EECE 5639, Extra Credit Project:
Stereo

Sreejith Sreekumar

December 11, 2018

Abstract

In traditional stereo vision, two cameras, displaced horizontally from one another are used to obtain two differing views on a scene. By comparing these two images, the relative depth information can be obtained in the form of a disparity map, which encodes the difference in horizontal coordinates of corresponding image points. The values in this disparity map are inversely proportional to the scene depth at the corresponding pixel location. Building on these ideas we try to estimate the depth of a scene.

1 Description of Algorithms

In this section we list and describe all of the algorithms implemented in the completion of this project.

1.1 Corner Detection

Harris Corner Detector is a corner detection operator that is commonly used in computer vision algorithms to extract corners and infer features of an image. This algorithm takes the differential of the corner score into account with reference to direction directly. The main equations in corner detection are given by:

- Computation of M from gradient components

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- Intensity change in shifting window eigenvalue analysis:

$$E(u,v) = \begin{bmatrix} u & v \end{bmatrix}$$

- Corner Response Measure

$$R = detM - k\ trace^2(M)$$

- Threshold R to determine the corners.

1.2 Non-Max Suppression

The output of harris corner detection could be thought of as a heat map, with a region of high magnitudes around a single corner. In order to produce corners from this map, we define the local maxima within a region as a corner, if it is above a certain threshold. To find these local maxima we slide a 3x3 window over the image. Every pixel in the window that is less then than maximum value in the window is set to zero.

1.3 Normalized Cross Correlation

We use the corners as features to find the interesting points in each image and determine pair-wise point correspondences. One of the methods to compare these interesting regions is Normalized Cross Correlation.

Normalized Cross Correlation is defined by the equation:

$$N_{fg} = \sum_{x,y} \hat{f}(i,j)\hat{g}(i,j)$$

where

$$\hat{f} = \frac{f}{||f||} = \frac{f}{\sqrt{\Sigma_{[i,j]\in R} f^2(i,j)}}$$

$$\hat{g} = \frac{g}{||g||} = \frac{g}{\sqrt{\Sigma_{[i,j]\in R} g^2(i,j)}}$$

In practice, we perform a brute force search between all possible combinations of points from both images. We then sort the results of this search by magnitude of correlation value. The N correspondences with the highest normalized correlation values are maintained. The resulting correspondences are then further pruned by realistic geometric constraints. In our case, we know the images represent a small lateral rotation. Correspondences with vertical translations greater than a small threshold, or lateral translations greater than a large threshold are said to not fit this constraint, and are deleted. The result is approximately 60 robust correspondences between the two images.

1.4 Fundemental Matrix

The fundamental matrix is a 3 x 3 matrix which relates corresponding points in stereo images. In epipolar geometry with homogeneous image coordinates, x and x' of corresponding points in a stereo image pair, F * x describes a line on which the corresponding point x' must line (epipolar line).
The equation relating x, x' and F can be written as

$$x'^T F x = 0$$

1.5 Disparity Map

2 Experiments

This section describes the intermediate steps and experiments performed in the completion of this project.

2.1 Reading the image inputs

2.2 Detecting Corners in the Images

For both images, we apply the harris corner detection algorithm and then perform non max supression and thresholding to get the co-ordinates of the corners from the images.

Iterating through the corners detected for each image, we create cropped neighborhoods for every corner with it at the center. This way, we'd have N template patches corresponding to the N points in the image which were detected to be corners.

Following this procedure, we do *normalized cross correlation* (subsection 1.3) on every template from the first image with every other template of the second image, constituting a brute force search of all point neighborhood correlation combinations between the images. The results of this search are sorted by cross correlation magnitude and all matches above a tuned correlation threshold are kept.

The resulting matches are then pruned by realistic geometric constraints. In our case, we know the images are both taken from the same viewpoint from a slightly different angle. Thus we eliminate correspondences with vertical translations above a small threshold and lateral translations above a large threshold, to produce a set of robust and realistic correspondences.

The correspondences which were found hence are visualized below using lines of different colors.

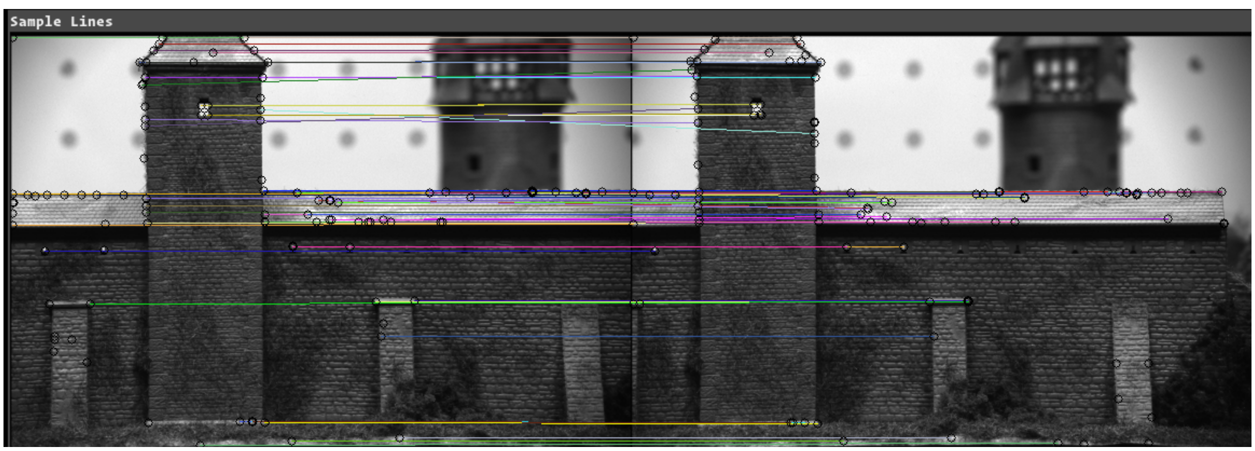

Figure 1: Color Coded Corner correspondences

3 Conclusion

4 Appendix