Unmanned Obstacle Avoidance using Monocular Vision

Carter Vavra, Jennifer Pedraza, Sarah Evans

December 2023

Abstract

In this research project, our objective is to implement automated obstacle avoidance for drone navigation. We address obstacles both within the drone's camera view and those that have moved out of sight, assuming static object conditions. Our methodology employs monocular depth estimation to ascertain the distance of objects from the camera. The identification of approaching obstacles is based on positional data gathered across multiple frames. While the development of a model for avoiding obstacles outside the camera's view is ongoing, we have successfully built a depth map using Intel's MiDaS monocular depth estimation models.

1 Introduction

The proliferation of unmanned aerial vehicles (UAVs) has underscored the need for robust obstacle avoidance systems to ensure safe and efficient drone navigation. In this research project, our primary objective is to implement automated obstacle avoidance, addressing challenges posed by obstacles both and within and outside the drone's immediate visual field. We operate under the assumption of state object conditions, employing monocular depth estimation, specifically Intel's MiDaS models, to discern object distances. The identification of approaching obstacles is based on the meticulous analysis of positional data across multiple frames. While our focus remains on navigating obstacles within the camera's view, we concurrently strive to develop a comprehensive model for obstacles within the camera's view, we concurrently strive to develop a comprehensive model for obstacle avoidance beyond this perspective. The successful construction of a depth map stands as a noteworthy achievement in our ongoing pursuit of advancing drone navigation capabilities.

From our research, the DJI Tello does not have advanced obstacle avoidance features like some higher-end DJI drones. Instead, it relies on its sensors for basic stability and control. The Tello drone is equipped with downward-facing vision sensors that help it maintain a stable position when flying indoors and close to the ground. Obstacle avoidance technology typically involves the use of sensors such as cameras, ultrasonic sensors, and infrared sensors to detect obstacles in the drone's path and adjust its flight accordingly. However, the Tello Drone's obstacle avoidance capabilities are limited compared to more advanced drones in the DJI lineup.

2 Methods

2.1 Risk Calculation for in-view Geometry

To achieve unmanned obstacle avoidance, we adopted a multi-step approach leveraging monocular depth estimation and risk assessment. The primary method involved generating a depth map using the Intel MiDaS small model, providing essential information on the distance of objects from the drone's camera.

Subsequently, a "depth mask" was generated, incorporating parameters such as the drone's speed and velocity vector. This mask focuses on a cropped region, excluding values outside the designated area. The cropped region is then utilized to calculate the "risk" of the drone colliding with objects through a risk calculation process. This is necessary to avoid inaccurate risk calculations via pixels with high depth values which do not pose a threat to the viewer. In the following figure, we see an image whose depth value approaches the 50% threshold despite posing no risk of collision with the camera.

The risk calculation process involves two key components: the MiDaS-Small engine, responsible for rendering the initial depth image from a monocular image buffer, and the Risk Calculation Threshold (RCT). The depth image, created by the MiDaS engine, served as the basis for the depth mask generation. An average of the values in the depth buffer (0.0-1.0) are computed against the RCT. The RCT, specifically for objects in the drone's immediate view, is currently set at a preliminary threshold of 50% (or 0.5), subject to further testing and optimization for varying scenarios.

It is important to note that the depth map generation and risk calculation focused on obstacles within the drone's immediate visual field. Ongoing efforts include the development of a comprehensive model for obstacle avoidance beyond this perspective, exploring point cloud generation and worldspace translation via motion vectors. However, due to hardware limitations and timeline constraints, a testable prototype for this extended model is currently under development.

The experimental setup involved testing the obstacle avoidance system in controlled environments, considering different speeds and scenarios to assess the reliability and effectiveness of the proposed methodology.

Parameters:

- *Drone Speed*: Determines the size of the depth mask
- *Drone Velocity*: Determines the position of the depth mask
- *Risk Calculation Threshold (RCT):* 50% (subject to further testing)

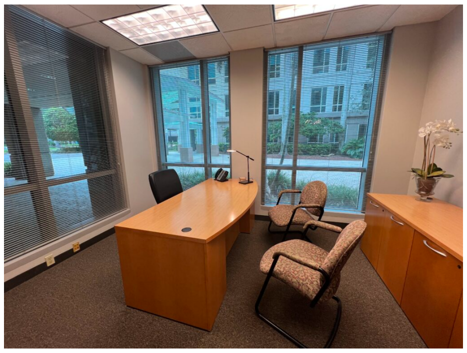

Figure 1: Sample image.

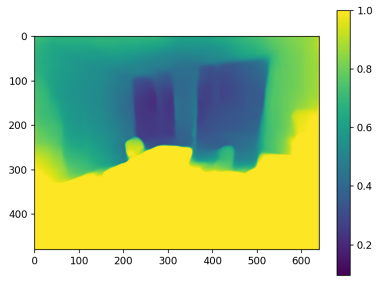

Figure 2: Depth map of sample image. Average pixel value: 73.2%

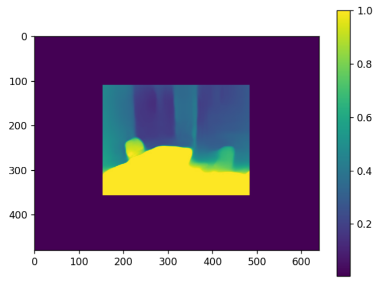

Figure 3: Depth map with static mask. Average pixel value: 37.0%

This methodology lays the foundation for a streamlined obstacle avoidance system; however, it is limited by the field of view of an unmanned drone with limited sensory input. In order to develop robust obstacle avoidance that relies on a single camera, a system of object remembrance is necessary in order to avoid objects no longer in monocular view.

2.2 Object Remembrance with Point Clouds

To optimize for the limited input criterion, we aim to explore point cloud generation for preserving information about out-of-view obstacles. This will involve further research into the implementation of point cloud technology and its integration into our obstacle avoidance system. Point cloud generation serves as a means of object remembrance, allowing the drone to navigate more effectively in dynamic environments beyond its immediate visual field.

We have successfully realized the generation of point clouds through monocular depth maps, employing the Intel MiDaS-Small MDE engine and Open3D; however, the intricate implementation of point clouds, specifically the translation of the point cloud relative to drone movement—a pivotal aspect for out-of-view obstacle avoidance—is currently in progress, primarily due to time constraints which disallowed exhaustive calibration testing. Consequently, the definitive effectiveness of point clouds in the realm of object avoidance remains inconclusive at this juncture.

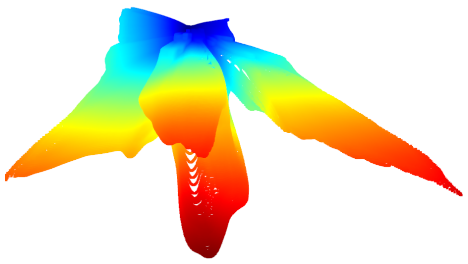
Figure 4. Point cloud generated from the image below.

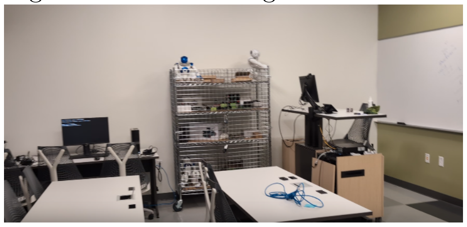
Figure 5. Sample image used for point cloud generation.

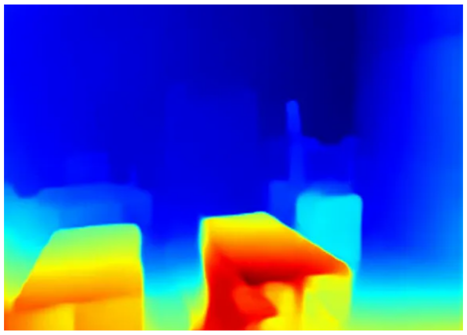

Figure 6. Depth map generated via MiDaS-Small before point cloud projection.

3 Implementation

3.1 Drone Communication (Async / Parallel)

We initially used the DJITelloPy library for Python for Tello drone communication; however, due to technical issues in the library that prevented error diagnostics and deterministic feedback, we developed a new implementation of the Tello protocol, allowing us to troubleshoot issues with both the original API and the drone hardware.

3.2 Asynchronous Video Stream (Async)

The Tello protocol sends video feedback from a drone to network port 11111, which is consumed by a UDP client. This image data is then sent to other software for image analysis.

The capabilities of both our library and the DJITelloPy API were affected due to various issues with video feedback. Much of the time with the drones was spent troubleshooting these issues. By the time of this report, we have yet to isolate any issues with the video feedback stream, and rely on an example script provided by the project supervisor to achieve stable video streaming via the DJITelloPy API.

Possible endemics of such issues include undocumented technical requirements for consuming video feedback as well as the usage of Tello EDU drones, which have been discontinued, and may thus affect the relevance of Tello documentation details.

3.3 Depth Estimation with MiDaS (Async / Batched)

Frame data is sent from the video feedback port to an external program which consumes the image data and performs monocular analysis to generate depth maps.

The depth maps generated by MiDaS are masked based on parameters mentioned in Section 2.

The unmasked depth maps are also sent to an external process for real-time point cloud generation.

3.4 Rendering (Async)

All image data is rendered for visual image analysis to pinpoint technical issues and software bugs related to our analysis software, and does not affect the responsiveness of the monocular image analysis pipeline.

4 Deprecated Methods

4.1 Object Remembrance via Neural Radiance Fields (NeRFs)

NeRFs make it easy to generate 3D scenes using a series of 2D monocular images. NeRF is largely applied in AR technologies, though we sought to leverage its capabilities for out-of-view object remembrance.

Initially, we had intended to leverage the NeRF model to generate a 3D scene of the room following scans aerial performed by a Tello drone; however, the limitations of our hardware – largely due to video RAM bottleneck – disallowed real-time scene generation, even for the lowest-quality NeRF models. As a result, we elected to find a more lightweight method of scene generation, which led to our developments towards point cloud generation.

4.2 Monocular Depth Estimation using Keras

We initially developed depth maps using Keras for Python; however, despite the effectiveness of our model, we could not get the model to perform on a machine with CUDA capabilities.

5 Results

The results of our research and implementation efforts are presented below, highlighting key findings and outcomes in the development of an unmanned obstacle avoidance system for the DJI Tello drone.

5.1 Obstacle Avoidance within Visual Field

The simplistic nature of risk calculation via monocular depth estimation makes visible obstacle avoidance trivial; however, calibration of the RCT and the depth mask still require extensive testing maximize the effectiveness of the MDE model.

5.2 Object Remembrance with Point Clouds

In our pursuit of comprehensive obstacle avoidance beyond the drone's visual field, we explored the implementation of point cloud technology for object remembrance. Figure 4 showcases a point cloud generated from a sample image, indicating progress in preserving information about out-of-view obstacles. The calibration of the parameters necessary to perfect the point cloud maps are still undergoing, and until then, the point cloud model is ineffective for object remembrance.

Furthermore, the translation of point clouds relative to drone movement, a critical aspect for out-of-view obstacle avoidance, is still in progress. Time constraints have limited exhaustive calibration testing, leaving the definitive effectiveness of point clouds in object avoidance inconclusive. Ongoing efforts focus on refining this aspect and conducting further experiments to validate the functionality of point clouds in dynamic environments.

5.3 Implementation Challenges and Solutions

Our implementation faced challenges, particularly in the areas of drone communication and asynchronous video streaming. Technical issues with the DJITelloPy library prompted the development of a new implementation of the Tello protocol, allowing for more effective troubleshooting. The challenges related to video feedback stream stability were addressed through collaboration with the project supervisor and reliance on a stable example script.

Despite these challenges, our depth estimation using Intel MiDaS, proves to be a robust component of our obstacle avoidance system by streamlining the generation of depth maps in real-time. Real-time computation of depth maps, risk assessment, and point cloud generation demonstrated the responsiveness and efficiency of our approach.

6 Conclusion

In conclusion, our research lays the groundwork for unmanned obstacle avoidance and remembrance for monocular sensory systems. Further testing is necessary to maximize the potential of our obstacle remembrance and avoidance model.

7 Future Goals

The results obtained from our research and implementation provide a solid foundation for the development of unmanned obstacle avoidance systems. Future directions include:

- Optimization and Testing: Further optimization of parameters, especially the Risk Calculation Threshold, and extensive testing in diverse scenarios to enhance the system's adaptability.

- Point Cloud Generation and Translation: Continued work on both parameter calibration to perfect the generation of point clouds, and refining the translation of point clouds relative to drone movement for effective out-of-view obstacle avoidance.
- Real-world Testing: Conducting real-world testing to validate the system's performance in dynamic and uncontrolled environments.

  In conclusion, our research project has made significant strides in the development of an obstacle avoidance system for the DJI Tello drone, laying the groundwork for safer and more efficient drone navigation in complex surroundings.

8 References and Sources

Stanford University. (2022). Depth map generation via monocular depth estimation. https://web.stanford.edu/class/cs231a/course_notes/08-monocular_depth_estimation.pdf

ISL - Intelligent Systems Lab. MiDaS: Monocular Depth Estimation using a Single Image. https://github.com/isl-org/MiDaS

Open3D. PointCloud - Open3D 0.13.1 documentation. http://www.open3d.org/docs/release/tutorial/geometry/pointcloud.html

Tancik, M. (2020). Neural Radiance Fields. https://www.matthewtancik.com/nerf

Keras. Depth estimation from a single image using the Keras example. https://keras.io/examples/vision/depth_estimation/

Ryze Tech. Tello SDK 2.0 User Guide. https://dl-cdn.ryzerobotics.com/downloads/Tello/Tello%20SDK%202.0%20User%20Guide.pdf

DJITelloPy. DJITelloPy Documentation. https://djitellopy.readthedocs.io/en/latest/

Evans, Vavra. tello-edu-py. https://github.com/zaruhev/tello-edu-py

9 Acknowledgments

Thank you Tomai, our Sole Proprietor, for providing us the ability to be able to work with the DJI Tello Drone and conduct research allowing us to gain insight on a different field in Computer Science.