C H A P T E R 29

Camera Matrices

29.1 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.

29.1.1 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 k dimensions is represented with k numbers. Now adopt the convention that a point in k dimensions is represented by $k+1$ 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

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

These coordinates are known as *homogeneous coordinates*, and will offer a particularly convenient representation of perspective projection.

> **Remember this:** *In homogeneous coordinates, a point in a k dimensional space is represented by $k+1$ coordinates $(X_1, \ldots, X_{k+1})$, together with the convention that*
>
> $$(X_1, \ldots, X_{k+1}) \equiv \lambda(X_1, \ldots, X_{k+1}) \text{ for } \lambda \neq 0.$$
>
> *The point $(0, 0, \ldots, 0)$ is meaningless in homogeneous coordinates.*

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

Example: 29.1 *Lines on the affine 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.

29.1.2 The projective line

In homogenous coordinates, we represent a point on a 1D space with two coordinates, so (X_1, X_2) (by convention, homogeneous coordinates are written with capital letters). Two sets of homogeneous coordinates (U_1, U_2) and (V_1, V_2) represent different points if there is no $\lambda \neq 0$ such that $\lambda(U_1, U_2) = (V_1, V_2)$. The set of all distinct points is known as a *projective line*. You should think of the projective line as an ordinary line (an *affine line*) with an "extra point". Every point on an affine line has a corresponding point on a projective line. A point on an affine line is given by a single coordinate x. This point can be identified with the point on a projective line given by $(X_1, X_2) = \lambda(x, 1)$ (for $\lambda \neq 0$) in homogeneous coordinates. The extra point has coordinates $(X_1, 0)$. These are the homogeneous coordinates of a single point (check this), but this point would be "at infinity" on the affine line.

There isn't anything special about the point on the projective line given by $(X_1, 0)$. You can see this by identifying the point x on the affine line with $(X_1, X_2) = \lambda(1, x)$ (for $\lambda \neq 0$). Now $(X_1, 0)$ is a point like any other, and $(0, X_2)$ is "at infinity". A little work establishes that there is a 1-1 mapping between the projective line and a circle (exercises).

29.1.3 The projective plane

The space represented by three homogeneous coordinates is known as a *projective plane*. 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*.

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

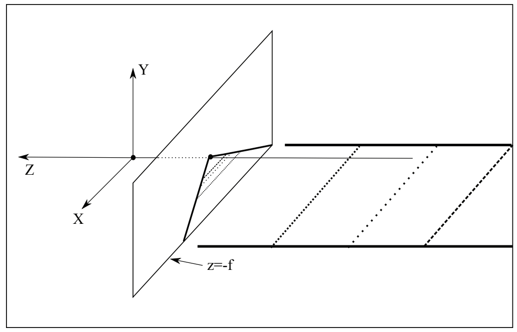

FIGURE 29.1: *The point at infinity is not just an abstraction: you can see it. Recall that lines that are parallel in the world can intersect in the image at a vanishing point. This vanishing point is the image of the point "at infinity" on the parallel lines. For example, on the plane $y = -1$ in the 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.*

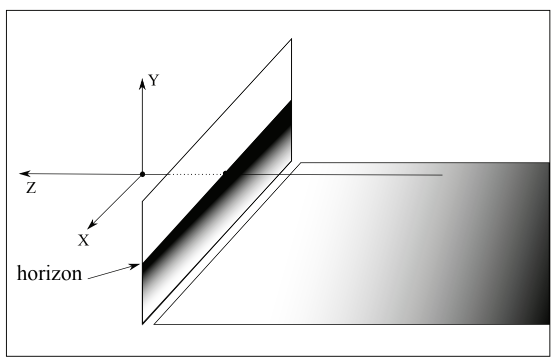

FIGURE 29.2: *The line at infinity is not just an abstraction: you can see it. Recall that, viewed in a perspective camera, planes have a horizon. This is the image of the line at infinity. For example, the plane $y = -1$ in the camera coordinate system has a horizon in the image as shown in the figure. The points on this plane can be written $(x, -1, z)$, and project to $(fx/z, -f/z, f)$. When z reaches infinity, we see the line $y = 0$ in the image plane. This is the image of the line at infinity.*

$aX_1 + bX_2 + cX_3 = 0$ as well.

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.*

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 $\mathbf{X}$.

Projective planes are unlike affine planes in one important respect. On a projective plane, every pair of distinct lines on the projective plane intersects in a unique point (on an affine plane, parallel lines do not intersect). 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.

Example: 29.2 *Where parallel lines intersect*

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*

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

29.1.4 Projective k-Spaces

Higher dimensional spaces follow the pattern above. In affine coordinates, a point in a k dimensional affine space (eg an *affine plane*; *affine 3D space*; etc) is given by k coordinates $(x_1, x_2, \ldots, x_k)$. The space described by $k+1$ homogeneous coordinates is a *projective space*. A point $(x_1, x_2, \ldots, x_k)$ in a k dimensional affine space can be identified with $(X_1, X_2, \ldots, X_{k+1}) = \lambda(x_1, x_2, \ldots, x_k, 1)$ (for $\lambda \neq 0$) in the k 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 $k-1$ dimensional space in homogeneous coordinates.

When $k = 3$, the points "at infinity" form a projective plane, and is known as the *plane at infinity*; the whole space is sometimes known as *projective 3-space*. Notice this means that 3D projective space is obtained by "sewing" a projective plane to the 3D affine space we 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 k dimensional space represented by $k+1$ homogeneous coordinates is a projective space. You can represent a point $(x_1, \ldots, x_k$ in affine k space in this projective space as $(x_1, \ldots, x_k, 1)$. Not every point in the projective space can be obtained like this – the points $(X_1, \ldots, X_k, 0)$ are "extra". These points form a projective $k-1$ space which is thought of as being "at infinity". Important cases are $k = 1$ (the projective line with a point at infinity); $k = 2$ (the projective plane with a line at infinity).*

29.1.5 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, we have that (a, b, c, d) are homogeneous coordinates for a plane in 3D. We 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 line $\mathbf{a}^T\mathbf{X} = 0$ 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. Check also that any three distinct planes intersect in a point.

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*

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

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

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

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 $\mathbf{a}_1$, $\mathbf{a}_2$.

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.

29.1.6 Homographies

Now assume we have a perspective camera viewing a plane in 3D. We 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 we 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 we have

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

This map is known as a *homography*. Recall that in Section 41.2, we 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 41.2).

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 $\mathcal{M}^{-T}\mathbf{a}$.*

29.1.7 Projective Transformations

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

29.2 CAMERA MATRICES AND TRANSFORMATIONS

29.2.1 Perspective and Orthographic Camera Matrices

In affine coordinates we wrote perspective projection as $(X, Y, Z) \to (X/Z, Y/Z)$ (remember, we will account for f later). This was in a left-hand coordinate system, which is a natural way to think of a camera (z increases as you move into the image) but is inconvenient otherwise. In a right-hand coordinate system, we flip the direction of the z axis, which yields the mapping $(X, Y, Z) \to (-X/Z, -Y/Z)$. Notice that doing so simply transforms the image *in the image plane*. As long as we remember to apply the minus signs, we can continue to work in our left-hand coordinate system.

In affine coordinates, the camera mapping is $(X, Y, Z) \to (X/Z, Y/Z)$ (remember: we account for f and the right-hand coordinate system 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

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

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 $\mathcal{C}_p$).

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*

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

29.2.2 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.

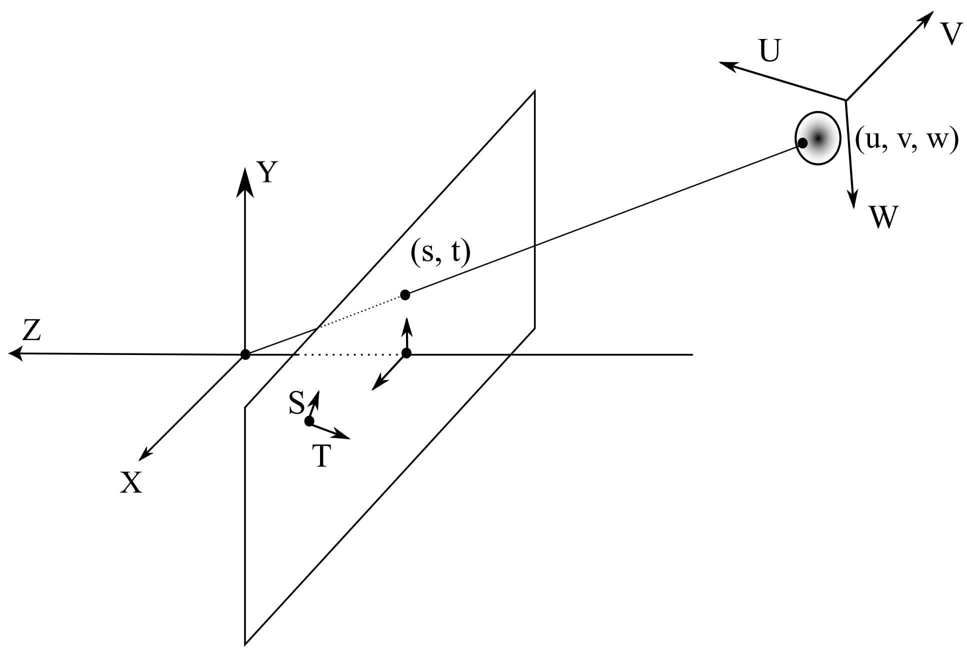

FIGURE 29.3: *A perspective camera (in its own coordinate system, given by X, Y and Z axes) views a point in world coordinates (given by (u, v, w) in the UVW coordinate system) and reports the position of points in ST coordinates. We must model the mapping from (u, v, w) to (s, t), which consists of 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.*

A general perspective camera transformation can be written as:

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

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

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.

29.2.3 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, 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. The exercises explore some properties of $\mathcal{T}_e$.

29.2.4 Camera Intrinsic Parameters

Camera intrinsic parameters must model a possible coordinate transformation in the image plane from projected world coordinates (write (x, y)) to pixel coordinates (write (u, v)), together with a possible change of focal length. This change is caused by the image plane being further away from, or closer to, the focal point. The coordinate transformation is not arbitrary (Figure 29.4).

Typically, the origin of the pixel coordinates is usually not at the camera center; instead, the camera center is at c'_x, c'_y in pixel coordinates. Write Δx for the step in the image plane from pixel (i, j) to $(i + 1, j)$ and Δy for the step to $(i, j + 1)$. These are vectors parallel to the camera coordinate axes. The vector Δx may not be perpendicular to the vector Δy, causing *skew*. For many cameras, $\| \Delta x \|$ is different from $\| \Delta y \|$ – such cameras have *non-square pixels*, and $\| \Delta x \| / \| \Delta y \|$ is known as the *aspect ratio* of the pixel. Furthermore, $\| \Delta x \|$ is not usually one unit in world coordinates. Finally, we have to apply the minus sign inherited from using a left-handed coordinate system.

There is one tricky point here. Rotating the world about the Z axis has an effect equivalent to rotating the camera coordinate system (Figure **??**). This means we cannot tell whether this rotation is the result of a change in the extrinsics (the world rotated) or the intrinsics (the camera coordinate system rotated). By convention, there is no rotation in the intrinsics *except* for the 180^o rotation required to handle the left-handed coordinate system, so a pure rotation of the image is always the result of the world rotating.

There are two possible parametrizations of camera intrinsics. Recall f is the focal length of the camera. Write (c'_x, c'_y) for the location of the camera center in pixel coordinates; a for the aspect ratio of the pixels ; and k' for the skew. Then $\mathcal{T}_i$ is parametrized as

$$\begin{bmatrix} -\|\Delta x\| & k' & c'_x \\ 0 & -\|\Delta y\| & c'_y \\ 0 & 0 & 1/f \end{bmatrix} \equiv \begin{bmatrix} -af\|\Delta y\| & fk' & fc'_x \\ 0 & -f\|\Delta y\| & fc'_y \\ 0 & 0 & 1 \end{bmatrix}$$

Notice in this case we are distinguishing between scaling resulting from $\|\Delta y\|$ and scaling resulting from the focal length. This is unusual, but can occur. More usual is to conflate these effects and parametrize the intrinsics as

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

where $s = f\|\Delta y\|$, $a = \|\Delta x\|\Delta y$, $k = fk'$, $c_x = fc'_x$, $c_y = fc'_y$.

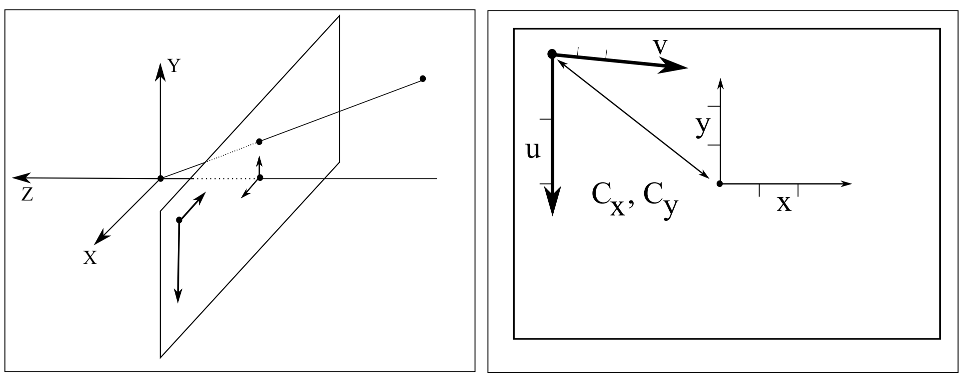

FIGURE 29.4: *The camera reports pixel values in pixel coordinates, which are not the same as world coordinates. The camera intrinsics represent the transformation between world coordinates and pixel coordinates. On the* **left***, a camera (as in Figure 26.1), with the camera coordinate system shown 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:*

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

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

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

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

$$\mathcal{K}\left[\mathcal{R}\mathbf{t}\right]$$

29.2.5 Cameras and Image to Image Mapping

PROBLEMS

29.1. 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 $(0, 0, t)$.

29.2. 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 $(u, v, 1) \equiv (U, V, W)$ in homogeneous coordinates. Show that the horizon of the plane is the set of points $\mathbf{u}$ in the image plane given by $\mathbf{l}^T\mathbf{u} = 0$, where $\mathbf{l} = [a_1, a_2, a_3]^T$.

29.3. 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 $c_1x(r, s) + c_2y(r, s) + c_3 = 0$.

29.4. 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 $\mathcal{M} = \left[\mathbf{m}_1^T; \mathbf{m}_2^T; \mathbf{m}_3^T\right]$ is applied to the projective plane. Show that the line whose coefficients are $\mathbf{v}_3$ maps to the line at infinity.

(c) Now write the homography as $\mathcal{M} = \left[\mathbf{m}'_1, \mathbf{m}'_2, \mathbf{m}'_3\right]$ (so $\mathbf{m}'$ are columns). Show that the homography maps the points at infinity to a line given in parametric form as $s\mathbf{m}'_1 + t\mathbf{m}'_2$.

(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 $s\mathbf{m}'_1 + t\mathbf{m}'_2$, 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.

29.5. We will show that there is no significant difference between choosing a right-handed camera coordinate system and a left-handed camera coordinate system. Notice that, in a right handed camera coordinate system (where the camera looks down the negative z-axis rather than the positive z-axis) the image plane is at $z = -f$.

(a) Show that, in a right-handed coordinate system, a pinhole camera maps

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

(b) Show that the argument in the text yields a camera matrix of the form

$$\mathcal{C}'_p = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/f & 0 \end{bmatrix}.$$

(c) Show that, if one allows the scale in $\mathcal{T}_i$ to be negative, one could still use

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

as a camera matrix.