Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Oscar Toomey, supervised by Dr Ciarán McGoldrick

LAIDLAW SCHOLARS

Incremental Pose Refinement from Pose Estimates for Structure-from-Motion and Visual SLAM Applications

Project Summary

The ability to extract 3D information from images is a problem that engineers have worked on for decades. The field is known as Structure-from-Motion (SfM) in computer graphics and *Visual Simultaneous Localisation and Mapping* (Visual SLAM) in robotics. The techniques developed have applications in medicine, architecture, autonomous vehicles, structural analysis and disaster relief. For example, it has been used to delimit flood risk areas (2008, I. Ion et al), medical imaging (2005, B. Starly et al.), deformation detection and structural health monitoring and to assist first responders and investigators in natural disaster and humanitarian crises (2021, R. P. Murphy).

While many algorithms exist, there is still a need to find techniques that are more performant. This is especially the case when processing power is limited or when the results are required almost immediately. For example, drones and autonomous cars that are navigating in an unknown environment will need to make decisions in fractions of a second based on what they can see.

The goal of this work was to determine to what extent physical sensor data could be used to reduce the amount of processing required to extract a device's position and orientation in its environment.

This work focused on whether the performance of pose extraction can be improved using an initial pose estimate captured using other sensors on the device. In this problem, the goal is to determine the pose of a camera relative to a reference camera, or the difference in position and rotation between the two. This is done by finding matching regions (or "features") in the image capture by each camera, determining the reprojection error between the two, and then minimizing it. The problem is described visually in Figure 1. The estimate for the position of the camera on the right is off such that the features, when reprojected, result in an error between the two; seen as e on the diagram. Pose refinement algorithms determine how to move the camera to minimize this error.

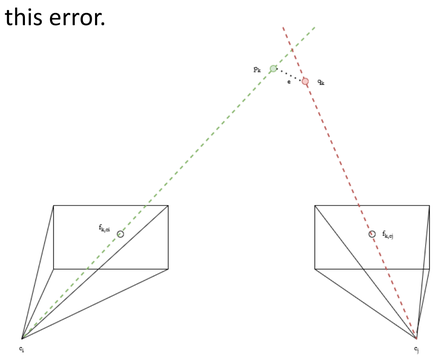

Figure 1: Improving the estimate of Camera c_j for the feature

Implementation

The processing graph of the pose refinement system developed in this work is shown in Figure 2. For testing purposes the plant was replaced with test. Feature extraction was performed using the Oriented FAST and Rotated BRIEF (ORB) algorithm (Rublee, 2011) and then matched using the distance ratio test (Lowe, 2004). Pose refinement is then performed on image pairs i and j, the first being the reference image (whose pose is taken as accurate) and the second being the process image (whose pose still needs to be refined).

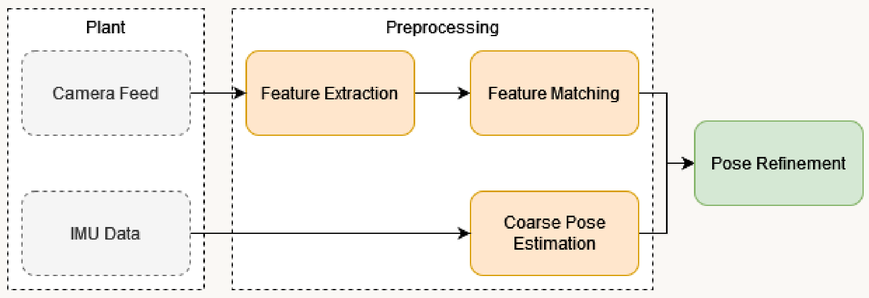

Figure 2: Pose refinement system implemented in this work

Inputs to pose refinement system are feature pairs f_{k,c_i} and f_{k,c_j} matched between each image, pose of the reference image T_i and pose *estimate* of the process image T_j. Each features lies on the 3D ray that passes from camera centre through feature's position on camera image plane into the scene. If pose estimate for process camera is perfectly accurate, then each feature ray pair intersects at the point where the feature is located. Otherwise, two estimates for the position exist (Figure 1) p_k and q_k.

The problem of pose refinement can be understood as the process of moving the process camera to eliminate the distance between these point pairs (Figure 3).

In this work, a technique from Image Based Visual Servo Control (IBVS) was used. On the process camera's image plane, the two feature position estimates are subtracted to form the error term, using the process camera's interaction matrix I_j:

$$e = I_j \cdot T_{c_j} \cdot p_k - f_{k,c_j}$$

Position and rotation of the camera is updated incrementally using control law:

$$\mathrm{v_c} = -\lambda \widehat{\mathrm{L_e^+}} \mathrm{e}$$

$\widehat{\mathrm{L_e^+}}$ is the interaction matrix that relates the screen space error of each feature pair to a correcting change in the camera's position and rotation. In this work it is the Moore-Penrose Inverse of L_k:

$$L_k = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{u_{k,c_i}}{Z} & u_{k,c_i}v_{k,c_i} & -\left(1+u_{k,c_i}^2\right) & v_{k,c_i} \\ 0 & -\frac{1}{Z} & \frac{v_{k,c_i}}{Z} & 1+v_{k,c_i}^2 & -u_{k,c_i}v_{k,c_i} & -u_{k,c_i} \end{bmatrix}$$

Where u_{k,c_i} and v_{k,c_i} are the x and y coordinate of the reference camera feature estimate in the process camera's view space, and Z is the distance of each feature estimate from the process camera.

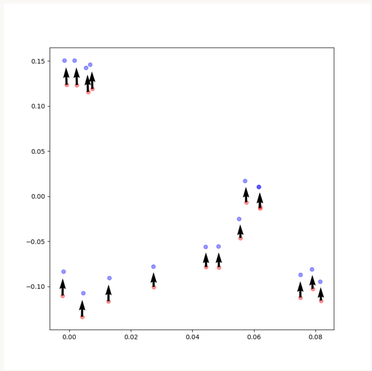

Figure 3: 16 feature pairs on the process camera image frame during pose refinement

Results

The algorithm implemented in this work is able to correct small errors in the rotation and position of the process camera. The time-to-convergence depends on both the initial pose error as well as the triangulation error, and as such must be adjusted for the target application. Empirically it was found using 25 iterations for errors within half a metre and 30 degrees was usually enough for convergence (Figure 3).

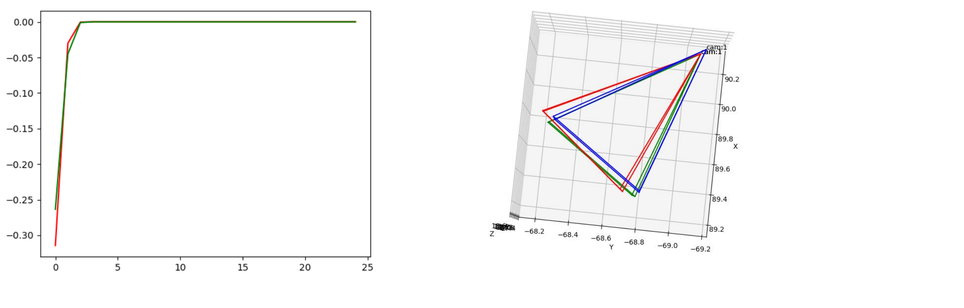

Figure 3: Left – Convergence of e on both the x axis (red) and y axis (green) for the case shown on the Right – The camera frustrum before post refinement (red), afterward (green) and the ground truth (blue)

However, there is a physical limit to the maximum refinement possible. There are a number of issues that hinder performance:

- Triangulation error: The depth of each point in the scene must be estimated for RGB cameras.
- Feature extraction error: The features matched by the feature matches often don't exactly line up between two image frames.
- Quantization error: For features that are further away, the maximum precision they can triangulated to is limited by the size of the pixels on the image frame.
- Positions close to the true value often result in a small screen space error.

Conclusions

The algorithm implemented in this work successfully demonstrated the ability to refine the pose of a camera given a reference camera and an initial pose estimate. However, due to the variable convergence time, state-of-the-art perspective and point algorithms, such as lambda-twist (Persson, 2018), which consistently converge in few iterations may be preferential for real time applications.

An area of future research is to determine whether pose estimates can be used to increase the performance of feature extraction and matching. This is an area that has received attention recently with models like LoFTR (Sun, 2021) and SuperGlue (Sarlin, 2019) using machine learning to find hundreds of matches between image pairs compared to the two or three dozen found by the ORB matcher used in this work. Using an initial pose estimate for the new image could potentially be used to reduce the search region for feature matches between image pairs.

References

1. Chaumette, S. Hutchinson, *Visual servo control, Part I: Basic approaches*, HAL Open Science, Institute of Electrical and Electronics Engineers, 2006
2. G. Lowe. *Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision*, 60(2):91–110, November 2004.
3. Persson, Nordberg, *Lambda Twist: An Accurate Fast Robust Perspective Three Point (P3P) Solver*. European Conference on Computer Vision (ECCV) 2018
4. Rublee, Rabaud, Konolige, R. Bradski, *ORB: An efficient alternative to SIFT or SURF*. International Conference on Computer Vision (ICCV) 2011: 2564-2571.
5. Sarlin, DeTone, Malisiewicz, Rabinovich, *SuperGlue: Learning Feature Matching with Graph Neural Networks*, Conference on Computer Vision and Pattern Recognition (CVPR) 2019
6. Sun, Jiaming and Shen, Zehong and Wang, Yuang and Bao, Hujun and Zhou, Xiaowei, *LoFTR: Detector-Free Local Feature Matching with Transformers*, Conference on Computer Vision and Pattern Recognition (CVPR) 2021

This work was conducted and funded as part of the Laidlaw Undergraduate Leadership and Research programme