CSci 5561 Assignment 5

Luis Guzman

Friday December 11, 2020

In this assignment, I implemented the stereo reconstruction algorithm for determining the depth of a scene using two images. My implantation works for the general case of two images of a scene, even when the camera orientations and positions are not known. I begin by using OpenCV for matching SIFT features between the two images. I used Lowe's ratio test to ensure a good match between features. The feature matching is shown in figure 1.

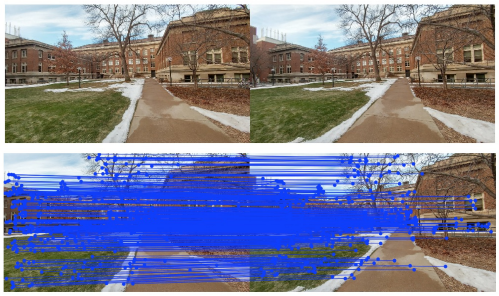

Figure 1: The original images and their matching SIFT features

Next I use the 8-point algorithm with RANSAC for estimating the fundamental matrix. SVC cleanup was performed to ensure that the fundamental matrix was rank-2. In figure 2 I plotted the epipolar lines to check the validity of my fundamental matrix.

Figure 2: The epipolar lines from the fundamental matrix

The fundamental matrix leads to four possible solutions for camera positions. For each of these solutions, I performed triangulation to determine the 3D location of each pixel using the equation

$$\begin{bmatrix} \begin{bmatrix} \mathbf{u} \\ 1 \end{bmatrix}_\times \mathbf{P}_1 \\ \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}_\times \mathbf{P}_2 \end{bmatrix} \begin{bmatrix} \mathbf{pts3D} \\ 1 \end{bmatrix} = \mathbf{0}$$

where the u and v vectors are represented by the 3x3 skew-symmetric matrix and P1, P2 are two 3x4 camera projection matrices. The product of the u skew-symmetric matrix with P1 leads to another 3x4

matrix, but the last row of this is linearly dependent on the first two. This row can be thrown away without any loss of information, so the result is a 2x4 matrix. Following the same procedure for v and stacking them, we get a 4x4 matrix. pts3D is then equal to the null space of this 4x4 matrix. I ended up having to approximate the null space by setting all singular values less than 0.1 of the largest singular value to zero. This worked out well for this application since the largest two singular values were on the order of 100 and the 3rd singular value was on the order of 10.

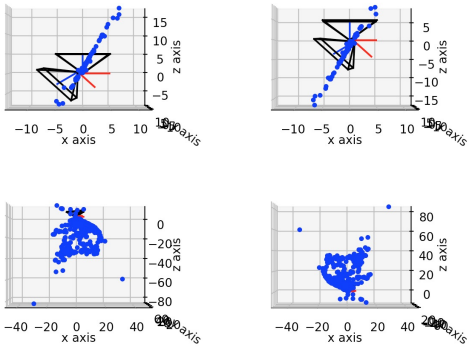

Figure 3: The four solutions and resulting 3D points from triangulation

Next I performed pose disambiguation by picking the solution that lead to the most points being in front of the camera. The four camera solutions and resulting point are shown in figure 3, and in this case the lower-right solution would be chosen since it has very few points behind the camera. This is known as the Cheirality condition:

$$r_3^T(X \quad C) > 0$$

where r_3^T is the 3rd row (z vector) of a 3x3 camera matrix and C is the camera center. The necessary homography transformation can then be calculated to warp each image to the same perspective, as shown in figure 4.

Figure 4: The images warped to the same perspective

Lastly, I performed dense SIFT feature matching on the warped images to determine the horizontal disparity between each matching feature in the images. Since horizontal disparity is inversely proportional to distance, I can use this disparity image to compute a depth map. I unfortunately was not able to get my depth map working correctly given my time constraints, but I've included a few of the attempts below.

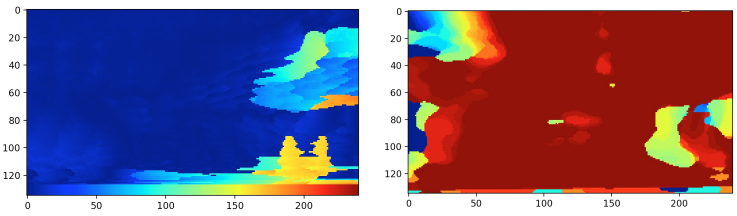

Figure 5: The final depth maps I created. Still needs some debugging