6

Camera Models

A camera is a mapping between the 3D world (object space) and a 2D image. The principal camera of interest in this book is *central projection*. This chapter develops a number of camera *models* which are matrices with particular properties that represent the camera mapping.

It will be seen that all cameras modelling central projection are specializations of the *general projective camera*. The anatomy of this most general camera model is examined using the tools of projective geometry. It will be seen that geometric entities of the camera, such as the projection centre and image plane, can be computed quite simply from its matrix representation. Specializations of the general projective camera inherit its properties, for example their geometry is computed using the same algebraic expressions.

The specialized models fall into two major classes – those that model cameras with a finite centre, and those that model cameras with centre "at infinity". Of the cameras at infinity the *affine camera* is of particular importance because it is the natural generalization of parallel projection.

This chapter is principally concerned with the projection of points. The action of a camera on other geometric entities, such as lines, is deferred until chapter 8.

6.1 Finite cameras

In this section we start with the most specialized and simplest camera model, which is the basic pinhole camera, and then progressively generalize this model through a series of gradations.

The models we develop are principally designed for CCD like sensors, but are also applicable to other cameras, for example X-ray images, scanned photographic negatives, scanned photographs from enlarged negatives, etc.

The basic pinhole model. We consider the central projection of points in space onto a plane. Let the centre of projection be the origin of a Euclidean coordinate system, and consider the plane $\mathrm{Z} = f$, which is called the *image plane* or *focal plane*. Under the pinhole camera model, a point in space with coordinates $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z})^\mathsf{T}$ is mapped to the point on the image plane where a line joining the point $\mathbf{X}$ to the centre of projection meets the image plane. This is shown in figure 6.1. By similar triangles, one quickly

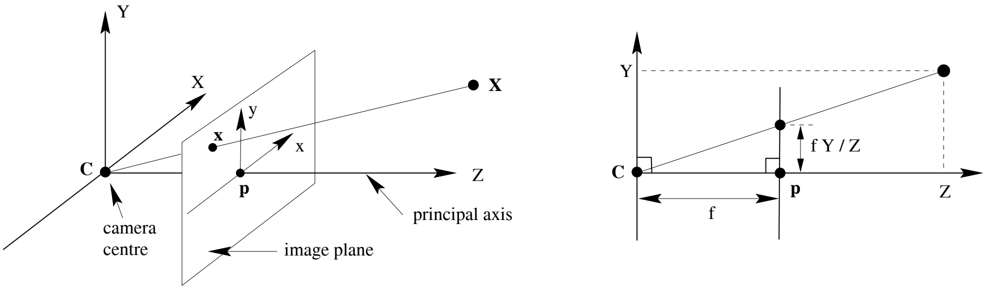

Fig. 6.1. **Pinhole camera geometry.** $\mathbf{C}$ *is the camera centre and* $\mathbf{p}$ *the principal point. The camera centre is here placed at the coordinate origin. Note the image plane is placed in front of the camera centre.*

computes that the point $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})^{\mathsf{T}}$ is mapped to the point $(f\mathrm{X}/\mathrm{Z}, f\mathrm{Y}/\mathrm{Z}, f)^{\mathsf{T}}$ on the image plane. Ignoring the final image coordinate, we see that

$$(\mathrm{X}, \mathrm{Y}, \mathrm{Z})^{\mathsf{T}} \mapsto (f\mathrm{X}/\mathrm{Z}, f\mathrm{Y}/\mathrm{Z})^{\mathsf{T}} \tag{6.1}$$

describes the central projection mapping from world to image coordinates. This is a mapping from Euclidean 3-space $\mathbb{R}^3$ to Euclidean 2-space $\mathbb{R}^2$.

The centre of projection is called the *camera centre*. It is also known as the *optical centre*. The line from the camera centre perpendicular to the image plane is called the *principal axis* or *principal ray* of the camera, and the point where the principal axis meets the image plane is called the *principal point*. The plane through the camera centre parallel to the image plane is called the *principal plane* of the camera.

Central projection using homogeneous coordinates. If the world and image points are represented by homogeneous vectors, then central projection is very simply expressed as a linear mapping between their homogeneous coordinates. In particular, (6.1) may be written in terms of matrix multiplication as

$$\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f\mathrm{X} \\ f\mathrm{Y} \\ \mathrm{Z} \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix}. \tag{6.2}$$

The matrix in this expression may be written as $\mathrm{diag}(f, f, 1)[\mathtt{I} \mid \mathbf{0}]$ where $\mathrm{diag}(f, f, 1)$ is a diagonal matrix and $[\mathtt{I} \mid \mathbf{0}]$ represents a matrix divided up into a 3×3 block (the identity matrix) plus a column vector, here the zero vector.

We now introduce the notation $\mathbf{X}$ for the world point represented by the homogeneous 4-vector $(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^{\mathsf{T}}$, $\mathbf{x}$ for the image point represented by a homogeneous 3-vector, and $\mathtt{P}$ for the 3×4 homogeneous *camera projection matrix*. Then (6.2) is written compactly as

$$\mathbf{x} = \mathtt{P}\mathbf{X}$$

which defines the camera matrix for the pinhole model of central projection as

$$\mathtt{P} = \mathrm{diag}(f, f, 1)\,[\mathtt{I} \mid \mathbf{0}].$$

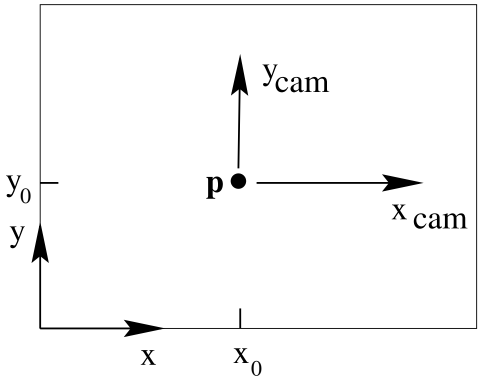

Fig. 6.2. *Image* (x, y) *and camera* $(x_{\text{cam}}, y_{\text{cam}})$ *coordinate systems.*

Principal point offset. The expression (6.1) assumed that the origin of coordinates in the image plane is at the principal point. In practice, it may not be, so that in general there is a mapping

$$(\mathrm{X}, \mathrm{Y}, \mathrm{Z})^{\mathsf{T}} \mapsto (f\mathrm{X}/\mathrm{Z} + p_x, f\mathrm{Y}/\mathrm{Z} + p_y)^{\mathsf{T}}$$

where $(p_x, p_y)^{\mathsf{T}}$ are the coordinates of the principal point. See figure 6.2. This equation may be expressed conveniently in homogeneous coordinates as

$$\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f\mathrm{X} + \mathrm{Z}p_x \\ f\mathrm{Y} + \mathrm{Z}p_y \\ \mathrm{Z} \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix}. \tag{6.3}$$

Now, writing

$$\mathtt{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \tag{6.4}$$

then (6.3) has the concise form

$$\mathbf{x} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]\mathbf{X}_{\text{cam}}. \tag{6.5}$$

The matrix $\mathtt{K}$ is called the *camera calibration matrix*. In (6.5) we have written $(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^{\mathsf{T}}$ as $\mathbf{X}_{\text{cam}}$ to emphasize that the camera is assumed to be located at the origin of a Euclidean coordinate system with the principal axis of the camera pointing straight down the Z-axis, and the point $\mathbf{X}_{\text{cam}}$ is expressed in this coordinate system. Such a coordinate system may be called the *camera coordinate frame*.

Camera rotation and translation. In general, points in space will be expressed in terms of a different Euclidean coordinate frame, known as the *world coordinate frame*. The two coordinate frames are related via a rotation and a translation. See figure 6.3. If $\widetilde{\mathbf{X}}$ is an inhomogeneous 3-vector representing the coordinates of a point in the world coordinate frame, and $\widetilde{\mathbf{X}}_{\text{cam}}$ represents the same point in the camera coordinate frame, then we may write $\widetilde{\mathbf{X}}_{\text{cam}} = \mathtt{R}(\widetilde{\mathbf{X}} - \widetilde{\mathbf{C}})$, where $\widetilde{\mathbf{C}}$ represents the coordinates of the camera

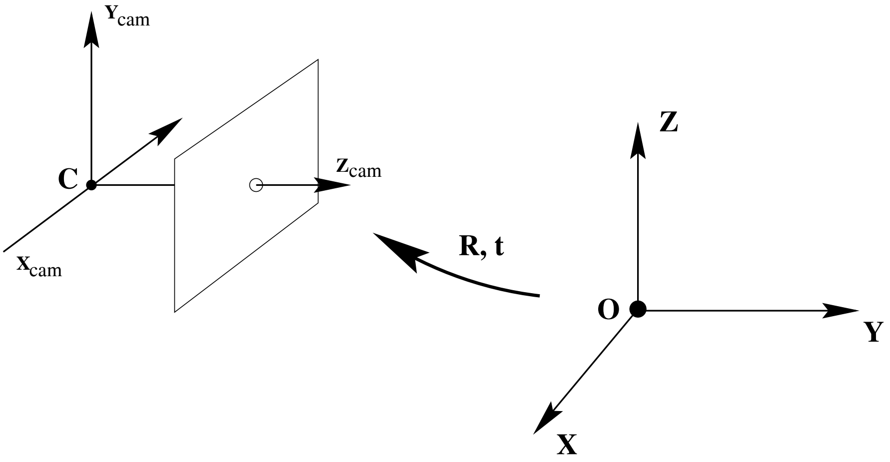

Fig. 6.3. *The Euclidean transformation between the world and camera coordinate frames.*

centre in the world coordinate frame, and $\mathtt{R}$ is a 3×3 rotation matrix representing the orientation of the camera coordinate frame. This equation may be written in homogeneous coordinates as

$$\mathbf{X}_{\text{cam}} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\widetilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\widetilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \mathbf{X}. \tag{6.6}$$

Putting this together with (6.5) leads to the formula

$$\mathbf{x} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]\mathbf{X} \tag{6.7}$$

where $\mathbf{X}$ is now in a world coordinate frame. This is the general mapping given by a pinhole camera. One sees that a general pinhole camera, $\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]$, has 9 degrees of freedom: 3 for $\mathtt{K}$ (the elements f, p_x, p_y), 3 for $\mathtt{R}$, and 3 for $\widetilde{\mathbf{C}}$. The parameters contained in $\mathtt{K}$ are called the *internal* camera parameters, or the *internal orientation* of the camera. The parameters of $\mathtt{R}$ and $\widetilde{\mathbf{C}}$ which relate the camera orientation and position to a world coordinate system are called the *external* parameters or the *exterior orientation*.

It is often convenient not to make the camera centre explicit, and instead to represent the world to image transformation as $\widetilde{\mathbf{X}}_{\text{cam}} = \mathtt{R}\widetilde{\mathbf{X}} + \mathbf{t}$. In this case the camera matrix is simply

$$\mathtt{P} = \mathtt{K}[\mathtt{R} \mid \mathbf{t}] \tag{6.8}$$

where from (6.7) $\mathbf{t} = -\mathtt{R}\widetilde{\mathbf{C}}$.

CCD cameras. The pinhole camera model just derived assumes that the image coordinates are Euclidean coordinates having equal scales in both axial directions. In the case of CCD cameras, there is the additional possibility of having non-square pixels. If image coordinates are measured in pixels, then this has the extra effect of introducing unequal scale factors in each direction. In particular if the number of pixels per unit distance in image coordinates are m_x and m_y in the x and y directions, then the transformation from world coordinates to pixel coordinates is obtained by multiplying (6.4) on the left by an extra factor diag$(m_x, m_y, 1)$. Thus, the general form of the calibration matrix of a CCD camera is

$$\mathtt{K} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \tag{6.9}$$

where $\alpha_x = fm_x$ and $\alpha_y = fm_y$ represent the focal length of the camera in terms of pixel dimensions in the x and y direction respectively. Similarly, $\tilde{\mathbf{x}}_0 = (x_0, y_0)$ is the principal point in terms of pixel dimensions, with coordinates $x_0 = m_x p_x$ and $y_0 = m_y p_y$. A CCD camera thus has 10 degrees of freedom.

Finite projective camera. For added generality, we can consider a calibration matrix of the form

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}. \tag{6.10}$$

The added parameter s is referred to as the *skew* parameter. The skew parameter will be zero for most normal cameras. However, in certain unusual instances which are described in section 6.2.4, it can take non-zero values.

A camera

$$\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}] \tag{6.11}$$

for which the calibration matrix $\mathtt{K}$ is of the form (6.10) will be called a *finite projective* camera. A finite projective camera has 11 degrees of freedom. This is the same number of degrees of freedom as a 3×4 matrix, defined up to an arbitrary scale.

Note that the left hand 3×3 submatrix of $\mathtt{P}$, equal to $\mathtt{KR}$, is non-singular. Conversely, any 3×4 matrix $\mathtt{P}$ for which the left hand 3×3 submatrix is non-singular is the camera matrix of some finite projective camera, because $\mathtt{P}$ can be decomposed as $\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]$. Indeed, letting $\mathtt{M}$ be the left 3×3 submatrix of $\mathtt{P}$, one decomposes $\mathtt{M}$ as a product $\mathtt{M} = \mathtt{KR}$ where $\mathtt{K}$ is upper-triangular of the form (6.10) and $\mathtt{R}$ is a rotation matrix. This decomposition is essentially the RQ matrix decomposition, described in section A4.1.1(*p*579), of which more will be said in section 6.2.4. The matrix $\mathtt{P}$ can therefore be written $\mathtt{P} = \mathtt{M}[\mathtt{I} \mid \mathtt{M}^{-1}\mathbf{p}_4] = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]$ where $\mathbf{p}_4$ is the last column of $\mathtt{P}$. In short

- *The set of camera matrices of finite projective cameras is identical with the set of homogeneous* 3×4 *matrices for which the left hand* 3×3 *submatrix is non-singular.*

General projective cameras. The final step in our hierarchy of projective cameras is to remove the non-singularity restriction on the left hand 3×3 submatrix. A *general projective* camera is one represented by an arbitrary homogeneous 3×4 matrix of rank 3. It has 11 degrees of freedom. The rank 3 requirement arises because if the rank is

Camera centre. The camera centre is the 1-dimensional right null-space $\mathbf{C}$ of $\mathtt{P}$, i.e. $\mathtt{P}\mathbf{C} = \mathbf{0}$.

- ◇ **Finite camera** ($\mathtt{M}$ is not singular) $\mathbf{C} = \begin{pmatrix} -\mathtt{M}^{-1}\mathbf{p}_4 \\ 1 \end{pmatrix}$
- ◇ **Camera at infinity** ($\mathtt{M}$ is singular) $\mathbf{C} = \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix}$ where $\mathbf{d}$ is the null 3-vector of $\mathtt{M}$, i.e. $\mathtt{M}\mathbf{d} = \mathbf{0}$.

Column points. For $i = 1, \ldots, 3$, the column vectors $\mathbf{p}_i$ are vanishing points in the image corresponding to the X, Y and Z axes respectively. Column $\mathbf{p}_4$ is the image of the coordinate origin.

Principal plane. The principal plane of the camera is $\mathbf{P}^3$, the last row of $\mathtt{P}$.

Axis planes. The planes $\mathbf{P}^1$ and $\mathbf{P}^2$ (the first and second rows of $\mathtt{P}$) represent planes in space through the camera centre, corresponding to points that map to the image lines $x = 0$ and $y = 0$ respectively.

Principal point. The image point $\mathbf{x}_0 = \mathtt{M}\mathbf{m}^3$ is the principal point of the camera, where $\mathbf{m}^{3\mathsf{T}}$ is the third row of $\mathtt{M}$.

Principal ray. The principal ray (axis) of the camera is the ray passing through the camera centre $\mathbf{C}$ with direction vector $\mathbf{m}^{3\mathsf{T}}$. The principal axis vector $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3$ is directed towards the front of the camera.

Table 6.1. *Summary of the properties of a projective camera* $\mathtt{P}$. *The matrix is represented by the block form* $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$.

less than this then the range of the matrix mapping will be a line or point and not the whole plane; in other words not a 2D image.

6.2 The projective camera

A general projective camera $\mathtt{P}$ maps world points $\mathbf{X}$ to image points $\mathbf{x}$ according to $\mathbf{x} = \mathtt{P}\mathbf{X}$. Building on this mapping we will now dissect the camera model to reveal how geometric entities, such as the camera centre, are encoded. Some of the properties that we consider will apply only to finite projective cameras and their specializations, whilst others will apply to general cameras. The distinction will be evident from the context. The derived properties of the camera are summarized in table 6.1.

6.2.1 Camera anatomy

A general projective camera may be decomposed into blocks according to $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$, where $\mathtt{M}$ is a 3×3 matrix. It will be seen that if $\mathtt{M}$ is non-singular, then this is a finite camera, otherwise it is not.

Camera centre. The matrix $\mathtt{P}$ has a 1-dimensional right null-space because its rank is 3, whereas it has 4 columns. Suppose the null-space is generated by the 4-vector $\mathbf{C}$, that is $\mathtt{P}\mathbf{C} = \mathbf{0}$. It will now be shown that $\mathbf{C}$ is the camera centre, represented as a homogeneous 4-vector.

Consider the line containing $\mathbf{C}$ and any other point $\mathbf{A}$ in 3-space. Points on this line may be represented by the join

$$\mathbf{X}(\lambda) = \lambda\mathbf{A} + (1 - \lambda)\mathbf{C} \ .$$

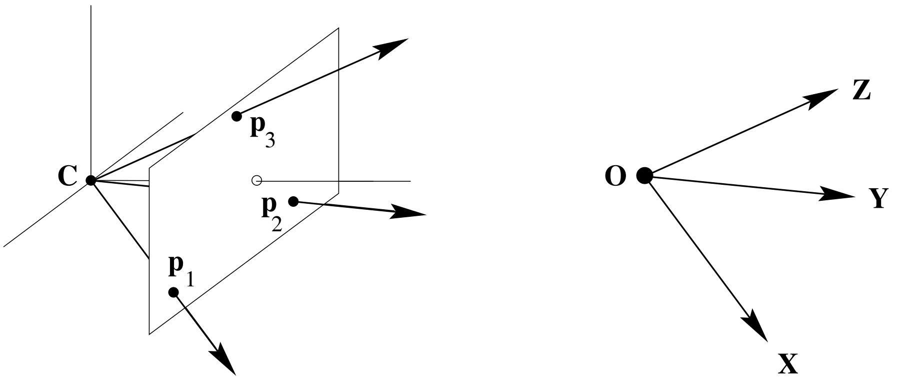

Fig. 6.4. *The three image points defined by the columns $\mathbf{p}_i, i = 1, \ldots, 3$, of the projection matrix are the vanishing points of the directions of the world axes.*

Under the mapping $\mathbf{x} = \mathtt{P}\mathbf{X}$ points on this line are projected to

$$\mathbf{x} = \mathtt{P}\mathbf{X}(\lambda) = \lambda \mathtt{P}\mathbf{A} + (1 - \lambda)\mathtt{P}\mathbf{C} = \lambda \mathtt{P}\mathbf{A}$$

since $\mathtt{P}\mathbf{C} = \mathbf{0}$. That is all points on the line are mapped to the same image point $\mathtt{P}\mathbf{A}$, which means that the line must be a ray through the camera centre. It follows that $\mathbf{C}$ is the homogeneous representation of the camera centre, since for all choices of $\mathbf{A}$ the line $\mathbf{X}(\lambda)$ is a ray through the camera centre.

This result is not unexpected since the image point $(0, 0, 0)^{\mathsf{T}} = \mathtt{P}\mathbf{C}$ is not defined, and the camera centre is the unique point in space for which the image is undefined. In the case of finite cameras the result may be established directly, since $\mathbf{C} = (\widetilde{\mathbf{C}}^{\mathsf{T}}, 1)^{\mathsf{T}}$ is clearly the null-vector of $\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]$. The result is true even in the case where the first 3×3 submatrix $\mathtt{M}$ of $\mathtt{P}$ is singular. In this singular case, though, the null-vector has the form $\mathbf{C} = (\mathbf{d}^{\mathsf{T}}, 0)^{\mathsf{T}}$ where $\mathtt{M}\mathbf{d} = \mathbf{0}$. The camera centre is then a point at infinity. Camera models of this class are discussed in section 6.3.

Column vectors. The columns of the projective camera are 3-vectors which have a geometric meaning as particular image points. With the notation that the columns of $\mathtt{P}$ are $\mathbf{p}_i, i = 1, \ldots, 4$, then $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ are the vanishing points of the world coordinate $\mathtt{X}$, $\mathtt{Y}$ and $\mathtt{Z}$ axes respectively. This follows because these points are the images of the axes' directions. For example the x-axis has direction $\mathbf{D} = (1, 0, 0, 0)^{\mathsf{T}}$, which is imaged at $\mathbf{p}_1 = \mathtt{P}\mathbf{D}$. See figure 6.4. The column $\mathbf{p}_4$ is the image of the world origin.

Row vectors. The rows of the projective camera (6.12) are 4-vectors which may be interpreted geometrically as particular world planes. These planes are examined next. We introduce the notation that the rows of $\mathtt{P}$ are $\mathbf{P}^{i\mathsf{T}}$ so that

$$\mathtt{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} \mathbf{P}^{1\mathsf{T}} \\ \mathbf{P}^{2\mathsf{T}} \\ \mathbf{P}^{3\mathsf{T}} \end{bmatrix}. \qquad (6.12)$$

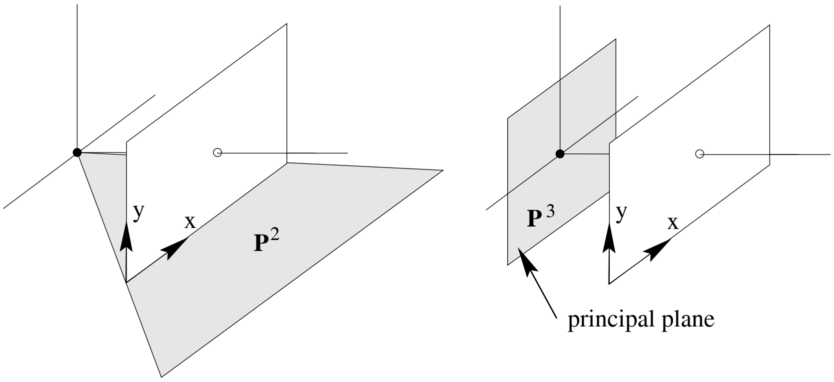

Fig. 6.5. *Two of the three planes defined by the rows of the projection matrix.*

The principal plane. The principal plane is the plane through the camera centre parallel to the image plane. It consists of the set of points $\mathbf{X}$ which are imaged on the line at infinity of the image. Explicitly, $\mathtt{P}\mathbf{X} = (x, y, 0)^\mathsf{T}$. Thus a point lies on the principal plane of the camera if and only if $\mathbf{P}^{3\mathsf{T}}\mathbf{X} = 0$. In other words, $\mathbf{P}^3$ is the vector representing the principal plane of the camera. If $\mathbf{C}$ is the camera centre, then $\mathtt{P}\mathbf{C} = \mathbf{0}$, and so in particular $\mathbf{P}^{3\mathsf{T}}\mathbf{C} = 0$. That is $\mathbf{C}$ lies on the principal plane of the camera.

Axis planes. Consider the set of points $\mathbf{X}$ on the plane $\mathbf{P}^1$. This set satisfies $\mathbf{P}^{1\mathsf{T}}\mathbf{X} = 0$, and so is imaged at $\mathtt{P}\mathbf{X} = (0, y, w)^\mathsf{T}$ which are points on the image y-axis. Again it follows from $\mathtt{P}\mathbf{C} = \mathbf{0}$ that $\mathbf{P}^{1\mathsf{T}}\mathbf{C} = 0$ and so $\mathbf{C}$ lies also on the plane $\mathbf{P}^1$. Consequently the plane $\mathbf{P}^1$ is defined by the camera centre and the line $x = 0$ in the image. Similarly the plane $\mathbf{P}^2$ is defined by the camera centre and the line $y = 0$.

Unlike the principal plane $\mathbf{P}^3$, the axis planes $\mathbf{P}^1$ and $\mathbf{P}^2$ are dependent on the image x- and y-axes, i.e. on the choice of the image coordinate system. Thus they are less tightly coupled to the natural camera geometry than the principal plane. In particular the line of intersection of the planes $\mathbf{P}^1$ and $\mathbf{P}^2$ is a line joining the camera centre and image origin, i.e. the back-projection of the image origin. This line will not coincide in general with the camera principal axis. The planes arising from $\mathbf{P}^i$ are illustrated in figure 6.5.

The camera centre $\mathbf{C}$ lies on all three planes, and since these planes are distinct (as the $\mathtt{P}$ matrix has rank 3) it must lie on their intersection. Algebraically, the condition for the centre to lie on all three planes is $\mathtt{P}\mathbf{C} = \mathbf{0}$ which is the original equation for the camera centre given above.

The principal point. The principal axis is the line passing through the camera centre $\mathbf{C}$, with direction perpendicular to the principal plane $\mathbf{P}^3$. The axis intersects the image plane at the principal point. We may determine this point as follows. In general, the normal to a plane $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)^\mathsf{T}$ is the vector $(\pi_1, \pi_2, \pi_3)^\mathsf{T}$. This may alternatively be represented by a point $(\pi_1, \pi_2, \pi_3, 0)^\mathsf{T}$ on the plane at infinity. In the case of the principal plane $\mathbf{P}^3$ of the camera, this point is $(p_{31}, p_{32}, p_{33}, 0)^\mathsf{T}$, which we denote $\widehat{\mathbf{P}}^3$. Projecting that point using the camera matrix $\mathtt{P}$ gives the principal point of the camera $\mathtt{P}\widehat{\mathbf{P}}^3$. Note that only the left hand 3×3 part of $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ is involved in this formula. In fact the principal point is computed as $\mathbf{x}_0 = \mathtt{M}\mathbf{m}^3$ where $\mathbf{m}^{3\mathsf{T}}$ is the third row of $\mathtt{M}$.

The principal axis vector. Although any point $\mathbf{X}$ not on the principal plane may be mapped to an image point according to $\mathbf{x} = \mathtt{P}\mathbf{X}$, in reality only half the points in space, those that lie in front of the camera, may be seen in an image. Let $\mathtt{P}$ be written as $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$. It has just been seen that the vector $\mathbf{m}^3$ points in the direction of the principal axis. We would like to define this vector in such a way that it points in the direction towards the front of the camera (the *positive* direction). Note however that $\mathtt{P}$ is only defined up to sign. This leaves an ambiguity as to whether $\mathbf{m}^3$ or $-\mathbf{m}^3$ points in the positive direction. We now proceed to resolve this ambiguity.

We start by considering coordinates with respect to the camera coordinate frame. According to (6.5), the equation for projection of a 3D point to a point in the image is given by $\mathbf{x} = \mathtt{P}_{\text{cam}}\mathbf{X}_{\text{cam}} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]\mathbf{X}_{\text{cam}}$, where $\mathbf{X}_{\text{cam}}$ is the 3D point expressed in camera coordinates. In this case observe that the vector $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3 = (0, 0, 1)^\mathsf{T}$ points *towards the front of the camera* in the direction of the principal axis, irrespective of the scaling of $\mathtt{P}_{\text{cam}}$. For example, if $\mathtt{P}_{\text{cam}} \rightarrow k\mathtt{P}_{\text{cam}}$ then $\mathbf{v} \rightarrow k^4\mathbf{v}$ which has the same direction.

If the 3D point is expressed in world coordinates then $\mathtt{P} = k\mathtt{K}[\mathtt{R} \mid -\mathtt{R}\widetilde{\mathbf{C}}] = [\mathtt{M} \mid \mathbf{p}_4]$, where $\mathtt{M} = k\mathtt{KR}$. Since $\det(\mathtt{R}) > 0$ the vector $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3$ is again unaffected by scaling. In summary,

- $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3$ *is a vector in the direction of the principal axis, directed towards the front of the camera.*

6.2.2 Action of a projective camera on points

Forward projection. As we have already seen, a general projective camera maps a point in space $\mathbf{X}$ to an image point according to the mapping $\mathbf{x} = \mathtt{P}\mathbf{X}$. Points $\mathbf{D} = (\mathbf{d}^\mathsf{T}, 0)^\mathsf{T}$ on the plane at infinity represent vanishing points. Such points map to

$$\mathbf{x} = \mathtt{P}\mathbf{D} = [\mathtt{M} \mid \mathbf{p}_4]\mathbf{D} = \mathtt{M}\mathbf{d}$$

and thus are only affected by $\mathtt{M}$, the first 3×3 submatrix of $\mathtt{P}$.

Back-projection of points to rays. Given a point $\mathbf{x}$ in an image, we next determine the set of points in space that map to this point. This set will constitute a ray in space passing through the camera centre. The form of the ray may be specified in several ways, depending on how one wishes to represent a line in 3-space. A Plücker representation is postponed until section 8.1.2(*p*196). Here the line is represented as the join of two points.

We know two points on the ray. These are the camera centre $\mathbf{C}$ (where $\mathtt{P}\mathbf{C} = \mathbf{0}$) and the point $\mathtt{P}^+\mathbf{x}$, where $\mathtt{P}^+$ is the pseudo-inverse of $\mathtt{P}$. The pseudo-inverse of $\mathtt{P}$ is the matrix $\mathtt{P}^+ = \mathtt{P}^\mathsf{T}(\mathtt{P}\mathtt{P}^\mathsf{T})^{-1}$, for which $\mathtt{P}\mathtt{P}^+ = \mathtt{I}$ (see section A5.2(*p*590)). Point $\mathtt{P}^+\mathbf{x}$ lies

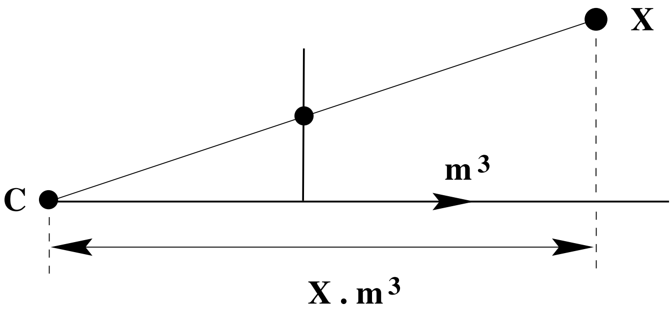

Fig. 6.6. *If the camera matrix* $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ *is normalized so that* $\|\mathbf{m}^3\| = 1$ *and* $\det \mathtt{M} > 0$*, and* $\mathbf{x} = w(x, y, 1)^\mathsf{T} = \mathtt{P}\mathbf{X}$*, where* $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$*, then* w *is the depth of the point* $\mathbf{X}$ *from the camera centre in the direction of the principal ray of the camera.*

on the ray because it projects to $\mathbf{x}$, since $\mathtt{P}(\mathtt{P}^+\mathbf{x}) = \mathtt{I}\mathbf{x} = \mathbf{x}$. Then the ray is the line formed by the join of these two points

$$\mathbf{X}(\lambda) = \mathtt{P}^+\mathbf{x} + \lambda\mathbf{C}. \tag{6.13}$$

In the case of finite cameras an alternative expression can be developed. Writing $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$, the camera centre is given by $\widetilde{\mathbf{C}} = -\mathtt{M}^{-1}\mathbf{p}_4$. An image point $\mathbf{x}$ back-projects to a ray intersecting the plane at infinity at the point $\mathbf{D} = ((\mathtt{M}^{-1}\mathbf{x})^\mathsf{T}, 0)^\mathsf{T}$, and $\mathbf{D}$ provides a second point on the ray. Again writing the line as the join of two points on the ray,

$$\mathbf{X}(\mu) = \mu\begin{pmatrix} \mathtt{M}^{-1}\mathbf{x} \\ 0 \end{pmatrix} + \begin{pmatrix} -\mathtt{M}^{-1}\mathbf{p}_4 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathtt{M}^{-1}(\mu\mathbf{x} - \mathbf{p}_4) \\ 1 \end{pmatrix}. \tag{6.14}$$

6.2.3 Depth of points

Next, we consider the distance a point lies in front of or behind the principal plane of the camera. Consider a camera matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$, projecting a point $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T} = (\widetilde{\mathbf{X}}^\mathsf{T}, 1)^\mathsf{T}$ in 3-space to the image point $\mathbf{x} = w(x, y, 1)^\mathsf{T} = \mathtt{P}\mathbf{X}$. Let $\mathbf{C} = (\widetilde{\mathbf{C}}, 1)^\mathsf{T}$ be the camera centre. Then $w = \mathbf{P}^{3\mathsf{T}}\mathbf{X} = \mathbf{P}^{3\mathsf{T}}(\mathbf{X} - \mathbf{C})$ since $\mathtt{P}\mathbf{C} = \mathbf{0}$ for the camera centre $\mathbf{C}$. However, $\mathbf{P}^{3\mathsf{T}}(\mathbf{X} - \mathbf{C}) = \mathbf{m}^{3\mathsf{T}}(\widetilde{\mathbf{X}} - \widetilde{\mathbf{C}})$ where $\mathbf{m}^3$ is the principal ray direction, so $w = \mathbf{m}^{3\mathsf{T}}(\widetilde{\mathbf{X}} - \widetilde{\mathbf{C}})$ can be interpreted as the dot product of the ray from the camera centre to the point $\mathbf{X}$, with the principal ray direction. If the camera matrix is normalized so that $\det \mathtt{M} > 0$ and $\|\mathbf{m}^3\| = 1$, then $\mathbf{m}^3$ is a unit vector pointing in the *positive* axial direction. Then w may be interpreted as the depth of the point $\mathbf{X}$ from the camera centre $\mathbf{C}$ in the direction of the principal ray. This is illustrated in figure 6.6.

Any camera matrix may be normalized by multiplying it by an appropriate factor. However, to avoid having always to deal with normalized camera matrices, the depth of a point may be computed as follows:

Result 6.1. *Let* $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\mathsf{T}$ *be a 3D point and* $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ *be a camera matrix for a finite camera. Suppose* $\mathtt{P}(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, \mathrm{T})^\mathsf{T} = w(x, y, 1)^\mathsf{T}$*. Then*

$$\mathrm{depth}(\mathbf{X}; \mathtt{P}) = \frac{\mathrm{sign}(\det \mathtt{M})w}{\mathrm{T}\|\mathbf{m}^3\|} \tag{6.15}$$

is the depth of the point $\mathbf{X}$ in front of the principal plane of the camera.

This formula is an effective way to determine if a point $\mathbf{X}$ is in front of the camera. One verifies that the value of $\text{depth}(\mathbf{X}; \mathtt{P})$ is unchanged if either the point $\mathbf{X}$ or the camera matrix $\mathtt{P}$ is multiplied by a constant factor k. Thus, $\text{depth}(\mathbf{X}; \mathtt{P})$ is independent of the particular homogeneous representation of $\mathbf{X}$ and $\mathtt{P}$.

6.2.4 Decomposition of the camera matrix

Let $\mathtt{P}$ be a camera matrix representing a general projective camera. We wish to find the camera centre, the orientation of the camera and the internal parameters of the camera from $\mathtt{P}$.

Finding the camera centre. The camera centre $\mathbf{C}$ is the point for which $\mathtt{P}\mathbf{C} = \mathbf{0}$. Numerically this right null-vector may be obtained from the SVD of $\mathtt{P}$, see section A4.4(*p*585). Algebraically, the centre $\mathbf{C} = (\text{X}, \text{Y}, \text{Z}, \text{T})^\mathsf{T}$ may be obtained as (see (3.5–*p*67))

$$\begin{aligned} \text{X} &= \det([\mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4]) \quad & \text{Y} &= -\det([\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_4]) \\ \text{Z} &= \det([\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4]) \quad & \text{T} &= -\det([\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]). \end{aligned}$$

Finding the camera orientation and internal parameters. In the case of a finite camera, according to (6.11),

$$\mathtt{P} = [\mathtt{M} \mid -\mathtt{M}\tilde{\mathbf{C}}] = \mathtt{K}[\mathtt{R} \mid -\mathtt{R}\tilde{\mathbf{C}}].$$

We may easily find both $\mathtt{K}$ and $\mathtt{R}$ by decomposing $\mathtt{M}$ as $\mathtt{M} = \mathtt{KR}$ using the RQ-decomposition. This decomposition into the product of an upper-triangular and orthogonal matrix is described in section A4.1.1(*p*579). The matrix $\mathtt{R}$ gives the orientation of the camera, whereas $\mathtt{K}$ is the calibration matrix. The ambiguity in the decomposition is removed by requiring that $\mathtt{K}$ have positive diagonal entries.

The matrix $\mathtt{K}$ has the form (6.10)

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

where

- α_x is the scale factor in the x-coordinate direction,
- α_y is the scale factor in the y-coordinate direction,
- s is the skew,
- $(x_0, y_0)^\mathsf{T}$ are the coordinates of the principal point.

The *aspect ratio* is α_y/α_x.

Example 6.2. The camera matrix

$$\mathtt{P} = \begin{bmatrix} 3.53553\,e{+}2 & 3.39645\,e{+}2 & 2.77744\,e{+}2 & -1.44946\,e{+}6 \\ -1.03528\,e{+}2 & 2.33212\,e{+}1 & 4.59607\,e{+}2 & -6.32525\,e{+}5 \\ 7.07107\,e{-}1 & -3.53553\,e{-}1 & 6.12372\,e{-}1 & -9.18559\,e{+}2 \end{bmatrix}$$

with $\mathtt{P} = [\mathtt{M} \mid -\mathtt{M}\widetilde{\mathbf{C}}]$, has centre $\widetilde{\mathbf{C}} = (1000.0, 2000.0, 1500.0)^{\mathsf{T}}$, and the matrix $\mathtt{M}$ decomposes as

$$\mathtt{M} = \mathtt{KR} = \begin{bmatrix} 468.2 & 91.2 & 300.0 \\ & 427.2 & 200.0 \\ & & 1.0 \end{bmatrix} \begin{bmatrix} 0.41380 & 0.90915 & 0.04708 \\ -0.57338 & 0.22011 & 0.78917 \\ 0.70711 & -0.35355 & 0.61237 \end{bmatrix}.$$

$\triangle$

When is $s \neq 0$? As was shown in section 6.1 a true CCD camera has only four internal camera parameters, since generally $s = 0$. If $s \neq 0$ then this can be interpreted as a skewing of the pixel elements in the CCD array so that the x- and y-axes are not perpendicular. This is admittedly very unlikely to happen.

In realistic circumstances a non-zero skew might arise as a result of taking an image of an image, for example if a photograph is re-photographed, or a negative is enlarged. Consider enlarging an image taken by a pinhole camera (such as an ordinary film camera) where the axis of the magnifying lens is not perpendicular to the film plane or the enlarged image plane.

The most severe distortion that can arise from this "picture of a picture" process is a planar homography. Suppose the original (finite) camera is represented by the matrix $\mathtt{P}$, then the camera representing the picture of a picture is $\mathtt{HP}$, where $\mathtt{H}$ is the homography matrix. Since $\mathtt{H}$ is non-singular, the left 3×3 submatrix of $\mathtt{HP}$ is non-singular and can be decomposed as the product $\mathtt{KR}$ – and $\mathtt{K}$ need not have $s = 0$. Note however that the $\mathtt{K}$ and $\mathtt{R}$ are no longer the calibration matrix and orientation of the original camera.

On the other hand, one verifies that the process of taking a picture of a picture does not change the apparent camera centre. Indeed, since $\mathtt{H}$ is non-singular, $\mathtt{HP}\mathbf{C} = \mathbf{0}$ if and only if $\mathtt{P}\mathbf{C} = \mathbf{0}$.

Where is the decomposition required? If the camera $\mathtt{P}$ is constructed from (6.11) then the parameters are known and a decomposition is clearly unnecessary. So the question arises – where would one obtain a camera for which the decomposition is not known? In fact cameras will be computed in myriad ways throughout this book and decomposing an unknown camera is a frequently used tool in practice. For example cameras can be computed directly by *calibration* – where the camera is computed from a set of world to image correspondences (chapter 7) – and indirectly by computing a multiple view relation (such as the fundamental matrix or trifocal tensor) and subsequently computing projection matrices from this relation.

A note on coordinate orientation. In the derivation of the camera model and its parametrization (6.10) it is assumed that the coordinate systems used in both the image and the 3D world are right handed systems, as shown in figure 6.1(p154). However, a common practice in measuring image coordinates is that the y-coordinate increases in the downwards direction, thus defining a left handed coordinate system, contrary to figure 6.1(p154). A recommended practice in this case is to negate the y-coordinate of the image point so that the coordinate system again becomes right handed. However, if the image coordinate system is left handed, then the consequences are not grave. The relationship between world and image coordinates is still expressed by a 3×4 camera matrix. Decomposition of this camera matrix according to (6.11) with $\mathtt{K}$ of the form (6.10) is still possible with α_x and α_y positive. The difference is that $\mathtt{R}$ now represents the orientation of the camera with respect to the negative $\mathtt{Z}$-axis. In addition, the depth of points given by (6.15) will be *negative* instead of positive for points in front of the camera. If this is borne in mind then it is permissible to use left handed coordinates in the image.

6.2.5 Euclidean vs projective spaces

The development of the sections to this point has implicitly assumed that the world and image coordinate systems are Euclidean. Ideas have been borrowed from projective geometry (such as directions corresponding to points on $\boldsymbol{\pi}_\infty$) and the convenient notation of homogeneous coordinates has allowed central projection to be represented linearly.

In subsequent chapters of the book we will go further and use a projective coordinate frame. This is easily achieved, for suppose the world coordinate frame is projective; then the transformation between the camera and world coordinate frame (6.6) is again represented by a 4×4 homogeneous matrix, $\mathbf{X}_{\text{cam}} = \mathtt{H}\mathbf{X}$, and the resulting map from projective 3-space $\mathbb{P}^3$ to the image is still represented by a 3×4 matrix $\mathtt{P}$ with rank 3. In fact, at its most general the projective camera is a map from $\mathbb{P}^3$ to $\mathbb{P}^2$, and covers the composed effects of a projective transformation of 3-space, a projection from 3-space to an image, and a projective transformation of the image. This follows simply by concatenating the matrices representing these mappings:

$$\mathtt{P} = [3 \times 3 \text{ homography}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} [4 \times 4 \text{ homography}]$$

which results in a 3×4 matrix.

However, it is important to remember that cameras are Euclidean devices and simply because we have a projective model of a camera it does not mean that we should eschew notions of Euclidean geometry.

Euclidean and affine interpretations. Although a (finite) 3×4 matrix can always be decomposed as in section 6.2.4 to obtain a rotation matrix, a calibration matrix $\mathtt{K}$, and so forth, Euclidean interpretations of the parameters so obtained are only meaningful if the image and space coordinates are in an appropriate frame. In the decomposition case a Euclidean frame is required for both image and 3-space. On the other hand, the interpretation of the null-vector of $\mathtt{P}$ as the camera centre is valid even if both frames are projective – the interpretation requires only collinearity, which is a projective notion. The interpretation of $\mathbf{P}^3$ as the principal plane requires at least affine frames for the image and 3-space. Finally, the interpretation of $\mathbf{m}^3$ as the principal ray requires an affine image frame but a Euclidean world frame in order for the concept of orthogonality (to the principal plane) to be meaningful.

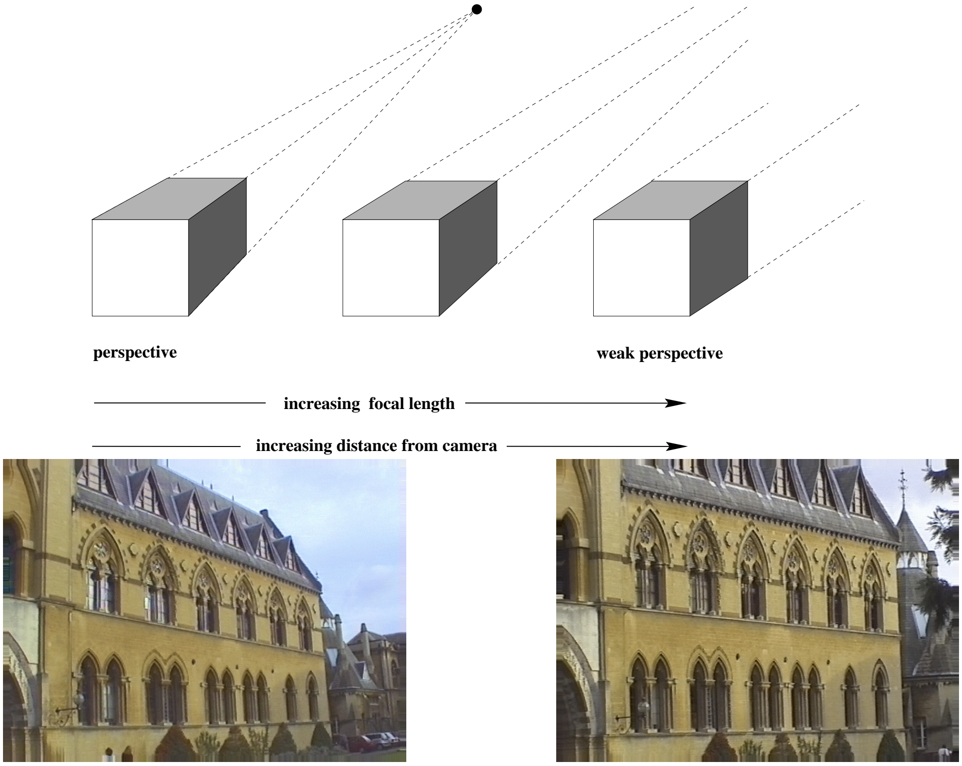

Fig. 6.7. *As the focal length increases and the distance between the camera and object also increases, the image remains the same size but perspective effects diminish.*

6.3 Cameras at infinity

We now turn to consider cameras with centre lying on the plane at infinity. This means that the left hand 3×3 block of the camera matrix $\mathtt{P}$ is singular. The camera centre may be found from $\mathtt{P}\mathbf{C} = \mathbf{0}$ just as with finite cameras.

Cameras at infinity may be broadly classified into two different types, *affine cameras* and *non-affine cameras*. We consider first of all the affine class of cameras which are the most important in practice.

Definition 6.3. An *affine* camera is one that has a camera matrix $\mathtt{P}$ in which the last row $\mathbf{P}^{3\mathsf{T}}$ is of the form $(0, 0, 0, 1)$.

It is called an affine camera because points at infinity are mapped to points at infinity.

6.3.1 Affine cameras

Consider what happens as we apply a cinematographic technique of tracking back while zooming in, in such a way as to keep objects of interest the same size[1]. This is illustrated in figure 6.7. We are going to model this process by taking the limit as both the focal length and principal axis distance of the camera from the object increase.

In analyzing this technique, we start with a finite projective camera (6.11). The

[1] See ‘Vertigo’ (Dir. Hitchcock, 1958) and ‘Mishima’ (Dir. Schrader, 1985).

camera matrix may be written as

$$\mathtt{P}_0 = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}] = \mathtt{K}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{3\mathsf{T}} & -\mathbf{r}^{3\mathsf{T}}\widetilde{\mathbf{C}} \end{bmatrix} \tag{6.16}$$

where $\mathbf{r}^{i\mathsf{T}}$ is the i-th row of the rotation matrix. This camera is located at position $\widetilde{\mathbf{C}}$ and has orientation denoted by matrix $\mathtt{R}$ and internal parameters matrix $\mathtt{K}$ of the form given in (6.10–p157). From section 6.2.1 the principal ray of the camera is in the direction of the vector $\mathbf{r}^3$, and the value $d_0 = -\mathbf{r}^{3\mathsf{T}}\widetilde{\mathbf{C}}$ is the distance of the world origin from the camera centre in the direction of the principal ray.

Now, we consider what happens if the camera centre is moved backwards along the principal ray at unit speed for a time t, so that the centre of the camera is moved to $\widetilde{\mathbf{C}} - t\mathbf{r}^3$. Replacing $\widetilde{\mathbf{C}}$ in (6.16) by $\widetilde{\mathbf{C}} - t\mathbf{r}^3$ gives the camera matrix at time t:

$$\mathtt{P}_t = \mathtt{K}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}(\widetilde{\mathbf{C}} - t\mathbf{r}^3) \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}(\widetilde{\mathbf{C}} - t\mathbf{r}^3) \\ \mathbf{r}^{3\mathsf{T}} & -\mathbf{r}^{3\mathsf{T}}(\widetilde{\mathbf{C}} - t\mathbf{r}^3) \end{bmatrix} = \mathtt{K}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{3\mathsf{T}} & d_t \end{bmatrix} \tag{6.17}$$

where the terms $\mathbf{r}^{i\mathsf{T}}\mathbf{r}^3$ are zero for $i = 1, 2$ because $\mathtt{R}$ is a rotation matrix. The scalar $d_t = -\mathbf{r}^{3\mathsf{T}}\widetilde{\mathbf{C}} + t$ is the depth of the world origin with respect to the camera centre in the direction of the principal ray $\mathbf{r}^3$ of the camera. Thus

- *The effect of tracking along the principal ray is to replace the (3,4) entry of the matrix by the depth d_t of the camera centre from the world origin.*

Next, we consider zooming such that the camera focal length is increased by a factor k. This magnifies the image by a factor k. It is shown in section 8.4.1(p203) that the effect of zooming by a factor k is to multiply the calibration matrix $\mathtt{K}$ on the right by $\mathrm{diag}(k, k, 1)$.

Now, we combine the effects of tracking and zooming. We suppose that the magnification factor is $k = d_t/d_0$ so that the image size remains fixed. The resulting camera matrix at time t, derived from (6.17), is

$$\mathtt{P}_t = \mathtt{K}\begin{bmatrix} d_t/d_0 & & \\ & d_t/d_0 & \\ & & 1 \end{bmatrix}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{3\mathsf{T}} & d_t \end{bmatrix} = \frac{d_t}{d_0}\mathtt{K}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{3\mathsf{T}} d_0/d_t & d_0 \end{bmatrix}$$

and one can ignore the factor d_t/d_0. When $t = 0$, the camera matrix $\mathtt{P}_t$ corresponds with (6.16). Now, in the limit as d_t tends to infinity, this matrix becomes

$$\mathtt{P}_\infty = \lim_{t\to\infty} \mathtt{P}_t = \mathtt{K}\begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & -\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{r}^{2\mathsf{T}} & -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}} \\ \mathbf{0}^{\mathsf{T}} & d_0 \end{bmatrix} \tag{6.18}$$

which is just the original camera matrix (6.16) with the first three entries of the last row set to zero. From definition 6.3 $\mathtt{P}_\infty$ is an instance of an affine camera.

6.3.2 Error in employing an affine camera

It may be noted that the image of any point on the plane through the world origin perpendicular to the principal axis direction $\mathbf{r}^3$ is unchanged by this combined zooming and motion. Indeed, such a point may be written as

$$\mathbf{X} = \begin{pmatrix} \alpha\mathbf{r}^1 + \beta\mathbf{r}^2 \\ 1 \end{pmatrix}.$$

One then verifies that $\mathtt{P}_0\mathbf{X} = \mathtt{P}_t\mathbf{X} = \mathtt{P}_\infty\mathbf{X}$ for all t, since $\mathbf{r}^{3\mathsf{T}}(\alpha\mathbf{r}^1 + \beta\mathbf{r}^2) = 0$.

For points not on this plane the images under $\mathtt{P}_0$ and $\mathtt{P}_\infty$ differ, and we will now investigate the extent of this error. Consider a point $\mathbf{X}$ which is at a perpendicular distance Δ from this plane. The 3D point can be represented as

$$\mathbf{X} = \begin{pmatrix} \alpha\mathbf{r}^1 + \beta\mathbf{r}^2 + \Delta\mathbf{r}^3 \\ 1 \end{pmatrix}$$

and is imaged by the cameras $\mathtt{P}_0$ and $\mathtt{P}_\infty$ at

$$\mathbf{x}_{\text{proj}} = \mathtt{P}_0\mathbf{X} = \mathtt{K}\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ d_0 + \Delta \end{pmatrix} \qquad \mathbf{x}_{\text{affine}} = \mathtt{P}_\infty\mathbf{X} = \mathtt{K}\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ d_0 \end{pmatrix}$$

where $\tilde{x} = \alpha - \mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}}$, $\tilde{y} = \beta - \mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}}$. Now, writing the calibration matrix as

$$\mathtt{K} = \begin{bmatrix} \mathtt{K}_{2\times 2} & \tilde{\mathbf{x}}_0 \\ \tilde{\mathbf{0}}^{\mathsf{T}} & 1 \end{bmatrix},$$

where $\mathtt{K}_{2\times 2}$ is an upper-triangular 2×2 matrix, gives

$$\mathbf{x}_{\text{proj}} = \begin{pmatrix} \mathtt{K}_{2\times 2}\tilde{\mathbf{x}} + (d_0 + \Delta)\tilde{\mathbf{x}}_0 \\ d_0 + \Delta \end{pmatrix} \qquad \mathbf{x}_{\text{affine}} = \begin{pmatrix} \mathtt{K}_{2\times 2}\tilde{\mathbf{x}} + d_0\tilde{\mathbf{x}}_0 \\ d_0 \end{pmatrix}$$

The image point for $\mathtt{P}_0$ is obtained by dehomogenizing, by dividing by the third element, as $\tilde{\mathbf{x}}_{\text{proj}} = \tilde{\mathbf{x}}_0 + \mathtt{K}_{2\times 2}\tilde{\mathbf{x}}/(d_0 + \Delta)$, and for $\mathtt{P}_\infty$ the inhomogeneous image point is $\tilde{\mathbf{x}}_{\text{affine}} = \tilde{\mathbf{x}}_0 + \mathtt{K}_{2\times 2}\tilde{\mathbf{x}}/d_0$. The relationship between the two points is therefore

$$\tilde{\mathbf{x}}_{\text{affine}} - \tilde{\mathbf{x}}_0 = \frac{d_0 + \Delta}{d_0}(\tilde{\mathbf{x}}_{\text{proj}} - \tilde{\mathbf{x}}_0)$$

which shows that

- *The effect of the affine approximation $\mathtt{P}_\infty$ to the true camera matrix $\mathtt{P}_0$ is to move the image of a point $\mathbf{X}$ radially towards or away from the principal point $\tilde{\mathbf{x}}_0$ by a factor equal to $(d_0 + \Delta)/d_0 = 1 + \Delta/d_0$.*

This is illustrated in figure 6.8.

Affine imaging conditions. From the expressions for $\tilde{\mathbf{x}}_{\text{proj}}$ and $\tilde{\mathbf{x}}_{\text{affine}}$ we can deduce that

$$\tilde{\mathbf{x}}_{\text{affine}} - \tilde{\mathbf{x}}_{\text{proj}} = \frac{\Delta}{d_0}(\tilde{\mathbf{x}}_{\text{proj}} - \tilde{\mathbf{x}}_0) \tag{6.19}$$

which shows that the distance between the true perspective image position and the position obtained using the affine camera approximation $\mathtt{P}_\infty$ will be small provided:

(i) The depth relief (Δ) is small compared to the average depth (d_0), and
(ii) The distance of the point from the principal ray is small.

The latter condition is satisfied by a small field of view. In general, images acquired using a lens with a longer focal length tend to satisfy these conditions as both the field of view and the depth of field are smaller than those obtained by a short focal length lens with the same CCD array.

For scenes at which there are many points at different depths, the affine camera is not a good approximation. For instance where the scene contains close foreground as well as background objects, an affine camera model should not be used. However, a different affine model can be used for each region in these circumstances.

6.3.3 Decomposition of $\mathtt{P}_\infty$

The camera matrix (6.18) may be written as

$$\mathtt{P}_\infty = \begin{bmatrix} \mathtt{K}_{2\times2} & \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} \\ \mathbf{0}^\mathsf{T} & d_0 \end{bmatrix}$$

where $\hat{\mathtt{R}}$ consists of the two first rows of a rotation matrix, $\hat{\mathbf{t}}$ is the vector $(-\mathbf{r}^{1\mathsf{T}}\widetilde{\mathbf{C}}, -\mathbf{r}^{2\mathsf{T}}\widetilde{\mathbf{C}})^\mathsf{T}$, and $\hat{\mathbf{0}}$ the vector $(0,0)^\mathsf{T}$. The 2×2 matrix $\mathtt{K}_{2\times2}$ is upper-triangular. One quickly verifies that

$$\mathtt{P}_\infty = \begin{bmatrix} \mathtt{K}_{2\times2} & \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} \\ \mathbf{0}^\mathsf{T} & d_0 \end{bmatrix} = \begin{bmatrix} d_0^{-1}\mathtt{K}_{2\times2} & \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$$

so we may replace $\mathtt{K}_{2\times2}$ by $d_0^{-1}\mathtt{K}_{2\times2}$ and assume that $d_0 = 1$. Multiplying out this product gives

$$\begin{aligned} \mathtt{P}_\infty &= \begin{bmatrix} \mathtt{K}_{2\times2}\hat{\mathtt{R}} & \mathtt{K}_{2\times2}\hat{\mathbf{t}} + \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} \mathtt{K}_{2\times2} & \hat{\mathbf{0}} \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} + \mathtt{K}_{2\times2}^{-1}\tilde{\mathbf{x}}_0 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathtt{K}_{2\times2} & \mathtt{K}_{2\times2}\hat{\mathbf{t}} + \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{0}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \end{aligned}$$

Thus, making appropriate substitutions for $\hat{\mathbf{t}}$ or $\tilde{\mathbf{x}}_0$, we can write the affine camera matrix in one of the two forms

$$\mathtt{P}_\infty = \begin{bmatrix} \mathtt{K}_{2\times2} & \hat{\mathbf{0}} \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} \mathtt{K}_{2\times2} & \tilde{\mathbf{x}}_0 \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{0}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \tag{6.20}$$

Consequently, the camera $\mathtt{P}_\infty$ can be interpreted in terms of these decompositions in

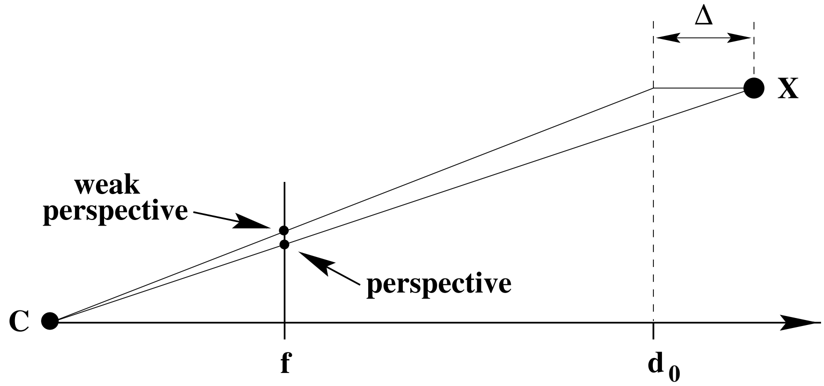

Fig. 6.8. **Perspective vs weak perspective projection.** *The action of the weak perspective camera is equivalent to orthographic projection onto a plane (at $Z = d_0$), followed by perspective projection from the plane. The difference between the perspective and weak perspective image point depends both on the distance Δ of the point $\mathbf{X}$ from the plane, and the distance of the point from the principal ray.*

one of two ways, either with $\tilde{\mathbf{x}}_0 = \mathbf{0}$ or with $\hat{\mathbf{t}} = \hat{\mathbf{0}}$. Using the second decomposition of (6.20), the image of the world origin is $\mathtt{P}_\infty(0,0,0,1)^\mathsf{T} = (\tilde{\mathbf{x}}_0^\mathsf{T}, 1)^\mathsf{T}$. Consequently, the value of $\tilde{\mathbf{x}}_0$ is dependent on the particular choice of world coordinates, and hence is not an intrinsic property of the camera itself. This means that the camera matrix $\mathtt{P}_\infty$ does not have a principal point. Therefore, it is preferable to use the first decomposition of $\mathtt{P}_\infty$ in (6.20), and write

$$\mathtt{P}_\infty = \begin{bmatrix} \mathtt{K}_{2\times 2} & \hat{\mathbf{0}} \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathtt{R}} & \hat{\mathbf{t}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \tag{6.21}$$

where the two matrices represent the internal camera parameters and external camera parameters of $\mathtt{P}_\infty$.

Parallel projection. In summary the essential differences between $\mathtt{P}_\infty$ and a finite camera are:

- *The parallel projection matrix* $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ *replaces the canonical projection matrix* $[\mathtt{I} \mid \mathbf{0}]$ *of a finite camera (6.5–p155).*
- *The calibration matrix* $\begin{bmatrix} \mathtt{K}_{2\times 2} & \hat{\mathbf{0}} \\ \hat{\mathbf{0}}^\mathsf{T} & 1 \end{bmatrix}$ *replaces* $\mathtt{K}$ *of a finite camera (6.10–p157).*
- *The principal point is not defined.*

6.3.4 A hierarchy of affine cameras

In a similar manner to the development of the finite projection camera taxonomy in section 6.1 we can start with the basic operation of parallel projection and build a hierarchy of camera models representing progressively more general cases of parallel projection.

Orthographic projection. Consider projection along the $\mathtt{Z}$-axis. This is represented by a matrix of the form

$$\mathtt{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.22}$$

This mapping takes a point $(\mathtt{X}, \mathtt{Y}, \mathtt{Z}, 1)^\mathsf{T}$ to the image point $(\mathtt{X}, \mathtt{Y}, 1)^\mathsf{T}$, dropping the $\mathtt{Z}$-coordinate.

For a general orthographic projection mapping, we precede this map by a 3D Euclidean coordinate change of the form

$$\mathtt{H} = \begin{bmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}.$$

Writing $\mathbf{t} = (t_1, t_2, t_3)^\mathsf{T}$, we see that a general orthographic camera is of the form

$$\mathtt{P} = \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \tag{6.23}$$

An orthographic camera has five degrees of freedom, namely the three parameters describing the rotation matrix $\mathtt{R}$, plus the two offset parameters t_1 and t_2. An orthographic projection matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{t}]$ is characterized by a matrix $\mathtt{M}$ with last row zero, with the first two rows orthogonal and of unit norm, and $t_3 = 1$.

Scaled orthographic projection. A scaled orthographic projection is an orthographic projection followed by isotropic scaling. Thus, in general, its matrix may be written in the form

$$\mathtt{P} = \begin{bmatrix} k & & \\ & k & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^\mathsf{T} & 1/k \end{bmatrix}. \tag{6.24}$$

It has six degrees of freedom. A scaled orthographic projection matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{t}]$ is characterized by a matrix $\mathtt{M}$ with last row zero, and the first two rows orthogonal and of equal norm.

Weak perspective projection. Analogous to a finite CCD camera, we may consider the case of a camera at infinity for which the scale factors in the two axial image directions are not equal. Such a camera has a projection matrix of the form

$$\mathtt{P} = \begin{bmatrix} \alpha_x & & \\ & \alpha_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \tag{6.25}$$

It has seven degrees of freedom. A weak perspective projection matrix $\mathtt{P} = [\mathtt{M} \mid \mathbf{t}]$ is characterized by a matrix $\mathtt{M}$ with last row zero, and first two rows orthogonal (but they need not have equal norm as is required in the scaled orthographic case). The geometric action of this camera is illustrated in figure 6.8.

The affine camera, $\mathtt{P}_{\mathrm{A}}$. As has already been seen in the case of $\mathtt{P}_\infty$, a general camera matrix of the affine form, and with no restrictions on its elements, may be decomposed as

$$\mathtt{P}_{\mathrm{A}} = \begin{bmatrix} \alpha_x & s & \\ & \alpha_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} \mathbf{r}^{1\mathsf{T}} & t_1 \\ \mathbf{r}^{2\mathsf{T}} & t_2 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}.$$

It has eight degrees of freedom, and may be thought of as the parallel projection version of the finite projective camera (6.11–*p*157).

In full generality an affine camera has the form

$$\mathtt{P}_{\mathrm{A}} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & t_1 \\ m_{21} & m_{22} & m_{23} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

It has eight degrees of freedom corresponding to the eight non-zero and non-unit matrix elements. We denote the top left 2×3 submatrix by $\mathtt{M}_{2\times3}$. The sole restriction on the affine camera is that $\mathtt{M}_{2\times3}$ has rank 2. This arises from the requirement that the rank of $\mathtt{P}$ is 3.

The affine camera covers the composed effects of an affine transformation of 3-space, an orthographic projection from 3-space to an image, and an affine transformation of the image. This follows simply by concatenating the matrices representing these mappings:

$$\mathtt{P}_{\mathrm{A}} = [3 \times 3 \text{ affine}] \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} [4 \times 4 \text{ affine}]$$

which results in a 3×4 matrix of the affine form.

Projection under an affine camera is a linear mapping on *inhomogeneous* coordinates composed with a translation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

which is written more concisely as

$$\tilde{\mathbf{x}} = \mathtt{M}_{2\times3}\widetilde{\mathbf{X}} + \tilde{\mathbf{t}} \quad . \tag{6.26}$$

The point $\tilde{\mathbf{t}} = (t_1, t_2)^\mathsf{T}$ is the image of the world origin.

The camera models of this section are seen to be affine cameras satisfying additional constraints, thus the affine camera is an abstraction of this hierarchy. For example, in the case of the weak perspective camera the rows of $\mathtt{M}_{2\times3}$ are scalings of rows of a rotation matrix, and thus are orthogonal.

6.3.5 More properties of the affine camera

The plane at infinity in space is mapped to points at infinity in the image. This is easily seen by computing $\mathtt{P}_{\mathrm{A}}(\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 0)^\mathsf{T} = (\mathrm{X}, \mathrm{Y}, 0)^\mathsf{T}$. Extending the terminology of finite projective cameras, we interpret this by saying that the principal plane of the camera is the plane at infinity. The optical centre, since it lies on the principal plane, must also lie on the plane at infinity. From this we have

(i) Conversely, any projective camera matrix for which the principal plane is the plane at infinity is an affine camera matrix.
(ii) Parallel world lines are projected to parallel image lines. This follows because parallel world lines intersect at the plane at infinity, and this intersection point is mapped to a point at infinity in the image. Hence the image lines are parallel.
(iii) The vector $\mathbf{d}$ satisfying $\mathtt{M}_{2\times 3}\mathbf{d} = \mathbf{0}$ is the *direction* of parallel projection, and $(\mathbf{d}^\mathsf{T}, 0)^\mathsf{T}$ the camera centre since $\mathtt{P}_{\text{A}} \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix} = \mathbf{0}$.

Any camera which consists of the composed effects of affine transformations (either of space, or of the image) with parallel projection will have the affine form. For example, *para-perspective* projection consists of two such mappings: the first is parallel projection onto a plane $\boldsymbol{\pi}$ through the centroid and parallel to the image plane. The direction of parallel projection is the ray joining the centroid to the camera centre. This parallel projection is followed by an affine transformation (actually a similarity) between $\boldsymbol{\pi}$ and the image. Thus a para-perspective camera is an affine camera.

6.3.6 General cameras at infinity

An affine camera is one for which the principal plane is the plane at infinity. As such, its camera centre lies on the plane at infinity. However, it is possible for the camera centre to lie on the plane at infinity without the whole principal plane being the plane at infinity.

A camera centre lies at infinity if $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ with $\mathtt{M}$ a singular matrix. This is clearly a weaker condition than insisting that the last row of $\mathtt{M}$ is zero, as is the case for affine cameras. If $\mathtt{M}$ is singular, but the last row of $\mathtt{M}$ is not zero, then the camera is not affine, but not a finite projective camera either. Such a camera is rather a strange object, however, and will not be treated in detail in this book. We may compare the properties of affine and non-affine infinite cameras:

	Affine camera	Non-affine camera
Camera centre on $\boldsymbol{\pi}_\infty$	yes	yes
Principal plane is $\boldsymbol{\pi}_\infty$	yes	no
Image of points on $\boldsymbol{\pi}_\infty$ on $\mathbf{l}_\infty$	yes	no in general

In both cases the camera centre is the direction of projection. Furthermore, in the case of an affine camera all non-infinite points are in front of the camera. For a non-affine camera space is partitioned into two sets of points by the principal plane.

A general camera at infinity could arise from a perspective image of an image produced by an affine camera. This imaging process corresponds to left-multiplying the

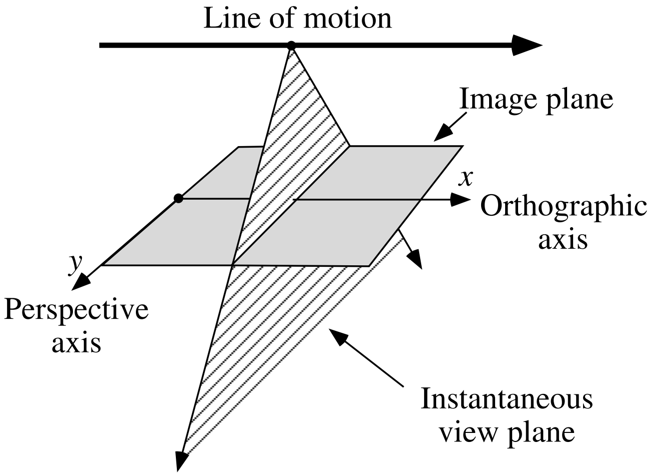

Fig. 6.9. *Acquisition geometry of a pushbroom camera.*

affine camera by a general 3×3 matrix representing the planar homography. The resulting 3×4 matrix is still a camera at infinity, but it does not have the affine form, since parallel lines in the world will in general appear as converging lines in the image.

6.4 Other camera models

6.4.1 Pushbroom cameras

The Linear Pushbroom (LP) camera is an abstraction of a type of sensor common in satellites, for instance the SPOT sensor. In such a camera, a linear sensor array is used to capture a single line of imagery at a time. As the sensor moves the sensor plane sweeps out a region of space (hence the name pushbroom), capturing the image a single line at a time. The second dimension of the image is provided by the motion of the sensor. In the linear pushbroom model, the sensor is assumed to move in a straight line at constant velocity with respect to the ground. In addition, one assumes that the orientation of the sensor array with respect to the direction of travel is constant. In the direction of the sensor, the image is effectively a perspective image, whereas in the direction of the sensor motion it is an orthographic projection. The geometry of the LP camera is illustrated in figure 6.9. It turns out that the mapping from object space into the image may be described by a 3×4 camera matrix, just as with a general projective camera. However, the way in which this matrix is used is somewhat different.

- *Let* $\mathbf{X} = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ *be an object point, and let* $\mathtt{P}$ *be the camera matrix of the LP camera. Suppose that* $\mathtt{P}\mathbf{X} = (x, y, w)^\mathsf{T}$*. Then the corresponding image point (represented as an inhomogeneous 2-vector) is* $(x, y/w)^\mathsf{T}$.

One must compare this with the projective camera mapping. In that case the point represented by $(x, y, w)^\mathsf{T}$ is $(x/w, y/w)^\mathsf{T}$. Note the difference that in the LP case, the coordinate x is not divided by the factor w to get the image coordinate. In this formula, the x-axis in the image is the direction of the sensor motion, whereas the y-axis is in the direction of the linear sensor array. The camera has 11 degrees of freedom.

Another way of writing the formula for LP projection is

$$\tilde{x} = x = \mathtt{P}^{1\mathsf{T}}\mathbf{X} \qquad \tilde{y} = y/z = \frac{\mathtt{P}^{2\mathsf{T}}\mathbf{X}}{\mathtt{P}^{3\mathsf{T}}\mathbf{X}} \tag{6.27}$$

where $(\tilde{x}, \tilde{y})$ is the image point.

Note that the $\tilde{y}$-coordinate behaves projectively, whereas the $\tilde{x}$ is obtained by orthogonal projection of the point $\mathbf{X}$ on the direction perpendicular to the plane $\mathtt{P}^1$. The vector $\mathtt{P}^1$ represents the sweep plane of the camera at time $t = 0$ – that is the moment when the line with coordinates $\tilde{x} = 0$ is captured.

Mapping of lines. One of the novel features of the LP camera is that straight lines in space are not mapped to straight lines in the image (they are mapped to straight lines in the case of a projective camera – see section 8.1.2). The set of points $\mathbf{X}$ lying on a 3D line may be written as $\mathbf{X}_0 + \alpha\mathbf{D}$, where $\mathbf{X}_0 = (\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1)^\mathsf{T}$ is a point on the line and $\mathbf{D} = (D_\mathrm{X}, D_\mathrm{Y}, D_\mathrm{Z}, 0)^\mathsf{T}$ is the intersection of this line with the plane at infinity. In this case, we compute from (6.27)

$$\begin{aligned} \tilde{x} &= \mathbf{P}^{1\mathsf{T}}(\mathbf{X}_0 + t\mathbf{D}) \\ \tilde{y} &= \frac{\mathbf{P}^{2\mathsf{T}}(\mathbf{X}_0 + t\mathbf{D})}{\mathbf{P}^{3\mathsf{T}}(\mathbf{X}_0 + t\mathbf{D})}. \end{aligned}$$

This may be written as a pair of equations $\tilde{x} = a+bt$ and $(c+dt)\tilde{y} = e+ft$. Eliminating t from these equations leads to an equation of the form $\alpha\tilde{x}\tilde{y} + \beta\tilde{x} + \gamma\tilde{y} + \delta = 0$, which is the equation of a hyperbola in the image plane, asymptotic in one direction to the line $\alpha\tilde{x} + \gamma = 0$, and in the other direction to the line $\alpha\tilde{y} + \beta = 0$. A hyperbola is made up of two curves. However, only one of the curves making up the image of a line actually appears in the image – the other part of the hyperbola corresponds to points lying behind the camera.

6.4.2 Line cameras

This chapter has dealt with the central projection of 3-space onto a 2D image. An analogous development can be given for the central projection of a plane onto a 1D line contained in the plane. See figure 22.1(*p*535). The camera model for this geometry is

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix} \begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} = \mathtt{P}_{2\times3}\mathbf{x}$$

which is a linear mapping from homogeneous representation of the plane to a homogeneous representation of the line. The camera has 5 degrees of freedom. Again the null-space, $\mathbf{c}$, of the $\mathtt{P}_{2\times3}$ projection matrix is the camera centre, and the matrix can be decomposed in a similar manner to the finite projective camera (6.11–*p*157) as

$$\mathtt{P}_{2\times3} = \mathtt{K}_{2\times2}\mathtt{R}_{2\times2}[\mathtt{I}_{2\times2} \mid -\tilde{\mathbf{c}}]$$

where $\tilde{\mathbf{c}}$ is the inhomogeneous 2-vector representing the centre (2 dof), $\mathtt{R}_{2\times2}$ is a rotation matrix (1 dof), and

$$\mathtt{K}_{2\times2} = \begin{bmatrix} \alpha_x & x_0 \\ & 1 \end{bmatrix}$$

the internal calibration matrix (2 dof).

6.5 Closure

This chapter has covered camera models, their taxonomy and anatomy. The subsequent chapters cover the estimation of cameras from a set of world to image correspondences, and the action of a camera on various geometric objects such as lines and quadrics. Vanishing points and vanishing lines are also described in more detail in chapter 8.

6.5.1 The literature

[Aloimonos-90] defined a hierarchy of camera models including para-perspective. Mundy and Zisserman [Mundy-92] generalized this with the affine camera. Faugeras developed properties of the projective camera in his textbook [Faugeras-93]. Further details on the linear pushbroom camera are given in [Gupta-97], and on the 2D camera in [Quan-97b].

6.5.2 Notes and exercises

(i) Let I_0 be a projective image, and I_1 be an image of I_0 (an image of an image). Let the composite image be denoted by I'. Show that the apparent camera centre of I' is the same as that of I_0. Speculate on how this explains why a portrait's eyes "follow you round the room." Verify on the other hand that all other parameters of I' and I_0 may be different.

(ii) Show that the ray back-projected from an image point $\mathbf{x}$ under a projective camera $\mathtt{P}$ (as in (6.14–*p*162)) may be written as

$$\mathtt{L}^* = \mathtt{P}^\mathsf{T}[\mathbf{x}]_\times \mathtt{P} \tag{6.28}$$

where $\mathtt{L}^*$ is the dual Plücker representation of a line (3.9–*p*71).

(iii) **The affine camera.**

(a) Show that the affine camera is the most general linear mapping on homogeneous coordinates that maps parallel world lines to parallel image lines. To do this consider the projection of points on $\boldsymbol{\pi}_\infty$, and show that only if $\mathtt{P}$ has the affine form will they map to points at infinity in the image.

(b) Show that for parallel lines mapped by an affine camera the ratio of lengths on line segments is an invariant. What other invariants are there under an affine camera?

(iv) **The rational polynomial camera** is a general camera model, used extensively in the satellite surveillance community. Image coordinates are defined by the ratios

$$x = N_x(\mathbf{X})/D_x(\mathbf{X}) \quad y = N_y(\mathbf{X})/D_y(\mathbf{X})$$

where the functions N_x, D_x, N_y, D_y are homogeneous *cubic* polynomials in the 3-space point $\mathbf{X}$. Each cubic has 20 coefficients, so that overall the camera has 78 degrees of freedom. All of the cameras surveyed in this chapter (projective, affine, pushbroom) are special cases of the rational polynomial camera. Its disadvantage is that it is severely over-parametrized for these cases. More details are given in Hartley and Saxena [Hartley-97e].

(v) A finite projective camera (6.11–p157) $\mathtt{P}$ may be transformed to an orthographic camera (6.22) by applying a 4×4 homography $\mathtt{H}$ on the right such that

$$\mathtt{PH} = \mathtt{KR}[\mathtt{I} \mid -\widetilde{\mathbf{C}}]\mathtt{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathtt{P}_{\text{orthog}} \ .$$

(the last row of $\mathtt{H}$ is chosen so that $\mathtt{H}$ has rank 4). Then since

$$\mathbf{x} = \mathtt{P}(\mathtt{HH}^{-1})\mathbf{X} = (\mathtt{PH})(\mathtt{H}^{-1}\mathbf{X}) = \mathtt{P}_{\text{orthog}}\mathbf{X}'$$

imaging under $\mathtt{P}$ is equivalent to first transforming the 3-space points $\mathbf{X}$ to $\mathbf{X}' = \mathtt{H}^{-1}\mathbf{X}$ and then applying an orthographic projection. Thus the action of any camera may be considered as a projective transformation of 3-space followed by orthographic projection.