Lecture 26:
3-D Reconstruction I:
Camera Geometry

Contents

Motivation

- The ultimate goal of computer vision is to extract information about a 3-D world from 2-D images.
- To understand this mechanism, we first have to investigate how 2-D images arise from a 3-D world.
- In the simplest case we consider a single camera and the so-called pinhole camera model. This scenario is called *monocular vision*. It requires some single view projective geometry.
- More advanced situations such as two cameras *(binocular vision, stereo vision)* and the corresponding *epipolar geometry* are treated in the next lecture.

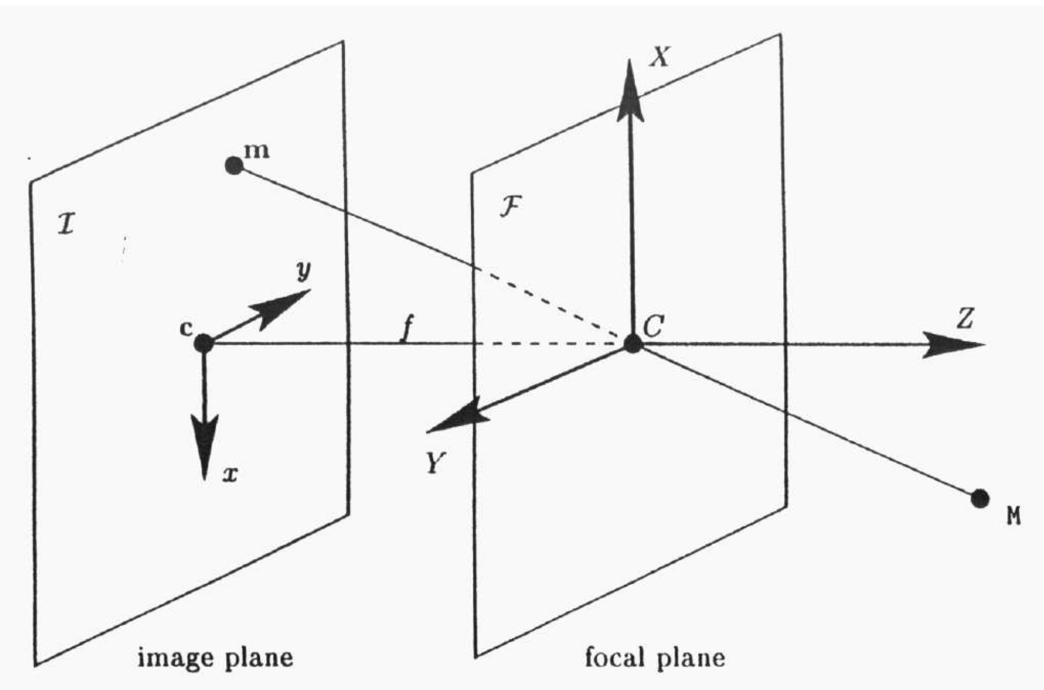
Motivation (2)

Albrecht Dürer, claiming to perform research on the formation of 2-D images from a 3-D world. This drawing is used on the cover page of a computer vision book by O. Faugeras (1993).

MIA
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

The Pinhole Camera Model (1)

The Pinhole Camera Model (Lochkameramodell)



Pinhole camera model. Authors: G. Xu, Z. Zhang (1996).

MIA
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

The Pinhole Camera Model (2)

- The pinhole camera model is a simple but fairly realistic model of a camera system.
- perspective projection of the 3-D space onto a 2-D image plane
- maps 3-D camera coordinates $\mathbf{M} = (X, Y, Z)^\top$ with centre $\mathbf{C}$ to 2-D image coordinates $\mathbf{m} = (x, y)^\top$ with centre $\mathbf{c}$
- Notations:
  $\mathbf{M}$: *scene point (Szenenpunkt)*
  $\mathbf{C}$: *focal point (Brennpunkt), optical centre*: location of the pinhole
  $\mathbf{m}$: *image point (Bildpunkt)*
  $\mathcal{I}$: *image plane (Bildebene)*
  $\mathcal{F}$: *focal plane (Brennebene)*: coplanar to image plane, contains focal point $\mathbf{C}$
  optical axis (optische Achse): orthogonal to focal plane, passes through $\mathbf{C}$
  optical ray (Sichtstrahl): passes through $\mathbf{M}$ and $\mathbf{C}$
  f: *focal distance (Brennweite)*: distance image plane $\mathcal{I}$ – optical centre $\mathbf{C}$
  $\mathbf{c}$: *principal point (Hauptpunkt)*: intersection between image plane and optical axis

The Pinhole Camera Model (3)

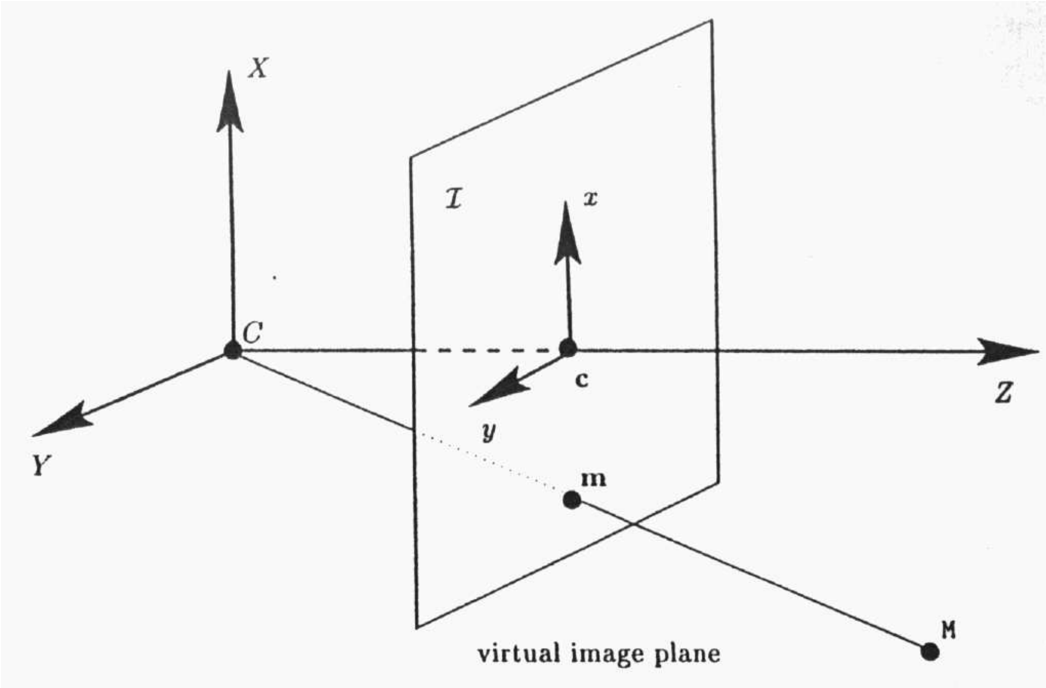
Drawback of this Representation:

- 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.
- In this way, objects that are oriented upwards in the real world appear upwards in the image.

The Pinhole Camera Model (4)



Pinhole camera model with virtual image plane in front of the focal plane. Authors: G. Xu, Z. Zhang (1996).

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

The Pinhole Camera Model (5)

Basic Geometric Relation

- Theorem of intersecting lines (Strahlensatz) yields

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

- Note the nonuniqueness:
  All points on the optic ray are mapped onto the same image point:
  Two points $\mathbf{M_1} := (X, Y, Z)^\top$ and $\mathbf{M_2} := (wX, wY, wZ)^\top$ are mapped to the same image point $\mathbf{m} = (x, y)^\top$.
  Thus, the depth information is lost.

- The preceding basic equation describes the projective geometry, but is unpleasant to work with:
  It is a nonlinear transformation between $(X, Y, Z)^\top$ and $(x, y)^\top$, that involves divisions.

Is there a simple remedy ?
As often in image analysis, a suitable transformation can help.

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

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 an 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

Example

- The basic equation

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

  states that

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

- Rewriting this in homogeneous coordinates can be done by introducing new variables $u := wx$ and $v := wy$ with a scaling factor $w := Z$:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

  The occurring matrix is called *projection matrix* P.

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

- Short notation:

$$w\,\tilde{\mathbf{m}} \;=\; P\,\tilde{\mathbf{M}}$$

where the tilde denotes a vector supplemented with the additional component 1:

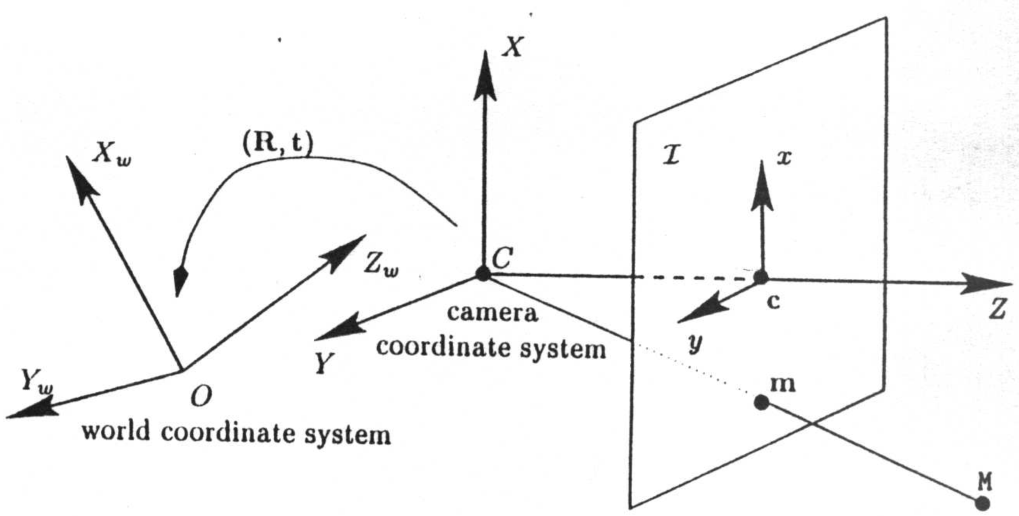
$$\tilde{\mathbf{m}} \;:=\; \begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

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

- Note that the tilde describes a transition from standard to homogeneous coordinates.

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

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 transitions between these coordinate sytems depend on the position of the camera and its internal characteristics. Extrinsic and intrinsic camera parameters characterise these transformations. Authors: G. Xu, Z. Zhang (1996).

Extrinsic and Intrinsic Parameters (2)

Extrinsic Camera Parameters

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

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

- Rotation of the world coordinates with a 3×3 matrix (r_{ij}) is characterised by

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Extrinsic and Intrinsic Parameters (3)

- Example: Rotation around the Z axis with angle ϕ:

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

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

$$TR = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \neq RT$$

- 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 the *extrinsic camera parameters*.

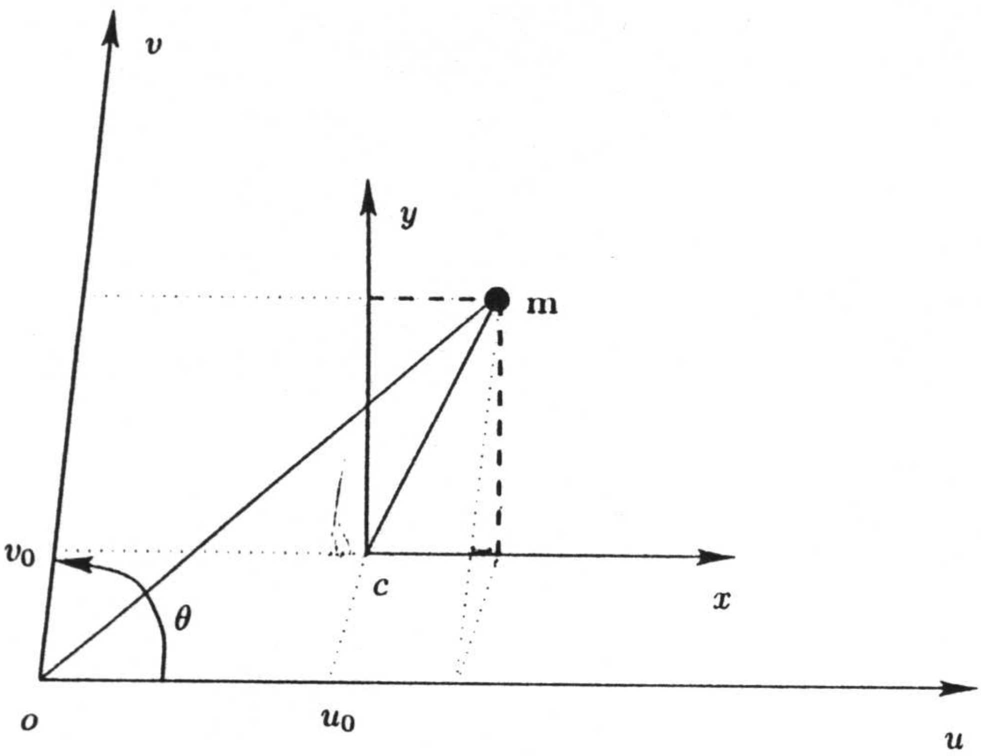
Intrinsic Camera Parameters

- 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 dimensions k_u and k_v.
  - In the worst case, the coordinate axis may have an angle $\theta \neq \pi/2$.
- One can show that these 5 intrinsic parameters lead to a matrix

$$H = \begin{pmatrix} k_u & -k_u \cot\theta & u_0 \\ 0 & k_v/\sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

that decribes the transition from the ideal camera coordinates to the measured coordinates.

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



Intrinsic camera parameters. Authors: G. Xu, Z. Zhang (1996).

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

Extrinsic and Intrinsic Parameters (6)

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 coordinates $(u, v, w)^\top$:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \underbrace{\begin{pmatrix} k_u & -k_u \cot\theta & u_0 \\ 0 & k_v / \sin\theta & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\text{intrinsic matrix}} \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\text{projection matrix}} \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{extrinsic matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

$$=: \underbrace{\begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}}_{\text{full projection matrix}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- 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.

Camera Calibration

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 on 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 (1)

MIA 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

Projective Mappings

Projective Mapping in Nonhomogeneous Coordinates

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

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix},$$

we obtain a nonlinear transformation in nonhomogeneous coordinates:

$$\begin{aligned} x &= \frac{u}{w} = \frac{p_{11}X_w + p_{12}Y_w + p_{13}Z_w + p_{14}}{p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}}, \\ y &= \frac{v}{w} = \frac{p_{21}X_w + p_{22}Y_w + p_{23}Z_w + p_{24}}{p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}}. \end{aligned}$$

Projective Mappings (2)

MIA 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

Properties of the Perspective Projection

- Convex sets in 3-D are mapped to convex sets in 2-D.
- Parallel Lines in 3-D intersect in 2-D in the *vanishing point (Fluchtpunkt)*.
- The centre of gravity in 3-D is *not* mapped to the centre of gravity in 2-D.
- nonlinear transformation that can be ill-posed: 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 ?

Affine Mappings

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

$$Z = w = p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}.$$

  Example: satellite images.

- In this case the denominator $p_{31}X_w + p_{32}Y_w + p_{33}Z_w + p_{34}$ in the nonlinear projective mapping can be approximated by a constant c.
  Thus, three entries the perspective projection matrix can be set to zero:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & c \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

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

- This leads to the *linear* transformation

$$x = \frac{u}{w} = \frac{p_{11}X_w + p_{12}Y_w + p_{13}Z_w + p_{14}}{c},$$

$$y = \frac{v}{w} = \frac{p_{21}X_w + p_{22}Y_w + p_{23}Z_w + p_{24}}{c}.$$

Properties of the Affine Approximation

- Parallel lines in 3-D are parallel in 2-D.
- at first glance, only 8 degrees of freedom instead of 11, but:
- 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.

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

Summary

- The pinhole camera model decribes a perspective projection.
- Homogeneous coordinates allow a linear description of this nonlinear transformation. They use an 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 reasonable linear simplification of the projective model, if the relative depth variation of the object is small. It contains only 6 free parameters.

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

Literature

- 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.)

Assignment P7 – Programming Work

Please download the required files from the webpage

`http://www.mia.uni-saarland.de/Teaching/ipcv07.shtml`

into your own directory. You can unpack them with the command `tar xvzf Ex07.tgz`.

Problem 0 (Evaluation) (0)

Please go to the Clix webpage and evaluate this class by Friday, February 8:

`http://frweb.cs.uni-sb.de/03.Studium/08.Eva/`

Remember that this is the only possibility to improve lectures, so please state what you like and what you dislike. Even if you like everything to be kept the way it is, it is important to have your opinion, since things might be changed otherwise.

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

Problem 1 (Optic Flow) (12+8)

(a) The programme `hsTemplate.c` should be extended in such a way that it implements a simple Horn-Schunck method. Here "simple" means that the image sequence is not presmoothed by Gaussian convolution, and the optic flow is visualised by its magnitude only.
Please supplement the subroutine `flow` with the missing code and make sure that the image boundaries are treated correctly. You can compile the programma via

```
gcc -O2 -o hsTemplate hsTemplate.c -lm .
```

(b) The *filling-in-effect* that is characteristic for variational methods can be studied with the Danish pig images `pig1.pgm` and `pig2.pgm`. To this end, investigate the results for different number of iterations. What is a good value for the regularisation parameter α in this case ?

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

Assignment P7 (3)

Submission

Please remember that up to three people from the same tutorial group can work and submit their results together. For submitting the files rename the main directory `Ex07` to `Ex07_<your_name>` and use the command

```
tar czvf Ex07_<your_name>.tgz Ex07_<your_name>
```

to pack the data. The directory that you pack and submit should at least contain the following files:

- the source code for computing the Horn-Schunck method in the subroutine `flow`
- flow images for at least three different values of α
- flow images for at least three different iteration numbers (for a fixed value of α)
- a text file README that
  - states the regularisation parameter for the flow results
  - contains information on all people working together for this assignment.

Please make sure that only your final version of the programmes and images are included.
Submit the file via e-mail to your tutor via the address:

```
ipcv-xx@mia.uni-saarland.de
```

where `xx` is either `t1`, `t2`, `t3`, `t4`, `w1` or `w2` depending on your tutorial group.

Deadline for submission: Tuesday, February 12, 10 am (before the lecture)