

Assignment 1: Title of the assignment

Vincent Matthys

vincent.matthys@ens-paris-saclay.fr

QIA.1: (i) Why is it important to have a similarity co-variant feature detector? (ii) How does this affect the descriptors computed at these detections? (iii) How does this affect the matching process?

- (i) It is important to have a similarity co-variant feature detector in order to extract the same regions of interest regardless of translations, rotations or scales, *i.e.* regardless of viewpoints.
- (ii) The fact that the feature detector is co-variant causes the descriptors computed at these detections to undergo the same transformations than the image.
- (iii) Having feature detectors with similarity co-variance and feature descriptors with this and that other property affects the matching process in this way.

QIA.2: Show the detected features in the two images for three different values of the peakThreshold option

The detected features for $\text{peakThreshold} \in \{0.0001, 0.001, 0.01\}$ are shown in Figure 1 for the all_souls.000002.jpg image, and in Figure 2 for the all_souls.000015.jpg image. As expected, the number of keypoints decreases with the threshold. Moreover the keypoints are fewer in the darker regions, which leads to asymmetric density in the Figure 2, especially because of the building shadow.

QIA.3: Note the change in spatial density of detections across images, for a given value of peakThreshold. (i) Is the density uniform? If not, why? (ii) Which implications for image matching can this have? (iii) How can it be avoided?

- (i) In Figures 1a and 2a. The density is clearly not uniform between the two images. In both, the top-center of the image is very dense, but, in the first one, the grass admits no detection, which leads to a gap in the density, with some detections in the bottom limit of the grass.
- (ii) This can lead to error matching the correspondances with local features
- (iii) To avoid it we can proceed to a spatial verification relying on global geometric relations.

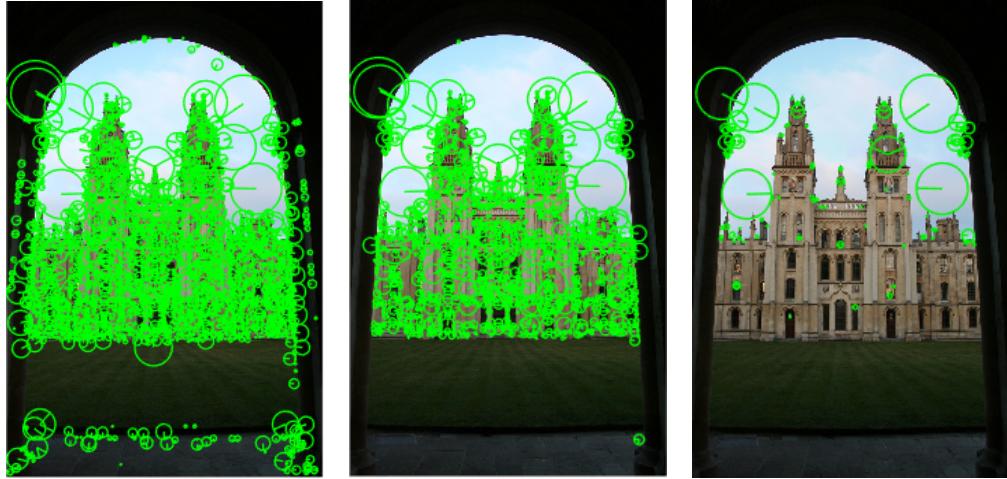


Figure 1: Feature detectors with three different values of $peakThreshold$ for all_souls_000002.jpg

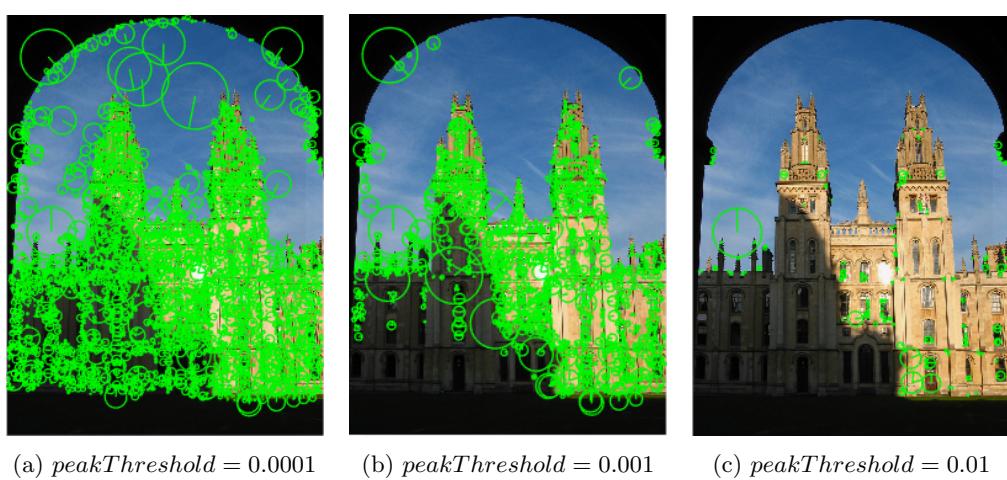
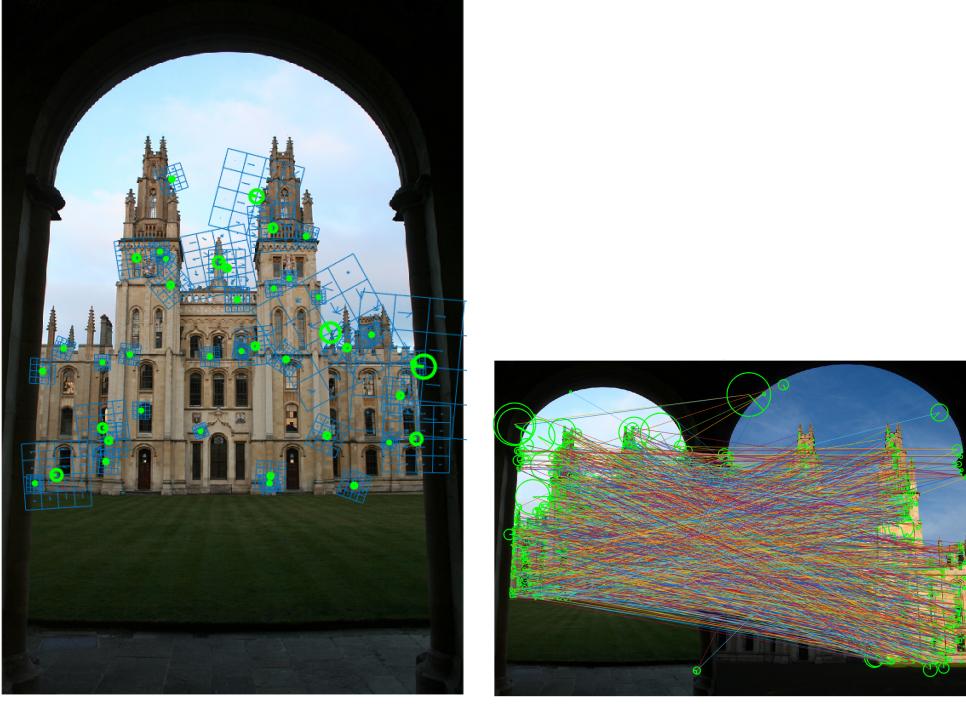


Figure 2: Feature detectors with three different values of $peakThreshold$ for all_souls_000015.jpg

QIB.1: Note the descriptors are computed over a much larger region (shown in blue) than the detection (shown in green). Why is this a good strategy?

The larger regions are shown in Figure 3a. It is a good strategy to be insensitive to small geometric deformations, and small variations of illumination, by taking the average over a larger region



(a) 4x4 grid around the keypoint

(b) first NN match

Figure 3: Computation of SIFT descriptors and matching

QIB.2: Examine carefully the mismatches and try to understand the different causes of mismatch. (i) In your report, present at least 3 of them and explain why the mistakes are being made. For example, is the change in lighting a problem? (ii) What additional constraints can be applied to remove the mismatches?

- (i) Three types of mismatch are presentend in Figure 4, and are extracted from Figure 3b
 - repetitive pattern (in blue)
 - constant local change in lighting leading to similarity between two different key-points (in red)
 - ambiguity between a keypoint his first two nearest neighbors (in yellow)

A global change in lighting is irrelevant for matching, because of the gradient nullifying the constant component

- (ii) Geometric verification with global constraints can discard those mismatches. Another method, presented below, is to look at the second nearest neigbour to remove ambiguous matches.



Figure 4: Three mismatches with SIFT features descriptors, subgroup of matches extracted from 3b

QIC.1: Illustrate and comment on improvements that you can achieve with this step. Include in your report figures showing the varying number of tentative matches when you change the $nnThreshold$ parameter.

As expected, the percentage of matches kept increases with $nnThreshold$. The results obtained are shown in Figures 5 6 7. Until $nnThreshold = 0.7$, there is no substantial mismatches, but the number of matches kept is below 10 %. If more matches are kept, a non-negligible part of new matches is missmatches, because the ambiguity between the first and the second nearest neighbour is no longer removed. In Figure 7b we can see the exponential growth of the number of matches with $nnThreshold$. A good compromise would be to take the threshold before this exponential growth.

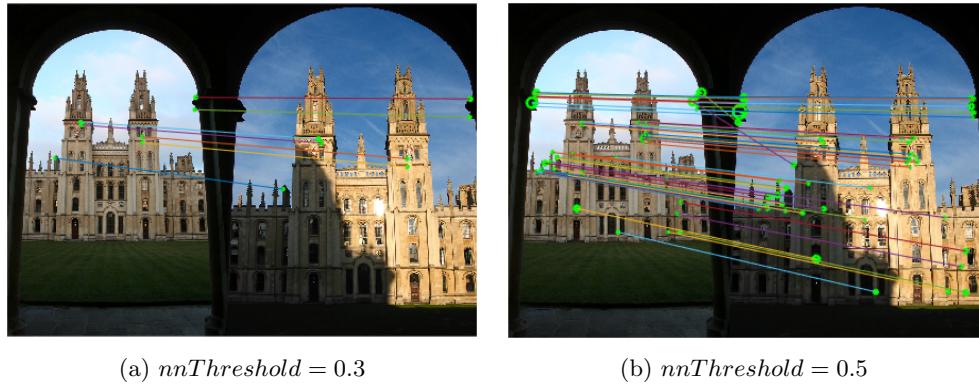
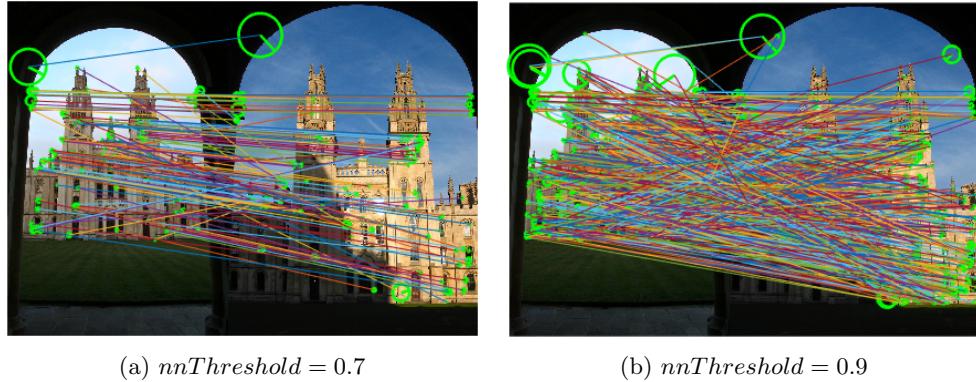


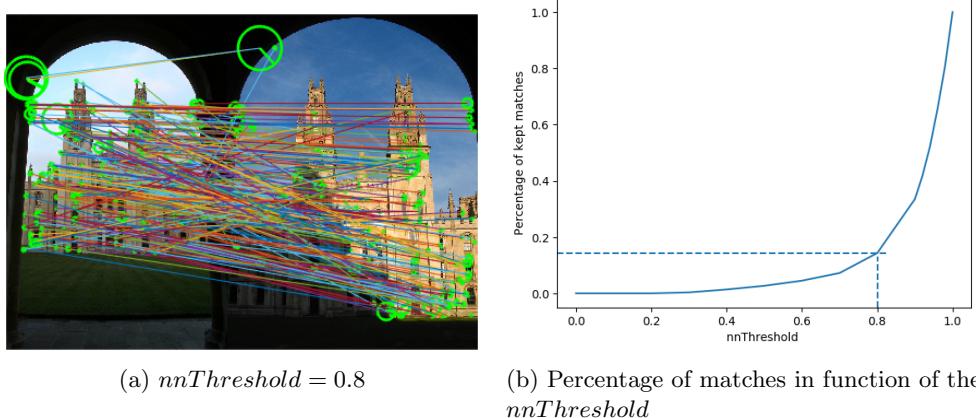
Figure 5: Matching using Lowe's second nearest neighbour - part1



(a) $nnThreshold = 0.7$

(b) $nnThreshold = 0.9$

Figure 6: Matching using Lowe's second nearest neighbour - part2



(a) $nnThreshold = 0.8$

(b) Percentage of matches in function of the $nnThreshold$

Figure 7: Matching using Lowe's second nearest neighbour - part3

QID.1: Work out how to compute this transformation from a single correspondence. Hint: recall from Stage I.A that a SIFT feature frame is an oriented circle and map one onto the other.

QID.2: Illustrate improvements that you can achieve with this step.

In Figure 8a, it's clear there is no mismatch anymore: every match corresponds to a unique geometric transformation. The number of matches dropped from 293 to 123. It means only 42% of the matches given by the Lowe's method were inliers. In Figure 8b we can observe the almost linear growth of matches with $nnThreshold$ with the geometric verification, in comparison with the exponential growth of the Lowe's method alone.

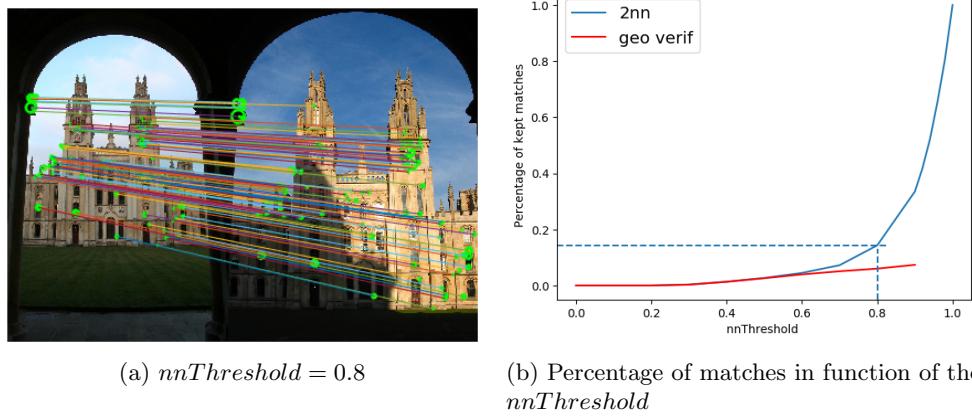


Figure 8: Improving SIFT matching with Lowe’s 2nd nn and geometric verification

QII.1: Include in your report images and a graph showing the number of verified matches with changing viewpoint. At first there are more similarity detector matches than affine. Why?

At first, introducing affine detectors is introducing a second spatial dimension for each blob, which leads to one more constraint to verify in the geometric transformation. This results discarding some outliers at the very beginning.

QIII.A.1: In the above procedure the time required to convert the descriptors into visual words was not accounted for. Why the speed of this step is less important in practice?