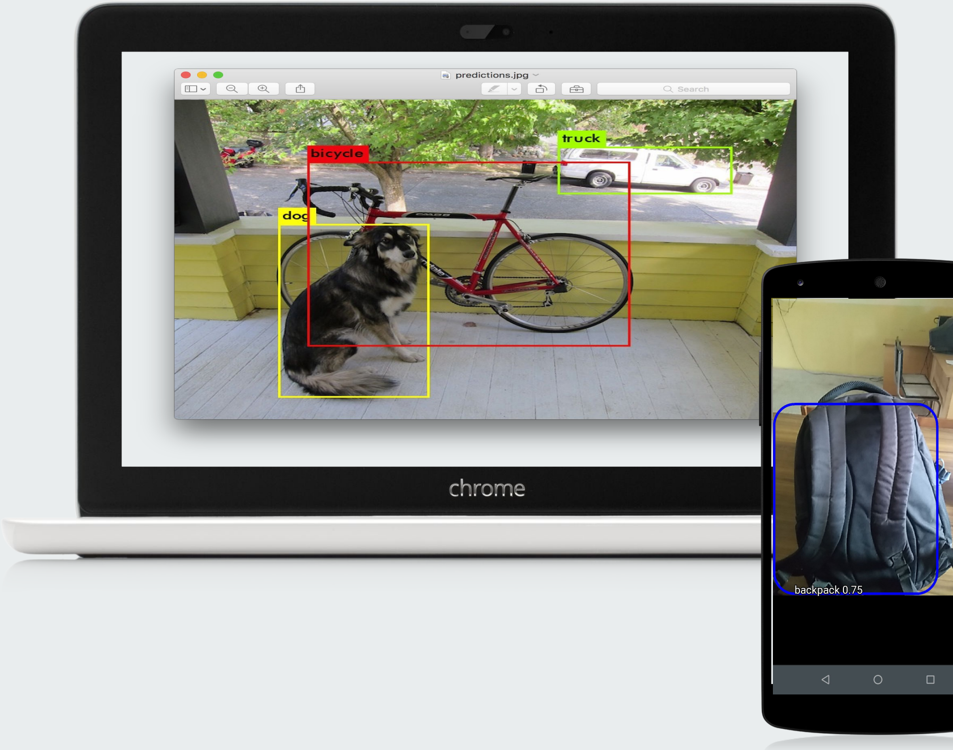
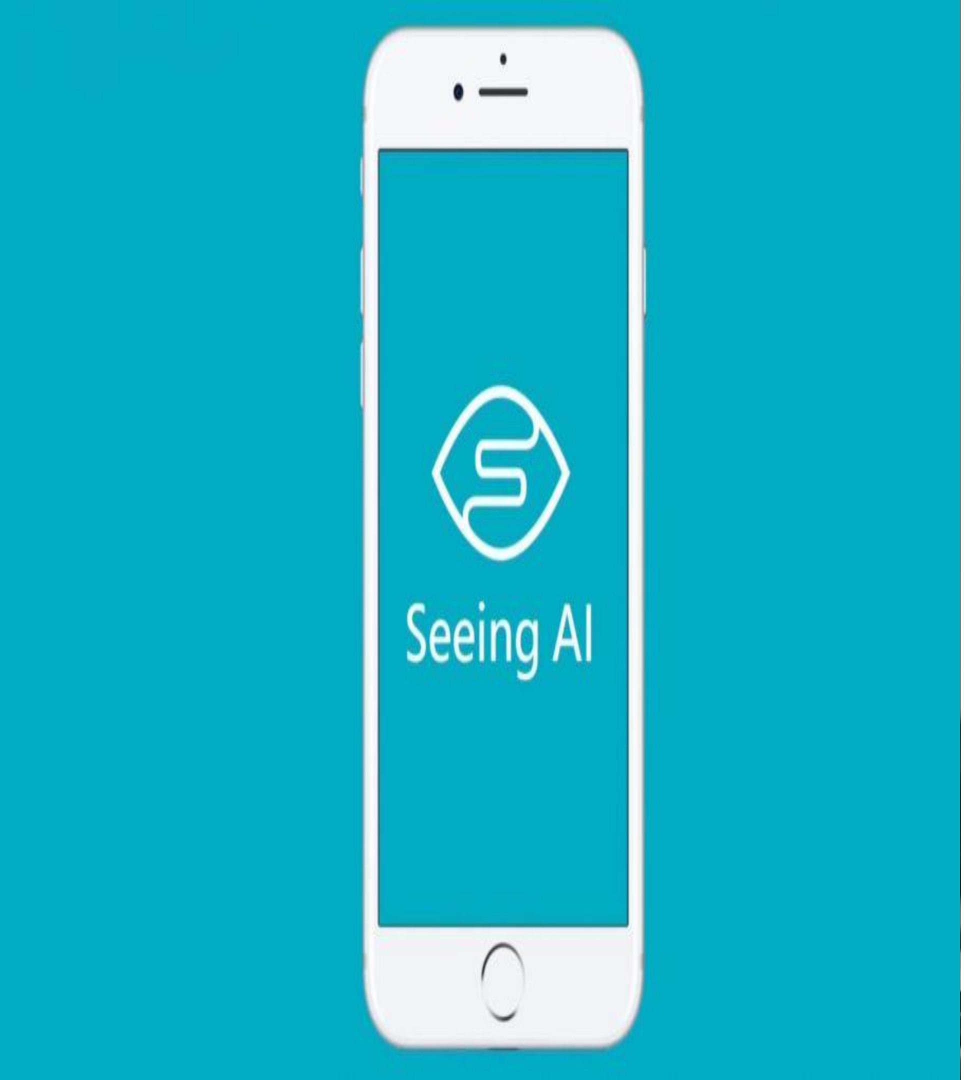
Automated Blind Support System

Made By:-
Shubham Mathur (BE/25005/15)
Apoorv Gaurav Agarwal (BE/25034/15)
Prageet N. Gupta (BE/25068/15)

Mentor: Dr. Shripal Vijayvargiya





Our Aim:

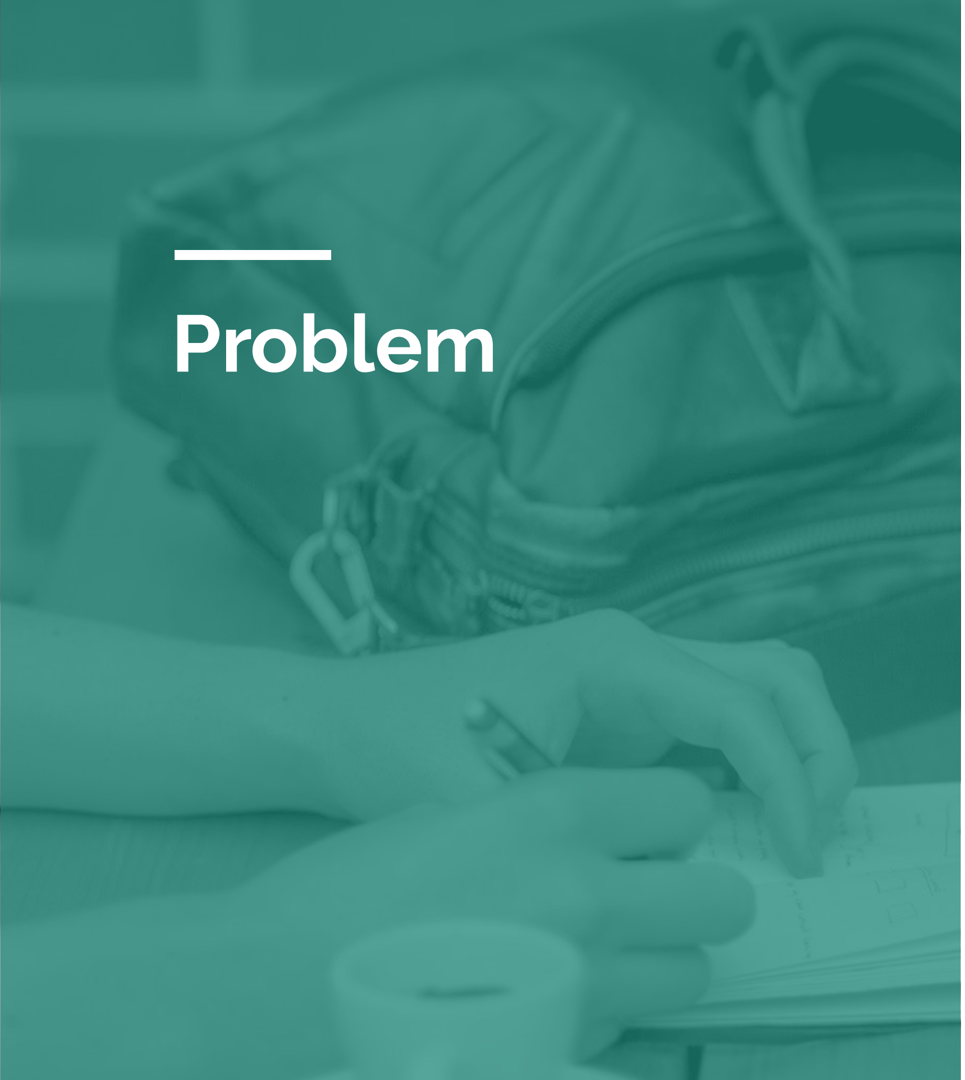
To build an app for the visually impaired inspired from Seeing AI.

Seeing AI is a Microsoft research project that uses the power of the cloud and AI to deliver an intelligent app designed to help the visually challenged people to navigate their day. It operates by pointing the phone’s camera at the desired object , selecting a channel, and hearing a description of what the AI has recognized around that person.

Major Modules

- Object Detection & Recognition
- Distance Estimation
- Speech Recognition

These modules have primarily been implemented in the project using the Android platform application which provides portability to the whole system.A previous version of the same was developed on the laptops.



Problem

The plan was to use Python Scripting tools for Android like BeeWare, Kivy or Pyqtdeploy for shifting the system to Android.

But

Processing each captured frame
+
Running computations for 60,000 models
+
4GB RAM (of which 2 GB is available)
+
Using a tool running python scripts
+
Building a real-time application
=
BAD IDEA

We Observed:

- ➔ Lag as long as 10 seconds which is generally not acceptable for a real time application
- ➔ It was observed that the scripting tool reserved some RAM by the android system which was used for running the scripts
- ➔ Scripting tool was using more RAM than the scripts themselves
- ➔ We decided to implement the modules on android from scratch , this way the pertinent issue of portability could also be resolved upto a certain extent.

Object Detection & Recognition

Using TensorFlow, training on COCO Model having 3.3 lakhs images of general objects categorised into 90 classes.

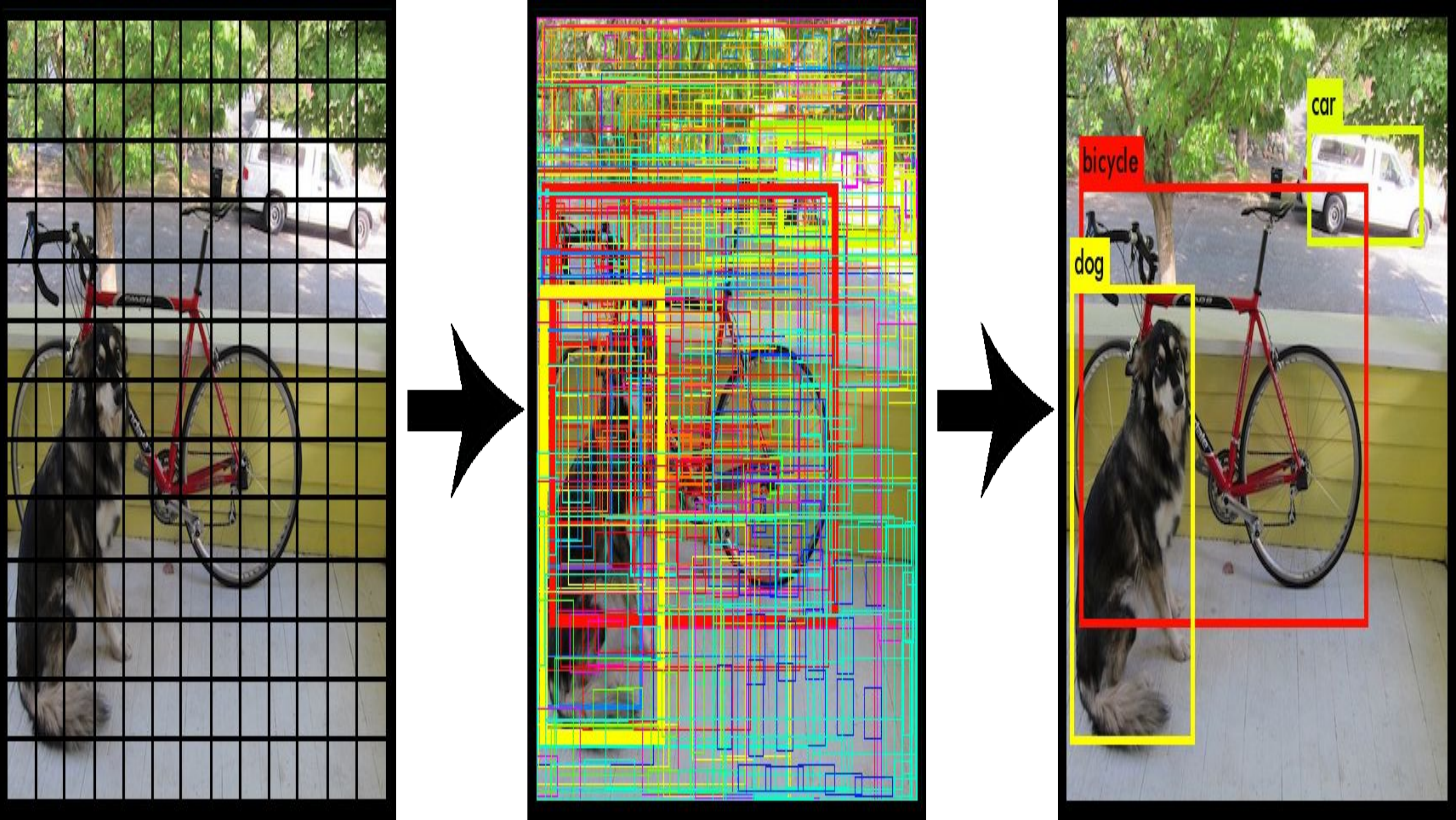
OBJECT DETECTION

Object detection is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.It is widely used in computer vision task such as face detection, face recognition, video object co-segmentation. It is also used in tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, tracking a person in a video.Every object class has its own special features that helps in classifying the class – for example all circles are round. Object class detection uses these special features. For example, when looking for circles, objects that are at a particular distance from a point (i.e. the center) are sought. Similarly, when looking for squares, objects that are perpendicular at corners and have equal side lengths are needed.

TENSORFLOW

Tensorflow is Google’s Open Source Machine Learning Framework for dataflow programming across a range of tasks. It is based on the concept of graphical representation of the problem where nodes in the graph represent mathematical operations, while the graph edges represent the multi-dimensional data arrays (tensors) communicated between them.

Tensors are just multidimensional arrays, an extension of 2-dimensional tables to data with a higher dimension. There are many features of Tensorflow which makes it appropriate for Deep Learning which is a technique for implementing Machine Learning. It uses neural networks to learn, sometimes, using decision trees may also be referred to as deep learning, but for the most part deep learning involves the use of neural networks.

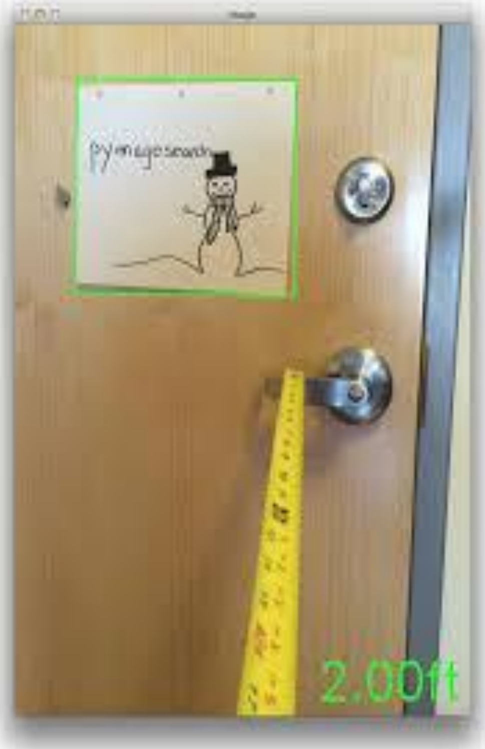

car
bicycle
dog

COCO (Common Objects in Context) MODEL

Common Objects in Context or COCO model is a state-of-the-art technique in object recognition which places the question of object recognition in the context of the broader question of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. Objects are labeled using per-instance segmentations to aid in precise object localization. The dataset contains photos of 90 objects types that would be easily recognizable by people of all ages . With a total of 2.5 million labeled instances in 328k images, the creation of this dataset drew its dataset information from various sources for category detection, instance spotting and instance segmentation. Finally, it provides a better baseline performance analysis for bounding box and segmentation detection results using a Deformable Parts Model.

Distance estimation

Distance to any object from a camera is obtained by measuring the size of the object from a predefined distance and then according to the current visible size calculating the relative distance.

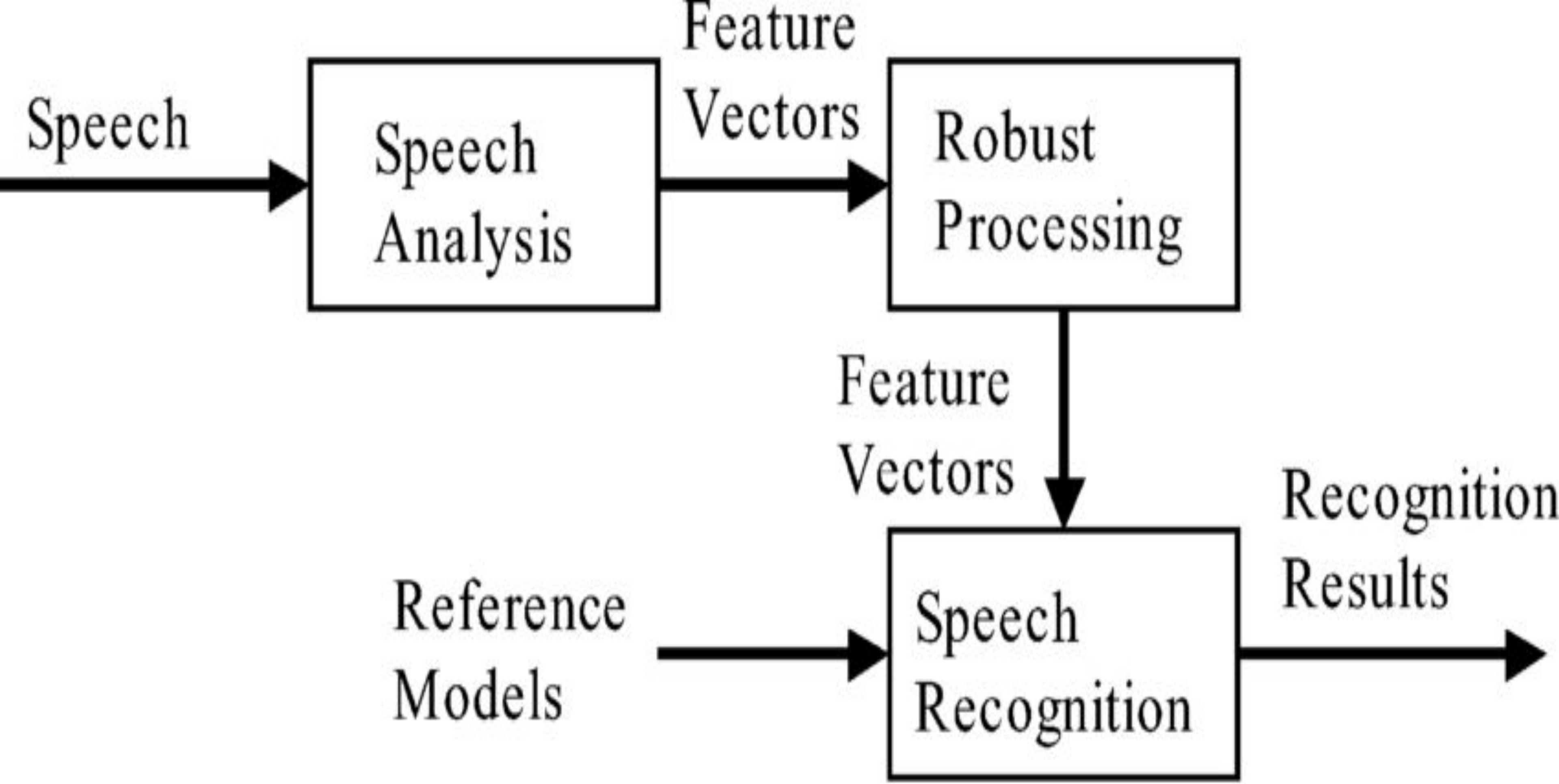


Distance Estimation

The calculation of distance from a single source camera is very much different from using a sensor. The approach is very basic. We have used the predefined size of the object on the compared the new size of the object which is being extracted from the stream. Now according to the new size , if the new size has more width and height from usual size of the object in that case the object must be nearer, else the object if far or at good distance from the user. The Model we have used contains more than 90 objects and the model is trained for all these objects so as to optimise the distance.

Speech Recognition

Speech is recognized using pre-defined speech patterns and then most probable word is given as output. These words are compared using their speech frequency.



Speech

Speech Analysis

Feature Vectors

Robust Processing

Feature Vectors

Reference Models

Speech Recognition

Recognition Results

Hidden Markov Model

Thanking You