Computer vision aided Obstacle detection and distance estimation for visually impaired people

Authors : Raj Bhensadadia* (201501011), Bansi Gajera* (201501028), Dhwani Mehta* (201501054), Kanan Vyas* (201501121)

Email Id : raj.b.btechi15@ahduni.edu.in, bansi.g.btechi15@ahduni.edu.in, dhwani.m.btechi15@ahduni.edu.in, kanan.v.btechi15@ahduni.edu.in

*School of Engineering and Applied Science, Ahmedabad University

***Abstract*—Millions of people live in this world with incapacities of understanding the environment due to visual impairment. Although they can develop alternative approaches to deal with daily routines, they also suffer from certain navigation difficulties as well as social awkwardness. For example, it is very difficult for them to find a particular room in an unfamiliar environment.Computer vision technologies have been rapidly developed in recent years. It is promising to use the state-of-art computer vision techniques to help people with vision loss. Our project proposes a model for real time Obstacle Detection and position estimation, with the goal of informing the visually impaired users about the surrounding obstacles and their spatial position using navigations.**

***Keywords*— Computer vision, distance estimation, object detection, object recognition, Path Navigation**

I. Introduction

The World Health Organization estimates there are about 314 million vision impaired people in the world, of which about 45 million are blind. The leading causes of blindness are cataract, uncorrected refractive errors, glaucoma, and macular degeneration. Many people who are seriously vision impaired use a white cane and/or a guide dog to avoid obstacles. Moving through an unknown environment becomes a real challenge when we can't rely on our own eyes. Since dynamic obstacles usually produce noise while moving, blind people develop their sense of hearing to localize them. However they are reduced to their sense of touch when the matter is to determine where an inanimate object exactly is.

Nowadays, most of the commercial solutions for visually impaired localization and navigation assistance are based on the Global Positioning System (GPS). However, these solutions are not suitable for the visually impaired community mainly due to low accuracy, signal loss and the impossibility to work on indoor environments. Moreover, GPS cannot provide local information about the obstacles in front of or in the near surroundings of the person. Furthermore, other commercial products available in the market present limited functionalities have low scientific value and are not widely accepted by the users.

Computer vision-based approaches offer substantial advantages with respect to those systems and constitute a promising alternative to address these problems. But till now there is no existing tools that satisfies major needs of blind persons. So here we are propose a new framework for blind assisting system. Also we took a survey of some use full methods. By means of visual Simultaneous Localization and Mapping (SLAM) techniques it is possible to build an incremental map of the environment, providing at the same time the location and spatial orientation of the user within the environment. In addition, compared with other sensory modalities, computer vision can also provide a very rich and valuable perception information of the environment such as obstacle detection [1] or 3D scene understanding [2]

I. Literature review

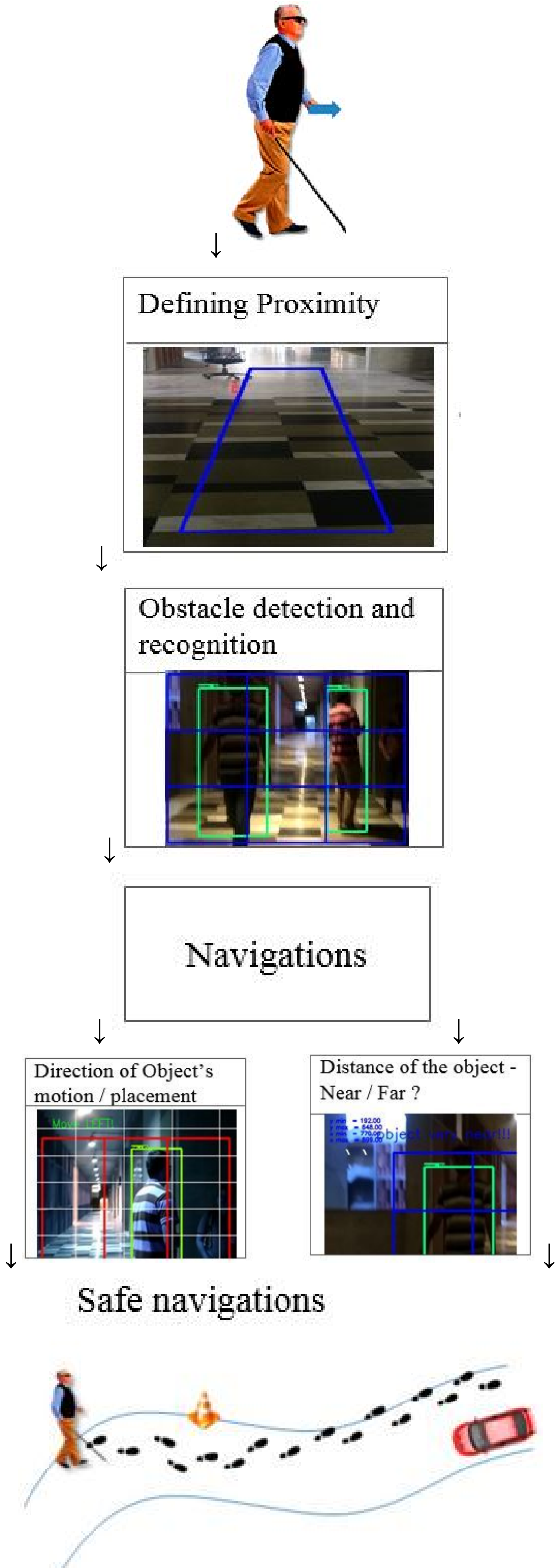
With the goal of advancing in the state-of-art in object recognition, Microsoft has provided a dataset called Common Objects in Context (COCO). It is collection of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance segmentations to aid in precise object localization. Dataset contains images of 91 objects with a total of 2.5 million labelled instances in 328k images.

You Only Look Once algorithm (YOLO) frames object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. YOLO learns very general representations of objects.

SSD discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. SSD is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stages and encapsulates all computation in a single network.

Advances like SPPnet and Fast R-CNN have reduced the running time of object detection networks, exposing region proposal computation as a bottleneck. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection.

II. Block-Diagram



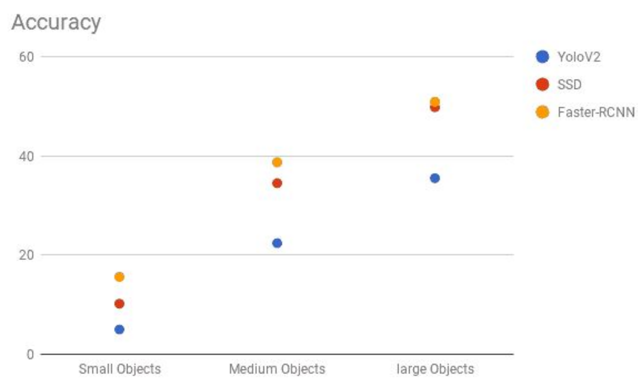
I. Object detection and recognition

Various computer vision algorithms can be used for object detection. SIFT can be used to detect known objects. Trained objects can then be searched for in the environment using common SIFT feature matching and classification methods. When using local features such as SIFT and SURF, it is only possible to detect specific, known objects (i.e., existent in the database) with a distinctive texture.

Using pretrained models for object detection and recognition is a better approach as:

1. It is faster
2. Sufficient number of trained classes
3. Better accuracy

Some of the pretrained models are Faster RCNN, YOLO, SSD and many more. We are using Faster RCNN for object detection.



Comparative analysis of YOLO, SSD and faster RCNN in terms of accuracy in object detection.

From this graph. We can see that SSd and Faster R-CNN works better on large images in terms of accuracy.

Faster RCNN : Faster R-CNN has two networks: region proposal network (RPN) for generating region proposals and a network using these proposals to detect objects. While RCNN approach works well in terms of accuracy, it is very costly to compute since the Neural Network has to be evaluated for each ROI. Fast R-CNN addresses this drawback by only evaluating most of the network (to be specific: the convolution layers) a single time per image.

SSD : Key idea here is single network (for speed) and no need for region proposals instead it uses different bounding boxes and then adjust the bounding box as part of prediction. Different bounding box predictions is achieved by each of the last few layers of the network responsible for predictions for progressively smaller bounding box and final prediction is union of all these predictions.

II. Proximity Defining algorithm

In any video frame, we need to define a particular proximity where object detection model will be applied. This proximity will be containing the obstacles for VI user. Initially, when user is walking, the center part of the video part can be considered as video frame because most of the times, the camera will be move like center portion containing the path.

Algorithm:

1.Define a region of interest which is in proximity of the camera.(left,right,front)

2. If the object comes under the region of interest, deploy object detection algorithm within that region.

3.Estimate detection class and detection score of the respective object in the defined region.

4.If the detection score is greater than or equal to a particular threshold (detection threshold is taken to be 70%), consider the object

5.This way the object will be tracked when it comes near to the user and alerts the user.

III. PATH NAVIGATION

Path navigation will be done in two ways. The first part of navigation is by detecting where the obstacle and navigate using the coordinate of obstacle. The second part of the navigation is send navigate alert for the object is near and then send the notification for the far object.

1.Left-Right Navigation:Left-Right Navigation will help user to plan his/her movement. The first step would be to further divide the region of interest into 3 vertical sections each determining the left, right and front direction respectively.

As the object comes in left section, the coordinates of the bounding box of object will lie in the left segment. At that time, user will be navigated to move towards right. When the object comes in right section, coordinates will lie in right segment and in this situation , user will be instructed to move towards left.

A. Divide the region of interest into 3 segments vertically (left, center and right)
B. When the coordinates of object is in the right segment, it will be navigated to Left and when the coordinates of object is in the left segment, it will be navigated to Right.

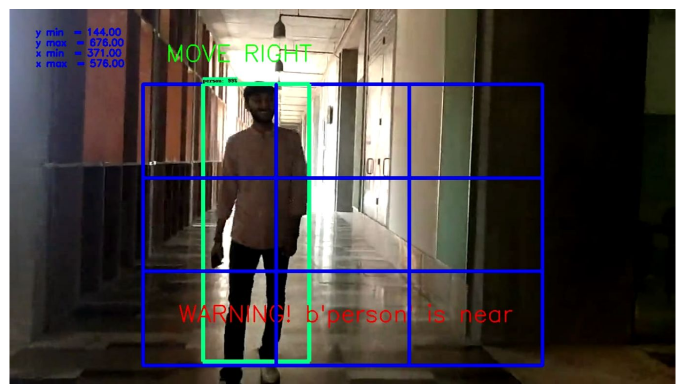
2. Distance Navigation: Distance Navigation will be used to make user aware of spatial position of the objects in terms of distance to help him/her plan his movement For this, the region of interest is divided into 2 horizontal sections each representing the far and near area of the path. When an object falls into far section, its coordinates will also lie in the far segment, based on which user will be informed that the object is far. And when object comes near to fall under near section, a proximity alert will be generated to make user aware of its presence and help avoid collision.

A.Divide the region into 2 segments horizontally (far and near)

B.When the coordinates of the object is in far segment, user will be navigated about the object's distance as far and when the object's coordinates are in near segment, user will get warned about the object's distance as near and proximity alert will be generated.

IV. RESULTS

We implemented obstacle detection followed by path navigation algorithm on a video frames and got about 70% accurate navigation and detection results. The model also works accurate in occlusion conditions and blurred videos.



In this test image, we can see the person is identified as obstacle and the user is navigated to the Right, since the person is on the left side. Also, the obstacle is very near to the user, warning has been generated based on the distance navigation algorithm.

V. CONCLUSION

In this project, we have successfully implemented Obstacle recognition using Tensorflow object recognition API and then implemented distance estimation of the object and navigation so that it can help visually impaired people to navigate based on the obstacles. As a part of future work, this recognition and navigation technique will be implemented in Real time android application. We will also focus on Navigation optimization for better acoustic feedback. In real time implemention, we also need to take care about how much time it will take frames per second.

VI. REFERENCES

[1] Godard, Clément, Oisin Mac Aodha, and Gabriel J. Brostow. "Unsupervised monocular depth estimation with left-right consistency." CVPR. Vol. 2. No. 6. 2017.

[2]Wedel, Andreas, et al. "Real time depth estimation and obstacle detection from monocular video." Joint Pattern Recognition Symposium. Springer, Berlin, Heidelberg, 2006.

[3]Mohan, Dinesh and Dr. A. Ranjith Ram. “A Review on Depth Estimation for Computer Vision Applications.”, positions 4.11 (2015).

[4]Rodríguez, Alberto, et al. "Assisting the visually impaired: obstacle detection and warning system by acoustic feedback." Sensors 12.12 (2012): 17476-17496.

[5]Tapu, Ruxandra, et al. "A smartphone-based obstacle detection and classification system for assisting visually impaired people." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2013.