Outline

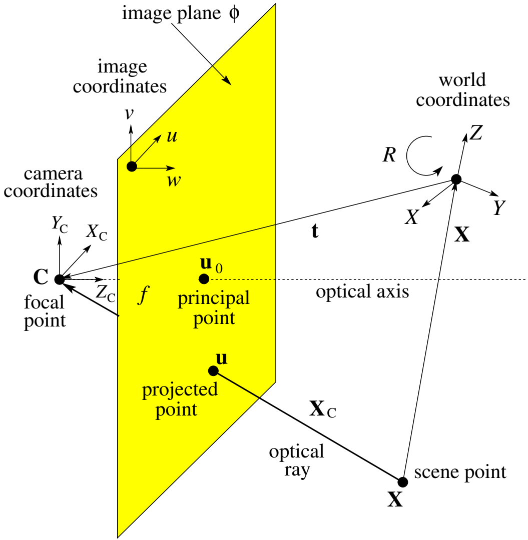
1. Camera Models
   - Perspective (pin-hole) camera
   - Weak-perspective camera
   - Comparison of camera models
   - Back-projection to 3D space
2. Homography
   - Homography estimation
   - Non-linear estimation by minimizing geometric error
3. Camera Calibration
   - Calibration by a spatial object
   - Calibration using a chessboard
   - Radial distortion
4. Summary

Outline

Gemoetric Imaging Models

- We introduce different **geometric** models
  - General perspective camera
  - Simplified camera models
- Perspective camera model equivalent to **pin-hole camera.**
  - camera obscura
- Pin-hole camera is close to real optics
  - → simple model of a thin optics
  - → Physical models are significantly complicated.
- However, a perspective camera is a very good **geometric approximation.**
- We address separately the following issues:
  - radiometric properties (brightness, colors)
  - geometric distortions

Perspective camera model



Notations: coordinates and transformations

- Coordinates

$$\mathbf{X} = [X, Y, Z]^T \qquad \text{world}$$
$$\mathbf{X}_c = [X_c, Y_c, Z_c]^T \qquad \text{camera}$$
$$\mathbf{u} = [u, v]^T \qquad \text{image plane}$$

- Homogeneous coordinates

$$\mathbf{X} = [X, Y, Z, 1]^T \qquad \text{world}$$
$$\mathbf{X}_c = [X_c, Y_c, Z_c, 1]^T \qquad \text{camera}$$
$$\mathbf{u} = [u, v, 1]^T \qquad \text{image plane}$$

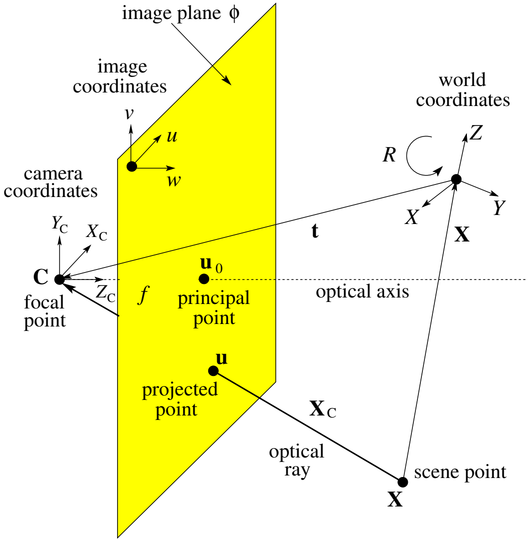
- Transformations
  - **R**: rotation (matrix)
  - **t**: translation (vector)

Notations: camera

$\mathbf{C}$	ϕ	f	$\mathbf{u}_0 = [u_0, v_0]^T$
focal point	*image plane*	*focal length*	*principal point*

- $\mathbf{C}$ focal point: central projection
- Optical ray: it connects a 3D point and focal point $\mathbf{C}$
- Optical axis: Contains the focal point $\mathbf{C}$ and perpendicular to image plane ϕ
- Focal length: distance between $\mathbf{C}$ and ϕ.
- Principal point: the point in image plane where optical axis intersects ϕ

Perspective camera model



Translation and rotation

- World $\longrightarrow$ Camera
- Euclidean coordinates

$$\mathbf{X}_c = R(\mathbf{X} - \mathbf{t}) \tag{1}$$

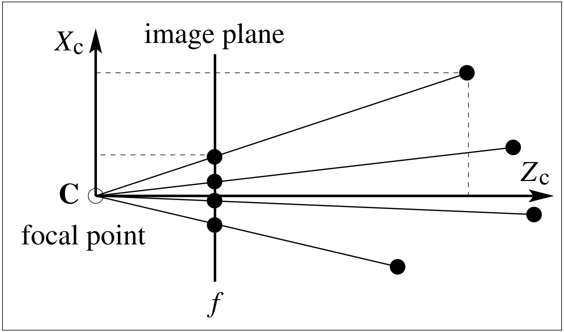
- Homogeneous coordinates

$$\mathbf{X}_c = R\,[\mathrm{I}| - \mathbf{t}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \tag{2}$$

  - I is a 3×3- **identity matrix**
  - $[\mathrm{I}| - \mathbf{t}]$ is a 3×4 -matrix

  $\rightarrow$ I completed by colums $-\mathbf{t}$

Projection to an image plane



$$u = \frac{fk_u}{Z_c} X_c + u_0 \quad (3)$$

$$v = \frac{fk_v}{Z_c} Y_c + v_0 \quad (4)$$

- k_u, k_v is the horizontal/vertical pixel size.

→ their unit is *pixel/length*.

- Usually, $k_u = k_v = k$.

Projection using homogeneous coordinates

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \mathbf{K}\mathbf{X}_c \tag{5}$$

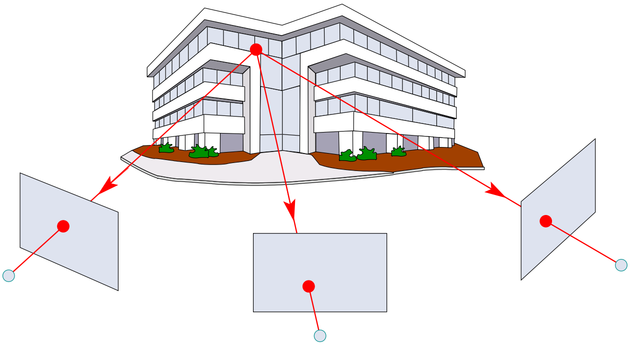
- $\sim$ homogeneous division yields scale ambiguity
- K is the (intrinsic) **calibration matrix**

$$K = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

  - upper triangular matrix
  - consists of 5 parameters, but only four are realistic

$\rightarrow$ fk_u, fk_v, u_0, v_0

Multi-view projection of a spatial point



- Locations of the same spatial point differ in images.
- Locations should be detected and/or tracked in the images.
  - $\rightarrow$ They are called *correspondences*.

Perspective camera model

- Goal: to determine the location of the projected 3D points in camera images.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \mathbf{K}\mathbf{R}\left[\mathrm{I} \,|\, -\mathbf{t}\right] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = P \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} \tag{7}$$

- $P \doteq \mathbf{K}\mathbf{R}\left[\mathrm{I} \,|\, -\mathbf{t}\right]$ is the **projection matrix**
  - consists of 11 parameters

  $\rightarrow$ 5 in K, 3 in R, another 3 in $\mathbf{t}$.

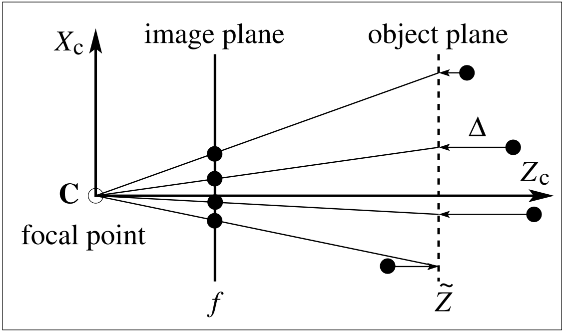
Outline

Weak-perspective projection 1/2

- It is assumed that the object is not 'too close' from the camera
  - change in depth is significantly smaller than the camera-object distance
- Object plane is parallel to the image plane
  - it is ideal if object center contains the center of gravity of the object.
- Objects are orthogonally projected into the object plane
- Then perspective projection is applied
  - as there is no difference in depth, location of principal point does not matter.

  $\rightarrow$ for the sake of simplicity, $u_0 = v_0 = 0$.

Weak-perspective projection 2/2



$$u = \frac{fk}{\tilde{Z}_c} X_c + u_0 \qquad (8)$$

$$v = \frac{fk}{\tilde{Z}_c} Y_c + v_0 \qquad (9)$$

- If pixel is a square, $k_u = k_v = k$
- It is also assumed that $Z_c \gg \Delta$

$\rightarrow$ $Z_c \approx \tilde{Z}_c$, where $\tilde{Z}_c$ is the common depth

$\rightarrow$ scaled orthographic projection

Weak-perspective camera model 1/2

- Translation and rotation in conjunction with weak-perspective projection:

$$u = q\mathbf{r}_1^T(\mathbf{X} - \mathbf{t}) + u_0 \tag{10}$$

$$v = q\mathbf{r}_2^T(\mathbf{X} - \mathbf{t}) + v_0, \quad \text{where} \tag{11}$$

$$q \doteq \frac{fk}{\widetilde{Z}_c}$$

- $\mathbf{r}_1^T$ and $\mathbf{r}_2^T$ are the first and second row vectors of rotation matrix R.
- $\mathbf{u}_0$ represents offset: $\longrightarrow u_0 = v_0 = 0$

$$u = q\mathbf{r}_1^T(\mathbf{X} - \mathbf{t}) \tag{12}$$

$$v = q\mathbf{r}_2^T(\mathbf{X} - \mathbf{t}) \tag{13}$$

Weak-perspective camera model 2/2

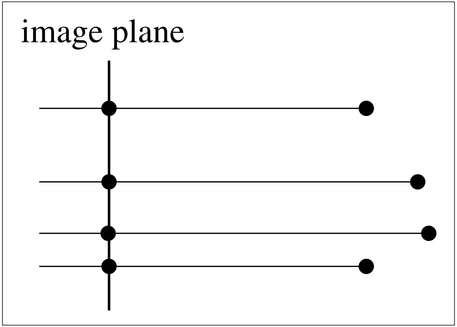
- Projection can be written with the help of a weak-perspective camera matrix:

$$\begin{bmatrix} u \\ v \end{bmatrix} = [M|\mathbf{b}] \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix}, \quad \text{where} \tag{14}$$

$$M \doteq q \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \end{bmatrix}, \qquad \mathbf{b} \doteq - \begin{bmatrix} q\mathbf{r}_1^T\mathbf{t} \\ q\mathbf{r}_2^T\mathbf{t} \end{bmatrix}$$

- Model has 6 degree of freedom (DoF)
  - if $k_u \neq k_v$, DOF=7
- There is no scale ambiguity.

orthographic projection



- Orthogonal projection can be applied if object
  - is far from the camera
  - depth is relatively static
- Model has 5 degree of freedom (DoF)
  - R, t_1, t_2

Outline

Affine camera

- **General affine camera**

$$\mathbf{u} = M_{2\times 3}\mathbf{X} + \mathbf{t}$$

  - 8 degrees of freedom
  - $M_{2\times 3}$ is a 2×3matrix with rank two

- Hierarchy of affine cameras
  - general affine camera

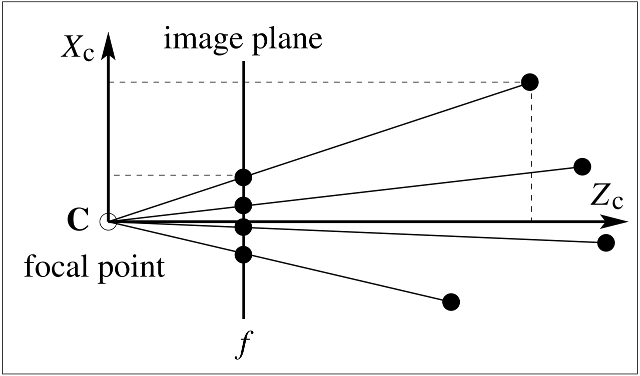
$$\Downarrow$$

  - more constraints,
  - less DoFs

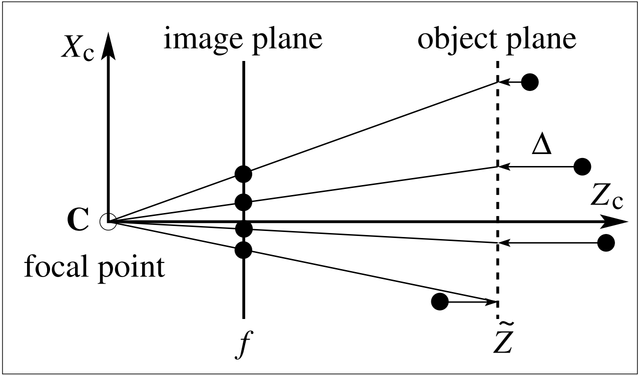
Herarchy of affine camera models

- **Weak-perspective projection**
  - 7 degrees of freedom ($k_u \neq k_v$)
- **Scaled orthographic projection**
  - six degrees of freedom
  - orthogonal projection $+$ isotropic scale
  - $\rightarrow$ if $k_u = k_v$, it is a scaled orthographic projection
- **Orthogonal projection**
  - five degrees of freedom

Perspective projection



Weak-perspective projection



Applicability of weak-perspective projection

- Projection error of weak-perspective projection

$$\mathbf{x}_c^{\text{weak}} - \mathbf{x}_c^{\text{proj}} = \frac{\Delta}{\widetilde{Z}_c}\mathbf{x}_c^{\text{proj}}$$

$\rightarrow$ with respect to real location

- Δ: distance betwen point and object plane
- $\widetilde{Z}_c$ mean of depth values

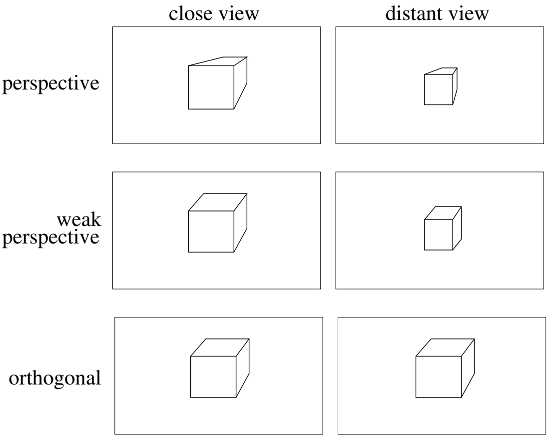
$\rightarrow$ Weak-perspective projection applicable if

- $\Delta \ll \widetilde{Z}_c$

Benefits & disadvantages of weak-persective projection

- Advantages over real perspective projection
  - no scale ambiguity
  - less parameters to be estimated
  - $\rightarrow$ accuracy of estimation can be better
  - simpler
  - $\rightarrow$ closed-form solutions exist for several problems
- Disadvantages
  - It is only an approximation of real projection
  - $\rightarrow$ less accurate if conditions do not hold

Comparison of the projection models



effect	perspective	weak-persp.	orthogonal
change in sizes	yes	yes	no
persp. distortion	yes	no	no

Outline

Back-projection of a point 1/2

- Projection

$$u = \frac{fk_u}{Z_c} X_c + u_0$$
$$v = \frac{fk_v}{Z_c} Y_c + v_0$$

- Back projection by expressing spatial coordinates:

$$X_c = \frac{Z_c}{fk_u}(u - u_0)$$
$$Y_c = \frac{Z_c}{fk_v}(v - v_0)$$
$$Z_c = Z_c$$

Back-projection of a point 2/2

- Matrix form

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = Z_c \begin{bmatrix} \frac{1}{fk_u} & 0 & -\frac{u_0}{fk_u} \\ 0 & \frac{1}{fk_v} & -\frac{v_0}{fk_v} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = Z_c K^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{15}$$

  - where the calibration matrix is as follows:

$$K = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

- a 3D point is ambiguous w.r.t. depth
  $\rightarrow$ a point in image represents a line in 3D space

Back projection by homogeneous coordinates

- Projection of 3D point to an image plane:

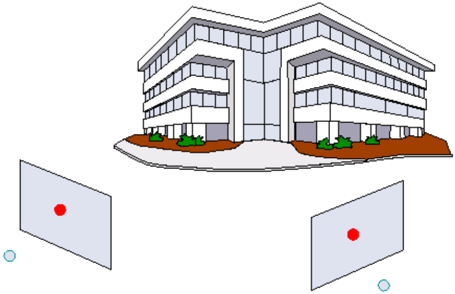
$$\mathbf{u} = P\mathbf{X}$$

- Back-projection yields a line:

$$\mathbf{X}(\lambda) = (1-\lambda)P^{+}\mathbf{u} + \lambda\mathbf{C}$$

  - It is a line written by a parameter λ
- P^{+} is the pseudo-inverse of P
  - $PP^{+} = \mathrm{I}$ (I : identity matrix)
  - $P^{+} = P^{\mathrm{T}}\left(PP^{\mathrm{T}}\right)^{-1}$
- The line contains
  - point $P^{+}\mathbf{u}$ ($\lambda = 0$)
  - Focal point **C** of camera
    - **C** is the null-vector of matrix P

Back projection and triangulation



- To estimate a 3D points
  - at least two calibrated cameras
  - and two corresponding points in the images are required.
- Estimation of 3D coordinates is called **triangulation** in computer vision.