Junyu Yan, jy420, 01895990, Yibing Liu, yl11820, 01939400

1. Collect Data

Task part In this part, we first collect 20 pictures (FD) with different camera's positions. FD contains 10 pictures with only 3D grid and the other 10 pictures with object in the grid. Then, we collect 10 pictures (HG) by changing the zoom to 1.5 factor and slightly rotating the camera clockwise or anti-clockwise with 10 to 20 degrees. The whole data is shown in the Appendix.1 and Appendix.2.

2. Keypoint correspondences between images

Task part In this part, we first use the MATLAB function `ginput(n)` to manually click on 10 corresponding points in two HG images. Then, we use SURF descriptor to detect keypoints and match correspondences automatically in the same pair of images. For automatic matching, in order to avoid matches on the grid rather than the object, we adopt the `ROI` method to specify the object range before matching. The results are shown in Fig.1 and Fig.2.

To evaluate the accuracy of manual and automatic methods, we use 20 manually generated correspondences to obtain the homography matrix, and reproject the keypoints. After that, the Mean Distance Error (average distance) between the reprojective correspondences and the manually and automatically captured correspondences is calculated to compare the quality of different methods. We conducted three experiments with different pairs of images, and the error is shown in Table.1.

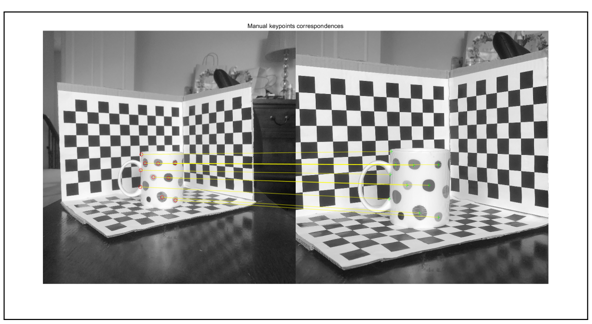


Figure 1. T2.1 Manually keypoints correspondences

For the quantity of correspondences, the number of keypoints produced by manual method is equal to the arbitrary integer n specified in `ginput(n)`, while the SURF detectors will automatically determines the number of keypoints based on the parameters, like 'MetricThreshold', 'MatchThreshold', 'Metric', etc[1]. In one experiment shown in the Fig.1 and Fig.2, we set $n = 10$ to obtain 10 manual correspondence, while the SURF descriptor obtains 81 with $MathchThreshold = 5$, $Metric = SAD$ and other parameters as default. If we lower the 'MatchThreshold' or 'MetricThreshold', we will get smaller quantity for SURF. For the quality of correspondences, as shown in Fig.1 and Fig.2, the manual method has better performance. More specifically, the manual method has a lower error than the automatic one as listed in Table.1.

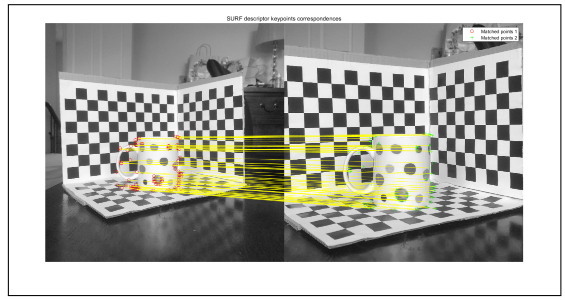


Figure 2. T2.2 Automatically keypoints correspondences by SURF descriptor

Experiment	Manual Error (pixel)	SURF Error (pixel)
Trial1	19.1953	37.4042
Trial2	10.7618	27.5919
Trial3	13.4885	28.9612

Table 1. T2.1 Results of Mean Distance Error between reprojective and original correspondences with manual and automatic methods

The Table.1 shows that the errors between the ground truth and actual correspondences are in the range of $10 - 40$ pixels, which is not a significantly small number. There are two main reasons. Firstly, the keypoints in the used images may not be clear enough to identify due to the small size or blur of images. Secondly, the error in the manual ground truth correspondences will result in imperfect homography matrix, therefore, the reprojective correspondences may not ideally accurate. The error with the same pair of images, but different number and chosen correspondences for ground truth generation is shown in Appendix, which can support the second reason.

3. Camera calibration

Task part 1) In this part, we first uses 13 FD images with only 3D grid for camera calibration. The images used are listed in Appendix.3. The calibration is executed with Single Camera Calibrator App in MATLAB. We create the checkboard pattern grid with size of $25mm \times 25mm$ by code, however, due to the distortions of the printer, the actual square size is $23mm \times 23mm$. After the first time calibration, the Mean Reprojection Error (MRE) of our calibration is 0.58. Then we remove two images with the largest Reprojection Error to improve the accuracy. The final MRE of our calibration is 0.3633, which is below 0.5 and illustrates our result is valid. The camera intrinsic and extrinsic parameters combined with their errors are listed in Fig.3 and Fig.4.

```
Intrinsics
----------
Focal length (pixels):   [ 3092.1530 +/- 31.7699     3111.6181 +/- 31.2745 ]
Principal point (pixels):[ 1449.4641 +/- 2.7077      1559.8991 +/- 7.0726  ]
Radial distortion:       [    0.2462 +/- 0.0121        -1.0586 +/- 0.1404  ]
```

Figure 3. T3.1 Camera Intrinsic Parameter

```
Extrinsics
----------
Rotation vectors:
                         [   -0.0492 +/- 0.0045       0.1859 +/- 0.0039       1.5786 +/- 0.0004  ]
                         [   -0.1053 +/- 0.0051       0.1018 +/- 0.0050       1.6278 +/- 0.0004  ]
                         [   -0.2427 +/- 0.0021       0.1101 +/- 0.0020       1.5280 +/- 0.0003  ]
                         [   -0.1334 +/- 0.0036       0.0381 +/- 0.0036       1.6330 +/- 0.0003  ]
                         [   -0.1627 +/- 0.0020       0.2216 +/- 0.0017       1.5979 +/- 0.0003  ]
                         [   -0.2310 +/- 0.0024       0.1315 +/- 0.0020       1.5805 +/- 0.0002  ]
                         [   -0.1425 +/- 0.0031       0.0198 +/- 0.0030       1.5651 +/- 0.0002  ]
                         [   -0.2873 +/- 0.0020       0.2429 +/- 0.0017       1.5993 +/- 0.0003  ]
                         [   -0.3193 +/- 0.0032      -0.0341 +/- 0.0020       1.5459 +/- 0.0003  ]
                         [   -0.1601 +/- 0.0022       0.0912 +/- 0.0022       1.6013 +/- 0.0002  ]
                         [   -0.2527 +/- 0.0044       0.1493 +/- 0.0036       1.6883 +/- 0.0004  ]

Translation vectors (millimeters):
                         [  165.5326 +/- 1.4484     -66.6854 +/- 3.7648    1669.1360 +/- 17.0254 ]
                         [   -0.1300 +/- 1.4689      -9.2567 +/- 3.8000    1672.7804 +/- 17.2225 ]
                         [  -77.2669 +/- 1.5254    -268.8298 +/- 3.8517    1738.3472 +/- 17.5608 ]
                         [   -5.4421 +/- 1.0085      -7.8067 +/- 2.6058    1148.2093 +/- 11.7953 ]
                         [  131.6432 +/- 1.0816    -127.5377 +/- 2.7579    1238.4203 +/- 12.5338 ]
                         [   29.1637 +/- 0.9377     -22.4414 +/- 2.4355    1074.5434 +/- 10.9263 ]
                         [   -2.2147 +/- 0.9250     -37.8249 +/- 2.3817    1053.2862 +/- 10.7899 ]
                         [   90.6878 +/- 0.9999     -37.1691 +/- 2.5827    1143.4256 +/- 11.4917 ]
                         [ -102.8367 +/- 1.0270    -119.8894 +/- 2.6278    1166.8111 +/- 11.6566 ]
                         [   84.8620 +/- 1.2128    -170.1115 +/- 3.1001    1393.8069 +/- 14.1194 ]
                         [  105.8890 +/- 1.3731     165.5028 +/- 3.6048    1569.7813 +/- 16.1255 ]
```

Figure 4. T3.2 Camera Extrinsic Parameter

Task part 2) According to intrinsic parameters, the camera we used has some radial distortion, and this can be clarified in Fig.5. It shows that the original straight line, specified by the blue line, is bend in the image, which is the pincushion distortion.

4. Transformation estimation

Task part 1) We operates on HG images with manually generated 20 correspondences to obtain the homography matrix. More specifically, we first generate the matrix $\mathbf{A}$ with the coordinates of correspondences, then derive the matrix based on the eigenvalues obtained from the SVD of $\mathbf{A}$. Multiple experiments were conducted with different

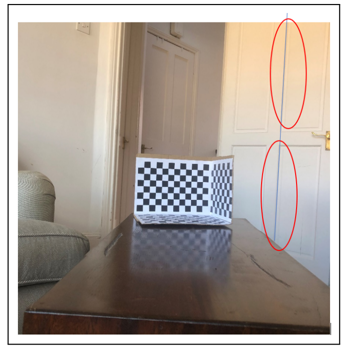


Figure 5. T3.1 The image illustrating the camera distortion

pairs of HG images or different number of correspondences for more reasonable results. The obtained homography matrix is listed in Appendix.4, and each experiment with same images and sufficient large number of correspondences provides similar results. With homography matrix, we can reproject the keypoints onto the other image and obtain the new correspondences. The results are shown in Fig.6 and Fig.7

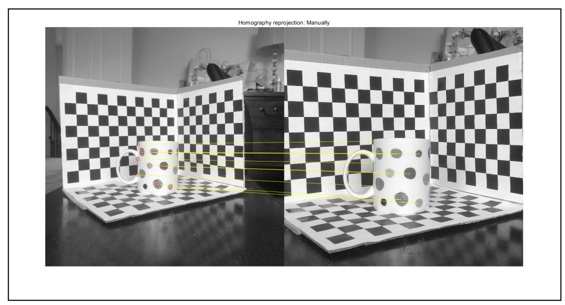


Figure 6. T4.1 The reprojective correspondences with manually selected 10 keypoints

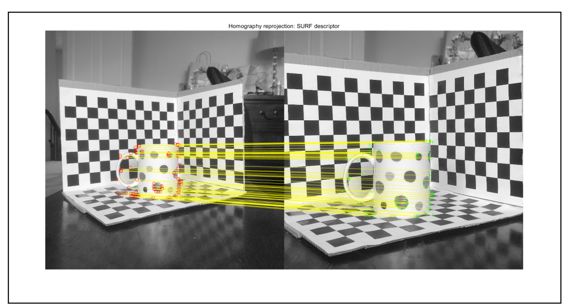


Figure 7. T4.2 The reprojective correspondences with SURF selected 81 keypoints

Note that when choosing the number of manually selected keypoints for matrix generation, there is a trade-off. More keypoints will make the homography matrix closer to the ideal one, however, a large number of them will cause higher complexity and human-clicking errors, which may results in a worse homography matrix. After multiple tests on 4 points, 8 points, 10 points, 20 points, 30 points as well as 35 points, we finally find using 20 points for generating matrix is more proper. Compared with figures in Task2, the reprojective keypoints and correspondences with hompgraphy matrix are accurate enough. The matching degree of selected correspondences for generating the homography matrix is quite important for the matrix accuracy. Some cases with badly selected correspondences are shown in the Appendix.4.

Task part 2) To obtain the fundamental matrix, we first use the SURF descriptor to capture the correspondences, then use MATLAB function `estimateFundamentalMatrix` to calculate the matrix. We adopt the RANSAC method to estimate the result using data without outliers. After the fundamental matrix is obtained, we use the MATLAB function `epipolarLine` and `lineToBoarderPoints` to obtain the epipolar lines and keypoints. The intersect point of epipolar lines is the epipole. Moreover, the vanishing point is obtained by extending parallel lines in the image. The 3D checkboard grid can produce many vanishing points, and connecting two of them gives the horizon. The results are shown in Fig.8 and Fig.9. The fundamental matrix obtained is shown in Appendix.

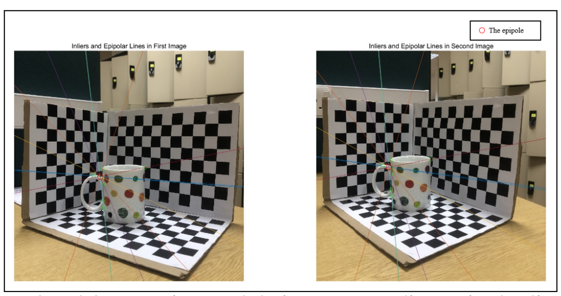


Figure 8. T4.3 Keypoints and their corresponding epipolar lines

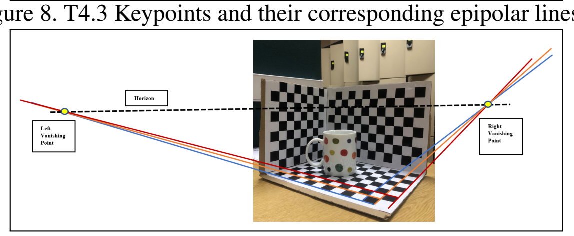


Figure 9. T4.4 Vanishing points and horizon with FD images

When obtaining the fundamental matrix, from different experiments, we found that the randomness in RANSAC leads to different matrix result each time, but the values are similar. Moreover, the figures and features used are also of effect, and more accurate correspondences will produce a more accurate result.

For the epipole and vanishing point, some pairs of images may not have them in the image due to different positions of cameras.

5. 3D geometry

Task part1 In this part, we use the method of Uncalibrated Stereo Image Rectification with MATLAB to rectify the stereo images. It first generates correspondences between two images with SURF descriptor, then obtain the fundamental matrix with correspondences and use it to remove outliers using epipole constraint, and finally uses `estimateUncalibratedRectification` function to rectify images. After rectification, both images have slight change since the distortion has been calibrated. The epipolar lines are almost parallel and coherent in the two images, as shown in Fig.10.

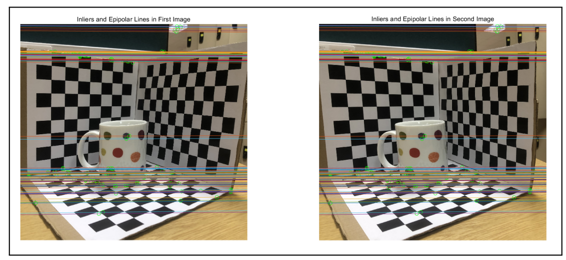


Figure 10. T5.1 Epipolar lines on rectified images

Task part2 To get the depth map, we use the Semi-Global Matching (SGM) method operating on rectified images to get disparity map, and then convert it into depth map. The result is shown in Fig.11. We also tried a more smoothed method, as Semi-Global Block Matching (SGBM), the result is listed in Appendix.

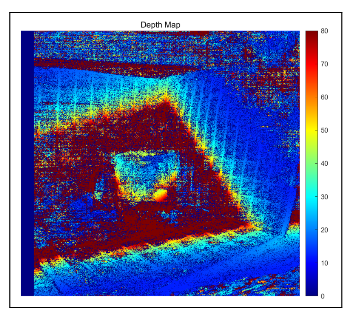


Figure 11. T5.2 Depth map of the object

6. Conclusion

In this paper, we illustrate the method for generating correspondences, homography and fundamental matrix, operating camera calibration and stereo rectification. With these methods, the 3D model of the object can be constructed.

7. Appendix

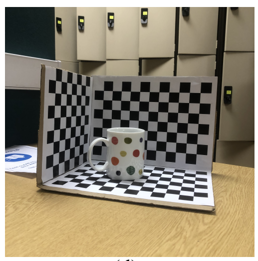
7.1. The FD images

This section lists the images produced by changing the position of camera with no rotation and zooming. Fig.7.1 lists FD images with objects, while Fig.13 lists images without the object.


(a)


(b)


(c)


(d)


(e)

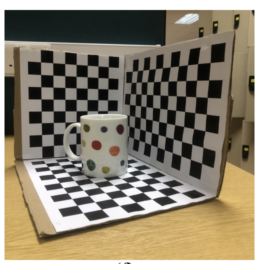

(f)

Figure 12. The FD images with objects

7.2. The HG images

This section lists the images produced by zooming or rotating the camera, while keep the same camera's location. Fig.14 lists HG images by zooming, while Fig.15 lists images created by rotation.

7.3. Calibration images

For the calibration, since the checkboard pattern square we used is small, the distance between the object and camera should be large enough (i.e,. larger than $1m$). Also, since the MATLAB can only detect the 2D grid, the images we take should only focus on one side of the 3D grid model. The example images we used for calibration is shown in Fig.16.


(a)


(b)


(c)


(d)

Figure 13. The FD images with only grids


(a)


(b)


(c)


(d)

Figure 14. The HG images by zooming

7.4. Homography matrix

We tried to uses different number of manually selected keypoints to generate the homography matrix with the same pair of images, After that, we reproject the SURF keypoints to evalute the accuracy of our homography matrix, The obtained homography matrix is give by Table2. The reprojective images is shown in Fig.17

From reporjective images, we find that as the number of keypoints chosen increases from 4 to 20, the accuracy of reprojection increases, but it shows bad performance when the number becomes 35. According to the Table.3, the ho-

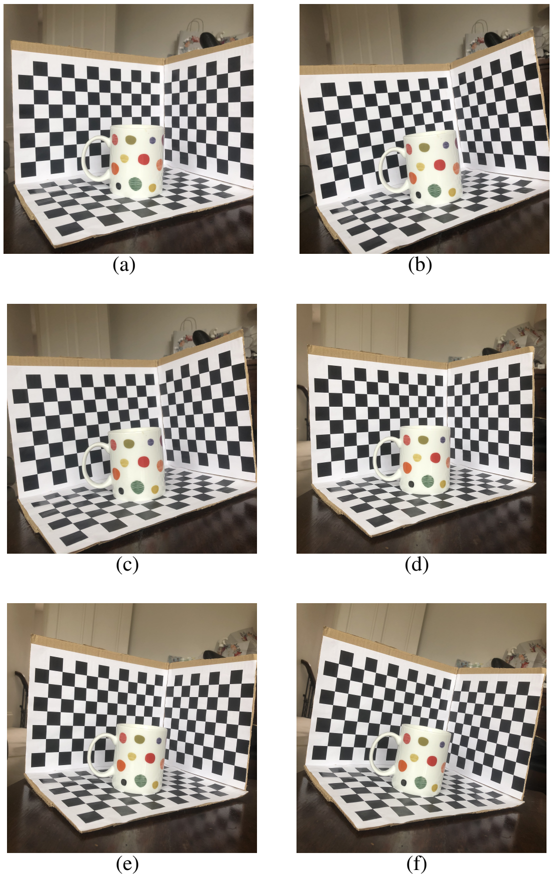


Figure 15. The HG images with rotation

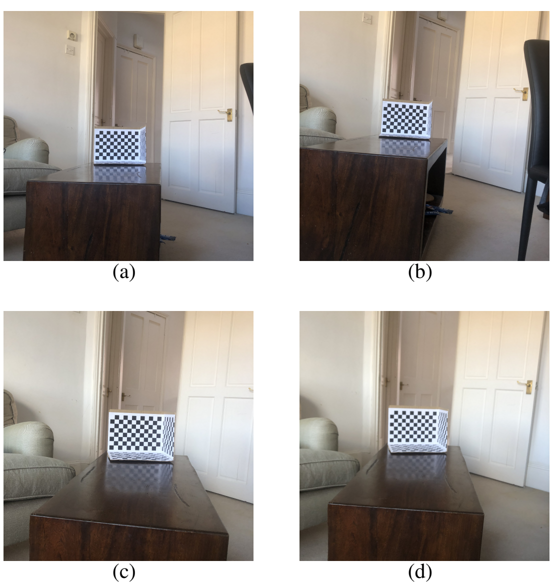


Figure 16. The example images for calibration

Experiment	Homography matrix
Trail1: 4 keypoints	$\begin{bmatrix} 6.0163 & 0.4673 & -4757.6402 \\ 1.9929 & 4.9763 & -5797.3638 \\ 0.0000 & 0.0003 & 1 \end{bmatrix}$
Trail2: 10 keypoints	$\begin{bmatrix} 1.9506 & 0.3206 & -1300.3793 \\ 0.0040 & 2.3898 & -1653.1797 \\ 0.0000 & 0.0002 & 1 \end{bmatrix}$
Trail3: 15 keypoints	$\begin{bmatrix} 2.3276 & 0.4571 & -1708.9962 \\ 0.1246 & 2.9393 & -2317.5688 \\ 0.0000 & 0.0003 & 1 \end{bmatrix}$
Trail4: 20 keypoints	$\begin{bmatrix} 2.3889 & 0.1269 & -1430.9226 \\ 0.4730 & 2.2547 & -1907.2378 \\ 0.0002 & 0.0000 & 1 \end{bmatrix}$
Trail5: 35 keypoints	$\begin{bmatrix} 2.8403 & 0.5727 & -2258.3022 \\ 0.2935 & 3.4267 & -2943.6770 \\ 0.0001 & 0.0004 & 1 \end{bmatrix}$

Table 2. The homography matrix generated with different number of manual chosen correspondences

mography matrices from 8 keypoints to 20 points are similar. Therefore, we conclude that the 20 manual-clicking correspondences will be the most proper.

The obtained homography with experiments of different pairs of images is given in Table3, we mark 20 keypoints in both trial1 and trial2. It is clear that for different images taken with different camera's positions, the homomgraphy matrices are different.

Experiment	Homography matrix
Trail1	$\begin{bmatrix} -3.0918 & -2.0198 & 5104.2241 \\ -1.2167 & -4.5652 & 6010.3339 \\ -0.0007 & -0.0011 & 1 \end{bmatrix}$
Trail2	$\begin{bmatrix} 1.9506 & 0.3206 & -1300.3793 \\ 0.0040 & 2.3898 & -1653.1797 \\ 0.0000 & 0.0002 & 1 \end{bmatrix}$

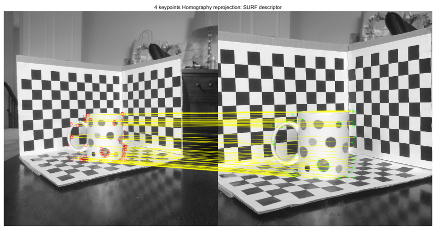
Table 3. The homography matrix generated with different pairs of images

7.5. Fundamental Matrix

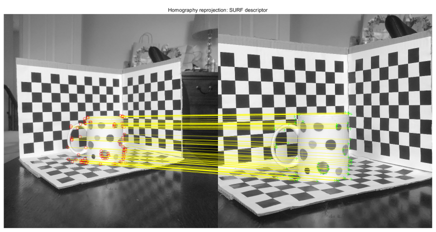
We first generate the images with the same pair of images, and repeat multiple times. The results are shown in Table.4. Then, we try on different images to compare the differences. This results are shown in Table.5.

Experiment	Fundamental matrix
Trail1	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0030 \\ 0.0000 & 0.0000 & -0.0231 \\ 0.0019 & 0.0213 & 0.9995 \end{bmatrix}$
Trail2	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0025 \\ 0.0000 & 0.0000 & -0.0260 \\ 0.0016 & 0.0238 & 0.9994 \end{bmatrix}$
Trail3	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0022 \\ 0.0000 & 0.0000 & -0.0216 \\ 0.0014 & 0.0195 & 0.9996 \end{bmatrix}$

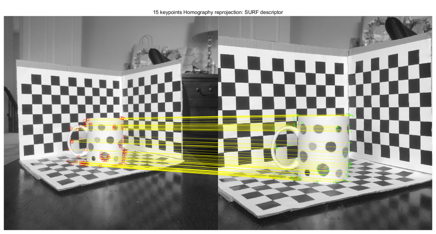
Table 4. The fundamental matrix generated with the same pair of images



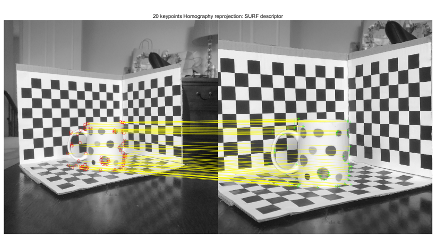
(a) 4 keypoints reprojective image



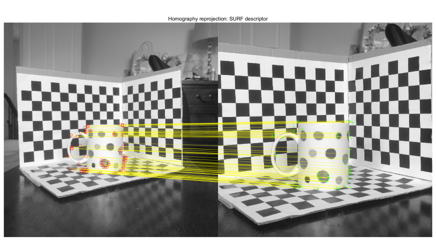
(b) 10 keypoints reprojective image



(c) 15 keypoints reprojective image



(d) 20 keypoints reprojective image



(e) 35 keypoints reprojective image

Figure 17. The reprojective images with homography matrix generated with different number of manual-clicking correspondences

From the results, the fundamental matrix are very similar for different experiments. It is a 3×3 rank 2 matrix independent of scene structures. The minor error is due to the mismatch of correspondences.

Experiment	Fundamental matrix
Trail1	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0020 \\ 0.0000 & 0.0000 & 0.0173 \\ 0.0010 & -0.0178 & 0.9997 \end{bmatrix}$
Trail2	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0025 \\ 0.0000 & 0.0000 & -0.0260 \\ 0.0016 & 0.0238 & 0.9994 \end{bmatrix}$
Trail3	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0005 \\ 0.0000 & 0.0000 & 0.0084 \\ 0.0001 & -0.0092 & 0.9999 \end{bmatrix}$

Table 5. The fundamental matrix generated with the different pairs of images

7.6. Depth Map

The rectified stereo image and the disparity map of the SGM method are shown in Fig.18 and Fig.19. Compared with the depth map, they are inversely proportional. The

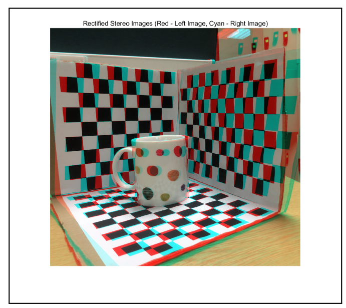


Figure 18. The rectified stereo image

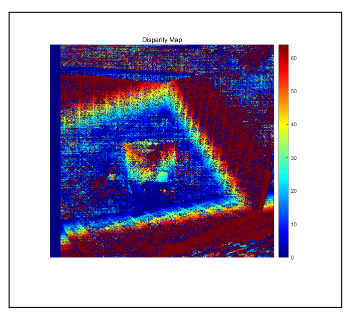


Figure 19. The disparity map of the object with SGM

depth map produced by SGBM method is shown in Fig.20 We also try images from different views, and the rectified stereo image and depth map are shown in Fig.21 and Fig.22, respectively.

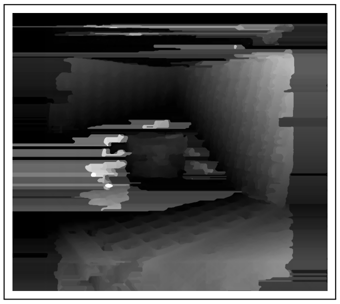


Figure 20. The depth map generated with SGBM

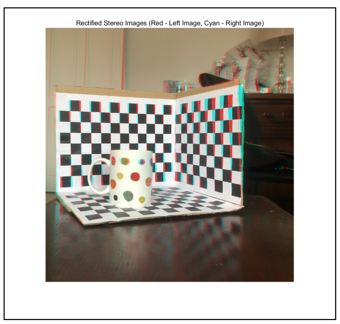


Figure 21. Rectified stereo image from different views

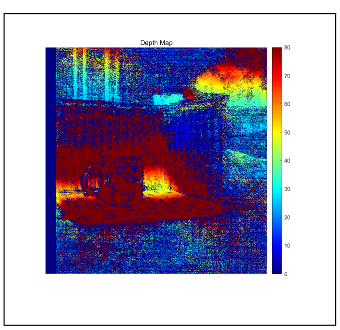


Figure 22. Depth map from different views

References

[1] Mathworks.com. Find matching features - matlab matchfeatures., 2021.