ENPM673 - Perception for Autonomous Robots
Project 3 - Stereo Vision System

Saurabh Palande
Masters in Robotics
University of Maryland, College Park
UID: 118133959
Email: spalande@umd.edu

***Abstract*—In this project, we implement the concept of Stereo Vision. We are given 3 different datasets, each of them contains 2 images of the same scenario but taken from two different camera angles. By comparing the information about a scene from 2 vantage points, we need to obtain the 3D information by examining the relative positions of objects.**

I. Stereo Vision System

A. Calibration

Steps to perform the Calibration step:

1) Find features
   First, we need to compare the two images in each dataset and select a set of matching features.

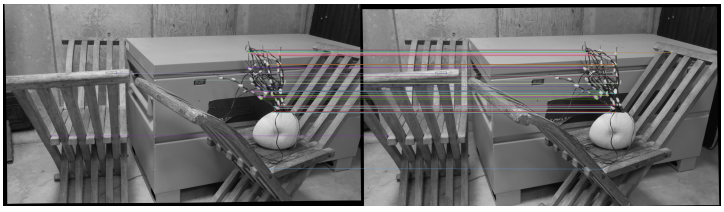

Fig. 1: Curule feature matching

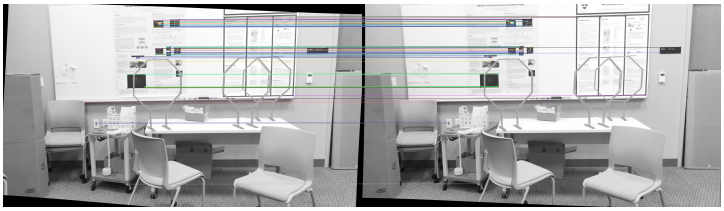

Fig. 2: Octagon feature matching

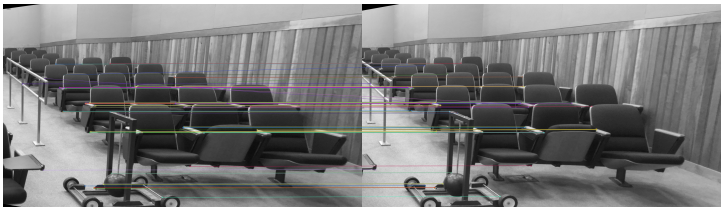

Fig. 3: Pendulum feature matching

2) Estimate Fundamental matrix
   After selecting the feature points, we need to compute the fundamental matrix. The F matrix is only an algebraic representation of epipolar geometry and can be found by both geometrically (contructing the epipolar line) and arithematically. As a result we obtain:
   ${x_i}^T F x_i = 0$
   where i=1,2,....,m. This is known as epipolar constraint or correspondance condition (or Longuet-Higgins equation). Since, F is a 3×3 matrix, we can set up a homogenrous linear system with 9 unknowns: In F matrix estimation, each point only contributes one constraints as the epipolar constraint is a scalar equation. Thus, we require at least 8 points to solve the above homogenous system. That is why it is known as Eight-point algorithm. The value of $Ax = 0$ can be found vy using SVD. When applying SVD to matrix A, the decomposition USV^T would be obtained with U and V orthonormal matrices and a diagonal matrix S that contains the singular values. The singular values σ_i are positive and are in decreasing order with $\sigma_9 = 0$ since we have 8 equations for 9 unknowns. However, due to noise in the correspondences, the estimated F matrix can be of rank 3. So, to enfore the rank 2 constraint, the last singular value of the estimated F must be set to zero. If F has a full rank then it will have an empty null-space i.e. it won't have any point that is on entire set of lines. Thus, there wouldn't be any epipoles.
   To find the best estimate of Fundamental Matrix we use RANSAC. Out of all the feature correspondences, we select 8 random points and find the fundamental matrix with the maximum number of inliers.
3) Estimate Essential Matrix
   Essential matrix is another 3×3 matrix, but with some additional properties that relates the corresponding points assuming that the cameras obeys the pinhole model (unlike F). More specifically, $E = K^T F K$ where K is the camera calibration matrix or camera intrinsic matrix. Clearly, the essential matrix can be extracted from F and K. As in the case of F matrix computation, the singular values of E are not necessarily (1,1,0) due to the noise in K. This can be corrected by reconstructing it with (1,1,0) singular values.
4) Estimate Camera pose

The camera pose consists of 6 degrees-of-freedom (DOF) Rotation (Roll, Pitch, Yaw) and Translation (X, Y, Z) of the camera with respect to the world. Since the E matrix is identified, the four camera pose configurations: (C1,R1),(C2,R2),(C3,R3) and (C4,R4) where C is the camera center and R is the rotation matrix, can be computed.

$E = UDV^T$ and W = $\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

C1 = U(:,3) and $R1 = UWV_T$
C2 = -U(:,3)and $R2 = UWV_T$
C3 = U(:,3) and $R3 = UW^TV^T$
C4 = -U(:,3)and $R4 = UW^TV^T$

If det(R)= -1, the camera pose must be corrected i.e. C = -C and R = -R.

5) Calculating correct pose using Triangulation and cheirality check. Given two camera poses, (C1,R1) and (C2,R2), and correspondences, x1 and x2 we triangulate 3D points using linear least squares. To check the cheirality condition, triangulate the 3D points (given two camera poses) using linear least squares to check the sign of the depth Z in the camera coordinate system w.r.t. camera center. A 3D point X is in front of the camera if: r3(X-C)¿0 where r3 is the third row of the rotation matrix (z-axis of the camera). Not all triangulated points satisfy this coniditon due of the presence of correspondence noise. The best camera configuration, (C,R,X) is the one that produces the maximum number of points satisfying the cheirality condition.

B. Rectification

We apply perspective transformation to make sure that the epipolar lines are horizontal for both the images. For this part, we use cv2.stereoRectifyUncalibrated which gives H1 and H2. After finding the homography matrices, the images are warped respectively and then the epilines are plotted on the images.

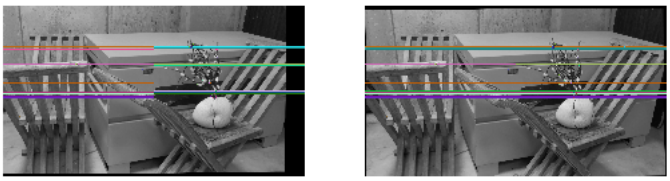

Fig. 4: Curule Epilines

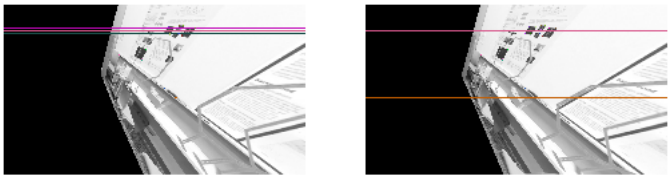

Fig. 5: Octagon Epilines

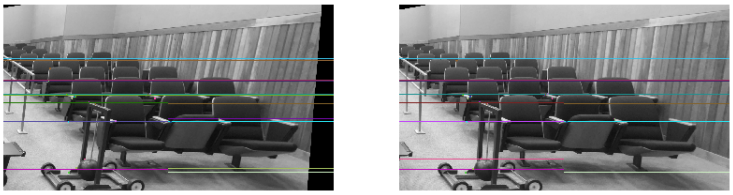

Fig. 6: Pendulum Epilines

Estimate Fundamental matrix

C. Correspondence - Calculate Disparity map

To calculate the disparity, we apply the matching windows concept by using the following steps:

1. For a pixel in the left image, select the pixels in its neighborhood specified as block size from the left image.

2. Compute SSD by comparing each block from the left image (same size as block size) and each block selected from the search block in the right image. Slide block on the right image by one pixel within the search block. Record all the SSD scores.

3. Find the highest pixel SSD score from the previous step. For the block with highest SSD, return the pixel location at the center of the block as the best matching pixel.

4. If xl is the column index of the left pixel, and the highest SSD score was obtained for a block on the right image whose center pixel has column index xr, we will note the disparity value of $|xl - xr|$ for the location of left image pixel.

5. Repeat the matching process for each pixel in the left image and note all the disparity values for the left image pixel index.

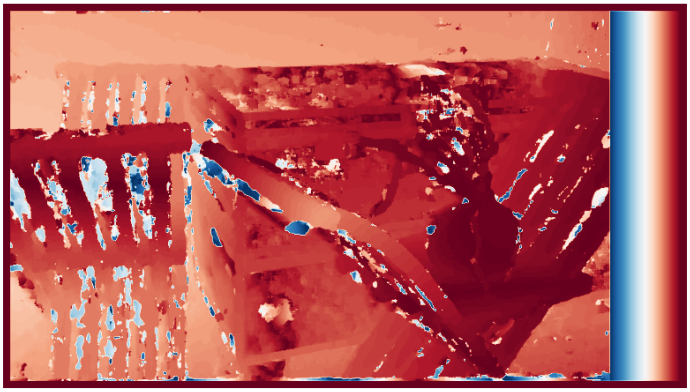

Fig. 7: Curule Disparity Map

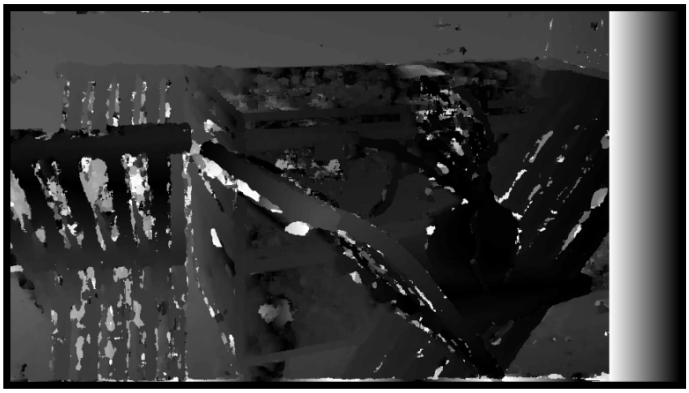

Fig. 8: Curule Disparity Map - Grayscale

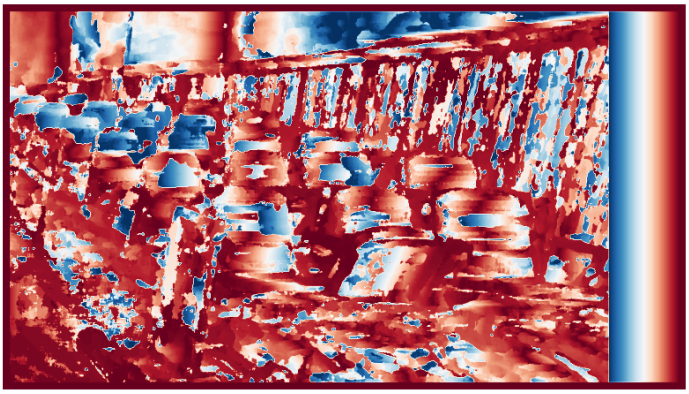

Fig. 11: Pendulum Disparity Map

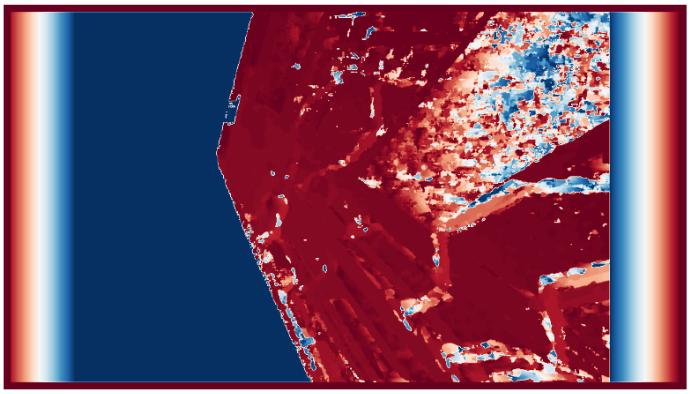

Fig. 9: Octagon Disparity Map

Fig. 12: Pendulum Disparity Map - Grayscale

D. Depth Map

Using the disparity information obtained above, we compute the depth information for each image pixel by using the below equation
distance = focal length $*$ baseline distance / disparity

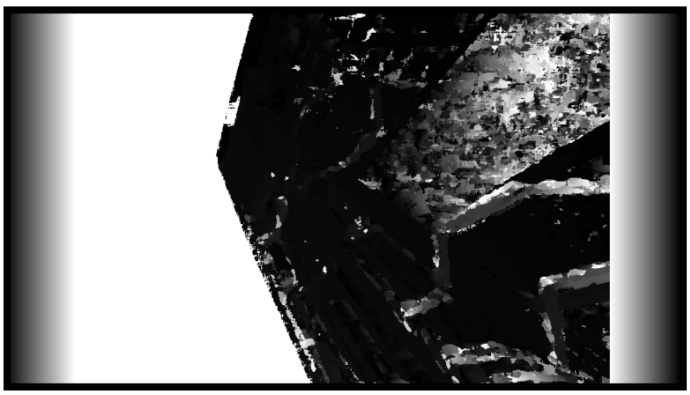

Fig. 10: Octagon Disparity Map - Grayscale

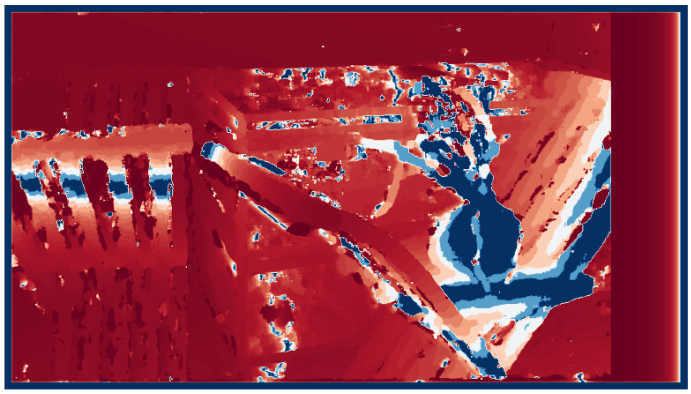

Fig. 13: Curule Depth Map

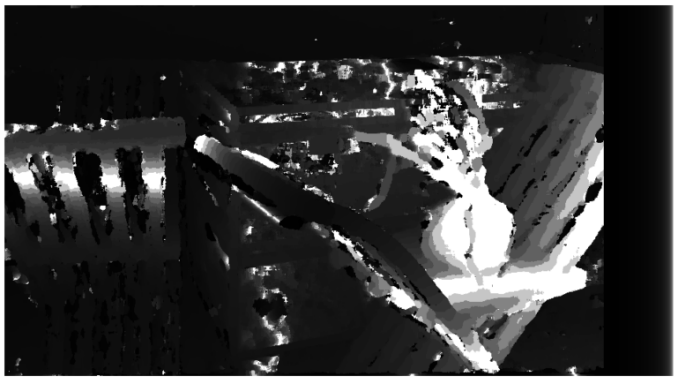

Fig. 14: Curule Depth Map - Grayscale

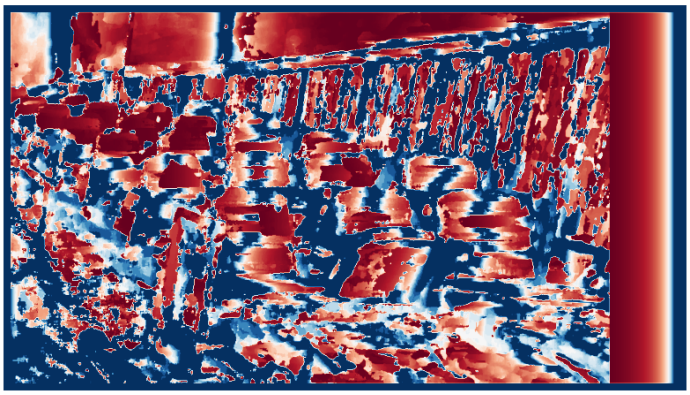

Fig. 17: Pendulum Depth Map

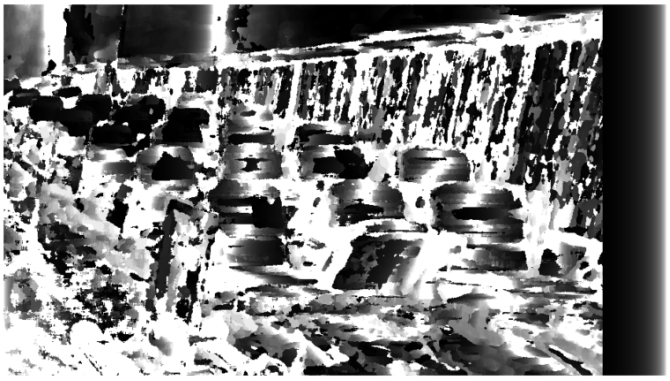

Fig. 18: Pendulum Depth Map - Grayscale

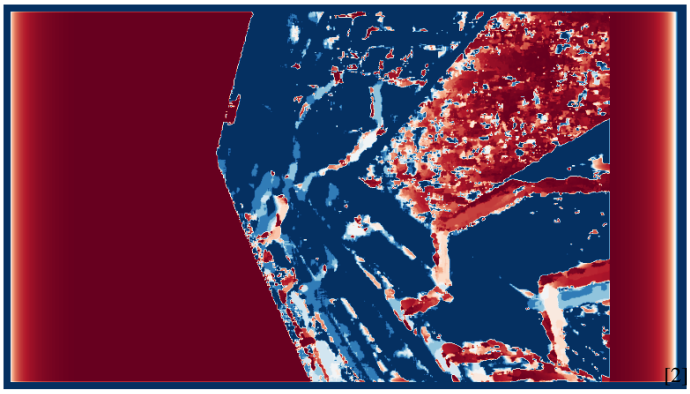

Fig. 15: Octagon Depth Map

Fig. 16: Octagon Depth Map - Grayscale

REFERENCES

[1] https://www.cs.cmu.edu/ 16385/s17/Slides/13.2$_{S}tereo_{M}atching.pdf$

https://pramodatre.github.io/2020/05/17/stereo-vision-exploration/