ECE661 Fall 2024: Homework 4

By – Aayush Bhadani

abhadani@purdue.edu

The goal is to perform automated interest point detection and correspondence search for a given pair of images of the same scene.

Methods used for detection:

- Harris Corner Detection
- SIFT
- CNN-based SuperPoint

To establish correspondence between the two views, the following methods were employed:

- SSD (Sum of Squared Differences) & NCC (Normalized Cross Correlation) as feature similarity measure.
- GNN-based SuperGlue feature matching network

3. Theory Question:

Let an image be represented as $f(x,y)$ and its $\sigma - smoothed$ version be $ff(x,y,\sigma)$ with

$ff(x,y,\sigma) = \iint_{-\infty}^{\infty} f(x',y')g(x-x',y-y')dx'dy'$ where $g(x,y)$ is the Gaussian.

We can leverage an important result from state space theory for images to show the relationship between LoG and DoG, suggesting that LoG can be approximated using DoG. This result is given below:

$$\frac{\partial}{\partial\sigma} ff(x,y,\sigma) = \sigma\nabla^2 ff(x,y,\sigma)$$

$$\left(\begin{array}{c} The\ proof\ for\ the\ above\ shows\ that: \\ \nabla^2 ff(x,y,\sigma) = f(x,y) * h(x,y,\sigma) \quad ; \quad \frac{\partial}{\partial\sigma} ff(x,y,\sigma) = \sigma f(x,y) * h(x,y) \\ \text{where } h(x,y,\sigma) = -\frac{1}{2\pi\sigma^4}\left(2 - \frac{x^2+y^2}{\sigma^2}\right) e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \end{array} \right)$$

$\nabla^2 ff(x,y,\sigma)$ is the LoG of $f(x,y)$, which means $LoG\big(f(x,y)\big) = \frac{\partial}{\partial\sigma} ff(x,y,\sigma)$

LoG of an image $f(x,y)$ can be approximated by subtracting $(\sigma + \delta\sigma)$ smoothed version of $f(x,y)$ from $\sigma -$ smoothed version. This difference of two Gaussian-smoothed versions of $f(x,y)$ for two different values of σ is the DoG.

It is computationally more efficient to compute LoG of an image as a DoG for the same value of σ.

This is because the Gaussian $g(x, y)$ is separable in x and y, thus each of the 2-D smoothing for DoG can be carried out by two applications of 1-D smoothing. However, this separation of x and y is not possible in LoG operator. Additionally, we are likely to use a smaller size operator for DoG than for LoG.

4.

4.1 Task 1

<u>Harris Corner detector:</u>

It is used for interest point detection. We start by converting our color image to grayscale and normalize it. Next, we create Haar wavelet filters at scale σ . These filters are of size MxM, where M is the smallest even integer greater than 4σ . We convolve these x & y-oriented Haar filters with our image to get d_x and d_y. Next, we construct C matrix in 5σX5σ neighbourhood of the pixel.

$$C = \begin{bmatrix} \sum d_x^2 & \sum d_x d_y \\ \sum d_x d_y & \sum d_y^2 \end{bmatrix}$$

Here the summation is over all the pixels in the 5σX5σ neighbourhood.

If λ_1 and λ_2 are the eigenvalues of matrix C where $\lambda_1 \geq \lambda_2$ we get:

$$Tr(C) = \Sigma d_x^2 + \Sigma d_y^2 = \lambda_1 + \lambda_2$$

$$det(C) = \Sigma d_x^2 \, \Sigma d_y^2 - \Sigma d_x d_y \Sigma d_x d_y = \lambda_1 \lambda_2$$

The ratio $\frac{det(C)}{[Tr(C)]^2} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$ can be used directly thresholded for corner detection.

The response function is calculated as:

$$R = \det(C) - k\big(Tr(C)\big)^2 \; ; \; with\; k = 0.04 - 0.06$$

For threshold, we can use $R_{threshold} = 0.01 \cdot R_{max}$

Lastly, we do non-maximum suppression to thin the responses and identify the most significant corners. We do this by taking the max response in a given window.

In the correspondence figures, I am using only top p points.

After extracting corners from images f_1, f_2 of an image pair, we can establish correspondence by directly comparing gray levels in a window around the corner pixel in one image with gray levels in a similar window around the corresponding pixel in the other image. For this, we use the following two metrics:

<u>SSD (Sum of Squared Differences):</u>

$$SSD = \sum_i \cdot \sum_j |f_1(i,j) - f_2(i,j)|^{\,2}$$

Limiting the number of correspondence lines in the output image by using the p percentile(30) of SSD.

NCC (Normalized Cross Correlation):

$$NCC = \frac{\Sigma\Sigma(f_1(i,j) - m_1)(f_2(i,j) - m_2)}{\sqrt{\left(\Sigma\Sigma(f_1(i,j) - m_1)\right)^2 \left(\Sigma\Sigma(f_2(i,j) - m_2)\right)^2}}$$

Here m_1, m_2 are the means in f_1, f_2 respectively.

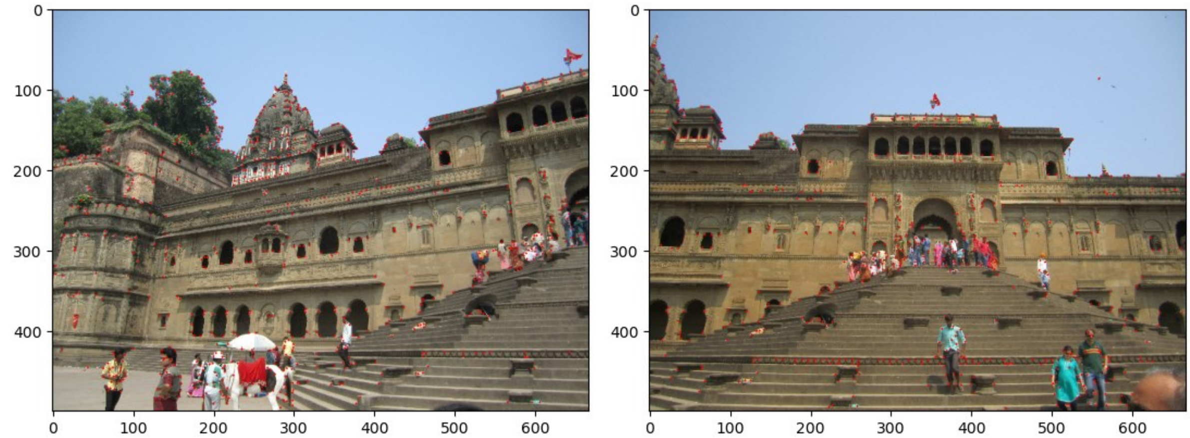
To limit the number of correspondence lines in the output image, I am using a threshold(0.8) on NCC.

4.1.1

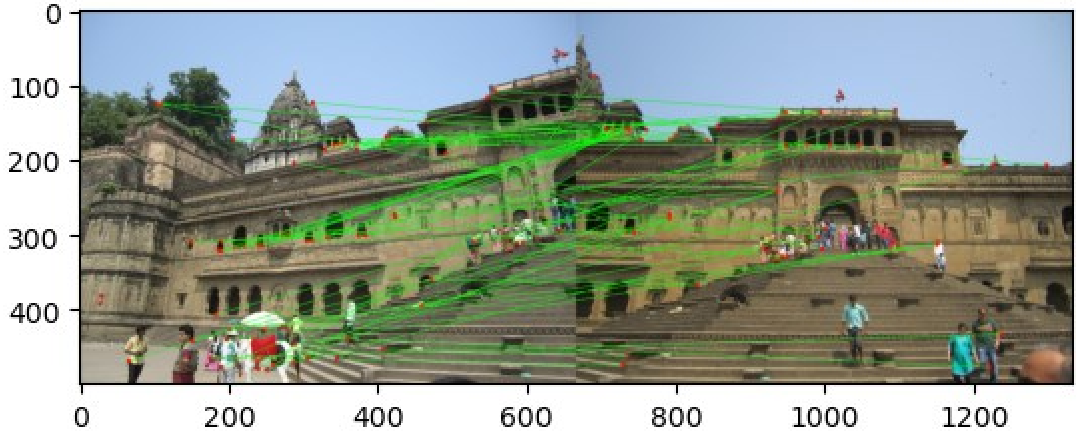
Image pair 1:

Harris detector for $\sigma = 0.8$

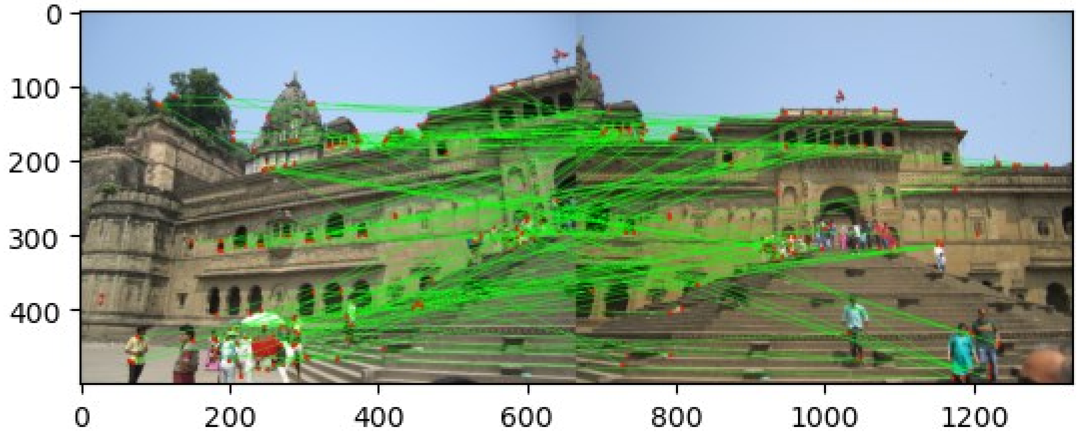
Points:



Correspondences using SSD:

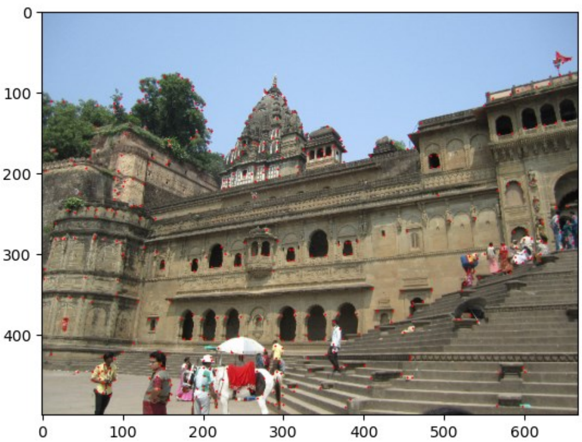


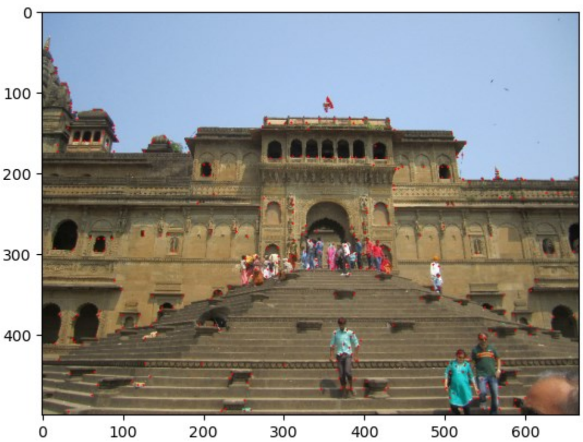
Correspondences using NCC:



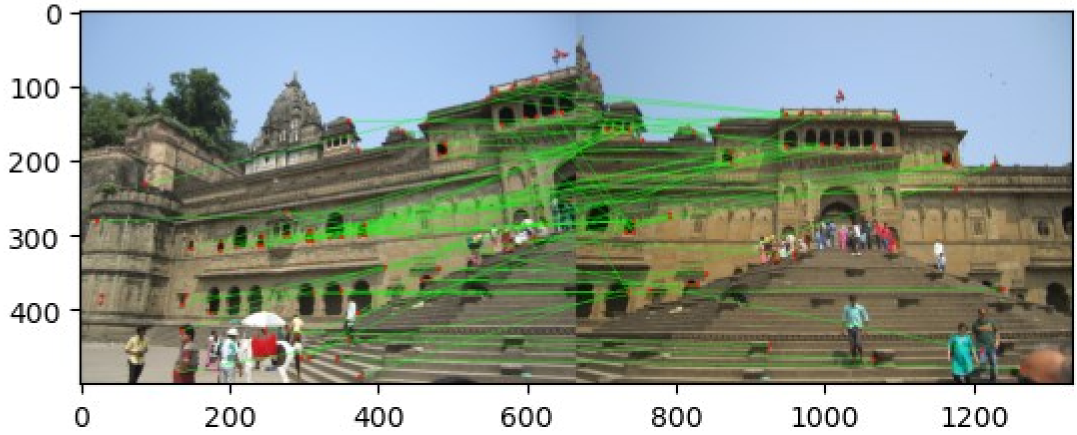
<u>Harris detector for $\sigma = 1.0$</u>

Points:

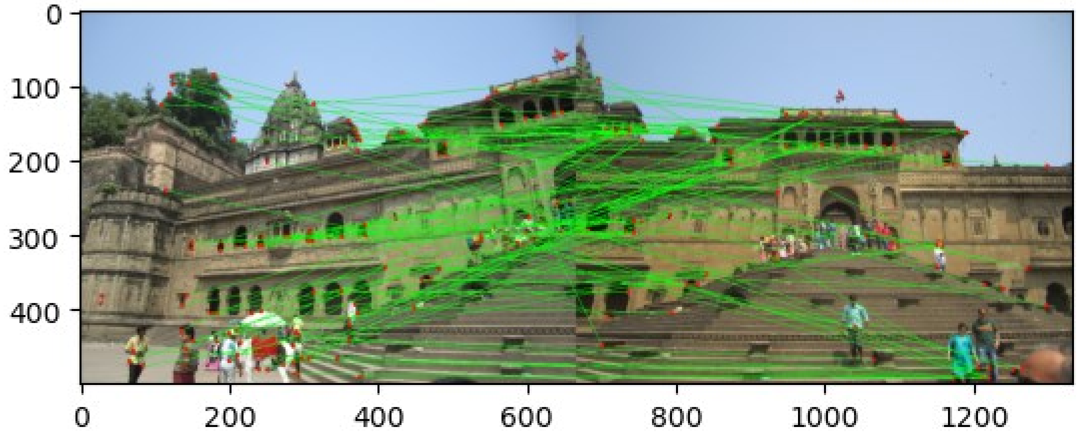




Correspondences using SSD:

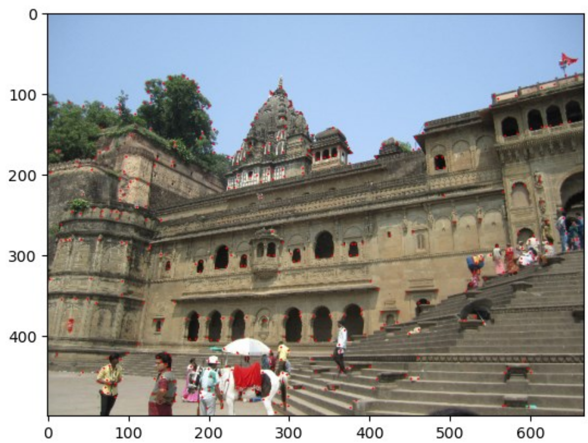


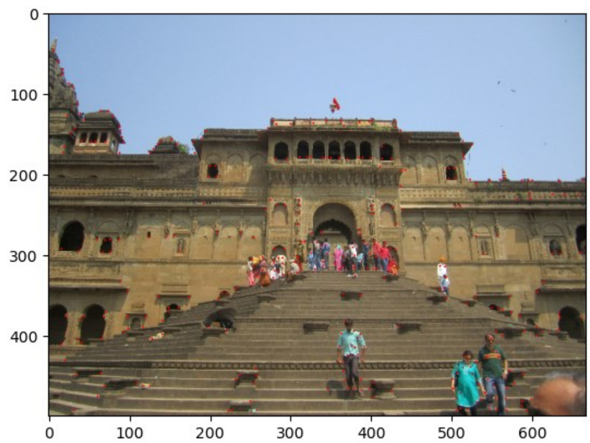
Correspondences using NCC:



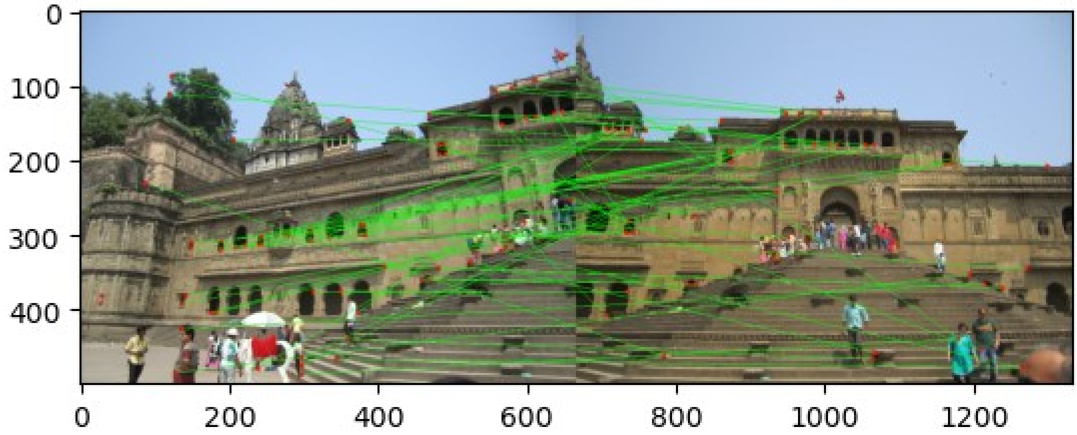
<u>Harris detector for $\sigma = 1.2$</u>

Points:

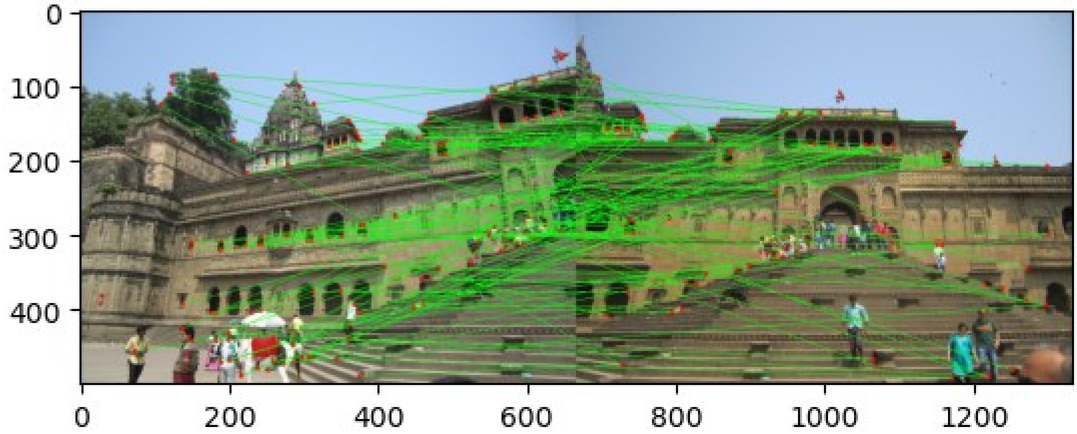




Correspondences using SSD:

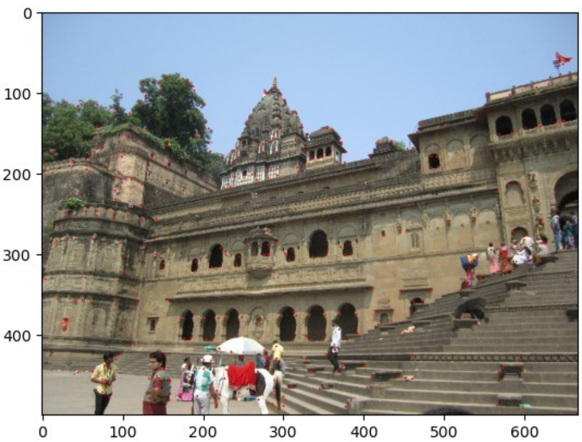


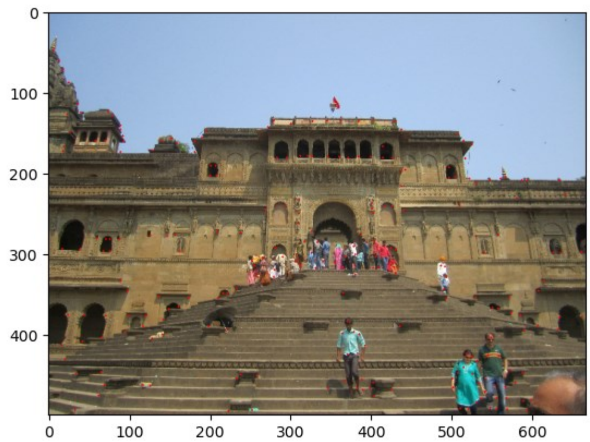
Correspondences using NCC:



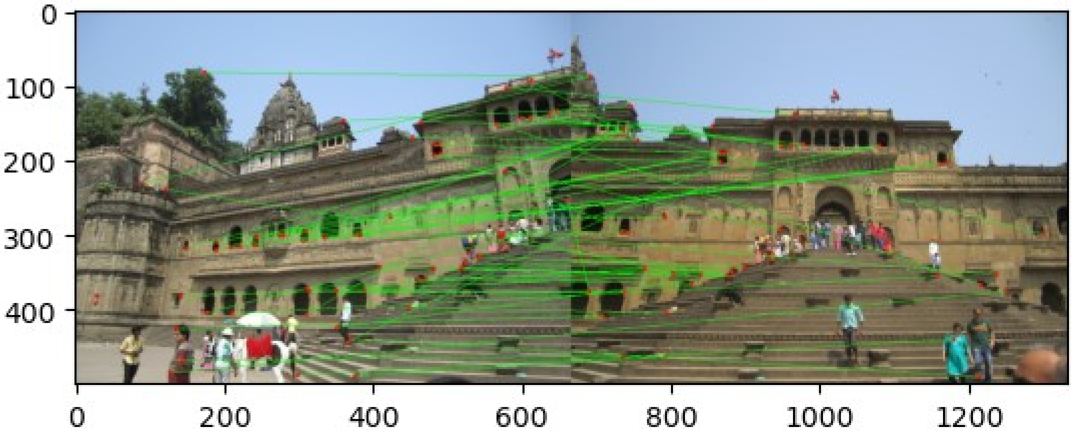
<u>Harris detector for $\sigma = 1.4$</u>

Points:

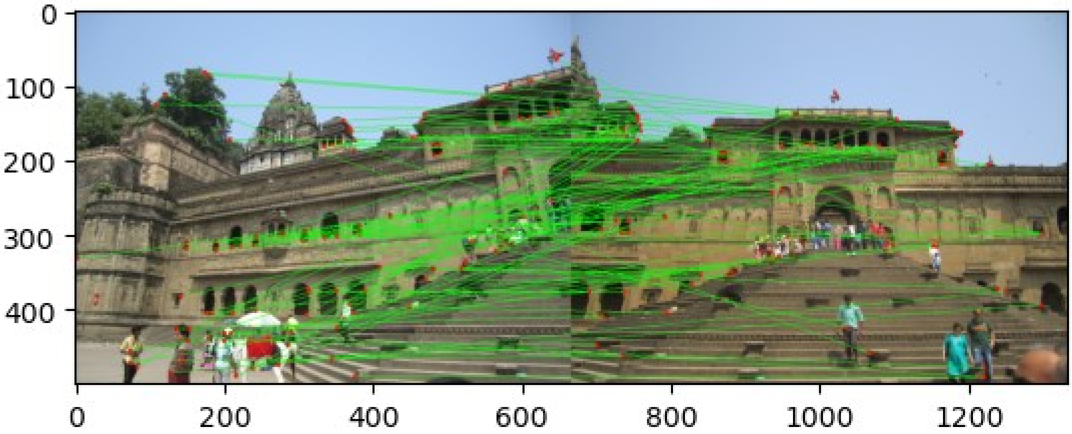




Correspondences using SSD:

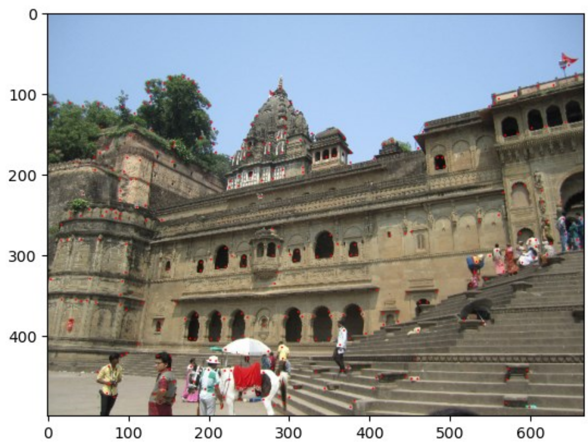


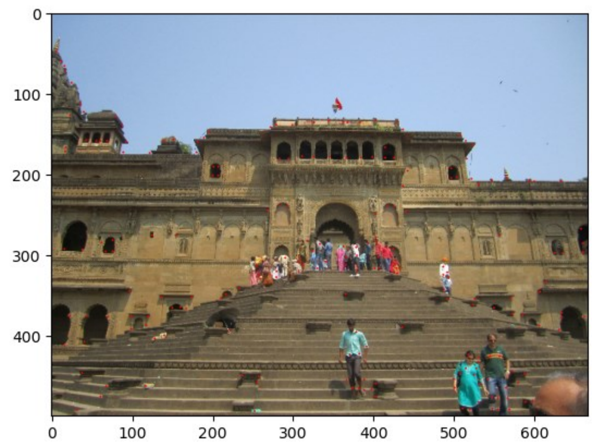
Correspondences using NCC:



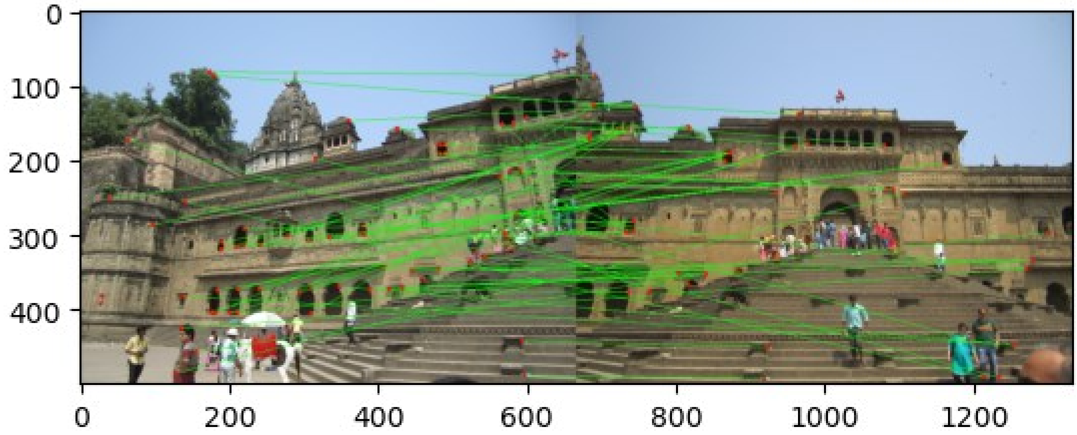
Harris detector for $\sigma = 1.6$

Points:

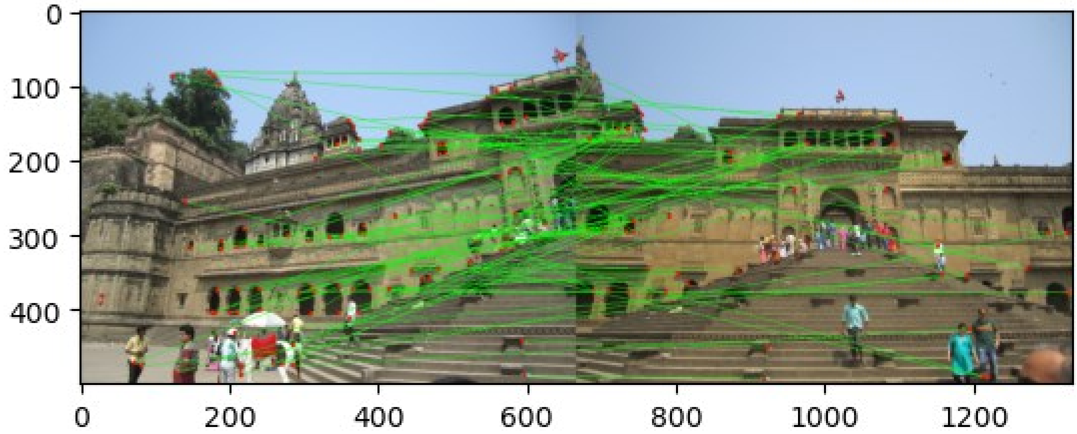




Correspondences using SSD:

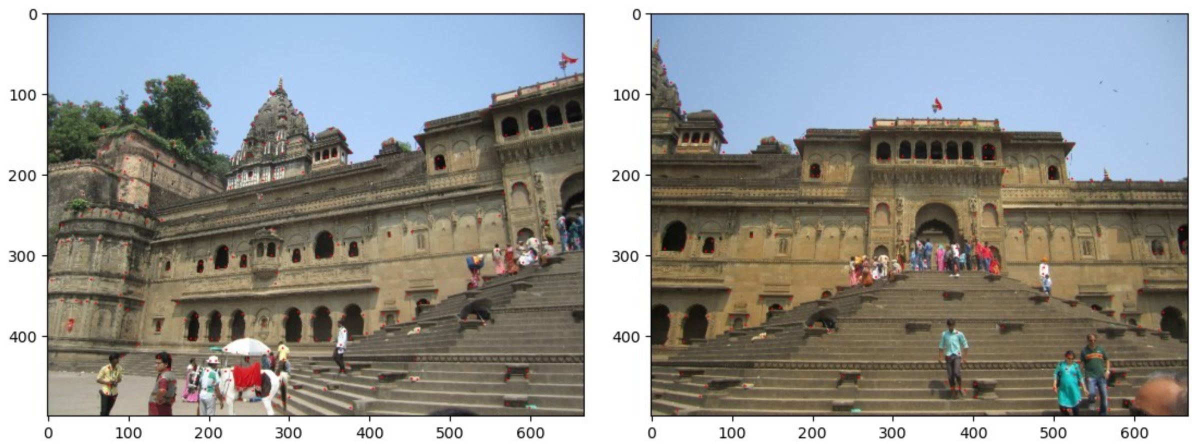


Correspondences using NCC:

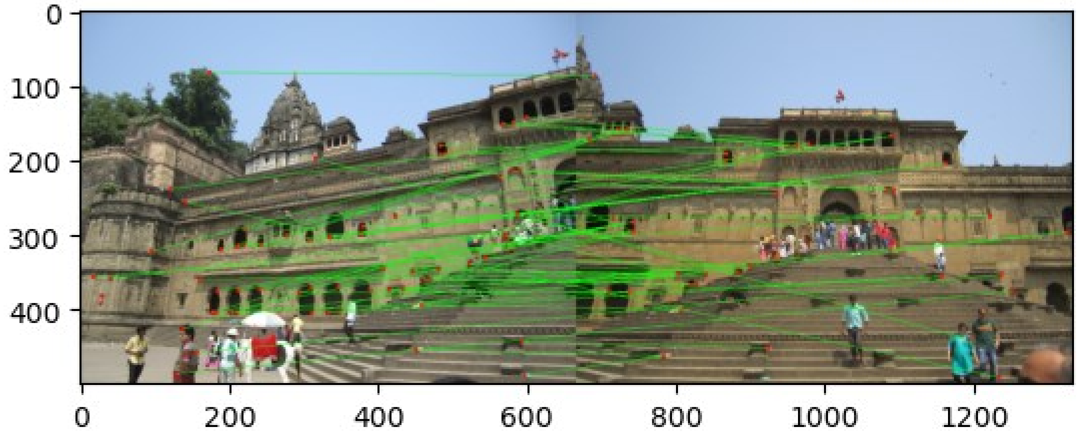


Harris detector for $\sigma = 1.8$

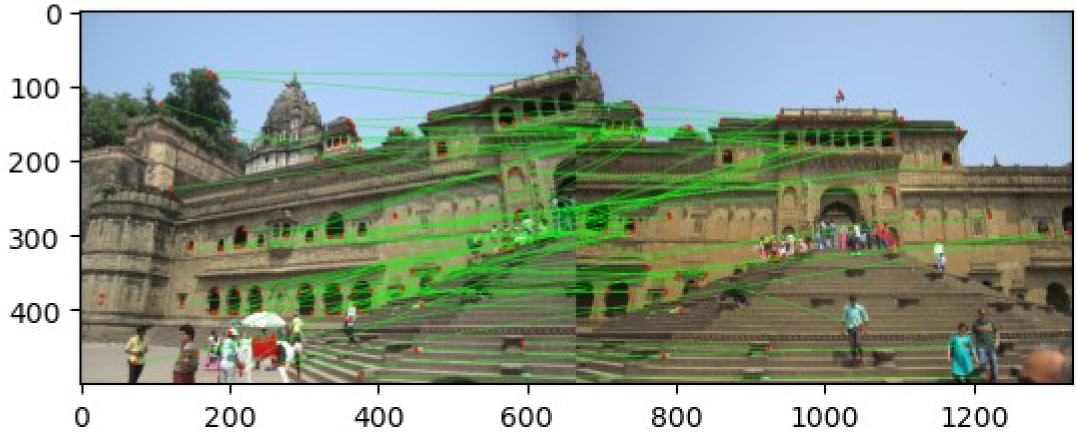
Points:



Correspondences using SSD:

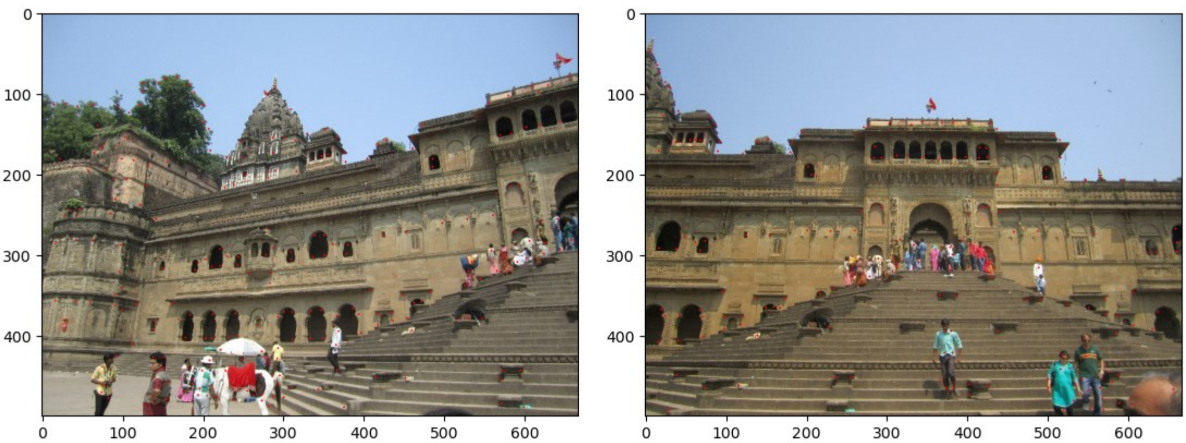


Correspondences using NCC:

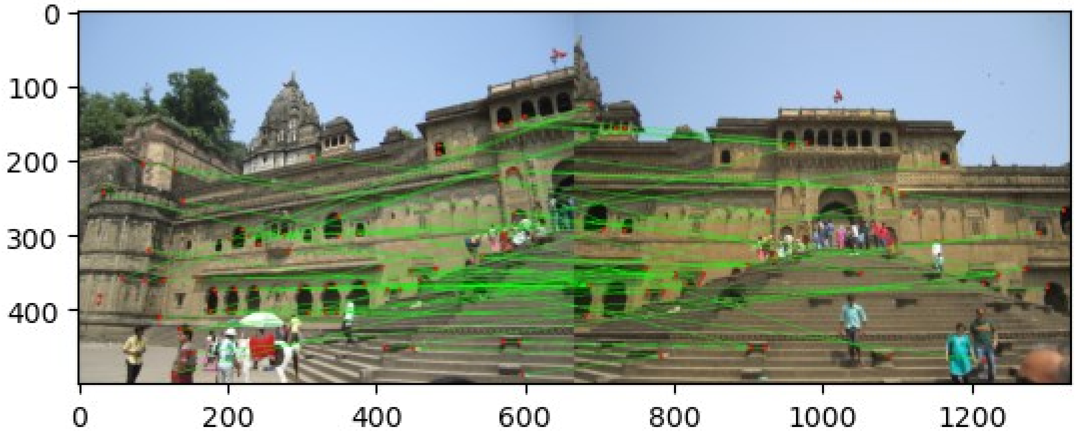


<u>Harris detector for $\sigma = 2.0$</u>

Points:



Correspondences using SSD:



Correspondences using NCC:

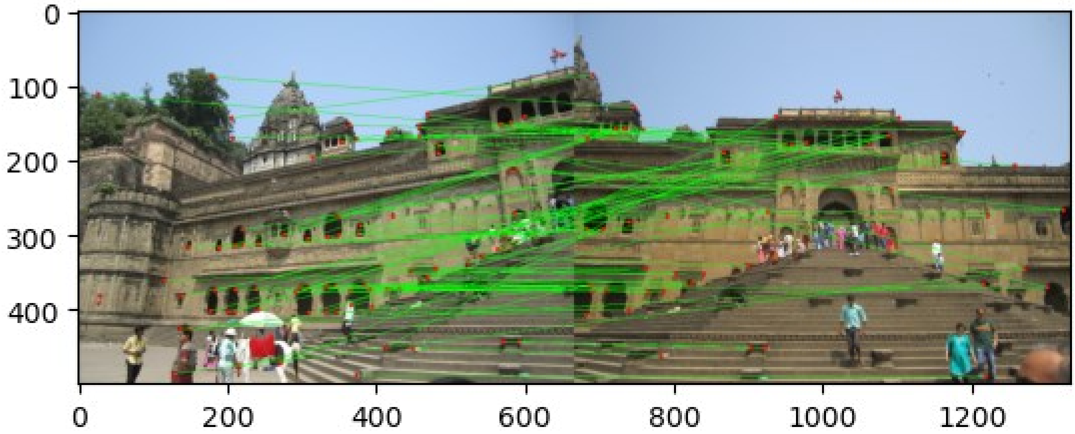
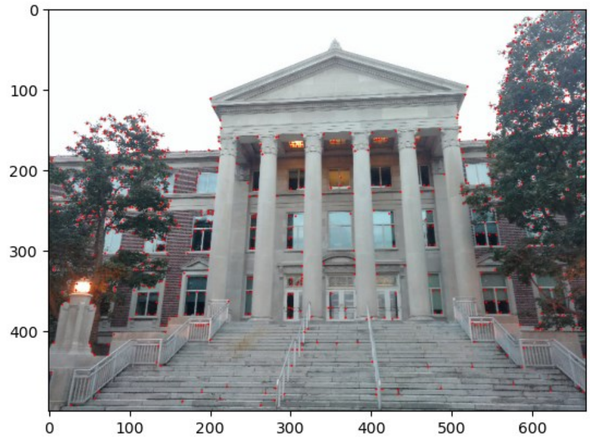


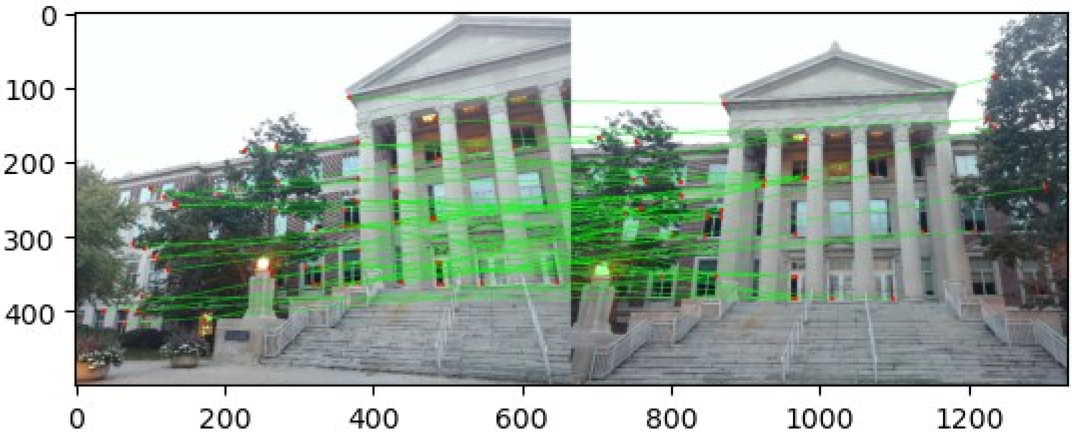
Image pair 2:


Harris detector for $\sigma = 0.8$

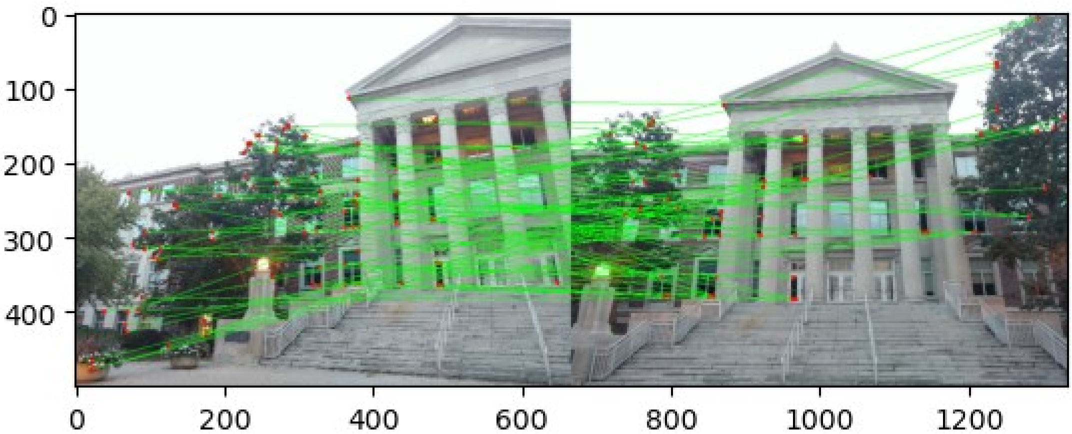
Points:



Correspondences using SSD:

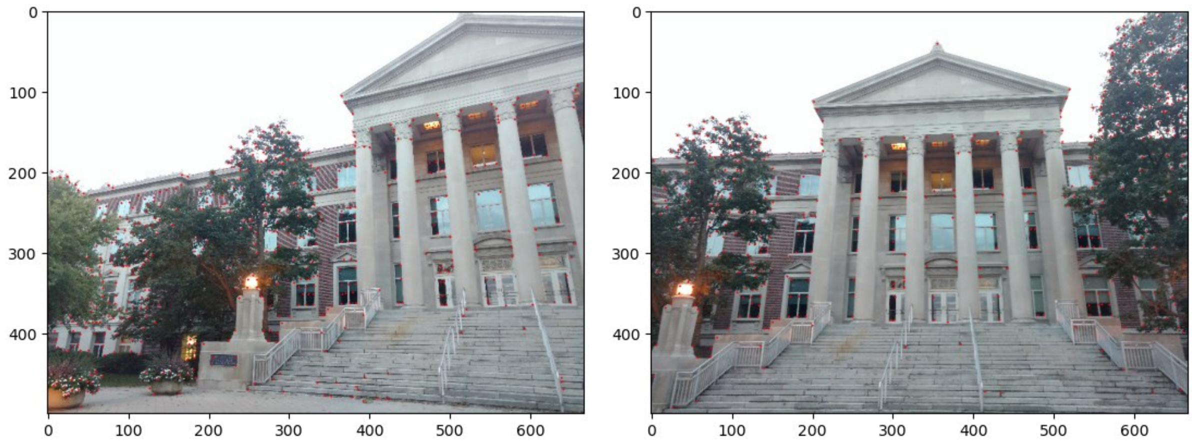


Correspondences using NCC:

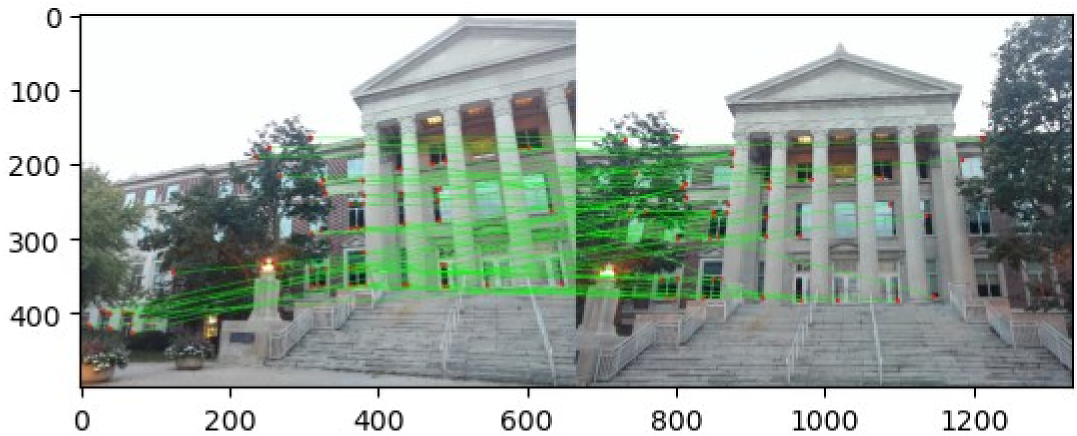


Harris detector for $\sigma = 1.0$

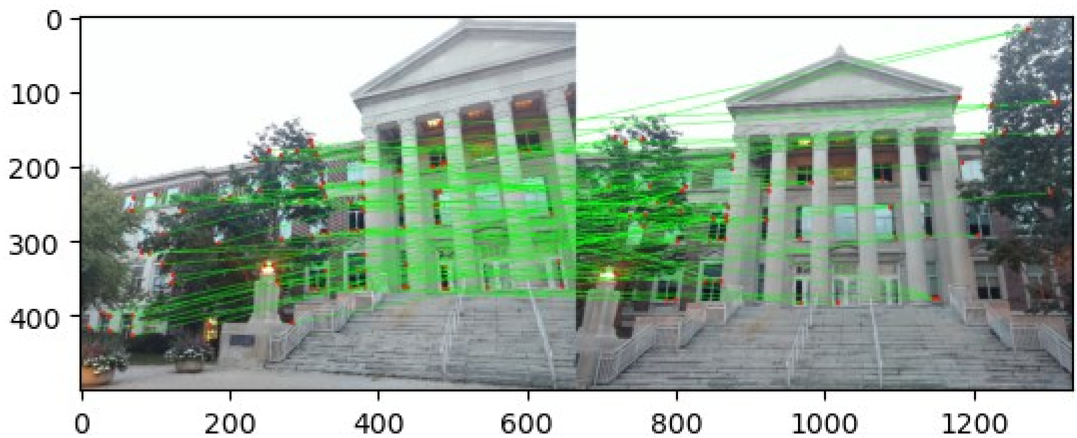
Points:



Correspondences using SSD:

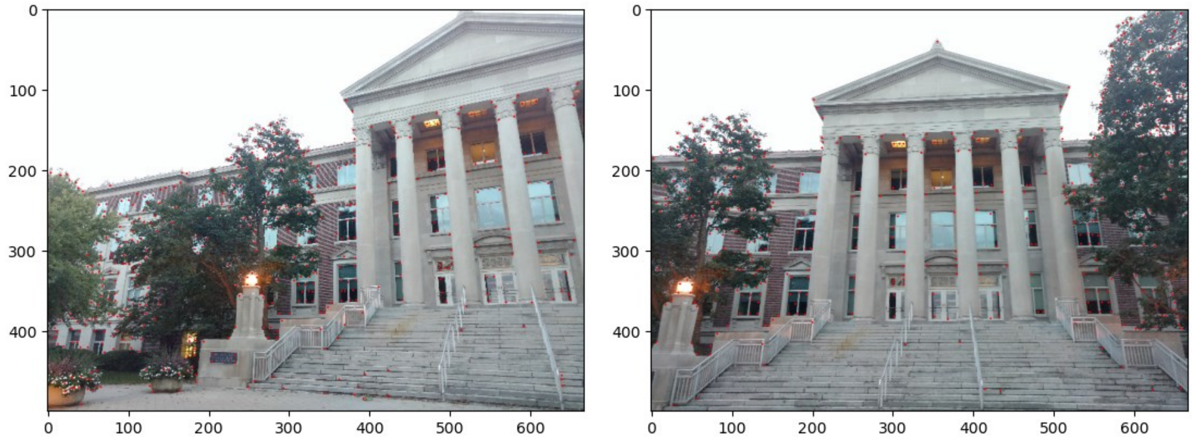


Correspondences using NCC:

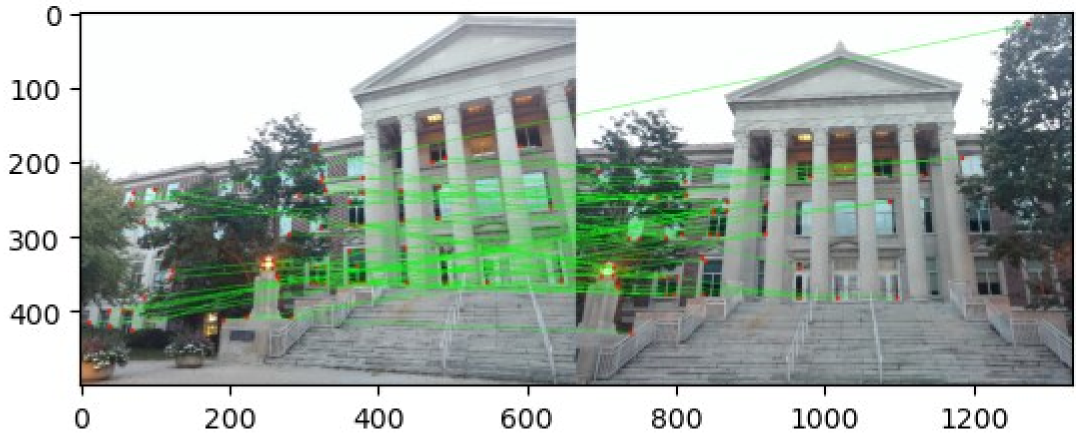


<u>Harris detector for $\sigma = 1.2$</u>

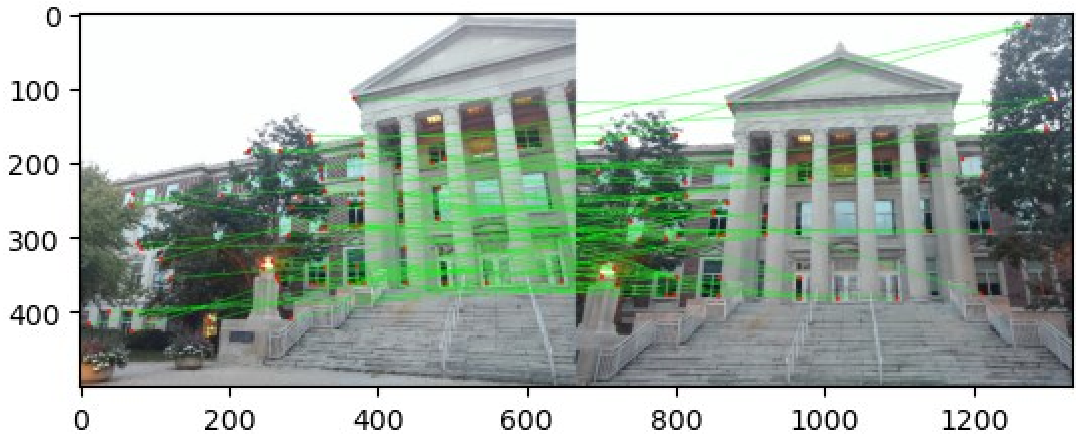
Points:



Correspondences using SSD:

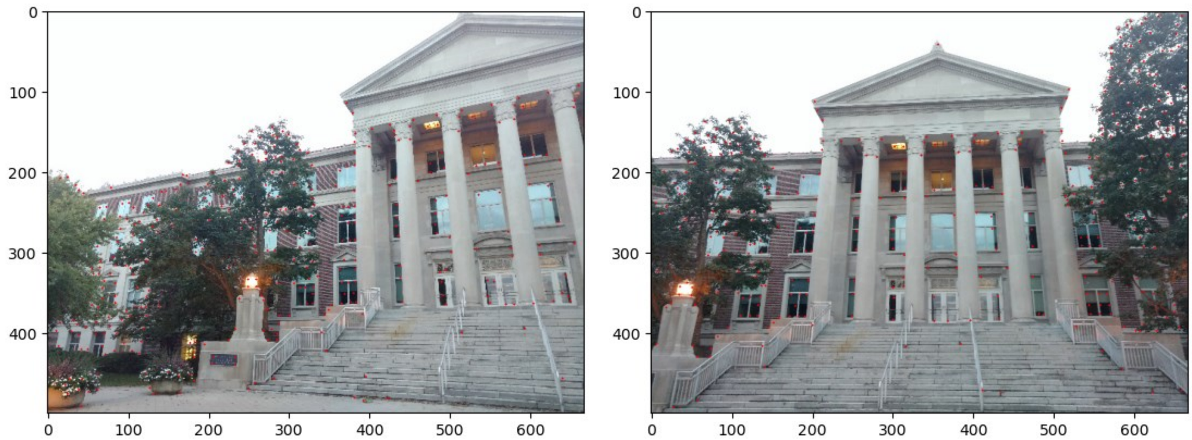


Correspondences using NCC:

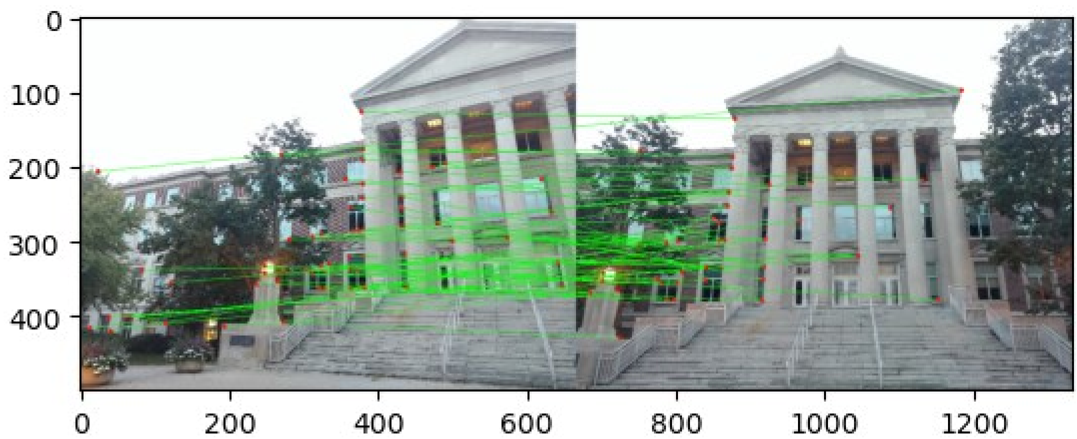


Harris detector for $\sigma = 1.4$

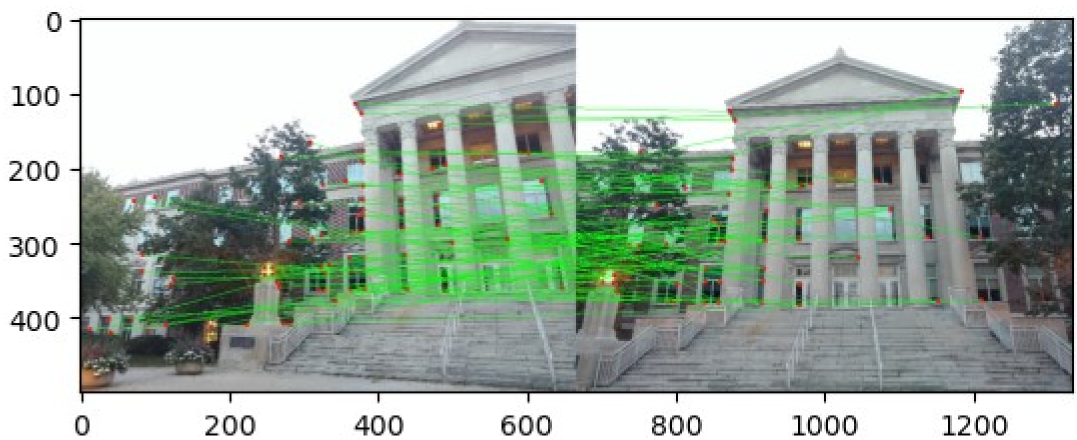
Points:



Correspondences using SSD:

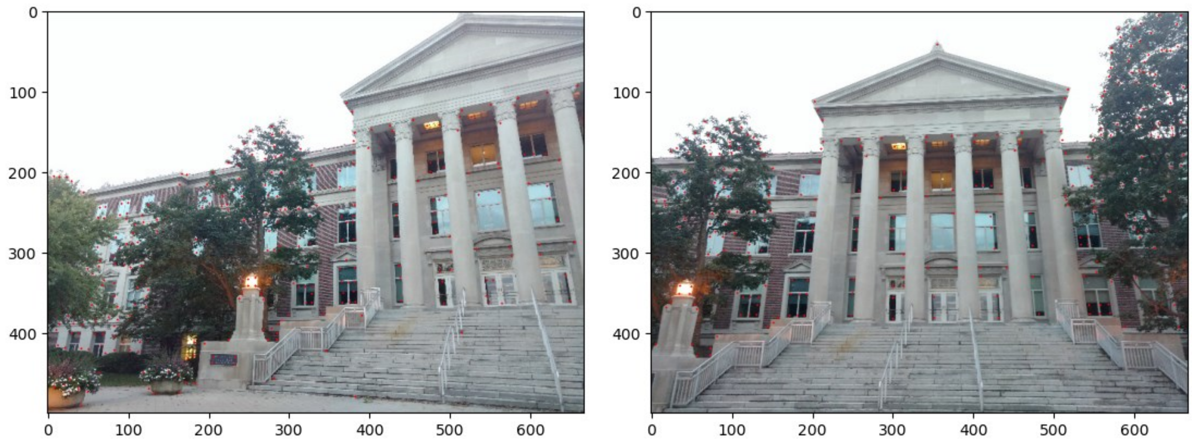


Correspondences using NCC:

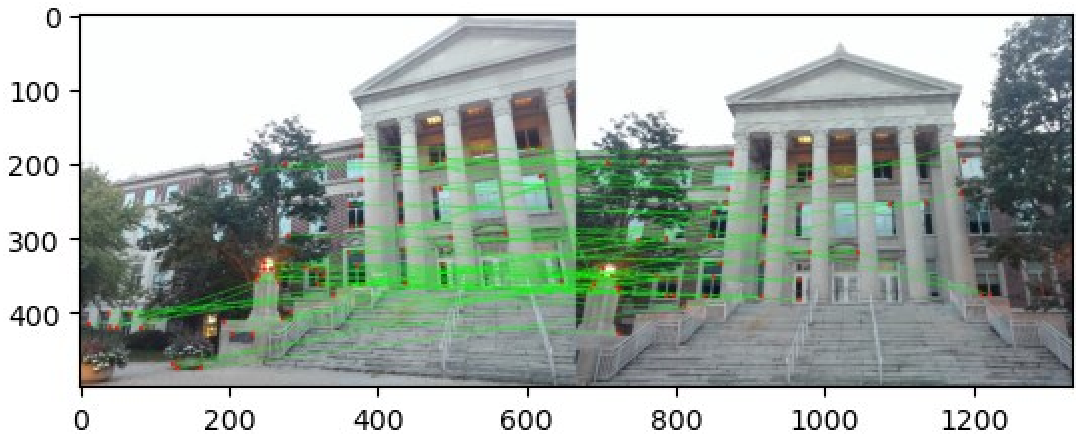


Harris detector for $\sigma = 1.6$

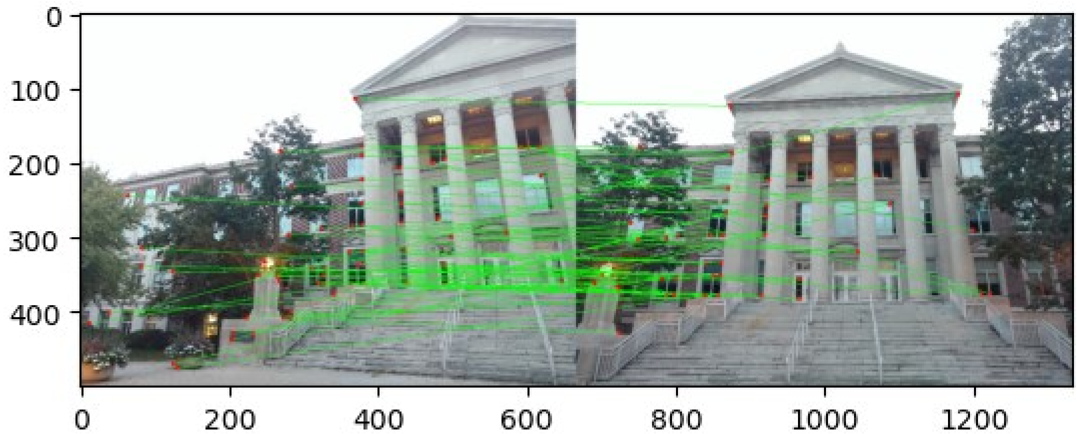
Points:



Correspondences using SSD:

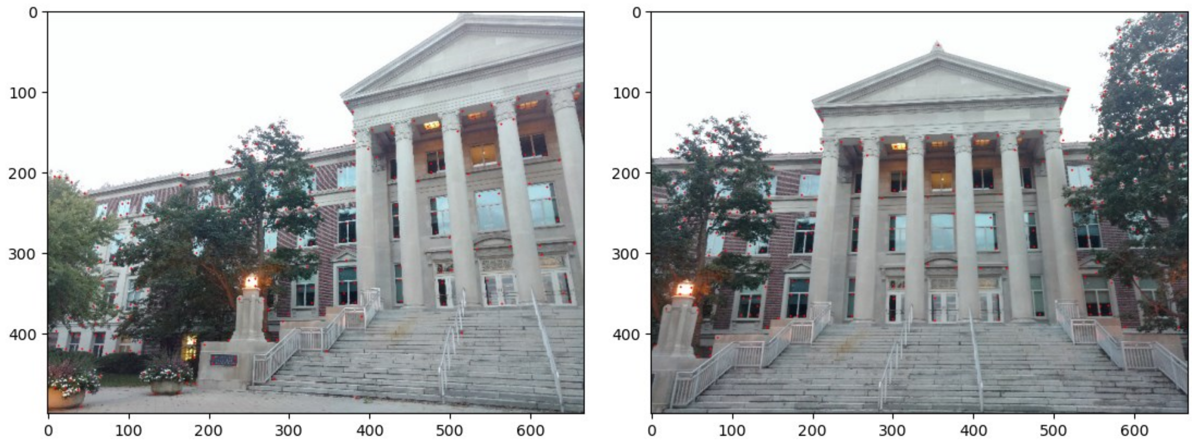


Correspondences using NCC:

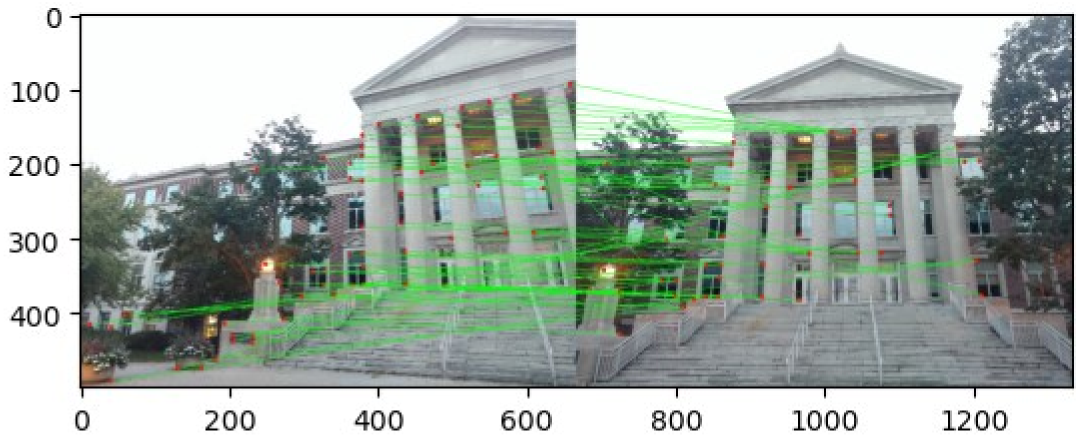


Harris detector for $\sigma = 1.8$

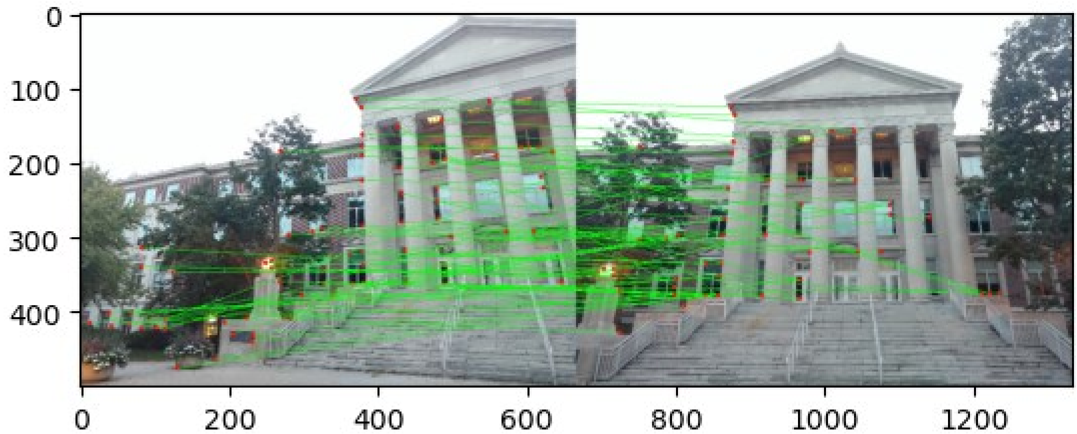
Points:



Correspondences using SSD:

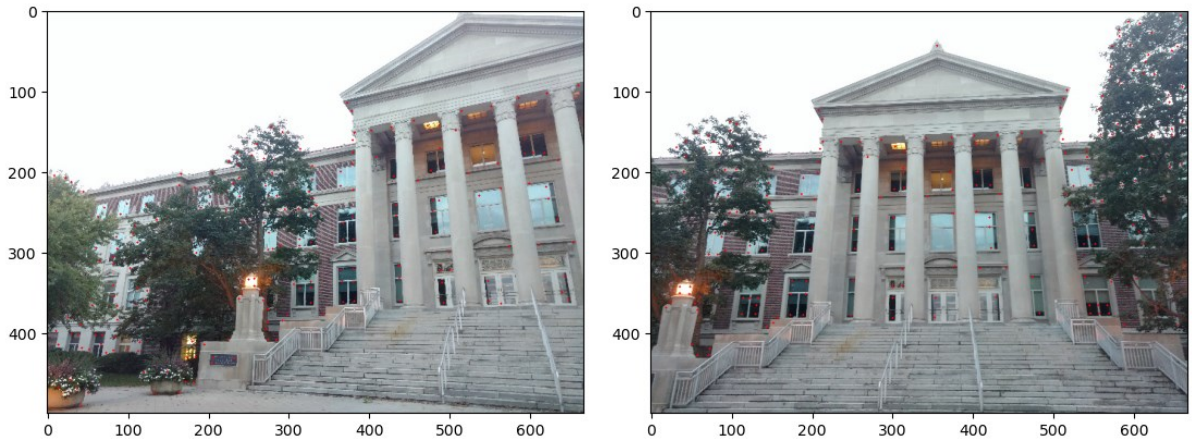


Correspondences using NCC:

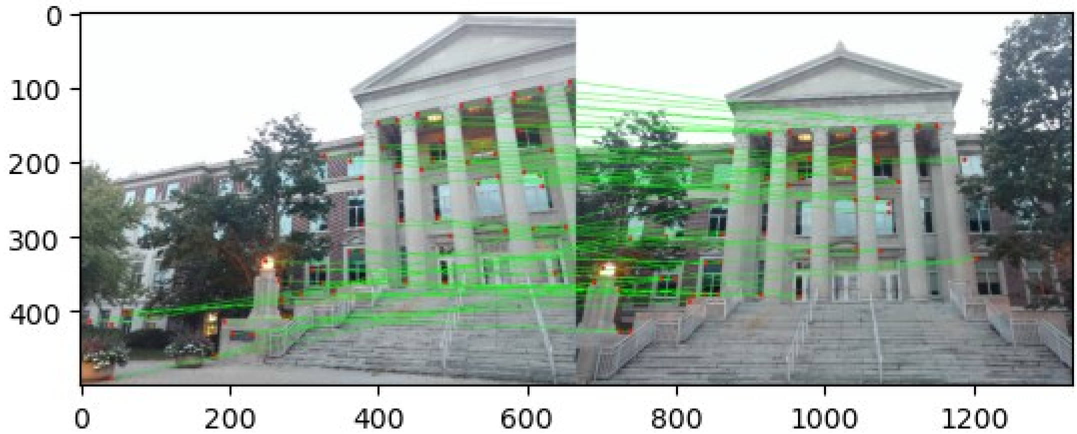


<u>Harris detector for $\sigma = 2.0$</u>

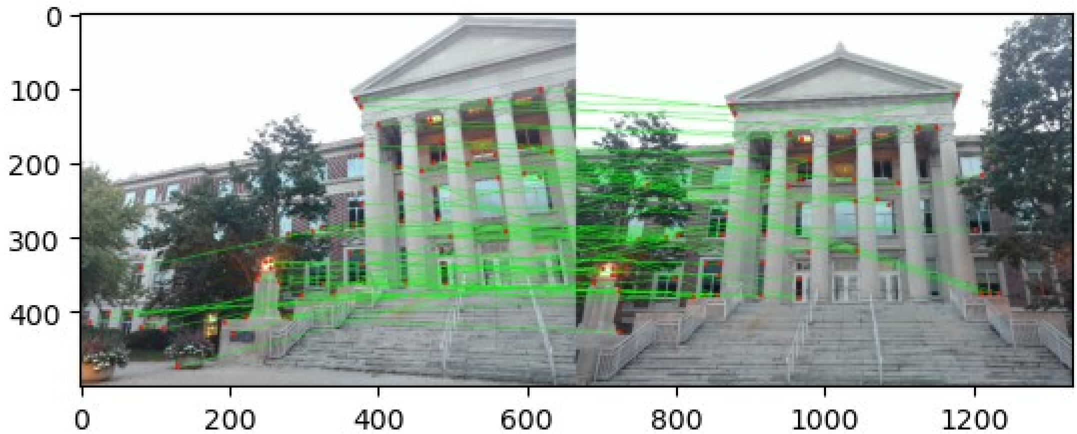
Points:



Correspondences using SSD:



Correspondences using NCC:



4.1.2 Using SIFT

(top p=100 matches):

Image pair 1 output:

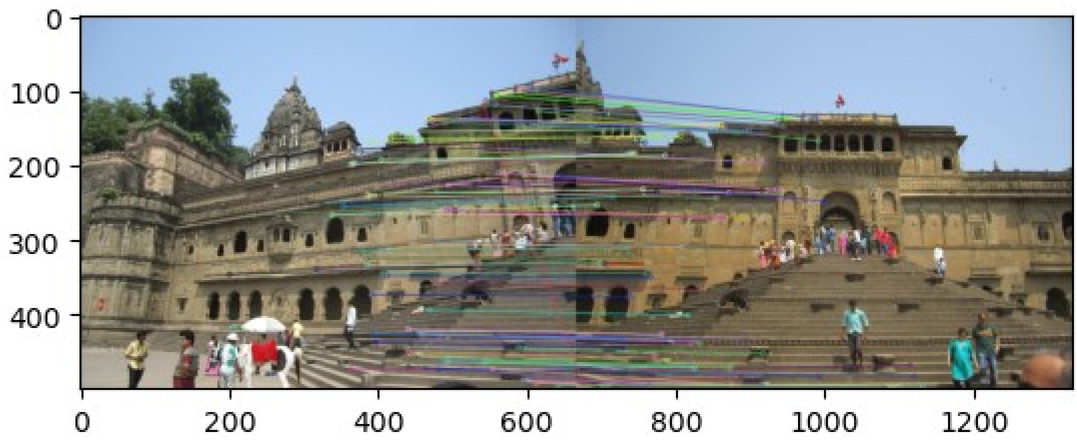
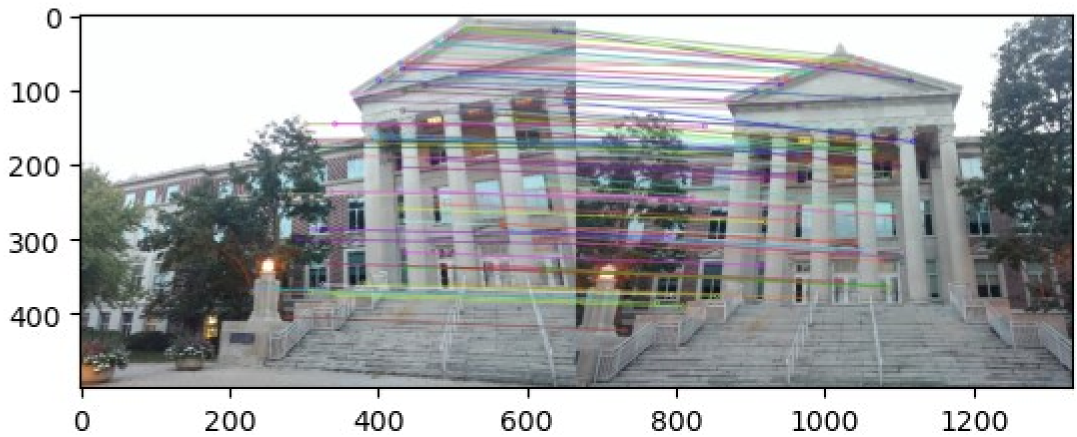


Image pair 2 output:



4.2

Using SuperPoint and SuperGlue

Image Pair 1 output:

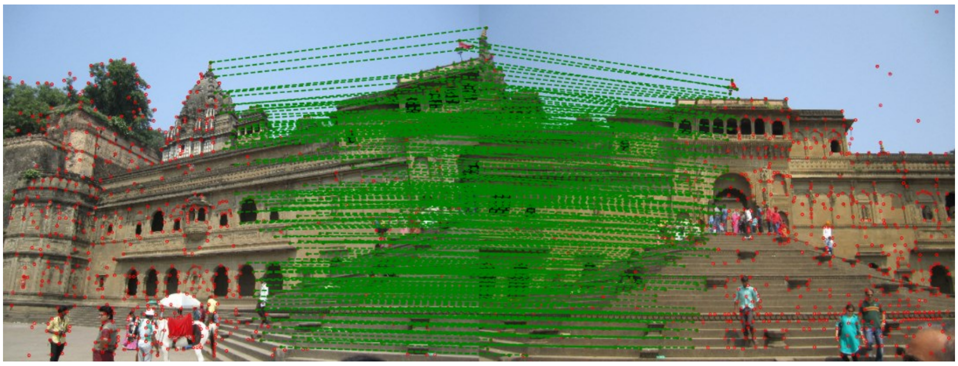
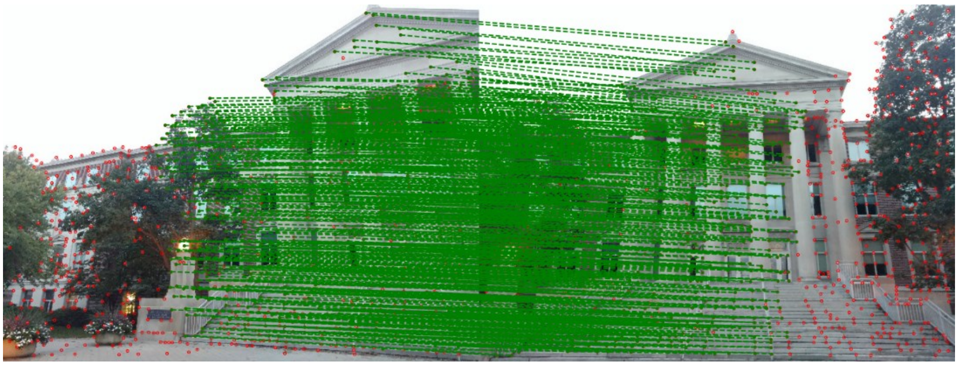


Image pair 2 output:



4.2 Task 2

Using other image pairs.

Image pair 3:

Harris detector for $\sigma = 0.8$

Points:

Correspondences using SSD:

Correspondences using NCC:

<u>Harris detector for $\sigma = 1.2$</u>

Points:

Correspondences using SSD:

Correspondences using NCC:

<u>Harris detector for $\sigma = 1.6$</u>

Points:

Correspondences using SSD:

Correspondences using NCC:

Harris detector for $\sigma = 2.0$

Points:

Correspondences using SSD:

Correspondences using NCC:

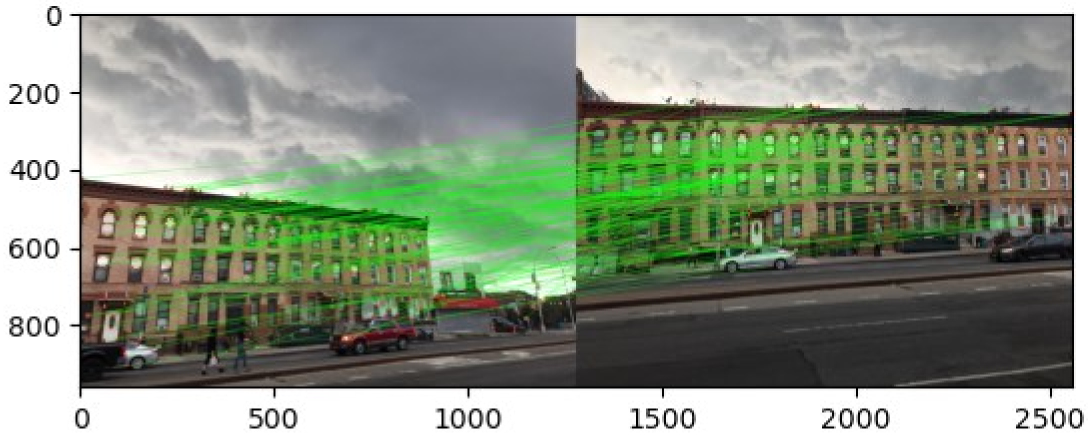
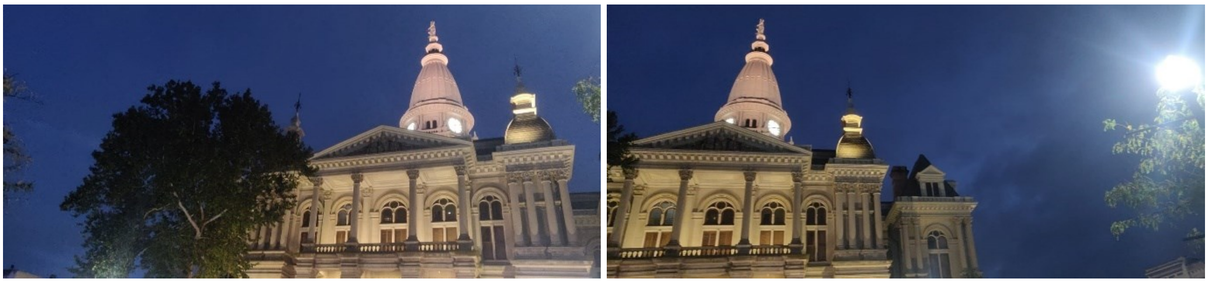
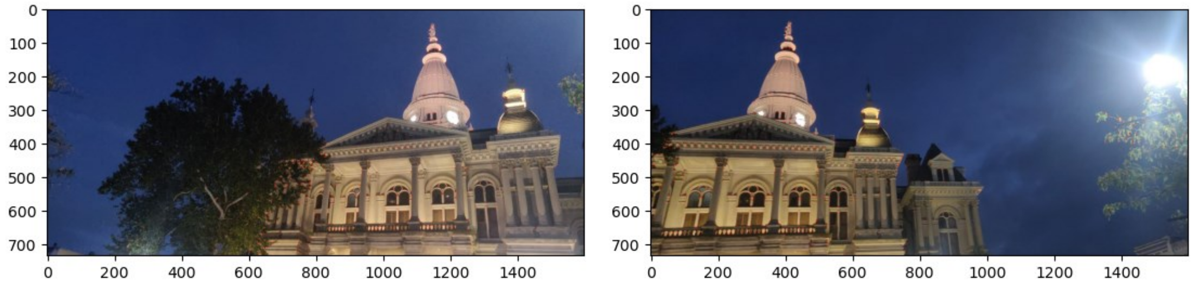


Image pair 4:

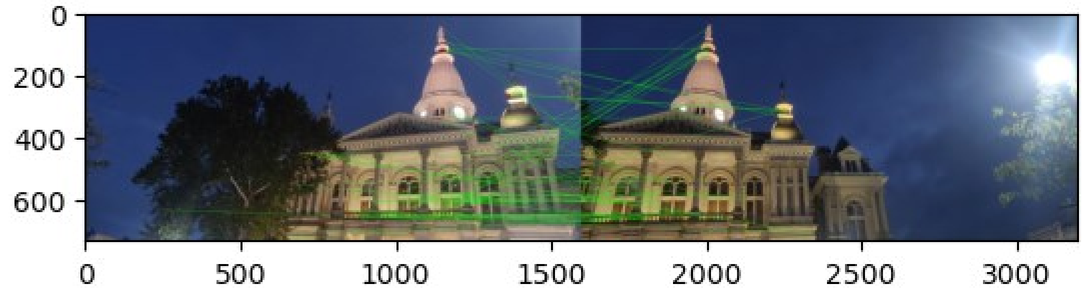


Harris detector for $\sigma = 0.8$

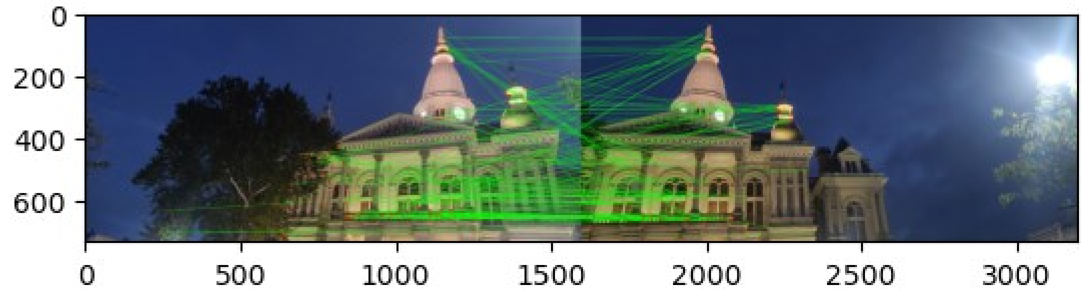
Points:



Correspondences using SSD:

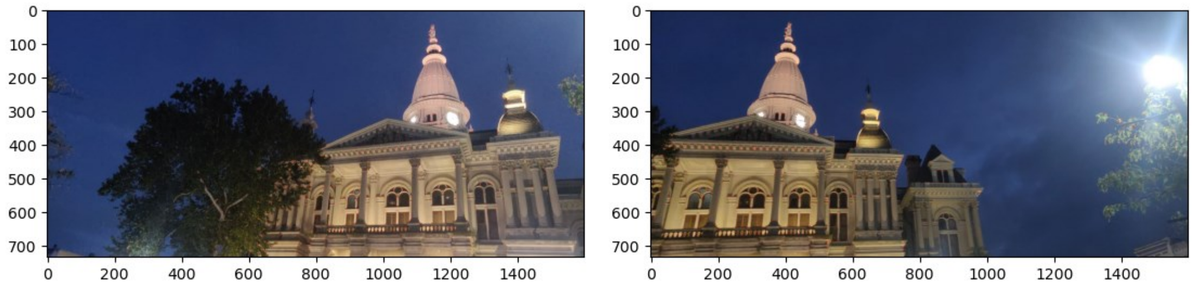


Correspondences using NCC:

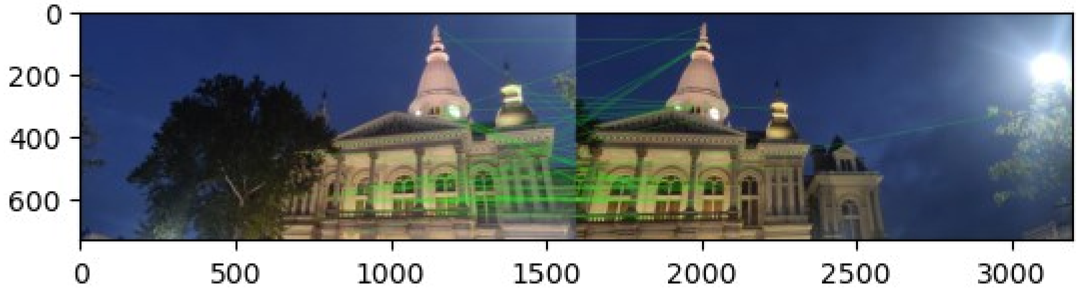


Harris detector for $\sigma = 1.2$

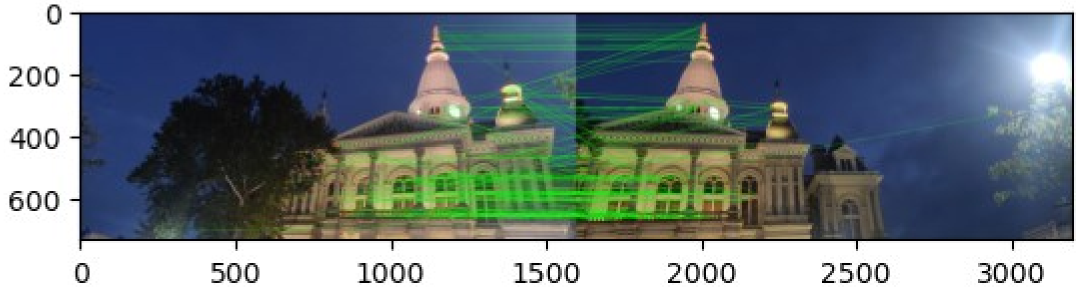
Points:



Correspondences using SSD:

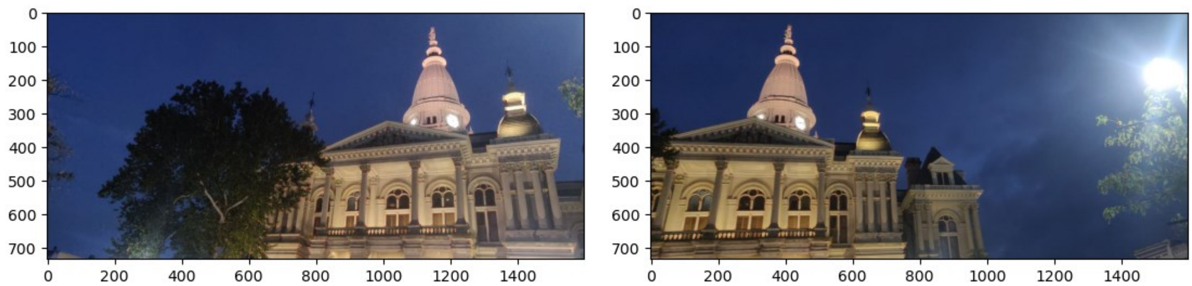


Correspondences using NCC:

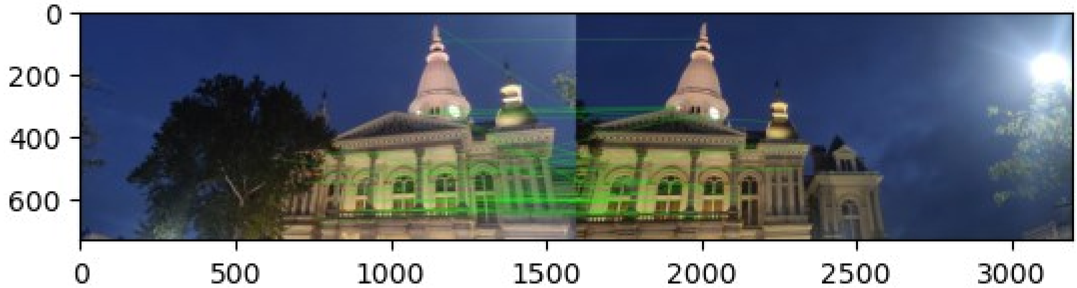


<u>Harris detector for $\sigma = 1.6$</u>

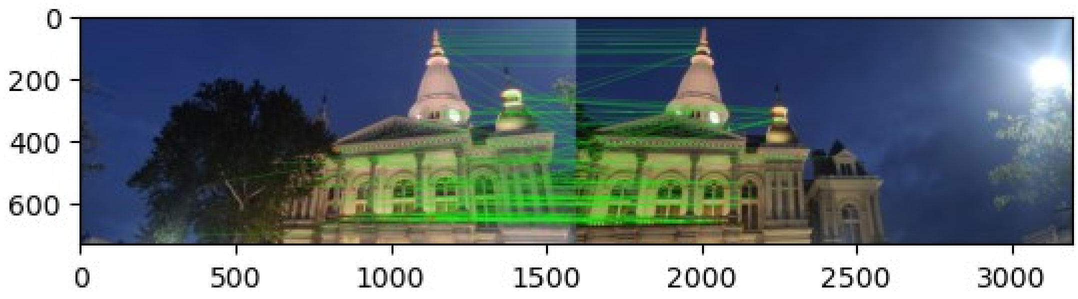
Points:



Correspondences using SSD:

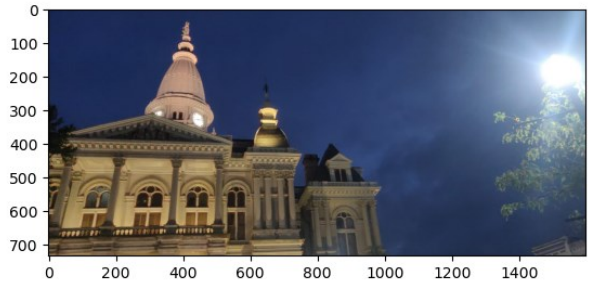


Correspondences using NCC:

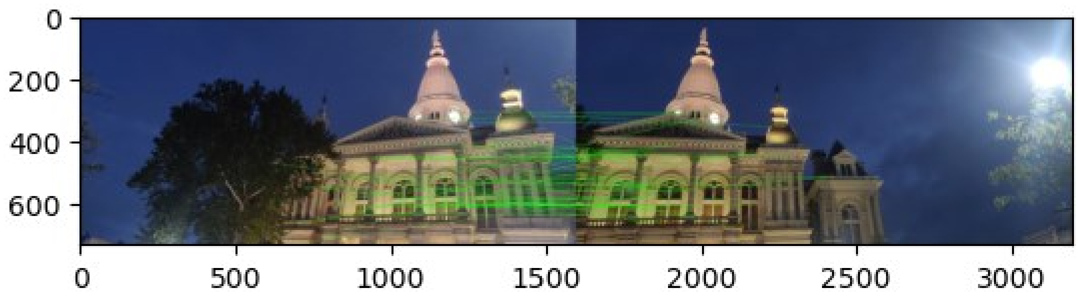


<u>Harris detector for $\sigma = 2.0$</u>

Points:



Correspondences using SSD:



Correspondences using NCC:

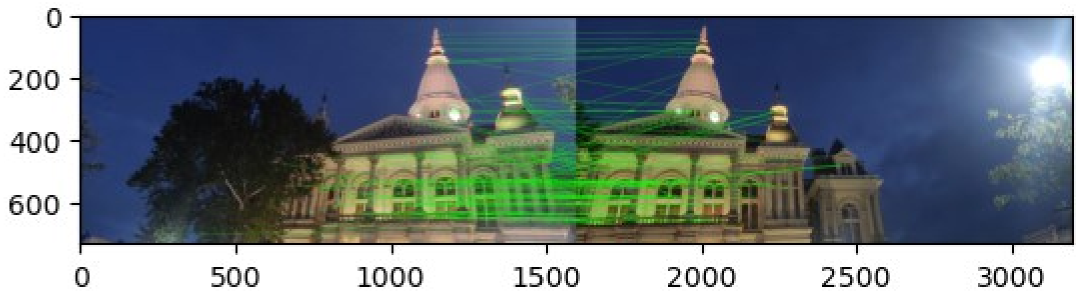
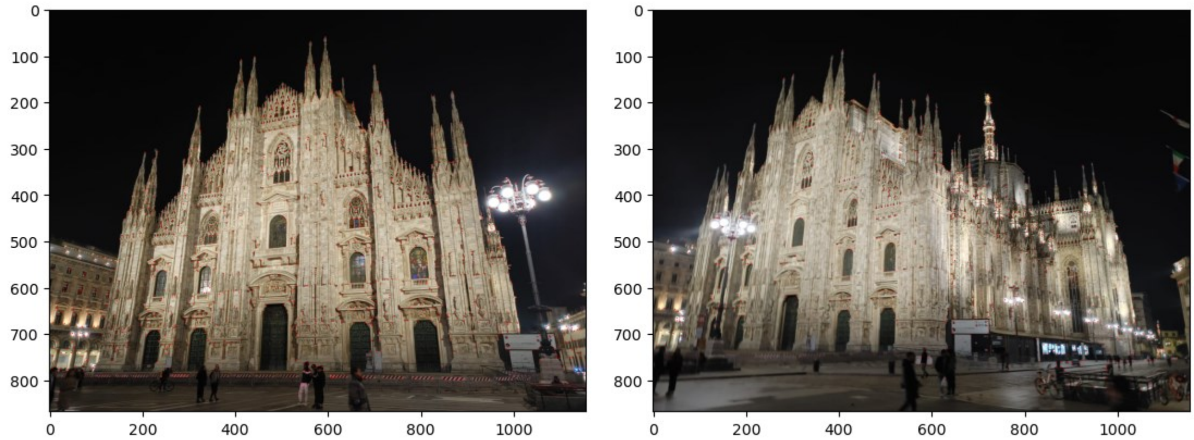


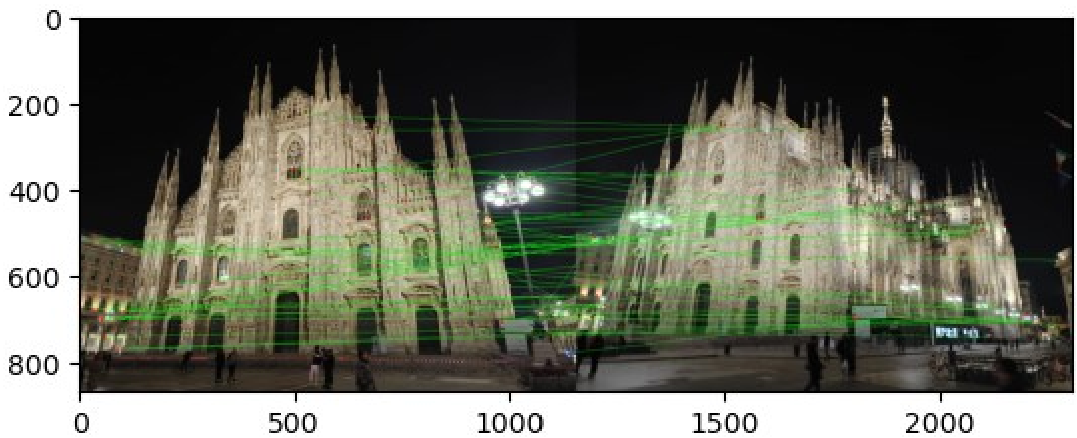
Image pair 5:

<u>Harris detector for $\sigma = 1.2$</u>

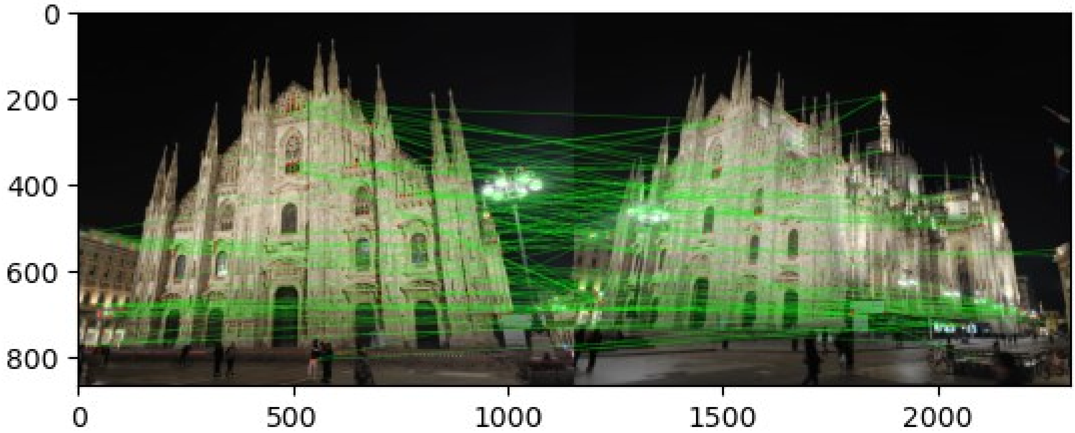
Points:



Correspondences using SSD:



Correspondences using NCC:



Using SIFT

Image pair 3 output:

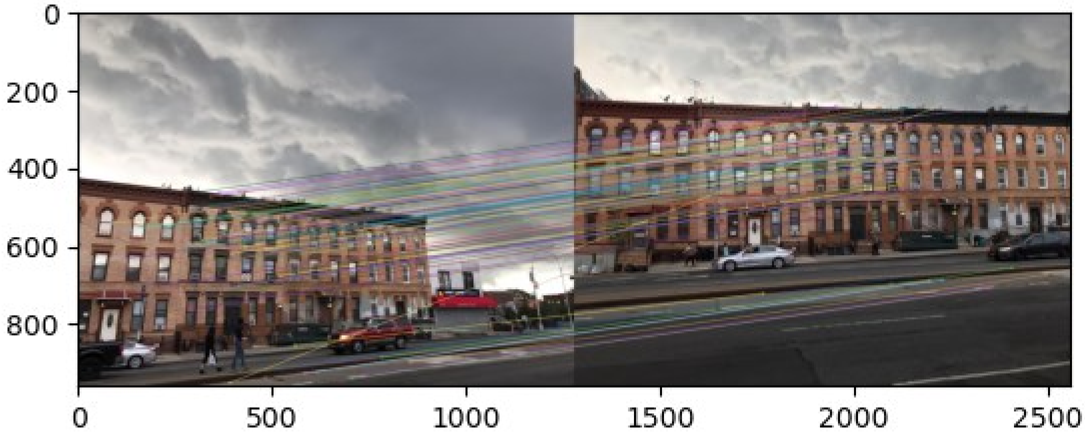


Image pair 4 output:

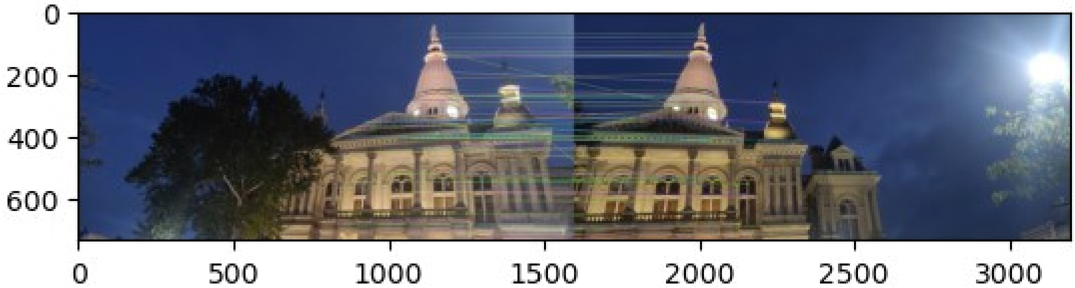
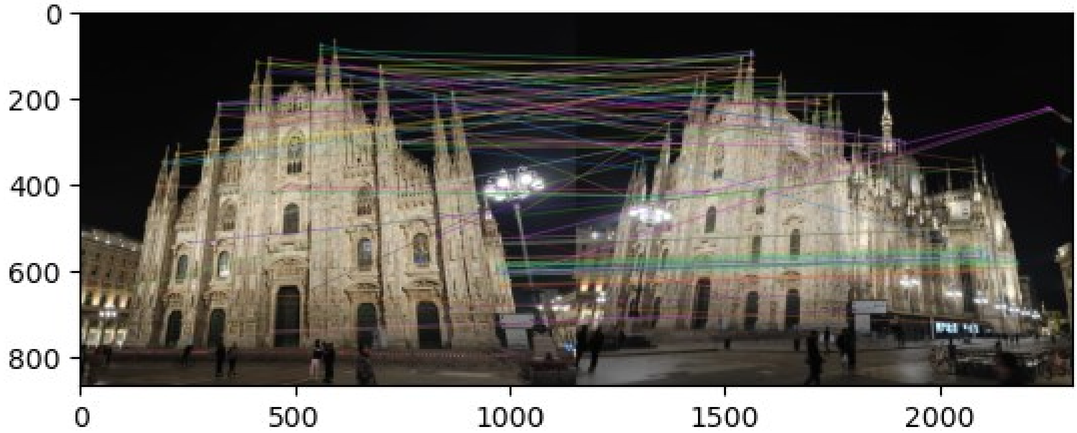


Image pair 5 output:



Using SuperPoint and SuperGlue

Image pair 3 output:

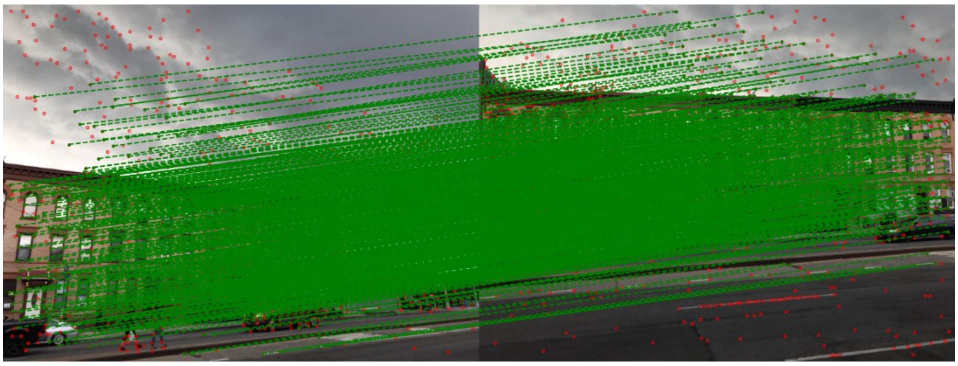


Image pair 4 output:

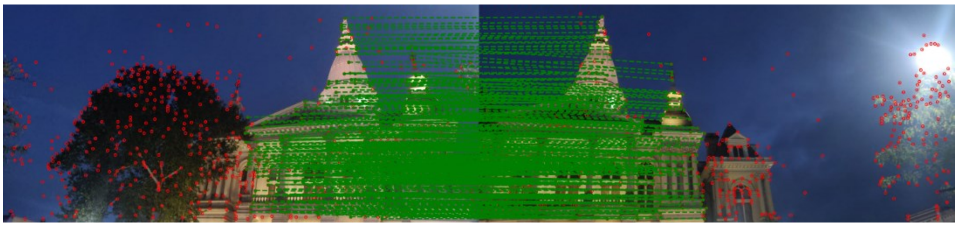
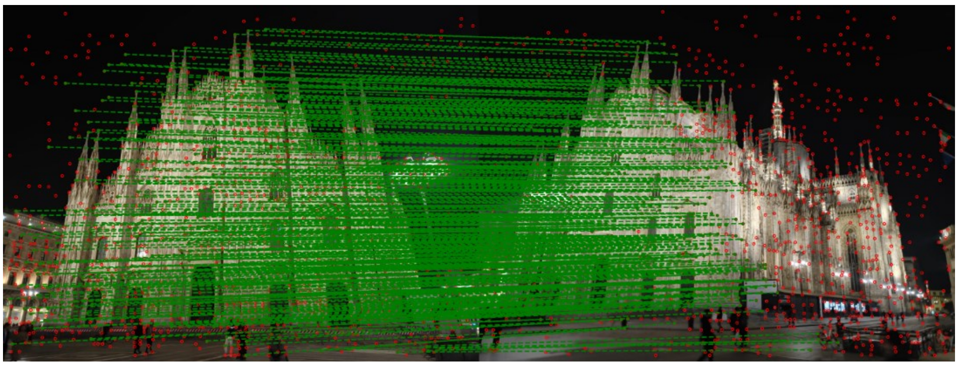


Image pair 5 output:



Observations:

In the Harris corner detection method, choice of σ played an important role in the extraction of interest points. Smaller σ is able to catch finer details leading to more points, whereas a larger value of σ catches more pronounced features thus resulting in a fewer number of points. As a result, when we perform point-to-point correspondences we get better matching results (fewer mismatches) with larger σ value. Comparing the results from the two metrics: SSD and NCC, I found SSD correspondences to be better with fewer spurious matches. SIFT performed better than the Harris detection method and is scale invariant. SIFT detected more points and had better point-to-point correspondences. Although these correspondences still had quite some mismatches. SuperPoint with SuperGlue was the best performer among all with almost perfect correspondences results.