Visual Odometry

Approach

The approach used for the final submission is described below. **I only used opencv for feature detection and matching. Implemented the rest of the code myself**.

1. Feature Matching

- Take both image frames and compute key points and feature descriptors using cv2 ORB.
- Use cv2 knn match to get feature matches with k = 2. This means that each keypoint of the first image is matched with each keypoint of the other image and only the top 2 best matches are kept. The best 2 matches are determined by the distance.
- Perform **Lowe's Ratio Test** to disregard bad matches with ratio constant **0.75**. This means that all features matches which are too close will be disregarded because in theory each keypoint should have a one good match and not two. So if a point has two close matches then those matches are disregarded. But if they are apart enough, then the first match is taken . This step improves the feature matching and helps improve the final results. Then out of all these matches the top **100** matches are taken as 2D - 2D correspondences.

2. Estimating Fundamental and Essential Matrix (No openCV)

- First compute F using normalized 8 point algorithm
- Compute inlier points using the estimated F. The inlier points are computed using the eq abs(x'Fx) < t . The threshold t = 0.01. So if the points x and x' obey this equation then they are marked as inlier points.
- If the number of inlier points are >= 8, then F is estimated again from inlier points using the normalized 8 point algorithm.
- Estimate E using the eq E = K(transpose) F K.
- Enforce constraints on E i.e decompose it using SVD (E = USVt) and make the diagonals 1,1,0 in S and then get E as E = USVt

This approach is similar to RANSAC. However, I did not use RANSAC because the results I was getting with RANSAC and this above approach were similar and my implementation of RANSAC was very slow.

3. Recover Pose (No OpenCV)

- Decompose E using Hartley and Zisserman Method and to get 4 estimates of R and t.
- Check determinants.
  - If Determinant(R) < 0
    - R,t = -R,-t
- Check chirality conditions for each of the 4 pose estimates on all inlier 2D points. The pose estimate that produces most points in front of both cameras is chosen as the pose estimate. The chirality condition is check as follows
  - For each pose estimate T = [R|t]
    - Get camera projection matrices P1=K[I | 0] and P2=K[R|t].
    - Triangulate points using P1 and P2 to get 3d points in front of the first camera.
    - Multiply current pose T with triangulated points to get 3d points in front of the second camera.
    - To count the points in front of both cameras, check z coordinates of both sets of 3d points. If the z coordinate of a point is greater than 0, it is in front of the camera.
  - Choose the pose T which produces the maximum number of points in front of both cameras.

4. Get Current Pose and Path

- Scale the estimated **t.**
- Form the Transformation matrix T = [R|t] of shape 4x4 by concatenating 0s and 1.
- If this is the first image then current pose = T else current pose is the inverse of this T. We have to take the inverse of this transformation because we have computed R,t in terms of how the points move. Since we are estimating the path of the camera in visual odometry, we have to inverse this.
- Current position is now chosen as the last column of the inverted transform.

SCORE: -214

Improvements

The following things helped improve the score

- Changing the initial number of features computed from each image from cv2 ORB to 3000. The default value was 500.
- Using Lowe's ratio test to disregard bad matches.
- Improving the estimate of F using the inlier points helped.
- Recovering pose using only the inlier points.
- Speed of this whole was improved by switching from RANSAC to this approach. RANSAC was taking too much time and not producing better results so I disregarded that.

Before applying these improvements my score was around -290 but after these improvements I managed to get -214 and beat the baseline. I also achieved an MSE of around 70 on the given video.