# Object Detection and their Localization for Visually Impaired Users

Twinkle Motwani, Sejal Gianani, Abhishek Mehta, Rohan Shende Computer Engineering, V.E.S.I.T. Mumbai, India

> Sharmila Sengupta Professor, Computer Engineering Dept, V.E.S.I.T. Mumbai, India

# ABSTRACT

Visual impairment is the functional loss of the eye or eyes in which a person's eyesight cannot be corrected up to a normal level. Visually impaired people face many problems in their day to day life and the most common inconvenience is to find their personal items which are misplaced in their indoor space. With advancement in technology, normal conduction of tasks can be made possible for such cases. This paper describes a system which is a user friendly human computer interface enhanced with computer vision technology which will be able to support the visually impaired to localize and pick up objects used in their daily life. Several efforts have been taken in the past few years to improve the quality of life of such people. The aim of our system is to bridge the gap between what the disabled person wants to do and what the existing social infrastructure allows them to do.

#### **Keywords**

Visually impaired; Object detection; Object positioning; Deep Learning.

## 1. INTRODUCTION

According to the World Health Organization, 253 million people live with visual impairment, 36 million people are blind, 217 people have moderate to severe vision impairment. Computer vision can be delineated as the field of science and artificial intelligence for building artificial systems that obtain knowledge from images or multi-dimensional data and give computers a visual understanding of the world. Some of the latest technologies developed for the blind include - smart glasses [1] that can read and recognize faces, Finger reader, Blind reader, Co-robotic cane- GPS based assistive device [2] etc. These existing systems succeed in object recognition, navigation [3] but haven't achieved much in determining the exact position of an object so that the user grabs it. We have arrived at a method for determining the position of an object with respect to user's hand. This method guides the user to move his hand and reach the object. It uses combined methods like deep learning based object detection, neural networks, image processing techniques (feature extraction, morphology).

The project helps visually impaired people by increasing their confidence level and making them more independent. The relevance of this system plays a vital role in the society. It can be used in blind training institutes, non - profit organizations and any indoor environments like home, office, schools.

# 2. METHODOLOGY

#### 2.1. System Design

There are several modules in the system design which are depicted in Figure 1. First we have the user with the device i.e. an integrated hardware unit which is developed. The image is captured via live video stream through a camera module. The video is divided into frames which are resized and these frames are converted into BLOBs which is a collection of binary data stored as a single entity in a database management system. This data is then given to the deep learning neural network. Deep learning neural networks are different from single hidden layer neural networks because of their depth i.e the number of node layers through which data passes in a process of multiple steps. Previously neural networks had one input and one output layer, and at most one hidden layer in between. More than three layers including input and output is considered as deep learning. This is followed by the object detection module using OpenCV which detects all objects included in our dataset. Approaching further, we have the object mapping and positioning module. After detection of objects the system will assist the user what objects are there in front of him. The user via voice commands tell the system which object he wants to grab, his speech is converted into text. Next the required object is mapped with the real time input and it is the only object that is shown in the frame. Then distance calculations are done and the angle through which the user should move his hand to grab the object is computed. As our system is designed for objects within a vicinity of 1 meter, the system checks if the object is within an arm's length (1 meter) and accordingly text to speech conversion takes place. The system then guides the user by giving appropriate instructions via audio commands (i.e. angle and distance) for grabbing the object.



Fig. 1: System Design

#### 2.2. Object Detection

Object Detection has been implemented using OpenCV which also supports deep learning frameworks, and an open source library for programming functions in python. The problem with traditional algorithms is that they do not give high accuracy results [4]. Hence the use of neural networks was initiated. SSD (Single Shot Detector) framework combined with MobileNet architecture was used to arrive at faster and efficient deep learning based method for object detection [5]. This network architecture was then trained to detect common objects that the blind person comes across in his daily life.

Figure 2 shows the mapped output of a bottle which is detected. Mapped output means that from a certain range of objects, only those objects appear in the frame which the user wants to grab.



Fig. 2: Detection of dissimilar objects



Fig. 3: Detection of similar objects

Figure 3 depicts detection of different objects present in front of the user. The user can now tell the system via voice commands which particular object he wants to grab. Figure 3 shows detection of similar objects in which bottles of different sizes and shapes are detected. The system tells the user how many bottles are present in front of him and also tells the distance and angle of each bottle with respect to user's hand. The user can now grab the bottle which is at least distance from his hand.

#### 2.3. Band Detection

In order to give a unique identification to the user, a band of specific unique colour is used. This band is worn by the user on his wrist every time he wishes to grab an object. The restriction that comes into picture is that there should not be any other object having the same colour as that of the band, otherwise the system takes into consideration the object with larger surface area.

The live input stream is taken continuously and then divided into several frames. Further, these frames are converted from RGB to HSI. The reason for the conversion to HSI is that when light falls on a particular object the intensity changes and thus leads to change in the RGB value, hence it is not easily recognizable. Range of that unique band colour is provided. Minimum range to maximum range is given for the band. As a range is specified, all the objects (or colors) that lie in the frame On the image of the band detected, an open operation followed by close operation is then performed using 5\*5 mask. It is used to enhance and aggregate all the separable objects and make it appear as one entity. Number of contours is counted in that frame. If the number of contours appeared to be 0 it indicates that the hand is not in the frame. It indicates the user by saying "Hand is not in frame". And if it appears to be greater than 0 then the hand is in the frame and it is detected. After the detection the center of the contour is found out, this tends to be the user band. After calculating the center position the distance to all the objects is then calculated by keeping it as the reference coordinate.

Figure 4 shows the band which is detected along with its centroid coordinates as (320, 222).



Fig. 4: Band Detection

#### 2.4. Object Positioning

After detection of the object and the band (worn by user on his wrist), we have to position the object with respect to users hand. The position of the object is instructed to the user by telling him whether the object is to his right, left or center, at what angle and distance is the object from user's hand.

Among the captured band image and the mapped number of objects the centroid coordinates of each object is calculated. A line is drawn connecting the centroids and the Euclidean distance is computed. The result is obtained in pixel which is converted into centimeters and then multiplied by a scaling factor to get the real life distance. The scaling factor is computed for x and y axis by conducting an experiment. Using these values and trigonometry, scaling factor at any other angle was found out. Angle was found out using trigonometry and centroid coordinates.

Table 1 shows the results of an experiment performed. An experiment to find out the scaling factor between the image of an object on the paper and the object as it appears in real life was performed. Figure 5 shows the setup for our experiment in which an object was moved across different positions i.e. right, left, forward, and backward and the distance was noted from a fixed position.



Fig. 5: Experimental setup

Table 1. Experimental Results	
Position	Image to Real Life Distance
Centre to right	1 cm in image = 4.59 cm in real
Centre to left	1 cm in image = 4.5 cm in real
Centre to forward	1 cm in image = 18.89 cm in real
Centre to backward	1 cm in image = 12.5 cm in real

Scaling factor for x axis: (4.59+4.5)/2 = 4.5

Scaling factor for y axis: (12.5+18.89)/2 = 15.7

#### 2.4.1 Finding the Angle

Let (ax, ay) and (bx, by) be the centroid coordinates of the object and the band. The following conditions were tested:

1. (bx-ax) not equal to zero: If the object lies on any other line other than vertical line then its slope is calculated using two-point form. This gives the angle with respect to horizontal axis. Then we find the angle with respect to vertical axis by subtracting it from ninety degrees.

2. (by-ay) equal to zero: If the object lies on horizontal axis, then it simply returns angle as 90 degrees.

3. (bx-ax) equal to zero: If the object lies on vertical axis, then it simply returns angle as zero degrees.



Fig. 6: Angle Error Calculation

Figure 6 shows the difference in angle computed in real life and by our system for different cases.

#### 2.4.2 Finding the Distance

An image based technique had been developed earlier to compute the depth and lateral distance of different types of objects but it was restricted to few objects of shape rectangular, cylindrical and triangular [6]. This paper describes another method for distance calculations in an easier way. Euclidean distance between two points a and b is the length of the line segment joining them. If a = (ax, ay) and b = (bx, by) then the distance is given by

$$d(a, b) = \sqrt{(ax - bx)^2 + (ay - by)^2}$$

The conditions mentioned above for angle computations are again tested for distance and the distance is multiplied by appropriate scaling factors to get the actual distance.

Actual distance = Euclidean distance\*scaling factor\* 0.0265 (for pixel to centimeters)



Near vo carculated

Fig.7: Distance Error Calculation

Figure 7 shows the difference in distance measured in real life and by our system for different cases.

Figure 8 shows the prototype of our system consisting of a raspberry pi, raspberry pi camera module, wrist band, earphones along with microphones. This is a lightweight system which is easily wearable by the user.



Fig. 8: Prototype of the system

## **3. CHALLENGES**

The distance accuracy is one of the major challenges faced. As we move from a three dimensional space to a two dimensional space, there is no concrete algorithm which exists for three dimensional to two dimensional mapping. The system does not take into consideration the height of the object and height at which our hand is from the base. In case of a taller or shorter object, the user will have to adjust the height of his hand to grab the object. The user should have basic distance and trigonometry knowledge in order to interpret how much he should move his hand after hearing the angle and distance.

The band worn by user on his wrist is detected based on its unique colour. If any other object appears in the frame with the same unique colour, the system takes into consideration the unique colour object which has a larger surface area. This leads to misinterpretation by the system and hence affects further calculations.

## 4. FUTURE SCOPE AND CONCLUSION

This project has a great scope for further development. Adding sensors to the system could help in detecting the temperature of the object from which we can conclude object is hot or cold to touch. This would prevent any harm to the user. This project can be further extended for navigation purpose by adding GPS feature. The system can also be used in robotics which can act and behave the same like human and also can be further extended for the outdoor environment purpose in which it can not only detect objects but also detect any obstacles that come in their way [7].

Technology transforms people's lives. The visually impaired group of people isn't able to make full utilization of the technology. There is a need and responsibility for the society and community to explore technologies to help them. This paper has described the design of a prototype which is a small step towards helping the visually impaired group of people. Overall our system is low cost, lightweight, simple to use and works well with the objects with the definite shape.

#### 5. REFERENCES

- [1] Jinqiang Bai, Shiguo Lian, Member, IEEE, Zhaoxiang Liu, Kai Wang, Dijun Liu. "Smart Guiding Glasses for Visually Impaired People for Indoor Environment", IEEE Transactions on Consumer Electronics, Vol. 63, No. 3, August 2017.
- [2] Majid Al Shamsi, Mahmoud Al-Qutayri, Jeedella Jeedella, "Blind Assistant navigation system," in 1st Middle East Conference on Biomedical Engineering, 2011.
- [3] Ananth Noorithaya , Kishore Kumar, Dr.Shreedevi A."Voice Assisted Navigation System for the Blind", Proceedings of the International Conference on Circuits, Communication, Control and Computing.
- [4] Andy Lee, "Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics,". Available: <u>https://www.cs.swarthmore.edu/~meeden/cs81/f15/papers/</u>

https://www.cs.swarthmore.edu/~meeden/cs81/115/papers/ Andy.pdf

- [5] Guangxing Han, Xuan Zhang, Chongrong Li. "single shot object detection with top-down refinement", Image Processing (ICIP), 2017 International Conference.
- [6] Ashfaqur Rahman, Abdus Salman, Partha Sarker, Mahfuzul Islam."An Image Based Approach to Compute Distance", International Journal of Computational Intelligence Systems, Vol 1, No 4 (December, 2008).
- [7] Larisa Dunai Dunai, Ismael Lengua Lengua, Ignacio Tortajada, Fernando Brusola Simon. "Obstacle detectors for visually impaired people", Optimization of Electrical and Electronic Equipment, 2014 International Conference.