Geometric calibration of a camera or a stereoscopic vision sensor

Jean-José ORTEU

28.03.2023

Table des matières

I.Presentation

Module :

Imagerie

Author(s)

Jean-José ORTEU[1] - École des Mines d'Albi

Professeur des Ecoles des Mines. Directeur adjoint du centre de recherche CROMeP de l'École des Mines d'Albi. Directeur adjoint du GDR CNRS « Mesures de champs et Identification en Mécanique des Solides » Responsable du groupe de recherche « Mesure, Contrôle et Surveillance » de l'IGM. Domaines de recherche : métrologie (3D) par vision artificielle, photomécanique, surveillance de procédés, CND par vision. Domaines d'enseignement : automatique, signaux et systèmes, calcul numérique, photomécanique.

Site Web de l'auteur : *http://www.orteu.fr/*[2]

Abstract :

The geometric calibration of a camera consists in determining the existing mathematical relation between the coordinates of the 3D points of the scene observed and the 2D coordinates of their projection in the image (image-points). This step of calibration represents the initial point for several applications of the artificial vision, such as for instance: recognition and objects localization, control procedure for the dimensional measurement of cast parts, reconstruction of the environment for the navigation of a mobile robot, etc. Calibrating a camera is about choosing a model of camera a priori and then determining the parameters of that model. We will describe the main models of camera used as well as the main methods suggested in order to determine the parameters of the model chosen. To get tridimensional information, it is necessary to combine two cameras in order to constitute a stereoscopic vision sensor. The calibration of such a sensor is a specific problem and is also described.

Keywords :

Geometric models of camera, camera calibration, geometric distortions, stereoscopic vision.

Prerequisites :

Basics in linear algebra.

Pedagogical objectives :

Upon completion of module, students should be able to basically explain how to mathematically model a camera and how to experimentally determine the parameters of the model chosen.

Course overview :

- Course
- Modelling and calibration of a camera
- Modelling and calibration of a stereoscopic vision system

Design & production :

Le Mans Université

License :

1 - jean-jose.orteu@mines-albi.fr
2 - http://www.orteu.fr/
3 - http://www.gnu.org/licenses/fdl.txt

II.Course

The geometric calibration of a camera (1.)🚗 consists in determining the existing mathematical relation between the coordinates of 3D points and the coordinates in 2D of their projection in the image (image point) (see Figure 1). This step of calibration is the initial point for several applications of **A**rtificial **V**ision, such as the recognition and the location of objects, the control procedure for the dimensional measurement of cast parts, the reconstruction of the environment for a mobile robot navigation, etc.

The calibration of a camera is particularly important when you want to obtain metric information, for the application of dimensional measurements, from acquired images.To obtain precise dimensional measurements, it is essential to take into account the geometric distortions induced by the optical system used.

The calibration of a camera involves choosing a model of camera a priori and then determining parameters of this model.

We will describe the main models of cameras used as well as the main methods developed in order to help determining the parameters of the model chosen.

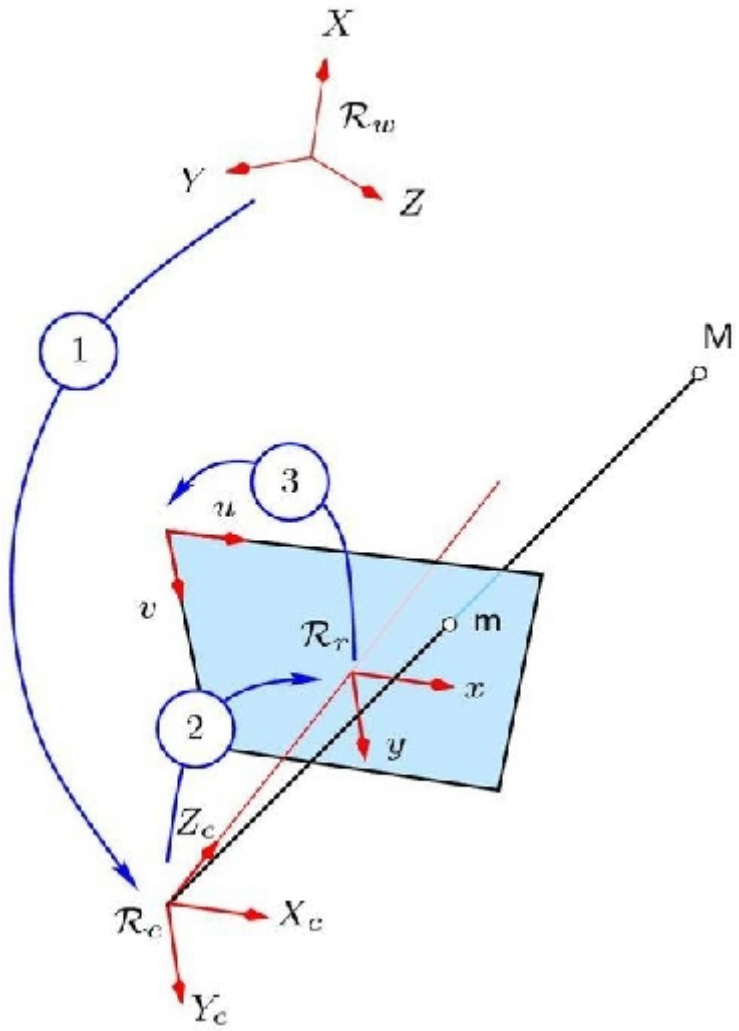
To obtain tridimensional information, it is necessary to combine two cameras in order to constitute a sensor for stereoscopic vision. The calibration of such a sensor is a specific issue that is to be described too.

1. Modelling and calibration of a camera

In this section, first we will describe the usual pinhole model and then the different models which enable to take the distortions into account: parametric and non-parametric approaches.

1.1. The projection model

The pinhole model [1 [[01]], 2 [[02]], 3 [[03]], 4 [[04]]] helps modeling a camera by a perspective projection. This model tranforms a 3D point M into an image-point m and may be split into three successive basic transformations (see Figure 2).



1.2. Use of homogeneous coordinates

In computer vision, **homogeneous coordinates** are often used [1 [[01]], 5 [[05]], 3 [[03]], 4 [[04]]]:

in 2D:

$$m = \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{Euclidean\ coordinates} \Rightarrow \tilde{m} = \underbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}_{Homogeneous\ coordinates}$$

in 3D:

$$M = \underbrace{\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}}_{Euclidean\ coordinates} \Rightarrow \tilde{M} = \underbrace{\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}}_{Homogeneous\ coordinates}$$

There are several advantages to that. For instance, we will see in the "*Transformation between the camera reference frame and the sensor reference frame (retinal plane)*" section that it enables expressing the pinhole model with a linear relation.

1.3. Transformation between the world reference frame and the camera reference frame

As indicated on Figure 2, ① represents a transformation between the world reference frame R_w (arbitrarily chosen) and the camera reference frame R_c (which origin is located in the optical center of the camera). This rigid transformation consists of a rotation $[R]$ and a translation $[t]$. The parameters of this transformation are called extrinsic parameters of the camera.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = [\mathbf{R}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} + \mathbf{t} = \begin{bmatrix} \mathbf{R} & t \\ \mathbf{0^t} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = [\mathbf{T}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

with:

$$\mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} ; [\mathbf{R}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

T is a 4×4 matrix.

Remarque

The representation of a rotation by the nine parameters $r_{\ddot{U}}$ is not minimal. Indeed, three parameters are enough to represent a rotation (instantaneous rotation vector, Euler angles, Bryant angles etc.).

1.4. Transformation between the camera reference frame and the sensor reference frame (retinal plane)

The second transformation, referred as ② on Figure 2 binds the camera reference frame R_c to the sensor reference frame R_r (retinal plane). This is a perspective projection (3×4 matrix, referred as $[P]$) that transforms a 3D point $\begin{pmatrix} X_c & Y_c & Z_c \end{pmatrix}$ into an image point (x, y) (in metric units).

$$s. \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = [\mathbf{P}] \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

where f refers to the focal length of the lens used.

Remarque

Equation (2) that shows the perspective projection is to be written:

$$x = f\frac{X_c}{Z_c}$$

$$y = f\frac{Y_c}{Z_c}$$

These equations are **non-linear ones.**

The use of homogenous coordinates makes it possible to write the perspective projection (and the complete pinhole camera model) under a **linear** form (see equation (2)).

1.5. Transformation between the sensor reference frame and the image reference frame

The third and last transformation, referred as ③ on Figure 2, describes the process that transforms image coordinates (x, y) (in metric units) into discrete image coordinates (u, v) (pixels).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & k_x \cot(\theta) & c_x + c_y \cot(\theta) \\ 0 & \frac{k_y}{\sin(\theta)} & \frac{c_y}{\sin(\theta)} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = [\mathbf{A}] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where:

- c_x and c_y (in pixels) represent the coordinates of the optical axis intersection with the image plane (situated, in theory, at the center of the image)
- k_x and k_y represent the number of pixels per unit of length along directions x and y of the sensor respectively ($k_x = k_y$ in the case of square pixels)
- θ takes into account the possible non-orthogonality of the lines and the columns in the image. In practice, θ is very close from to $\pi/2$. This parameter is referred as "skew factor".

It is often considered that the "skew factor" is negligible $\theta = \frac{\pi}{2}$ and equation (3) is then simplified as below:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & c_x \\ 0 & k_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \left[\mathbf{A}_{\text{simplifiée}}\right] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

1.6. Full pinhole model

The composition of transformations ①, ② and ③ may be summarized by the equation shown in Figure 3.

$$\begin{pmatrix} X & Y & Z \end{pmatrix} \xrightarrow{\mathbf{T}} \begin{pmatrix} X_c & Y_c & Z_c \end{pmatrix} \xrightarrow{\mathbf{P}} \begin{pmatrix} x & y \end{pmatrix} \xrightarrow{\mathbf{A}} \begin{pmatrix} u & v \end{pmatrix}$$

That leads to the equation of the pinhole camera model:

$$\tilde{m} = \underbrace{\mathbf{AP}}_{\mathbf{K}} \mathbf{T} \tilde{M}$$

with:

$$\mathbf{K} = \mathbf{AP} = \begin{bmatrix} k_x & k_x \cot(\theta) & c_x + c_y \cot(\theta) \\ 0 & \frac{k_y}{\sin(\theta)} & \frac{c_y}{\sin(\theta)} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} f_x & f_x \cot(\theta) & c_x + c_y \cot(\theta) & 0 \\ 0 & \frac{f_y}{\sin(\theta)} & \frac{c_y}{\sin(\theta)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

where $f_x = f k_x$ and $f_y = f k_y$ represent the focal length of the camera in pixels along directions x and y respectively.

The five parameters $\begin{pmatrix} c_x & c_y & f_x & f_y & \theta \end{pmatrix}$ of the matrix K are called **intrinsic parameters** of the camera.

In the end, the pinhole camera model is described by five intrinsic parameters $\begin{pmatrix} c_x & c_y & f_x & f_y & \theta \end{pmatrix}$ and six extrinsic parameters (three for the rotation and three for the translation).

Remarque

In the case of a neglected "skew factor", the pinhole camera model, which binds 3D coordinates $\begin{pmatrix} X & Y & Z \end{pmatrix}$ of a point written in the world reference frame with 2D coordinates $\begin{pmatrix} u & v \end{pmatrix}$ of its projection in the image-point (image point= pixel), is often described as below:

$$u = f_x \frac{r_{11}X + r_{12}Y + r_{13}Z + t_x}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + c_x$$

$$v = f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_y}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + c_y$$

We sometimes refer to this relation as **colinearity relations**.

1.7. Taking the distortions into account

Rappel

The pinhole camera model modelizes an ideal camera (simple perspective projection) and does not take into account the eventual geometric distortions induced by the optical system used. Several authors [6 [[06]], 7 [[07]]] proved that it was indispensable to take those distortions into account within applications of dimensional metrology in order to correct them.

(Usual) parametric approach

The usual parametric approach consists in the modeling of distortion by the improvement of the pinhole camera model with additional terms (thus the model becomes non linear). With this approach, the model draws its inspiration from the geometric aberration theory for a centered (lens) system by adding some corrective terms corresponding to different types of distortions: radial distortion, prismatic distortion, decentering distortion [8 [[08]], 9 [[09]], 10 [[10]]].

From the pinhole camera model, the distortions effects can be taken into account by a fourth transformation D, binding the "ideal" retinal coordinates (2.) $m_r = \begin{pmatrix} x & y \end{pmatrix}$ to the "real" retinal coordinates $\breve{m}_r = \begin{pmatrix} \breve{x} & \breve{y} \end{pmatrix}$:

$$\begin{pmatrix} X & Y & Z \end{pmatrix} \xrightarrow{T} \begin{pmatrix} X_c & Y_c & Z_c \end{pmatrix} \xrightarrow{P} \boxed{\begin{pmatrix} x & y \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \breve{x} & \breve{y} \end{pmatrix}} \xrightarrow{A} \begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$$

$$\breve{m}_r = D(m_r) = m_r + \delta(m_r)$$

$$= m_r + \overbrace{\underbrace{\delta_r(m_r)}_{radial} + \underbrace{\delta_d(m_r)}_{décentrage} + \underbrace{\delta_p(m_r)}_{prismatique}}$$

Several authors showed that the following model, often called **R3D1P1**(3.), is more than enough for most of lens objectives with a focal length superior to $5mm$:

$$D(m_r) = m_r(1 + r_1(x^2 + y^2) + r_2(x^2 + y^2)^2 + r_3(x^2 + y^2)^3) + \begin{pmatrix} d_1(3x^2 + y^2) + 2d_2xy + p_1(x^2 + y^2) \\ 2d_1xy + d_2(x^2 + 3y^2) + p_2(x^2 + y^2) \end{pmatrix}$$

where $d = \begin{pmatrix} r_1 & r_2 & r_3 & d_1 & d_2 & p_1 & p_2 \end{pmatrix}$ is the vector of the **distortion parameters**.

Remarque

It is often enough to use a radial model (from order 1 to 3). Writing down: $\rho = \sqrt{x^2 + y^2}$, the model R3 is often written as:

$$D(m_r) = m_r(1 + r_1\rho^2 + r_2\rho^4 + r_3\rho^6)$$

We call k the vector of the intrinsic parameters defined by matrix K, and d the vector of the distortion coefficients (which are also intrinsic to the camera):

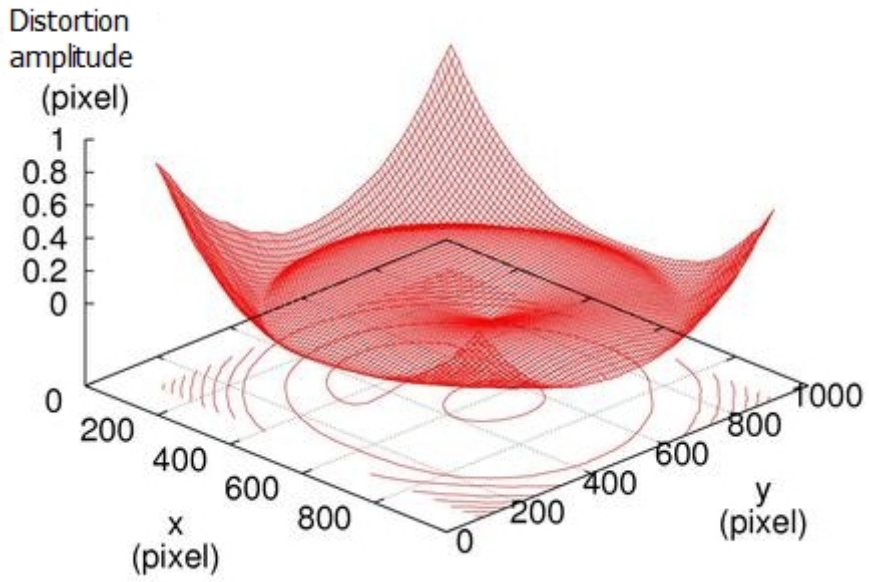
$$\mathbf{k} = \begin{pmatrix} c_x & c_y & f_x & f_y & \theta \end{pmatrix}$$

$$\mathbf{d} = \begin{pmatrix} r_1 & r_2 & r_3 & d_1 & d_2 & p_1 & p_2 \end{pmatrix}$$

The model of camera is **non-linear** and may be written as a vector function F :

$$\check{m} = \mathbf{F}(\mathbf{k}, \mathbf{d}, \mathbf{R}, \mathbf{t}, M)$$

For instance, Figure 5 shows a distortion map (amplitude of the distortion in each pixel of the image) obtained during the calibration of a camera equipped with a $25mm$ lens objective. This figure perfectly reveals that the main component is the radial distortion (the more we move away from the center of the image, the more the distortion is important). In this example, the distortion is pretty low (the magnitude of the distortion is in the order of 1 pixel at the corners of the image) but it can reach several pixels (or even around 10 pixels) for objectives with lower focal length.



a) Correction of the distortion

It is sometimes essential to know the ideal pixel coordinates $\begin{pmatrix} u & v \end{pmatrix}$, i.e. non-distorted, corresponding to the distorted ones $\begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$.

$$\begin{aligned} u &= c_x + f_x x \\ v &= c_y + f_y y \end{aligned} \quad \text{soit} \quad \begin{aligned} x &= \frac{u - c_x}{f_x} \\ y &= \frac{v - c_y}{f_y} \end{aligned}$$

Hence, we can deduce that:

$$\breve{u} = c_x + f_x\breve{x} = c_x + f_x(x + \delta_x(x, y)) = c_x + (u - c_x) + f_x\delta_x(x, y)$$

$$\breve{v} = c_y + f_y\breve{y} = c_y + f_y(y + \delta_y(x, y)) = c_y + (u - c_y) + f_y\delta_y(x, y)$$

From equations (12) and (13), we notice that it is possible to express $\begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$ according to $\begin{pmatrix} u & v \end{pmatrix}$ and the intrinsic parameters k and d of the camera:

$\breve{u} = f_u(u, v, \mathbf{k}, \mathbf{d})$

$\breve{v} = f_v(u, v, \mathbf{k}, \mathbf{d})$

In the general case, the distortion model given by equation (14) cannot be inverted and it is therefore necessary to use a numerical method to estimate the ideal coordinates $\begin{pmatrix} u & v \end{pmatrix}$ of the image-point that would have been obtained with a camera exempt from distortion.

Let $\breve{m} = \begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$ be a pixel of the distorted image. We seek the non distorted pixel $m = \begin{pmatrix} u & v \end{pmatrix}$.

The pixel $\begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$ corresponds to point $\breve{m}_r = \begin{pmatrix} \breve{x} & \breve{y} \end{pmatrix}$ in the retinal plane:

$$\breve{x} = \frac{\breve{u} - c_x}{f_x}$$

$$\breve{y} = \frac{\breve{v} - c_y}{f_y}$$

We seek the point $m_r = \begin{pmatrix} x & y \end{pmatrix}$ such as $D(m_r) = \breve{m}_r, i.e.$:

$\breve{m}_r - D(m_r) = 0$

Equation (15) can be resolved thanks to the Newton method applied to the function $f(m_r) = \breve{m}_r - D(m_r)$ that we initialize with $m_r^{(0)} = \breve{m}_r$.

The common iteration is given by the formula: $m_r^{(i)} = m_r^{(i-1)} - d_N^{(i-1)}$ avec $d_N^{(i-1)}$ solution of $\mathbf{J_f} d = f(m_r^{(i-1)})$

Namely:

$m_r^{(i)} = m_r^{(i-1)} + \mathbf{J_D}(m_r^{(i-1)})^{-1}(\breve{m}_r - D(m_r^{(i-1)}))$

with:

$$\mathbf{J_D} = \begin{pmatrix} \frac{\partial \breve{x}}{\partial x} & \frac{\partial \breve{x}}{\partial y} \\ \frac{\partial \breve{y}}{\partial x} & \frac{\partial \breve{y}}{\partial y} \end{pmatrix}_{|(x,y)=m_r^{(i-1)}}$$

The stop criterion may be based on the value of the error module ε after each iteration. In practice, convergence is reached after only few iterations (≈ 4).

After obtaining the retinal coordinates $\begin{pmatrix} x & y \end{pmatrix}$, the equation (12) is used to enable the calculation of the sought coordinates $\begin{pmatrix} u & v \end{pmatrix}$.

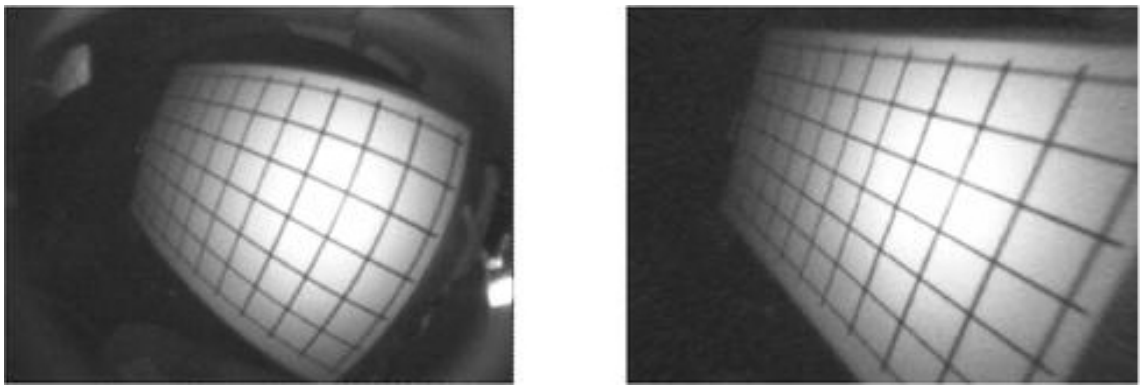
Remarque

To calculate a corrected image from a distorted images (which is a different problem from correcting only one point), it is not necessary to use a numerical method in order to invert the distortion model. You just have to use the direct model given by equation (14) and to fill the image-to-draw by sweeping u and v from the destination image.

For a given pixel $\begin{pmatrix} u & v \end{pmatrix}$ (in integer coordinates) in the destination image, equation (14) is used to enable the calculation of the coordinates $\begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$ of the corresponding point in the source image. Usually, these coordinates are not integer and an interpolation is required in order to calculate the intensity value (grayscale) which must be copied in the destination image at the position $\begin{pmatrix} u & v \end{pmatrix}$ (cf. [11 [[11]]]).

Other correction methods of distortion have been suggested in the literature. For a review, see in particular the "perspective camera inverse model" section in [12 [[12]]].

For instance, Figure 6 shows a distorted image (on the left) and the corrected image (on the right).

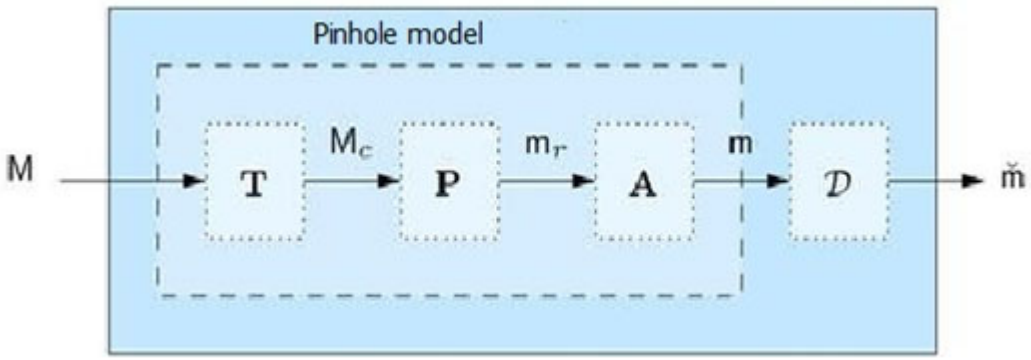


b) Non-parametric approach

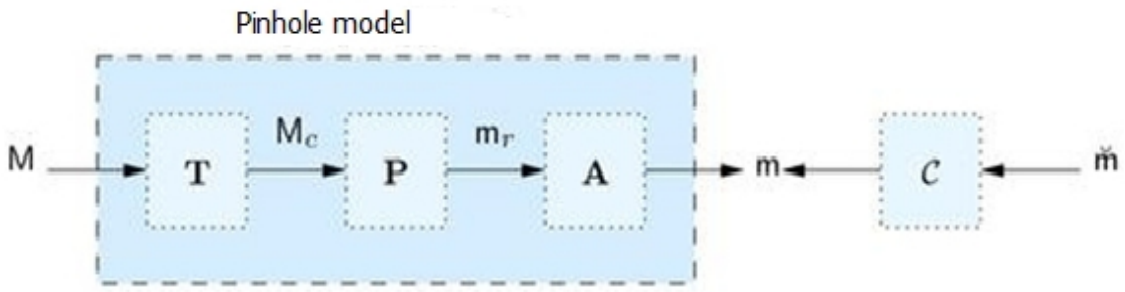
In the case of complex optical systems, some authors [13 [[13]]] have proved that it is recommended to modelize the distortion in a non-parametric way using spline functions [14 [[14]]].

In that case, it is a purely mathematical modeling (kind of "black box" approach) aiming to determine the distortion function that shows the best the way in which the ideal image is distorted [15 [[15]], 16 [[16]]].

As it is a purely mathematical modeling (4.)⌓, it is not a problem to adopt the layout of Figure 7 and to seek the distortion function D that links the ideal image-coordinates $\begin{pmatrix} u & v \end{pmatrix}$ of point m to the image-coordinates $\begin{pmatrix} \breve{u} & \breve{v} \end{pmatrix}$ of point $\breve{m}$.



In this approach, to be able to correct the distortion easily, it is better to choose the model schematically presented on Figure 8, in which the reciprocal function C of distortion correction is used instead of the distortion function D. Indeed, the distortion can be corrected directly using C hereas it may require a lot of time to do all the calculations in the case of inverting function D , particularly when its reciprocal function C cannot be analytically determined. Moreover, the field of definition of the spline function for the correction of C is *a priori* known and determined by the dimension of the images, whereas the spline function for the distortion of D has an *a priori* unknown definition field since it is expressed in the retinal plane which is defined by the calibration.



The equation of the distortion correction becomes:

$m = D^{-1}(\breve{m}) = C(\breve{m})$

The estimation of the function of distortion correction C consists in approximating horizontal (following the axis x) and vertical (following the axis y) components of the distortion correction field by two spline surfaces S_y and S_y [13 [[13]], 17 [[17]]].

The function of distortion correction permits to correct the points (or a full image) from their distortion. The corrected points are linked to the 3D entry points by a classical pinhole model the parameters of which are easy to estimate.

We are done with establishing the linear (5) and non-linear (11) models of a camera, we are now going to deal with the calibration methods that permit to estimate the parameters of these models.

1.8. Calibration of a camera

Calibrating a camera consists in estimating the parameters of the model chosen to represent it. It is a kind of "**parametric estimation**".

For the pinhole model (with or without distortion), this means estimating the intrinsic parameters of the camera, its position and its orientation relative to the world reference frame that has been chosen (extrinsic parameters).

Remarque

Actually, the calibration of a camera is used particularly to determine its intrinsic parameters, which, as their name indicates, are intrinsic to the camera and do not change if the camera is moved. Specific methods (called localization methods) have been developed in order to determine the position of a camera according to a work reference frame when its intrinsic parameters are already known.

Usually, this calibration problem can be solved by using a specific calibration device (calibrating target) that provides known 3D points in the world reference frame.

Several calibration methods have been suggested. Throughout the years, those methods have become increasingly sophisticated to lead to a more precise and easier-to-implement calibration.

In the following part, we will describe the method that is considered nowadays as the most efficient one.

The method consists in obtaining n images of a calibration target (plane (5.)) composed of p points. The target is freely moved (rotations and translations) in the field of view of the camera (see Figure 9). The calibration method is said to be of a photogrammetric type. It helps to estimate all the camera model parameters at the same time as well as the tridimensional points of the calibration target. Therefore, the geometry of the calibration target does not need to be known with precision *a priori*.

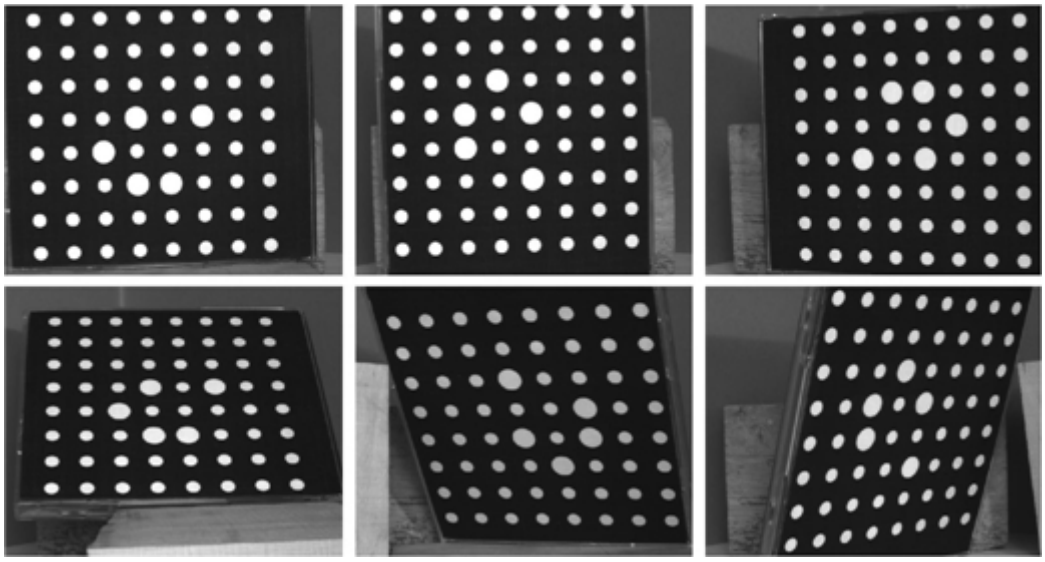
Remarque

During the shifting of the calibration target, it is important to cover well the whole field of view of the camera in order to correctly calibrate the distortion (which is generally bigger at the images' edge rather than at their centers).

The target points may be the intersection nodes of horizontal and vertical lines (grid), the corners of a target in the shape of a checkerboard or the centers of circular patches.

The target points, which are extracted by specific image processing procedures, provide the measurements.

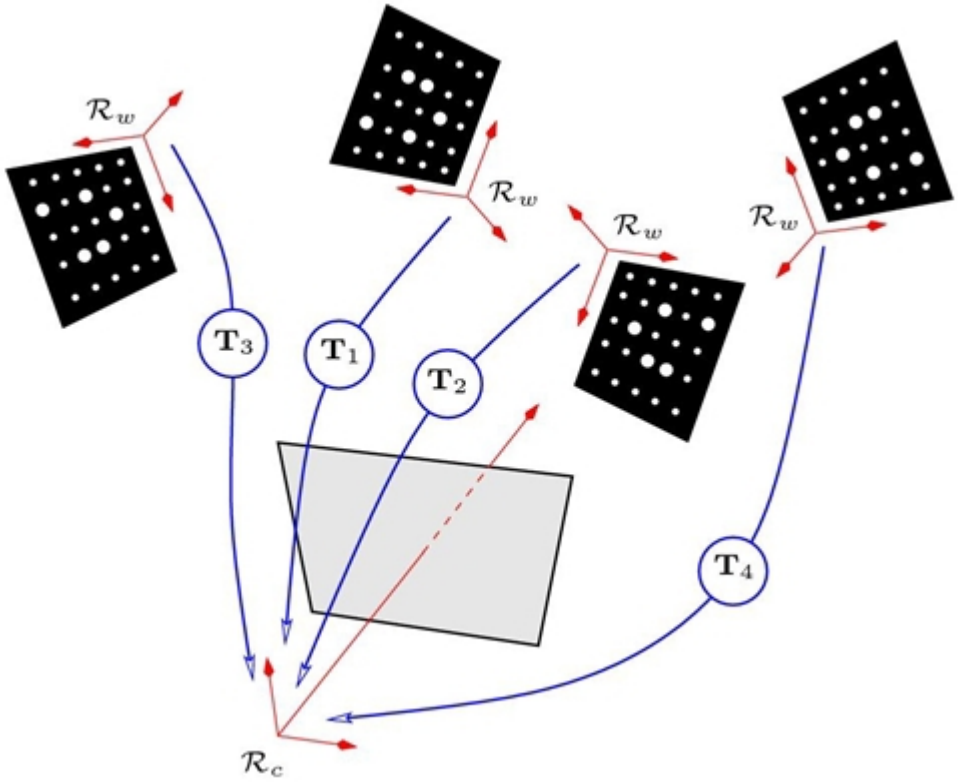
Figure 9 illustrates a 6-image sequence of a calibration target composed of 64 circular patches.



We write $\breve{m}_i^j$ down, the coordinates of the projection of the j_th point $M_j(j = 1...p)$ of the $i-th$ view $(i = 1...n)$ on the camera image plane. If the distortion is taken into account, we can write from (11):

$$\breve{m}_i^j = \mathbf{F}(\mathbf{k}, \mathbf{d}, \mathbf{R_i}, \mathbf{t_i}, M_j)$$

In those equations, the target reference frame is used for each view as the world reference frame. (See Figure 10).



By using (18), each projection of a tridimensional point provides two equations. Therefore, there are $2np$ equations.

Let's count the unknown parameters left to estimate: five intrinsic parameters, seven distortion parameters in the case of a R3D1P1 model, $6n$ extrinsic parameters (three for the rotation and three more for the translation of each rigid transformation T_i) and $3p$ tridimensional coordinates. It gives us a total of $12 + 6n + 3p$ unknowns.

Consequently, there are $2np$ equations and $12 + 6n + 3p$ unknowns. If n and p are big enough (for instance if $n = 6$ and $p = 64$, there are 768 equations for 240 unknowns)) we can estimate all the parameters minimizing the sum of the distances between the projection of the $j - th$ point of the $i - th$ view onto the image and the point $\breve{m}_i{}^j$ extracted in the image:

$$\theta = \arg\min_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{p} \|\breve{m}_i{}^j - \mathbf{F}(\mathbf{k}, \mathbf{d}, \mathbf{R_i}, \mathbf{t_i}, M_j)\|^2$$

with $\theta = (\mathbf{k}, \mathbf{d}, \mathbf{R}_{1\ldots n}, \mathbf{t}_{1\ldots n}, M_{1\ldots p})$.

Minimizing (19) is a matter of non-linear optimization (called bundle adjustment) [18 [[18]]].

The problem is usually solved using the Levenberg-Marquardt algorithm [19 [[19]]], with the rotations R_i expressed under a minimal form (instantaneous rotation vector, Euler angles, Bryant angles, etc.).

In order to converge, the minimization algorithm needs an **initial estimate of the sought parameters**: the estimation of the five intrinsic parameters of the pinhole model and of the extrinsic parameters may be obtained by analytic methods described in [20 [[20]], 21 [[21]]]. The distortion parameters are generally initialized to zero. The initial tridimensional coordinates of the target points are those of its model which served to its creation. Given that they will be estimated again, these coordinates do not need to be known with precision, which constitutes an advantage compared with the methods requiring a precise knowledge of the calibration target used.

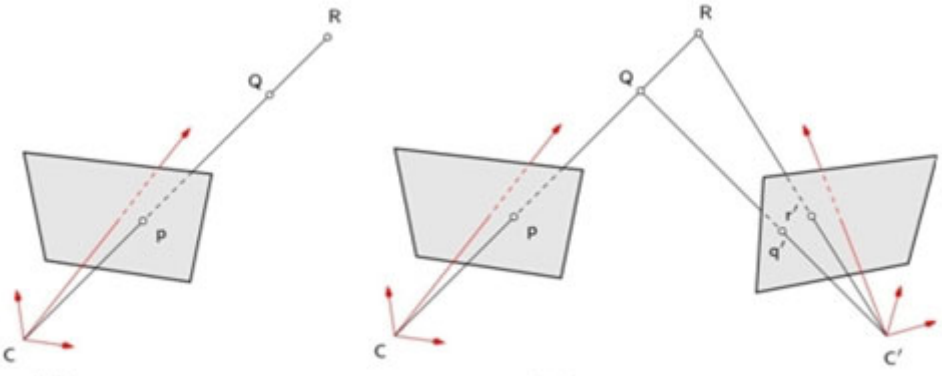
The minimization of (19) leads to a solution defined up to a scale factor. This factor can be determined providing the distance measured in space between two given points (2 particular points of the target for instance).

2. Modelling and calibration of a stereoscopic vision system

In this section we will focus on the modeling of a sensor composed of two rigidly linked cameras: a **stereoscopic vision sensor**, also called a **stereovision sensor**.

To begin with, we will briefly justify the use of a second camera in order to perceive the environment in three dimensions and then, we will enumerate the different reference frames involved and the transformations linking those reference frames. [10 [[10]], 4 [[04]]]

2.1. Why using two cameras?

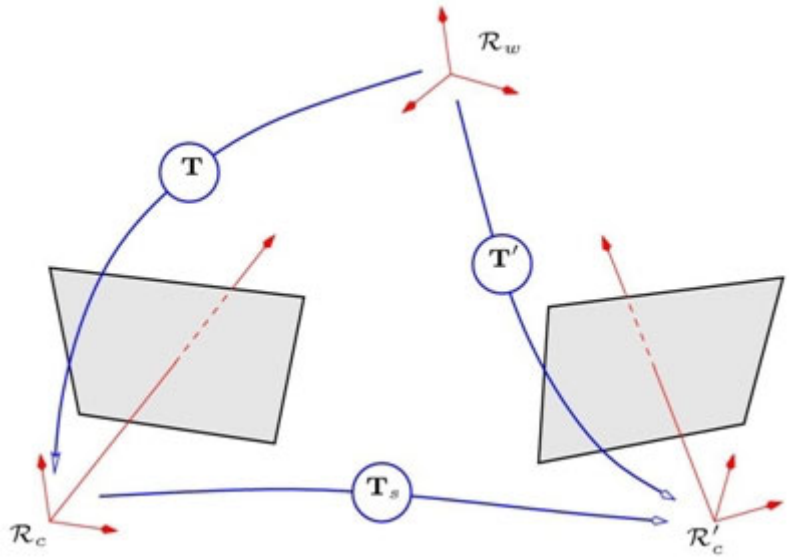


If we reason from a geometric point of view, a camera is a sensor which transforms every "visible point" of the tridimensional space into a point of the image's bidimensional space. Therefore, this transformation suppresses the third dimension and is, thus, irreversible. It is graphically expressed by Figure 11.a: both points Q and R of the space project themselves on the image plane in a single and only point p because they are on the same projection line $(\mathbf{C},\mathbf{p})$, C is called the **projection center** or **optical center**. It means that given an image point p, there is an infinity of tridimensional points that can be the projection of it. By using two cameras, as shown on figure 11.b, it is possible to determine the point tridimensional position by **triangulation**. Indeed, there is only one point of space corresponding to the pair of projected points $\mathrm{p,q'}$ and only one corresponding to $\mathrm{p,r'}$. Therefore, the triangulation consists in determining the intersection in the space of two projection lines. Thus, it is necessary to express those two lines according to a common reference frame, the one of the left camera for instance. To reach it, we will try to express a geometric relation between both cameras.

2.2. Reference frames and changes of reference frames

The tridimensional space of the scene is fitted with its orthonormal reference frame R_w. Each of the two cameras has its own orthonormal reference frame: we will call them left camera reference frame R_c and right camera reference frame R'_c. Figure 12 illustrates those three

reference frames as well as the rigid transformations allowing the expression of a point in another reference frame.



With those conventions, we can write the following relations down:

$\tilde{M}_c \cong \mathbf{T}\tilde{M}$

$\tilde{M}'_c \cong \mathbf{T}'\tilde{M}$

$\tilde{M}'_c \cong \mathbf{T}_s\tilde{M}_c$

These equations show us that the three transformations are not independent since we can determine one of them by using the two others:

$\mathbf{T} \cong \mathbf{T}_s^{-1}\mathbf{T}'$

$\mathbf{T}' \cong \mathbf{T}_s\mathbf{T}$

$\mathbf{T}_s \cong \mathbf{T}'\mathbf{T}^{-1}$

When a point M of the scene is simultaneously visible by both cameras, it gives us two points: m for the left camera and m' for the right one. Using the geometric model of the camera and the relation of dependence between the three reference frames R_w, R_c and R'_c, we can write the relations of m and m' according to M:

$\tilde{m} \cong \mathbf{K}\mathbf{T}\tilde{M}$

$\tilde{m'} \cong \mathbf{K}'\mathbf{T}'\tilde{M} \cong \mathbf{K}'\mathbf{T}_s\mathbf{T}\tilde{M}$

2.3. Triangulation

Equations (24) and (25) express the coordinates of both points $m = \begin{pmatrix} u & v \end{pmatrix}$ and $m' = \begin{pmatrix} u' & v' \end{pmatrix}$, respectively the left and right projections of a point M of the scene.

Asking:

$$T = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } T' = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & t'_x \\ r_{21} & r'_{22} & r'_{23} & t'_y \\ r'_{31} & r'_{32} & r'_{33} & t'_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We can write the following four-equation system:

$$u = f_x \frac{r_{11}X + r_{12}Y + r_{13}Z + t_x}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + C_x \qquad v = f_y \frac{r_{21}X + r_{22}Y + r_{23}Z + t_y}{r_{31}X + r_{32}Y + r_{33}Z + t_z} + C_y$$

$$u' = f'_x \frac{r'_{11}X + r'_{12}Y + r'_{13}Z + t'_x}{r'_{31}X + r'_{32}Y + r'_{33}Z + t'_z} + C'_x \qquad v' = f'_y \frac{r'_{21}X + r'_{22}Y + r'_{23}Z + t'_y}{r'_{31}X + r'_{32}Y + r'_{33}Z + t'_z} + C'_y$$

If we know both image points m and m' and if the stereoscopic sensor is calibrated (6.), then, by resolving this four-equation overdetermined system, we can determine the three unknowns which are the tridimensional coordinates of point M [22 [[22]]].

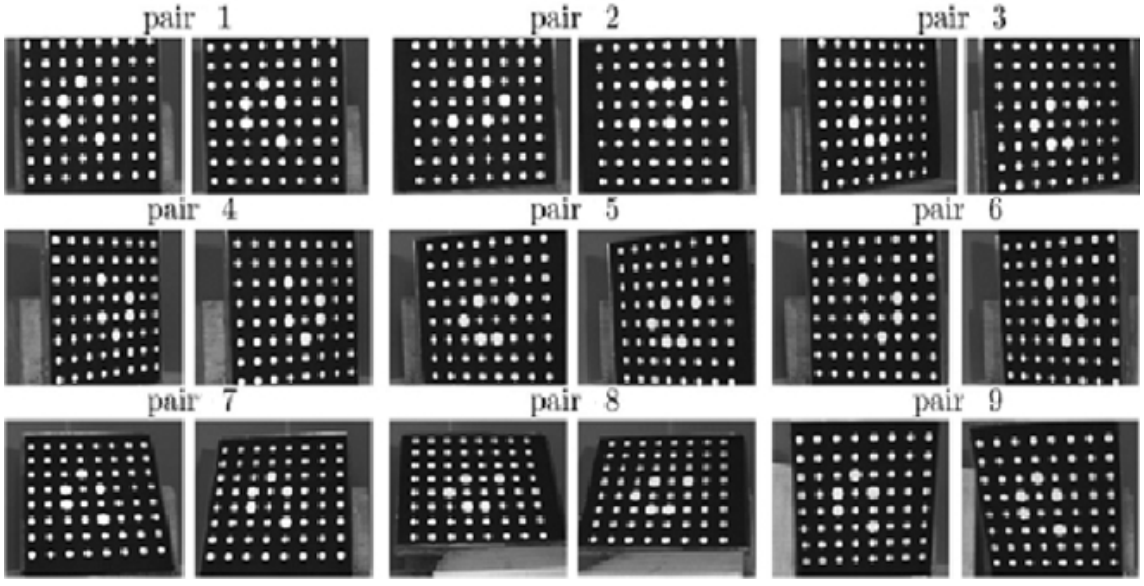
2.4. Calibration of a stereoscopic vision sensor

When we calibrate a single camera, we mainly focus on the intrinsic parameters defined by the matrix K and, if desired, on the extrinsic parameters defined by the rigid transformation T (localization of the camera according to the world reference frame). When we calibrate a stereoscopic vision sensor, we focus on both groups of intrinsic parameters defined by both matrix K and K' and to the relative position and orientation of both cameras defined by the rigid transformation T_s.

The aim of this sensor calibration is to allow the reconstruction of the tridimensional points observed by both cameras and is therefore very important for all of those who want to reach precise tridimensional measurements.

On a practical way, the process for the calibration of a stereoscopic vision sensor is similar to the process described in section 3 (*Calibration of a camera*) for the calibration of a camera. A target is placed in the field of view, common to both cameras, and a series of images of that target, viewed under different orientations, is taken by each camera.

For instance, Figure 13 shows a series of 9 pairs of images of a target which served to the calibration of a stereovision sensor.

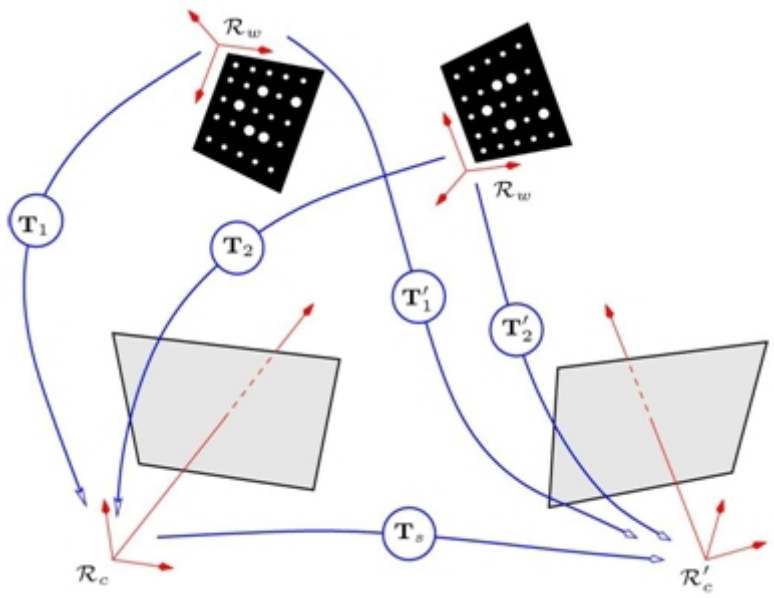


We will write down the rigid transformations T_i and T'_i respectively for the left and right camera, as below:

$$T_i = \begin{bmatrix} R_i & t_i \\ 0^T & 1 \end{bmatrix} \text{ et } T'_i = \begin{bmatrix} R'_i & t'_i \\ 0^T & 1 \end{bmatrix}$$

They link the i-th view of the target respectively to the left camera reference frame and to the right camera one. For each position of the target, we have the following relation (see Figure 14) according to (23):

$$T_s T_i = T'_i$$



Different methods permit the estimation of transformation T_s.

The method which is usually used consists in the calibration of each camera independently, using the method described in section 3 (*Calibration of a camera*), in order to determine the intrinsic parameters and the coefficients of distortion of both cameras. Then, both groups $\{T_i\}$ and $\{T'_i\}$ which are the matrices of the extrinsic parameters R_s and t_s can be calculated using any pair $k \in \{1...n\}$ of matrices of extrinsic parameters using the equation (23):

$$T_s = T'_k T_k^{-1}$$

The choice of the pair of matrices of the extrinsic parameters T_k and T'_k is delicate and several heuristics are possible, such as:

- Always choosing arbitrarily the k-th pair of matrices of the experiment, for instance the first pair T_1 and T'_1 ;
- Taking the pair of matrices that corresponds to the lowest global error of reprojection of the target points in both images.

All those heuristics have the inconvenience of not using the redundancy provided by the simultaneous use of all the pairs of matrices of the extrinsic parameters to estimate the transformation $\boldsymbol{T}_s$.

Dorian Garcia [10 [[10]]] suggested a method which enables to estimate R_s and t_s by using all the matrices of the extrinsics $\{T_i\}$ and $\{T'_i\}$, and showed that it permits more precision in the calibration.

His method consists in directly calculating R_s and t_s minimizing a functional of the form:

$$\theta = \arg\min_{\theta} \sum_{i=1}^{n} \sum_{j=1}^{p} \| \breve{m}_i^j - F(k, k', d, d', R_i, t_i, R_s, t_s \; ; M_j) \|^2$$

with:

$$\theta = (k, k', d, d', R_{1...n}, t_{1...n}, R_s, t_s, M_{1...P})$$

$$\breve{m} = (\breve{m} \;\; \breve{m}')$$

$$= \text{ vector containing the group of measurements provided by both cameras}$$

This problem of non-linear optimization is solved using the Levenberg-Marquardt algorithm.

III.Exercices

1. Knowledge test and application exercises

Question 1

[Solution n°1 p 21]

Must the intrinsic parameters of a camera be estimated again if we move the camera?

Question 2

[Solution n°2 p 21]

We use a camera CCD $1/2'$ which sensor size equals $6.4 \times 4.8mm$. Its resolution equals 800×600 pixels. We use a $25mm$ prime lens objective.

1. Calculate the size of the pixels.
2. Calculate the value of k_x.
3. Calculate the value of f_x(that) the calibration must provide.

Question 3

[Solution n°3 p 21]

We want to estimate the model R3D1P1 of a camera (12 intrinsic parameters). Can we calibrate this camera with the photogrammetric approach using three views of a 9-point target? Explain.

Question 4

[Solution n°4 p 21]

The calibration of a camera the resolution of which is 800× 600 pixels (sensor with square pixels) provided the following results:

$c_x = 225\ ; c_y = 280\ ; f_x = 3250\ ; f_y = 4510$

What enables us to suspect that the calibration is wrong?

Question 5

[Solution n°5 p 21]

You must show that in the case of a model of parametric distortion type 1 (1st order radial distortion), equation (14) of the lesson linking the distorted and non-distorted coordinates enables to calculate the non-distorted coordinates (correction from the distortion) by the resolution of a third degree equation of the form:

$A(u - c_x)^3 + (u - c_x) - B = 0$

Calculate both coefficients A and$B$.

Solution des exercices

>Solution n°1 *(exercice p. 20)*

NO (except if we shake the camera too much changing the set up of the focal length).

>Solution n°2 *(exercice p. 20)*

1. square pixels with a size of: $8\mu m$.
2. $k_x = k_y = 125mm^{-1}$
3. $f_x = f_y = 3125$ pixels

>Solution n°3 *(exercice p. 20)*

NO
$2np = 54$ equations ; $12 + 6n + 3p = 57$ unknowns
More unknowns than equations!

>Solution n°4 *(exercice p. 20)*

we must find $(c_x, c_y) \simeq (400, 300)$ (i.e. the center of the sensor) and since the pixels are square ones, we must find $f_x \simeq f_y$.

>Solution n°5 *(exercice p. 20)*

Distortion $R1$:
$\breve{x} = x(1 + r_1(x^2 + y^2))$
$\breve{y} = y(1 + r_1(x^2 + y^2))$
According to equation (13) of the lesson:
$\breve{u} = c_x + f_x\breve{x} = c_x + f_x x(1 + r_1(x^2 + y^2))$
$\breve{v} = c_y + f_y\breve{y} = c_y + f_y y(1 + r_1(x^2 + y^2))$

$$\breve{u} = c_x + (u - c_x)\left[1 + r_1\left(\frac{(u - c_x)^2}{f_x^2} + \frac{(c - c_y)^2}{f_y^2}\right)\right]$$

$$\breve{v} = c_y + (v - c_y)\left[1 + r_1\left(\frac{(u - c_x)^2}{f_x^2} + \frac{(c - c_y)^2}{f_y^2}\right)\right]$$

According to (1):

$$\frac{\breve{u} - c_x}{u - c_x} = \frac{\breve{v} - c_y}{v - c_y}$$

Namely:

$$v - c_y = (u - c_x)\frac{\breve{v} - c_y}{\breve{u} - c_x}$$

Reporting in the 1st equation (1) comes:

$$\breve{u} - c_x = (u - c_x) + \frac{r_1}{f_x^2}(u - c_x)^3 + \frac{r_1}{f_y^2}(u - c_x)^3\left(\frac{\breve{v} - c_y}{\breve{u} - c_x}\right)^2$$

Namely:

$$\frac{r_1}{f_x^2}(u-c_x)^3\left(1+\frac{f_x^2}{f_y^2}\left(\frac{\breve{v}-c_y}{\breve{u}-c_x}\right)^2\right)+(u-c_x)-(\breve{u}-c_x)=0$$

Of the form:

$$A(u-c_x)^3+(u-c_x)-B=0$$

The equation (6) enables calculating u, and then the equation (3) makes it possible to calculate v.

Glossaire

1.

We call ”camera” the set composed of the sensor and the associated optical system

2.

Which would be provided by an ideal camera exempt from distortion and following the pinhole model

3.

3rd order radial distortion, 1st order decentering distortion and first order prismatic distortion

4.

As seen in section "*Taking the distortions into account*", a physical approach to the distortion mode leads to a distortion of the points after projection on the retinal plane, before applying the transformation **A** which produces the discrete image (pixels map). In that case, the corrective terms are applied to the retinal coordinates, expressed in the sensor reference frame (retinal plane) rather than discrete coordinates. Compared to Figure 7, this means that the distortion function D is located between the transformation **P** and the transformation **A** (see Figure 4).

5.

The target does not need to be flat but, in practice, it is easy to draw it on a sheet of paper and to paste it on a rigid or almost plane surface. The fact that the focal point is almost flat permits easily providing an initial estimation of the parameters (X Y Z) of the calibration points M to the bundle adjustment process (see below). The method could work with any calibration object provided that an approximate 3D model of the object is known.

6.

It means that we know the values of R, R', t, t', fx, $f'x$, fy, $f'y$, cx, $c'x$, cy and $c'y$.

Bibliographie

[[01]] Olivier Faugeras, *Three-Dimensional Computer Vision : A Geometric Viewpoint*, The MIT Press, 1993, ISBN 0-262-06158-9.

[[02]] R. Horaud and O. Monga, *Vision par ordinateur : outils fondamentaux*, Hermès, 2nd edition, 1995.

[[04]] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN : 0521540518, 2nd edition, 2004.

[[05]] O. Faugeras, Q. T. Luong, and T. Papadopoulos, *The Geometry of Multiple Images*, The MIT Press, 2001, ISBN 0-262-06158-9.

[[06]] D. C. Brown, *Close-range camera calibration*, Photometric Engineering, 37(8) :855–866, 1971.

[[07]] Horst A. Beyer, *Accurate Calibration of CCD cameras*, In Proc. International Conference on Computer Vision and Pattern Recognition (CVPR'92), pages 96–101, Urbana Champaign (USA), June 1992.

[[08]] J. Weng, P. Cohen, and M. Herniou, *Camera Calibration with Distorsion Models and Accuracy Evaluation*, Pattern Analysis and Machine Intelligence (PAMI), 14(10) :965–980, October 1992.

[[09]] Cyril Zeller, *Calibration projective, affine et euclidienne en vision par ordinateur et application à la perception tridimensionnelle*, Thèse de doctorat, École Polytechnique (France), February 1996.

[[10]] Dorian Garcia, *Mesure de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images*, Thèse de doctorat, Institut National Polytechnique de Toulouse (France), December 2001.

[[11]] J.-J. Orteu, *Application de la vision par ordinateur à l'automatisation de l'abattage dans les mines*, PhD thesis, Université Paul Sabatier de Toulouse, 18 novembre 1991.

[[12]] Joan Solà, *Towards Visual Localization, Mapping and Moving Objects Tracking by a Mobile Robot : a Geometric and Probabilistic Approach*, PhD thesis, Institut National Polytechnique de Toulouse (France), February 2007.

[[13]] N. Cornille, D. Garcia, M.A. Sutton, S.R. McNeill, and J.-J. Orteu, *Calibrage d'imageurs avec prise en compte des distorsions*, Instrumentation, Mesure, Métrologie (I2M), 4(3-4/2004) :105–124, September 2005, ISSN 1631-4670.

[[14]] H. B. Nielsen, *Cubic splines*, Lecture notes, Department of Mathematical Modelling, Technical University of Denmark, 1998.

[[15]] P. Brand, R. Mohr, and P. Bobet, *Distorsions optiques : correction dans un modèle projectif*, In Actes du 9ème Congrès AFCET (RFIA'94), pages 87–98, Paris (France), January 1994.

[[16]] Bernard Peuchot, *Utilisation de détecteurs subpixels dans la modélisation d'une caméra*, In Actes du 9ème Congrès AFCET (RFIA'94), pages 691–695, Paris (France), January 1994.

[[17]] Nicolas Cornille, *Accurate 3D Shape and Displacement Measurement using a Scanning Electron Microscope*, PhD thesis - Thèse de doctorat en co-tutelle, University of South Carolina (USA) et Institut National des Sciences Appliquées de Toulouse (France), June 2005.

[[18]] Bill Triggs, Philip F. McLauchlan, Richard I. Hartley and Andrew W., *Fitzgibbon Bundle Adjustment - A Modern Synthesis Proceedings of the International Workshop on Vision Algorithms (ICCV'99)*, Springer-Verlag, London, UK, 2000, pages 298-372, ISBN 3-540-67973-1.

[[19]] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in C — The Art of Scientific Computing*, Cambridge University Press, 2nd edition, 1992.

[[20]] O. Ravn, N. A. Andersen, and A. T. Sorensen, *Auto-calibration in Automation Systems using Vision*, In 3rd International Symposium on Experimental Robotics (ISER'93), pages 206–218, Kyoto (Japan), 1993.

[[21]] Zhengyou Zhang, *A Flexible New Technique for Camera Calibration*, Technical Report MSR-TR-98-71, Microsoft Research, December 1998, Updated on March, 1999.

[[22]] R. I. Hartley and P. Sturm, *Triangulation*, In Computer Vision and Image Understanding (CVIU'97), volume 68 :2, pages 146–157, November 1997.

Webographie

[[03]] Marc Pollefeys, Visual 3D Modeling from Images - A Tutorial. Available online : http://www.cs.unc.edu/~marc/tutorial/.