# Oregon State University

## College of Engineering

### SENIOR DESIGN CAPSTONE  FINAL REPORT

MAY 29, 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

## 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

Web-cam made by Logitech. [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 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 CURRENT STATE

### 4.0.1 Proposed Design

The Logitech Brio webcam provides a high-resolution, two-dimensional image but lacks depth perception. The LEDDAR M16 provides accurate depth measurement in a horizontal dimension but lacks vertical perspective beyond a 40-degree spread. 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.

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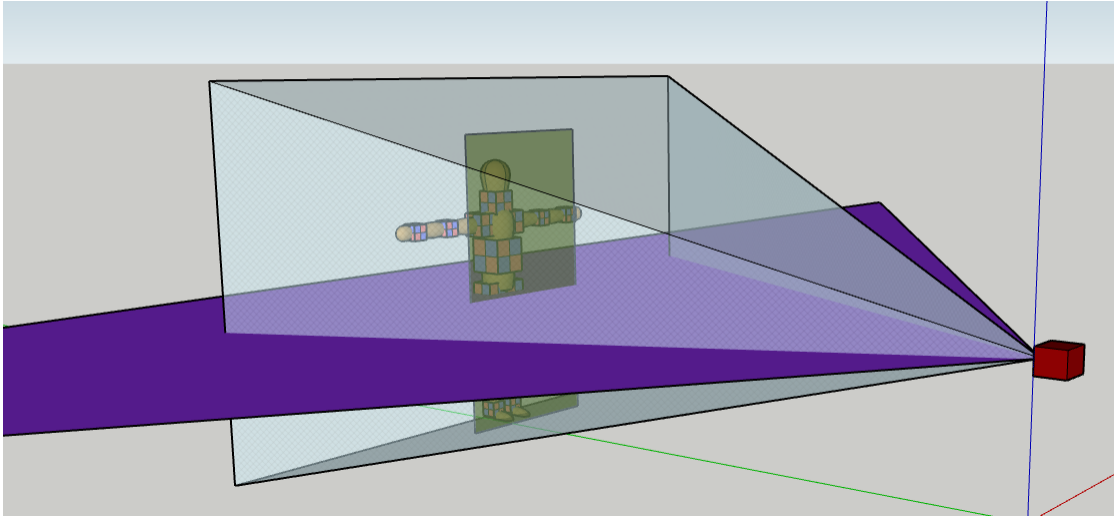
Fig. 1. Visualizing different dimensions measured by the M16 Lidar and Brio Webcam.

### 4.1  Computer Vision

Over the past term, my contribution towards the final design was creating a reliable image-recognition model. 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.

Our 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.
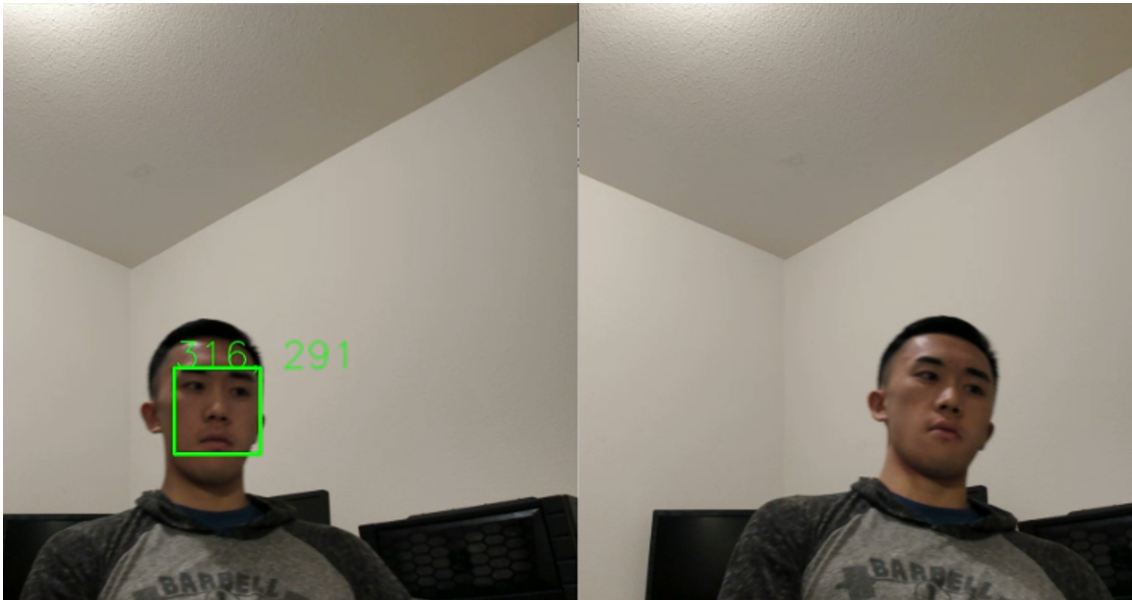


Fig. 2. SVM face classification (Left) fails when subject slightly turns their head (Right)

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. Our pre-trained Tensorflow model can reliably detect multiple subjects

Using this pre-trained Tensorflow model, our project is now able to accurately outline and label over 90 subjects as they come into view of the webcam. In addition to compiling this Tensorflow model, I have also prepared our code for when we eventually create the depth-sensing component of our project. The current state of the code will enable us to selectively edit 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 our list of material requisitions for our physical expo demo, moving calculations to the GPU greatly improves the output video quality, frame rate, resolution, and classification speed. [9]

# 5 TEN-WEEK TERM RETROSPECTIVE

| Week | Positives | Deltas | Actions |
|---|---|---|---|
| 1 | - | 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 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 | 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<br>I believe we have not gone astray. What I have written is relevant in either option. I still believe we can accomplish 1).<br>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.<br>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. |
| 2 | Meeting with kevin over webex last Friday transcribed notes<br>May or may not have a partner<br>Usable up to 50 meters, 450 spread<br>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<br>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 | - | 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 |

| | | | |
|---|---|---|---|
| 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 doesnt 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. | - | Successfuly implimented a pretrained 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. Fortunatly I had it cloned on my laptop so nothing was lost. | Need to discuss work schedule with Kevin and Lucian. |