COSC 342 Assignment 1 – Image Mosaicing

Report for Beckham Wilson ID #7564267

Introduction

The experiments conducted within this report demonstrate the effects of different choices on parts of the image stitching process, specifically the feature matching and homography estimation stages. In the first experiment, we will compare two feature matchers used by OpenCV. The second experiment will compare the choice of RANSAC threshold used in homography estimation. Both experiments will measure the accuracy of the homography via the reprojection error.

What is a Reprojection Error:

Reprojection error calculates the Euclidean distance (measured in pixels) between points in the source image transformed by the homography to their corresponding points in the destination image. It measures how well the homography aligns the source and destination points. OpenCV's findHomography() function computes the reprojection error and employs the RANSAC threshold for outlier filtering. The reprojection error formula is below.

$\mathbf{d}$ 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 use the same image dataset, comprising of three scene types: outdoor/greenery, the exterior of buildings, and the interior of buildings. Image pairs vary by camera rotations or translations, particularly in planar scenes.

Image Numbers	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 contain all three 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 feature matches.
4. Use RANSAC to estimate the homography between the images.
5. Compute the reprojection error for both inlier and outlier points.

Data Filtering:

The Tukey fence method was chosen for its resistance to extreme outlier values, ensuring that the data points considered for analysis are within a reasonable range of variation. Applying this filter criterion excludes data lying beyond 1.5 times the interquartile range from the lower quartile, and upper quartile. This method removes the influence of erroneous or extreme measurements, such as reprojection errors in the thousands or tens of thousands of pixels caused by obviously inaccurate feature matches.

Experiment 1: Feature Matching

Hypothesis/Question:

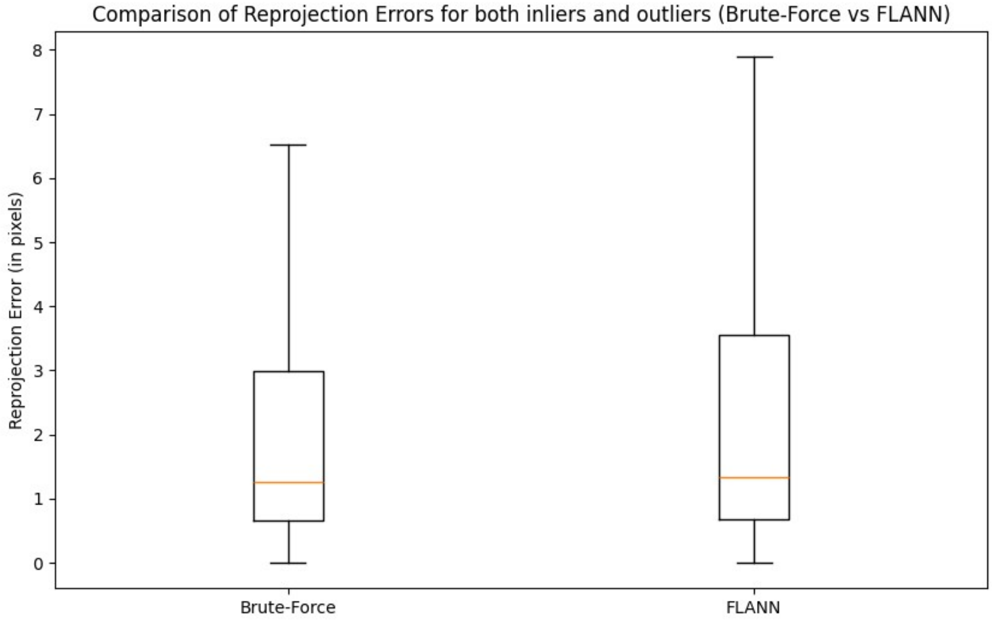
Does the Brute-Force feature matcher result in a more accurate homography than the FLANN feature matcher?

Experimental Design:

We will follow the outlined experiment process above but vary the feature matcher between Brute-Force and FLANN.

Results:

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



Below are the summary statistics for reprojection errors (in pixels) for both inliers and outliers points.

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 is 2.93, 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 of the distribution. This, along with the standard deviation, explains FLANN's higher mean of 4.59 compared to Brute-Force's 2.93.

The Brute-Force matching generally exhibits slightly 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 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 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.

Experiment 2: RANSAC for Homography Estimation

Hypothesis/Question:

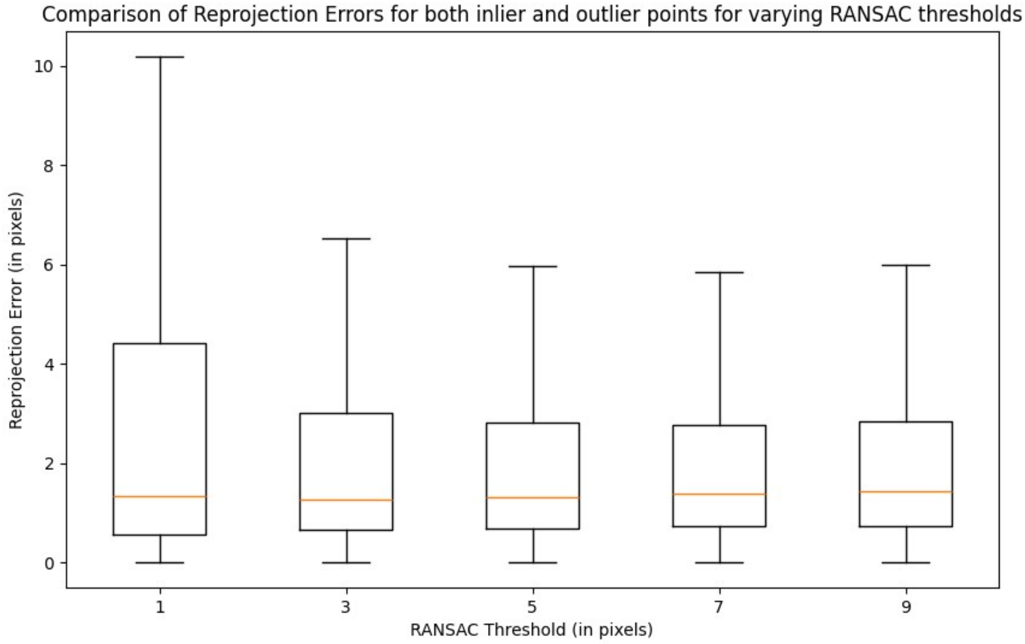
How does the choice of RANSAC threshold affect the accuracy of the homography?

Experimental Design:

We will follow the experiment process as mentioned at the start of the report, but only use the Brute-Force feature matcher and then vary the RANSAC threshold variable.

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 172,242)

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

The Lower Quartile (LQ) increases as the RANSAC threshold increases. This suggests that points closer to inlier points exhibit higher reprojection errors as RANSAC threshold increases. The largest increase in LQ occurs from a RANSAC threshold of 1 to 3.

By increasing the RANSAC threshold, the Upper Quartile (UQ), Mean, and Standard Deviation will decrease 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.

The median RANSAC threshold exhibits a decreasing trend for threshold 1 is 1.32 pixels, and for threshold 3 is 1.26 pixels. 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 pixels 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 reprojection error reaches its lowest value at a RANSAC 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, the choice of RANSAC threshold and feature matcher influences homography accuracy.

Experiment 1 demonstrated that the feature matcher Brute-Force outperformed FLANN, exhibiting lower reprojection errors overall, resulting in 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 the RANSAC threshold on homography accuracy. Results showed that increasing the threshold led to a rise in reprojection 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.

Limitations of the study include the restricted diversity in scene types. The experiments conducted within this report show the effects of different choices on parts of the image stitching process, specifically the feature matching and homography estimation stages. Furthermore, the unequal distribution of images across certain pixel resolutions and scene types within each resolution group. Additionally, due to prolonged processing times associated with the larger 3000 x 4000 pixels images, many images were resized to smaller resolutions.

Future research could explore additional scene types, address image resolution disparities, and analyze inlier and outlier points separately. This can help develop insights into homography estimation accuracy, especially in experiment 2 where we inferred how the inlier and outlier points behave as we change the RANSAC threshold. Also, for experiment two, we should have exclusively utilized the feature matches from Brute Force for computing reprojection errors, while employing the respective FLANN and Brute Force feature matchers for homography estimation.

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