OBJECT RECOGNITION AND FACE RECOGNITION FOR VISUALLY IMPAIRED PEOPLE

Manoj Sharma V
UG Scholar
Computer Science & Engineering
REVA University
Bangalore,India
R18CS218@cit.reva.edu.in
Krishnapuram Kalyan Kumar Reddy
UG Scholar
Computer Science & Engineering
REVA University
Bangalore,India
R18CS191@cit.reva.edu.in
Manthan N R
UG Scholar
Computer Science & Engineering
REVA University
Bangalore,India
R18CS219@cit.reva.edu.in
Bindushree D C
Assistant Professor
Computer Science & Engineering
REVA University
Bangalore,India
Bindushree.DC@reva.edu.in

Lalith Kumar Nitesh Kumar Mehtha

UG Scholar

Computer Science & Engineering

REVA University

Bangalore,India

R18CS200@cit.reva.edu.in

Abstract

The technology shown here is designed for visually impaired people to recognize objects and faces in their environment, allowing users to walk around securely without colliding with them. Object identification is related to the video captured by the camera. OpenCV, YOLO, and FaceNet are used to recognize faces and things in the video recorded. When a human face is spotted, the algorithm matches the name to the individual. The user will next be given with an audio version of their identity. Likewise, items spotted in the area will be delivered to the user in audio format along with their names.

Keywords— Visually Impaired, YOLO, OpenCV, Object identification and Face recognition..

I. INTRODUCTION

Many people are disabled, both temporarily and permanently.

There are a lot of blind people all throughout the world. According to the World Health Organization (WHO), about 390 lakh people are fully blind, while another 2850 lakh are purblind, or visually impaired [5]. Many supporting or guiding systems have been produced or are being developed to help people navigate from one place to another in their daily lives.

Common tasks required medium sight, interaction, reading and writing of certain sight, analyzing the surrounding space, and activities involving distant sight are all affected by poor eyesight. Furthermore, any task needs constant eye concentration.

Emerging developments in computer vision technology have prompted researchers to focus their efforts on providing solutions for persons with visual impairments. These devices are designed to assist the user in moving about securely.

The prospect of deploying technologies to detect objects [7] and individuals in the immediate surroundings was investigated in this study. The detection result is delivered to the user in digital audio. The capabilities of vision and hearing are similar in many respects. A real-time object detection and facial recognition system is detailed with the goal of making a user aware of the objects and people in his immediate surroundings.

II. LITERATURE SURVEY

The most challenging situations with respect to the blind or visually impaired population is to fight with unemployment [1]. Many schools adapt the existing Braille to educate them but it becomes unachievable due to the high demanding expenses.

Out of 12 million visually impaired people, only 10% of them makes an effort to learn Braille. Using computer vision to read any text in any format and lighting condition a non-expensive wearable device was designed using Raspberry Pi along with a camera to record content around and translate the same to the blind in their choice of language and a sensor to alert the user about the distance with an object. The system is designed including of image processing, machine learning and speech synthesis techniques. The accuracy recorded with both optical character recognition and the object recognition algorithms was found to be 84% [1].

Smart spec [2] produces a voice output for the visually impaired persons using text detection. Specs comprises of an inbuilt camera to capture images and is further analyzed using Tesseract-Optical Character recognition (OCR). Text is converted to speech with open-source software speech synthesizer, eSpeak. Further the headphones produce the speech by TTS. Raspberry Pi acts as an interface between camera, sensors, IP, and controls the peripheral units.

In the field of electronic travel aids (ETA) which comes with sensor technology and signal processing, it greatly improves the mobility of visually impaired persons in dynamic conditions. Results are achieved in the field of like integrated environment for assisted movement, acoustical virtual reality (AVR), bioinspired solutions [3].

In this paper, many computers vision technology has been developed to assist blind or visually impaired people. Wayfinding, navigation, and finding daily necessities have all been made easier with the help of camera-based systems [9]. The observer's movement causes all scene objects, whether stationary or non-stationary, to move. As a result, detecting moving objects with a moving observer is critical [9].

It is described how a CNN-based correlation algorithm can help visually impaired persons. Given the wealth of information that can be derived from pictures captured, adding a visual processing unit in the framework of systems that aid persons with visual impairments is urgently important, regardless of the version presented. This research describes a correlation technique that uses cellular neural networks (CNNs) to improve the characteristics of helping systems and provide more information from the surroundings to visually impaired people [4]. Parallel processing can handle the majority of the operations (calculations) in the suggested approach. As a result, the computing time may be reduced, and the computing time does not rise proportionately with the size of the template pictures [4].

III. PROBLEM DEFINITION

Visually Impaired are unable to move because they are unable to recognize the terrain and surroundings [6]. In your daily life, you will repeatedly require assistance and walking support systems. Without vision, it can be difficult for the visually handicapped to navigate a room or hallway without running into objects. Even with assistance, such as a walking stick, avoiding obstacles can be difficult, uncomfortable, and possibly inaccurate.

IV. METHODOLOGY

Identifying several object in a picture is called object detection, and it includes both object localization and object categorization. A first basic method would be to slide a window with variable dimensions and use a network trained on cropped photos to predict the content class each time. Convolutions can be used to automate this procedure, which has a significant computational cost.

The main principle behind YOLO is to place a grid on an image (typically 19x19) in which just one cell, the one holding the center/midpoint of an object, is responsible for identifying that object.

The image recorded will be broken into little grids in this method. The midpoint will be determined using these grids. The midpoint of the bounding box will be bx and by, as well as the width and height will be bw and bh. The confidents will be determined from this, and if the probability of mid-point is equal to or greater than confidents, the object or person name to which the confidents level matched will be predicted.

Intersection Over Union method to assess item localization, which quantifies the overlap between two bounding boxes. Many outputs may be generated when estimating the bounding box of a given object in a particular cell of the grid; Non-Max Suppression helps you identify the object just once. It chooses the box with the highest probability and ignores the other boxes with a lot of overlap (IOU).

Block Diagram

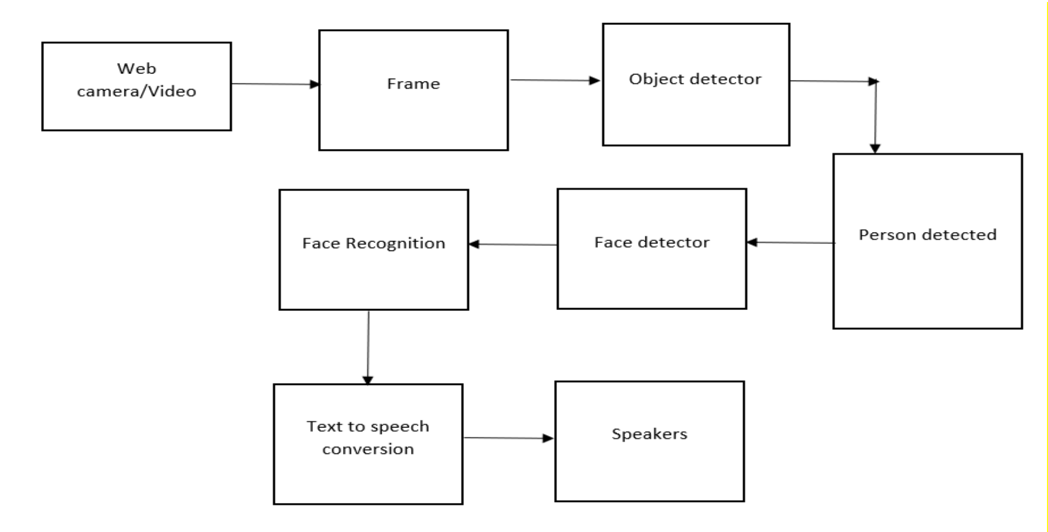

Fig 1: Illustration of the developed framework

The camera will capture live video, and the frames will be drawn from the footage. The objects as well as the person's face will be identified. The object's name and confidentiality will be determined. The audio output is created after the text to speech conversion.

I. SYSTEM IMPLEMENTATION

Python libraries include Scikit-learn for machine learning, OpenCV for computer vision, TensorFlow for neural networks, and more. Real-time computer vision tasks are performed with OpenCV. YOLO provides a framework for object detection in near-real time. Keras is a TensorFlow and other frameworks-compatible deep neural network library.

It's user-friendly, and it makes neural network-based machine learning models extremely straightforward to train. Keras is a useful toolbox for a number of applications since it contains a variety of neural network add-on features including as layers, optimizers, and activation functions.

Some hardware made use of a camera for live video capture and a headphone for audio output.

II. EXECUTION AND RESULTS

The following image shows the final detection results.

Fig 2: Face Recognition

Figure 2 shows an example output of facial recognition on the system, complete with bounding box, name of person detected, and confidence score.

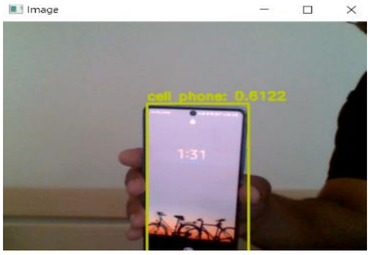

Fig 3: Object Recognition of Bottle

Fig 4: Object Recognition of Cell Phone

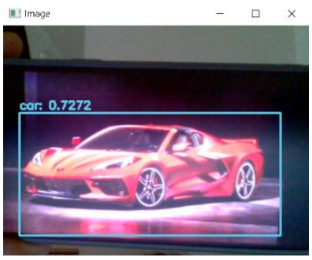

Fig 5: Object Recognition of Car

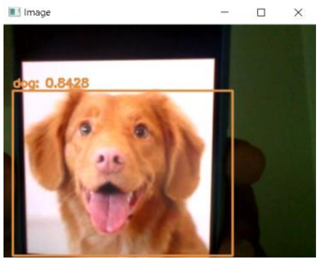

Fig 6: Object Recognition of Dog

Figures 3, 4, 5, and 6 show the sample output, which shows identified objects with bounding boxes, labels, and confidence ratings. These photos were captured for the purpose of object detection.

Table 1: Face Recognition testing

Live Input	Expected Output	Live Output	Status
Kalyan Face(clear background)	Face should recognize as Kalyan	Kalyan-Face	Correct
Kalyan Face(Normal room)	Face should recognize as Kalyan	Kalyan-Face	Correct

The live input, expected output, live output, and status of the face recognition tests are detailed in Table 1.

Input	Expected output	Live Output	Confidence score of YOLO	Status
Bottle	Bottle detected	Bottle	0.7298	Correct
Cell phone	Cell phone detected	Cell phone	0.6122	Correct
Car	Car detected	Car	0.7272	Correct
Dog	Dog detected	Dog	0.8428	Correct

Table 2: Object Recognition testing

Table 2 shows the live input, expected output, live output, YOLO confidence score, and face recognition testing status.

III. OBJECTIVE

The goal of this thesis is to create an object recognition system that can distinguish between 2D and 3D objects in a picture. The characteristics utilized and the classifier used for recognition determine the performance of the object recognition system. This study aims to present a new feature extraction approach for extracting global features and getting local features from the study area. In addition, the study work aims to combine classical classifiers in order to recognize the item.

IV. APPLICATIONS

It is both free and accessible. Real-time results are provided. Reliability, the visually impaired user may rely on the system to provide accurate results. The difference between various objects, such as a chair, a table, etc. may be clearly distinguished depending on the video quality.

V. CONCLUSION

Object categorization and localization within a scene are two of the most challenging aspects of object detection. The application of deep neural networks has aided in the identification of objects. However, implementing such strategies necessitates a significant amount of computing and memory resources. As a consequence, utilizing deep neural network designs for object detection, such as YOLO, produces positive results, demonstrating that they may be utilized for real-time object identification and face recognition, which can benefit the visually impaired.

VI. FUTURE ENHANCEMENT

For object detection at night, the camera's night vision mode should be accessible as an integrated feature. For visual monitoring, the scale of the design remains constant. When the size of the monitored object decreases over time, the background takes precedence over the tracked object. In this case, the item may not be traceable. Splitting and merging with a single camera is not possible in all cases, resulting in a loss of content from a 3D object projection in 2D images.

VII. REFERENCES

[1] M.P. Arakeri, N.S. Keerthana, M. Madhura, A. Sankar, T. Munnavar, “Assistive Technology for the Visually Impaired Using Computer Vision”, International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, pp. 1725-1730, sept. 2018.

[2] R. Ani, E. Maria, J.J. Joyce, V. Sakkaravarthy, M.A. Raja, “Smart Specs: Voice Assisted Text Reading system for Visually Impaired Persons Using TTS Method”, IEEE International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT), Coimbatore, India, Mar. 2017.

[3] V. Tiponuţ, D. Ianchis, Z. Haraszy, “Assisted Movement of Visually Impaired in Outdoor Environments”, Proceedings of the WSEAS International Conference on Systems, Rodos, Greece, pp.386-391, 2009.

[4] L. Ţepelea, A. Gacsádi, I. Gavriluţ, V. Tiponuţ, “A CNN Based Correlation Algorithm to Assist Visually Impaired Persons”, IEEE Proceedings of the International Symposium on Signals Circuits and Systems (ISSCS 2011), pp.169-172, Iasi, Romania,2011.

[5] P. Szolgay,L. Ţepelea, V. Tiponuţ, A. Gacsádi, “Multicore Portable System for Assisting Visually Impaired People”, 14th International Workshop on Cellular Nanoscale Networks and their Applications, pp. 1-2, University of Notre Dame, USA, July 29-31, 2014.

[6] E.A. Hassan, T.B. Tang, “Smart Glasses for the Visually Impaired People”, 15th International Conference on Computers Helping People with Special Needs (ICCHP), pp. 579-582, Linz, Austria, 2016.

[7] M. Trent, A. Abdelgawad, K. Yelamarthi, “A Smart Wearable Navigation System for Visually Impaired”, 2nd EAI international Conference on Smart Objects and Technologies for Social Good (GOODTECHS), pp. 333-341, Venice, Italy, 2016.

[8] Jae Sung Cha, Dong Kyun Lim and Yong-Nyuo Shin, “De,sign and Implementation of a Voice Based Navigation for Visually Impaired Persons”, International Journal of Bio-Science and Bio-Technology, Vol. 5, No. 3, pp.61-68, June 2013.

[9] S. Khade, Y.H. Dandawate, “Hardware Implementation of Obstacle Detection for Assisting Visually Impaired People in an Unfamiliar Environment by Using Raspberry Pi”, Smart Trends In Information Technology And Computer Communications, SMARTCOM 2016, vol. 628, pp. 889-895, Jaipur, India, 2016.

[10] R. C. Gonzalez, R. E. Woods and S. L. Eddins, “Digital Image Processing using MATLAB”, Pearson Education, 2004.