1 Homographies

Q1.1 Homography

$\mathbf{P}_1,\ \mathbf{P}_2 \in \mathbb{R}^{3\times 4}$ are camera projection matrices, transforming from 3D to 2D, since the additional dimension is for the homogeneous coordinates. Let $\mathbf{X} \in \mathbb{R}^3$ be the focus point in the plane Π, such that $\mathbf{x}_1 = \alpha \mathbf{P}_1 \mathbf{X}$ and $\mathbf{x}_2 = \beta \mathbf{P}_2 \mathbf{X}$, for scalar scales $\alpha, \beta \in \mathbb{R}$. We have $\mathbf{x}_2 = \beta \mathbf{P}_2 \mathbf{X} \implies \dfrac{1}{\beta}\mathbf{P}_2^{+}\mathbf{x}_2 = \mathbf{X}$, where A^+ is the (left) pseudo-inverse of matrix A, which is defined for $A = \mathbf{P}_2$ since $\mathbf{P}_2 \in \mathbb{R}^{4\times 3}$ has linearly independent components formed by the basis for the projection matrix.

Combining, $\mathbf{x}_1 \equiv \underbrace{\mathbf{P}_1\mathbf{P}_2^{+}}_{\mathbf{H}}\mathbf{x}_2$, up to some scale $(\dfrac{\alpha}{\beta})$, where homography $\mathbf{H} \in \mathbb{R}^{3\times 3}$ exists.

Q1.2 Correspondences

1. $\mathbf{H}$ represents a projective transform, which is an affine transform (having a degree of freedom of 6) plus 2 more possible warpings after the affine transform in x- and y-directions, so it has 8 degrees of freedom; hence $\mathbf{h}$ being reshaped from $\mathbf{H}$ share the 8 degrees of freedom.

2. Each point pair supplies two equations, when there are 8 unknowns from $\mathbf{h}$, so it needs 4 equations, that is 4 point pairs to work.

3.

$$
\begin{aligned}
\mathbf{x}_1^i \equiv \mathbf{H}\mathbf{x}_2^i &\overset{\text{def}}{\iff} \begin{bmatrix} x_1^i \\ y_1^i \\ 1 \end{bmatrix} = \alpha \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x_2^i \\ y_2^i \\ 1 \end{bmatrix} \\
&\iff \begin{cases} x_1^i = \alpha(h_1 x_2^i + h_2 y_2^i + h_3) \\ y_1^i = \alpha(h_4 x_2^i + h_5 y_2^i + h_6) \\ 1 = \alpha(h_7 x_2^i + h_8 y_2^i + h_9) \end{cases} \\
&\iff \begin{cases} x_1^i(h_7 x_2^i + h_8 y_2^i + h_9) = (h_1 x_2^i + h_2 y_2^i + h_3) \\ y_1^i(h_7 x_2^i + h_8 y_2^i + h_9) = (h_4 x_2^i + h_5 y_2^i + h_6) \end{cases} \\
&\iff \begin{cases} h_7 x_1^i x_2^i + h_8 x_1^i y_2^i + h_9 x_1^i - h_1 x_2^i - h_2 y_2^i - h_3 = 0 \\ h_7 x_2^i y_1^i + h_8 y_1^i y_2^i + h_9 y_1^i - h_4 x_2^i - h_5 y_2^i - h_6 = 0 \end{cases} \\
&\iff \underbrace{\begin{bmatrix} -x_2^i & -y_2^i & -1 & 0 & 0 & 0 & x_1^i x_2^i & x_1^i y_2^i & x_1^i \\ 0 & 0 & 0 & -x_2^i & -y_2^i & -1 & x_2^i y_1^i & y_1^i y_2^i & y_1^i \end{bmatrix}}_{\mathbf{A}_i} \mathbf{h} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
&\iff \mathbf{A}_i \mathbf{h} = \mathbf{0}_{2\times 1}
\end{aligned}
$$

4. $\mathbf{h} = \mathbf{0}_{9\times 1}$ would be a trivial solution.

   $\mathbf{A} \in \mathbb{R}^{8\times 9}$ which stacks $\mathbf{A}_i, i \in \{1,2,3,4\}$ together, is full rank, because it contains 8 independent equations (corresponding to x- and y- coordinates of 4 independent data point pairs), $\textit{rank } \mathbf{A} = 8 = \min(8,9)$.

   This implies that $\textit{rank } \mathbf{A}^\top\mathbf{A} = \textit{rank } \mathbf{A} = 8$. However, $\mathbf{A}^\top\mathbf{A} \in \mathbb{R}^{9\times 9}$ so $\textit{rank } \mathbf{A}^\top\mathbf{A} < 9$, it is not full rank, and the null space has a dimension of 1. There is exactly one zero eigenvalue when we eigendecompose $\mathbf{A}^\top\mathbf{A}$, so there is one corresponding eigenvector for the zero eigenvalue, this eigenvalue solves $\mathbf{h}$ as a non-trivial solution to $\mathbf{Ah} = \mathbf{0}$.

Q1.3 Homography under rotation

Given $\mathbf{x}_1 = \mathbf{K}_1 \left[\mathbf{I} \mid \mathbf{0}\right] \mathbf{X}$ and $\mathbf{x}_2 = \mathbf{K}_2 \left[\mathbf{R} \mid \mathbf{0}\right] \mathbf{X}$, we have

$$\begin{cases} \mathbf{K}_1^{-1}\mathbf{x}_1 &= \left[\mathbf{I} \mid \mathbf{0}\right] \mathbf{X} \\ \mathbf{K}_2^{-1}\mathbf{x}_2 &= \mathbf{R}\left[\mathbf{I} \mid \mathbf{0}\right] \mathbf{X} \end{cases} \implies \mathbf{R}\mathbf{K}_1^{-1}\mathbf{x}_1 = \mathbf{K}_2^{-1}\mathbf{x}_2 \implies \mathbf{x}_1 = \underbrace{\mathbf{K}_1\mathbf{R}^{-1}\mathbf{K}_2^{-1}}_{\mathbf{H}}\mathbf{x}_2,$$

where $\mathbf{H}$ always exists because $\mathbf{K}_1, \mathbf{K}_2$ are upper triangular matrices with nonzero diagonals and $\mathbf{R}$ is a rotation matrix, which both are always invertible.

Q1.4 Understanding homographies under rotation

Define $\mathbf{x}_0$ be the camera view initially of $\mathbf{X}$, $\mathbf{x}_\theta$ be the camera view of $\mathbf{X}$ after rotating θ and $\mathbf{x}_{2\theta}$ be the camera view of $\mathbf{X}$ after rotating 2θ. From Q1.3, we get the following equations.

$$\mathbf{x}_0 = \underbrace{\mathbf{K}\mathbf{R}_\theta^{-1}\mathbf{K}^{-1}}_{\mathbf{H}}\mathbf{x}_\theta, \qquad \mathbf{x}_\theta = \mathbf{K}\mathbf{R}_\theta^{-1}\mathbf{K}^{-1}\mathbf{x}_{2\theta}$$

Combining both equations, we get

$$\begin{aligned} \mathbf{x}_0 &= \underbrace{\mathbf{K}\mathbf{R}_\theta^{-1}\mathbf{K}^{-1}\mathbf{K}\mathbf{R}_\theta^{-1}\mathbf{K}^{-1}}_{\mathbf{H}^2}\mathbf{x}_{2\theta} \\ &= \mathbf{K}(\mathbf{R}_\theta^2)^{-1}\mathbf{K}^{-1}\mathbf{x}_{2\theta} \\ &= \mathbf{K}\mathbf{R}_{2\theta}^{-1}\mathbf{K}^{-1}\mathbf{x}_{2\theta}, \end{aligned}$$

which shows $\mathbf{H}^2$ corresponds to a rotation of 2θ.

Q1.5 Limitation of the planar homography

A homography is a transformation between two projective planes. If the scene image cannot be represented (approximately) by a plane, homography would not work.

Q1.6 Behavior of lines under perspective projections

Suppose we have $\mathbf{x}_0 = \mathbf{P}\mathbf{X}_0$ for some initial points $(\mathbf{x}_0, \mathbf{X}_0)$. Consider the line $\mathbf{X}_0 + \lambda\mathbf{T}$ in 3D, where $\lambda > 0$ and $\mathbf{T}$ is some constant vector in 3D, we then have

$$\mathbf{P}(\mathbf{X}_0 + \lambda\mathbf{T}) = \mathbf{x}_0 + \lambda \underbrace{\mathbf{P}\mathbf{T}}_{\text{constant}},$$

which represents a line in 2D.

2 Computing Planar Homographies

Q2.1.1 FAST Detector

FAST detector is similar to Harris corner detector in the sense that they both are using corners as keypoints. However, the corner detection is FAST is less rigorous, and is susceptible to noise and false-positives because it is only using a very simple test. However, because of the simple test in FAST, it is several times faster than Harris corner detector.

Q2.1.2 BRIEF Descriptor

BRIEF compares the intensity values of random pair of points, that the descriptor is binary, while the filter bank method considers the responses of the image after going through the filters. It is possible to use the responses of the images to a particular filter as a descriptor.

Q2.1.3 Matching Methods

We can match a point p in one image with points in the other images using nearest neighbour search of p in Hamming space. Using Hamming distance provides a more intuitive meaning because what it is measuring is basically the number of pairs of points in the image with different ordering of intensities. Using Euclidean distance, however, does not provide any intuitive meanings.

Q2.1.4 Feature Matching

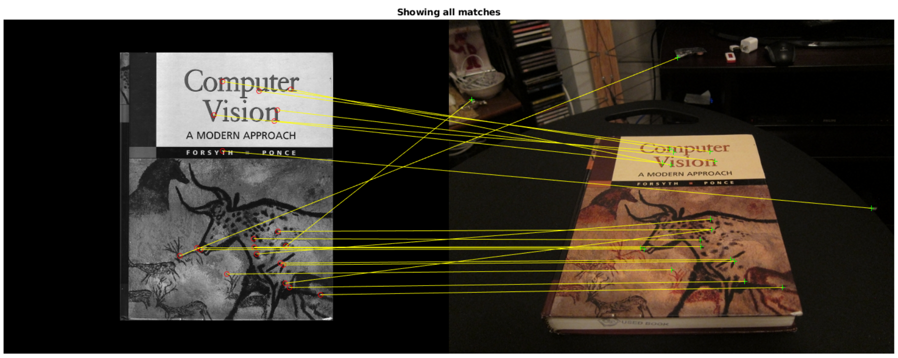

Figure 1: Matched FAST corners using BRIEF

Q2.1.5 BRIEF and Rotations

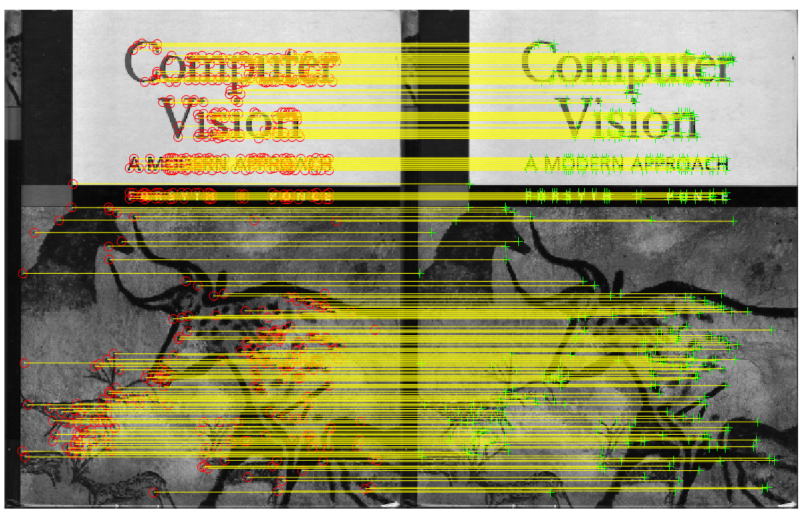

Figure 2: Matched FAST corners using BRIEF at $\theta = 0$.

Figure 3: Matched FAST corners using BRIEF at $\theta = 2\pi/3$.

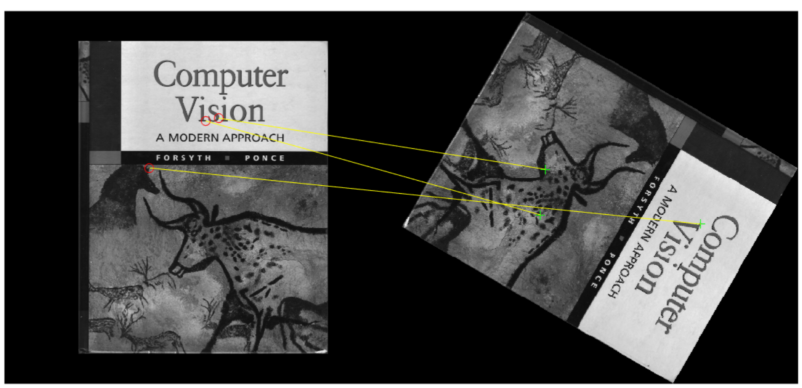

Figure 4: Matched FAST corners using BRIEF at $\theta = 4\pi/3$.

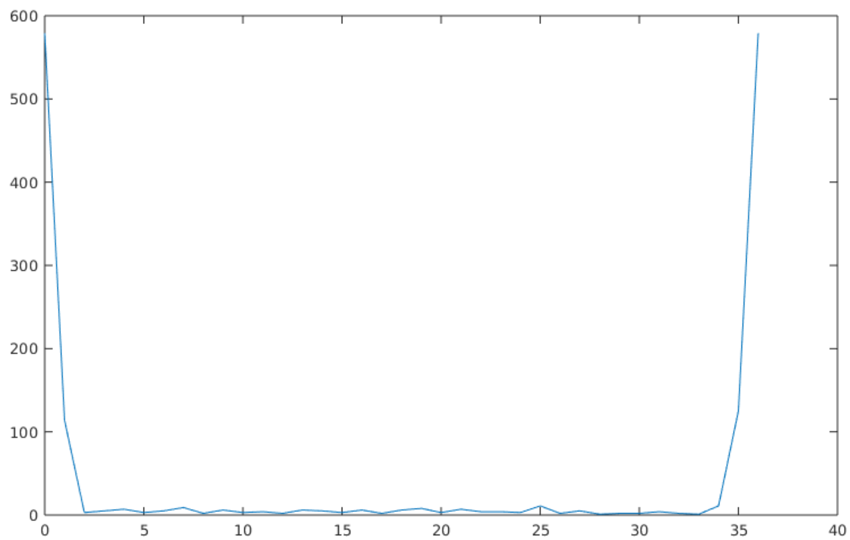

Figure 5: Histogram on number of matched points with FAST and BRIEF.

Figure 6: Matched SURF corners using BRIEF at $\theta = 0$.

Figure 7: Matched SURF corners using BRIEF at $\theta = 2\pi/3$.

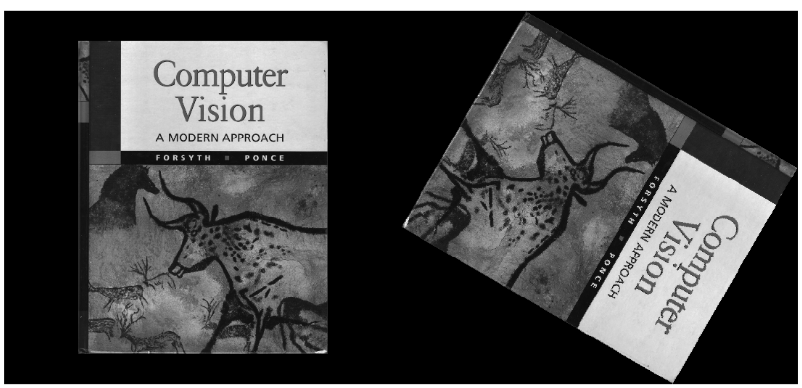

Figure 8: Matched SURF corners using BRIEF at $\theta = 4\pi/3$.

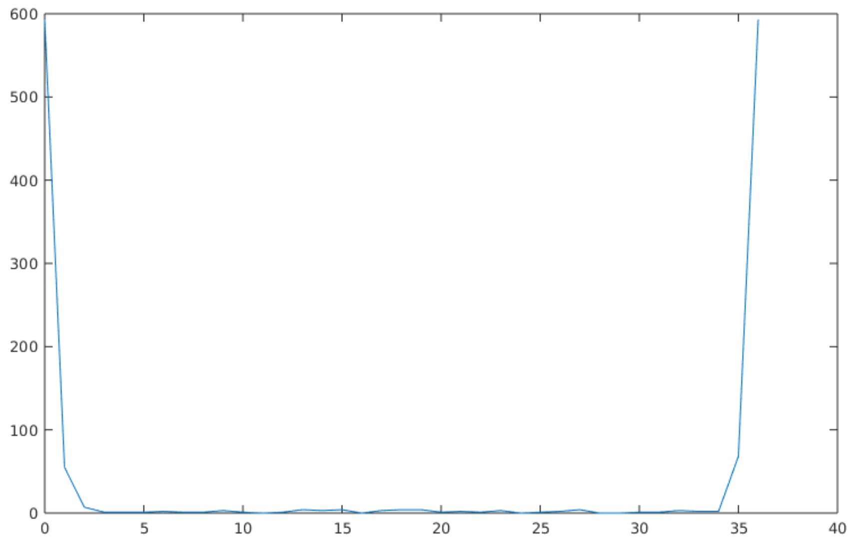

Figure 9: Histogram on number of matched points with SURF and BRIEF.

In both histograms, we see the number of matched points drop significantly as θ increases. This is probably because BRIEF descriptors are not rotational invariant. Swapping FAST with SURF does not improve the situation (arguably it is a bit worse). That is because as long as the descriptors are not rotational invariant, the points would not match well, i.e. it does not relate to the performance of the detectors.

Q2.2.4 Putting it together

After step 3, although the image `hp_cover.jpg` is properly warped to the correct location, it does not fill up the same space as the book because it does not have the same size as `cv_cover.jpg`. Resizing `hp_cover.jpg` to the dimensions of `cv_cover.jpg` by scaling the image width and height would fix the issue.

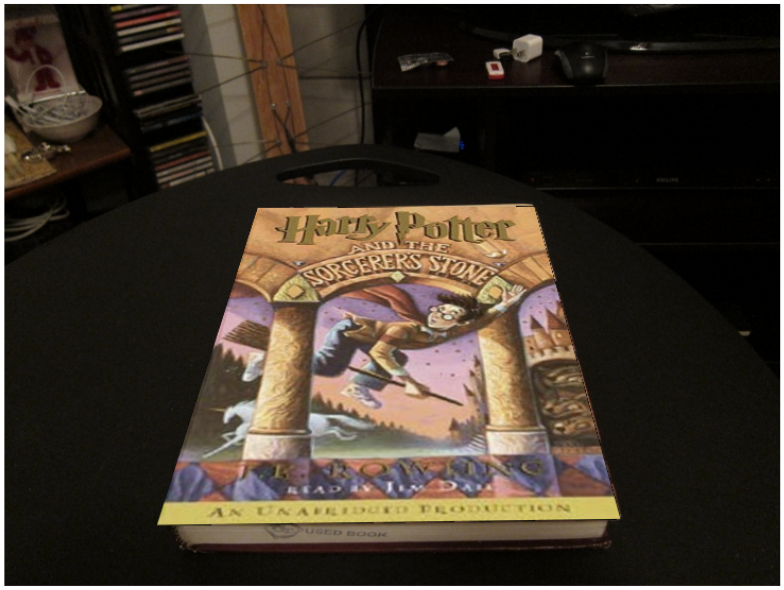

Figure 10: Composite image of Harry Potter cover on CV book.

4 Extra Credit

Q4.1x Make your AR real-time

Note to grader: Please run the code twice to get accurate result in fps. That is because the initialisation of `VideoPlayer` in the first iteration interferes with the timer. This implementation achieves 30-35 fps on our local machines.

Q4.2x Create a Simple Panorama

Figure 11: Original Image 1 (left)

Figure 12: Original Image 2 (right)

Figure 13: Panorama