CSE-483: Mobile Robotics **Monsoon 2019**

Epipolar Geometry I

Prepared by: Mohsin Mustafa, Shyamgopal Kartik

1 Need for Stereo

Given views of an object or a group of objects, we aim to retrieve their original 3D structure, or a relation that maps points from one view to that of another view. We first see two basic approaches two do this and examine the inherent limitations present in them.

1.1 Single View

Recovering structure from single view involves estimating the 3D coordinates of the objects present in a single image. This is difficult however due to the intrinsic ambiguity that arises out of mapping to lower dimensional coordinates — at least one dimension is *flattened*. In the case of 3D to 2D, a dimension worth of information, crucial in estimating the depth of a point on the image, is lost; this can be seen as the orange vector in Figure 1. Photographic illusions can be constructed by taking advantage of this, including popular tourist poses near landmarks like the Leaning Tower of Pisa.

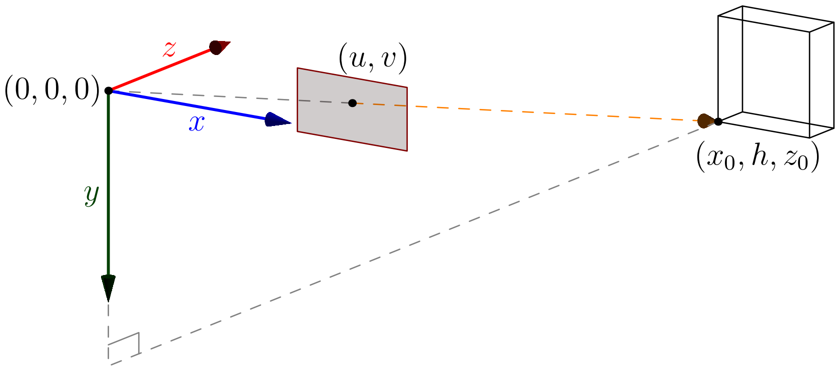


Figure 1: Given only the image plane (brown shaded rectangle), a single point (u, v) can correspond to any point along a vector $((x_0, h, z_0)$ as shown by the orange ray).

1.2 Appearance Based

Using multiple views of the same object, one method could be to take points of interest in both images, and map *similar* points of interest (based on SIFT/SURF features) across both images to each other. This has a two major limitations: one, it does not take advantage of the geometry of the camera and the images. Second, similar points of interest can be erroneously mapped to each other due to existing symmetric or pattern-like structures in the images. One such example is shown in Figure 2, where the landmark is symmetric around its vertical axis; in such a case, appearance is unable to map the windows to each other correctly.

This shows that we need to devise a method that takes into account the positions and orientations of the multiple cameras as well.

Figure 2: The windows in the left view can be erroneously mapped to any of the *wrong* windows in the right view, using just appearance-based features.

2 Epipolar Geometry

We start with Figure 3 of two cameras located at O_1 and O_2. We take the camera O_1 to be located at origin. O_2 can be represented in terms of O_1 as:

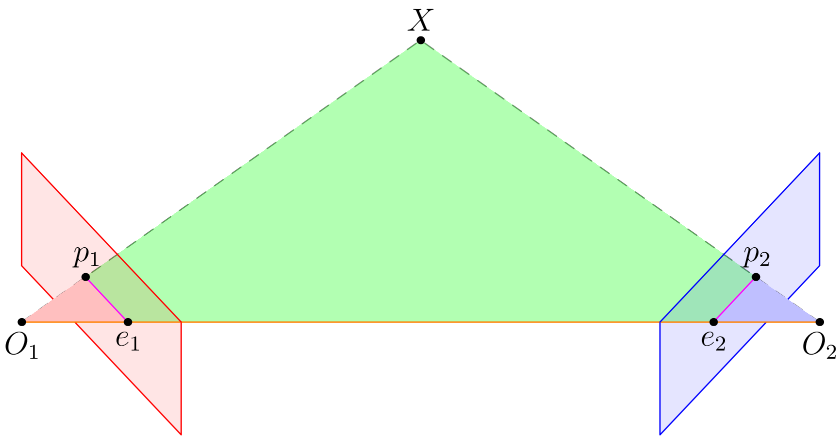
$$O_2 = \begin{bmatrix} \boldsymbol{R}_{12} & \boldsymbol{t}_{12} \end{bmatrix}_{3\text{x}4} O_1$$

The projection of O_2 on O_1's image plane is denoted by e_1, and the projection of O_1 on O_2's image plane is denoted by e_2. These two points, e_1 and e_2, are called the *epipoles* or *epipolar points*.

The line $\overline{O_1X}$ is seen as a point at O_1, as it is directly in line with O_1. However, this line can be seen on the image plane of O_2 as the line $\overline{e_2p_2}$; this line $\overline{e_2p_2}$ in camera O_2 is called an *epipolar line*. Similarly, the line $\overline{O_2X}$ seen at O_2 as a point is seen as the epipolar line $\overline{e_1p_1}$ by O_1.

Any epipolar line has to pass through the respective image plane's epipole. The epipolar lines in the two image planes are a function of X; as X varies, the epipolar lines change as well. Note that the epipoles remain constant as long as the cameras do not move.

Figure 3: The shaded triangles (red, green and blue) together make up the epipolar plane. The points e_1 and e_2 are the epipoles. The magenta lines $\overline{e_1p_1}$ and $\overline{e_2p_2}$ are the epipolar lines.



The triangle $\Delta O_1 X O_2$ is known as the epipolar plane. The epipolar planes intersect the projections of X and the epipoles on both image planes.

2.1 Epipolar Constraints

If the relative positions of the cameras are known, we lead to two observations:

1. If p_1 and the epipolar line e_2p_2 is known, then the corresponding projection p_2 on the second image plane is contained in the e_2p_2 epipolar line. This means that every X_i that projects to p_1 on the first image plane will be projected onto points on the epipolar line e_2p_2 on the second plane.
2. If p_1 and p_2 are known, then we known the projection lines of these points too. This implies that we can find the 3D coordinate of X via *triangulation*.

These constraints can be concisely written using the Essential (or Fundamental) Matrix as shown in the next section.

3 Deriving the Essential Matrix

As p_1 is the projection of X on the O_1's image plane,

$$\lambda_1 \vec{p}_1 = \boldsymbol{K} \begin{bmatrix} I & \boldsymbol{0} \end{bmatrix} \vec{X}_{4\text{x}1} \tag{1}$$

$$\implies \lambda_1 \boldsymbol{K}^{-1} \vec{p}_1 = \vec{X}_{4\text{x}1} \tag{2}$$

Note that $\lambda_1 \vec{p}_1$ need not pass through X; it passes through a linear transformation, *i.e* the perspective projection, of X onto O_1's image plane. $\boldsymbol{K}^{-1}\vec{p}_1$ is the direction vector of the ray from the camera origin O_1 to the 3D point X. $\boldsymbol{K}^{-1}\vec{p}_1$ gives us the normalized image coordinates that has removed effects of the camera, and is equivalent to assuming the use of calibrated cameras.

$$\boldsymbol{K}^{-1} \vec{p}_1 = \vec{x}_1 \tag{3}$$

Similarly, for O_2,

$$\lambda_2 \vec{p}_2 = \boldsymbol{K} \begin{bmatrix} \boldsymbol{R}_{21} & \boldsymbol{t}_{21} \end{bmatrix} \vec{X}_{4\text{x}1} \tag{4}$$

$$\implies \lambda_2 \boldsymbol{K}^{-1} \vec{p}_2 = \begin{bmatrix} \boldsymbol{R}_{21} & \boldsymbol{t}_{21} \end{bmatrix} \vec{X}_{4\text{x}1} \tag{5}$$

$$\implies \lambda_2 \underbrace{\vec{x}_2}_{\text{normalized coordinate}} = \begin{bmatrix} \boldsymbol{R}_{21} & \boldsymbol{t}_{21} \end{bmatrix} \vec{X}_{4\text{x}1} \tag{6}$$

Taking the epipolar plane $\Delta O_1 O_2 X$ (note that the following calculations are from O_1 frame, unless specified),

$$\overrightarrow{O_1 P} \cdot \left(\overrightarrow{O_1 O_2} \times \overrightarrow{O_2 X} \right) = 0 \tag{7}$$

$$\implies \lambda_1 \vec{x}_1 \cdot (\boldsymbol{t}_{12} \times \boldsymbol{R}_{12} \lambda_2 \vec{x}_2) = 0 \tag{8}$$

$$\implies \lambda_1 \lambda_2 \vec{x}_1 \cdot (\boldsymbol{t}_{12} \times \boldsymbol{R}_{12} \vec{x}_2) = 0 \tag{9}$$

We can represent a cross product operation as a matrix multiplication as follows:

$$\vec{a} \times \vec{b} = \begin{bmatrix} 0 & -\vec{a}_z & -\vec{a}_y \\ \vec{a}_z & 0 & -\vec{a}_x \\ -\vec{a}_y & \vec{a}_x & 0 \end{bmatrix} \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \begin{bmatrix} \vec{a}_\times \vec{b} \end{bmatrix}$$

where $a_\times$ is the skew-symmetric matrix of $\vec{a}$. Using this formulation in Eqn. 9, we obtain

$$\vec{x}_1^T \left[\boldsymbol{t}_{12\times}\right] \boldsymbol{R}_{12}\vec{x}_2 = 0 \tag{10}$$
$$\implies \quad \vec{x}_1^T \left[\boldsymbol{E}_{12}\right] \vec{x}_2 = 0 \tag{11}$$

$\boldsymbol{E}_{12}$ is the **Essential Matrix** that relates the coordinates in the second image to the corresponding points in the first one. Similarly, $\vec{x}_2^T \left[\boldsymbol{E}_{21}\right] \vec{x}_1 = 0$ is the equation which relates the first image to the second.

4 Deriving the Fundamental Matrix

Now we try to find a similar equation which can relate the non-normalized coordinates of the two images with each other, *i.e* p_1 and p_2. Starting from Eqn. 10,

$$\vec{x}_1^T \left[\boldsymbol{t}_{12\times}\right] \boldsymbol{R}_{12}\vec{x}_2 = 0$$
$$\implies \left[\boldsymbol{K}^{-1}p_1\right]^T \left[\boldsymbol{t}_{12\times}\right] \boldsymbol{R}_{12}\boldsymbol{K}^{-1}p_2 = 0 \tag{12}$$
$$\implies \quad p_1^T\boldsymbol{K}^{-T} \left[\boldsymbol{t}_{12\times}\right] \boldsymbol{R}_{12}\boldsymbol{K}^{-1}p_2 = 0 \tag{13}$$
$$\implies \quad p_1^T\boldsymbol{F}_{12}p_2 = 0 \tag{14}$$

$\boldsymbol{F}_{12} = \boldsymbol{K}^{-T} \left[\boldsymbol{t}_{12\times}\right] \boldsymbol{R}_{12}\boldsymbol{K}^{-1}$ is the **Fundamental Matrix** that relates the coordinates in the second image to the corresponding points in the first one. These coordinates are in terms of pixel coordinates in homogeneous form, *i.e* $\vec{p}_1 = \begin{bmatrix}\vec{p}_{1x} & \vec{p}_{1y} & 1\end{bmatrix}^T$.

The Essential Matrix and the Fundamental Matrix can be related as

$$\boldsymbol{K}^T\boldsymbol{F}_{12}\boldsymbol{K} = \boldsymbol{E}_{12} \tag{15}$$
$$\implies \quad \boldsymbol{K}^T\boldsymbol{F}\boldsymbol{K} = \boldsymbol{E} \tag{16}$$
$$\implies \quad \boldsymbol{F} = \boldsymbol{K}^{-T}\boldsymbol{E}\boldsymbol{K}^{-1} \tag{17}$$

5 Properties of Essential Matrices

- E is a singular matrix
- E has 2 equal non-singular values
- $Ee' = 0$ and $E^T e = 0$ where e is the epipole
- $E^T x$ is the epipolar line associated with x

References

[1] Multi-View Geometry Hartley and Zisserman Chapter 8, Link: "`https://www.robots.ox.ac.uk/~vgg/hzbook/hzbook1/HZepipolar.pdf`"

[2] CS231A Course Notes 3: Epipolar Geometry: Kenji Hata and Silvio Savarase, Link: "`https://web.stanford.edu/class/cs231a/course_notes/03-epipolar-geometry.pdf`"