EECE 5639, Extra Credit Project:
Stereo

Sreejith Sreekumar

December 12, 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/ Concepts Involved

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 Stereo Vision and Epipolar Geometry

Stereo vision involves the vision dynamics involved when two cameras take pictures of the same scene, but they are separated by a distance. The shifted amount is called the disparity. The disparity at which objects in the image best match is used by the computer to calculate their depth information.

Epipolar geometry is the geometry describing stero vision. The essential matrix is a 3 x 3 marix that encodes the epipolar geometry of the two views. Given a point in an image, multiplying by the Essential Matrix, will tell us the epipolar line in the second image where the corresponding point must be.

1.5 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

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

1.6 Disparity and Depth

Disparity refers to the difference in image location of an object seen by the left and right camera images. The disparity of features between two stereo images are usually computed as a shift to the left of an image feature when viewed in the right image. Stereo images may not always be correctly aligned to allow for quick disparity calculation. For example, the set of cameras may be slightly rotated off level. Through a process known as image rectification, both images are rotated to allow for disparities in only the horizontal direction.
Disparity information can be useful in depth/distance calculation. Disparity and distance from the cameras are inversely related. As the distance from the cameras increases, the disparity decreases. This allows for depth perception in stereo images.

2 Experiments

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

2.1 Reading the image inputs

The input for the depth map computation are two images of a toy castle. The images are read and converted into grayscale for processing. For detecting the corner features, we add a padding of 20 pixels around the image around its borders.

2.2 Detecting Corners in the Images

Harris corner detection algorithm is applied on the images followed by a 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 correspondences which were found hence are visualized below using lines of different colors.

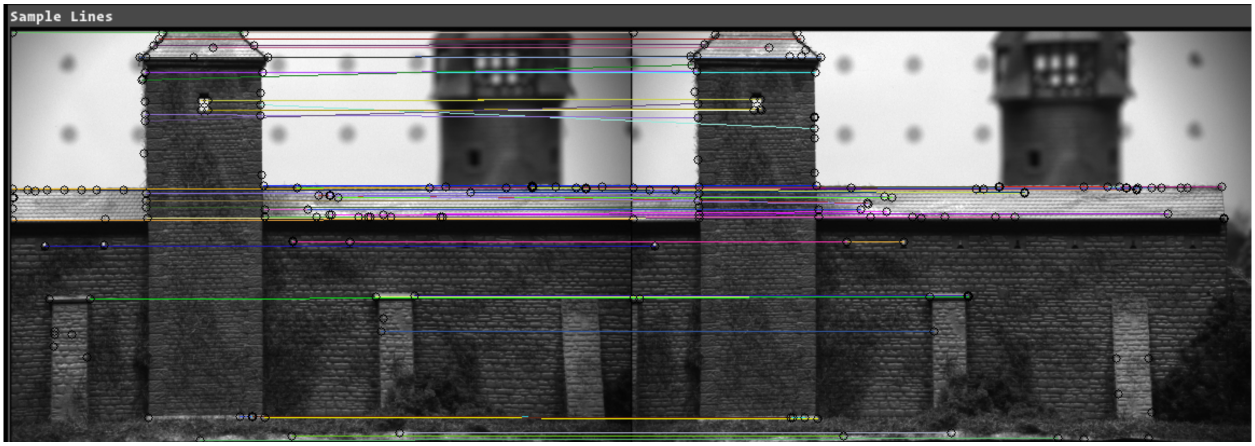

Figure 1: Color Coded Corner Correspondences

2.3 Computation of Fundemental Matrix - 8 point algorithm

The fundemental matrix is constructed from the mapped points using the 8 point algorithm. The corresponding coordinates of every mapping is taken and we construct a matrix, A as:

$$A = \begin{bmatrix} x_1x_1' & x_1y_1' & x_1 & y_1x_1' & y_1y_1' & y_1 & x_1' & y_1' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx_m' & x_my_m' & x_m & y_mx_m' & y_my_m' & y_m & x_m' & y_m' & 1 \end{bmatrix}$$

Now we find the SVD of A, written by:

$$A = UDV^T$$

The columns of V are the eigenvectors of A^TA; the last one corresponds to the smallest eigenvalue.

2.4 Computing the disparity maps

3 Conclusion

4 Appendix

4.1 Code