ENPM 673: Perception for Autonomous Robots - Project 3

Aswath Muthuselvam
aswath@umd.edu

18th April 2022

1   Calibration

Our task is to compute the depth of a scene with given 2 photos of a scene taken from different view points. First, we find the Essential matrix of the camera poses. By using epipolar geometry, we formulate the problem as shown in the below equation, where $\mathbf{x}$ and $\mathbf{x}'$ i s the world point and F is the Fundamental matrix:

$$\mathbf{x}F\mathbf{x}' = 0$$

We rewrite the matrix in a linear equation form where the unknown elements of F are coefficients of products of $\mathbf{x}$ and $\mathbf{x}'$.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = 0$$

Then, using SVD, I decomposed F and find the column of V which corresponds to the least eigenvalue in diagonal matrix σ.

I calculated the Essential matrix E by decomposing F multiplied by camera matrix K.

$$E = K^T F K$$

Further decomposing E gives us $R1$, $R2$ and T.

- I obtain $R1$ by computing $R1 = UWV^T$.

- I obtain $R2$ by computing $R2 = UW^TV^T$. Where W^T

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Translation t is obtained from = U[:,3].

Four relative poses are possible: (R1, t), (R1, -t), (R2, t), (R2, -t). Left camera is taken as origin or reference pose with $R = I_{33}$, $t = [0, 0, 0]$. We perform chierality check and find out that only one out of the 4 poses are possible because only one point lies in-front of the camera. I triangulate the points and check which set of poses gives the maximum number of positive depth component(Z) in the 3D points.
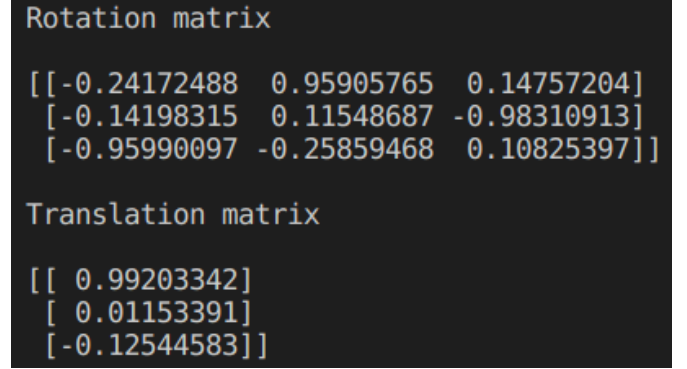

Figure 1: Rotation and Translation of curule image

2   Rectification

I computed the epipolar lines and plotted it for both the images.

Figure 2: Depth image of octagon

Using homography matrix obtained for each images, I used warpPerspective to change the scene to straight view, shown in Figure 3.
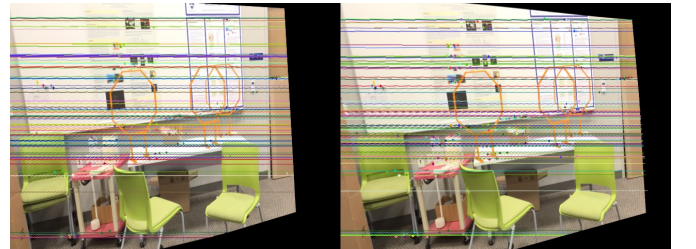

Figure 3: Depth image of octagon

3   Correspondence

Before I calculate the disparity, I noticed that I had to align both images, I had to translate one of them. However, the scale was also off as well, the x axis of the image

was squished. Hence, I did another feature matching both the rectified images using homography as shown in Figure 4.
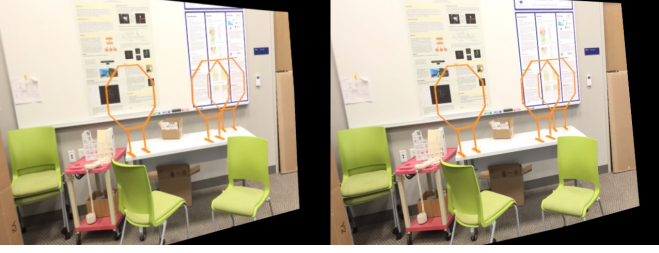

Figure 4: Aligned image of octagon

Then, I used window matching method to find the least difference to locate a pixel in its righ-hand side image counterpart. The window matching formula I used was $p = min|w_l - w_r|)$, where p is the (x, y) coordinates of the best pixel match. I tested various window sizes ranging from 5 to 20 pixels. Using a larger window size gave a reliable but dilated depth map, I used window size of 20 for curule image as shown in Figure 7. Using smaller window size give a finer resolution as shown in Figure 9. For octagon, I used a smaller window size of 5, because of the thin edges of the octagon, it can be better located.
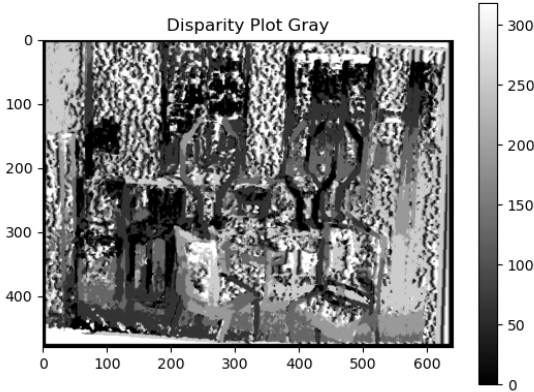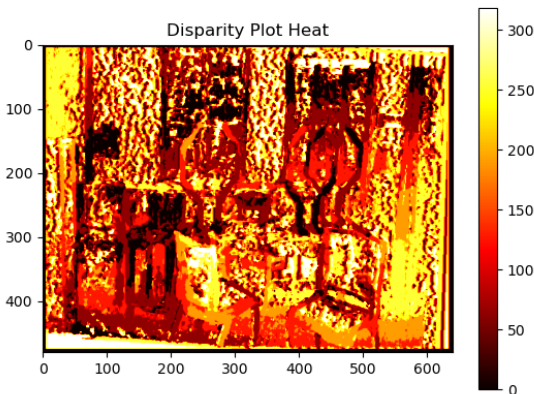

Figure 5: Disparity gray image of octagon

Figure 6: Disparity color image of octagon

4   Compute Depth Image

I computed the depth using the following equation, using baseline B, focal point f_x and disparity d :

$$depth = \frac{Bf_x}{d}$$

From Figure 9, we can see that the chair is in the front(very light orange), followed by the lone octagon(darker shade of orange), followed by 2 other octagons(dark purple). We can also see that the left most chair is further away than the middle chair, and this region of the image is different from the disparity image in Figure 6.
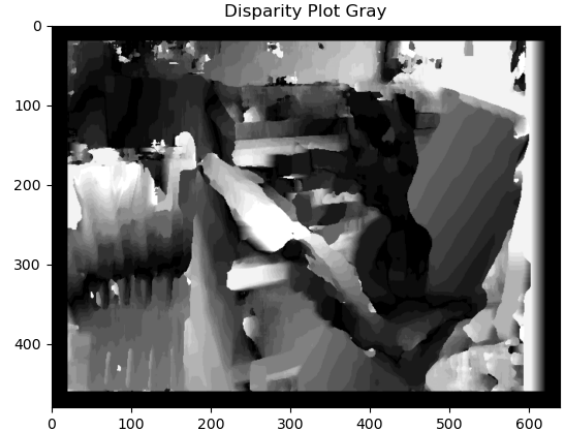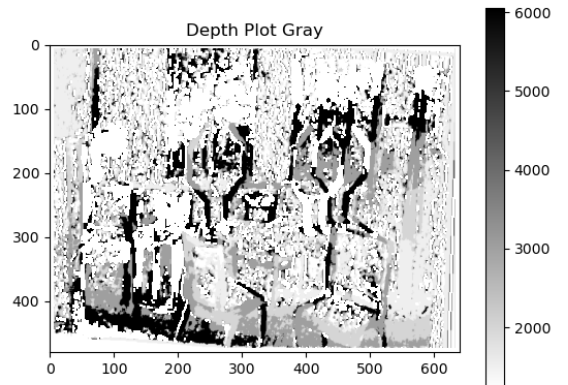

Figure 7: Depth image of curule
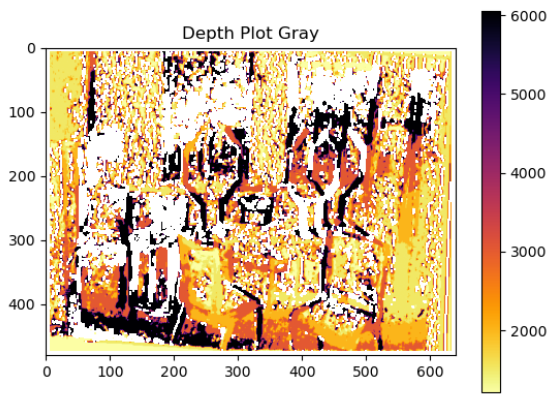

Figure 8: Depth gray image of octagon

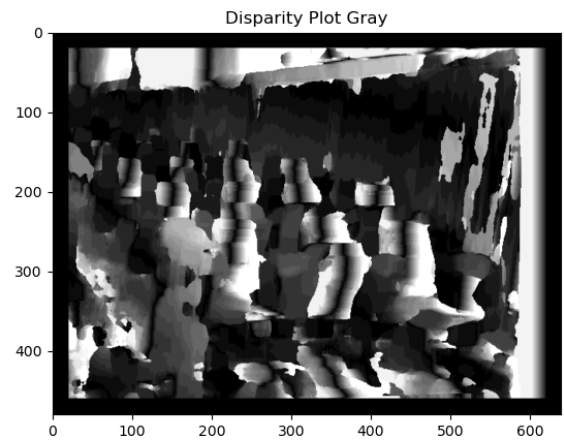Figure 9: Depth color image of octagon

Figure 10: Depth image of pendulum

5   Conclusion

I calculated the dense depth map of images taken from different poses. This project taught me topics about Fundamental and Essential Matrix.