Smart Coat for Visually Impaired using Machine Learning

Anoushka K. AMC Engineering College Department of Electronics and Communication Final Year B E Harshita K. R. AMC Engineering College Department of Electronics and Communication Final Year B E Harshitha C. S. AMC Engineering College Department of Electronics and Communication Final Year B E

Aruna Ramalingam AMC Engineering College Department of Electronics and Communication Associate Professor

ABSTRACT

Vision is god's gift to all the beings, it is dominant feature of life. However there are many human beings who are deficient with eyesight. These visually impaired have to face many changes in their daily life. Only few navigation systems are available for making their lifestyle easier and they are guide cane, trained pet dog and eye substitution devices etc. These systems do help them but not to large extent. In this paper we present a navigation device called as SMART COAT for visually impaired using machine learning. This device helps them to travel with ease and confidence. Smart Coat has 4 Machine Learning Modules ,GPS,GSM of Raspberry-pi and smoke sensor to detect smoke.

General Terms

Object recognition, face recognition, pothole detection, text to speech, YOLO algorithm, convolutional neural network

Keywords

Tesseract, raspberry pi, gps ,gsm, smoke sensor

1. INTRODUCTION

According to "Global Data on Visual Impairments by World Health Organization (WHO), the populace of visually impaired all over the world is 285 million, out of which 39 million are blind and 246 million have very low partial sight and among them 15 million people are from India. Hence India becomes a home for large number of visually impaired. For achieving dreams, goals and living a normal life, independence plays an important role. Even though visually impaired have sharp haptic sensitivity, they find it difficult in fields like education, employment and leading a life independently on daily basis is their most toughest and challenging task.

To assist the blind there are few aids like white Cane, trained pet dog and mobile applications. But they prefer the ancient white cane as the basic tool for their assistance. Hence keeping this fact in mind, efforts were made to upgrade this white cane with latest technology known as smart cane which had sensors in it guide them .Further as the technology improvised different kinds of devices known as Electronic Travel Aids were developed especially for mobility of visually impaired. The feature of this device is to get information about the present location and with this information they will be able to reach their destination but these devices help them in mobility to a smaller extent. The Smart Coat does not wave off the use of white cane. It simplifies their life by supporting them to a larger extent. The smart coat has raspberry pi integrated with four modules- face recognition, object recognition, pothole detection and text to speech with a smoke sensor. All these module's output is obtained through speech output.

2. PROPOSED METHODOLOGY



Fig 1: Methodology

Camera captures the image and the image undergoes preprocessing and required features are extracted. Depending upon the features extracted the image is sent to the respective classification module like object recognition, face recognition, pothole detection or text to speech. Once the captured image has been recognised by the respective module , speech output of the respective module will be obtained. Raspberry-Pi has inbuilt GPS and GSM which provides location and shares the information to the guardian through GSM. Smoke sensor is used in case of fire. The complete methodology is shown in fig 1.

3. BLOCK DIAGRAM **Face Matches** Raspberry -P **Face Recognition Object Matches** Ye Pi-cam processes **Object Recognition** the Speech images and Pothole Detecte Output Classifies based or application Pothole Detection Text to Speech Output PS.GSM.Sm Smoke sensor Output

Fig 2: Block Diagram

The above diagram depicts the basic functioning of our project. The figure 2 is the integration of object recognition, face recognition, text to speech ,pothole detection and smoke sensor with raspberry pi. The raspberry pi has inbuilt camera. The camera captures the image, pre-process it and extracts the required features, later they are classified into different classifiers based on applications that is object recognition ,face recognition, text to speech or pothole detection. Based on the results from these modules we get speech output. GPS provides location and this location is sent to guardian using GSM. Smoke sensor detects smoke or fire and gives the speech output if detected. Raspberry pi is a sequence of small single board of computers. It can be plugged into TV or computer monitor. Raspberry 3 B model comes with an on board 802 .11n wireless LAN adapter,1.Four GHZ 64 bit quad processor .In smart coat we are integrating face recognition, object recognition, pothole detection and text to speech into raspberry pi which will process the images taken by the camera and give output in speech.

3.1 Object Recognition



Fig 3: Flowchart of YOLO Algorithm

In object recognition we are using You Only Look Once (YOLO) algorithm which is already trained to recognise the objects. YOLO algorithm is convolutional neural network for object recognition in real time. It applies single neural network to the full image and then divides the image into regions and predicts the bounding boxes and probabilities for each region. The bounding boxes are weighted by the predicted probabilities. YOLO trains on full image and directly optimises detection performance. The flowchart of YOLO algorithm is shown above in fig 3. COCO names, COCO weights and YOLO config are the files which should be in same folder where the code is present. The system is already trained with various objects for recognition. Initially the camera will switch on and detects the object and compares it with the trained objects. If the trained object and detected object is matched then the object will be recognised with its respective name and speech output will be obtained of the recognised object.

3.2 Face Recognition



Fig 4:Face Model Algorithm

In face recognition we are using face models-Open CV-LBPH face recogniser. Local Binary Pattern Histograms is the algorithm which can recognise both front face and side face. Along with this we have used Haar cascade frontal face.xml file, which contains the features of a frontal face. There are three phases in this face recognition as shown in fig 4. Phase one is data gathering, phase two is to train the recogniser and phase three is the recognition. The first phase of the code will ask for id and name. In second phase we give the id and name, the camera switches on and captures our face and will train the recogniser .Once training is done, next is the third phase where we are testing , the camera will switch on and capture the face , if the captured face matches with the trained face then face is recognised with the respective name and id, then speech output will be obtained.

3.3 Pothole Detection:



Fig 5:Convolutional Neural Network

Pothole detection is done using convolutional neural network. Building blocks of CNN are

- Convolutional Layer-This is the first layer and holds the pixel values of training image. It performs convolution and extracts the high level features.
- Pooling Layer-This layer reduces the spatial size of the image and operates on each feature map.

Max pooling is done which reduces the number of pixels in output from previous layers.

- ReLU Layer-This layer refers to rectified unit. This is the activation function which transforms the weighted input from node into activation of node output. Sigmoid activation function is also used.
- Fully Connected Layer-This layer takes the output of all three layers and use them to classify the image into a label. Then they pass forward to output layer in which every neuron represents a classification label.
- Loss Layer- This is the output layer .It specifies how training penalizes the deviation between predicted and true labels.

CNN uses convolution as basic mathematical operation in their layers. CNN model will be trained with images of potholes and non potholes during training phase. As the model is trained with dataset of potholes and non potholes, when a visually impaired is walking and if there is a pothole or something similar to it, then automatically the camera will capture the image and load the image for processing and classifies as pothole or non potholes as shown in fig 5.

3.4 Text to Speech:



Fig 6 : Working of Text to Speech

Text to speech is a speech synthesis application that converts text from images to voice output. It uses optical character recognition which is electronic conversion of text from images to machine encoded text.. OCR is a technology that will recognize the alphabet through optical mechanism. It reads the printed text by scanning the text optically character by character and converts them into machine encoded text.. OCR is always implemented with machine learning. It specifically uses convolutional neural network .The following procedure outlines procedure for OCR:

- Obtain image
- Perform pre-processing on the image
- Apply algorithm for character recognition
- Post processing

Character recognition is obtained by segmentation, feature extraction and classification. Tesseract OCR is type of OCR engine with matrix matching. Tesseract is an open source command line optical character recognition engine that support for Unicode and comes with an ability to recognize more than 100 languages for various operating systems. It does image processing operations internally before doing the actual OCR. Tesseract will recognize and read the text embedded in the image. It can be trained to work in other languages. The camera captures the image with the text and preprocessing of the image is done as shown in fig 6 to get the text in the image with speech output.

4. **RESULTS**

4.1 Object Recognition:



Fig 7: Object Recognized by Camera and Object Name is Displayed



Fig 8:Scissor Recognised Through Camera and is Displayed



Fig 9: Cat Recognised Through Camera and is Displayed



Fig 10: Pizza Recognised Through Camera and is Displayed



Fig 11: Bus Recognised Through Camera and is Displayed



Fig 12:Chair Recognised Through Camera and is Displayed



Fig 13: Bed Recognised Through Camera and is Displayed

Using YOLO algorithm we have recognized few objects like cell teddy bear doll, scissor, cat, pizza, bus, chair and bed and their results are shown in fig 7, fig 8, fig 9, fig 10, fig11, fig 12 and fig 13. The objects recognised are displayed and speech output is obtained .

4.2 Face Recognition



Fig 14: Camera Capturing Face

141 Reduction Face	
the second second	
1al dire	
Enter Cholce: 3	
101	
10	
140. 20. 20. 20. 20. 20. 20. 20. 20. 20. 2	1.64
	1.22
27, 27, 27, 27, 10, 17, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20	
21, 20, 21, 21, 21, 21, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20	1.20
20, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,	140
In the late has he	1. 1
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 17, 13, 13, 15, 16, 16, 16, 16, 17, 15, 17, 17, 17, 17,	44.5
	100
10, 20, 30, 10, 10, 10, 40, 50, 50, 10, 10, 10, 10, 20, 20, 20, 10, 10, 10,	140
AR,	-14,
13, 13, 12, 13, 13, 13, 13, 13, 13, 2, 2, 3, 2, 3, 3, 3, 3, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,	121
2, 2, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 2, 3, 2, 3, 2, 3, 3, 3, 2, 3, 3, 3, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,	2. 2
2, 2, 3, 3, 4, 5, 1, 2, 2, 2, 3, 3, 3, 4, 5, 3, 3, 5, 6, 6, 6, 6, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8,	8.
	1.
	5277
Do ho ho ho ho ho he he ha	NI

Fig 15: Training of Faces



Fig 16: Recognising of Trained Face



Fig 17: Face Recogized as Harshitha with the id



Fig 18: Face Recognized as Anoushka with id



Fig 19: Face Recognized as Aruna with id



Fig 20: Face Recognized as Dolly with id

In face Recognition first the name of the person and id is taken then the camera captures face as shown in fig 14. Then the recogniser is trained with different faces as shown in fig 15, using Haar cascade frontal face. After training the camera recognises the face with name and id as shown in fig 16. Other faces recognised are shown in fig 17, fig 18, fig 19 and fig 20.

4.3 Pothole Detection



Fig 21: Image Containing Pothole

2 0	0 [[0.3362726 0.6637274]] >>> given image detected path hole >>> RESTART: F:\Frojects 2019\Embedded
1	Fig 22: Output of Pothole Detection

Using Convolutional neural network, pothole detection is done. It takes the image from camera as shown in fig 21 and compares it with the trained dataset and if trained dataset matches with the image captured then "zero and given image is detected pothole " is given in idle as shown in fig 22. If pothole is not detected then "one with given image not detected as pothole" is given in idle. Speech output is obtained.

4.4 Text to Speech



Fig 23: Text Recognised From the Image



Fig 24: Text from the Image is Displayed

Text to speech is obtained using OCR technique and Tesseract software. Camera captures the image and detects the text in the image as shown in fig 23. The recognised text is printed in idle as shown in fig 24. .Speech output is obtained.

Table 1. Details of Samples Tested

Modules	Samples	Results
Object Recognition	80 Objects	55 Object-100%
Recognition	Objects	25 Objects-90%
Face Recognition	15 Faces	100%
Pothole	10	82%
Detection	Potholes	0270
Text to Speech	5 Texts	90%

5. CONCLUSION

Nowadays everyone are dependent on technology for improvising there quality of life .Technology is getting better and better every day. In every field like education, entertainment ,medical, security, agriculture technology has played a huge role for standardization. Many aids and devices have come up to make our life much easier than before.To simplify the lifestyle of visually impaired, we have come up with smart coat . The smart coat has raspberry-Pi, a inbuilt camera which performs different applications through image processing – object recognition, face recognition, pothole detection and text to speech output.It is also helpful in detecting smoke or fire . Raspberry-Pi has inbuilt GPS and GSM for location tracking and providing information to the guardians.

6. **REFERENCES**

- "Object detection using Haar-cascade Classifier", Sander Soo Institute of Computer Science ,University of Tartu, sander92@ut.ee.
- [2] V .Jones, "Rapid object detection using a boosted cascade of simple features", Computer Vision and Pattern Recognition, 2001.
- [3] "Detecting potholes using simple image processing techniques and real world footage", S NEINABER, MJ BOOYSEN AND R S KROON
- [4] Danti, A., Kulkarni, J. & Hiremath, P., 2012. "An Image Processing Approach to Detect Lanes, Pot Holes and Recognize Road Signs in Indian Roads" International Journal of Modeling and Optimization, 2(6), pp. 658-662
- [5] Ultrasonic blindstick with GPS tracking-Vishnu Srinivasan.B.S, Anup Murali.M, Prakash.P,Mr.N.Krishna Prasad
- [6] "Navigation System Using GPS and GSM for Visually Impaired Personals with Nerve Stimulation Feedback" Sakhawat Hossen Rakib, Atika Farhana, A.H.M. Zadidul Karim, Galib Hashmi
- [7] Real-TimeVisualRecognitionwithResultsConvertedto 3DAudio.Rui (Forest) Jiang Earth Science, Stanford forestj,Qian Lin Applied Physics, Stanford ,Shuhui Qu Civil and Environmental Engineering, Stanford
- [8] "David Brown, Tom Macpherson, and Jamie Ward,Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device"Perception, 40(9):1120–1135, 2011.