Hardware design is carried out in four phases,

1. Mathematical model.

2. Simulation and verification of algorithm.

3. Calibration of sensor.

4. Hardware Design.

0.1 Mathematical model

The mathematical model serves the most important role throughout the project, as it is intended to solve the issues that persisted in previous research.

This model is further divided into two sub-parts,

0.1.1 Listener tracking model

The listener tracking model is a combination of face detection and stereo vision technique for estimating depth.

Face detection

We must classify and sort out the entities from the rest of the objects from the surroundings to align the speakers properly.

In our case, these entities are people who are listening to the system. To classify them from other objects from surroundings, we implement the Harr cascade face detection algorithm to sort and cluster out these entities.

Harr cascades is a cascade classifier that implements a machine learning approach based on the Adaboost meta-algorithm.

The rectangular shape of the face is meaningful in initializing the classifier. Further, the algorithm focuses on the property that the eyes region is often darker than the face and nose region. The second feature proposes that the eyes are darker than the bridge of the nose. Similarly, this approach finds the entity's possible relations and features and records the features for further prediction.

Once the face is detected, we can obtain the face location from the origin (center of the image).

Stereo Vision

Stereo vision compares the information about a scene from two vantage points and examining relative positions of objects in the two panels.

An image can be termed as a grid of pixels within some range of indices. Using face detection, we narrowed down the object's position in the grid of pixels (x, y).

Stereo vision gives two images of the same scene from different positions in the same plane. Each image gives the disparity of the face from the origin of that respective image.

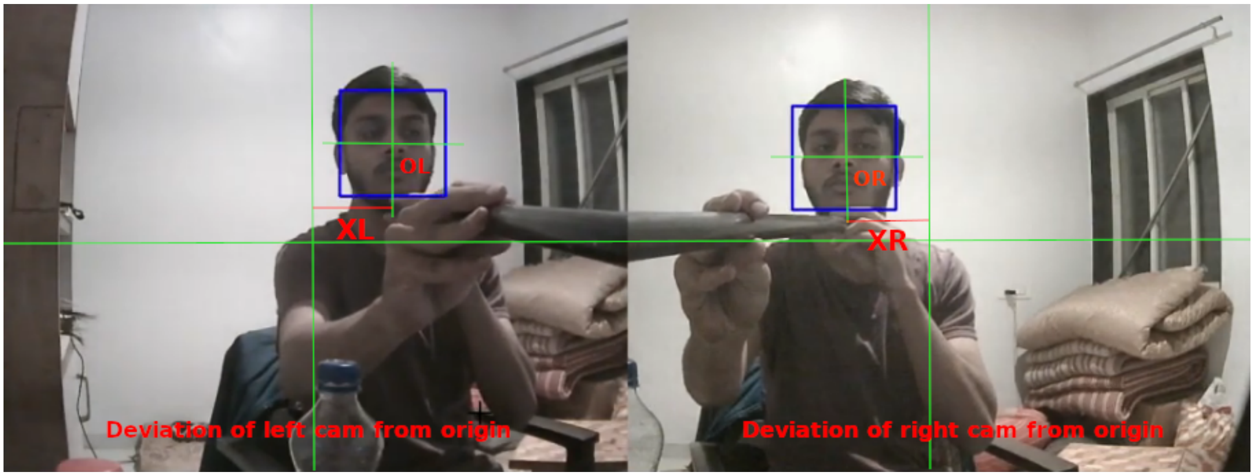
OL
XL
Deviation of left cam from origin
OR
XR
Deviation of right cam from origin

Figure 1: Stereo vision showing the disparity from origins of both the cameras

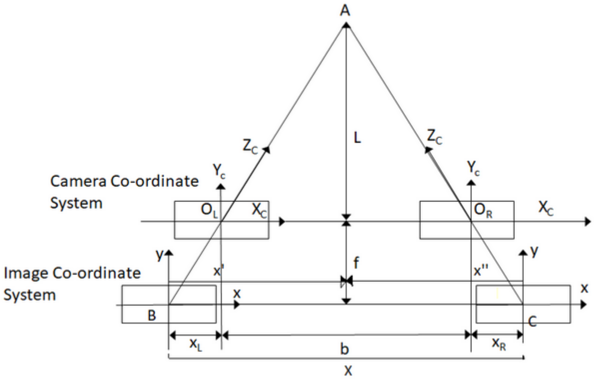

Figure 2: Depth sensing geometrical model

Figure 5.1 shows us the geometry behind the stereo vision method for depth sensing.

Here,

x = Distance between two webcams.

f = Focal length of the webcams.

X_L = Disparity of the image from the origin of the left webcam.

X_R = Disparity of the image from the origin of the right webcam.

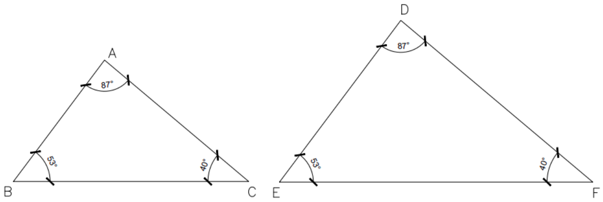

Figure 3: AA symmetry criterion

Stereo vision depth-sensing works on the principle of angle-angle symmetry (AA symmetry criterion) of two triangles. Using the AA symmetry criterion, we can state that if angle of the two triangles are congruent, then the third angle of both triangles must be the same. Hence, the ratio of each parallel side of triangles is equal.

i.e.,

$$\frac{AB}{DE} = \frac{BC}{EF} = \frac{AC}{DF} \tag{1}$$

Hence from figure 5.2, we can prove that,

$$\frac{f}{z} = \frac{X_L}{X'} \tag{2}$$

Where, $z = f + L$,

Similarly,

$$\frac{f}{z} = \frac{X_R}{X''} \tag{3}$$

From equation 5.2 and 5.3, we can say that,

$$x' = \frac{z \times X_L}{f} \tag{4}$$

$$x'' = \frac{z \times X_R}{f} \tag{5}$$

From figure 5.2 we can say that $x = x' + x''$,

$$\therefore x = \frac{z \times X_L}{f} + \frac{z \times X_R}{f}$$

$$\therefore x = \frac{z}{f} \times (X_L + X_R)$$

Finally, we get depth (z),

$$\boxed{z = \frac{x \times f}{X_L + X_R}} \tag{6}$$

Focal length