Chapter 3

Camera model and multiple view geometry

Before discussing how 3D information can be obtained from images it is important to know how images are formed. First, the camera model is introduced; and then some important relationships between multiple views of a scene are presented.

3.1 The camera model

In this work the perspective camera model is used. This corresponds to an ideal pinhole camera. The geometric process for image formation in a pinhole camera has been nicely illustrated by Dürer (see Figure 3.1). The process is completely determined by choosing a perspective projection center and a retinal plane. The projection of a scene point is then obtained as the intersection of a line passing through this point and the center of projection $\mathtt{C}$ with the retinal plane $\mathcal{R}$.

Most cameras are described relatively well by this model. In some cases additional effects (e.g. radial distortion) have to be taken into account (see Section 3.1.5).

3.1.1 A simple model

In the simplest case where the projection center is placed at the origin of the world frame and the image plane is the plane $Z = 1$, the projection process can be modeled as follows:

$$x = \tfrac{X}{Z} \quad y = \tfrac{Y}{Z} \tag{3.1}$$

For a world point (X, Y, Z) and the corresponding image point (x, y). Using the homogeneous representation of the points a linear projection equation is obtained:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.2}$$

This projection is illustrated in Figure 3.2. The optical axis passes through the center of projection $\mathtt{C}$ and is orthogonal to the retinal plane $\mathcal{R}$. It's intersection with the retinal plane is defined as the principal point $\mathtt{c}$.

3.1.2 Intrinsic calibration

With an actual camera the focal length f (i.e. the distance between the center of projection and the retinal plane) will be different from 1, the coordinates of equation (3.2) should therefore be scaled with f to take this into account.

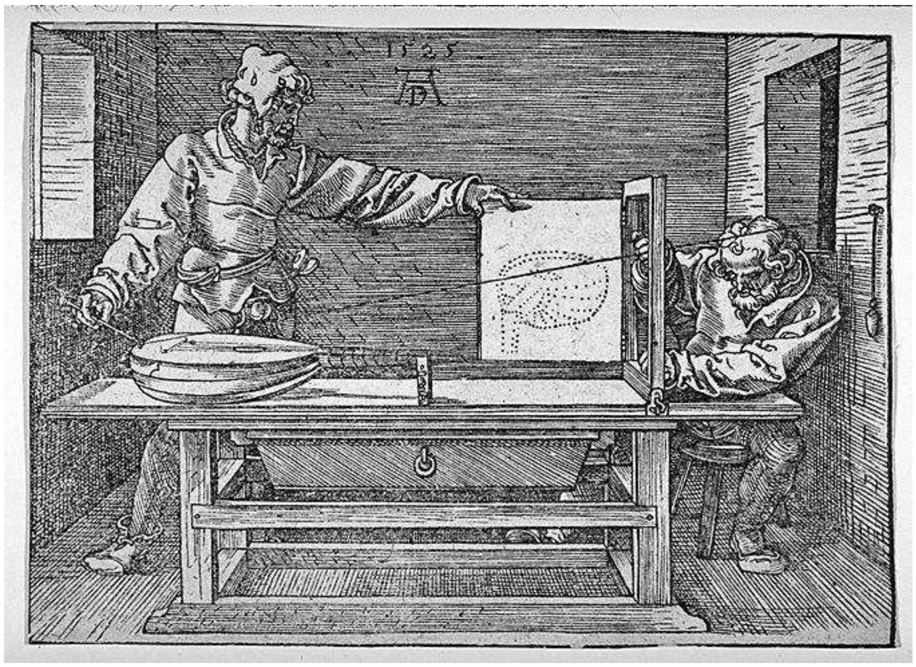

Figure 3.1: *Man Drawing a Lute (The Draughtsman of the Lute), woodcut 1525, Albrecht Dürer.*

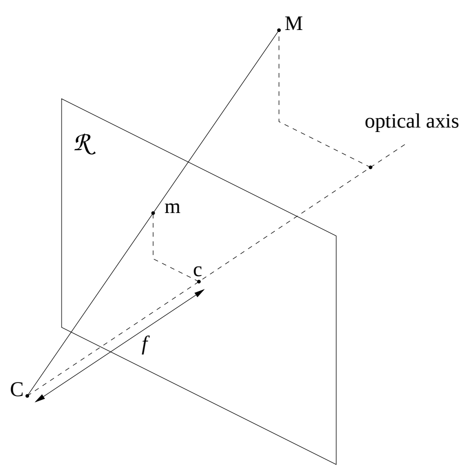

Figure 3.2: *Perspective projection*

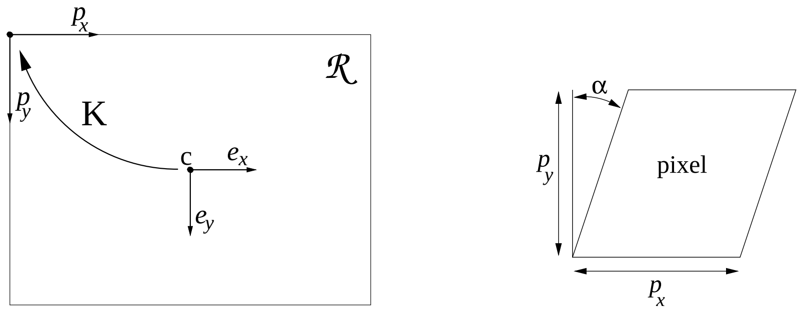

Figure 3.3: *From retinal coordinates to image coordinates*

In addition the coordinates in the image do not correspond to the physical coordinates in the retinal plane. With a CCD camera the relation between both depends on the size and shape of the pixels and of the position of the CCD chip in the camera. With a standard photo camera it depends on the scanning process through which the images are digitized.

The transformation is illustrated in Figure 3.3. The image coordinates are obtained through the following equations:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{p_x} & (\tan\alpha)\frac{f}{p_y} & c_x \\ & \frac{f}{p_y} & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x_{\mathcal{R}} \\ y_{\mathcal{R}} \\ 1 \end{bmatrix}$$

where p_x and p_y are the width and the height of the pixels, $\mathtt{c} = [c_x \; c_y \; 1]^\top$ is the principal point and α the skew angle as indicated in Figure 3.3. Since only the ratios $\frac{f}{p_x}$ and $\frac{f}{p_y}$ are of importance the simplified notations of the following equation will be used in the remainder of this text:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \begin{bmatrix} x_{\mathcal{R}} \\ y_{\mathcal{R}} \\ 1 \end{bmatrix} \tag{3.3}$$

with f_x and f_y being the focal length measured in width and height of the pixels, and s a factor accounting for the skew due to non-rectangular pixels. The above upper triangular matrix is called the *calibration matrix* of the camera; and the notation $\mathbf{K}$ will be used for it. So, the following equation describes the transformation from retinal coordinates to image coordinates.

$$\mathtt{m} = \mathbf{K}\mathtt{m}_{\mathcal{R}} \quad . \tag{3.4}$$

For most cameras the pixels are almost perfectly rectangular and thus s is very close to zero. Furthermore, the principal point is often close to the center of the image. These assumptions can often be used, certainly to get a suitable initialization for more complex iterative estimation procedures.

For a camera with fixed optics these parameters are identical for all the images taken with the camera. For cameras which have zooming and focusing capabilities the focal length can obviously change, but also the principal point can vary. An extensive discussion of this subject can for example be found in the work of Willson [173, 171, 172, 174].

3.1.3 Camera motion

Motion of scene points can be modeled as follows

$$\mathtt{M}' = \begin{bmatrix} \mathbf{R} & \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \mathtt{M} \tag{3.5}$$

with $\mathbf{R}$ a rotation matrix and $\mathtt{t} = [t_x t_y t_z]^\top$ a translation vector.

The motion of the camera is equivalent to an inverse motion of the scene and can therefore be modeled as

$$\mathtt{M}' = \begin{bmatrix} \mathbf{R}^\top & -\mathbf{R}^\top \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \mathtt{M}\,, \tag{3.6}$$

with $\mathbf{R}$ and $\mathtt{t}$ indicating the motion of the camera.

3.1.4 The projection matrix

Combining equations (3.2), (3.3) and (3.6) the following expression is obtained for a camera with some specific intrinsic calibration and with a specific position and orientation:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}^\top & \text{-}\mathbf{R}^\top \mathtt{t} \\ 0_3^\top & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

which can be simplified to

$$\mathtt{m} \sim \mathbf{K}[\mathbf{R}^\top \ \text{-}\mathbf{R}^\top \mathtt{t}]\mathtt{M} \tag{3.7}$$

or even

$$\mathtt{m} \sim \mathbf{P}\mathtt{M}\ . \tag{3.8}$$

The 3×4 matrix $\mathbf{P}$ is called the *camera projection matrix*.

Using (3.8) the plane corresponding to a back-projected image line $\mathtt{l}$ can also be obtained: Since $\mathtt{l}^\top \mathtt{m} \sim \mathtt{l}^\top \mathbf{P}\mathtt{M} \sim \Pi^\top \mathtt{M}$,

$$\Pi \sim \mathbf{P}^\top \mathtt{l} \tag{3.9}$$

The transformation equation for projection matrices can be obtained as described in paragraph 2.1.3. If the points of a calibration grid are transformed by the same transformation as the camera, their image points should stay the same:

$$\mathtt{m} \sim \mathbf{P}'\mathtt{M}' \sim \mathbf{P}\mathbf{T}^{-1}\mathbf{T}\mathtt{M} \sim \mathbf{P}\mathtt{M} \tag{3.10}$$

and thus

$$\mathbf{P} \mapsto \mathbf{P}' \sim \mathbf{P}\mathbf{T}^{-1} \tag{3.11}$$

The projection of the outline of a quadric can also be obtained. For a line in an image to be tangent to the projection of the outline of a quadric, the corresponding plane should be on the dual quadric. Substituting equation (3.9) in (2.17) the following constraint $\mathtt{l}^\top \mathbf{P}\mathbf{Q}^* \mathbf{P}^\top \mathtt{l} = 0$ is obtained for $\mathtt{l}$ to be tangent to the outline. Comparing this result with the definition of a conic (2.10), the following projection equation is obtained for quadrics (this results can also be found in [65]). :

$$\mathbf{C}^* \sim \mathbf{P}\mathbf{Q}^*\mathbf{P}^\top\ . \tag{3.12}$$

Relation between projection matrices and image homographies

The homographies that will be discussed here are collineations from $\mathcal{P}^2 \rightarrow \mathcal{P}^2$. A homography $\mathbf{H}$ describes the transformation from one plane to another. A number of special cases are of interest, since the image is also a plane. The projection of points of a plane into an image i can be described through a homography $\mathbf{H}_{\Pi i}$. The matrix representation of this homography is dependent on the choice of the projective basis in the plane.

As an image is obtained by perspective projection, the relation between points $\mathtt{M}_\Pi$ belonging to a plane Π in 3D space and their projections $\mathtt{m}_{\Pi i}$ in the image is mathematically expressed by a homography $\mathbf{H}_{\Pi i}$. The matrix of this homography is found as follows. If the plane Π is given by $\Pi \sim [\pi^\top\ 1]^\top$ and the point $\mathtt{M}_\Pi$ of Π is represented as $\mathtt{M}_\Pi \sim [\mathtt{m}_\Pi^\top\ 1]^\top$, then $\mathtt{M}_\Pi$ belongs to Π if and only if $0 = \Pi^\top \mathtt{M}_\Pi = \pi^\top \mathtt{m}_\Pi + 1$. Hence,

$$\mathtt{M}_\Pi \sim \begin{bmatrix} \mathtt{m}_\Pi \\ 1 \end{bmatrix} = \begin{bmatrix} \mathtt{m}_\Pi \\ -\pi^\top \mathtt{m}_\Pi \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3\times3} \\ -\pi^\top \end{bmatrix} \mathtt{m}_\Pi\ . \tag{3.13}$$

Now, if the camera projection matrix is $\mathbf{P}_i = [\mathbf{A}_i|\mathtt{a}_i]$, then the projection $\mathtt{m}_{\Pi i}$ of $\mathtt{M}_\Pi$ onto the image is

$$\begin{aligned} \mathtt{m}_{\Pi i} \sim \mathbf{P}_i \mathtt{M}_\Pi &= [\mathbf{A}_i|\mathtt{a}_i] \begin{bmatrix} \mathbf{I}_{3\times 3} \\ -\pi^\top \end{bmatrix} \mathtt{m}_\Pi \\ &= [\mathbf{A}_i - \mathtt{a}_i \pi^\top] \mathtt{m}_\Pi \ . \end{aligned} \quad (3.14)$$

Consequently, $\mathbf{H}_{\Pi i} \sim \mathbf{A}_i - \mathtt{a}_i \pi^\top$.

Note that for the specific plane $\Pi_{\mathtt{REF}} = [0\,0\,0\,1]^\top$ the homographies are simply given by $\mathbf{H}_{\mathtt{REF}i} \sim \mathbf{A}_i$.

It is also possible to define homographies which describe the transfer from one image to the other for points and other geometric entities located on a specific plane. The notation $\mathbf{H}^\Pi_{ij}$ will be used to describe such a homography from view i to j for a plane Π. These homographies can be obtained through the following relation $\mathbf{H}^\Pi_{ij} = \mathbf{H}_{\Pi j}\mathbf{H}^{-1}_{\Pi i}$ and are independent to reparameterizations of the plane (and thus also to a change of basis in $\mathcal{P}^3$).

In the metric and Euclidean case, $\mathbf{A}_i = \mathbf{K}_i\mathbf{R}_i^\top$ and the plane at infinity is $\Pi_\infty = [0001]^\top$. In this case, the homographies for the plane at infinity can thus be written as:

$$\mathbf{H}^\infty_{ij} = \mathbf{K}_i \mathbf{R}^\top_{ij} \mathbf{K}_i^{-1} \,, \quad (3.15)$$

where $\mathbf{R}_{ij} = \mathbf{R}_i^\top \mathbf{R}_j$ is the rotation matrix that describes the relative orientation from the j^{th} camera with respect top the i^{th} one.

In the projective and affine case, one can assume that $\mathbf{P}_1 = [\mathbf{I}_{3\times3}|0_3]$ (since in this case $\mathbf{K}_i$ is unknown). In that case, the homographies $\mathbf{H}_{\Pi 1} \sim \mathbf{I}_{3\times 3}$ for all planes; and thus, $\mathbf{H}^{\mathtt{REF}}_{1i} = \mathbf{H}_{\mathtt{REF}i}$. Therefore $\mathbf{P}_i$ can be factorized as

$$\mathbf{P}_i = [\mathbf{H}^{\mathtt{REF}}_{1i}|\mathtt{e}_{1i}] \quad (3.16)$$

where $\mathtt{e}_{1i}$ is the projection of the center of projection of the first camera (in this case, $[0\,0\,0\,1]^\top$) in image i. This point $\mathtt{e}_{1i}$ is called the *epipole*, for reasons which will become clear in Section 3.2.1.

Note that this equation can be used to obtain $\mathbf{H}^{\mathtt{REF}}_{1i}$ and $\mathtt{e}_{1i}$ from $\mathbf{P}_i$, but that due to the unknown relative scale factors $\mathbf{P}_i$ can, in general, not be obtained from $\mathbf{H}^{\mathtt{REF}}_{1i}$ and $\mathtt{e}_{1i}$. Observe also that, in the affine case (where $\Pi_\infty = [0001]^\top$), this yields $\mathbf{P}_i = [\mathbf{H}^\infty_{1i}|\mathtt{e}_{1i}]$.

Combining equations (3.14) and (3.16), one obtains

$$\mathbf{H}^\Pi_{1i} = \mathbf{H}^{\mathtt{REF}}_{1i} - \mathtt{e}_{1i}\pi^\top \quad (3.17)$$

This equation gives an important relationship between the homographies for all possible planes. Homographies can only differ by a term $\mathtt{e}_{1i}[1 - \pi']^\top$. This means that in the projective case the homographies for the plane at infinity are known up to 3 common parameters (i.e. the coefficients of π_∞ in the projective space).

Equation (3.16) also leads to an interesting interpretation of the camera projection matrix:

$$\mathtt{m}_1 \sim [\mathbf{I}_{3\times3}|0_3] \begin{bmatrix} \mathtt{m} \\ 1 \end{bmatrix} = \mathtt{m} \quad (3.18)$$

$$\mathtt{m}_i \sim [\mathbf{H}^{\mathtt{REF}}_{1i}|\mathtt{e}_{1i}] \begin{bmatrix} \mathtt{m} \\ 1 \end{bmatrix} = \mathbf{H}^{\mathtt{REF}}_{1i}\mathtt{m} + \mathtt{e}_{1i} \quad (3.19)$$

$$= \lambda \mathbf{H}^{\mathtt{REF}}_{1i}\mathtt{m}_1 + \mathtt{e}_{1i} = \mathbf{P}_i(\lambda \begin{bmatrix} \mathtt{m}_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0_3 \\ 1 \end{bmatrix}) \quad (3.20)$$

In other words, a point can thus be parameterized as being on the line through the optical center of the first camera (i.e. $[0001]^\top$) and a point in the reference plane $\Pi_{\mathtt{REF}}$. This interpretation is illustrated in Figure 3.4.

3.1.5 Deviations from the camera model

The perspective camera model describes relatively well the image formation process for most cameras. However, when high accuracy is required or when low-end cameras are used, additional effects have to be taken into account.

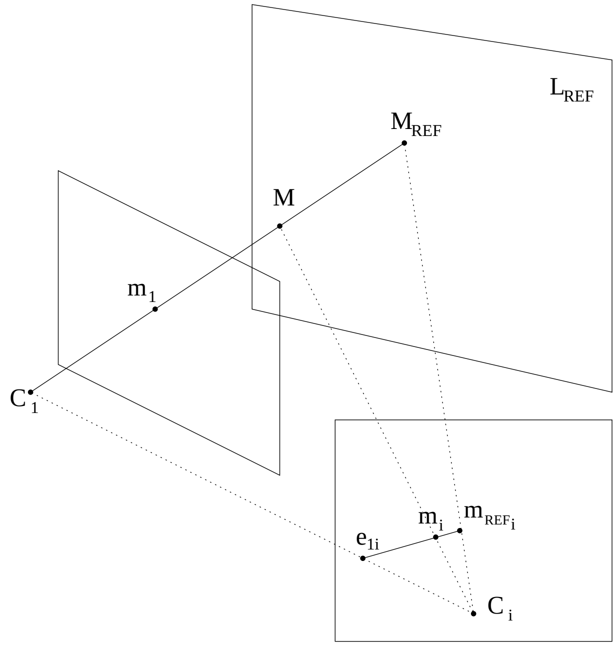

Figure 3.4: *A point* $\mathtt{M}$ *can be parameterized as* $\mathtt{C}_1 + \lambda \mathtt{M}_{\mathtt{REF}}$. *Its projection in another image can then be obtained by transferring* $\mathtt{m}_1$ *according to* $\Pi_{\mathtt{REF}}$ *(i.e. with* $\mathbf{H}_{1i}^{\mathtt{REF}}$*) to image* i *and applying the same linear combination with the projection* $\mathtt{e}_{1i}$ *of* $\mathtt{C}_1$ *(i.e.* $\mathtt{m}_i \sim \mathtt{e}_{1i} + \lambda \mathbf{H}_{1i}^{\mathtt{REF}} \mathtt{m}_1$*).*

The failures of the optical system to bring all light rays received from a point object to a single image point or to a prescribed geometric position should then be taken into account. These deviations are called aberrations. Many types of aberrations exist (e.g. astigmatism, chromatic aberrations, spherical aberrations, coma aberrations, curvature of field aberration and distortion aberration). It is outside the scope of this work to discuss them all. The interested reader is referred to the work of Willson [173] and to the photogrammetry literature [137].

Many of these effects are negligible under normal acquisition circumstances. Radial distortion, however, can have a noticeable effect for shorter focal lengths. Radial distortion is a linear displacement of image points radially to or from the center of the image, caused by the fact that objects at different angular distance from the lens axis undergo different magnifications.

It is possible to cancel most of this effect by warping the image. The coordinates in undistorted image plane coordinates (x, y) can be obtained from the observed image coordinates (x_o, y_o) by the following equation:

$$\begin{aligned} x &= x_o + (x_o - c_x)(K_1 r^2 + K_2 r^4 + \ldots) \\ y &= y_o + (y_o - c_y)(K_1 r^2 + K_2 r^4 + \ldots) \end{aligned} \qquad (3.21)$$

where K_1 and K_2 are the first and second parameters of the radial distortion and

$$r^2 = (x_o - c_x)^2 + (y_o - c_y)^2 \ .$$

Note that it can sometimes be necessary to allow the center of radial distortion to be different from the principal point [174].

When the focal length of the camera changes (through zoom or focus) the parameters K_1 and K_2 will also vary. In a first approximation this can be modeled as follows:

$$\begin{aligned} x &= x_o + (x_o - c_x)(K_{f1} \tfrac{r^2}{f^2} + K_{f2} \tfrac{r^4}{f^4} + \ldots) \\ y &= y_o + (y_o - c_y)(K_{f1} \tfrac{r^2}{f^2} + K_{f2} \tfrac{r^4}{f^4} + \ldots) \end{aligned} \qquad (3.22)$$

Due to the changes in the lens system this is only an approximation, except for digital zooms where (3.22) should be exactly satisfied.