COSC 342 Assignment 1 – Image Mosaicing

Report for Beckham Wilson ID #7564267

Introduction

This report aims to perform experiments to see the effects of different choices on parts of the image stitching process specifically the feature matching and homography estimation stages. First experiment we will compare two feature matches used by OpenCV and second experiment vary the choice of RANSAC threshold used in homography estimation. Both experiments will measure the accuracy of the homography via the reprojection error.

Reprojection Error calculates the Euclidean distance between transformed points in the source image to points in the destination image. It measures how well the homography aligns the source and destination points. The same reprojection error is used by OpenCV findHomography() for filtering by the ransacThreshold. The image below shows how it is calculated.

$\mathbf{d}$ point in the image we want to align to

$\mathbf{s}$ point in the image we want to transform using homography H

Reprojection error = $||\mathbf{d} - H\mathbf{s}||$

Image Dataset:

Both experiments utilized the same image dataset, comprising a variety of scene types: outdoor/greenery, exterior of buildings, and interior of buildings. Image pairs vary due to camera rotations or translations, particularly in planar scenes.

Image Number	Resolution/Image type	Source
1-40	1000x750 / JPG	https://github.com/tlliao/Single-perspective-warps/tree/master [1]
41-59	3000x4000 / JPG	Myself
60-104	1500x2000 / JPG	Myself

Images 1-40 depict scenes exclusively featuring the exteriors of buildings, with a combination of outdoor and greenery elements.

Images 41-104 encompass various scene types, including 40 interior shots, 16 exterior shots of buildings, and 8 outdoor/greenery scenes.

Experiment Process:

For each image pair:

1. Generate SIFT features from both images.
2. Use a feature matcher like Brute-Force or FLANN to find corresponding features between the images.
3. Apply Lowe's ratio test to remove unreliable matches.
4. Use RANSAC to estimate the homography between the images.
5. Compute the reprojection error for both inlier and outlier points.

Data Filtering:

Reprojection error data was filtered where all data points must be between LQ − 1.5 IQR and UQ + 1.5 IQR.

LQ (Lower Quartile)

UQ (Upper Quartile)

IQR (InterQuartile Range)

Experiment 1: Feature Matching

Hypothesis/Question:

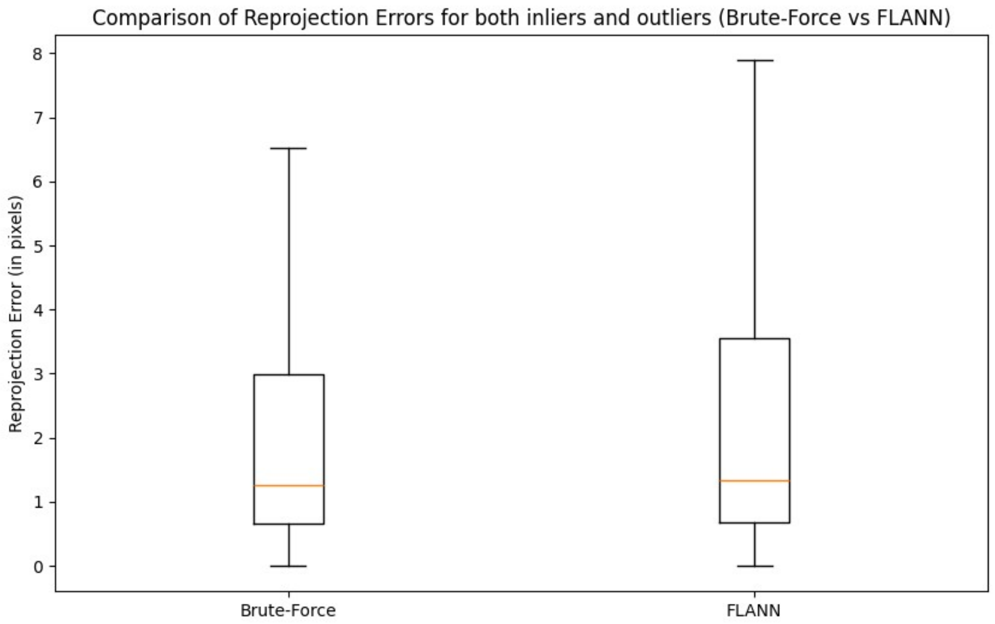
Brute-Force feature matcher results in a more accurate homography than the FLANN feature matcher.

Experimental Design:

We will follow the Experiment Process above but vary the feature matcher between Brute-Force and FLANN.

Results:

The below boxplot shows the distributions for the Reprojection error for both inlier and outlier points for the Brute-Force and FLANN feature matcher.



Below is the summary statistics for reprojection errors for both inliers and outliers

Feature Matcher:	Brute-Force	FLANN
Count before data filtering	172242	181106
Count after data filtering	140447	140495
Mean	2.93	4.59
Standard Deviation	4.15	9.07
Min	0.01	0
Lower Quartile (25%)	0.65	0.67
Median (50%)	1.26	1.33
Upper Quartile (75%)	3	3.56
Max	22.54	67.65

Standard Deviation: Brute-Force's standard deviation stands at 4.15, slightly lower than FLANN's 4.59, indicating FLANN's higher variability in reprojection errors.

Quartiles: Brute-Force and FLANN have similar lower quartile and median values. However, FLANN's upper quartile is higher at 3.56 compared to Brute-Force's 3, suggesting FLANN's tendency for higher reprojection errors at the upper end. This, along with the standard deviation, explains FLANN's higher mean of 4.59 compared to Brute-Force's 2.93.

Overall, Brute-Force matching generally exhibits lower reprojection errors across all summary statistic metrics. The difference is more evident at the higher end of the distribution, with FLANN showing higher variability. FLANN tends to have slightly higher maximum errors and higher reprojection errors in the upper quartile range, indicating more significant errors from its outliers compared to Brute-Force. However, for inlier points (below the RANSAC threshold of 3), the reprojection error difference between the two methods is minimal.

Discussion/Conclusions:

The results support our hypothesis that the Brute-Force feature matcher generates a more accurate homography compared to FLANN due to its lower reprojection error. This aligns with expectations as FLANN employs an approximate method, resulting in less precise matches. Hence this is why we see a very noticeable difference in the reprojection error for outliers as the matching wasn't accurate.

Experiment 2: RANSAC for Homography Estimation

Hypothesis/Question:

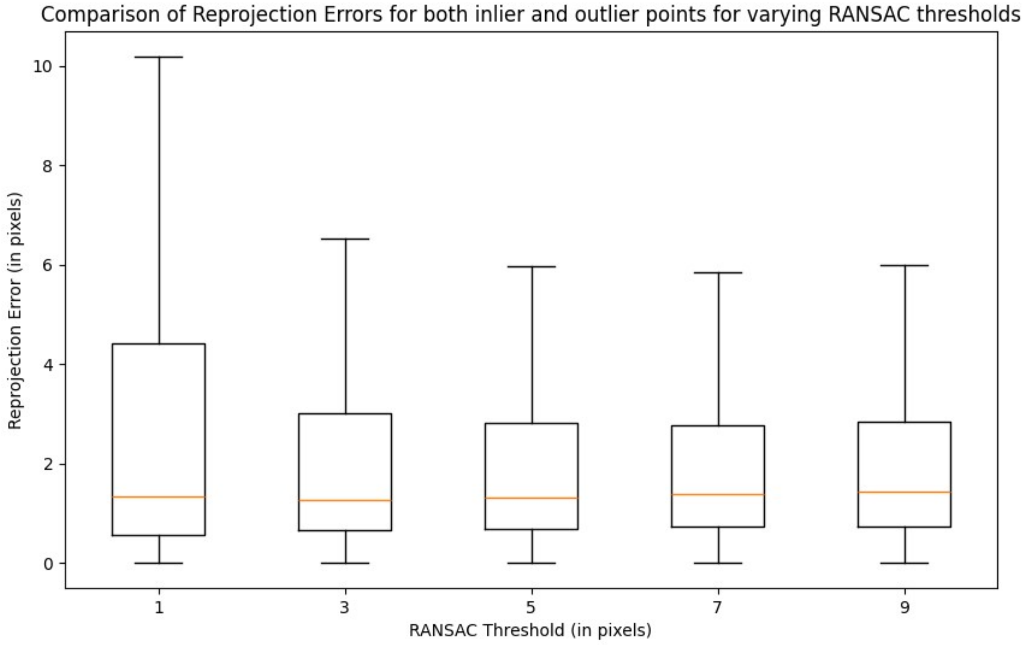
How does the choice of RANSAC threshold affect the accuracy of the homography?

Experimental Design:

We will follow the Experiment Process mentioned at the start but only use the Brute-Force feature matcher and then only vary the RANSAC threshold.

Results:

The below boxplot shows the distributions for the Reprojection error for both inlier and outlier points for the varying RANSAC thresholds.



Below are the summary statistics for reprojection errors for both inliers and outliers. (The number of feature matches before data filtering was 172242)

RANSAC Threshold:	1	3	5	7	9
Count after data filtering	142468	140447	138779	138698	138704
Mean	4.14	2.93	2.55	2.42	2.38
Standard Deviation	6.35	4.15	3.32	2.93	2.74
Min	0	0	0	0	0
Lower Quartile (25%)	0.56	0.65	0.68	0.72	0.73
Median (50%)	1.32	1.26	1.32	1.39	1.43
Upper Quartile (75%)	4.41	3	2.8	2.77	2.83
Max	33.17	22.54	19.87	17.75	16.44

Lower Quartile (LQ) Increase: As the RANSAC threshold increases, the Lower Quartile (LQ) for reprojection error also increases. This suggests that points closer to inlier points exhibit higher reprojection errors. The largest increase in LQ occurs from a RANSAC threshold of 1 to 3.

Upper Quartile (UQ), Mean, and Standard Deviation Decrease: Increasing the RANSAC threshold leads to a decrease in the Upper Quartile (UQ), Mean, and Standard Deviation of the reprojection error. This indicates that as the threshold increases, points closer to outlier points show decreased reprojection errors. The most significant decrease occurs from a RANSAC threshold of 1 to 3.

Median RANSAC Threshold: The median RANSAC threshold exhibits a decreasing trend from 1.32 for a threshold of 1 to 1.26 for a threshold of 3. However, there is a consistent increase in the median RANSAC threshold for threshold values of 5, 7, and 9, reaching 1.32, 1.39, and 1.43 respectively.

Discussion/Conclusions:

As the RANSAC threshold increases, errors tend to rise for points closer to inliers but decrease for those closer to outliers. The median RANSAC threshold reaches its lowest value at a threshold of 3. This implies that while errors for inliers tend to rise and errors for outliers tend to decrease with higher thresholds, there's an optimal balance at a threshold of 3 where the median reprojection error is minimised evenly for both inliers and outliers. Therefore we notice that the choice of RANSAC threshold does play a noticeable part in the accuracy of the homography.

Final Remarks

In conclusion, both experiments shed light on the impact of the choice of feature matcher and the impact of the RANSAC threshold on the accuracy of homography estimation in image mosaicing.

Experiment 1 demonstrated that the feature matcher Brute-Force outperformed FLANN, exhibiting lower reprojection errors overall and hence more accurate homography. This aligns with expectations, given Brute-Force's precision in matching features compared to FLANN's approximate method.

Experiment 2 focused on the impact of RANSAC threshold variation on homography accuracy. Results showed that increasing the threshold led to a rise in errors for inliers but a decrease for outliers. Surprisingly, the optimal balance, reflected by the lowest median reprojection error, was observed at a threshold of 3.

Combining the two experiments we see that the choice of RANSAC threshold significantly influences overall accuracy compared to the choice of feature matcher used.

Limitations include having only three scene types, as well as unbalanced pixel resolutions and scene type distributions within each resolution group. Additionally, I resized many images to smaller resolutions due to lengthy processing times with the 3000x4000 images, with some resulting in over 100,000 features.

Future research could explore additional scene types, address image resolution disparities, and analyze inlier and outlier points separately for deeper insights into homography estimation accuracy especially in experiment 2 where we inferred how the inlier and outlier points behave as we change RANSAC threshold.

References

[1] Liao, T., & Li, N. (2020). Single-Perspective Warps in Natural Image Stitching. *IEEE Transactions on Image Processing*, *29*, 724–735. https://doi.org/10.1109/TIP.2019.2934344