



College of Engineering

SENIOR DESIGN CAPSTONE INDIVIDUAL WINTER PROGRESS REPORT

MARCH 20, 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 Project Purpose

4 Current State

4.1	Computer Vision
-----	---------------------------

5 Partner Evaluation

6 Conclusion

7 Ten-Week Term Retrospective

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>

2 DEFINITIONS

2.1 IR

IR refers to the infrared light spectrum.

2.2 IR Depth Sensor

A device that calculates distances by emitting infrared patterns.

2.3 lidar

Light Detection And Ranging - A method that uses lasers to measure distance

2.4 Microsoft Kinect

A product that uses an IR Depth sensor to measure distances. Referred to as a benchmark comparison for the purpose of this project.

2.5 Logitech Brio Webcam

Web-cam used for this project 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

Commercial infrared-based depth sensors such as the model used in Microsoft's Kinect can quickly calculate distances in indoor scenarios. However, IR depth sensors can be confused by other infrared emitters 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 in varying conditions.

Depth Sensing with Computer Vision and lidar proposes combining the power of computer vision with the reliability of lidar technology. Instead of strobing IR beams, lidar uses a pulsing laser to measure relative distance. The Leddar M16 is a high-end solid-state lidar that can acquire distance data without the use of moving parts. (2.7) We'll be combining this with a high-end Logitech Brio Webcam. (2.5) This project proposes bridging the utility of both devices by securing them in stationary positions, then using software to combine their outputs. The completed project shall present a conglomeration of computer-vision based object recognition married to the high-accuracy of lidar. This involves using the RPLidar's library to get depth sensing information, and using computer vision to recognize objects.

4 CURRENT STATE

4.1 Computer Vision

My primary focus this term was creating a reliable image-recognition model. I first started with OpenCV's pre-trained facial/pedestrian SVM classifier. Our design specifications envision a system that quickly tracks multiple subjects. This SVM does not fulfill this requirement. As seen in 1, the SVM model is limited to faces and not very accurate. If the human subject were to turn their head or move too quickly, the SVM will have difficulty tracking their body. Additionally, calculations are very CPU-intensive, resulting in low frame-rate and resolution.

Recognizing the SVM's weaknesses, I started experimenting with Tensorflow's open source object detection classifier. [4] Unsurprisingly, Tensorflow's powerful network is several magnitudes more accurate than OpenCV's static SVM. 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.



Fig. 1. SVM face classification (Left) fails when subject slightly turns their head (Right)

5 PARTNER EVALUATION

My partner, Lucian Tamno, displays excellent work ethic, enthusiasm, dedication to our project, and an absolute commitment to do his best. I am very proud of my partner's performance, and I am confident we can achieve our project goals despite our late project start. Lucian's professional demeanor sets him apart and above many of our peers, as he is able to self-assign tasks and quickly comprehend technical concepts. I believe our individual experiences in professional working environments facilitates our team chemistry and communication. It is extremely refreshing to witness my project partner assume responsibilities and take ownership of his role. Where others may attribute their failures to others or blame external influences, Lucian displays level-headed and long-term thinking that will benefit him in his professional career. While we sometimes struggle with a small language barrier, I am certain this won't be an issue as we learn each other's mannerisms. I believe we have a lot to teach each other, and I look forward to our final term as we build this project to reflect our best efforts.

6 CONCLUSION



Fig. 2. Tensorflow pre-trained model can detect multiple subjects.

7 TEN-WEEK TERM RETROSPECTIVE

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.