JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Jhalak: Object Detection for Assistive Vision

Akhil Pal
Dept. Computer Engineering
Universal College of Engineering
Mumbai, India
palakhil197@gmail.com

Dhiraj Choudhary
Dept. Computer Engineering
Universal College of Engineering
Mumbai, India
dhirajchd19@gmail.com

Hrishabh Sahu
Dept. Computer Engineering
Universal College of Engineering
Mumbai, India
00hrishabh14@gmail.com

Mrs. Neha Saxena
Dept. Computer Engineering
Universal College of Engineering
Mumbai, India
neha.saxena@universal.edu.in

***Abstract*—**Importance of visual awareness has been highlighted since past decade.

According to WHO, 43 million of the world population are victims of permanent blindness and 295 million have visual impairments.

Our proposed model helps enable the visually impaired to recognize day to day objects more accurately. The proposed tool, "Jhalak" catches the objects in the said models' frame and notifies the user about it. It uses a set of pre trained database/models and compares it with real time objects.

"Jhalak" can be used indoors and specifically stable areas outdoors.

Keywords*—*yolo, coco, object detection, raspberry pi 4b, pi camera, wide angle.

I. Introduction to Jhalak

The sheer amount of struggle any blind human has to put up is unimaginable. Assistance from another person or object is inevitable and needed. Besides knowing the critical information, as a human having the rest 5 senses, they are curious about the beauty of the earth, the joy of exploring the world, and knowing what is happening in front of them. Or they can have their own goods without needing anyone. Every living being on Earth has a set of such sensors, these sensors help them find and hunt for food, water, shelter, etc. Among all the sensors, vision is one of the most important sensors. The main advantages of vision as a sensor, compared to others is its massive range (long distance visibility, at a wide angle), ability to provide complex data that can further be processed to obtain information about the object such as its color, shape etc.

II. Literature Survey

Survey for existing and commercially accessible vision assisting gadgets has been finished. Correlation depends on input mechanics, versatility in the encompassing climate, and usefulness of gadgets. The review stretched out to cutting edge research depends on software engineering also picture handling. The combination of item identification methods in assistive gadgets for outwardly disabled in explicit applications has been considered.

The discoveries from this study recommend that a couple assistive gadgets are accessible monetarily. The monetarily accessible assistive gadgets affected by the mentalities of outwardly hindered and monetary advantages. A blend of a few gadgets can be productive in taking out the limit of an accessible gadget. More than adequate examination has been done for route, correspondence, object identification, and item acknowledgment to help visual debilitated (or not).

[1] The Smart Glass could be a useful tool for persons who are blind or visually handicapped, and it could improve their quality of life. The smart glass is designed for people who want to travel independently or have no other form of assistance, while remaining socially comfortable and secure. It is founded on the basic premise that blind persons do not want to draw attention to themselves when utilizing assistive technology. This paper focuses on the work that has been done in the field of wearable electronics, as well as the functionalities that are available as upgrades. The Smart Glass uses ultrasonic sensors to detect obstacles in front of it in real time and passes the data to the Raspberry for analysis, whether it's an obstacle or a human. It can also tell the user if the object is approaching at a fast pace.

[2] Despite the fact that the idea is identical to the others, this study provides text information to speech engines. Getting an audio warning for an approaching thing can be quite useful for a visually impaired individual. As a result, for our product, we use the pyttsx3 package.

[3] This is a great example of how object detection may assist the vision impaired, but it appears to be too big to wear on a hat. The hat may also be unsuitable for indoor use and may not match the outfit.

[4] RFID-based methods, sonar-based methods, image processing-based methods, and computer vision-based methods are some of the assistive methods (vision substitution) that have been created so far. For object

recognition, computer vision-based technologies appear to be the most promising.

[5] The COCO dataset from Microsoft is the best quality level for estimating the presentation of state-of-the-art PC vision models. The COCO dataset is less notable among general experts, notwithstanding its broad use in the PC vision research field. The COCO dataset, which represents Common Objects in Context, is expected to address a wide

[6] Transmitter and Receiver modules make up the ultrasonic sensor. The beat is shot out by the transmitter and got by the recipient. Assuming an obstacle is put before the sensor, the sent soundwave is reflected back to the sensor. The recipient segment gets the reflected soundwave. It is determined how long passes among transmission and gathering. This data is utilized to process distance. This can be utilized to decide the distance away the recognized things are.

[7] The Google Text to Speech API, or gTTS API, is a well-known and widely used API. The tool is simple to use and comes with several built-in features, including the ability to save a text file as an mp3 file. We don't need to employ a neural network or train a model to convert the file to voice because that is likewise difficult. Instead, we'll finish a task using these APIs. The gTTS API allows you to convert text files into a variety of languages, including English, Hindi, German, Tamil, French, and many others. We can also change the speed of the audio speaking. However, as of the most recent update, we are unable to alter the speech file; it will be generated automatically.

[8] This study gives a minimal expense, restricted handling equipment-based object recognition and stance assessment procedure. The calculation is partitioned into two sections: the recognition stage utilizes the recognition stage involves usage of CNN (convolutional neural network) - explicitly, MobileNet SSD - to perceive and follow targeted objects. The posture assessment stage utilizes sound system backing 3D reproduce the spatial directions of various ORB highlights designated in the box of the perceived article. The last place of the article in not entirely set in stone by taking a weighted normal of similar spatial directions of the sound system compared central issues, with the loads corresponding to the degree of ORB sound system coordinating. This approach was tried on inserted frameworks and viewed as like best in class, GPU-subordinate profound learning calculations for assessing object present. They put our strategy under serious scrutiny in a controlled setting where the item removal (in either the X or Y bearings) could be handily observed. They contrasted our assessment results and a profound gaining based strategy utilizing objects from the Yale-CMU-Berkeley (YCB) dataset, even though our methodology isn't bound to this object determination.

[9] Significant headway has been made in visual identification as of late, and a plenty of magnificent models have been created. Nonetheless, even the most exceptional article recognition networks have a few deficiencies with regards to identifying little targets. Because of their colossal intricacy, they as often as possible neglect to run on versatile gadgets or installed frameworks. An ongoing item identification model called Tiny Fast You Only Look Once (TFYOLO) is made in this workpaper for use in an implanted framework. To start, the kmeans++ technique is utilized to group the dataset, bringing about better priori boxes for the objectives. Second, the system in YOLOv3 is effectively redesigned and streamlined by 3 scales to identify the recently separated highlights, propelled by the multiscale expectation thought in the Feature Pyramid Networks (FPN) method. The reconfigured network is so touchy to little targets. The recommended TFYOLO method is a more modest, quicker, and more proficient organization model that works on start to finish preparing and ongoing item recognizable proof for an assortment of gadgets, as indicated by test information.

[10] Object identification is one area of PC vision where there has been huge turn of events. Finding a particular item among a few articles in a scene is one of the most troublesome and key difficulties in object location. With the development of convolutional brain organizations, exemplary identification strategies for recognizing objects were supplanted. Profound learning-based highlight extraction calculations have been utilized for include extraction starting around 2012, and this has brought about huge advances in this field. This study gives a far-reaching outline of current advances and results in object location utilizing profound learning calculations. Viola-Jones (VJ), histogram of arranged inclination (HOG), a single shot and two-shot finders, benchmark datasets, assessment measurements, accelerate approaches, and present status of-the-craftsmanship object locators are among the subjects covered. On GPU-based inserted frameworks, definite conversations on specific fundamental applications in the article identification fields, like passer-by location, swarm discovery, and continuous item recognition have been introduced.

[11] A computer vision model can perform tasks like providing scene captions, detecting objects, and recognising faces. They have been demonstrated in this research. While supporting people who are visually impaired, these tasks promote their autarchy and independence, there are questions about bias, privacy and possible utility.

[12] The objective of this experiment is to take a look at the ability of one of the most notable Arduino sensors, the ultrasonic HC-SR04 sensor. This sensor is comprised of a transmitter and beneficiary that work in the sonar recurrence band. It is like the Radar guideline in that it is utilized to decide the distance between objects. In view of the discoveries of the examinations, another component called distance goal may be added to the information sheet of this sensor. The trial discoveries uncover that this sensor has goal ability designed in the scope of 7 cm to 11cm. Likewise, the article prescribes raising the sensor opening to work on this capacity for applications that utilize the sensor to segregate between adjoining objects.

[13] Arduino, being an open source IOT device, helps takes inputs (light from a device, finger on any button, text messages) from its sheets and converts them to yields. Basic programming and equipment are needed for Arduino to deliver its best performance.

For instance - switching on an LED, distributing anything on the web by giving a bunch of guidelines to the board's microcontroller (instructions on how it must function) are doable with the aid of programming language and its IDE (Arduino software) on Arduino. Huge number of activities have utilized Arduino enabling the peer feedback and critics. Ranging from familial errands to muddled logical mechanical assembly.

This open-source stage has backers involving worldwide producers, craftsmen, understudies, software engineers and top hierarchical experts. Efforts thrown on this stage by such

experts is unquantifiable in addition to the resources and learning material brought together.
In the long run, Arduino has been the father of millions of undertakings, from daily items, artifacts to complex logical instruments. The available open information that is put through by the experts, local area of creators and specialists are useful for the future generations as well as to aid new research that might require aid of this open-source stage.

[14] The rise of Artificial Intelligence has filled in as a moving force around innovation. Presently we can develop things that we thought were only an idea. One such creation is self-driving vehicles. The days are soon to come when an individual can take care of his business or rest in a vehicle and without contacting the guiding wheel, the gas pedal can in any case arrive at his objective securely. This study proposes a viable model of self-driving vehicles that can drive starting from one spot then onto the next or allude to various kinds of tracks, for example, curved tracks, straight tracks, and straight tracks followed by curved tracks. The camera module is attached to the highest point of the vehicle and the Raspberry Pi computer sends pictures in real time to the Convolutional Neural Network that predicts one of the accompanying bearings. for example, right, left, forward or vertical followed by a message sent by the Arduino for the vehicle controller and therefore the vehicle is moving in the ideal course without human intervention. We can utilize a similar mix to refine our framework.

[15] Although Discovery is a respected profession at the moment, it is a work in progress. It plays an important role in many uses such as image identification, automatic image annotation, and perception. To eliminate visual impairment in visually impaired people, the proposed work can be used effectively to identify objects and their design patterns accurately and to identify among many different objects in each inserted photograph with high accuracy and professionalism. roaming, using a Specific model XY aircraft by calculating its percentage of precision detection and supporting input of images converted into speech. Object discovery also provides your results for many things as well as a variety of ways to find art objects, identify and combine each step to produce.

[16] A creative infrared sensor information-based framework to assist outwardly disabled clients with tracking down deterrents in their way while exploring autonomously inside the presented house. The program is intended for the recently sent off Google Project Tango Tablet Development Kit furnished with a strong imaging processor and a couple of sensors that permit it to follow its development and position in a three-dimensional situation continuously. It utilizes the inherent elements of the Unity motor in the Tango SDK to make a three-dimensional reproduction of the climate, and afterward associates the Unity collider with the client and utilizations it to decide its connection with the recreated cross section to recognize impediments. The client is made aware of any deterrents that might be recognized by sound notices. Notwithstanding, such asset hungry projects require costly hardware to work and subsequently increment complete expenses.

[17] This project uses the Semantic segmentation to get tangible help. Although traditional assisted systems depend on multiple depth sensors and monocular detectors, semantic isolation allows for the solving of multiple navigation vision problems simultaneously and thus quickly applied to visual aids. R-CNN was used directly in understanding the surrounding content. Semantic classification was also used to address cross-street perceptions such as cross-access detection, sidewalks, and blindness. In addition, it has received growing interest and a variety of programs from the field. However, both traditional sensory-based and segregated methods cannot effectively address the physical barriers. Old-fashioned visual aids are turning to multi-sensor integration, eg, integration of RGB-D cameras with ultrasonic sensors have evolved a multimodal stereo matching algorithm to overcome problems encountering regular obstacles such as glass objects, French windows, French doors. To improve the depth measurement of indicating objects with two depth sensors. Polarized signals and sophisticated focus objects are also often tested for visual acuity.

III. PROPOSED SYSTEM

Fig. 3.1 shows the detailed system architecture. Our system works on Raspberry Pi 4B (4GB variant) as the main computer.
A 64GB micro-sd card is used as the primary boot device, on which Raspbian OS legacy (32-bit) is installed.
A wide-angle camera is used as the input source for the object detection algorithm.
This system is only restricted to places where proper lighting is available.
The current iteration of "Jhalak" is equipped with 80 datasets enabling it to detect up to 80 objects.

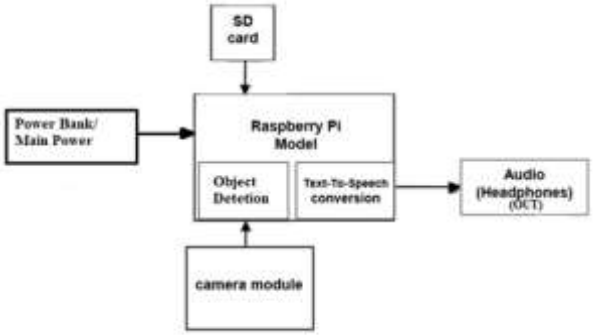

FIG 3.1. system architecture.

The entire program was compiled in python 3.7 using open cv2, imutils, gtts, and essential tools like putty.

We have used the YOLO algorithm paired with a pretrained COCO dataset, so that the detection is fairly quick and accurate.

Wide angle camera (pi camera) is used as an input source to detect objects in front of the user. The detected objects are then conveyed to the user via an audio output.

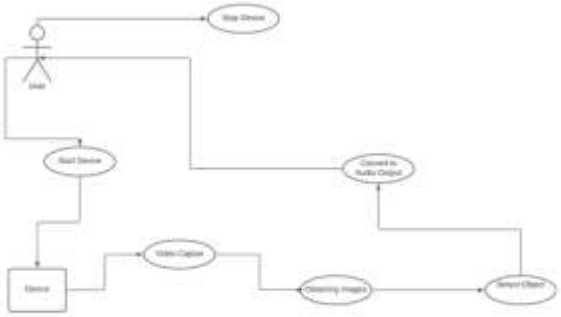

fig.3.2 use case diagram.

Fig.3.2 shows the use case diagram, how the user interacts with the device and how the output is delivered.

Further along, it also shows how information is being processed in the system and the critical stages where information exchange takes place.

Fig3.3 Assembly of Raspberry pi 4B with camera

Fig3.3 Shows us the general assembly of the computing unit. An acrylic case is used to assemble all the components into one unit. This unit will then be mounted on a pair of glasses to make it easy to use.

IV. Analysis and Results

After the literature survey has been completed and research has been done, "Jhalak" was put into simulations by comparing it with models that are available globally.
Models that are available globally, are projects that have the motive to fulfill financial and economic goals. "Jhalak" instead is budget friendly and isn't intended for the market.
Fair pricing comes with its own advantages and disadvantages. "Jhalak" isn't intended to work at night and having 80 pre-trained data only factors in the general objects.
"Jhalak" is made in such a way that it's susceptible to modification to increase its limits given the corresponding costs of modification is tackled.
As mentioned earlier, “Jhalak” can successfully detect upto 80 objects with high accuracy. This can further be increased if the hardware used is more robust.

fig4.1 Actual model

Fig4.1 Shows the end result of the actual model prepared.

V. Future Work

A. *RADAR*

Ultrasonic sensor can be used to create an image map and determine the object distance from the user. This may also help to increase the accuracy of the object acquisition model, as the ultrasonic sensor can also detect the direction of the object.

B. *OCR*

An OCR (for text) can be added to the system to help and ease the life of the user. This may be turn out to be more resource hungry and hence requires better hardware.

C. *FACE RECOGNITION*

A face recognition algorithm can be used to determine who is coming to the user. This may provide a sense of familiarity to the user as it will notify the user of any known / unknown person coming to the user.

Conclusion

A 2021 report shows that worldwide there are 295 million people struggling to make ends meet. "Jhalak" is a good example that aims to make the lives of such people comfortable. Unexplained vision and direction increase the risk factor that can be tested using "Jhalak."
The built-in prototype device is slightly portable, and the device's camera acts as a fake for visually impaired people. This activity helped to draw closer to creating a new "Jhalak" rebellion.
With the ever-growing population, the number of visually impaired people is increasing. India itself has a population of blind people amounting to 12 million, and there is no effective solution for such people.
With the official launch of the International Agency for the Prevention of Blindness Vision Atlas, prototypes like “Jhalak” are a step closer to becoming a reality.

References

[1] Intelligent Smart Glass for Visually Impaired using Deep Learning Machine Vision Techniques and Robot Operating System (ROS). (2015)
[2] Assistive Object Recognition System for Visually Impaired. (2020)
[3] Oculus– Virtual eye for the Blind. (2019)
[4] Assistive System for Visually Impaired using Object Recognition. (2015)
[5] Microsoft COCO: Common Objects in Context. (2014)
[6] Distance Measurement using Ultrasonic Sensor and Arduino. (2017)
[7] Text to Speech (TTS) Conversion in Python. (2016)
[8] Object Detection and Pose Estimation Using CNN in Embedded Hardware for Assistive Technology. (2020)
[9] TF-YOLO: An Improved Incremental Network for Real-Time Object Detection. (2019)
[10] Investigations of Object Detection in Images/Videos Using Various Deep Learning Techniques and Embedded Platforms—A Comprehensive Review. (2020)
[11] Implications of Computer Vision Driven Assistive Technologies Towards Individuals with Visual Impairment. (2019)
[12] Investigating the resolution ability of the HC-SRO4 ultrasonic sensor. (2019)
[13] The working principle of an Arduino. (2018)
[14] Working model of Self-driving car using Convolutional Neural Network, Raspberry Pi and Arduino. (2018)