Image Stitching

Github user: F-Riv

1 Implementation Details

The code consists in the following steps:

1. Convert the two images to gray scale and find peaks (keypoints) in the Harris corner response, every area of 2 * min_distance + 1 pixel has a peak.
2. calculate SIFT descriptor for every keypoint, compare descriptors and pick the best matches.
3. run RANSAC on the best matches for a fixed number of iterations to find the homography that best aligns the matched keypoints, filtering out outliers.
4. Use the homography matrix to warp the second image to align with the first image.
5. Compare the warped image with the original image using the metrics described in the following sections and eventually save the best parameters in a json file.

Tests on other images can be found at the end of the notebook. In the next sections I delve into the results on the assignment images.

2 Parameters

Different combination of parameters have been tried, results will be showed for the values in brackets.

- Top n values: Number of couples of source and destination keypoints that are selected from each image according to the lower euclidean distance between SIFT descriptors [100, 200].
- RANSAC iterations: [1000, 2000].
- RANSAC Inlier threshold: Maximum distance between a source point and its transformation for being considered an inlier [2, 4].
- Peak min distance: Minimum distance between peaks when selecting Harris Corners [3, 5].
- k: parameter k in corner_harris, lower values filter out points with less shifts [0.01, 0.05].

2.1 Non-used parameters

The following parameters has been explored but they are fixed or not used:

- Type_of_distance: with SIFT descriptors euclidean distance is more efficient, normalized correlation has been tried but does not lead to any result.
- Minimum number of inliers vs. max. nubers of iterations in RANSAC: instead of running RANSAC for a fixed number of iterations it is possible to look for a minimum number of inliers. This parameter has not been used because a fair choice of a value would depend on other parameters.
- SIFT descriptor range size: smaller size permits the descriptor to describe the area more accurately since the descriptor shape is fixed (16x16)

3 Sensitivity Analysis

3.1 Metrics

For sensitivity analysis gray scale patches has been extracted from the surrounding around the inliers.The original points in the second image (destination points) are transformed according to the inverse of the homography matrix and the area surrounding the original destination points in the original second image is compared to the transformed destination points on the warped image to verify the consistency of the transformation within the destination image itself. If the following metrics provide good scores then it means that the final image mantains the local features and appears natural:

1. Norm. MSE(Normalized Mean Square Error): MSE is normalized by maximum distance in the gray scale values, lower values indicate similarity ranges between 0 and 1.
2. SSIM (Structural Similarity Index): combines luminance, contrast and structures scores. It represents better than MSE the quality perceived by people. Higher values indicates more similarity it ranges between -1 and 1.

Moreover some other metrics obtained during RANSAC are considered:

1. Average Euclidean Inlier Residuals (Normalized on RANSAC maximum distance threshold): euclidean distance between the coordinates of a source point transformed according to the affinity matrix and the correspondent destination point. The maximum distance threshold considered for normalization (RANSAC inlier threshold) is the higher one among all the possible combination of parameters in the current run (so in the parameters considered 4). Lower values indicates a better result.
2. ratio inliers/top_n keypoints: this metric is only used to assess that an image has enough inliers to consider the other metrics valid.

3.2 Results

According to the several parameter combinations tried the best ones are:

```
Higher average SSIM score in inlier patches 0.88984658057137  obtained with parameters:

Distance type: eucl
Top n values: 100
RANSAC iterations: 1000
RANSAC inlier threshold: 2
Peak min distance: 5
SIFT Patch size: 15
k Harris: 0.01
with ratio inliers/keypoints: 0.85

Lower MSE score in inlier patches 0.0824441489361702  obtained with parameters:

Distance type: eucl
Top n values: 100
RANSAC iterations: 2000
RANSAC inlier threshold: 2
Peak min distance: 3
SIFT Patch size: 15
k Harris: 0.01
with ratio inliers/keypoints: 0.94

Lower Euclidean residual inliers average (norm) 0.18921085484347896  obtained with parameters:
```

```
Distance type: eucl
Top n values: 100
RANSAC iterations: 2000
RANSAC inlier threshold: 2
Peak min distance: 3
SIFT Patch size: 15
k Harris: 0.01
with ratio inliers/keypoints: 0.94
```

The best parameters appears to be the same for every metrics with an exception for SSIM that obtains a better output with 2000 RANSAC iterations. In figure 1 the stitched image for best result for the last two metrics mentioned (SSIM score is high as well in this combination), in figure 2 are plotted some of the patches used for calculating SSIM and normalized MSE. The result images from the previous steps (images with keypoints detected by Harris corners, images with areas considered by SIFT descriptors, images with inliers scattered) a can be found in the notebook.

Figure 1: Stiched image with inliers

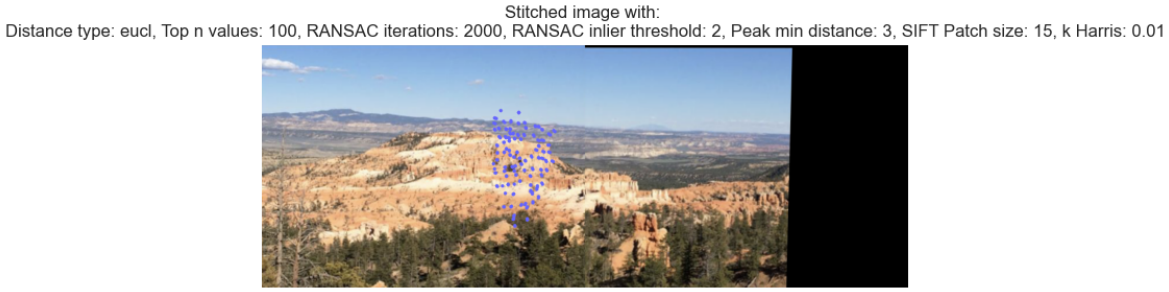
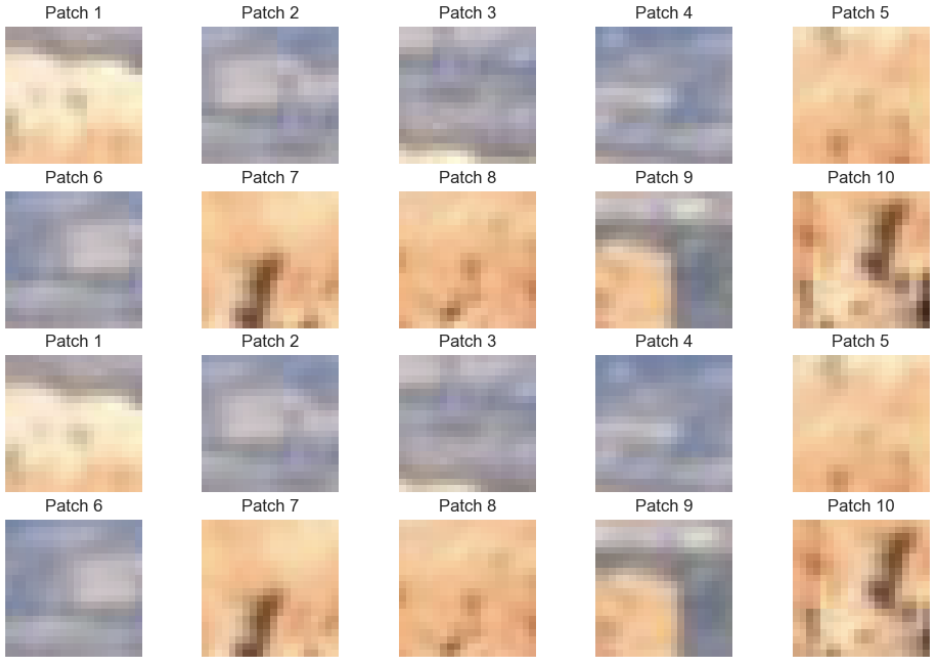
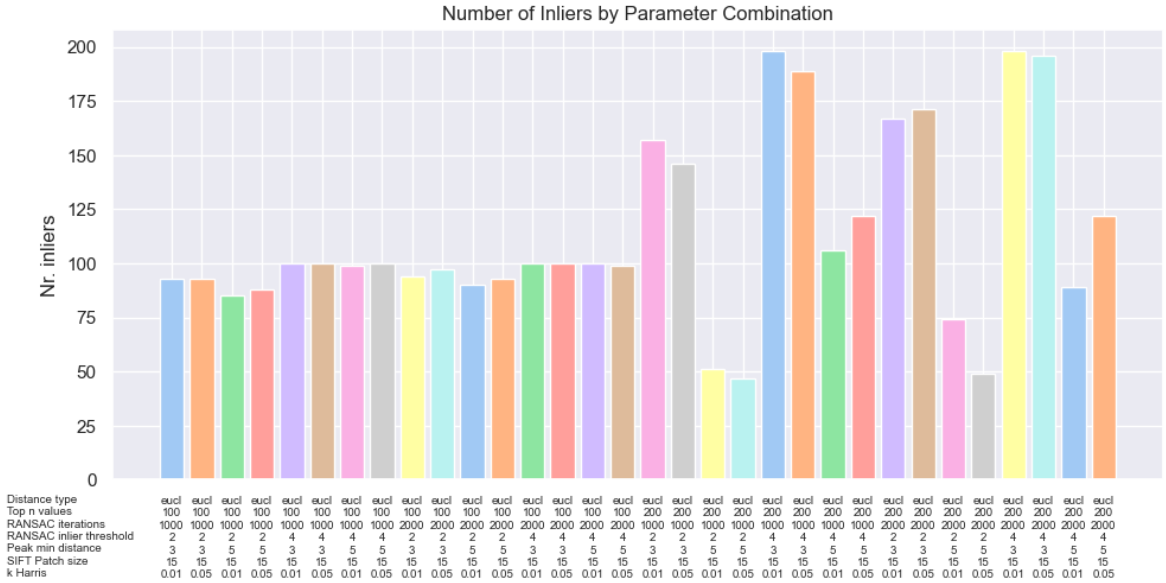


Figure 2: Patches destination points vs. transformed destination points



In figure 3 it is possible to notice that if there are only the 100 selected keypoints (top_n) used in RANSAC then RANSAC for every combination of parameters can manage to detect all of them (or almost all of them) as inliers. With top_n 200 the output is more sparse.

Figure 3: Number of inliers



In figure 4 and 5 it is possible to compare the results for the other metrics.

Figure 4: SSIM and Normalized MSE

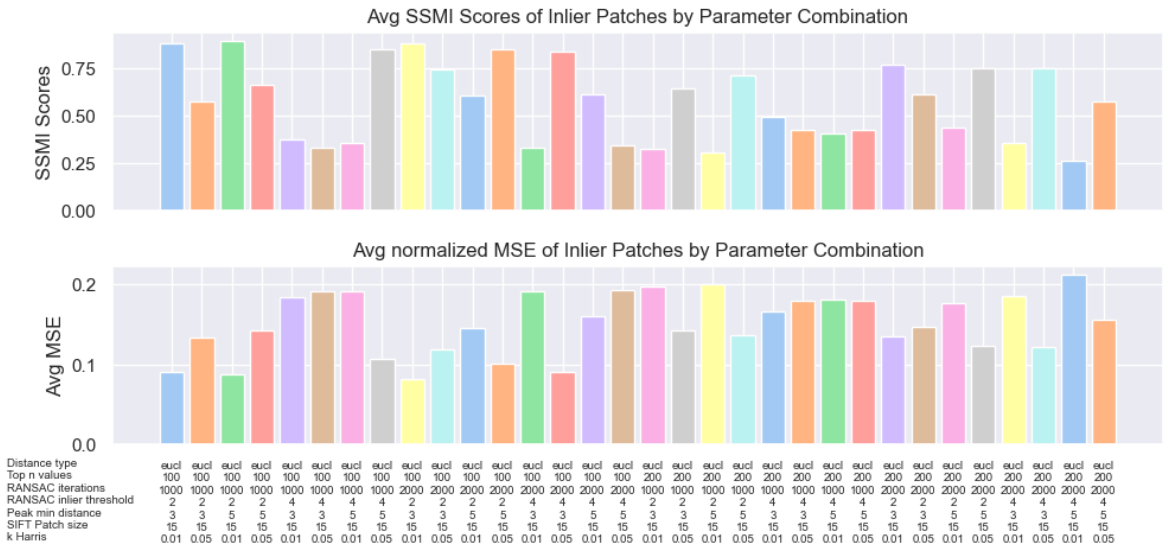
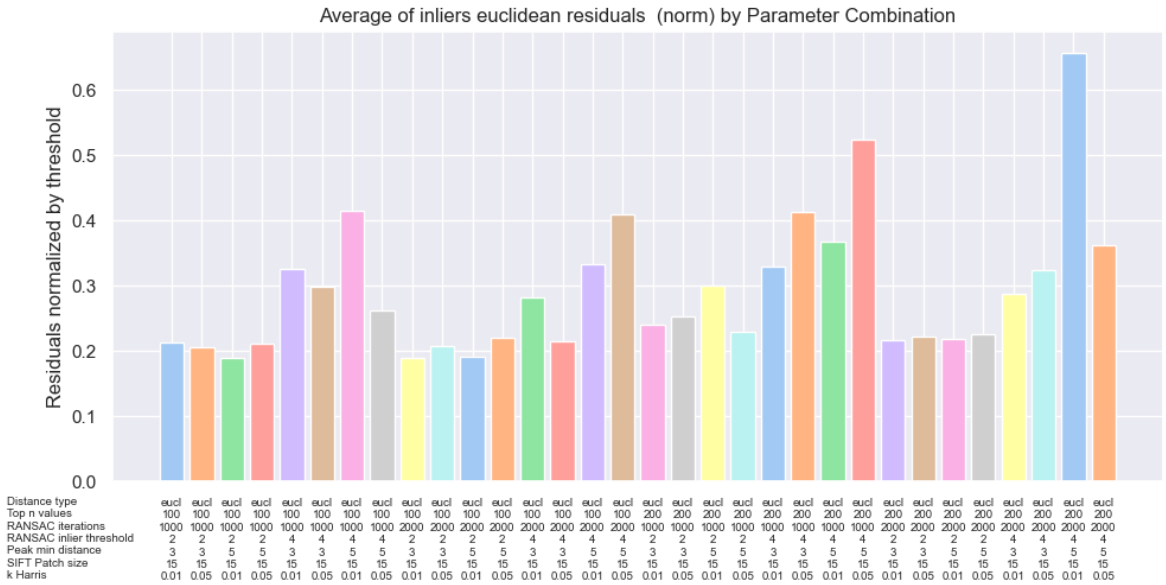


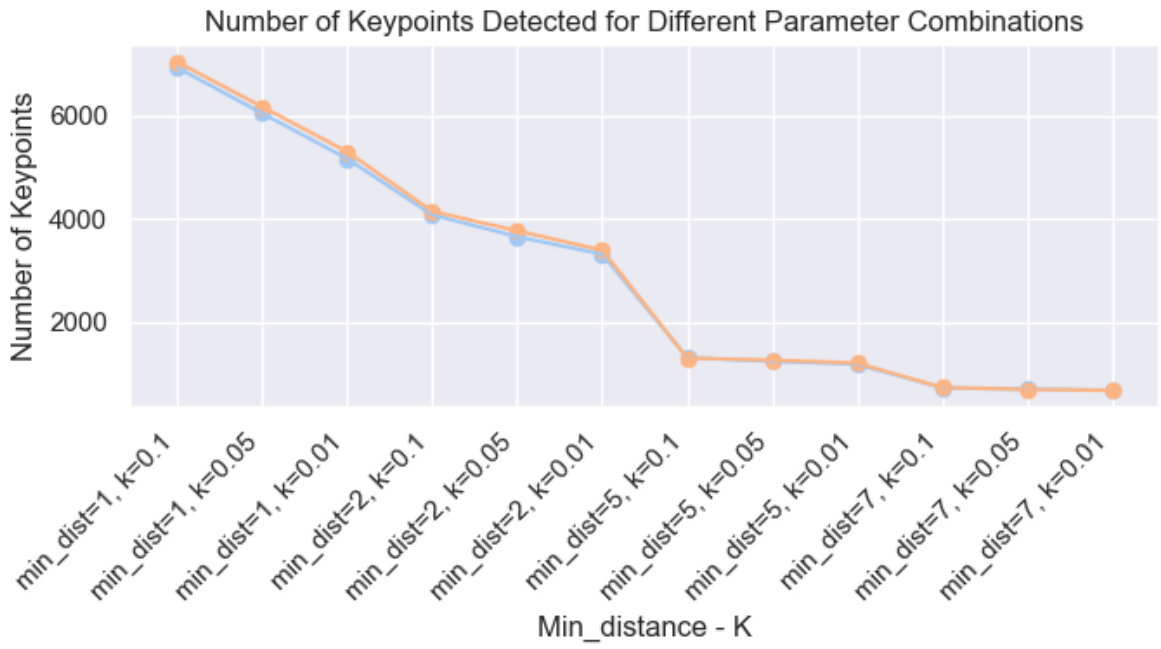
Figure 5: Average of inlier residual score (normalized by RANSAC inlier threshold = 4



4 Observations

Finding a good tradeoff of min_distance, k and a fair amount of keypoints is important in order to reduce the run time, otherwise too many keypoints will be found and calculating the distances between all descriptors might take a lot of time. On the other hand, if there are not enough keypoints it might be difficult to obtain a good result. While I keep the best parameters of previous executions saved in a .json file, the fact that some parameters work nicely on a particular image does not mean that they suit all the image stitching tasks. For instance, for high resolution images setting a lower k and a lower min_distance might be appropriate to reduce the number of keypoints (and consequently the time of computation). In figure 6 an analysis of the number of keypoints for assignment images with different k min_distance values.

Figure 6: keypoints according to K and min_distance



4.1 Possible improvements

In corner_peaks is possible to define regions of the image where to look for more (or less) peaks, knowing in advance the side where the two images have to be stitched it is possible to focus the key points just on the chosen portion of images. Another parameter that would be worth explore is the threshold in corner_peaks. Metric results for certain parameters might slightly vary in different runs thus for a more consistent output the mean of the the result of different iterations should be considered.