Computer Vision Guided Autonomous Docking Rover

By David Lackner, ECE '19

Introduction

A typical stereo camera consists of two lenses with two distinct photo sensors. Comparing scene information from two separate vantage points allows for depth measurements of objects. Embedded graphic processing units (GPU) have led to faster image processing algorithms, reducing the latency error from camera input to control output. Stereo cameras are a viable option in autonomous vehicle navigation. This paper explorers the application of stereo cameras in autonomous vehicles, the challenges of stereo camera navigation, and a proposition to obtain 360° sensory input using multiple stereo cameras.

Stereoscopic Cameras and Robotic Navigation

"*Stereopsis* meaning *solid-appearance*, is the perception of a 3-dimensional object using visual input derived from binocular vision." [9] The visual cortex of animal's and human brains use the disparity of two slightly different images projected onto the retinas of the eyes to obtain depth information. The human brain is impressively fast at finding corresponding points in our left and right eyes. Evolution and years of practice have trained our brains to calculate this disparity and determine how far away an object is from the viewer.

Obtaining 3 dimensional measurements in a camera system is accomplished using a similar algorithm. Stereo vision requires two photo sensors to be placed a defined distance from one another. Light from an object of interest enters through the pinhole cameras and is absorbed by a photo sensor. Each photo sensor receives slightly different images. An onboard processor uses digital signal processing to find corresponding pixels from one image to another. Simple Euclidian geometry is then used to estimate the distance to the object of interest.

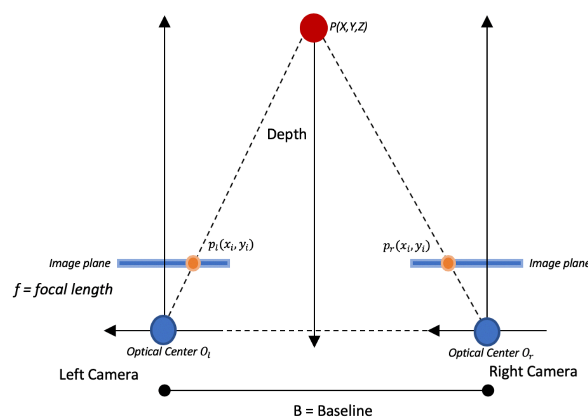

Fig 1: Stereo Camera Geometry

The concept is simple in theory, but requires a solid understanding of three key concept:

1. *Image acquisition*. Every pixel in an image defines a 3-dimensional ray. If we identify a real-world point in both images, the 3-dimensional position can be found through triangulation outlined in Figure 1.

2. *Camera Calibration*. To find matched points in two images, the cameras must be calibrated and the distance between lenses must be known. Imperfections in the manufacturing of lenses, PCB boards, and semiconductor doping can introduce variations in each photo sensor. Stereo cameras must be calibrated, and these discrepancies must be factored into the digital signal processing of the image. Without calibration, measurements can be significantly skewed.

3. *Image Processing*. The correspondence problem refers to the problem of determining which parts of one image correspond to which parts of another image. Differences between the two images are due the camera's movement, elapsed time, or movement of objects within the photos. This is a computationally expensive task and can be very slow. However, techniques such as Epipolar geometry [13] can narrow the search window. Instead of searching across the entire image for similar pixels, the scope is narrowed down to a one-dimensional array of pixels. This process can be expedited by taking advantage of GPU's. GPU's take specialized workloads and make them easier. Image processing requires a significant amount of 3-dimensional matrix arithmetic, 3-dimensional plane equations and matrix solving. GPU's approach the problem as a throughput problem and use their parallel processing capabilities to execute a large number of calculations quickly.

Stereo Camera Application

Autonomous vehicles rely on a robust sensor package to feed scene input into an extended Kalman filter. Sensor data can come from a variety of sources. The goal of these sensor packages is to obtain information about the vehicle's surroundings so that it can react to its environment.

The term *autonomous vehicle* is typically associated with passenger cars and roadable vehicles. However, stereo cameras can be used to navigate a variety of vehicles including submarines, boats, swarm bots, and fixed wing aircrafts.

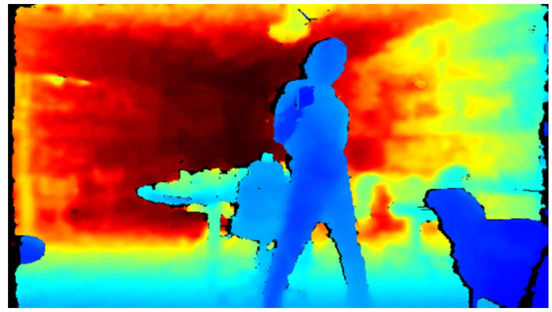

Fig 2: Stereo Camera image with post processing of Tufts University Senior Design Lab

The key advantage of stereo cameras in autonomous navigation is that we obtain two data inputs for the price of one: scene information and distance. Not only do stereo cameras allow for a system to identify an object, they also can provide localization of the autonomous vehicle in space.

[11]

Figure 3: Intel Real Sense D435 Stereo Camera

Stereo cameras allow autonomous vehicles to avoid obstacles and path plan. Figure 2 is the video output of Figure 3, Intel Real Sense D435 Stereo Camera. The video input is fed into an embedded processing unit (Nvidia Jetson TX1) where post image processing and coloring is overlaid onto the image. Blue coloring represents objects that are close to the vehicle. Red coloring represents objects that are further away from the autonomous vehicle. As the autonomous vehicle moves through the scene two maps are created; a global map and a local map. The global map is the initial path or waypoints the autonomous vehicle must take in order to reach the target goal.

In the case of our Senior Design Project: *Computer Vision Guided Autonomous Docking Rover,* the target would be a scaled version of Terrafugia's TF-2 Aircraft [10]. The autonomous vehicle uses the local map to react to the immediate environment around it; for example, a person walking in front of the vehicle. Drones, boats, and other autonomous vehicles can utilize stereo cameras in similar method to navigate their environment.

Challenges

There are a variety of challenges facing stereo cameras that must be addressed before they can be implemented ubiquitously across autonomous vehicles. The issues that needs to be overcome are as follows:

1. *Environmental Conditions*

Think back to the last time you were driving a vehicle through a rain storm or on a foggy morning. It is difficult to see the road. Objects that you usually rely on for navigation, such as lane markers or signs, may be unavailable. Computer vision navigation requires fair lighting conditions. Environmental factors that can impact navigation include clouds, sleet, snow, rain, murky water, and many more.

2. *Efficient image processing for real time analysis*

Autonomous vehicles will need to navigate and react to their environments at high speeds. Systems will need to acquire scene information, analyze images, and output a control action in time to avoid an obstacle, navigate a road, or land a quadcopter. Embedded GPU's are tackling this problem. As computer vision continues to expand into a variety of applications (including autonomous vehicle navigation), the need for parallel processing units with significant off-chip memory, and high-speed buses to maximize data rate transfers will be required.

3. *Difficulty in non – illuminated and varying light conditions*

In low light conditions, stereo cameras struggle to produce reliable depth estimates. Matching pixels in dark scenes can cause false positives. In addition, high intensity light aimed directly into the photo sensors can overwhelm the sensors causing them to become useless.

360° Stereo Camera System for Autonomous vehicles

One of the principle advantages of LIDAR is that it is able to obtain a 360-degree map of its surrounding. Stereo cameras are limited by its field of view. However, this does not mean that a collection of stereo cameras and software cannot compensate for this disparity.

The model outlined in Figure 4 is a proposition for an autonomous driving rover to obtain 360-degree view of its environment using four stereo cameras, each with its own embedded GPU processing unit. The computational load is distributed across four units allowing for faster processing and lower latency. Control commands are then sent via the MAVLINK protocol to a centralized Kalman Filter.

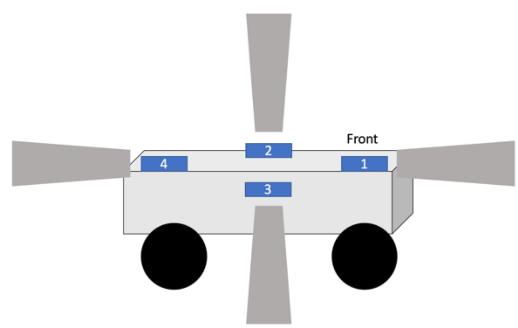

Figure 4: 360° Stereo Camera System for Autonomous Rover

Conclusion

Stereo cameras are a viable option for autonomous vehicles but still have limitations. Before stereo cameras can replace traditional sensor packages, we must address resolution, environmental conditions, and latency issues.

Our semester long Senior Design Project: *Computer Vision Guided Autonomous Docking Rover* is using a stereo camera to dock a 1/10th scale rover to a Terrafugia TF-2 flight vehicle. The rover is using OpenCV, ROS, and MAVROS to locate a scaled TF-2 flight vehicle and autonomously dock. Docking with the TF-2 requires repeatable precision.

Humans do not have the precision to accomplish this task. Using stereo cameras to dock a rover in an efficient and reliable process removes the need for human operators and reduces the risk to passengers in the docking process.

This research topic has provided insight into the challenges and limitations of stereo cameras.

References

1. Z. Cook, Lin Zhao, J. Lee and Woosoon Yim, "Unmanned aerial vehicle for hot-spot avoidance with stereo FLIR cameras," *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence*

2. Rostam Affendi Hamzah, Hasrul Nisham Rosly and S. Hamid, "An obstacle detection and avoidance of a mobile robot with stereo vision camera," *2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, Kuala Lumpur, 2011, pp. 104-108.

3. J. Kim, J. Yoo and J. Koo, "Road and Lane Detection Using Stereo Camera," *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, Shanghai, 2018, pp. 649-652.
   doi: 10.1109/BigComp.2018.00117

4. S. A. Rawashdeh and M. Aladem, "Toward autonomous stereo-vision control of micro aerial vehicles," *2016 IEEE National Aerospace and Electronics Conference (NAECON) and Ohio Innovation Summit (OIS)*, Dayton, OH, 2016, pp. 151-155.

5. A. H. A. Hasan, R. A. Hamzah and M. H. Johar, "Region of Interest in Disparity Mapping for Navigation of Stereo Vision Autonomous Guided Vehicle," *2009 International Conference on Computer Technology and Development*, Kota Kinabalu, 2009, pp. 98-102.
   doi: 10.1109/ICCTD.2009.205

6. L. Pei and Z. Rui, "The Analysis of Stereo Vision 3D Point Cloud Data of Autonomous Vehicle Obstacle Recognition," *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, 2015, pp. 207-210.
   doi: 10.1109/IHMSC.2015.192

7. M. Samadi and M. F. Othman, "Using fish-eye stereo vision system for autonomous vehicles," *2017 11th Asian Control Conference (ASCC)*, Gold Coast, QLD, 2017, pp. 1331-1334.
   doi: 10.1109/ASCC.2017.8287364

8. https://en.wikipedia.org/wiki/Correspondence_problem

9. https://en.wikipedia.org/wiki/Stereopsis

10. https://terrafugia.com/

11. https://realsense.intel.com/stereo/

12. https://enwikipedia.org/wiki/Epipolar_geometry