

# Xuanyu Lai, 02448599 , Shutuo Guo, 01857692

## 1. Collect Data

**Task 1** For most computer vision tasks, selecting an appropriate subject for detection is crucial. From experiments, we observed that the size and the texture significantly influenced the outcome. Ultimately, was selected a medicine bottle as the target. The Canon R10 was employed as the imaging device for this experiment.

The images utilized for this coursework is detailed in the appendix section. The calibration grid employed has a uniform square size of 20mm for each black and white square.

## 2. Keypoint orrespondences between images

**Task a&b** Manual & Automatic Figure 1 presents the outcome of manual correspondence established through ten selected points. Figure 2 illustrates the automated approach employing the SURF algorithm for point matching. The automated method yields a higher count of correspondences than the manual technique, though it is not devoid of occasional mismatches.

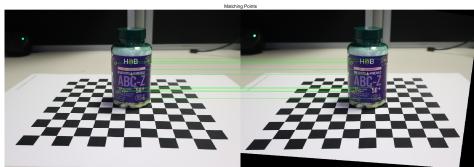


Figure 1. Manual selecting points

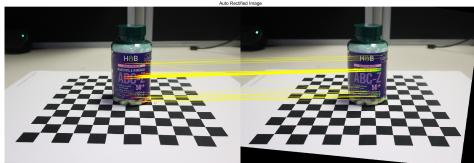


Figure 2. Automatic selecting points with SURF algorithms

**Analyze** In this task we tried different number of manual corresponding points and different automatic matching point algorithms: ORB SIFT SUFT Harris, etc., and concluded that we found the following advantages and disadvantages.

### Quality of Correspondences:

Manual: User-controlled validation provides accurate matches, but there are often errors at the pixel level, as it is difficult for the human eye to be pixel accurate.

Automatic: Uses algorithms (e.g., SIFT, SURF, ORB) for matching, generally effective but can mistakenly match similar features in different areas, especially in repetitive textures.

### Quantity of Correspondences:

Manual: Time-intensive and yields fewer matches, not suitable for large-scale or dense matching needs.

Automatic: Quickly detects numerous keypoints, ideal for applications needing dense matching, due to its ability to produce huge numbers of correspondences.

Additionally, it was observed that manual methods may lack consistency. Reproducing an identical set of correspondences poses a challenge. On the other hand, automated methods deliver consistent and reproducible results, provided the parameters and data remain constant.

## 3. Camera calibration

**Task 1** Matlab's toolbox has a wealth of analysis tools, in this paper, we choose the Camera Calibration tool to analyse the camera parameters, and its estimated parameters are shown in table 1

Parameter	Quantity & Errors
Focal Length ( $F_x$ )	$4981.1520 \pm 4.4746$
Focal Length ( $F_y$ )	$4991.5030 \pm 4.2318$
Principal Point ( $P_x$ )	$2986.5406 \pm 3.2510$
Principal Point ( $P_y$ )	$2008.8690 \pm 4.3134$
Skew	$-5.3552 \pm 0.5051$

Table 1. Camera calibration parameters with uncertainties.

Figure 3 illustrates camera reprojection errors and visualizations of extrinsic parameters. It is observable that the reprojection errors remain under 1.2, and the complete artifact planes confirm the camera's effective calibration.

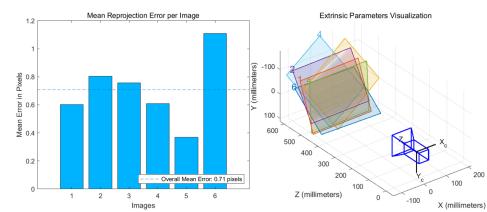


Figure 3. Visualization of calibration

**Task 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 2 below

According to table 1, it can be found that the camera lens has minor radial and tangential distortions. The low values suggest that the image quality is relatively unaffected by these distortions, and they could be corrected relatively easily through calibration.

Parameter	Quantity & Errors
Radial distortion ( $R_1$ )	$-0.0074 \pm 0.0021$
Radial distortion ( $R_2$ )	$0.0366 \pm 0.0133$
Tangential distortion ( $T_1$ )	$-0.0005 \pm 0.0002$
Tangential distortion ( $T_2$ )	$-0.0001 \pm 0.0002$

Table 2. Camera distortion parameters with uncertainties.

Distortion effects are evidenced through the comparison of the two images in figure 4. On the left is the camera's original image, the calibrated image is on the right.

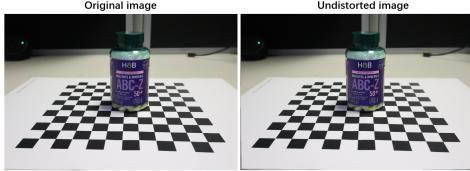


Figure 4. Comparation of original and undistorted image

**Analyze** In figure 4, the image has only a slight distortion in the bottom corners. The reason for the high quality of the images is that we use optical lenses. Traditional optical systems use complex designs to correct aberrations. In contrast, mobile phone lenses, limited by size and cost, may not be able to integrate correction elements. Therefore, they can only rely on correction algorithms, which introduce new errors in the image output process.

#### 4. Transformation estimation

**Task 1.a** The homography matrix transforms the appearance of images to account for changes in perspective. The estimated homography matrix is:

$$H = \begin{bmatrix} 0.9670 & 0.1665 & -0.0000 \\ -0.1675 & 0.9729 & -0.0000 \\ 407.6838 & -583.9127 & 1.0000 \end{bmatrix}$$

The values within this matrix suggest a modest scale and rotation adjustment, combined with a more pronounced perspective shift between the images. ORB features were instrumental in identifying keypoints for this transformation. RANSAC further refined these matches to isolate the most consistent points—termed inliers—and discard mismatches or outliers. The strength of this homography estimation is evidenced in Figures 5 and 6, where inliers clearly delineate the coherent structural links between the images.



Figure 5. Keypoint Matching. The number of inliers is 203, while the number of outliers is 1056, illustrating the initial disparity before RANSAC filtering.

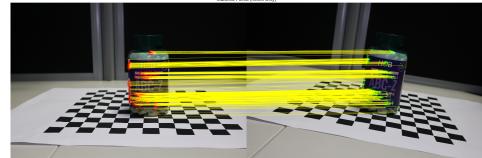


Figure 6. Keypoint Matching with inliers only.

**Task 2.a** The fundamental matrix contains the relationship between two images taken from different perspectives. The matrix estimated for the given pair of images is:

$$F = \begin{bmatrix} 7.077 \times 10^{-8} & 4.094 \times 10^{-7} & -8.036 \times 10^{-4} \\ -4.081 \times 10^{-7} & 7.066 \times 10^{-8} & 1.147 \times 10^{-3} \\ 3.185 \times 10^{-4} & -1.377 \times 10^{-3} & 1.000 \end{bmatrix}$$

The estimated fundamental matrix reveals slight shifts and rotations paired with notable translation effects in the image sequence. ORB features enable efficient and reliable matching, with RANSAC refining the process to consistent, accurate correspondences. The epipolarLine function uses the fundamental matrix to compute the position of these lines, which are then drawn using the lineToBorderPoints function to extend them to the edges of the images.

The alignment of inliers with their respective epipolar lines in the chosen figures confirms the fundamental matrix's effectiveness. These observations support the practical application of the matrix to understand scene geometry and inform subsequent 3D reconstruction.

**Task 2.b** Figures 7,8,9 show the key points with corresponding epipolar lines, the epipolar point, and the vanishing point with the horizon respectively.

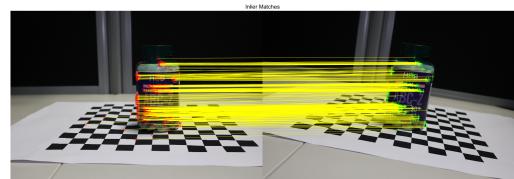


Figure 7. Keypoint Matching with inliers only. The number of inliers is 697, while the number of outliers is 500.

Figure 9 illustrates the detected vanishing points with the horizon line emphasized by a connecting green line. This is achieved by first applying a Sobel filter to the preprocessed image, enhancing the visibility of structural boundaries. The subsequent application of the Hough Transform algorithm identifies and extrapolates the prominent linear features within the image. Intersections of these lines are then computed and marked. The delineation of the horizon is represented by the line connecting these intersections, underscoring the scene's depth and perspective.

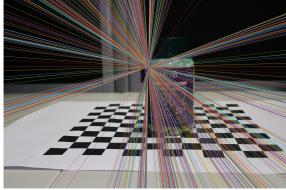


Figure 8. HG Image with epipoles inside the image

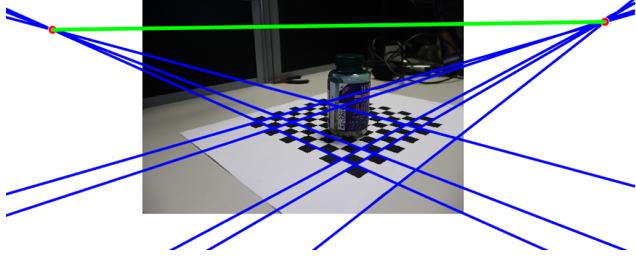


Figure 9. Vanishing Point

**Task 3** Experiments were carried out by manually adding noise to the keypoint matches and observing how the number of inliers changes. The result could be found in figure 10. The graph suggests that there's a certain threshold of noise standard deviation (around 1) beyond which the number of inliers plateaus and does not significantly decrease.

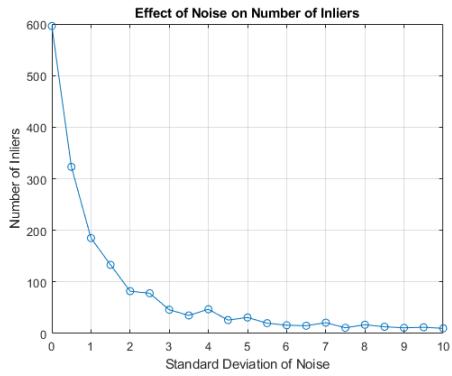


Figure 10. Effect of Noise on Number of Inliers.

## 5. 3D geometry

**Task 1.a** The method from last task was used to estimate the fundamental matrix. Following this, uncalibrated stereo rectification was implemented, utilizing the fundamental matrix along with the inlier feature points to warp and align the images along the epipolar lines. The rectified images with epipolar lines are shown in 11. The rectification process successfully aligned the images, as can be seen from the parallel epipolar lines.

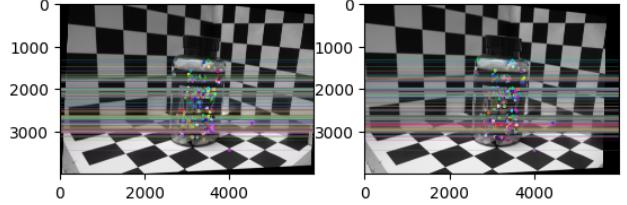


Figure 11. The stereo rectification aligned the images along the epipolar lines.

**Task 1.b** A disparity map was computed using the Semi-Global Matching (SGM) algorithm, which provided an estimation of the depth for each pixel. This depth estimation revealed the varied spatial relationships within the scene, translating them into a disparity map highlighting the observed environment's three-dimensional structure.

The depth map displayed in figure 12 adeptly outlines the scene's object depths. In this representation, a color gradient transitions from warm to cool tones to signify distance—contrary to intuition, here warm colors mark objects that are farther from the camera, whereas cool colors highlight those closer. This reversal in color coding effectively emphasizes the medicine bottle's contour and spatial information, affirming the reliability of our image rectification and depth quality.

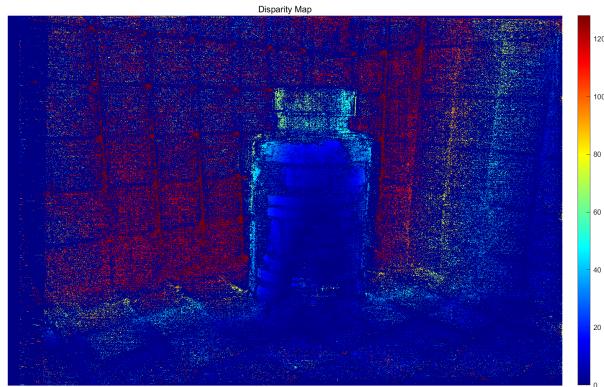


Figure 12. The disparity map of the object

## 6. Appendix

**Task 1 FD and HG image** The figures 13, 14, 15, 16 below display a portion of data collected for the HG and FD images used in this coursework. We employed a professional camera and stabiliser to minimize errors from mobile phone correction algorithms and human shake.

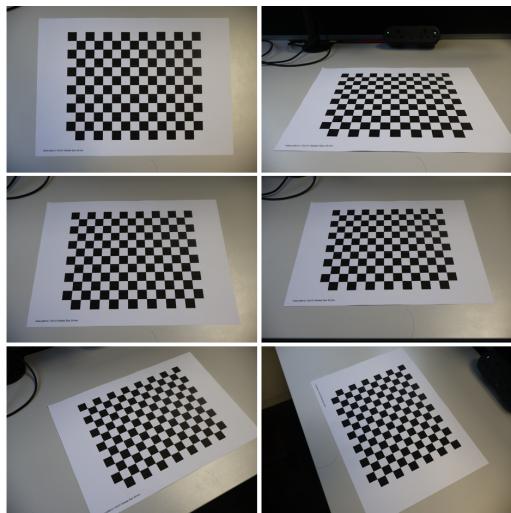


Figure 13. FD images without object



Figure 15. HG images 1

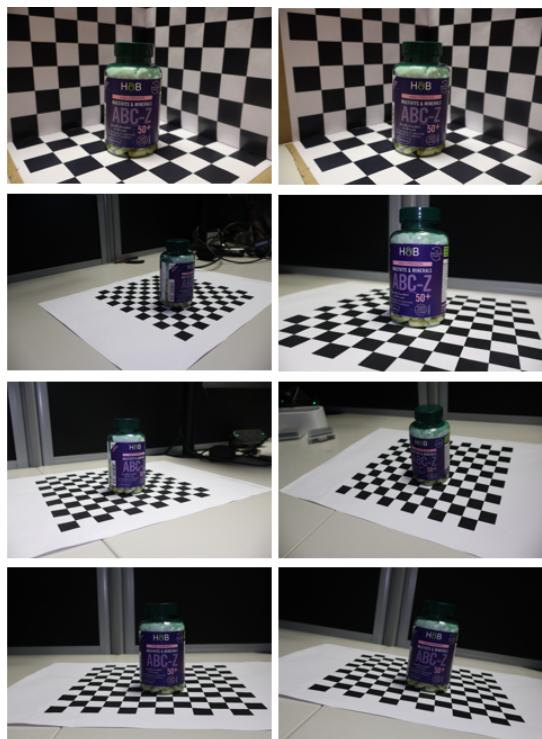


Figure 14. FD images with object

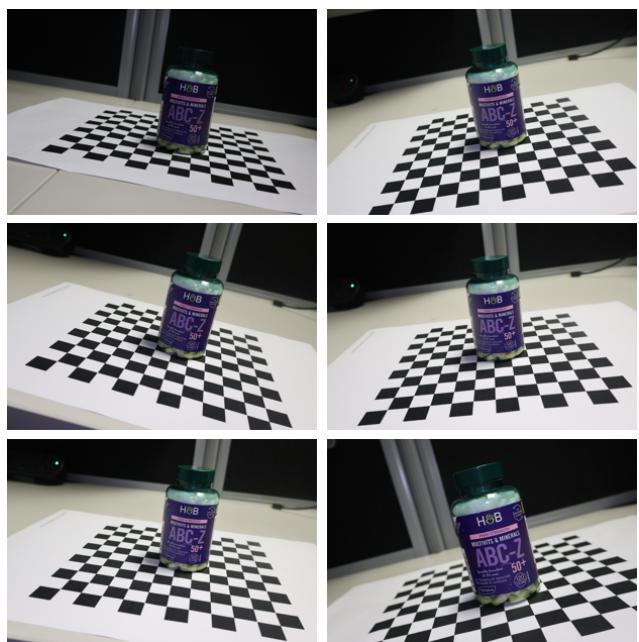


Figure 16. HG images 2