Camera Geometry for One Camera

D.A. Forsyth, UIUC

Homogeneous coordinates

- Add an extra coordinate and use an equivalence relation
- for 2D
  - equivalence relation
    k*(X,Y,Z) is the same as (X,Y,Z)
- for 3D
  - equivalence relation
    k*(X,Y,Z,T) is the same as (X,Y,Z,T)
- "Ordinary" or "non-homogeneous" coordinates
  - properly called affine coordinates
  - in 3D, affine -> homogeneous

$$(x, y, z) \to k * (x, y, z, 1)$$

  - in 3D, homogeneous to affine

$$(X, Y, Z, T) \to \left(\frac{X}{T}, \frac{Y}{T}, \frac{Z}{T}\right)$$

Homogenous coordinates

- Notice $(0,0,0,0)$ is meaningless (HC's for 3D)
  - also $(0,0,0)$ in 2D
- Basic notion
  - Possible to represent points "at infinity" by careful use of zero
  - Where parallel lines intersect
    - eg

$$(tX, tY, tZ, 1) \quad \text{and} \quad (tX + a, tY + b, tZ + c, 1)$$

intersect at $(X, Y, Z, 0)$

  - Where parallel planes intersect (etc)
- Can write the action of a perspective camera as a matrix

Lines on the plane

Lines on the affine plane form one important example of homogeneous coordinates. A line is the set of points (x, y) where $ax + by + c = 0$. We can use the coordinates (a, b, c) to represent a line. If $(d, e, f) = \lambda(a, b, c)$ for $\lambda \neq 0$ (which is the same as $(d, e, f) \equiv (a, b, c)$), then (d, e, f) and (a, b, c) represent the same line. This means the coordinates we are using for lines are homogeneous coordinates, and the family of lines in the affine plane is a projective plane. Notice that encoding lines using affine coordinates must leave out some lines. For example, if we insist on using $(u, v, 1) = (a/c, b/c, 1)$ to represent lines, the corresponding equation of the line would be $ux + vy + 1 = 0$. But no such line can pass through the origin – our representation has left out every line through the origin.

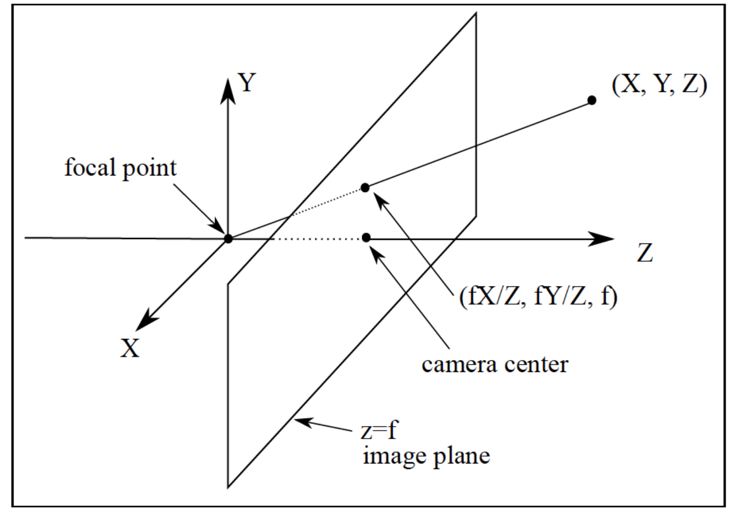
Remember this: *Most practical cameras can be modelled as a pinhole camera. A pinhole camera with focal length f maps*

$$(X, Y, Z) \rightarrow (fX/Z, fY/Z).$$

Figure 2.1 shows important terminology (focal point; image plane; camera center).

The camera matrix

- Turn previous expression into HC's
  - HC's for 3D point are (X,Y,Z,T)
  - HC's for point in image are (U,V,W)



In Camera coordinate system:

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

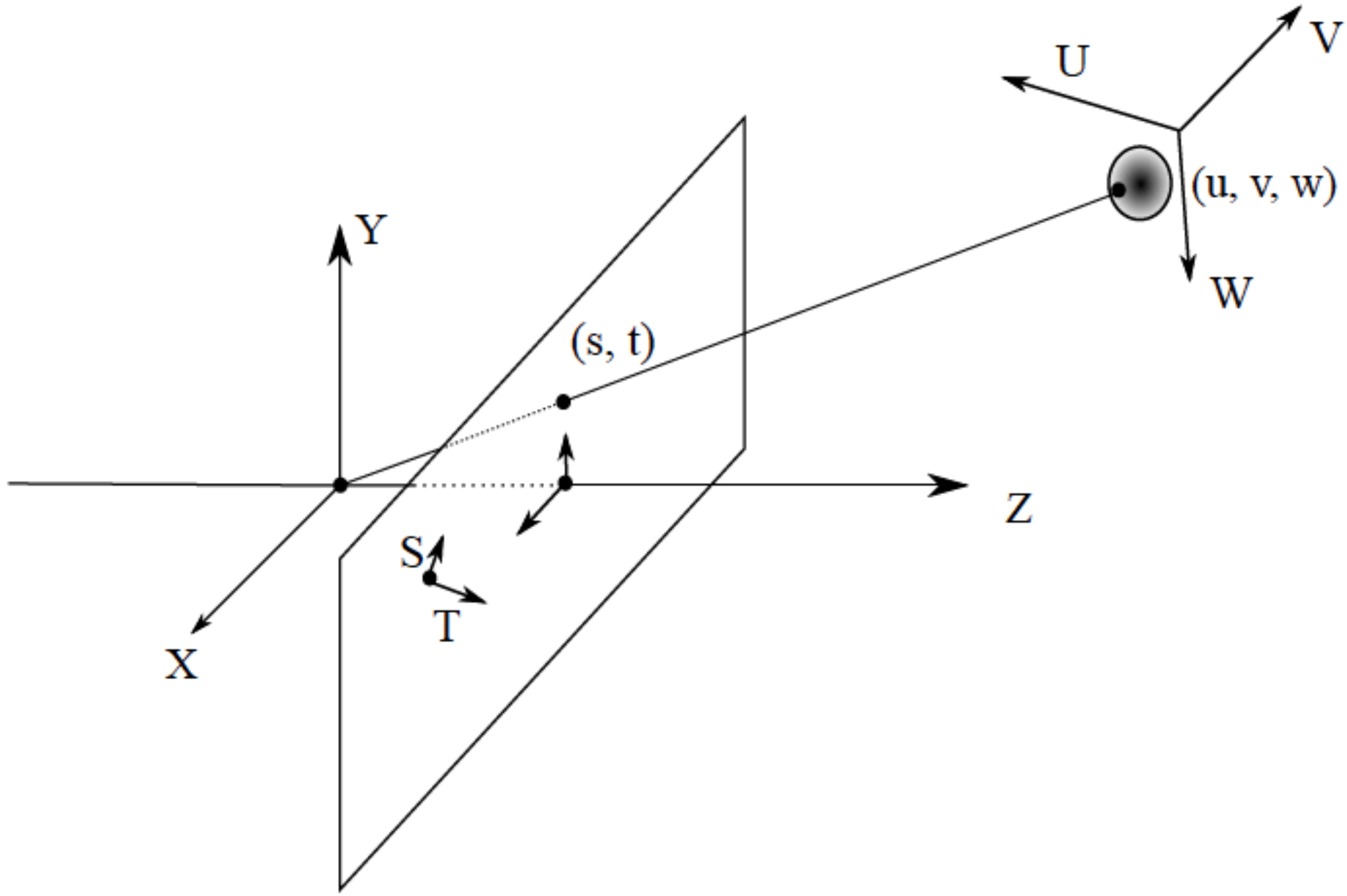


FIGURE 2.7: *A perspective camera (in its own coordinate system, given by X, Y and Z axes) views a point in world coordinates (given by (u, v, w) in the UVW coordinate system) and reports the position of points in ST coordinates. We must model the mapping from (u, v, w) to (s, t), which consists of a transformation from the UVW coordinate system to the XYZ coordinate system followed by a perspective projection followed by a transformation to the ST coordinate system.*

The camera matrix - II

- Turn previous expression into HC's
  - HC's for 3D point are (X,Y,Z,T)
  - HC's for point in image are (U,V,W)

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

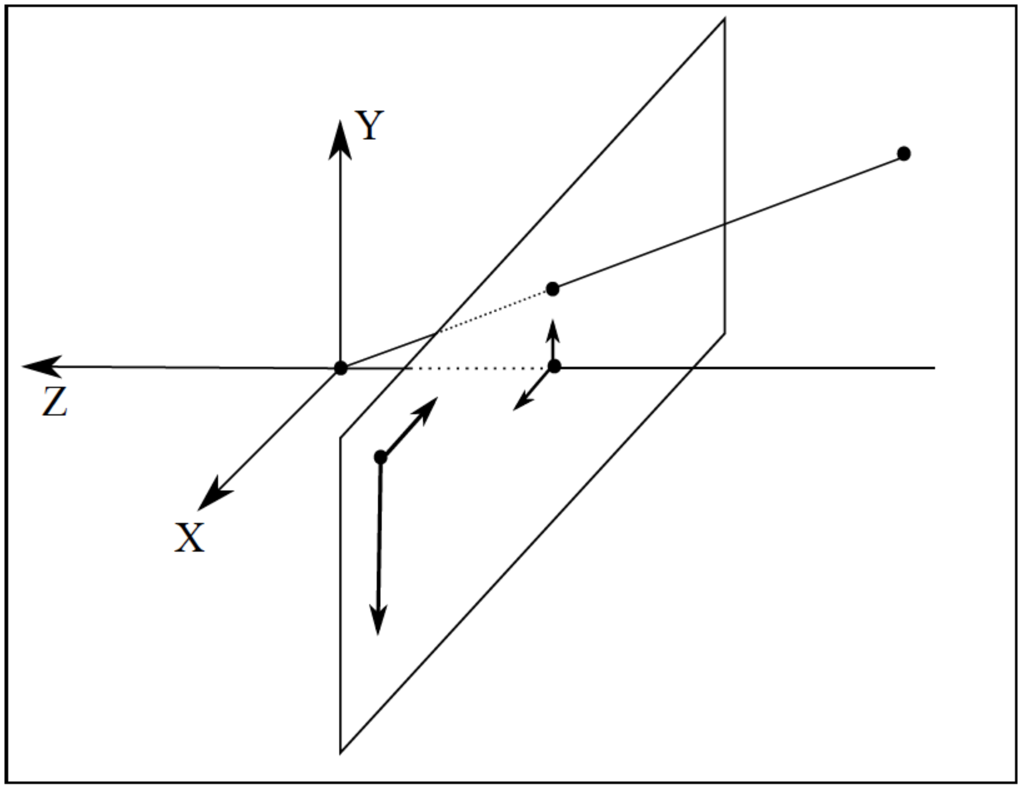
Camera extrinsics

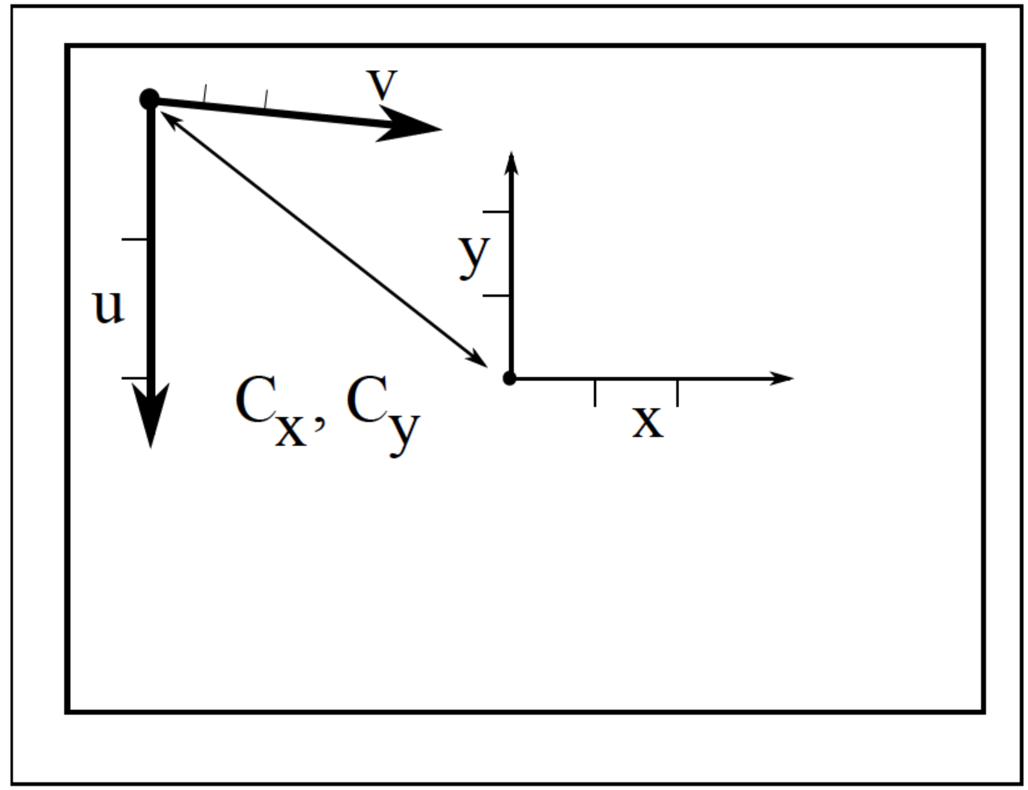
$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \text{Transformation} \\ \text{mapping image} \\ \text{plane coords to} \\ \text{pixel coords} \end{bmatrix} \mathcal{C}_p \begin{bmatrix} \text{Transformation} \\ \text{mapping world} \\ \text{coords to camera} \\ \text{coords} \end{bmatrix}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

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

Rotation matrix - orthornormal, det=1

Camera intrinsics





Camera intrinsics

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

$$\begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1/f \end{bmatrix}$$

Skew

Pixel aspect ratio

$$\begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1/f \end{bmatrix}$$

Location of the camera center (where the z-axis pierces the image plane)

A scale

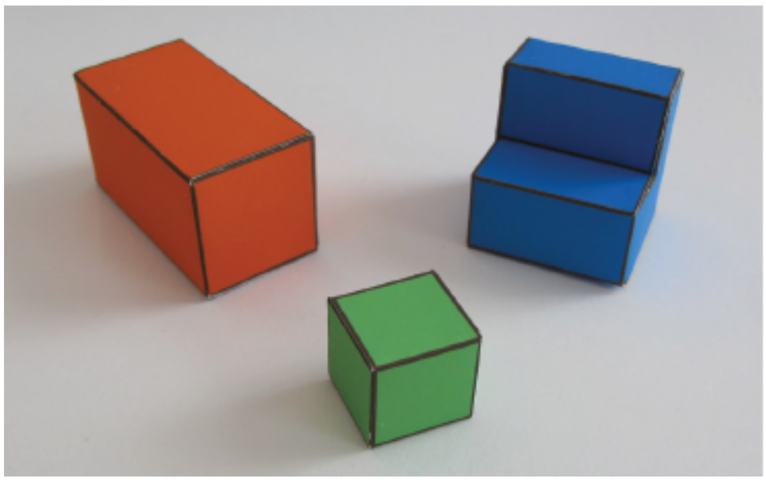
Camera matrix for orthographic projection

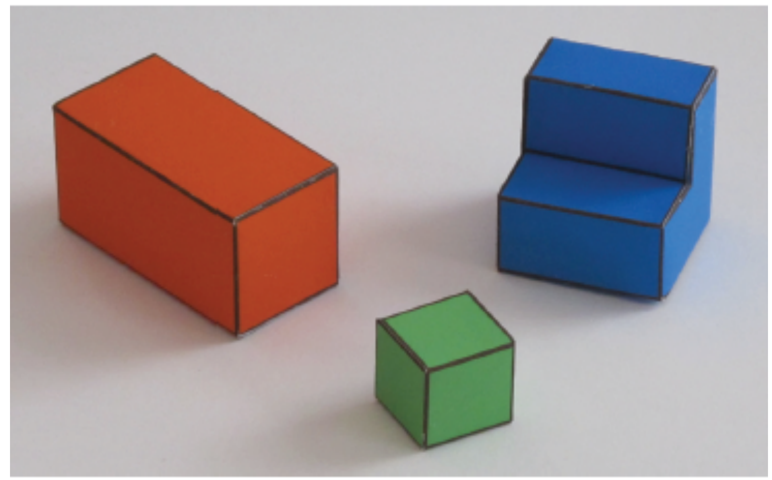
- Almost never encounter orthographic projection

> **Remember this:** *Scaled orthographic projection maps*
>
> $$(X, Y, Z) \rightarrow s(X, Y)$$
>
> *where s is some scale. The model applies when the distance to the points being viewed is much greater than their relief. Many views of people have this property.*

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \mathcal{C} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathcal{W} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Which projection?





Lift from Antonio Torralba's slides