IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Blind Assistance Using Machine Learning

Dr. Rajesh K S	Abhishek	Akarsh Gagan	Mahesh	Sachin
Associate Professor and HOD, AIML	BE, CSE	BE, CSE	BE, CSE	BE, CSE
RRCE, Bengaluru	RRCE, Bengaluru	RRCE, Bengaluru	RRCE, Bengaluru	RRCE, Bengaluru

Abstract — In day-to-day life, people with visually impaired suffer more and face lot of inconveniences to move in and out without any proper assistance or help from others. Hence it becomes mandate to provide an user friendly device to assist them. Thus a solution for such people is proposed by an efficient technique like machine learning algorithm. In order to access machine learning technique, the necessary data input are obtained using a technique called Image Classification. The objects in the vicinity where blind people around are captured as images using camera. It can detect every object precisely within some prior distance .The captured images are then converted into audio signal for providing convenient means for assisting the blind people. Thus, a flexible guiding mechanism with user friendliness is developed for helping the blind people.

Index Terms – **blind, machine learning, image to audio conversion, python.**

I. INTRODUCTION

It's undeniably true that all around the world that the outwardly weakened (somewhat or totallyvisually impaired) individuals face a great deal of troubles in perusing, recognizing an item, and keeping away from the impediments. As per the advancement in the present innovation towards the PC vision, computerized camera and compact PCs it is achievable to foster a camera-based innovation that joins PC vision innovation with other business items like OCR frameworks. Perusing is exceptionally fundamental in the present society. Wherever the printed text is as Reports, bank articulations, receipts, eatery menu's and so on so the visually impaired clients face a trouble in perusing these structures. To diminish the disappointed issue, the technique Text to Voice Adaption Using Portable Camera is alluded. The technique which is as of now existed a conveys significant disadvantage in size and not compact. To lessen this disadvantage,we pick an installed stage raspberry pi (Model 3) which goes about as a smaller than normal Stick the camera is conveyed to the raspberry pi board and the got pictures is taken care of to the Raspberry pi board. profit from beginning capital speculation technique is used to restrict and see the text. The text codes from the ROI are seen by the Optical Character affirmation (OCR) and the got picture is dealt with using the python programming language. The text from the OCR is differentiated and the text in the Open CV library to perceiving the headings and edge pixels. In this way, they got pictures are changed over in to the text. The text codes are taken care of to the pytxx library and it is result to astonish clients in talk. Besides, in addition, we add a ultrasonic sensor to caution the outwardly debilitated clients by the speaker in avoiding the tangles.

II. PROBLEM STATEMENT

To build a reliable smart perception system capable of giving description about the vicinity of blind and visually impaired people in a variety of indoor and outdoor environment using GPT-2 algorithm.

III. LITERATURE SURVEY

Image to Audio Conversion using Portable Camera Author:Baskaran and Laxmi

Abstract: This strategy proposes a camera based assistive text perusing for the outwardly debilitated people (to some degree or totally ignorant concerning) read the text in the mark andfrom the items bundling in their regular routines. This proposes a capable and convincing based procedure to depict the area of curiosity (ROI) in which things can be isolated from the confused establishments. This ROI eliminates the text restriction and affirmation to get text information. Limiting the text from the thing ROI is done using the first text limitation estimation by learning the features of work up bearings and edge pixels in the advancement support model. The text characters in the confined text areas are seen by optical individual affirmation (OCR) programming. The codes from the OCR are seen and it is an outcome to the outwardly weakened client in talk.

Object Detection and Identification for Blind People in VideoScene Author: Hanen Jabnoun I, Faouzi Benzarti, Hamid Amiri

Abstract: Vision is one of the extremely fundamental human faculties and it assumes the main part in human discernment about general climate. Thus, more than large number of papershave been distributed regarding these matters that propose an assortment of PC vision items and administrations by growing new electronic guides for the visually impaired. This paper intends to present a proposed framework that re-establishes a focal capacity of the visual framework which is the distinguishing proof of encompassing articles. This technique dependson the nearby element's extraction idea. The reproduction results utilizing SFIT calculation andcentral issues matching showed great exactness for identifying objects. Along these lines, our commitment is to introduce the possibility of a visual replacement framework in view of highlights extractions and matching to perceive and find objects in pictures.

A Novel Elastic Window for Face Detection and Recognitionfrom Video

Author: Shiladitya Chowdhury, Aniruddha Dey, Jamuna KantaSing, Dipak Kumar Basu, Mita Nasipuri

Abstract: Since the beyond quite a long while, face acknowledgment from video has gottencritical consideration because of wide scope of business and law implementation applications,like reconnaissance frameworks, shut circuit TV (CCTV) checking, and so on Human face recognition is the first and significant errand in a unique climate, like video, where commotionconditions, enlightenments, areas of subjects and posture can shift fundamentally starting withone edge then onto the next outline. In this paper, an original versatile window, which doesn'tmake any suspicion about the posture, demeanor or earlier confinement of a face in a video outline is introduced for tracking down limit of face area. The window finds the conceivable face limits by flexibly extending its size utilizing nearby picture inclinations. Before this, a video-outline goes through in a few pre-handling assignments to eliminate clamour, foundation, and so forth and creating slight twofold picture addressing just conceivable face limits and dissipated commotions. Subsequent to distinguishing faces from video outlines, we remove discriminant facial elements from these trimmed face pictures. A multi-class SVM is utilized as a classifier for face acknowledgment in view of these facial elements.

IV. SYSTEM DESIGN ARCHITECTURE

Part I

Camera interface - The photos are captured using the Raspberry Pi camera module, which can be linked to the Raspberry Pi through CSI. It's flexible enough to accommodate increased data rates and identify and process information quickly. It has a clear image quality rating of 11 pixels and can capture HD video at 360 frames per second.

Part II

Raspberry Pi - the Raspberry Pi system is equipped with a quad-core 64-bit Wide Cam 2711 CPU with cortex A72 processor. It also has 2.4GHz Wi-Fi and USB connections, with two USB 2.0 connectors and two USB 3.0 ports. A camera port is available for connecting a Raspberry Pi camera through CSI. Because an SD card does not have a built-in hard disc or memory, it is used for storage.

Part III

Audio jack - The audio jack is utilized to extract the module's audio output and transmit it to the blind people's headphones. On one side, it has a USB port, while on the other, it has a female jack for connecting earphones.

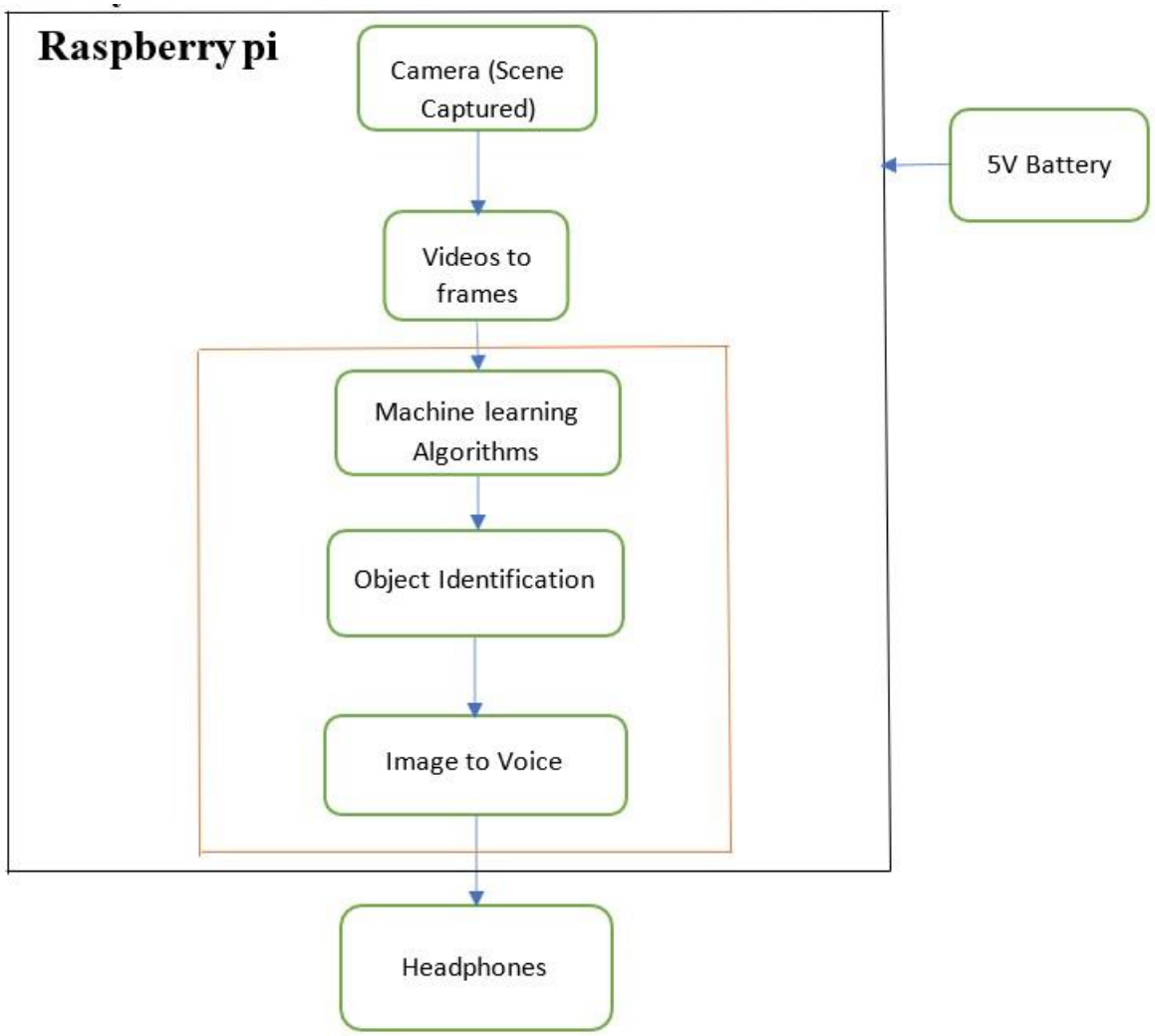


Figure 4.1: System Architecture Design

Part IV

Installing the operating system, connecting all essential parts, and turning on the electricity supply are the main steps in this methodology. This version of the Raspberry Pi employs Python programming and installs all of the essential packages to run the system. The camera, which is connected to the Pi board, is switched on and begins to take pictures. The algorithm that is employed begins to operate in the background and recognizes each object within a predetermined distance before producing a spoken output. The output is sent to the audio jack, which is where the headphones are plugged in.

V. IMPLEMENTATION

The Raspberry pi camera is used to capture the images and it can be interfaced by using the Camera Slot Interface. The Algorithm which is used starts to run in background and detects every object within a prior distance and produces theoutput through speech. The output produced is redirected to the audio jack in which the earphones are connected.

1. GPT-2 Inference Module

Firstly In this module the image will be taken as input from the camera. Then the image will be converted into the text format.

2. Text to Speak Module

In this module the text will be taken from GPT-2 Inference module and then it will be converted into audio format.

3. Main Module

Firstly the image will be taken from the camera then the image will be converted to the text format using GPT-2 module here the GPT-2 uses the concept weights. In this main module all the modules will be imported and called.

4. Audio Play Module

In this module the text will be taken as input and the it will be converted into the audio.

5. Training Module

Input will be taken as image and output will be in text format, here training will take place for visual GPT-2 architecture with loss function as accuracy and optimizer as adam.

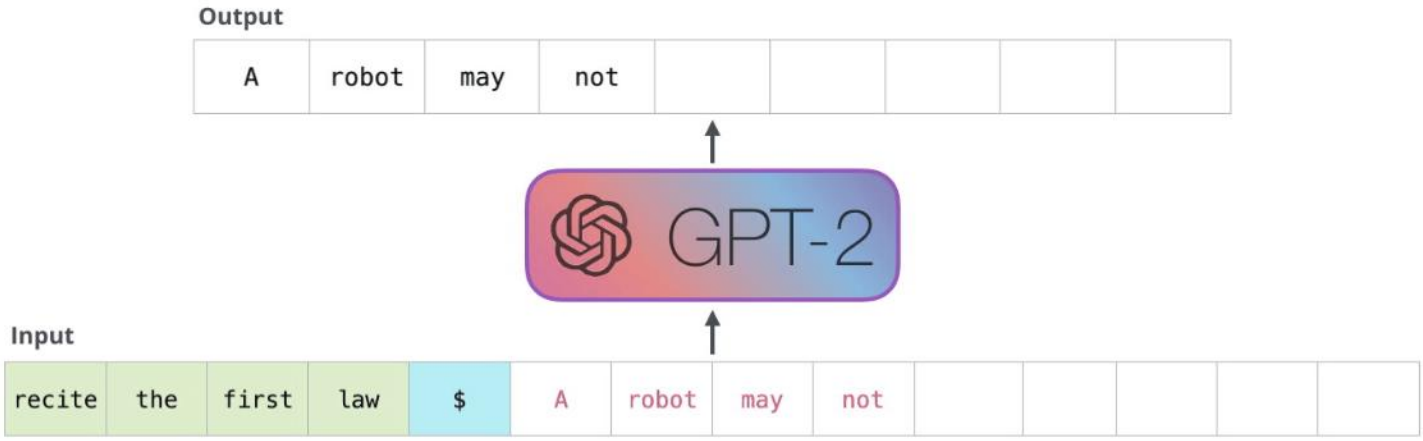
ALGORITHM

GPT-2

Generative Pre-trained Transformer 2 (GPT-2) is an open-source artificial intelligence created by openAI in February 2019.The GPT-2 can generate words, build sentences and paragraphs that are indistinguishable from human-generated content.Trained on 40 GB of textual data, GPT-2 is a very large model containing a massive amount of compressed knowledge from a cross-section of the internet.GPT-2 has a lot of potential use cases. It can be used to predict the probability of a sentence. This, in turn, can be used for text autocorrection.The GPT-2 is built using transformer decoder blocks. GPT-2 model outputs one token at a time.Let's for example prompt a well-trained GPT-2 to recite the first law of robotics:

The way this model actually work is that after each token is produced, that token is added to the sequence of input. And that new sequence becomes the input to the model in its next step. This is an idea called "auto-regression". The unsupervised pre-training was performed using books corpus a dataset of over 7,000 unpublished fiction books from various genres; while other models this dataset was chosen in part because its long passages of continuous text conditioned the model to handle long-range information.

Other available datasets, while larger, were rejected on the basis that they lacked this long-range structure , The fifty library was used to clean .



VI. RESULT AND CONCLUSION

```
Processing..  6  Image
100%|                                                  | 1/1 [00:08<00:00,  8.23s/it]
A tree in the middle of a forest with a sky background.
41.67872929573059
High Performance MPEG 1.0/2.0/2.5 Audio Player for Layers 1, 2 and 3
        version 1.26.4; written and copyright by Michael Hipp and others
        free software (LGPL) without any warranty but with best wishes


Terminal control enabled, press 'h' for listing of keys and functions.

Playing MPEG stream 1 of 1: audio.mp3 ...

MPEG 2.0 L III cbr32 24000 mono

[0:04] Decoding of audio.mp3 finished.
```

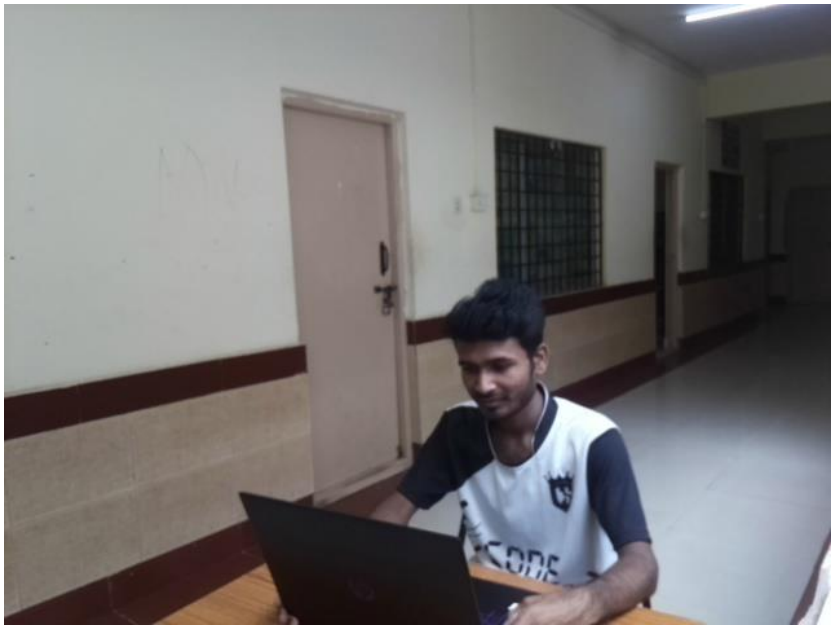
```
Processing..  7  Image
100%|                                                  | 1/1 [00:06<00:00,  6.13s/it]
A view of a building with a cloudy sky.
37.57713317871094
High Performance MPEG 1.0/2.0/2.5 Audio Player for Layers 1, 2 and 3
        version 1.26.4; written and copyright by Michael Hipp and others
        free software (LGPL) without any warranty but with best wishes


Terminal control enabled, press 'h' for listing of keys and functions.

Playing MPEG stream 1 of 1: audio.mp3 ...

MPEG 2.0 L III cbr32 24000 mono

[0:02] Decoding of audio.mp3 finished.
```



```
Processing..  3  Image
100%|                                                  | 1/1 [00:06<00:00,  6.89s/it]
A man is using a laptop computer on a table.
37.62597703933716
High Performance MPEG 1.0/2.0/2.5 Audio Player for Layers 1, 2 and 3
        version 1.26.4; written and copyright by Michael Hipp and others
        free software (LGPL) without any warranty but with best wishes


Terminal control enabled, press 'h' for listing of keys and functions.

Playing MPEG stream 1 of 1: audio.mp3 ...

MPEG 2.0 L III cbr32 24000 mono

[0:03] Decoding of audio.mp3 finished.
```

```
Processing..  16  Image
100%|                                                  | 1/1 [00:07<00:00,  7.06s/it]
A row of motorcycles parked in front of a building.
40.804279088974
High Performance MPEG 1.0/2.0/2.5 Audio Player for Layers 1, 2 and 3
        version 1.26.4; written and copyright by Michael Hipp and others
        free software (LGPL) without any warranty but with best wishes


Terminal control enabled, press 'h' for listing of keys and functions.

Playing MPEG stream 1 of 1: audio.mp3 ...

MPEG 2.0 L III cbr32 24000 mono

[0:03] Decoding of audio.mp3 finished.
```

VII. CONCLUSION & FUTURE WORK

The main aim of the project is to provide information for the blind people which gives a sense of vision by providing the information about their surroundings and objects around them.The proposed system has been successfully tested and has shown quick audio response to the captured input image. It can conclude that this system will greatly help blind people to manage their daily activities more comfortably.

VII. REFERENCE

1. Chen X, Yang J, Zhang J, Waibel A (2004) Automatic detection and recognition of signs from natural scenes. IEEE Transactions on image processing, Vol. 13, No. 1, pp 87-99.

2. Bilgaiyan S, James S, Bhonsle SS, Shahdeo S, Keshavamurthy (2016) Android Based Signboard Detection using Image and Voice Alert System, IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology

3. Chowdhury S, Dey A, Sing JK, Basu DK, Nasipuri M (2014) A Novel Elastic Window for Face Detection and Recognition from Video. International Conference on Computational Intelligence and Communication Networks.

4. Basavaraju R, Chetana H (2015) Traffic Signal Time Analysis and Voice Based App for Visually Impaired Pedestrians,IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology. Text to Audio

5. Baskaran S and Lakshmi D, (2016), Text Conversion using Portable Camera', International Conference on Engineering and Technology (ICET) 6. Hanen Jabnoun, Faouzi Benzarti. Hamid Amiri, 'Object Idetection and identification for blind people in video scene', (2015) 15th International Conference on Intelligent Systems Design and Applications (ISDA)

7. Durette, B., Louveton, N., Alleysson, D., and H'erault, J, 2008. Visuoauditory sensory substitution for mobility assistance: testing the VIBE, Workshop on Computer Vision Applications for the Visually Impaired, Marseille, France.

8. Abboud, S., Hanassy, S., Levy-Tzedek, S., Maidenbaum, S... and Amedi, A., (2014). EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution, Restorative neurology and neuroscience.

9.Jabnoun, H., Benzarti, F. and Amiri, H., (2014). Visual substitution system for blind people based on SIFT description. International Conference of Soft Computing and Pattern Recognition. Tunsia, pp. 300-305.

10. Jafri, R., Ali, SA., Arabnia, HR., and Fatima, S., (2013). Computer vision-based object recognition for the visually impaired in an indoors environment: a survey, The Visual Computer, Springer Berlin Heidelberg, pp. 1-26.