

CSE803 HW3

Javen W. Zamojcin
zamojci1@msu.edu

2024, October 3

1 Question 1: RANSAC

1.1 Part A: Fitting a Line

1.1.1

Any given line is composed of at least two points, therefore the number of points randomly selected which constitutes a random sample is two. $s = 2$

1.1.2

Probability of selecting at least one outlier in a given sample $= 1 - (1 - e)^s = 1 - (1 - 0.1)^2 = 0.19 = 19\%$ chance of failure.

1.1.3

Number of samples required for success with probability of 0.95 $= \frac{\log(1-p)}{\log(1-(1-e)^s)}$
 $= \frac{\log(1-0.95)}{\log(1-(1-0.1)^2)} = \text{ceil}(1.803) = 2.$

1.2 Part B: Fitting Transformations

1.2.1

Degrees of freedom $= 2^n = 2^2 = 4$

M requires at least four 2D samples.

1.2.2

$\text{argmin}_m ||Am - b||^2 = \text{argmin}_m \sum_{i=1}^N (y_i - mx_i - b)^2 = (A^T A)^{-1} (A^T Y)$

1.2.3

$S = \begin{bmatrix} 2.04098127 & -3.0696843 \\ -1.01644528 & 0.94335775 \end{bmatrix}, t = [-1.87154926 \quad -3.05145812]$

1.2.4

See Figure 1.

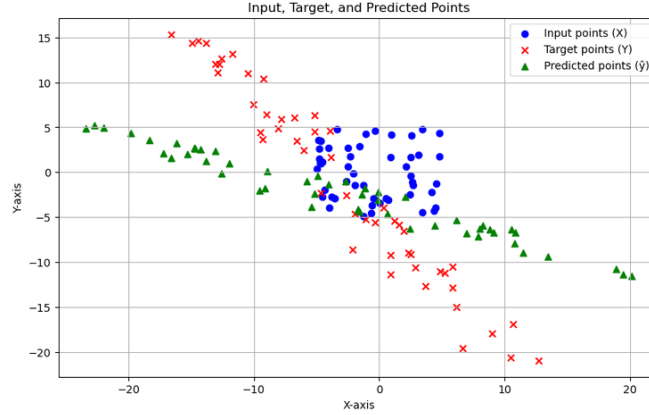


Figure 1: Q1-P4 X, Y, Y-Hat

1.2.5

Matrix S represents how the input points (x) are transformed through scaling and rotation (linear transformation). Each input point is multiplied by matrix S , changing its size and direction. Vector t represents how the linearly transformed input points are positionally shifted (translation).

For this particular fitting result, we see that while the scale or variance, and perhaps the translation too, of the predicted distribution is quite good, the rotation is significantly off.

1.2.6

See Figure 2.

1.2.7

See Figure 3.

1.2.8

All of the homographic transformations seem to make sense except for one, case 4. These transformations are designed to map one set of points in a 2D plane to another set of points in a different plane. The H matrix captures the relationship between two views of the same planar surface under different camera perspectives. In all of the cases except 4, we see the transformed X -points

Case #0 H:	Case #4 H:
[[1.00555949e+00 1.61370672e-03 -1.35143989e-01	[[-1.00000000e+00 5.46837533e-14 7.60000000e+01
[2.56045861e-03 6.22536404e-01 -7.35872070e-01	[-5.53591078e-13 -1.00000000e+00 1.01000000e+02
[4.51704286e-05 3.59823762e-05 1.00000000e+00]]	[-1.01342064e-14 4.55531693e-15 1.00000000e+00]]
Case #1 H:	Case #5 H:
[[6.20394776e-01 1.50050345e-03 -6.60977645e-01	[[4.01387046e-01 6.36777993e-01 -1.79445757e+00
[4.62565656e-05 1.00232229e+00 -4.06369835e-02	[6.40534201e-01 3.96745666e-01 -1.76059696e+00
[3.66035139e-07 2.25931463e-05 1.00000000e+00]]	[2.21034153e-04 1.39932850e-04 1.00000000e+00]]
Case #2 H:	Case #6 H:
[[6.46216419e-01 1.01850154e-02 -1.45233663e+00	[[7.70056604e-01 4.54255201e-02 -6.34008816e-02
[1.39156925e-02 6.46332905e-01 -1.60285342e+00	[-7.87253712e-02 8.63813194e-01 5.49671671e+00
[3.11050183e-04 2.76954685e-04 1.00000000e+00]]	[-3.05270237e-04 -6.37121535e-05 1.00000000e+00]]
Case #3 H:	Case #7 H:
[[2.01392556e-14 1.00000000e+00 -3.97173241e-13	[[8.41165566e-01 -4.86333708e-02 4.30457199e+00
[1.00000000e+00 3.34001622e-16 -9.19382468e-14	[-4.77384685e-02 9.86999947e-01 3.56055033e+00
[1.14114340e-17 2.31083165e-18 1.00000000e+00]]	[-4.55556648e-04 1.00171485e-04 1.00000000e+00]]

(a) Cases 0-3

(b) Cases 4-7

Figure 2: Q1-P6 H-Matrices

approximating their respective Y-points. In 4, the predicted points match the flipped Y-image orientation, but appear diagonal positionally to the Y-image.

2 Question 2: Image Stitching

2.0.1

See Figure 4.

2.0.2

The SIFT detector used to generate the images for this part was initialized with parameters of 0.15 for contrast threshold and 7 for edge threshold. I found that these parameters produced a balanced number of keypoint and descriptors. See Figure 5.

2.0.3

For calculating the pair-wise descriptor distances, the system I settled on was first normalizing both descriptors to have zero mean and unit standard deviation, then simply calculating the euclidean distance for each pair combination to form the distance matrix. This approach for the tower images produced a distance matrix with dimensions (491, 444), or 218004 total pair distances.

2.0.4

For selecting the putative matches based on pair-wise descriptor distances, my approach was to implement a threshold filter system. Any key-point pair whose distance was below the given threshold was selected as a putative match. The threshold value I ended up settling on was a value of 8.0, which for the tower images produced 500 matches or 0.02% of the total distance pairs.

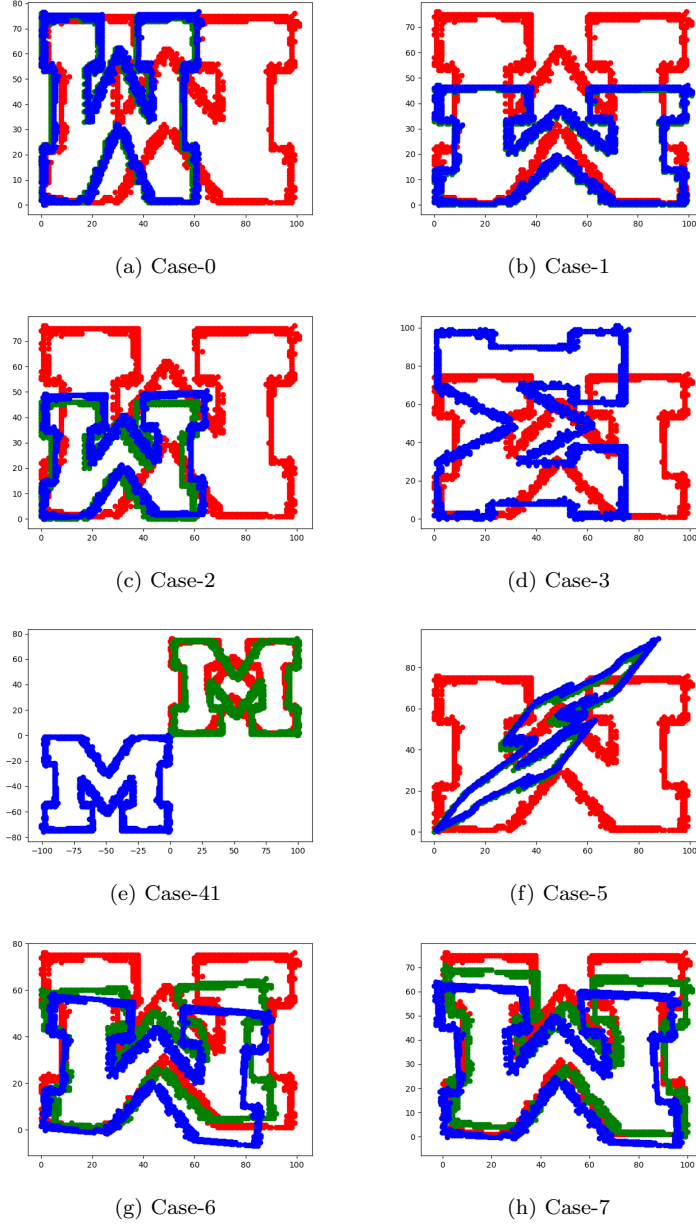


Figure 3: Q1-P7 Homographic Transformations

2.0.5

My RANSAC algorithm implementation in its final version used the parameters of 4 for the number of samples each iteration, 10,000 iterations, and an epsilon

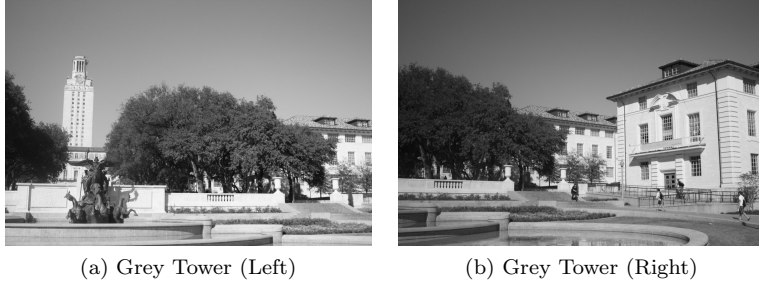


Figure 4: Q2-P1 Greyscale Tower Images

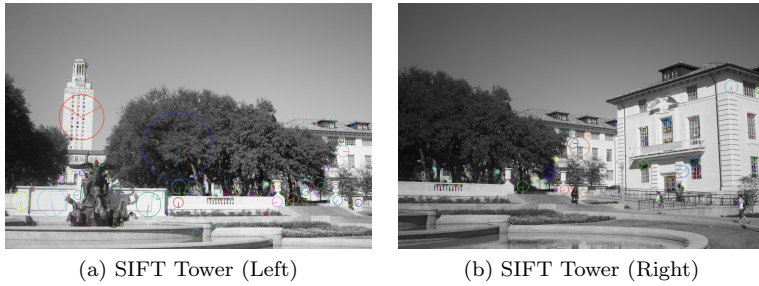


Figure 5: Q2-P2 SIFT Tower Images

error threshold value of the standard deviation of the Y-points divided by 2. For each iteration, the data matrix (X-Y) is shuffled and 4 (p) samples are selected. I then calculate the H (M?) matrix and the transformed X-points (\hat{Y}) using the `fit_homography()` `homography_transform()` functions implemented earlier, respectively. The residual, or error, is then calculated as the L2-norm between each Y and \hat{Y} point; any point pair with error less than epsilon is determined to be an inlier point.

The best model produced using this configuration had 27 inlier points (see Figure 6), an average residual of 454.07, and the following H-matrix values:

$$\begin{bmatrix} 8.87476543e-01 & 1.68410835e-01 & 4.15649007e+02 \\ -1.11381327e-01 & 1.07615150e+00 & 3.67730546e+01 \\ -1.62874767e-04 & 1.26696814e-04 & 1.00000000e+00 \end{bmatrix}$$

There were other configurations I experimented with that actually produced a lower average residual and higher number of inlier points, but for one reason or another this configuration produced a better quality panorama image of the towers as shown later on.

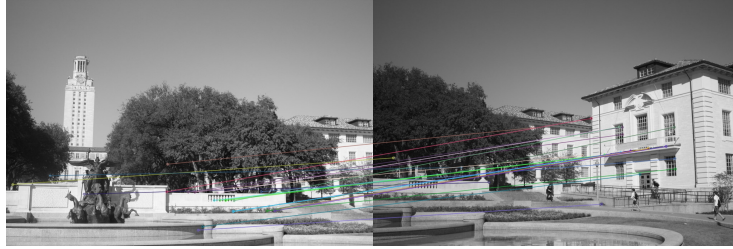


Figure 6: Q2-P5 RANSAC Tower Inliers

2.0.6

For the image warping and RANSAC calculations, I implemented a right-to-left image approach - meaning the right image was warped and the right keypoints were passed as the X-matrix for RANSAC. The dimensions for the warped image were calculated as a combination of the two given image widths, and the maximum image height. See Figure 7.



Figure 7: Q2-P6 Warped Tower (Right)

2.0.7

Creating the panorama is fairly straightforward, assuming the warping was calculated correctly. The warped image dimensions were already sufficiently large enough for the left image (non-warped) to be sliced into warped image starting at the beginning of both axis. Because this model configuration produced a good quality warped right image, it wasn't necessary to adjust it at all. See Figure 8.

2.0.8

Figures (9, 10, 11) attached below display the results of each stage of stitching together the BBB images.

Using the same configuration and parameters as previously, a pairwise distance matrix with dimensions (330, 324), or a total size of 106,926 points, was produced and 1192 points were selected as putative matches - substantially more



Figure 8: Q2-P7 Panorama Tower

than the previous images. The best model produced by RANSAC had 117 inliers and an average residual of 200.316.

As clearly displayed in the BBB panorama figure, this panorama is significantly worse quality than the tower panorama. Some of the factors that may have led to this were the provided images had different dimensions, and the angles were significantly different between the two images. This would require further experimentation to configure appropriate hyperparameters.



Figure 9: Q2-P8 SIFT BBB Images

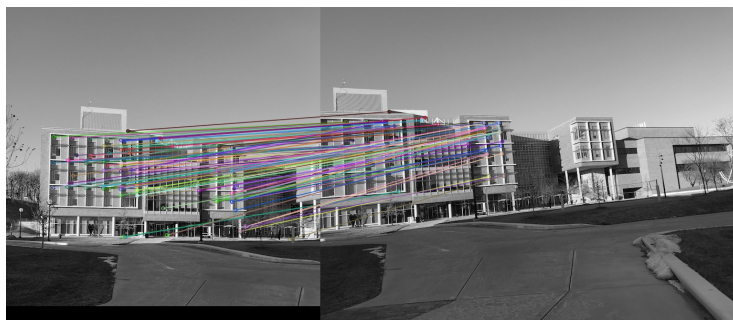


Figure 10: Q2-P8 RANSAC BBB Inliers



(a) Warped BBB



(b) Panorama BBB

Figure 11: Q2-P8 Warped and Stitched Images