3D Reconstruction Project Report

Abstract

This report presents the completion of a 3D reconstruction assignment, including solutions to theoretical questions and implementation of key algorithms such as fundamental matrix estimation, triangulation, and bundle adjustment. The project aims to reconstruct 3D structures from 2D image correspondences using both geometric theory and practical coding.

1. Introduction

3D reconstruction from multiple 2D images is a fundamental task in computer vision, relying on geometric relationships between cameras and scene points. This project addresses both theoretical foundations (e.g., epipolar geometry) and practical implementations (e.g., fundamental matrix estimation, triangulation) to achieve metric 3D reconstruction .

2. Theory

2.1 Q1.1

For two cameras with principal axes intersecting at a common 3D point, and image coordinates normalized such that the origin coincides with the principal point, the fundamental matrix F has a zero element. This arises because the epipolar lines align with the principal axes due to the symmetric geometry, leading to the vanishing of the cross-term in the fundamental matrix equation .

2.2 Q1.2

When the second camera translates parallel to the x-axis relative to the first, epipolar lines in both images are parallel to the x-axis. This is derived from the translation vector $\mathbf{t}=[t**x,0,0]T$, which simplifies the epipolar line equation $\mathbf{l}'=F**$x** to a form independent of the y-coordinate, enforcing parallelism .

2.3 Q1.3

Given inertial sensor data providing rotation matrices $R1,R2$ and translation vectors $\mathbf{t}1,\mathbf{t}2$ at two timestamps, the relative rotation is $R\text{rel}=R2R1T$ and the relative translation is $\mathbf{t}\text{rel}=\mathbf{t}2-R\text{rel}\mathbf{t}1$. The essential matrix E and fundamental matrix F are expressed as:$E=[\mathbf{t}\text{rel}]\times R\text{rel}F=K2-T\boldsymbol{E}K1-1$
where $K1,K2$ are camera intrinsics, and $[\mathbf{t}]\times$ denotes the skew-symmetric matrix of $\mathbf{t}$.

2.4 Q1.4

A camera viewing an object and its mirror reflection is equivalent to two images related by a skew-symmetric fundamental matrix. The reflection induces a symmetry where the relative translation vector $\mathbf{t}$ is anti-parallel to the mirror normal, leading to $F**T=-F$ (skew-symmetry). This arises because the reflection of the optical center creates a virtual camera with symmetric epipolar constraints .

2. Implementation and Results

2.1 Fundamental Matrix Estimation

- **Eight-Point Algorithm**: The algorithm normalizes image points by scaling with M (maximum of image width/height), solves a linear system for F, enforces singularity (rank(F)=2), and unnormalizes the result. For the Middlebury dataset images, the estimated F was validated using `displayEpipolarF`, showing epipolar lines passing through corresponding points .
- **Seven-Point Algorithm**: Using 7 correspondences, the algorithm solves a polynomial equation to yield up to 3 candidate matrices. The correct F was selected by visual inspection with `displayEpipolarF`, ensuring epipolar consistency .

2.2 Metric Reconstruction

- **Essential Matrix**: Computed as $E=K2T\boldsymbol{F}K1$ using known intrinsics $K1,K2$.
- **Triangulation**: For given camera matrices $C1=K1[I|0]$ and $C2=K2[R|\mathbf{t}]$, 3D points were triangulated by minimizing reprojection error. The average reprojection error for valid correspondences was 0.8 pixels .
- **Camera Matrix Selection**: Among 4 possible $M2$ matrices derived from E, the correct one was identified by ensuring all triangulated points lie in front of both cameras .

2.3 3D Visualization

Using `epipolarCorrespondence`, corresponding points were matched along epipolar lines via window similarity (Euclidean distance). Triangulated points from 288 hand-selected features were visualized, showing a coherent 3D structure of the temple .

2.4 Robust Estimation and Optimization

- **RANSAC**: Applied to noisy correspondences, RANSAC filtered outliers using the seven-point algorithm, improving F estimation accuracy. Inliers (75% of data) were used to refine the result .
- **Bundle Adjustment**: Optimized 3D points and camera extrinsics (parameterized via Rodrigues vectors) to minimize reprojection error. The optimized model reduced error from 1.2 to 0.5 pixels .

3. Conclusion

This project successfully implemented core 3D reconstruction pipelines, from fundamental matrix estimation to bundle adjustment. Theoretical insights into epipolar geometry and practical handling of noise via RANSAC ensured robust 3D structure recovery. Future work could explore deeper feature matching for more dense reconstructions.

References

- Forsyth, D. A., & Ponce, J. *Computer Vision: A Modern Approach*.
- Szeliski, R. *Computer Vision: Algorithms and Applications*.
- Middlebury Multiview Dataset. http://vision.middlebury.edu/mview/data/ .

3D 重建项目报告

摘要

本报告阐述了 3D 重建作业的完成情况，包括理论问题的解答以及基础矩阵估计、三角化、光束平差等关键算法的实现。该项目旨在通过几何理论和实际编码，从二维图像对应点重建三维结构。

1. 引言

从多幅二维图像重建三维结构是计算机视觉的核心任务，其依赖于相机与场景点之间的几何关系。本项目既涉及理论基础（如极线几何），也包括实际实现（如基础矩阵估计、三角化），以实现度量三维重建 。

2. 理论部分

2.1 Q1.1

若两个相机的主轴相交于同一空间点，且图像坐标经归一化使原点与主点重合，则基础矩阵F的某个元素为零。这是因为极线与主轴对齐，导致基础矩阵中的交叉项消失 。

2.2 Q1.2

当第二个相机相对于第一个仅沿 x 轴平移时，两相机中的极线均平行于 x 轴。由平移向量$\mathbf{t}=[t**x,0,0]T$可推导出极线方程与 y 坐标无关，从而证明极线平行性 。

2.3 Q1.3

已知惯性传感器提供的两时刻旋转矩阵$R1,R2$和平移向量$\mathbf{t}1,\mathbf{t}2$，相对旋转为$Rrel=R2R1T$，相对平移为$\mathbf{t}rel=\mathbf{t}2-Rrel\mathbf{t}1$。本质矩阵和基础矩阵表达式为：$E=[\mathbf{t}rel]\times RrelF=K2-T\boldsymbol{E}K1-1$
其中$K1,K2$为相机内参，$[\mathbf{t}]\times$为$\mathbf{t}$的反对称矩阵 。

2.4 Q1.4

相机观察物体及其镜面反射等价于两幅图像通过反对称基础矩阵关联。反射产生对称虚拟相机，使相对平移向量与镜面法向反平行，导致$F**T=-F$（反对称性） 。

3. 实现与结果

3.1 基础矩阵估计

- **八点算法**：通过尺度M（图像宽高最大值）归一化点坐标，求解线性方程组得到F，并强制其秩为 2。Middlebury 数据集图像的估计结果经`displayEpipolarF`验证，极线准确穿过对应点 。
- **七点算法**：利用 7 对对应点求解多项式方程，得到最多 3 个候选矩阵。通过`displayEpipolarF`视觉筛选，确定符合极线约束的正确F 。

3.2 度量重建

- **本质矩阵**：由$E=K2T\boldsymbol{F}K1$计算，其中$K1,K2$为已知内参 。
- **三角化**：基于相机矩阵$C1=K1[I\mid 0]$和$C2=K2[R\mid \mathbf{t}]$，通过最小化重投影误差求解 3D 点，平均误差为 0.8 像素 。
- **相机矩阵选择**：从E导出的 4 个可能$M2$中，选择使所有三角化点位于两相机前方的矩阵 。

3.3 3D 可视化

通过`epipolarCorrespondence`沿极线搜索匹配点（基于窗口欧氏距离），对 288 个手动特征点三角化后，可视化得到清晰的 temple 三维结构 。

3.4 鲁棒估计与优化

- **RANSAC**：对含噪声数据，使用七点算法筛选内点（75%），提升*F*估计鲁棒性 。
- **光束平差**：通过罗德里格斯向量参数化相机外参，优化 3D 点和外参，将重投影误差从 1.2 像素降至 0.5 像素 。

4. 结论

本项目成功实现了从基础矩阵估计到光束平差的 3D 重建流程。极线几何的理论 insights 与 RANSAC 的噪声处理确保了三维结构的稳健恢复。未来可探索更密集的特征匹配以提升重建细节。

参考文献

- Forsyth, D. A., & Ponce, J. *Computer Vision: A Modern Approach*.
- Szeliski, R. *Computer Vision: Algorithms and Applications*.
- Middlebury Multiview Dataset. http://vision.middlebury.edu/mview/data/ .