CSE-483: Mobile Robotics **Monsoon 2019**

Stereo Geometry

Prepared by: Ishaan Khare and Amarthya Sasi

1 Introduction

Stereo geometry is the extraction of 3D information from digital images. When we compare the information from two distinct locations we can find out 3D orientations by comparing the relative positions of the objects in the images.

In these notes we will touch upon these topics -

1. Disparity

2. Rectification

3. Point cloud generation

2 Disparity

1. The difference in image location of the same 3D point when projected under perspective to two different cameras.

   ```
   d = xleft - xright
   ```

   It means that there is a shift in the position of the 3D points when viewed by a different camera.

2. There are two cases in which we can calculate the disparity:

   ```
   1. Ideal case(Stereo setup)
   2. General case
   ```

   Figure 1 and 2 illustrates the two cases.

3. Let us find the disparity in both the cases -

   Proof. For case 1
   Taking Origin at O1 and P = (Xp, Yp, Zp). Using same ratio properties of similar triangles.
   $\frac{f}{Zp} = \frac{x1}{Xp}$
   Also,
   $\frac{f}{Zp} = \frac{-x2}{b - Xp}$
   Now, solve the equations to get Zp -

   $$Zp = \frac{bf}{x1 - x2}$$

   In the above equation the disparity is x1 - x2. Figure 3 illustrates this derivation.

   □

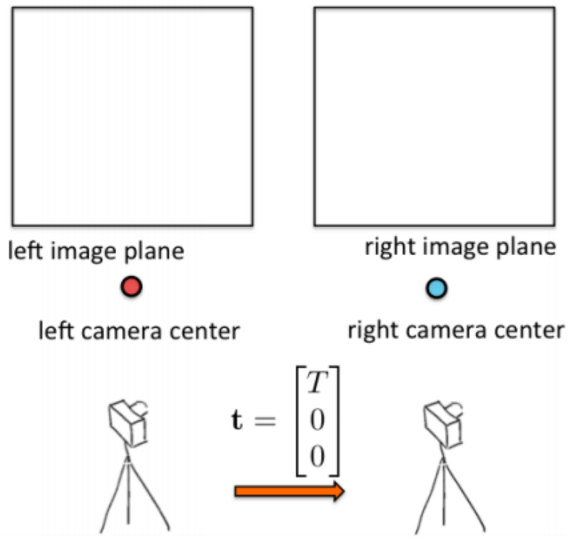

Figure 1: Case 1: Ideal stereo setup

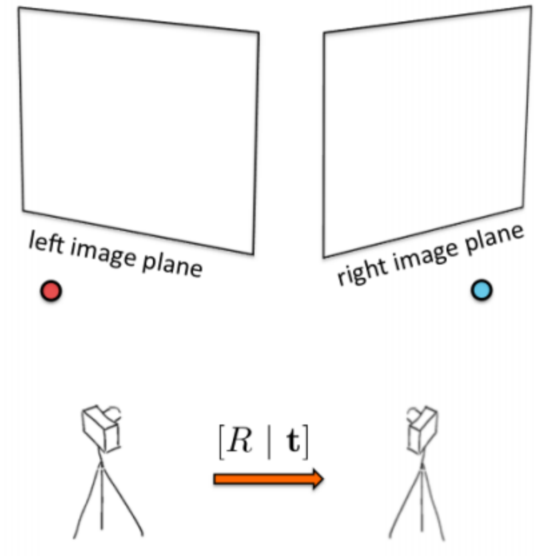

Figure 2: Case 2: General setup

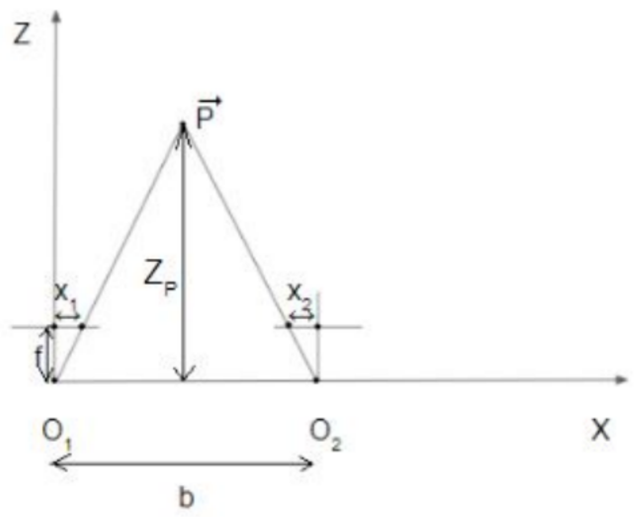

Figure 3: Diagram for case 1

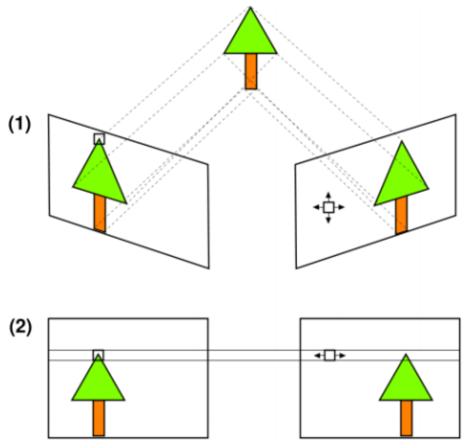

Figure 4: Rectification

Proof. For case 2

Taking Origin at O1 and our 3D point X = (Xp, Yp, Zp). Assumption, the rays formed by extending the rays x1 and x2 intersect at X.

Camera 1:

Projection matrix, P_1 = K1(I|0)

Camera 2:

Projection matrix, P_2 = K2(R|T)

$$\lambda_1 x_1 = P_1 X$$

$$\lambda_2 x_2 = P_2 X$$

The above two lines represent two epipolar lines which are parallel. The cross product of two parallel lines will be zero.

Hence,

$$x_1 \times P_1 X = 0$$

$$x_2 \times P_2 X = 0$$

Now we need to find X(our 3D point) by using the above equations. Now, we can see that each of the above equation gives us 2 equations(one from x-coordinate and other from y-coordinate). So, we have 4 equations and 3 variables(Xp, Yp, Zp).

Now solve for X by using SVD and taking the vector corresponding to the smallest eigen value. After that compute the disparity by using the equation derived in Case 1 -

$$Disparity = \frac{bf}{Zp}$$

□

3 Rectification

1. If the two cameras are aligned to be coplanar, the search for corresponding points is simplified to one dimension - a horizontal line parallel to the baseline between the cameras. This search is simplified because now we know that the disparity is only in horizontal direction and there is no disparity in the vertical direction. Figure 4 shows a how the search space reduces to 1 dimension.

   Proof. Let,

   Projection matrix of camera 1 - $\mathrm{P}_1 = \mathrm{K}_1 R_1(\mathrm{I}|\text{-T1})$

   Projection matrix of camera 2 - $\mathrm{P}_2 = \mathrm{K}_2 R_2(\mathrm{I}|\text{-T2})$

   $P_1 = \mathrm{KR}(\mathrm{I}|\text{-T1})$

   $P_2 = \mathrm{KR}(\mathrm{I}|\text{-T2})$

   Let , $R = [r_1^T, r_2^T, r_3^T]^T$

   Now since we know that baseline is parallel to the X-axis we can say that -

   $r_1^T = \dfrac{T1 - T2}{\mid T1 - T2 \mid}$,where T1 - T2 is the baseline.

   Now let us take Y-axis perpendicular to x-axis and z-axis of the old camera and find R.

   $r_2^T = Rx \times r_1^T$

   $r_3^T = r_1^T \times r_2^T$

   The new calibration matrix can be taken as -

   $\mathrm{K} = \dfrac{K_1 + K_2}{2}$

   So, now the camera pixels in the rectified image will be -

   $\lambda_1\ \mathrm{x}_1 = KR(\mathrm{I}|\text{-T1})X$

   $\lambda_2\ \mathrm{x}_2 = KR(\mathrm{I}|\text{-T2})X$

   □

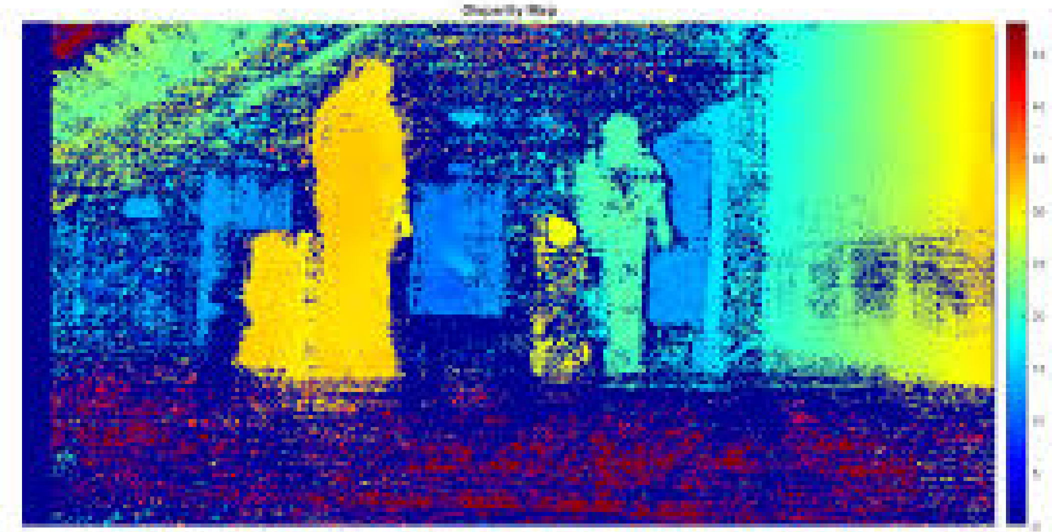

Figure 5: Disparity Map

4 Point cloud

As we know the depth is given by-

$$Zp = \frac{bf}{x1 - x2}$$

In order to construct a disparity map find out the disparity in each of the corresponding points. Then use the focal length, disparity and position to find out the disparity map.

$$x = \frac{b(x1 + x2)}{2(x1 - x2)}$$

$$y = \frac{b(y1 + y2)}{2(y1 - y2)}$$

Now if we know the disparities in the corresponding images we can find out the depth of a scene and also create a depth map.

References

[1] http://www.sci.utah.edu/ gerig/CS6320-S2012/Materials/CS6320-CV-F2012-Rectification.pdf

[2] lecture slides