Project 2: Building Built in Minutes: Phase 1

Dhrumil Kotadia
Robotics Engineering Department
Worcester Polytechnic Institute
Worcester, Massachusetts
Using One Late Day

Dhiraj Kumar Rouniyar
Robotics Engineering Department
Worcester Polytechnic Institute
Worcester, Massachusetts
Using One Late Day

Introduction

In this report, we will provide a comprehensive analysis of our implementation of the monocular camera structure from motion (SFM). Our SFM pipeline consists of several key steps, including feature matching with SIFT, estimating the fundamental matrix with epipolar constraints, deriving the essential matrix from the fundamental matrix, estimating the camera pose from the essential matrix, triangulating points linearly and nonlinearly while checking for chierality constraints, solving for perspective-n-points using linear and nonlinear optimization, and performing bundle adjustment for all input images.

A. *Feature Matching*

We begin by establishing feature matches between each pair of images. We employed the Scale-Invariant Feature Transform (SIFT) algorithm for this task. Before identifying feature matches, we first determined the camera's intrinsic matrix via calibration and rectified the raw images to correct for distortion. Figure 1 illustrates the feature matches discovered between the first and second images.

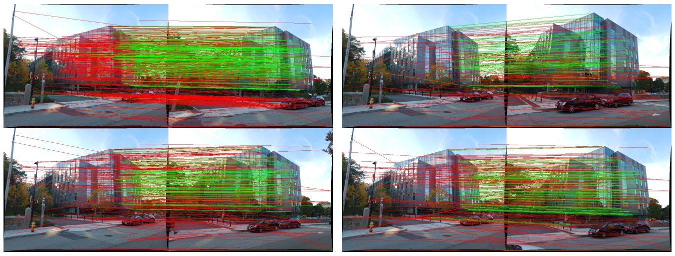

Fig. 1. Feature Matching and Outlier Rejection

B. *Fundamental Matrix Calculation*

Next, we estimate the fundamental matrix(8 point algorithm with SVD cleanup), represented as F, which is a 3x3 matrix with a rank of 2. It serves to establish the relationship between corresponding sets of points in two images captured from different viewpoints, also known as stereo images. The fundamental matrix can be calculated by

$$\begin{bmatrix} x_i' & y_i' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

To ensure the selection of only reliable correspondences, RANSAC procedure was conducted. This process aimed to isolate pure inliers from the sets of correspondences. Subsequently, these refined points were utilized to compute a more precise Fundamental Matrix. The resulting fundamental matrix is as follows:

$$F = \begin{bmatrix} 3.2126e-08 & -3.1944e-05 & 1.3178e-02 \\ 3.4375e-05 & 3.0271e-06 & -3.4118e-02 \\ -1.5041e-02 & 3.2359e-02 & 1 \end{bmatrix}$$

The green matches in Figure 1 shows the matches after 8-point algorithm RANSAC.

C. *Essential Matrix and camera pose Calculation:*

Using the estimated fundamental matrix found in the previous step, we then calculated the Essential Matrix between an image pair as $E = K^T F K$, where K is the camera intrinsic matrix. However, since there might still be noise in the calculated Essential Matrix, we have to enforce the epipolar constraint by reconstructing the Essential Matrix with singular value decomposition with rank reduction. $E = UDV^T$ then substitute D with

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The essential matrix generated is obtained as follows:

$$E = \begin{bmatrix} 0.0046 & -0.6339 & 0.1146 \\ 0.6821 & 0.0513 & -0.7147 \\ -0.1639 & 0.7606 & 0.0260 \end{bmatrix}$$

Using this, we now have 4 different combinations of R and C giving 4 different camera poses with 2 values for each - R and C. Out of these 4 poses, there is only one valid pose for which, the world points lie in front of both of the cameras(chirality condition). Using this, we obtained the actual camera pose.

D. *Linear and Non Linear Triangulation with Chirality:*

After decomposing the essential matrix, linear triangulation with Chirality constraints was employed to eliminate three implausible camera pose configurations. Upon determining the actual camera pose configuration, we conduct non-linear optimization to minimize the reprojection error and enhance the estimation of the features' world coordinates. Figure 2 compares the outcomes of linear and non-linear triangulation, highlighting the considerable enhancement in accuracy achieved through non-linear triangulation.

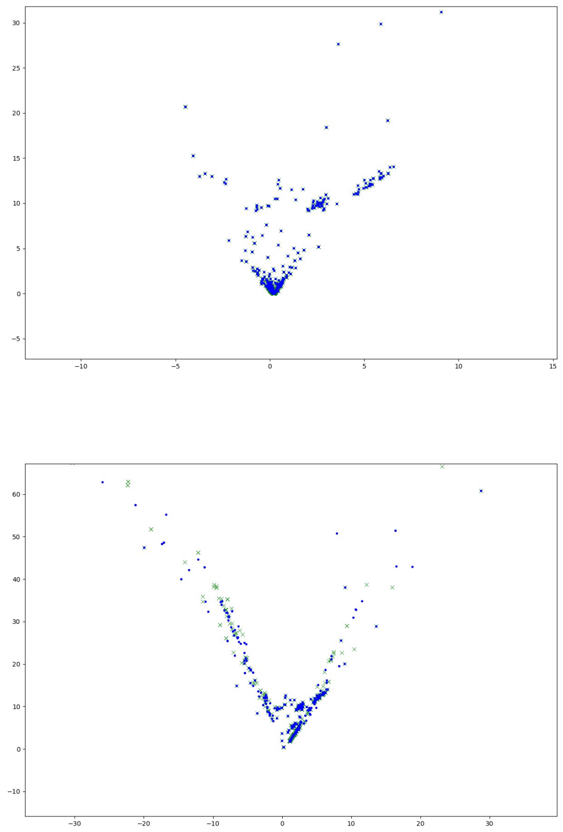

Fig. 2. Linear and Non-Linear Triangulation for Images 2_3 and 2_1

E. Perpective-n-Points:

To incorporate features from a third or more images into the constructed scene, we initially employed PnP RANSAC to eliminate outlier features from the third image. This process aimed to minimize the algebraic error between the measured and reprojected features. Utilizing the remaining inlier features, we computed the estimated camera pose for the third image.

However, due to the inherent nonlinearity in the division and reprojection within our system, the estimated camera poses are prone to inaccuracies. To refine these estimates, we utilized the estimated camera poses from linear PnP as initial guesses and conducted non-linear optimization. This optimization minimized the geometric reprojection error, resulting in more precise estimations of the camera poses.

F. Bundle Adjustment:

The next step is to refine the poses and 3D points simultaneously through bundle adjustment. Utilizing the camera poses and 3D points obtained from the previous steps as initial estimates, we minimize the reprojection error using least_squares. Figure 3 displays the outcome of the bundle adjustment process.

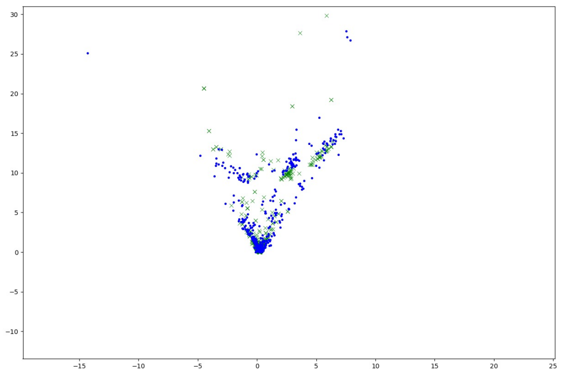

Fig. 3. Bundle Adjustment

All the reprojection errors are mentioned in the following table:

TABLE I
REPROJECTION ERRORS

Method	**Reprojection Error**
Linear Triangulation	26.04
Nonlinear Triangulation	25.7
Linear PnP	2720.6
Nonlinear PnP	381.09