Maria Stamatopoulou ms2520 CID:01977521 Michal Olak mro20 CID: 01985975

1. Data Collection

Task part The data acquired for this work is in form of images capturing calibration grid with and without an object of interest inside of it. Two datasets were acquired, namely FD and HG. The FD dataset contained images taken from different positions and orientations of the camera, whereas the HG dataset contained images taken from the same position, but different orientation of the camera. The hardware used was back camera of IPhone SE.

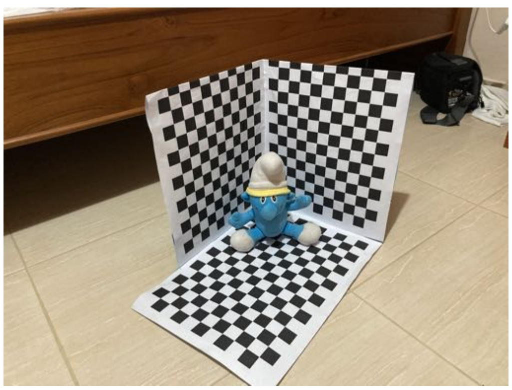

Figure 1: T1.a. Example image from FD part of the dataset. Each grid consists of 6x8 squares with 25mm side length.

2. Keypoint correspondences between images

Two methods are applied to detect the keypoints of the images.

Automatic method All images are converted into grey scale images to make the neighbour detection more accurate, and to reduces computational requirements. The detector used was SURF, as it was found to be the most successful in feature detection. A match threshold is set to 0.1, to avoid overestimation of points. The keypoints from the images are extracted and then matched to each other. To ensure that correspondences are accurately matched, the sum of squared difference metric condition is applied when matching the features.

Manual method The images were displayed and then the Data Tool was used, several corresponding points were manually detected in both images. The points were stored in vectors and then displayed using the same function as in the automatic case. Points in the edges are selected, as they are the easiest to detect and won't cause any confusion in matching.

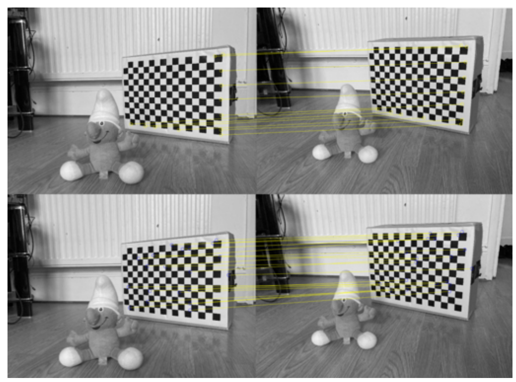

Figure 2: Feature Matching: a. Automatic, b.Manual

Evaluation Quantity wise, automatic detection is more efficient. Manual detection is rigorous while automatic can detect many features in a matter of seconds.

Qualitative evaluation was determined manually. The automatic methods line number was reduced, but setting a stricter threshold and then manually evaluated for correctness. For the given pair of images, the automatic detection was reduced to 112 lines, out of which 82 were successful in corresponding points, giving 73.2 % accuracy. Hence, manual detection is more accurate.

Method	Quality (%)	Quantity
Manual	100	12
Automatic	73.2	112

Table 1: T1.2 Manual vs Automatic Feature Detection

3. Camera Calibration

The goal of camera calibration is the estimation of the camera parameters, i.e. intrinsics, extrinsics, and distortion coefficients, to then remove distortion from the images. Calibration was performed by loading the 2D-images in the MATLAB ”Camera Calibrator” App, setting the size of the checkerboard square to 25mm and then comparing them to generated checkerboard 3-D world points of the same size. There were 21 images used, bpth with and without the presence of the object, to ensure accurate parameter estimation.

Intrinsic Parameters	Values(mm)
Intrinsic Matrix	$\begin{bmatrix} 3321 & 0 & 0 \\ 0 & 3332 & 0 \\ 2023 & 1459 & 10 \end{bmatrix}$
Focal Length	[3320 3332]
Principal Point	[2023 1459]
Image Size	[3024 4032]

Extrinsic Parameters	Values(mm)
Rotation	$\begin{bmatrix} -0.8738 & -0.0436 & 0.4844 \\ -0.0291 & -0.9895+03 & -0.1417 \\ 0.4855 & -0.1379 & 0.8633 \end{bmatrix}$
Translation	$\begin{bmatrix} [281.6682 & 66.766 & 674.8967] \end{bmatrix}$

Distortion Parameters	Values(mm)
Radial Distortion	[0.1519 -0.3924]
Tangential Distortion	[0 0]
Skew	0

Table 2: Camera Parameters

Images before and after applying the camera parameters are depicted below. The reported extrinsic parameters are based on the first displayed image. Distortion is corrected, as it can be seen from the edges of the image frame and checkerboard lines.

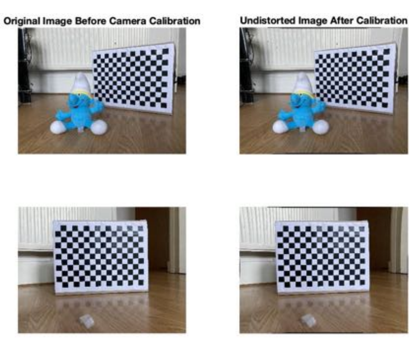

Figure 3: Camera Calibration Resulted Images

4. Transformation estimation

For estimation of both matrices the correspondences obtained in manual way were used. The reasoning behind that decision was that the performance of the automatic method is relatively poor. At the same time, the minimum number of required correspondences is quite small, respectively 4 and 7 for homography and fundamental matrices, therefore the amount of time required to obtain them is negligible. It is to be noted that in an automated system the correspondence matching would have to be improved to be consequently used for further applications. However, for the interest of time, this was deemed as further work.

Homography matrix Homography is a 6 degrees of freedom transformation that links two images of the same planar surface. In order to find its matrix coefficients, a minimum of 4 correspondences between two images are needed. Those were fed into OpenCV function findHomography, which outputted the following matrix:

$$H = \begin{bmatrix} 0.92 & 0.26 & -234.89 \\ -0.32 & -0.90 & 897.36 \\ 0 & 0 & 2 \end{bmatrix}$$

The matrix was used to transform the source image in order to match the destination image. The result of that can be seen in Figure 4

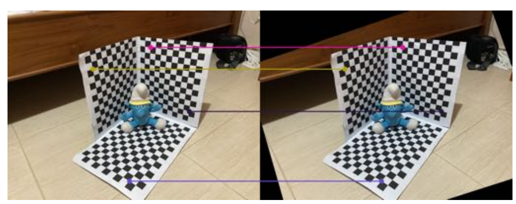

Figure 4: Keypoints and their correspondences projected from the other image. Left: destination image, Right: source image after applying Homography transformation to it

Fundamental matrix Corresponding points in stereo images are related by fundamental matrix. It is a 8 degrees of freedom transformation that requires a minimum of 7 correspondences to be calculated. For better accuracy, 10 correspondences were manually found and fed into OpenCV function findFundamentalMat. This obtained matrix is as follows:

$$F = \begin{bmatrix} 4.83e-08 & 22.34e-08 & -14.60e-04 \\ -16.53-08 & -0.85e-08 & -12.49e-04 \\ 11.67e-04 & 10.08e-04 & 1 \end{bmatrix}$$

Epilines visible in Figure 5 are showing convergence effect.

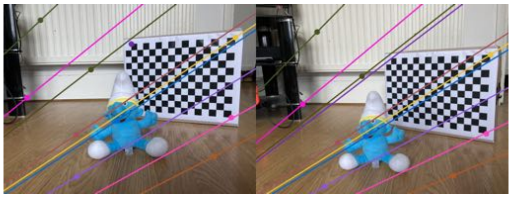

Figure 5: Keypoints and their corresponding epipolar lines in the other image

In both images they seem to converge towards a common point located south-west from the image. The point where the epilines meet is called epipole and thus the epipoles for both images are located away from the frame, towards south-west direction. Figure 6 presents the right image from figure Figure 5, however with vanishing point lines (light orange) and horizon (light green) lines added. Note that the exact location of those is unknown, as the vanishing point is located away from the frame, towards west direction.

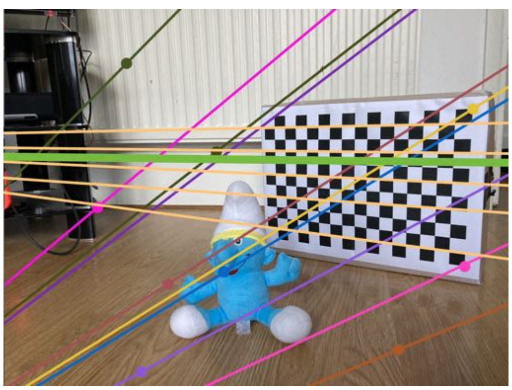

Figure 6: Stereo rectified pair of images with parallel epipolar lines. Top: original output of rectification, Bottom: after rectification the images were rotated approximately parallel to the original horizontal line

5. 3D geometry

Rectification Rectification is a transformation process used to project images onto a common image plane. It requires the correspondence points as well as fundamental matrix, which are fed into OpenCV function stereoRectifyUncalibrated. The images are rotated according to the function outputs. When stereo images are rectified, their epipolar lines become parallel, as presented in Figure 7.

Depth map Depth map is an image containing information relating to the distance of the surfaces of scene objects from a viewpoint, in this case camera position. The

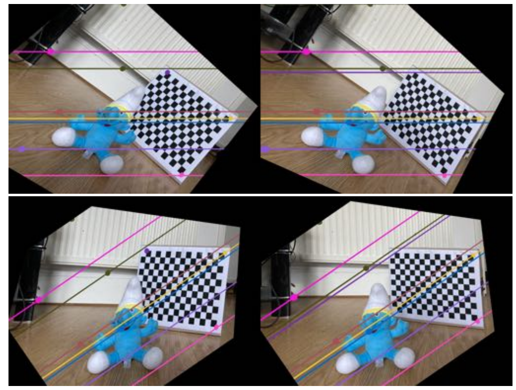

Figure 7: Stereo rectified pair of images with parallel epipolar lines. Top: output of rectification, Bottom: after rectification the images were rotated approximately parallel to the original horizontal line

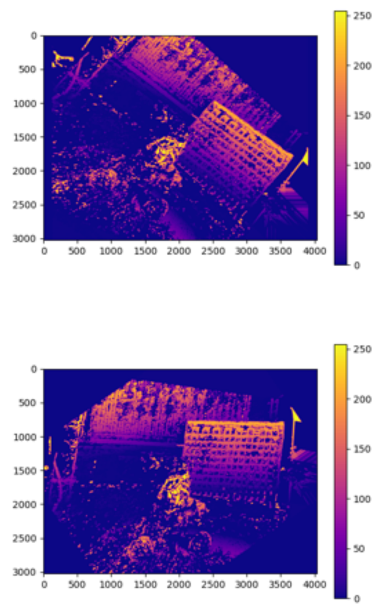

Figure 8: Depth map of the scene estimated from two views. Top: original depth map, Bottom: depth map for rotated image

depth is calculated based on disparity, which is difference in distance of corresponding image points and their camera centers. The bigger the disparity, the closer the object is. The depth map was obtained using OpenCV function StereoSGBM-compute , which was supplied with two rectified images in grayscale. The result can be seen in Figure 8

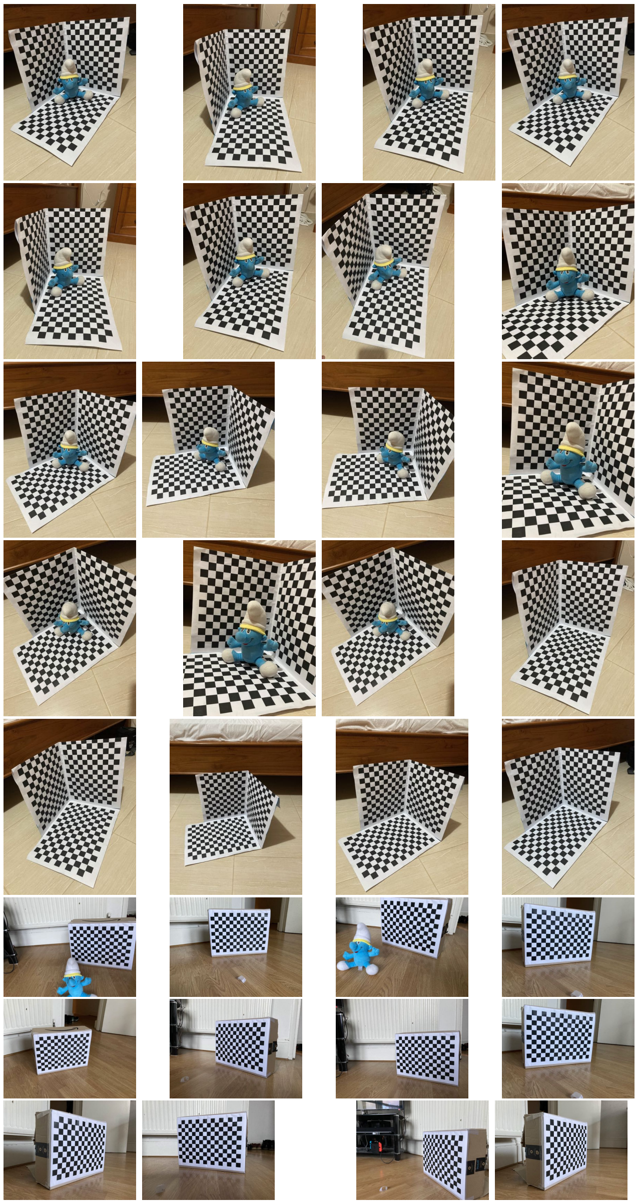

Figure 9: Task1:FD Images: Different Camera Positions

6. Appendix

Additional results are included here:

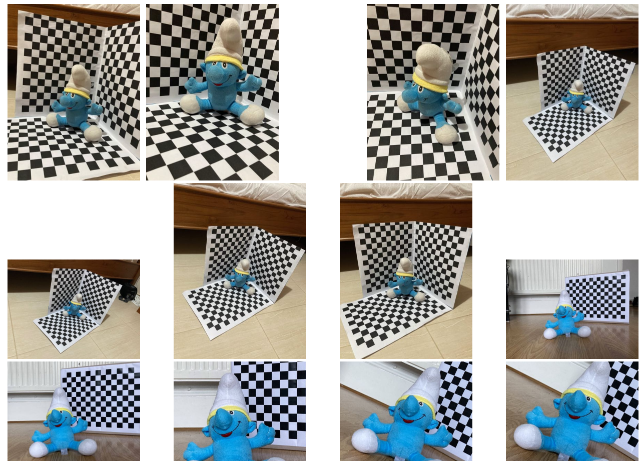

Figure 10: Task1:HG Images: Different Zoom and Camera Angle

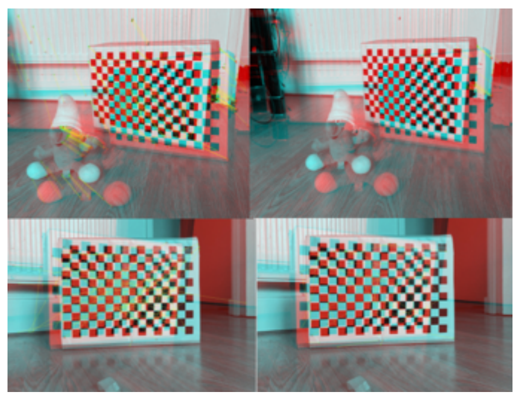

Figure 11: Task2: Matched Features for Images With and Without Object

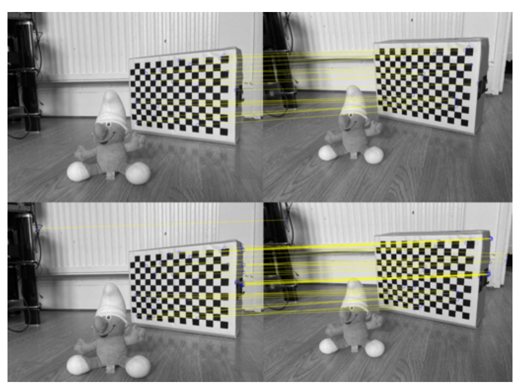

Figure 12: Task2: Matched Features: a. With Metric Application, b.Without Metric Application

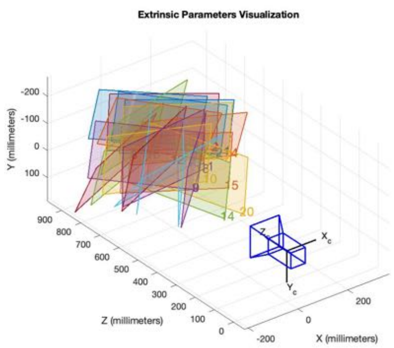

Figure 13: Task3: Estimated Extrinsic Parameters

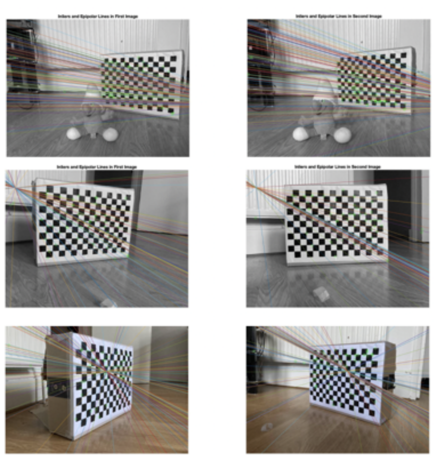

Figure 14: Task4: Images with Epipolar Lines Within and Out of the Image

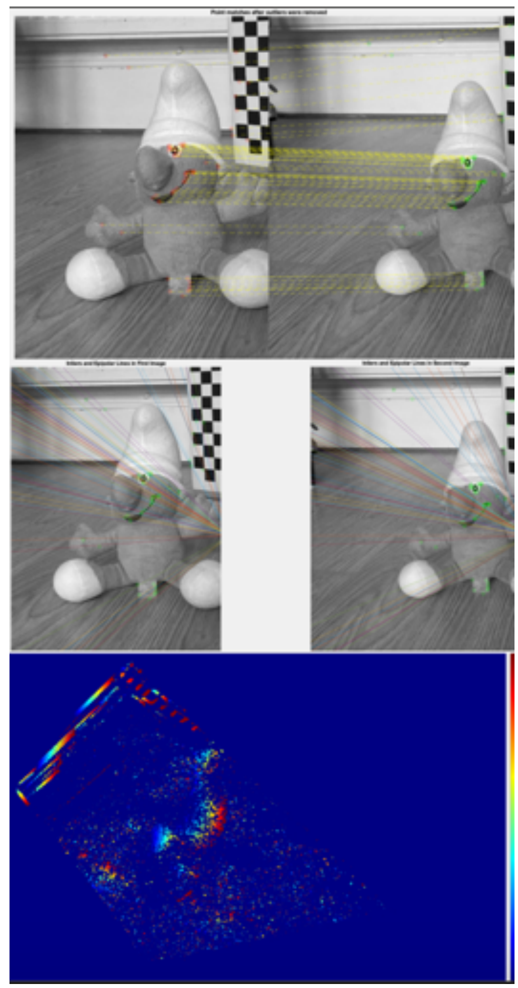

Figure 15: Task5: Disparity Map of Object