actual cost' during traversal of the vegetation and, using the temporarily stored vegetation attributes, will update the regression model through stochastic gradient descent.

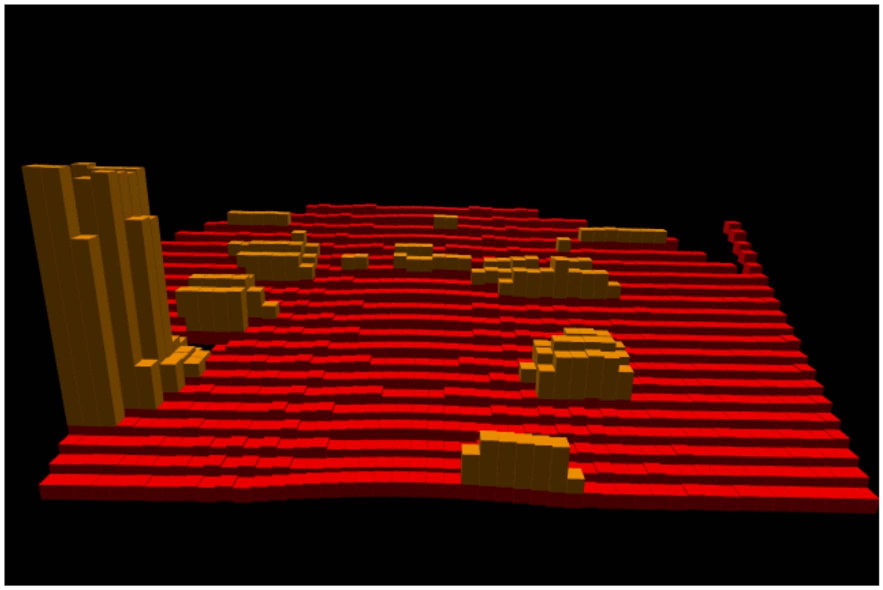

Figure 7

5 Conclusion

This paper presents research on the current technologies for off-road robotics when traversing through vegetation and suggests a more robust system for future implementation in autonomous mobile robots. Currently the majority of "off-road" autonomous vehicles are either restricted to premade paths free of vegetation, or treats all vegetation equally as "traversable" or "non-traversable" classes. Even the most advanced autonomous systems only have a simple cost system associated to a few vegetation classes. The current simplification of vegetation costs greatly hinders a robot's capabilities and efficiency due to its path planning being built around overly-cautious movement.

Instead of building one deep neural network regression model to compute predicted vegetation resistance, it was decided to expand the number of vegetation classes and have a simple regression model for each. This method is believed to have better accuracy, give more flexibility in tasking options for future robots, and is more in line with the future of AI technology.

The robot will first semantically segment a scene and remove paths with the highest resistive forces based solely on class and simple size measurements. The remaining vegetation will be evaluated using image and lidar features for their respective class regression models. A path will be selected which minimizes total energy and impact force based on vegetation features and other environmental factors. Data will be collected when traversing over selected vegetation, which can then be used to update the regression models in real-time through unsupervised online learning.

6 Future Work

Naturally, the next step for this project would be to build the actual system that is described in this paper as a proof of concept. The regression dataset will be created from controlled outdoor scenes with a limited amount of vegetation types, no slopes, consistent soil type, and constant speed. The classification dataset can start with open source annotated data with vegetation but will likely need to be expanded upon with newly annotated data for our specific purpose. Quantity of vegetation classes and types of vegetation to test on for a proof of concept will be carefully deliberated.

The dataset will need to have annotated images of various vegetation as well as sensor data from the robot from when it traversed through the vegetation. This proprioceptive feedback can come through a force sensing bar on the front of the robot and current sensors on the motor as described above. The process of collecting enough data to accurately model these classification and regression models will take a notable amount of time and effort.

Once enough data is collected, the semantic segmentation neural network will be built starting with an appropriate pre-trained model as a base via transfer learning. Various regression models will be deliberated, constructed, and tested with inputs from lidar and stereo imaging. It may be decided to structure the regression models in two ways; with only a few vegetation classes as a baseline and with a larger number of classes as presented in this paper. The comparison of the two dataset structures will shed light on how much accuracy we gain and efficiency we lose when increasing the number of classes and subclasses. Accuracy will be computed for vegetation similar to what was traversed during training and vegetation never-before-seen by the robot.

References

[1] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR. (2017)

[2] Wu, Z., Shen, C., van den Hengel, A.: Wider or deeper: Revisiting the resnet model for visual recognition. arXiv:1611.10080 (2016)

[3] Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.W.: Understanding convolution for semantic segmentation. arXiv:1702.08502 (2017)

[4] Redmon, Joseph, and Ali Farhadi, "YOLO9000: better, faster, stronger," arXiv preprint arXiv:1612.08242, 2016.

[5] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497 (2015)

[6] H Zhao, Xiaojuan Qi, Xiaoyong Shen, Jianping Shi, Jiaya Jia, "ICNet for Real-Time Semantic Segmentation on High-Resolution Images," .arXiv preprint arXiv:1704.08545, 2017

[7] Caesar, H., Uijlings, J., Ferrari, V., Coco-stuff: Thing and stuff classes in context, .arXiv:1612.03716 (2016)

[8] Kirillov, A., He, K., Girshick, R., Rother, C., Dollar, P.: Panoptic segmentation. In: arXiv ´ preprint arXiv:1801.00868. (2018)

[9] J. Kiefer, J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," Ann. Math. Statist., vol. 23, no. 3, pp. 462-466, 1952.

[10] Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **70,** 489–501 (2006).

[11] Liu, Y., Fang, J. & Liu, C. Research of terrain recognition for off-road robot based on extreme learning theory. *2016 International Conference on Advanced Mechatronic Systems (ICAMechS)* (2016).

[12] Bradley, D. M. *et al.* Scene understanding for a high-mobility walking robot. *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2015).

[14] M. Bajracharya, J. Ma, A. Howard and L. Matthies, "Real-Time 3D Stereo Mapping in Complex Dynamic Environments", Int. Conf on Robot. and Autom.-Semantic Mapping Percept. and Explor. (SPME) Workshop 2012.

[15] Bajracharya, M. *et al.* High fidelity day/night stereo mapping with vegetation and negative obstacle detection for vision-in-the-loop walking. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2013).

[16] Hussein, A., Marin-Plaza, P., Martin, D., de la Escalera, A. & Armingol, J. M. Autonomous Off-road navigation using STEREO-vision and laser-rangefinder fusion for outdoor obstacles detection. *2016 IEEE Intelligent Vehicles Symposium (IV)* (2016).

[17] Konolige, K. *et al.* Outdoor mapping and navigation using stereo vision. *Experimental Robotics* 179–190

[18] Online learning for offroad robots: Using spatial label propagation to learn long-range traversability. *Robotics* (2008).

[19] Reina, G., Milella, A. & Worst, R. Lidar and stereo combination for traversability assessment of off-road robotic vehicles. *Robotica* **34,** 2823–2841 (2015).

[20] Suger, B., Steder, B. & Burgard, W. Traversability analysis for mobile robots in Outdoor Environments: A semi-supervised learning approach based on 3D-LIDAR data. *2015 IEEE International Conference on Robotics and Automation (ICRA)* (2015).

[21] Artem Provodin, Liila Torabi, Beat Flepp, Yann LeCun, Michael Sergio, Lawrence D. Jackel, Urs Muller, Jure Zbontar: Fast Incremental Learning for Off-Road Robot Navigation. CoRR abs/1606.08057 (2016)

[22] Schilling, F., Chen, X., Folkesson, J. & Jensfelt, P. Geometric and visual terrain classification for Autonomous Mobile Navigation. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017).

[23] Holder, C. J., Breckon, T. P. & Wei, X. From on-road to off: Transfer learning within a deep convolutional neural network for segmentation and classification of off-road scenes. *Lecture Notes in Computer Science* 149–162 (2016).

[24] Suleymanov, T., Paz, L. M., Pinies, P., Hester, G. & Newman, P. The path less taken: A fast variational approach for scene segmentation used for closed loop control. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2016).

[25] Fleischmann, P., Kneip, J. & Berns, K. An adaptive detection approach for Autonomous Forest Path following using Stereo Vision. *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (2016).

[26] Giusti, A. *et al.* A machine learning approach to visual perception of forest trails for Mobile Robots. *IEEE Robotics and Automation Letters* **1,** 661–667 (2016).

[27] Pan, Y. *et al.* Agile autonomous driving using end-to-end deep imitation learning. *Robotics: Science and Systems XIV* (2018).

[28] Lee, H., Kwak, K. & Jo, S. An incremental nonparametric bayesian clustering-based traversable region detection method. *Autonomous Robots* **41,** 795–810 (2016).

[29] Bernuy, F. & Solar, J. R. Semantic mapping of large-scale outdoor scenes for autonomous off-road driving. *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)* (2015).

[30] McCullough, M., Jayakumar, P., Dasch, J. & Gorsich, D. The next generation NATO reference mobility model development. *Journal of Terramechanics* **73,** 49–60 (2017).

[31] Durst, P. J., Anderson, D. T., Bethel, C. L., Carruth, D. W. & Monroe, G. A history and overview of mobility modeling for Autonomous Unmanned Ground Vehicles. *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything* (2018).

[32] Durst, P. J., Goodin, C. T., Anderson, D. T. & Bethel, C. L. A reference autonomous mobility model. *2017 Winter Simulation Conference (WSC)* (2017).

[33] Bergerman, M., Billingsley, J., Reid, J. & van Henten, E. Robotics in agriculture and Forestry. *Springer Handbook of Robotics* 1463–1492 (2016).

[34] Durmus, H., Gunes, E. O., Kirci, M. & Ustundag, B. B. The design of General Purpose Autonomous Agricultural Mobile-Robot: “agrobot.” *2015 Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics)* (2015).

[35] Bawden, O. *et al.* Robot for weed species plant-specific management. *Journal of Field Robotics* **34,** 1179–1199 (2017).

[36] Tang, L., Ding, X., Yin, H., Wang, Y. & Xiong, R. From one to many: Unsupervised traversable area segmentation in off-road environment. *2017 IEEE International Conference on Robotics and Biomimetics (ROBIO)* (2017).

[37] Matthies, L., Bergh, C., Castano, A., Macedo, J. & Manduchi, R. Obstacle detection in foliage with Ladar and Radar. *Springer Tracts in Advanced Robotics* 291–300 (2005).

[38] Ordonez, C. *et al.* Characterization and traversal of pliable vegetation for robot navigation. *Springer Proceedings in Advanced Robotics* 293–304 (2020).

[39] Morris, D. D. Obstacles and foliage discrimination using LIDAR. *SPIE Proceedings* (2016).

[40] Angelova, A. N. Visual prediction of rover slip: Learning algorithms and field experiments. (2008).