Camera geometry and Stereo

C306
Fall 2001
Martin Jagersand

The camera matrix

- Homogenous coordinates for 3D
  - four coordinates for 3D point
  - equivalence relation (X,Y,Z,T) is the same as (k X, k Y, k Z,k T)
- Turn previous expression into HC's
  - HC's for 3D point are (X,Y,Z,T)
  - HC's for point in image are (U,V,W)

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} \qquad (U,V,W) \rightarrow \left(\frac{U}{W}, \frac{V}{W}\right) = (u,v)$$

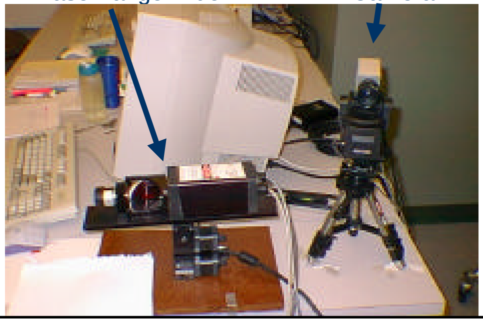
Camera parameters

- Issue
  - camera may not be at the origin, looking down the z-axis
    - extrinsic parameters
  - one unit in camera coordinates may not be the same as one unit in world coordinates
    - intrinsic parameters - focal length, principal point, aspect ratio, angle between axes, etc.

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} \text{Transformation} \\ \text{representing} \\ \text{intrinsic parameters} \end{pmatrix} \begin{pmatrix} \text{Transformation} \\ \text{representing} \\ \text{projection model} \end{pmatrix} \begin{pmatrix} \text{Transformation} \\ \text{representing} \\ \text{extrinsic parameters} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

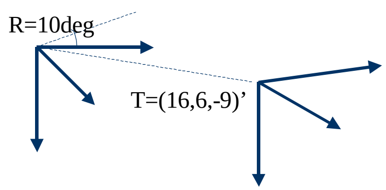
Review: Real camera

- Laser range finder
- Camera



Relative location Camera-Laser

- Camera
- Laser



In homogeneous coordinates

- Rotation:

$$\mathbf{R} = \begin{bmatrix} \cos -10 & 0 & \sin -10 \\ 0 & 1 & 0 \\ -\sin -10 & 0 & \cos -10 \end{bmatrix}$$

- Translation

$$T = \begin{pmatrix} 1 & 0 & 0 & 16 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Full perspective projection model

- Camera internal parameters
- Camera projection

$$p_{camera} = \begin{bmatrix} 1278.6657 & 0 & 256 \\ 0 & 1659.5688 & 240 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

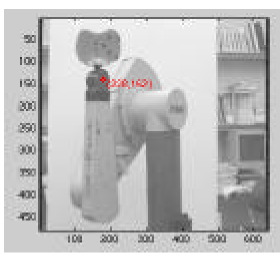
$$\begin{pmatrix} 0.985 & 0 & -0.174 & 0 \\ 0 & 1 & 0 & 0 \\ 0.174 & 0 & 0.985 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 16 \\ 0 & 1 & 0 & 6 \\ 0 & 0 & 1 & -9 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.6612 \\ -10.55 \\ 108.0 \\ 1 \end{pmatrix} = \begin{pmatrix} 22262 \\ 16755 \\ 97.47 \end{pmatrix}$$

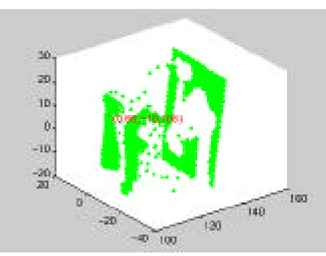
3D world point P

2D image point p

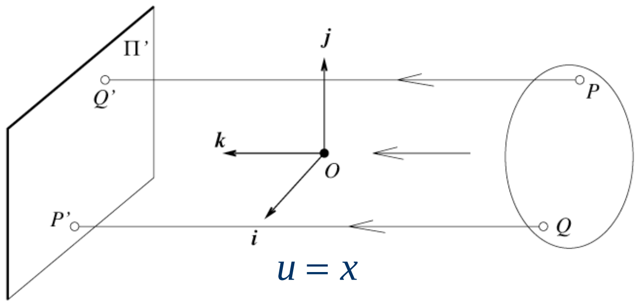
Result

- Camera image
- Laser measured 3D structure





Orthographic projection



$$u = x$$
$$v = y$$

The fundamental model for orthographic projection

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Perspective and Orthographic Projection

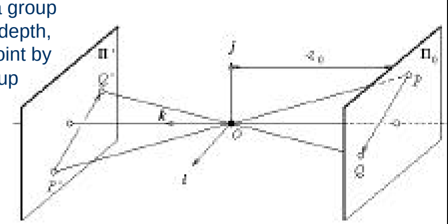
perspective

Orthographic (parallel)

Weak perspective

$$u = Tx$$
$$v = Ty$$
$$T = f / Z$$

- Issue
  - perspective effects, but not over the scale of individual objects
  - collect points into a group at about the same depth, then divide each point by the depth of its group
  - Adv: easy
  - Disadv: wrong



The fundamental model for weak perspective projection

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f/Z^* \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

Camera Model Structure

Assume R and T express camera *in world coordinates, then*

$$^{c}p = \begin{pmatrix} & R' & & -R'T \\ 0 & 0 & 0 & 1 \end{pmatrix} {}^{w}p$$

Combining with a weak perspective model (and neglecting internal parameters) yields

$$^{c}u = M\,{}^{w}p = \begin{pmatrix} -R'_x & R'_x T \\ -R'_y & R'_y T \\ 0 & \dfrac{R_z(\bar{P} - T)}{f} \end{pmatrix} {}^{w}p$$

Where $\bar{P}$ is the nominal distance to the viewed object

Other Models

- The *affine camera* is a generalization of weak perspective.
- The *projective camera* is a generalization of the perspective camera.
- Both have the advantage of being linear models on real and projective spaces, respectively.

Recall: Intrinsic Parameters

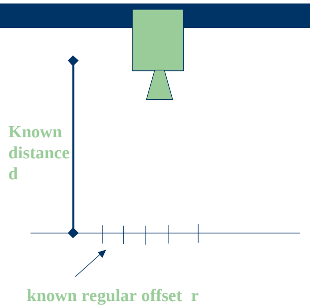
Intrinsic Parameters describe the conversion from metric to pixel coordinates (and the reverse)

$$x_{mm} = -(x_{pix} - o_x)\, s_x$$
$$y_{mm} = -(y_{pix} - o_y)\, s_y$$

or

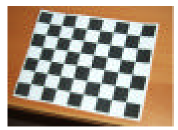
$$\begin{pmatrix} x \\ y \\ w \end{pmatrix}_{pix} = \begin{pmatrix} -1/s_x & 0 & o_x \\ 0 & -1/s_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix}_{mm} = M_{\text{int}}\, p$$

CAMERA CALIBRATION: A WARMUP



$$\frac{rk_i}{d} = (x_i - o_x)s_x$$

$$\frac{r}{d} = (x_{i+1} - x_i)s_x$$



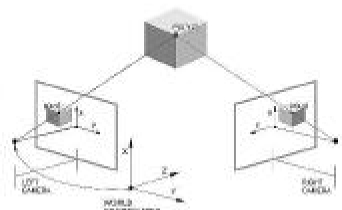
A simple way to get scale parameters; we can compute the optical center as the numerical center and therefore have the intrinsic parameters

Camera calibration

- Issues:
  - what are intrinsic parameters of the camera?
  - what is the camera matrix? (intrinsic+extrinsic)
- General strategy:
  - view calibration object
  - identify image points
  - obtain camera matrix by minimizing error
  - obtain intrinsic parameters from camera matrix
- Error minimization:
  - Linear least squares
    - easy problem numerically
    - solution can be rather bad
  - Minimize image distance
    - more difficult numerical problem
    - solution usually rather good, but can be hard to find
      - start with linear least squares
  - Numerical scaling is an issue

Stereo Vision

- **GOAL:** Passive 2-camera system for triangulating 3D position of points in space to generate a depth map of a world scene.
- Humans use stereo vision to obtain depth



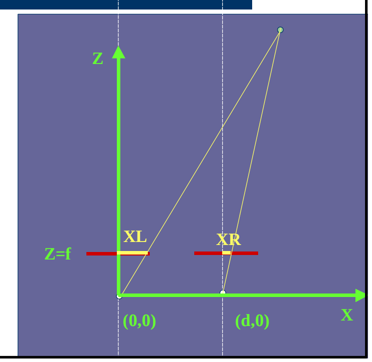
Stereo depth calculation: Simple case, aligned cameras

DISPARITY=
(XL - XR)

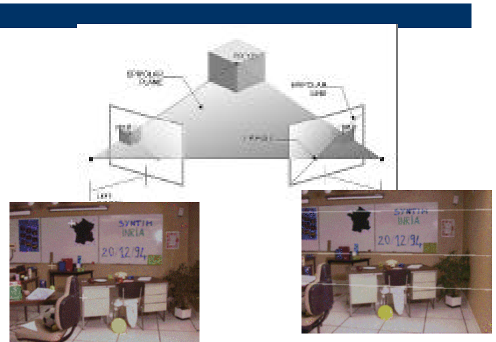
Z = (f/XL) X
Z= (f/XR) (X-d)

(f/XL) X = (f/XR) (X-d)
X = (XL d) / (XL - XR)

$$Z = \frac{df}{(XL - XR)}$$

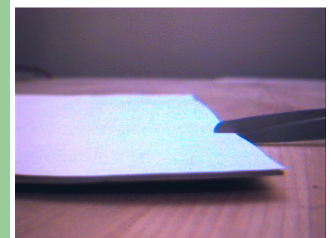


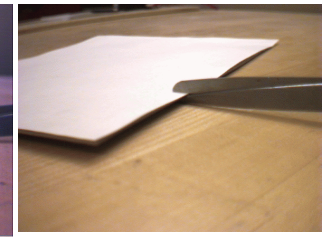
Epipolar constraint



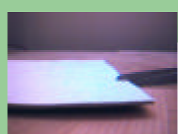
Special case: **parallel cameras** – epipolar lines are parallel and aligned with rows

Visual ambiguity





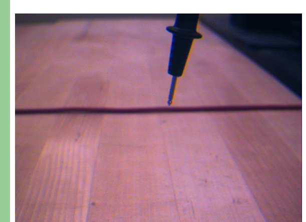
- Will the scissors cut the paper in the middle?

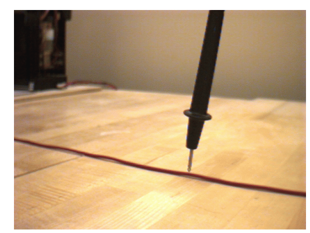


Ambiguity

- Will the scissors cut the paper in the middle? **NO!**

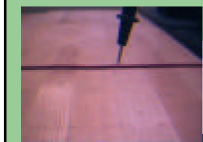
Visual ambiguity

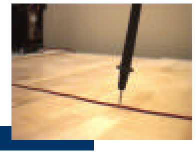


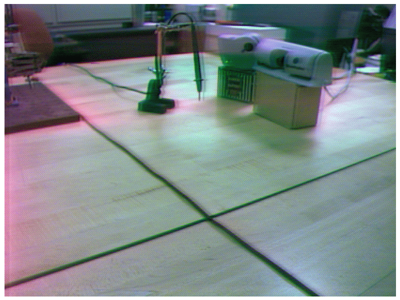


- Is the probe contacting the wire?

Ambiguity

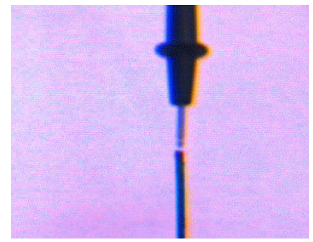


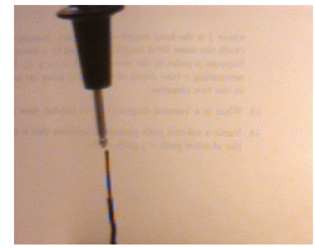




- Is the probe contacting the wire? **NO!**

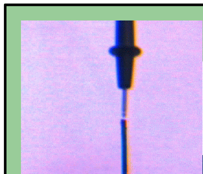
Visual ambiguity

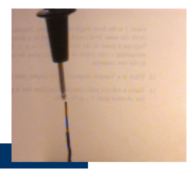


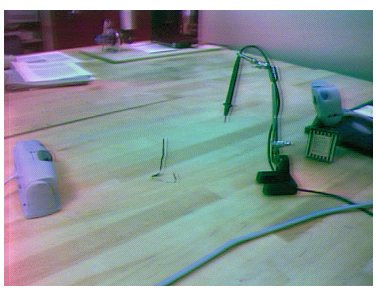


- Is the probe contacting the wire?

Ambiguity







- Is the probe contacting the wire? **NO!**

Camera Models

- Internal calibration:
- Weak calibration:
- Affine calibration:
- Stratification of stereo vision:
  - characterizes the reconstructive certainty of weakly, affinely, and internally calibrated stereo rigs

C_{sim} up to a similarity (scaled Euclidean transformation)

C_{aff} up to an affine transformation of task space

C_{proj} up to a projective transformation of task space

C_{inj} reconstruction up to a bijection of task space

Visual Invariance