Stereo Disparity

Antonio Terán Espinoza
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

1 Similar Triangles

The relationship between the pixel measurements, i.e., screen coordinates, and the three-dimensional location of the observed point can be deduced via similar triangles found in a calibrated stereo rig. Such a stereo setup is depicted in Figure 1.

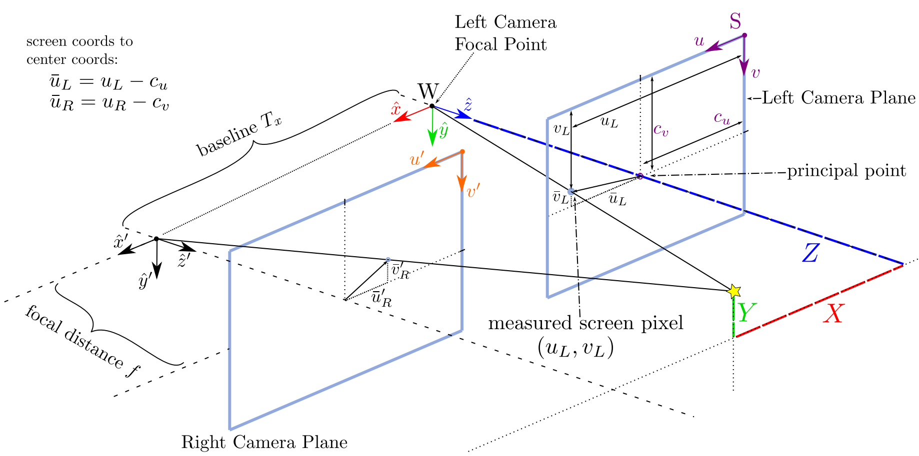


Figure 1: 3D view of calibrated stereo rig configuration.

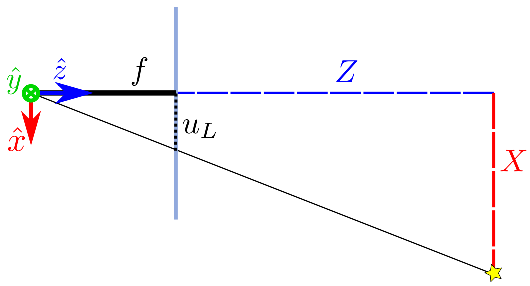
Conveniently, we can choose the origin of both cameras' frames of reference to be located at their focal points (left camera focal point explicitely shown). Such a choice yields a distance of length f between the origin and the corresponding camera plane; f represents the camera's focal length.

Within each camera plane, there are two important frames of reference: the 'global' screen frame S — with coordinates (u, v) — and a similarly aligned frame P — with coordinates $(\bar{u}, \bar{v})$ — whose origin is located at the camera's principal point. The principal point is defined as the point in the camera plane through which the principal axis (blue, $\hat{z}$) crosses. Its location, with respect to S, is given by the coordinates (c_u, c_v).

Using these definitions, the coordinates of any measured screen pixel with respect to the principal point frame P can be expressed as

$$\bar{u} = u - c_u, \quad \text{and} \quad \bar{v} = v - c_v.$$

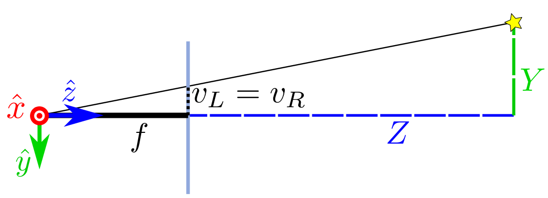
Similar similar triangles (hehe) in the $\hat{x}\hat{z}$ and the $\hat{y}\hat{z}$ plane can be formed in order to solve for the world X and Y coordinates of any point (generic point shown as star in Figure 1). These triangles, as well as the distance relationships, are shown in Figures 2 and 3.



$$\frac{Z}{f} = \frac{X}{\bar{u}_L} \implies X = \frac{Z}{f}\bar{u}_L$$

$$\implies X = \frac{T_x}{d}\bar{u}_L$$

Figure 2: Similar triangles in XZ plane.



$$\frac{Z}{f} = \frac{Y}{\bar{v}_L} \implies Y = \frac{Z}{f}\bar{v}_L$$

$$\implies Y = \frac{T_x}{d}\bar{v}_L$$

Figure 3: Similar triangles in YZ plane.

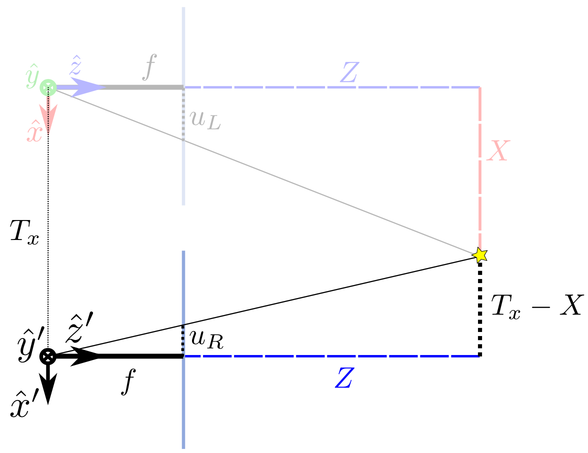
Initially, both coordinate values can only be expressed as a function of the depth of the point, Z. To solve for this last unknown, a similar triangle can be formed using the stereo rig's baseline T_x (distance between camera frames) and the right camera. This is shown in Figure 4, in which the difference between the left and right camera's $\bar{u}$ coordinate is defined as *disparity* d; it can be seen that the depth of a point is inversely proportional to the disparity.

Putting everything together, the following expressions for a 3D point expressed in the left camera frame W are obtained:

$$X = \frac{T_x}{d}\bar{u}_L, \quad Y = \frac{T_x}{d}\bar{v}_L, \quad \text{and} \quad Z = \frac{fT_x}{d}.$$

2 From Screen to World Coordinates (2D to 3D)

Case: Know screen coords (u_L, v_L) and disparity d, want $\mathbf{p} = [X, Y, Z]^\top$.



$$\frac{Z}{f} = \frac{T_x - X}{-\bar{u}_R} = \frac{T_x - \frac{\bar{u}_L Z}{f}}{-\bar{u}_R}$$

$$\implies Z = \frac{f T_x}{\bar{u}_L - \bar{u}_R}$$

$$\implies Z = \frac{f T_x}{d}$$

$$\bar{u}_L - \bar{u}_R \equiv d : \text{disparity}$$

Figure 4: Similar triangles used to solve for the depth of the observed point.

Going from the screen frame S — 2D coordinates (u, v) — to world coordinates — 3D coordinates (X,Y,Z) — with respect to the left camera frame W (for world) can be handled by the use of the so-called reprojection matrix $\mathbf{Q}$. Making use of the stereo rig's intrinsic (f, c_u, c_v) and extrinsic (T_x) parameters, the reprojection matrix takes the form

$$\mathbf{Q} = \begin{bmatrix} 1 & \cdot & \cdot & -c_u \\ \cdot & 1 & \cdot & -c_v \\ \cdot & \cdot & 0 & f \\ \cdot & \cdot & -\frac{1}{T_x} & \frac{c_u - c_v}{T_x} \end{bmatrix}.$$

Using the left camera's pixel coordinates (u_L, v_L) and the corresponding calculated disparity d to work in homogeneous coordinates, the transformation can be expressed as

$$\mathbf{Q} \cdot \mathbf{x} = \begin{bmatrix} 1 & & & -c_u \\ & 1 & & -c_v \\ & & 0 & f \\ & & -\frac{1}{T_x} & \frac{c_u - c_v}{T_x} \end{bmatrix} \cdot \begin{bmatrix} u_L \\ v_L \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X/W \\ Y/W \\ Z/W \\ W \end{bmatrix} = \begin{bmatrix} \mathbf{p}' \\ W \end{bmatrix} = \mathbf{x}',$$

where $\mathbf{p}' = [X/W,\ Y/W,\ Z/W]^\top$. In order to fully recover the 3D coordinates $\mathbf{p}$ of the the point being considered, the last component of the resulting homogeneous coordinate $\mathbf{x}'$ — W — is to be used to multiply $\mathbf{p}'$, that is,

$$\mathbf{p} = W \cdot \mathbf{p}'.$$

3 From World to Screen Coordinates (3D to 2D)

Case: Know $\mathbf{p} = [X, Y, Z]^\top$, want screen coords (u_L, v_L).

Taking a similar approach (based on the obtained similar triangles), it is straightforward to rearrange the equations' terms in order to solve for the screen coordinates (u_L, v_L) of the projection of a 3D point $\mathbf{p}$ onto the camera plane; such projection is depicted in Figure 5. Screen coordinates are thus expressed as

$$u_L = \frac{f}{Z} \cdot X + c_u, \quad \text{and} \quad v_L = \frac{f}{Z} \cdot Y + c_v.$$

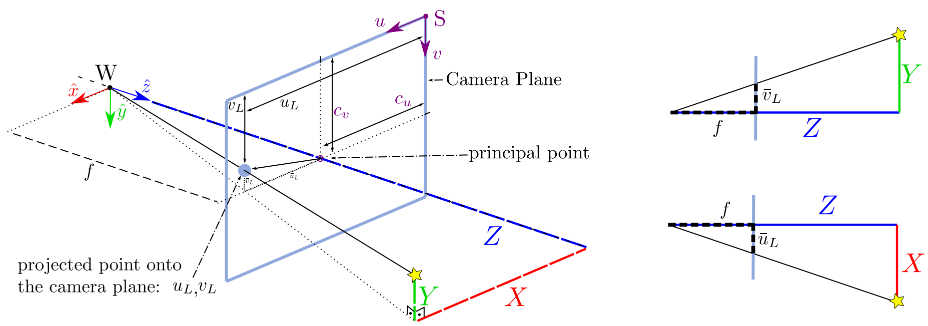


Figure 5: Similar triangles used to solve for the screen coordinates to project a 3D point onto the camera plane.