

# Depth and Zoom Estimation for PTZ Camera Using Stereo Vision

*Report submitted in*

*May 2014*

*to the department of*

**Computer Science and Engineering**

*of*

**National Institute of Technology Rourkela**

*in partial fulfillment of the requirements*

*for the degree of*

**Bachelor of Technology**

*by*

**A Kiran Kumar[110CS0125]**

*and*

**G Ram Pratheek[110CS0515]**

*under the supervision of*

**Pankaj Kumar Sa**



**Department of Computer Science and Engineering  
National Institute of Technology Rourkela  
Rourkela, Odisha, 769 008, India**



Department of Computer Science and Engineering  
**National Institute of Technology Rourkela**  
Rourkela-769 008, Odisha, India.

## Certificate

This is to certify that the work in the thesis entitled *Depth and Zoom Estimation for PTZ Camera Using Stereo Vision* by *A Kiran Kumar* and *G Ram Pratheek*, bearing roll numbers *110CS0125* and *110CS0515*, is a record of an original research work carried out by them under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Place: NIT Rourkela  
Date: May 2014

**Pankaj Kumar Sa**  
Assistant Professor, CSE Department  
NIT Rourkela, Odisha

## Acknowledgement

We take this opportunity to express our profound gratitude and deep regards to our guide Dr. Pankaj Kumar Sa for his exemplary guidance, monitoring and constant encouragement throughout the course of this project. He motivated and inspired us through the entire duration of work, without which this project could not have seen the light of the day.

We convey our regards to all the faculty members of Department of Computer Science and Engineering, NIT Rourkela for their valuable guidance and advices at appropriate times. We would also like to thank Mr. Rahul Raman, Mr. Anshuman, Mr. Prakash and the other research scholars of the Image Processing Laboratory for their invaluable support and all our friends for their help and assistance all through this project.

Last but not the least, we express our profound gratitude to the Almighty and our parents for their blessings and support without which this task could have never been accomplished.

## **Authors Declaration**

I hereby declare that all the work contained in this report is my own work unless otherwise acknowledged. Also, all of my work has not been previously submitted for any academic degree. All sources of quoted information have been acknowledged by means of appropriate references.

A Kiran Kumar, G Ram Pratheek  
NIT Rourkela

## Abstract

Depth perception comes naturally to humans. However in a Computer Vision scenario estimation of distance between object and camera is an area still under research. This thesis aims to use binocular stereo vision to reconstruct a 3D scene from 2D images of the scene taken by a pair of cameras and use it to estimate the distance of the object from the camera. Further, this estimated distance is used to calculate the Zoom of a PTZ camera. Although there are various ways to determine the distance of an object from the camera using Sensors, Lasers and other such external devices, the method used in this thesis is independent of the use of such external devices and uses only image processing techniques to determine the distance. Results obtained and the process are clearly outlined.

Keywords: Stereo Vision, Depth Estimation, Disparity.

# Contents

<b>Certificate</b>	<b>ii</b>
<b>Acknowledgement</b>	<b>iii</b>
<b>Authors Declaration</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Stereopsis and Stereo Vision . . . . .	1
1.2 PTZ Camera . . . . .	2
1.3 Objectives . . . . .	3
<b>2 Methodology</b>	<b>4</b>
2.1 Epipolar Geometry . . . . .	4
2.2 Camera Calibration . . . . .	8
2.3 Image Rectification . . . . .	10
2.4 Disparity Calculation . . . . .	11
<b>3 Results</b>	<b>17</b>
<b>4 Conclusion</b>	<b>32</b>

# List of Figures

1.1	Stereo Vision Setup . . . . .	2
2.1	The principle of triangulation in stereo imaging[13]. . . . .	5
2.2	The epipolar constraint[13]. . . . .	6
2.3	The epipolar line along which the corresponding point for X must lie[13]. . . . .	6
2.4	The relationship between the two coordinate systems[13]. . . . .	7
2.5	Checkerboard used for Camera Calibration . . . . .	9
2.6	Typical Camera Configuration . . . . .	11
2.7	Distance estimation by Triangulation. . . . .	15
3.1	Left and Right image pair#01 . . . . .	17
3.2	Left and Right image pair#02 . . . . .	18
3.3	Left and Right image pair#03 . . . . .	18
3.4	Left and Right image pair#04 . . . . .	18
3.5	Left and Right image pair#05 . . . . .	19
3.6	Left and Right image pair#06 . . . . .	19
3.7	Left and Right image pair#07 . . . . .	19
3.8	Left and Right image pair#08 . . . . .	20
3.9	Left and Right image pair#09 . . . . .	20
3.10	Left and Right image pair#10 . . . . .	20
3.11	Corner Detection in Left and Right image pair#01 . . . . .	21
3.12	Corner Detection in Left and Right image pair#02 . . . . .	21
3.13	Corner Detection in Left and Right image pair#03 . . . . .	22
3.14	Corner Detection in Left and Right image pair#04 . . . . .	22
3.15	Corner Detection in Left and Right image pair#05 . . . . .	22

3.16	Corner Detection in Left and Right image pair#06	23
3.17	Corner Detection in Left and Right image pair#07	23
3.18	Corner Detection in Left and Right image pair#08	23
3.19	Corner Detection in Left and Right image pair#09	24
3.20	Corner Detection in Left and Right image pair#10	24
3.21	Rectified Left and Right image pair#1	25
3.22	Rectified Left and Right image pair#2	25
3.23	Rectified Left and Right image pair#3	25
3.24	Rectified Left and Right image pair#4	26
3.25	Rectified Left and Right image pair#5	26
3.26	Rectified Left and Right image pair#6	26
3.27	Rectified Left and Right image pair#7	27
3.28	Rectified Left and Right image pair#8	27
3.29	Rectified Left and Right image pair#9	27
3.30	Rectified Left and Right image pair#10	28
3.31	Disparity Map#01	29
3.32	Disparity Map#02	30
3.33	Disparity Map#03	31



# Chapter 1

## Introduction

### 1.1 Stereopsis and Stereo Vision

The process of judging the depth and distance of objects from our eyes by comparing the images on our retinas works so well that many of us who are not ophthalmologists don't realize the importance of this ability. If you want to know the importance of this process, try driving a car or riding a bicycle or try playing tennis for even a few minutes with one eye closed. Everyone goes to 3-D movies these days. Ever wonder how the special glasses you wear work?

All the above scenarios and many more depend on the concept of stereopsis.

If we consider the images that are cast on our retinae, they are two-dimensional in nature, but in reality we look out onto a three-dimensional world. So determining the object's three-dimensional shape and estimating the depth & distance is a problem worth solving.

**Stereo Vision:** As described above, Stereopsis is the biological system of our eyes. Stereovision can be seen as an artificial rendition of the Stereopsis system. So what exactly is Stereovision? Stereo vision is the extraction of three-dimensional information from a pair of digital images, obtained by two cameras. Information about a scene from two different vantage points can be compared and 3D information can be extracted by examining the relative positions of objects in the two images. This has been modelled based on the process of stereopsis.

In the case of a biological scenario where the binocular vision of eyes is used for recon-

structuring the 3D scene, the image obtained on the retina is processed by the intraparietal sulcus of the brain. This ability of the brain to construct a clear mental image of a scene on the basis of two different images received from the two eyes, overcomes the need for any pre-processing of the images.

A traditional stereo vision system has two cameras placed at a distance horizontally from each other which are used to obtain two different viewpoints of a particular scene, similar to the manner in which the human binocular vision works. Comparison of the two images yields the relative depth in the form of disparity maps. The disparities are found to be inversely proportional to the differences in distances of the object.

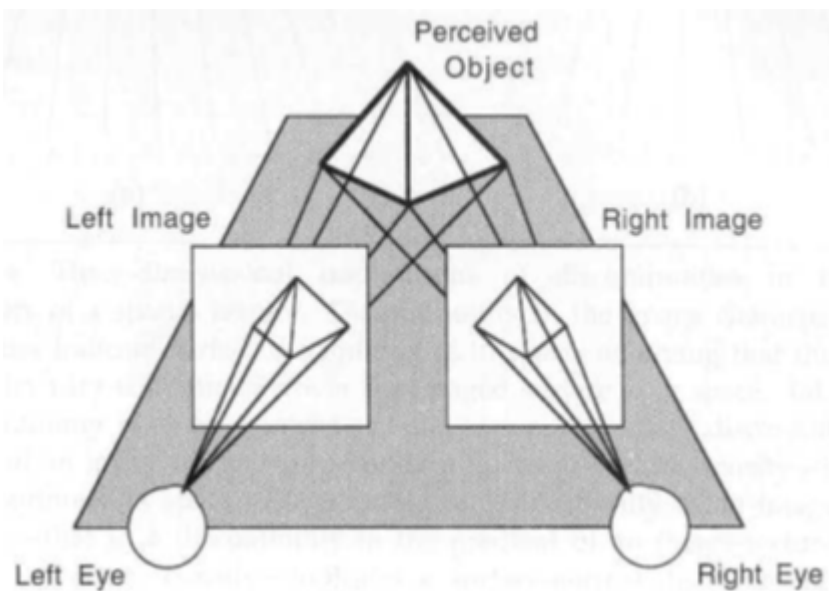


Figure 1.1: Stereo Vision Setup

A real camera system unlike the human binocular eye system requires various pre-processing steps to be performed on the images obtained from the two cameras. The various pre-processing steps are discussed in the forthcoming chapters.

## 1.2 PTZ Camera

A **PTZ** camera is an IP-camera capable of directional and zoom control. It is a high-resolution camera which zooms and pans into portions of the image based on the param-

eters Pan, Tilt and Zoom.

**Pan:**

Pan refers to the rotation of a camera in a horizontal plane. It results in a camera motion similar to that of someone shaking his head from one side to another or a yaw rotation performed by an aircraft.

**Tilt:**

Tilt is a technique in which the camera rotates in the vertical plane. It is similar to Pan but in a different planar context. Tilting of a camera results in a motion similar to nodding of the head indicating "yes" / no or can be said to be similar to the pitch rotation performed by an aircraft.

**Zoom:**

It is the optical zoom that is achieved by the use of zoom lens. A zoom lens is an assembly of lens for which the focal length and in turn the angle of view can be varied.

The final objective of this thesis is to calculate the Zoom parameter of the PTZ camera and use it in a real time scenario to zoom and focus the PTZ camera on the desired object.

### 1.3 Objectives

- 1) Obtaining the Extrinsic and Intrinsic parameters by calibrating the cameras with a checkerboard.
- 2) Rectification of the Calibrated images.
- 3) Constructing the Disparity Map from the Rectified Images.
- 4) Estimating the Distance between an object and the PTZ Camera.
- 5) Using the distance estimated to operate a PTZ camera by calculating the parameters.

# Chapter 2

## Methodology

### 2.1 Epipolar Geometry

This section gives a detailed account of Epipolar Geometry which is the mathematical basis for Stereo vision. Following are the definitions of some of the building blocks of Epipolar Geometry which will be used extensively throughout this section.

**Epipole:** "the point of intersection of the line joining the optical centres of the camera which is called the baseline with the image plane". Thus the epipole can be said to be "the image, in one camera, of the optical centre of the other camera".

**Epipolar Plane:** "the plane which is defined by a three-dimensional point  $M$  and the optical centres of the two cameras  $C$  and  $C'$ ".

**Epipolar Line:** "the line of intersection of the epipolar plane with the image plane". It can be termed as "the image in one camera of a ray through the optical centre and image point in the other camera". The Epipole is the point of intersection of all the Epipolar Lines.

Epipolar Geometry becomes useful in extracting the 3D structure of a scene or an object from a pair of images. There are two ways for this:

- 1) In the first method we calibrate both cameras with respect to a world coordinate system, and then calculate the geometry by calculating the essential matrix of the system, and from this the three-dimensional structure of the scene is constructed.
- 2) In the second method a fundamental matrix is calculated from image correspondences, and this is used to determine the three-dimensional structure of the scene.

The Essential and the Fundamental matrices have the following properties:

- Both the intrinsic and the extrinsic parameters of the camera are encapsulated in the Fundamental Matrix  $\mathbf{F}$ , whilst the essential matrix contains only the extrinsic parameters.
- The essential matrix  $\mathbf{E}$  is a  $3 \times 3$  matrix. In order to estimate this for corresponding image points the intrinsic parameters of both cameras must be known.
- Corresponding Epipolar Lines are mapped by the matrix  $\mathbf{F}$  to the image points.

The underlying principle in both the above approaches of binocular vision is triangulation. Any object point that is visible in an image must lie on the straight line passing through the centre of projection of the camera and the image of the object point (see figure 2.1). So Triangulation can be defined as "The determination of the intersection of two such lines generated from two independent images".

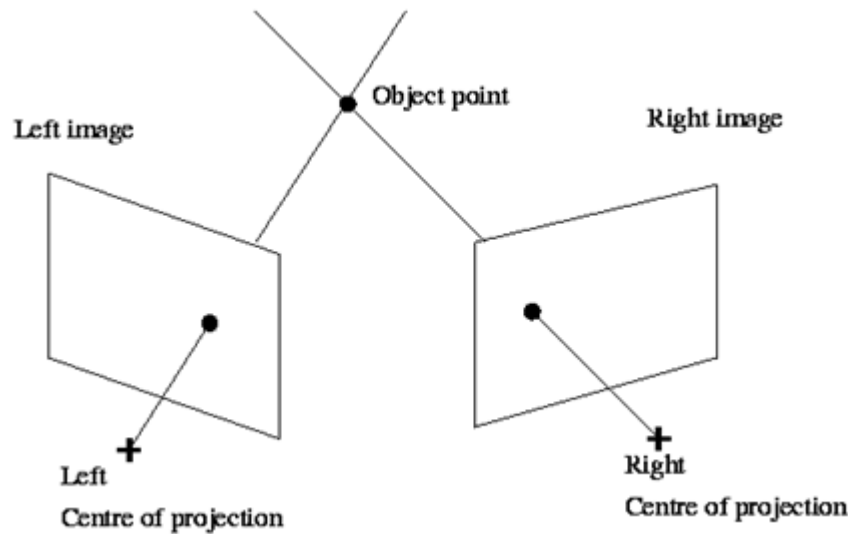


Figure 2.1: The principle of triangulation in stereo imaging[13].

It is a prerequisite to determine the three dimensional scene position of an object point using the triangulation method and matching the image location of the object point in one image to the location of the same object point in the other image. This process of finding such matches between the points of a pair of images of the same scene is called

*correspondence*, and will be dealt with in the following sections.

The epipolar geometry constraint reduces the correspondence problem to a search problem in a single line. Consider the following figure which tries to explain this point.

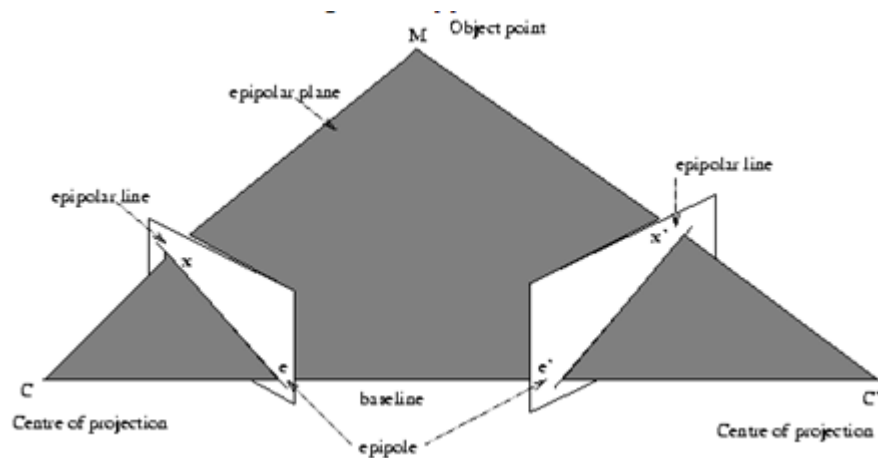


Figure 2.2: The epipolar constraint[13].

So every point  $x$  in one of the images generates a line in the other image on which lies the point  $x$ . Hence the search for correspondences is reduced from a region to a line.

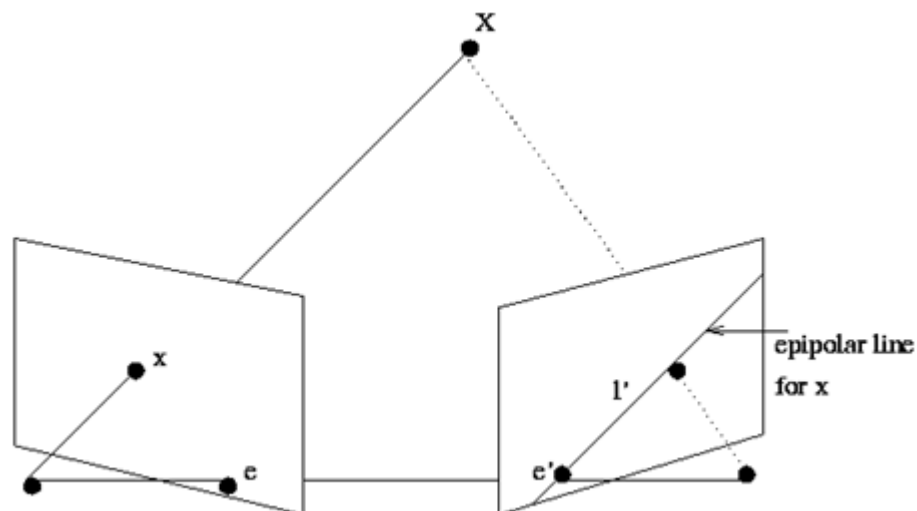


Figure 2.3: The epipolar line along which the corresponding point for  $X$  must lie[13].

In order to calculate the depth information from a pair of images we need to compute the epipolar geometry. In this thesis we have used a calibrated environment so we start by calculating the *essential* matrix and proceed by showing the process for calculating

the *fundamental* matrix.

The following equation gives the relation of how the coordinate system of the two cameras is related by rotation and translation:

$$x' = Rx + T[13]$$

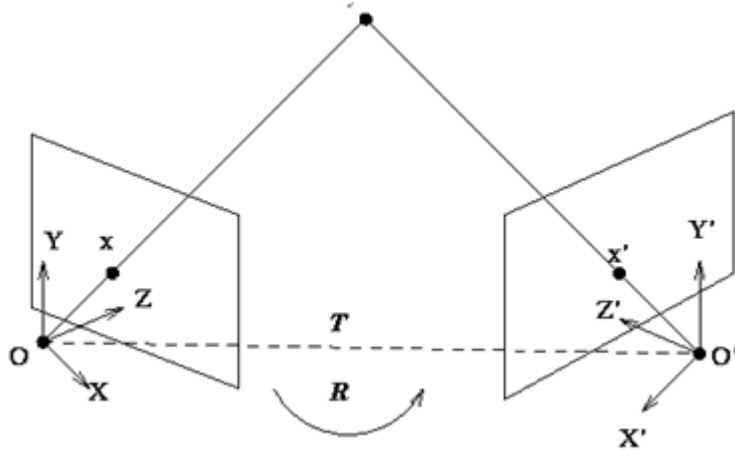


Figure 2.4: The relationship between the two coordinate systems[13].

Taking the vector product with  $\tau$ , followed by the scalar product with  $x$  we obtain

$$x'.(\tau \wedge Rx) = 0[13] \quad (2.1)$$

The above equations expresses the fact that the vectors  $Cx$ ,  $CX$  and  $CC$  are coplanar. This equation can be rewritten as:

$$x'^T E x = 0; [13] \quad (2.2)$$

Where,

$$E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} .R[13]$$

is the essential matrix, and Equation( 2.2) is the epipolar geometry for the calibrated system used in this thesis and this matrix relates corresponding image points expressed in the coordinate system of the cameras.

So Epipolar geometry for both the calibrated and un-calibrated systems in the form of Fundamental and Essential Matrices is used in this thesis for the pre-processing of images.

## 2.2 Camera Calibration

Camera Calibration gives the relation between known points in the world and points in an image. It is one of the requisites of Computer Vision. A calibrated camera can essentially be seen as a sensor which can be used to measure the distances between camera and objects. It is employed in many applications in order to recover 3D constraints of an observed scene from the 2D images. Various properties of the object such as the distance of the object from the camera, what is the height of the object? Etc. can be measured from a calibrated camera. To get the image points one must first acquire a series of known world points. These world points are then used to find the intrinsic and extrinsic parameters of the cameras.

### Extrinsic Parameters:

- Rotation.
- Translation.
- These relate the coordinate system of the camera to an ideal fixed world coordinate system. The position of the coordinate system and its orientation are also given.

### Intrinsic parameters:

- Intrinsic parameters are specific to a camera.
- These relate the coordinate system of the camera to the ideal fixed world coordinate system.
- The Focal Length ( $f_x, f_y$ ) and Optical Center (Center of the image ( $c_x, c_y$ )) are represented as a matrix called Camera Matrix.
- It depends on the camera, so once calculated it can be stored for future purposes.

$$\text{Camera Matrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$



A Good calibration is the basis for reconstructing the 3D world from 2D images as the Extrinsic Parameters determine the transformation matrix for aligning the Camera coordinate system with the world coordinate system.

One of the most common methods used for calibration is to use planar objects and apply various affine transformation to them with respect to the camera to develop an independent series of data points. Usually a Checkerboard pattern such as the one shown in the figure is used for this purpose.

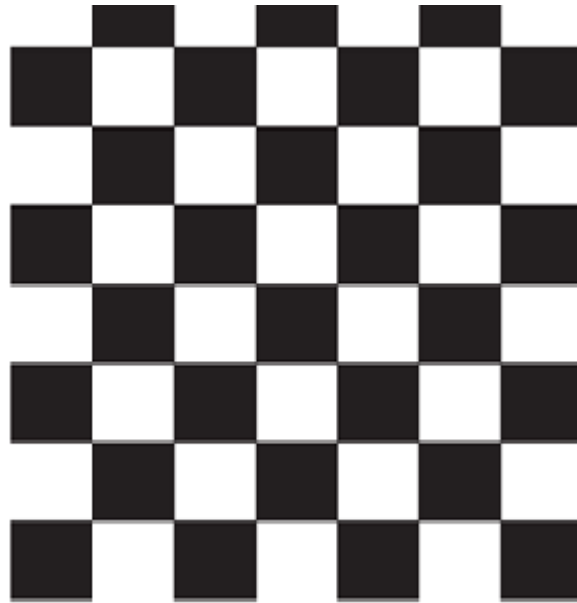


Figure 2.5: Checkerboard used for Camera Calibration

**Steps of Calibration:**

- 1) Corners of the checkerboard are found using edge detection.
- 2) Lines are fit to the data points.
- 3) The intersections of the lines are computed in order to find the corners.

The corners are transformed into world points by assuming the top left corner of the checkerboard as the origin and then imposing that the distance between each square between neighbouring corners remain constant.

The process of calibration first detects the corners of the inner squares and then uses them the corners to calibrate an image. The extrinsic and intrinsic parameters of the cameras are calculated from the calibrated image.

## 2.3 Image Rectification

Image rectification is a transformation process used to project two-or-more images onto a common image plane. It corrects image distortion by transforming the image into a standard coordinate system.

- One of the main reasons for the application of rectification is to reduce the complexity of the correspondence problem by finding matching points between the images.
- It is used to merge images taken from multiple viewpoint perspectives into a common coordinate system.

---

### Algorithm 1 :Rectification

---

- 1: First both camera matrices are separately factorized into three parts:
  - The internal camera matrix  $A_0$ .
  - The rotational matrix  $R_0$  that gives the rotation between the camera's frame of reference and the world frame of reference.
  - The translation matrix  $t_0$  that gives the translation between the camera's frame of reference and the world frame of reference.

$$\text{This means } P_0 = A_0[R_0 \ j \ t_0]$$

- 2: The new rotational matrix  $R_n$  is constructed. This matrix makes sure that in the new cameras reference systems, the x-axis is parallel to the baseline. The baseline is simply the line between the two optical centers, which can be retrieved from  $P_0$ .
  - 3: The new internal camera matrix  $A$  is constructed. This can be chosen arbitrarily and in this algorithm the mean between both original inner camera matrices ( $A_{01}$ ;  $A_{02}$ ) is used.
  - 4: The new camera matrices are now  $P_{n1}=A[R \ | \ -R_{c1}]$  and  $P_{n2}=A[R \ | \ -R_{c2}]$ . The optical centers are unchanged.
  - 5: A mapping is created that maps the original image planes of  $P_{01}$ ;  $P_{02}$  to the new  $P_{n1}$ ;  $P_{n2}$ . Because in general the pixels in the new images don't correspond to integer positions in the old images, bilinear interpolation is used to ll up the gaps
-

## 2.4 Disparity Calculation

Pairs of cameras are used in almost all high-end robotics systems to provide depth perception. While in humans depth perception is taken for granted, judging depth is difficult for computers, and remains a subject of ongoing research.

The fundamental process can be viewed as two different problems. The problem of finding corresponding points in two different views of the same scene and a simple triangulation can then be used to determine the distance.

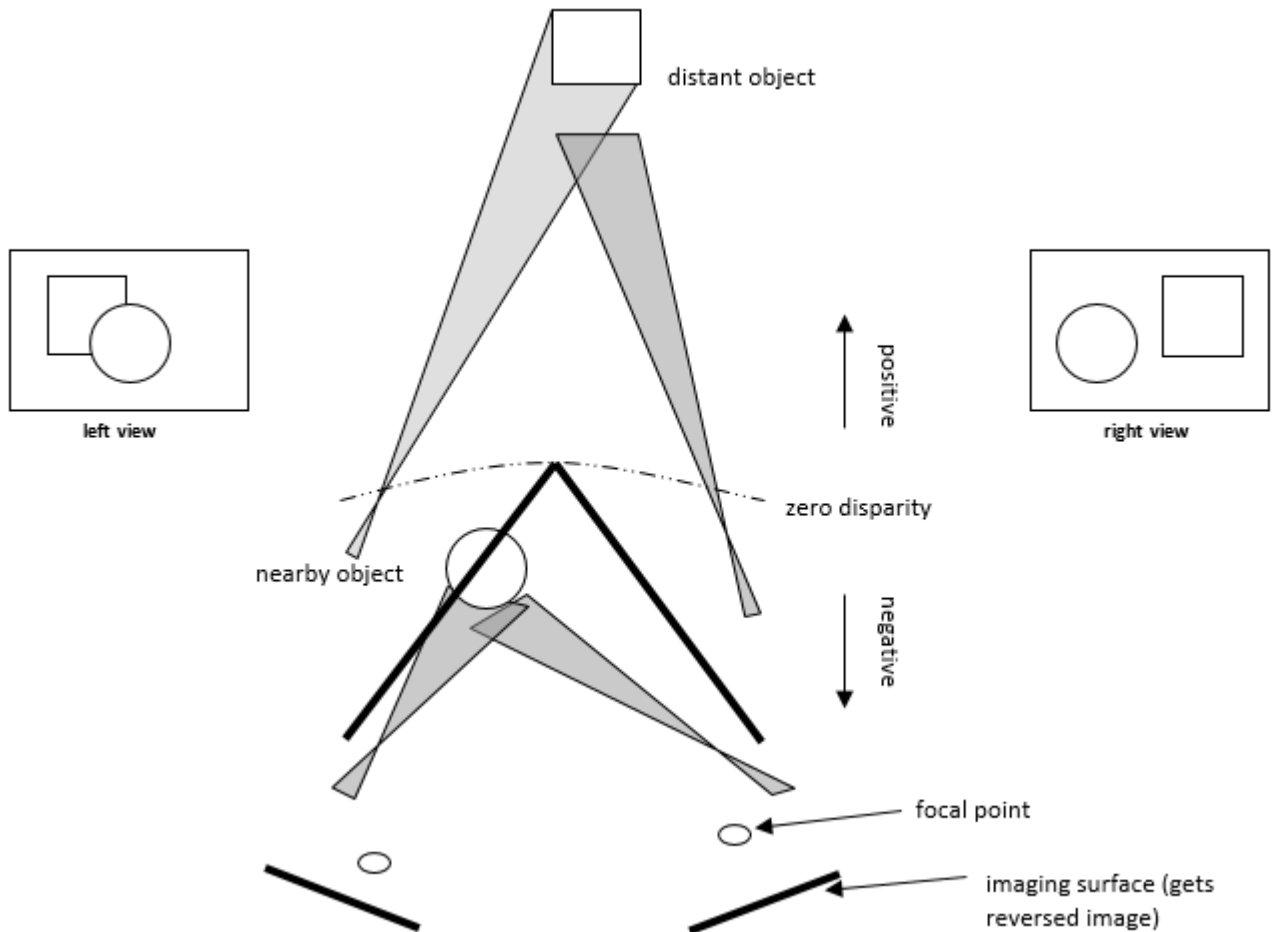


Figure 2.6: Typical Camera Configuration

The above figure depicts a typical camera configuration, with the cameras pointing somewhat inwards. Suppose the coordinates of two corresponding points are  $(x_{left}, y_{left})$  and  $(x_{right}, y_{right})$ . For cameras that are properly aligned,  $y_{left} = y_{right}$ . The disparity is defined to be  $x_{right} - x_{left}$ . This value can be positive or negative, depending on the angle of the cameras as well as the distance to the object.

The most common approach in both stereo disparity calculations and motion compensation is to slide a block taken from one image over a second image. This approach is known as the Semi Global Block Matching Algorithm, which is the algorithm used in our system.

#### **SGBM Algorithm:**

The Semi-Global Block matching algorithm has been implemented as an optimization algorithm using Dynamic programming to calculate the minimum of energy evaluation function. The SGBM algorithm gives a much clearer disparity map than the BM algorithm.

For the SGBM algorithm these steps are filled in as follows:

**Cost computation** can be done from the intensity or color pixel values. Radiometric differences also have to be kept in mind, so the gradient can also be calculated.

**Cost aggregation** connects the costs of the neighbourhood of pixels to find the cheapest (thus matching) pixel in the compare image. This is done through a global energy function from all directions through the image.

**Disparity computation** calculates the real disparity from the previously calculated energy through a winner-takes-all implementation

---

**Algorithm 2** :Semi Global Block Matching[14]

---

**1: Cost Calculation:**

$$C_{BT}[(x - \min X) * \max D - \min D + (d - \min D)]$$

**2: Cost Aggregation:**

$$E(D) = \sum_p (C_{BT}(p, D_p) + \sum_{q \in N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2 T[|D_p - D_q| > 1])$$

**3: Disparity Calculation:** For every base image  $I_b$ , the base disparity map  $D_b$  is calculated by choosing the disparity with the least cost for every pixel. If  $D_m$  be the disparity map of the match image from  $I_m$ , it is calculated by searching on the epipolar line of the pixel in the match image. In order to enhance the quality of the disparity map, calculation of disparities can be done with  $I_m$  as base image and  $I_b$  as match image. After the calculation of  $D_m$  and  $D_b$ , a consistency check is done between the two. If the consistency between the two is too high the disparity is set to invalid.

$$L_r(p, d) = C_{BT}(p, d) + \min(L_r(p - r, d), L_r(p - r, d - 1) + P_1, L_r(p - r, d + 1) + P_1, \min_i L_r(p - r, i) + P_2) - \min_k L_r(p - r, k)$$

---

**Distance Estimation:**

This sections deals with the estimation of distance of an object from the camera which is one of the objectives of this thesis. This can be considered as the Triangulation problem of Stereo vision.

**Triangulation Problem:** Triangulation in stereo analysis can be defined as the task of computing the 3D position of points in the images, given the disparity map and the geometry of the stereo setting.

The relationship between the world coordinates of a point  $P(X, Y, Z)$  and the coordinates of the point in the image plane  $(x, y)$  of a camera can be given by:

$$x = f * \frac{X}{Z}$$
$$y = f * \frac{Y}{Z}$$

where  $f$  is the focal length of the camera.

The 3D position  $(X, Y, Z)$  of the point  $P$  can be reconstructed by taking the perspective

projection of the point P on the image planes of the cameras once the relative position and orientation of the two cameras is known.

We choose the left camera reference system to be the 3D world reference system. The right camera is translated and rotated with respect to the left one.

The simplest case of transformation arises when the optical axes of two cameras are parallel, and the translation of the right camera is only along the X axis.

When both the translation vector and rotation matrix, T and R respectively are given, and they describe the transformation from the left camera coordinate system to the right camera coordinate system , the equation that gives the stereo triangulation is:

$$p' = R^T(p - T)$$

where , p and p' are the coordinates of P in the left and right camera coordinate systems respectively, and  $R^T$  is the transpose matrix of R.

So once both the coordinate systems are aligned to each other, the distance estimation becomes a trivial problem of using the properties of triangles to calculate the distance.

The following figure describes the situation and the equations that follow propose a method for this calculation.

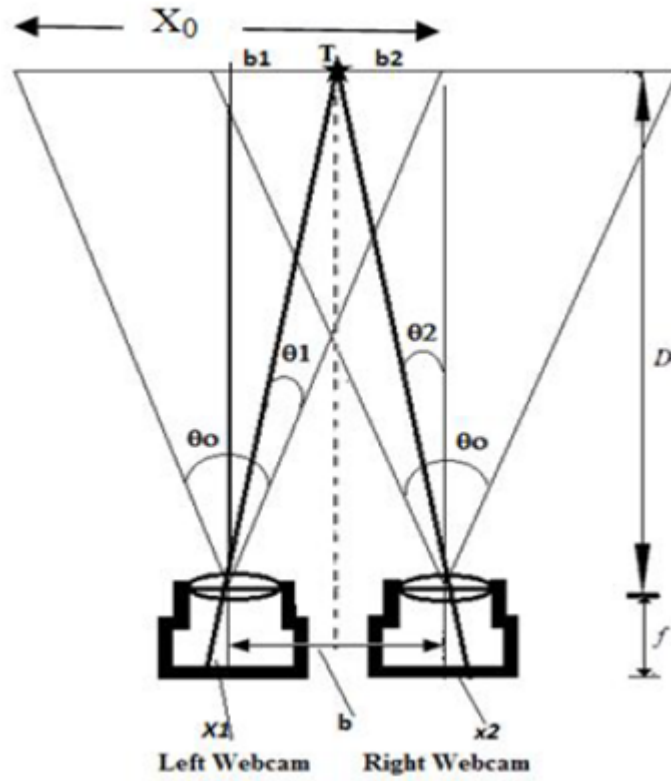


Figure 2.7: Distance estimation by Triangulation.

$$\frac{b_1}{D} = \frac{-x_1}{f}$$

$$\frac{b_2}{D} = \frac{-x_2}{f}$$

since

$$b = b_1 + b_2, \text{ then,}$$

$$b = \frac{d}{f} * (x_2 - x_1)$$

$$D = \frac{bf}{(x_2 - x_1)}$$

$$\tan \frac{\theta_0}{2} = \frac{x_0}{D} = \frac{x_1}{f}$$

$$f = \frac{x_0}{2 \tan \frac{\theta_0}{2}}$$

$$D = \frac{bx_0}{2 \tan \frac{\theta_0}{2}(x_2 - x_1)} \quad (2.3)$$

In equation( 2.3) above,  $x_0$  is the width of the image in pixels,  $x_2-x_1$  is the disparity between the two images in pixels. Since the distance between both webcams  $b$ , the webcam angle view  $\theta_0$  and image width are constant for 3D webcam, then it seems the distance  $D$  is inversely proportional to disparity ( $x_2-x_1$ ). In order to compensate alignment errors, the view angle must be corrected by adding another term  $\phi$  within the tangent in equation( 2.3) as follows:

$$D = \frac{bx_0}{2 \tan(\frac{\theta_0}{2} + \phi)(x_2 - x_1)} \quad (2.4)$$

The alignment compensation term  $\phi$  must be found experimentally. Therefore, equation( 2.4) can be written as a power relation:

$$D = k.x^z \quad (2.5)$$

where  $k$  is constant given by,

$$k = \frac{bx_0}{2 \tan(\frac{\theta_0}{2} + \phi)}$$

$x=x_2-x_1$ , and  $z$  is a constant to be found out experimentally.

The distance estimated after finding the disparity is used in the calculation of these parameters by using the relations between these parameters and the distance as follows.



# Chapter 3

## Results

The Cameras used to conduct this experiment with stereo vision were Logitech C920 Pro Webcams. The input images and the resulting outputs of each phase of pre-processing are shown in this section.

### Camera Calibration:



Figure 3.1: Left and Right image pair#01

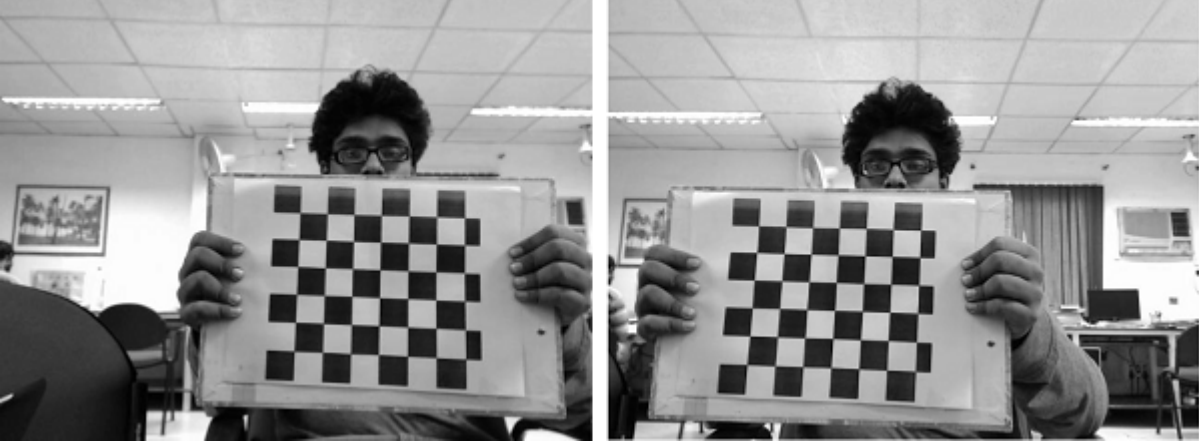


Figure 3.2: Left and Right image pair#02



Figure 3.3: Left and Right image pair#03

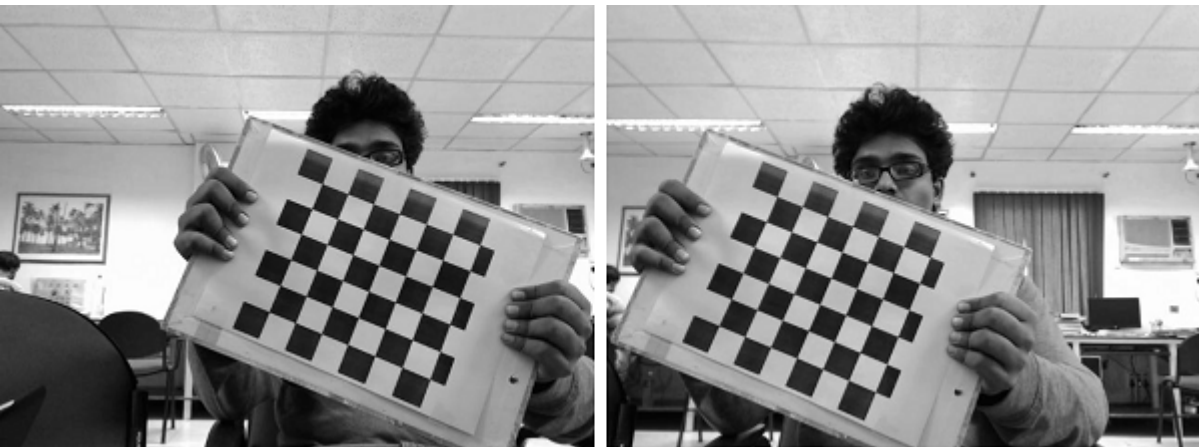


Figure 3.4: Left and Right image pair#04

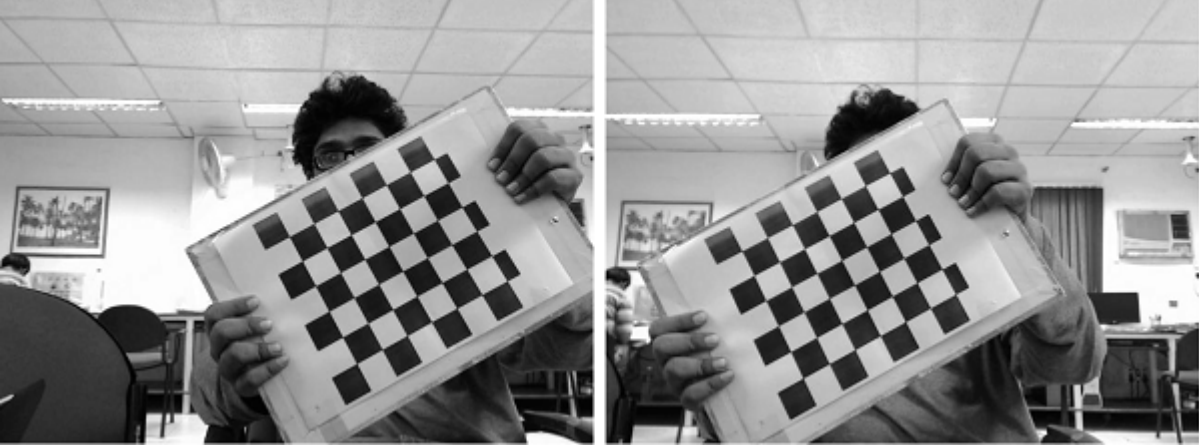


Figure 3.5: Left and Right image pair#05

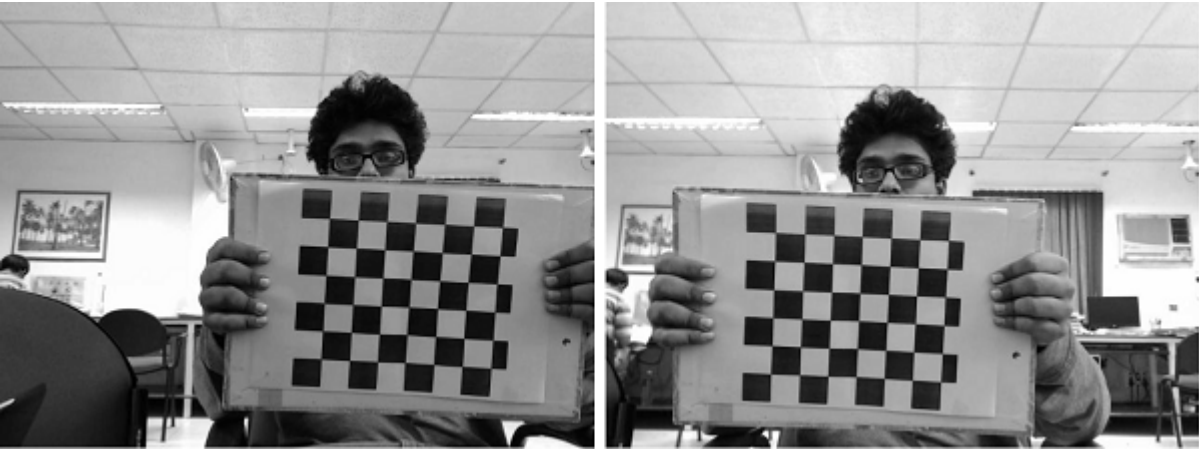


Figure 3.6: Left and Right image pair#06

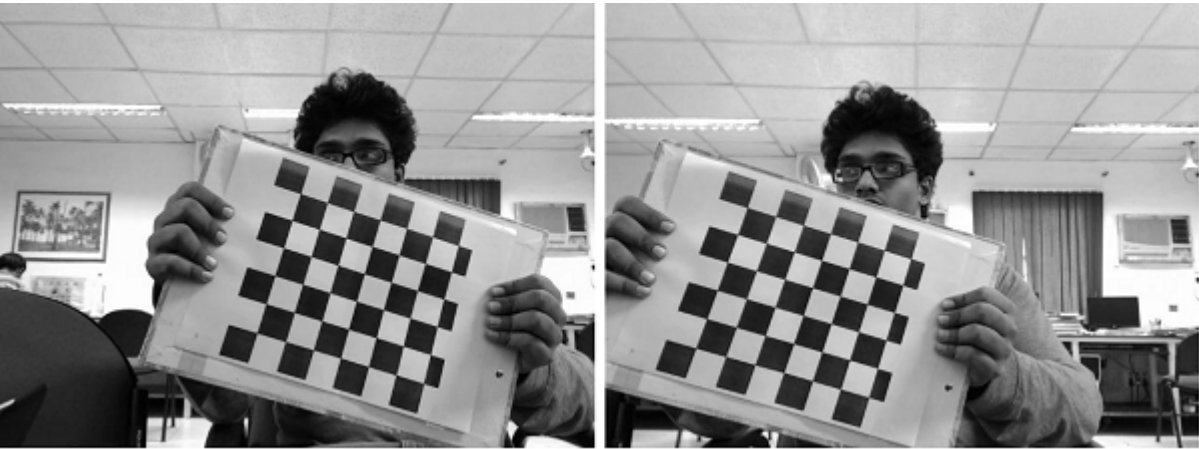


Figure 3.7: Left and Right image pair#07

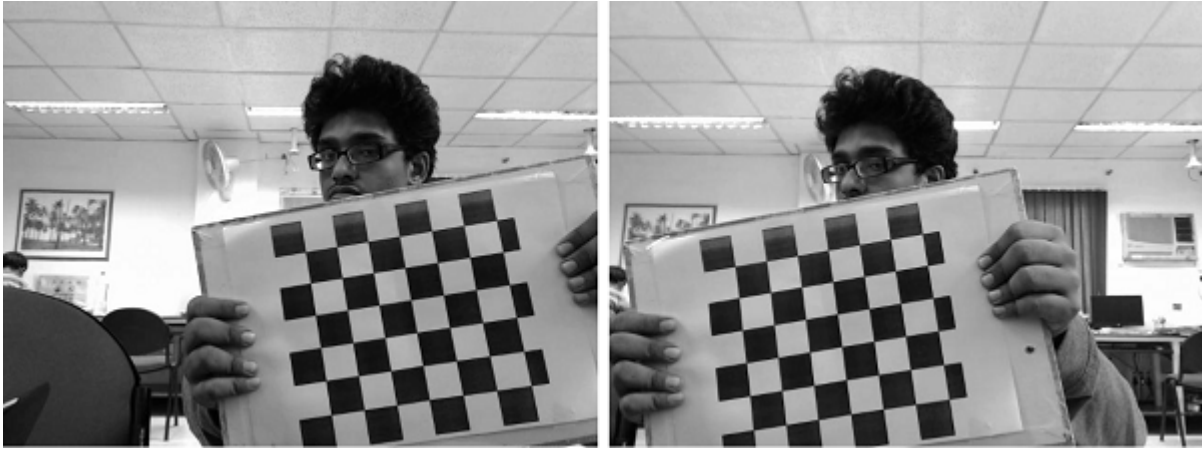


Figure 3.8: Left and Right image pair#08



Figure 3.9: Left and Right image pair#09

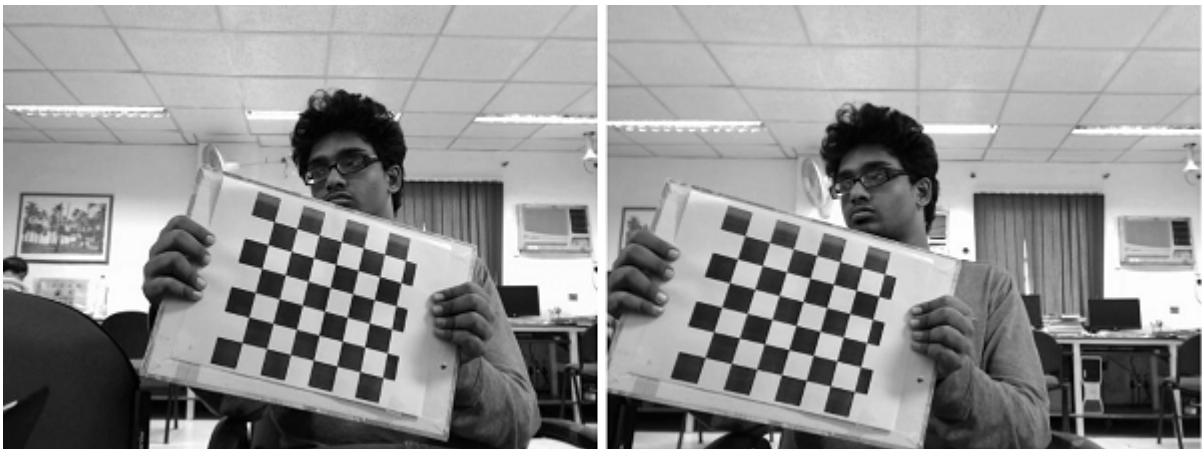


Figure 3.10: Left and Right image pair#10

Corner Detection:

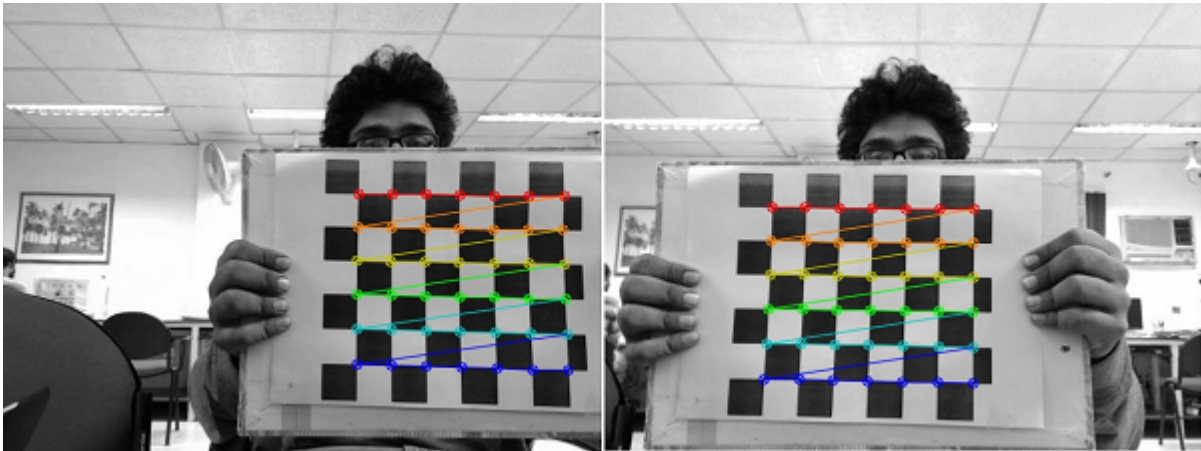


Figure 3.11: Corner Detection in Left and Right image pair#01

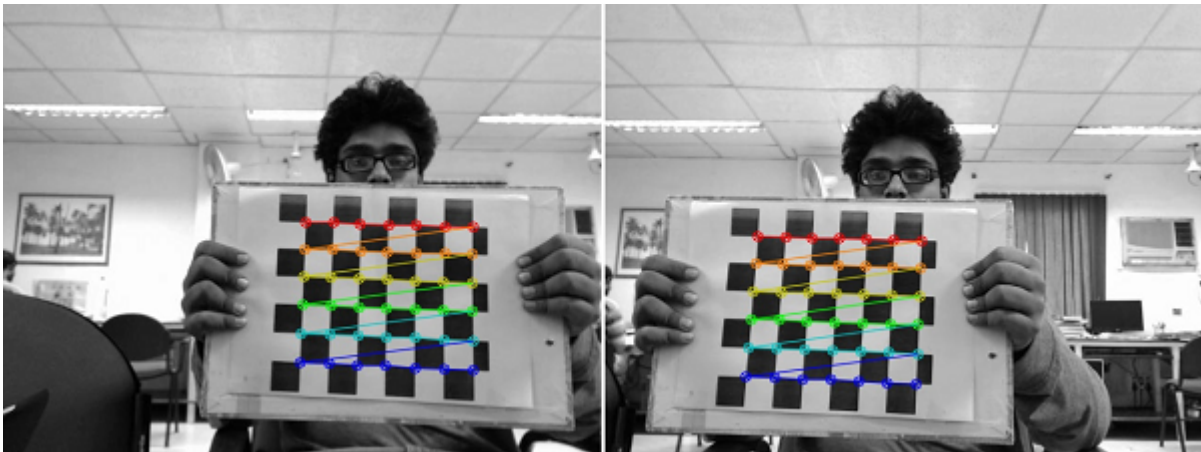


Figure 3.12: Corner Detection in Left and Right image pair#02

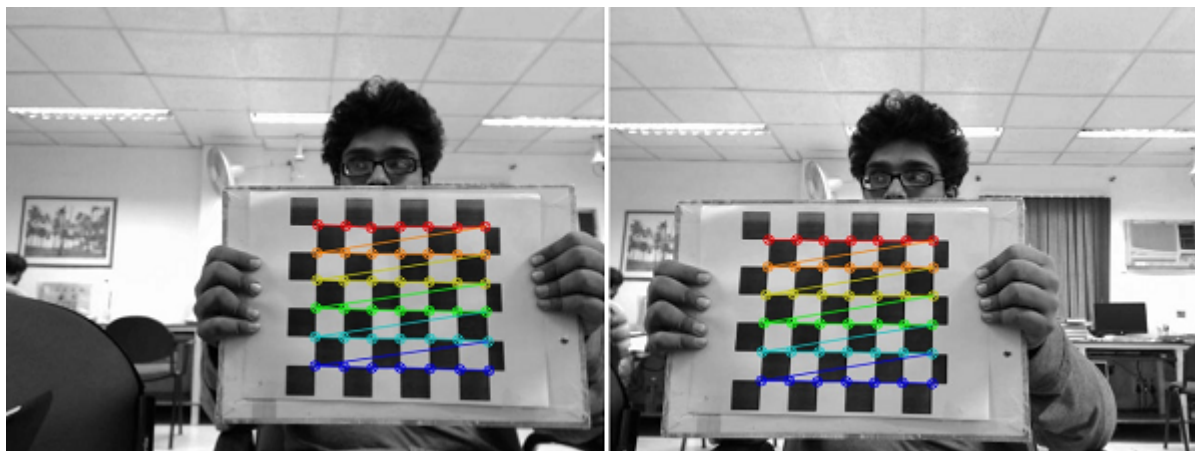


Figure 3.13: Corner Detection in Left and Right image pair#03

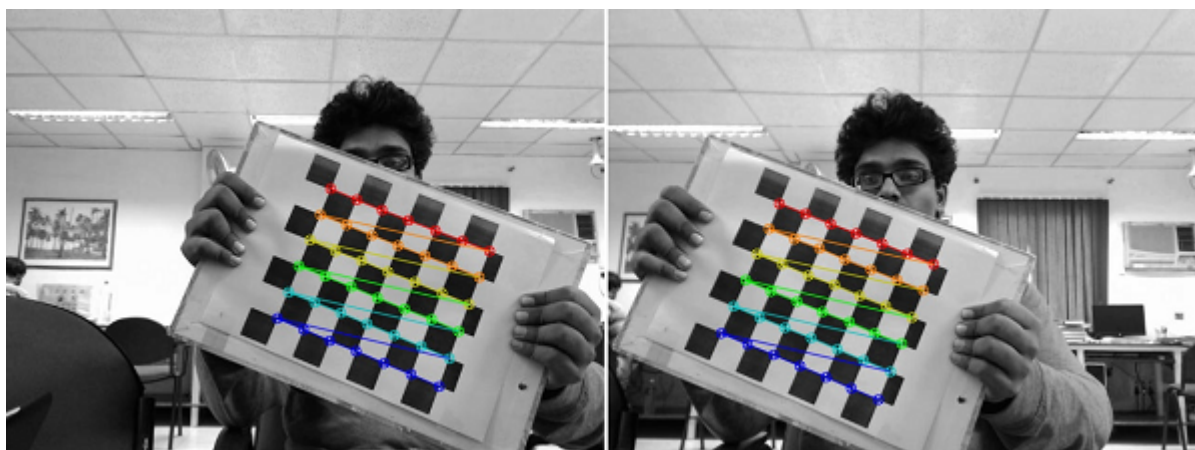


Figure 3.14: Corner Detection in Left and Right image pair#04

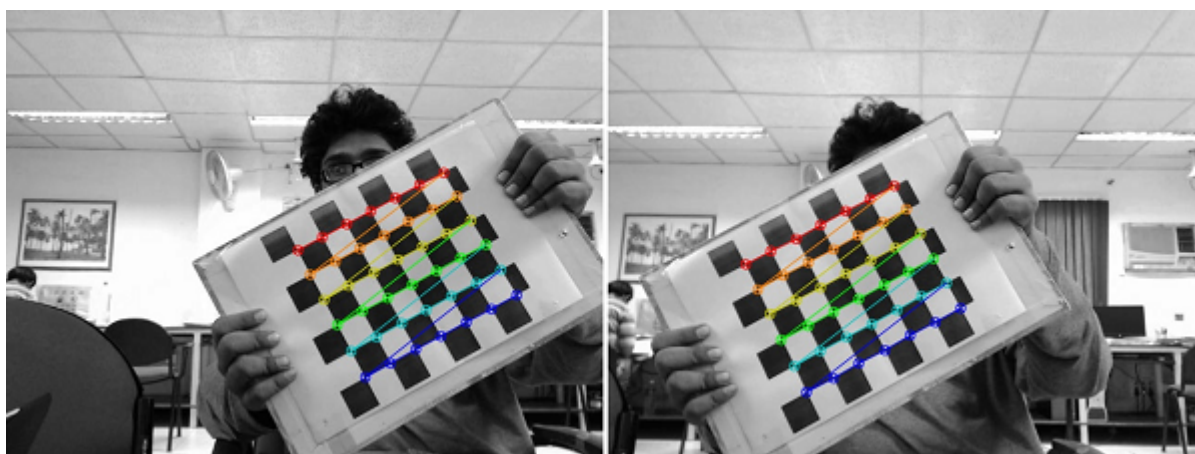


Figure 3.15: Corner Detection in Left and Right image pair#05

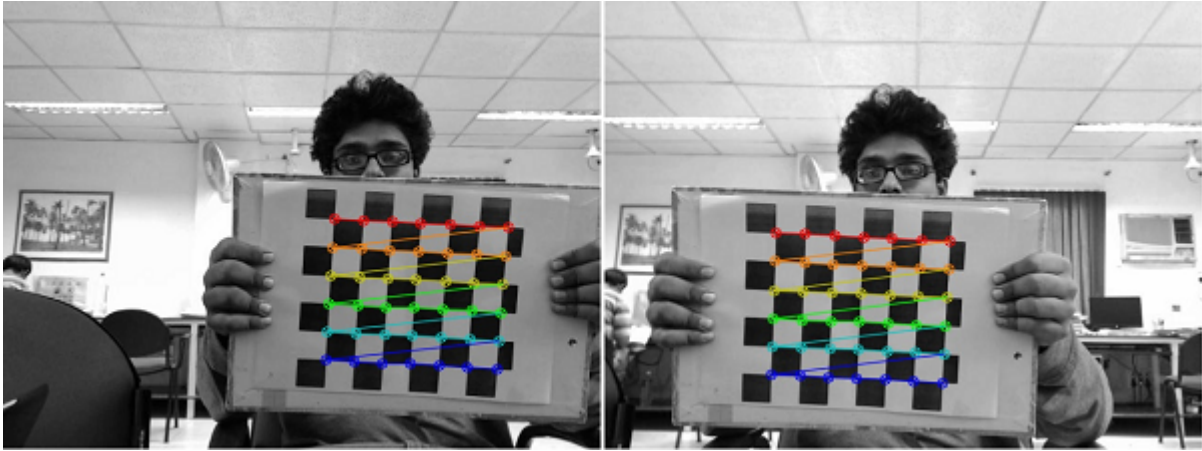


Figure 3.16: Corner Detection in Left and Right image pair#06

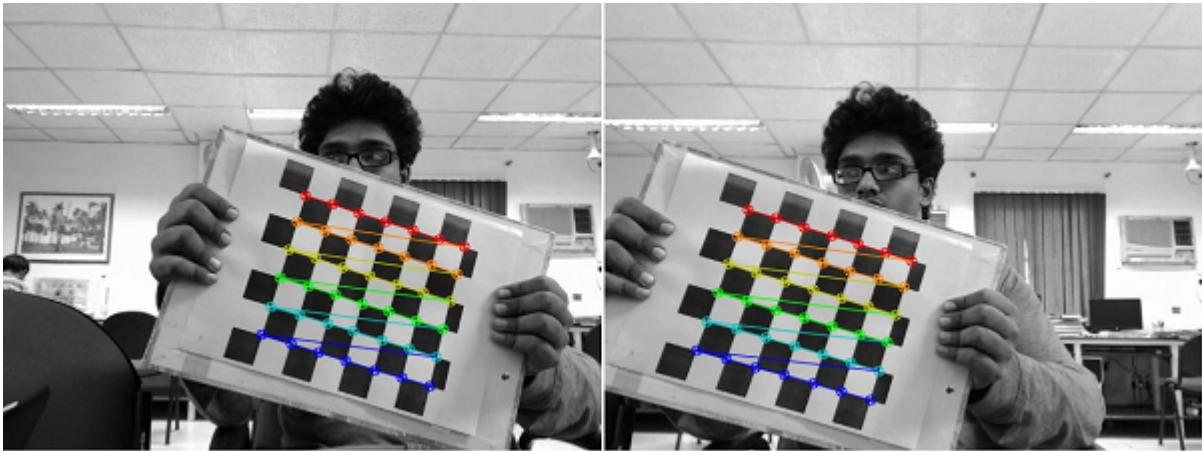


Figure 3.17: Corner Detection in Left and Right image pair#07



Figure 3.18: Corner Detection in Left and Right image pair#08

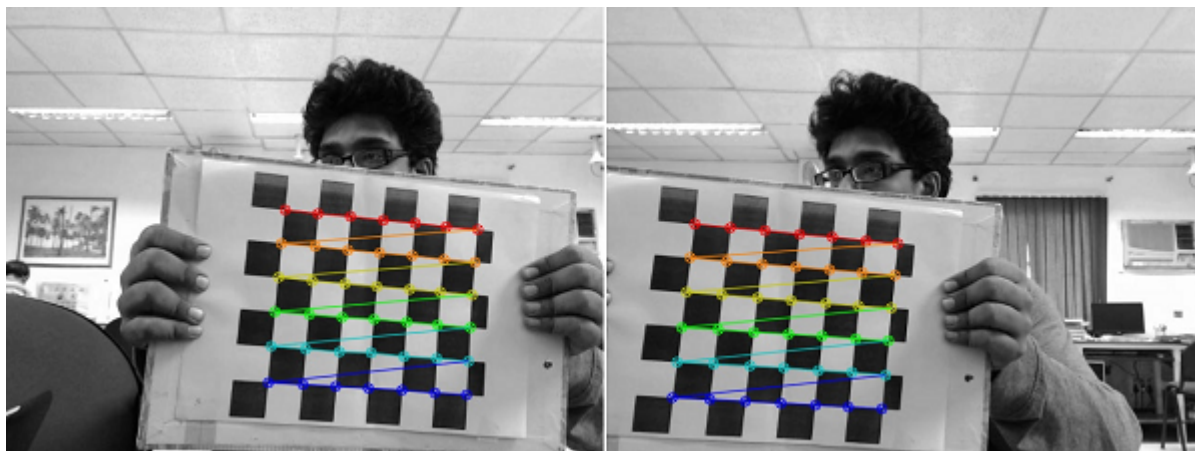


Figure 3.19: Corner Detection in Left and Right image pair#09

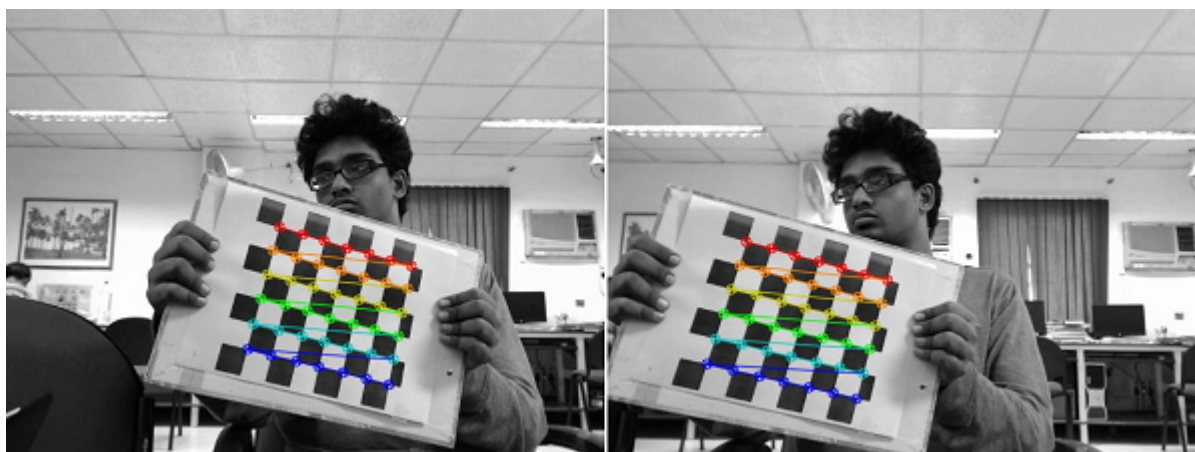


Figure 3.20: Corner Detection in Left and Right image pair#10



Rectified Images:

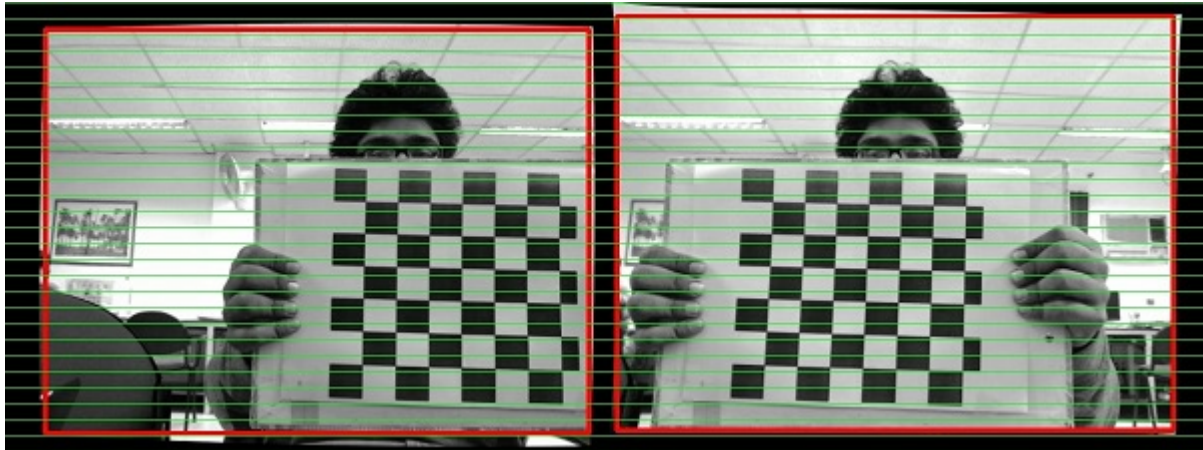


Figure 3.21: Rectified Left and Right image pair#1

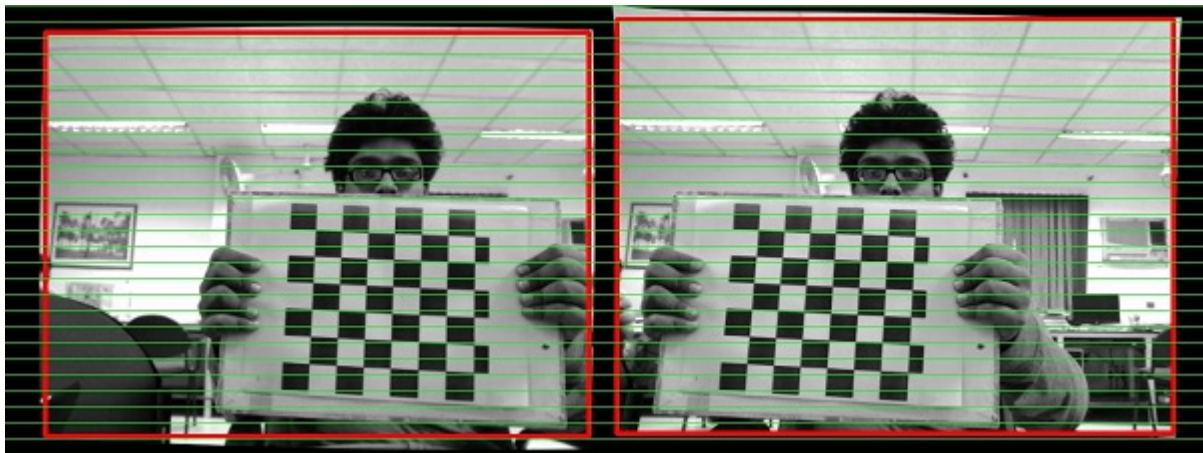


Figure 3.22: Rectified Left and Right image pair#2

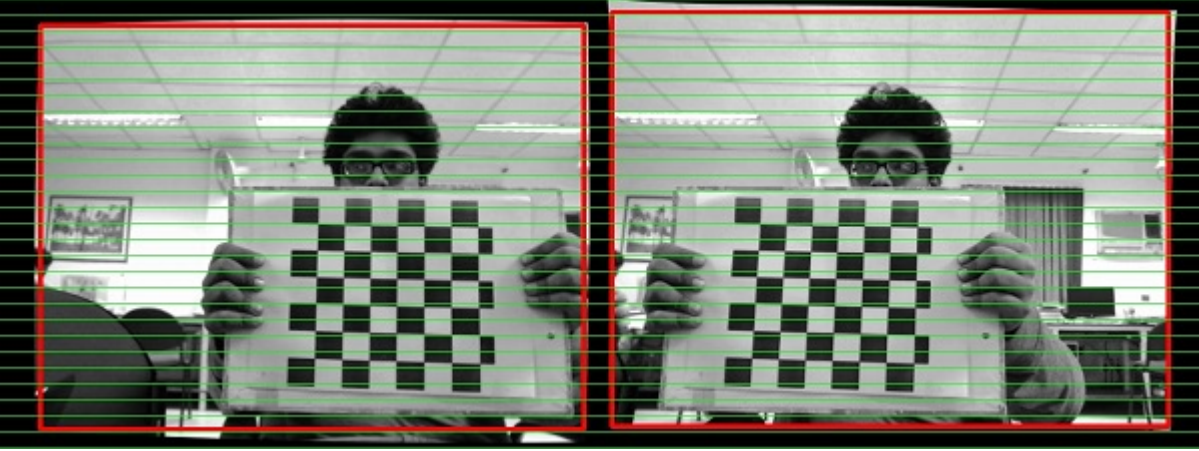


Figure 3.23: Rectified Left and Right image pair#3

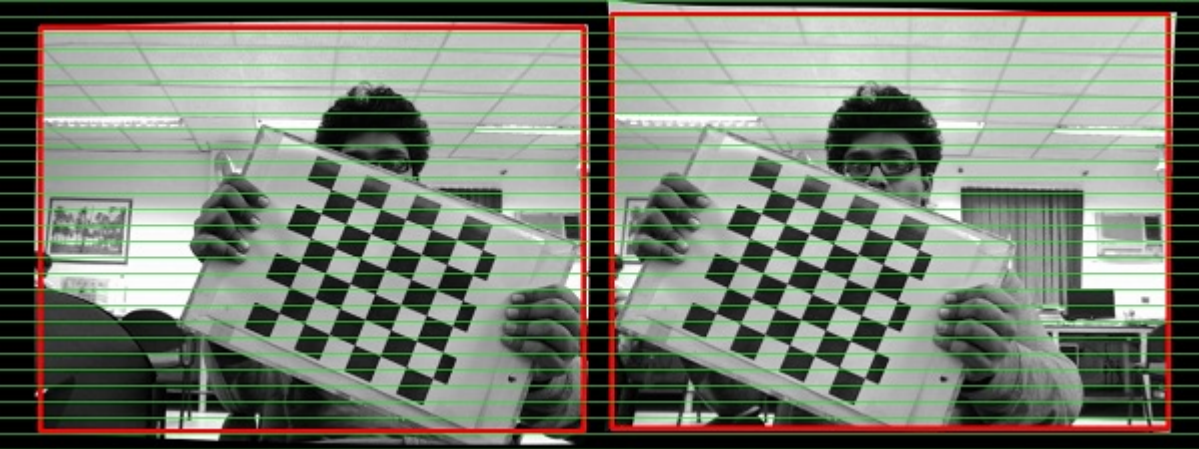


Figure 3.24: Rectified Left and Right image pair#4



Figure 3.25: Rectified Left and Right image pair#5

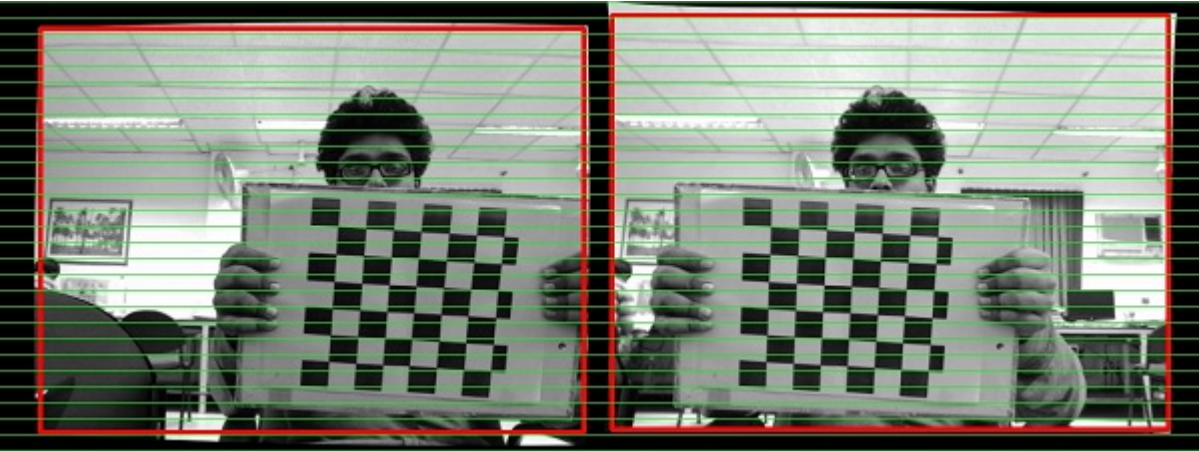


Figure 3.26: Rectified Left and Right image pair#6

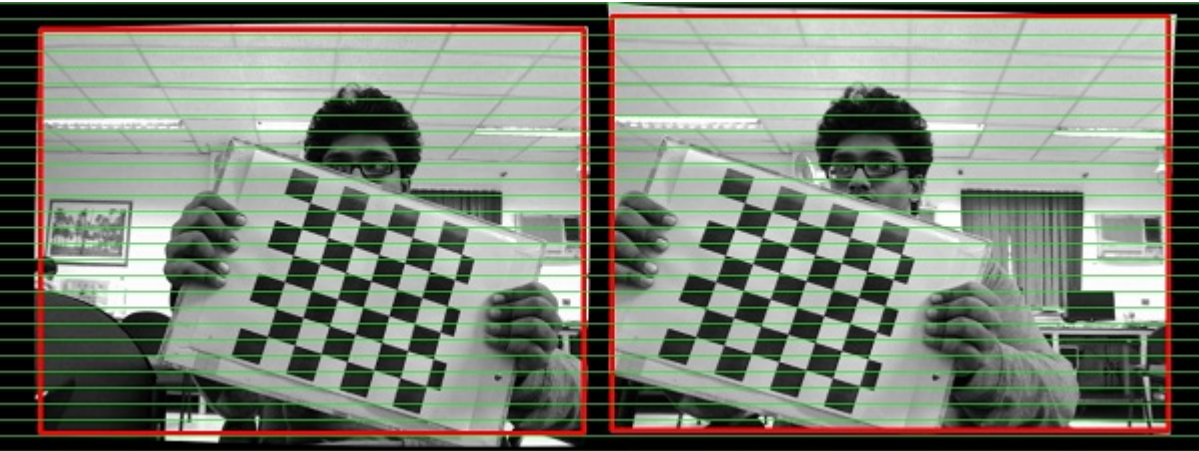


Figure 3.27: Rectified Left and Right image pair#7



Figure 3.28: Rectified Left and Right image pair#8

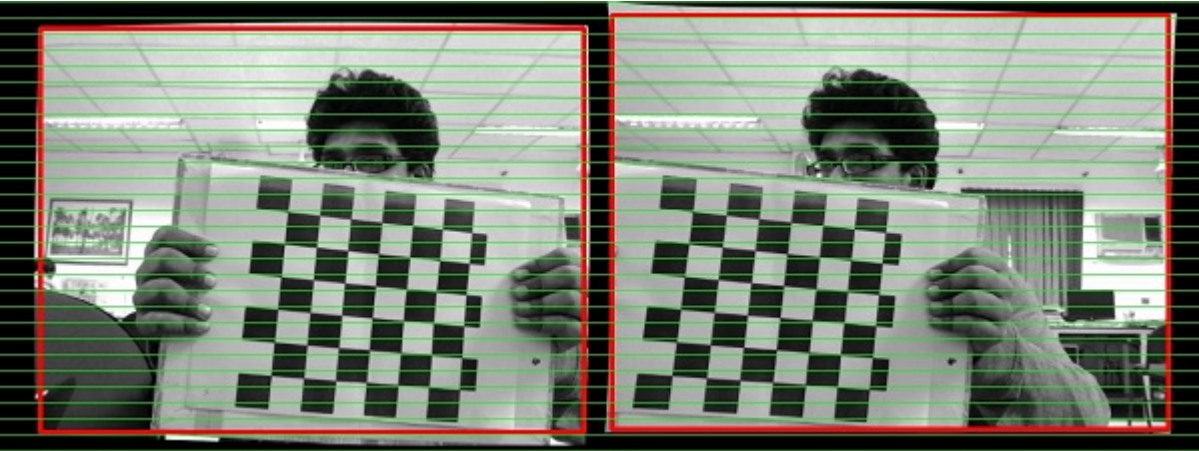


Figure 3.29: Rectified Left and Right image pair#9

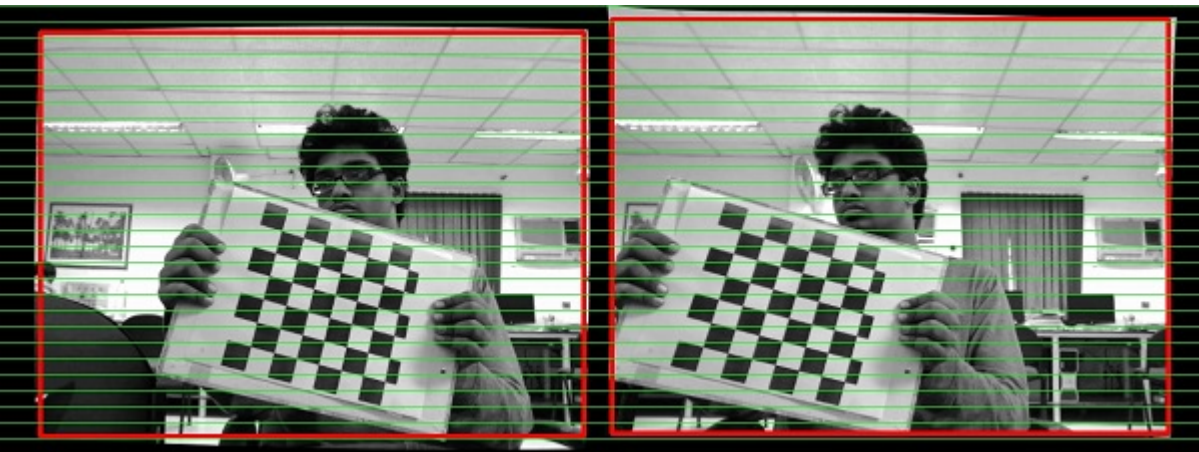


Figure 3.30: Rectified Left and Right image pair#10

Disparity Calculation:

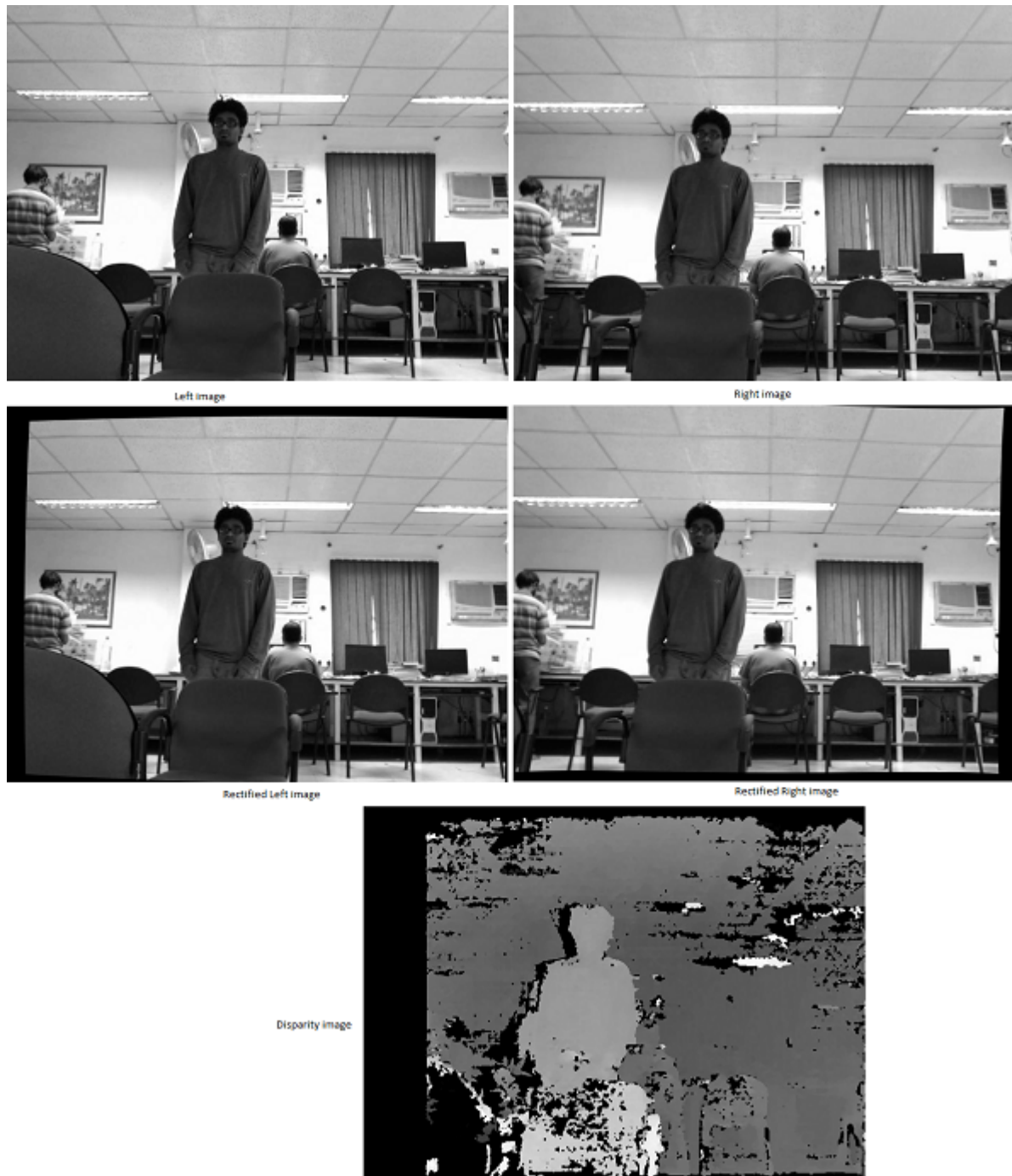


Figure 3.31: Disparity Map#01



Figure 3.32: Disparity Map#02

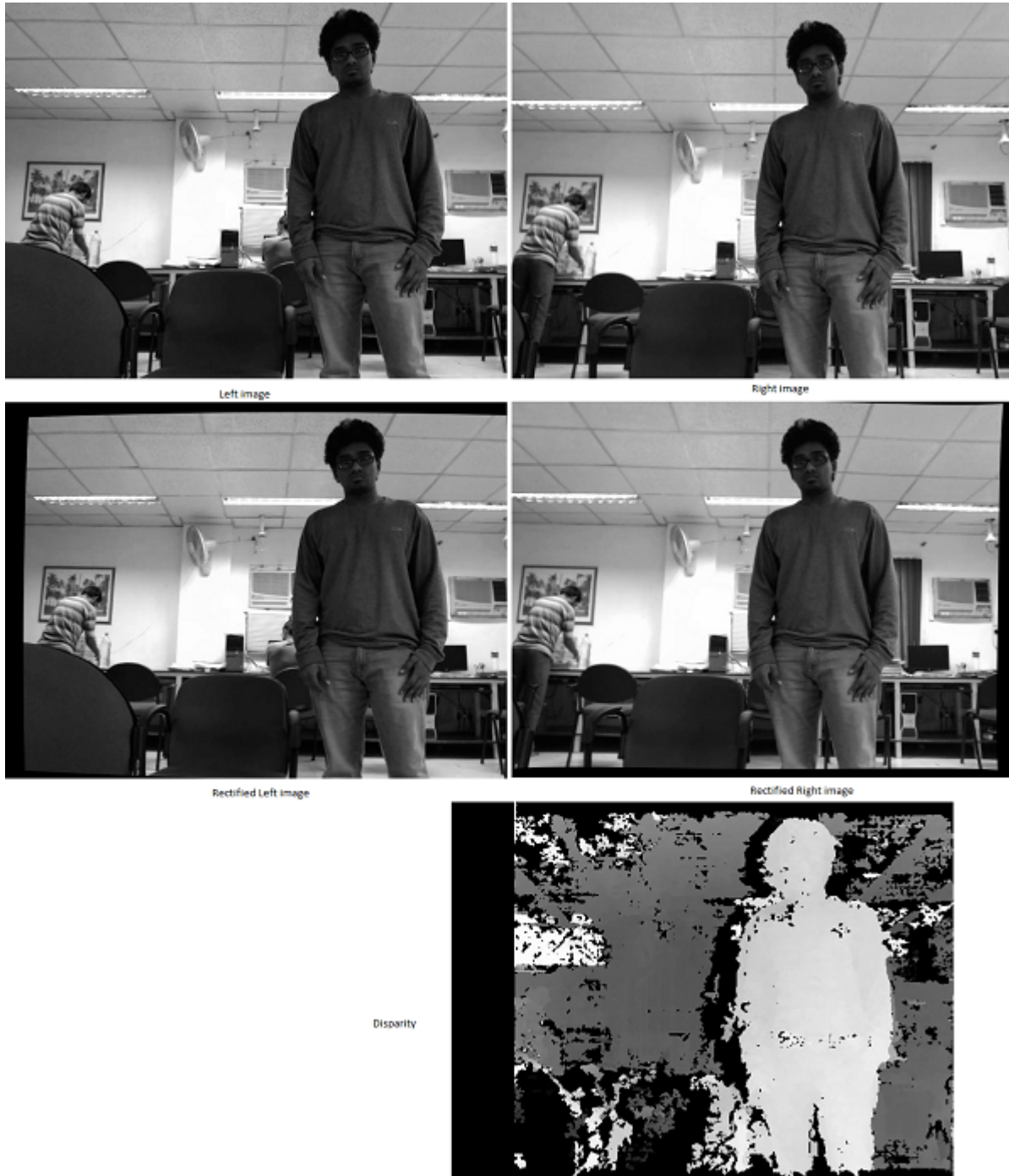


Figure 3.33: Disparity Map#03

# Chapter 4

## Conclusion

This system estimates the distance between the camera and a point in the image which lies on the object that is of interest. The disparity map calculated has been implemented using SGBM and BM algorithms with SGBM giving the better results. The speed of the system can be improved further by developing better algorithms to tackle the Correspondence problem. The disparity calculated using the algorithms is used to calculate the distance. After estimating the distance it is used to calculate the zoom parameter of the PTZ camera.

During this particular experiment two webcams with Auto-Focus were used which hindered our progress as the focal length of the camera which was pivotal in determining the distance by Triangulation was changing for every image. If this is rectified, using webcams with constant focal length the distance can be easily estimated using our implementation which can in turn be used to calculate the Pan, Tilt and Zoom of the PTZ camera.

And furthermore this system is a binocular stereovision system which limits the performance of the system. It estimates the distance between the camera and a point in the image which lies on the object that is of interest only if the object is in the field of vision of the cameras. If due the object is obstructed from the even one of the cameras, the system cannot function. To overcome this problem a system of many stereo pairs can be used. If the object is obstructed from one pair the other pairs are used to estimate the distance. But this further brings in a problem of finding a common coordinate system for all the camera pairs. So all these problems can be taken up to extend the system and to improve its performance



# Bibliography

- [1] Roger Y. Tsai, A Versatile Camera Calibration Techniaue for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses, IEEE Journal Of Robotics And Automation, VOL. RA-3, NO. 4, AUGUST 1987
- [2] A. Fusiello, E. Trucco and A. Verri, A compact algorithm for rectification of stereo pairs, Machine Vision Appl., 12 (1) (2000), pp: 16-22.
- [3] Meredith Drennan, An Implementation of Camera Calibration Algorithms, Department of Electrical and Computer Engineering Clemson University
- [4] Bouguet, Jean-Yves, Camera Calibration Toolbox for Matlab, [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/) (2008)
- [5] Cheng-Yuan Ko, Chung-Te Li, Chen-Han Chung, and Liang-Gee Chen, High Accuracy Users Distance Estimation by Low Cost Cameras, Proceedings of 3DSA2013, Selected paper 3
- [6] OpenCV: Open source Computer Vision library, <http://docs.opencv.org/>.
- [7] Loop C, Zhang Z, Computing rectifying homographies for stereo vision. In: CVPR99, Fort Collins, CO,(1999) pp I:125131
- [8] J. Mrovlje and D.Vranci, Distance measuring based on stereoscopic pictures, 9th International PhD Workshop on Systems and Control: Young Generation Viewpoint, Izola, Slovenia, Oct.1- 3, (2008).
- [9] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, (2008)
- [10] Xinghan Luo, Robby T. Tan, Remco C. Veltkamp, Multiple People Tracking Based on Dynamic Visibility Analysis
- [11] P. Swapna , N. Krouglicof and R. Gosine, The question of accuracy with geometric camera calibration, Proc. CCECE, pp.541 -546 (2009)

- [12] Manaf A. Mahammed, Amara I. Melhum, Faris A. Kochery, Object Distance Measurement by Stereo Vision, International Journal of Science and Applied Information Technology (IJSAIT), Vol.2 , No.2, Pages : 05-08 (2013)
- [13] [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/OWENS/LECT10/node3.html](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT10/node3.html)
- [14] Fengjun HU, Yanwei ZHAO, "Comparative Research of Matching Algorithms for Stereo Vision", Journal of Computational Information Systems 9: 13 (2013) 54575465