

3D Vision for Multiple or Moving Cameras

Lab Evaluation

Ahmet Furkan Ün

INTRODUCTION

I. OBTENTION OF THE INTRINSIC PARAMETERS OF A CAMERA

In this section, I calibrate the internal parameters of a chosen camera using two different calibration patterns: a checkerboard displayed on a screen and a custom-designed star pattern. I follow Zhang's calibration method, utilizing homographies and multiple views to compute the intrinsic parameters matrix.

A. Camera Calibration using Checkerboard Pattern

To simulate a physical checkerboard, I used a 1080×1080 pixel checkerboard image enlarged on a full-screen monitor. The physical width of the projected checkerboard image was measured as **150 mm**. I captured 7 images from different views and distances. Each captured image has a resolution of **3472x3472 pixels**.

The internal parameters matrix **A** computed using Zhang's method is:

$$\mathbf{A}_{\text{checkerboard}} = \begin{bmatrix} 3434.09 & -30.29 & 1750.04 \\ 0 & 3437.10 & 1691.91 \\ 0 & 0 & 1 \end{bmatrix}$$

Quantitative Analysis:

- Pixel Squareness:** The focal lengths in x and y directions are $f_x = 3434.09$ and $f_y = 3437.10$, respectively. The relative difference is:

$$\frac{f_x}{f_y} \approx 0.999 \Rightarrow \text{pixels are essentially square.}$$

- Principal Point Offset:** The image center is at $(1736, 1736)$ assuming a 3472×3472 resolution. The estimated principal point is at $(1750.04, 1691.91)$, which is offset by $\Delta x = 14.04$ and $\Delta y = -44.09$. This shows small deviation from image center but it is highly accurate.
- Axis Orthogonality:** The skew coefficient is $s = -30.29$, which is small relative to the focal lengths. This suggests the axes are close to orthogonal.

B. Camera Calibration using Star Pattern

In this calibration setup, I used a custom-designed star pattern displayed on a computer screen. Similar to the checkerboard pattern, this calibration target was also generated digitally. I defined the real-world coordinates of

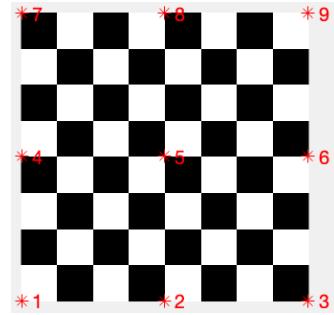


Fig. 1. Calibration Pattern

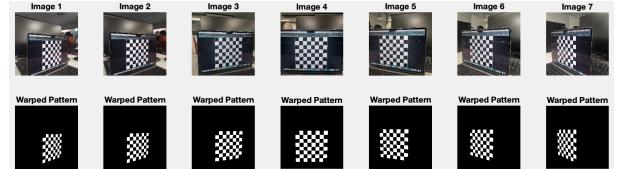


Fig. 2. Homography alignment results between each input image and the reference checkerboard pattern.

ten selected points manually, based on physical measurements of the star pattern in a unit image. These coordinates were specified within a custom MATLAB function, `get_real_points_star`, which returns the 2D coordinates in millimeters and a rendered image of the star.

The calibration was performed using 7 captured images of the screen displaying the star pattern, processed similarly to the checkerboard calibration. The physical width of this pattern on the screen was **150 mm**. The image resolution for all captures was **3472×3472 pixels**.

The internal calibration matrix obtained through Zhang's method was:

$$\mathbf{A}' = \begin{bmatrix} 3667.54 & -12.33 & 1924.86 \\ 0 & 3638.18 & 1807.76 \\ 0 & 0 & 1 \end{bmatrix}$$

Quantitative Analysis:

- Pixel Squareness:** The focal lengths in x and y directions are $f_x = 3667.54$ and $f_y = 3638.18$, respectively. The relative difference is:

$$\frac{f_x}{f_y} \approx 1.008 \Rightarrow \text{pixels are essentially square.}$$

- Principal Point Offset:** The image center is at $(1736, 1736)$ assuming a 3472×3472 resolution. The estimated principal point is at $(1924.86, 1807.76)$, which

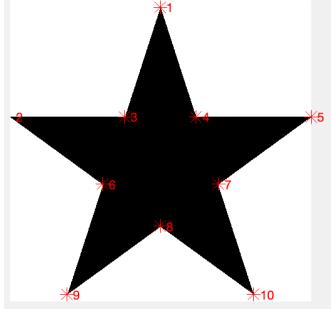


Fig. 3. Calibration Pattern

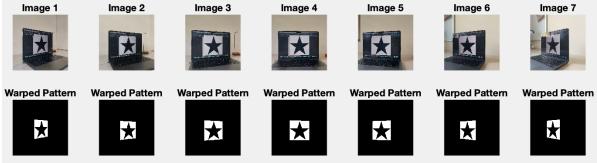


Fig. 4. Homography alignment results between each input image and the reference star pattern.

is offset by $\Delta x = 188.86$ and $\Delta y = 71.76$. This shows a bigger deviation from image center but still not that bad.

- **Axis Orthogonality:** The skew coefficient is $s = -12.33$, which is small relative to the focal lengths. This suggests the axes are close to orthogonal.

Comparison and Discussion: Both calibrations produce consistent estimates of the internal parameters, though with slight deviations, possibly due to differences in measurement noise or point accuracy in the custom pattern. These results reinforce the flexibility of Zhang's method when reliable real-world correspondences are available.

The differences in the matrices are theoretically expected and practically acceptable. However, the intrinsic matrix obtained from the checkerboard will be used in the rest of this project to ensure consistency.

II. DETECTION AND MATCHING OF KEYPOINTS ACROSS VIEWS

A. Scene Capture and Dataset Description

I constructed a 3D scene consisting of four everyday textured objects: a toothpaste tube, a calculator, a matchbox, and a powerbank. The chosen objects exhibit varying surface geometries and textures to challenge both feature detection and descriptor robustness.

The scene was captured under three lighting conditions:

- **Light 1 (Fully Bright):** Uniform lighting with minimal shadows.
- **Light 2 (Side Shadows):** Room light turned off; illumination from one side, introducing strong directional shadows.
- **Light 3 (Low Light):** Weak ambient light to challenge keypoint detection under poor visibility.

For each lighting setup, four different views of the scene were captured by rotating the camera around the scene. Camera poses were kept consistent across lighting conditions to facilitate comparative evaluation.



Fig. 5. Mosaic of the four captured views under different light conditions.

B. Evaluation of Detection, Description, and Matching Setups

I selected SIFT for keypoint detection and description based on its proven robustness to scale and illumination changes. I evaluated four different parameter sets under three lighting conditions:

TABLE I
EVALUATED SIFT PARAMETER SETS

Set	ContrastThreshold	EdgeThreshold	Sigma
1	0.01	10	1.6
2	0.03	10	1.6
3	0.01	5	1.2
4	0.02	15	2.0

I tested 3 consecutive combinations of view pairs (1-2, 2-3 and 3-4) under each lighting condition using the four parameter sets, resulting in 36 configurations. For each configuration, I recorded:

- Number of matched keypoints
- Inlier counts for homography and fundamental matrix
- Homography reprojection error
- Mean Sampson error

Sampson Error: The Sampson error provides a first-order approximation of geometric error in fundamental matrix estimation:

$$E_S = \frac{(x_2^\top F x_1)^2}{(F x_1)_1^2 + (F x_1)_2^2 + (F^\top x_2)_1^2 + (F^\top x_2)_2^2} \quad (1)$$

where x_1 and x_2 are corresponding points in homogeneous coordinates and F is the fundamental matrix. This error

measures the geometric distance from a point to its epipolar line, normalized by the gradient, and is more robust than direct algebraic error.

Quantitative Summary: Table II, Table III, and Table IV show the performance results across different lighting conditions.

Discussion on Scene and Method Performance:

- Under Light 1 (bright and even), SIFT performs robustly across all parameter sets. Set 1 gives the best balance between high match count and reasonable error.
- Under Light 2 (directional), shadows degrade feature reliability, but performance remains acceptable. Sets 2 and 4 provide reasonable compromises.
- Under Light 3 (low light), all configurations degrade, especially in Sampson error. Set 4 still performs slightly better in match count.

Homography and Fundamental Matrix Quality:

- Homographies show low reprojection errors across all conditions, indicating good inlier selection.
- Sampson error is highly sensitive to noise and outliers. Large errors in Light 3 suggest difficulty in accurate fundamental matrix estimation.
- The scene's texture and structure (flat surfaces with strong edges) help SIFT, but shadows and illumination drop performance.

C. Selected Best Pair and Setup

Among the evaluated setups, **Parameter Set 2 under Light 1** condition yielded the best compromise between inlier ratio and reprojection error, while also exhibiting the lowest Sampson error across scenes. For this reason, later this parameter set and light condition will be used to re-construct the 3D scene.

As seen in Table V, the best pair of views for this light condition and parameter set is 2-3.



Fig. 6. Visual correspondences for the best performing pair (Scene 2-3, Light 1, Param Set 2).

• Estimated Homography:

$$H = \begin{bmatrix} 0.9167 & -0.0293 & -0.0000 \\ 0.0106 & 0.9706 & 0.0000 \\ 3.8451 & 11.6641 & 1.0000 \end{bmatrix}$$

• Warped Image Result:



Fig. 7. Visual correspondences for the best performing pair (Scene 2-3, Light 1, Param Set 2).



Fig. 8. Warped image using the estimated homography.

• Estimated Fundamental Matrix:

$$F = \begin{bmatrix} 9.28 \times 10^{-9} & -3.14 \times 10^{-7} & 0.00016 \\ 9.59 \times 10^{-7} & -1.77 \times 10^{-8} & -0.0168 \\ -0.0010 & 0.01587 & 0.99973 \end{bmatrix}$$

• Epipolar Geometry Visualization (vgg-gui_F.m):



Fig. 9. Screenshot from vgg-gui_F.m displaying the epipolar lines and correspondences.

III. 3D RECONSTRUCTION AND CALIBRATION

In this section, I use the intrinsic parameters of the camera and the feature point matches obtained in Section ?? to construct a 3D reconstruction of the captured scene. The reconstruction pipeline follows the strategies explored in

TABLE II
QUANTITATIVE RESULTS FOR LIGHT 1 CONDITION

Param Set	Avg Matches	Avg Inliers (F)	Reproj. Error (H)	Mean Sampson Error
1	1076.33	538.33	0.569	1111.73
2	722.67	361.67	0.564	144.05
3	592.33	296.33	0.538	1722.83
4	1046.00	523.33	0.564	727.97

TABLE III
QUANTITATIVE RESULTS FOR LIGHT 2 CONDITION

Param Set	Avg Matches	Avg Inliers (F)	Reproj. Error (H)	Mean Sampson Error
1	1261.67	631.33	0.520	1983.56
2	907.67	454.33	0.467	809.12
3	729.67	365.00	0.517	3897.32
4	1353.33	677.33	0.529	1246.99

TABLE IV
QUANTITATIVE RESULTS FOR LIGHT 3 CONDITION

Param Set	Avg Matches	Avg Inliers (F)	Reproj. Error (H)	Mean Sampson Error
1	452.67	226.67	0.680	3602.06
2	332.00	166.67	0.625	3365.76
3	273.00	136.67	0.660	9877.65
4	521.67	261.67	0.691	2742.30

TABLE V
RESULTS FOR LIGHT 1 CONDITION AND PARAM SET 2

Pair of Views	Num. Matches	Num. Inliers (F)	Reproj. Error (H)	Mean Sampson Error
1-2	440	220	0.6420	40.8932
2-3	853	427	0.5278	29.8190
3-4	875	438	0.5217	361.42

Unit III of the course. Our implementation consists of the following steps:

A. Consistent N-View Matching

I first compute consistent feature point correspondences across four selected views, one from each camera position under the first lighting condition. Based on the evaluation in Section 1, the SIFT detector and descriptor provided the most reliable and repeatable matches under appearance changes. I thus extract SIFT keypoints and descriptors and apply the provided `n_view_matching` function to obtain consistent correspondences.

Figure 10 shows the detected interest points used for the N-view matching across the four views. The matching is performed with a descriptor matching ratio of 0.5 using the SSD metric.



Fig. 10. Detected interest points for N-view matching across four views.

B. Initial Projective Calibration from Two Views

I estimate the Fundamental matrix and an initial projective reconstruction using views 1 and 4. Figure 11 illustrates the

matched interest points between these views. Using these matches, I apply the `MatFunProjectiveCalib` function to compute the projection matrices and a set of 3D points.



Fig. 11. Matched keypoints used for Fundamental matrix estimation between view 1 and view 4.

The mean reprojection error obtained using the 8-point algorithm is **0.2662** pixels. The corresponding histogram is shown in Figure 12.

C. Projective Bundle Adjustment

To refine the initial reconstruction, I resection the intermediate cameras (views 2 and 3) using the previously triangulated 3D points and perform projective bundle adjustment over all views.

The reprojection error after resectioning increases significantly to **35.3106** pixels, likely due to inconsistencies

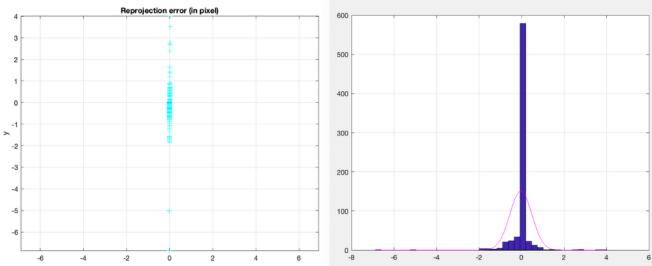


Fig. 12. Histogram of reprojection error for the initial projective calibration (2-view).

between the estimated 3D points and intermediate camera viewpoints. After projective bundle adjustment, the reprojection error improves to **22.5849** pixels. The respective histograms are shown in Figures 13 and 14.

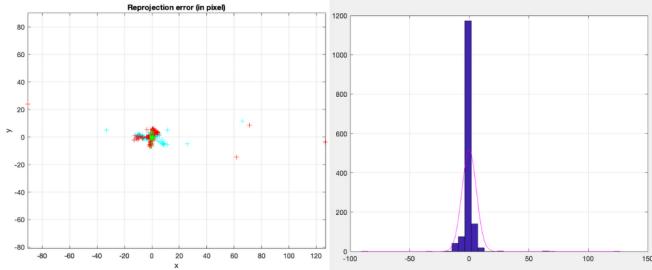


Fig. 13. Histogram of reprojection error after resectioning.

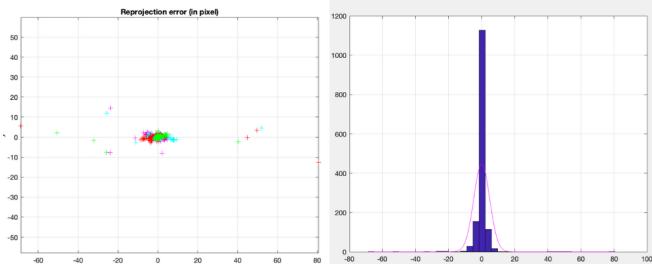


Fig. 14. Histogram of reprojection error after projective bundle adjustment.

Despite the improvement, the relatively high error post-adjustment suggests the presence of inconsistencies in point correspondences or unmodeled scene deformation, especially under varying lighting and specular reflections.

D. Euclidean Upgrade via Essential Matrix Factorization

I compute the Essential matrix using the fundamental matrix and the camera intrinsic matrix from Section ??:

$$E = K^T F K \quad (2)$$

Using the `factorize_E` function, I retrieve four possible combinations of rotation and translation. A cheirality check identifies the valid solution, ensuring positive depth in both views.

I then upgrade the projection matrices to Euclidean form, setting the first camera as the reference and computing the remaining using resectioning.

E. Final Evaluation and Visualization

The reprojection error after Euclidean reconstruction increases significantly to **190.5258** pixels. This dramatic increase suggests a drift in the estimated 3D structure due to accumulated geometric error, sensitivity to point accuracy, or incorrect correspondences. The histogram is shown in Figure 15.

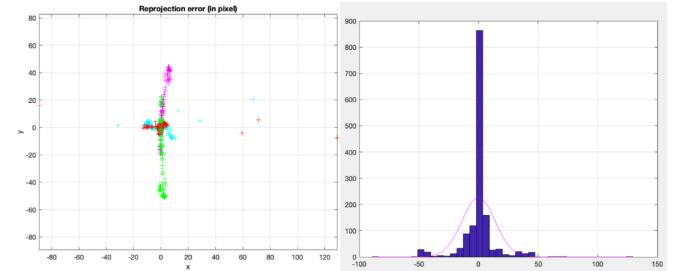


Fig. 15. Histogram of reprojection error after Euclidean reconstruction.

Finally, I visualize the reconstructed 3D point cloud and camera centers using the `draw_scene` function. An overview of front and top views provided in Figures 16 and 17.

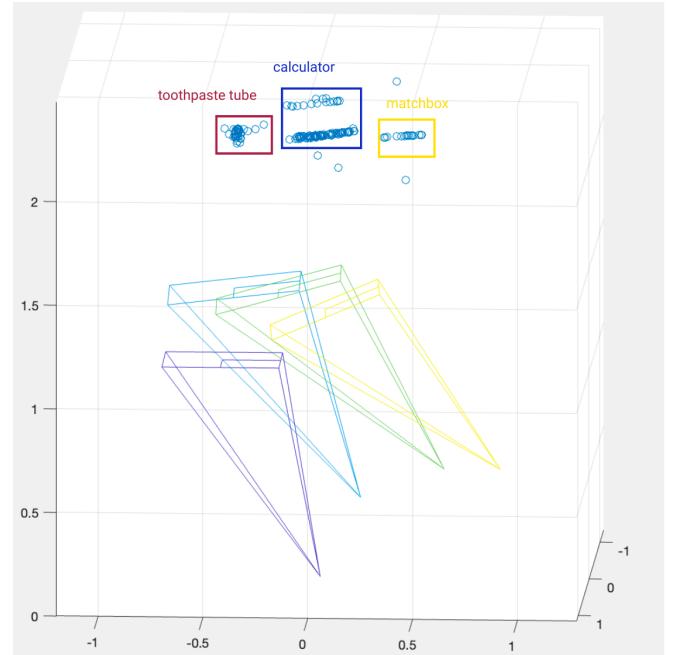


Fig. 16. 3D scene from top

F. Discussion

Figure 16 and Figure 17 present the final 3D reconstruction obtained from the consistent multi-view matching and projective-to-Euclidean pipeline. Notably, the reconstructed point cloud includes only three of the four objects in the scene: the toothpaste, the calculator box, and the matchbox. The powerbank is entirely missing from the reconstruction. This omission is attributed to the absence of matched keypoints on the powerbank across the selected views. While

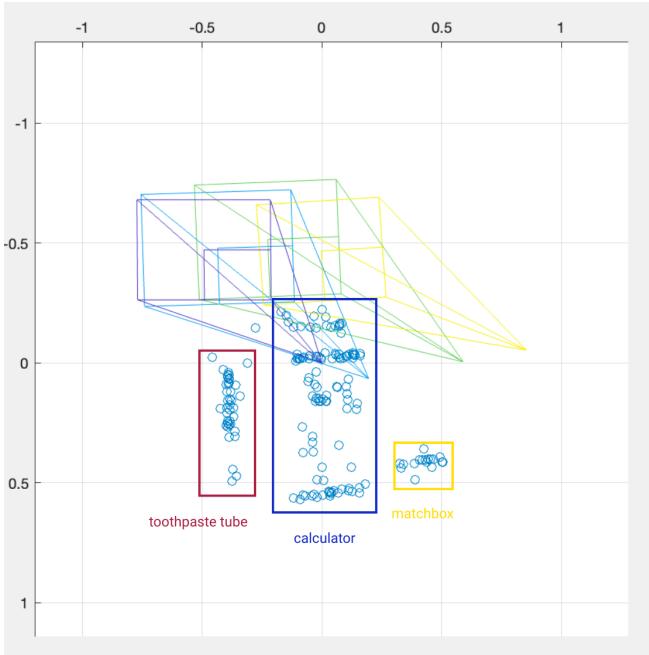


Fig. 17. 3D scene from front

this could potentially be addressed by relaxing the descriptor matching parameters (e.g., increasing the ratio threshold in Lowe's ratio test), such changes would likely reduce the overall precision and introduce unreliable correspondences. Therefore, I chose to preserve match quality over completeness in this case.

Interestingly, the reconstruction successfully captures the structural detail of the calculator box, including its retail display feature (a flat, 2D extension at the top, typically used for hanging the product in stores). This thin, planar section is visible in both the original images and the reconstructed geometry, highlighting the fidelity of the projection and triangulation steps. The reconstruction's ability to preserve such fine and thin elements demonstrates robustness in recovering weakly textured yet geometrically consistent parts of the scene.

However, reprojection errors are high in the later stages, particularly after the Euclidean upgrade. This may be attributed to:

- Ambiguities in feature correspondences due to textureless regions or lighting changes.
- Potential mismatches propagated during N-view matching.
- Errors introduced in triangulation and resectioning steps.
- Limitations of the intrinsic matrix estimation.

Further improvements could involve better outlier rejection in matching, multi-scale keypoint detection, or the use of robust bundle adjustment with outlier-aware loss functions.