

Object Detection and Recognition for Visually Impaired People

Sasikala Rani K, Jeevitha D

Abstract: Good vision is an expensive gift but now a day's loss of vision is becoming common issue. Blind or visually impaired people does not have any conscious about the danger they are facing in their daily life. To help the blind people the visual world has to be transformed into the audio world with the potential to inform them about objects. Various challenges are faced by visually impaired patients even in the familiar environment. Visually impaired individuals are at drawback due to lack of sufficient information about their familiar environment. This project employs a Convolution Neural Network for recognition of pre-trained objects. This project employs in deep learning a deep Neural Network (DNN) for recognizing the object which is captured from the real world. The captured image is compared with some pre trained objects that is stored in dataset .The comparative of the object is based on the shape and size of an objects. In deep neural network, TensorFlow package using a model called Mobile Net SSD that is comparing the real time capture image with pre trained object based on shape, size of the object. If the image is matched with that trained object, it will display the name of the object. Then the name of the object is converted into audio output with the help of gTTS. This will helps to identify and detect what object is present in front of blind people and give output as audio.

Keywords: Visual impairment, Smart vision, Video processing, Segmentation, Deep neural network, MobileNet model

I. INTRODUCTION

Object detection plays a very important role in many applications such as image retrieval, surveillance, robot navigation, etc. Object detection is a technology that emphasizes image processing as well as computer vision. The technology involves in instance detection of objects including buildings, car, etc. Based on the World Health Organization (WHO), there were about 161 million visually impaired people around the world in 2002, about 10% of the total population. From the total population 124 million had low vision and 37 million were blind. In the globe millions of visually impaired people's faces challenges in understanding the environment. These peoples go with alternative approaches for managing their daily routines but they face certain issues regarding navigation difficulties. Various assistive technologies were developed to help and assist visually impaired peoples. The voice vision technology for the totally blind offers to detect the real world objects and converting to sound.

Revised Manuscript Received on March 16, 2020.

* Correspondence Author

Sasikala Rani k *, Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India.
Email: 18epcs001@skcet.ac.in

Jeevitha D, Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, India.

To store some pre trained images .Using webcam to detect the real world object and given the name of an object through audio output to the user. This will helps to the user identify what object present in front of them.

II. LITERATURE REVIEW

A. Intensity range based background subtraction for effective object detection

K. K. Hati, P. K. Sa explains an intensity range based object detection scheme for videos with fixed background and static cameras. The scheme suggests two different algorithms; the first one models the background from initial few frames and the second algorithm extracts the objects based on local thresholding. The strength of the scheme lies in its simplicity and the fact that, it defines an intensity range for each pixel location in the background to accommodate illumination variation as well as motion in the background. The efficacy of the scheme is shown through comparative analysis with competitive methods. Both visual as well as quantitative measures show an improved performance and the scheme has a strong potential for applications in real time surveillance.

B. Multiple objects tracking using extended kalman filter, gmm and mean shift algorithm

H. Santosh explains the current image and the background image is differentiated by approaching conventionally in image processing. Image subtraction based algorithms are mainly used in extracting features of moving objects and take the information in frames. Here three algorithms namely Extended Kalman Filter, Gaussian Mixture Model (GMM), Mean Shift Algorithm are compared in the context of multiple object tracking.

The comparative results show that GMM performs well when there are occlusions. Extended Kalman filter fails because of abnormal behavior in the distribution of random variables when there is nonlinear transformation. It cannot identify multiple objects when there are occlusions. Mean shift algorithm is best suitable for single object tracking and is very sensitive to window size which is adaptive. three algorithms namely Extended Kalman Filter, Gaussian

C. Robust foreground segmentation using improved gaussian mixture model and optical flow

H. fradi and J.L. Dugrlay says in this context, Gaussian Mixture Model (GMM) background subtraction has been widely employed. It is based on a probabilistic approach that achieves satisfactory performance thanks to its ability to handle complex background scenes. However, the background model estimation step is still problematic; to achieve an improved overall performance, motion cue could provide a rich source of information about the scene.

Therefore, in this paper, we proposed a new approach based on incorporating a uniform motion model into GMM background subtraction. By considering these both cues, high accuracy of foreground segmentation is obtained.

D. An effective background reconstruction method for video object detection

Li Cao and Yi Jiang says the background subtraction is an important method to detect the moving objects, and effective background reconstruction is the key for the background subtraction. Based on the idea that the pixel average appearing with high frequency in an image series is the background points, the pixel intensity classification (PIC) algorithm can reconstruct background accurately. In this paper, a new background reconstruction method based on the PIC algorithm is proposed. Through normalization, quantitative statistic, quantization range extension of the pixels of the chosen image sequence is used to reconstruct the background. Based on the background reconstructed by the improved PIC algorithm, we build a moving object detection system with OpenCv.

E. Object detection system

Seema Udgirkar says Visual impairment and blindness caused by various diseases has been hugely reduced, but there are many people who are at risk of age-related visual impairment. Visual information is the basis for most navigational tasks, so visually impaired people are at disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology it is possible to extend the support given to people with visual impairment during their mobility. In this context we propose a system, named Smart Vision, whose objective is to give blind users the ability to move around in unfamiliar environment, whether indoor or outdoor, through a user friendly interface. This paper is focused mainly in the development of the computer vision module of the Smart Vision system.

F. Deep reinforcement learning with parameterized action space for object detection

Zheng Wu, Naimul Mefraz Khan presents Object detection is a fundamental task in computer vision. With the remarkable progress made in big visual data analytics and deep learning, Reinforcement Learning (RL) is becoming a promising framework to model the object detection problem since the detection procedure can be cast as a Markov decision process (MDP). The proposed system uses an active agent exploring in a scene to identify the location of a target object, and learns a policy to refine the geometry of the agent by taking simple actions in parameterized space, which integrates the discrete actions and its corresponding continuous parameters. We then optimize the representation of the generated region proposals with the discriminative multiple canonical correlation analysis (DMCCA).

III. EXISTING SYSTEM

Existing model correctly detects the objects, within a range of 2m to 5m away. When the objects are not present within the range will be unrecognized. In existing model the single image contains multiple objects that cannot be recognized and time delay is very high to detecting an object. This is very uncomfortable to detecting an object for the blind people. Convolutional neural networks can work reliably for object

recognition, but the problem is manual preparation of input data for the learning process. Manual input is the images are stored in local disk. This model does not detect a real world object. Very large quantity of input samples need to be prepared, to overcome this problem using MobilNet SSD to detecting and recognizing an object without specifying any range and without any time delays. Using a webcam can detect an object from real world.

A. PROBLEM IN EXISTING SYSTEM

The problem of existing system is when two object names overlap then the user cannot be able to distinguish the objects. When the person is detected as speaking, then the program cannot be able to speak out until completion of 5 seconds. This is only an optimal which does not require to deal with variety of objects present in the same class. Similarly when two persons are present in the frame then, it is necessary to information about the presence of more than one person. But there is no need of intimating the same person again and again. Tracking objects that present between the frames using overlapping boundary box is still under research. This does not detect and recognize the real world object. It will help only the objects that are stored in local memory.

B. PROBLEM STATEMENT

Various disability and assistive technologies are developed in order to assist blind and visually impaired peoples. Technology has given several facilities for the peoples with disabilities including visual impairment, people with motion disabilities etc. Most of the navigational tasks requires visual information, so visually impaired people at critical situation due to lack of necessary information regarding the surrounding environment in order to do navigational tasks. Visually impaired individuals are at disadvantage due to the lack of sufficient information about their surrounding environment. The recent advances in inclusive technology, peoples are provided with extended support for visual impairment peoples during their mobility. They are very unaware of the objects even in very familiar environment. To overcome this, the voice vision technology for the totally blind provides sophisticated image-to-sound renderings by using a camera.

The camera is used to detect the object and recognized sent the voice output to the user. High quality of images of objects are required by convolutional networks for performing training process. Multiple object detection can be used in convolutional networks with a single image detection which is partially visible and consists of high time delay. This paper employs a deep Neural Network (DNN) for recognizing the object which is captured from real world. The captured image is compared with some pre-trained objects that are stored in dataset.

IV. PROPOSED SYSTEM

The proposed system employs a deep learning (DL) in Deep neural network (DNN) to recognize objects input from a camera and processing on a portable computer system. Capturing an image with multiple objects in real world. That images are compared with pre-trained dataset. If the captured image is matched with trained dataset. The dataset will display the name of an object. The object name is converted into the form of audio output using gTTS.

This will help to blind people to identify the object which is present in front of them. We illustrate our method on the task of image detection in snap. Detection algorithm is used that runs at 0.025 seconds per 320×240 image. The speed up factor of this detection algorithm is 2.5 when compared to our previous image detector. It gives a real time system which can be used for applications in order to help the blind or visually impaired.

V. MODULE DESCRIPTION

Segmentation

Computer vision technology incorporates image segmentation involves in the process of partitioning a single digital image into various segments. Main aim of segmentation is to make simple and modify the representation of actual object or image into a significant segments which is easier to analyse. Boundaries present in the image are relocated through segmentation. Segmentation result provides set of segments that collectively cover the entire image. In other words the result is a set of contours has been extracted from the image using edge detection. The pixel present in the region are similar in the aspect of few characteristics and computed property that includes colour, intensity etc. Adjacent regions differs based on the characteristics which is applied to the image.

Object extraction

Object detection is referred as the process of identifying real world instances for example building etc. instances are detected as either image or video from. Extracted features have been used in object detection algorithm in order to recognize the instances from various object categories. To detect an object that are capture from real world. It is commonly used in many application.

Identification of object

To identifying an object is very important first capture images are compared and identified based upon size, shape and coordinate values. Images are converted to grayscale which means the colour image RGB is converted into greyscale. RGB images are 32-bit which is hard for manipulation so it is converted to greyscale. Where grayscale images are 16-bit, which easier for manipulation while compared to RGB images. Binary images are much easier for manipulation. It is a black and white images which encloses zeroes and ones. Both grayscale images are subtracted which eliminates the same portions in both images. Only contrast portions are obtained which is called object extraction.

Detecting an object

Capturing images have some moving objects using background subtraction to detecting a moving object. To perform obtain background subtraction, the first step carried out is to model the background. Then, the obtained incoming frame is subtracted out from the background model. With the background model, detecting a moving object is easier and it detects the moving object. This algorithm is called as “Background Subtraction”. The efficiency of a background subtraction technique correlates with three important steps such as modelling, noise removal and data validation. Background modeling, is the backbone of the Background Subtraction algorithm.

SYSTEM ARCHITECTURE

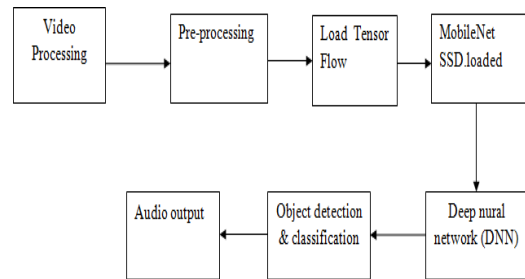


Fig. 1. Video Processing System Architecture

A. Video processing

Video processing is a part of signal processing, in particular image processing, which often employs video filters and where the input and output signals are video files or video streams. Video processing is a part of signal processing which involves in processing the images using video filters. The obtained output through video processing will be the form of video streams or video files. Video processing techniques have been used in various applications including video players, television etc.

B. Preprocessing

Image pre -processing is processing of images using mathematical operations. The output of image pre-processing will be in the form of set of characteristics or will be form of images which is related to original file. It work based on size and shape of an image.

C. Load tensorflow

TensorFlow is considered as most famous deep learning library in the world. **TensorFlow** is created for Python by the Google Brain team which is an open source library. TensorFlow compiles many different algorithms and models together. This is achieved through enabling the user to implement deep neural networks for use in tasks like image recognition/classification and natural language processing.

D. Mobilenetssd

To train the dataset, we use the MobileNet model. This model involves in dividing each images into boundary boxes (S x S grid). N bounding boxes are predicted for each grid with confidence. Accuracy of boundary box is predicted based on the confidence obtained. The confidence achieved also predicts the objects present in boundary box regarding class. The input image is run by SSD through convolutional network which allows to determine and calculate the feature map. A small 3×3 sized convolutional kernel is run on this feature map in order to predict the bounding boxes as well as classification probability. Anchor boxes are used in various aspects with higher accuracies and high-speed outputs as well as higher computation time.

E. Deep neural network

A DNN (Deep Neural Network) is considered as ANN (Artificial Neural Network) consists of several layers between the input layer and output layer. All mathematical manipulation is considered as a layer.

Multiple layers have been present in complex deep neural network hence it is called “Deep” networks. The first perceptions will be shallow which is composed of one input and one output layer and all other layers are present in between. More than three layers (including input and output) qualify as “deep” learning. Deep networks are known for better accuracy of picture class, photograph recovery and item detection. Image segmentation is obtained even for the handwritten digits are also obtained through Deep networks. Deep learning is composed of large neural networks that incorporated teaching machines for automating the works done by visual systems. The teaching machines offers supervised knowledge which highly speed and requires only less computational energy when compared to other strategies used in electronic gadgets. It has a drawback regarding global packaging but this can be avoided by deep gaining knowledge through networks. Unsupervised gaining knowledge can be obtained through deep network which minimizes its down side effects.


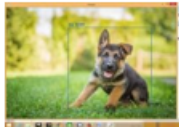

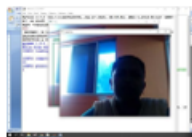
F. Object detection and classification

Object detection is referred as computer technology that incorporates image processing as well as computer vision to detect the semantic instance of objects of particular class from videos and digital images (such as humans, buildings etc.). Face detection, video object segmentation, facial recognition are the computer vision tasks which apparently uses object detection and classification techniques. The technique also used for object tracking. For example, in a football match, tracking the movement of ball, tracking the bat and its movement during cricket match and so on. It also used for tracking the movement of person. Class of each object consist of its special features which facilitates to classify the class of that particular object. Image classification is the processing of assigning label to particular image. Object localization takes place after image classification that involves in drawing another objects for an image as well as drawing bounding box. Object detection involves in combining two processes or tasks to draw a bounding box for objects. Object detection is a challenging task which is difficult to perform. The bounding box that have been drawn for each objects are assigned with class label. All the process put together and referred as object detection.

G. Audio output

The real world object are detected and recognized finally send the audio output to the blind people what object is present in front of the user.

VI. RESULT

objects	description	output
1. Showing an object	It will show saved object in a computer.	
2. Detecting an object	It will detecting an object based on size of the object	
3. Detecting an object with output	It will detecting an object with audio output.	
4. Detecting real time object with audio output	It will capturing an object from Real world using camera and producing sound as a output. it will helps to visually impaired people to identify what object present in front of them.	

VII. CONCLUSION

Recent years, few solutions have been devised to aid blind people or visual impaired in recognizing generic objects in their environment. In this prototype, we investigate the need from blind and visually impaired people. Base on the impetus of the CNN, we develop a blind visualization system that helps blind people better explore the surrounding environment. Among these tools being utilized for this purpose, only a minority is now used daily. This will definitely navigate the blind people to detect the obstacles in front of them. Even there are plenty of techniques are available to guide them but this work efficiently help the blind people to guide and navigate them. In this project implementation as the real time object Recognition & classification using deep learning algorithms after that Classification Object output hear to the blind People to hear Sound what categories of the object present in near we are find it.

SHOWING AN OBJECT

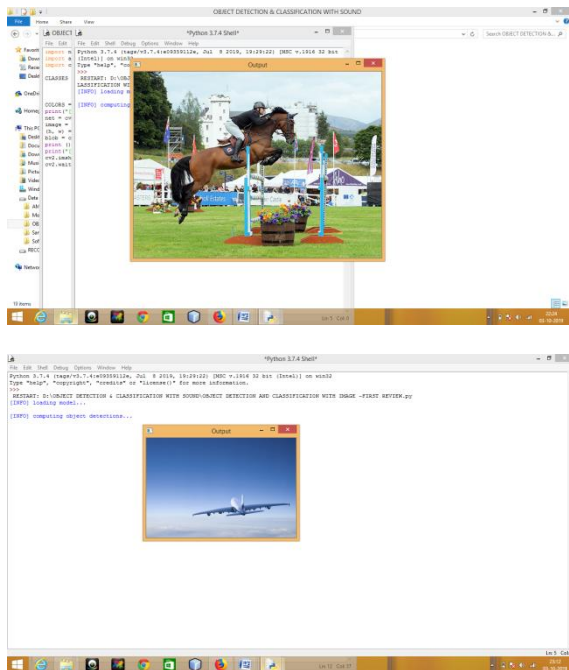


Fig 2. Showing an object

Detecting Real Time Object With Audio Output

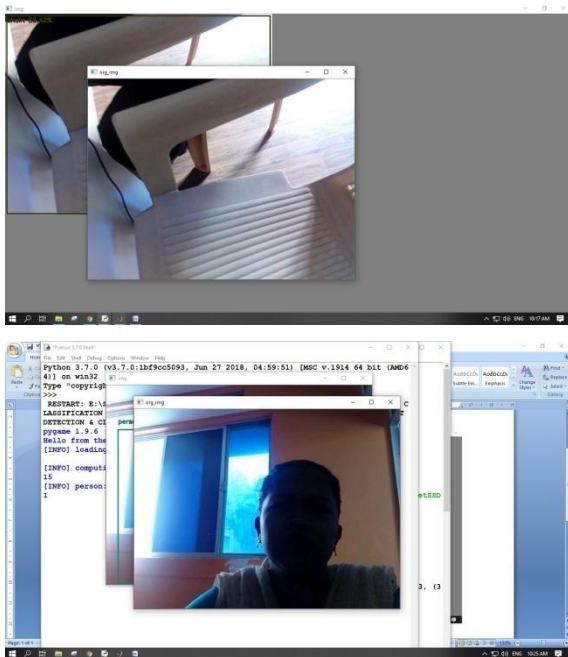


Fig 3. Detecting real time object with audio output.

REFERENCES

1. K. K. Hati, P. K. Sa and B. Majhi, "Intensity Range Based Background Subtraction for Effective Object Detection," in *IEEE Signal Processing Letters*, vol. 20, no. 8, pp. 759-762, Aug. 2013
2. D. H. Santosh and P. G. K. Mohan, "Multiple objects tracking using Extended KalmanFilter, GMM and Mean Shift Algorithm – A comparative study," *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, Ramanathapuram, 2014, pp.1484-1488
3. H. fradi and J.L. Dugrlay, "Robust foreground segmentation using improved Gaussian mixture model and optical flow" *2012 International Conference on Informatics, Electronics & Vision (ICIEV)*, Dhaka, 2012, pp.248-25
4. Li Cao and Yi Jiang "an effective background reconstruction method

for video objects detection" 2012 third international conference on networking and distributed computing.

5. International Journal of Innovative Research in Computer and Communication Engineering, Seema Udgirkar, (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 9, September 2016 Copyright to IJIRCCDOI: 10.15680/IJIRCC.2016. 040904015918 "Object Detection System for Blind People".
6. IEEE International Symposium on Multimedia (ISM) 2018 " Deep Reinforcement Learning with Parameterized Action Space for Object Detection " Zheng Wu, Naimul Mefraz Khan, Lei Gao, and Ling Guan

AUTHORS PROFILE



Dr. K. Sasi Kala Rani, is currently a Professor and head, Dept. of Computer Science and Engineering, Krishna College of Engineering and Technology, Coimbatore. She has completed her B.E in Computer Science and Engineering in 2000, M.E in Computer Science and Engineering in 2005 and Ph.D, in 2014. She has around 17 years of teaching experience. Her area of research includes Networking, Database Systems Modeling and Simulation, Soft Computing Techniques, Evolutionary strategies. She has published around 20 papers in International journals and has published 2 patents. She is a life member of Institution of Engineers, Indian Society for Technical Education, IEEE, CSI. She has published 4 books in her area of interest.



Jeevitha D, is currently pursuing Master degree of computer science and engineering in Sri Krishna College of Engineering and Technology in the year of 2020. She received B.E in CSE from Sri Krishna College of Engineering and Technology with honors during 2014 and 2018 respectively. she attended the more number of conferences in the field of computer science. Her area of interest is Internet Of Things and Machine learning.