GROUP 4 – PROJECT PROPOSAL

1. GOAL STATEMENT: Face Recognition and Object detection System to assist the Visually Impaired.
   1. INPUT:

Some Sample Image Captures from the input video.

**Label:** Tree **Label:** Pole

Our project would have the following constraints:

- Resolution of the camera should be at least 720p.

- The camera should be facing upright, perpendicular to the floor and parallel to the user’s body.

- The camera must be either held by the user or held by a sling.

- Weather conditions like rain and snow may reduce the accuracy for object detection considerably, unless trained or resolved.

- Lighting conditions would also reduce the accuracy of the application, unless trained or resolved.

- Facial Recognition would work for faces that are a meter away from the camera.

- Recognition requires faces without the mask.

* 1. OUTPUT: The result of the detection is delivered to the user using the mobile phone in the form of an audio speech.

1. Development Platform and Target AVD and Device we are testing it on: Will target API 26, Android 3.1.24, Using OpenCV version 4.5.3 in Android Studio and will test on Xiaomi Poco F1.
2. ALGORITHMS:

3.1) OVERVIEW:

We will be using the following components in our Project: Datasets, YOLO for Object Detection and FACENET for Face Recognition and Voice API for converting Text to Speech.

3.2) DATASET:

We will be using two Datasets - COCO Dataset for Object Detection and our dataset which contains multiple images of our own.

3.3) YOLO for Object Detection:

Over the years, there have been many object detection algorithms like R-CNN, Fast R-CNN,

however, YOLO has proven to be the most efficient one. YOLO (You Only Look Once) is an object

detection algorithm, which, as the name suggests, only parses the image once.

It involves dividing the image into an S x S grid. Each grid-cell is responsible for detecting B

bounding boxes. The grid-cells which contain the center of the objects are the ones where this

process starts from.

The algorithm involves the following steps:

- Drawing an S x S grid on the image

- Each cell would define B bounding boxes, so a total of S x S x B boxes.

- The box which has the center of the object, would define dimensional parameters, to define

the bounding boxes.

- If the bounding boxes exceed the current grid-cell, the same process would repeat for the grid

cells into which the bounding boxes extend.

- After all the bounding boxes are mapped, they go through 2 processes, one for predicting

whether the box has something significant in it and other to predict its class.

- Both processes involve giving a confidence score to the bounding boxes.

- We later combine and filter out the boxes which have a confidence score less than 30%.

- The resulting image contains the bounding boxes denoting the objects of significance.

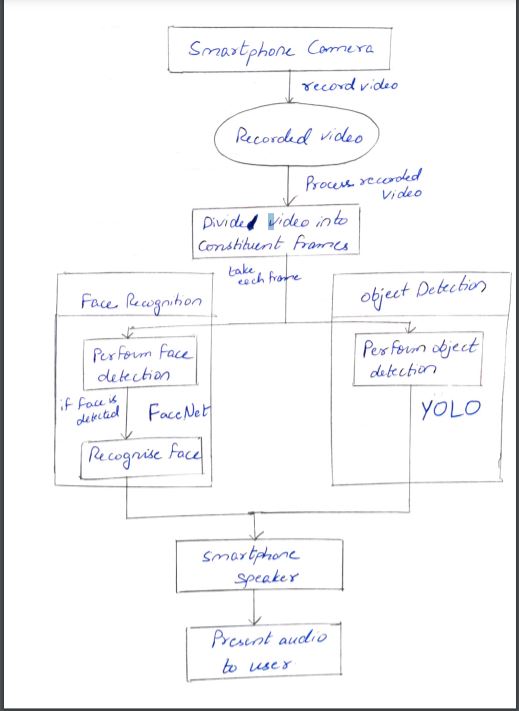
3.4) FaceNet for Face Recognition:

Face recognition is divided into multiple steps:  
-Face detection - Detecting one or more faces in an image.  
-Feature extraction - Extracting the important features from an image of the face.  
-Face classification - Classifying the face based on extracted features.  
-FaceNet takes an image of the person’s face as input and outputs a vector that represents the most important features of a face. In machine learning, this vector is called embedding.  
Embeddings of similar faces are also similar. One possible way of recognizing a person on an unseen image would be to calculate its embedding, calculate distances to images of known people and if the face embedding is close enough to embeddings of a particular person, we say that this image contains the face of that person.  
-First, we take multiple pictures of different people. We create a dedicated directory for each person with their images in it.   
-Then we train the FaceNet. At the beginning of training, it generates random vectors for every image which means the images are scattered randomly when plotted. It randomly selects an anchor image, then selects another image which might be the same person or different person, Adjusts the FaceNet network parameters so that the image of the same person is closer to the anchor than the image with a different person.   
-These steps are repeated until there are no more changes. With these images of the same person will come together and form like a group.  
-Then it recognizes a person on an unseen image by calculating its embedding, calculating the distances to images of known people and if the face embedding is close enough to embeddings of a particular person, we say that this image contains the face of that person. Classification is done by calculating the embedding distances between a new face and known faces.

3.5) Voice API for converting Text to Speech:   
 The Smartphone Speaker needs to deliver the output to the User. We will be converting the   
 text which is our result into speech. The voice API for Android will be added to ensure that  
 the results are converted into audio format.

3.6) REPORTING RESULTS: We will present the user with an audio output. The first part is object  
 detection after detecting the object it presents the user with a speech, If the object is a  
 familiar person, then it recognizes the person and produces audio with his name.

1. GUI INTERFACE:



Flow Diagram

: