Proceedings of The National Conference
On Undergraduate Research (NCUR) 2020
Montana State University, Bozeman MT
March 26-28, 2020

Real-Time Mobility Assistance for the Legally Blind

Raunak Khaitan
Computer Science
University of Wisconsin, Milwaukee
Milwaukee, Wisconsin 53211 USA

Faculty Advisor: Dr. Mohammad Habibur Rahman

Abstract

With increasing autonomous technology around us, this research aims at bringing vision to the legally blind. We are currently using a rover to develop and test the technology. It uses ultrasonic waves to detect objects at a certain distance and results in a five-bit sequence of 1/0 where 1 represents an object on path and 0 represents a clear path forward. This five-bit sequence divides the forward 180 degrees vision into five angles. This is done for both forward and backward motion of the rover. We are also using a Light Detection and Ranging (LiDAR) sensor that maps the surrounding and in conjunction with the rover inputs, it is capable of accurately detecting obstacles in the path. A webcam is used that recognizes objects using neural nets and its implementation is in progress. The webcam along with the machine learning model will be able of classifying objects as stationary and in motion and differentiate particularities like if the traffic sign for the pedestrian is on or off. All of these technologies will be integrated to provide a cohesive and holistic experience to the legally blind person to navigate on the streets independently in real-time. The information will be converted in the form of audio commands and fed to the user to follow them. The long-term perspective of the research is to eliminate every assistive measure being currently used by blind people while navigating and shrink down the technology to smart glasses that are both easy to wear and adapt and fashionable at the same time. Haptic and braille language feedbacks are also a part of long-term ways to impart the sensory information to the user seamlessly in real-time.

Keywords: Blind, Machine Learning, CNN

1. Introduction

New technology. Comfortable life. Connected world. Advancing economy. What's next? Who is being benefited by this? Does this benefit only the healthy and wealthy? Or does the pace of evolving technologies are also addressing the needs of people with different abilities? We are currently in the third decade of the 21st century and we still see people with disabilities facing trouble with their daily life activities like commuting places, communicating with people, and living an independent life. They are always burdened by the need of a babysitter or a caretaker. Can we not change this scenario?

"The problems relating to the handicapped people are in a cyclic order in relation to physical, educational, economic, social, and psychological aspects"[8]. Every time a problem is encountered, to solve them, numerous other problems stand in place. This makes the solution even more difficult. According to the World Health Organization, 2.2 billion people are affected globally with some sort of vision impairment and half of the same population suffer up to severe impairment mainly because of most of the times, the problems are not traced on time and if traced, it's too late for a corrective measure or surgery to take place. Other challenges that make the population more vulnerable to complete blindness or severe impairment is how developed the country is in terms of medical infrastructure. In the same report by the World Health Organization, it has been noticed that countries that fall under the low-medium income category are four times more prone to vision impairment as compared to high-income countries, primarily African countries.

1.1 Global numbers on vision impairment and blindness

Globally, the numbers from The International Agency for the Prevention of Blindness [7] speaks the frightening and grim future the world is headed towards. As per IAPB, a whopping US$8 billion loss is levied on the world economy from Trachoma, a disease that eventually leads to blindness. Also, 1 out of 3 living in the USA suffer from some type of diabetic retinopathy and it is approximated that 1 in 10 is also prone to some sort of vision impairment. Another daunting number from IAPB puts into focus that "70 million people worldwide are at risk of sight loss from Diabetic Retinopathy by 2040" [7]. With little over 28% of the world's population affected by Myopia by 2010, the numbers are predicted to reach a 50% population by 2050.

1.2 Numbers from North and South America

With 3.4 million Americans being legally blind or visually impaired [4], regions like Africa and Latin America not having proper healthcare facilities are at a higher risk of disease contraction as per World Economic Forum [9]. As per the National Institute of Health, it is expected that visual impairment and blindness cases in the U.S. are expected to double by 2050 with increasing cases of glaucoma, which has a high potential of leading to blindness [10]. A study shows that among the Latin American countries, "there are 5.2 ophthalmologists per 100,000 population" [6], which means that these countries have a higher chance of seeing more cases of blindness and impairment.

1.3 Existing technology and market

The market for assistive technologies will hit a record of $6,105.7 Million by 2025 [1]. One of the most commonly used technique by the blind is the braille system. Braille system is available across the globe, from switches to elevators, these help the legally blind and people with poor visibility to understand the buttons they are about to press or the door they are about to enter. Now, that is limited to a certain space and size. Therefore, to combat that, other assistive technologies that use smart glasses to provide real-time guidance through cameras are used. Companies like aira.io use smart glasses to assist people with vision impairment where the user needs to connect to a call center where someone guides the user over a phone call [2]. Where this technology is a great way to assist the people in need in real-time, it is not feasible by all strata of society. Its dependency on a call center representative adds to the complexity of user-friendliness.

Another great technology is the BeMyEyes portal that uses a similar interface but here a volunteer receives a video call from the user and provides guidance accordingly [3]. This technology sets its limit to objects in a confined space or territory. If someone needs to walk on the street independently, just like a normal person, these would not be very much assistive, especially in low-income countries where the infrastructure development is not as good as we witness in the UK or the USA.

2. Proposal and Methodology

To address the challenges faced by people with poor vision and no vision, to give them the independence to live life with comfort and not be dependent on assistive measures like a walking stick, a help dog or other form of technological assistance, we bring our model that we assure to be self-reliant and give independence of mobility to the people in need, irrespective of their geographical location and the time of the day they want to use it. Our goal is to bring a smart glass device that does mainly two important things: collect details around the user and fed that information through audio.

As of now, we have worked on the setup of a small rover that uses a Light Detection and Ranging (LiDAR) sensor to map the surrounding in real-time, a camera for image recognition, and ultrasonic sensors to detect immediate obstacles. Once tested, we plan to shrink this technology into a small form factor like a smart glass that can be both easy to adapt to and fashionable at the same time. This will eliminate the need for other assistive measures. As the information is collected every time, a precise location of that data, the type of object, and if the object is in motion or stationary is being determined and in real-time, the information is converted to an audio signal which the user can use while navigation on the streets.

We have approached this from a rover's perspective and the steps from start until the finished product is:

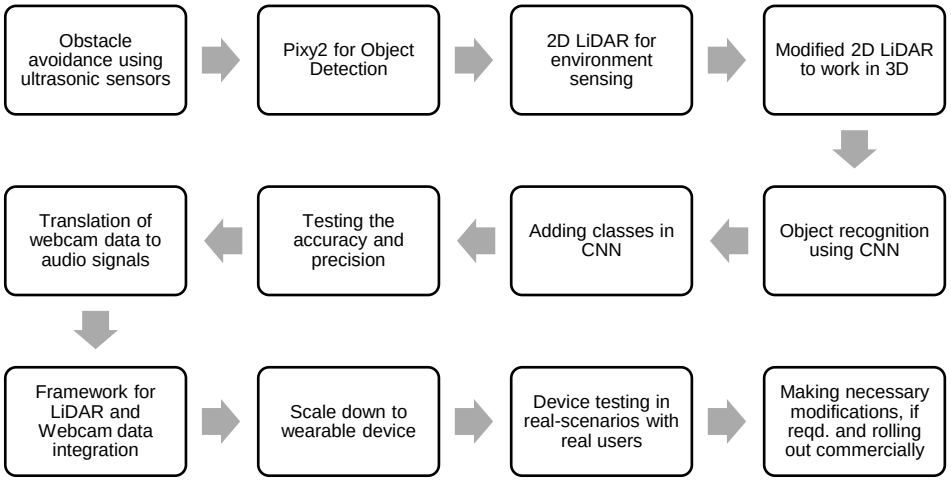

Figure 1. Stepwise approach

The ultrasonic sensors are attached to the rover that it detects obstacles in both forward and reverse motion. The sensor moves 180º in forwarding motion, where the entire angle is divided into five zones: Right (0º), Right Diagonal(45º), Center(90º), Left Diagonal(135º), and Left(180º). A five-bit sequence is generated when continuously when the rover is in motion. The five-bit sequence of 1's and 0's represent an obstacle with 1 representing an obstacle and 0 representing no obstacle. For example, a sequence 00100 would represent an obstacle in the center only and a sequence 01100 would represent an obstacle in the center and left diagonal only. Therefore, at this stage, the rover would stop, scan again, and proceed towards an obstacle-free path.

```
begin str=01100
begin str=11111
begin str=11111
begin str=11111
begin str=00100
begin str=00000
begin str=01110
```

Figure 2. Ultrasonic Sensor Output

The pixy2 camera was then used for image recognition at the early stages where the job was satisfactorily met but it was limited to a certain number of objects that can be identified. In this picture, the bounding boxes represent s=3, i.e. the third image it was taught to recognize where the ceiling lights. But in the real-time scenario where thousands of activities are happening around us, being versatile is the main challenge, and therefore the ability to recognize all possible actions and happenings was the drawback of this device.

Figure 3. Pixy2 camera output

A 2-dimensional LiDAR sensor is used in this work so far. Because a 2D LiDAR only captures one perspective of the environment, we modified it to give a 3D depth of the environment. This sensor revolves 360º, with frequencies adjustable up to 12Hz. While the sensor would revolve in a circular motion, the modifications would move it 60º above and below the line of actual sight. When modified to work in 3D, for every object scanned above and below the line of actual sight, the angle that is created changes the distance at which the obstacle is present. Therefore, to combat that difference, the distance will be modified using the formula from trigonometric ratios: $\cos\theta$ = base/hypotenuse.

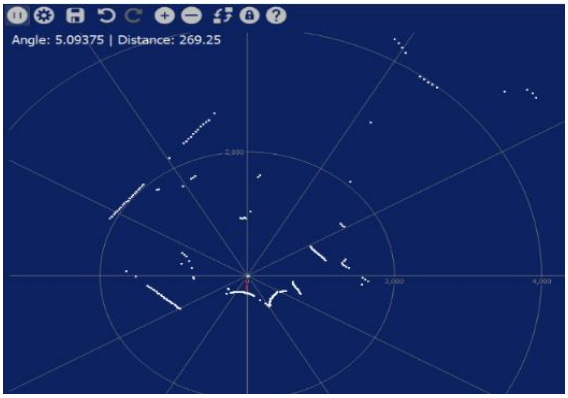

Figure 4. 2D LiDAR output

The next step included using machine learning to train a model to identify objects in real-time. This is one concrete step towards the entire process and goal of this research. For training the first model, a convolutional neural network was trained on a dataset provided by drive.ai, accessed from Coursera [5]. In this dataset, there are 80 classes and the YOLO architecture is used (You Only Look Once). In this architecture, the basic concept is that the algorithm looks at the picture just once and generates output. It requires only one pass in the forward direction, thus saving computational time and increasing accuracy. The input image size is (608,608,3) and 5 anchor boxes were used, meaning each model would predict 19x19x5 = 1805 boxes.

Figure 5. CNN output of two images, detecting cars and traffic lights through the bounding boxes.

The next model was trained on the CIFAR10 dataset. This dataset was obtained from the University of Toronto's computer science department's website [12]. It is a subset of 80 million tiny images dataset collected by Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. It consists of 60,000 images with 10 classes, each class having 60,000 images ranging from airplanes to animals to automobiles. There are 50,000 images in the training set and 10,000 images in the test set. This dataset was directly imported to the IDE from Keras documents. The images are in color red, green and blue, measuring 32x32 pixel squares each. In this model, a vector of integers from 0 and 1 for each class defines the output and therefore OneHotEncoder was used to convert them into a binary matrix. In this dataset, features are the 10 classes with pictures from different segments. For example, the first class is an airplane, the second class is an automobile, etc. These classes are mutually exclusive, meaning that no two classes have the same images. For example, automobiles and trucks are two different classes but since the truck can be considered in the automobile, none of the images overlap. There are no missing values or attributes in this dataset. Based on the features or the classes, targets are random images from these classes that need to be classified accurately with one of the class labels. There are no outliers or irregular cardinality in this dataset. Stochastic gradient descent (SGD) was used in training this machine learning model because the dataset is enormous and using simple gradient descent to find the lowest point would be computationally expensive. Therefore, SGD was used which in each iteration picks random data points from the entire dataset, thus saving time. I used two layers each of 32 filters, 64 filters, and 128 filters. In addition to that, there was one layer each of 512 filters and 1024 filters, all with filter size 3x3. The same padding was used as a design decision meaning that the outside boundary was padded with zeros. ReLU activation function was used and MaxPooling was used so that only the maximum of the incoming node was passed on. Because it is a classification task, therefore softmax function was used at the end. Maxnorm is a regularization technique that enforces the weight vector magnitude to not exceed a certain limit and categorical_crosentropy was used as loss for classification with stochastic gradient descent for optimizer and accuracy for the metrics while compiling the CNN.

3. Results

As seen before, the ultrasonic sensor gives 100% accurate obstacle notification through the five-bit sequence in its line of sight but the pixy2 camera fails to act in real-world challenges. The modified LiDAR serves the purpose but integrating a 3D LiDAR system with the webcam image recognition would be the best complete model that would suffice the needs. From the object detection on two different datasets, the initial dataset does a great job of identifying objects using bounding boxes from real-life pictures and in case of the CIFAR10 dataset, with restricted access to CPU and computational power, different filters and epochs make a difference in the accuracy in which the model performs.

Table 1. Output comparison from CIFAR10 dataset with different configurations.

Method	Accuracy
2-layer CNN without SGD	<30% (Epoch = 50)
3-layer CNN with SGD	Around 70% (Epoch = 30)
4-layer CNN without SGD	Around 35% (Epoch = 50)
8-layer CNN with SGD	Around 80% (Epoch = 75)
8-layer CNN with SGD	Around 70% (Epoch = 10)
10-layer CNN with SGD (2 layers each of 32, 64, 128, 512 and 1024 filters)	Around 68% (Epoch = 50)

4. Future Work

As this research progresses, the next steps are to generate new machine learning models with higher accuracy that can run irrespective of the computational power they require, integrate the platform with an actual 3D LiDAR, and mount everything on the rover. The next step would be the conversion of raw information to human-understandable language that the user would use to navigate on the street. These steps would then follow up to the amount of accuracy level the technology works on in real-time and having everything near perfect, the job to shrink down the platform from the size of a rover to a wearable would take place.

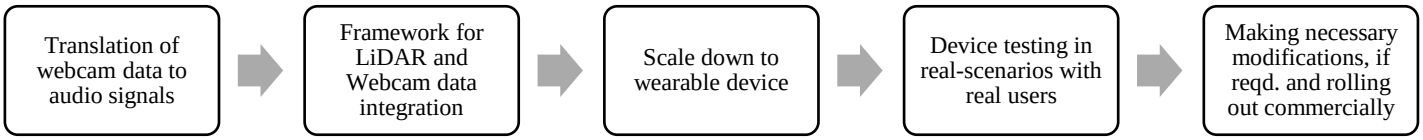

Figure 6. Future steps in this research

5. Acknowledgments

Funding for this research was possible through the University of Wisconsin Milwaukee's Support for Undergraduate Research program. The author would also like to thank the BioRobotics Lab of the university for providing the expertise and guidance in this relatively new field for the author.

6. References

1. A. M. Research, "Assistive Technologies for Visually Impaired Market to hit $6,105.7 Million by 2025-Insights on Competitive Strategies, Key Trends, Restraints, Demand, Drivers, Products, End Users, and Business Opportunities: Adroit Market Research," GlobeNewswire News Room, https://www.globenewswire.com/news-release/2019/09/16/1915732/0/en/AssistiveTechnologies-for-Visually-Impaired-Market-to-hit-6-105-7-Million-by-2025-Insights-onCompetitive-Strategies-Key-Trends-Restraints-Demand-Drivers-Products-End-Users-andBu.html
2. Aira, "Pricing," https://aira.io/pricing.
3. Be My Eyes - See the world together , "Be My Eyes - See the world together," https://www.bemyeyes.com/.
4. Centers for Disease Control and Prevention, "CDC – Burden of Vision Loss – About Vision Health – Vision Health Initiative (VHI)", https://www.cdc.gov/visionhealth/basic_information/vision_loss_burden.htm.
5. Coursera, "Deep Learning Specialization," https://www.coursera.org/specializations/deep-learning.
6. H. Hong, O. J. Mújica, J. Anaya, V. C. Lansingh, E. López, and J. C. Silva, "The Challenge of Universal Eye Health in Latin America: distributive inequality of ophthalmologists in 14 countries," BMJ Open, Nov. 2016.
7. IAPB Vision Atlas, "Discover the IAPB world of eye health," http://atlas.iapb.org/.

8. “Identification of the problems of the disabled,” https://sg.inflibnet.ac.in/bitstream/10603/5117/8/08_chapter%203.pdf
9. J. A. G. Ramirez, “These are the 5 health challenges facing Latin America,” World Economic Forum, https://www.weforum.org/agenda/2016/06/these-are-the-5-health-challenges-facinglatin-america/.
10. National Institutes of Health, “Visual impairment, blindness cases in U.S. expected to double by 2050”, https://www.nih.gov/newsevents/news-releases/visual-impairment-blindness-cases-us-expected-double-2050.
11. World Health Organization, “Blindness and vision impairment,” https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment.
12. The University of Toronto, “The CIFAR-10 dataset,” https://www.cs.toronto.edu/~kriz/cifar.html.