Report for CVPR Coursework

Ziyue Tong, zt421, CID 02077487; Yangyang Cao, yc2321, CID 02077610

1. Task 1 Collect Data

Collect Data Figure 1, the target object is a toy with various texture features, which is suitable for feature detection. In this coursework, all photos were taken by the camera boarded by iPhone 11.

Figure 1. The toy used for target object in this coursework

The data used in this coursework will be presented in the appendix session. The calibration grid adapted is 22mm in length for each black or white box.

2. Task 2 Keypoint correspondence between images

Task 2 Part 1 A set of 15 corresponding points are selected manually with the rectified image is in Figure 2.

Task 2 part 2 Three algorithms(SURF,Harris,SIFI) are applied to detect and match points automatically and their performances are compared with that of manual methods. The results of all the methods are analyzed by their Mean Square Error per pixel. The homography matrix is used to project the keypoints of the original image onto the rectified ones. Therefore with the values generated by the two methods are regarded as estimate values, we can compute their corresponding Mean Square Error(MSE). SIFI has shown the most accuracy with the least MSE per pixel among all the methods. SURF method computes the result with the least time required. All the automatic algorithms have shown wrong matches between keypoints. This error hardly occurs if points are selected manually, but can raise other problems such as reducing accuracy and increasing the time consumed.

Method	MSE per pixel
Manual	4.3009×10^-3
SURF	3.8318×10^-3
Harris	$6.1317 \times 10 - 4$
SIFI	2.3054×10^-5

Table 1. MSE comparison

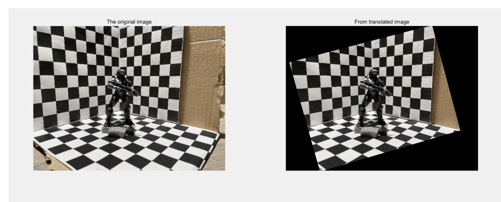

Figure 2. Manual selecting points and results

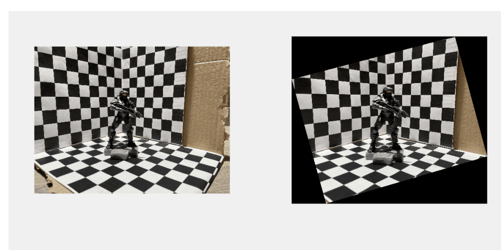

Figure 3. Automatic selecting points and results

3. Task 3 Camera calibration

Task 3 part 1 The parameter parameters were acquired by using Camera Calibrator App in MATLAB [1]. In table 2, the intrinsic parameters of the camera used will be presented.

From the table above, it can be discovered that the parameters of focal length estimated in both x and y direction are approximately the same, indicating the reliability of the parameter. The data of principle point estimated

Parameter	Quantity	Errors
Focal length (F_x)	3086.8191	+/-15.4495
Focal length (F_y)	3088.3032	+/-15.0282
Principal point (P_x)	2038.7043	+/-10.3000
Principal point (P_y)	1423.7761	+/-19.8792
Skew (s)	3.8325	+/- 2.1123

Table 2. Parameters of the camera used

agreed with the central point of the image (with resolution 4032*3024), although minor error exits. The estimated value of skew is not reliable referring to the ratio between its error and estimated value, which may caused by not only original misalignment from manufacture, but also some damages to the camera.

For the corrected image, though the edges on the calibration board were straighten, the whole image can be still distorted since there were distortions when the chess board was placed on the card board.

Task 3 part 2 Distortions of the camera used can also be acquired by the Camera Calibrator toolbox in MATLAB. The estimated distortion will be presented in table 3 below.

Type	Quantity	Errors
Radial distortion (X)	0.1731	+/-0.0087
Radial distortion (Y)	-0.1921	+/-0.0219
Tangential distortion(X)	-0.0202	+/-0.0013
Tangential distortion(X)	0.0047	+/-0.0015

Table 3. Parameters of the camera used

The distortion can be illustrated by comparing the two images presented below. The left part of the figure below is the original image taken by the camera, and on the right is the one after calibration (MATLAB function undistortImage was used to show the rectified image). The changes between these two images illustrate the distortions of the camera.

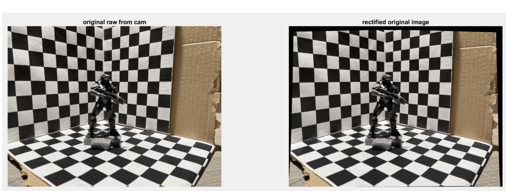

Figure 4. The comparison between original image (left) and rectified image (right)

A significant calibration for the skew angle, radial, and tangential distortion can be discovered by observing the figure above. The data set used for calibrating the camera and the extrinsic parameters visualization may refer to the appendix.

4. Task 4 Transformation estimation

Task 4 part 1 In this task, the homography matrix between a pair of images from HG, which is a similar sequence of the object with three planar of calibration grid as background. In this way, the position of the centre line of the camera was maintained nearly steady, while minor changes were applied, including rotation and zoom in. Therefore, a homography matrix can define the transformation process in between these two images from HG. In figure 5, the keypoints and their correspondences were visualized.

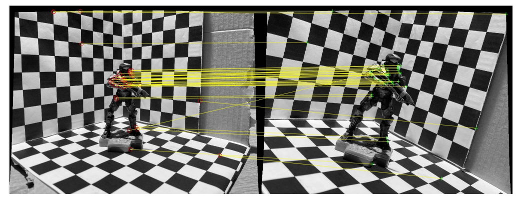

Figure 5. The comparison between original image (left) and rectified image (right)

The accuracy of the keypoints on the object is higher than that of the chess board. Generally, the keypoints are matched accurately. Based on the figure above, the homography matrix can be estimated, which will be presented below.

$$\begin{bmatrix} 1.1148968 & 0.29303497 & -1.9025118e-05 \\ -0.39301369 & 1.1039032 & -3.5306584e-05 \\ 282.06497 & -825.51031 & 1 \end{bmatrix}$$

Task 4 part 2 A pair of images from FD was used for this task. These two images were taken in slightly different positions and rotated by a minor angle. The fundamental matrix estimated would be presented below.

$$\begin{bmatrix} 1.7681e-08 & 1.2041e-07 & -2.6200e-04 \\ 1.1518e-07 & 5.4416e-08 & -3.7962e-04 \\ -2.5833e-04 & -2.0376e-04 & 0.9999 \end{bmatrix}$$

The fundamental matrix estimated can be verified by comparing with the equation presented below, in which X is for the coordinate of matching point coordinates and f is the fundamental matrix above. The result is -0.00485, which is close to 0, meaning that the fundamental matrix is valid though minor errors existed.

$$X^T \times f \times X = 0$$

The keypoints and their corresponding epipolar lines of the image pair will be presented in figure 6.

The vanish point of the image can be accomplished by conducting edge detection. Since the object was placed on a calibration board, the edges can be effectively detected. By applying Hough transform to the image, the lines indicating

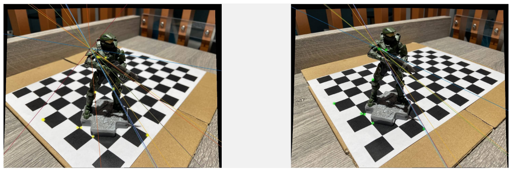

Figure 6. The keypoints and corresponding epipolar lines of two images

the extension of edges can be estimated. In the figure 8, The vanish point was generated.

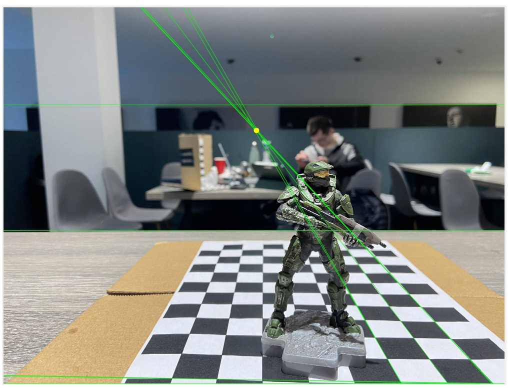

Figure 7. Vanishing point of an image, the vanishing point is the yellow dot

Task 4 part 3 A method has been proposed to calculate the number of outliers the method can tolerate. Average distances between every outlier and the nearest epipolar line is computed to observe the trend of mean error as the number of outliers increase. From the figure it is clear that after 80 outliers the mean error is beyond tolerance.

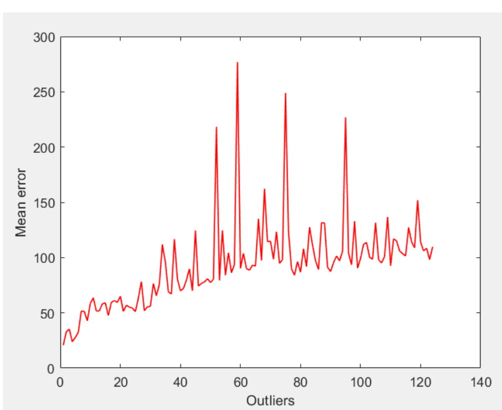

Figure 8. Mean error as outliers increase

5. Task 5 3D geometry

Task 5 Part 1 To generate a stereo vision with inputs of similar images, which were taken in two slightly different positions, feature detection was conducted and rectification was completed based on the data of feature points [3]. The stereo rectified pair of the images used in this section will be presented in the figure 9, in which the epipolar lines were labeled. It can be discovered that the epipoles of the rectified images are located at infinite far away (which means that epipolar lines are parallel).

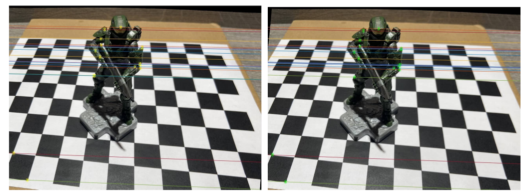

Figure 9. The rectified image pair with parallel epipolar lines

Task 5 Part 2 The depth map acquired by the stereo image pair presented in the figure above can generate a depth map of the object. By processing the rectified images from the previous part, the parameter of the depth dimension can be calculated [2]. In figure 10, the depth map is being displayed.

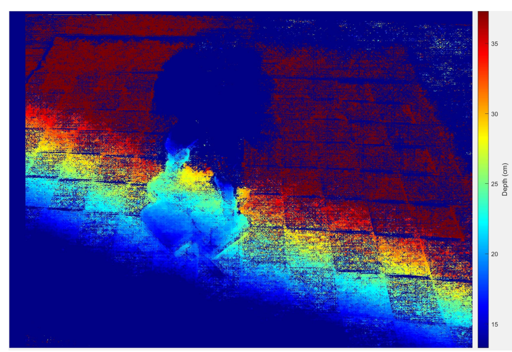

Figure 10. The depth map of the object. Colors close to red represent large distances and color close to blue represent short distances from the camera. Details shown on color bar

According to the depth map presented above, the map acquired can distinguish the base, legs, arms, and, gun of the toy. Moreover, the depth of all parts of the toy to the camera can be specified. Particularly, the best result of depth map was presented in the appendix.

References

[1] MathWorks. Camera calibrator, [online] https://uk.mathworks.com/help/vision/ref/cameracalibrator-app.html.

[2] MathWorks. Depth estimation from stereo video, [online] https://uk.mathworks.com/help/vision/ug/depth-estimation-from-stereo-video.html.

[3] MathWorks. undistortimage, [online] https://uk.mathworks.com/help/vision/ref/undistortimage.html.

6. Appendix

Task 1 FD and HG sequence In the figures below the data collected for HG and FD sequences would be presented.

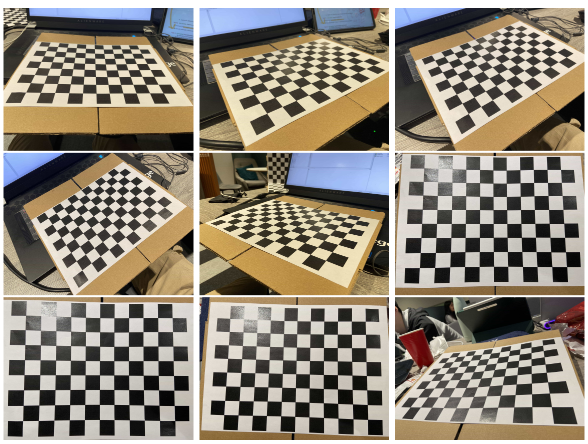

Figure 11. The FD sequence without an object

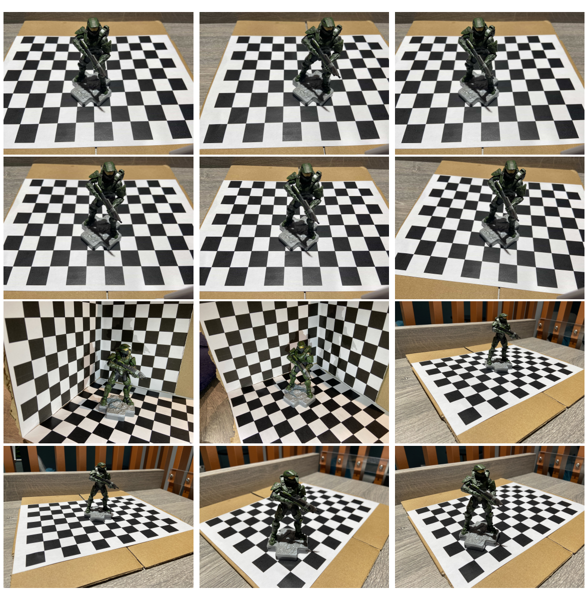

Figure 12. The FD sequence with an object

The HG sequence is used to generate the homography metrix, which would require the centre line of the camera maintain stable while the original and translated images were taken.

Figure 13. The HG sequence

Task 5 Depth map The best results of the depth map would be shown below

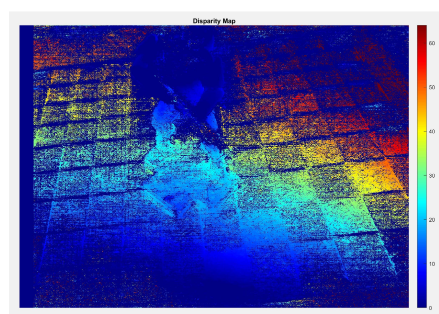

Figure 14. The best result of depth map, however, the color bar was not defined