Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 3): pp. 4469-4477
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2020.03.735599

Integrated Smart Glass For Visually Impaired Mobility Using Yolo & Deep Sort

H. Kavietha[1] , S. Abilash[2] , I. Ramakrishna[3] , S. Balaji[4,]

Assistantprofessor[1],

Students[2,3,4]Department of Computer Science and Engineering Sri
Sai Ram Institute of technology Sai Leo Nagar, Chennai-44

Abstract- Every visually impaired people has one dream common that is to do stuff without the help of any human. So, we are introducing vision. Vision is a simple glass that has an AI assistant called "vision" this assistant will communicate with the user(visually impaired)like it can chat with them, recognize objects (such as car, obstacle, and day to day life objects) and many more. The main idea is a mini camera, that is fixed with the frame of the glass and the camera is connected to an AI and these communicate with each other output in a speaker which is also connected in the frame.

Keywords- AI-ML ·YOLO · DEEP-SORT.

1.1 INTRODUCTION

Starting with the most crucial part of human psychology, Vision is an essential need of any person. Considering the fact given by the World Health Organization (WHO), in 2018 nearly 1.3 billion people in the world are suffered from visual problems. Among them, about 39 million people areblind.

Visual impaired persons usually depend on others to fulfill their daily needs, and many times suffer or compromise a lot due to this illness. Visually impaired persons need some help to make their life easier and better. So that we want to make something for them that would help them become independent. The main goal of our project is to help blind people who have vision difficulties by introducing newtechnology.

1.2 SCOPE

All about our research we take care of one problem that is a visual disability. To make a solution we did this low-cost project. We believe that this project will spread all around society and convert disable to able. This is our hope, to consider these glasses as a smart eye for the visual impairments. The main scope of this project is to help visually impaired people and develop a different system that will be more efficient and reliable to blindpeople.

2.1 LITERATURE SURVEY

1. Visually Impaired Assistive System, KanchanM.Varpe,M.P.Wankhade

In this author focuses on the independent portability of blind people who travel in an unfamiliar environment without any manual assistance. The system includes on the server-side Zigbee transceiver for wireless conversation, RFID reader with an integrated microcontroller, Zigbeetransmitter, and TTS for playing information to the user.

The VIAS(Visually Impaired Assistive System) can be used by visually impaired or blind users in the system-implemented environment such as organization campus which can be school, college, hospitals, shopping mart, bus stands,etc

2. Navigation Assistance for the Visually Impaired Using RGB-D Sensor With Range Expansion, A.Aladren, G.Lopez-Nicolas, Luis Puig, and Josechu J.Guerrero

In this paper, a new system for NAVI is presented based on visual and range information. Rather of using multiple sensors, we choose one device, a consumer RGB-D camera, and take advantage of both range and visual information. Inappropriate, the combination of depth information with image intensities, resulting in the robust expansion of the range-based floor segmentation. Our system detects the main structural elements of the scene using range data.

3.1 PROBLEMSTATEMENT

Without vision it can be challenging for a visually impaired person to navigate through a room or a hallway without bumping into obstacles. Even with aid, such as walking stick, it can be sometimes inconvenient, uncomfortable, and perhaps inaccurate in avoiding obstacles.

3.2 FUNCTIONAL REQUIREMENTS Functional requirements specify which output file should be produced from the given file they describe ther elationship between the input and outputof the system, for each functional requirement a detailed description of all data inputs and their source and the range of valid inputs must be specified.

1) Detectingobjects

2) Alert message is generated viaspeaker

3.3 NON-FUNCTIONALREQUIREMENTS

1. Describe user-visible aspects of the system that are not directly related with the functional behaviour of thesystem.

2. Non-Functional requirements include quantitative constraints, such as response time or accuracy.

3.3.1 Usability

This section includes all of those requirements that affect usability. Our system is plug and play type. It does not require any special skills or training.

3.3.2 Performance

The performance characteristics of the system are outlines here:

1. Response time (average,maximum)
2. Throughput (frames processed persecond)
3. Accuracy

4. Resource utilization(memory,disk, camera)

4.1 SYSTEM REQUIREMENT SOFTWARE:

1. Operatingsystem:Windows7andabove / Linux
2. Coding Language :Python
3. Package : OpenCV,Tensorflow
4. Storage :MYSQL

HARDWARE:

1. System : Raspberrypi
2. Hard Disk : 16GB
3. Ram : 1GB
4. Others : Sensor & Camera,Speaker

4.2 SYSTEM DESIGN

We are using Python for this project, the camera is initialized by using OpenCV library and the camera starts capturing frames with the rate of 30 frames per second to the algorithm. Then the system uses YOLO which is trained on the COCO dataset to identify the object kept before the user. The object identified is later converted to an audio segment using gTTs which is a python library. The audio segment is the output of our system that gives the spatial location and name of the object to the person. Now by using this information the person can have a visualization of the objects around him. The proposed system will even protect the person from colliding to the objects around will secure him from injuries.

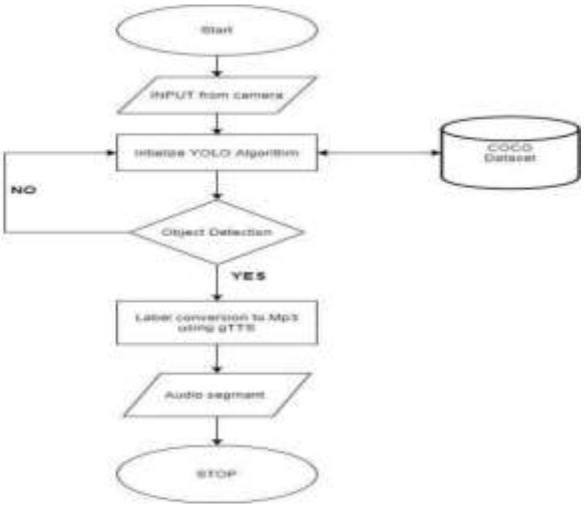


Fig:4.1 Flowchart

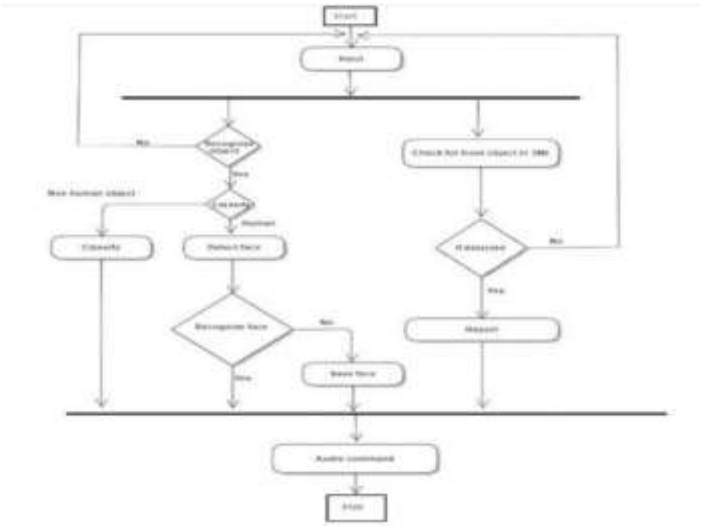


Fig:4.2 Activity Digram

5.1 PYTHONLIBRARIES

5.1.1 OpenCV

OpenCV was designed for computational efficiency and having a high focus on real-time image detection. OpenCV is coded with optimized C and can take work with multicore processors. One of OpenCV's goals is to provide a simple-to- use computer vision infrastructure which helps people to build highly sophisticated vision applications fast. The OpenCV library, containing over 500 functions, spans many areas in vision. Because computer vision and machine learning often goes hand-in-hand, OpenCV also has a complete, general-purpose, Machine Learning Library(MLL).

5.1.2 Tensorflow

TensorFlow, an interface for expressing machine learning algorithms, is utilized for implementing ML systems into fabrication over a bunch of areas of computer science, including sentiment analysis, voice recognition, geographic information extraction, computer vision, text summarization, information retrieval, computational drug discovery and flaw detection to pursue research.

5.1.3 gTTS (Google Text toSpeech)

gTTS, a Python library and CLI tool to interface with Google Translates text-to speech API. Converts from document to a spoken mp3 information, a record like article (bytestring) for additional sound control, or stdout. It highlights adaptable pre-preparing and tokenizing, just as programmed recovery of bolstered dialects.

5.1.4 Matplotlib

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.

5.1.5 Easydict

Easydict is equivalent to a substitute for a dictionary in Python. It is more concise and convenient to use. easyDict allows to access dict values as attributes (works recursively)

5.1.6 pillow

Python Imaging Library (expansion of PIL) is the de facto image processing package for Python language. It incorporates lightweight image processing tools that aids in editing, creating and saving images. Pillow supports a large number of image file formats including BMP, PNG, JPEG, and TIFF.

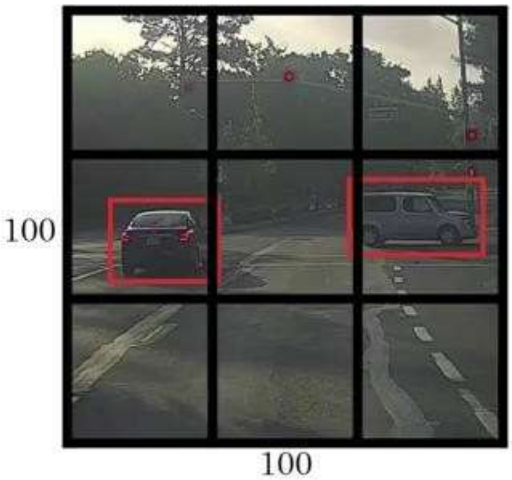
6.1 YOLO

In this project we have used YOLO v3 which is faster than the prior version. It works three times faster, at 320 × 320 YOLO runs 22ms at 28.2 map. It has a similar performance but 3.8× faster. The most notable characteristic of v3 is that it makes 3 distinct scales of detections. YOLO is a fully convolutional neural network and it generates its resultant output by applying a 1 x 1 kernel to a feature map. In YOLO, the recognition is obtained by implementing 1 x 1 detection kernels to three-size feature maps at three different regions in the network.

Within each boundary the network predicts 4 coordinates tx,ty,tw,th. Whereas if a cell is offset in the upper left corner of the image by (cx,cy) and prior bounding boxes has pw, ph width and height respectively then the prediction is done.

6.1.1 How ItWorks?

1. It gonna divide the input image into grid and image classification and localization (i.e) roi are Applied on each grid and it then predict the bounding box and class probability forobject.
2. For training we gonna give labeled data with bounding box and classname



y =	pc
	bx
	by
	bh
	bw
	c1
	c2
	c3

Fig6.1:Grid

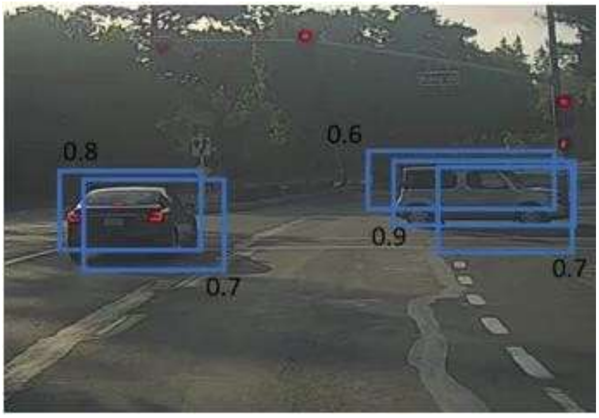
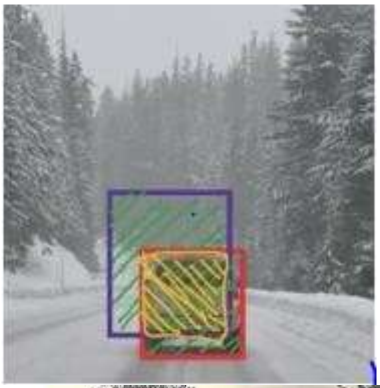
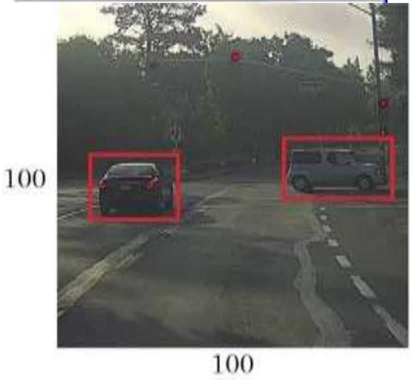


Fig6.2: It gonna remove all the predicted bounding box of an object which has less than the probability and take the maximum probability bounding as predicted bounding box





6.2.1 Distance Metric:

It use squared Mahalanobis distance (effective metric when dealing with distributions) to incorporate the uncertainties from the Kalman filter. Thresholding this distance can give us a very good idea on the actual associations. This metric is more accurate than say, euclidean distance *as w*e are effectively measuring distance between 2 distribution.

$$d(\vec{x},\vec{y}) = \sqrt{(\vec{x}-\vec{y})^{\mathsf{T}}\mathbf{S}^{-1}(\vec{x}-\vec{y})}.$$

Fig 6.3: It gonna calculate the iou between predicted bounding and original bounding box

6.2 DEEPSORT

The most popular and one of the most widely used, elegant object tracking Framework is Deep SORT, an extension to SORT (Simple Real time Tracker).

6.2.1 Kalmanfilter:

1. Kalman filter is a crucial component in deep SORT. Our state contains 8 variables, (u,v,a,h,u',v',a',h') where (u,v) are centres of the bounding boxes, a is the aspect ratio and h, the height of the image. The other variables are the respective velocities of the variables.
2. The variables have only absolute position and velocity factors, since we are assuming a simple linear velocity model. The Kalman filter helps us factor in the noise in detection and uses prior state in predicting a good fit for boundingboxes.
3. For each detection, we create a “Track”, that has all the necessary state information. It also has a parameter to track and delete tracks that had their last successful detection long back, as those objects would have left the scene .Also, to eliminate duplicate tracks, there is a minimum number of detections threshold for the first fewframes.

6.2.2 Efficientalgorithm:

The Hungarian method is a combinational optimization algorithm that solves the Assignment Problem in Polynominal Time and which anticipated later Primal-Dual Method.

6.2.3 Appearance featurevector:

1. This is used to solve theocclusion
2. It is a classifier with deep learning model to extract the feature map of theframe.
3. we just need to pass all the crops of the detected bounding box from the image to this network and obtain the “128 X 1” dimensional feature vector.
4. Now, the updated distance metric will be:

D=Lambda∗Dk+(1−Lambda)∗Da

Dk is the Mahalanobis distance and DaDa is the cosine distance between the appearance feature vectors and Lambda is the weighting factor**.**

6.3 Object Distance

We can say it is the distance between from the predicted object and the user

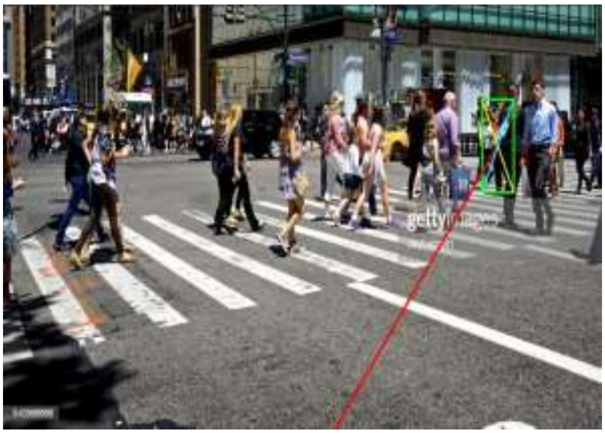


Fig:6.3.1 An example of Object Distance

6.3.1 How Does It Work?

1. We have used basic Euclidean formula for calculating the distance between From the object and theuser.
2. The Bounding box of predicted output is feed inside this algorithm calculate the Distance using the forumula . (Simple Right!)

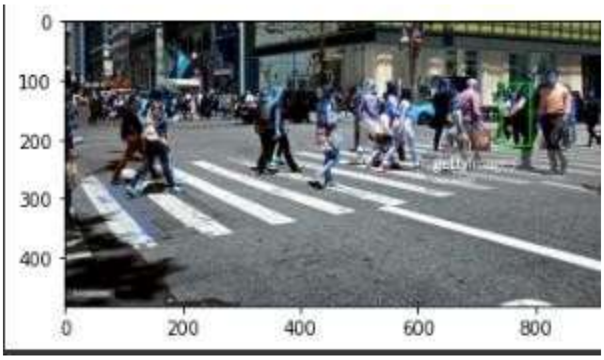


Fig: 6.3.2 An Example of Bounding Box After Object Detection

References:

1. Joseph Redmon, SantoshDivvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, University of Washington, Allen Institute forAI ,Facebook AI Research, 2016.

2. J. Redmon and A. Farha, Yolov3: An incremental improvement.arXiv,2018.

3. J. Redmon and A. Farhadi. Yolo9000: Better, faster, stronger. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 6517–6525.IEEE,2017.

4. ABradski, Gary, and Adrian Kaehler. "OpenCV." Dr. Dobb's journal of software tools 3 (2000).Learning OpenCV 3by Gary Bradski, Adrian Kaehler Publisher: O'Reilly Media, Inc. Release Date: December2016.

5. KedarPotdar, ChinmayPai and SukrutAkolkar, "A Convolutional Neural Network based Live Object Recognition System as Blind Aid", arXiv:1811.10399v1 [cs.CV] 26 Nov 2018https://arxiv.org/pdf/1811.10399.pdf

6. Liam Betsworth, Nitendra Rajput, SaurabhSrivastava,and Matt Jones. Audvert: Using spatial audio to gain a sense of place. InHuman- Computer Interaction–INTERACT 2013, pages 455–462. Springer,2013

7. Evanitsky, Eugene."Portable blind aid device." U.S. Patent No. 8,606,316, 10 Dec. 2013.

8. A.Culjak, D.Abram, T. Pribanic, H. Dzapo and M. Cifrek, A brief introduction to OpenCV,"2012 Proceedings of the 35thInternational Convention MIPRO, Opatija, 2012, pp. 1725-1730.

9. https://pysource.com/2021/01/28/object-tracking-with-opencv-and-python/

10. https://www.analyticsvidhya.com/blog /2018/10/a-step-by-step-introduction-to-the-basic-object-detection-algorithms-part-1/

11. https://www.analyticsvidhya.com/blog /2018/12/practical-guide-object-detection-yolo-framewor-python/

12. https://manivannan-ai.medium.com/object-tracking-referenced-with-the-previous-frame-using-euclidean- distance-4911