

# CS543 Assignment 3

Your Name: Zong Fan

Your NetId: zongfan2

## Part 1: Homography estimation

**Describe your solution, including any interesting parameters or implementation choices for feature extraction, putative matching, RANSAC, etc.**

I employed both SIFT and Harris detector to detect the key points on both left and right image.

**SIFT:** When using SIFT descriptor, it extracted a 128-dimension histogram of gradient orientation from each neighboring window of each key point. Then pairwise euclidean distance was computed among all key points based on the SIFT feature. 100 pairs of key points from left image and right image with lowest distance were used as the candidates for putative matching. RANSAC was employed to find the optimal homography mapping matrix. With total 100 iteration, 4 random pairs were selected to compute a putative homography matrix. Then these 100 points on the left image were transformed with this matrix to get the putative positions on the right image. Squared euclidean distance between these predicted points and corresponding paired key points on the right image. All the point pairs within distance of 16 were treated as the inliers. Iterated these steps and kept the homography matrix that had the highest number of inliers. This optimal matrix was used to warp the left image, then the warped image were merged with original right image.

**Harris:** After getting the locations of key points via Harris detector, the surrounding windows of them with window size as 16 were set to extract the pixel intensity values. These values were flattened to 256-d vectors as the descriptor features of corresponding key points. These features were normalized to 0 mean and unit standard deviation. The following steps were similar, while key points on the two images whose distances less than 75 were set as the matched key points. Also, if the transformed points whose distance to the original paired key points on the right image were less than 64, they were treated as inliers. The iteration times in RANSAC were set as 10000.

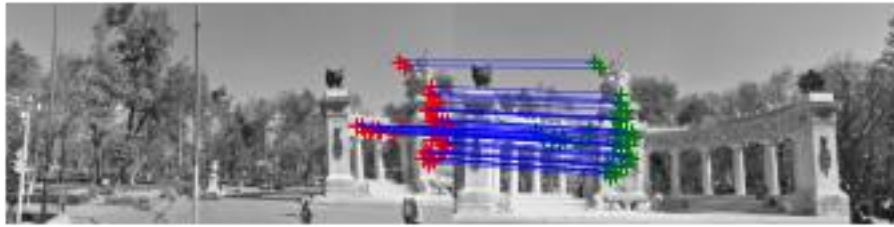
**For the image pair provided, report the number of homography inliers and the average residual for the inliers.**

**SIFT:** 45 inliers (total 100), average residual is 1.7

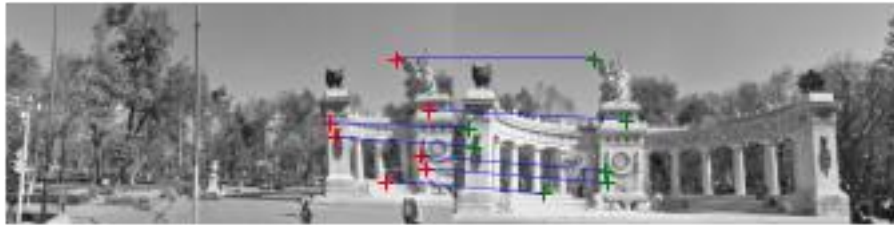
**Harris:** 7 inliers (total 31), average residual is 21.4

**Also, display the locations of inlier matches in both images.**

**SIFT:**

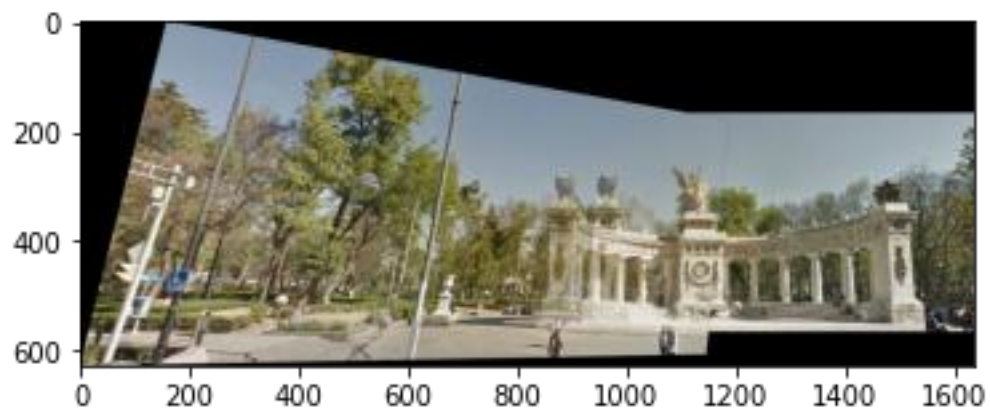
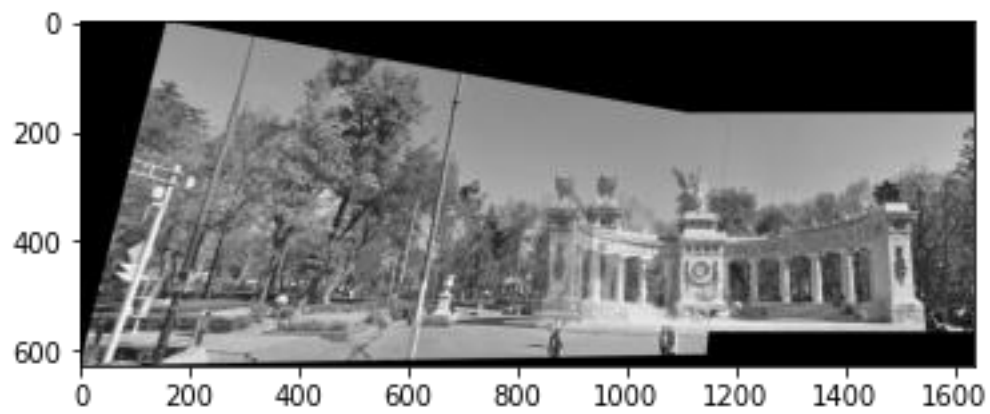


Harris:

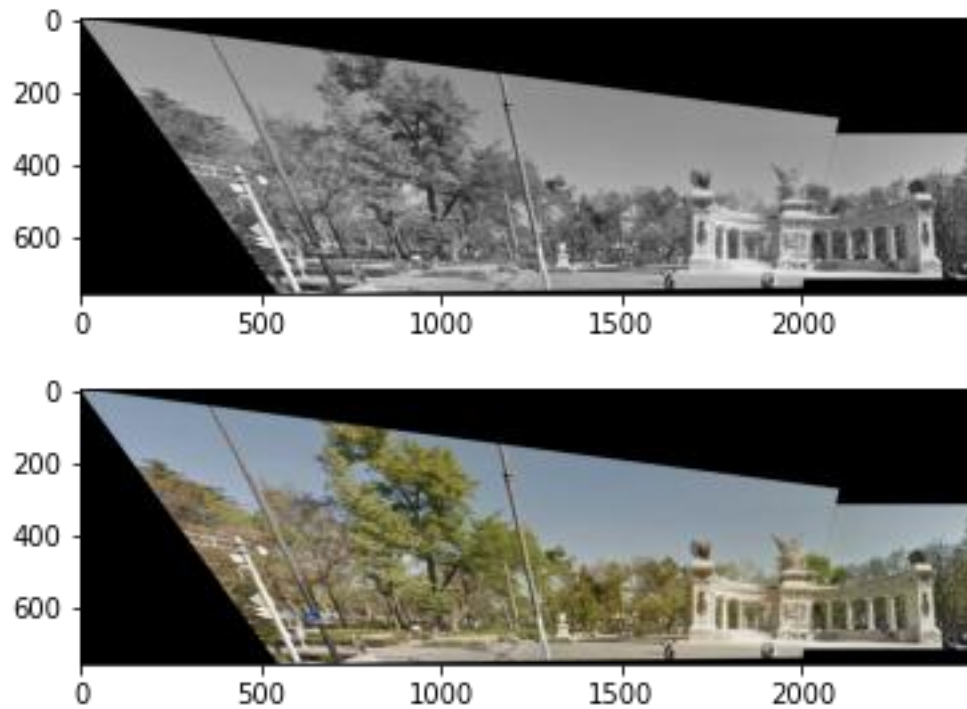


Display the final result of your stitching.

SIFT:



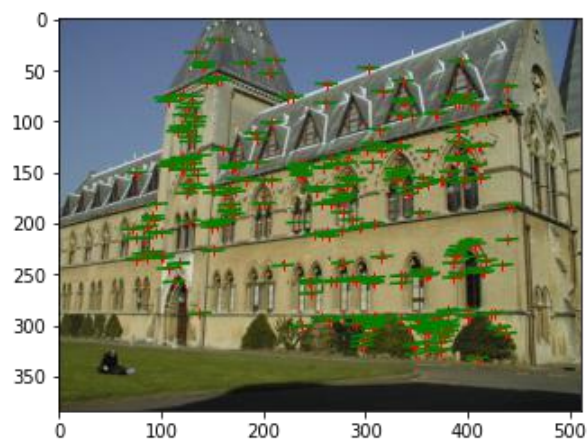
Harris:



## Part 2: Fundamental Matrix Estimation, Camera Calibration, Triangulation

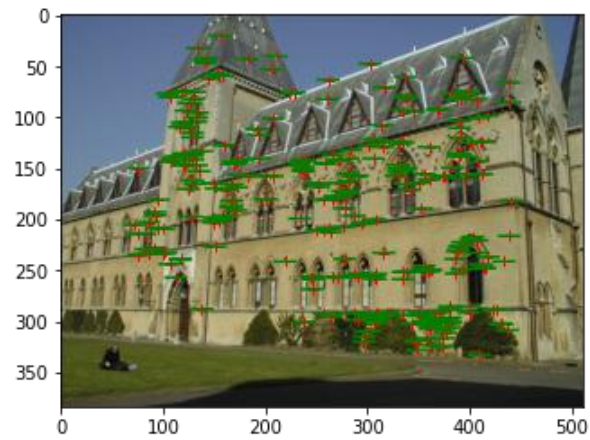
For both image pairs, for both unnormalized and normalized fundamental matrix estimation, display your result (points and epipolar lines) and report your residual.

**Library unnormalized:**



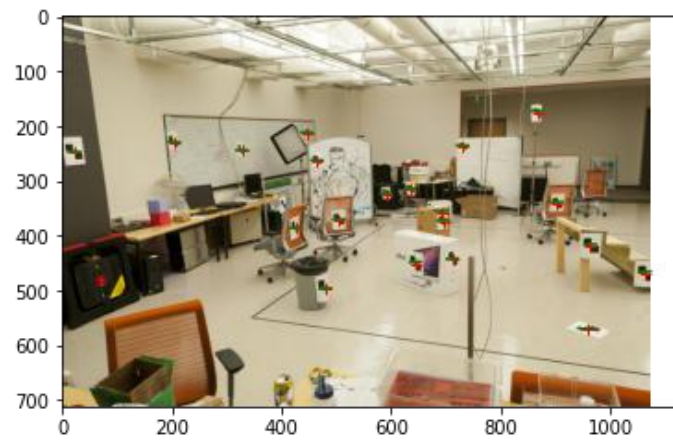
Residual: 0.34

**Library normalized:**



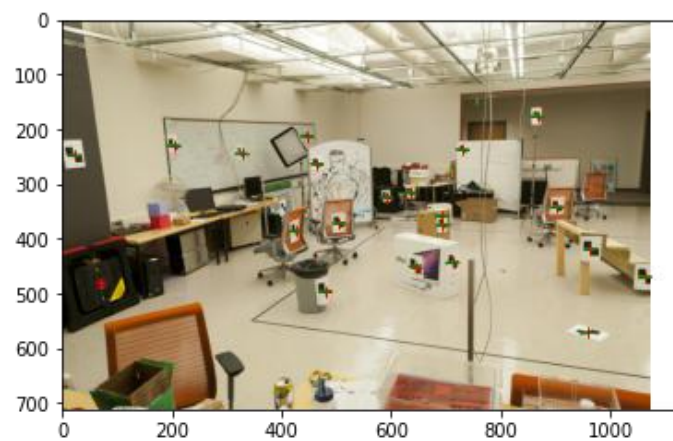
Residual: 0.18

**Lab unnormalized:**



Residual: 2.24

**Lab normalized:**



Residual: 0.62

**For the lab image pair, show your estimated 3x4 camera projection matrices.**

**Left camera projection matrix:**

```
[[ -3.09963996e-03, -1.46204548e-04,  4.48497465e-04,  9.78930678e-01]
 [ -3.07018252e-04, -6.37193664e-04,  2.77356178e-03,  2.04144405e-01]
 [ -1.67933533e-06, -2.74767684e-06,  6.83964827e-07,  1.32882928e-03]]
```

**Right camera projection matrix:**

```
[[ -3.09963996e-03, -1.46204548e-04,  4.48497465e-04,  9.78930678e-01]
 [ -3.07018252e-04, -6.37193664e-04,  2.77356178e-03,  2.04144405e-01]
 [ -1.67933533e-06, -2.74767684e-06,  6.83964827e-07,  1.32882928e-03]]
```

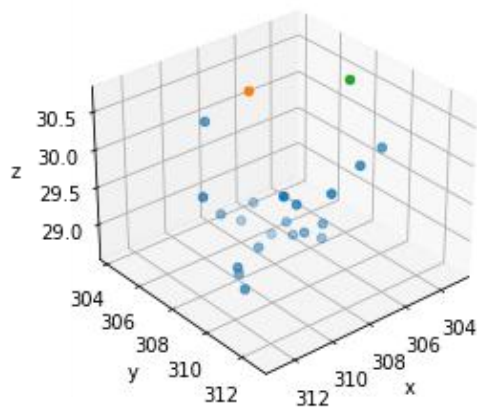
**Report the residual between the projected and observed 2D points.**

**Lab:** residual=1.03

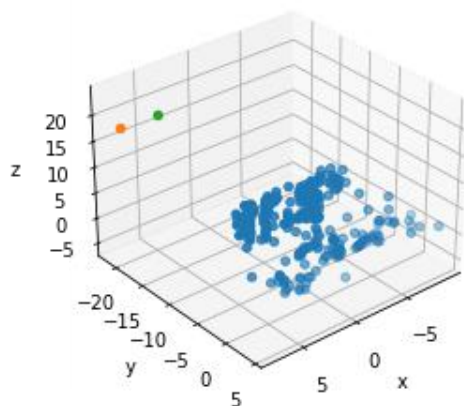
**Library:** residual=0.17

**For both image pairs, visualize 3D camera centers and triangulated 3D points.**

**Lab:**



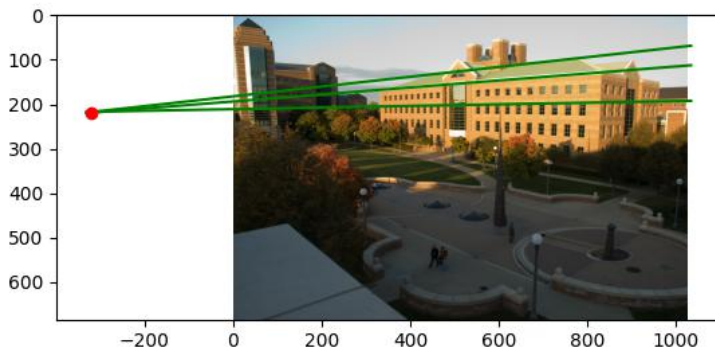
**Library:**



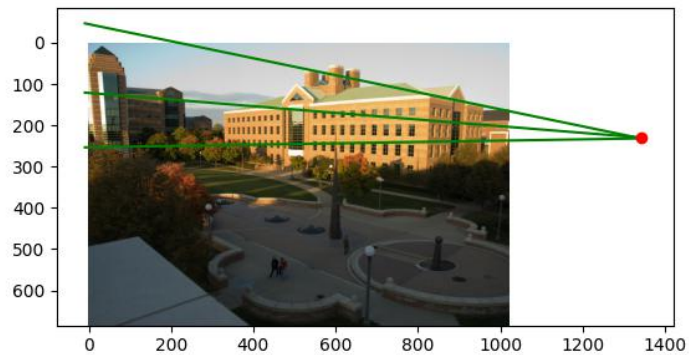
## Part 3: Single-View Geometry

Plot the VPs and the lines used to estimate them on the image plane using the provided code.

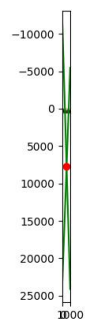
**VPx:**



**VPy:**



**VPz:**



Specify the VP pixel coordinates.



**VPs coordinates: left column - VPx, mid column - VPy, right column - VPz**

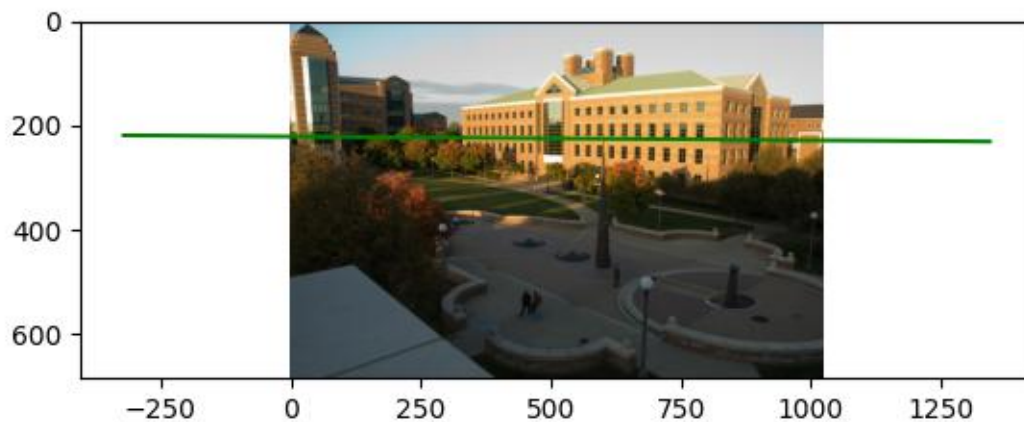
$[[-3.21387197e+02, 1.34317946e+03, 5.41450444e+02]$

$[ 2.18547020e+02, 2.30503363e+02, 7.68057890e+03]$

$[ 1.00000000e+00, 1.00000000e+00, 1.00000000e+00]]$

**Plot the ground horizon line and specify its parameters in the form  $a * x + b * y + c = 0$ .  
Normalize the parameters so that:  $a^2 + b^2 = 1$ .**

$a = -7.18267095e-03, b = 9.99974204e-01, c = -2.20849801e+02$



**Using the fact that the vanishing directions are orthogonal, solve for the focal length and optical center (principal point) of the camera. Show all your work.**

Focal length: -822.92138174; Principle points: [594.34131321, 317.09180841]

**Compute the rotation matrix for the camera.**

Rotation matrix:

$[[ 0.74142118, -0.671002, 0.00713823],$

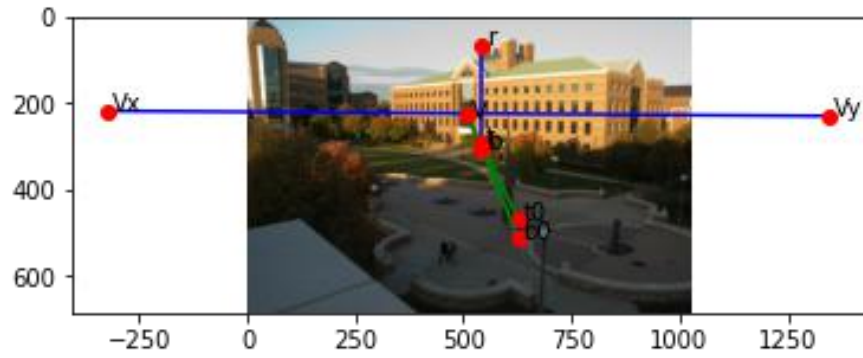
$[ 0.07978696, 0.07758822, -0.99378776],$

$[ 0.66627973, 0.73738483, 0.11106276]]$

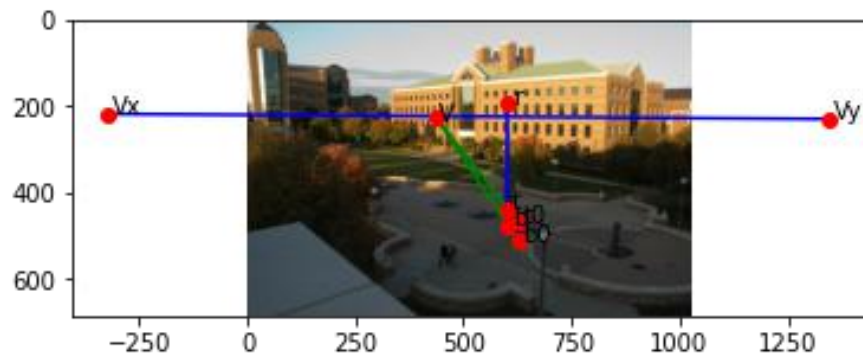
**Estimate the heights of (a) the CSL building, (b) the spike statue, and (c) the lamp posts assuming that the person nearest to the spike is 5ft 6in tall. In the report, show all the lines and measurements used to perform the calculation.**

When height of reference person is 1.6764m, then the height of CSL building is 31.717m, spike statue is 12.057m, lamp posts is 4.907m.

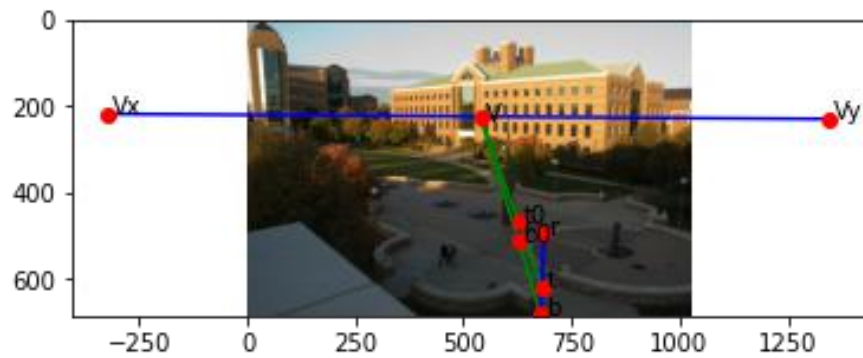
For CSL building,



For spike statue,



For lamp posts,



**How do the answers change if you assume the person is 6ft tall?**

When height of reference person is 1.8288m, then the height of CSL building is 34.6m, spike statue is 13.153m, lamp posts is 5.353m. They increased with same ratio of reference height.



# Extra Credit

Don't forget to include references, an explanation, and outputs to receive credit. Refer to the assignment for suggested outputs.

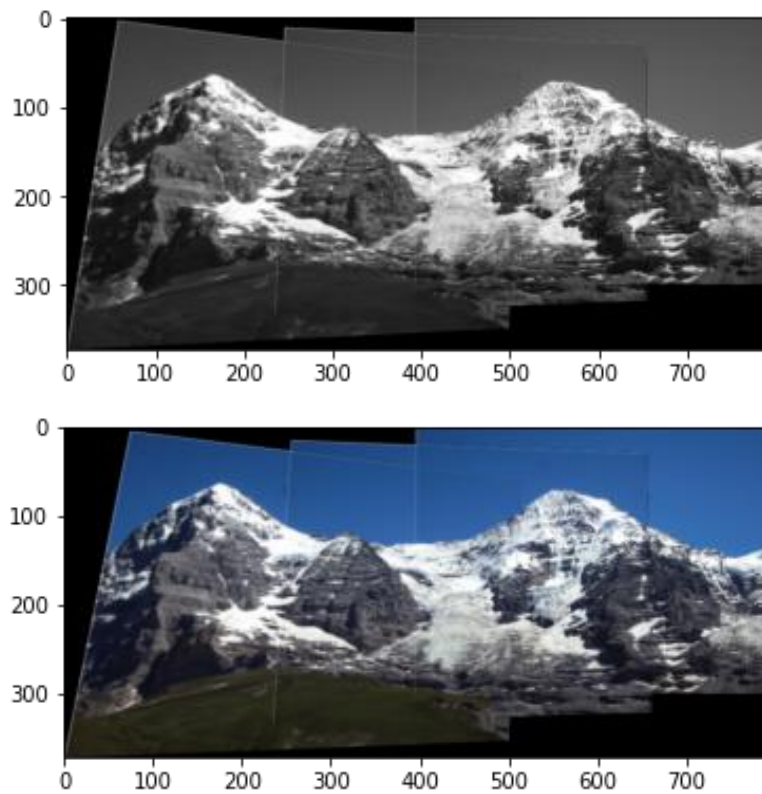
## Part 1

### 1. multiple image homography estimation

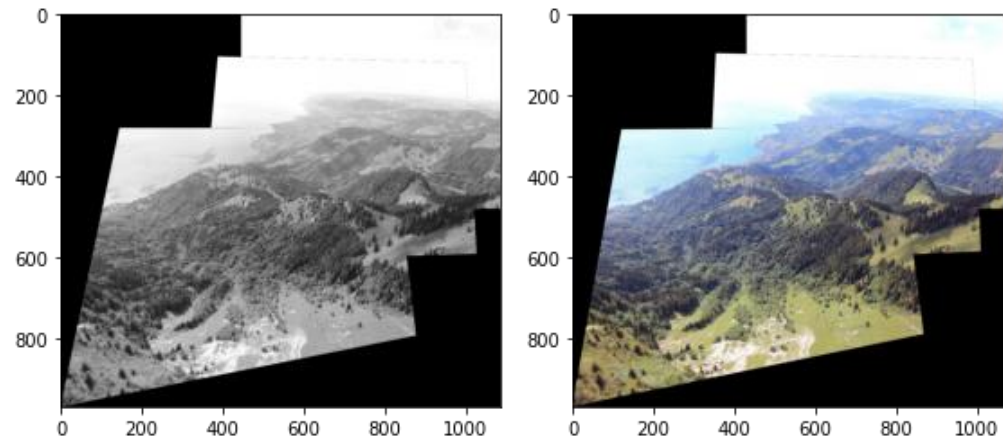
It's similar to two image estimation in part 1. First, use left and middle image to get the merged image as part1. Then use the merged image and original right image to get the final merged image.

Here are the results:

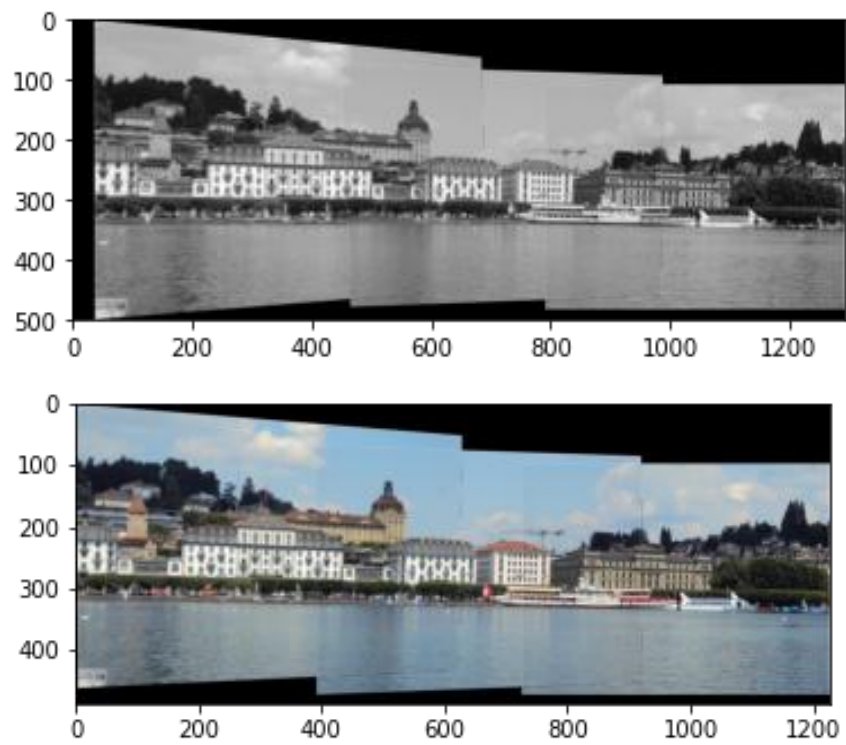
**Hill:**



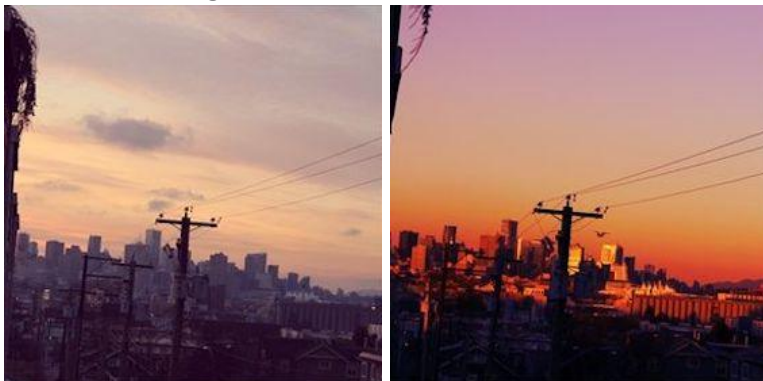
**Ledge:**



**Pier:**



## 2. Difficult image pair estimation



<https://www.pinterest.co.uk/pin/364650901052890279/?d=t&mt=login>

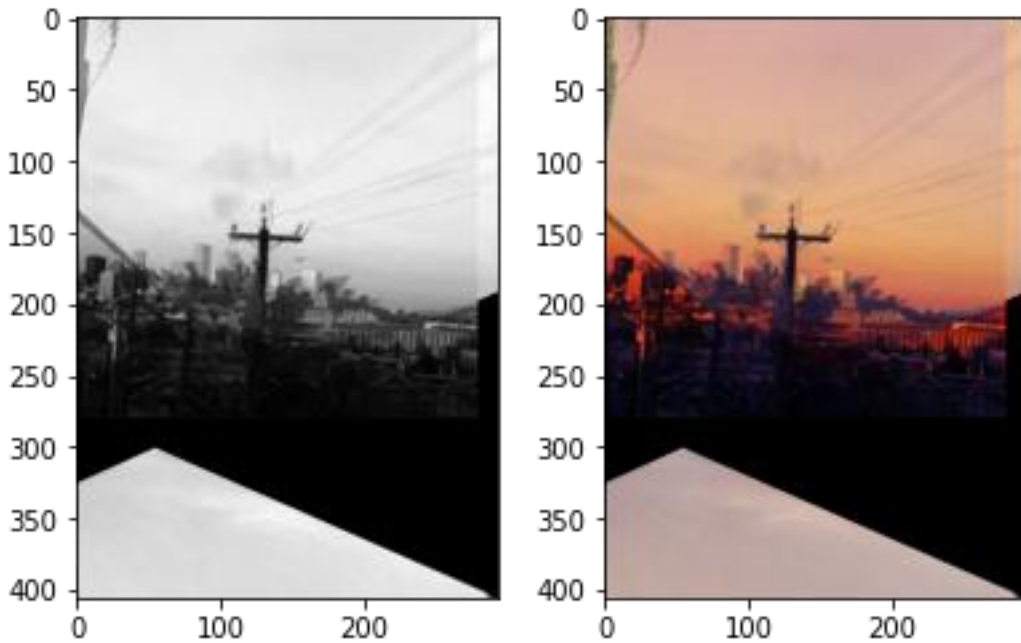
Here are the two images of the same object taken from different times. The severe lighting change poses challenge on the matching between key points, since the descriptor of each matched key points also becomes very different.



To make them still matched, we loose the inlier threshold to 80. Also, we find the key points close the the edge causes the confusion. So we removed all key points on both images which were within the boundary of  $0.15 \times \text{image width}$ . Now the inlier matching becomes like this.

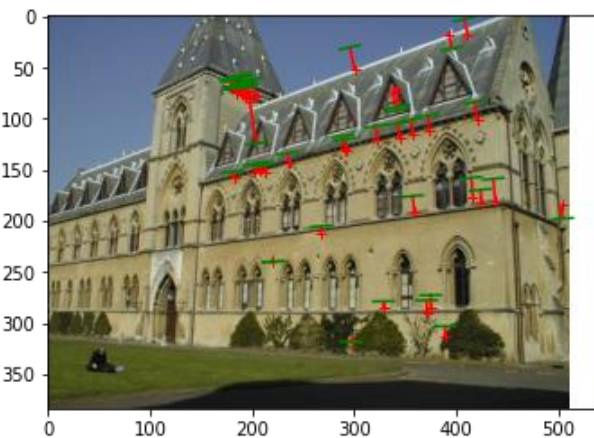


The optimal homography matrix could produce 25 inliers and the residual is 9.5. Here are the merged image.



## Part 2

As part 1, we use SIFT detector and descriptor to calculate the matchness between the key points on the two images. 50 matched pairs with lowest euclidean distance were selected as the candidates. Iterate 50 times and point pairs with distance  $< 1.5$  were treated as inliers. 47 inliers were obtained and the residual was 12.21

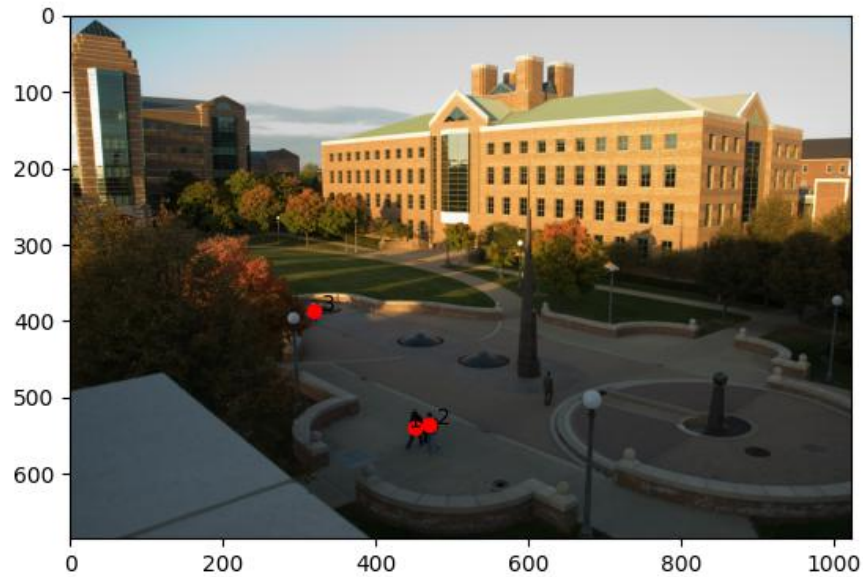


Compared to ground truth matches, the SIFT detected matches has larger noise. Also, the epipolar lines are biased to the top right corner.

## Part 3

### 1. Additional measurements

Measure another 3 persons on the image.



When the reference person is 1.68m, then person 1 is 1.87m, person 2 is 1.72m, person 3 is 1.81m. It seems the person 1 is tallest. However, since the person 3 occupies quite few pixels on the image, so pixel picking step may inject much noise. Basically, the bigger the object or the closer the object to the reference, then the estimation is more accurate.