CSC420 A4 Report

Sailing Ni

Apr 2023

Part I: Theoretical Problems

Question 1: RANSAC

$$N = \frac{\log(1-p)}{\log\left(1-(1-e)^s\right)}$$

p = desired probability that we get a good sample
s = number of points in a sample $s = 4.$
N = number of samples (we want to compute this)
e = probability that a point is an outlier

(1) Given visual estimation that at least 70% of the matches are correct, we set inlier ratio p to 0.7 & outlier ratio e to 0.3. Since we want prob of success greater than 99.5%, we set $P = 0.995$. Then, we need at least 4 correspondances for homography, so $S = 4$.

Therefore:

$$N = \frac{\log(1-0.995)}{\log(1-(1-0.3)^4)} \approx 19.29$$

We should round up to 20.

In conclusion, we need at least 20 RANSAC itterations to fit such homography

(2) Fitting an affine transformation would require fewer RANSAC iterations than fitting a homography due to fewer degree of freedom. An affine transformation has 6 degrees of freedom while a homography has 8 d.o.f. Therefore, it's easier to find a good estimate of transformation for affine transformation, which means fewer iterations are needed.

Question 2: Camera Model

(1) We know $\vec{p} = \vec{p_0} + t\vec{d}$ are points on line l and $\vec{P_0} = [x_0, y_0, z_0]^T$

Then $\vec{p} = \begin{pmatrix} wx \\ wy \\ w \end{pmatrix} = K\vec{P} = K \begin{pmatrix} X_0 + td_x \\ Y_0 + td_y \\ Z_0 + td_z \end{pmatrix}$ and $K = \begin{pmatrix} f & 0 & P_x \\ 0 & f & P_y \\ 0 & 0 & 1 \end{pmatrix}$

Where f is the camera focal length and (P_x, P_y) is the principal point

Then $\begin{pmatrix} f & 0 & P_x \\ 0 & f & P_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_0 + td_x \\ Y_0 + td_y \\ Z_0 + td_z \end{pmatrix} = \begin{pmatrix} f(X_0 + td_x) + P_x(Z_0 + td_z) \\ f(Y_0 + td_y) + P_y(Z_0 + td_z) \\ Z_0 + td_z \end{pmatrix}$

$$= (Z_0 + td_z) \begin{pmatrix} P_x + f\dfrac{X_0 + td_x}{Z_0 + td_z} \\ P_y + f\dfrac{Y_0 + td_y}{Z_0 + td_z} \\ 1 \end{pmatrix}$$

To get the coordinates of vanishing point, take $t \to \infty$ as limit

$$\lim_{t\to\infty} \left[P_x + f\frac{X_0 + td_x}{Z_0 + td_z}\right] = \lim_{t\to\infty} \left[P_x + \frac{fX_0 + ftd_x}{Z_0 + td_z}\right]$$

$$= P_x + \frac{fd_x}{d_z}$$

$$\lim_{t\to\infty} \left[P_y + f\frac{Y_0 + td_y}{Z_0 + td_z}\right] = P_y + \frac{fd_y}{d_z}$$

Therefore, the pixel coordinates of the vanishing point is

$$\begin{bmatrix} P_x + f\dfrac{d_x}{d_z} \\ P_y + f\dfrac{d_y}{d_z} \end{bmatrix} \quad \text{if } d_z \neq 0.$$

If $d_z = 0$, then there's no vanishing point since the line is parallel to the projection plane.

(2)

Since all lines on the plane are perpendicular to the normal vector $\vec{n}$

$$\vec{n}\cdot\vec{d} = 0 \Rightarrow \quad n_x dx + n_y dy + n_z dz = 0$$

$$n_x \frac{dx}{dz} + n_y \frac{dy}{dz} + n_z = 0$$

$$\frac{dx}{dz} = -\frac{n_z}{n_x} - \frac{n_y}{n_x}\frac{dy}{dz}$$

In part (1), we proved the vanishing point of a line is $\left[P_x + f\frac{dx}{dz},\ P_y + f\frac{dy}{dz}\right]^T$

denote the vanishing point as (V_x, V_y)

$$V_x = P_x + f\frac{dx}{dz} = P_x + f\left(-\frac{n_z}{n_x} - \frac{n_y}{n_x}\frac{dy}{dz}\right)$$

$$= P_x - f\frac{n_z}{n_x} - f\frac{n_y}{n_x}\frac{dy}{dz} - P_y\frac{n_y}{n_x} + P_y\frac{n_y}{n_x}$$

$$= P_x - \frac{n_y}{n_x}\left(P_y + f\frac{dy}{dz}\right) + P_y\frac{n_y}{n_x} - f\frac{n_z}{n_x}$$

$$= P_x - \frac{n_y}{n_x}V_y + P_y\frac{n_y}{n_x} - f\frac{n_z}{n_x}$$

$$V_x = P_x - \frac{n_y}{n_x}V_y + P_y\frac{n_y}{n_x} - f\frac{n_z}{n_x}$$

$$n_x V_x + n_y V_y = n_x P_x + n_y P_y - f n_z$$

Express it as $n_x V_x + n_y V_y - (n_x P_x + n_y P_y - f n_z) = 0$

which is a equation of a line.

Therefore, the vanishing points of all the lines lying on the plane will form a projection line.

Question 3: Homogenous Coordinates

Using the homogeneous coordinates:

1. **(15 marks)** (a) Show that the intersection of the 2D line l and l' is the 2D point $p = l \times l'$.
   (here $\times$ denotes the cross product)

Let the two lines l, l' be:

$$l: \ a_1 x + b_1 y + c_1 = 0$$

$$l': \ a_2 x + b_2 y + c_2 = 0$$

Solve for the intersection point of l, l'

$$x = \frac{b_1 c_2 - b_2 c_1}{a_1 b_2 - a_2 b_1}, \quad y = \frac{a_2 c_1 - a_1 c_2}{a_1 b_2 - a_2 b_1}$$

In homogeneous coordinates, scaling doesn't affect anything. $\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} w \cdot x \\ w \cdot y \\ w \end{bmatrix}$

take $w = a_1 b_2 - a_2 b_1$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} w \cdot x \\ w \cdot y \\ w \end{bmatrix} = \begin{bmatrix} b_1 c_2 - b_2 c_1 \\ a_2 c_1 - a_1 c_2 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

Take cross product.

$$l \times l' = [a_1, b_1, c_1]^T \times [a_2, b_2, c_2]^T = \begin{vmatrix} i & j & k \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix}$$

$$= \begin{bmatrix} b_1 c_2 - c_1 b_2 \\ a_2 c_1 - a_1 c_2 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}$$

$$= \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Therefore, we have shown the intersection of 2D lines l and l' is exactly the 2D point $p = l \times l'$

2. **(15 marks)** (b) Show that the line that goes through the 2D points p and p' is $l = p \times p'$.

Let p, p' be two arbitrary points s.t. $p = (x_1, y_1)^T$, $p' = (x_2, y_2)^T$

Let l be the line $ax + by + c = 0$ that goes thru p, p'

Then it satisfies:
$$\begin{cases} ax_1 + by_1 + c = 0 \\ ax_2 + by_2 + c = 0 \end{cases}$$

Solve for the system:
$$a = \frac{c(y_2 - y_1)}{x_2 y_1 - x_1 y_2}$$

$$b = \frac{c(x_1 - x_2)}{x_2 y_1 - x_1 y_2}$$

Use homogeneous coordinates, let $w = \frac{x_2 y_1 - x_1 y_2}{c}$

express the line vector:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \frac{c(y_2 - y_1)}{x_2 y_1 - x_1 y_2} \\ \frac{c(x_1 - x_2)}{x_2 y_1 - x_1 y_2} \\ c \end{pmatrix} = \begin{pmatrix} y_2 - y_1 \\ x_1 - x_2 \\ x_2 y_1 - x_1 y_2 \end{pmatrix}$$

Take cross product:

$$p \times p' = (x_1, y_1)^T \times (x_2, y_2)^T = \begin{vmatrix} i, & j, & k \\ x_1, & y_1, & 1 \\ x_2, & y_2, & 1 \end{vmatrix}$$
$$= \begin{pmatrix} y_1 - y_2 \\ x_2 - x_1 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}$$

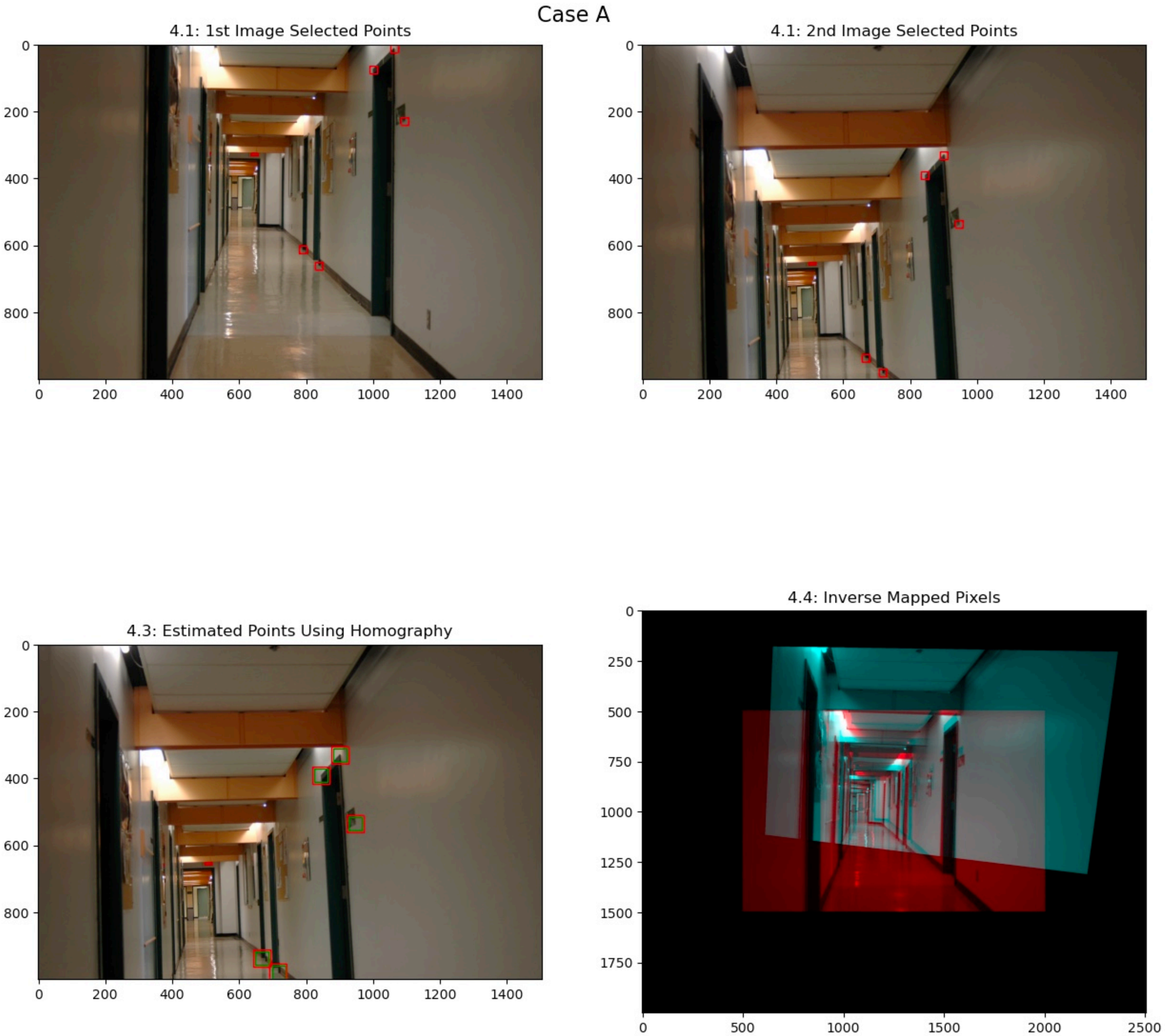
which is homogenous coordinates with $(a, b, c)^T$

Therefore, the line l that goes thru p, p' is $l = p \times p'$

Part II: Implementation Tasks

Question 4: Homography

Case A:

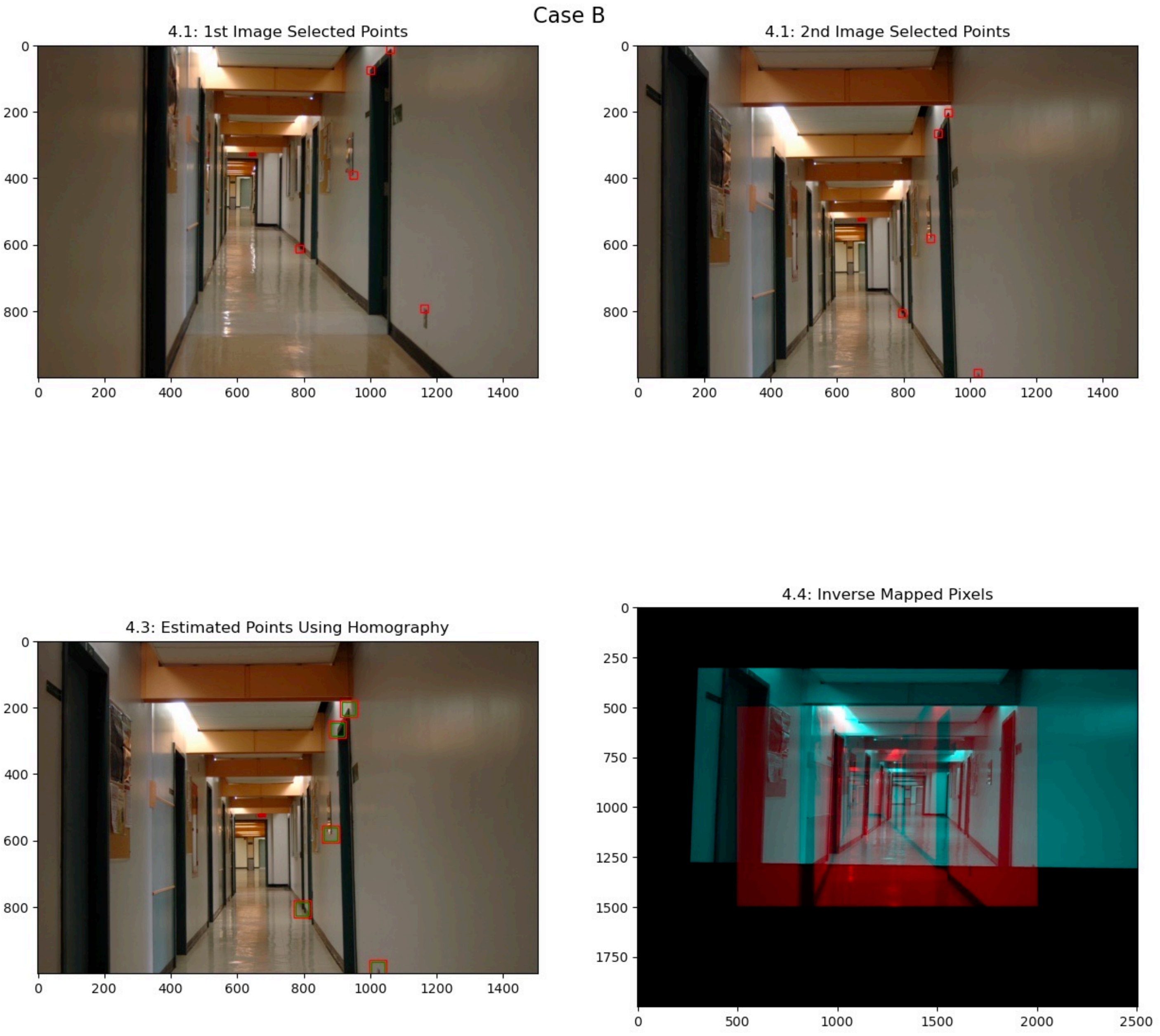


4.2: Estimated H:

array([[1.18303955e+00, 1.93511977e-02, -2.38565321e+02],
       [4.55095974e-02, 1.05588002e+00, 3.11969371e+02],
       [1.23191781e-04, -5.50027735e-05, 1.00000000e+00]])

Comment:
- Translation: the translation vector is approximately (-238.57, 311.97), indicating that the transformed image has been shifted to the left and up relative to the original image.
- Scale: the scaling factors are 1.18 and 1.055 respectively for x and y directions, meaning the image is slightly up-scaled.
- Shear: very small figures that can be ignored
- Rotation: very small figures.

Case B:



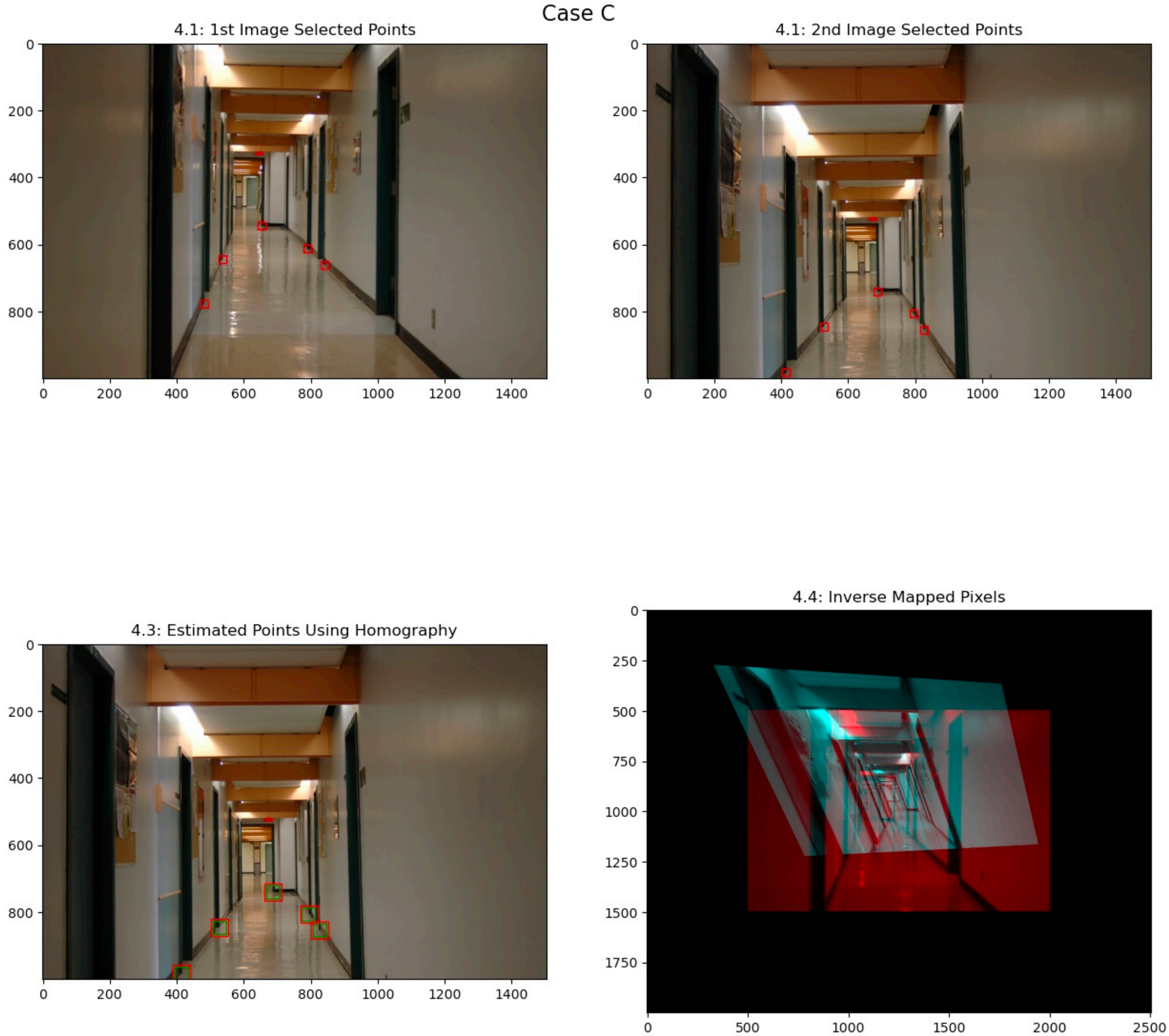
4.2: Estimated H:

```
array([[ 5.97617390e-01,  1.23336140e-02,  3.11952364e+02],
       [ 1.33765930e-03,  9.91258403e-01,  1.92806221e+02],
       [ 1.05551705e-05, -2.26071996e-05,  1.00000000e+00]])
```

Comment:

- Translation: the translation vector is approximately (311.95, 192.81), indicating that the transformed image has been shifted to right and up.
- Scale: the scaling factors are 0.598 and 0.991 respectively for x and y directions, meaning the image is down-sampled.
- Shear: shearing factors (0.0123, 0.0013) means the figure is slightly sheared in both directions.
- Rotation: very small figures that can be ignored

Case C:



4.2: Estimated H:

```
array([[ 7.87370586e-01, -4.58530493e-01,  3.33140808e+02],
       [-1.07243197e-01,  7.87024972e-01,  2.85354298e+02],
       [-1.04032401e-04, -1.14696412e-04,  1.00000000e+00]])
```

Comment:

- Translation: the translation vector is approximately (333.14, 285.35), indicating that the transformed image has been shifted to the right and up relative to the original image.
- Scale: the image has been downsampled by 0.787 in each direction.
- Rotation: transformed image has been rotated counterclockwise
- Shear: -0.4585 and -0.1072 represent shear to the left and down

Overall:

4.3: Estimated Points

- Green squares are the estimated points using homography; red squares are those selected in the second image
- In all cases, all estimated points overlap with those selected.

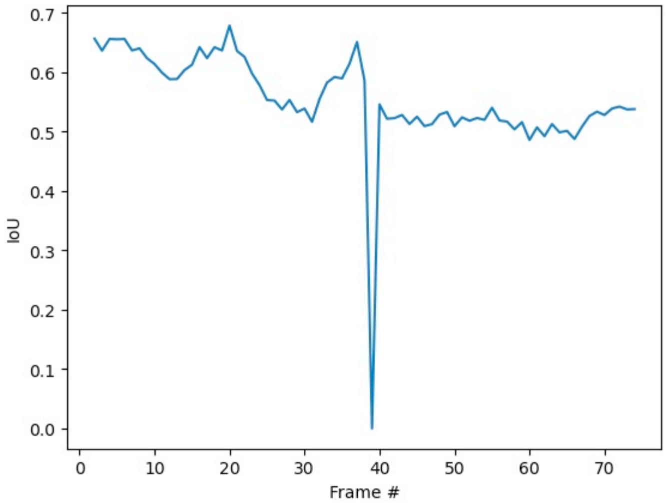
4.4: Inverse Mapped Pixels

- Homography contains perspective transformations consist of scale, translation, shear, and rotation
- Camera position is about the same in image1 and image2, with camera orientation up in image2.
- Camera positions are different in image1 and image3, also camera orientation is slightly counter-clockwisely rotated.
- The floor is of less Lambertian reflectance while the right wall has more Lambertian reflectance. As we can observe from the three images, the floor has a "mirror-like" reflections while the right wall is more of matte texture.

Question 5: Mean Shift Tracking

5.1: Performance Evaluation

- **Plot of IoU over time**:



- **Sample Frames:** Highest IoU vs Lowest IoU

- *Red box*: mean shift tracked box.

- *Green box*: Viola-Jones detected box

- Notice the lowest IoU is 0% where the green box is missing. This was when Viola-Jones detection misses faces that are fully visible.

- **Percentage of frames in which IoU is larger than 50%:** 93.24%

- **Below is Analysis of low IoU frames:**
  - Inspecting the histogram, we notice only 1 IoU is lower than 10%, which is 0% indeed. To effectively analyze the lower IoUs, we will pick 50% as the threshold.
  - There are 4 frames with IoU under 50%.

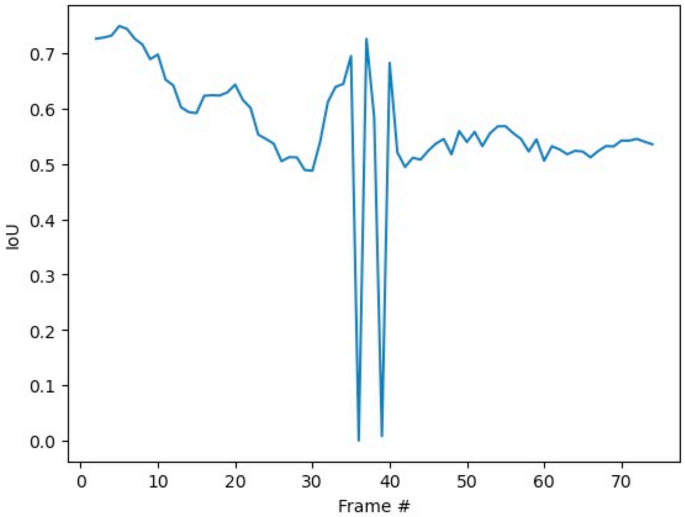
- **Comment:**

- Comparing the four pairs of boxes, we can see green box of Viola-Jones detection is more accurate in facial detection than mean shift tracking.

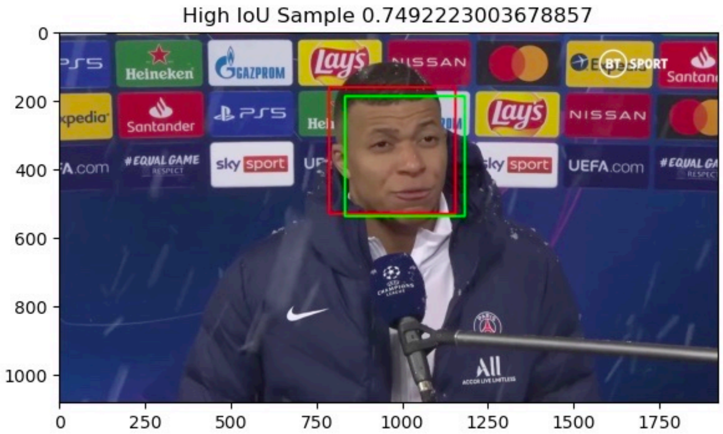
- Viola-Jones is a trained machine learning approach to detect faces, while mean shift tracking relies on appearance and motion cues to track objects (we used hue histogram here). Mean shift tracking is more sensitive to hue changes, and above red squares are examples where hair is detected as part of face.

5.2: Implement a Simple Variation

- **Plot of IoU over time**:



- **Sample Frames:** High IoU > 70% vs Low IoU < 50%



- **Percentage of frames in which IoU is larger than 50%:** 93.24%
- **Percentage of frames in which IoU is larger than 70%:** 12.16%

- **Comment**:
  - Compare to the standard mean shift tracking, the modified version performs slightly better.
  - Considering IoU over 70%, the standard mean shift tracking's highest IoU is 67.9%, while the modified version has 12% of frames with IoU
  - Since pixels around hair/forehead in the video are fairly consistent throughout the image, gradient direction is the same. Using a gradient angle instead of hue can partly overcome the sensitivity to hue in standard mean shift tracking.