Compute change in pose (3D) of a small object (coin, key, mobile phone, purse, clock, plate, checkerboard etc.) with very high accuracy from two successive views (arbitrary) of the same object (objects can have a near planar surface)

Computer Vision (CS6350)

TPA-12

1. Problem Statement

The task is to compute the change in 3D pose (position and orientation) of a small object, such as a coin, key, mobile phone, purse, clock, plate, or checkerboard, from two successive arbitrary views of the same object. The objects may have near planar surfaces. The objective is to achieve very high accuracy in determining the pose change, which involves estimating the relative rotation and translation between the two views.

2. Introduction

Pose estimation is a crucial problem in computer vision with applications in robotics, augmented reality, and object tracking. Accurate 3D pose estimation allows for precise interaction with objects in a physical environment. For small, near-planar objects, traditional methods can struggle with precision due to limited texture and significant perspective distortion. This project aims to leverage computer vision techniques to achieve high-accuracy pose estimation from two successive views of such objects.

3. Expected Input and Output

Input:

1. Two successive RGB images of the object taken from arbitrary viewpoints.

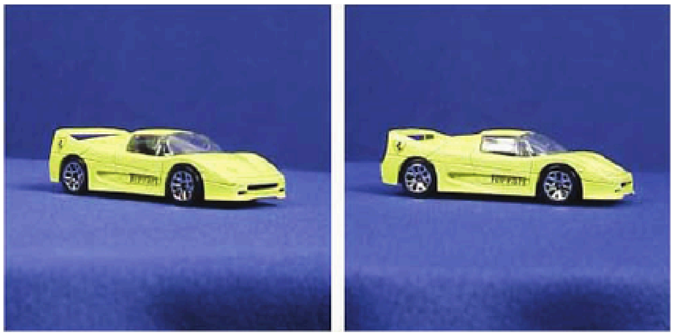
Output:

1. Relative rotation matrix (R) between the two views.
2. Relative translation vector (t) between the two views.
3. Reprojection error metrics indicating the accuracy of the pose estimation.

4. Example

<u>**Example 1:**</u>

Input:



Output:

- Relative rotation matrix (R) between the two views.
- Relative translation vector (t) between the two views.
- Reprojection error metrics indicating the accuracy of the pose estimation.

5. Dataset

- **KITTI Dataset:** Although not specifically for tiny objects, the KITTI dataset provides stereo images with ground truth for various tasks, including 3D object detection and pose estimation.
- **TUM RGB-D Dataset:** This dataset includes stereo image sequences with ground truth camera poses. While it's more focused on general scenes, it can be useful.

6. References

To provide a solid foundation and methodology for this problem, the following references can be useful:

1. **Feature Detection and Matching**:
    - Lowe, D.G., "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, 2004.
    - Rublee, E., Rabaud, V., Konolige, K., Bradski, G., "ORB: An efficient alternative to SIFT or SURF," ICCV, 2011.
2. **Fundamental and Essential Matrix Estimation**:

- Hartley, R., Zisserman, A., "Multiple View Geometry in Computer Vision," Cambridge University Press, 2003.
- Nister, D., "An efficient solution to the five-point relative pose problem," PAMI, 2004.

3. **Pose Estimation and Triangulation**:
   - Sturm, P., Triggs, B., "A factorization based algorithm for multi-image projective structure and motion," ECCV, 1996.
   - Yuxing Mao; Ching Y. Suen; Caixin Sun; Chunhua Feng, ,"Pose Estimation Based on Two Images from Different Views", WACV, 2007.
   - Zhang, Z., "A flexible new technique for camera calibration," PAMI, 2000.
   - Yujing Sun, Caiyi Sun, Yuan Liu, Yuexin Ma, Siu Ming Yiu, "Extreme Two-View Geometry From Object Poses with Diffusion Models", arXiv 2024.
4. **Self-Supervised Learning for Pose Estimation**:
   - Wang, Q., Zhao, P., Zhu, H., Tang, H., "Self-Supervised Learning for Accurate 3D Pose Estimation," CVPR, 2020.

Updated on 29th July, 2024.