Outdoor Navigation Assistant for Visually Impaired People

Mansi Phute
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
mansi.phute18@vit.edu

Ashutosh Lahoti
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
ashutosh.lahoti18@vit.edu

Swarali Purandare
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
swarali.purandare18@vit.edu

Yash Nikhare
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
yash.nikhare18@vit.edu

Khushi Bharti
Electronics and Telecommunication
Vishwakarma Institute of Technology
Pune, India.
khushi.bharti18@vit.edu

Abstract*—The computer vision based assistive technology for the blind and visually impaired is a developing area. The assistive technology helps the visually impaired by providing them with a greater independence. By enabling them with their day-to-day activities like indoor and outdoor navigation, obstacle detection, locating the doors and lost objects, etc. In this paper, we present an outdoor mobility assistive system for visually impaired people based on the detection of objects around them and directing them to the correct path. This system consists of three main components: Detection of objects such as potholes, trees, poles, rods, garbage, people , signposts, dogs , cats, car, stone.The system includes an algorithm using depth sensing for detection of potholes and uneven surfaces present in roadsides. Navigation ,The end user (Visually impaired People) of the system will be directed to a clear path over roadsides .Text – Speech conversion feature in system will make the output of system user-friendly by giving audio99881 to end users.

***Keywords*—component; formatting; style; styling; insert (key words)**

I. Introduction

Computer vision, or the ability of artificially intelligent systems to "see" like humans, has been a subject of increasing interest and rigorous research for decades now.There are a large variety of computer vision applications coming in the domain of autonomous vehicles ,sports, gait analysis, robotics, navigation, defense, security, public safety, medical devices, pharmaceuticals, automotive industries, food & drink.computer vision equipped tools and technology has rolled out for an ever-increasing number of uses. Sight is arguably the most important sense for safety and self-preservation. According to the WHO report globally, at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion have a vision impairment that could have been prevented or has yet to be addressed.This 1 billion people includes those with moderate or severe distance vision impairment or blindness due to unaddressed refractive error, as well as near vision impairment caused by unaddressed presbyopia.

This paper discusses the development of a computer vision based system aimed at assisting the visually impaired people in obstacle detection and avoidance in outdoor environments.

There are various approaches to help the visually disabled. Broadly The approaches are A) Using Convolutional Neural Networks B) Using basic Computer Vision Algorithms C) Using lasers and other hardware like IR sensors.

II. Literature Review

There are various approaches to help the visually disabled. Broadly The approaches are using Convolutional Neural Networks, using basic Computer Vision Algorithms or using lasers and other hardware like IR sensors. Many of the papers deal with Convolutional Neural Networks. Two types of deep learning modules are usually used for object detection. The first module is an object position detection module, such as an R-

CNN module (based on region proposals), the module you only look once (YOLO) (based on regression) or a single shot detector (SSD) module (based on regression). The second is an image classification module, such as an Inception series module, a residual network (ResNet) module, or a mobile net module. Apart from these we also came across a system that includes projecting laser patterns, recording the patterns through a monocular video, analyzing the patterns to extract features and then providing path cues for the blind user. Hough Transform and Histogram of Intersections (HoI) is used to detect laser projected lines and to get the accumulator Matrix, k-means clustering and Canny edge detection is also used. It can be used at night times also. But ambient light may blend with the laser and interfere with the scan's accuracy.

III. PREPROCESSING IMAGES

1. GAUSSIAN BLUR OPERATOR

Gaussian Blur is used as a pre-processing technique as a tool to smooth the edges of objects or minimize the noise present in the image. It is implemented in the form of a low pass filter that filters out high frequency signals to minimize the noise.

2. HISTOGRAM OF ORIENTED GRADIENTS

This feature extraction algorithm reduces the extra information in the images which is not necessary like illumination and colors in image. This algorithm is better than other edge detection algorithms as it does not only detect edges but also provide edge direction. This is done by extracting the gradient and orientation of edges.

3. SOBEL OPERATOR

Sobel is an edge detection algorithm. This is a differentiation operator which calculates the approximate gradient of image intensity function. G_x and G_y are used to detect horizontal and vertical derivatives respectively.

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

4. CANNY EDGE DETECTION

Canny Edge Detection algorithm uses Sobel operator but the difference in the two is that in canny there is elimination of noise using gaussian filtering which is not there in Sobel edge detection. Also a double threshold is applied to determine potential edges. If a pixel's gradient value – based on the Sobel differential – is above the high threshold value, it is considered a strong candidate for an edge. If the gradient is below the low threshold value, it is turned off. If the gradient is in between, the pixel is considered a weak candidate for an edge pixel. Therefore we found out that there is less noise in the final output of canny as compared to sobel operator.

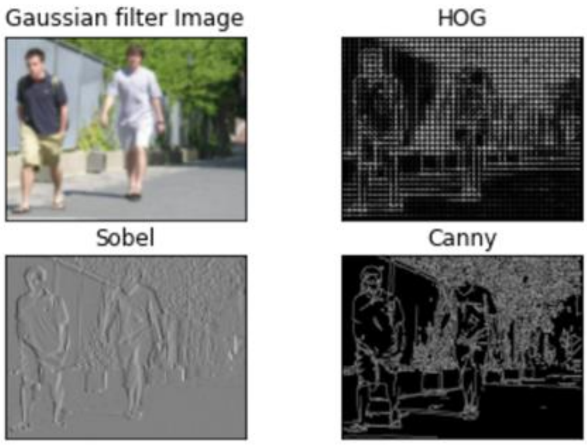

Fig. 1. Sobel, Canny Edge and HOG and Gaussian filter applied on an image with the aim of detecting people

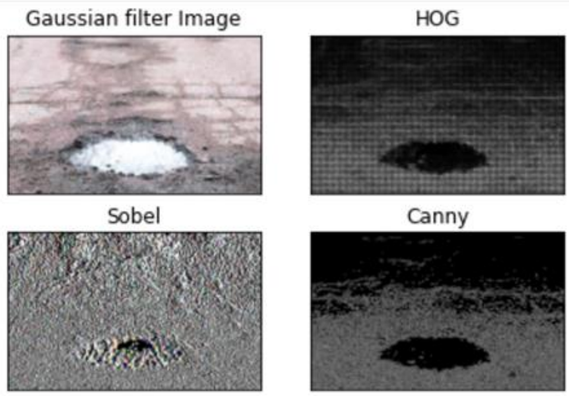

Fig. 2. Sobel, Canny Edge and HOG and Gaussian filter applied on an image with the aim of detecting potholes

IV. OBJECT DETECTION ALGORITHM

1. SVM + HOG

Histogram of Oriented Gradients (HOG) is a method that was proposed by Dalal. In the proposed method the detector provides an output indicating whether the desired object is present within a particular region of the image. The image received by the detector is firstly converted to a set of gradients that are spatially discretized. Regions from the image are selected and converted to the equivalent feature vector, i.e., HOG. The feature vector is provided as input to a binary support vector machine (SVM) classification system. This system gives us the output in binary format. Thus, can be used for classification of the images. This method can also be used to detect the location of the item in the image and draw bounding boxes around them by using mean-shift density estimation algorithm to estimate the position of the object.

In our project we had to use detect the presence of multiple classes as opposed to simply a single class, multiclass SVM is used instead of binary SVM.

2. Logistic Regression

Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes.We have used logistic regression to predict whether the image has a pothole or not.This classifier works best on binary classification rather than multi classification .

3. Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.The algorithm itself learns the features of object.There are three types of layers in a convolutional neural network: convolutional layer, pooling layer, and fully connected layer. Each of these layers has different parameters that can be optimized and performs a different task on the input data.It is computationally efficient .We have used cnn model for multiclass object detection , gave best results from above algorithms .As it is tough to obtain complex relationships using logistic regression. More powerful and compact algorithms such as Neural Networks can easily outperform this algorithm.

4. Sliding Windows

Regions of the images are selected based on parameters such as window shape, size and the step size. After such patches are selected the classifier created is applied to the patches to identify presence of the desired object in the patch. The resultant detections over the image are combined and depicted by a bounding box which surrounds the detected object. Sliding window is used to identify the object in various locations in the image, and when combined with an image pyramid it can also be used to identify the scale of the detected objects.

5. YOLO

YOLO ("you only look once") is one of the popular algorithms because it achieves high accuracy along with being able to run in real-time. The algorithm "only looks once" at the image, i.e. it requires only one forward propagation pass through the network so that it can make predictions. After non-max suppression, it gives the name of the recognized. The algorithm applies a single neural network to the full image, and then divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities. After non-max suppression, it gives the name of the recognized object along with the bounding boxes around them. By using Bounding boxes for object detection, only one object can be identified by a grid. So, for detecting more than one object we go for Anchor box. The steps in non-maximum suppression are:

• Out of the left boxes select the box that has the highest score.

• Compute its overlap with all other boxes, and discard the boxes that overlap it more than IoU value.

• Go back to the step 1 and iterate until there are no more boxes with a less scores than the current selected box.

This discards all boxes and only the best box remains in the last.

6. Faster RCNN

Faster RCNN eliminates the selective search algorithm unlike RCNN and Fast RCNN, it lets the network learn the region proposals. The image is provided as an input to a convolutional network which provides a convolutional feature map. Instead of using selective search algorithm on the feature map to identify the region proposals, a separate network is used to predict the region proposals that is RPN (Region Proposal Network) is small neural network sliding on the last feature map of the convolution layers and predict whether there is an object or not and also predict the bounding box of those objects .The predicted region proposals are then reshaped using a RoI pooling layer which is then used to classify the image within the proposed region and predict the offset values for the bounding boxes. To train this architecture, we use SGD to optimize convolution layers filters, RPN weights and the last fully connected layer weights. The Faster R-CNN is jointly trained with 4 losses : RPN classification (Object foreground/background),RPN regression (Anchor → ROI),Fast RCNN Classification (object classes),Fast RCNN Regression (ROI → Bounding Box).

V. Results

In the pre-processing steps, the Canny Edge Detection did not improve much, but the output from Sobel operator shows a lot of improvement when it comes to the surrounding mild rough surface. Edge detection performed better when paired with a Gaussian filter on the pothole dataset. It was also observed that unlike edge detection, HoG on blurred images performed worse when it came to identifying the pothole region. CNN classifier worked better on the dataset than both the SVM+HOG and Logistic regression models. Both the models work well in almost all lighting conditions. YOLO stands for You Only Look Once. In practical it runs a lot faster than faster RCNN due it's simpler architecture. Unlike faster RCNN, it's trained to do classification and bounding box regression at the same time. Run Speed of YOLO v5 small(end to end including reading video, running model and saving results to file) — 52.8 FPS! Run Speed of Faster RCNN ResNet 50(end to end including reading video, running model and saving results to file) —21.7 FPS. These results are evaluated on NVIDIA 1080 Ti.The orientation of the camera affects the performance of the model.

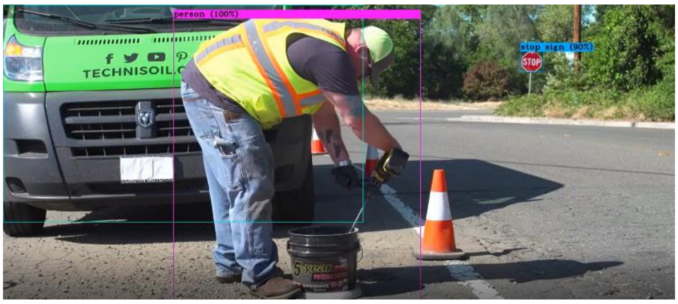

Fig. 3. Object detection in video input using YOLO. In this image the classes of person, car, stop sign have been detected as obstacles

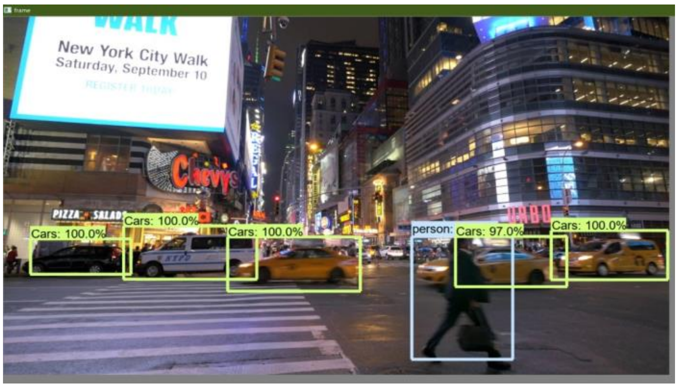

Fig. 4. Object detection in video input using Faster RCNN. In this image the classes of person, car have been detected as obstacles

VI. FUTURE SCOPE

The future work will require research on more efficient object detection techniques so that the model does not fail to detect or mis classify the object . Another major concern is computational efficiency as it is implemented in real time, Latency should be minimum . Gyroscopes can be added , so that orientation of the camera does not affect the performance of the model .Optimisation in algorithm to increase the speed and accuracy of the model ,along with identifying the hyperparameters of the model .

VII. CONCLUSION

We would say both the models i.e. YOLO v5 and faster RCNN struggle to detect people in the distance as they walk into the corridor. This could be attributed to low light and smaller objects. When the crowd gets closer to the camera, both are able to pick up overlapping people. The final comparison b/w the two models shows that YOLO v5 has a clear advantage in terms of run speed. The small YOLO v5 model runs about 2.5 times faster while managing better performance in detecting smaller objects. The results are also cleaner with little to no overlapping boxes.

Acknowledgment

We thank Prof Jyoti Madake for her valuable guidance an assistance throughout the course of the project.

References

[1] João José, Miguel Farrajota, João M.F. Rodrigues, J.M. Hans du Buf Vision Laboratory, Institute for Systems and Robotics (ISR),University of the Algarve (FCT and ISE), Faro, Portugal May 2011International Journal of Digital Content Technology and its Applications 5(5):362-375

[2] Hsueh-Cheng Wang ; Robert K. Katzschmann ; Santani Teng ; Brandon Araki ; Laura Giarré ; Daniela Rus " Enabling Independent Navigation for Visually Impaired People through a Wearable Vision-Based Feedback System " 2014 IEEE International conference of Robotics and Automations (ICRA) ,Singapore 29 May-3 June 2017 DOI: 10.1109/ICRA.2017.7989772

[3] Sylvie Treuillet, Eric Royer, Thierry Chateau, Michel Dhome, Jean-Marc Lavest "Body Mounted Vision System For Visually Impaired Outdoor And Indoor Wayfinding Assistance " Proceedings of the Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments: Assistive Technology

[4] Shankar Sivan , Gopu Darsan ICCCNT '16: Proceedings of the 7th International Conference on Computing Communication and Networking TechnologiesJuly 2016 Article No.: 41 Pages 1–9.

[5] Sylvie Treuillet, Eric Royer. Outdoor/Indoor Vision Based Localization for Blind Pedestrian Navigation Assistance. International Journal of Image and Graphics, World Scientific Publishing, 2010,

[6] Paulo Costa, Hugo Fernandes, Paulo Martins, Joao Barroso, Leontios J. Hadjileontiadis. Obstacle detection using stereo imaging to assist the navigation of visually impaired people. Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012) pg 83 – 93

[7] Lee, Young Hoon & Medioni, Gérard. RGB-D Camera Based Navigation for the Visually Impaired. Computer Vision and Image Understanding volume 149 August 2016 pp 3-20

[8] P. Duh, Y. Sung, L. Fan Chiang, Y. Chang and K. Chen, "V-Eye: A Vision-based Navigation System for the Visually Impaired," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2020.3001500.

[9] Shufei Lin, Ruiqi Cheng , Kaiwei Wang and Kailun Yang :Visual Localizer: Outdoor Localization Based on ConvNet Descriptor and Global Optimization for Visually Impaired Pedestrians

[10] Hugo Fernandes , Paulo Costa , Vítor M Filipe , Vítor M Filipe Stereo vision in blind navigation assistance, Conference: World Automation Congress (WAC), 2010

[11] Simona Caraiman, Otilia Zvoristeanu , Adrian Burlacu and Paul Herghelegiu: Stereo Vision Based Sensory Substitution for the Visually Impaired Article Sensors

[12] Computer Vision for the Visually Impaired: the Sound of Vision System Simona Caraiman, Anca Morar, Mateusz Owczarek, Adrian Burlacu1, Dariusz Rzeszotarski3, Nicolae Botezatu, Paul Herghelegiu, Florica Moldoveanu ,Pawel Strumillo and Alin Moldoveanu

[13] Ruxandra Tapu, Bogdan Mocanu, Titus Zaharia ARTEMIS Department: A Computer Vision System that Ensure the Autonomous Navigation of Blind People in The 4th IEEE International Conference on E-Health and Bioengineering - EHB 2013

[14] Van-Nam Hoang, Thanh-Huong Nguyen, Thi-Lan Le1, Thanh-Hai Tran1, Tan-Phu Vuong2, Nicolas Vuillerme3 Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile Kinect

[15] Aravinda S. Rao, Jayavardhana Gubbi, Marimuthu Palaniswami and Elaine Wong:A Vision-Based System to Detect Potholes and Uneven Surfaces for Assisting Blind People in IEEE ICC 2016 SAC E-Health

[16] Yusuke Kajiwara and Haruhiko Kimura: Object Identification and Safe Route Recommendation Based on Human Flow for the Visually Impaired article sensors

[17] Matteo Poggi, Stefano Mattoccia: A Wearable Mobility Aid for the Visually Impaired based on embedded 3D Vision and Deep Learning in IEEE Workshop on ICT solutions for eHealth 2016

[18] Wafa Elmannai and Khaled Elleithy : Sensor-Based Assistive Devices for Visually-Impaired People: Current Status, Challenges, And Future Directions

[19] Rabia Jafri , Rodrigo Louzada Campos, Syed Abid Ali, And Hamid R. Arabnia: Visual And Infrared Sensor Data-Based Obstacledetection For The Visually Impaired Using The Google Project Tango Tablet Development Kit And The Unity Engine

[20] Theodora S. Brisimi , Christos G. Cassandras, Chris Osgood, Ioannis Ch. Paschalidis And Yue Zhang: Sensing And Classifying Roadway Obstacles In Smart Cities: The Street Bump System

[21] Amita Dhiman And Reinhard Klette: Pothole Detection Using Computer Visionand Learning In IEEE Transactions On Intelligent Transportation Systems

[22] Wan-Jung Chang , Liang-Bi Chen , Chia-Hao Hsu , Jheng-Hao Chen , Tzu-Chin Yang , And Cheng-Pei Lin: Medglasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning For Visually Impaired Chronic Patients

[23] P. E. Rybski, D. Huber, D. D. Morris and R. Hoffman, "Visual classification of coarse vehicle orientation using Histogram of Oriented Gradients features," *2010 IEEE Intelligent Vehicles Symposium*, San Diego, CA, 2010, pp. 921-928, doi: 10.1109/IVS.2010.5547996.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[25] A. A. Fathima, V. Vaidehi, N. Rastogi, R. M. Kumar and S. Sivasubramaniam, "Performance analysis of multiclass object detection using SVM classifier," *2013 International Conference on Recent Trends in Information Technology (ICRTIT)*, Chennai, 2013, pp. 157-162, doi: 10.1109/ICRTIT.2013.6844198