SLAM and Object Detection for Search and Rescue Using UAVs in Indoor Environments

Bryant Hong, Department of Computer Science
Mentor: Dr. Subodh Bhandari
Kellogg Honors College Capstone Project

Abstract

Innovations in unmanned aerial vehicle (UAV) technology have grown significantly to reduce the need for traditionally human-operated vehicles to perform dull and dangerous tasks. Due to their lower cost, agility, and small size, UAVs can be cost-effectively used in many applications including search and rescue operations in indoor environments during natural disasters. This project discusses about the research being done for the three-dimensional (3-D) mapping of GPS-denied indoor environments for autonomous navigation of UAVs using Simultaneous Localization and Mapping (SLAM) algorithms. SLAM refers to a method of generating a map of an unknown environment by detecting distinct features in that environment while being able to keep track of where the vehicle running this algorithm has been within the same map. Intel's RealSense Visual SLAM algorithm was used in this research due to its ability to produce accurate 3-D maps of indoor environments with Inertial Measurement Unit (IMU) data and computer vision techniques. Target detection was accomplished using the NVIDIA Inference platform and You Only Look Once (YOLOv3) deep learning algorithm to accurately detect and classify objects in the view of the onboard camera. This project will show the results of these techniques including the results in simulation environments and the results of individual component testing. Accuracy and limitations of these algorithms will be discussed. Once found satisfactory in ground testing, the algorithms will be tested in flight.

Background

UAVs have been gaining popularity these past few years and they have been getting more sophisticated than ever before. UAVs can be used for many applications such as preventing natural disasters, agriculture, and rescue missions. Currently UAVs use Global Positioning System (GPS) as the main source of information to know where the plane is in real time. However, GPS is not always reliable and, in some cases, futile if the environment is obstructed, collapsed or remote. The purpose of this research is to evaluate different SLAM algorithms to effectively perform autonomous flight.

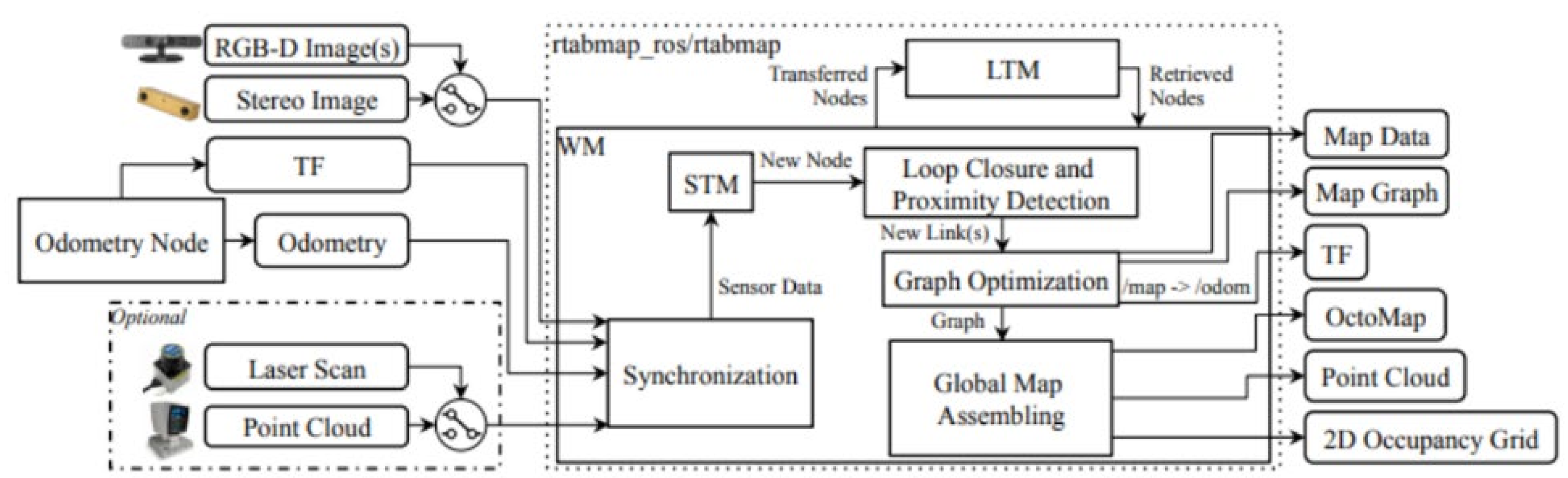
In the past, SLAM algorithms could only work with 2D environments because equations at the time were not sophisticated enough. Probabilistic distribution was the first method to solve the SLAM problem. This method uses data that is produced by noise and attitude caused by a robot in motion to solve location but is unable to produce a map. SLAM robots initially used laser and sonar as the main onboard sensors to produce a map but because the information provided by these sensors was underwhelming, the performance of the SLAM algorithm was restricted. 3D mapping would not come along until particle filtering was used to solve the SLAM algorithm. However, this method uses a lot of data and information from its environment which means that most computers using this method are overloaded and are unable to produce a map in real time. Kalman filtering was applied in the late 2000s to solve the issue of abundant information. Kalman filtering handles the noise produced by a moving robot then uses multiple inputs from the robot and uses these inputs to then provide a least squared error of a process state. Visual imagery then became popular in combination with SLAM algorithms because they provide much more information that is sent to the algorithm to display a much more accurate map.

SLAM has proven itself to be an effective technique to solve autonomous navigation without the use of GPS. However, further research needs to be done to make SLAM more accurate as it is not the perfect solution to autonomy. The results presented in this paper come from the latest SLAM algorithms developed using visual imagery from cameras as the only sensor onboard. As mentioned before, a drone with SLAM implemented in its software does not require GPS and can produce its own map (indoor or outdoor) in real time that can be sent to a ground control station. This information can be used in applications such as search and rescue (SAR), land surveying and many others. In addition to SLAM being used to being used for generating maps in the surrounding area during SAR, object detecting algorithms can be used to identify objects, specifically humans, in the area, and pinpoint them in the map area.

SLAM Algorithm

RTAB-Map was mainly observed throughout this research due to its ability to produce accurate 3D maps using stereoscopic images. As opposed to other SLAM algorithms, RTAB-Map can identify most unique features in the surrounding environment and save it throughout its map generation process. Considering the targeted application being for drone operation, RTAB-Map was most suitable for its ability to adjust its detection rate to accommodate for drone travel speed (faster travel speeds require faster detection rates). In terms of performance, using a graph-based approach allows detected features to be efficiently created and processed as nodes which are subject to a weighting mechanism that ranks their importance (based on frequency) to map creation and disposes of low-ranking nodes for optimization.

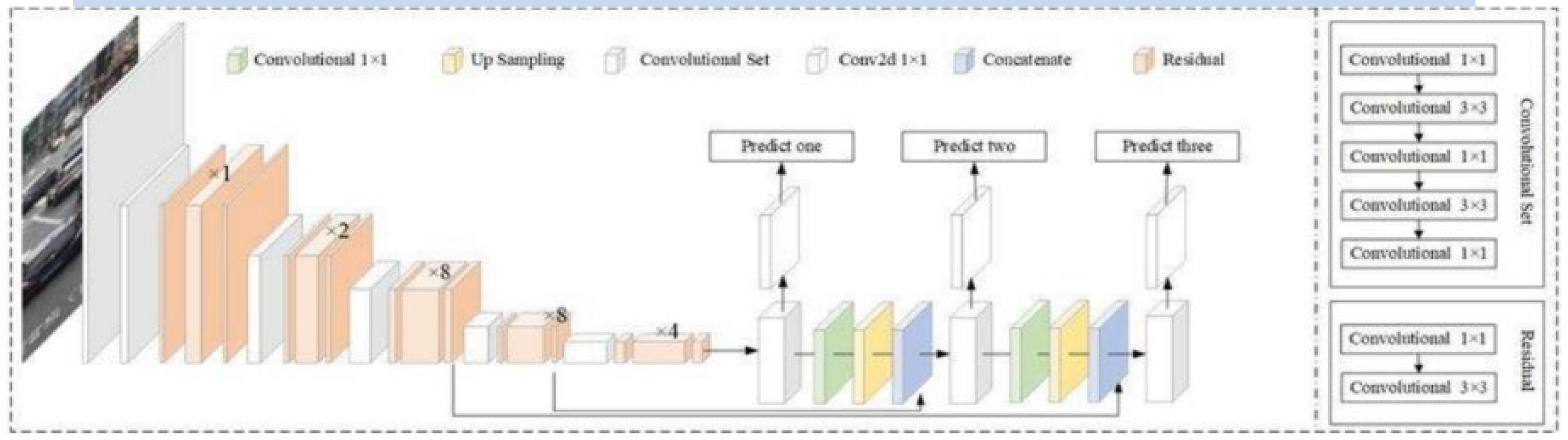
The algorithm first starts with a stereo image is fed into working memory (WM). Within the WM, images are converted into sensor data which is passed into short-term memory (STM) and becomes a new node in the system. The newly acquired node is evaluated using loop closing detection and if it satisfies this condition, it will be transferred into the new map data as new point cloud information for the final 3D map output. Once a node successfully finishes its cycle, it gets transferred into long-term memory and remains until a loop closure is detected.



Object Detection Algorithm

YOLOv3 is an object detection algorithm that is built using a Convolutional Neural Network (CNN) capable of identifying and classifying objects in real-time. CNN's function as classifier-based systems that process inputted images by converting them to structured arrays of data to identify patterns between them. YOLO has gained its popularity for its ability to quickly output detections while maintaining accuracy, which is an important consideration when requiring quick turnarounds during drone operation.

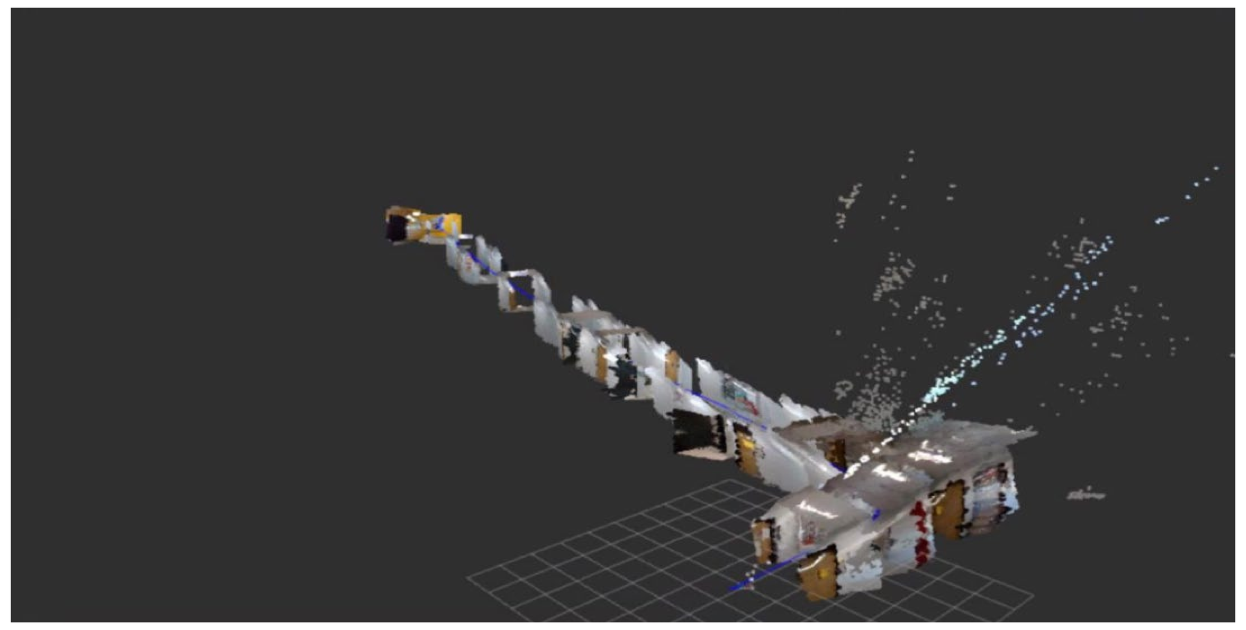
YOLOv3 first divides the image into a grid and each grid cell is responsible for predicting some number of boundary boxes surrounding objects that have a high confidence value corresponding to an already predefined class that the model is trained to recognize. These bounding boxes are generated based on the dimensions of the highly graded boxes and are compared to the algorithm's dataset to find commonalities in shapes and sizes to produce a classification for the object.

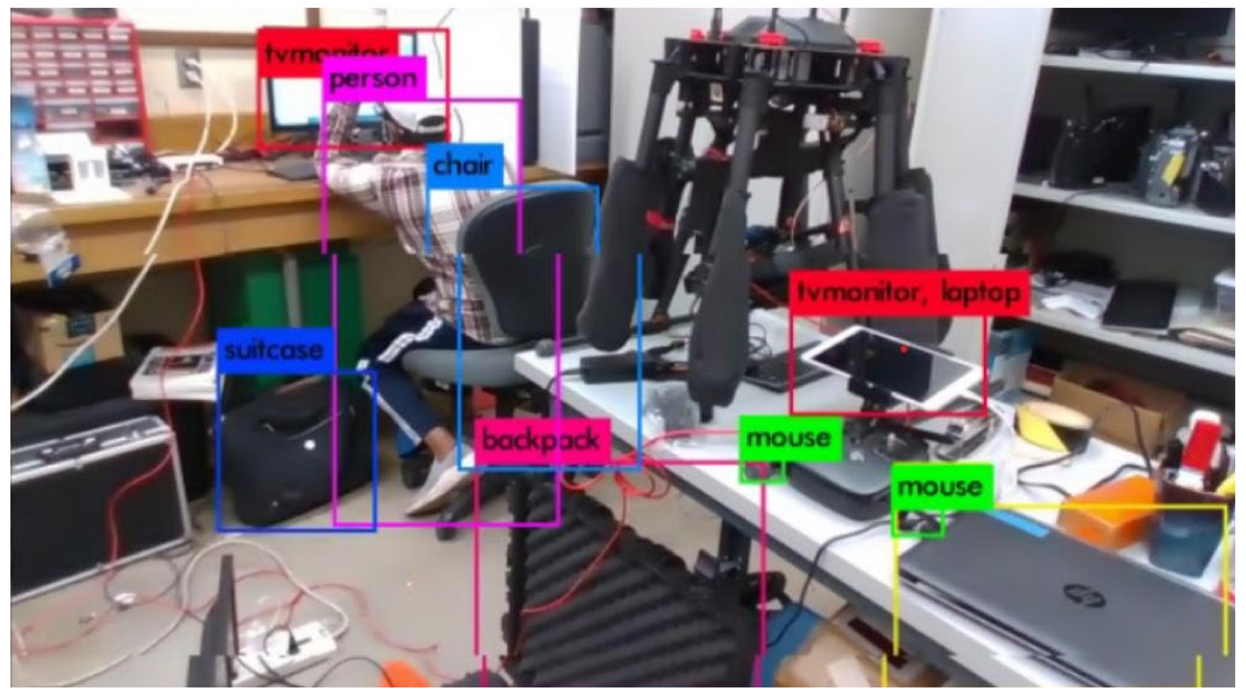


Results and Conclusions

The map produced by RTAB-Map proved to create an accurate 3-D map of indoor environments but had noticeable shortcomings with mapping featureless areas such as blank walls. Mapping was performed while slowly navigating in a narrow hallway environment and was successfully completed except for featureless walls being omitted towards the end of the long hallway segment of the map. Room for improvements can be made in the drone's computing hardware to better handle mapping for long durations since it was common for the hardware to reach high operating temperatures.

YOLOv3 performed well in both indoor and outdoor environments with considerations for lighting and the number of targets within view. Even with the limitation of two bounding boxes per grid cell being more noticeable in cluttered environments, the number and quality of detections are sufficient for indoor search and rescue applications.





References

Labbé, Mathieu, and François Michaud. "RTAB-Map As An Open-Source Lidar And Visual Simultaneous Localization And Mapping Library For Large-Scale And Long-Term Online Operation". Journal Of Field Robotics, vol 36, no. 2, 2018, pp. 416-446. Wiley, doi:10.1002/rob.21831.

Mu, Lili et al. "Research On SLAM Algorithm Of Mobile Robot Based On The Fusion Of 2D Lidar And Depth Camera". IEEE Access, vol 8, 2020, pp. 157628-157642. Institute Of Electrical And Electronics Engineers (IEEE), doi:10.1109/access.2020.3019659.