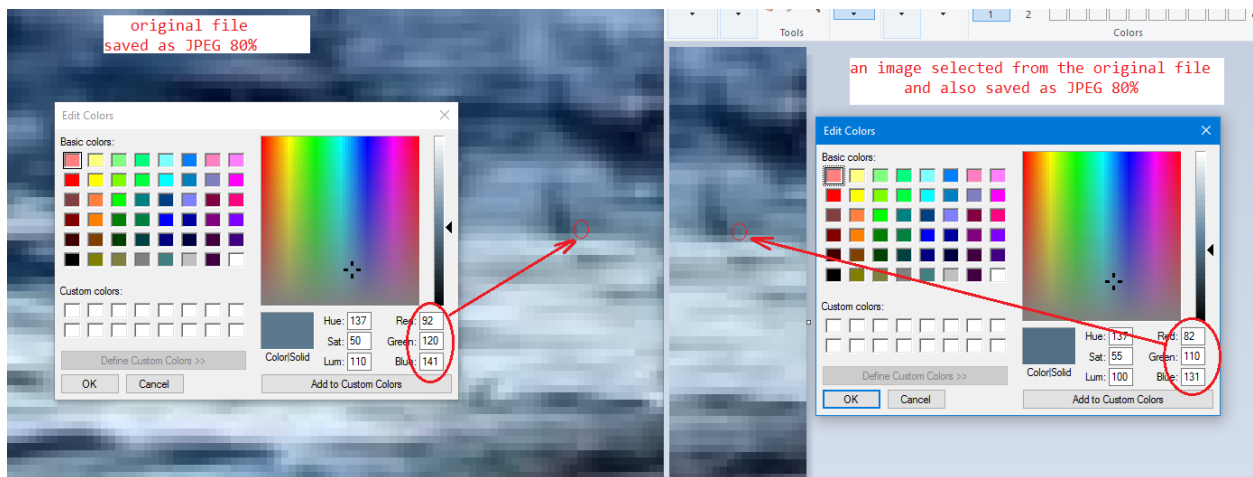


Overview

The original problem of *finding a location of cropped image inside an original one* is complicated by the fact that we have to deal with JPEG images. There is a good explanation of JPEG format and how the original image gets compressed: <https://en.wikipedia.org/wiki/JPEG>.

After studying the most interesting part - the steps involved into compression, it's easy to see that an image gets distorted when being saved as a JPEG file. The distortion depends in big part on the "quality" chosen. In fact, even if one chooses the **same** quality for the original and cropped images, the RGB colors (after loading and decompression) will be quite different in the original image region and the cropped version. Even more than that, different editors may compress images slightly different even if the same quality is chosen (due to potential 'downsampling' phase). And, last but not least, the color pallet of the image itself may affect the distortions during compression. Here is an example illustrating the point:



Without going into details, the searching algorithm should take into consideration the possibility of cropped image to be quite different in RGB color space.

Current solution

So, the current solution is straight forward and perhaps can be improved. It tries to find a location of sub-image, given some 'similarity' measure. The similarity here is defined as a maximum of Euclidian distances among all the RGB points of 2 images compared 1-to-1.

The algorithm basically:

- finds the biggest image first (which may be an original) just by looking at dimensions of images
- iterates over a bigger image, first checking for matching a 0-point (given the similarity threshold)
- if 0-point matches, then the full comparison is performed at that location

Testing

Besides a simple unittest, there is a full permutation test developed. I've picked some image from internet (of a 833 x 470 size), and first saved it as BMP image (to have a reference point), then selected 2 regions from it (also saved as BMP) and then slightly distorted 1 of those samples in 2 different ways (also saved as BMP).

So I had:

- *img/img-01/img-01.bmp* (original image)
- *img/img-01/img-01-smpl-01.bmp* (sample 1)
- *img/img-01/img-01-smpl-02.bmp* (sample 2)
- *img/img-01/img-01-smpl-04-n.bmp* (distorted sample)
- *img/img-01/img-01-smpl-03-n.bmp* (distorted sample)

Those images I've converted into JPEGs with different compression values. If you look into *img/img-01* folder, you'll see the file names follow the naming convention: ORIGINAL_BMP_NAME-COMPRESSION.jpg.

After that, the permutation test run, which tried to find the sampled images in the original ones. The results can be found in:

- *results/TestImg01Permutations-similarity-0.95.html* - tried to find subimages with 0.95% similarity
- *results/TestImg01Permutations-similarity-0.90.html* - tried to find subimages with 0.90% similarity
- *results/TestImg01Permutations-similarity-0.85.html* - tried to find subimages with 0.85% similarity

Summary

As it is easy to expect, the lower you set the similarity parameter, the more you find higher-compressed sub-images. On the other hand, there is a drawback: at some point you may run into situation where a not actual sub-images gets found, since it's close enough in terms of the distance measure.

For example, for my worst-quality images I've seen:

```
D:\Files\waldo-ag\subimage>python src/main/subimage.py img/img-01/img-01-050.jpg img/img-01/img-01-smpl-01-050.jpg -o target/out.jpg -s .89
found 0.8914 similarity at (x,y) = (413,147)
```

```
D:\Files\waldo-ag\subimage>python src/main/subimage.py img/img-01/img-01-050.jpg img/img-01/img-01-smpl-01-050.jpg -o target/out.jpg -s .8
found 0.7985 similarity at (x,y) = (413,146)
```

Notice how the Y coordinate differs and is slightly wrong at the second line. In this case that I had, the original image was compressed at JPEG 50%, and the sub-image was also at JPEG-50%. In order to find the sub-image I had to set the similarity measure to be 0.85.

In a normal case (when images are compressed at about 95%), a default similarity of 0.95 is sufficient enough.

Here is the conceptual look at permutation runs with different similarity measures:

TridaghtPermissions:

[illegible]

Bridge Inspection											
Bridge ID	Bridge Name	Location	Span Type	Material	Year Built	Inspector	Inspection Date	Inspection Type	Inspector	Inspection Date	Inspection Type
1	Bridge 1	Location 1	Span Type 1	Material 1	Year Built 1	Inspector 1	Inspection Date 1	Inspection Type 1	Inspector 1	Inspection Date 1	Inspection Type 1
2	Bridge 2	Location 2	Span Type 2	Material 2	Year Built 2	Inspector 2	Inspection Date 2	Inspection Type 2	Inspector 2	Inspection Date 2	Inspection Type 2
3	Bridge 3	Location 3	Span Type 3	Material 3	Year Built 3	Inspector 3	Inspection Date 3	Inspection Type 3	Inspector 3	Inspection Date 3	Inspection Type 3
4	Bridge 4	Location 4	Span Type 4	Material 4	Year Built 4	Inspector 4	Inspection Date 4	Inspection Type 4	Inspector 4	Inspection Date 4	Inspection Type 4
5	Bridge 5	Location 5	Span Type 5	Material 5	Year Built 5	Inspector 5	Inspection Date 5	Inspection Type 5	Inspector 5	Inspection Date 5	Inspection Type 5
6	Bridge 6	Location 6	Span Type 6	Material 6	Year Built 6	Inspector 6	Inspection Date 6	Inspection Type 6	Inspector 6	Inspection Date 6	Inspection Type 6
7	Bridge 7	Location 7	Span Type 7	Material 7	Year Built 7	Inspector 7	Inspection Date 7	Inspection Type 7	Inspector 7	Inspection Date 7	Inspection Type 7
8	Bridge 8	Location 8	Span Type 8	Material 8	Year Built 8	Inspector 8	Inspection Date 8	Inspection Type 8	Inspector 8	Inspection Date 8	Inspection Type 8
9	Bridge 9	Location 9	Span Type 9	Material 9	Year Built 9	Inspector 9	Inspection Date 9	Inspection Type 9	Inspector 9	Inspection Date 9	Inspection Type 9
10	Bridge 10	Location 10	Span Type 10	Material 10	Year Built 10	Inspector 10	Inspection Date 10	Inspection Type 10	Inspector 10	Inspection Date 10	Inspection Type 10
11	Bridge 11	Location 11	Span Type 11	Material 11	Year Built 11	Inspector 11	Inspection Date 11	Inspection Type 11	Inspector 11	Inspection Date 11	Inspection Type 11
12	Bridge 12	Location 12	Span Type 12	Material 12	Year Built 12	Inspector 12	Inspection Date 12	Inspection Type 12	Inspector 12	Inspection Date 12	Inspection Type 12
13	Bridge 13	Location 13	Span Type 13	Material 13	Year Built 13	Inspector 13	Inspection Date 13	Inspection Type 13	Inspector 13	Inspection Date 13	Inspection Type 13
14	Bridge 14	Location 14	Span Type 14	Material 14	Year Built 14	Inspector 14	Inspection Date 14	Inspection Type 14	Inspector 14	Inspection Date 14	Inspection Type 14
15	Bridge 15	Location 15	Span Type 15	Material 15	Year Built 15	Inspector 15	Inspection Date 15	Inspection Type 15	Inspector 15	Inspection Date 15	Inspection Type 15
16	Bridge 16	Location 16	Span Type 16	Material 16	Year Built 16	Inspector 16	Inspection Date 16	Inspection Type 16	Inspector 16	Inspection Date 16	Inspection Type 16
17	Bridge 17	Location 17	Span Type 17	Material 17	Year Built 17	Inspector 17	Inspection Date 17	Inspection Type 17	Inspector 17	Inspection Date 17	Inspection Type 17
18	Bridge 18	Location 18	Span Type 18	Material 18	Year Built 18	Inspector 18	Inspection Date 18	Inspection Type 18	Inspector 18	Inspection Date 18	Inspection Type 18
19	Bridge 19	Location 19	Span Type 19	Material 19	Year Built 19	Inspector 19	Inspection Date 19	Inspection Type 19	Inspector 19	Inspection Date 19	Inspection Type 19
20	Bridge 20	Location 20	Span Type 20	Material 20	Year Built 20	Inspector 20	Inspection Date 20	Inspection Type 20	Inspector 20	Inspection Date 20	Inspection Type 20
21	Bridge 21	Location 21	Span Type 21	Material 21	Year Built 21	Inspector 21	Inspection Date 21	Inspection Type 21	Inspector 21	Inspection Date 21	Inspection Type 21
22	Bridge 22	Location 22	Span Type 22	Material 22	Year Built 22	Inspector 22	Inspection Date 22	Inspection Type 22	Inspector 22	Inspection Date 22	Inspection Type 22
23	Bridge 23	Location 23	Span Type 23	Material 23	Year Built 23	Inspector 23	Inspection Date 23	Inspection Type 23	Inspector 23	Inspection Date 23	Inspection Type 23
24	Bridge 24	Location 24	Span Type 24	Material 24	Year Built 24	Inspector 24	Inspection Date 24	Inspection Type 24	Inspector 24	Inspection Date 24	Inspection Type 24
25	Bridge 25	Location 25	Span Type 25	Material 25	Year Built 25	Inspector 25	Inspection Date 25	Inspection Type 25	Inspector 25	Inspection Date 25	Inspection Type 25
26	Bridge 26	Location 26	Span Type 26	Material 26	Year Built 26	Inspector 26	Inspection Date 26	Inspection Type 26	Inspector 26	Inspection Date 26	Inspection Type 26
27	Bridge 27	Location 27	Span Type 27	Material 27	Year Built 27	Inspector 27	Inspection Date 27	Inspection Type 27	Inspector 27	Inspection Date 27	Inspection Type 27
28	Bridge 28	Location 28	Span Type 28	Material 28	Year Built 28	Inspector 28	Inspection Date 28	Inspection Type 28	Inspector 28	Inspection Date 28	Inspection Type 28
29	Bridge 29	Location 29	Span Type 29	Material 29	Year Built 29	Inspector 29	Inspection Date 29	Inspection Type 29	Inspector 29	Inspection Date 29	Inspection Type 29
30	Bridge 30	Location 30	Span Type 30	Material 30	Year Built 30	Inspector 30	Inspection Date 30	Inspection Type 30	Inspector 30	Inspection Date 30	Inspection Type 30
31	Bridge 31	Location 31	Span Type 31	Material 31	Year Built 31	Inspector 31	Inspection Date 31	Inspection Type 31	Inspector 31	Inspection Date 31	Inspection Type 31
32	Bridge 32	Location 32	Span Type 32	Material 32	Year Built 32	Inspector 32	Inspection Date 32	Inspection Type 32	Inspector 32	Inspection Date 32	Inspection Type 32
33	Bridge 33	Location 33	Span Type 33	Material 33	Year Built 33	Inspector 33	Inspection Date 33	Inspection Type 33	Inspector 33	Inspection Date 33	Inspection Type 33
34	Bridge 34	Location 34	Span Type 34	Material 34	Year Built 34	Inspector 34	Inspection Date 34	Inspection Type 34	Inspector 34	Inspection Date 34	Inspection Type 34
35	Bridge 35	Location 35	Span Type 35	Material 35	Year Built 35	Inspector 35	Inspection Date 35	Inspection Type 35	Inspector 35	Inspection Date 35	Inspection Type 35
36	Bridge 36	Location 36	Span Type 36	Material 36	Year Built 36	Inspector 36	Inspection Date 36	Inspection Type 36	Inspector 36	Inspection Date 36	Inspection Type 36
37	Bridge 37	Location 37	Span Type 37	Material 37	Year Built 37	Inspector 37	Inspection Date 37	Inspection Type 37	Inspector 37	Inspection Date 37	Inspection Type 37
38	Bridge 38	Location 38	Span Type 38	Material 38	Year Built 38	Inspector 38	Inspection Date 38	Inspection Type 38	Inspector 38	Inspection Date 38	Inspection Type 38
39	Bridge 39	Location 39	Span Type 39	Material 39	Year Built 39	Inspector 39	Inspection Date 39	Inspection Type 39	Inspector 39	Inspection Date 39	Inspection Type 39
40	Bridge 40	Location 40	Span Type 40	Material 40	Year Built 40	Inspector 40	Inspection Date 40	Inspection Type 40	Inspector 40	Inspection Date 40	Inspection Type 40
41	Bridge 41	Location 41	Span Type 41	Material 41	Year Built 41	Inspector 41	Inspection Date 41	Inspection Type 41	Inspector 41	Inspection Date 41	Inspection Type 41
42	Bridge 42	Location 42	Span Type 42	Material 42	Year Built 42	Inspector 42	Inspection Date 42	Inspection Type 42	Inspector 42	Inspection Date 42	Inspection Type 42
43	Bridge 43	Location 43	Span Type 43	Material 43	Year Built 43	Inspector 43	Inspection Date 43	Inspection Type 43	Inspector 43	Inspection Date 43	Inspection Type 43
44	Bridge 44	Location 44	Span Type 44	Material 44	Year Built 44	Inspector 44	Inspection Date 44	Inspection Type 44	Inspector 44	Inspection Date 44	Inspection Type 44
45	Bridge 45	Location 45	Span Type 45	Material 45	Year Built 45	Inspector 45	Inspection Date 45	Inspection Type 45	Inspector 45	Inspection Date 45	Inspection Type 45
46	Bridge 46	Location 46	Span Type 46	Material 46	Year Built 46	Inspector 46	Inspection Date 46	Inspection Type 46	Inspector 46	Inspection Date 46	Inspection Type 46
47	Bridge 47	Location 47	Span Type 47	Material 47	Year Built 47	Inspector 47	Inspection Date 47	Inspection Type 47	Inspector 47	Inspection Date 47	Inspection Type 47
48	Bridge 48	Location 48	Span Type 48	Material 48	Year Built 48	Inspector 48	Inspection Date 48	Inspection Type 48	Inspector 48	Inspection Date 48	Inspection Type 48
49	Bridge 49	Location 49	Span Type 49	Material 49	Year Built 49	Inspector 49	Inspection Date 49	Inspection Type 49	Inspector 49	Inspection Date 49	Inspection Type 49
50	Bridge 50	Location 50	Span Type 50	Material 50	Year Built 50	Inspector 50	Inspection Date 50	Inspection Type 50	Inspector 50	Inspection Date 50	Inspection Type 50
51	Bridge 51	Location 51	Span Type 51	Material 51	Year Built 51	Inspector 51	Inspection Date 51	Inspection Type 51	Inspector 51	Inspection Date 51	Inspection Type 51
52	Bridge 52	Location 52	Span Type 52	Material 52	Year Built 52	Inspector 52	Inspection Date 52	Inspection Type 52	Inspector 52	Inspection Date 52	Inspection Type 52
53	Bridge 53	Location 53	Span Type 53	Material 53	Year Built 53	Inspector 53	Inspection Date 53	Inspection Type 53	Inspector 53	Inspection Date 53	Inspection Type 53
54	Bridge 54	Location 54	Span Type 54	Material 54	Year Built 54	Inspector 54	Inspection Date 54	Inspection Type 54	Inspector 54	Inspection Date 54	Inspection Type 54
55	Bridge 55	Location 55	Span Type 55	Material 55	Year Built 55	Inspector 55	Inspection Date 55	Inspection Type 55	Inspector 55	Inspection Date 55	Inspection Type 55
56	Bridge 56	Location 56	Span Type 56	Material 56	Year Built 56	Inspector 56	Inspection Date 56	Inspection Type 56	Inspector 56	Inspection Date 56	Inspection Type 56
57	Bridge 57	Location 57	Span Type 57	Material 57	Year Built 57	Inspector 57	Inspection Date 57	Inspection Type 57	Inspector 57	Inspection Date 57	Inspection Type 57
58	Bridge 58	Location 58	Span Type 58	Material 58	Year Built 58	Inspector 58	Inspection Date 58	Inspection Type 58	Inspector 58	Inspection Date 58	Inspection Type 58
59	Bridge 59	Location 59	Span Type 59	Material 59	Year Built 59	Inspector 59	Inspection Date 59	Inspection Type 59	Inspector 59	Inspection Date 59	Inspection Type 59
60	Bridge 60	Location 60	Span Type 60	Material 60	Year Built 60	Inspector 60	Inspection Date 60	Inspection Type 60	Inspector 60	Inspection Date 60	Inspection Type 60
61	Bridge 61	Location 61	Span Type 61	Material 61	Year Built 61	Inspector 61	Inspection Date 61	Inspection Type 61	Inspector 61	Inspection Date 61	Inspection Type 61
62	Bridge 62	Location 62	Span Type 62	Material 62	Year Built 62	Inspector 62	Inspection Date 62	Inspection Type 62	Inspector 62	Inspection Date 62	Inspection Type 62
63	Bridge 63	Location 63	Span Type 63	Material 63	Year Built 63	Inspector 63	Inspection Date 63	Inspection Type 63	Inspector 63	Inspection Date 63	Inspection Type 63
64	Bridge 64	Location 64	Span Type 64	Material 64	Year Built 64	Inspector 64	Inspection Date 64	Inspection Type 64	Inspector 64	Inspection Date 64	Inspection Type 64
65	Bridge 65	Location 65	Span Type 65	Material 65	Year Built 65	Inspector 65	Inspection Date 65	Inspection Type 65	Inspector 65	Inspection Date 65	Inspection Type 65
66	Bridge 66	Location 66	Span Type 66	Material 66	Year Built 66	Inspector 66	Inspection Date 66	Inspection Type 66	Inspector 66	Inspection Date 66	Inspection Type 66
67	Bridge 67	Location 67	Span Type 67	Material 67	Year Built 67	Inspector 67	Inspection Date 67	Inspection Type 67	Inspector 67	Inspection Date 67	Inspection Type 67
68	Bridge 68	Location 68	Span Type 68	Material 68	Year Built 68	Inspector 68	Inspection Date 68	Inspection Type 68	Inspector 68	Inspection Date 68	Inspection Type 68
69	Bridge 69	Location 69	Span Type 69	Material 69	Year Built 69	Inspector 69	Inspection Date 69	Inspection Type 69	Inspector 69	Inspection Date 69	Inspection Type 69
70	Bridge 70	Location 70	Span Type 70	Material 70	Year Built 70	Inspector 70	Inspection Date 70	Inspection Type 70	Inspector 70	Inspection Date 70	Inspection Type 70
71	Bridge 71	Location 71	Span Type 71	Material 71	Year Built 71	Inspector 71	Inspection Date 71	Inspection Type 71	Inspector 71	Inspection Date 71	Inspection Type 71
72	Bridge 72	Location 72	Span Type 72	Material 72	Year Built 72	Inspector 72	Inspection Date 72	Inspection Type 72	Inspector 72	Inspection Date 72	Inspection Type 72
73	Bridge 73	Location 73	Span Type 73	Material 73	Year Built 73	Inspector 73	Inspection Date 73	Inspection Type 73	Inspector 73	Inspection Date 73	Inspection Type 73
74	Bridge 74	Location 74	Span Type 74	Material 74	Year Built 74	Inspector 74	Inspection Date 74	Inspection Type 74	Inspector 74	Inspection Date 74	Inspection Type 74
75	Bridge 75	Location 75	Span Type 75	Material 75	Year Built 75	Inspector 75	Inspection Date 75	Inspection Type 75	Inspector 75	Inspection Date 75	Inspection Type 75
76	Bridge 76	Location 76	Span Type 76	Material 76	Year Built 76	Inspector 76	Inspection Date 76	Inspection Type 76	Inspector 76	Inspection Date 76	Inspection Type 76
77	Bridge 77	Location 77	Span Type 77	Material 77	Year Built 77	Inspector 77	Inspection Date 77	Inspection Type 77	Inspector 77	Inspection Date 77	Inspection Type 77
78	Bridge 78	Location 78	Span Type 78	Material 78	Year Built 78	Inspector 78	Inspection Date 78	Inspection Type 78	Inspector 78	Inspection Date 78	Inspection Type 78
79	Bridge 79	Location 79	Span Type 79	Material 79	Year Built 79	Inspector 79	Inspection Date 79	Inspection Type 79	Inspector 79	Inspection Date 79	Inspection Type 79
80	Bridge 80	Location 80	Span Type 80	Material 80	Year Built 80	Inspector 80	Inspection Date 80	Inspection Type 80	Inspector 80	Inspection Date 80	Inspection Type 80
81	Bridge 81	Location 81	Span Type 81	Material 81	Year Built 81	Inspector 81	Inspection Date 81	Inspection Type 81	Inspector 81	Inspection Date 81	Inspection Type 81
82	Bridge 82	Location 82	Span Type 82	Material 82	Year Built 82	Inspector 82	Inspection Date 82	Inspection Type 82	Inspector 82	Inspection Date 82	Inspection Type 82
83	Bridge 83	Location 83	Span Type 83	Material 83	Year Built 83	Inspector 83	Inspection Date 83	Inspection Type 83	Inspector 83	Inspection Date 83	Inspection Type 83
84	Bridge 84	Location 84	Span Type 84	Material 84	Year Built 84	Inspector 84	Inspection Date 84	Inspection Type 84	Inspector 84	Inspection Date 84	Inspection Type 84
85	Bridge 85	Location 85	Span Type 85	Material 85	Year Built 85	Inspector 85	Inspection Date 85	Inspection Type 85	Inspector 85	Inspection Date 85	Inspection Type 85
86	Bridge 86	Location 86	Span Type 86	Material 86	Year Built 86	Inspector 86	Inspection Date 86	Inspection Type 86	Inspector 86	Inspection Date 86	Inspection Type 86
87	Bridge 87	Location 87	Span Type 87	Material 87	Year Built 87	Inspector 87	Inspection Date 87	Inspection Type 87	Inspector 87	Inspection Date 87	Inspection Type 87
88	Bridge 88	Location 88	Span Type 88	Material 88	Year Built 88	Inspector 88	Inspection Date 88	Inspection Type 88	Inspector 88	Inspection Date 88	Inspection Type 88
89	Bridge 89	Location 89	Span Type 89	Material 89	Year Built 89	Inspector 89	Inspection Date 89	Inspection Type 89	Inspector 89	Inspection Date 89	Inspection Type 89
90	Bridge 90	Location 90	Span Type 90	Material 90	Year Built 90	Inspector 90	Inspection Date 90	Inspection Type 90	Inspector 90	Inspection Date 90	Inspection Type 90
91	Bridge 91	Location 91	Span Type 91	Material 91	Year Built 91	Inspector 91	Inspection Date 91	Inspection Type 91	Inspector 91	Inspection Date 91	Inspection Type 91
92	Bridge 92	Location 92	Span Type 92	Material 92	Year Built 92	Inspector 92	Inspection Date 92	Inspection Type 92	Inspector 92	Inspection Date 92	Inspection Type 92
93	Bridge 93	Location 93	Span Type 93	Material 93	Year Built 93	Inspector 93	Inspection Date 93	Inspection Type 93	Inspector 93	Inspection Date 93	Inspection Type 93
94	Bridge 94	Location 94	Span Type 94	Material 94	Year Built 94	Inspector 94	Inspection Date 94	Inspection Type 94	Inspector 94	Inspection Date 94	Inspection Type 94
95	Bridge 95	Location 95	Span Type 95	Material 95	Year Built 95	Inspector 95	Inspection Date 95	Inspection Type 95	Inspector 95	Inspection Date 95	Inspection Type 95
96	Bridge 96	Location 96	Span Type 96	Material 96	Year Built 96	Inspector 96	Inspection Date 96	Inspection Type 96	Inspector 96	Inspection Date 96	Inspection Type 96
97	Bridge 97	Location 97	Span Type 97	Material 97	Year Built 97	Inspector 97	Inspection Date 97	Inspection Type 97	Inspector 97	Inspection Date 97	Inspection Type 97
98	Bridge 98	Location 98	Span Type 98	Material 98	Year Built 98	Inspector 98	Inspection Date 98	Inspection Type 98	Inspector 98	Inspection Date 98	Inspection Type 98
99	Bridge 99	Location 99	Span Type 99	Material 99	Year Built 99	Inspector 99	Inspection Date 99	Inspection Type 99	Inspector 99	Inspection Date 99	Inspection Type 99
100	Bridge 100	Location 100	Span Type 100	Material 100	Year Built 100	Inspector 100	Inspection Date 100	Inspection Type 100	Inspector 100	Inspection Date 100	Inspection Type 100

As you can see, the matching becomes unstable once you start lowering the similarity measure.

Potential for improvements

The current runtime order is about $O((n-m)^2 * m^2)$ where 'n' refers to width or height of 1 "bigger" image, and "m" refers to width or height of the "smaller" image. This is, obviously, too long. Unfortunately, the time constraints don't allow further investigation of this problem. There are multiple things can be tried such as:

- different pre-processing schema of images
- organizing image info into tree/hash-like structure (very similar to how JPEG actually gets compressed with "Discrete cosine transform")
- perhaps may try to use ML (like a NN) to improve a similarity measure of comparing 2 equi-sized JPEGs (but I'm not sure there will be any improvement in performance).
- etc