CHAPTER 30

Camera Matrices

30.1 SIMPLE PROJECTIVE GEOMETRY

Draw a pattern on a plane, then view that pattern with a perspective camera. The distortions you observe are more interesting than are predicted by simple rotation, translation and scaling. For example, if you drew parallel lines, you might see lines that intersect at a vanishing point – this doesn't happen under rotation, translation and scaling. *Projective geometry* can be used to describe the set of transformations produced by a perspective camera.

30.1.1 Homogeneous Coordinates

The coordinates that every reader will be most familiar with are known as *affine coordinates*. In affine coordinates, a point on the plane is represented by 2 numbers, a point in 3D is represented with 3 numbers, and a point in d dimensions is represented with d numbers.

In *homogenous coordinates*, you represent a point in d dimensions using $d+1$ numbers *not all of which are zero*. Two representations $\mathbf{X}_1$ and $\mathbf{X}_2$ represent the same point (write $\mathbf{X}_1 \equiv \mathbf{X}_2$) if there is some $\lambda \neq 0$ so that

$$\mathbf{X}_1 = \lambda \mathbf{X}_2.$$

> **Remember this:** *In homogeneous coordinates, a point in a d dimensional space is represented by $d+1$ coordinates $(X_1, \ldots, X_{d+1})$. Write $\mathbf{X}_1 \equiv \mathbf{X}_2$ when $\mathbf{X}_1$ and $\mathbf{X}_2$ represent the same point. For any $\lambda \neq 0$,*
>
> $$\mathbf{X}_1 \equiv \lambda \mathbf{X}_1$$
>
> *The point $(0, 0, \ldots, 0)$ is meaningless in homogeneous coordinates.*

30.1.2 Projective Spaces

The space represented by $d+1$ homogeneous coordinates is different from the space represented by d affine coordinates in important but subtle ways.

Example: Lines on the affine plane form one important example of a projective space. A line on the affine plane is the set of points (x, y) where $ax+by+c=0$. Use the coordinates (a, b, c) to represent a line. Now $\lambda(a, b, c)$ for $\lambda \neq 0$ represents the same line. In turn, if you represent a line by (a, b, c) to represent a

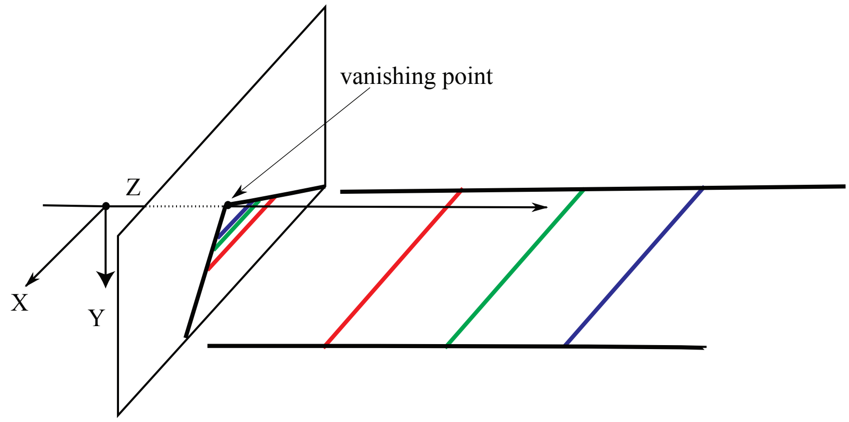


FIGURE 30.1: *The point at infinity is not just an abstraction: you can see it. Recall that lines that are parallel in the world can intersect in the image at a vanishing point. This vanishing point is the image of the point "at infinity" on the parallel lines. For example, on the plane* $y = -1$ *in the camera coordinate system, draw two lines* $(1, -1, t)$ *and* $(-1, -1, t)$ *(the lines shown in the figure). Now these lines project to* $(f1/t, f(-1/t), f)$ *and* $(f(-1/t), f(-1/t), f)$ *on the image plane, and their vanishing point is* $(0, 0, f)$*. This vanishing point occurs when the parameter* t *reaches infinity — it is the image of the point at infinity.*

line, you are using homogeneous coordinates. If you try to use affine coordinates to represent lines, you will leave out some lines. For example, if you insist on using $(u, v, 1) = (a/c, b/c, 1)$ to represent a line, 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. As another example, if you represent lines as $y = ax + b$ – equivalently, $(-a, 1, -b)$ – you will omit vertical lines.

Example: The *projective line* is represented by two homogenous coordinates, (X_1, X_2) (by convention, homogeneous coordinates are written with capital letters). Almost all points on the projective line can be written in the form $(u, 1)$. You can parametrize the whole of the *affine line* (the usual line) with one parameter u, so every point on an affine line has a corresponding point on a projective line. But there is an "extra point" on the projective line, given by $(X_1, 0)$. This point would be "at infinity" on an affine line. This point is like any other point on the line, as you can see by applying the projective transformation that takes (X_1, X_2) to (X_2, X_1). Check that this transformation is one-to-one, and that it maps the point at infinity to the origin and the origin to the point at infinity **exercises** . Furthermore, you can see the point at infinity (Figure 30.1).

Example: The *projective plane* is the space represented by three homogeneous coordinates. You can map an *affine plane* (the usual plane, with coordinates x, y) to a projective plane by writing $(X_1, X_2, X_3) = (x, y, 1)$. Notice that there are points on the projective plane — the points where $X_3 = 0$ — that are missing. These points form a projective line (check this!). This line is often referred to as the *line at infinity*. You can see the line at infinity, too (Figure 30.2).

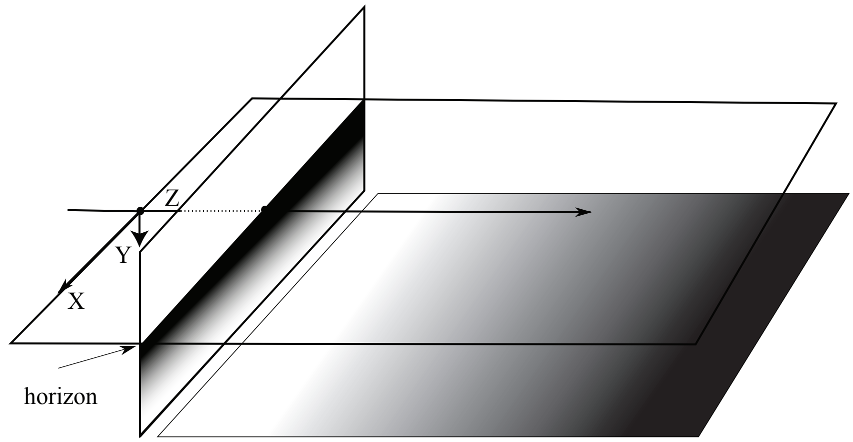


FIGURE 30.2: *The line at infinity is not just an abstraction: you can see it. Recall that, viewed in a perspective camera, planes have a horizon. This is the image of the line at infinity. For example, the plane $y = -1$ in the camera coordinate system has a horizon in the image as shown in the figure. The points on this plane can be written $(x, -1, z)$, and project to $(fx/z, -f/z, f)$. When z reaches infinity, you see the line $y = 0$ in the image plane. This is the image of the line at infinity.*

30.1.3 Lines on the Plane

Recall that a line on an affine plane is the set of points x, y such that $ax + by + c = 0$ for some a, b, c. Map the points on this line to homogeneous coordinates to get $(x, y, 1) = (X_1/X_3, X_2/X_3, 1) \equiv (X_1, X_2, X_3)$. If $ax + by + c = 0$, then $aX_1 + bX_2 + cX_3 = 0$ as well. Notice an interesting point here. The family of lines on a projective plane can be described by three homogenous coordinates, and so is itself a projective plane. This plane is *dual* to the original projective plane. You can interpret the equation of a line $\mathbf{a}^T\mathbf{X} = 0$ as either a description of all points that lie on the line with homogeneous coordinates $\mathbf{a}$ *or* as a description of all lines that pass through the point with homogeneous coordinates $\mathbf{X}$.

> **Remember this:** *A line on the projective plane is the set of points* $\mathbf{X}$ *such that*
> $$\mathbf{a}^T\mathbf{X} = 0.$$
> *Here* $\mathbf{a}$ *is a vector of homogeneous coordinates specifying the particular line.*

The intersection of two lines $\mathbf{a}_1$ and $\mathbf{a}_2$ is the point $\mathbf{X}$ such that $\mathbf{a}_1^T\mathbf{X} = 0$ and $\mathbf{a}_2^T\mathbf{X} = 0$. This is a point in homogeneous coordinates, and it always exists, even if the lines are parallel. On an affine plane, parallel lines do not intersect.

Example: On the affine plane, the two lines given by $x = 1$ and $x = 2$ do not intersect – they are parallel. Corresponding lines on the projective plane are $X_1 - X_3 = 0$ and $X_1 - 2X_3 = 0$. These two lines intersect at the point with homogeneous coordinates $(0, 1, 0)$.

Remember this: *Write* $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *for two points on the projective plane that are represented in homogeneous coordinates and are different. The line through these two points is given by*

$$\mathbf{a} = \mathbf{P}_1 \times \mathbf{P}_2$$

(check $\mathbf{a}^T\mathbf{P}_1 = 0$ *and* $\mathbf{a}^T\mathbf{P}_2 = 0$*). A parametrization of this line is given by*

$$U\mathbf{P}_1 + V\mathbf{P}_2.$$

30.1.4 Projective d-Spaces

Higher dimensional spaces follow the pattern above. In affine coordinates, a point in a d dimensional affine space (eg an *affine plane*; *affine 3D space*; etc) is given by d coordinates $(x_1, x_2, \ldots, x_d)$. The space described by $d + 1$ homogeneous coordinates is a *projective space*. A point $(x_1, x_2, \ldots, x_d)$ in a d dimensional affine space can be identified with $(X_1, X_2, \ldots, X_{d+1}) = \lambda(x_1, x_2, \ldots, x_d, 1)$ (for $\lambda \neq 0$) in the d dimensional projective space. The points in the projective space given by $(X_1, X_2, \ldots, 0)$ have no corresponding points in the affine space. Notice that this set of points is a $d - 1$ dimensional space in homogeneous coordinates.

When $d = 3$, the points "at infinity" form a projective plane which is known as the *plane at infinity*; the whole space is sometimes known as *projective 3-space***exercises** . Notice this means that 3D projective space is obtained by "sewing" a projective plane to the 3D affine space you are accustomed to. The piece of the projective space "at infinity" isn't special, using the same argument as above. The particular line (resp. plane) that is "at infinity" is chosen by the homogeneous coordinate you divide by. There is an established convention in computer vision of dividing by the last homogeneous coordinate and talking about the line at infinity and the plane at infinity.

Remember this: *The d dimensional space represented by $d+1$ homogeneous coordinates is a projective space. You can represent a point $(x_1, \ldots, x_d$ in affine d space in this projective space as $(x_1, \ldots, x_d, 1)$. Not every point in the projective space can be obtained like this – the points $(X_1, \ldots, X_d, 0)$ are "extra". These points form a projective $d-1$ space which is thought of as being "at infinity". Important cases are $d=1$ (often called the projective line, with a point at infinity); $d=2$ (the projective plane, with a line at infinity); and $d=3$ (projective space, with a plane at infinity).*

30.1.5 Planes in Projective 3-Space

Planes in projective 3-space work rather like lines on the projective plane. The locus of points (x, y, z) where $ax+by+cz+d=0$ is a plane in affine 3-space. Because (a, b, c, d) and $\lambda(a, b, c, d)$ give the same plane, (a, b, c, d) are homogeneous coordinates for a plane in 3D. You can write the points on the plane using homogeneous coordinates to get

$$(x, y, z, 1) = (X_1/X_4, X_2/X_4, X_3/X_4, 1)$$

or equivalently

$$(X_1, X_2, X_3, X_4) \text{ where } X_1 = xX_4, X_2 = yX_4, X_3 = zX_4.$$

Substitute to find the equation of the corresponding plane in projective 3-space $aX_1 + bX_2 + cX_3 + dX_4 = 0$ or $\mathbf{a}^T\mathbf{X} = 0$. A set of four homogenous coordinates can be used to describe either a point in projective 3-space or a plane in projective 3-space, so points in projective 3-space are dual to planes in projective 3-space.

You can interpret the equation of a plane $\mathbf{a}^T\mathbf{X} = 0$ as either a description of all points that lie on the plane with homogeneous coordinates $\mathbf{a}$ *or* as a description of all planes that pass through the point with homogeneous coordinates $\mathbf{X}$. Every pair of distinct planes in projective 3-space intersects in a unique line. The intersection of two planes $\mathbf{a}_1$ and $\mathbf{a}_2$ is a line, formed by the set of points $\mathbf{X}$ such that $\mathbf{a}_1^T\mathbf{X} = 0$ and $\mathbf{a}_2^T\mathbf{X} = 0$. Check that this is a line in homogeneous coordinates **exercises** . Check also that any three distinct planes intersect in a point **exercises** .

Remember this: *A plane in projective 3D is the set of points $\mathbf{X}$ such that*

$$\mathbf{a}^T\mathbf{X} = 0.$$

Here $\mathbf{a}$ is a vector of homogeneous coordinates specifying the particular plane.

Remember this: *Write $\mathbf{P}_1$, $\mathbf{P}_2$ and $\mathbf{P}_3$ for three points in projective 3D that are represented in homogeneous coordinates, are different points, and are not collinear. From the exercises, the plane through these points is given by*

$$\mathbf{a} = \textit{NullSpace}\left(\begin{bmatrix} \mathbf{P}_1^T \\ \mathbf{P}_2^T \\ \mathbf{P}_3^T \end{bmatrix}\right).$$

From the exercises, a parametrization of this plane is given by

$$U\mathbf{P}_1 + V\mathbf{P}_2 + W\mathbf{P}_3.$$

Example: Two distinct planes in projective 3-space intersect in a line. Write $\mathbf{a}_1^T\mathbf{X} = 0$ for the equation of the first plane, $\mathbf{a}_2^T\mathbf{X} = 0$ for the equation of the second. Then the line is the set of all points that cause both equations to vanish. Notice that many different pairs of planes will give the same line. As long as $a_{11}a_{22} - a_{21}a_{12} \neq 0$, the pair $a_{11}\mathbf{a}_1 + a_{12}\mathbf{a}_2$ and $a_{21}\mathbf{a}_1 + a_{22}\mathbf{a}_2$ specifies the same line as the pair $\mathbf{a}_1$, $\mathbf{a}_2$. **exercises**

Example: Three planes in projective 3-space could: be the same plane; lie on a shared line; or intersect in a single point. Write $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ for the coefficients of the three different planes. Then check that the common points of these planes are given by the null space of the 3×4 matrix

$$\begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{pmatrix}.$$

In the usual case, the null space is one dimensional, and so is a point (remember that scaling a set of homogeneous coordinates doesn't change the point they represent). If this matrix has a two dimensional null space, the planes share a line, and if it is three dimensional they are all the same plane.**exercises**

30.1.6 Homographies

Now assume you have a perspective camera viewing a plane in 3D. You parametrize this plane by (s, t), and the points on the plane are given by

$$\begin{pmatrix} X(s,t) \\ Y(s,t) \\ Z(s,t) \end{pmatrix} = \begin{pmatrix} a_{11}s + a_{12}t + a_{13} \\ a_{21}s + a_{22}t + a_{23} \\ a_{31}s + a_{32}t + a_{33} \end{pmatrix}$$

where $a_{11}, \ldots a_{33}$ are parameters that choose the plane and its parametrization. The perspective camera maps the point in 3D (X, Y, Z) to the point $(X/Z, Y/Z)$ on the image plane. This means that the point *on the plane* given by (s, t) is mapped to

$$\begin{pmatrix} \frac{X(s,t)}{Z(s,t)} \\ \\ \frac{Y(s,t)}{Z(s,t)} \end{pmatrix} = \begin{pmatrix} \frac{a_{11}s+a_{12}t+a_{13}}{a_{31}s+a_{32}t+a_{33}} \\ \\ \frac{a_{21}s+a_{22}t+a_{23}}{a_{31}s+a_{32}t+a_{33}} \end{pmatrix}$$

Now write this out in homogeneous coordinates. Write (S, T, U) for the coordinates on the world plane, where $S/U = s$ and $T/U = t$. Write (X, Y, Z) for the coordinates on the image plane. Then

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} S \\ T \\ U \end{pmatrix}.$$

This map is known as a *homography.* Recall that Section 21.3 showed how to fit a homography to a set of corresponding points. Now assume you see (say) an image of a tiled floor. With some simple correspondence information to support the fitting process, you can recover the image of the tiling *as it looks like from above* (Figure 21.3).

Any homography will map every line to a line. Write $\mathbf{a}$ for the line in the projective plane whose points satisfy $\mathbf{a}^T\mathbf{X} = 0$. Now apply the homography $\mathcal{M}$ to those points to get $\mathbf{V} = \mathcal{M}\mathbf{X}$. Notice that

$$\mathbf{a}^T\mathcal{M}^{(-1)}\mathbf{V} = \mathbf{a}^T\mathbf{X} = 0,$$

so that the line $\mathbf{a}$ transforms to the line $\mathcal{M}^{(-T)}\mathbf{a}$. Homographies are easily inverted.

Remember this: *A homography is a mapping from the projective plane to the projective plane. Assume $\mathcal{M}$ is a 3×3 matrix with non-zero determinant.*

- *The homography represented by $\mathcal{M}$ maps the point with homogeneous coordinates $\mathbf{X}$ to the point with homogeneous coordinates $\mathcal{M}\mathbf{X}$.*
- *The two matrices $\mathcal{M}$ and $\lambda\mathcal{M}$ represent the same homography.*
- *The inverse of this homography is represented by $\mathcal{M}^{-1}$.*
- *The homography represented by $\mathcal{M}$ will map the line represented by $\mathbf{a}$ to the line represented by $\mathcal{M}^{-T}\mathbf{a}$.*

30.1.7 Projective Transformations

Write $\mathbf{X} = (X_1, X_2, \ldots X_{d+1})$ for the coordinates of a point in projective d-space. Now consider $\mathbf{V} = \mathcal{M}\mathbf{X}$, where $\mathcal{M}$ is a $d + 1 \times k + 1$ matrix with non-zero determinant. You can interpret $\mathbf{V}$ as a point in projective d-space. In fact, $\mathcal{M}$ is a mapping from projective d-space to itself.

There is something to check here. Write $\mathcal{M}(\mathbf{X})$ for the point that $\mathbf{X}$ maps to, etc. Because $\mathbf{X} \equiv \lambda\mathbf{X}$ (for $\lambda \neq 0$), we must have that $\mathcal{M}(\mathbf{X}) \equiv \mathcal{M}(\lambda\mathbf{X})$ otherwise one point would map to several points. But

$$\mathcal{M}(\mathbf{X}) = \mathcal{M}\mathbf{X} \equiv \lambda\mathcal{M}\mathbf{X} = \mathcal{M}(\lambda\mathbf{X})$$

so $\mathcal{M}$ is a mapping. Such mappings are known as *projective transformations*. It should be pretty clear that this is a general version of a homography.

You should check that $\mathcal{M}^{(-1)}$ is the inverse of $\mathcal{M}$, and is a projective transformation. You should check that $\mathcal{M}$ and $\lambda\mathcal{M}$ represent the same projective transformation.

Remember this: *A projective transformation is a mapping from projective d-space to projective d-space. A projective transformation can be represented by $\mathcal{M}$, a $d+1 \times k+1$ matrix with non-zero determinant.*

- *The projective transformation represented by $\mathcal{M}$ maps the point with homogeneous coordinates $\mathbf{X}$ to the point with homogeneous coordinates $\mathcal{M}\mathbf{X}$.*
- *The two matrices $\mathcal{M}$ and $\lambda\mathcal{M}$ represent the same projective transformation.*
- *The inverse of this projective transformation is represented by $\mathcal{M}^{-1}$.*
- *The projective transformation represented by $\mathcal{M}$ will map the line represented by $\mathbf{a}$ to the line represented by $\mathcal{M}^{-T}\mathbf{a}$.*

30.2 CAMERA MATRICES AND TRANSFORMATIONS

30.2.1 Perspective and Orthographic Camera Matrices

A canonical perspective camera has its focal point at the origin, looks down the z-axis, and has its image plane at $z = 1$ (Figure 30.3). In affine coordinates, the camera mapping is $(x, y, z) \to (x/z, y/z)$. 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

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

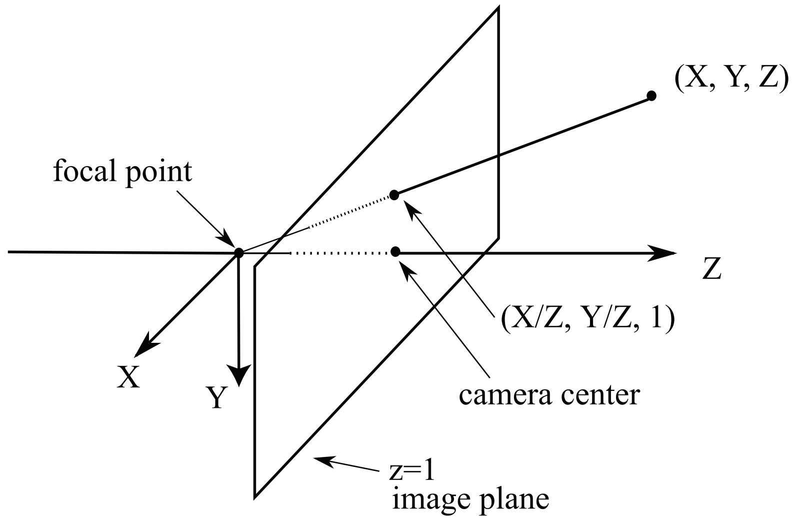


FIGURE 30.3: *A canonical pinhole camera is like the usual geometric abstraction, but has image plane $z = 1$.*

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 $\mathcal{C}_p$).

Remember this: *The perspective camera matrix is*

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

Recall the focal point of a camera cannot be imaged because you can't construct a unique line through the focal point and the focal point. The focal point of our camera is at $(0, 0, 0, T)$ in homogeneous coordinates (here $T \neq 0$). Notice that the perspective camera matrix maps this point to the point $(0, 0, 0)$ in homogeneous coordinates — but this point is meaningless. You should check no other point maps to $(0, 0, 0)$.

In affine coordinates, in the right coordinate system and assuming that the scale is chosen to be one, scaled orthographic perspective can be written as $(x, y, z) \rightarrow$

(x, y). Following the argument above, we obtain in homogeneous coordinates

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

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

Remember this: *The orthographic camera matrix is*

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

30.2.2 Cameras in World Coordinates

The camera matrix describes a perspective (resp. orthographic) projection for a camera in a specific coordinate system – the focal point is at the origin, the camera is looking backward down the z-axis, and so on. In the more general case, the camera is placed somewhere in world coordinates looking in some direction, and we need to account for this. Furthermore, the camera matrix assumes that points in the camera are reported in a specific coordinate system. The pixel locations reported by a practical camera might not be in that coordinate system. For example, many cameras place the origin at the top left hand corner. We need to account for this effect, too.

A general perspective camera transformation can be written as:

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

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.

30.2.3 Camera Extrinsic Parameters

The transformation $\mathcal{T}_e$ represents a rigid motion (equivalently, a *Euclidean transformation*, which consists of a 3D rotation and a 3D translation). In affine coordinates,

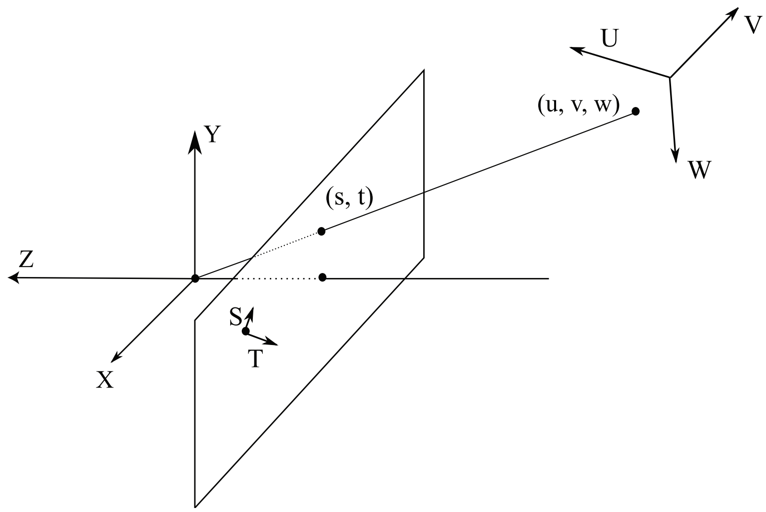


FIGURE 30.4: *A perspective camera (in its own coordinate system, given by X, Y and Z axes) views a point in world coordinates (given by (u, v, w) in the UVW coordinate system) and reports its position (s, t) in ST coordinates. The mapping from (u, v, w) to (s, t) consists of a transformation from the UVW coordinate system to the XYZ coordinate system, followed by a perspective projection, followed by a transformation to the ST coordinate system.*

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** .

30.2.4 Camera Intrinsic Parameters

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

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

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

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

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

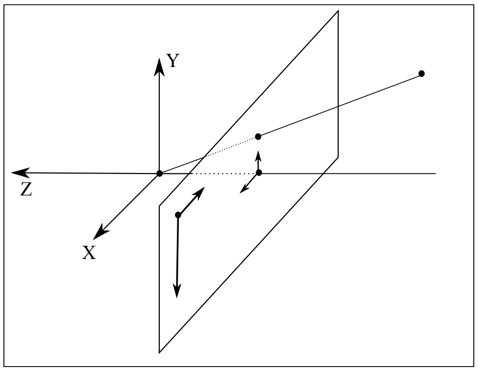
$$(\frac{X}{Z}, \frac{Y}{Z}, 1) \rightarrow (s_x\frac{X}{Z} + k\frac{Y}{Z} + c_x, s_y\frac{Y}{Z} + c_y, 1)$$

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.

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*.



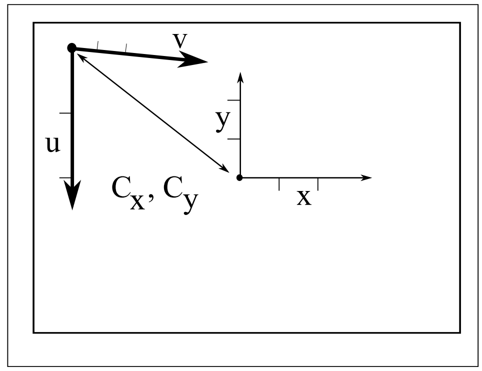


FIGURE 30.5: *The camera reports pixel values in pixel coordinates, which are not the same as world coordinates. The camera intrinsics represent the transformation between world coordinates and pixel coordinates. On the* **left***, a camera (as in Figure 27.1), with the camera coordinate system shown in heavy lines. On the* **right***, a more detailed view of the image plane. The camera coordinate axes are marked* (u, v) *and the image coordinate axes* (x, y)*. It is hard to determine* f *from the figure, and we will conflate scaling due to* f *with scaling resulting from the change to camera coordinates. The camera coordinate system's origin is not at the camera center, so* (c_x, c_y) *are not zero. I have marked unit steps in the coordinate system with ticks. Notice that the* v*-axis is not perpendicular to the* u*-axis (so* k *- the* skew *- is not zero). Ticks in the* u, v *axes are not the same distance apart as ticks on the* x, y *axes, meaning that* s *is not one. Furthermore,* u *ticks are further apart than* v *ticks, so that* a *is not one.*

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

$$\begin{aligned}
\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} &= \mathcal{T}_i \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \\
&= \begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}
\end{aligned}$$

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

By the arguments above, a general orthographic camera transformation can be written as:

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

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

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

30.2.5 From Camera Matrix to Intrinsics and Extrinsics

Section 30.2.2 states that a camera matrix $\mathcal{C}$ can be expressed as an intrinsic matrix times a canonical camera matrix times an extrinsic matrix. It is straightforward to recover $\mathcal{T}_i$ and $\mathcal{T}_e$ from $\mathcal{C}$.

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

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

30.3 YOU SHOULD

30.3.1 remember these facts:

30.3.2 be able to:

- Find the vanishing point for a set of lines in a perspective image.
- Find the horizon for a plane in a perspective image.
- Express a point in homogeneous coordinates.
- Derive a canonical perspective camera matrix.
- Derive an orthographic camera matrix.
- Explain the three terms (intrinsic, camera, extrinsic) that make up a general camera matrix.

EXERCISES

QUICK CHECKS

30.1. The coordinates of points on the projective line are (X_1, X_2). Show the transformation that takes (X_1, X_2) to (X_2, X_1) is one-to-one.

30.2. The coordinates of points on the projective line are (X_1, X_2). Show the transformation that takes (X_1, X_2) to (X_2, X_1) takes the point at infinity to the origin, and the origin to the point at infinity.

30.3. Section 30.1.2 says: " Furthermore, you can see the point at infinity (Figure 30.1)." Explain.

30.4. Confirm that the points "at infinity" in a projective 3 space form a projective plane (Section 30.1.4).

30.5. Check that two distinct planes $\mathbf{a}_1$ and $\mathbf{a}_2$ in projective 3 space *always* intersect in a projective line.

30.6. $\mathbf{a}_1$ and $\mathbf{a}_2$ are two planes in projective 3 space that intersect in a line. Show that, as long as $a_{11}a_{22} - a_{21}a_{12} \neq 0$, the pair $a_{11}\mathbf{a}_1 + a_{12}\mathbf{a}_2$ and $a_{21}\mathbf{a}_1 + a_{22}\mathbf{a}_2$ specifies the same line.

30.7. Check that three planes $\mathbf{a}_1$ and $\mathbf{a}_2$ in projective 3 space can intersect in: a plane; a line; or a point, depending on appropriate conditions. What are the conditions for each case?

30.8. Show that

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

represents a Euclidean transformation in homogenous coordinates *if* $\mathcal{R}$ is a rotation matrix.

30.9. Section 30.2.4 has: "This could only be a camera matrix if $\mathcal{M}$ had full rank." Explain.

30.10. Section 30.2.4 has: "This means that, if $\mathcal{T}_i$ is *not* upper triangular, an appropriate choice of rotation would make it upper triangular." Explain.

30.11. Given a 3×4 matrix $\mathcal{C}$ with rank 3, apply the procedure of 30.1 to obtain $\mathcal{T}_i$ and $\mathcal{T}_e$. Check that $\mathcal{C} \equiv \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e$.

30.12. Assume the imaging sensor is not perpendicular to the lens axis by a small amount. Why would this result in k not being zero? (**Hint:** think about the relationship between an imaginary sensor that is perpendicular to the lens axis, and the true sensor. This question isn't easy, but is quick.)

LONGER PROBLEMS

30.13. We show that any square matrix of full rank can be factored into a product of an upper triangular matrix and a rotation matrix. Start with a 2×2 matrix $\mathcal{M}$. Write

$$\mathcal{M} = \begin{bmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \end{bmatrix} \text{ and } \mathcal{U} = \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} \text{ and } \mathcal{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \end{bmatrix}$$

and assume $\mathcal{M} = \mathcal{U}\mathcal{R}$.

(a) Show that $u_{22} = \sqrt{\mathbf{m}_2^T \mathbf{m}_2}$ and $\mathbf{r}_2 = (1/u_{22})\mathbf{m}_2$. Show there are two possible choices for u_{22}, differing only by sign.

(b) Show that $\mathbf{m}_1^T \mathbf{r}_2 = u_{12}$, and that $(\mathbf{m}_1 - (\mathbf{m}_1^T \mathbf{r}_2)\mathbf{r}_2) = u_{11}\mathbf{r}_1$. Explain how to recover u_{11} and $\mathbf{r}_1$. Is there a sign ambiguity?

(c) What happens to this procedure of $\mathcal{M}$ does not have full rank?
(d) Show that any square matrix of full rank can be factored into a product of an upper triangular matrix and a rotation matrix, using induction on the dimension of the matrix.

30.14. We construct the vanishing point of a pair of parallel lines in homogeneous coordinates.
(a) Show that the set of points in homogeneous coordinates in 3D given by $(s, -s, t, s)$ (for s, t parameters) form a line in 3D.
(b) Now image the line $(s, -s, t, s)$ in 3D in a standard perspective camera with focal length 1. Show the result is the line $(s, -s, t)$ in the image plane.
(c) Now image the line $(-s, -s, t, s)$ in 3D in a standard perspective camera with focal length 1. Show the result is the line $(-s, -s, t)$ in the image plane.
(d) Show that the lines $(s, -s, t)$ and $(-s, -s, t)$ intersect in the point $(0, 0, t)$.

30.15. We construct the horizon of a plane for a standard perspective camera with focal length 1. Write $\mathbf{a} = [a_1, a_2, a_3, a_4]^T$ for the coefficients of the plane, so that for every point $\mathbf{X}$ on the plane we have $\mathbf{a}^T\mathbf{X} = 0$.
(a) Show that the plane given by $\mathbf{u} = [a_1, a_2, a_3, 0]$ is parallel to the plane given by $\mathbf{a}$, and passes through $(0, 0, 0, 1)$.
(b) Write the points on the image plane $\mathbf{x} = (x, y, 1) \equiv (X, Y, Z)$ in homogeneous coordinates. Show that the horizon of the plane is the set of points $\mathbf{x}$ in the image plane given by $\mathbf{l}^T\mathbf{x} = 0$, where $\mathbf{l} = [a_1, a_2, a_3]^T$.

30.16. A pinhole camera with focal point at the origin and image plane at $z = f$ views two parallel lines $\mathbf{u} + t\mathbf{w}$ and $\mathbf{v} + t\mathbf{w}$. Write $\mathbf{w} = [w_1, w_2, w_3]^T$, etc.
(a) Show that the vanishing point of these lines, on the image plane, is given by $(f\frac{w_1}{w_3}, f\frac{w_2}{w_3})$.
(b) Now we model a family of pairs of parallel lines, by writing $\mathbf{w}(r, s) = r\mathbf{a} + s\mathbf{b}$, for any (r, s). In this model, $\mathbf{u} + t\mathbf{w}(r, s)$ and $\mathbf{v} + t\mathbf{w}(r, s)$ are the pair of lines, and (r, s) chooses the direction. First, show that this family of vectors lies in a plane. Now show that the vanishing point for the (r, s)'th pair is $(f\frac{ra_1+sb_1}{ra_3+sb_3}, f\frac{ra_2+sb_2}{ra_3+sb_3})$.
(c) Show that the family of vanishing points $(f\frac{ra_1+sb_1}{ra_3+sb_3}, f\frac{ra_2+sb_2}{ra_3+sb_3})$ lies on a straight line in the image. Do this by constructing $\mathbf{c}$ such that $\mathbf{c}^T\mathbf{a} = \mathbf{c}^T\mathbf{b} = 0$. Now write $(x(r, s), y(r, s)) = (-f\frac{ra_1+sb_1}{ra_3+sb_3}, -f\frac{ra_2+sb_2}{ra_3+sb_3})$ and show that $c_1 x(r, s) + c_2 y(r, s) + c_3 = 0$.

30.17. All points on the projective plane with homogeneous coordinates $(U, V, 0)$ lie "at infinity" (divide by zero). As we have seen, these points form a projective line.
(a) Show this line is represented by the vector of coefficients $(0, 0, C)$.
(b) A homography $\mathcal{M} = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3]$ is applied to the projective plane. Show that the line whose coefficients are $\mathbf{m}_3$ maps to the line at infinity. (**Hint:** show that if you map the projective plane using the homography $\mathcal{M}$, then the line whose coefficients are $\mathbf{a}$ goes to the line whose coefficients are $\mathcal{M}^{-1}\mathbf{a}$.)
(c) Show that the homography $\mathcal{M}$ (above) maps the points at infinity to a line given in parametric form as $U\mathbf{m}_1 + V\mathbf{m}_2$.
(d) Now write $\mathbf{n}$ for a non-zero vector such that $\mathbf{n}^T\mathbf{m}_1 = \mathbf{n}^T\mathbf{m}_2 = 0$. Show that, for any point $\mathbf{x}$ on the line given in parametric form as $U\mathbf{m}_1 + V\mathbf{m}_2$, we have $\mathbf{n}^T\mathbf{x} = 0$. Is $\mathbf{n}$ unique?

(e) Use the results of the previous subexercises to show that for any given line, there are some homographies that map that line to the line at infinity.

(f) Use the results of the previous subexercises to show that for any given line, there are some homographies that map the line at infinity to that line.