OBJECT DETECTION AND VOICE GUIDANCE FOR VISUALLY IMPAIRED PERSONS

A Major Project Report Submitted
In partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology in
Information Technology

by

Sandyana Pasunuri **18N31A12C8**

Under the esteemed guidance of

Mr.R.Chandrasekhar
Associate Professor

Department of Information Technology

Malla Reddy College of Engineering & Technology

(Autonomous Institution- UGC, Govt. of India)

(Affiliated to JNTUH, Hyderabad, Approved by AICTE, NBA &NAAC with
'A'Grade)Maisammaguda, Kompally, Dhulapally, Secunderabad – 500100 website:
www.mrcet.ac.in

2018-2022

Malla Reddy College of Engineering & Technology

(Autonomous Institution- UGC, Govt. of India)

(Affiliated to JNTUH, Hyderabad, Approved by AICTE, NBA &NAAC with ‘A’
Grade)Maisammaguda, Kompally, Dhulapally,Secunderabad – 500100
website: www.mrcet.ac.in

CERTIFICATE

This is to certify that this is the bonafide record of the project entitled “Object detection and voice guidance for visually impaired persons”, submitted by Sandyana Pasunuri (18N31A12C8) of B. Tech in the partial fulfillment of the requirements for the degree of Bachelor of Technology in Information Technology, Department of IT during the year 2021-2022. The results embodied in this project report have not been submitted to any other University or Institute for the award of any degree or diploma.

Internal Guide
Mr.R.Chandrasekhar
Associate Professor

Head of the Department
Dr. G.Sharada
Professor

External Examiner

DECLARATION

We hereby declare that the project titled "**Object detection and voice guidance for visually impaired persons**" submitted to Malla Reddy College of Engineering and Technology (UGC Autonomous), affiliated to Jawaharlal Nehru Technological University Hyderabad (JNTUH) for the award of the degree of Bachelor of Technology in Information Technology is a result of original research carried-out in this report. It is further declared that the project report or any part thereof has not been previously submitted to any University or Institute for the award of degree or diploma.

P. Sandyana- 18N31A12C8

ACKNOWLEDGEMENT

We feel honored to place our warm salutation to our college Malla Reddy College of Engineering and Technology (UGC-Autonomous) for giving us an opportunity to do this Project as part of our B.Tech Program. We are ever grateful to our Director Dr. VSK Reddy and Principal Dr.S.Srinivas Rao who enabled us to have experience in engineering and gain profound technical knowledge.

We express our heartiest thanks to our HOD, Dr. G. Sharada for encouraging us in every aspect of our course and helping us realize our full potential.

We would like to thank our Project Guide Mr.R.Chandrasekhar for his regular guidance, suggestions and constant encouragement. We are extremely grateful to our Project Coordinator Mr.V.Narsing Rao for his/her continuous monitoring and unflinching cooperation throughout project work.

We would like to thank our Class Incharge Mr.V.Narsing Rao who in spite of being busy with his/her academic duties took time to guide and keep us on the correct path.

We would also like to thank all the faculty members and supporting staff of the Department of IT and all other departments who have been helpful directly or indirectly in making our project a success.

We are extremely grateful to our parents for their blessings and prayers for the completion of our project that gave us strength to do our project.

With regards and gratitude

P. Sandyana- 18N31A12C8

ABSTRACT

Object detection plays a very important role in many applications such as image retrieval, surveillance, robot navigation, way-finding, etc. In this proposed work, an effective method to propose different approaches to detect indoor signage, stairs and pedestrians. The proposed system detects objects in the web camera frame and alerts the user about its movement by tracking objects location. When an object movement is detected a voice alarm is generated for guiding the blind persons around.

The object detection is achieved by YOLO object detection model. This uses deep neural networks to learn and detect objects. The proposed work used GTTS module for text to voice conversion. This application is useful for surveillance the objects around the visually impaired people. Experimental setup shown that this system was quite fast for object detection and giving voice messages.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 PURPOSE AND OBJECTIVES

According to the World Health Organization, there are approximately 285 million people who are visual impairments, 39 million of them are blind and 246 million have a decrease of Visual acuity. Almost 90% who are visually impaired are living in lowincome countries. In this context, Tunisia has identified 30,000 people with visual impairments; including 13.3% of them are blind. These Visual impairment present severe consequences on certain capabilities related to visual function:

− The daily living activities (that require a vision at a medium distance)

− Communication, reading, writing (which requires a vision closely and average distance)

− Evaluation of space and the displacement (which require a vision far)

− The pursuit of an activity requiring prolonged maintenance of visual attention.

In the computer vision community, developing visual aids for handicapped persons is one of the most active research projects. Mobility aids are intended to describe the environment close to the person with an appreciation of the surrounding objects. These aids are essential for fine navigation in an environment described in a coordinate system relative to the user. An overview of vision substitution modalities and their functionalities.

OBJECTIVE OF WORK

Visual impairment and blindness caused by various diseases has been hugely reduced, but there are many people who are at risk of age-related visual impairment. Visual information is the basis for most navigational tasks, so visually impaired people are at disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology it is possible to extend the support given to people with visual impairment during their mobility. In this context we propose a system, named Smart Vision, whose objective is to give blind users the ability to move around in unfamiliar environment, whether indoor or outdoor, through a user friendly interface.

Applications

This application is mainly designed and proposed to help visually impaired people when they move around on traffic and also at residential places.

This application is aimed at real time alert system for visually impaired persons.

1.2 EXISTING AND PROPOSED SYSTEM

EXISTING SYSTEM

- ❖ Many studies have investigated navigation for blind people. According to these studies, devices and recognition methods can be divided into the following three categories: electronic travel aids (ETAs), electronic orientation aids (EOAs), and position locator devices (PLDs).

- ❖ ETAs are general assistant devices to help visually impaired people avoid obstacles. The sensing inputs of ETAs are mainly classified into depth camera, general camera, radio frequency identification (RFID), ultrasonic sensor, and infrared sensor.

- ❖ EOAs are designed to aid visually impaired people for finding their way in an unknown environment. EOA systems usually need much environmental information to analyze the scope of unknown environment. A combination of a camera and other multiple sensors is usually used to get more information to draw the shapes of passageway and obstacles

- ❖ PLDs are used to determine the precise position of its holder such as devices that use global positioning system (GPS) and geographic information system (GIS) technologies. GPS and GIS-based guiding systems for blind people with user input interfacing (such as voice) intellectually find the current location and give the alert to the blind people if he arrives at his destination area. A pure GPS and

GIS-based navigation systems for general people are easily used to guide user from current location to destination.

Disadvantages

- The above systems are all uses hardware components making it difficult for users to use
- ETA's are too expensive for most families and schools, and mobility specialists aren't trained to use them (or allowed ethically without advanced training). They also set off the sonic wall of burglar alarms.
- Sometimes GPS may fail thanks to certain reasons and therein case you would like to hold a backup map and directions. If you're using GPS on A battery operated device, there could also be A battery failure and you'll need an external power supply which isn't always possible. GPS doesn't penetrate solid walls or structures. it's also suffering from large constructions or structures.

PROPOSED SYSTEM

- This application detection moving objects using Yolo trained model, it uses recurrent neural networks for object detection.
- The objects are identified and its co-ordinates are arrived. Comparing it with previous frames we find the object movement.
- The object class identified is converted to voice by google voice converter.
- The input is video stream from webcam.

Advantages

- This application is cheap, a simple and user friendly
- It is implemented as guidance system to improve the mobility of both blind and visually impaired people in a specific area.

1.3 SCOPE OF THE PROJECT

In computer vision, object detection refers to finding and identifying an object in an image or video. The main steps involved in object detection include feature extraction, feature processing, and object classification. Object detection achieved excellent performance with many traditional methods that can be described from the following four aspects: bottom feature extraction, feature coding, feature aggregation, and classification. 'e feature extraction plays an essential role in the object detection and recognition process. There will be more redundant information which can be modeled to achieve better performance than previous point-of-interest detection. Previously used scale-invariant feature transformations (SIFT) and histogram of oriented gradients (HOG) belong to this category.

The object detection is critical in different applications, such as surveillance, cancer detection, vehicle detection, and underwater object detection. Various techniques have been used to detect the object accurately and efficiently for different applications.

The application developed can detect the objects in the user's surroundings. It can alert the user of the obstacles in his pathway and this way helps the user to navigate from one place to another saving him from tripping anywhere. It will also solve the problem of keeping a special device or a walking stick. However, these proposed methods still have problems with a lack of accuracy and efficiency. To tackle these problems of the object detection, machine learning and deep neural network methods are more effective in correcting object detection.

The need of navigation help among blind people and a broader look at the advanced technology becoming available in today's world motivated us to develop this project. Technology is something which is there to ease tasks for human beings. Hence, in this project, we use technology to solve the problems of visually impaired people. The project aims to help users in navigation with the use of technology. The proposed system consists of 3 modules; object detection, depth estimation, and text to speech. The camera will capture photos, which will be sent to the object detection and depth estimation modules simultaneously.

The detected objects and their corresponding depths will be constructed into a sentence and fed to the text to speech module. The output generated will be an audio file, which will guide the user about the nearby obstacles. The information about the distant obstacles will not be conveyed to the user. This application is mainly designed and proposed to help visually impaired people when they move around on traffic and also at residential places. This application is aimed at real time alert system for visually impaired persons. The object recognition system can be applied in the area of surveillance system, face recognition, fault detection, character recognition etc. The objective of this thesis is to develop an object recognition system to recognize the 2D and 3D objects in the image. The performance of the object recognition system depends on the features used and the classifier employed for recognition. This research work attempts to propose a novel feature extraction method for extracting global features and obtaining local features from the region of interest. Also the research work attempts to hybrid the traditional classifiers to recognize the object. The object recognition system developed in this research was tested with the benchmark datasets like COIL100, Caltech 101, ETH80 and MNIST. The object recognition system is implemented in MATLAB 7.5It is important to mention the difficulties observed during the experimentation of the object recognition system due to several features present in the image. To tackle these problems of the object detection, machine learning and deep neural network methods are more effective in correcting object detection.

This application is mainly designed and proposed to help visually impaired people when they move around on traffic and also at residential places. This application is aimed at real time alert system for visually impaired persons.

CHAPTER 2

LITERATURE SURVEY

This chapter gives the overview of literature survey. This chapter represents some of the relevant work done by the researchers. Many existing techniques have been studied by the researchers on object detection, few of them are discussed below.

Title: Low-altitude small-sized object detection using lightweight feature-enhanced convolutional neural network

Description: Unauthorized operations referred to as "black flights" of unmanned aerial vehicles (UAVs) pose a significant danger to public safety, and existing low-attitude object detection algorithms encounter difficulties in balancing detection precision and speed. Additionally, their accuracy is insufficient, particularly for small objects in complex environments. To solve these problems, we propose a lightweight featureenhanced convolutional neural network able to perform detection with high precision detection for low-attitude flying objects in real time to provide guidance information to suppress black-flying UAVs. The proposed network consists of three modules. A lightweight and stable feature extraction module is used to reduce the computational load and stably extract more low-level feature, an enhanced feature processing module significantly improves the feature extraction ability of the model, and an accurate detection module integrates low-level and advanced features to improve the multiscale detection accuracy in complex environments, particularly for small objects. The proposed method achieves a detection speed of 147 frames per second (FPS) and a mean average precision (mAP) of 90.97% for a dataset composed of flying objects, indicating its potential for low-altitude object detection. Furthermore, evaluation results based on microsoft common objects in context (MS COCO) indicate that the proposed method is also applicable to object detection in general.

Deep Learning Based Object Detection and Recognition of Unmanned Aerial Vehicles

In this study, the methods of deep learning-based detection and recognition of the threats, evaluated in terms of military and defense industry, by unmanned aerial vehicles (UAV) are presented. In the proposed approach, firstly, the training for machine learning on the objects is carried out using convolutional neural networks, which is one of the deep learning algorithms. By choosing the Faster-RCNN and YoloV4 architectures of the deep learning method, it is aimed to compare the achievements of the accuracy in the training process. In order to be used in the training and testing stages of the recommended methods, data sets containing images selected from different weather, land conditions and different time periods of the day are determined. The model for the detection and recognition of the threatening elements is trained, using 2595 images. The method of detecting and recognizing the objects is tested with military operation images and records taken by the UAVs. While an accuracy rate of 93% has been achieved in the Faster-RCNN architecture in object detection and recognition, this rate has been observed as 88% in the YoloV4 architecture

Description: One of the active research topics that maintains its popularity in the field of Computer Vision is the problem of object detection in autonomous cars. Since object detection is a difficult problem, high performance solutions do not work very quickly. Similarly, real-time solutions make compromise on performance. However, due to the nature of autonomous driving, object detection systems must perform in real time and high performance. In this study, Tiny YOLOv3, one of the most successful object detection architectures, was combined with one of the classical object tracking methods, the Kalman filter. A small and real-time object detection system, which increases the model's accuracy without losing its speed, is proposed.

Computer Vision is the branch of the science of computers and software systems which can recognize as well as understand images and scenes. Computer Vision is consist of various aspects such as image recognition, object detection, image generation, image super-resolution and many more. Object detection is widely used for face detection, vehicle detection, pedestrian counting, web images, security systems and self-driving cars. In this project, we are using highly accurate object detection-algorithms and methods such as R-CNN, Fast-RCNN, Faster-RCNN, RetinaNet and fast yet highly accurate ones like SSD and YOLO.

Using these methods and algorithms, based on deep learning which is also based on machine learning require lots of mathematical and deep learning frameworks understanding by using dependencies such as TensorFlow, OpenCV, imageai etc, we can detect each and every object in image by the area object in an highlighted rectangular boxes and identify each and every object and assign its tag to the object. This also includes the accuracy of each method for identifying objects.

Title: A Comparative Study on the Maritime Object Detection Performance of Deep Learning Models

Description: With the increasing volume of maritime traffic, the need for maritime surveillance is also increasing. In this situation, unlike the land environment where various data sets are built, the marine environment lacks data, so the progress of technology research is insufficient. Several object detection methods have been proposed and show excellent performance in areas where sufficient data is secured, but performance verification of existing techniques has not been performed on marine images. In this paper, we compare the performance of marine object detection in various marine images with the latest object detection methods and propose an object detection method suitable for marine environments.

Title: Enhancement and Fusion of Multi-Scale Feature Maps for Small Object Detection

Description: In recent years, deep convolutional neural networks have made breakthrough progress in object recognition and object detection tasks in the field of computer vision, and have achieved great results both in accuracy and speed. However, the detection of small objects is still difficult in the field of object detection, and the accuracy on the common dataset MS COCO is very low. This paper briefly reviews some work in multi-scale object detection algorithms, and then proposes a method of feature enhancement and fusion based on multi-scale feature maps, improving detection accuracy of small objects on MS COCO.

Assistive Object Recognition System for Visually Impaired

The issue of visual impairment or blindness is faced worldwide. According to statistics of the World Health Organization (WHO), globally, at least 2.2 billion people have a vision impairment or blindness, of whom at least 1 billion are blind. In terms of regional differences, the prevalence of vision impairment in low- and middle-income regions is four times higher than in high-income regions.[6] Blind people generally have to rely on white canes, guide dogs, screen-reading software, magnifiers, and glasses to assist them for mobility, however, To help the blind people the visual world has to be transformed into the audio world with the potential to inform them about objects as well as their spatial locations.

Therefore, we propose to aid the visually impaired by introducing a system that is most feasible, compact, and cost- effective. So, we implied a system that makes use of

Raspberry Pi in which you only look once (YOLO v3) machine learning algorithm trained on the coco database is applied. The experimental result shows YOLO v3 achieves state-of-the-art results of 85% to 95% on overall performance, 100% (person, chair, clock, and cell-phone) recognition accuracy. This system not only provides mobility to the visually impaired with that it provides the term that ahead is an XYZ object rather than a sense of obstacle.

CHAPTER 3 SYSTEM ANALYSIS

3.1 HARDWARE REQUIREMENTS:

The development of the application requires the following general and specific minimum requirements

RAM	4GB
Hard Disk	4GB
Output device	VGA and High-Resolution Monitor.

SOFTWARE REQUIREMENTS:

The development of the application requires the following general and specific minimum requirements

Operating System	Windows 7
Programming	Python 3.6 and related libraries

3.2 SYSTEM REQUIREMENT SPECIFICATIONS

Introduction

Computer vision is a rapidly growing dynamic area of research these days. The recent researchers in machine learning promise the improved accuracy of computer vision and artificial intelligence. Here the computers are enabled to think by developing intelligence by learning. There are many types of Machine Learning Techniques and deep learning which are used to achieve computer vision.

Requirement Analysis

Software Requirement Specification (SRS) is the starting point of the software developing activity. As system grew more complex it became evident that the goal of the entire system cannot be easily comprehended. Hence the need for the requirement phase arose. The software project is initiated by the client needs. The SRS is the means of translating the ideas of the minds of clients (the input) into a formal document (the output of the requirement phase.)

Under requirement specification, the focus is on specifying what has been found giving analysis such as representation, specification languages and tools, and checking the specifications are addressed during this activity.

The Requirement phase terminates with the production of the validate SRS document. Producing the SRS document is the basic goal of this phase.

The purpose of the Software Requirement Specification is to reduce the communication gap between the clients and the developers. Software Requirement Specification is the medium though which the client and user needs are accurately specified. It forms the basis of software development. A good SRS should satisfy all the parties involved in the system.

Python

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of its variant implementations. CPython is managed by the non-profit Python Software Foundation.

Functional Requirements

The proposed application should be able to alert user for moving objects, this application is much helpful for the blind people.

3.2.1.1.Product Perspective

The application is developed in such a way that any future enhancement can be easily implementable. The project is developed in such a way that it requires minimal maintenance. The software used are open source and easy to install. The application developed should be easy to install and use.

3.2.1.2. Product features

This application detection moving objects using Yolo trained model, it uses recurrent neural networks for object detection.

The objects are identified and its co-ordinates are arrived. Comparing it with previous frames we find the object movement.

The object class identified is converted to voice by google voice converter.

The input is video stream from webcam.

3.2.1.3. User characteristics

Application is developed in such a way that its users are

- Easy to use
- Error free
- Minimal training or no training

- ❖ Patient regular monitor

3.2.1.4. Assumption & Dependencies

It is considered that the dataset taken fulfils all the requirements.

3.2.1.5. Domain Requirements

This document is the only one that describes the requirements of the system. It is meant for the use by the developers, and will also by the basis for validating the final delivered system. Any changes made to the requirements in the future will have to go through a formal change approval process.

3.2.1.6. User Requirements

User needs to sit and observe the object movement

User needs to move object focusing the webcam

3.2.2. Non-Functional Requirements

- ➢ Video stream from webcam
- ➢ The mike or speaker should be enabled
- ➢ Voice files are stored

3.2.2.1. Product Requirements

The following functionalities need to be checked.

Efficiency: Less time for detection and price forecast for five days

Reliability: Maturity, fault tolerance and recoverability

Portability: can the software easily be transferred to another environment, including install ability.

Usability: How easy it is to understand, learn and operate the software system

3.2.2.2. Organizational Requirements:

Do not block some available ports through the windows firewall. Internet connection should be available

3.2.2.2.1. Implementation Requirements

Quandl.com authentication key for dataset collection, internet connection to install related libraries.

3.2.2.2.2. Engineering Standard Requirements

User Interfaces

User interface is developed in python, which gets input such stock symbol.

Hardware Interfaces

Ethernet

Ethernet on the AS/400 supports TCP/IP, Advanced Peer-to-Peer Networking (APPN) and advanced program-to-program communications (APPC).

ISDN

To connect AS/400 to an Integrated Services Digital Network (ISDN) for faster, more accurate data transmission. An ISDN is a public or private digital communications network that can support data, fax, image, and other services over the same physical interface. can use other protocols on ISDN, such as IDLC and X.25.

3.2.2.3. Operational Requirements

- **Economic**

The developed product is economic as it is not required any hardware interface etc.

- **Environmental**

Statements of fact and assumptions that define the expectations of the system in terms of mission objectives, environment, constraints, and measures of effectiveness and suitability (MOE/MOS).

The customers are those that perform the eight primary functions of systems engineering, with special emphasis on the operator as the key customer.

- **Health and Safety**

The software may be safety-critical. If so, there are issues associated with its integrity level. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.

Systems with different requirements for safety levels must be separated. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

SYSTEM STUDY

Three key considerations involved in the feasibility analysis are

- Economical feasibility
- Technical feasibility
- Social feasibility

ECONOMICAL FEASIBILITY:

- This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY:

- This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY:

- The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system

CHAPTER 4 SYSTEM DESIGN

4.1 DESCRIPTION:

MODULES

- ❖ Object detection
- ❖ Find object location
- ❖ Find object movement
- ❖ Text to speech conversion

This application detection moving objects using Yolo trained model, it uses recurrent neural networks for object detection. The objects are identified and its co-ordinates are arrived. Comparing it with previous frames we find the object movement. YOLO is

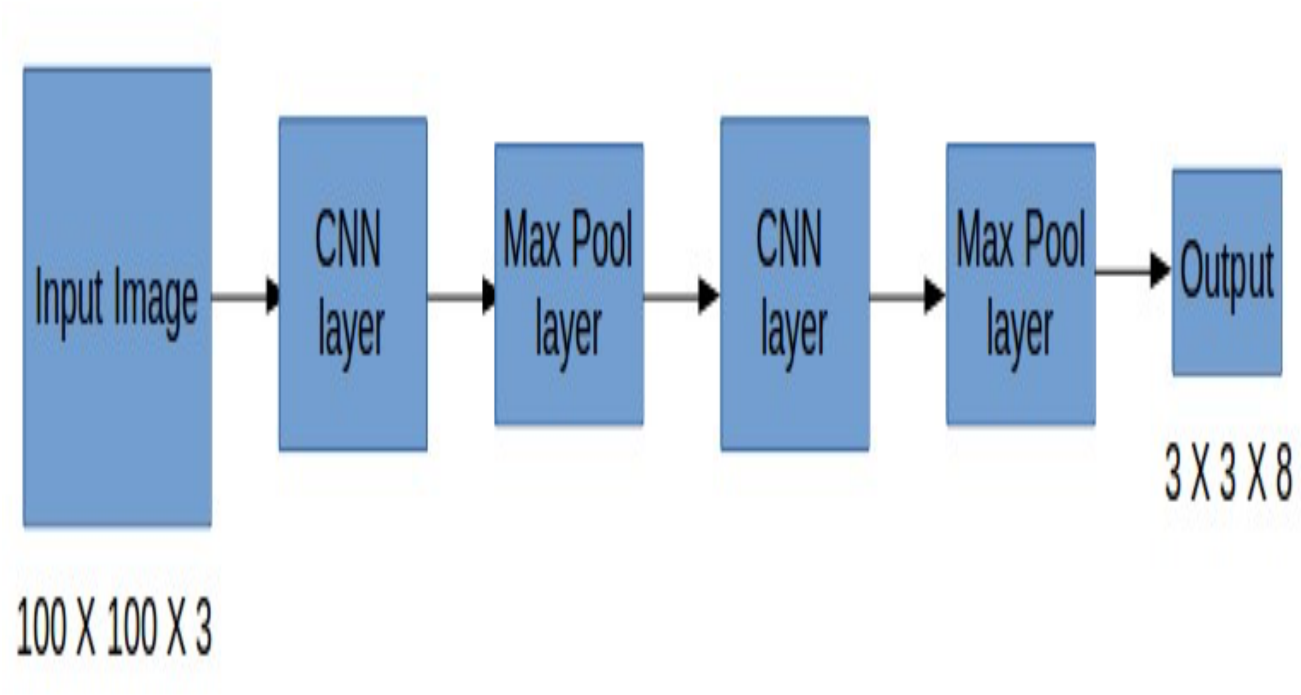
Fast. Good for real-time processing. Predictions (object locations and classes) are made from one single network. YOLO detects one object per grid cell. It enforces spatial diversity in making prediction.

The object class identified is converted to voice by google voice converter. These modules are further described below.

- **Object detection**

  In object detection, module, the input of video stream is considered. The input video stream is identified for objects using R-CNN techniques and it primarily use regions to localize the objects within the image. The network does not look at the entire image, only at the parts of the images which have a higher chance of containing an object.

  YOLO first takes an input image from video stream. The framework then divides the input image into grids. Image classification and localization are applied on each grid. YOLO then predicts the bounding boxes and their corresponding class probabilities for object.



- **Find object location**

Object localization from the input video stream can be identified. Image classification from object detection, which goes through a ConvNet that results in a vector of features fed to a softmax to classify the object. Neural network have a few more output units that encompass a bounding box. In particular, we add four more numbers, which identify the x and y coordinates of the upper left corner and the height and width of the box (bx, by, bh, bw)

Intersection over Union is a popular metric to measure localization accuracy and calculate localization errors in object detection models.

To calculate the IoU with the predictions and the ground truth, we first take the intersecting area between the bounding boxes for a particular prediction and the ground truth bounding boxes of the same area. Following this, we calculate the total area covered by the two bounding boxes—also known as the *Union*.

The intersection divided by the Union, gives us the ratio of the overlap to the total area, providing a good estimate of how close the bounding box is to the original prediction.

- **Find object movement**

The Python's built in deque datatype to efficiently store the past N points the object has been detected and tracked at. Libraries imutils also used by collection of OpenCV and Python convenience functions.

By checking the X,Y co-ordinates values and tracking them, the object movement can be detected.

- **Text to speech conversion**

We identify the Object_Class from above modules and this object class name is given to voice module. Here we use, google text to speech is used and we create a mp4 file. This file output the name of the object moving around the blind person.

4.2 ARCHITECTURE

This chapter gives overview of architecture design, dataset for implementation, algorithm used and UML designs.

SYSTEM ARCHITECTURE

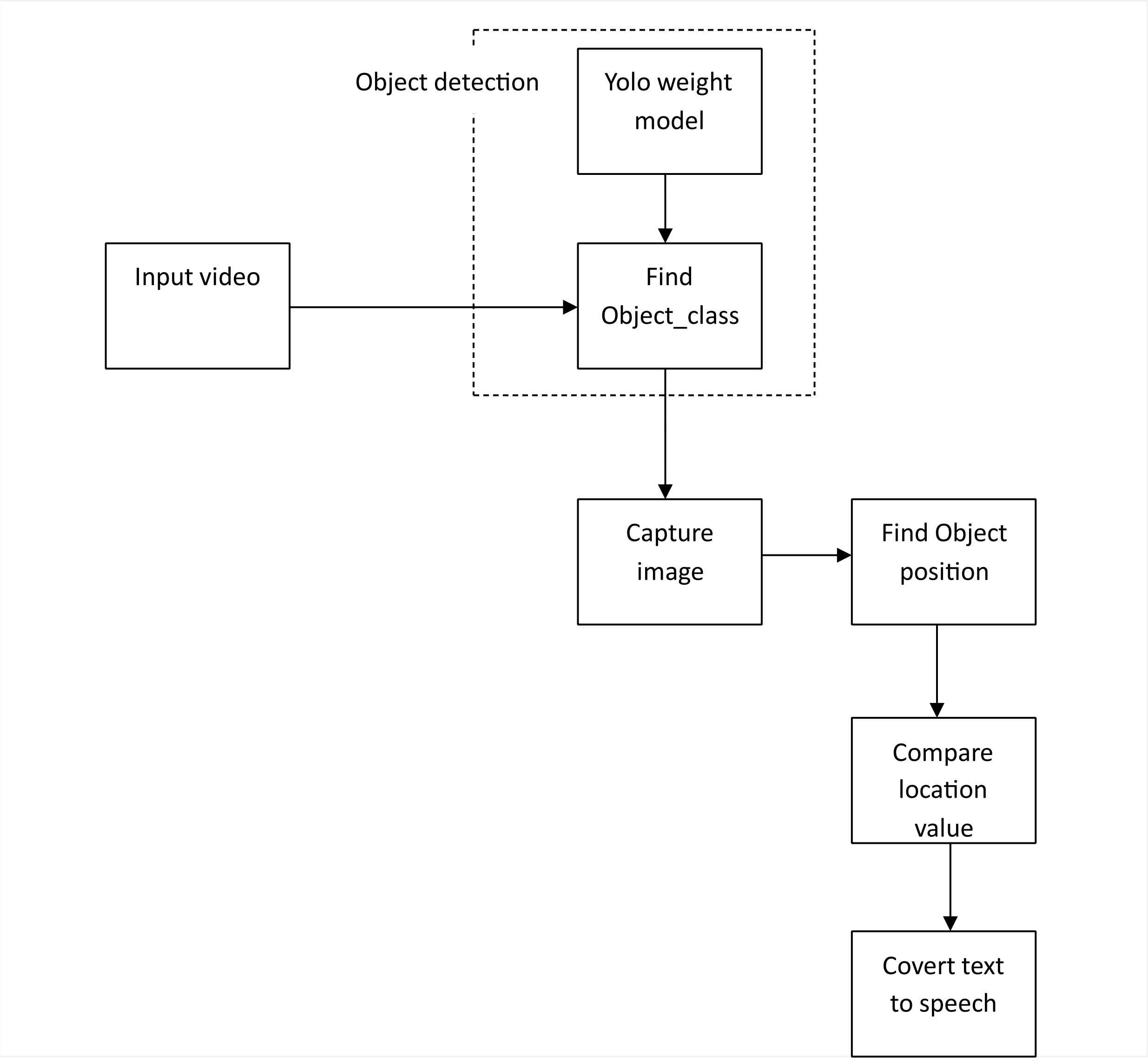


Figure: System Architecture

The above figure represents system architecture of proposed system, where we are showing the process of object detection and tracking the object.

OBJECT DETECTION- AN OVERVIEW

Object detection is a phenomenon in computer vision that involves the detection of various objects in digital images or videos. Some of the objects detected include people, cars, chairs, stones, buildings, and animals.

The major concept of YOLO is to build a CNN network to predict a (7, 7, 30) tensor. It uses a CNN network to reduce the spatial dimension to 7×7 with 1024 output channels at each location. YOLO performs a linear regression using two fully connected layers to make 7×7×2 boundary box predictions. To make a final prediction, we keep those with high box confidence scores (greater than 0.25) as our final predictions.

Compared to other region proposal classification networks (fast RCNN) which perform detection on various region proposals and thus end up performing prediction multiple times for various regions in a image.

Yolo architecture is more like FCNN (fully convolutional neural network) and passes the image (nxn) once through the FCNN and output is (mxm) prediction.

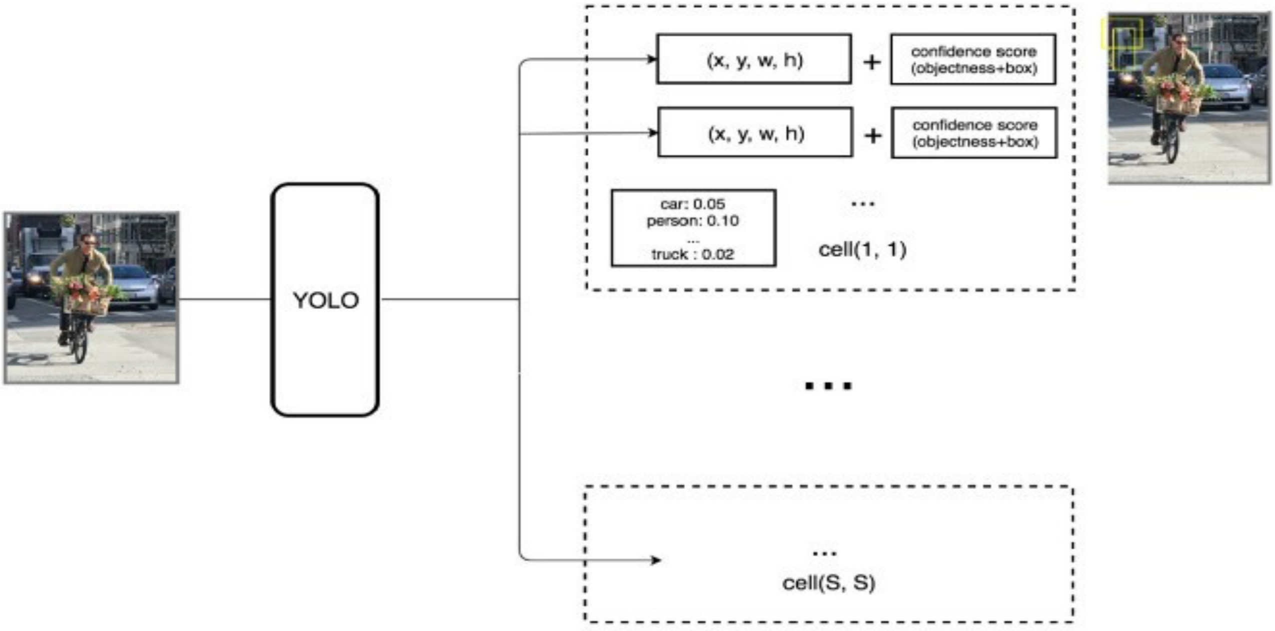


Figure: Overview- YOLO detection

This the architecture is splitting the input image in mxm grid and for each grid generation 2 bounding boxes and class probabilities for those bounding boxes.

Our network uses features from the entire image to predict each bounding box. It also predicts all bounding boxes across all classes for an image simultaneously. This means our network reasons globally about the full image and all the objects in the image. The YOLO design enables end-to-end training and real time speeds while maintaining high average precision.

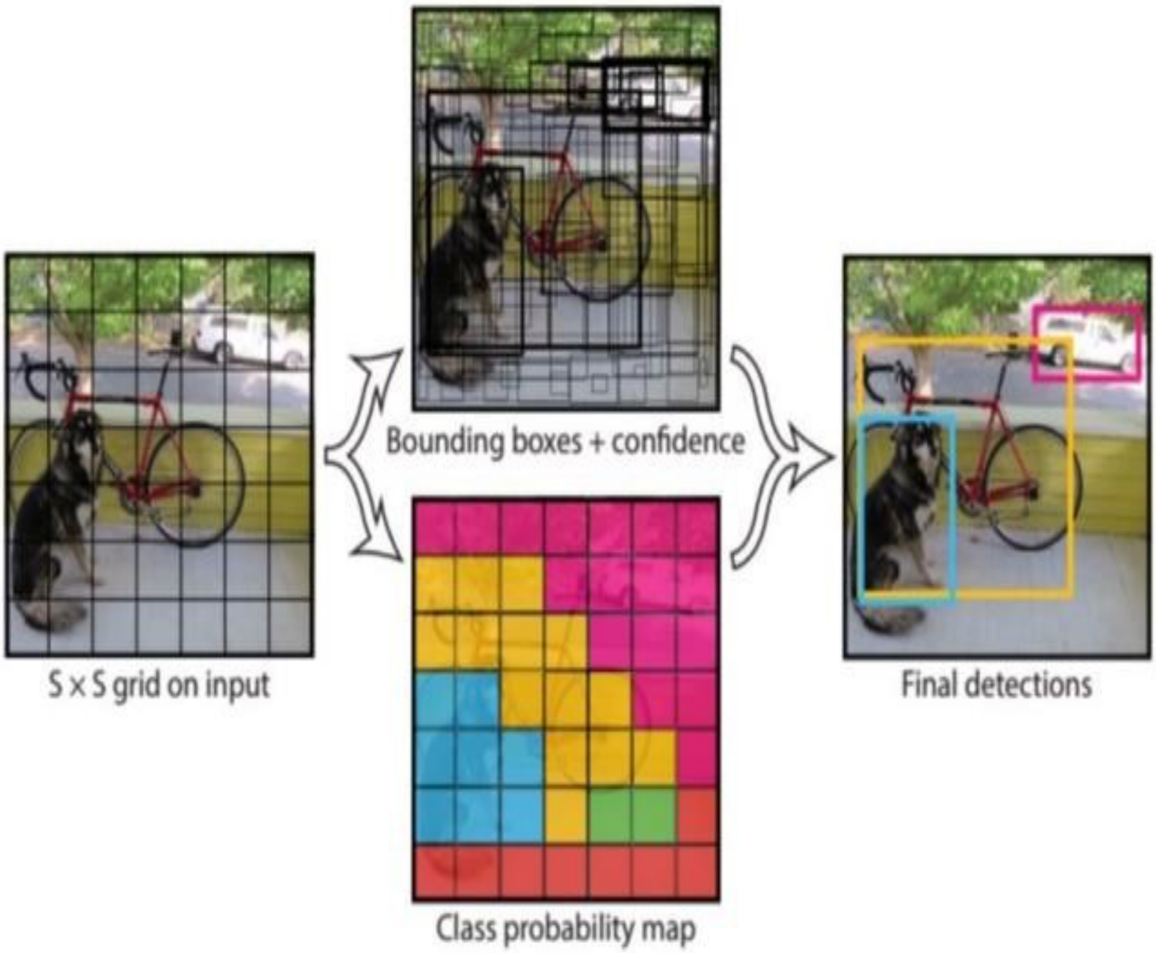


Figure: Objects with bounding boxes

YOLO algorithm employs convolutional neural networks (CNN) to detect objects in real-time. As the name suggests, the algorithm requires only a single forward propagation through a neural network to detect objects. This means that prediction in the entire image is done in a single algorithm run. The CNN is used to predict various class probabilities and bounding boxes simultaneously.

First, the image is divided into grid cells. Each grid cell forecasts B bounding boxes and provides their confidence scores. The cells predict the class probabilities to establish the class of each object. For example, we can notice at least three classes of objects: a car, a dog, and a bicycle. All the predictions are made simultaneously using a single convolutional neural network.

Intersection over union ensures that the predicted bounding boxes are equal to the real boxes of the objects. This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of the objects (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly. For example, the car is surrounded by the pink bounding box while the bicycle is surrounded by the yellow bounding box. The dog has been highlighted using the blue bounding box.

Advantages

YOLO is Fast. Good for real-time processing. Predictions (object locations and classes) are made from one single network. Can be trained end-to-end to improve accuracy.

YOLO is more generalized. It outperforms other methods when generalizing from natural images to other domains like artwork.

Region proposal methods limit the classifier to the specific region. YOLO accesses to the whole image in predicting boundaries. With the additional context, YOLO demonstrates fewer false positives in background areas. YOLO detects one object per grid cell. It enforces spatial diversity in making prediction.

4.3 UML DIAGRAMS:

The design is a plan or drawing produced to show the look and function or workings of an object before it is made. Unified Modeling language (UML) is a standardized modeling language enabling developers to specify, visualize, construct and document artifacts of a software system. Thus, UML makes these artifacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software development. It uses graphic notation to create visual models of software systems.

The different types of UML diagram are as follows.

- Use Case Diagram
- Class Diagram
- Activity Diagram
- Sequence Diagram

USE CASE DIAGRAM

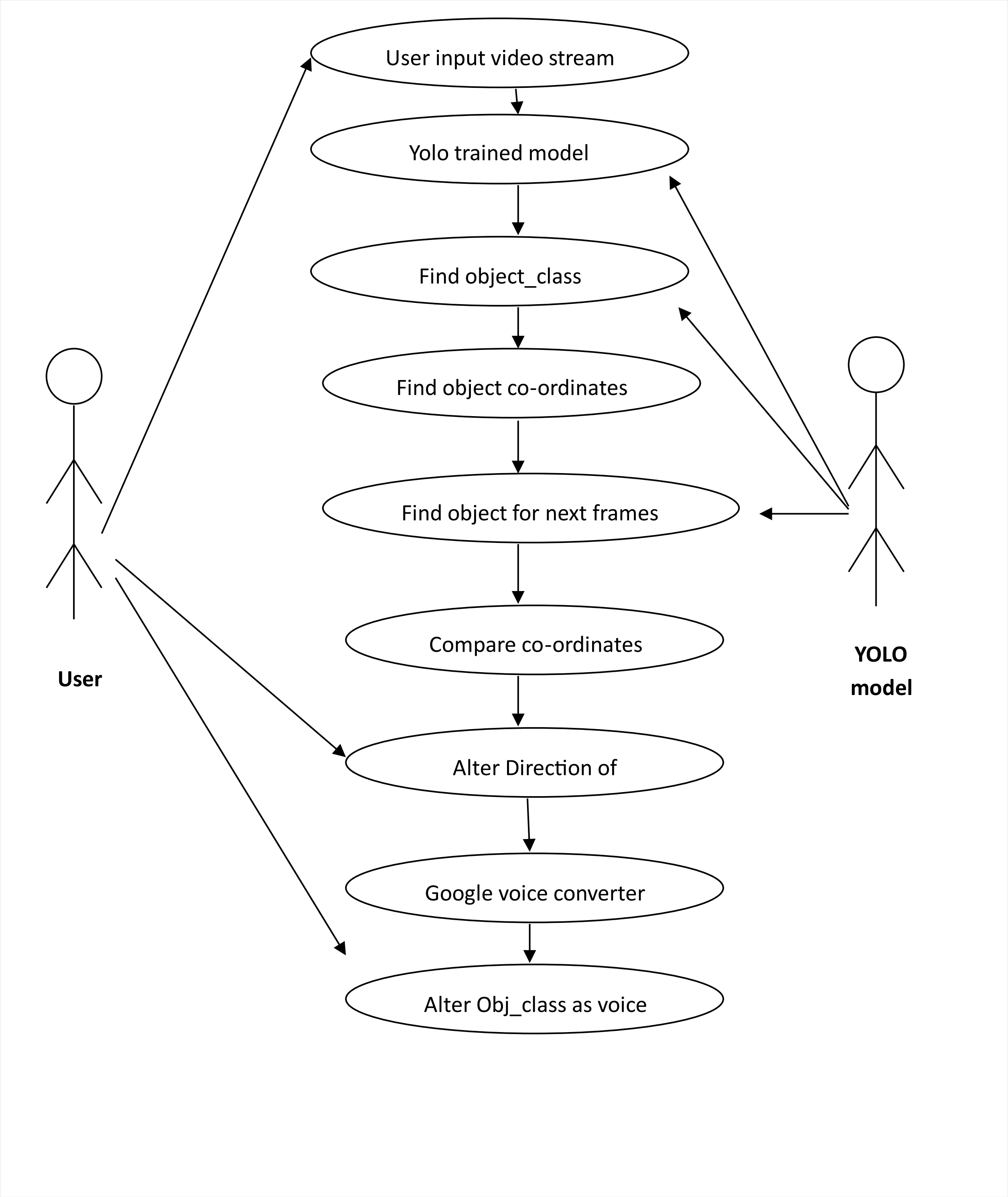


Figure: Use case Diagram

The above figure represents use case diagram of proposed system, where user inputs video frame, the algorithm work to generate the identified output. The actor and use case is represented. An eclipse shape represents the use case namely input image, preprocess, segmentation, recognition and output. Use case diagram is used to represent the dynamic behavior of a system. It encapsulates the system's functionality by incorporating use cases, actors, and their relationships. It models the tasks, services, and functions required by a system/subsystem of an application. It depicts the high-level functionality of a system and also tells how the user handles a system. The main purpose of a use case diagram is to portray the dynamic aspect of a system. It accumulates the system's requirement, which includes both internal as well as external influences. It invokes persons, use cases, and several things that invoke the actors and elements accountable for the implementation of use case diagrams. It represents how an entity from the external environment can interact with a part of the system.

SEQUENCE DIAGRAM

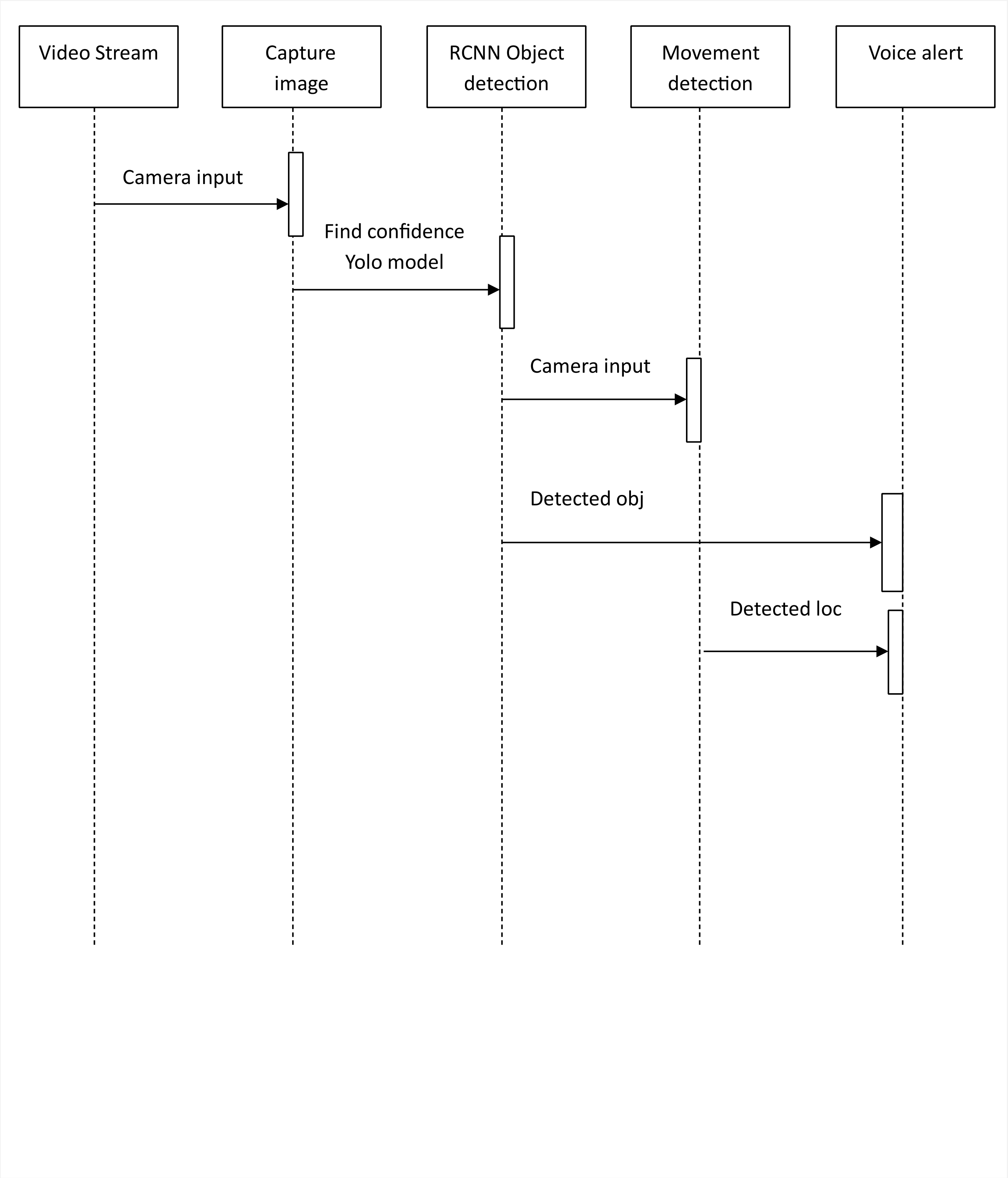


Figure: Sequence Diagram

A sequence diagram shows a parallel vertical lines, different processes or objects that live simultaneously, and as horizontal arrows, the messages exchanged between them, in order in which they occur. The above figure represents sequence diagram, the proposed system's sequence of data flow is represented.

Sequence diagram is a communication outline that shows how objects work with each other and in what order. It is a develop of a message sequence chart. A sequence diagram indicates question cooperation masterminded in time arrangement. It portrays the items and classes engaged with the situation and the arrangement of messages traded between the articles expected to complete the usefulness of the situation. Sequence diagrams are normally connected with use case realizations in the Logical View of the framework a work in progress. Sequence diagrams are ordinarily connected with utilize case acknowledge in Logical View of the framework being worked on. Sequence diagrams are sometimes called event diagrams or event scenarios.

ACTIVITY DIAGRAM

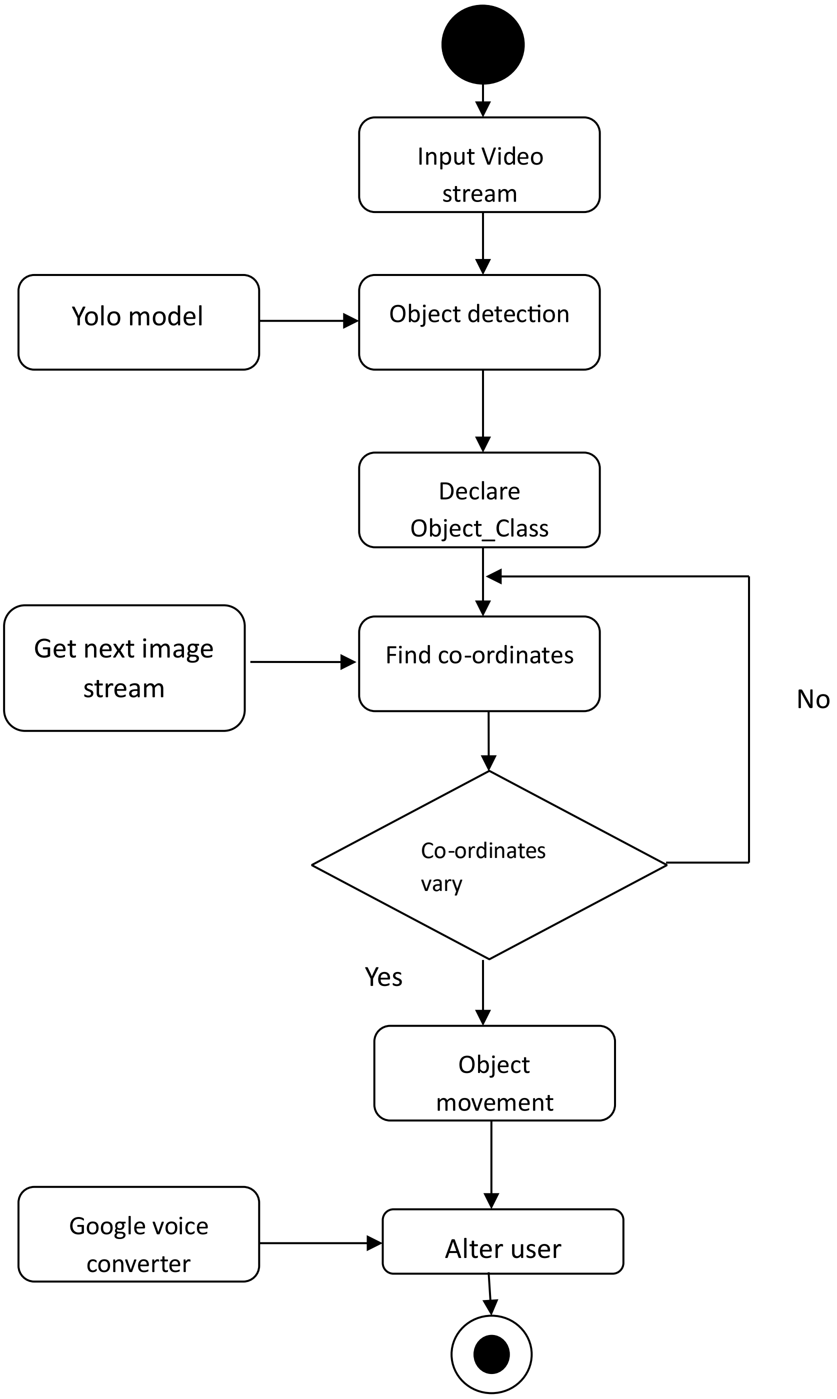


Figure: Activity Diagram

The above figure shows the activity diagram of the proposed system, where we represented the identified activities and its functional flow.

REFERENCES

[1] Joseph Redmon and Anelia Angelova, Real-Time Grasp Detection Using Convolutional Neural Networks (ICRA), 2015.

[2] P. Tang "Object Detection in Videos by High Quality Object Linking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 12721278, 1 May 2020, doi: 10.1109/TPAMI.2019.2910529.

[3] Saurabh Gupta, Ross Girshick, Pablo Arbelaez and Jitendra Malik, Learning Rich Features from RGBD Images for Object Detection and Segmentation (ECCV), 2014.

[4] X. Chen "High-Quality R-CNN Object Detection Using Multi-Path Detection Calibration Network," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 2, pp. 715-727, Feb. 2021, doi: 10.1109/TCSVT.2020.2987465.

[5] Andrej Karpathy Deep VisualSemantic Alignments for Generating Image Descriptions (CVPR), 2015.

[6] C. Baoyuan, L. Yitong and S. Kun, "Research on Object Detection Method Based on FF-YOLO for Complex Scenes," in IEEE Access, vol. 9, pp. 127950-127960, 2021, doi: 10.1109/ACCESS.2021.3108398.

[7] Liam Betsworth, Nitendra Rajput, Saurabh Srivastava, and Matt Jones. Audvert: Using spatial audio to gain a sense of place. In Human-Computer Interaction–INTERACT 2013, pages 455–462. Springer, 2013.

[8] Kevin Ramdath, Manor, Dharmdeo Sigh. A low cost outdoor assistive navigation system for blind people. In Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on, pages 828–833. IEEE, 2013.