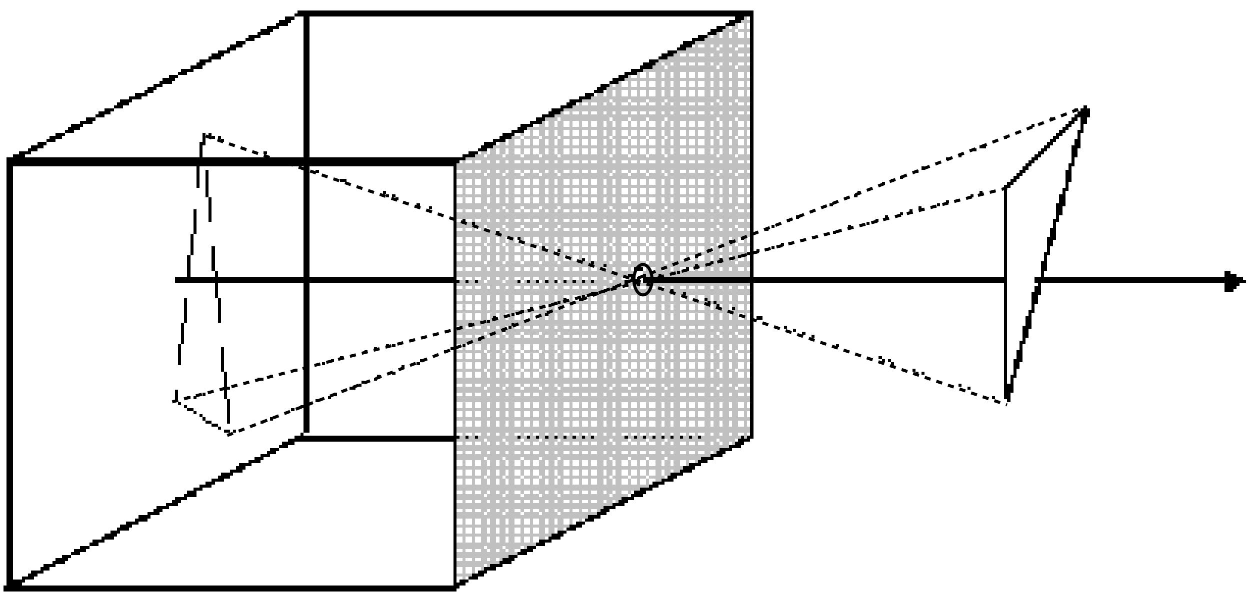
COMPSCI 773 S1C

Camera Calibration

Pinhole Camera

Let a small hole has been punched in the fore screen. Through this hole, some of the light rays, emitted or reflected by objects, form an inverted image of these objects on the back screen. The operation, which creates this inverted image, is called a **perspective projection**.

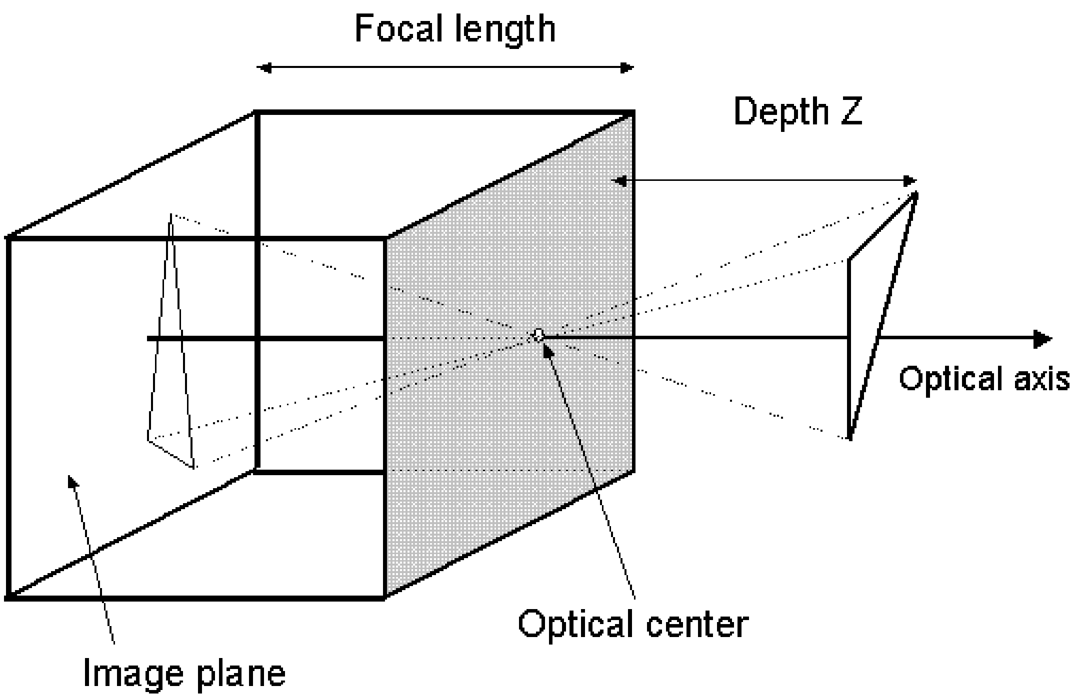


Geometric Pinhole Model

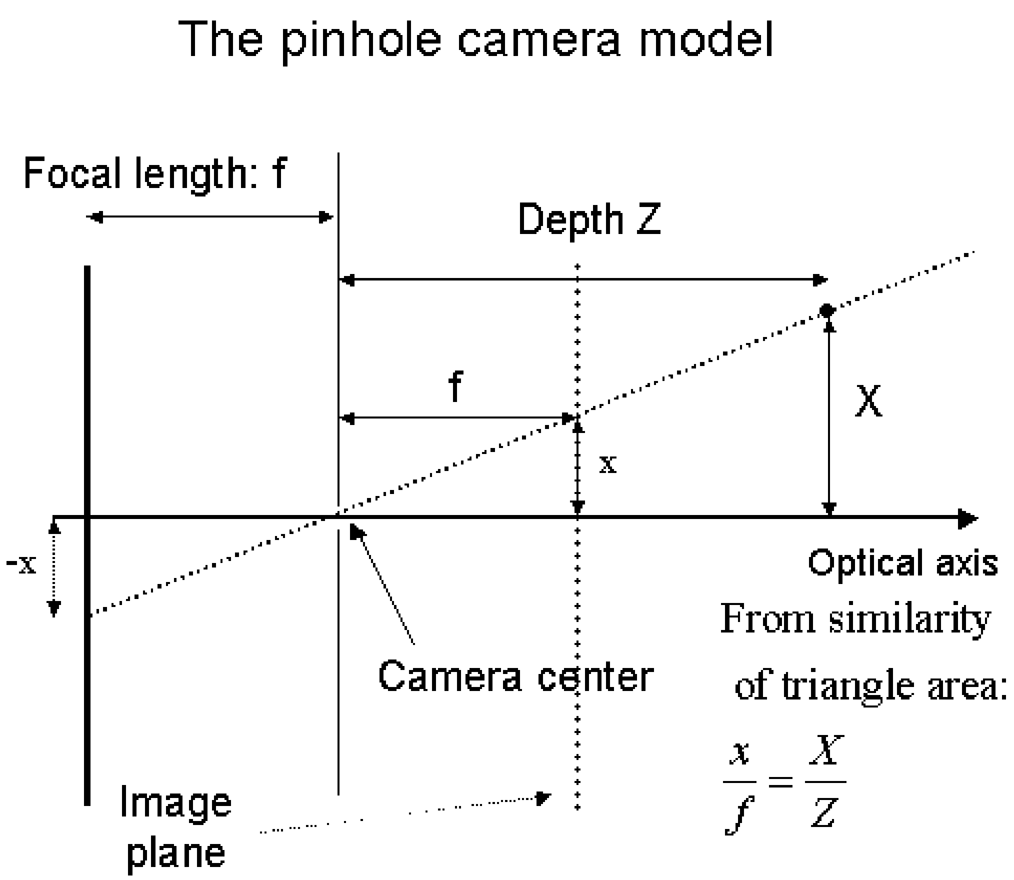
An image is formed in the retinal (or image) plane **R** through a perspective projection. The optical centre **C** at a distance **f** (the focal length of the optical system).

A 3D point **M** at a depth **Z** from **C** is projected onto **R** by intersecting the ray **CM** with **R**. Then, an inverted image of the object is formed.

The pinhole camera model



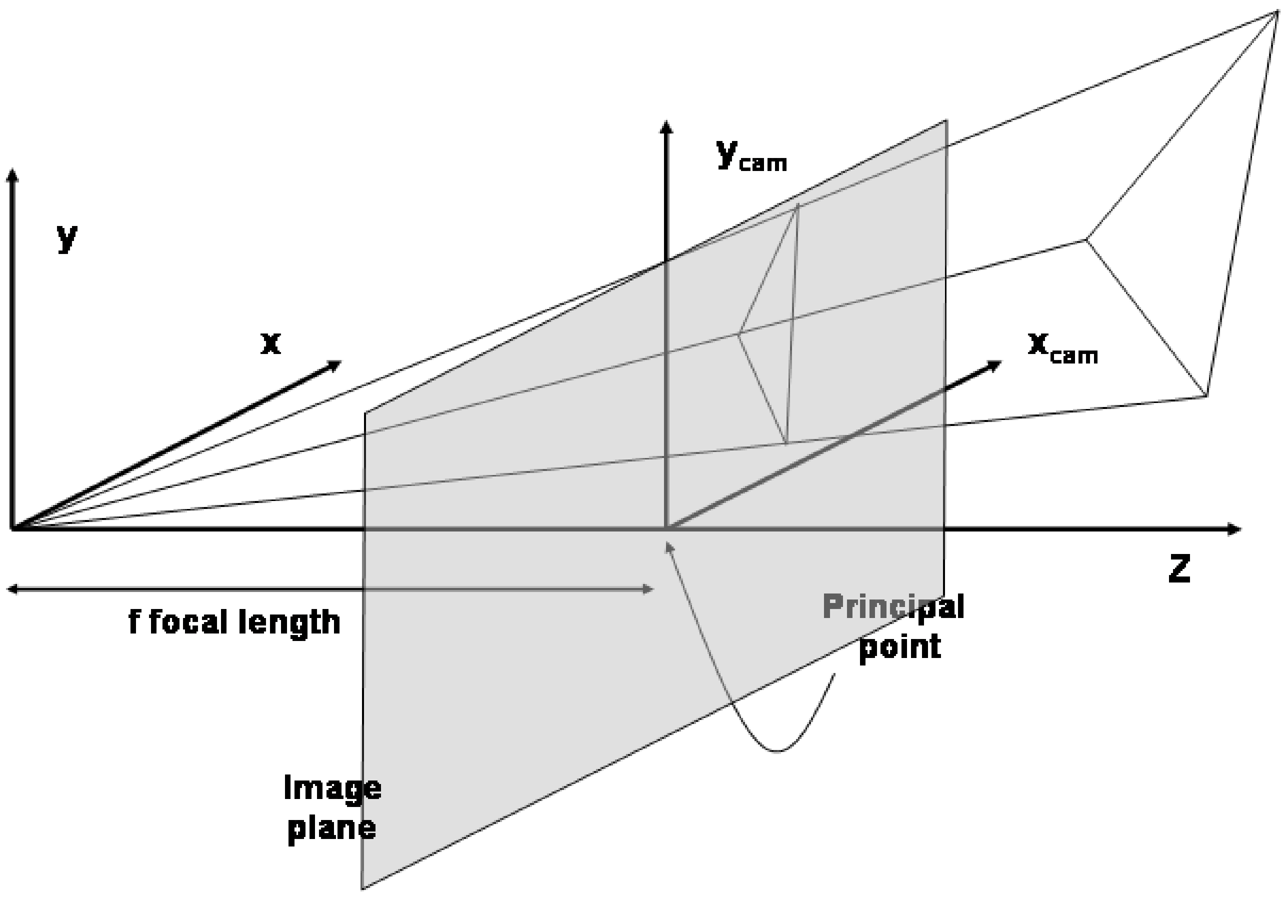
The projection of any object in a plane in front of the focal plane at the focal distance is similar to the rectified image obtained in the retinal plane.



Traditionally, the image plane is drawn in front of the focal point.

.

Projective geometry in 3D



And when you look in 3D.

Camera calibration: definition and purpose

Calibrating a camera means determining the parameters of the imaging process that maps a 3D point in a reference frame onto its 2D image position in pixel units coordinates.

External, or extrinsic, parameters relate to a transformation of the world reference frame into the camera reference frame

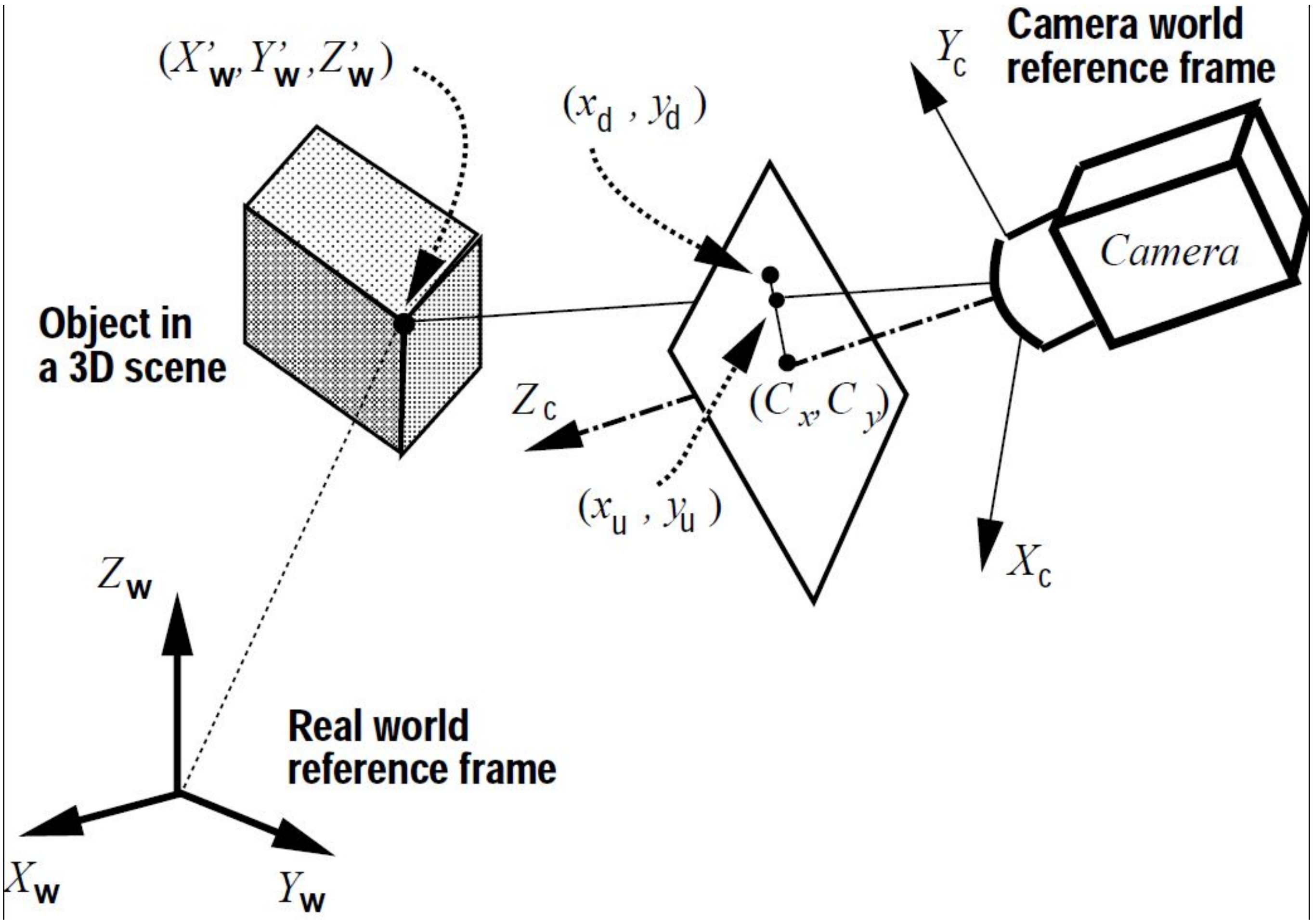
Internal, or intrinsic, parameters relate to the optics and the physics of the camera.

From image pixels, rays can be back projected in the 3D space:

- Respective positions of two pixels provide the angle between the two rays.
- A ray and a depth value provide the position of a 3D point.
- Rays traced from two coupled cameras intersect at the 3D position of any point displayed in the both images.

From the world to the camera

In the 3D world reference frame **WRF** (whatever it might be), a point is measured in pm, mm, inches, light years, parsecs. In the image reference frame, points are measured in pixels, which has the size of a photosensitive receptor of the camera for CCD camera. We can find the affine transform, which gives the position of points in one reference frame to their positions in another one.



Transform Matrix

A 3D point from the world reference frame **WRF**, $(x, y, z, 1)^T$ in homogeneous coordinates, is transformed into its counterpart $(x', y', z', 1')^T$ in the camera reference frame **CRF**:

$$\begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} R_{3*3} & T_{3*1} \\ 0_{1*3} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

The 4×4 transform matrix combines rotation and translation in homogeneous coordinates.

The matrices R and T describe the position and orientation of the camera with respect to the new world reference frame. Their coefficients are called the *extrinsic* parameters of the camera.

Projection into the image plane

3D points in the camera reference frame are projected using the pinhole camera model into the image plane situated at focal distance f of the camera reference frame. Homogeneous coordinates respective to the centre of the image derived from perspective geometry transform of a 3D point onto a plane is given by:

$$\begin{pmatrix} sx_{cam} \\ sy_{cam} \\ s \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, s \neq 0$$

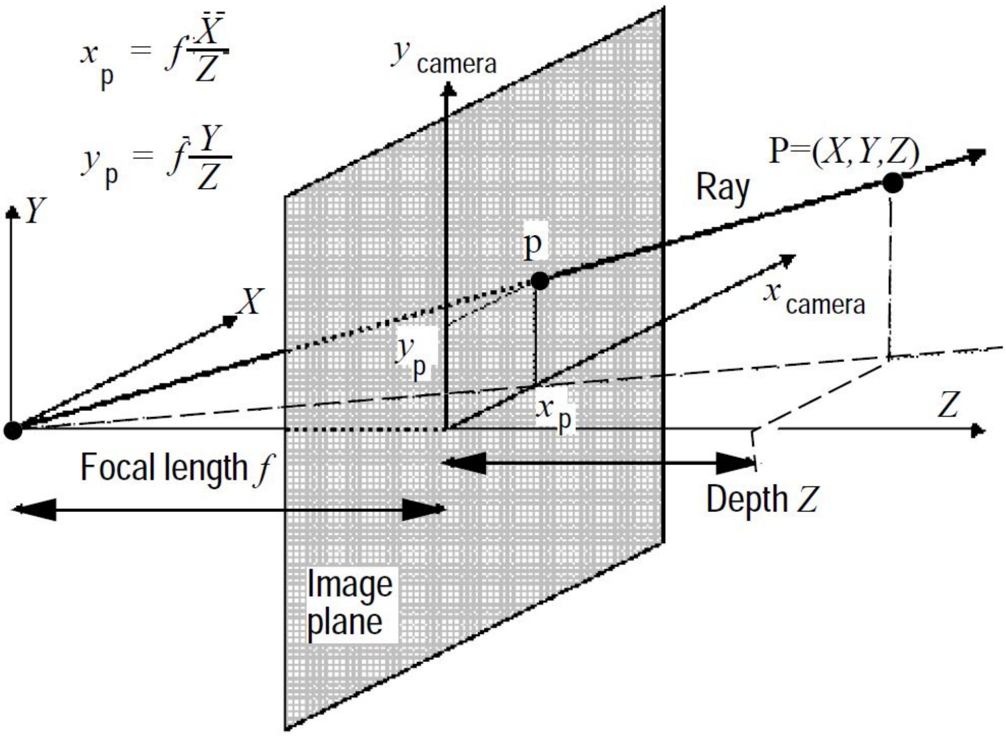
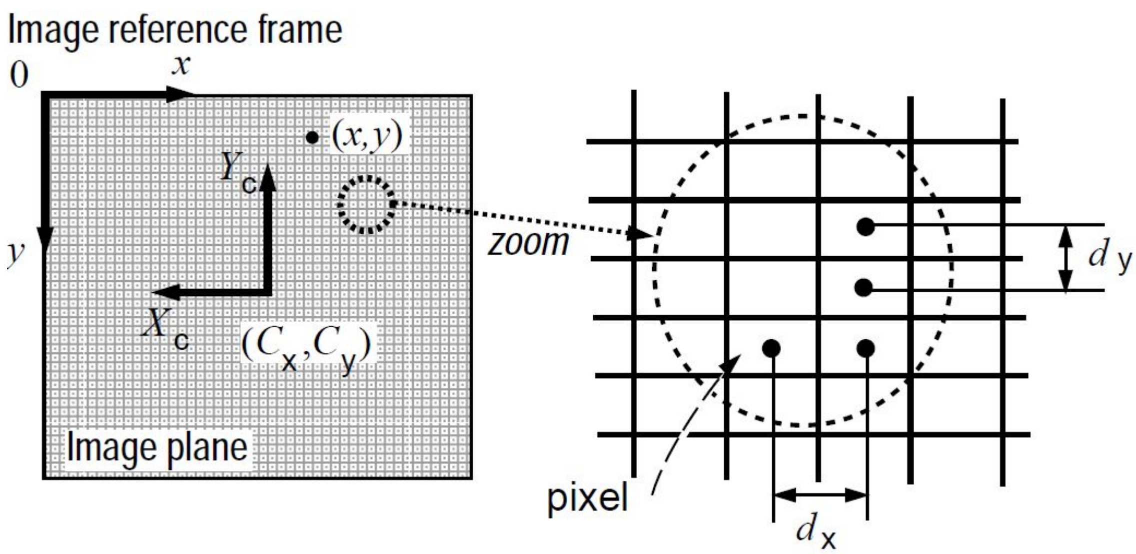


Image processing specification

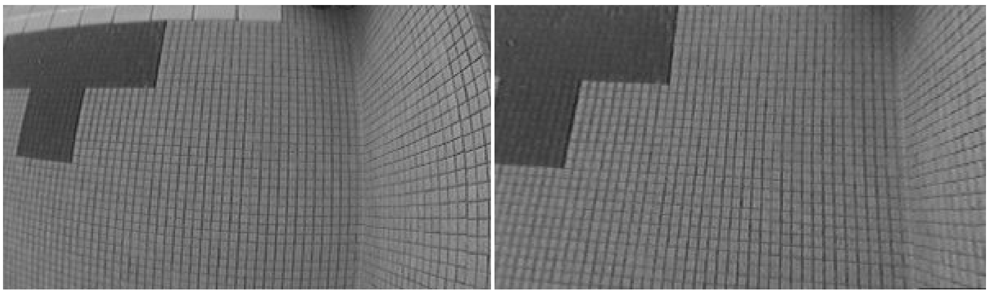
In image processing, the origin of the reference frame is at the top left of the image, x-axis pointing rightward, y-axis downward. Pixels are the unit of measurement and correspond to the size of the photosensitive receptors, which transform the light rays into electrical or logical signals. They are not necessarily square. Coordinates of a point into the pixel-based reference frame is given by:

$$\begin{pmatrix} x_{image} \\ y_{image} \\ 1 \end{pmatrix} = \begin{pmatrix} -{d_x}^{-1} & 0 & x_c \\ 0 & -{d_y}^{-1} & y_c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

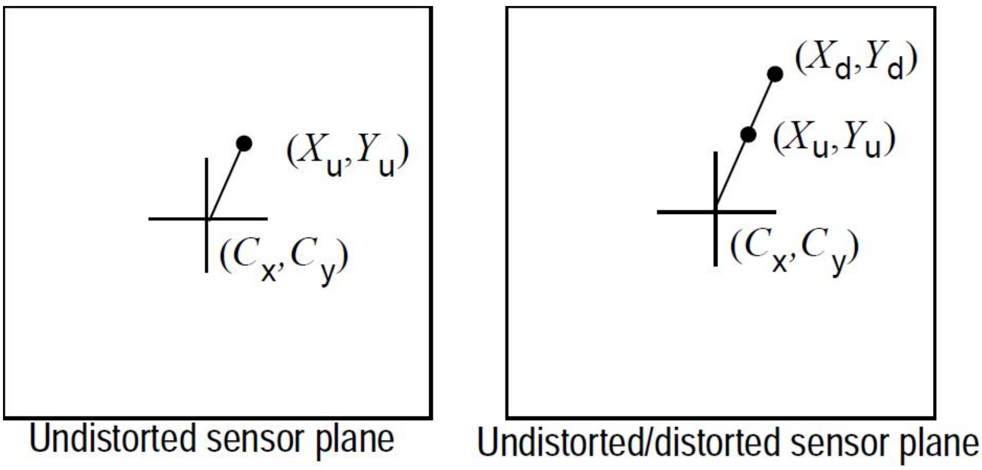


Distortion

Most existing calibration methods assume that the camera design follows (optically and geometrically) the pinhole camera model. Still, some methods consider additional parameters that model image distortions in cameras w.r.t. the pinhole model.



Left: distortion effects on a sub-marine camera. Right: image after correction of distortion



Taking into account κ_1, first order radial lens distortion coefficient (standard procedure), the distorted (or true) image coordinates are given by non-linear equations:

$$X_d(1 + \kappa_1({X^2}_d + {Y^2}_d)) = X_u$$
$$Y_d(1 + \kappa_1({X^2}_d + {Y^2}_d)) = Y_u$$

Camera Specifications

CCD cameras have different building specifications regarding sensors size, numbers, sampling rate, etc. A 1% error can lead to up to 3 to 5 pixels-error for a full frame image (Tsai).

$$\begin{aligned} X_f &= s_x d_x'^{-1} X_d + C_x \quad \text{with } d_x' = d_x \frac{N_c x}{(N_f x)} \\ Y_f &= d_y^{-1} Y_d + C_y \end{aligned}$$

- d_x center-to-center distance between adjacent horizontal sensor elements.

- d_y center-to-center distance between adjacent vertical sensor elements.

- N_{cx} number of sensor elements in the horizontal direction.

- N_{fx} number of pixels in a line as sampled by the computer.

- s_x, uncertainty scale factor for x, due to TV camera scanning and acquisition timing error to account for any uncertainty due to framegrabber horizontal scanline resampling or pixels per line count.

d_x and d_y are usually provided by the manufacturer for CCD cameras. d_y can be doubled if even or odd field used instead of full size frame.

CC procedure

- Determine the 3×4 theoretical calibration matrix
  - Rigid body transform from the world reference frame to the camera reference frame (R, t)
  - projection onto the image plane (f)
  - integrating optics and physics specifications (d'_x, d_y, κ_1)
- Obtain metrics for a set of real points and their corresponding pixel location in the image plane
- Compute the calibration matrix coefficients (point correspondence)
- Compute the intrinsic and extrinsic calibration parameters (optional)
- Estimate your error ratio:
  - Radius of ambiguity in ray tracing: *distance between the back projected points (on the test plane) and the ideal point in the same plane.*
  - Accuracy of 3D coordinate measurement obtained through stereo triangulation using the calibrated camera parameters.

Calibration Summary

- We can use homogeneous coordinates to map the pinhole projection camera model of 3D points (real values) to the image plane (positive integer values) by:

$$x_{\mathbb{N}^3} = PX_{\mathbb{R}^4}$$

  - Transform the 3D points to the camera reference frame (meter to meter)
  - Project the points onto the image plane (3D to 2D)
  - Map the points on the camera image plane to the image plane coordinates (meter to pixels)

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{-1}{d_x} & 0 & x_c \\ 0 & \frac{-1}{d_y} & y_c \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

- P is a 3×4 matrix which can be written as:

$$P = K \left[\begin{array}{ccc} R & | & t \end{array} \right]$$

where K is the camera calibration matrix for the intrinsic (internal) parameters, **R** representing the rotation matrix describing the orientation of the camera with respect to the **WRF** and **t** the translation vector representing the translation of the camera origin with respect to **WRF**

Solving the system

All characteristics of the camera can be expressed by the following 3×4 matrix which transforms a 3D point $(x, y, z)^T$ into its pixel counterpart $(u, v)^T$ in the image:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} s \neq 0$$

Unknowns are the elements of the matrix, which can be related to the intrinsic and extrinsic parameters of the camera. Let us rewrite the system as:

$$\begin{cases} su = ax + by + cz + d \\ sv = ex + fy + gz + h \\ s = ix + jy + kz + 1 \end{cases}$$

Solving for u and v:

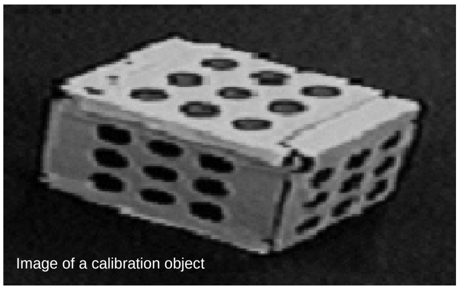
$$\begin{cases} u = ax + by + cz + d - ixu - jyu - kzu \\ v = ex + fy + gz + h - ixv - jyv - kzv \end{cases}$$

yields the following linear system for a given set of points M_i, $i \epsilon \{1, ..., N\}$, whose $3D$ coordinates $(x_i, y_i, z_i)^T$ and corresponding 2D pixels (u_i, v_i) are known (or measured):

$$\begin{pmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1u_1 & -y_1u_1 & -z_1u_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1v_1 & -y_1v_1 & -z_1v_1 \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -x_2u_2 & -y_2u_2 & -z_2u_2 \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -x_2v_2 & -y_2v_2 & -z_2v_2 \\ x_3 & y_3 & z_3 & 1 & 0 & 0 & 0 & 0 & -x_3u_3 & -y_3u_3 & -z_3u_3 \\ 0 & 0 & 0 & 0 & x_3 & y_3 & z_3 & 1 & -x_3v_3 & -y_3v_3 & -z_3v_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \\ j \\ k \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{pmatrix}$$

Calibration tools

- Design a Calibration object with either planar or non-coplanar points.



- Geometry and measurements of the 3D points should be precisely determined.

- Extract (manually or automatically) feature points (patches, corners or predefine patterns) from the image.

- Do the calibration (Tsai or else).

- Assess the quality of your calibration (discrepancy between the real 3D coordinates of test points and their computed coordinates after calibration)

- Refine your calibration process