ENPM 673

Perception for Autonomous Robots

Project 3

Instructor:
S. Charifa

Student:
Pulkit Mehta
UID: 117551693

Semester:
Spring 2022

April 18, 2022

Contents

1 Calibration **2**
1.1 Feature Detection and Matching 2
1.2 Estimating Fundamental Matrix 3
1.2.1 Normalizing points 3
1.2.2 Fundamental Matrix 3
1.2.3 RANSAC 3
1.3 Estimating Essential Matrix 4
1.4 Estimating Camera Pose from Essential Matrix 4

2 Rectification **4**

3 Correspondence **5**
3.1 SSD 6
3.2 Block Matching 6
3.3 Disparity 7

4 Computing Depth Information **8**

1 Calibration

1.1 Feature Detection and Matching

The features in both the images of a given dataset were detected using ORB and then Brute Force matching was performed with Lowe's ratio as 0.7.

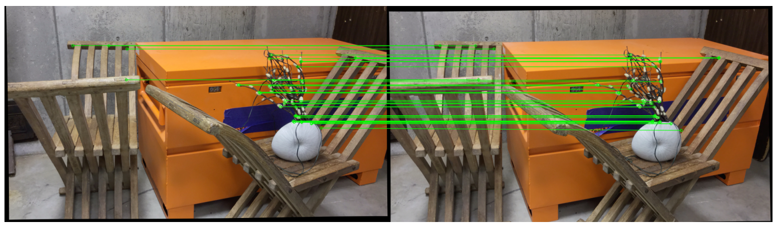

Figure 1: Matched features: curule

Figure 2: Matched features: pendulum

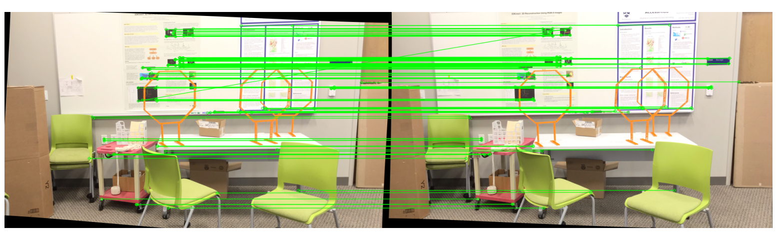

Figure 3: Matched features: octagon

1.2 Estimating Fundamental Matrix

1.2.1 Normalizing points

The points are transformed according to $\hat{x}_i = Tx_i$ and $\hat{x}'_i = T'x'_i$, where T and T' are normalizing transformations consisting of translation and scaling. The points originally have the origin at the top left corner of image so the coordinates in each image are translated so as to bring the centroid of all set of points to the origin. The coordinates are also scaled form homogeneous coordinates.

1.2.2 Fundamental Matrix

The fundamental matrix is defined by the equation:

$$x'_i F x_i = 0 \tag{1}$$

for any pair of corresponding points x_i and x'_i in the two images.
The equation can thus be written as,

$$Af = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_nx_n & x'_ny_n & x'_n & y'_nx_n & y'_ny_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{2}$$

The solution for F is obtained using singular value decomposition(svd) and taking the eigenvector corresponding to the minimum eigenvalue. The rank of F matrix must be 2, thus the rank 2 constraint is applied, i.e., the last eigenvalue found using svd of F is made zero and the F matrix is reconstructed.

Denormalizing is performed using the equation:

$$F = T'^T \hat{F}' T \tag{3}$$

where T' is te transformation of the right image and T of the left image.

1.2.3 RANSAC

A random sample of eight points(the minimum number for solving F) is taken from the feature points and the fundamental matrix is calulated. The error is calculated using the equation 1.2.2. The threshold value for error is taken as 0.01 and the probability is 0.99. Thus, the set of pair of inlier points from both the images are found along with the best estimate of the fundamental matrix.

1.3 Estimating Essential Matrix

The Essential matrix can be calculated using the equation

$$E = K^T F K \tag{4}$$

The eigenvalues are set as 1,1 and 0 after performing svd and the E matrix is reconstructed.

1.4 Estimating Camera Pose from Essential Matrix

The Essential matrix is decomposed using svd as:

$$E = UDV^T \tag{5}$$

Now, the rotational matrices R1, and R2 are obtained as:

$$R1 = UWV^T; R2 = UW^T V^T \tag{6}$$

where

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

and the translational vector t is the third column of matrix U. The four camera pose configurations are: (R1, t), (R1, -t), (R2, t) and (R2, -t). If the determinant of the rotational matrix is negative then both R and t are multiplied by a factor of -1.

In order to find the correct camera pose from the above four poses, we need to remove the disambiguity by checking the cheirality condition, i.e., the reconstructed points must be in front of the cameras. Linear triangulation is performed using svd to check the sign of depth Z in camera coordinate system. The best camera configuration is the one that produces maximum number of points satisfying the cheirality condition.

2 Rectification

The estimated fundamental matrix is used to check the epipolar lines in the image. The homography matrices H1 and H2 for the left and right images respectively are computed using the inlier points obtained above. For this purpose the inbuilt function $cv2.stereoRectifyUncalibrated$ is used.

I had problem performing rectification on the octagon dataset, but the other two datasets were fine.

Figure 4: rectified images: curule

Figure 5: rectified images: pendulum

3 Correspondence

After rectification is performed on both the images, the corresponding points along the epipolar lines are matched using block matching

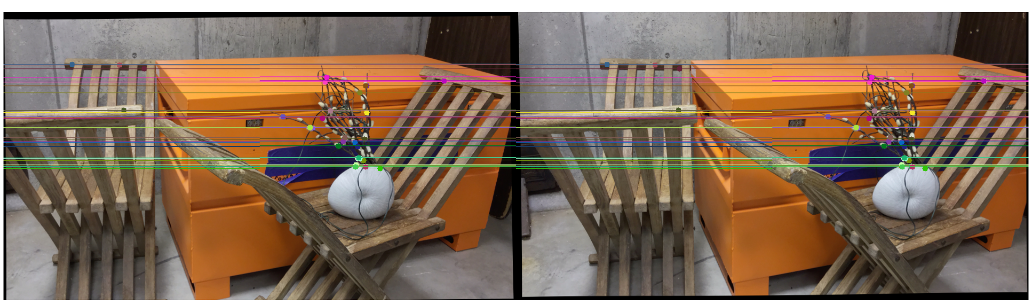

Figure 6: epipolar lines: curule

This may be due to the ORB feature descriptor with Brute force matching as I wasn't able to obtain properly warped images and epipolar lines even with the opencv inbuilt fundamental matrix calculator method.

Figure 7: epipolar lines: pendulum

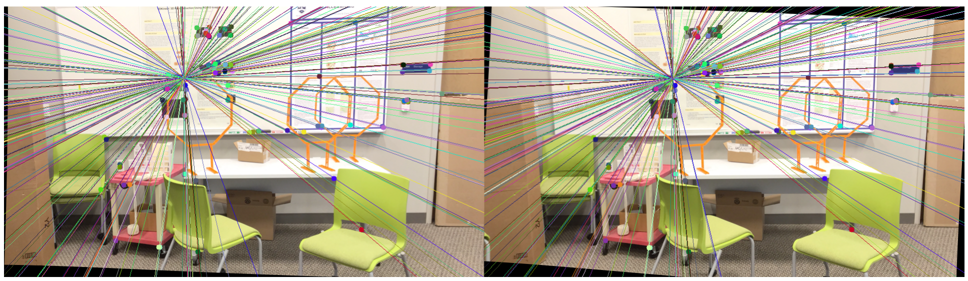

Figure 8: epipolar lines: octagon

3.1 SSD

The sum of squared differences(ssd) is computed as:

$$ssd = \Sigma_{[i,j]\epsilon R}(f(i,j) - g(i,j))^2 \tag{8}$$

3.2 Block Matching

A search range is defined along with a window size for block matching along the epipolar lines. The block with the minimum value of ssd is considered to be the matching block and the x value in right image corresponding to the matching x-coordinate in the left image is found.

3.3 Disparity

The disparity d is calculated by getting the absolute difference of the x-values of the matched coordinates and thus the disparity map is constructed by performing block matching along each pixel of the left and right image.

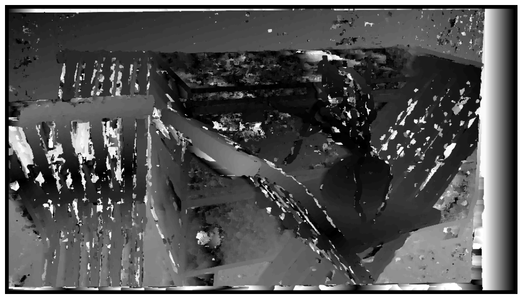

Figure 9: Disparity map grayscale: curule

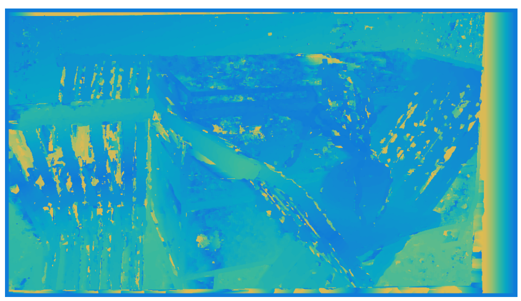

Figure 10: Disparity map colored: curule

Figure 11: Disparity map grayscale: pendulum

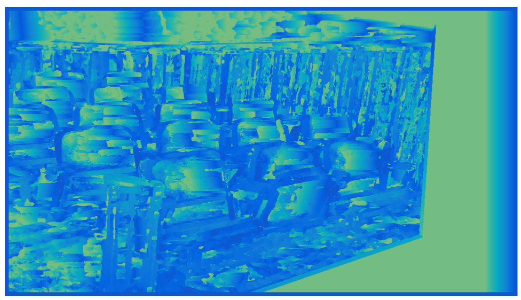

Figure 12: Disparity map colored: pendulum

4 Computing Depth Information

The depth, z for a point can be calculated from the disparity map by using the below equation:

$$z = \frac{b * f}{d} \tag{9}$$

where, d is the disparity for the point, b is the baseline distance between the camera centers and f is the focal length of the camera.

Figure 13: Depth map grayscale: curule

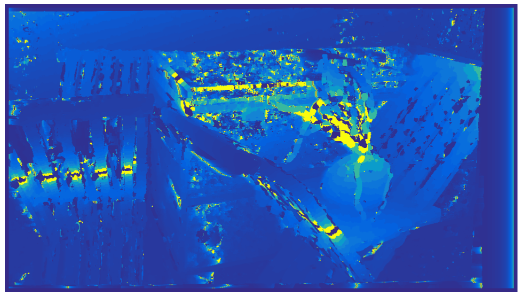

Figure 14: Depth map colored: curule

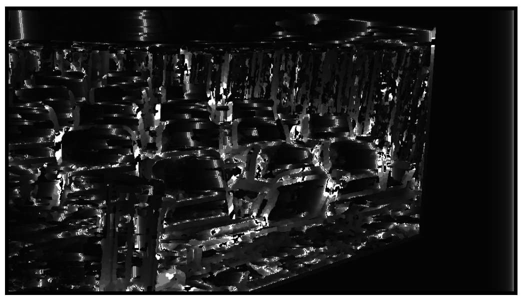

Figure 15: Depth map grayscale: pendulum

Outputs can be seen through the link below: `https://drive.google.com/drive/folders/1ntggIKRFNjfYSpgayttl08k-8D5Kt627?usp=sharing`

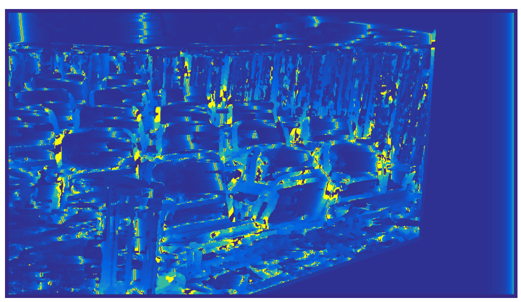

Figure 16: Depth map colored: pendulum