Visual DUX: A Low cost wearable device for guiding the Blind

Vamsi Krishna Mekala[a], Sowmya Gamidi[b], Babu Rao Markapudi[c], Kavitha Chaduvula[d]

[a,b] B.Tech. Student, [c,d] Professor and Head
[a,b,c] Department of CSE, [d]Department of IT
[a]VR Siddhartha Engineering College, Vijayawada, AP
[b,c,d] Gudlavalleru Engineering College, Gudlavalleru, AP

Abstract: Amongst the world's total population there exist 285 million visually impaired people. There exist several types of equipment which can lead them to a safe way but are restricted to indoor areas and are high priced. In this paper, a wearable device called VISUAL DUX is developed that helps the blind by detecting the objects in their path and directing them through the freeway either indoor or outdoor in a cost-efficient way. As object detection works better with the deep learning algorithms we have used the YOLO V3 dataset as input to the model. The YOLO method uses convolutional neural networks to detect objects by preparing bounding boxes. After getting the bounding boxes, classes of the objects are identified based on the probabilities of the bounding boxes. After finding the class of the object the direction of the obstacle is detected to direct the people with no vision with a safer path. Compared to the existing devices our device gives better accuracy and not only detecting the obstacle it also gives the safer direction to move as an audio output.

Keywords: Visual DUX, wearable device, Object detection, direction prediction, the blind, YOLO

_____

1. Introduction

The most difficult thing for blind people is to fulfil their daily needs independently. Blind people recognize obstacles on their way by sensing them with hands, which is not safe all the time. To make their things happen maverick they seek the help of the devices which guide them. There is a strong need for a system or a device that detects the moving objects and gives the direction to pass that objects safely. There exist devices that detect or predict the object up to the indoor level. Dynamic object detection in the path is much needed while they walk or move. The system must capable to detect the moving objects in the scene.

Though some systems exist for helping the visually impaired, there are a set of limitations associated with their functional approach. Some need more amount of light on the object to detect, some need internet, some need high-end processors to carry out the classification with large image datasets. Some are robust, perform low when in noise, manual setup, and cost of operating is high. [5]Some take a high amount of time for detection and are built with GPUs. The classes of objects that these methods can detect are less in number and some simply say about the objects that are much near to them. Some are limited to night time and some simply give a vibrating signal that there is an obstacle.

Our system is profoundly significant in distinguishing the real-time objects and also the position of the object that assists in giving the direction to the person. Our device work in offline mode with a power source. This uses multiple pre-trained object detection models randomly basing the requirement to give accurate prediction and direction. [7]Computer vision is implemented in python using the modules like OpenCV, gTTS, multithreading, and Imutils. [3]Raspberry Pi included with an external camera module executes the model immediately when it is powered up. V. Kulyukin [8] proposed a robot assisted path finding for the blind with a restriction to indoor environment. The speech synthesis was discussed in [9] which convert text into speech. Jyothi Anusha [10] used object detection based on background subtraction. But our device has compatibility options such as Bluetooth to the external audio device or earphones that can be used.

2. Related Work

The model is proposed by Zhongen Li, Fanghao Song is used to detect some objects at the indoor level [1]. In this model, there are three phases to detect the object: 1. Detecting the object using RGB-D 2. Finding the Orientation of the object 3. Recognizing the object. The raw image of the object is given as the input from the Red, Blue, Green, Depth camera which then produces the depth image with black holes and noises. Later the image undergoes preprocessing to reduce the noises, Blackholes and get the enhanced depth image in Red, Blue, Green colors. And later the orientation of the obstacle is determined by using the IMU Sensors. Later these are passed to the CNN Algorithm for the obstacle to be compared and detected. If there exists an object similar to the obstacle then, the object is detected and there appears a text that in this much distant there is an object (name of the object) else not. The main drawback in this model is the setup isn't portable and also the accuracy of the object

to be detected is not very high. The below Figures 1(a) is the raw input to the RGB - D camera, and 1(b) is the output for the input.
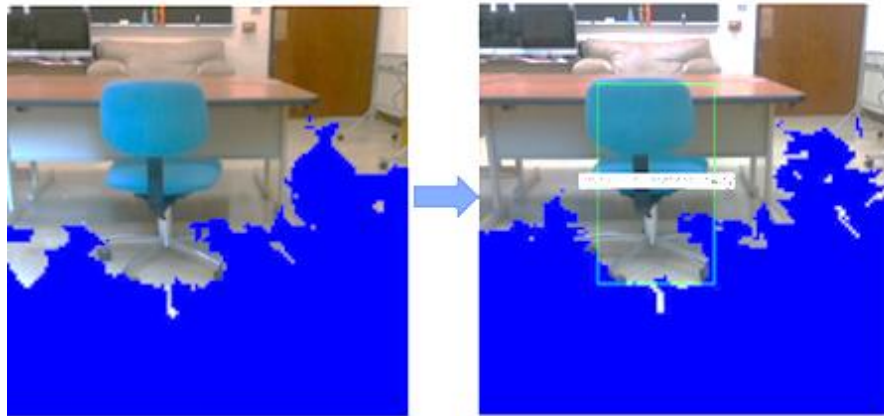

Fig 1(a): Raw Image   **Fig 1(b):** Output for the input image

The Smart stick [2] is used for the detection of objects which are present in a smaller range. The Blind stick has different sensors for the identification of different types of objects. The components used in the Smart stick are 1. Sensors 2. Microcontrollers 3. Speech Warning Message 4. Water Sensor 5. Vibration Sensor 6. Calling Stick. This model consists of 2 types of Sensors named Infrared and Ultrasonic sensors. 2 Ultrasonic sensors are placed one at the top to detect the objects on the upper region and the other at the bottom to detect the objects on the lower region. The Infrared sensors were used to detect the smaller objects whereas the ultrasonic sensors used are to identify the larger ones. The microcontroller used in the blind stick is the PIC microcontroller, which is used to control the blind stick and it acts as the simplified computer with memory, support for the peripherals, and is used as a simplified processor. Recorded voice messages were used as the outputs when an obstacle is detected. Presence of water in their path was detected by using Water sensors. The vibration motor gives a buzzer on detection of the obstacle. Calling Stick is used for the wireless communication of Blind Stick and the visually impaired person using Frequency Modulation. The main drawback with the blind stick is it can't predict the safer path and the obstacle is not specified in particular. Fig 2 is the Smart stick used by the visually impaired people.

Fig 2: Smart stick

3. Visual DUX

The model functions and gives the direction to the blind as audio even offline as shown in Fig.2. The model is built in 3 phases of development, where at each stage the hardware is involved in capturing, detecting, and delivering audio with the software implementation. The hardware developed using computer vision and deep learning uses a model that takes in a video stream and gives out an audio output that voice out your next course of action i.e. whether to move left or right. This process is achieved by using OpenCV for image data handling, a predefined YOLOv3 model for result prediction along with a set of hardware and software modules as depicted in Fig.4.
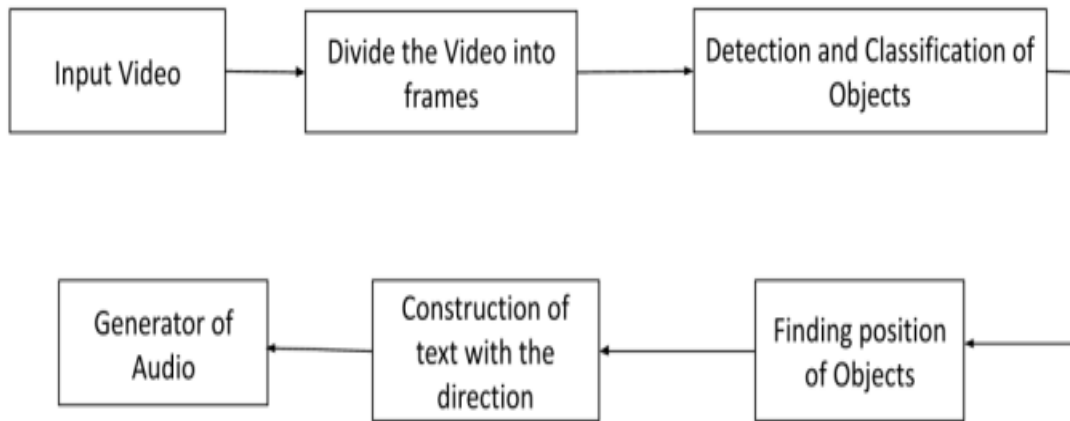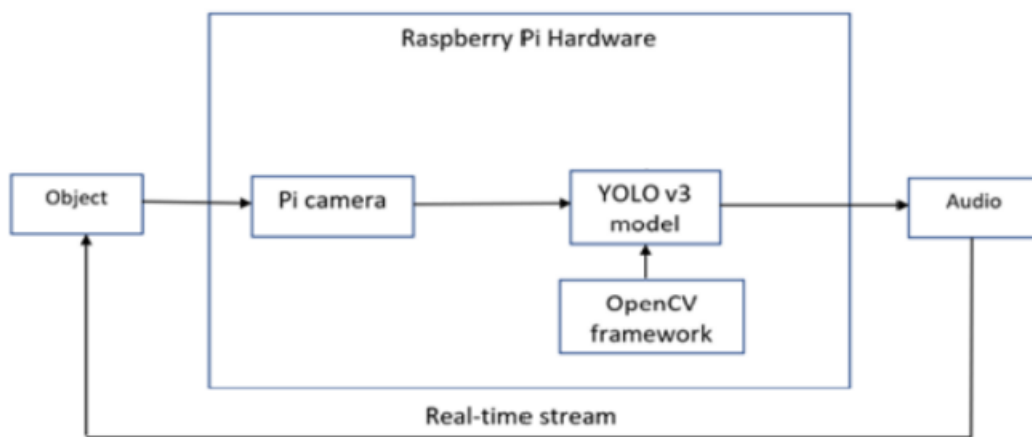
Fig 3: Flow diagram of VISUAL DUX

Fig 4: Hardware functionality of VISUAL DUX

3.1 Yolo V3 Dataset

YOLO is very useful model having its expanded form as "You Only Look Once" is one among the object detecting algorithms and works fast. Although it might not be the object detecting algorithm with the highest accuracy, it is very much useful in real-time object detection with one of the best accuracy. First, YOLO v3 [6] uses another version of Darknet, that has basically a trained network of 53 layers on Imagenet. To complete the task of detection, additional 53 layers are added to provide a fully connected 106 convolutional layers which is the architecture of YOLO v3. The Fig. 5 represents the general flow for a single input image from YOLO V3.
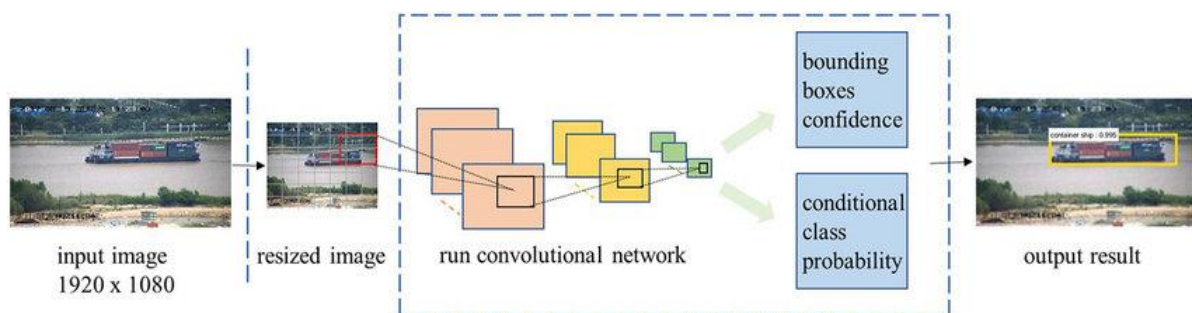

Fig 5: Working of YOLO V3 model

The algorithm in YOLO takes the image frames and converts them into SxS (frame size) frames and each frame cell is provided with the bounding boxes. The object in each bounding box is identified and would be compared with the existing ones and based on the confidence of their identity the object is predicted and also based on the position of the obstacle the direction to move would also be identified. If the object is not identified as any existing object the confidence that would be returned is zero. By using bounding boxes we can identify only a single object i.e. if there exists only a single image in the whole frame. If there are several elements to be identified in a frame then comes the Anchor box.

3.2 Predicting the direction based on object position:

After the different set of objects were detected by YOLO from the given frame, those objects' position is calculated according to the (X, Y) coordinates and also the height and breadth of the bounding box around the object. From this, the converse side of the frame i.e., part with no objects is detected, and by referring to the user's side, the direction is predicted as RIGHT or LEFT. Then all the information about the direction and also the list of objects is turned into text format of a string. The below Fig 6 is used to determine the position of the obstacle from the input stream.

```
if len(idxs) > 0:
    # loop over the indexes we are keeping
    for i in idxs.flatten():
        # find positions
        centerX, centerY = centers[i][0], centers[i][1]

        if centerX <= W/3:
            W_pos = "Move right "
        elif centerX <= (W/3 * 2):
            W_pos = "Move right or left "
        else:
            W_pos = "Move left "
        Addtxt=" identified"


        texts = LABELS[classIDs[i]] + Addtxt + W_pos
```

Fig 6: Code for calculating the position of obstacle

3.3 Audio Output

Soon after the completion of the required semantic string construction, that string is converted into an audio file of the extension .MP3. This is a file saved in the current directory and then with the help of the gTTS module, it is played and heard through the earphones by the user, which helps him crossing the obstacle safely without any hitting or danger.

For every new frame of input from the live stream, this audio file is updated with the current string to be played. This ultimately enhances the overall idea and goal of the project in assisting the visually impaired one to walk autonomously.

4. Experimental Setup

The different hardware components that were used in the development of Visual Dux are:

Raspberry pi 4:

Fig 7: Raspberry pi 4

The Fig. 7 represents the Raspberry pi 4 whose size is equivalent to credit card works like a mini computer with an internal processor and can be interfaced by connecting to any monitor or computer. The RAM of the Raspberry pi that we used is 4GB. This Raspberry Pi has many sets of external connection pins for various purposes. It contains 2 Universal type ports for connecting the Pi to any monitor and a LAN port with 4 USB ports.

Raspberry Pi Camera module:

Fig 8: Raspberry Pi Camera module

The Fig. 8 represents a camera module of Raspberry Pi. It is a light weight and portable device. This device uses serial interface protocol of MIPI camera for communicating with the Pi. The general purpose of this camera module is image processing and video capturing with an appreciable amount of resolution that is generally needed.

TYPE-C power cable:

Fig 9: TYPE-C power cable

Then a USB 3.2 power cable with TYPE-C the pin, the above Fig 9 is used to connect the Raspberry pi to the external power source.

MICRO SD card of 16GB class10:

Fig 10: MICRO SD card of 16GB class10

This is memory card shown in Fig. 10 is used as the secondary storage for the Raspberry Pi that stores the RASBIAN OS which is required for the operation of Raspberry Pi and also the source code of the application, the trained model is been stored in this Memory card. To avoid functional and operational issues, a memory card of type, class10 is preferably used.

Earphones of 3.5mm jack:

Fig 11: Earphones of 3.5mm jack

Earphones/headset of 3.5mm jack presented in Fig. 11 is compatible with the taken model of Raspberry Pi. From there, the person can hear the output from the Raspberry Pi.

The different software used in the model are:

The programming language used: Python
Packages and Modules:
1. OpenCV
2. NumPy
3. gTTS API
4. Multithread
5. Picam
6. Imutils
7. IPython
Model Used: YOLO v3
IDE used: THONNY 1.2

Firstly, the Raspberry Pi is loaded with the RASBIAN OS which has been downloaded from the official site and is stored on the memory card. Then after the setup of the Raspberry Pi is completed, Pi is powered by using the TYPE-C cable. The Pi camera module is plugged into the Raspberry Pi and the required ports are enabled in the Pi configuration. [9]By using the THONNY IDE in Raspberry Pi, the python source code of the project is dumped in. Then the required python modules and the pre-trained YOLO model have been installed along with the source code. When executed, the real-time video is captured through the Pi camera. Frames are obtained from the video, those frames are then fed to the classifier to detect the objects in the frame. Then basing on the position of the object, the direction is attached to the string with all the objects in the frame. Now, this string is converted into audio by the gTTS module in python and stored as an mp3 file. Finally, that file is played to the person through the audio jack or earphones. To make this process iterative, the on boot loader is been added with the python file path. So, it starts detecting the object, soon after the Raspberry Pi is powered up.

5. Results & Discussions

The set of 80 classes of objects can be detected and given as output by the YOLO V3 model. The classes are presented in Fig.12.

person	traffic light	sheep	tie	surfboard	banana	donout	mouse	book	laptop
bicycle	fire hydrant	cow	suitcase	tennis racket	apple	chair	remote	clock	cake
motorbike	stop sign	elephant	frisbee	bottle	sandwich	sofa	keyboard	vase	bowl
Car	parking meter	bear	skis	wine glass	orange	Potted plant	cell phone	scissors	skateboard
aeroplane	bench	zebra	snowboard	cup	broccoli	bed	microwave	teddy bear	handbag
bus	bird	giraffe	sports ball	fork	carrot	Dining table	oven	hair drier	horse
train	cat	backpack	kite	knife	hot dog	toilet	toaster	toothbrush	boat
truck	dog	umbrella	baseball bat	spoon	pizza	Tv monitor	sink	refrigerator	bike

Fig. 12: Table of classes which are detected by the model

The model could exhibit an appreciable accuracy when tested with different datasets. The accuracy is shown in Fig.13.

MODEL USED	FPS(Frames per Second)	Accuracy
YOLO-V3	10	90%
MOBILENET	8	86%
YOLO-TINY	12	89%
CAFFENET	8	85%

Fig. 13: Accuracy for different FPS

The highest notable accuracy is observed with YOLO-V3 i.e., 90% at a frame rate of 10 FPS. Thus, the cause is fulfilled of helping the visually impaired with this device that works offline. The final model looks as the below Fig. 14

Fig. 14: VISUAL DUX

6. Conclusion

The device Visual Dux helps in detecting the objects from real-time video stream with at most accuracy, which enhances the people with no vision to move in both indoor and outdoor independently. In the future view, we are developing an app that can be installed in android devices such that the external device can be neglected for the one who can operate mobiles. Thus leading and applying the science for the welfare of disabled people, to live a happy life.

References

1. ZHONGEN LI, FANGHAO SONG, BRIAN C.CLARK, DUSTIN R. GROOMS, and CHANG LIU, "A Wearable Device for Indoor Imminent Danger Detection and Avoidance with Region-Based Ground Segmentation", Digital Object Identifier 10.1109/ACCESS.2020.3028527, Oct.2020.
2. Ayat Nada, Samia Mashali, Mahmoud Fakhr, "Effective Fast Responce Smart Stick for Blind People", ICABEE, Apr.2015.
3. S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das, and S. Gupta, ''Design, development, and clinical evaluation of the electronic mobility cane for vision rehabilitation,'' IEEE Trans. Neural Syst. Rehabil. Eng., vol. 22, no. 6, pp. 1148–1159, Nov. 2014.
4. H. Zhang and C. Ye, ''An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired,'' IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no. 9, pp. 1592–1604, Sep. 2017.
5. S. Shoval, J. Borenstein, and Y. Koren, ''The NavBelt—A computerized travel aid for the blind based on mobile robotics technology,'' IEEE Trans. Biomed. Eng., vol. 45, no. 11, pp. 1376–1386, Nov. 1998.
6. Sakshi Gupta, Dr. T. Uma Devi, "YOLOv2 based Real Time Object Detection", IJCST-Volume 8 Issue 3, May.2020.
7. M. Brock and P. O. Kristensson, ''Supporting blind navigation using depth sensing and sonification,'' in Proc. ACM Conf. Pervas. Ubiquitous Comput. Adjunct Publication UbiComp Adjunct, New York, NY, USA: Association for Computing Machinery, 2013, pp. 255–258.
8. V. Kulyukin, C. Gharpure, J. Nicholson, and G. Osborne, ''Robot-assisted way finding for the visually impaired in structured indoor environments,'' Auto. Robots, vol. 21, no. 1, pp. 29–41, Aug. 2006.
9. History and Development of Speech Synthesis, Helsinki University of Technology, Retrieved on November 4, 2006.
10. Jyothi Anusha, Ch. Kavitha, M.Baburao, "Efficient Background Subtraction Using improved Multilayered codebook", published in International Journal of Computer & Organization Trends(IJCOT) –Volume28 Number1– January 2016, PP:14-21.