International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

AI Assistance for Visually Challenged People

Pranav Jadhav[1]

[1]*Student, Department of Computer Science, Vishwakarma University, Pune, Maharashtra, India*

Abstract - *This research paper serves the AI/ML model which measures the distance of specific objects from the host without any additional sensors other than the camera's own sensors. It also detects the Indian denominations and report them through hearing aid. Model is made with CNN and OpenCV which integrates the computer vision concepts. Custom models are built which are exported to pkl file for runtime access. We use around 1500 images of Indian notes and augmented them for further processing. The model predicts the denomination from real time video and gives the audio response. It classifies notes with a 92 percent accuracy and is much better for a real time video input. This model can be directly implement for assisting visually challenged people. The model handles the error with 0.4 adjust score. It even works best with just 15fps camera. This makes it cost effective and easily scalable*

***Key Words*: Visually Challenged, Safety Research, Work Distance, Measurement Notes Classification, Video Application**

1. INTRODUCTION

Distance calculation is a major concept used in various day to day events. It's the backbone of various technologies like self-driving cars, defense automation system, robotics industry, navigation system, submarines etc. Here distance is calculated by sensors (majorly ultrasonic sensors). But sensors are not effective in cases where you need to process the input data as per need. It increases the size of the model and also limit the work as only distance is send as a response. So it's not that effective when it comes for helping visually challenged people. This work can be optimized by computer vision. By removing sensors, the work of distance calculation is done by computer vision and the response send by it is processed by AI/ML technology which makes the model more flexible and scalable. In response various parameters are received like object, coordinates, appearance and information related to positions which helps to process the data and get important insights that can help visually challenge people in carrying day to day activity and know the things that happening around them.

We have removed the sensors for distance calculation and use camera to calculate the distance from the host. CMOS sensors [1] which is present in every camera is used for this purpose. We defined the reference objects which help to find the distance. We have made a notes classifier which helps to identify the denomination and tell the user through hearing aid. The to the point results of our model are send as an audio response. It contains three different responses with highest priority for Notes Classification which can halt any other process. Distance calculated by model sends a response at every 10 sec intervals along with object type. Audio response is given which is rendered on hearing aid which will help the person to know the things happening around them. The model is built by combining OpenCV and Machine Learning to coordinate together on the concept of computer vision. The notes classification can be called by saying 'check note' command which then recognizes the notes placed in front of camera.

It's a primitive model build with these use cases at early stage. It can be upgrade to various new tasks and features by using the response received by our model.

2. SOFTWARE ARCHITECTURE

Our model is made with python language. We used two different domains of computer vision. OpenCV for processing the input and deep learning for taking insights from the processed data.

We have used the optical concept of lens for distance calculation. Majorly we have used three formulas for distance calculation. When the object is placed at d distance apart from lens (CMOS sensor rendering camera lens), it forms the image at focus of double convex lens. When the object moves m distance, the image still forms at focal length but with change of size. In Fig. 1 the illustration shows the image formation with respect to CMOS sensor [2]. The ratio of height and distance is equate to get focal length. First we calculated the focal length of CMOS sensor. We physically calculated the distance from defined location. Objects were placed at a distance of 46 inch. Then multiple pictures were taken from camera. Width of the object is calculated manually. This act as a reference for other entities. Then the width of the reference image is calculated through OpenCV. Ratio is taken as state in Fig. 2(a). This gives focal length of CMOS sensor [3]. Then the distance is calculated by just multiplying the focal length with ratio of physically measured width and width calculated by OpenCV from the input frame as can be seen in Fig. 2(b). This gives the distance from the host or camera. It gives score of correctness at every second which can be used to find adjust score Fig. 2(c). Adjust score tells with which proportion the model handles the wrong parameters. We set the confidence value of 35% which state that the result will only be render when the model conforms the object with at least 35% score.

The detection code is fetch with custom weights which build up the yolo model for object detection. It is made with inclination towards greedynms [4] [5]. The objects that a model recognizes is exported in doc format which is used to identify classes of different objects.

A custom CNN model is built for notes classification. It has 5 base layers, 5 maxpool layer and two dense layer. It

identifies the notes with around 92% accuracy at runtime. It contains max 2,097,664 trainable parameters. This model is exported in pkl file for runtime use. It gets rendered at the start of the program but only gets called when the user says 'check note' command. This CNN model is made in the WSGI server which operates in conda environment. The CUDA [6] [7] environment is also establish for fast retrieval.

The architecture of our complete model is shown in Fig. 3. Our model only relies on an input of video frames and don't need any other external integration. The model starts with establishing venv environment. It checks for the audio response from user. If a person commands 'check note' then CNN model track is rendered else yolov4 model. When the Yolo model renders, it fetches the custom weights and populates the model with defined constraints. Here the object type is identified and class is annotate for further processing. With the identification of object, frames are continuously pass for distance calculation. Here the reference data is fed and cfg file [8] of yolo model is fetch. The reference images are processed and coordinates are drawn for taking ratio with coordinates received from frames of the real-time video. Here reference processing image function operates along with focal length function. They calculate the distance and process the coordinates with different orientation, perspective, frame of reference which helps to increase the adjust score. It results a single output with object type and distance from camera. At this level, other integration can also be made like fatalness of object, its speed and various other features. But to keep it easy for user to navigate we have limit the model functionality to distance only. For distance calculation continuous input of frames is necessary [9] (irrespective whether the object is in static or in motion). The calculated distance is send to hearing aid at every 10 sec intervals. The output of 7/10th to 8/10th frame is rendered at every 10 sec. The model operates at each frame but the output is limited to the result of 7th and 8th frame only. This process continuously iterates till the highest priority command i.e. 'check note' receives. This processes shifts it control to CNN model when receives the above command. After receiving it, the model continuously captures the 3 next frames. The 2nd frame is used for processing. If it's not meaningful then 3rd frame is used. This help in keeping the model active and effective. The frame captured are exported to jpg format. The image is preprocess here. Its matrix is convert to array and dimension is expanded to vertical stack. This matrix is passed through CNN model as an input parameter. It gives the probabilities of different denomination. The max probability is passed as an output to audio module. This notes classifier model can be called from any place or while any other process is ongoing. It can halt the audio response, distance calculation, or object recognition process. The audio response is given by using pyttsx [10] module. Audio response is imported from Microsoft female voice directory.

3. Functionality

Our model provides following use cases: Object Detection, Distance Calculation, Speech response, Voice Commands, Notes Classification. It' s workflow can be seen in Fig 4. First the object is recognized through yolonet module and then the distance is calculated by using reference objects. It activates the voice engine and sends the speech response while checking the voice command 'check note'. If matches the command, then renders the notes classifier CNN model and then activates the pyttsx module for speech response. The sequence following of model working is shown in Fig 5.

4. Impact

This model can be directly implemented to assist Visually challenged people for directing their route. They not only know the distance of objects but also gets the idea of object type. This helps them to become cautious and know whether they are surrounded by some danger or not. It also helps them to know the denomination which makes them independent in doing transactions. They can use money without asking anyone. This allow them to trade money easily and prevent fraud assistance from people. The application of this model also servers in providing a self-help to them and do things independently without using others help.

This model can be used in camps for training the needy person which will help them to teach blind people to sense things around them and walk accordingly. This not only helps the blind people but also the trainee. Young generation suffering from eyesight can easily get hold to it and can learn to sense things around them easily. It's easy to use and follow.

Organization like Indian Association for the blind [11] works to teach blind people about new technology and provide them education that a common person takes. Study state that blind people feel isolated and lacks the self-respect as they rely on other people [12] [13] [14]. This model helps them to overcome it and feels the sense of responsible as the model assist them at every move.

This model is majorly made for Visually challenge people and serves its functionality for assisting them. The removal of sensor can alter the industry in various way. It's not effective in conditions like underwater, rainy season, humid weather or conditions where the frame gets affected. But servers an important aspect when need to process the input data. Like in robotics industry [15], these models can be directly implement to make the robots response to stimulus [16]. This will not only reduce the weight and size of robot but also the complexity of the algorithm. It also helps to optimize the model and provide new functionality to them.

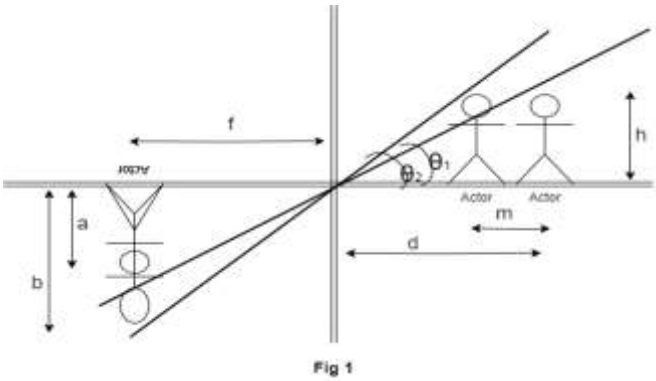

Fig -1: CMOS Sensor response for object movement

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

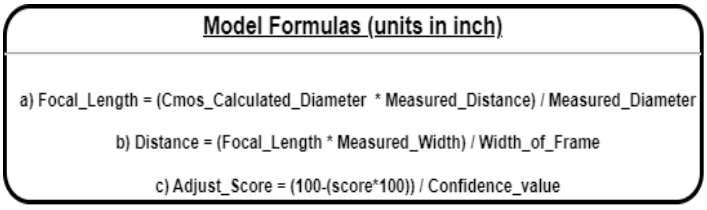

Fig -2: Distance Calculation formulas

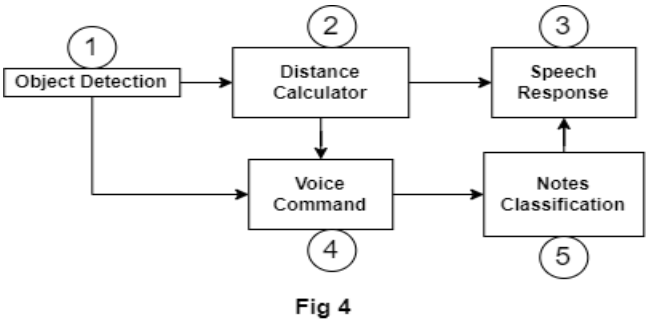

Fig -4: Use Case Diagram

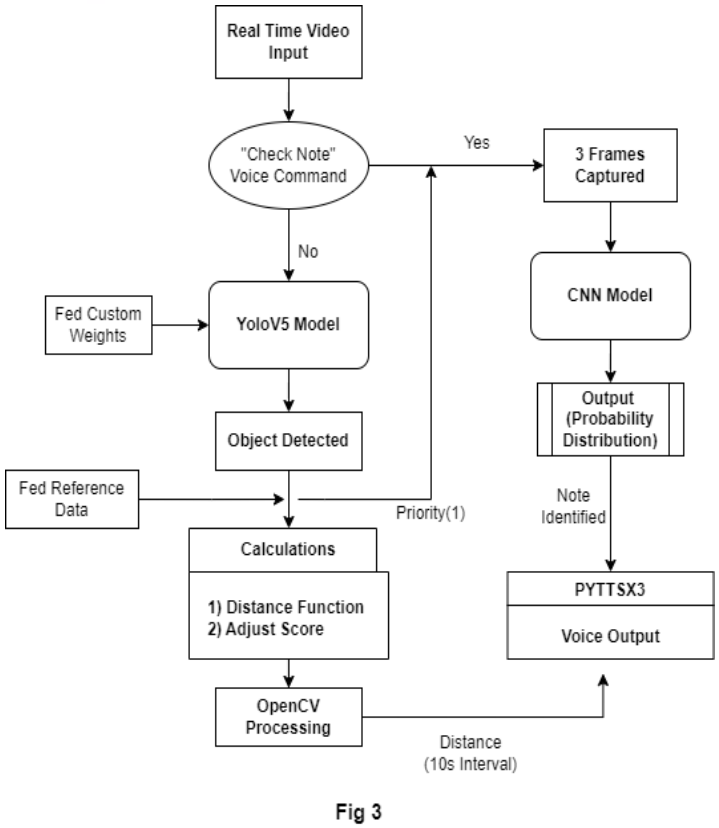

Fig -3: Software Architecture and Data Flow

5. Discussion and future improvement

Our model returns many parameters which can be used to do further tasks. It includes coordinates of the objects which can be used for suggesting the path to follow through. Fatalness of the object can be suggested based on the activity of the object. Speed can be calculate by processing the consecutive frames. These all work can be done without using any sensors. This will help to utilize the space effectively. Various ML models can be implement which

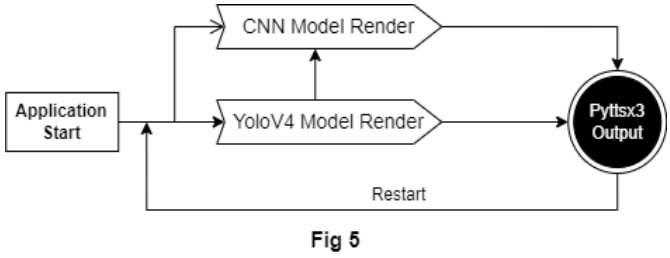

Fig -5: Sequence Flow of OpenCV and Deep Learning

will help blind people to know the entities they encounter while moving. Like a dataset of known people can be made which will identify them and tell their name and information through hearing aid. So whenever they are in public and encounters the know person, the model will automatically detect and gives the information to host user. This model can also be integrated with third party apps like google maps etc. which will help user to navigate around and call their families, relatives or their close one in case they are lost in public. As the model is built in python and use operating system module [17] so can be easily integrated with mobile phones, cloud technology or other online services. The routes are already establish in our model so doesn't need much work to alter the code for new functions. It supports speech and voice interaction and thus can act as an assistance for the user. The model is built in virtual machine venv and conda environment and thus can be easily configured on cloud. Its environment can be easily embedded on hardware.

Pre-trained models of keras application can be used for image processing and to serve transfer learning for creating of class for each and every object present in the universe [18]. This will help blind people to know every important object around them.

6. CONCLUSIONS

In conclusion, our model removes external sensors and their work is done with the camera's own sensor. With this we were able to give extra functionality to our model. The space is optimized and even the complexity of code is reduced. All the code is made in python and thus the architecture is clearly notable. We have merged two different aspects of computer vision i.e. OpenCV and Deep Learning. Each one is use for the purpose its best in. OpenCV for processing and deep learning for model building. The seamless transfer among these two technology makes our model exclusive and builds foundation for connecting the different domains. This interaction makes our model best for assisting Visually Challenged people.

REFERENCES

[1] M. Bigas,E. Cabruja J. Forest , J. Salvi. Review of CMOS image sensors https://www.sciencedirect.com/science/article/abs/pii/S0026269205002764

[2] A. El Gamal, H. Eltoukhy.CMOS image sensors https://ieeexplore.ieee.org/document/1438751

[3] B. Davari. IBM Microelectronics Division, Semiconductor and Research Development Center. CMOS technology: Present and future https://ieeexplore.ieee.org/document/797216

[4] PSRR-MaxpoolNMS: Pyramid Shifted MaxpoolNMS with Relationship Recovery Tianyi Zhang; Jie Lin; Peng Hu; Bin Zhao; Mohamed M. Sabry Al https://ieeexplore.ieee.org/document/9578280

[5] Yu Liu, Lingqiao Liu, Hamid Rezatofighi, Thanh-Toan Do. Learning Pairwise Relationship for Multi-object Detection in Crowded Scenes https://www.researchgate.net/publication/330382642_Learning_Pairwise_Relationship_for_Multi-object_Detection_in_Crowded_Scenes

[6] David Luebke. CUDA: Scalable parallel programming for high-performance scientific computing https://ieeexplore.ieee.org/document/4541126/

[7] G. Quéméner, S. Salvador -nvidia cuda https://www.sciencegate.app/keyword/355688

[8] Mai Zheng, Om Gatla, Duo Zhang, Tabassum Mahmud. Understanding configuration dependencies of file. https://dl.acm.org/doi/abs/10.1145/3538643.3539756

[9] Alexander Conway, Ian Durbach, Alistair, Mcinnes, Rober Harris. Frame-by-frame annotation of video recordings using deep neural networks https://www.researchgate.net/publication/349768936_Frame-by-frame_annotation_of_video_recordings_using_deep_neural_networks

[10] Ravivanshikumar Sangpal; Tanvee Gawand; Sahil Vaykar; Neha Madhavi. JARVIS: An interpretation of AIML with integration of gTTS and Python. https://ieeexplore.ieee.org/document/8993344

[11] Indian Association for the blind: Career and Skill Training, https://theiab.org/pages/what-we-do-career

[12] Karst M.P. Hoogsteen, Sarit Szpiro- A holistic understanding of challenges faced by people with low vision https://www.sciencedirect.com/science/article/abs/pii/S0891422223000951

[13] AbdulQayyum Khan, Mohammad Baqar Abbas, M.K.A. Sherwani, Mohammad Jesan Khan, Naiyer Asif , Danish Kama. Orthopaedic problems in the blind. https://www.sciencedirect.com/science/article/abs/pii/S0976566223001698

[14] Jiaji Wang, Shuihua Wang , Yudong Zhang - Artificial intelligence for visually impaired. https://www.sciencedirect.com/science/article/pii/S0141938223000240

[15] Peng Li, Xiangpeng Li Common Sensors in Industrial Robots: A Review https://www.researchgate.net/publication/334510807_Common_Sensors_in_Industrial_Robots_A_Review

[16] Abu Rahyan. Artificial Intelligence In Robotics: From Automation To Autonomous Systems https://www.researchgate.net/publication/372589771_ARTIFICIAL_INTELLIGENCE_IN_ROBOTICS_FROM_AUTOMATION_TO_AUTONOMOUS_SYSTEMS

[17] Generic Operating System Services-Miscellaneous operating system interfaces. https://docs.python.org/3/library/os.html

[18] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, Qing He. A Comprehensive Survey on Transfer Learning https://arxiv.org/abs/1911.02685