CSE344: Computer Vision

Assignment-3

Divyajeet Singh (2021529)

April 14, 2024

Theory

Question 1.

Since the epipoles are the null space and left null space of the Essential matrix $\mathbf{E} = [\mathbf{t}_\times]\mathbf{R}$, we have

$$\mathbf{E}\mathbf{e}_1 = 0 \implies [\mathbf{t}_\times]\mathbf{R}\mathbf{e}_1 = 0$$

which means $\mathbf{R}\mathbf{e}_1$ is parallel to $\mathbf{t}$. So (using the fact that $\mathbf{R}$ is orthogonal),

$$\mathbf{R}\mathbf{e}_1 = \lambda_1\mathbf{t} \implies \mathbf{e}_1 = \lambda_1\mathbf{R}^\top\mathbf{t} \quad (\lambda_1 \in \mathbb{R})$$

Similarly, for the second epipole, we have (using the fact that $[\mathbf{t}_\times]$ is skew-symmetric),

$$\mathbf{E}^\top\mathbf{e}_2 = 0 \implies \mathbf{R}^\top[\mathbf{t}_\times]\mathbf{e}_2 = 0 \quad \because [\mathbf{t}_\times]^\top = -[\mathbf{t}_\times]$$

which means $\mathbf{e}_2$ is parallel to $\mathbf{t}$, since $\mathbf{R}^\top$ is a full rank matrix. So, the solution is that $\mathbf{e}_2 = \lambda_2\mathbf{t}$, $\lambda_2 \in \mathbb{R}$.

Question 2.

We are given a stereo camera setup with no relative rotation, i.e. $\mathbf{R} = \mathbf{I}$, and purely horizontal translation, i.e. $\mathbf{t} = \begin{bmatrix} t_x & 0 & 0 \end{bmatrix}^\top$. So, we construct the Essential matrix as follows

$$\mathbf{E} = [\mathbf{t}_\times]\mathbf{R} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \mathbf{I} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix}$$

We now try to find the relate corresponding points in this setup. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be the images of the same 3D point in the two images. Then, we have

$$\begin{aligned} \mathbf{x}_2^\top\mathbf{E}\mathbf{x}_1 &= 0 \\ \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -t_x \\ 0 & t_x & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} &= 0 \\ \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -t_x \\ t_x y_1 \end{bmatrix} &= 0 \\ -t_x y_2 + t_x y_1 &= 0 \\ \implies y_1 &= y_2 \end{aligned}$$

Note that it is crucial that $t_x \neq 0$ for this to hold - there must be some horizontal translation between the two cameras. Hence, we have found that the corresponding points in this setup must have the same y-coordinate. This makes sense, as the cameras are only translating horizontally and no rotation is involved.

Question 3.

We are required to find a rotation matrix $\mathbf{R}_{\text{rect}}$ that rectifies the stereo pair, i.e. maps the epipoles to infinity along the x-axis. We first decomose the (assumed to be given) Essential matrix $\mathbf{E}$ into the translation vector $\mathbf{t}$ and the rotation matrix $\mathbf{R}$ using Singular Value Decomposition. Taking reference from [1], we have

$$\mathbf{E} = [\mathbf{t}_\times]\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \mathbf{U} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{V}^\top \quad \because \textsc{Rank}(\mathbf{E}) = 2$$

If $\mathbf{U} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$, then $u_3 = \pm\mathbf{t}$ because $\mathbf{t}$ lies on the epipolar baseline, as

$$\mathbf{P}_2 \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} \mid \mathbf{t} \end{bmatrix}_{3\times 4} \begin{bmatrix} \mathbf{0}_3 \\ 1 \end{bmatrix} = \mathbf{t} \quad \text{and} \quad \mathbf{t}^\top \mathbf{E} = 0 \implies \mathbf{t} = \pm u_3$$

since the left null space of $\mathbf{E}$ is the epipole in the second image; where $\mathbf{P}_2$ is the perspective projection matrix of the second camera, and we assume that the first camera's origin coincides with the world coordinate frame, i.e. $\mathbf{P}_1 = \begin{bmatrix} \mathbf{I}_{3\times 3} \mid \mathbf{0}_3 \end{bmatrix}$. Now, we find $\mathbf{t}$ and $\mathbf{R}$ as follows

$$\mathbf{E} = \left(\mathbf{U} \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{U}^\top \right) \left(\mathbf{U}\mathbf{Y}\mathbf{V}^\top \right) = [\mathbf{t}_\times]\mathbf{R}$$

where

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{Y}$$

Now, we have $\mathbf{R}$ and $\mathbf{t}$, and we can find $\mathbf{R}_{\text{rect}}$. Let

$$\mathbf{R}_{\text{rect}} = \begin{bmatrix} r_1^\top \\ r_2^\top \\ r_3^\top \end{bmatrix}$$

Using the given slide deck as reference, we have

$$\begin{aligned} r_1 &= \mathbf{e}_1 = \hat{\mathbf{t}} = \frac{\mathbf{t}}{||\mathbf{t}||} \\ r_2 &= \frac{1}{\sqrt{t_x^2 + t_z^2}} \begin{bmatrix} -t_y \\ t_x \\ 0 \end{bmatrix} \\ r_3 &= r_1 \times r_2 \end{aligned}$$

where $\mathbf{t} = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^\top$ is the translation vector between the two cameras and $\mathbf{e}_1$ is the epipole in the first image, i.e. $\mathbf{E}\mathbf{e}_1 = 0$.

Panorama Generation

Keypoint Detection

We use the SIFT Algorithm to detect keypoints and descriptors in the first two images. Figure 1 shows the detected keypoints in the images. It is easy to verify that keypoints detected on sharp corners are correct.

Figure 1: Keypoints detected in the first two images

Feature Matching

We use the Brute-Force Matcher and the FLANN-based Matcher to match the keypoints in the two images. The matches are shown in Figure 2. We only display the matches that have a distance ratio less than 0.75 for the Brute-Force Matcher and a distance ratio less than 0.7 for the FLANN-based matcher. The matches are displayed on grayed images for better visibility.

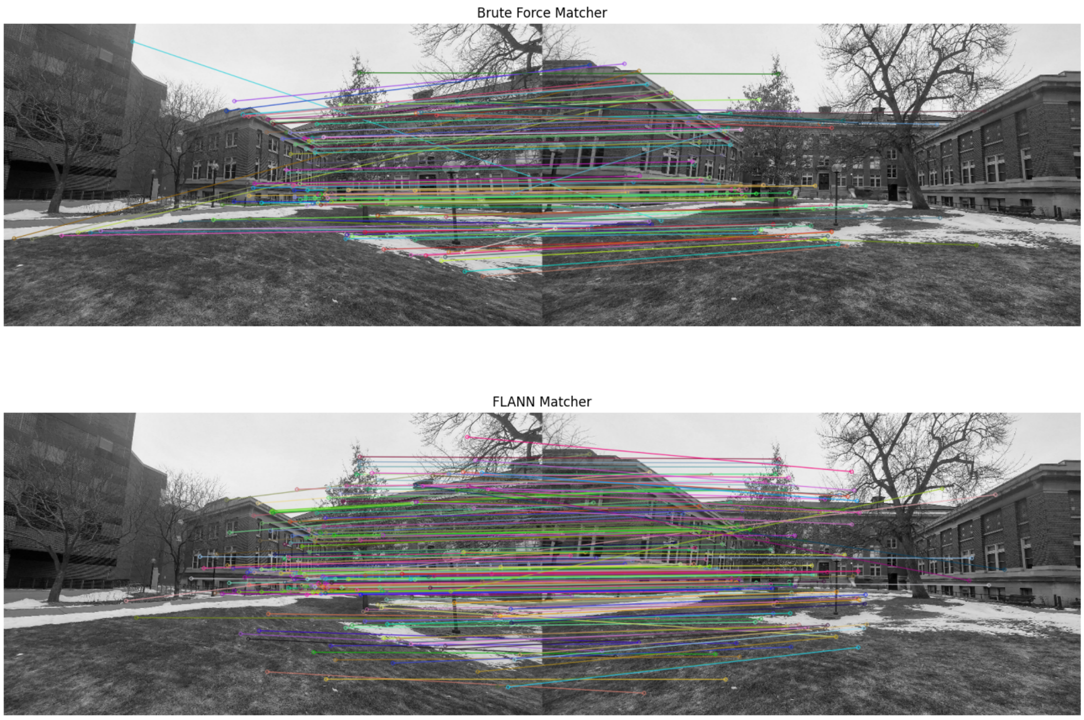

Figure 2: Matches between the keypoints in the first two images

Homography Estimation

We then estimate the homography matrix using the RANSAC algorithm. The homography matrix is used to warp one image onto the other. The homography matrix, rounded up to 2 decimal places, is

$$\mathbf{H} = \begin{bmatrix} -52.37 & 1.12 & 18955.0 \\ -15.91 & -32.21 & 9880.81 \\ -0.05 & -0.001 & 1.0 \end{bmatrix}$$

Perspective Warping

We use the homography matrix to warp the second image onto the first image onto the viewpoint of the second image. We warp the first image to look like it was taken from the viewpoint of the second image. The left and right side of the image is shown in Figure 3.

Figure 3: Warped image of the first image to the viewpoint of the second image

Stitching

We next create a panorama by stitching together the warped image and the second image. Figure 4 shows the generated panorama without any cropping or blending.

Figure 4: Panorama generated without any cropping or blending

Figure 5 shows the final panorama after cropping and blending the images.

Figure 5: Final panorama after cropping and blending

Multi-Stitching

Finally, we perform multi-stitching using `cv2.Stitcher()` and stitch all 8 images into a single panorama. Figure 6 shows the final panorama after multi-stitching.

Figure 6: Final panorama after multi-stitching

Referneces

1. *Lecture 10* in EE290T (Fall 2019) - Advanced topics in Signal Processing: 3D Image Processing & Computer Vision, University of California, Berkeley