The University of Melbourne - Department of Mechanical Engineering

Path generation for off-road navigation using aerial mapping

Tuan Khoi Nguyen (1025294) Fawad Ahmed Malik (1004550) Sean Breukel (1080348)
Victor Bouchet-Hibbert (915058)

Introduction

Since the turn of the century, autonomous systems and Artificial Intelligence (AI) at large have experienced exponential growth, with global investment in the autonomous vehicle industry alone exceeding US $80 billion in 2017, and continues to increase with the presence of competitors [1]. Another major application of AI in autonomy is through Unmanned Aerial Vehicles (UAVs). Presently, they have value in operations such as the inspection of construction sites [2] and agricultural areas [3], or the mapping of geographical features (ex. drainage ditches) [4]. In this report, we propose an application of AI that focuses more on exploration. Specifically, an AI model will be suggested to generate an optimum path in an off-road area for an autonomous unmanned ground vehicle (UGV) from real-time aerial map created by a tracking UAV for essential exploration purposes such as aiding disaster response.

1 Context

The proposed idea can be broadly split into two tasks: mapping an area via sensors from an aerial vantage point, and generating a path through the map towards a specified location.

1.1 Aerial Mapping

While creating a map of an area might appear as simple as taking a photo, numerous studies regarding the uses of UAVs greatly expand upon the topic, particularly with regards to generating maps through onboard sensors. LiDAR data is one such type of sensor, useful for its ability to generate 3D data, which can include elevation changes across terrain and between objects, as well as the straight-line distances, and its high penetration, which allows the mapping of even heavily-wooded areas [5]. Object identification, however, can be a tricky process with LiDAR data, given that it consists of multitudes of point clouds, and, as such, generally requires prior knowledge about the types of objects that can be found in a given environment. On the other hand, aerial photographs excel at allowing machine learning models to identify and classify obstacles. Its primary weakness however lies in that aerial photographs inherently lack a sense of scale, thus requiring information such as camera pose and position relative to the ground to determine critical information (e.g. distance between objects).

1.2 Path Generation

Primarily a computer vision-based task, path generation involves identifying possible routes between a starting point and an end point on a given map, and determining which would be best for an UGV to follow. The presumed best path however depends entirely on the type of data available about the environment, methods used to generate cost maps, and the evaluation metric a model might be using, such as distance, energy, or time [5]. Before a path can be generated however, an algorithm must be capable of identifying and locating the endpoint, whether it be a person, an object, or a set of coordinates, and localising the UGV in relation to said endpoint.

1.3 Existing Integrations

The quality of these relative coordinates though, is heavily dependent on the quality and quantity of information provided by the inputted map. By integrating aerial mapping obtained via UAVs, the path generation for UGVs can be made much more effective. For example, in the case of attempting to get a rover to traverse part of the Negev Desert [5], traversable terrain was established by identifying obstacles such as large rocks, structures, plants, bodies of water, and segmenting them from areas with roads or with uniform features. The segmentation happens on both the macro and micro level as with two convolutional neural networks (CNNs) utilising U-NET architecture processing low and high resolution maps obtained from high and low altitudes, respectively [6]. Based on the existing obstacles or lack thereof, map pixels are given traversibility values from 0 to 255, with 0 being the most easy to traverse. This naturally provides an easy-to-interpret mapping to calculate the path with the lowest cost, in this case the A-Star (A*) algorithm.

The aforementioned segmentation are necessary to create a terrain and cost map from the aerial data that determines the optimality of a chosen motion. There are approaches to create terrain maps that do not require neural networks; a Hue-Saturation-Value (HSV) classifier [7] for example. Based on the terrain and cost maps, pre-existing search algorithms such as A* [5] or D* [7] can then be used to obtain an optimal path for the ground vehicle to follow. However, they require the evaluations, i.e. heuristics to be self-formed by the user, causing uncertainty if the environment is not well-known. Neural network solutions on the other hand can be employed in the classification of aerial data into multiple classes and/or segments, with each having associated cost/rewards based on traversability that the model can either self-learn [8] or have information fed in by user, either partially or fully [9]. This can in turn greatly improve flexibility and autonomy for the operation of the specified tasks of interest.

2 Idea

While the above example primarily focused on applications relating to exploration, there are also additional practical uses in the form of disaster relief. When a disaster occurs, such that search-and-rescue operations are required, it is imperative that they be completed as quickly and safely as possible. To that end, the use of UAVs and UGVs combined not only can drastically decrease the dedicated manpower needed by locating rescuees and calculating optimal paths to navigate towards them, but also increase overall safety [10] and appraise disaster relief services with up-to-date information on the physical environmental impact caused by volatile disasters such as fires, earthquakes, or floods. However, existing studies that implement the use of UAVs and UGVs for the aforementioned search-and-rescue use cases [10, 11], generally focus on open areas that are relatively straightforward to map compared to unexplored and densely-forested environments [4]. Current well-known robotics and AI competitions held internationally such as EURATHLON and ELROB aim to improve the search and rescue use cases [12], yet their settings are also mostly limited to urban area simulations.

As such, we aim to propose an AI model system that uses a tracking UAV to produce aerial mapping of an off-road or unexplored area (that might be forested) in real-time, and subsequently generate an optimal path for an UGV to traverse in hazardous environments, relating to possible search-and-rescue scenarios.

3 Implementation

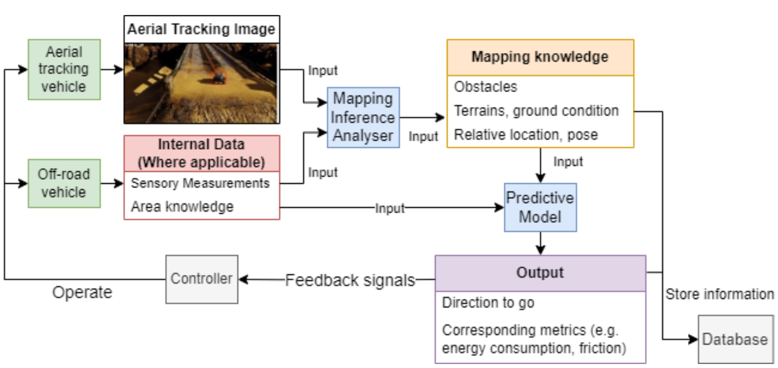

Figure 1: Proposed flowchart of the system. Image taken from KAUST dataset [13]

3.1 General Summary

For system implementation, the following final outputs can be of interest: the vehicle's decision of the next heading orientation at any given time instance, and additional corresponding information including energy consumption. These outputs are in the form of continuous variables, making it possible to be processed via neural networks, in the form of a multivariate regression problem. The model will be trained on different aerial tracking images of a vehicle at different instances, where map information and headings can be manually observed from the images, and previous statistical data can either be observed from internal vehicle sensor measurements, or determined from the travel process the vehicle takes before requiring maintenance. From the model predictions, other than storing data from the exploration, the vehicle can receive information to calculate its next heading autonomously. A feedback system can be created, where the efficacy of the recommended step can be assessed by comparing against the desired direction. A diagram summarizing the system flow is displayed in Figure 1.

3.2 Model Architecture

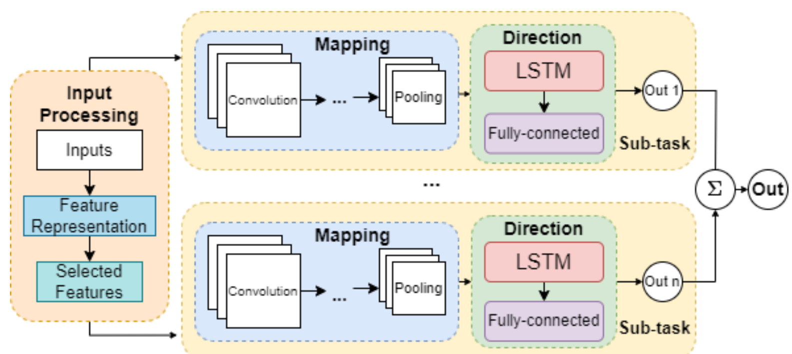

Figure 2: An example diagram of the suggested architecture

The proposed model architecture has to be capable of incorporating multiple steps as part of its process. This includes extracting and fusing important information from collected data to form mapping information. Image data of an operating vehicle can give many important details, such as the terrain being traversed [14, 15], the surrounding obstacles, or environment conditions (ex. rocky, wet, sandy, etc.) [16]. Being able to extract such features from images would require the important segments of details to be discovered, making CNN architectures being the feasible choice at this stage, which can either be developed from scratch, or adopt existing compact architectures such as VGG [17], U-NET [6] or MobileNet [18]. Another task that the model needs to perform is to be able to provide the desired estimations from the given processed information. With the extracted features being collected over time, they can be considered sequential data, allowing the following process to utilise time series neural network architectures such as Recurrent Neural Network (RNN) or Long-Short Term Memory (LSTM).

The suggested neural network architecture has to consider all possible data sources, even when their effect on determining the output is not yet known. One possible method to aid this is to use representation learning, where each features can be plotted on a multi-dimension space, thus discovering important relationships [19, 20]. Such an implementation would be able to perform initial feature selection, thus reducing the neural network's total training time and resource usage [20].

While the ultimate aim is to have the vehicle moving efficiently, environmental factors can be incredibly diverse, requiring multiple sub-problems on a variety of specific conditions to be solved in transit. One feasible solution is to consider multi-task learning, where the neural network can learn multiple relevant sub-tasks simultaneously, and combine together to achieve improved prediction results [21]. For instance, the model can learn separately, but simultaneously, the best way for an UGV to traverse snowy conditions and landslides, then combine the two in order to achieve greater results for avalanche rescue. This has been considered in similar rescue missions, where a multi-task learning model was successfully implemented in earthquake rescue by combining disaster classification and victim identification [22]. Figure 2 shows one of the possible model structures for the task, displaying the complete methodology mentioned in this section.

3.3 Data Source

As mentioned in Section 3.1, the input for the model suggested will be tracking images of moving vehicle or terrains, which are potentially aerial images, as it can capture many details of the surrounding environment from a safe distance. In order to collect data, possible sources can be both self-produce and collected via existing studies.

Many datasets of aerial images, both taken or generated from simulations, are available for various purposes and thus would be sufficient for a model to train and validate the extraction of many useful features. Some common examples include terrain segmentation [23], 3D-generated vehicle tracking images [13], or localized aerial images [24]. While this method may not satisfy specific aims, they are readily available and can be used to leverage more desired features.

Additional aerial data can also be collected via 3D simulation or on-site testing. This method adds an extra step to produce data, and requires specialised software or devices. That being said, it can produce data that is more tailored to the purpose of the research, improving the training process of the suggested model.

4 Potential Expansion Directions

Whilst our implementation describes one deep learning approach to the chosen task, there are still numerous ways that said implementation could be further explored.

4.1 Method Development

One such way would be to expand upon the methodology chosen to tackle the task. This could involve adding or exchanging which sensors are utilised for gathering input data for the neural network. RADAR could be incorporated, which may be more robust in adverse environmental conditions, such as reflective ground (e.g. snow or wetlands). This is because LiDAR and cameras are both light based sensors, where extreme illumination or weather can easily impair their ability to capture reliable data [25]. This expansion approach however offers unique challenges, as the compatibility of additional data sources would need to be further investigated. The neural network model itself could also be further developed; namely in the architecture's set up and whether any changes would need to be made to accommodate data from different sensor types. An additional consideration would be scalability concerning the unmanned vehicles, increasing the number of UAVs to aid in acquiring data. Intuitively, we may be able to leverage having multiple cameras apart from each other to provide a triangulation of the depth, similar to stereoscopy. It would also provide extra data from different viewpoints, providing the network an overall more detailed idea of the environment that it is mapping.

4.2 Application expansion

The application space can also be increased. This means extending the context in which our implementation can be applied, for instance, underwater or space exploration. Another possibility is to expand the capabilities of the UGV itself, for example by including actuators that allow it to impact the environment around it, such as shovelling or clearing a path.

5 Conclusion

Overall, AI is becoming more and more widely applied to the field of autonomous vehicles, though development is still progressing slowly in off-road navigation, due to lack of exploration and potentially hazardous conditions of the surrounding environments. Our idea proposes the use of neural networks and aerial image inputs to retrieve mapping information, as well as helping off-road vehicles to traverse an area efficiently, by combining architectures of CNN, RNN, LSTM, etc. and additional enhancing procedures of representation learning and multi-task learning. This can potentially reduce exploration time and prevent human exposure to hazards, boosting the efficiency of off-road navigation applications in critical areas requiring rescue operations. When a minimum viable product is developed, additional research can focus on improving the existing system or extend into more specialized applications, such as space exploration.

REFERENCES

[1] Cameron F Kerry and Jack Karsten. Gauging investment in self-driving cars. *Brookings Institution, October*, 16, 2017.

[2] Adrian Zbiciak and Tymon Markiewicz. A new extraordinary means of appeal in the polish criminal procedure: the basic principles of a fair trial and a complaint against a cassatory judgment. *Access to Justice in Eastern Europe*, 6(2):1–18, March 2023.

[3] Lala El Hoummaidi, Abdelkader Larabi, and Khan Alam. Using unmanned aerial systems and deep learning for agriculture mapping in dubai. *Heliyon*, 7(10):e08154, October 2021.

[4] William Lidberg, Siddhartho Shekhar Paul, Florian Westphal, Kai Florian Richter, Niklas Lavesson, Raitis Melniks, Janis Ivanovs, Mariusz Ciesielski, Antti Leinonen, and Anneli M. Ågren. Mapping drainage ditches in forested landscapes using deep learning and aerial laser scanning. *Journal of Irrigation and Drainage Engineering*, 149(3), March 2023.

[5] Raphael Prinz, Rizwan Bulbul, Joahnnes Scholz, Matthias Eder, and Gerald Steinbauer-Wagner. Off-road navigation maps for robotic platforms using convolutional neural networks. *AGILE: GIScience Series*, 3:1–7, June 2022.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[7] John Peterson, Haseeb Chaudhry, Karim Abdelatty, John Bird, and Kevin Kochersberger. Online aerial terrain mapping for ground robot navigation. *Sensors*, 18(2):630, February 2018.

[8] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning, 2013.

[9] Thomas Fournier, Arnaud Lallouet, Télio Cropsal, Gaël Glorian, Alexandre Papadopoulos, Antoine Petitet, Guillaume Perez, Suruthy Sekar, and Wijnand Suijlen. A deep reinforcement learning heuristic for SAT based on antagonist graph neural networks. In *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, October 2022.

[10] Jeffrey Delmerico, Elias Mueggler, Julia Nitsch, and Davide Scaramuzza. Active autonomous aerial exploration for ground robot path planning. *IEEE Robotics and Automation Letters*, 2(2):664–671, April 2017.

[11] D. C. Guastella, L. Cantelli, C. D. Melita, and G. Muscato. A global path planning strategy for a UGV from aerial elevation maps for disaster response. In *Proceedings of the 9th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications, 2017.

[12] Frank E. Schneider, Dennis Wildermuth, and Hans-Ludwig Wolf. ELROB and EURATHLON: Improving search & rescue robotics through real-world robot competitions. In *2015 10th International Workshop on Robot Motion and Control (RoMoCo)*. IEEE, July 2015.

[13] Matthias Mueller, Neil Smith, and Bernard Ghanem. A benchmark and simulator for UAV tracking. In *Computer Vision – ECCV 2016*, pages 445–461. Springer International Publishing, 2016.

[14] Yang Chen, Pengfei Zhang, Shifeng Wang, and Dawei Gong. Image feature based machine learning approach for road terrain classification. In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*. IEEE, August 2018.

[15] T. Kubota, T. Misu, T. Hashimoto, and K. Ninomiya. Image-based topographic recognition on natural terrain. In *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*. IEEE.

[16] Yuka Toyoshima, Yasuhiro Hayashi, and Yasushi Kiyoki. A composition-based image retrieval method for environment-visualization with images and spatio-temporal information. In *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*. IEEE, October 2018.

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.

[18] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *ArXiv*, abs/1704.04861, 2017.

[19] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, August 2013.

[20] Bernhard Scholkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, May 2021.

[21] Simon Vandenhende, Stamatios Georgoulis, Wouter Van Gansbeke, Marc Proesmans, Dengxin Dai, and Luc Van Gool. Multi-task learning for dense prediction tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2021.

[22] Yi Jie Wong, Mau-Luen Tham, Ban-Hoe Kwan, Ezra Morris Abraham Gnanamuthu, and Yasunori Owada. An optimized multi-task learning model for disaster classification and victim detection in federated learning environments. *IEEE Access*, 10:115930–115944, 2022.

[23] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, July 2017.

[24] András L Majdik, Charles Till, and Davide Scaramuzza. The zurich urban micro aerial vehicle dataset. *The International Journal of Robotics Research*, 36(3):269–273, March 2017.

[25] Ziyang Hong, Yvan Petillot, Andrew Wallace, and Sen Wang. Radar slam: A robust slam system for all weather conditions, 2021.