![UCL logo]

COMP0130 - ROBOT VISION AND NAVIGATION

Coursework 3: SLAM Systems

Name: **Yinji Zhu** ID: **18062537** Email: **ucaby22@ucl.ac.uk**

21 Apr 2019

Part A: The description of the tracking thread of the system

The tracking thread of the SLAM system is made of 4 steps, ORB Extraction, Pose Prediction via Previous Frame and Global Relocalization, Track Local Map and New Keyframe Decision, which is shown in Figure 1 [1].

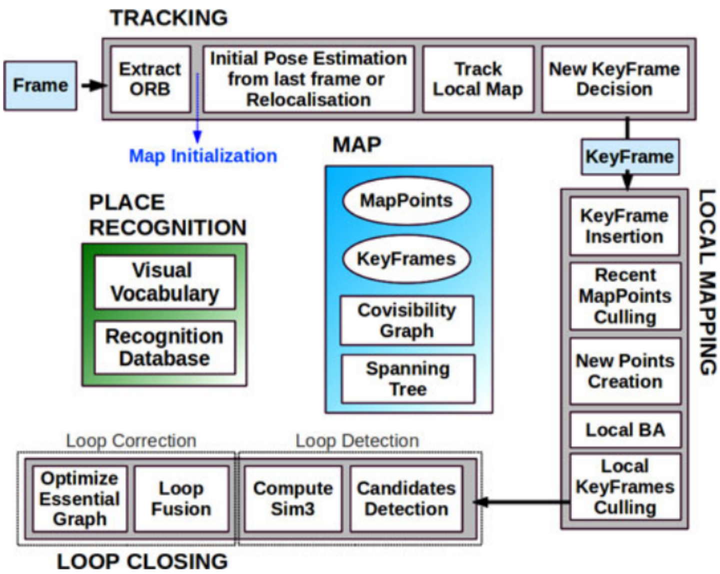

Figure 1: The thread of the ORB-SLAM2 system [1]

ORB-SLAM is the abbreviation of Oriented Fast and Brief SLAM. Firstly, when frames are transmitted to the tracking thread, FAST algorithm extracts feature points by comparing the current point with the surrounding points. If the current point is different with most surrounding points, it can be regarded as a corner. The algorithm detects corners at 8 scale levels and extracts default 1000 corners which has at least 5 scale levels. After these feature points are determined, features between two frames are compared through descriptor. Knowing that the same features always have different directions, sizes, intensities, etc. If just comparing their surrounding points, features will not regarded as the same. Consequently, ORB descriptor is used rather than Brief descriptor, which meets the repeatability of feature descriptor[2].

In the second part, the judgement of inliers and outliers are based on RANSAC, which can exclude the potential bad matches[2]. If the number of outliers exceeds one value, the frame will be regarded as tracking unsuccessful. If the last frame is tracked successful, it will be used to predict the camera pose. If not, the key frame will be used to predict the camera pose. The camera's current pose is calculated by updating with feature matching and position of features. Using two frames, the nine parameters in essential matrix can be calculated from equation: $p_2^T E p_1 = 0$, which can be solved by SVD. By decomposing essential matrix, R_k and t_k can be obtained using SVD as well[2]. In this case, the current camera pose can be obtained.

If the estimation of the camera pose is obtained, the Motion-only BA is used to optimize rotation and translation using the equation to obtain the minimum projection error between key points and global map, where the initial camera pose is the predicted camera pose:

$$\{\mathbf{R}, \mathbf{t}\} = \underset{\mathbf{R},\mathbf{t}}{\operatorname{argmin}} \sum_{i\in\mathcal{X}} \rho\left(\left\|\mathbf{x}^i_{(\cdot)} - \pi_{(\cdot)}\left(\mathbf{R}\mathbf{X}^i + \mathbf{t}\right)\right\|^2_\Sigma\right) [3]$$

By changing the value of rotation $\mathbf{R}$ and translation $\mathbf{t}$, the energy can be minimized.

The final step in tracking is new key frame decision. The decision way of key frame obeys some rules like: the new key frame can only be created after several frames of last global re-localization since the adjacent key frames should have differences, the frame needs to track larger than a specific number of feature points to ensure the quality of key frame, and so on[1].

Part B: The results of the evaluations

In this part, the ORB-SLAM2 system was evaluated on the sequences, f2_xyz and f3_long_office. The ORB-SLAM2 system was run by the combination of the RGB data and depth data of the sequences and the results were Camera Trajectory data and KeyFrame Trajectory data. In this case, an evaluation tool, "evaluate_ate.py", was used to evaluate the SLAM system by the results Trajectories of the ORB-SLAM2 system and the ground truth in the sequences.

By this evaluation tool, the Absolute Trajectory Error (ATE) will be worked out and evaluated. In the evaluation function, the translational error of per point in the two trajectories will be worked out as a $1 \times n$ matrix, where the n is the number of points in the shorter trajectory. The returned values of this function are the root mean square value of the n translational errors, the mean value of the n translational errors, the median value of the n translational errors, the standard deviation of the n translational errors, the maximum value in the n translational errors and the minimum value in the n translational errors.

Task 1

In this task, firstly, the ORB-SLAM system was run with the 2 sequences, which is shown in Figure 2 and 3.

After the SLAM system finishes the working, the results, Camera Trajectories, are used to evaluate the SLAM system with the ground truth.

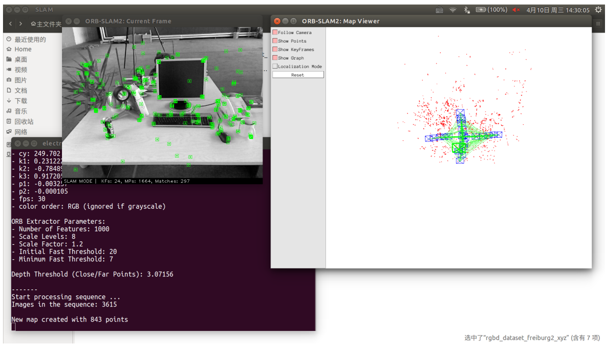

Figure 2: The running by f2_xyz

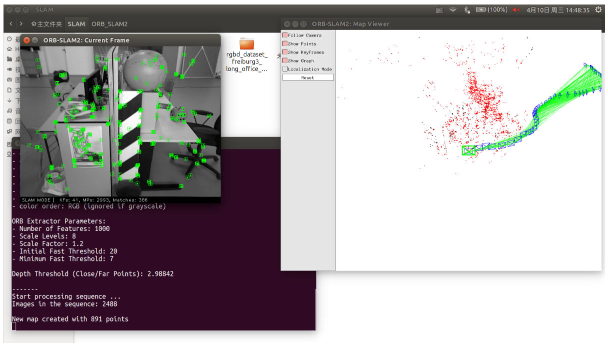

Figure 3: The running by f3_long_office

For the sequence, f2_xyz, the plot of the ATE evaluation is shown in Figure 4, where the black line is the ground truth, the blue line is the estimated trajectory and the red line is the error between them.

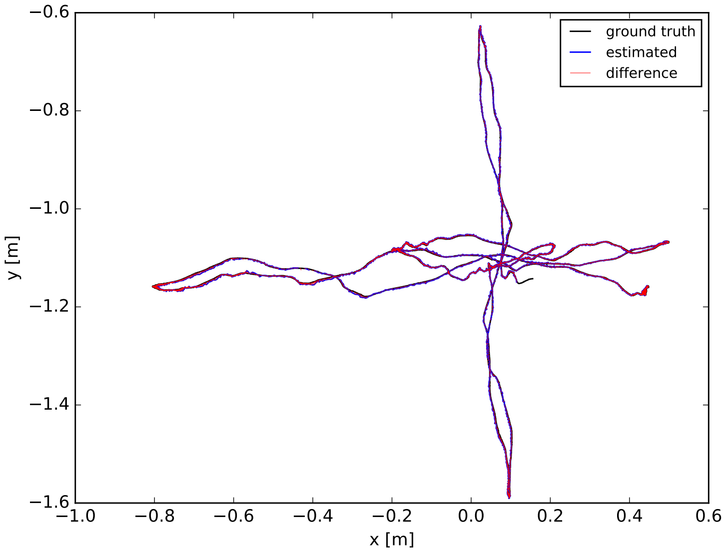

Figure 4: The plot of the ATE evaluation with f2_xyz

The root mean square value with the sequence, f2_xyz, is 0.003691 m. The m is the unit, meter. The mean value is 0.003202 m. The median value is 0.002865 m. The standard deviation is 0.001836 m. The minimum value is 0.000121 m. The maximum value is 0.012120 m.

For the sequence, f3_long_office, the plot of the ATE evaluation is shown in Figure 5.

The root mean square value with the sequence, f3_long_office, is 0.008767 m. The mean value is 0.007939 m. The median value is 0.007448 m. The standard deviation is 0.003717 m. The minimum value is 0.000520 m. The maximum value is 0.025449 m.

From the figures, the camera trajectories are quite closed to the ground truth. It can be proved from both the root mean square values and standard deviations are small as well. In this case, the ORB-SLAM2 system works stably and tracks accurately.

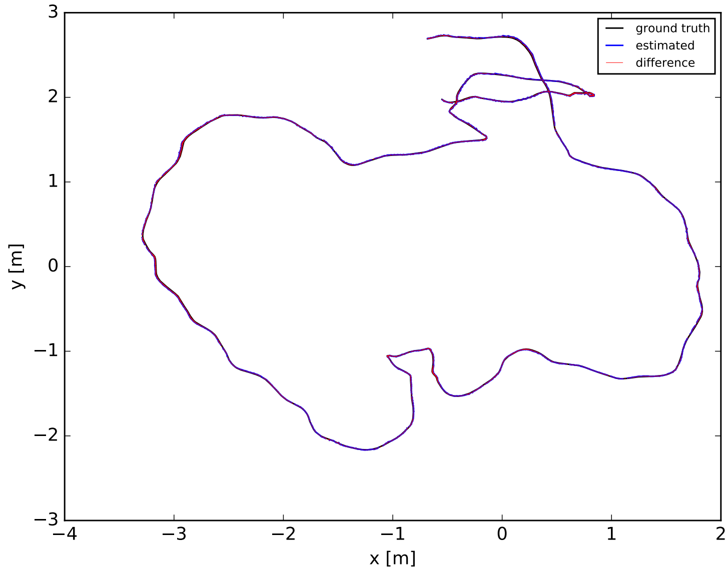

Figure 5: The plot of the ATE evaluation with f3_long_office

Task 2

To adjust the number of ORB features, 2 parameters in the files, "TUM2.yaml" and "TUM3.yaml", with the same name can be changed. Both of the default numbers of the ORB features in the 2 files are 1000.

Following the requirement, reducing the number of ORB features with choosing 3 levels, the 3 levels of the number of features were set to 500, 750 and 1000.

For the sequence, "f2_xyz", the plots of the ATE evaluations with 3 different ORB features are shown in Figure 6, 7 and 8.

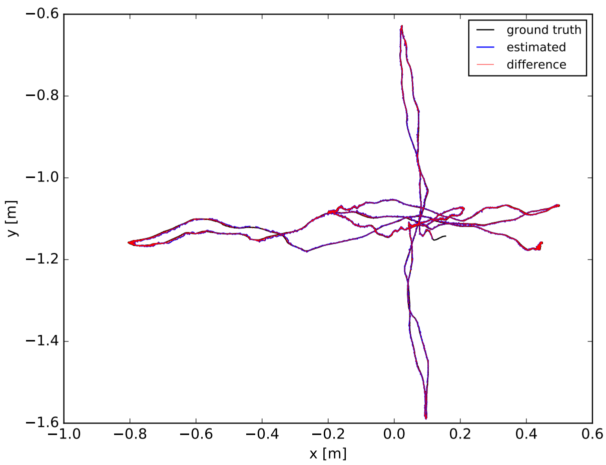

Figure 6: The plot, f2_xyz, 500 feature points

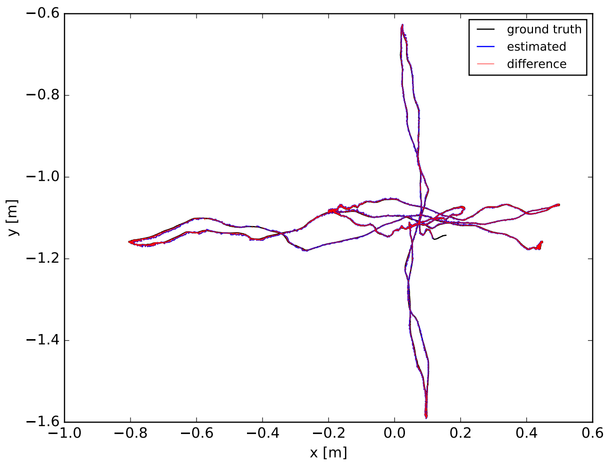

Figure 7: The plot, f2_xyz, 750 feature points

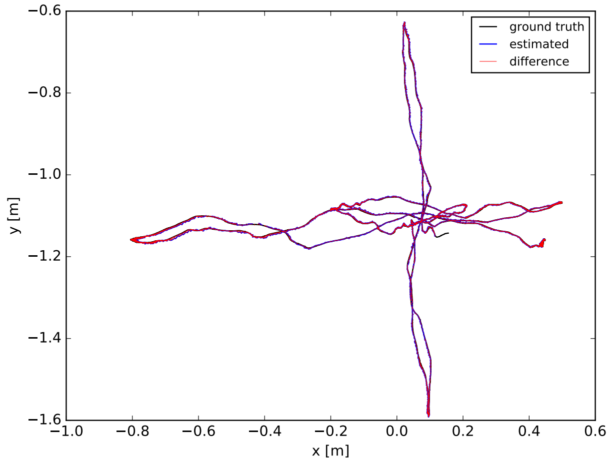

Figure 8: The plot, f2_xyz, 1000 feature points

The ATE evaluation results are shown in Table 1.

Table 1: The ATE evaluation results with f2_xyz

	rmse	mean	median	std	min	max
500	0.004086	0.003633	0.003367	0.001870	0.000142	0.013197
750	0.003723	0.003298	0.003067	0.001729	0.000238	0.011916
1000	0.003533	0.003107	0.002855	0.001680	0.000155	0.011317

From the figures, there is no obvious difference. However, in the table, with the increasing of the feature points, the root mean square value will be decreased. Furthermore, the standard deviation becomes smaller as well. Therefore, under the situation, the sequence "f2_xyz", with the feature points being more, the ORB-SLAM system becomes more stable and accurate.

For the sequence, "f3_long_office", the plots of the ATE evaluations with 3 different ORB features are shown in Figure 9, 10 and 11.

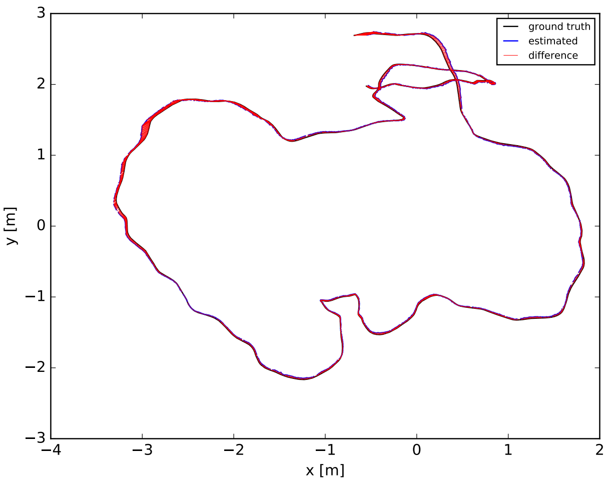

Figure 9: f3_long_office, 500 feature points

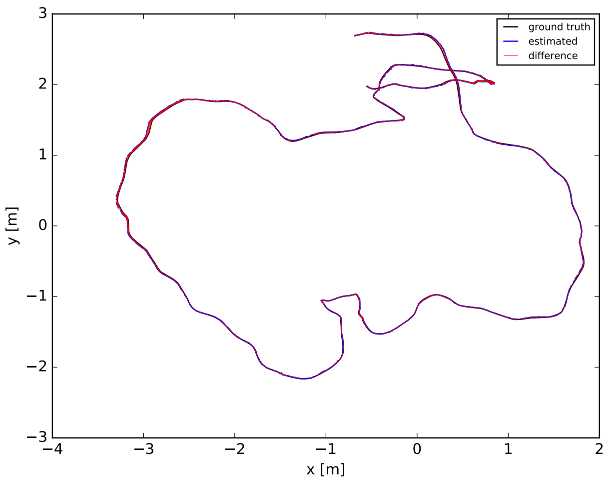

Figure 10: f3_long_office, 750 feature points

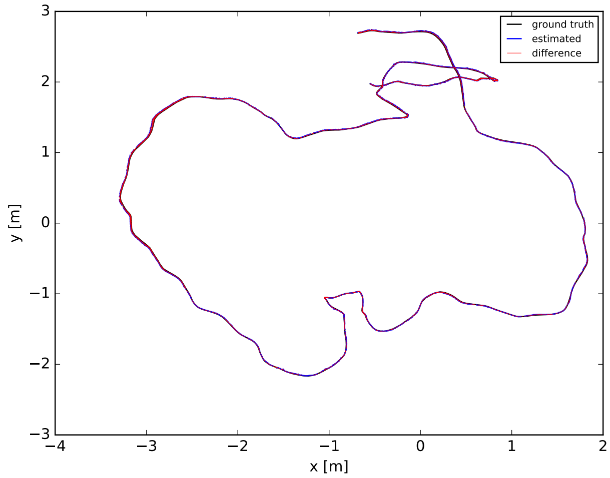

Figure 11: f3_long_office, 1000 feature points

The ATE evaluation results are shown in Table 2.

Table 2: The ATE evaluation results with f3_long_office

	rmse	mean	median	std	min	max
500	0.038206	0.029115	0.020480	0.024739	0.001351	0.146037
750	0.013743	0.011830	0.009804	0.006994	0.000860	0.040431
1000	0.008767	0.007939	0.007448	0.003717	0.000520	0.025449

From the figures, the difference is not large, but it can be observed. It can be seen that fig.11 has smaller width than fig.10 and fig.9, so the result of 1000 feature points is more accurate than the result of 500 points. Moreover, in the table, with the increasing of the feature points, the root mean square value will be decreased. Furthermore, the standard deviation becomes smaller as well. Therefore, under the situation, the sequence "f3_long_office", with the feature points being more, the ORB-SLAM system becomes more stable and accurate as well.

Task 3

To turn off the outlier rejection stage, the code in the file, "Optimizer.cc", was adjusted. At the line 397 and the line 426, change the "true" to "false". At the line 398 and the line 427, change the 1 to 0. At the line 399 and the line 428, remove the "nBad++;". The idea is to remove the function of finding out the outliers.

In this part, the number of feature points is set to 1000. During the running of the ORB-SLAM in this case, the number of key frame become much smaller. The reason is that because the outlier rejection stage is removed, the SLAM system algorithm 'thinks' there is no outliers and the quality of the tracking is good. Therefore, there will be much less key frames. Conversely, if the outliers rejection stage is kept, there will be many outliers being found out and the SLAM algorithm will 'think' the quality of the tracking is bad. In this case, there will be many key frames.

For the sequence, "f2_xyz", the plot of the ATE evaluations without the outlier rejection stage is shown in Figure 12.

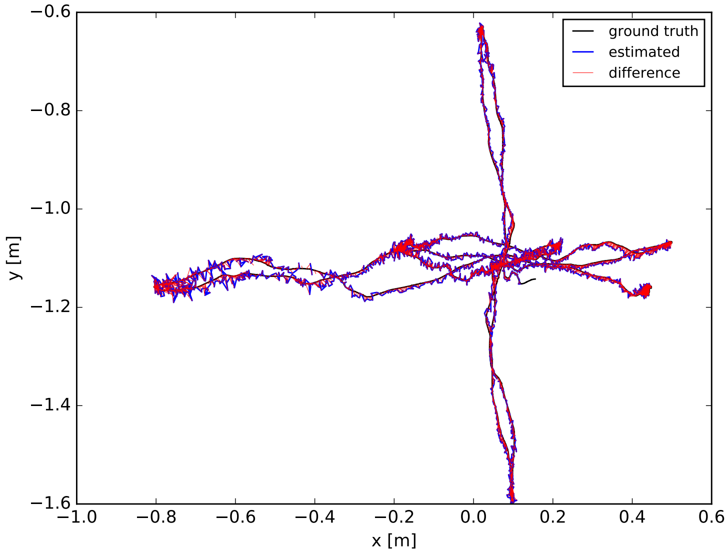

Figure 12: f2_xyz, without the outlier rejection stage

From the figure, without the outlier rejection stage, the camera trajectory has large oscillations. About the ATE values, the results data are shown in Table 3.

Table 3: The ATE evaluation results of f2_xyz with & without outlier rejection stage

The outlier rejection stage	rmse	mean	median	std	min	max
with	0.003691	0.003202	0.002865	0.001836	0.000121	0.012120
without	0.011535	0.010175	0.009261	0.005434	0.000466	0.043395

For the sequence, "f3_long_office", the plot of the ATE evaluations without the outlier rejection stage is shown in Figure 13.

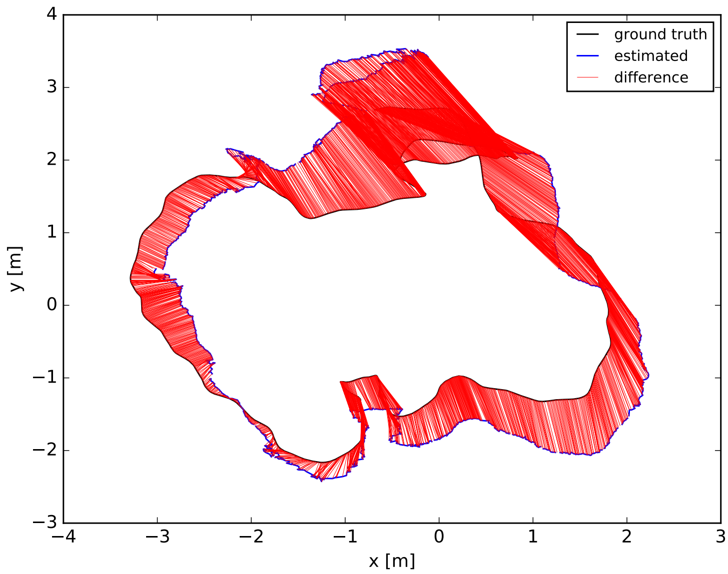

Figure 13: f3_long_office, without the outlier rejection stage

From the figure, without the outlier rejection stage, the camera trajectory has large oscillations. Also, the difference between the camera trajectory and ground truth becomes too large. About the ATE values, the results data are shown in Table 4.

Table 4: The ATE evaluation results of f3_long_office with & without outlier rejection stage

The outlier rejection stage	rmse	mean	median	std	min	max
with	0.008767	0.007939	0.007448	0.003717	0.000520	0.025449
without	1.009113	0.868707	0.698210	0.513476	0.032984	1.974704

From the results, after the outlier rejection stage is removed, the whole system cannot work normally and it is quite unstable. Compared with the situation with outliers rejection stage, the root mean square value of the error and the standard deviation becomes quite huge. The reason is the effect of outliers is huge and they are not removed. Without the outlier rejection stage, the bad pairs and the large error feature points will be kept. In this case, there will be large errors in the results.

References

[1] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[2] L. A. Simon Julier. (2019) Compgx04: Robotic vision and navigationengineering a slam system. [Online]. Available: https://moodle-1819.ucl.ac.uk/pluginfile.php/368986/mod_resource/content/1/Slides_Robot_Vision.pdf

[3] R. Mur-Artal and J. D. Tards, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct 2017.