# Lecture 25:
# 3-D Reconstruction I:
# Camera Geometry

**Contents**









# Motivation

- One central goal of computer vision:
  Extract information about a 3-D world from 2-D images.
- To this end, we first have to investigate how 2-D images arise from a 3-D world.
- First we consider a single camera *(monocular vision).*
  We model the imaging process with the so-called pinhole camera model.
  This requires some (single view) *projective geometry.*
- The next lecture deals with two cameras *(binocular vision, stereo vision).*
  The corresponding stereo geometry is called *epipolar geometry*.

Projective geometry has a long tradition in art. Already in 1525, the painter Albrecht Dürer published the monograph "Underweysung der Messung mit dem Zirckel und Richtscheyt", in which he performed studies on concepts from projective geometry. His book became surprisingly popular, not only among painters. Source: `http://www.aphilia.de/kunst-albrecht-duerer-05-perspektive.html`



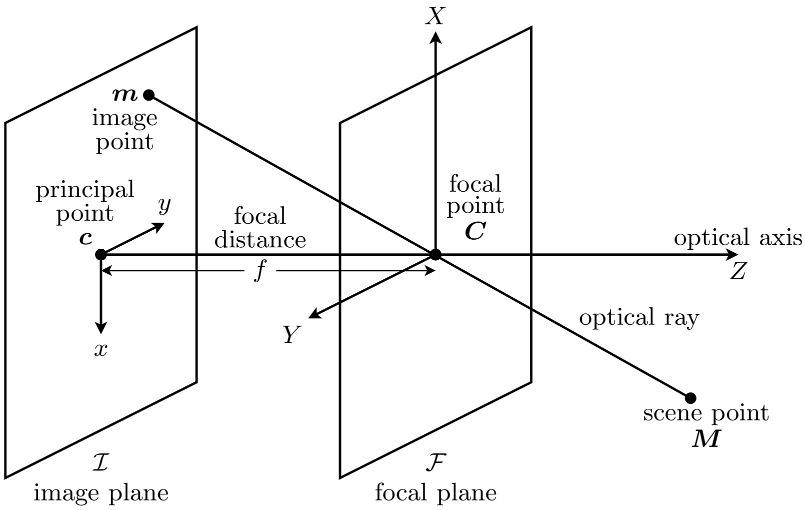


# The Pinhole Camera Model (Lochkameramodell)

- simple but fairly realistic model of a camera system
- describes perspective projection of the 3-D space onto a 2-D image plane
- maps 3-D camera coordinates $\boldsymbol{M} = (X, Y, Z)^\top$ with centre $\boldsymbol{C}$ to 2-D image coordinates $\boldsymbol{m} = (x, y)^\top$ with centre $\boldsymbol{c}$

MI A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

## The Pinhole Camera Model (2)




Pinhole camera model. Author: M. Mainberger. (2010). Notations:

| | |
|---|---|
| $\boldsymbol{M}$: *scene point (Szenenpunkt):* | location of the object |
| $\boldsymbol{C}$: *focal point (Brennpunkt), optical centre:* | location of the pinhole |
| $\mathcal{F}$: *focal plane (Brennebene):* | specifies camera orientation, contains focal point $\boldsymbol{C}$ |
| *optical axis (optische Achse):* | orthogonal to focal plane, passes through $\boldsymbol{C}$ |
| $\mathcal{I}$: *image plane (Bildebene):* | parallel to focal plane |
| $f$: *focal distance (Brennweite):* | distance between focal plane $\mathcal{F}$ and image plane $\mathcal{I}$ |
| $\boldsymbol{c}$: *principal point (Hauptpunkt):* | intersection between image plane and optical axis |
| *optical ray (Sichtstrahl):* | passes through $\boldsymbol{M}$ and $\boldsymbol{C}$ |
| $\boldsymbol{m}$: *image point (Bildpunkt):* | intersection between optical ray and image plane |



## The Pinhole Camera Model (3)

### Drawback of this Representation

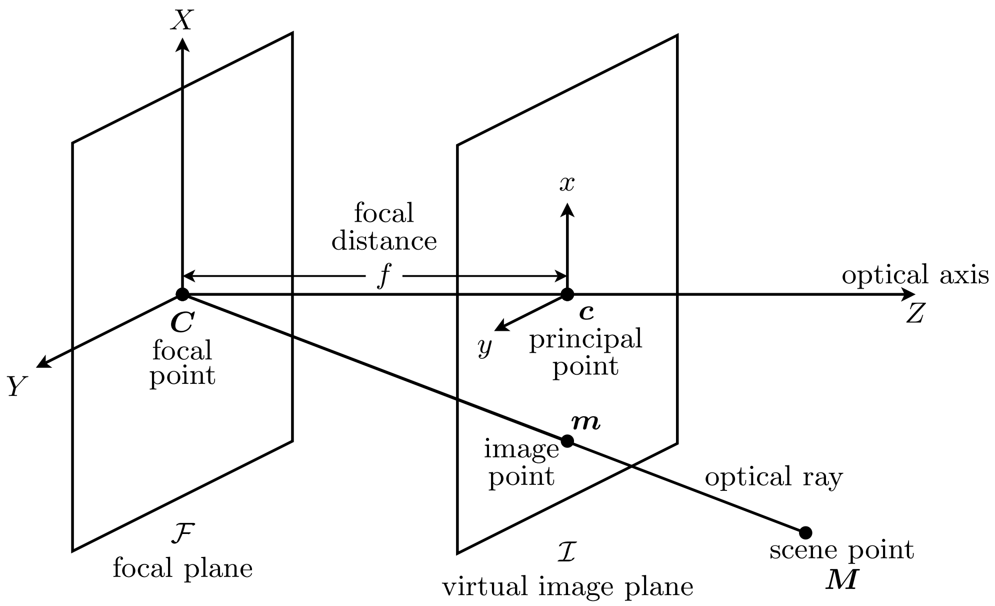
- The projective mapping inverts the orientation:
  Objects oriented upwards in the real world appear downwards in the image.

### Simplification

- Instead of the image plane behind the focal plane,
  consider a *(virtual) image plane* in front of the focal plane (with distance $f$).
- Then objects oriented upwards in the real world appear upwards in the image.

## The Pinhole Camera Model (4)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28




Pinhole camera model with virtual image plane in front of the focal plane. Author: M. Mainberger.

## The Pinhole Camera Model (5)

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

### Basic Geometric Relation

- The theorem of intersecting lines (Strahlensatz) yields the *basic equation*

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z}.$$

- It fully describes the projective geometry behind the pinhole model.

### Problems with this Formulation

- *Nonuniqueness:*
  Two points $\boldsymbol{M}_1 := (X, Y, Z)^\top$ and $\boldsymbol{M}_2 := (wX, wY, wZ)^\top$ are mapped to the same image point $\boldsymbol{m} = (x, y)^\top$. Thus, the depth information is lost.

- *Nonlinearity:*
  The mapping from the 3-D camera coordinates $(X, Y, Z)^\top$ to the 2-D image coordinates $(x, y)^\top$ is a nonlinear transformation that involves divisions:

$$x = f\,\frac{X}{Z}, \qquad y = f\,\frac{Y}{Z}.$$

Introducing so-called homogeneous coordinates addresses both problems.

## The Pinhole Camera Model (6)

The fact that all scene points on the optical ray are mapped to the same image point creates ambiguities.
Sources: `taz.de`, `www.spiegel.de`.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

## Homogeneous Coordinates (1)

# Homogeneous Coordinates

### Motivation and Definition

- Homogeneous coordinates are an elegant tool for describing the *nonlinear* projective camera geometry by means of matrices, i.e. *linear* mappings.
- The price one has to pay for this is one additional coordinate.
- Transformation from standard coordinates to homogeneous coordinates:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} wx \\ wy \\ w \end{pmatrix}$$

  with some arbitrary scaling factor $w \neq 0$.
- For the backtransformation, divide the first two coordinates by the third.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

## Application to the Basic Equation

- The basic equation

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z}$$

states that

$$\begin{aligned} Zx &= fX, \\ Zy &= fY. \end{aligned}$$

- Reformulation with homogeneous coordinates:
Introduce new variables $u := wx$ and $v := wy$ with a scaling factor $w := Z$:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\boldsymbol{P}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

Note that in homogeneous coordinates, the basic equation is now *linear*.
The matrix $\boldsymbol{P}$ is called *projection matrix*.



M I A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

- Short notation of the basic equation:

$$w\,\tilde{\boldsymbol{m}} = \boldsymbol{P}\,\tilde{\boldsymbol{M}}$$

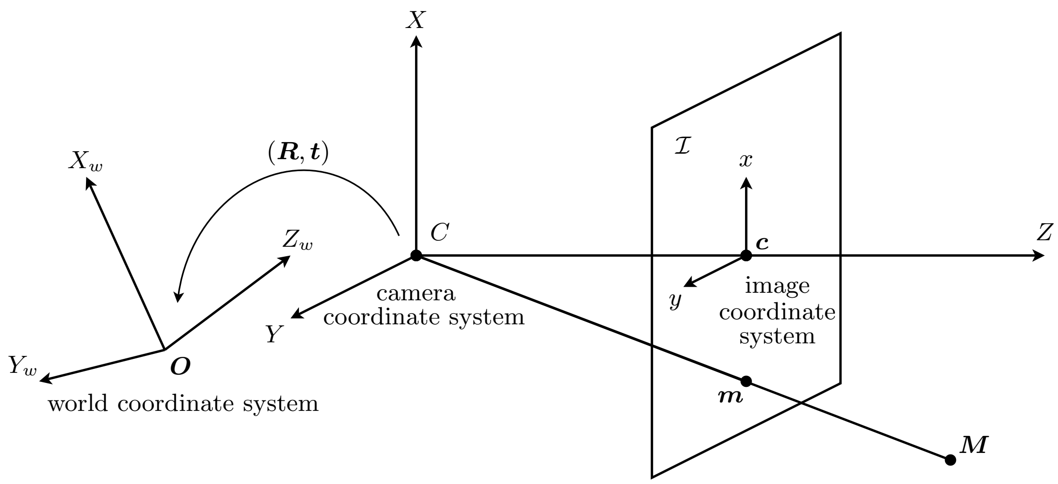
where the tilde notation extends a vector with the additional component $1$:

$$\tilde{\boldsymbol{m}} := \begin{pmatrix} \boldsymbol{m} \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$\tilde{\boldsymbol{M}} := \begin{pmatrix} \boldsymbol{M} \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

# Extrinsic and Intrinsic Parameters




World coordinate system $(X_w, Y_w, Z_w)$, camera coordinate system $(X, Y, Z)$, and image coordinate system $(x, y)$. The position of the world coordinate system relative to the camera coordinate system can be described by six extrinsic parameters: three for rotation, and three for translation. Author: M. Mainberger.



M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

## Extrinsic Camera Parameters

- describe the position of the world coordinate system relative to the camera coordinate system
- In homogeneous coordinates, 3-D transformations such as translations and rotations can be expressed by multiplications with $4 \times 4$ matrices.
- Translation of world coordinates by $(t_1, t_2, t_3)^\top$ is described by multiplication with

$$\boldsymbol{T} = \begin{pmatrix} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

- Rotation of world coordinates with a $3 \times 3$ matrix $(r_{i,j})$ is characterised by

$$\boldsymbol{R} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & 0 \\ r_{2,1} & r_{2,2} & r_{2,3} & 0 \\ r_{3,1} & r_{3,2} & r_{3,3} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

## Extrinsic and Intrinsic Parameters (3)

- Example: Rotation around the $Z$ axis with angle $\phi$:

$$\boldsymbol{R} = \begin{pmatrix} \cos\phi & -\sin\phi & 0 & 0 \\ \sin\phi & \cos\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- The matrices can be concatenated, but in general they do not commute:

$$\boldsymbol{TR} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \neq \boldsymbol{RT}.$$

- For the transformation between two 3-D coordinate systems, we have 6 free parameters: 3 for translation and 3 for rotation (1 angle per axis).

- Since they only depend on the camera orientation, but not on internal camera specifics, they are called *extinsic camera parameters*.



## Extrinsic and Intrinsic Parameters (4)

### Intrinsic Camera Parameters

- characterise the geometry of the image plane inside the camera

- Problems:
  - Origin of the image plane can be located in another point than the principal point, e.g. at the top left.
    Let the principal point in this coordinate system be located in $(u_0, v_0)^\top$.
  - Pixels may have different sizes $h_u$ and $h_v$.
  - In the worst case, the coordinate axes might even have an angle $\theta \neq \frac{\pi}{2}$.

- One can show that these five intrinsic parameters lead to a matrix

$$\boldsymbol{H} = \begin{pmatrix} h_u & -h_u \cot\theta & u_0 \\ 0 & h_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It describes the transition from ideal image coordinates to real (pixel) coordinates.

## General Projective Mapping in Homogeneous Coordinates

- Concatenating extrinsic, projection and intrinsic matrices gives the *full projective mapping*.
- It maps a 3-D point in homogeneous world coordinates $(X_w, Y_w, Z_w, 1)^\top$ to a 2-D image point with homogeneous pixel coordinates $(u, v, w)^\top$:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} h_u & -h_u \cot\theta & u_0 \\ 0 & h_v/\sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic matrix}} \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection matrix}} \underbrace{\begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{extrinsic matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$=: \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix}}_{\text{full projection matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- 12 parameters in total, but with one free scaling parameter:
  Usually the length scale is chosen such that the focal length $f$ is normalised to 1.
  This leaves 11 degrees of freedom: 6 extrinsic and 5 intrinsic parameters.



MI A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

# Camera Calibration

- denotes the estimation of the 5 intrinsic and 6 extrinsic camera parameters
- Many algorithms have been proposed in the literature.
- Basic idea: Investigate image of an object of known size and shape.
- Each identified point correspondence gives 2 constraints.
- Thus, for estimating 11 parameters, one has to find 6 corresponding points.
- Taking into account more point correspondences (e.g. in a least squares sense) makes the estimation less sensitive w.r.t. errors.

# Projective Mappings

## Projective Mapping in Nonhomogeneous Coordinates

- From the *linear* representation of the projective mapping in homogeneous coordinates,

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix},$$

we obtain a *nonlinear* transformation in nonhomogeneous coordinates:

$$x = \frac{u}{w} = \frac{p_{1,1}\,X_w + p_{1,2}\,Y_w + p_{1,3}\,Z_w + p_{1,4}}{p_{3,1}\,X_w + p_{3,2}\,Y_w + p_{3,3}\,Z_w + p_{3,4}},$$

$$y = \frac{v}{w} = \frac{p_{2,1}\,X_w + p_{2,2}\,Y_w + p_{2,3}\,Z_w + p_{2,4}}{p_{3,1}\,X_w + p_{3,2}\,Y_w + p_{3,3}\,Z_w + p_{3,4}}.$$



M I A

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

## Properties of the Perspective Projection

- Convex sets in 3-D are mapped to convex sets in 2-D.
- Parallel lines in 3-D converge in 2-D to the *vanishing point (Fluchtpunkt)*.
- The centre of gravity in 3-D is *not* mapped to the centre of gravity in 2-D.
- It is a nonlinear transformation that can be ill-posed:
  If the denominator $w$ approaches $0$, small perturbations in the 3-D data can create large deviations in the projected image.

Are there special cases that are more convenient to handle ?

Example image illustrating that under perspective projection, parallel lines in 3-D converge in the 2-D image to the vanishing point. Photo: J. Weickert.





# Affine Mappings

- affine approximation to the nonlinear perspective projection
- realistic if we have only a small relative variation of the depth

$$Z \;=\; w \;=\; p_{3,1}\,X_w + p_{3,2}\,Y_w + p_{3,3}\,Z_w + p_{3,4}\,.$$

  Example: satellite images.

- In this case the denominator

$$w \;=\; p_{3,1}\,X_w + p_{3,2}\,Y_w + p_{3,3}\,Z_w + p_{3,4}$$

  in the nonlinear projective mapping can be approximated by a constant $c$.

- Thus, three entries of the perspective projection matrix can be set to $0$:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ 0 & 0 & 0 & c \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}.$$

## Affine Mappings (2)

- This leads to the *affine* transformation

$$x = \frac{u}{w} = \frac{p_{1,1}\,X_w + p_{1,2}\,Y_w + p_{1,3}\,Z_w + p_{1,4}}{c},$$

$$y = \frac{v}{w} = \frac{p_{2,1}\,X_w + p_{2,2}\,Y_w + p_{2,3}\,Z_w + p_{2,4}}{c}.$$

### Properties of the Affine Approximation

- Parallel lines in 3-D are parallel in 2-D.
- At first glance, it seems to have 8 degrees of freedom instead of 11.
  However, the centre of gravity in 3-D is mapped to the centre of gravity in 2-D.
  This additional constraint reduces the degrees of freedom to 6.
- mathematically well-posed:
  Small perturbations in 3-D create small deviations in 2-D.

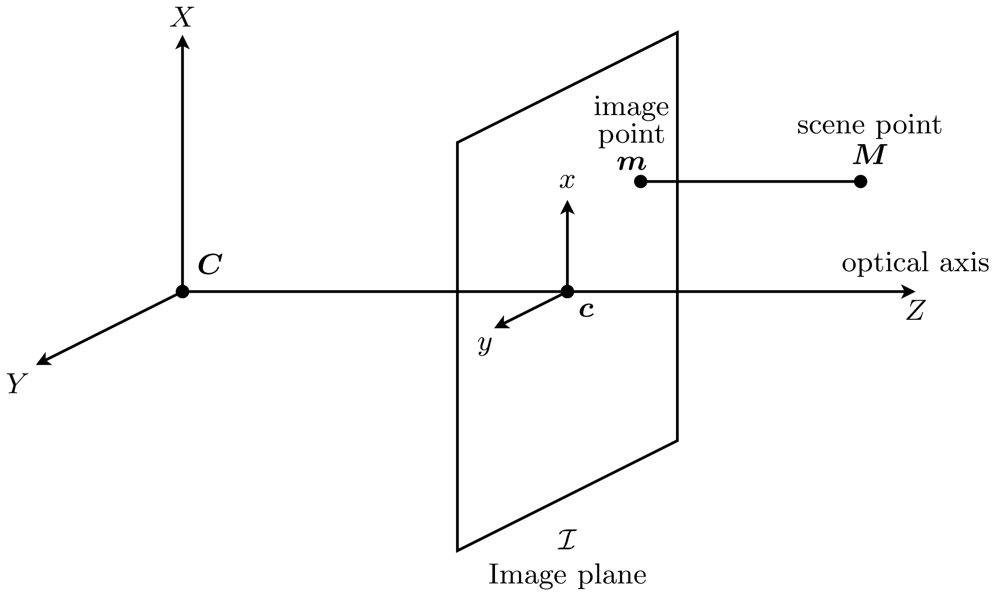


## Affine Mappings (3)

A satellite image that has a small depth variation compared to the object distance. This justifies to replace the projective camera model by its affine approximation. Note that in the affine model, parallel lines in the real world remain parallel in the image. Source: Google Earth / Digital Globe.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

# Orthographic Mappings




An orthographic mapping is much simpler than the perspective projection model of a pinhole camera. It simply maps a camera coordinate $\boldsymbol{M} = (X, Y, Z)^\top$ to an image coordinate $\boldsymbol{m} = (x, y)^\top$ by setting $x := X$ and $y := Y$. Author: M. Mainberger.



M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

- simplest projection model
- maps a camera coordinate $\boldsymbol{M} = (X, Y, Z)^\top$ to an image coordinate $\boldsymbol{m} = (x, y)^\top$ by setting $x := X$ and $y := Y$.
- completely ignores the depth information
- too crude for most applications
- still popular for specific applications, e.g. shape from shading (Lecture 27).

# Summary

- The pinhole camera model decribes a perspective projection.
- Homogeneous coordinates give a linear description of this nonlinear transformation. They use one additional coordinate.
- The general projective model has 5 intrinsic and 6 extrinsic parameters. The focal length is normalised to 1.
- The extrinsic parameters consists of 3 translation and 3 rotation parameters.
- The 5 intrinsic parameters comprise
  – the centre of the image coordinate system (2 parameters),
  – the pixel size (2 parameters),
  – and the angle between the coordinate axes.
- The estimation of the intrinsic and extrinsic parameters is called camera calibration.
- The affine camera model is a simplification of the projective model. It is valid if the relative depth variation of the object is small. It contains only 6 free parameters.
- The orthographic projection model simply ignores the depth.





# References


- G. Xu, Z. Zhang: *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Kluwer, Dordrecht, 1996.
  *(well-written description of the pinhole camera model)*
- R. Klette, K. Schlüns, A. Koschan: *Computer Vision: Three-Dimensional Data from Images*. Springer, Singapore, 1998.
  *(see Section 2.1, also for camera calibration)*
- O. Faugeras: *Three-dimensional Computer Vision*. MIT Press, Cambridge, MA, 1993.
  *(Chapter 2 deals with projective geometry.)*
- O. Faugeras, Q.-T. Luong, T. Papadopoulo: *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.
  *(Chapter 2 on projective, affine and Euclidean geometry)*
- R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.
  *(Part I deals with single view geometry.)*


MI A 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28