Slimane LARABI

Computer Vision

From Bidimensional Images to Three Dimensional Scene

January 02, 2025

Springer Nature

Chapter 1

Camera Calibration and its Applications

1.1 Introduction

Camera calibration is a foundational step in computer vision and image processing that ensures accurate mapping between the three-dimensional world and its two-dimensional representation. By estimating the internal and external parameters of a camera, calibration enables precise reconstruction of spatial relationships, measurement of real-world dimensions, and correction of lens distortions. This process plays a critical role in a variety of applications that require high accuracy and reliability in geometric understanding and need some skills in geometry and algebra in order to find accurately the mathematical model corresponding to the image formation.

There are many applications where the camera need to be calibrated in order to get the 3D information. We can cite:

- **3D Reconstruction**: Accurate calibration is crucial for reconstructing 3D models of objects and scenes from multiple images, enabling applications in archaeology, architecture, and virtual reality. Figure 1.1 illustrates a stereoscopic system where two images, acquired by two calibrated cameras, allow for scene reconstruction.

- **Robotics**: Robots rely on calibrated cameras for navigation, object detection, and interaction within their environments. Camera calibration ensures precise localization and manipulation. Figure 1.2 shows a robot equipped with a stereoscopic system that can navigate a scene with obstacles using a map computed from 3D scene information provided by the embedded cameras.

- **Augmented Reality (AR)**: By aligning virtual objects with the real world, camera calibration improves the realism and accuracy of AR experiences.

- **Autonomous Vehicles**: In self-driving cars, calibration ensures that cameras capture accurate spatial data for tasks like lane detection, object tracking, and depth estimation.

- **Photogrammetry**: For mapping and surveying, calibrated cameras provide precise measurements, enabling the creation of detailed maps and terrain models.

- **Medical Imaging**: In fields like endoscopy and surgery, calibration ensures accurate spatial measurements and alignment with preoperative data.

Starting from two or more images of 3D scene acquired scene by a camera, we can recover the three dimensional structure only if the internal and external camera parameters are known. The internal parameters are:

- Focal length: f
- Position of impact of optical axis (Ox, Oy)
- Dimensions of the pixel (mm) ex, ey.

and external parameters are :

- The position and Orientation of the camera coordinate frame relatively to the world coordinate frame (OXYZ).

In this chapter, we study the following elements:

- The linear camera model for image formation,
- The camera calibration, extracting intrinsic and extrinsic matrices,
- The simple stereo,
- The depth computation.

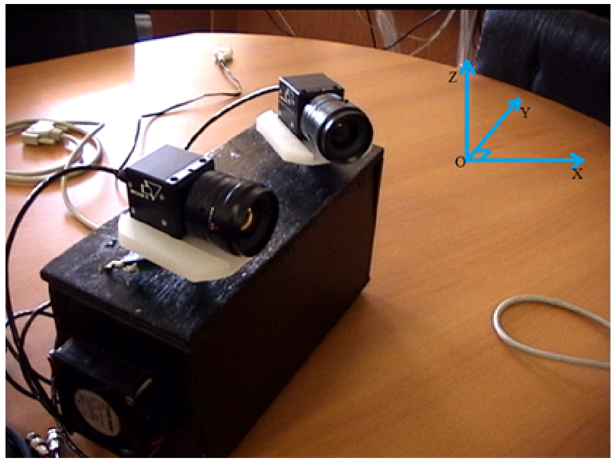

Fig. 1.1 The stereoscopic system developed in my laboratory with two rotating camera.

Fig. 1.2 A Pioneer 3DX, embedded with a stereoscopic system.

1.2 Linear camera model for image formation

1.2.1 Integrating Internal Parameters into the Image Formation Model

It is easy to choose a world coordinate frame, denoted as $(W\hat{X}_w\hat{Y}_w\hat{Z}_w)$, attached to an object in the scene. This allows for determining the coordinates of any $3D$ point of the object with respect to that coordinate system. Figure 1.3 shows an example of a selected point P_W with coordinates $(0,2,3)$.

We assume that the camera coordinate frame $(C\hat{X}_c\hat{Y}_c\hat{Z}_c)$ is fixed such that the plane defined by the two axes $(C\hat{X}_c, C\hat{Y}_c)$ is parallel to the image plane, and $(C\hat{Z}_c)$

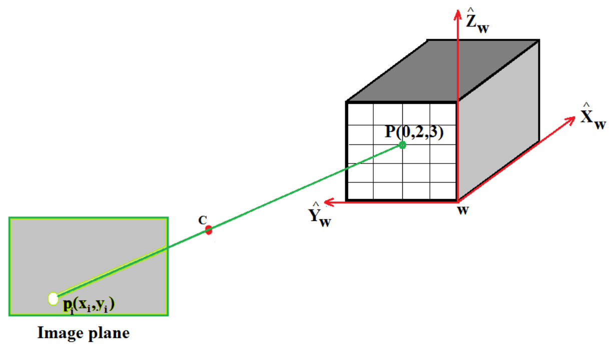

Fig. 1.3 Forward imaging model: 3D to 2D. The three-dimensional point $P(x_w, y_w, z_w)$ is projected through the projection center C onto the two-dimensional point $p_i(x_i, y_i)$.

coincide with optical axis (see Figure 1.4). If the coordinates of the point $P_c = P(x_c, y_c, z_c)$ with respect to camera coordinate frame $(C\hat{X}_c\hat{Y}_c\hat{Z}_c)$, are known (see Figure 1.4), then the coordinates of its projection $p_i(x_i, y_i)$ onto the image plane may be computed using geometric properties.

First, we give the expression of points p_i, P_C, P_W using the homogeneous coordinates.

$$\tilde{p_i} = \begin{bmatrix} \tilde{x_i} \\ \tilde{y_i} \\ \tilde{z_i} \end{bmatrix}, \tilde{P_c} = \begin{bmatrix} \tilde{x_c} \\ \tilde{y_c} \\ \tilde{z_c} \end{bmatrix}, \tilde{P_w} = \begin{bmatrix} \tilde{x_w} \\ \tilde{y_w} \\ \tilde{z_w} \end{bmatrix} \tag{1.1}$$

If we apply Thales's Theorem on triangles of figure 1.4, we obtain:

$$\frac{x_i}{f} = \frac{x_c}{z_c}, \frac{y_i}{f} = \frac{y_c}{z_c} \tag{1.2}$$

Therefore, we obtain:

$$x_i = f \times \frac{x_c}{z_c}, y_i = f \times \frac{y_c}{z_c} \tag{1.3}$$

If we assume that e_x, e_y are the pixel densities (pixels/mm) in $\hat{x}$ and $\hat{y}$ directions (see Figure 1.5), the coordinates of the pixel are (x_i, y_i) where:

$$x_i = e_x \times f \times \frac{x_c}{z_c} \tag{1.4}$$

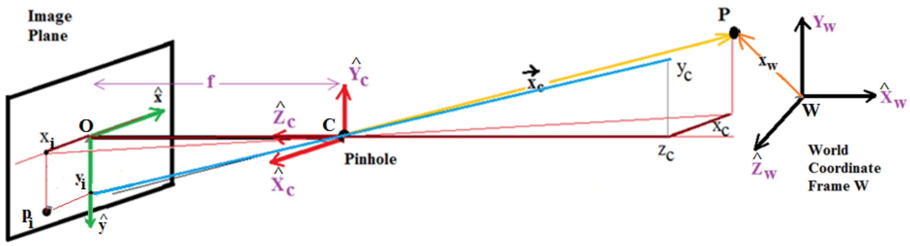

Fig. 1.4 Geometric modeling of image formation.

$$y_i = e_y \times f \times \frac{y_c}{z_c} \tag{1.5}$$

Let (O_x, O_y) be the coordinates of the principle point O with respect to the top left corner of image plane (see figure 1.5). We can the write:

$$x_i = e_x \times f \times \frac{x_c}{z_c} + O_x \tag{1.6}$$

$$y_i = e_y \times f \times \frac{y_c}{z_c} + O_y \tag{1.7}$$

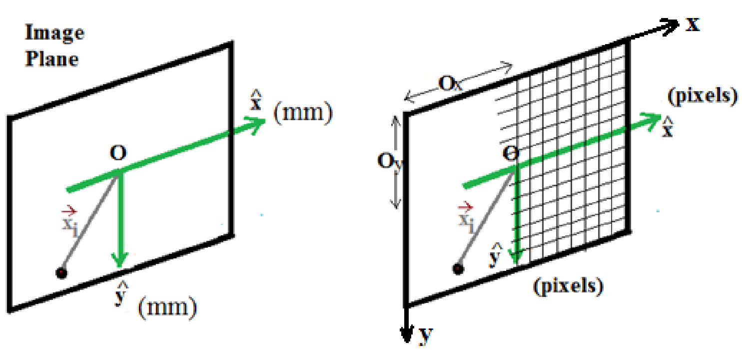

Fig. 1.5 Integrating the impact point coordinates into the x_i, y_i mathematical expression.

Let $f_x = e_x \times f$, $f_y = e_y \times f$ be the focal lengths in pixels in $\hat{x}$ and $\hat{y}$ directions. We can then write the non linear equations for perspective projection:

$$x_i = f_x \times \frac{x_c}{z_c} + O_x, y_i = f_y \times \frac{y_c}{z_c} + O_y \tag{1.8}$$

The intrinsic parameters of the camera are: f_x, f_y, O_x, O_y.

The Homogeneous coordinates of a $2D$ point $p_i(x_i, y_i)$ is a 3D point $(\tilde{x_i}, \tilde{y_i}, \tilde{z_i})$ such that

$$x_i = \frac{\tilde{x_i}}{\tilde{z_i}}, y_i = \frac{\tilde{y_i}}{\tilde{z_i}} \tag{1.9}$$

From figure 1.6, each point belonging to (L) has the Homogeneous coordinates of the $2D$ point $p(x_i, y_i)$.

$$p \equiv \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} \tilde{z_i} x_i \\ \tilde{z_i} y_i \\ \tilde{z_i} \end{bmatrix} \equiv \begin{bmatrix} \tilde{x_i} \\ \tilde{y_i} \\ \tilde{z_i} \end{bmatrix} \equiv \tilde{p_i} \tag{1.10}$$

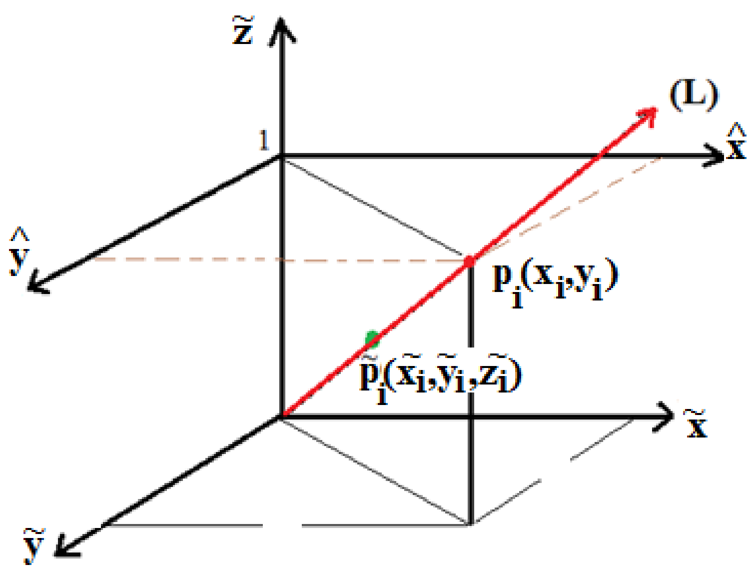

Fig. 1.6 Homogeneous coordinates of the point p_i.

Rewriting the equations of perspective projection using the homogeneous coordinates of (x_i, y_i), we obtain:

$$p_i \equiv \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x x_c + z_c O_x \\ f_y y_c + z_c O_y \\ z_c \end{bmatrix} = \begin{bmatrix} f_x & 0 & O_x & 0 \\ 0 & f_y & O_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \tag{1.11}$$

The matrix including internal parameters is called intrinsic matrix M_{int} and is given by:

$$M_{int} = \begin{bmatrix} f_x & 0 & O_x \\ 0 & f_y & O_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1.12}$$

It is then possible to compute the coordinates (in pixels) of the image point p_i if the intrinsic matrix has been estimated a priori and the $3D$ coordinates (x_c, y_c, z_c) of the point P_C are known. However, the coordinates of the point $P(x_c, y_c, z_c)$ is unknown with respect to the camera coordinate frame $(C\hat{X}_c\hat{Y}_c\hat{Z}_c)$, because the position of the projection center C (Pinhole) is unknown with respect to the world coordinate frame $(W\hat{X}_w\hat{Y}_w\hat{Z}_w)$ (see figure 1.4). Consequently, we need to write the coordinate of P_W (known with respect to world coordinate frame) with respect to camera coordinate frame. We study in the next subsection, how we can do this task.

1.2.2 Integrating external parameters in image formation model

As seen previously, to get the the coordinate of P_C (whose coordinates are known with respect to world coordinate frame) with respect to camera coordinate frame, we need to estimate the following external parameters:

- The vector $\overrightarrow{C_W}$ which indicates the position of C with respect to the world coordinate frame (see figure 1.4).

- The orientation of the camera (with respect to the world coordinate frame) defined by the rotation matrix R:

- r_{11}, r_{12}, r_{13} is the direction of $\hat{x}_c$ in the world coordinate frame,

- r_{21}, r_{22}, r_{23} is the direction of $\hat{y}_c$ in the world coordinate frame,
- r_{31}, r_{32}, r_{33} is the direction of $\hat{z}_c$ in the world coordinate frame.

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{1.13}$$

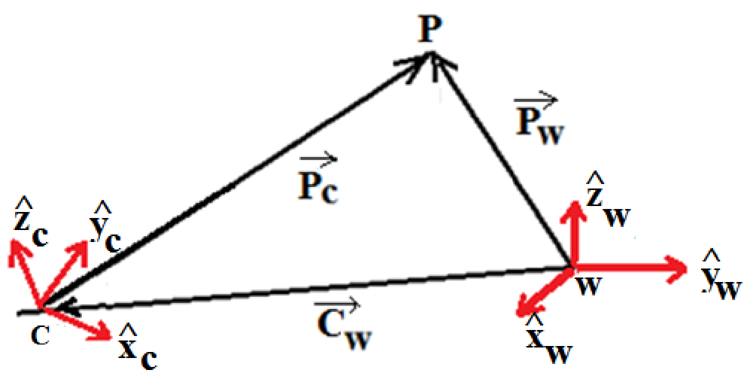

Fig. 1.7 Geometric relationship between C, P and W.

From Figure 1.7, we can write the expression of P'_c with coordinates x, y, z with respect to the world coordinates frame using the following equation:

$$P'_c = RP_c = R(P_w - C_w) = RP_w - RC_w = RP_w + t \tag{1.14}$$

where: $t = -RC_w$

$$P'_c = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{1.15}$$

We can write this last equation as follow:

$$\tilde{P_c'} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1.16}$$

This expression including the matrix of rotation and translation parameters, is called the external matrix of calibration, we write the previous equation as follow:

$$\tilde{P_c'} = M_{ext}\tilde{P_w} \tag{1.17}$$

Projection matrix Pr is then computed as follow:

$$\tilde{p_i} = M_{int}\tilde{P_c} = M_{int}M_{ext}\tilde{P_w} = P\tilde{P_w} \tag{1.18}$$

$$Pr = M_{int}M_{ext} \tag{1.19}$$

We have now expressed the $2D$ coordinates of the image point $p_i(x_i, y_i)$ using the projective matrix and the three coordinates of the corresponding $3D$ point P. The values of x_i, y_i can only be estimated once the internal and external parameters of the camera, which define the projection matrix, are determined. The method used to perform this task is presented in the following section

1.3 Camera calibration

Camera calibration means the estimation of the projection matrix Pr. For this, we need to operate as follow:

- Capture an image of an object with known geometry.
- Identify the correspondence between $3D$ scene points $P_W(x_w^i, y_w^i, z_w^i)$ and their image points.
- For each pair of corresponding points (scene-image) we get two equations:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} Pr_{11} & Pr_{12} & Pr_{13} & Pr_{14} \\ Pr_{21} & Pr_{22} & Pr_{23} & Pr_{24} \\ Pr_{31} & Pr_{32} & Pr_{33} & Pr_{34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{1.20}$$

We obtain then:

$$x_i = \frac{Pr_{11}x_w + Pr_{12}y_w + Pr_{13}z_w + Pr_{14}}{Pr_{31}x_w + Pr_{32}y_w + Pr_{33}z_w + Pr_{34}}, y_i = \frac{Pr_{21}x_w + Pr_{22}y_w + Pr_{23}z_w + Pr_{24}}{Pr_{31}x_w + Pr_{32}y_w + Pr_{33}z_w + Pr_{34}} \tag{1.21}$$

We can write these equations for n pairs of points and obtain a system of equations noted $A.Pr = 0$, as follow:

$$\begin{pmatrix} x_w^1 & y_w^1 & z_w^1 & 1 & 0 & 0 & 0 & 0 & -x_1x_w^1 & -x_1y_w^1 & -x_1z_w^1 & -x_1 \\ 0 & 0 & 0 & 0 & x_w^1 & y_w^1 & z_w^1 & 1 & -y_1x_w^1 & -y_1y_w^1 & -y_1z_w^1 & -y_1 \\ x_w^2 & y_w^2 & z_w^2 & 1 & 0 & 0 & 0 & 0 & -x_2x_w^2 & -x_2y_w^2 & -x_2z_w^2 & -x_2 \\ 0 & 0 & 0 & 0 & x_w^2 & y_w^2 & z_w^2 & 1 & -y_2x_w^2 & -y_2y_w^2 & -y_2z_w^2 & -y_2 \\ .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. & .. \\ x_w^n & y_w^n & z_w^n & 1 & 0 & 0 & 0 & 0 & -x_nx_w^n & -x_ny_w^n & -x_nz_w^n & -x_n \\ 0 & 0 & 0 & 0 & x_w^n & y_w^n & z_w^n & 1 & -y_nx_w^n & -y_ny_w^n & -y_nz_w^n & -y_n \end{pmatrix} \begin{bmatrix} Pr_{11} \\ Pr_{12} \\ Pr_{13} \\ Pr_{14} \\ Pr_{21} \\ Pr_{22} \\ Pr_{23} \\ Pr_{24} \\ Pr_{31} \\ Pr_{32} \\ Pr_{33} \\ Pr_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.22}$$

The projection matrix Pr is defined only up to a scale because Pr and $k.Pr$ produce the same homogeneous coordinates. Scaling Pr implies simultaneously the scaling of the world and camera which not change the image.

$$\begin{bmatrix} Pr_{11} & Pr_{12} & Pr_{13} & Pr_{14} \\ Pr_{21} & Pr_{22} & Pr_{23} & Pr_{24} \\ Pr_{31} & Pr_{32} & Pr_{33} & Pr_{34} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = Pr.\tilde{p}_w = k.Pr.\tilde{p}_w \tag{1.23}$$

In order to determine the projection matrix P, we have two choices:

- Set the scale so that $Pr_{34} = 1$,
- Set the scale so that $||Pr||^2 = 1$

We search the eigenvector p with smallest value λ of the matrix $A^T A$ which minimize the loss function $L(p, \lambda) = A^T A - \lambda p$.

1.4 Simple Stereo System

A simple stereo is characterized by the planarity of the two image planes of the two cameras (left and right) as shown by figures 1.8, 1.9.

In this case, we can compute the 3D coordinates (x, y, z) of the scene point if the cameras are calibrated as follow:

For the left camera:

$$x_l = f_x \times \frac{x}{z} + O_x, y_l = f_y \times \frac{y}{z} + O_y \tag{1.24}$$

$$x_r = f_x \times \frac{x - b}{z} + O_x, y_r = f_y \times \frac{y}{z} + O_y \tag{1.25}$$

We assume that the parameters f_x, f_y, O_x, O_y, b are known, b is the distance between the two centers of projection of the two cameras.

Using the expressions of x_l, y_l and x_r, y_r, we can write the following expressions:

$$zx_l = xf_x + zO_x, zy_l = yf_y + zO_y \tag{1.26}$$

$$zx_r = (x - b)f_x + zO_x, zy_r = yf_y + zO_y \tag{1.27}$$

Fig. 1.8 Example of simple stereo device: Fujifilm FinePix REAL 3D W3

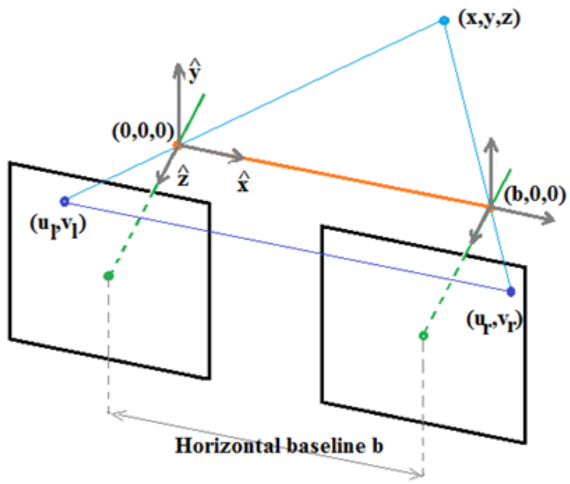

Fig. 1.9 Geometric Modeling of a simple stereo

$$z = \frac{b f_x}{x_l - x_r} \tag{1.28}$$

Applying some transformations, we get:

$$x = b\frac{x_l - O_x}{x_l - x_r}, y = b f_x \frac{x_l - O_x}{x_l - x_r} \tag{1.29}$$

We call the disparity: $d = (x_l - x_r)$. Note that the depth z is inversely proportional to the disparity and the disparity d is proportional to the baseline b.

1.4.1 Computing depth map from images of simple stereo

Figure 1.10 shows two images taken with a simple stereo and constitute a sample from the Middlebury dataset [7]. From the equations of y_l and y_r, the images of scene point lie on the same horizontal line.

For each point, we search in the same horizontal line the best match using template matching technique. We can use as similarity measure: Minimum Sum of Absolute Differences (SAD) , Minimum Sum of Squared Differences (SSD) or Maximum Normalized Cross-Correlation (NCC).

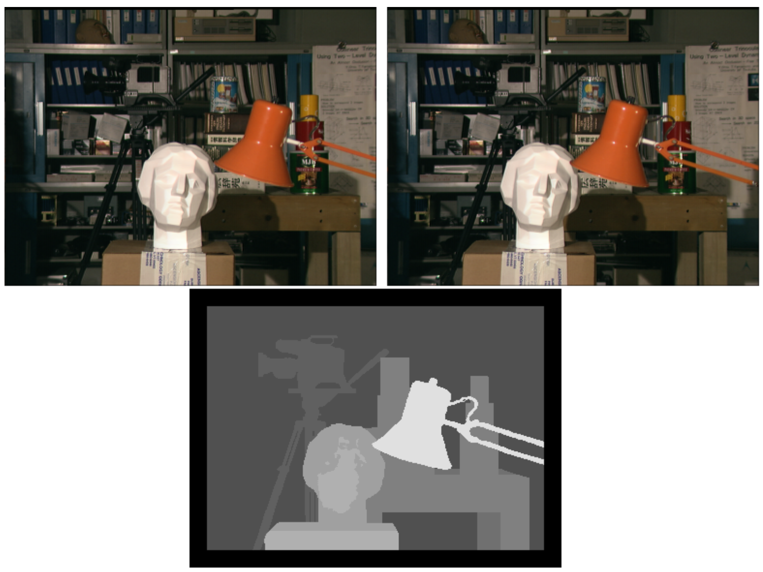

Fig. 1.10 Two images of Middlebury dataset and the depth map (ground truth).

Figure 1.11 shows some results of depth computation using SSD measure, and a window of size 21 for template matching,

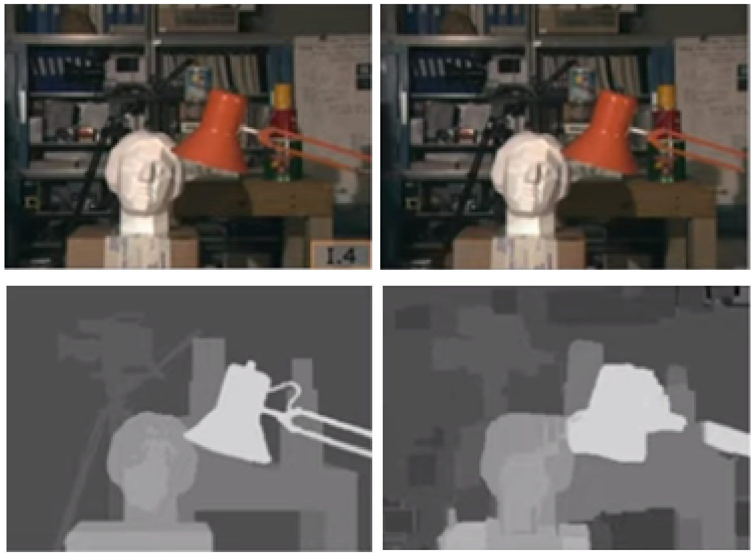

Fig. 1.11 Depth computation two images of Middlebury dataset. (Top) The left and right image, (Bottom) The ground truth and the computed depth.

1.5 Method of Zhengyou Zhang for Camera Calibration [8]

Zhengyou Zhang in [8] proposed a flexible new technique for camera calibration with the modeling of radial lens distortion. The basic principle of this method is to orient the camera towards a planar pattern and acquire at least two images from different orientations. Either the camera or the planar pattern can be freely moved. There is no need to measure the motion of the camera. The proposed procedure consists of a closed-form solution, followed by a nonlinear refinement based on the maximum likelihood criterion [8].

This method is implemented in OpenCV. Figure 1.12 shows a sample of pattern in our tests. Note that the distance between two corners is considered equal to 1 unit and the $z-$coordinates is equal to zero of all points of the pattern. Consequently, from the top to bottom, from left to right, the coordinates of corners are equal to:

$(0, 0, 0), \ldots (0, m-1, 0)$,

$(1, 0, 0), \ldots., (1, m-1, 0)$,

....

$(n-1, 0, 0), ..., (n-1, m-1, 0)$.

where n is the number of corners located on the column, and m is the number of corners located on the row.

The calibration matrix estimated for the used camera is:

$$M_{int} = \begin{bmatrix} 2.193e+03 & 0.00 & 3.94e+02 \\ 0.00 & 1.184e+03 & 1.27e+02 \\ 0.00 & 0.00 & 0.00 \end{bmatrix} \tag{1.30}$$

This method estimates also the translation and rotation vectors for each camera.

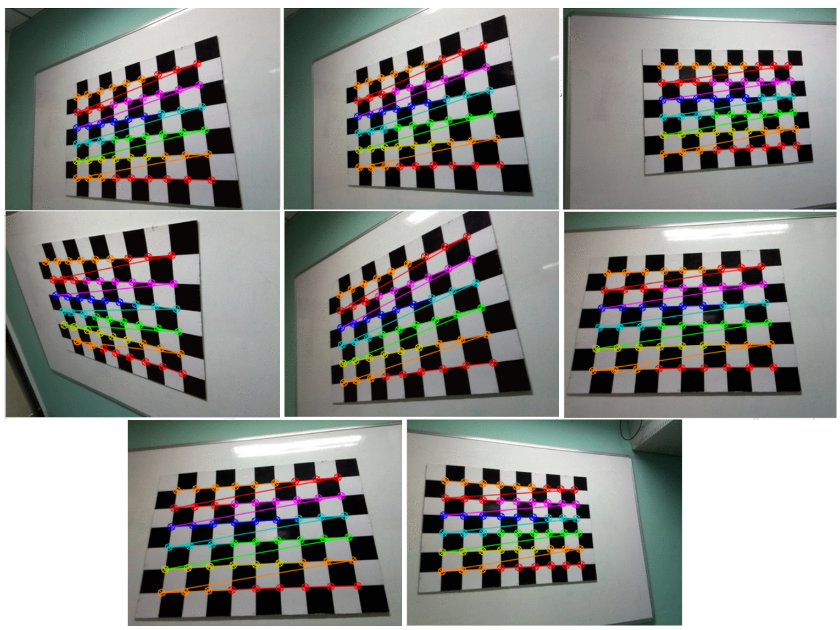

Fig. 1.12 Images of the pattern and for each image, corners are located as the link between black and white squares. For this pattern, $n = 6$ and $m = 9$.

1.6 Conclusion

Camera calibration for computer vision, provides the relationship between the three-dimensional real world and its two-dimensional image representation.

In this chapter we presented in the first the practical applications of camera calibration in depth estimation and 3D reconstruction. By leveraging calibrated cameras, tasks such as obstacle navigation in robotics and augmented reality rendering become achievable. In the next we explored the fundamental concepts of image formation based on the integration of internal and external parameters. We detailed how to calibrate a camera based on the internal and external matrices. We saw also that the depth map may be obtained using a simple stereo where two cameras are used such that the two image planes are in the same plane. At the end of this chapter, we introduced Zhengyou Zhang's method for camera calibration, which remains a robust and efficient approach widely adopted in modern computer vision systems. The next chapter, we study how can we obtain the 3D structure using uncalibrated cameras.

1.7 A Quiz Exercises

A Quiz

Select (one or many) correct answer for the following questions:

1- Camera calibration means :

- Calculate the intrinsic and extrinsic parameters of the camera
- Calculate the intrinsic parameters
- Find the transformation 2D ¡-¿ 3D
- Calculate the orientation of the camera with respect to the world coordinates.

2- In the camera calibration process :

- We need to know the 3D coordinates of scene points with respect to the camera frame.
- We need to know the 3D coordinates of scene points with respect to the world frame.
- The coordinates of image points are sufficient.
- We need to know in addition to the 3D coordinates of scene points with respect to the camera frame, the coordinates of their projection on the image plane.

3- In the camera calibration process :

- We need to know the orientation and relative position of the camera coordinate frame with respect to the world coordinate frame.
- We need to know the orientation and relative position of the world coordinate frame with respect to the camera coordinate frame.
- We need to know only the orientation of the world coordinate frame with respect to the camera coordinate frame.
- There is no need to know anything.

4- The internal parameters of the camera are:

- The pixel densities (pixels/mm): m_x, m_y.
- The coordinates of the principle point (O_x, O_y).

- The focal length and pixel densities (pixels/mm): m_x, m_y.
- The focal length and pixel densities (pixels/mm): m_x, m_y and the coordinates of the principle point (O_x, O_y)

  e- f_x, f_y and the coordinates of the principle point (O_x, O_y).

5- In the writing of calibration camera equations, homogeneous coordinates are used in order to:

- Get a linear system of equations.
- Be able to express linearly the orientation and relative position of the world coordinate frame with respect to the camera coordinates frame.
- Be able to express linearly the orientation and relative position of the camera coordinate frame with respect to the world coordinates frame.

6- What does mean the following equation:

- The scene point $P_c(x_c, y_c, z_c)$ is written with respect to the world coordinate frame.
- The scene point $P_w(x_w, y_w, z_w)$ is written with respect to the camera coordinate frame.
- This is a writing of the coordinate of P_w with respect to the camera coordinate frame using the rotation (r_{ij}) and translation (t_x, t_y, t_z).
- It gives the position of the camera with respect to the world coordinate frame.

7- What does mean the following equation:

- The image point (u,v) is written with respect to the camera coordinate frame.
- The image point (u,v) is written using the intrinsic calibration matrix and the coordinates of the scene point with respect to the camera coordinate frame.
- The image point (u,v) is written using the extrinsic calibration matrix and the coordinates of the scene point with respect to the camera coordinate frame.
- It gives the image formation model.

8- The projection matrix P is calculated using:

- The addition of the intrinsic and extrinsic matrices.

- The product of the intrinsic and extrinsic matrices,
- Only with the image points.

9- To calibrate a camera:

- We need at least the 3D coordinates of 8 scene points and their images.
- We need no more than the 3D coordinates of 4 scene points and their images.
- We need at least the 3D coordinates of 6 scene points and their images.

10- Camera calibration needs:

- Known non planar 3D coordinates of scene points with a static camera.
- Planar 3D coordinates of scene points with dynamic camera.
- Known non planar 3D coordinates of scene points with a dynamic camera.
- Planar 3D coordinates of scene points with static camera.

11- A stereo system of vision is composed by:

- Two cameras such that the image planes are planar.
- Two cameras such that their image planes are not planar.
- Two cameras.

12- A simple stereo system is composed by:

- Two cameras such that the image planes are planar.
- Two cameras such that their image planes are not planar.
- A moving camera along the image plane.

13- In case of simple stereo system, for depth computation:

- We need to know the calibration matrix.
- The disparity is sufficient.
- We need to know the stereo correspondence of image points.

Exercise 1

Capture a pair of images with a camera that has undergone a horizontal translation by a known distance b. After calibrating the camera to obtain its intrinsic parameters, calculate the $3D$ coordinates of points in the scene by identifying corresponding pairs using SIFT.

Exercise 2

We have pairs of images taken by a camera with axial translation movement. The images are given by the middlebury dataset: https://vision.middlebury.edu/stereo/data/ Calculate the disparity map for a pairs of images using a matching algorithm.

References

1. Seymourt Papert. The Summer Vision Project. Artificial Intelligence Group, MIT, 1966. https://dspace.mit.edu/handle/1721.1/11589
2. Bob Sumner. Augmented Creativity, TEDxZurich, 19th of January, 2015. https://www.youtube.com/watch?v=AJJOWemfOYI
3. Sachiko Iwase and Hideo Saito. Tracking soccer players based on homography among multiple views", Proc. SPIE 5150, Visual Communications and Image Processing 2003, (23 June 2003); https://doi.org/10.1117/12.502967
4. Martin A. Fischler, Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, Volume 24, Issue 6, pp 381–395, https://doi.org/10.1145/358669.358692
5. "The Eye of a Robot: Studies in Machine Vision at MIT" and "TX-O Computer". Courtesy of MIT Museum.
6. Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
7. D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202, Madison, WI, June 2003.
8. Zhengyou Zhang. A Flexible New Technique for Camera Calibration. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 11, NOVEMBER 2000
9. A computer algorithm for reconstructing a scene from two projections. Nature volume 293, pages 133–135 (1981)
10. OD Faugeras, QT Luong, SJ Maybank. Camera self-calibration: Theory and experiments. Second European Conference on Computer Vision ECCV'92, 1992
11. Olsson, Carl; Enqvist, Olof, Stable Structure from Motion for Unordered Image Collections Scandinavian Conference on Image Analysis, 2011.
12. Marc Pollefeys and Luc Van Gool. 3-D Modeling from Images. COMMUNICATIONS OF THE ACM July 2002/Vol. 45, No. 7.
13. Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Xin ; O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis , July 2017. ACM Transactions on Graphics, 36(4):1-11
14. M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, A. Eriksson. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. 2019. arXiv:1901.06802 [cs.CV]. https://arxiv.org/pdf/1901.06802.pdf

15. A. Podlozhnyuk, S. Pirker, C. Kloss. Efficient implementation of superquadric particles in Discrete Element Method within an open-source framework. Computational Particle Mechanics 4(1), 2016. DOI: 10.1007/s40571-016-0131-6
16. M.A. Turk et al. Face recognition using eigenfaces,CVPR 1991.
17. S.K.Nayar et al., Real-Time 100 Object Recognition System, ICRA, 1996
18. C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese. 3dr2n2: A unified approach for single and multi-view 3d object reconstruction. In European conference on computer vision, pages 628–644. Springer, 2016
19. R. Girdhar, D. F. Fouhey, M. Rodriguez, and A. Gupta. Learning a predictable and generative vector representation for objects. In European Conference on Computer Vision, pages 484–499. Springer, 2016
20. D. J. Rezende, S. A. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess. Unsupervised learning of 3d structure from images. In Advances in Neural Information Processing Systems, pages 4996–5004, 2016
21. S. R. Richter and S. Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1936–1944, 2018
22. C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 1(2):4, 2017.
23. Y. Liao, S. Donne, and A. Geiger. Deep marching cubes: ´ Learning explicit surface representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2916–2925, 2018
24. H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., volume 2, page 6, 2017
25. Mateusz Michalkiewicz et al. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. arXiv:1901.06802v1 [cs.CV], 21 Jan 2019.
26. Roberts, Lawrence G. 1963. Machine perception of three-dimensional solids. Outstanding PhD dissertations in the computer sciences. Garland Publishing, New York. 1963.
27. Adolfo Guzman-Arénas, Computer Recognition of Three-Dimensional Objects In a Visual Scene. PhD Dissertations in the computer sciences, MIT 1968.
28. M.B. Clowes. On seeing things. Artificial Intelligence, Volume 2, Issue 1, Spring 1971, Pages 79-116.
29. David L. Waltz. GENERATING SEMANTIC DESCRIPTIONS FROM DRAWINGS OF SCENES WITH SHADOWS. MIT Artificial Intelligence Laboratory. Technical Report 271, November 1972.

30. Kemp, M. Julesz's joyfulness. Nature 396, 419 (1998). https://doi.org/10.1038/24753
31. Julesz B. Foundations of Cyclopean Perception. Chicago University Press; Chicago, IL, USA: 1971.
32. David Marr. Vision, A Computational Investigation into the Human Representation and Processing of Visual Information. Originally published: San Francisco : W. H. Freeman, c1982.
33. Arthur Coste. Affine Transformation, Landmarks registration, Non linear Warping. https://www.sci.utah.edu/ãcoste/uou/Image/project3/ArthurCOSTE_Project3.pdf October 2012.
34. Sid Yingze Bao, Mohit Bagra, Yu-Wei Chao, Silvio Savarese. Semantic Structure From Motion with Points, Regions, and Objects. CVPR 2012.
35. Marco Crocco, Cosimo Rubino, Alessio Del Bue. Structure from Motion with Objects, CVPR 2016.
36. D.P. Robertson and R. Cipolla. Structure from Motion. In Varga, M., editors, Practical Image Processing and Computer Vision, John Wiley, 2009.
37. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. International Journal of Computer Vision, 9(2):137–154, 1992.
38. P. F. Sturm andW. Triggs. A factorization based algorithm for multi-image projective structure and motion. In European Conference on Computer Vision (ECCV'96), pages 709–720, 1996.
39. F. Schaffalitzky, A. Zisserman, R. I. Hartley, and P. H. S. Torr. A six point solution for structure and motion. In European Conference on Computer Vision (ECCV'00), pages 632–648, 2000.
40. W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In W. Triggs, A. Zisserman, and R Szeliski, editors, Vision Algorithms: Theory and Practice, LNCS, pages 298–375. Springer Verlag, 2000.
41. D. C. Brown. The bundle adjustment - progress and prospects. International Archives of Photogrammetry, 21(3), 1976.