CS 381V Visual Recognition
Coding Assignment 1
Ambika Verma
(av28944)

Algorithm Steps:

1. The 'object template' image as well as several test 'scene' images are read as input.
2. Each image is converted from RGB to grayscale before moving further.
3. Keypoints and SIFT descriptors for 'object template' image are computed.
4. Steps 5 to 11 are repeated for each test 'scene' image.
5. Keypoints and SIFT descriptors for each 'scene' image are computed.
6. Euclidean distance of each descriptor from scene image with respect to each descriptor from the template image is computed.
7. Scene descriptors at minimum distance from template descriptors are saved separately in a match matrix.
8. Apply a raw threshold (0.8*mean distance) on the minimum Euclidean distance obtained
9. Apply Lowe's ratio test
10. Apply Random Sample Consensus (RANSAC) to eliminate outliers and obtain the best affine transformation parameters
11. Based on a threshold on the number of matches remaining after preceding steps decide if the object is present or not in the test scene image, draw a bounding box if the object is present

Results:

1.

The following results are obtained by applying the following thresholds successively:

- Raw threshold on Euclidean distance = 0.8*mean distance
- Lowe's ratio threshold = 0.6 (ratio of distance to nearest neighbor/distance to second nearest neighbor)
- RANSAC (wherein inlier threshold distance is 1, i.e. a match is considered an inlier for an affine transformation if the transformed and actual x,y positions are less than 1 unit apart).

Test 1:

Test one comprises of using object-template.jpg and object-template-rotated.jpg. Both images are shown below (in grayscale) -

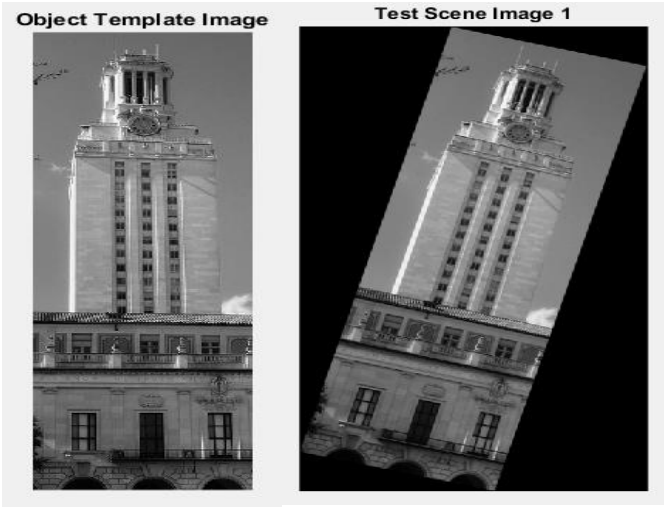

Fig.1 (Left) Object Template image, (Right) Scene Image 1

When SIFT is applied to both the images we obtain 510 and 562 features for object template and scene image respectively.

For each of the 510 template features we find the nearest neighbor match in the scene image. These 510 matches are shown below in Fig.2 (Top Left). A number of incorrect matches can clearly be seen.

Applying the raw threshold on nearest neighbors helps in removing the completely wrong matches (which can be seen in Fig.2 (top left) as the ones running diagonally across the images).

The Lowe's ratio test and RANSAC further filter these matches, though the filtering is not very drastic as the scene image is only the rotated version of the object template and thus a number of consistent invariant features are obtained.

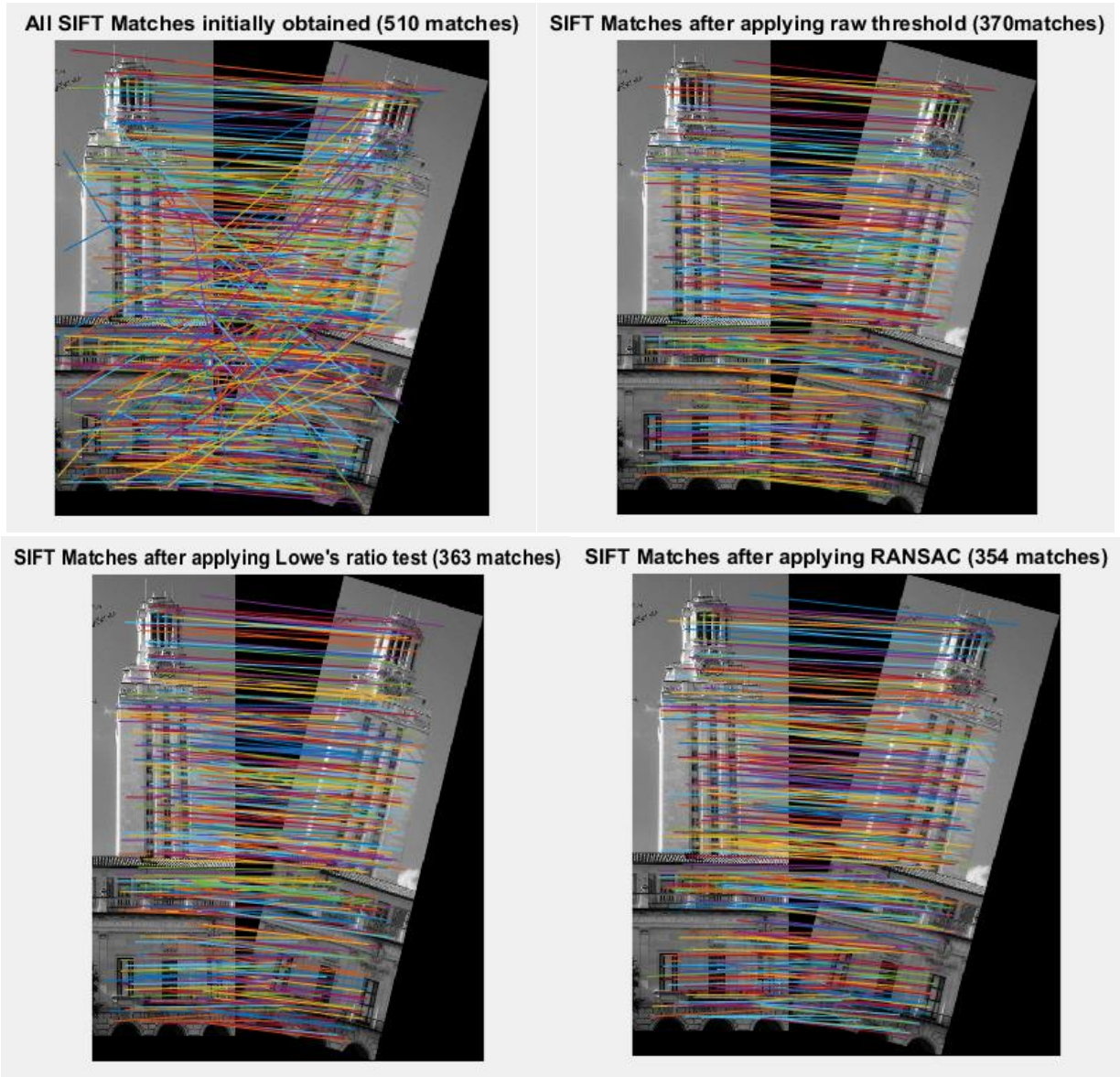

Fig.2 (Top Left) Initial SIFT matches, (Top Right) SIFT matches after thresholding nearest neighbors, (Bottom Left) SIFT matches after applying Lowe's ratio test, (Bottom Right) SIFT matches i.e. inliers after applying RANSAC

Test 2:

Test 2 comprises of using object-template.jpg and scene1.jpg. Both images are shown below (in grayscale) -

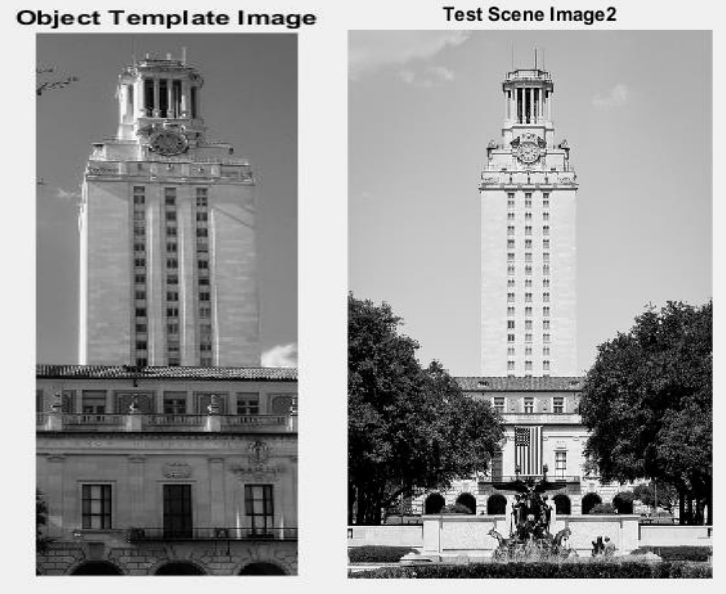

Fig.3 (Left) Object Template image, (Right) Scene Image 2

When SIFT is applied to both the images we obtain 510 and 893 features for object template and scene image respectively.

For each of the 510 template features we find the nearest neighbor match in the scene image. These 510 matches are shown in Fig.4 (Left). A number of incorrect matches can clearly be seen.

Applying the raw threshold on nearest neighbors helps in removing some of the wrong matches, shown in Fig.4 (Right), but is not able to remove all the incorrect matches effectively.

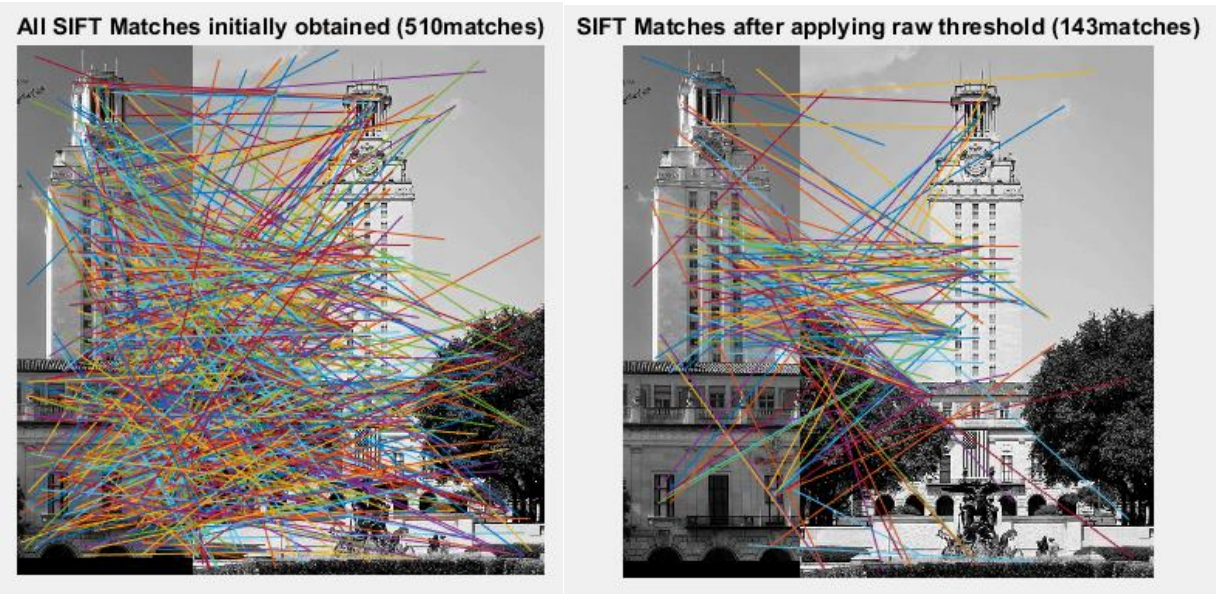

Fig.4 (Left) Initial SIFT matches, (Right) SIFT matches after thresholding nearest neighbors

The Lowe's ratio test and RANSAC further filter these matches, and their advantage can easily be seen in the following figures.

Thus, we can say that raw thresholding on nearest neighbors does provide effective robustness in case of simple transformation of the object template itself. But, it suffers when more complex conditions such as viewpoint, illumination, occlusion, scale change etc. occur. Lowe's ratio test and RANSAC provide the required robustness to these variations, thus making SIFT considerably invariant to the same. In this case, it is really the Lowe's ratio test which filters out incorrect matches extensively.

Additionally, RANSAC further filters for correct matches through spatial verification process and removes outliers which do not agree with other matches for a particular affine transformation (for example, an incorrect match between the building and fountain is removed).

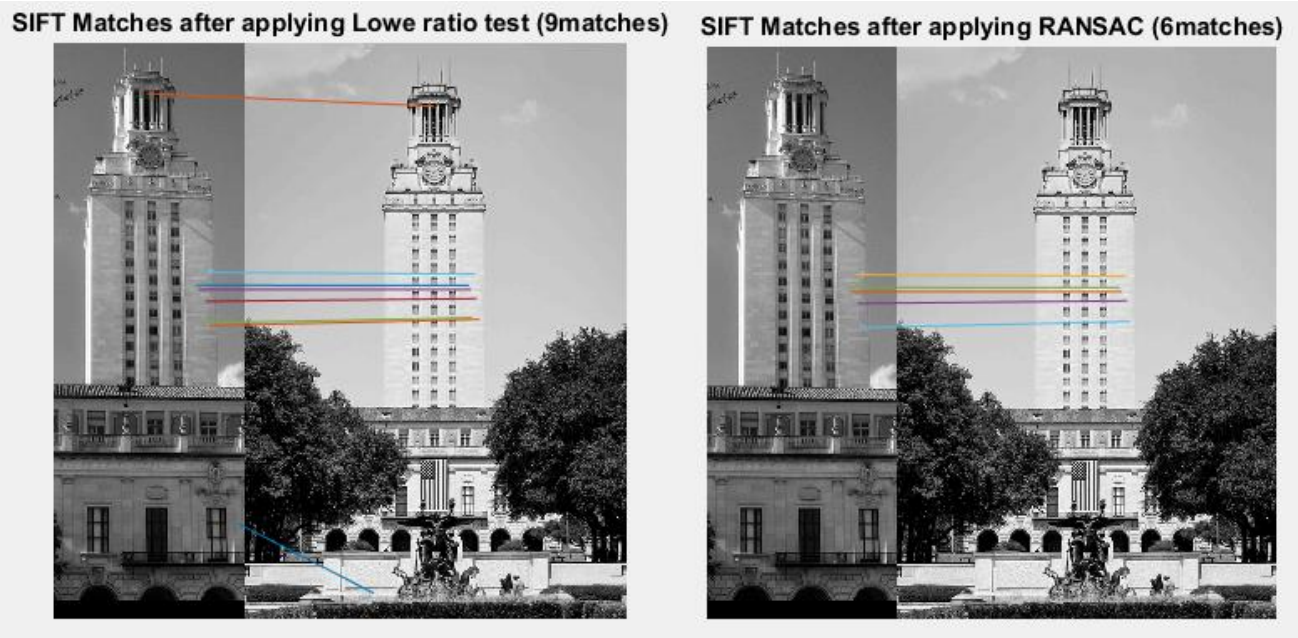

Fig.5 (Left) SIFT matches after applying Lowe's ratio test, (Right) SIFT matches i.e. inliers after applying RANSAC

Test 3:

Test 3 comprises of using object-template.jpg and scene2.jpg. Both images are shown below (in grayscale) -

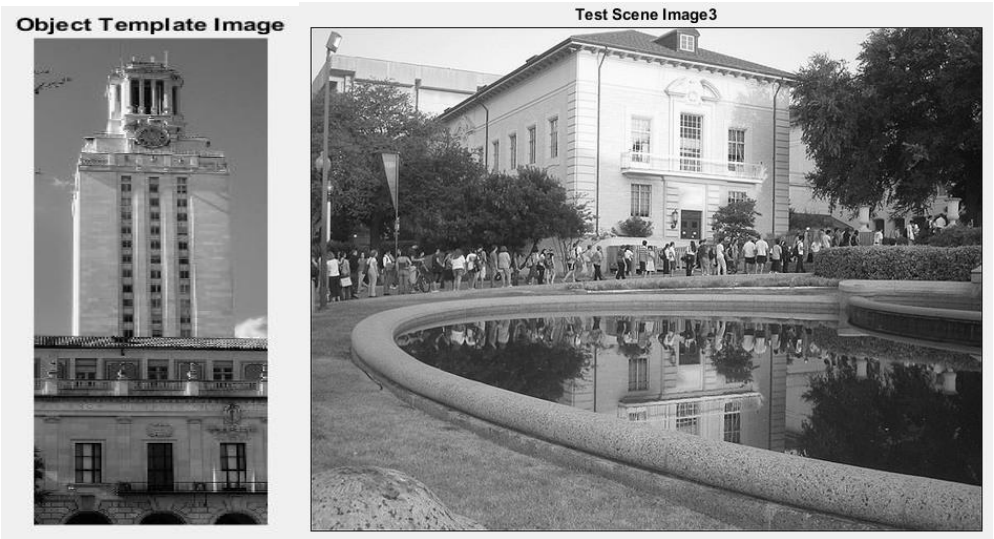

Fig.6 (Left) Object Template image, (Right) Scene Image 3

When SIFT is applied to both the images we obtain 510 and 1057 features for object template and scene image respectively.

For each of the 510 template features we find the nearest neighbor match in the scene image. These 510 matches are shown below in Fig.7 (Top Left). A number of incorrect matches can clearly be seen.

Applying the raw threshold on nearest neighbors helps in removing some of the wrong matches, but there are still some matches left even though the object is not present in the scene.

The Lowe's ratio test further cuts down on the number of matches. But, we cannot base our detections just based on the number of matches after applying Lowe's ratio test, since we still see 6 matches (which is comparable to 9 matches we saw in previous case) even though the object is not present in the scene.

This limitation is overcome by using RANSAC for spatial verification since an affine transformation should exist if the object is present in the scene. As can be seen below, only 3 matches survive after applying RANSAC and since our detection threshold is set to be greater than 3, the object is not detected in the given scene.

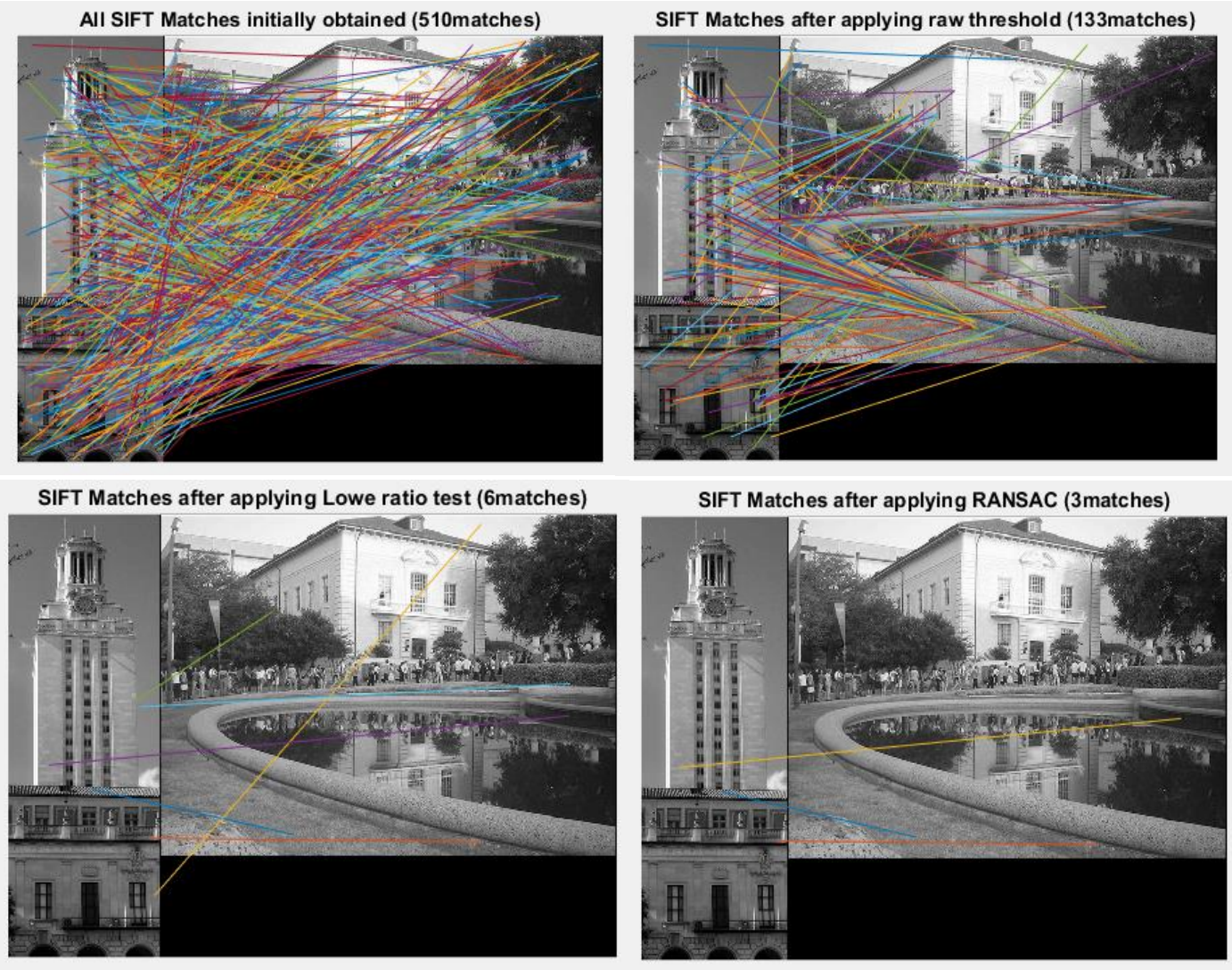

Fig.7 (Top Left) Initial SIFT matches, (Top Right) SIFT matches after thresholding nearest neighbors, (Bottom Left) SIFT matches after applying Lowe's ratio test, (Bottom Right) SIFT matches i.e. inliers after applying RANSAC

2.

A threshold of 3 is used on the number of inliers (after applying RANSAC) to classify an object as being present in the scene images. This threshold was arrived at by empirical observations made.

The identification is almost perfect for the object-template-rotated.jpg scene (Fig.8 (Top Left)).

For scene1.jpg (pictured in Fig. 8 (Top Right)) the bounding box constructed is not perfect, which seems reasonable since SIFT provides us with local features, which under image variations would not result in the perfect affine transformation matrix (which is intrinsically global in nature).

For scene2.jpg, since the number of inliers after RANSAC is 3 which is lower than our object detection threshold, the object is not detected and a bounding box is not generated.

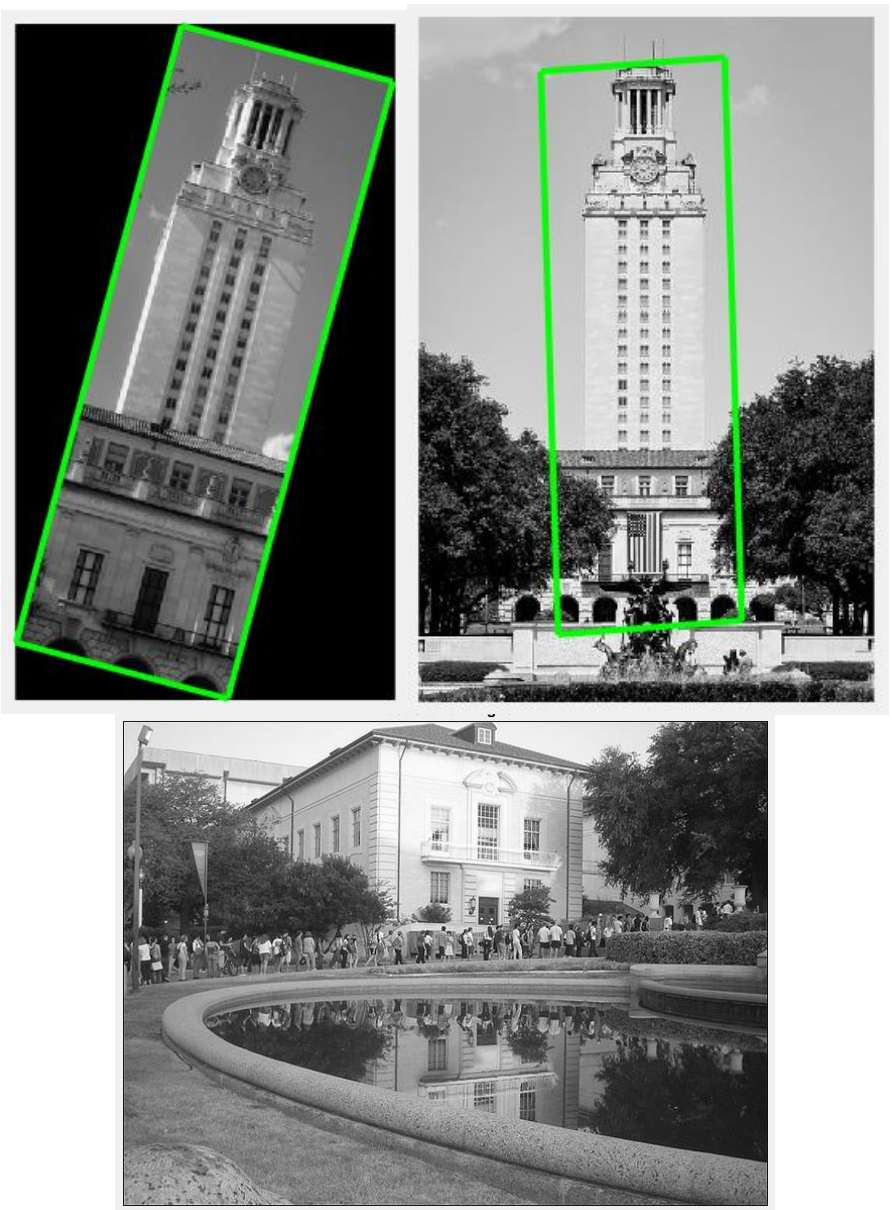

Fig.8 (Top Left) Object detected in object-template-rotated.jpg, (Top Right) Object detected in scene1.jpg, (Bottom center) Object not detected in scene2.jpg (as expected)

Extra Credits:

1.

Below are a set of images of Christ Church which I used for further testing SIFT. I believe the difficulty/ complexity of the test images In terms of extensive scale, rotation, translation, illumination etc. variations increases successively.

Fig. 9 shows the results from applying the SIFT procedure to a fairly less variation in terms of illumination, rotation and scale.

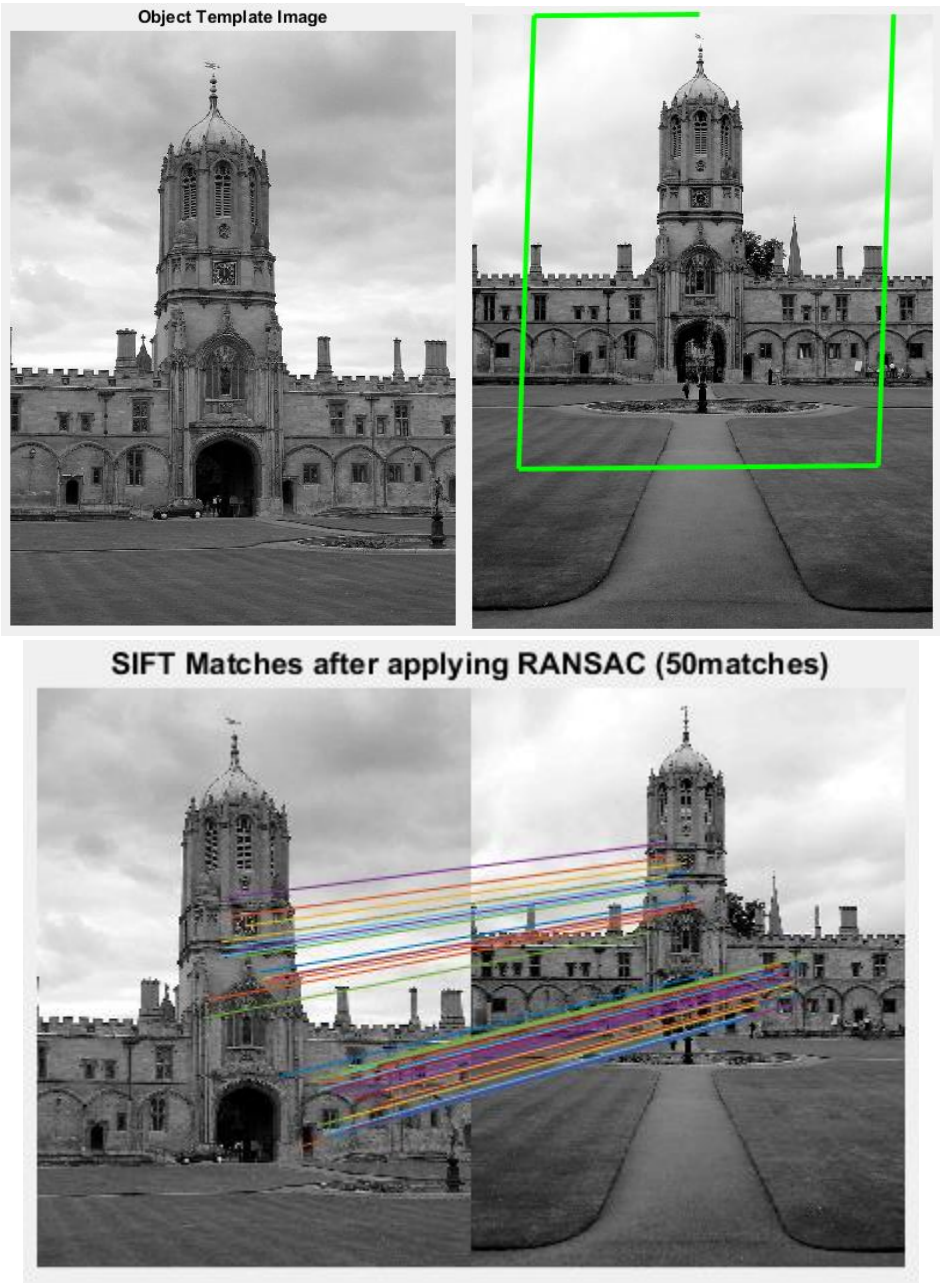

Fig.9 (Top Left) Object template of Christ Church used, (Top Right) Object detected in test image 1, (Bottom) Number of SIFT matches after applying the thresholds and RANSAC successively

Fig. 10 shows the results from applying the SIFT procedure to a much larger variation in terms of illumination and rotation though the scale remains approximately the same. We are still able to detect the object, thus showing the robustness of SIFT and RANSAC.

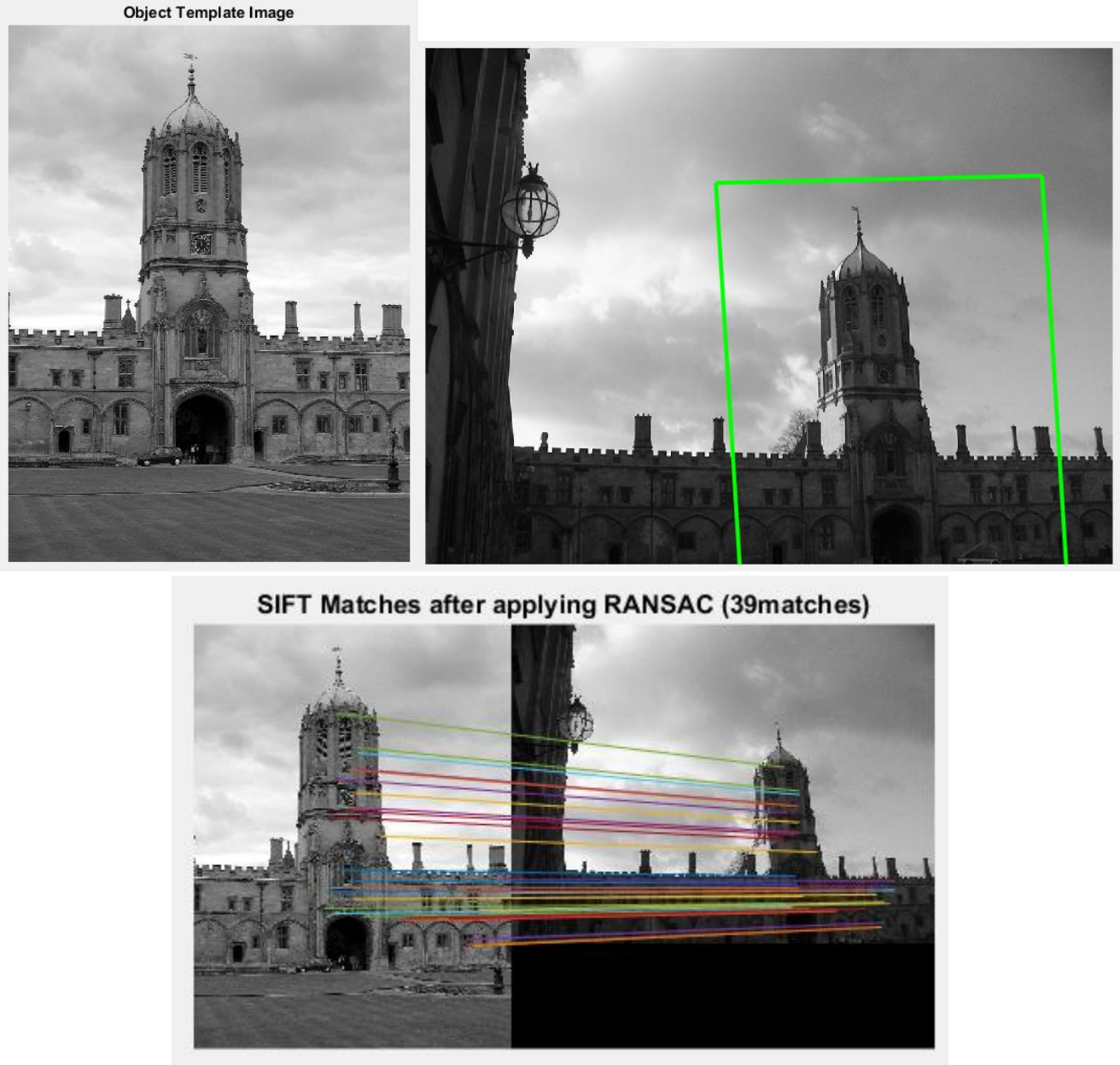

Fig.10 (Top Left) Object template of Christ Church used, (Top Right) Object detected in test image 2, (Bottom) Number of SIFT matches after applying the thresholds and RANSAC successively

Fig. 11 shows the results from applying the SIFT procedure to more drastic variations in terms of illumination, rotation and scale, as well as a person is also part of the view. It is interesting to note that even after such extreme variations we are still able to detect the object.

Though it is worthy to note that as the degree and severity of variations increase the number of SIFT matches obtained keeps on decreasing, thus pointing us to the observation that SIFT is only partially invariant to such variations and there exists a limit after which the designed process breaks down. Figures 12 and 13 shown below (which also contain fairly severe variations in terms of scale, illumination and rotation) further provide evidence for the mentioned observation. *Cases shown in fig. 12 and 13 are really borderline scenarios where the affine transformation obtained is not always as expected.

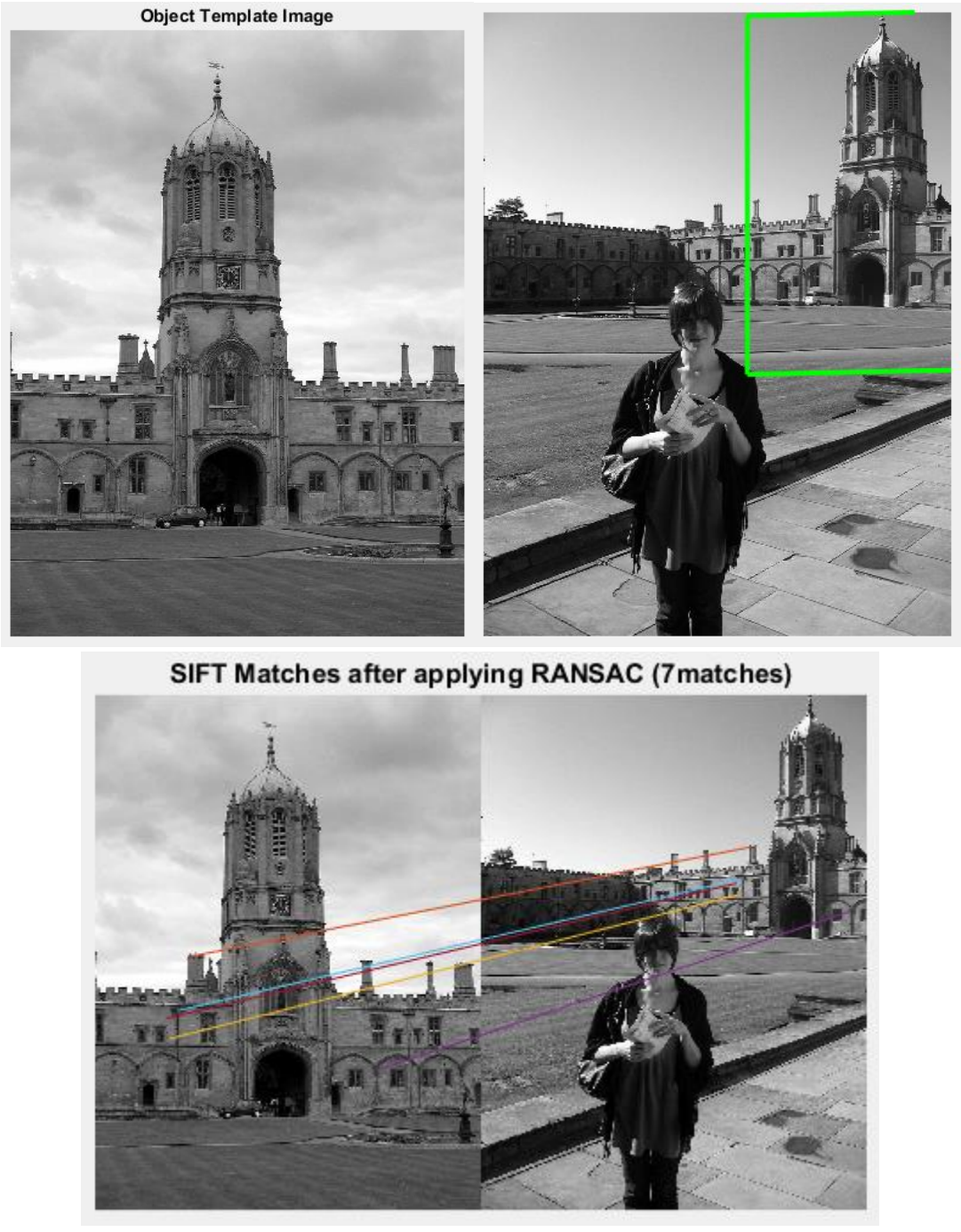

Fig.11 (Top Left) Object template of Christ Church used, (Top Right) Object detected in test image 3, (Bottom) Number of SIFT matches after applying the thresholds and RANSAC successively

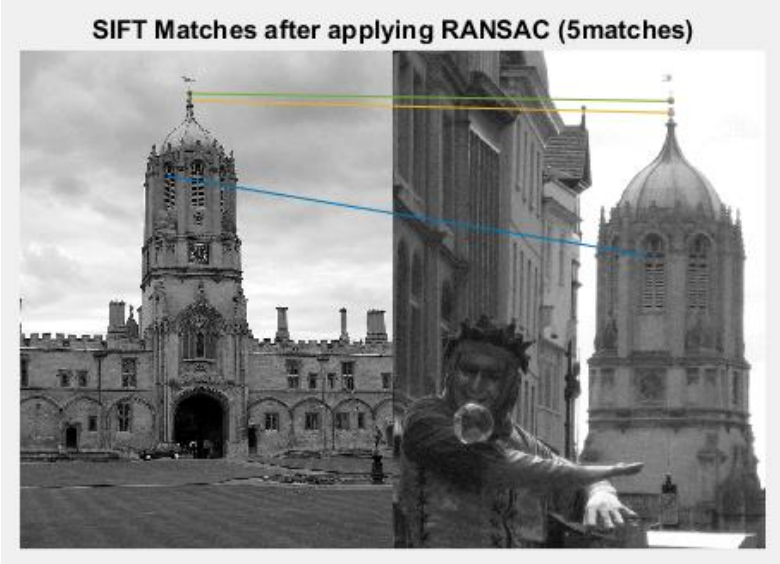

Fig.12 Number of SIFT matches after applying the thresholds and RANSAC successively

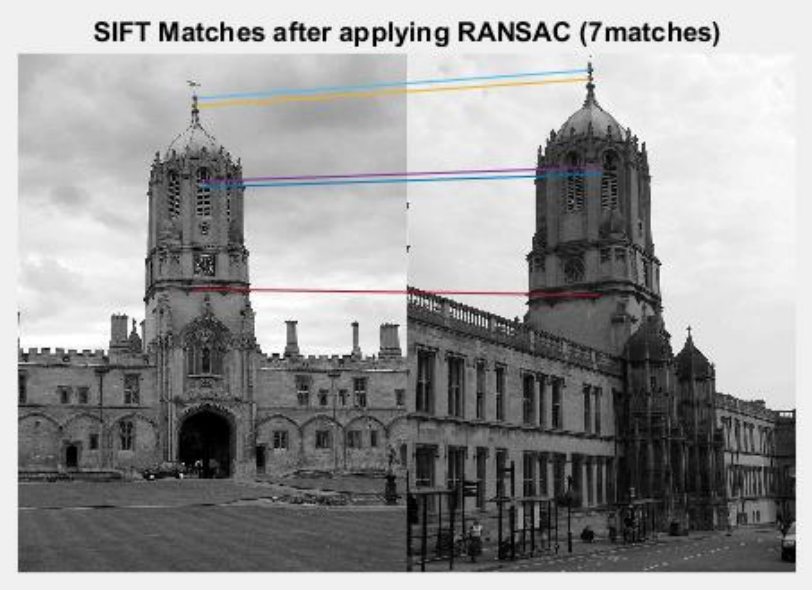

Fig.13 Number of SIFT matches after applying the thresholds and RANSAC successively

Next, we see some scenarios where the SIFT method applied fails to detect the object in the given scene images. Figure 14 is one such case where the illumination change is very drastic (changes from day to night).

Though the number of inliers after RANSAC is more than 3 (which satisfies the threshold we have specified for object detection), these matches as shown in the figure do not correspond to the same aspects of the object and hence result in a very inaccurate affine transformation.

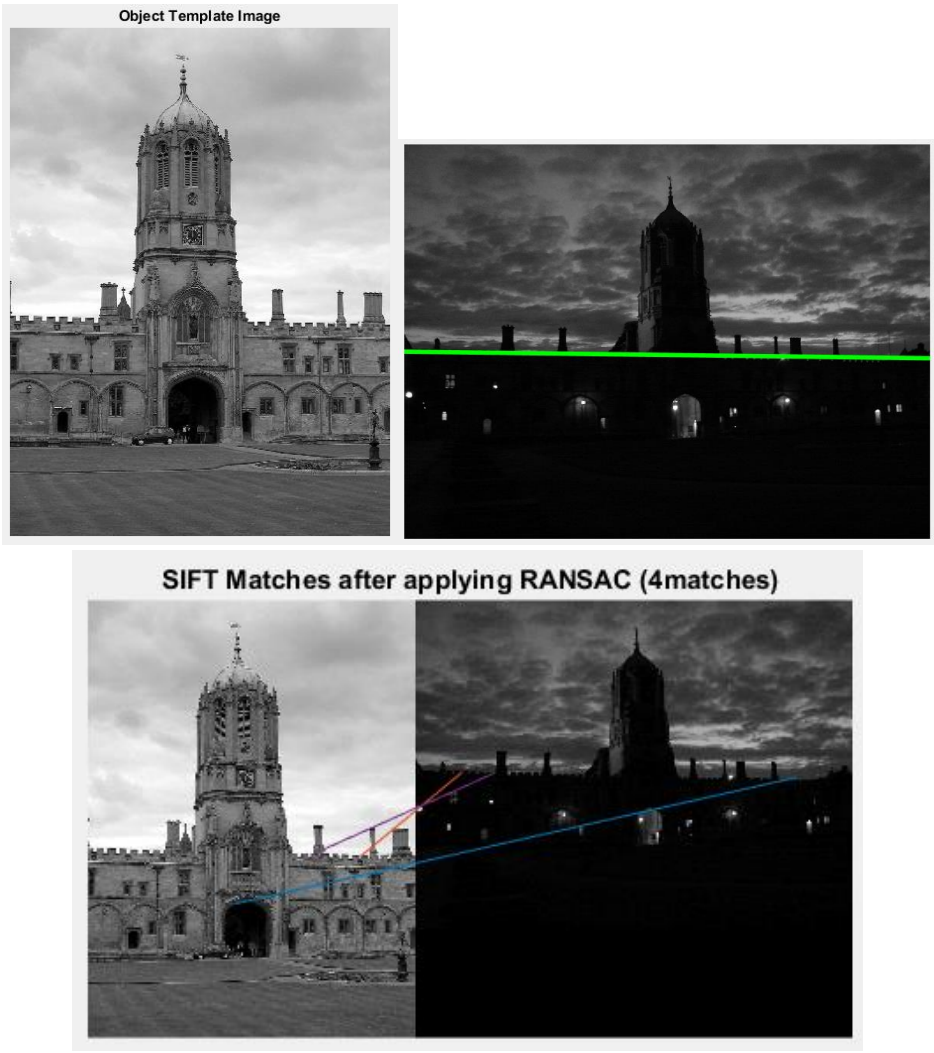

*Fig.14 (Top Left) Object template of Christ Church used, (Top Right) Object **NOT** detected in test image 3, (Bottom) Number of SIFT matches after applying the thresholds and RANSAC successively (**incorrectly matched**)*

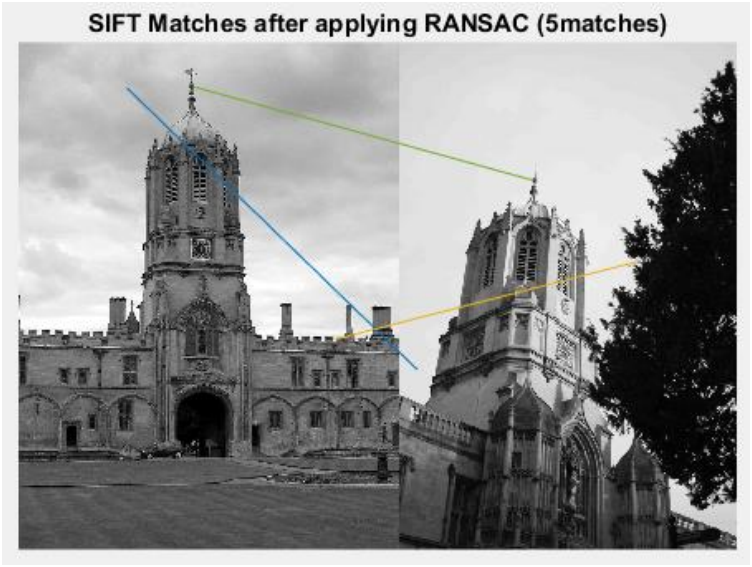

*Fig.15 Number of SIFT matches after applying the thresholds and RANSAC successively (**incorrectly matched**)*

2.

Below is the case shown where the number of ratio test neighbors is high even though the object is not present in the scene image. Figure 16 and 17 use the same Christ Church object template image but the scene images are now of very different structures (which clearly are not the object we are looking for).

In both cases a high number of SIFT matches are obtained (51 and 45 respectively) after applying Lowe's ratio test. If our object detection was solely based on Lowe's ratio test we would have erroneously detected the object in these scenes.

These, examples highlight the effectiveness of spatial verification (implemented here as RANSAC) which prevents us from making false detections. As the number of inliers after applying RANSAC is only 3, which does not satisfy our object detection threshold set, resulting in no false detections.

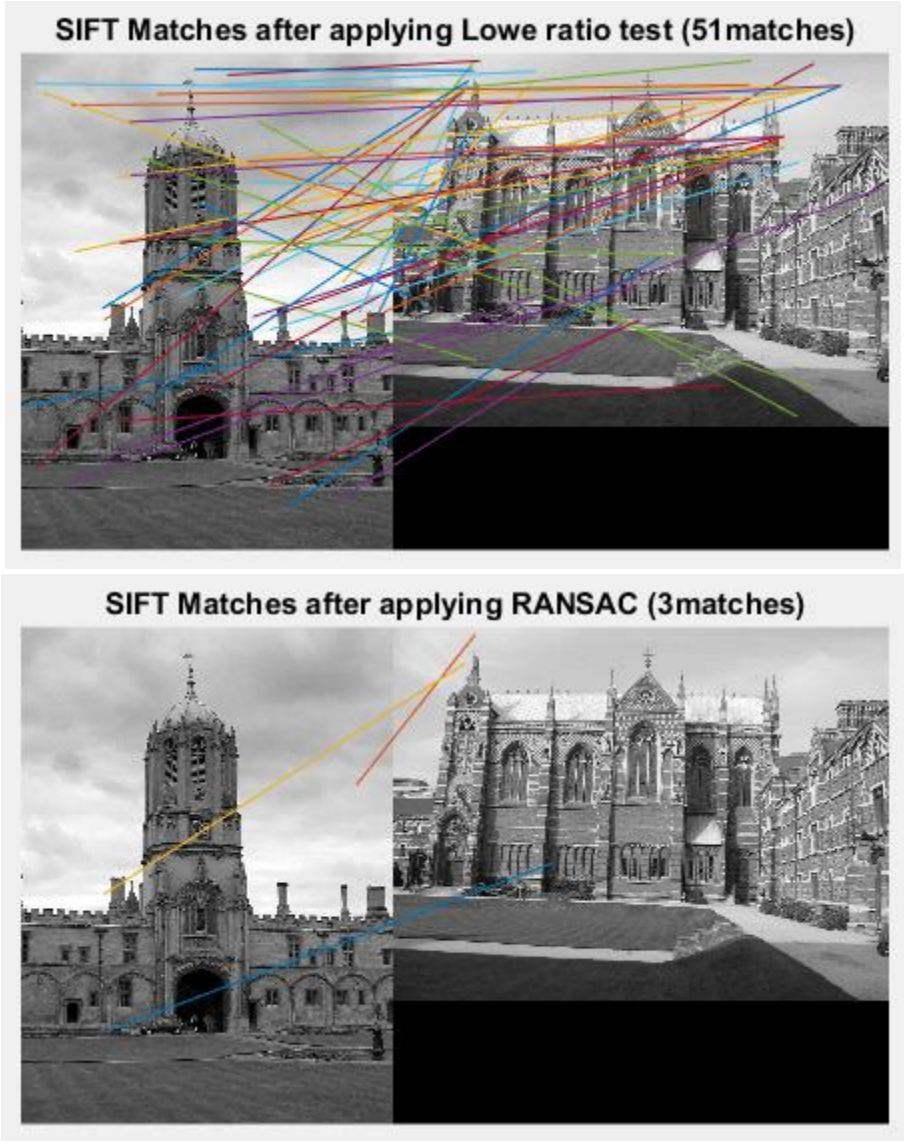

Fig.16 (Top) SIFT matches remaining after Lowe's ratio test (Bottom) Number of SIFT matches after applying RANSAC

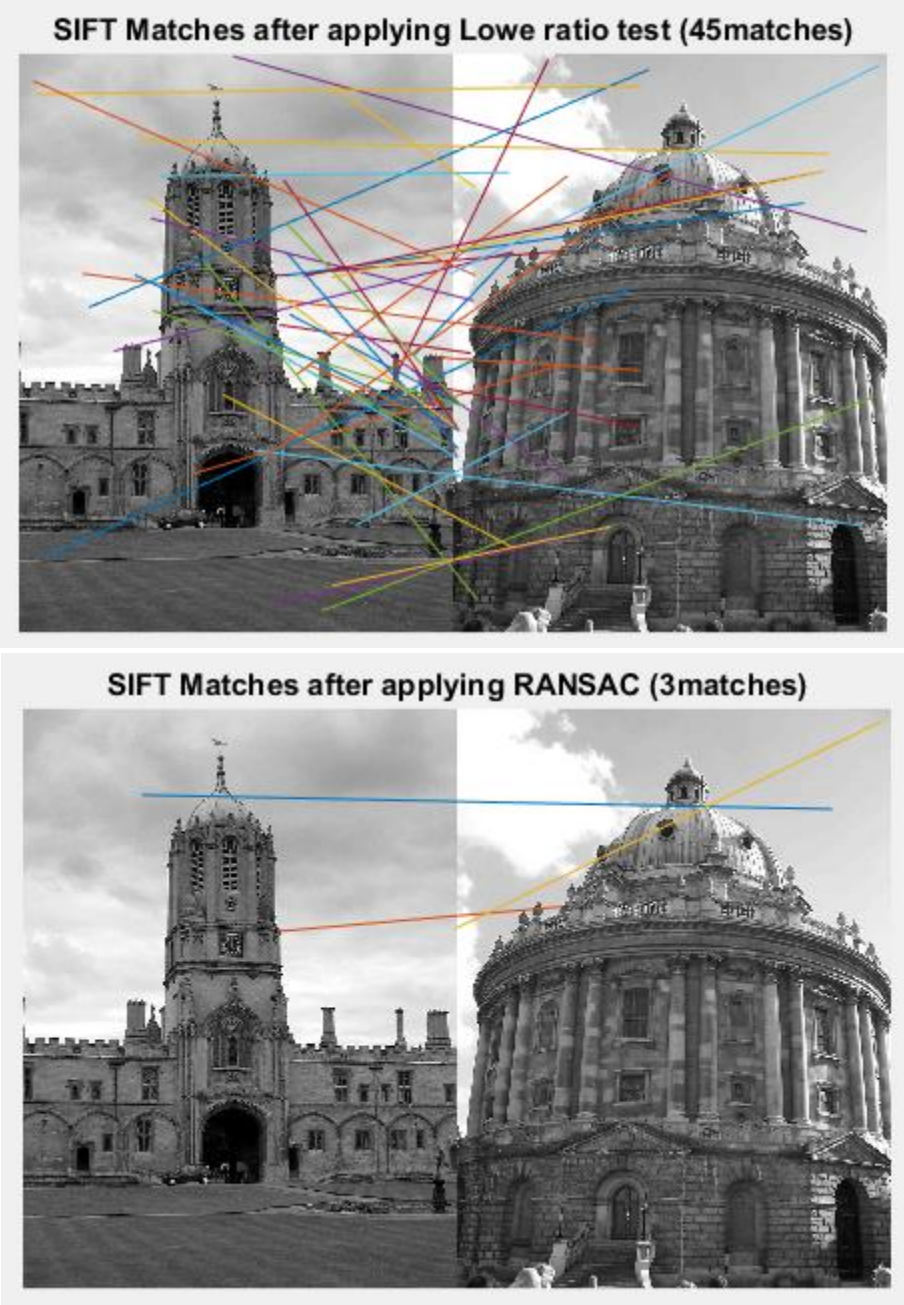

Fig.17 (Top) SIFT matches remaining after Lowe's ratio test (Bottom) Number of SIFT matches after applying RANSAC

References:

- Object Recognition from Local Scale-Invariant Features, Lowe, ICCV 1999
- CS 381V Visual Recognition course slides
- http://www.robots.ox.ac.uk/~vgg/practicals/instance-recognition/index.html
- VLFeat: An Open and Portable Library of Computer Vision Algorithms, A. Vedaldi and B. Fulkerson, 2008, http://www.vlfeat.org/