Semantic Guided Monocular Dynamic Obstacle Avoidance for Aerial System

Jerrin Bright
Researcher, Department of Aerospace Engineering,
Indian Institute of Science, Bangalore, India.

October 2021

Abstract

A certain level of autonomy is imperative to the future of aerial vehicles. The research revolves around a robust real-time aerial system framework that can evade obstacles, plan and navigate in indoor or outdoor environments robustly. The process workflow includes three modules: perception, planning and control. The perception module entails the depth algorithm and semantic segmentation (task specific). Semantic segmentation pixelates dynamic classes and cross-task guidance is established which optimizes the depth map to effectually detect near-range obstacles from the UAV's Field of View (FOV). The planning module includes the configuration space formation, cost function and hybrid A* search for a kinodynamic feasible path. B-Spline optimization and iterative time adjustments find the most feasible path. The control module entails the desired navigation expected from the UAV to evade obstacles and reach the setpoint. Stabilization is achieved using position and altitude PID controllers. The system is planned to test with various operating speeds, and compare with state-of-the-art avoidance algorithms.

Objective

This research was aimed to bring in a computationally efficient and robust aerial navigation system which is capable of running in any environment. Some of the major factors aimed to achieve via this research include Eradicating the need for static world assumptions by pixelating the dynamic classes and the usage of spatial-temporal correspondence to perceive the scene alike humans, Small object avoidance by extrapolating the depth map using semantic guided features, Solving lack of ego-motion resulting in insignificant structural change, Simultaneous kinematics and dynamics constraints resulting in agile motion planning.

Preliminary Literature Review

The sole of the perception module is an extension of SGDepth [8], where they proposed a guided depth segmentation approach derived from the research progression of [3]. Some of the benchmarked depth algorithms evaluated includes monodepth [5] that estimates depth maps from left-right consistency based reconstruction, monodepth2 [6] that works on the downside of [5] thereby reducing visual artifacts and re-projection error, densedepth [1] deals with high resolution depth maps using transfer learning and augmentation, and PyD-net [9] capable to run with CPU in real-time. There has been some interesting research works by [2], [7] and [10] that uses monocular vision to do obstacle avoidance using depth maps and HSV, depth map and collision networks, depth and confidence based pCNN respectively. An kinodynamic path planning [4] version proposed by [11], leverages hybrid A*, B-Spline and iterative time adjustment methods.

Methodology

The proposed system has three major modules - perception, planning and control. In figure 1, the overall architecture of our proposed system is shown.

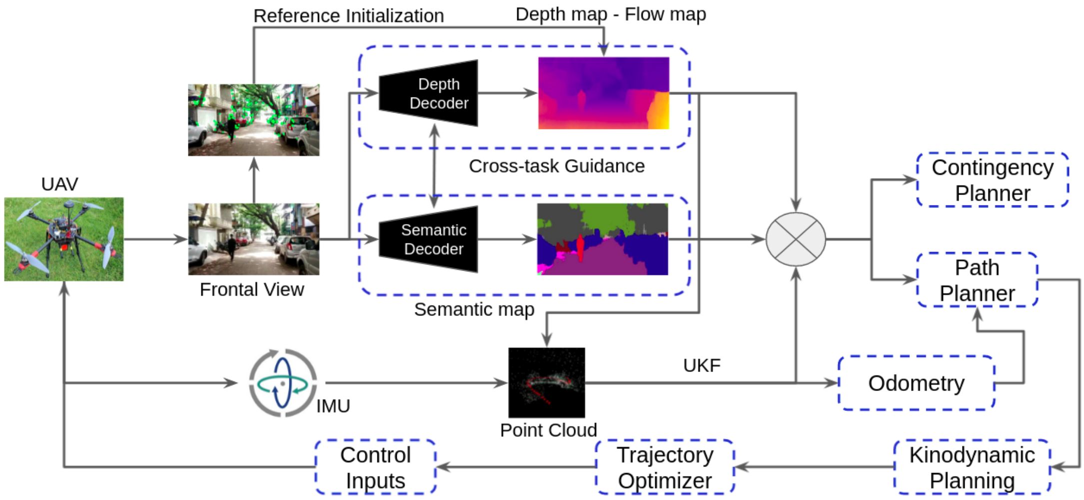

Figure 1: Proposed Architecture

The perception module is split into two sub-modules- depth and semantic map. Real-time Semantic Segmentation is done over the video frames in a supervised manner, which is sent as semantic guidance to the depth encoders, which processes the pixel-wise information, resulting in a depth map (self-supervised). The generated depth maps are processed through spatio-temporal correspondence using matching and optical flow approach to obtain a flow map which are combination of depth information isolated for dynamic classes enabling perception similar to a human.

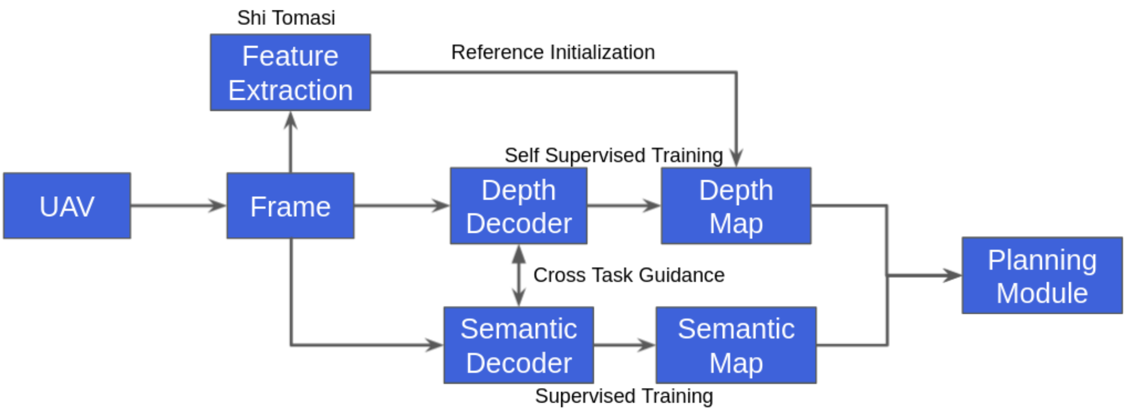

Figure 2: Perception Module

ORB-SLAM based sparse mapping is used to localize the UAV and Unscented Kalman Filters (UKF) is used to fuse the localization data obtained with IMU. ORB features are also used for reference initialisation to determine the absolute scale values of the depth maps obtained in the perception module. Then, point clouds are extrapolated and are stored in a Vector Field Histogram (VFH) on which RRT* path planning algorithm is used to explore the histogram. This in turn finds the most effectual path and way points are generated which is then fed to the trajectory planner to sketch the trajectory for navigation.

Figure 3: Frame - Segmentation - Guided Depth Map

Figure 3 summarizes a clear visualization of the perception module. When obstacles are very close to the drone, the potential of all algorithms deviated catastrophically, but this algorithm uses information from semantic map and guides the depth map, resulting in no deviations.

Figure 4: Depth Map of thin objects (wires)

Figure 4 depicts depth map of thin obstacles like cable wires and branch side which wasn't detected in any of the previous algorithms included [6]. Thus, the semantic guided approach enables detection of thin objects and can be enhanced line detectors algorithms like canny-edge.

Figure 5: Point Cloud from depth map

Scale dependent point clouds were estimated from the semantically guided depth maps obtained in the perception module. The point clouds are assigned voxels in an occupancy grid based on probability using Bayes Rule, thus creating the configuration space of the environment that the UAV perceives. ORB SLAM's localization data is used for global frame initialization of the voxels inside the occupancy grid. Now the map generated will be used by the planning module to find the control points for efficient navigation.

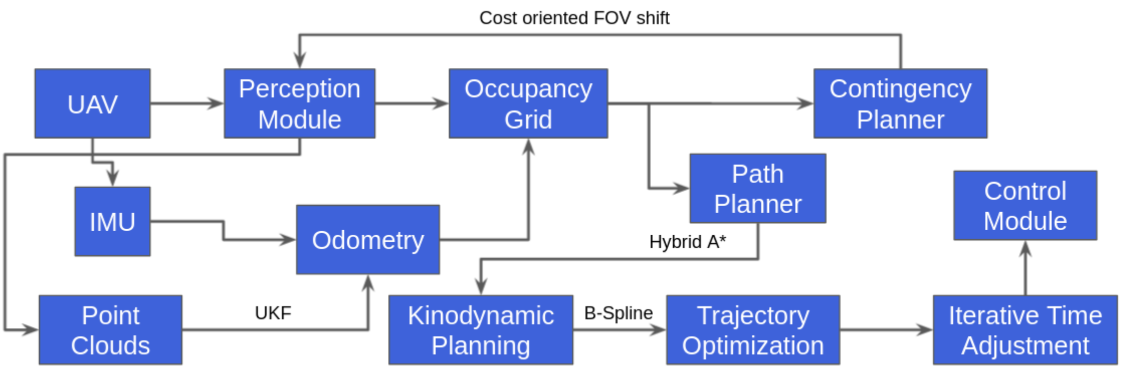

Figure 6: Planning Module

Kinodynamic path planning approach is adapted to navigate from the information perceived from the perception model. B-spline optimization is done to improve the smoothness and clearance of the estimated trajectory. Hybrid State A* algorithm is adapted to search a feasible path between the current pose and desired goal pose as a graph based problem over the generated configuration space. Cost is assigned for each generated trajectory and heuristics enforce faster search.

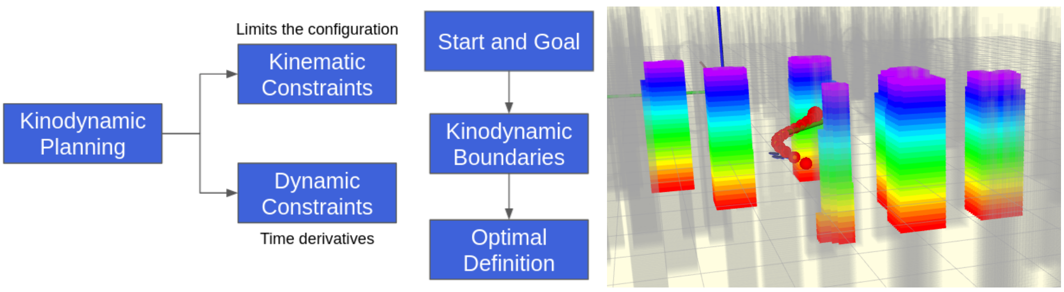

Figure 7: Kinodynamic Path Planning

Deriving the mathematical model of the UAV includes understanding the reference frames, deriving kinematic and dynamic equations of the UAV. The inertial frame, body frame and the vehicle frame upon transformation can be shifted to a common reference frame. Here formulation of equation of motion is done on the body frame considering inertial measurements are in this frame (symmetry of UAV tends to simplify mathematical equations). Euler angles are derived in the vehicular frames and angular rate from body frame.

The control module majorly consist of Model Predictive Controller (MPC) for navigating effectually from the information and commands understood from the perception and the planning modules. Altitude and Position are studied as outputs by the PID controller. Genetic algorithm (optimization and search technique) based tuning is done for the PID controller coefficients.

Preliminary Results

Literature review of stable avoidance algorithms varying from supervised to self-supervised methods using various sensor types are studied. A completely teleoperable quadcopter is made ready (Tarot 680, Jetson Xavier NX, Pixhawk, Logitech C930e). Perception module successfully completed which encompasses depth and semantic map establishing a guided approach. Compared with benchmarked algorithms including [5], [6], [1], [9], etc. and got significantly better results. Currently working on defining mathematical equations for kinodynamic path planning.

Future Works

Some of the most interesting add-ons to the proposed system are:

- Adding optical flow information to the guided depth map thereby making the optical flow map which can not just gives depth information but also gives moving obstacle information.
- The pr-processing module that is the guided depth estimation is planned to be tested with neuromorphic vision sensor (Dynamic and Active-pixel Vision Sensor (DAVIS) Camera). It processes frames at 4000 fps which is much advanced than traditional vision sensors.
- The hybrid A* method is planned to be replaced using a control barrier function which is computationally cheap and helps in reducing latency.
- Multi-trajectory formation using monte-carlo sampling techniques like Metropolis-Hastings Sampling, after which based on cost for control points, navigation trajectory can be chosen.

Conclusion

A robust and agile aerial system framework that can sense and avoid obstacles in a robust manner in terms of computation and accuracy. Texture independent system, lack of ego motion, small object detection and avoidance, fast and agile navigation are some imperative results of this robust system proposed via this research. The algorithm leverages guided segmentation approach to derive steadier depth maps and further extended to detect small objects including thin wires and poles. These small objects are marked as high potential regions and the UAV moves away from its direction (considering the depth information for small objects is not very reliable). As per my knowledge, this extension of segmentation map and the use of guided perception module is the first to ever be implemented in an avoidance system (aerial and ground). Hybrid A* search for the most feasible kinodynamic path enhanced by B-Spline and iterative time adjustment optimizations in the devised configuration space with an adjusted cost map entails the planning module. The system can be efficiently used in a lot of applications including search operations in disaster prone constructions, forests, delivery purposes in high risk environments, etc. to name a few.

References

[1] Ibraheem Alhashim and Peter Wonka. High quality monocular depth estimation via transfer learning. *CoRR*, abs/1812.11941, 2018.

[2] Kumar Bipin, Vishakh Duggal, and Madhava Krishna. Autonomous navigation of generic monocular quadcopter in natural environment. *IEEE International Conference on Robotics and Automation*, 2015:1063–1070, 06 2015.

[3] Po-Yi Chen and Yu-Chiang Frank Liu, Alexander H.and Wang. Towards scene understanding: Unsupervised monocular depth estimation with semantic-aware representation. June 2019.

[4] Bruce Donald, Patrick Xavier, John Canny, and John Reif. Kinodynamic motion planning. *J. ACM*, 40(5):1048–1066, November 1993.

[5] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency. *CoRR*, abs/1609.03677, 2016.

[6] Clément Godard, Oisin Mac Aodha, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. *CoRR*, abs/1806.01260, 2018.

[7] Kyle Hatch, John Mern, and Mykel J. Kochenderfer. Obstacle avoidance using a monocular camera. *CoRR*, abs/2012.01608, 2020.

[8] Marvin Klingner, Jan-Aike Termöhlen, Jonas Mikolajczyk, and Tim Fingscheidt. Self-supervised monocular depth estimation: Solving the dynamic object problem by semantic guidance. *CoRR*, abs/2007.06936, 2020.

[9] Matteo Poggi, Filippo Aleotti, Fabio Tosi, and Stefano Mattoccia. Towards real-time unsupervised monocular depth estimation on CPU. *CoRR*, abs/1806.11430, 2018.

[10] Xin Yang, Jingyu Chen, Yuanjie Dang, Hongcheng Luo, and Kwang-Ting. Fast depth prediction and obstacle avoidance on a monocular drone using probabilistic convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 22(1):156–167, 2021.

[11] Boyu Zhou, Chuhao, and Shaojie Shen. Robust and efficient quadrotor trajectory generation for fast autonomous flight. *IEEE Robotics and Automation Letters*, 4(4):3529–3536, 2019.