Project 3: Creation of a Stereo Vision System

Perception for Autonomous Robots
ENPM673 – Spring 2022

By: Yoseph Ayele Kebede

To: Professor Samer Charifa

Creation of a Stereo Vision System

Images are visual representations of the 3D world people live in. They, however, do not provide full detail of objects as they appear, i.e. when converting 3D instances into 2D information about depth is lost, as items placed in the same line of sight are depicted as one point in the image frame. Therefore, to solve this problem, stereo vision is used which utilizes two frames that are at different angle, pitched towards object of interest, to distinctly locate the objects that may have been hidden had been seen from only one camera position. Thus, through triangulation method depth can be calculated while via image rectification and correspondence techniques, depth map can be drawn to depict depth information thereby illustrating how the human brain perceives depth based on visual stimuli observed from eyes.

First, two images from a given dataset or images taken form a camera at an instance are loaded or saved. Then, the cameras were calibrated or their calibration information, mainly the focal point, principal points, baseline, and disparity, are deduced. Then, common feature matching libraries were implemented to detect areas from the image pairs that are identical, where Flann Based Matcher was used here shown in Figure 1. Once matching set of points are obtained, they were further filtered as outlier and inlier via utilization of RANSAC as well as Fundamental Matrix. The process involved obtaining 8 random pairs of points from both images, after which the Fundamental matrix was computed via the 8 point algorithm. Steps followed are as follows:

- Randomly pick 8 pts from point set 1 and 2
- Compute Fundamental matrix for 8 points
- For every matching point in the two sets find:
- Calculate |X2.T * F * X1|, where 2 is right and 1 is left image
- If above result is within threshold (in this case < 0.1), keep matching pairs
- Repeat process for up to selected number of iterations (M=1000 used here)

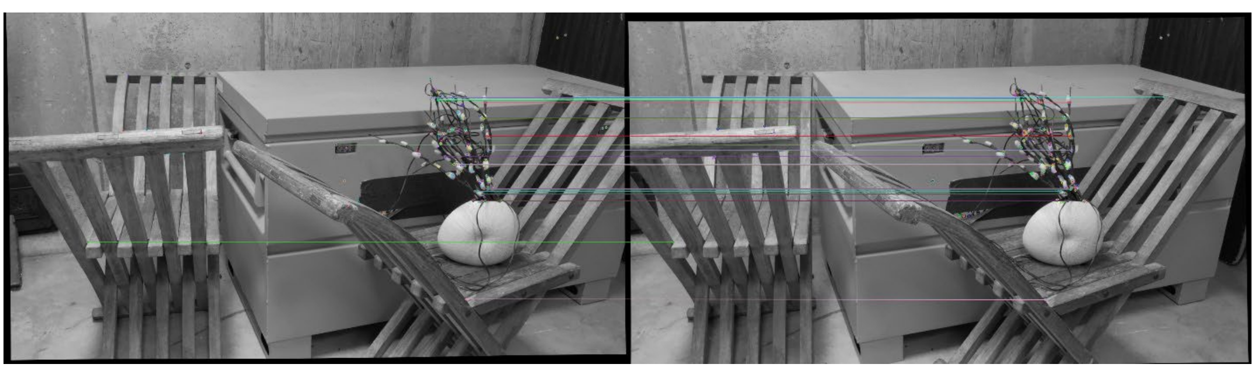


Figure 1. Feature Matching Between Images

Next, the essential matrix is calculated from the fundamental matrix and intrinsic matrix, which is followed by the computation of epipoles. These variables are key to finding the translational and rotational differences between the two images. That is, while the essential matrices are decomposed into the translation and rotational matrices, the epipole and fundamental points are used to compute the homography of each image. This homography pair were used to rectify both images to parallel vertical planes whose epilines point to infinity on the horizontal axis. Figure 2 shows the epilines before rectification, which here seem indifferent from the rectified images as the change in translation is minimal.

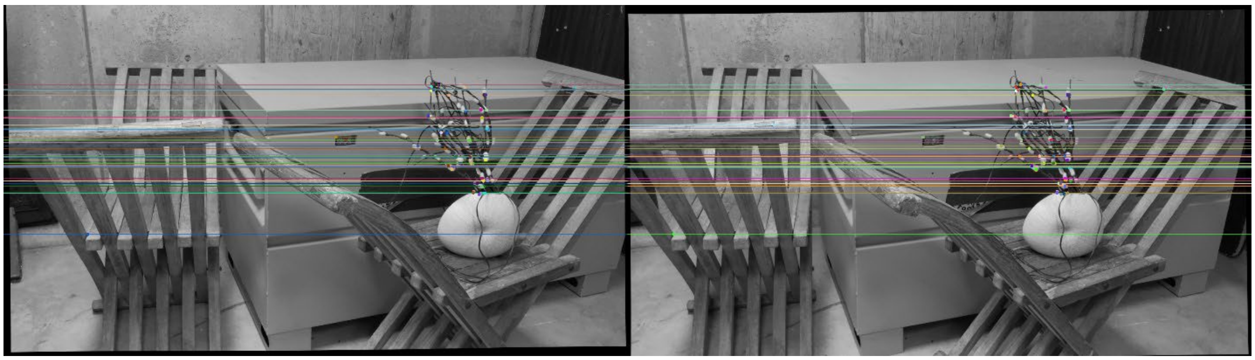


Figure 2. Epilines Drawn On Images Before Rectification from Built In Function F-matrix

When drawing the epipolar lines from Fundamental Matrix (F) computed by hand, there was an issue retrieving the matrix that yielded the epipolar lines from Figure 2. As a result, while the above figure was obtained from F computed from built in functions, Figure 3 below was yielded when drawing the lines from computed F.

Figure 3. Epilines Drawn On Images from computed F-matrix

Then, the two homography matrices are applied to warp both images so that each matching points are on the same epipolar line. Due to challenges with computing matching fundamental and homography matrices, accurate rectified or warped images were not obtained, as resulting images showed only the matching points without displaying the image as shown in Figure 4 below which display the expected images used with built in Fundamental matrix computing functions, while the actual result from hand computed F is depicted in Figure 5.

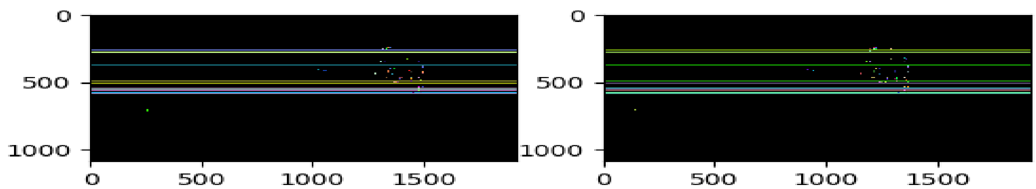


Figure 4. Warped Image Sets with F from Built In Function

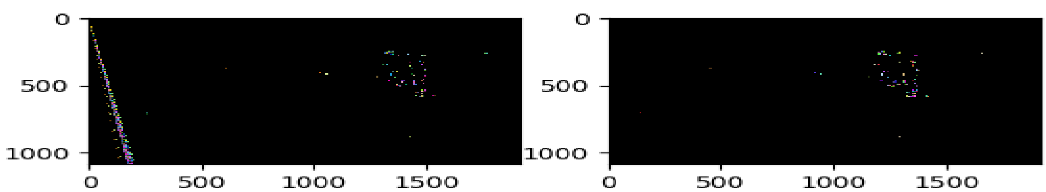


Figure 5. Warped Image Sets with F computed from 8-pt Algorithm

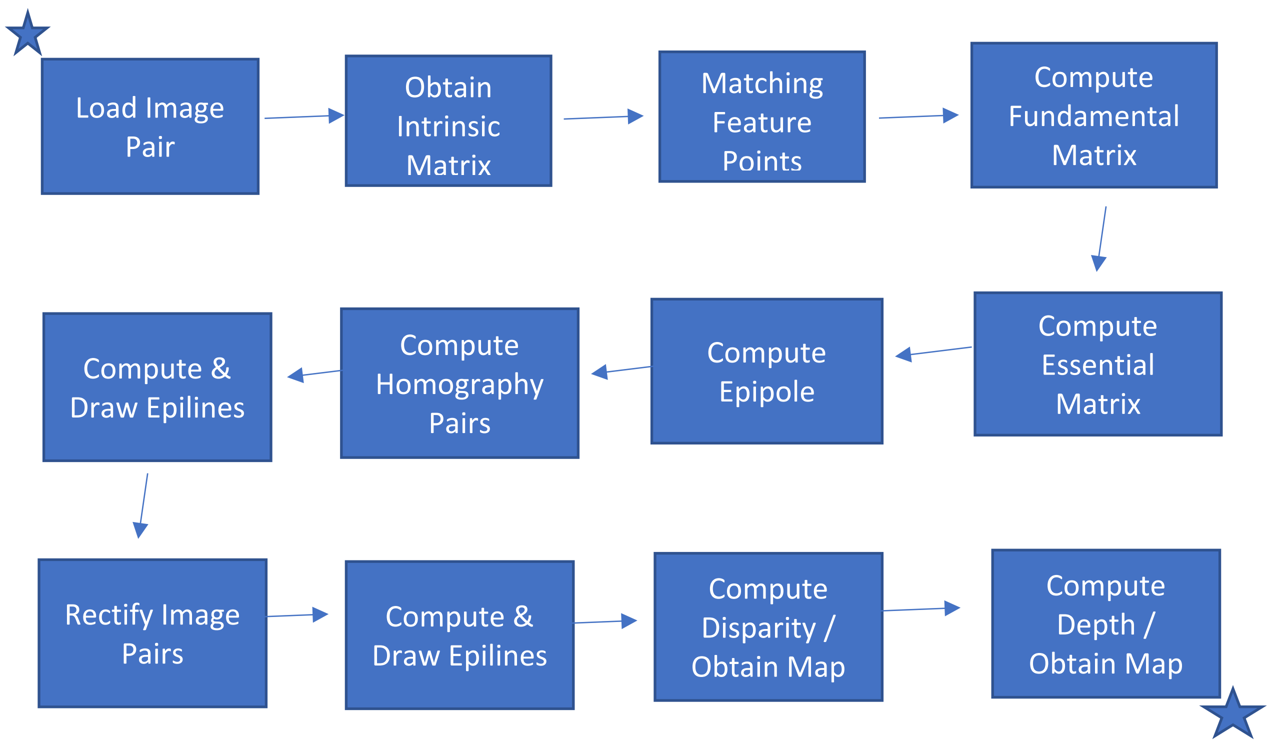
Now that the images are rectified and aligned based on their epipolar lines, assuming the correct warped images were obtained, disparity between the two images are computed. Since epipolar lines are aligned, the pixels on the left image can be compared with pixels on the right. As a result, using the matching windows concept, a window that contains a defined set of pixels, was sled across the horizontal axis along the same height on the right image. Thus, each window for a given height value is iterated through every window on the right side to find the window that mostly matches it. This is performed via Square of the Sum of Differences between the two windows. Thus, after computing the mean of the windows, the square of the differences was computed and stored in a list. Finally, the minimum value of the list will be extrapolated to yield the pixel coordinates from the right image which is declared as the window that matches the left window performing the search.

Moreover, now that the two matching pixel locations are obtained, the horizontal displacement between these windows is stored as disparity in a list. Later, each value in this list is scaled to the 0-255 intensity, which is stored in an image matrix. Draw disparity map in grayscale and heat map. Process summed as:

- Perform cross-correspondence and match window pixels
- Find horizonal displacement between matching pairs
- Scale disparity to intensity
- Draw grayscale and color heatmap of disparity image.

Finally, the depth map is computed from the disparity matrix. Depth for each pixel is computed by multiplying the focal point for each image with the baseline finally divided by the disparity for each pixel (Z = f * Baseline / disparity). The depth map is then created where the high depth values are assigned darker colors (lower intensities), while the lower depth values are illustrated with whiter colors (higher intensities).

In sum, the pipeline executed in this stereo vision system is:



Challenges: The two main challenges have been obtaining the fundamental matrix via the RANSAC method, which after numerous iterations did not yield the expected epilines. Secondly, for an unknown reason, the cv2.warpPerspectige function did not create a warped image of the original image pairs; but instead created a blank black image with only the matching points showing as colors as shown above. As a result, obtaining the disparity, disparity map, depth, and depth map have not been obtained as expected. The project code has, however, shown each module piece by piece one step at a time as shown in the pipeline above demonstrating the full attempt of yielding the desired results.