Camera Matrices

D.A. Forsyth,

University of Illinois at Urbana Champaign

Recall

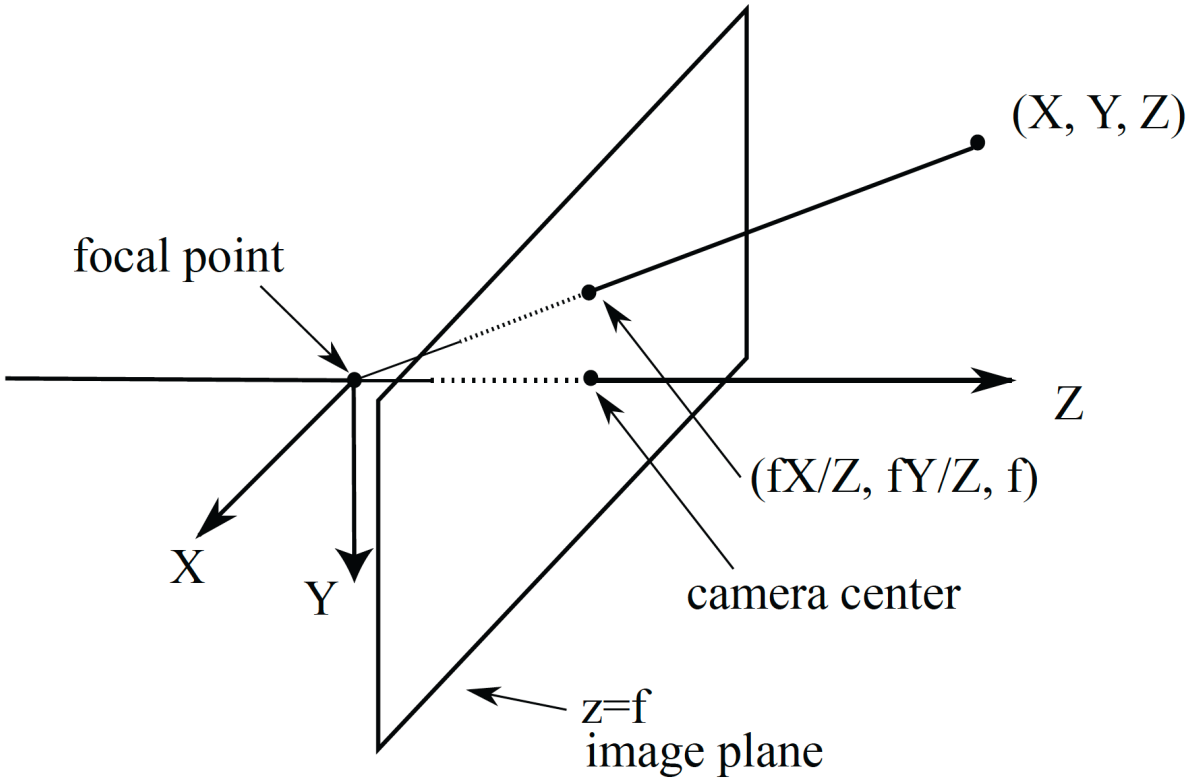


FIGURE 2.4: *The usual geometric abstraction of the pinhole model. The box doesn't affect the geometry, and is omitted. The pinhole has been moved to the back of the box, so that the image is no longer inverted. The image is formed on the plane* $z = f$*, by convention. Notice the y-axis goes down in the image. This allows me to use a* right handed *coordinate system and also have* z increase *as one moves into the image.*

Notice:

- Change in focal length is just a scaling
  - deal with this later
  - currently, f=1
- Now turn into homogenous coordinates

In 3D

Affine Homogenous

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cong \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

and

Camera

Affine Homogenous

$$\begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \cong \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

The perspective camera matrix

In affine coordinates, the camera mapping is $(X, Y, Z) \rightarrow (X/Z, Y/Z)$, and account for f later. Now write the 3D point in homogeneous coordinates as

$$\mathbf{X} = (X_1, X_2, X_3, X_4)$$

and the point in the image plane in homogeneous coordinates as

$$\mathbf{I} = (I_1, I_2, I_3).$$

Now we have

$$\mathbf{I} = (I_1, I_2, I_3) \equiv (X/Z, Y/Z, 1) \equiv (X, Y, Z) \equiv (X_1/X_4, X_2/X_4, X_3/X_4) \equiv (X_1, X_2, X_3).$$

This means that, in homogeneous coordinates, we can represent perspective projection as

$$(X_1, X_2, X_3, X_4) \rightarrow (X_1, X_2, X_3).$$

or

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

where the matrix is known as the *perspective camera matrix* (write $\mathcal{C}_p$).

Orthographic projection

27.1.2 Scaled Orthographic Projection

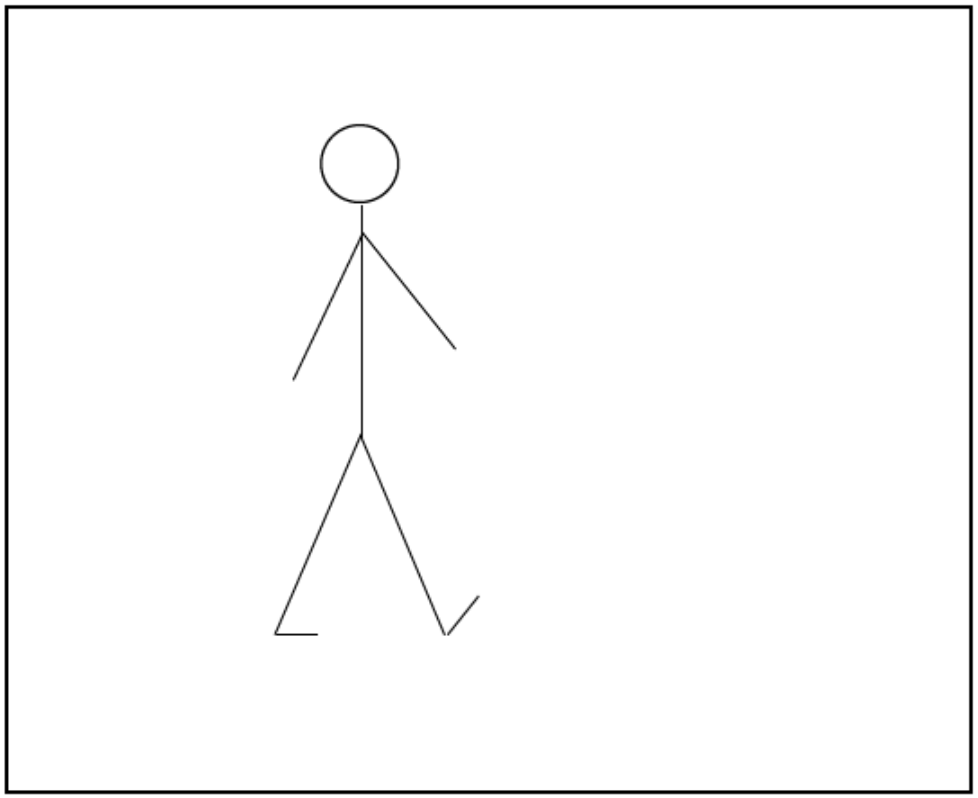
Under some circumstances, perspective projection can be simplified. Assume the camera views a set of points which are close to one another compared with the distance to the camera. Write $\mathbf{X}_i = (X_i, Y_i, Z_i)$ for the i'th point, and assume that $Z_i = Z(1+\epsilon_i)$, where ϵ_i is quite small. In this case, the distance to the set of points is much larger than the *relief* of the points, which is the distance from nearest to furthest point. The i'th point projects to $(fX_i/Z_i, fY_i/Z_i)$, which is approximately $(f(X_i/Z)(1-\epsilon_i), f(Y_i/Z)(1-\epsilon_i))$. Ignoring ϵ_i because it is small, we have the projection model

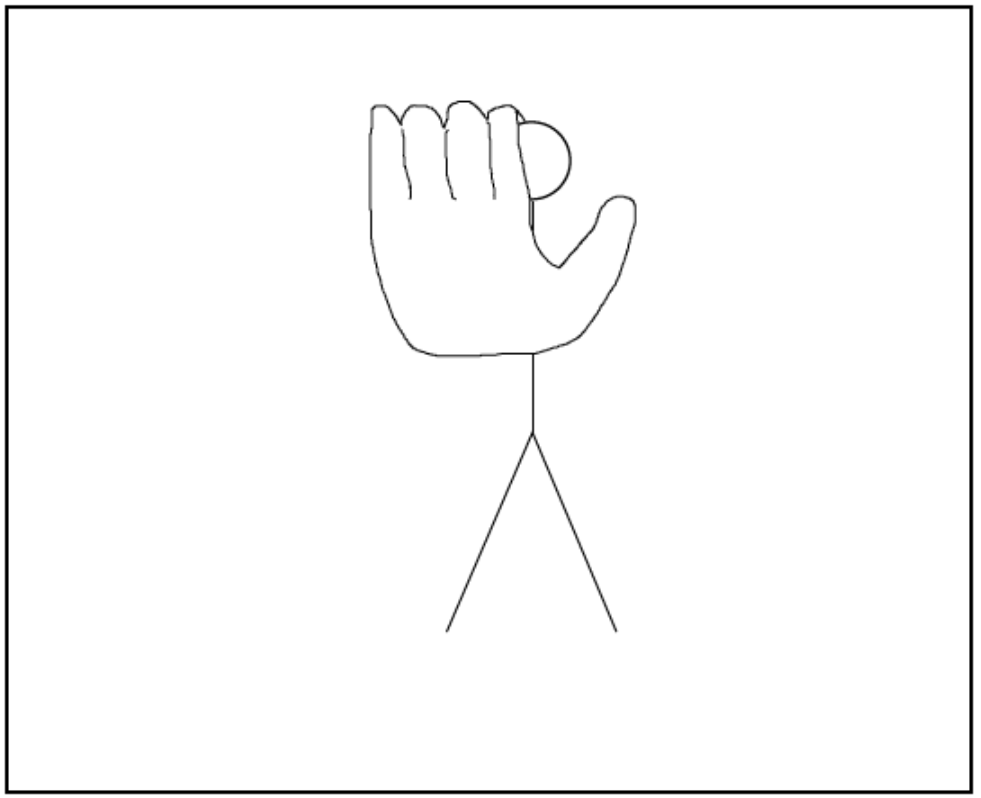
$$(X, Y, Z) \rightarrow (f/Z)(X, Y) = s(X, Y).$$

This model is usually known as *scaled orthograpic projection*. A geometric view of this model is that points in 3D "slide" down rays perpendicular to the image plane to form their image (Figure 27.8). It is an exercise to show that parallel lines do

Orthographic projection

- Often occurs when looking at people
  - criterion: difference in depth is small compared to depth





Orthographic projection

Orthographic projection maps

$$(X, Y, Z) \rightarrow (X, Y)$$

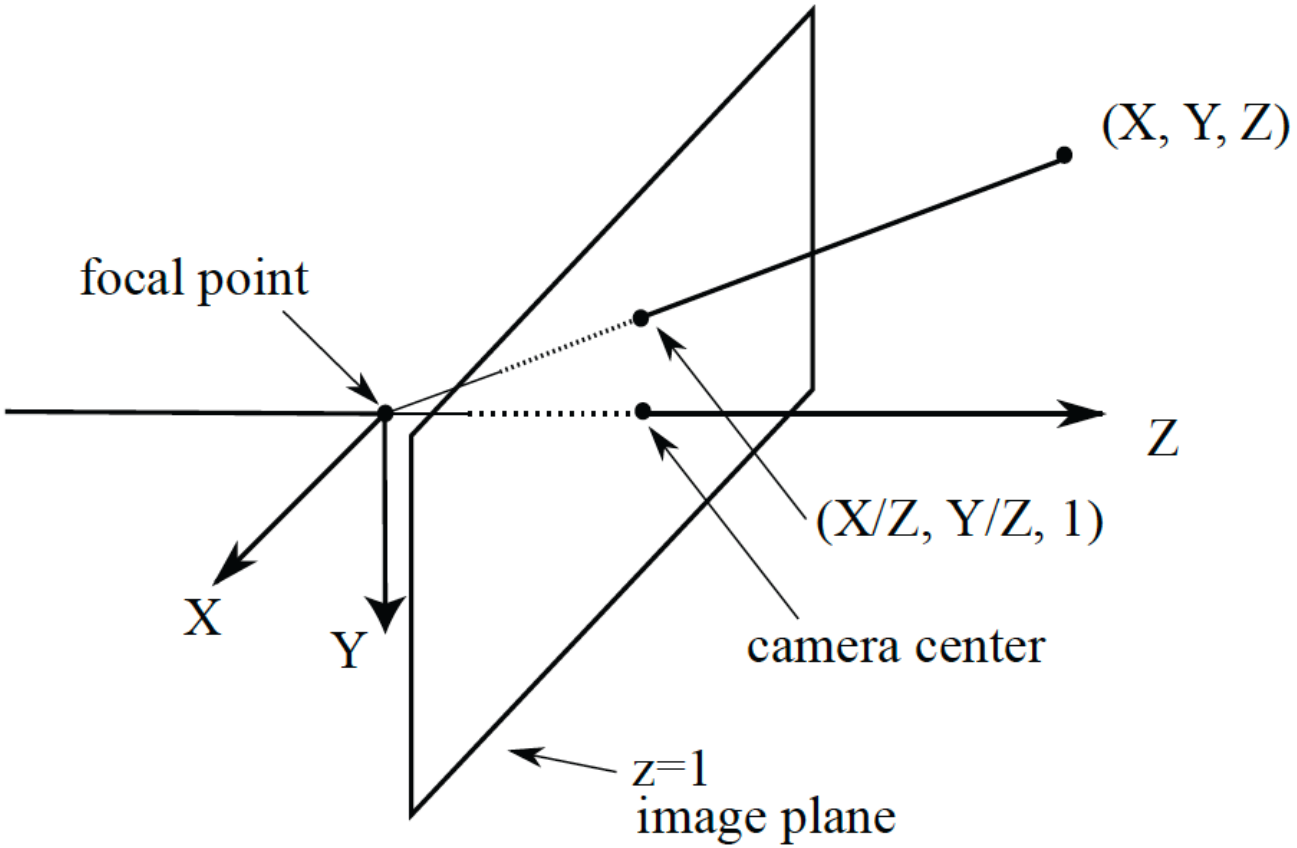
Camera matrix for orthographic

Orthographic projection maps

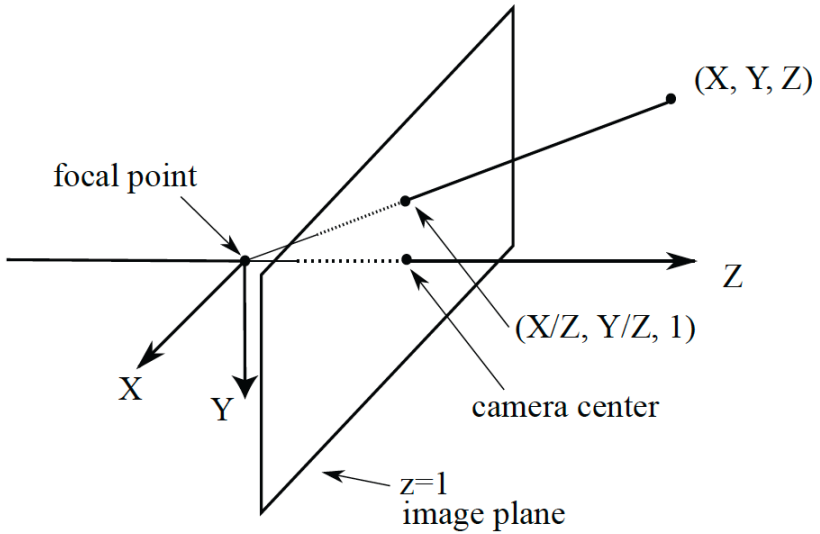
$$(X, Y, Z) \rightarrow (X, Y)$$

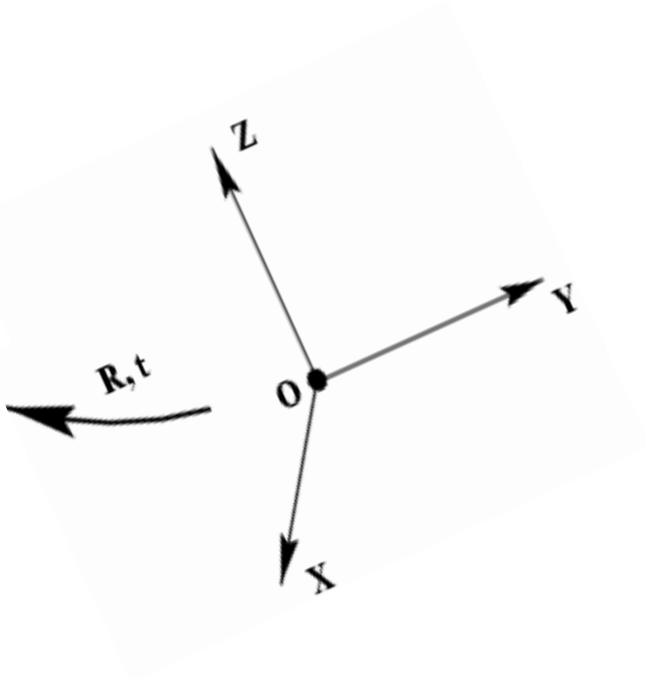
$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

Canonical camera

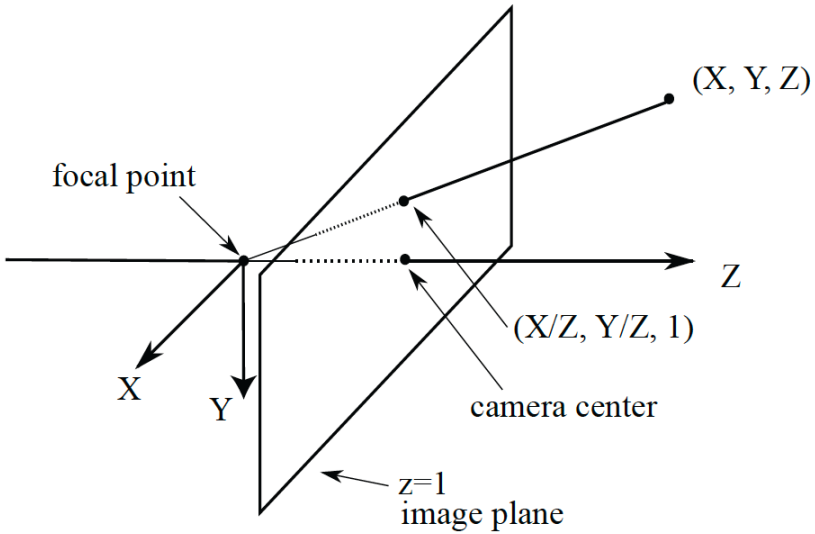


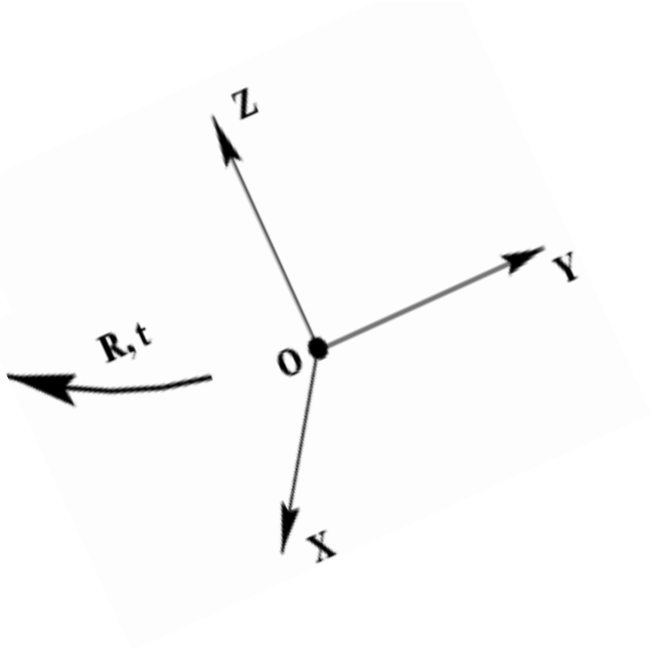
Actual camera





$$\begin{pmatrix} c_{11} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & c_{34} \end{pmatrix}$$





A general perspective camera transformation can be written as:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \text{Transformation} \\ \text{mapping image} \\ \text{plane coords to} \\ \text{pixel coords} \end{bmatrix} \mathcal{C}_p \begin{bmatrix} \text{Transformation} \\ \text{mapping world} \\ \text{coords to camera} \\ \text{coords} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$= \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$\mathcal{C}_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The parameters of $\mathcal{T}_i$ are known as *camera intrinsic parameters* or *camera intrinsics*, because they are part of the camera, and typically cannot be changed. The parameters of $\mathcal{T}_e$ are known as *camera extrinsic parameters* or *camera extrinsics*, because they can be changed.

Extrinsics

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \text{Transformation mapping image plane coords to pixel coords} \end{bmatrix} \mathcal{C}_p \begin{bmatrix} \text{Transformation mapping world coords to camera coords} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$= \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

Extrinsics

any Euclidean transformation maps the vector $\mathbf{x}$ to

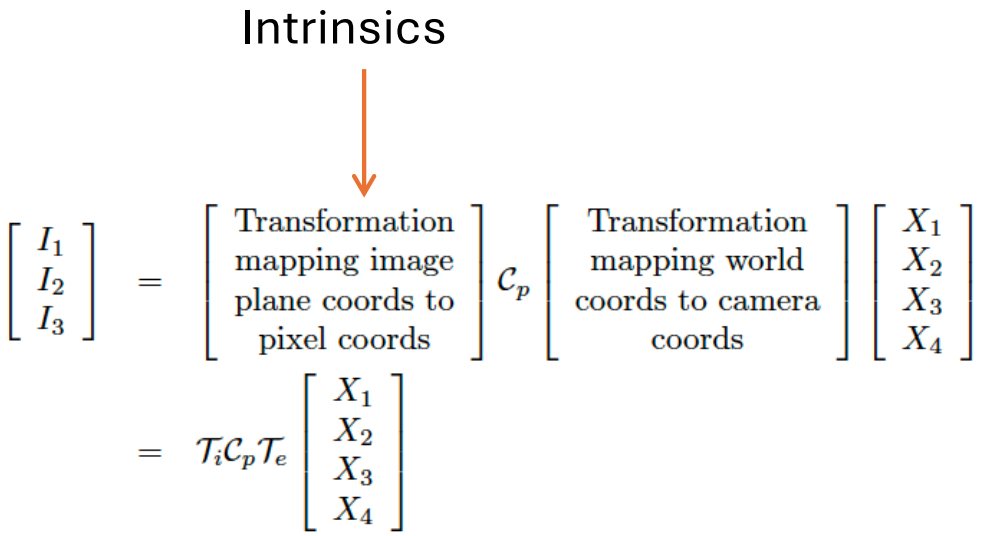
$$\mathcal{R}\mathbf{x} + \mathbf{t}$$

where $\mathcal{R}$ is an appropriately chosen 3D rotation matrix (check the endnotes if you can't recall) and $\mathbf{t}$ is the translation. Any map of this form is a Euclidean transformation. You should confirm the transformation that maps the vector $\mathbf{X}$ representing a point in 3D in homogeneous coordinates to

$$\lambda \begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{X}$$

represents a Euclidean transformation, but in homogeneous coordinates. It follows that any map of this form is a Euclidean transformation. Because $\mathcal{T}_e$ represents a Euclidean transformation, it must have this form **exercises** .

Intrinsics



Notice that

$$\mathcal{T}_i \mathcal{C}_p \mathcal{T}_e = \mathcal{T}_i \left[\mathcal{R} \mid \mathbf{t} \right]$$

which has a significant effect on the form of the intrinsic parameters. Any square matrix of full rank can be factored into a product of an upper triangular term and a rotation (**exercises**). Now assume you are *given* a general camera matrix

$$\mathcal{C} = \left[\mathcal{M} \mid \mathbf{v} \right].$$

Intrinsics

This could only be a camera matrix if $\mathcal{M}$ had full rank **exercises** . You can factor $\mathcal{M}$ into $\mathcal{U}\mathcal{Q}$, where $\mathcal{U}$ is upper triangular and $\mathcal{Q}$ is a rotation. The only ambiguities have to do with signs **exercises** . This means that, if $\mathcal{T}_i$ is *not* upper triangular, an appropriate choice of rotation would make it upper triangular **exercises** .

It is usual to work with an upper triangular $\mathcal{T}_i$. There are easy physical interpretations for the elements of $\mathcal{T}_i$. Write

$$\mathcal{T}_i = \begin{bmatrix} s_x & k & c_x \\ 0 & s_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

Interpreting intrinsics

$$\mathcal{T}_i = \begin{bmatrix} s_x & k & c_x \\ 0 & s_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

The bottom right element of $\mathcal{T}_e$ is 1, because you can scale the camera matrix without changing its effects – the camera matrix operates on homogenous coordinates. In turn, the mapping represented by $\mathcal{T}_e$ is

$$(\frac{X}{Z}, \frac{Y}{Z}, 1) \rightarrow (s_x \frac{X}{Z} + k\frac{Y}{Z} + c_x, s_y \frac{Y}{Z} + c_y, 1)$$

This mapping takes the camera center in world coordinates to $(c_x, c_y, 1)$, so c_x and c_y are given by the location of the camera center in camera coordinates. The parameter k is referred to as *skew*, and is usually 0. If the camera coordinate axes are not at right angles to one another, it might not be zero. The imaging device is usually perpendicular to the lens axis – if it has been knocked out of place slightly, k might not be zero. It is usual to assume $k = 0$ except in special cases.

Interpreting intrinsics

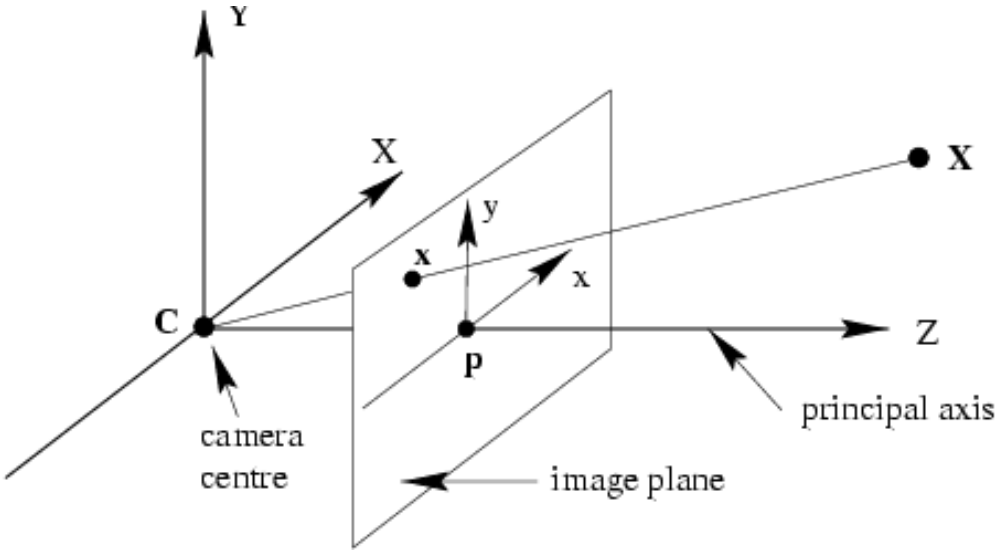
$$\mathcal{T}_i = \begin{bmatrix} s_x & k & c_x \\ 0 & s_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

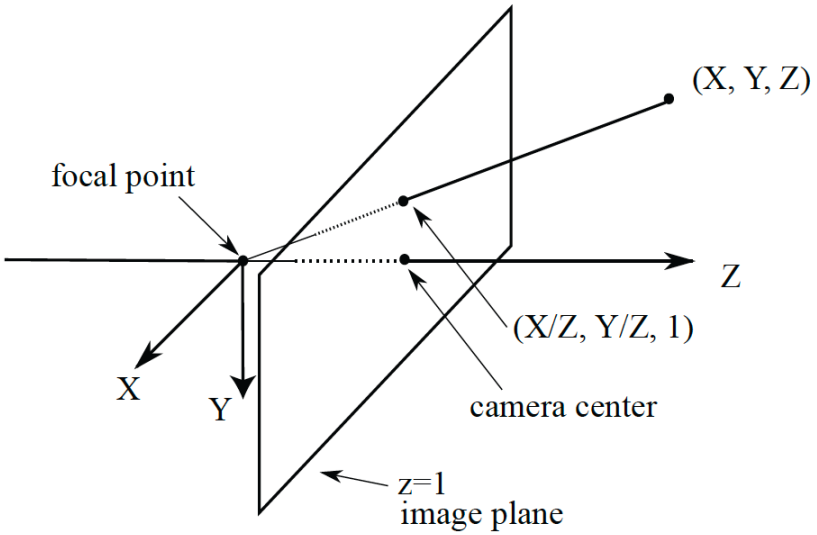
From any point in the image plane, take a unit step in the X direction on the image plane, where the size of this step is measured in world coordinates. In camera coordinates, the x-coordinate will change by s_x. You can interpret s_x as the scale of camera coordinates relative to world coordinates in the x-direction. For example, pixels in a camera sensor could be spaced a few micrometres apart. In this case, moving by 1000 pixels in the image plane might move the actual pixel location by a few millimetres. In some cameras, the spacing between pixels in the y-direction is different from that in the x direction, so s_y may be different from s_x. It is quite usual to use one scale s, and an aspect ratio a so that

$$\mathcal{T}_i = \begin{bmatrix} as & 0 & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

Here (c_x, c_y) is the location of the camera center; s is the *scale*; and a is the *aspect ratio*.

Different canonical cameras?





A camera views a plane

$$\mathbf{x} = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right] \begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{X}$$

- Put the plane at Z=0

$$\mathbf{X} = \begin{bmatrix} U \\ V \\ 0 \\ 1 \end{bmatrix}$$

- Simplify

A camera views a plane

- Transformation simplifies to

$$\mathbf{x} = \mathcal{K}\left[\mathbf{r}_1 \mathbf{r}_2 \mathbf{t}\right] \begin{bmatrix} U \\ V \\ 1 \end{bmatrix}$$

This is a homography

but may not be a general projective transformation

Camera rotates about its focal point

Camera 1

$$\mathbf{x} = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\mathbf{X}$$

Camera 2

$$\mathbf{x}' = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}\mathbf{X}$$

But t must be zero, because the focal point doesn't change (no translation)
Check – 0, 0, 0, 1 will be fp

Camera rotates about its focal point

Camera 1

$$\mathbf{x} = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\mathbf{X}$$

Camera 2

$$\mathbf{x}' = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\begin{bmatrix} \mathcal{R} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}\mathbf{X} = \mathcal{K}\left[\mathcal{R} \mid \mathbf{0}\right]\mathbf{X}$$

Camera rotates about its focal point

Camera 1

$$\mathbf{x} = \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\mathbf{X}$$

Camera 2

$$\begin{aligned}\mathbf{x}' &= \mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\begin{bmatrix}\mathcal{R} & \mathbf{0} \\ \mathbf{0}^T & 1\end{bmatrix}\mathbf{X} \\ &= \mathcal{K}\left[\mathcal{R} \mid \mathbf{0}\right]\mathbf{X} = \mathcal{K}\mathcal{R}\mathcal{K}^{-1}\mathcal{K}\left[\mathcal{I} \mid \mathbf{0}\right]\mathbf{X} \\ &= \mathcal{K}\mathcal{R}\mathcal{K}^{-1}\mathbf{x}\end{aligned}$$

Rotate about the focal point == apply a particular homography to image
Notice intrinsics matter here, entirely appropriate

Remember this: *A general perspective camera can be written in homogeneous coordinates as:*

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \mathcal{T}_i \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$= \begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

where $\mathcal{R}$ is a rotation matrix.

Remember this: *Alternative representations of perspective cameras are quite common. It is usual to write $\mathcal{K}$ for $\mathcal{T}_i$ (the intrinsic transformation). If you then write*

$$\begin{pmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}$$

for the extrinsic transformation, and multiply out, you get the quite common form

$$\mathcal{K}\,[\mathcal{R} \mid \mathbf{t}]$$

Procedure: 30.1 *Decomposing a general projective camera matrix*

Given a 3×4 camera matrix $\mathcal{C}$ with rank 3, decompose into

$$\mathcal{T}_i \mathcal{C}_p \mathcal{T}_e$$

as follows. Write $\mathcal{C} = [\mathcal{S} \mid \mathbf{p}]$. Now decompose $\mathcal{S}$ into an upper triangular matrix $\mathcal{U}$ and a rotation matrix $\mathcal{R}$. Then

$$\mathcal{T}_i = (1/u_{33})\mathcal{U} \text{ and } \mathcal{T}_e = \begin{bmatrix} \mathcal{R} & \mathcal{T}_i^{-1}\mathbf{p} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

I did not discuss...