Coursework on Image Matching & Geometry

Ashish Pandey
ashish.pandey17@imperial.ac.uk
CID-01383450; Login-ap8516

Ilias Chrysovergis
ilias.chrysovergis17@imperial.ac.uk
CID-01449042; Login-ic517

1. Image Matching

This section of our report makes use of Tsukuba images [1] for the demonstration of the implemented methods.

1.1 Manual Matching

A manual method for getting coordinates of corresponding interest points in two images using the `ginput` MATLAB built-in function has been implemented. A figure for five-point matching of two Tsukuba images using this method is presented in Appendix I.

1.2 Automatic Matching

Automatic Matching requires detection, description and finally matching of the features of two images as shown in figure 1. Therefore, an interest point (or corner) detector has been developed firstly, using the Harris-Stephens algorithm, which returns the most significant corner points of each image. Secondly, a descriptor (gradient orientation histogram) has been implemented, which characterizes each corner point of an image with a histogram of oriented gradients. Finally, a method which matches the histograms of two different pictures has been developed, using the nearest neighbor algorithm. The Euclidean distance metric was used, while 1-NN to 2-NN ratios & distance thresholds where introduced to eliminate errors.

Feature Detection
Feature Description
Feature Matching

Figure 1: The Automatic Matching Procedure.

Harris-Stephens algorithm is presented in page 48 of the lecture notes on Feature Detection [2]. The Gaussian filter is a 5X5 Gaussian matrix and the parameter of the cornerness function is $\alpha = 0.05$. The non-maximum suppression step uses a 3X3 window to indicate local maxima. The application of the Harris Detector algorithm to a Tsukuba image provides Figure 2.

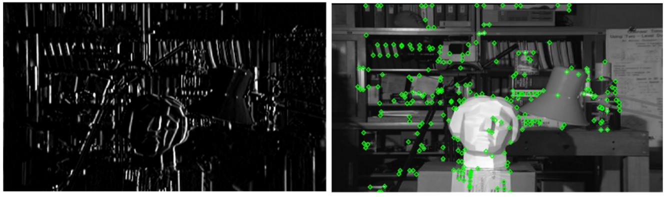

Figure 2: The Gradient Image and Harris interest points/corners.

The implemented feature description technique counts occurrences of gradient orientation in localized portions around the points/corners detected [3]. As in the Harris algorithm, the gradient images are calculated firstly, and then the 9-bins histograms of gradients of 8X8 cells are computed [4]. Finally, the 16X16 block normalization occurs and the final feature vector can be calculated. The application of the descriptor to the previous Tsukuba image produces the following figure:

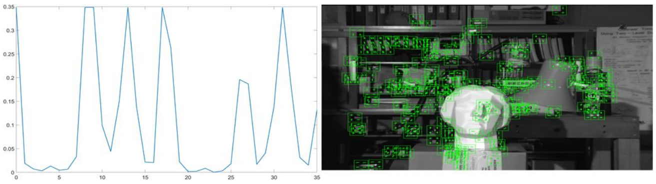

Figure 3: The oriented gradients of interest points & a histogram.

Applying the nearest neighbor matching algorithm, one gets the following figure for two images taken from the Tsukuba dataset.

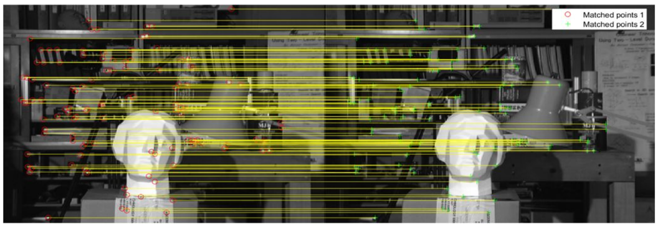

Figure 4: Nearest neighbor matching of two Tsukuba images.

1.3 Transformation Estimation

For transformation estimation two different matrices have been calculated. Firstly, the homography matrix relates the transformation between two planes, using 8 Degrees of Freedom (8-DoF) [5]. Secondly, the fundamental matrix (7-DoF) shows the relationship between any two images of the same scene, which constrains where the projection of points from the scene can occur in both images [6]. The algorithms are presented in the Image Matching lecture notes, while some references have been used too [7] [8]. The estimation of those two matrices requires at least 8 points (4 correspondences) to provide with a correct solution. Although increase of the corresponding points minimizes the error, outliers (incorrect matches) should be taken into consideration too. Therefore, the RANSAC algorithm (from Pattern Recognition course) was developed to keep most of the inliers, while dismissing the outlier points.

The homography and fundamental matrix using the set of corresponding point coordinates of figure 4 (output of the nearest neighbor matching algorithm) are the following:

$$\boldsymbol{H} = \begin{bmatrix} 0.58 & 0 & -0.0003 \\ 0 & 0.58 & -0.0004 \\ 0 & 0 & 0.58 \end{bmatrix}, \boldsymbol{F} = \begin{bmatrix} 0 & -0.0004 & 0.094 \\ 0.0004 & 0 & -0.88 \\ -0.093 & 0.88 & 0.23 \end{bmatrix}$$

The implementation of the Homography Projection (HP) of a point is presented in page 17 of the lecture notes on Matching, while the calculation of the epipolar lines requires the multiplication of the corresponding point with the fundamental matrix. The result defines the parameters a, b, c of the line, i.e. $ax + by + c = 0$. The epipolar lines and the inliers are shown in figure 5. The Homography (HA) and Fundamental Matrix (FA) accuracies have been calculated using the Euclidean distance since the images lie on the Euclidean space. Their values are:

$$HA = 0.005, FA = 0.0083,$$

and it is obvious that there is noise in the points/features for various reasons, that could not be eliminated.

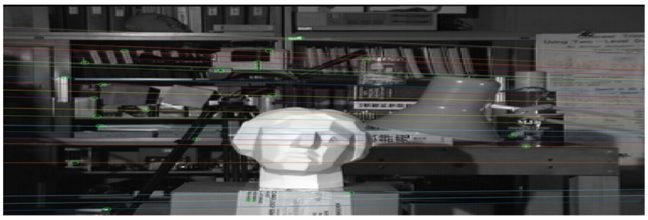

Figure 5: Inliers and Epipolar Lines in first image.

2. Image Geometry

2.1 Homography

A set of images hereafter referred to as **HG** was created by rotation and changing scale. Our **first objective** is to determine the effect of reducing size of original image to half and eventually to one-third on the operation of Harris Interest Point Detection and Matching. It can clearly be observed in figure 6, that reduction in size led to appearance of fewer interest points. We have **157, 93 and 47 interest points** respectively in our three HG images. This occurs due to that simple fact that the size of Gaussian window implemented in our Harris detector is invariant to scale changes. The table for HA error between same image at three different scales is given below. The HA errors are very low because there is only a change of scale between compared images and no translation or rotation exists. Also, only inliers have been used to compute homography matrix.

Table I. HA Error between Images

Image Pairs	Original: Scaled by 2	Original: Scaled by 3	Scaled by 2: Scaled by 3
HA Error	0.134	1.7x10^{-3}	0.234

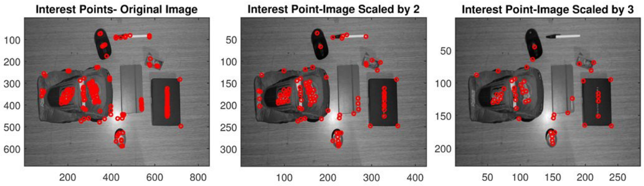

Figure 6: Visualization of interests points for our three images cases.

Our **second objective** is to compare two methods for interest point extraction: Automatic and Manual and to further analyse and compare geometric transformation parameters that can be derived from the two homographies. **Ten interest points were manually selected** and the same number of points were chosen randomly from amongst all automatically extracted ones. Another class of interest points referred to as **"ground-truth"** points were also chosen from amongst inliers of automatic interest points. From the interest points homography matrices for automatic (right) and manual (left) methods were obtained. These matrices were further used to project the points from one image to another image. The homography matrices are given below. Interpolation was used in provided matrices to eliminate inconsistencies in the rotational and scaling components before proceeding to compute geometric transformation parameters.

$$\boldsymbol{H_M} = \begin{bmatrix} 1.2044 & -0.6367 & 215.071 \\ 0.3133 & 1.2047 & -197.816 \\ 0 & 0 & 1.00 \end{bmatrix}, \boldsymbol{H_A} = \begin{bmatrix} 1.2023 & -0.6366 & 217.371 \\ 0.3153 & 1.2065 & -200.016 \\ 0 & 0 & 1.00 \end{bmatrix}$$

Solving simultaneous equations after eliminating inconsistencies for the automatic interest point extraction method we obtain **Scaling factor of 1.47** and **Rotation of 22.23°** for image transformation. Similarly, for manual interest point extraction method we obtain **Scaling factor of 1.57** and **Rotation of 26.23°** for image transformation.

To compare the obtained results, we need **ground-truth** values of geometric transformation parameters. We run RANSAC to obtain homography matrix again using all available points and set the projection error threshold to be only 1 pixel. Similar process is applied on the obtained ground truth homography matrix and **corresponding scaling factor of 1.51 and rotation factor of 23.67°** for image transformation has been obtained. Ground truth values are closer to those obtained from automatic method. To further establish superiority of automatic method over manual method we provide a figure below comparing these methods. The difference between expected and actual projection of points done by making use of homography matrices shows that indeed automatic method has lower error than manual method. A major reason for this can be attributed to the fact that it is very hard to select exactly the same pixel from two images using our eyes.

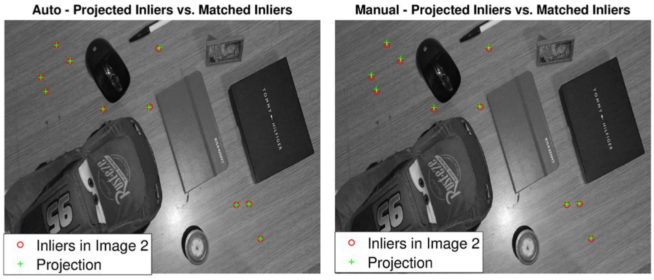

Figure 7: Projection of Interest points extracted using Automatic & manual method. Green points denote the projections of manually selected points using Homography matrix calculated from automatic and manual interest points respectively while the inliers (marked in red) are obtained from ground truth.

Our **third objective** for this sub-section is to understand the impact of number of available corresponding points on the accuracy of estimated homography. Ground truth set of 30 corresponding points was obtained by applying **RANSAC (Parameter setting: Iterations-1000; Threshold of Error-1.2).** We will obtain HA for each simulation by applying homography matrices from our simulations on these pair of ground-truth points. For our first simulation we randomly select 4 pair of points to compute homography matrix and obtain the HA as described above. The minimum number of points chosen is four because post RANSAC application, a correct homography matrix can be obtained as soon as are able to obtain four correct points from all available points. We gradually increase number of selected pair of points. Note that due to relatively small number of available interest points, the reference points will not be separated from the homography matrix training data. We plot the obtained result for Average HA as a function of number of correspondences in figure 8 (a). It can clearly be observed that average HA quickly converges to zero as number of available points increase. Probability of the event that out of all available points we have atleast four points that are part of previously determined ground truth increases with the number of available points. This in turn leads to the observed result of HA converging to zero with increase in number of correspondences. The outliers have been visualized in figure 8 (b). We had a total of 38 available pair of points out of which 8 were classified as outliers and 30 were chosen as ground truth. With our chosen value of threshold, the outliers might not appear as outliers to the human eye, but in reality, they are.

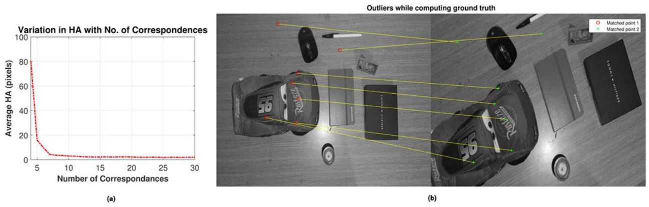

Figure 8: (a) Variation in HA with number of correspondences. (b) Outliers obtained while generating ground truth.

2.2 Stereo Vision

Fundamental matrices from automatic correspondences has been provided below:

$$\boldsymbol{F_1} = \begin{bmatrix} 0 & 0 & 0.0267 \\ 0 & 0 & 0.0008 \\ -0.0274 & -0.001 & 0.993 \end{bmatrix}, \boldsymbol{F_2} = \begin{bmatrix} 0 & 0 & -0.0277 \\ 0 & 0 & -0.0007 \\ 0.0269 & 0.0006 & 0.993 \end{bmatrix}$$

For the two images from set FD, two obtained epipoles for each of two images were: (1.0e+04 *0.0489, 1.0e+04 *-1.2113) and (1.0e+04 *0.02, 1.0e+04 *-1.373). In theory these epipoles should be located at infinity on the X-axis, however, our observations in this regard are somewhat different. FD set of images were supposed to have a horizontal translation of 20 cms but it appears that some vertical translation has crept in as well. Epipolar lines shown on images is given in figure 9 of this report. Since, we have epipoles located very far from the origin, it was not possible to fit them on the same image.

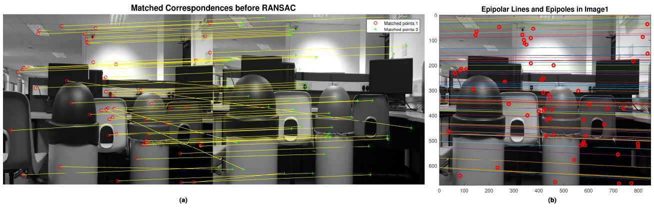

Figure 9: (a) Matched Correspondences before RANSAC (b) Epipolar Lines and Epipoles for Image A.

Since we observe that there has been some introduction of human error while taking these images we perform stereo rectification on our chosen set of images before proceeding with further tasks of this section. Stereo image rectification makes use of image projection onto a common image plane such that this image projection makes the image appear as though there is only horizontal translation between two images (images captured from parallel cameras). The stereo rectified pair of our images is given in figure 10(a). We now make use of stereo rectified images to calculate the disparity between image A and B. This disparity map is also presented in figure 10(b). The presence of black dots all over the disparity map hints at presence of some form of error. Possible reasons for this error can be: error in stereo rectification, algorithm used to compute disparity map.

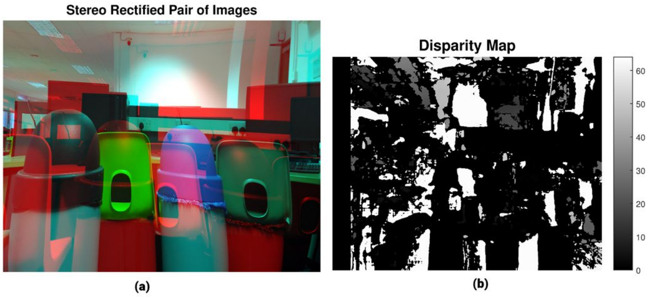

Figure 10: (a) Stereo Rectified Image (2.2(e)). (b) Disparity map between images A and B.

The FD images used are all 2D images, which can be converted to 3D images by reconstruction using our obtained disparity map. This resultant 3D reconstruction is known as depth map of our pair of images. High error areas in disparity map has led to poor reconstruction of corresponding areas in depth map. To obtain very accurate reconstruction, we need to positively ensure that error in disparity map is zero. The original depth map (a), depth map obtained after taking images by changing focal length by 2mm (b) and depth map obtained by addition of small random noise (Gaussian of size 2 pixels) to disparity map (c), are all illustrated in figure 11 of our report. Since between figure 11 (a) and (b) we have only changed the focal length, there is not much difference in the appearance of depth map. However due to addition of noise, there is marked difference between figures 11 (c) and 11 (a), (b) which are obtained from disparity maps without any element of noise added to them. The depth map could have been better in appearance if we had the resources to obtain optimal parameters through MATLAB's Stereo Camera Calibration App. For objects which are close to the camera, their 3D reconstruction is invariant to addition of small amount of noise, however, for objects which are located far away from the camera, we do not have reliable reconstruction and in the figure.

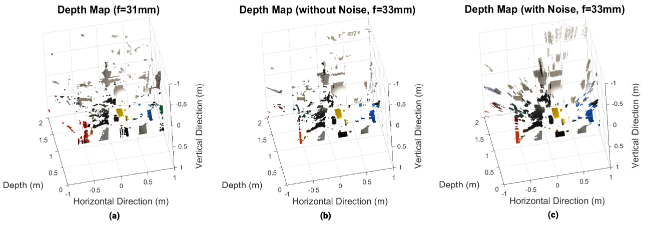

Figure 11: (a) Original 3D Depth Map (b) Depth Map (f=33mm) (c) Depth Map with Gaussian Noise added to Disparity map (f=33mm).

3. Conclusion

We have successfully implemented all the required objectives of our assignment. Due to unavailability of tripod and SLR camera some human error had crept in set of FD and HG images which has been a major roadblock in getting optimal results.

4. References

[1] "Tsukuba Images," [Online]. Available: http://vision.middlebury.edu/stereo/data/scenes2001/.

[2] K. Mikolajczyk, "Machine Learning for Computer Vision - Lecture Notes," Imperial College London, London, 2018.

[3] "Histogram of Oriented Gradients," [Online]. Available: https://en.wikipedia.org/wiki/Histogram_of_oriented_gradients.

[4] "Learn OpenCV," [Online]. Available: https://www.learnopencv.com/histogram-of-oriented-gradients/.

[5] "OpenCV - Homography," [Online]. Available: https://docs.opencv.org/3.4.1/d9/dab/tutorial_homography.html.

[6] "Wikipedia - Fundamental Matrix," [Online]. Available: https://en.wikipedia.org/wiki/Fundamental_matrix_(computer_vision).

[7] "Wikipedia - Eight Point Algorithm," [Online]. Available: https://en.wikipedia.org/wiki/Eight-point_algorithm.

[8] E. Dubrofsky, "Homography Estimation," Vancouver, 2009.

APPENDIX I

To support our result discussed in Section 1.1 of our report we present the outcome of manual matching of five-points on Tsukuba Images.

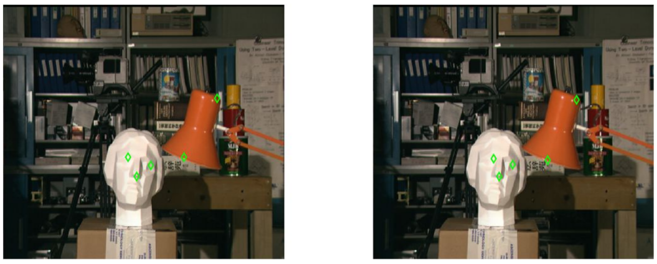

Figure I-1: Manually Matched Points on Tsukuba Images.

: