Part I

Theory

Chapter 2

Camera Model

2.1 Introduction

The goal of machine vision is to understand the visible world by inferring 3-D properties from 2-D images. This requires a model describing how a camera and its focalization lens form images.

In this chapter we will firstly present the functioning of the camera and its lens system (Section 2.2). Note that because we used only charge-coupled device (CCD) cameras in our work, we assume that the sensor is of this type. The functioning of the electronics associated with such a device is beyond the scope of the thesis, but the interested reader will find more details in [3].

We will then specify some vocabulary and notations (Section 2.3) and define the camera parameters (Section 2.4), useful to establish the passage equation between the image plane and the scene (Sections 2.5 and 2.6).

Section 2.7 closes this chapter with some considerations about the focalization lens.

Interesting reviews of the evolution of camera calibration methods and models are described in [4] and [5].

2.2 Camera and Lens Models

The 3-D scene visualized by the camera will be projected on the CCD sensor plane—termed *image plane*—by mean of a focalization lens. It is typically constructed from a series of individual spherical glass or plastic lenses and stops centered on a common axis, passing by the center of each lens element and called the *optical axis*. This makes the lens and all its imaging properties radially symmetric [6] [7].

A stop is an opaque element with a roughly circular opening to permit the passage of light. The element that most limits the angular spread of the bundle of rays that will pass unobstructed through the lens system is termed the *aperture stop*. Its size is typically set through the use of an adjustable diaphragm, and mostly serves to provide control over the quantity of light striking the image plane [6].

The composition of a lens system strongly varies with the type of objective. On Fig. 2.1 we present as an example a double-Gauss lens with one stop (i.e. the diaphragm). Moving the left group of lenses will adjust the zooming factor of the system. Focusing will be achieved by adjusting the right group of lenses.

The focal point of the system, that has by definition the property that any ray coming from it, or proceeding toward it, travels parallel to the optical axis after refraction, can be found by combination of each individual lens focal point.

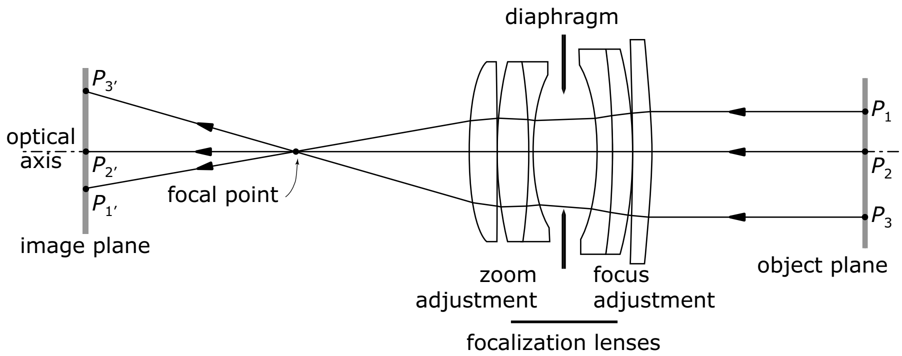

Fig. 2.1. Focalization lens, realistic model.

Only the scene points that are located on a specific plane—the *object plane*—parallel to the image plane are projected sharply (i.e. well focused) on the CCD matrix.

This system is complex to model so, in order to have computationally efficient, closed-form equations, we introduce simpler models, based on simplifications or abstractions of the true image-formation process of the lens [7]. The two most common abstract models are the *thin lens model* and the *pinhole model*.

2.2.1 The thin lens model

In the *thin lens model*, the optical system consists of an ideal (i.e. without any projection error) infinitely thin lens, with a focal length f, representing the perpendicular distance between the center of the lens and its focal point (see Fig. 2.2) [3].

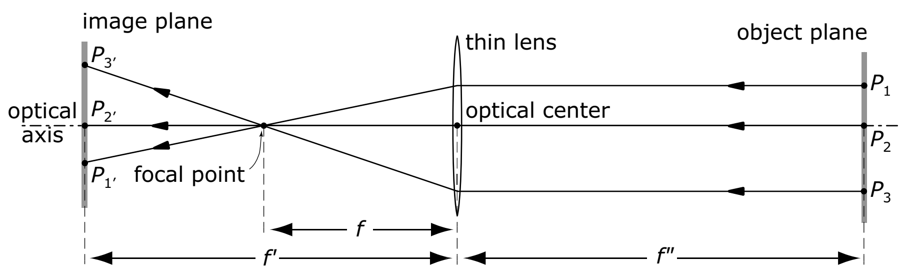

Fig. 2.2. Focalization lens, thin lens model.

The point at the intersection between the lens and the optical axis is defined as the *optical center*, at which the ray of light from a punctual source crosses the lens undeviated.

The object plane is situated at a distance f'', called *object width*, from the optical center. The distance f' between the optical center and the image plane is called *image width*.

The connection between f, f' and f'' is given by the famous Gaussian lens formula, holding that:

$$\frac{1}{f} = \frac{1}{f'} + \frac{1}{f''} \tag{2.1}$$

2.2.2 The pinhole model

If we focus the lens at infinity, $f'' = \infty$ and therefore the focal plane coincides with the image plane ($f = f'$). Otherwise $f \neq f'$ but in practice, f'' is much greater than f so that we can use the approximation $f \cong f'$.

It is thus not necessary to differentiate focal and image planes. This leads to the *pinhole model*, an ideal model supposing no lens (and thus no focusing) but an infinitesimally small hole situated on the optical center, through which light enters before forming an inverted image on the camera CCD sensor facing the hole. The mapping of three dimensions onto two is made by mean of a central projection (see Fig. 2.3) [8].

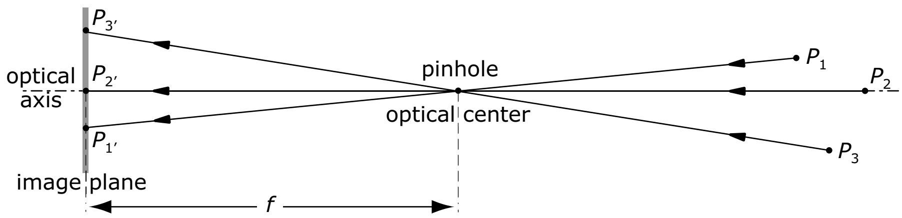

Fig. 2.3. Focalization lens, pinhole model.

Even though there is no focal plane in that model, for sake of easiness, we will continue to use the term *focal length* to mean the perpendicular distance from the image plane to the optical center, and will note it f.

So far we have considered the optical components of the focalization lens to be perfectly aligned on the same axis. In real lenses, things are not so easy. For a simple lens element we must differentiate the optical and mechanical axes of symmetry (see Fig. 2.4). The first one is the straight line joining the centers of curvature of the two lens surfaces, the second one is determined during the manufacture by the centerline of the machine used to grind the lens' edge.

Ideally the two axes coincide but in practice they are distant from an angle ω, called *decentration* [7].

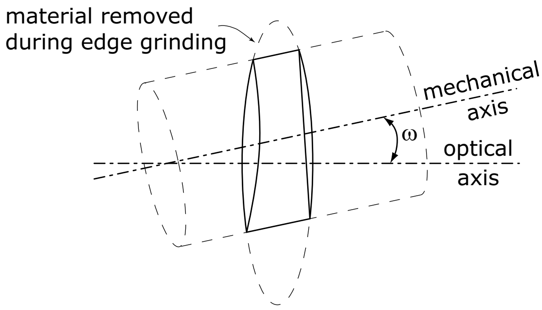

Fig. 2.4. *Decentration* for a simple lens.

Given this, the alignment between the elements mounted together to form the complete lens is generally not feasible. The consequence is that there is no 'ideal' optical axis for the lens.

Decentration and misalignment in the lens produce *distortion*, i.e. the displacement of an image point from the position that is predicted by a perfect pinhole or thin-lens model [7]. It can be either tangential or radial, but in practice only radial distortion will be considered (we will later see how), the tangential term being negligible.

2.3 Vocabulary and Notations

The different coordinate systems and notations are defined on Fig. 2.5.

We call[1]:

- *scene*: the camera field of view. Its associated orthonormal coordinate system is (x, y, and z), attached to the camera. Authors use to distinguish the latter from a second coordinate system (x_w, y_w, and z_w), attached to the corner of a room or to a particular object, but since we are focusing on ambulatory applications, it would not be pertinent. We thus use only one coordinate system (x, y, and z), and name it *world coordinate system* (WCS);

- *camera plane* : the (x, y) plane;

[1] Some of the following terms have already been defined. We remind them though, for the reader who would like to use this section as a glossary.

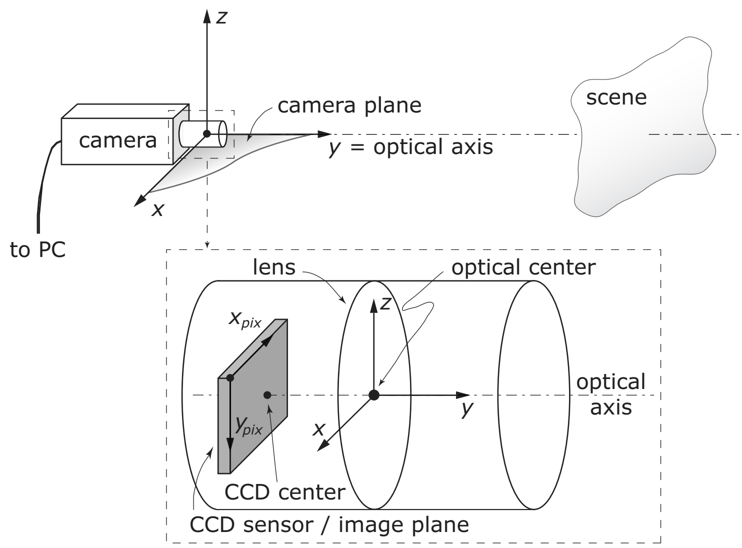

Fig. 2.5. Vocabulary and notations. 3-D view.

- *image plane* or *image*: the CCD sensor plane which is, after inversion, the image on the screen. Its discrete associated coordinate system is (x_{pix}, y_{pix}), or *image coordinate system* (ICS). y_{pix} is assumed to be parallel to the z axis in the WCS;
- *pixel*: the smallest discriminable entity in the image plane;
- *resolution*: the amount of *pixels* composing the CCD sensor in both x_{pix} and y_{pix} directions;
- *optical center*: the point at which the ray of light from a punctual source crosses the lens undeviated. We choose it as the origin of the WCS;
- *optical axis*: the y axis of the WCS, perpendicular to the CCD sensor and passing by the optical center;
- *CCD center* or *image center*: the point at the intersection between the optical axis and the image plane.

2.4 Camera Parameters

The parameters of our camera model are (see Fig. 2.6):

- the focal length f;
  Range of values: 5 . . . 200 mm.

- the dimensions a_x and a_y of a pixel in the x_{pix} and y_{pix} directions ($a_x = a_y$ if the pixels are square);
  Range of values: 5...10 μm.

- the normalized focal lengths f_x and f_y, obtained by dividing f by a_x and a_y respectively; they represent the same physical distance as f, but this time expressed in number of pixels in the x_{pix} and y_{pix} directions respectively ($f_x = f_y$ if the pixels are square). In practice we will use those two parameters instead of f, a_x and a_y;
  Range of values: 1000...30000 pixels, depending on the resolution of the camera.

- the CCD center coordinates $X_{pix,0}$ and $Y_{pix,0}$, expressed in number of pixels;
  Range of values: depends on the image size and on the focalization lens characteristics.

- the ICS' lack of orthogonality σ (usually close to zero). In practice we will rather use the skewness factor γ, expressed in number of pixels, that can be written as [9]:

$$\gamma = \frac{f}{a_y} \tan\sigma = f_y \tan\sigma \tag{2.2}$$

  Range of values for γ : -2...2 pixels.

- the two lens distortion coefficients k_1 and k_2 (still to be defined).
  Range of values: -0.5...0.5.

Eventually only seven parameters are sufficient to describe the camera and lens geometries : f_x, f_y, $X_{pix,0}$, $Y_{pix,0}$, γ, k_1 and k_2. These are called the *intrinsic parameters*.

2.5 Passage Equations

The *pinhole model* we are relying on is based on central projection, so that the passage equations between the WCS and the ICS will be derived from the similar triangles properties. On Fig. 2.7, where we would like to determine the y_{pix} coordinate of point P, from its WCS coordinates, we obtain the second relation in (2.3).

Applying the same principle on the x_{pix} coordinate, taking into account the skewness γ, provides us with the first equation in (2.3). These two relations constitute the camera ideal model [8] [10] [11]:

$$\begin{aligned} x_{pix} &= f_x \tfrac{x}{y} + \gamma \tfrac{z}{y} + X_{pix,0} \\ y_{pix} &= f_y \tfrac{z}{y} + Y_{pix,0} \end{aligned} \tag{2.3}$$

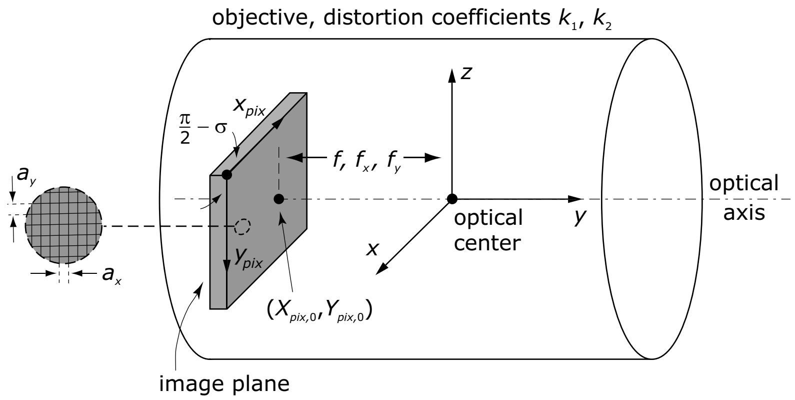

Fig. 2.6. Camera parameters.

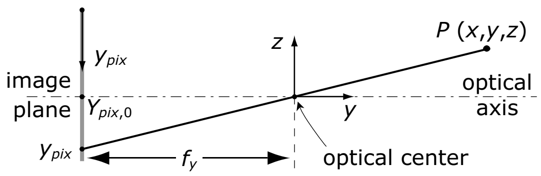

Fig. 2.7. y_{pix} determination. View in the (y, z) plane.

If we define the camera intrinsic matrix A as

$$A = \begin{pmatrix} f_x & X_{pix,0} & \gamma \\ 0 & Y_{pix,0} & f_y \\ 0 & 1 & 0 \end{pmatrix} \tag{2.4}$$

we obtain the matrix form of the camera model:

$$\begin{pmatrix} x_{pix} \\ y_{pix} \\ 1 \end{pmatrix} = \lambda A \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{2.5}$$

where λ is a normalization factor, equal to $1/y$.

However, this ideal model does not take into account the distortion introduced by the objective lens, which is mostly radial (terms of tangential distortion will be neglected). The classic approach models it using a parametric function d, defined as:

$$d = k_1 r^2 + k_2 r^4 \tag{2.6}$$

where k_1 and k_2 are the two distortion coefficients and

$$r^2 = \frac{x^2 + z^2}{y^2} \tag{2.7}$$

The complete model is then [11]:

$$\begin{aligned} x_{pix} &= f_x (1+d) \tfrac{x}{y} + \gamma (1+d) \tfrac{z}{y} + X_{pix,0} \\ y_{pix} &= f_y (1+d) \tfrac{z}{y} + Y_{pix,0} \end{aligned} \tag{2.8}$$

Note that for more complex systems—beyond the scope of this thesis—it has been shown in [12] that a nonparametric model (also called "the plane method") outperforms the previous technique.

Equations (2.8) are heavy to carry out and the only way to solve them is numerical. In practice, the passage from the image plane to the scene will thus be done in two separated operations: the first one will correct the considered pixel distortion, by applying an iterative procedure described in Section 2.6; the second one will make the transposition to the (x, y, and z) coordinate system.

The inverse passage (i.e. from the scene to the image plane) will be useful for theoretical aspects only, for which working on an ideal "undistorted" image is sufficient.

2.6 Correction of the Distortion

As said above, we use an iterative procedure to correct the distortion. The algorithm is to be applied on one particular pixel at a time, since reconstructing the whole undistorted image would be very complex, and unnecessary in our case (only a limited amount of pixels—those belonging to the laser curve(s)—are of interest).

At iteration i, we firstly calculate the normalized coordinates z/y and x/y of the considered pixel:

$$\begin{aligned} \left.\frac{z}{y}\right|_i &= \frac{\left.y^*_{pix}\right|_{i-1}}{f_y} \\ \left.\frac{x}{y}\right|_i &= \frac{\left.x^*_{pix}\right|_{i-1} - \gamma \left.\frac{z}{y}\right|_i}{f_x} \end{aligned} \tag{2.9}$$

where we set:

$$\begin{aligned} x^*_{pix} &= x_{pix} - X_{pix,0} \\ y^*_{pix} &= y_{pix} - Y_{pix,0} \end{aligned} \tag{2.10}$$

Distortion d is then evaluated [see (2.6)]:

$$\left.d\right|_i = k_1 \left(\left(\left.\frac{x}{y}\right|_i \right)^2 + \left(\left.\frac{z}{y}\right|_i \right)^2 \right) + k_2 \left(\left(\left.\frac{x}{y}\right|_i \right)^2 + \left(\left.\frac{z}{y}\right|_i \right)^2 \right)^2 \tag{2.11}$$

which allows us to update the considered pixel coordinates:

$$\begin{aligned} \left.x^*_{pix}\right|_i &= \left(f_x \left.\frac{x}{y}\right|_i + \gamma \left.\frac{z}{y}\right|_i \right) (1 + \left.d\right|_i) \\ \left.y^*_{pix}\right|_i &= f_y \left.\frac{z}{y}\right|_i (1 + \left.d\right|_i) \end{aligned} \tag{2.12}$$

To stop iterating, we choose a criterion of the type:

$$\left(\left.x^*_{pix}\right|_{i-1} - \left.x^*_{pix}\right|_i \right)^2 + \left(\left.y^*_{pix}\right|_{i-1} - \left.y^*_{pix}\right|_i \right)^2 \leq c \tag{2.13}$$

where constant c is fixed regarding the desired accuracy on the operation. We fixed it to 10^{-4}.

Fig. 2.8 shows a distorted image (left) and its undistorted version (right).

2.7 Considerations on the Focalization Lens

A few comments are to be made about the focalization lens, concerning focusing and chromatic aberration.

2.7.1 Focusing

Even if in the *pinhole model*, the focal plane coincide with the image plane, changing the lens focus implies a shift of the optical center regarding the CCD sensor. The focal length f thus changes.

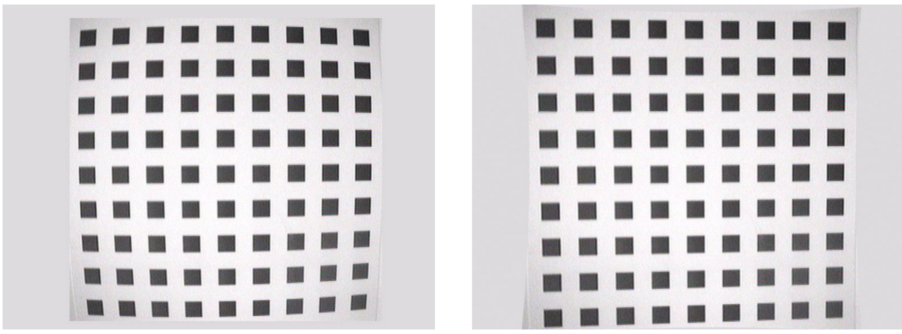

Fig. 2.8. Distorted image (left) and its undistorted version (right).

If we firstly focus the lens at a distance d_1 from the optical center, then at another distance d_2, the resulting focal length shift Δf will be, by the Gaussian lens law (2.1):

$$\Delta f = f\left(\frac{d_2}{f-d_2} - \frac{d_1}{f-d_1}\right) \tag{2.14}$$

Let us suppose $f = 8$ mm (we will see that it represents a common value for our applications) and an object-to-camera distance going from 30 cm to 2 m. The focal length shift could go up to 187 μm if we wanted to adjust the focus for each measurement. It represents more than 2% of the focal length, which is unacceptable.

Moreover, a focus length shift modifies the value of the distortion parameters, according to the approximate law [9]

$$k_1 = \frac{k_{f_1}}{f^2} \quad ; \quad k_2 = \frac{k_{f_2}}{f^4} \tag{2.15}$$

where k_{f_1} and k_{f_2} are two constants.

It is thus very important to not change the focus after having calibrated the camera (see Section 6.2). A good method will be to calibrate it with the lens focused at a distance situated at the middle of the desired range of measurement.

2.7.2 Chromatic aberration

The index of refraction of optical components varies as a function of the wavelength of the incident light. As a consequence, the different wavelengths of light are refracted to different degrees by the elements of the focalization lens. For instance, given a simple uncorrected thin lens with incident white light rays,

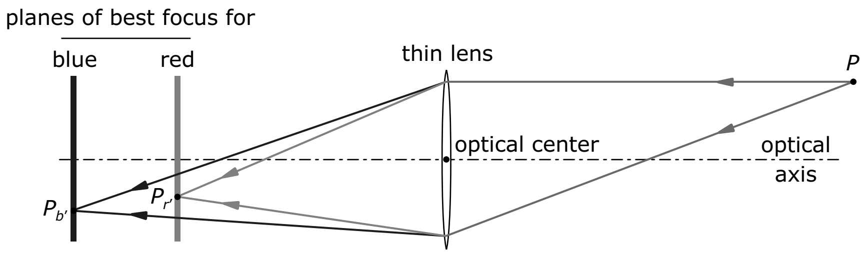

Fig. 2.9. Chromatic aberration in a thin lens.

their blue components will be brought to a focus closer to the lens than the red ones (see Fig. 2.9) [7].

This phenomenon is called chromatic aberration and is an intrinsic property of a camera lens. Its effects can be nonnegligible but we will nevertheless not consider them since—as we will see in Chapter 4—a narrow-band interference filter is systematically placed in front of the camera CCD sensor. The rays of light hitting the matrix are thus almost monochromatic (the typical filter bandwidth is of 10 nm).