**SIFT Detector Using FLANN:**

Object image: object.png

Scene image: scene.png

Detection method: sift

Matcher method: flann

Detecting image features...

Detected 1954 features in object in 0.827 seconds

Detected 7005 features in scene in 3.109 seconds

Found 1954 matches in 0.584 seconds

Finding good matches...

Found 720 good matches and 1440 good features from them in 0 seconds

Calculated the homography matrix in 0.01 seconds:

[0.7469067684415254, -0.4193402701464118, 788.8821724627741;

0.1669210682087139, 0.7067916864455451, 205.76387074164;

-9.509383300344636e-05, -0.0001846517030902413, 1]

Number of inliers: 420 out of 720 good matches.

Transformed Box Coordinates:

[788.882, 205.764]

[1311.88, 324.427]

[1118.19, 1162.94]

[509.867, 965.092]

**SIFT Detector Using BFM:**

Running object detection...

Object image: object.png

Scene image: scene.png

Detection method: sift

Matcher method: bfm

Detecting image features...

Detected 1954 features in object in 0.825 seconds

Detected 7005 features in scene in 3.009 seconds

Found 1954 matches in 4.295 seconds using BFM

Finding good matches...

Found 730 good matches and 1460 good features from them in 0 seconds

Calculated the homography matrix in 0.006 seconds:

[0.7643542851018954, -0.399449524550603, 784.8597640802167;

0.1779455817391118, 0.7339702943747934, 197.5394926825597;

-8.679102866451919e-05, -0.0001661868033253011, 1]

Number of inliers: 478 out of 730 good matches.

Transformed Box Coordinates:

[784.86, 197.539]

[1311.78, 321.024]

[1118.47, 1159.83]

[515.659, 964.864]

**ORB Detector Using FLANN:**

Object image: object.png

Scene image: scene.png

Detection method: orb

Matcher method: flann

Using ORB detector, setting matchers to use LSH index...

Detecting image features...

Detected 4998 features in object in 0.288 seconds

Detected 5000 features in scene in 0.85 seconds

Found 4998 matches in 2.456 seconds using FLANN

Finding good matches...

Found 400 good matches and 800 good features from them in 0 seconds

Calculated the homography matrix in 0.015 seconds:

[0.7670673388622371, -0.3934121555753614, 780.7922455852331;

0.1857554777052303, 0.7346164458659871, 195.7023262282622;

-7.97405541991974e-05, -0.000168220719939384, 1]

Number of inliers: 190 out of 400 good matches.

Transformed Box Coordinates:

[780.792, 195.702]

[1303.39, 322.59]

[1118.42, 1160.47]

[518.011, 965.333]

**ORB Detector Using BFM:**

Running object detection...

Object image: object.png

Scene image: scene.png

Detection method: orb

Matcher method: bfm

Using ORB detector, setting BFMatching to use Hamming...

Detecting image features...

Detected 4998 features in object in 0.291 seconds

Detected 5000 features in scene in 0.826 seconds

Found 4998 matches in 1.668 seconds using BFM

Finding good matches...

Found 48 good matches and 96 good features from them in 0 seconds

Calculated the homography matrix in 0.003 seconds:

[0.7712601721747161, -0.3860703668660186, 779.9561851950755;

0.1824239169079512, 0.7515230484930913, 190.6671868786122;

-8.502815401672418e-05, -0.000156950594959991, 1]

Number of inliers: 39 out of 48 good matches.

Transformed Box Coordinates:

[779.956, 190.667]

[1309.52, 316.256]

[1119.3, 1160.39]

[518.536, 965.49]

**Speed Test**

**Method:**

* This test used the given object and scene images from the skeleton codebase, it started by timing the time taken for each calculation needed for the detection using detector standards defined by the results from the accuracy test, i.e., the parameters that guaranteed an accurate match

***Used object.png and scene.png***

**====== Starting SIFT test ======**

Setting detector to SIFT detector...

Setting matcher to FLANN matcher...

Detecting image features...

Detected 1954 features in object in 0.805 seconds

Detected 7005 features in scene in 3.047 seconds

Matching descriptors...

Found 1954 matches in 0.625 seconds using FLANN

Finding good matches...

Found 720 good matches in 0 seconds

**# Repeating the test using BFM matcher #**

Detecting image features...

Detected 1954 features in object in 0.838 seconds

Detected 7005 features in scene in 2.944 seconds

Matching descriptors...

Found 1954 matches in 4.417 seconds using BFM

Finding good matches...

Found 730 good matches in 0 seconds

**====== Starting ORB test ======**

Setting detector to ORB detector...

Setting matcher to FLANN matcher...

Detecting image features...

Detected 500 features in object in 0.229 seconds

Detected 500 features in scene in 0.768 seconds

Matching descriptors...

Using ORB detector, setting matchers to use LSH index...

Found 500 matches in 0.144 seconds using FLANN

Finding good matches...

Found 87 good matches in 0 seconds

**# Repeating the test using BFM matcher #**

Detecting image features...

Detected 500 features in object in 0.21 seconds

Detected 500 features in scene in 0.774 seconds

Matching descriptors...

Using ORB detector, setting BFMatching to use Hamming...

Found 500 matches in 0.016 seconds using BFM

Finding good matches...

Found 44 good matches in 0 seconds

**====== Starting averaging test ======**

Average time taken for SIFT detector with FLANN matcher: 4.33565 seconds

Average time taken for SIFT detector with BFM matcher: 8.38425 seconds

Average time taken for ORB detector with FLANN matcher: 1.1367 seconds

Average time taken for ORB detector with BFM matcher: 1.00045 seconds

**Accuracy Test**

**Method:**

* To determine the true size of the object in the scene, OpenCV features were used to present the user with a photo of the object (for reference) and the corresponding scene image, the user would then select the visible corners of the object in scene, this gave the “ground truth” corners (with a user error margin of 5pixels), allowing the boundary box detection corners to be compared using Euclidean distance between the vectors. To minimize the user errors in corner selection, default corners were defined and used as so the variables remained constant across the different tests
* Tests were ran with different objects and scene combinations, 5 tests using the ORB (Orientated FAST and Rotated BRIEF) detector and 5 using the SIFT (Scale-Invariant Feature Transform). Each detector was given the same Lowe’s Ratio constraint and max feature cap, as well as the same image datasets for consistency.
* Pixel errors were calculated from the detected boundary box corners against the user defined corners of the object in scene. This gave an average “pixel error” for the image combination for each detector as well as average total errors over the test, this gave a standard to adjust detector parameters to determine an increase or decrease in accuracy

**Notes:**

For context here are true sizes of the image files:

object: 2171x2648

scene1: 4032x2531 (object needed to be scaled for this one)

scene2: 4032x3024 (object was naturally detected in this one)

Manually selecting the corners of the object in scene1 i got these corners;

Selected corner: (667, 1739)

Selected corner: (1021, 1478)

Selected corner: (1315, 1893)

Selected corner: (950, 2142)

The ORB detector found these corners based on the results I shared before;

(959.48, 1511.92)

(960.186, 1513.81)

(961.665, 1517.03)

(959.864, 1514.53)

Meaning that it only found a very small box on the scene image

ORB is good for speed, but **SIFT (Scale-Invariant Feature Transform)** is **better for scale differences**.

Hmm this is strange, so that test was trying to find an object in scene1, and when the test was run again for scene2 it gave these results;   
Selected corner: (969, 1849)   
Selected corner: (1446, 1430)   
Selected corner: (1969, 2044)   
Selected corner: (1473, 2471)   
Detected Object Corners:   
[951.61, 1846.42]   
[1451.84, 1446.22]   
[1952.81, 2035.38]   
[1452.54, 2468.44]   
Errors:   
Corner 0 Error: 17.5807 pixels   
Corner 1 Error: 17.2331 pixels

Corner 2 Error: 18.3453 pixels

Corner 3 Error: 20.6227 pixels

Average Error: 18.4455 pixels   
Which as you can see only gave an average error of 18.5 pixels which is a lot better, even the bounding box was drawn correctly around the object. Now the only thing changed was the scene image, and the only real difference between the scenes is scene2 is a bit of a closer image and thus the object is larger in the image. Obviously the object will be detected easier but the fact that its only a bit larger and its able to almost perfectly find the object whereas in the other scene it cannot find it at all is odd, concluded this is due to the scale difference of the object in scene 1 where the object is substantially smaller than scene 2. A hacky work around was to scale the object image file by a scale of 0.3x (only if the scene was “md-scene1”) which got much better results:

**ORB Detector Accuracy Test Results**

======= Test 1: md.jpg + md-scene1.jpg =======

Using default values for md-scene1.jpg

Selected corner: (675, 1736)

Selected corner: (1024, 1478)

Selected corner: (1328, 1883)

Selected corner: (961, 2154)

Detected 11572 features in object.

Detected 49506 features in scene.

Found 11572 matches using BFM.

Found 1106 good matches and 2212 good features from them.

Number of inliers: 783 out of 1106 good matches.

Detected Object Corners:

[679.5, 1732.14]

[1024.02, 1486.65]

[1311.18, 1884.42]

[946.574, 2146.72]  
Pixel Errors;

Corner 0 Error: 5.93037 pixels

Corner 1 Error: 8.65482 pixels

Corner 2 Error: 16.8757 pixels

Corner 3 Error: 16.1601 pixels

Average Error: 11.9052 pixels

======= Test 2: md.jpg + md-scene2.jpg =======

Using default values for md-scene2.jpg

Selected corner: (961, 1849)

Selected corner: (1454, 1430)

Selected corner: (1972, 2044)

Selected corner: (1471, 2463)

Detected 17740 features in object.

Detected 50000 features in scene.

Found 17740 matches using BFM.

Found 953 good matches and 1906 good features from them.

Number of inliers: 474 out of 953 good matches.

Detected Object Corners:

[969.494, 1848.99]

[1451.07, 1435.09]

[1962.21, 2037.25]

[1455.35, 2460.1]   
Pixel Errors;  
Corner 0 Error: 8.49439 pixels

Corner 1 Error: 5.8753 pixels

Corner 2 Error: 11.8944 pixels

Corner 3 Error: 15.9193 pixels

Average Error: 10.5458 pixels

======= Test 3: book.jpg + b-scene1.jpg =======

Using default values for b-scene1.jpg

Selected corner: (41, 1576)

Selected corner: (978, 1549)

Selected corner: (957, 2754)

Selected corner: (16, 2800)

Detected 37563 features in object.

Detected 50000 features in scene.

Found 37563 matches using BFM.

Found 6086 good matches and 12172 good features from them.

Number of inliers: 4219 out of 6086 good matches.

Detected Object Corners:

[48.7008, 1572.02]

[971.441, 1554.2]

[932.914, 2744.39]

[10.611, 2806.22]   
Pixel Errors;

Corner 0 Error: 8.66895 pixels

Corner 1 Error: 8.37035 pixels

Corner 2 Error: 25.9336 pixels

Corner 3 Error: 8.22852 pixels

Average Error: 12.8004 pixels

======= Test 4: book.jpg + b-scene2.jpg =======

Using default values for b-scene2.jpg

Selected corner: (1372, 1085)

Selected corner: (2132, 654)

Selected corner: (2735, 1638)

Selected corner: (1987, 2094)

Detected 37563 features in object.

Detected 50000 features in scene.

Found 37563 matches using BFM.

Found 4995 good matches and 9990 good features from them.

Number of inliers: 3666 out of 4995 good matches.

Detected Object Corners:

[1368.7, 1088.44]

[2134.5, 660.802]

[2736.19, 1641.63]

[1992.65, 2097.16]   
Pixel Errors;  
Corner 0 Error: 4.76814 pixels

Corner 1 Error: 7.2462 pixels

Corner 2 Error: 3.82283 pixels

Corner 3 Error: 6.47348 pixels

Average Error: 5.57767 pixels

======= Test 5: book.jpg + b-sceneD.jpg =======

Using default values for b-sceneD.jpg

Selected corner: (41, 1576)

Selected corner: (978, 1549)

Selected corner: (957, 2754)

Selected corner: (16, 2800)

Detected 37563 features in object.

Detected 15371 features in scene.

Found 37563 matches using BFM.

Found 2436 good matches and 4872 good features from them.

Number of inliers: 1570 out of 2436 good matches.

Detected Object Corners:

[48.0941, 1572.3]

[971.059, 1554.08]

[933.399, 2744.19]

[11.0263, 2806.82]   
Pixel Errors;  
Corner 0 Error: 8.00329 pixels

Corner 1 Error: 8.60062 pixels

Corner 2 Error: 25.5572 pixels

Corner 3 Error: 8.44303 pixels

Average Error: 12.651 pixels

**SIFT Detector Accuracy Test Results**

======= Test 2: md.jpg + md-scene2.jpg =======

Using default values for md-scene2.jpg

Selected corner: (961, 1849)

Selected corner: (1454, 1430)

Selected corner: (1972, 2044)

Selected corner: (1471, 2463)

Detected 3973 features in object.

Detected 20229 features in scene.

Found 3973 matches using BFM,.

Found 804 good matches and 1608 good features from them.

Number of inliers: 529 out of 804 good matches.

Detected Object Corners:

[966.016, 1848.62]

[1451.38, 1435.76]

[1958.97, 2037.28]

[1454.55, 2459.66]  
Pixel Errors;

Corner 0 Error: 5.03015 pixels

Corner 1 Error: 6.32687 pixels

Corner 2 Error: 14.6614 pixels

Corner 3 Error: 16.7858 pixels

Average Error: 10.701 pixels

======= Test 1: md.jpg + md-scene1.jpg =======

Using default values for md-scene1.jpg

Selected corner: (675, 1736)

Selected corner: (1024, 1478)

Selected corner: (1328, 1883)

Selected corner: (961, 2154)

Detected 2747 features in object.

Detected 21174 features in scene.

Found 2747 matches using BFM.

Found 570 good matches and 1140 good features from them.

Number of inliers: 412 out of 570 good matches.

Detected Object Corners:

[678.189, 1732.36]

[1024.19, 1485.51]

[1311.53, 1885.01]

[946.308, 2147.24]   
Pixel Errors;  
Corner 0 Error: 4.83615 pixels

Corner 1 Error: 7.50978 pixels

Corner 2 Error: 16.5899 pixels

Corner 3 Error: 16.1724 pixels

Average Error: 11.277 pixels

In running the initial tests, I struggled to get the ORB detector to find the Monopoly Deal card game in scene 1, this was most likely due to the fact that there was a large scale difference between the objects file size and the relative size of the object in the scene (object file size of 2171x 2648 compared to the object size of 446x505), I initially tried raising the feature cap of the detector to 50,000 features, and while this did help I ended up needing to scale the object by a scale of 0.4x before running it through the detector. Which was the minimum scale that allowed me to detect the object in the scene with any degree of accuracy.

======= Test 4: book.jpg + b-scene2.jpg =======

Using default values for b-scene2.jpg

Selected corner: (1372, 1085)

Selected corner: (2132, 654)

Selected corner: (2735, 1638)

Selected corner: (1987, 2094)

Detected 7023 features in object.

Detected 50000 features in scene.

Found 7023 matches using BFM.

Found 1820 good matches and 3640 good features from them.

Number of inliers: 1360 out of 1820 good matches.

Detected Object Corners:

[1367, 1088.04]

[2134.24, 660.96]

[2731.22, 1640.15]

[1992.59, 2096.82]

Pixel Error:

Corner 0 Error: 5.85317 pixels

Corner 1 Error: 7.31213 pixels

Corner 2 Error: 4.3527 pixels

Corner 3 Error: 6.25972 pixels

Average Error: 5.94443 pixels

======= Test 3: book.jpg + b-scene1.jpg =======

Using default values for b-scene1.jpg

Selected corner: (41, 1576)

Selected corner: (978, 1549)

Selected corner: (957, 2754)

Selected corner: (16, 2800)

Detected 7023 features in object.

Detected 50000 features in scene.

Found 7023 matches using BFM.

Found 1992 good matches and 3984 good features from them.

Number of inliers: 1499 out of 1992 good matches.

Detected Object Corners:

[48.7821, 1573.23]

[968.973, 1556.39]

[935.002, 2745.12]

[9.3464, 2809.16]

Pixel Error:

Corner 0 Error: 8.26169 pixels

Corner 1 Error: 11.6682 pixels

Corner 2 Error: 23.7236 pixels

Corner 3 Error: 11.3192 pixels

Average Error: 13.7432 pixels

======= Test 5: book.jpg + b-sceneD.jpg =======

Using default values for b-sceneD.jpg

Selected corner: (41, 1576)

Selected corner: (978, 1549)

Selected corner: (957, 2754)

Selected corner: (16, 2800)

Detected 7023 features in object.

Detected 4046 features in scene.

Found 7023 matches using BFM.

Found 821 good matches and 1642 good features from them.

Number of inliers: 479 out of 821 good matches.

Detected Object Corners:

[47.8987, 1572.47]

[969.7, 1555.63]

[933.833, 2743.56]

[9.9246, 2808.44]

Pixel Errors:

Corner 0 Error: 7.74744 pixels

Corner 1 Error: 10.6211 pixels

Corner 2 Error: 25.4124 pixels

Corner 3 Error: 10.401 pixels

Average Error: 13.5455 pixels

without capping the number of detected features of sift results in large set of possibly less robust features, filtering these out with a more lenient Lowes Ratio seems to account for the larger than usual pixel error, a secondary test was running with a Lowes Ratio of 0.7 and a cap at 10,000 maximum detectable features. One thing that was noticed was that with both detection methods on Test 4 (book.jpg + b-scene.jpg) resulted in a lower pixel average error than the other combinations, suggesting that the detection method’s work better at corner approximation than true corners.

After capping the max number of features to 5000 resulted in a lot of the detectors to max out on features in both scene and object images, but what was interesting was the slight increase in accuracy in detection using SIFT and the obvious drop using ORB, suggesting SIFT works better with a smaller feature count whereas ORB needs a lot more flexibility to work with a larger range of features. This was reflected in the dataset from a drop from an average of 2142 inliers across the test for the ORB detector to 453, whereas SIFT went from an average of 855 to 535, despite the ORB feature cap dropping from 50,000 to 5000, and SIFT was initially set to an infinite cap value, averaging >55,000 detected scene features, down to 5000.

I decided to experiment more with this and tightened the Lowes Ratio from 0.8 to a tighter 0.6, keeping both detector feature caps to 50,000 for a larger range of features. The results were that the SIFT detector performed slightly better with the tighter ratio, but worse than the test with the lower feature cap, determining that the SIFT detector is able to perform more accurately with a small feature dataset, though these pixel errors were negligible from 11.052 on the high end and 10.956 on the low end. Though there was a noticeable decrease in detection accuracy for T5, which is the same scene from T3 but the scene is darkened.

I then experimented with lowering the feature cap with differing results, while SIFT suffered an average increase of 2.8 pixel error offset for each corner, ORB pixel error offset increased by 9 pixels. Note that this average does not include the error results from T2 as the corner errors were as low as 350 pixels and as high as 1315 pixels, obviously failing the object detection entirely, I suspect this would’ve been due to the scale difference again as this test still included the workaround for T1 but not T2, and the difference between the scale in objects versus their in-scene counterparts aren’t as drastic for T3, T4, T5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Detector | Top-Left Corner | Top-Right Corner | Bottom-Right Corner | Bottom-Left Corner | Average Pixel Error |
| ORB | 7.17 | 7.75 | 16.82 | 11.05 | 10.70 |
| SIFT | 6.35 | 8.69 | 16.95 | 12.19 | 11.05 |

Test results using a lowes ratio of 0.8 and an unlimited feature cap for sift

**SIFT Detector Using FLANN:**

Object image: object.png

Scene image: scene.png

Detection method: sift

Matcher method: flann

Detecting image features...

Detected 1954 features in object in 0.863 seconds

Detected 7005 features in scene in 3.322 seconds

Found 1954 matches in 0.658 seconds using FLANN

Finding good matches...

Found 720 good matches and 1440 good features from them in 0 seconds

Calculated the homography matrix in 0.011 seconds:

[0.7469067684415254, -0.4193402701464118, 788.8821724627741;

0.1669210682087139, 0.7067916864455451, 205.76387074164;

-9.509383300344636e-05, -0.0001846517030902413, 1]

Number of inliers: 420 out of 720 good matches.

Transformed Box Coordinates:

[788.882, 205.764]

[1311.88, 324.427]

[1118.19, 1162.94]

[509.867, 965.092]

**SIFT Detector Using BFM:**

Object image: object.png

Scene image: scene.png

Detection method: sift

Matcher method: bfm

Detecting image features...

Detected 1954 features in object in 0.86 seconds

Detected 7005 features in scene in 3.22 seconds

Found 1954 matches in 4.409 seconds using BFM

Finding good matches...

Found 730 good matches and 1460 good features from them in 0 seconds

Calculated the homography matrix in 0.012 seconds:

[0.7643542851018954, -0.399449524550603, 784.8597640802167;

0.1779455817391118, 0.7339702943747934, 197.5394926825597;

-8.679102866451919e-05, -0.0001661868033253011, 1]

Number of inliers: 478 out of 730 good matches.

Transformed Box Coordinates:

[784.86, 197.539]

[1311.78, 321.024]

[1118.47, 1159.83]

[515.659, 964.864]