



Oregon State
UNIVERSITY

College of Engineering

SENIOR DESIGN CAPSTONE FINAL REPORT

JUNE 3, 2018

DEPTH SENSING WITH COMPUTER VISION AND LIDAR

PREPARED FOR

OREGON STATE UNIVERSITY

D. KEVIN MCGRATH

PREPARED BY

GROUP 69

KIN-HO LAM

Abstract

Depth Sensing with Computer Vision and LIDAR proposes combining computer vision and LIDAR to create a reliable depth sensor. This document details its project member's progress toward a final design.

1 TABLE OF CONTENTS

CONTENTS

1 Table of Contents

References

2 Definitions

| | |
|-----|--------------------------------|
| 2.1 | IR |
| 2.2 | IR Depth Sensor |
| 2.3 | LIDAR |
| 2.4 | Microsoft Kinect |
| 2.5 | Logitech Brio Webcam |
| 2.6 | RPLidar A1 |
| 2.7 | Leddar M16 |
| 2.8 | Computer Vision |

3 Introduction

| | |
|-------|---|
| 3.1 | Existing Depth Sensing Technology |
| 3.1.1 | Lidar |
| 3.1.2 | Infrared Depth Sensors |
| 3.1.3 | Radar |
| 3.2 | Project Purpose |

4 Design

| | |
|-----|---------------------------------|
| 4.1 | Physical Mount Design |
| 4.2 | Computer Vision |
| 4.3 | Final System Overview |

5 Cumulative Ten-Week Term Retrospective

| | |
|-----|-----------------------|
| 5.1 | Winter 2018 |
|-----|-----------------------|

REFERENCES

- [1] "Logitech brio webcam with 4k ultra hd video & rightlight 3 with hdr." [Online]. Available: <https://www.logitech.com/en-us/product/brio>
- [2] T. Huang, "Rplidar a1." [Online]. Available: <https://www.slamtec.com/en/lidar/a1>
- [3] "Leddar solid-state lidar: M16 multi-element sensor module." [Online]. Available: <https://leddartech.com/modules/m16-multi-element-sensor-module/>
- [4] "tensorflow/models." [Online]. Available: <https://github.com/tensorflow/models>
- [5] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and et al., "Speed/accuracy trade-offs for modern convolutional object detectors," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [6] "Coco - common objects in context." [Online]. Available: <http://cocodataset.org/>
- [7] "openimages/dataset," Jan 2018. [Online]. Available: <https://github.com/openimages/dataset>
- [8] [Online]. Available: <http://www.cvlibs.net/datasets/kitti/>

[9] "Tensorflow framework & gpu acceleration from nvidia data center." [Online]. Available: <https://www.nvidia.com/en-us/data-center/gpu-accelerated-applications/tensorflow/>

2 DEFINITIONS

2.1 IR

IR refers to the infrared light spectrum.

2.2 IR Depth Sensor

Device that calculates distances by emitting infrared patterns.

2.3 LIDAR

Light Detection And Ranging - Depth sensing technology that uses lasers to measure distance.

2.4 Microsoft Kinect

A product that uses an IR Depth sensor to measure distances.

2.5 Logitech Brio Webcam

A 4k Logitech web cam. [1]

2.6 RPLidar A1

A budget LIDAR device made by Slamtec. [2]

2.7 Leddar M16

A solid-state LIDAR device made by Leddar. [3]

2.8 Computer Vision

The methods for acquiring, processing, analyzing, and classifying digital images and extracting information.

3 INTRODUCTION

Depth Sensing with Computer Vision and LIDAR aims to create a reliable depth sensing system using computer vision and LIDAR. This project spanned 2018 Winter and Spring terms at Oregon State University. I, Kin-Ho Lam, researched, designed, and built this project in its entirety by myself. Much of this project would not have been possible without my project's sponsor, Kevin McGrath, who supplied equipment and technical guidance.

This project's motivation came from an annual robotics competition at Oregon State University, my client noted that all the robots had been designed and tested indoors. This later proved to be an issue because when the robots were brought outside for a competition, their guidance systems failed to work as the IR cameras were inherently subject to interference from the Sun's ambient light. As the prevalence of automation in our daily lives increases, such as autonomous cars or robots, a cheap and scalable depth sensing system that is resistant to common interferences is necessary. The goal of this project is to create a reliable depth sensing system using computer vision and LIDAR. Such a system that combines these two technologies will be more reliable than infrared depth sensing because cameras and LIDAR devices are less prone to interference from ambient light or anomalies.

3.1 Existing Depth Sensing Technology

3.1.1 Lidar

LIDAR stands for light detection and ranging. LIDAR is a distance finding system that uses a laser to determine distance and angle of objects relative to the sensor. Most LIDAR devices use a low powered 600-1000 nm laser which is mounted on a motor to sweep the laser in a circular motion. A receiver is used to measure the time it takes for the laser to bounce back to the sensor upon encountering an object. Commercial LIDAR systems are relatively low-cost. Their ability to work in varying lighting conditions and ability to detect small objects makes them well suited for use in self-driving cars or autonomous robots.

3.1.2 Infrared Depth Sensors

IR, or infrared depth sensors, strobe an infrared pattern on objects in front of it. This infrared pattern is picked up by a receiving camera, which uses the pattern and some basic geometry to calculate distance. While very accurate and cheap, IR depth sensors are not suited for self-driving cars or autonomous robots because they are highly sensitive to light conditions. Two IR depth sensors pointing at the same subject will overlap and confuse each other's sensors. Natural sunlight will also blind the sensor by washing out the IR pattern.

3.1.3 Radar

Radar stands for radio detection and ranging. Radar is an object-detection system that uses radio waves to determine the distance, angle, and velocity of objects relative to the sensor. A radar system consists of a transmitter producing electromagnetic waves, a transmitting antenna, a receiving antenna, and a processing system to interpret the data and determine the locations of objects. Radar works by sending out electromagnetic waves, then measuring the intensity of the reflection. Radar systems are widely used for military applications, airplanes, and weather forecasting. While radar systems have been tested to work extremely reliably in extreme weather conditions, unlike LIDAR, radar cannot detect small objects as easily. This is especially problematic when trying to apply a radar system to an autonomous car system; a radar system may not detect a narrow pole or small object in front of it.

3.2 Project Purpose

Infrared (IR) depth sensors such as the model used in Microsoft's Kinect 2.4 can quickly calculate distances in indoor scenarios. However, IR depth sensors can be confused by other infrared emitting sources such as other IR depth sensors or natural sunlight. For these reasons, IR depth sensors cannot be used in self-driving cars, outdoor robots, or any any device that requires high accuracy distance measurement in varying conditions. Depth Sensing with Computer Vision and LIDAR proposes combining computer-vision image classification with LIDAR to create a robust and reliable depth sensor.

4 DESIGN

The Logitech Brio webcam provides a high-resolution, two-dimensional image but lacks depth perception. The LIDAR provides accurate depth measurement in a horizontal dimension but lacks vertical perspective. This project proposes bridging the utility of both devices by securing them in stationary positions, then using software to combine their outputs. This involves using the M16 LIDAR to get depth sensing information and using computer vision to recognize objects. The result is a scalable and reliable depth sensor that will not be affected by natural light, and can be further improved by training a better computer vision model or adding more sensors. This project hopes to achieve a proof of concept design to be showcased in a live demo at Oregon State University's 2018 Undergraduate engineering expo. This live demo shall consist of the full system pointed at the project booth's audience.

Figure 1 illustrates different dimensions measured by the M16 LIDAR and Brio Webcam. The red cube represents the Logitech Brio webcam and M16 LIDAR secured in stationary positions. The flat purple triangle represents the M16 LIDAR's horizontal range detection. The transparent green rectangle in front of the person represents the computer vision model recognizing that there is a person in-front of the sensor. The transparent teal pyramid represents the Brio webcam's field-of-view.

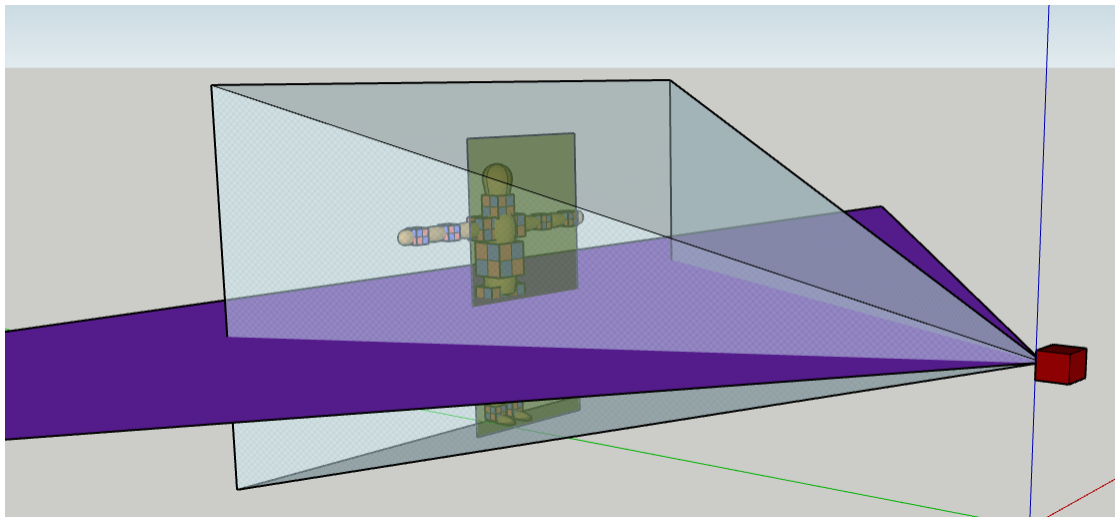


Fig. 1. Visualizing different dimensions measured by the M16 LIDAR and Brio Webcam.

4.1 Physical Mount Design

4.2 Computer Vision

I started with OpenCV's pre-trained facial/pedestrian support-vector-machine (SVM) classifier. This SVM is a combination of several other SVMs that detect the upper body, eyes, mouths, and noses. The combined SVM is intended to

detect faces with high accuracy. However, when applied to my design, I could not consistently replicate good results. This was due to several factors, namely the SVM used was meant to perform classification on still images where the camera's perspective is far from the subject.

My design specifications envision a system that quickly tracks multiple subjects in a crowded expo scenario. In an expo scenario, human subjects will be moving close or away from the camera, unpredictably shifting their positions, and moving in or out of the field-of-view. As seen in 2, the OpenCV SVM model does not perform to my specification. If the human subject were to turn their head or move too quickly, the SVM will have difficulty tracking their body. Additionally, the SVM performs intensive calculations on the computer's CPU, severely limiting the video output's frame-rate and resolution.

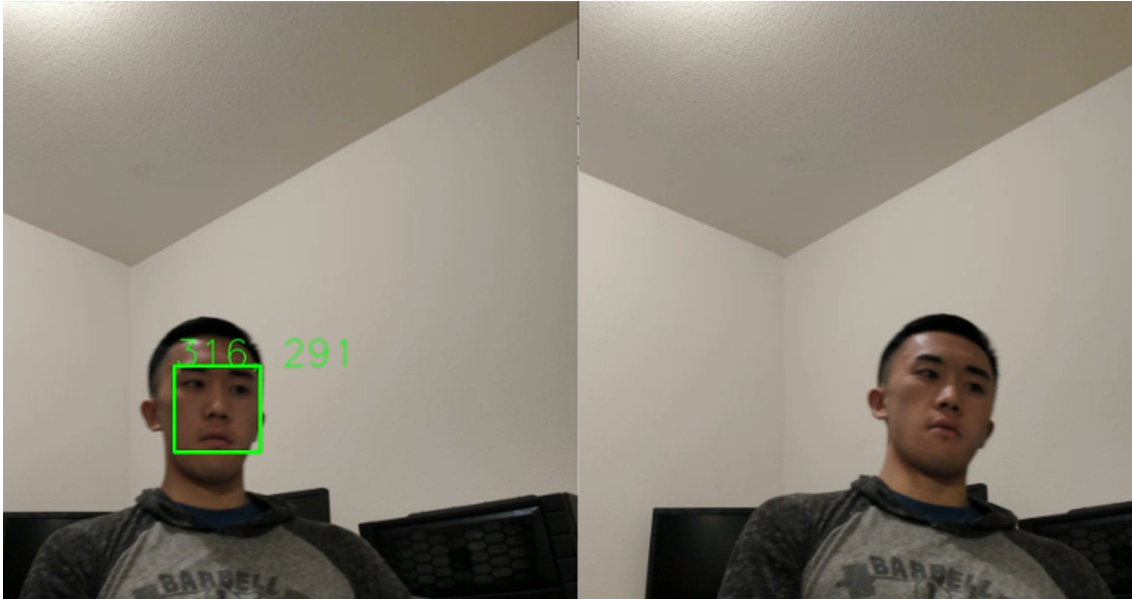


Fig. 2. SVM face classification (Left) fails when subject slightly turns their head (Right)

Recognizing the SVM's weaknesses, Tensorflow's open source object detection classifier presented a better computer-vision alternative. [4] The Tensorflow object recognition library is better suited for this project because its library has already been trained to recognize a large dataset of objects. [5] These pre-trained datasets in Tensorflow's library are sourced from other machine learning datasets including the COCO dataset, Kitti dataset, and the Open Images dataset. [6] [7] [8]



Fig. 3. Pre-trained Tensorflow model detecting multiple subjects

Using this pre-trained Tensorflow model, my project is now able to accurately outline and label over 90 subjects as they come into view of the webcam. At the end of Winter Term, the state of the code will enable selectively editing the output video frames to draw bounding boxes on subjects as they move in and out of the camera's field of view.

Tensorflow also enables us to take advantage of NVIDIA CUDA, a driver that moves intensive calculations to the GPU. While this increases the list of material requisitions for demo, moving calculations to the GPU greatly improves the output video quality, frame rate, resolution, and classification speed. [9]

4.3 Final System Overview

5 CUMULATIVE TEN-WEEK TERM RETROSPECTIVE

5.1 Winter 2018

| Week | Positives | Deltas | Actions |
|------|--|---|--|
| 1 | - | <p>Kim says we're moving too slowly compared to previous teams In my opinion making a website is a far simpler task than writing code that performs analysis and aggregates data in a system that is unproven and has never been done before</p> <p>Kim don't want us to work on filtering acceptable images even through that is what we stated we would do in our problem statement All problem statement drafts focused on filtering acceptable images None of our problem statements alluded to focusing on image analysis to interpret aerosol content</p> | <p>OUR TWO OPTIONS PRESENTED TO US: 1) Kim wants us to focus on a system to interpret aerosol content in the atmosphere based on images of the horizon, EXIF data from those images, and data from wunderground 2) (In her view) continue on this image filtering path we have started, overall this current project goal of a filter is not applicable to Aerolyzer</p> <p>I believe we have not gone astray. What I have written is relevant in either option. I still believe we can accomplish 1).</p> <p>Seems like she gave up on us (in my opinion) we are a lower priority given our pace and low confidence in our abilities The code I wrote is still relevant. It will still pick out the sky which is still very necessary, it will still need to be used no matter the option we choose.</p> <p>I am the only one on Github discussing anything about the project. Neither Logan or Daniel have communicated to me that the direction we were going was wrong I don't think Kim understands the scope of the code that is necessary for a filter or why a filter is necessary I don't think she is reading what I'm documenting on our github or how my code is used.</p> |
| 2 | <p>Meeting with kevin over webex last Friday transcribed notes</p> <p>May or may not have a partner</p> <p>Usable up to 50 meters, 450 spread</p> <p>Take feed from normal webcam and overlay Leddar point cloud on video With minimum range and maximum range Device: Leddar Lydar M16 Webcam: Logitech Brio webcam</p> <p>Most depth cameras don't work well outdoors, light sensors are overwhelmed by ambient light, overlaying point cloud on webcam video may be a way to resolve this problem</p> | - | <p>Research Leddar Lydar M16. 2D Lydar unit, see if I can overlay point cloud on webcam image. Ideas to try: Research dept/distance sensing fpr</p> |

| | | | |
|---|--|--|--|
| 3 | Talked with Lucian over email Unfortunately this time won't work for him, I want to keep this slot so we can at least talk, I'll leave it to him to find time to talk with you Gave him a quick rundown of what I understood the project to be He mentioned he was going to write up a problem statement and design document. | Not sure exactly how to help Lucian writing the documents. | Don't need ROS (Robot operating system, don't need to use it) Camera depth sensor doesn't work in sunlight, accompany with LIDAR unit Use realsense as point of comparison |
| 4 | Borrowed Logitech brio from Kevin Got it working with VLC on Windows 10, set it as capture device. | Alpha stage of project = visually complete interface Responsive code with placeholders for functionality | Experimenting with Brio webcam and opencv vision models. |
| 5 | Interface design? Suggestion: To start, draw red dot on video feed and output Construct internal pipeline first | Lucian and Kevin met in person for our weekly meeting, did not answer webex call so I'm not sure what they discussed. | Sent Kevin my update + screenshots and sketches of my proposed design. |
| 6 | Created opencv python script to overlay green box on webcam video feed Successfully overlaid pixels on camera output Output is stuck on 4:3 resolution, going to work with it now and write code to be scalable, I want to increase ratio to 16:9 sometime, see if it's because I'm running on Linux, try with Lucian's computer later today | So the algorithmic dilemma I'm struggling with: We want to transfer what the lydar detects in physical 3 dimensional space into a 2D representation overlaid on a video. Let this white bar I've drawn here represent an offset in the y dimension (2D space in video) and z dimension (3D physical space). As we move our point of interest to specified distances, this angle needs to remain constant and our y/z offset needs to grow or shrink accordingly. This means our camera and lydar need to be stationary to define some constant angle, once that is done we can just apply some pretty simple math. | Experiment with background subtraction and opencv SVM to capture human faces. |
| 7 | Lucian was talking about using some kind of equation in opencv to get distance, I think we need to get a better illustration to see how Kevin envisions us connecting the webcam feed to the data from the lydar | Sometimes I'm not sure what Lucian means. | Per my conversation with Kevin, I successfully used a trained SVM cascading front-face model to detect faces, not extremely accurate right now, thinking about using this to get face detection Specifically, this is needed because the Lydar will capture objects in an X-Y plane, we need the face detection to have some frame of reference in the Z direction. Our midterm progress report due date got pushed back, I drafted and finalized a few sections. Lucian wrote the other half. |
| 8 | Accidently missed our poster critique session, we're going to attend the extra credit session for credit. | - | Successfully implemented a pre-trained Tensorflow object detection model. This detects full or partial bodies with very high accuracy. After minimal tuning, the machine learning part is effectively done. |

| | | | |
|----|--|---|--|
| 9 | Poster Critique session had the following suggestions: Make sure correct poster format, put figure and captions, put x-y-z scale on main image, change layout, "visually connecting" layout to have presentation flow Doesnt need period at tagline Include contact Include Client | - | Made adjustments to poster design. |
| 10 | Dead week, we're busy writing the Winter report. | Lucian accidentally wiped the contents and history of our documents repo. Fortunately I had it cloned on my laptop so nothing was lost. | Need to discuss work schedule with Kevin and Lucian. |