Lesson 9: 3D Vision

1. Introduction
2. Camera Model
3. Calibration
4. Stereo Vision
5. Correspondence Search
6. Other 3D techniques

1. Introduction

- Cameras get a 2D projection of a 3D scene
- 3D reconstruction needs two or more images
  - 2 cameras: binocular stereo
  - 3 cameras: trinocular stereo
  - 1 single moving camera (SFM: structure from motion)
- Needs a model of the camera geometry and optics
- Main difficulty: search for correspondences
  - Given en element in one image (a point, edge, region,…), find it on the other images

2. Camera Model (Tsai)

- Projection of a point on an image:

$$\mathbf{u} = \text{Proj}\,(\mathbf{x}_w, \theta_c)$$

$$\mathbf{x}_w = \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix}$$ 3D point coordinates

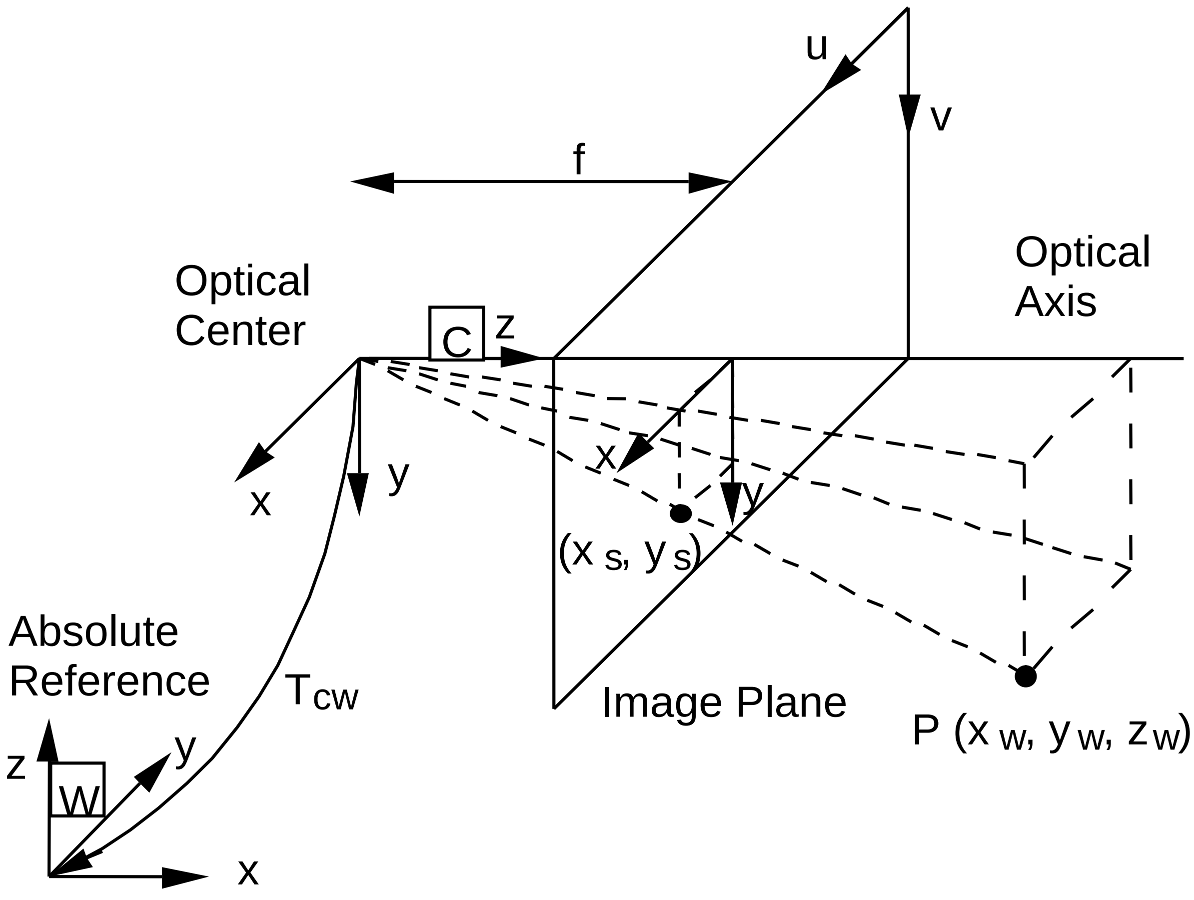
$$\mathbf{u} = \begin{pmatrix} u \\ v \end{pmatrix}$$ Coordinates on the image

$$\theta_c = \begin{pmatrix} \theta_{int} \\ \theta_{ext} \end{pmatrix}$$ Camera parameters

- Intrinsic camera parameters
  - Camera geometry and optics
  - Image acquisition system (for analog cameras)
- Extrinsic camera parameters
  - Camera 3D position and orientation
- Calibration: experimental determination of camera parameters

Tsai's Camera Model (1)

1) Transformation to camera reference



Coordinates relative to the Camera:

$$\mathbf{x}_c = \mathbf{R}_{cw}\mathbf{x}_w + \mathbf{t}_{cw}$$

$$\mathbf{x}_c = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$$

Position and Orientation of W relative to C:

$$\mathbf{t}_{cw} = \begin{pmatrix} x_{cw} \\ y_{cw} \\ z_{cw} \end{pmatrix}$$

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

Homogeneous Transformations

Position and Orientation of W relative to C

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{cw} & \mathbf{t}_{cw} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} = \mathbf{T}_{cw} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

Position and Orientation of C relative to W

$$\mathbf{T}_{wc} = \mathbf{T}_{cw}^{-1} = \begin{pmatrix} \mathbf{R}_{cw}^T & -\mathbf{R}_{cw}^T \mathbf{t}_{cw} \\ 0 & 1 \end{pmatrix}$$

- Orientation defined by 3 angles:

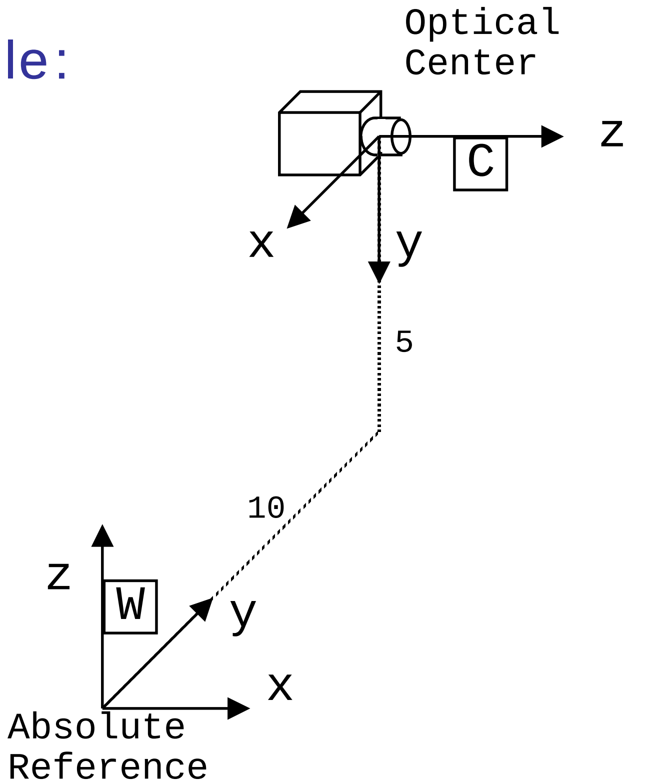
Rotation Order →

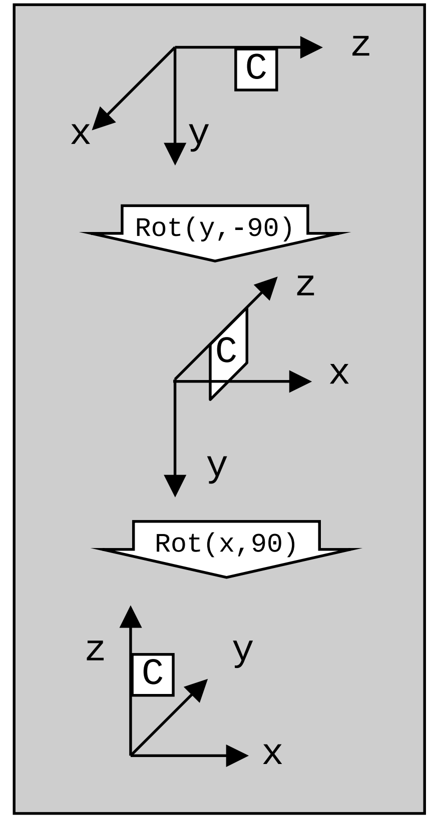
$$RPY(\phi, \theta, \psi) = Rot(z, \phi)\, Rot(y, \theta)\, Rot(x, \psi) =$$

$$RPY(\phi, \theta, \psi) = \begin{pmatrix} C\phi C\theta & C\phi S\theta S\psi - S\phi C\psi & C\phi S\theta C\psi + S\phi S\psi \\ S\phi C\theta & S\phi S\theta S\psi + C\phi C\psi & S\phi S\theta C\psi - C\phi S\psi \\ -S\theta & C\theta S\psi & C\theta C\psi \end{pmatrix}$$

Homogeneous Transformations

- Example:





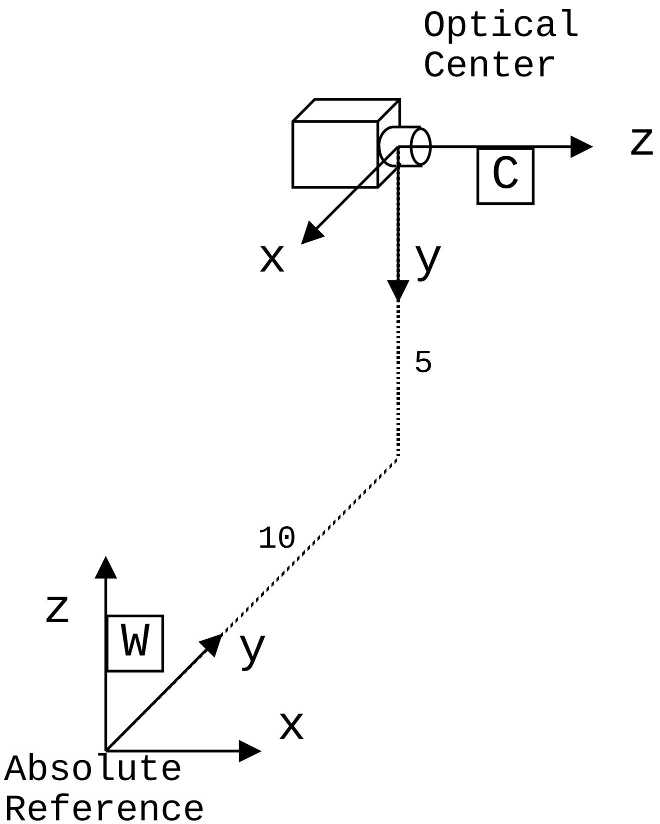
$$\mathbf{R}_{cw} = Rot(z,0)Rot(y,-90)Rot(x,90) = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}$$

x y z

Projection of W´s axis on reference C

Homogeneous Transformations

- Example:



$$\mathbf{T}_{cw} = \begin{pmatrix} 0 & -1 & 0 & 10 \\ 0 & 0 & -1 & 5 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

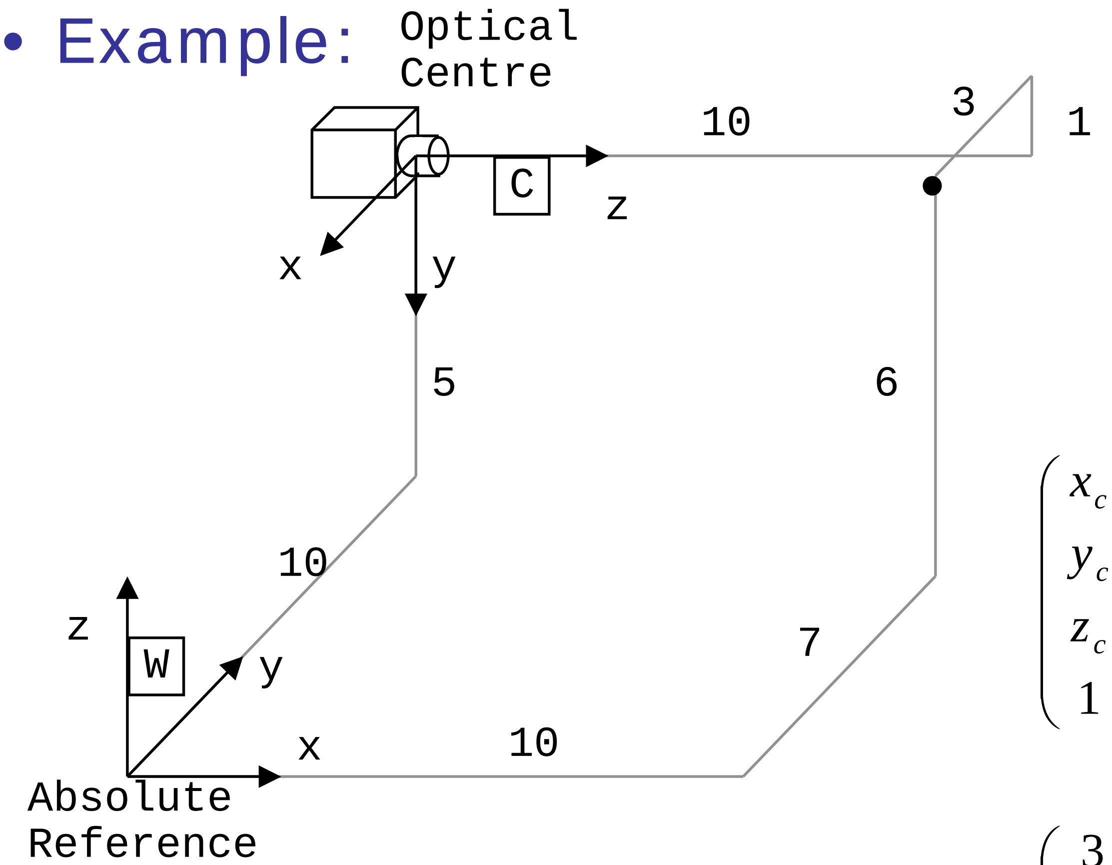
$$\mathbf{T}_{wc} = \mathbf{T}_{cw}^{-1} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 10 \\ 0 & -1 & 0 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

x y z

Projection of C´s axis on reference W

Homogeneous Transformations

- Example:

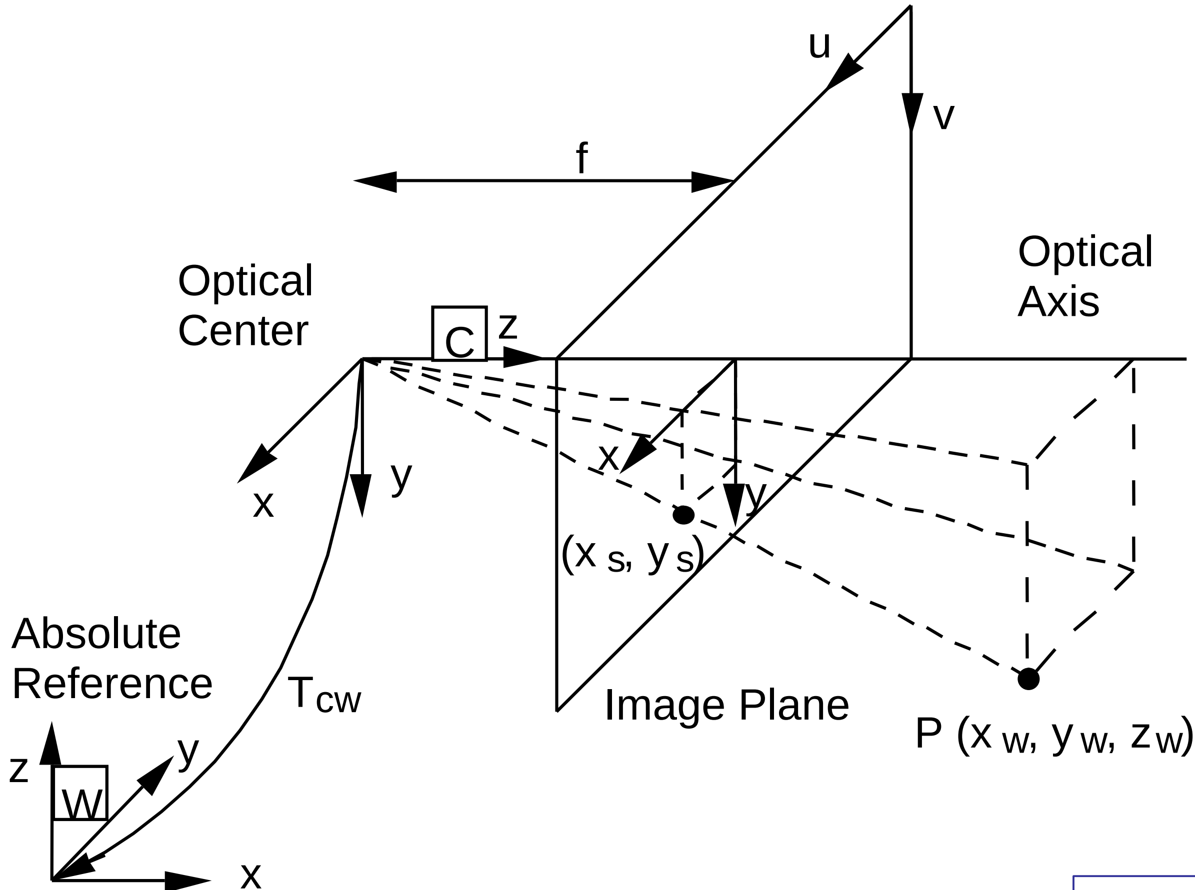


$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{cw} & \mathbf{t}_{cw} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} = \mathbf{T}_{cw} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 3 \\ -1 \\ 10 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 & 10 \\ 0 & 0 & -1 & 5 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 10 \\ 7 \\ 6 \\ 1 \end{pmatrix}$$

Tsai's Camera Model (2)

2) Theoretical projection on image sensor



$$\mathbf{x}_{st} = \begin{pmatrix} x_{st} \\ y_{st} \end{pmatrix}$$

$$x_{st} = f \frac{x_c}{z_c}$$

$$y_{st} = f \frac{y_c}{z_c}$$

Objective's focal length (mm)

Tsai's Camera Model (3)

3) Radial distortion

– Noticeable for wide lenses: f < 12 mm

$$\mathbf{x}_{sd} = \begin{pmatrix} x_{sd} \\ y_{sd} \end{pmatrix}$$

Distorted coordinates

Theoretical undistorted coordinates

Radial distortion coefficient

$$x_{st} = x_{sd}(1 + k_1 r^2)$$

$$y_{st} = y_{sd}(1 + k_1 r^2)$$

$$r^2 = {x_{sd}}^2 + {y_{sd}}^2$$

Tsai's Camera Model (4)

4) Transform to pixel coordinates

$$\mathbf{u} = \begin{pmatrix} u \\ v \end{pmatrix}$$

$$u = C_x + \frac{x_{sd}}{d_x} s_x$$

$$v = C_y + \frac{y_{sd}}{d_y}$$

Coordinates where the optical axis intersects the image plane (in pixels)

Horizontal correction factor (for analog cameras)

Pixel size on the sensor (mm)

- Example for a CCD analogic camera:

Acquiring all the lines

$$d_y = d_{cy}$$

or

Acquiring half the lines

$$d_y = 2d_{cy}$$

$$d_x = d_{cx} \frac{N_{sx}}{N_{tx}}$$

Pixels per line on the sensor

Pixels per line on memory

Tsai's Camera Model (4)

- Example: camera Pulnix TM-6 (analogic)
  - Pixels: 752(H) x 582 (V)
  - Cell size: 8.4μm x 8.2 μm

a) Sampling 512(H) x 512(V)

$$d_x = 8.4 \times 10^{-3} \frac{752}{512} = 12.34 \times 10^{-3} mm$$

$$d_y = 8.2 \times 10^{-3} mm$$

$$PAR = \frac{d_x}{d_y} = 1.5$$

PAR: Pixel Aspect Ratio

b) Sampling 768(H) x 512(V)

$$d_x = 8.4 \times 10^{-3} \frac{752}{768} = 8.2 \times 10^{-3} mm$$

$$d_y = 8.2 \times 10^{-3} mm$$

square pixel
PAR = 1

Tsai's Camera Model (summary)

- Projection model: $\mathbf{u} = \mathrm{Proy}(\mathbf{x}_w, \theta_c)$

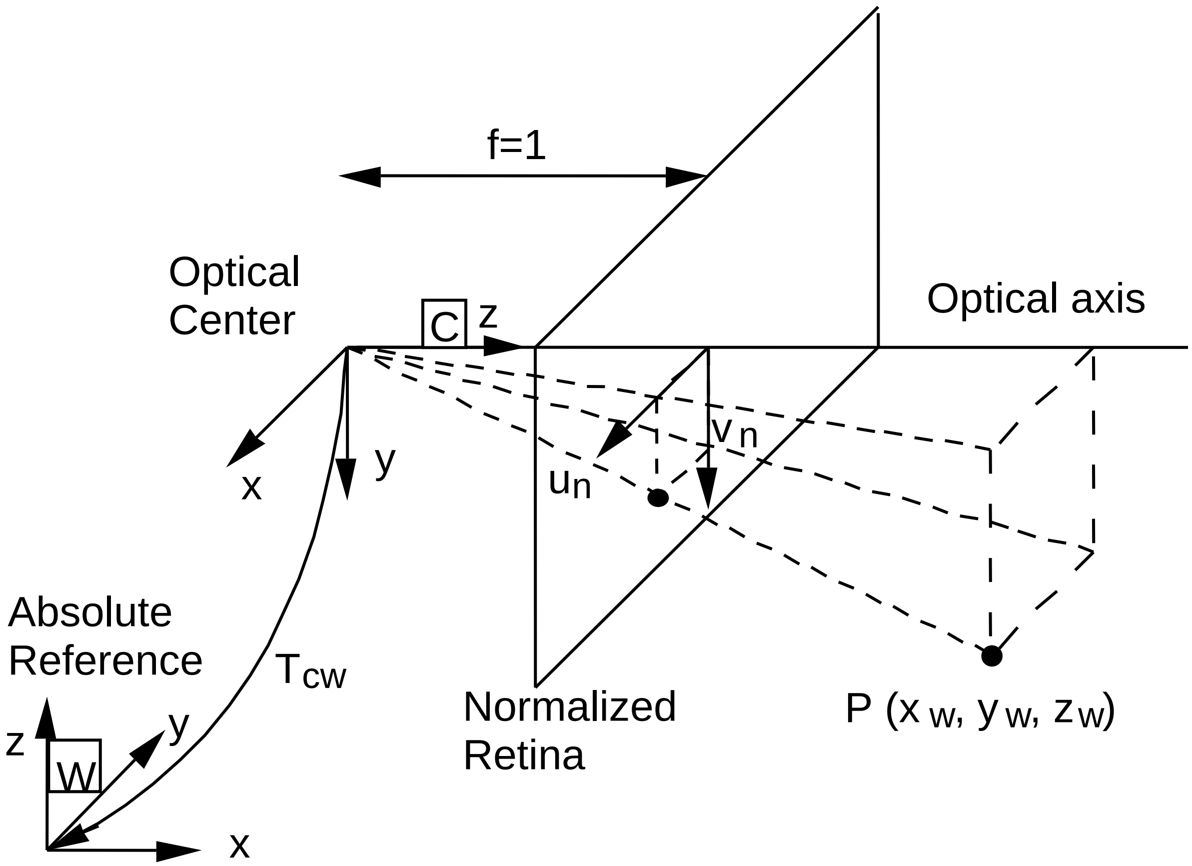
1. Transform to camera: $\mathbf{x}_c = \mathbf{R}_{cw}\mathbf{x}_w + \mathbf{t}_{cw}$
2. Theoretical projection: $x_{st} = f\dfrac{x_c}{z_c}$ $\quad y_{st} = f\dfrac{y_c}{z_c}$
3. Radial distortion: $x_{st} = x_{sd}(1 + k_1 r^2)$ $\quad y_{st} = y_{sd}(1 + k_1 r^2)$

   $r^2 = x_{sd}{}^2 + y_{sd}{}^2$
4. Transform to pixels: $u = C_x + \dfrac{x_{sd}}{d_x} s_x$ $\quad v = C_y + \dfrac{y_{sd}}{d_y}$

- Known parameters: d_x, d_y
- Parameter to Calibrate:
  - » Extrinsic: t_{cw}, R_{cw}
  - » Intrinsic: f, k_1, C_x, C_y, s_x

Normalized Retina

- Is an ideal retina (without distortion) placed at focal distance f = 1



The projection only depends on the extrinsic camera parameters:

$$\mathbf{u}_n = \mathrm{Norm\_proj}(\mathbf{x}_w, \theta_{ext})$$

$$\mathbf{x}_c = \mathbf{R}_{cw}\mathbf{x}_w + \mathbf{t}_{cw}$$

$$u_n = \frac{x_c}{z_c}$$

$$v_n = \frac{y_c}{z_c}$$

$$\mathbf{u}_n = \begin{pmatrix} u_n \\ v_n \end{pmatrix}$$

Normalized coordinates (mm)

Normalized Retina

- Each pixel can be transformed from the real retina to the normalized retina:

$$\mathbf{u}_n = \mathrm{Normal}(\mathbf{u}, \theta_{int})$$

$$x_{sd} = (u - C_x)\frac{d_x}{s_x} \qquad y_{sd} = (v - C_y)d_y$$

$$r^2 = x_{sd}^{\ 2} + y_{sd}^{\ 2}$$

$$x_{st} = x_{sd}(1 + k_1 r^2) \qquad y_{st} = y_{sd}(1 + k_1 r^2)$$

$$u_n = \frac{x_{st}}{f} \qquad v_n = \frac{y_{st}}{f}$$

3. Calibration

- Cameramodel: $\mathbf{u} = \mathrm{Proj}(\mathbf{x}_w, \theta_c)$

- Put a pattern with N points, in known positions:

$$(x_{w1}, \ldots x_{wN})$$

- Take an image and compute the point coordinates in the image: $(u_1, \ldots u_N)$

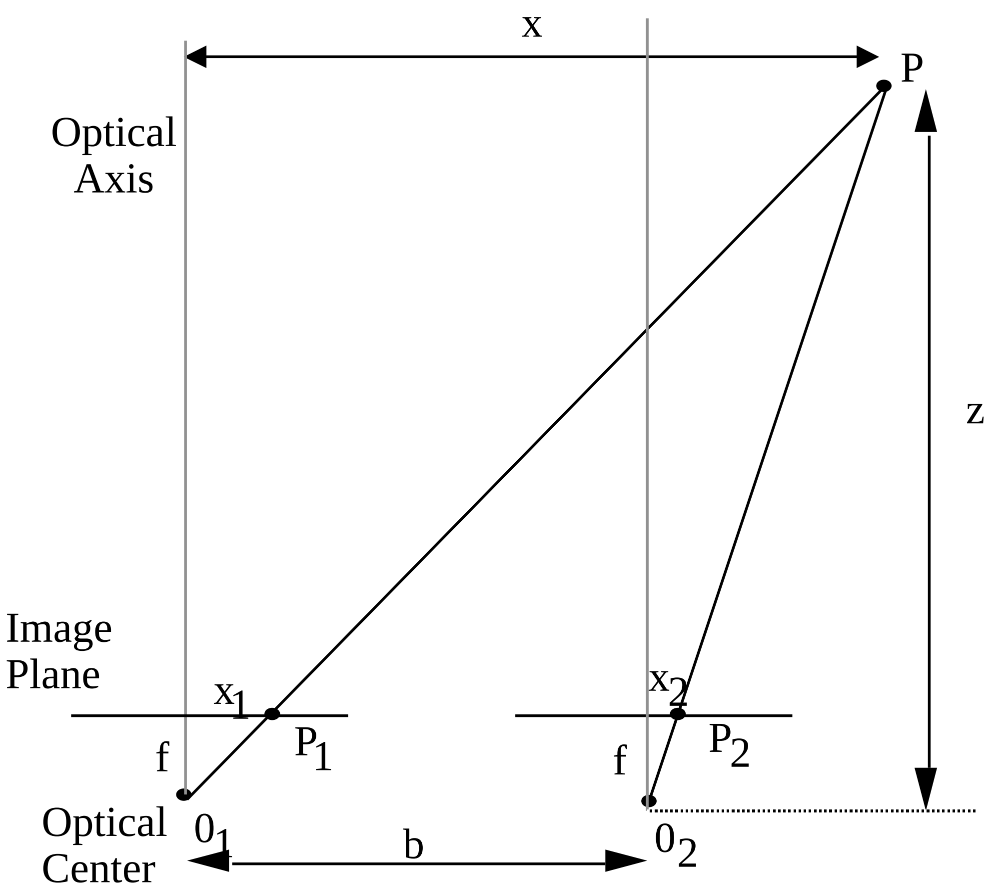
- Obtain the set of parameters θ_c that minimizes the squared error:

$$\sum_{i=1}^{N} \left[\mathbf{u}_i - \mathrm{Proj}(\mathbf{x}_{wi}, \theta_c)\right]$$

In order to compute all the parameters, the pattern must be 3D

4. Stereo Vision

- Simplest case: parallel cameras with no distortion



$$x_1 = f\,\frac{x_{c1}}{z_{c1}} = f\,\frac{x}{z}$$

$$x_2 = f\,\frac{x_{c2}}{z_{c2}} = f\,\frac{x-b}{z}$$

$$z = \frac{b\,f}{x_1 - x_2}$$

- Depth depends on the disparity: $x_1 - x_2$

Stereo Vision

- General case: real cameras in arbitrary positions
  - Using the normalized retinas:

$$\mathbf{x}_{c_k} = R_{c_k w}\mathbf{x}_w + \mathbf{t}_{c_k w} = \begin{pmatrix} r_{11k} & r_{12k} & r_{13k} \\ r_{21k} & r_{22k} & r_{23k} \\ r_{31k} & r_{32k} & r_{33k} \end{pmatrix}\mathbf{x}_w + \begin{pmatrix} x_{c_k w} \\ y_{c_k w} \\ z_{c_k w} \end{pmatrix}$$

$$u_{n_k} = \frac{x_{c_k}}{z_{c_k}} \qquad v_{n_k} = \frac{y_{c_k}}{z_{c_k}}$$

$$\mathbf{r}_{1k}\mathbf{x}_w + x_{c_k w} = u_{n_k}\left(\mathbf{r}_{3k}\mathbf{x}_w + z_{c_k w}\right)$$

$$\mathbf{r}_{2k}\mathbf{x}_w + y_{c_k w} = v_{n_k}\left(\mathbf{r}_{3k}\mathbf{x}_w + z_{c_k w}\right)$$

$$\begin{pmatrix} x_{c_k} \\ y_{c_k} \\ z_{c_k} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{1k}\mathbf{x}_w + x_{c_k w} \\ \mathbf{r}_{2k}\mathbf{x}_w + y_{c_k w} \\ \mathbf{r}_{3k}\mathbf{x}_w + z_{c_k w} \end{pmatrix}$$

where:

$$\mathbf{r}_{ik} = \left(r_{i1k} \quad r_{i2k} \quad r_{i3k}\right)$$

$$\mathbf{A}_k\mathbf{x}_w + \mathbf{b}_k = 0$$

$$\mathbf{A}_k = \begin{pmatrix} \mathbf{r}_{1k} - u_{n_k}\mathbf{r}_{3k} \\ \mathbf{r}_{2k} - v_{n_k}\mathbf{r}_{3k} \end{pmatrix}$$

$$\mathbf{b}_k = \begin{pmatrix} x_{c_k w} - u_{n_k} z_{c_k w} \\ y_{c_k w} - v_{n_k} z_{c_k w} \end{pmatrix}$$

Stereo Vision

- 3D point reconstruction from n images
  - Three unknowns: x_w, y_w, z_w
  - Two equations in each image k:

$$\mathbf{A}_k \mathbf{x}_w + \mathbf{b}_k = 0$$

  - System of linear equations (over-determined):

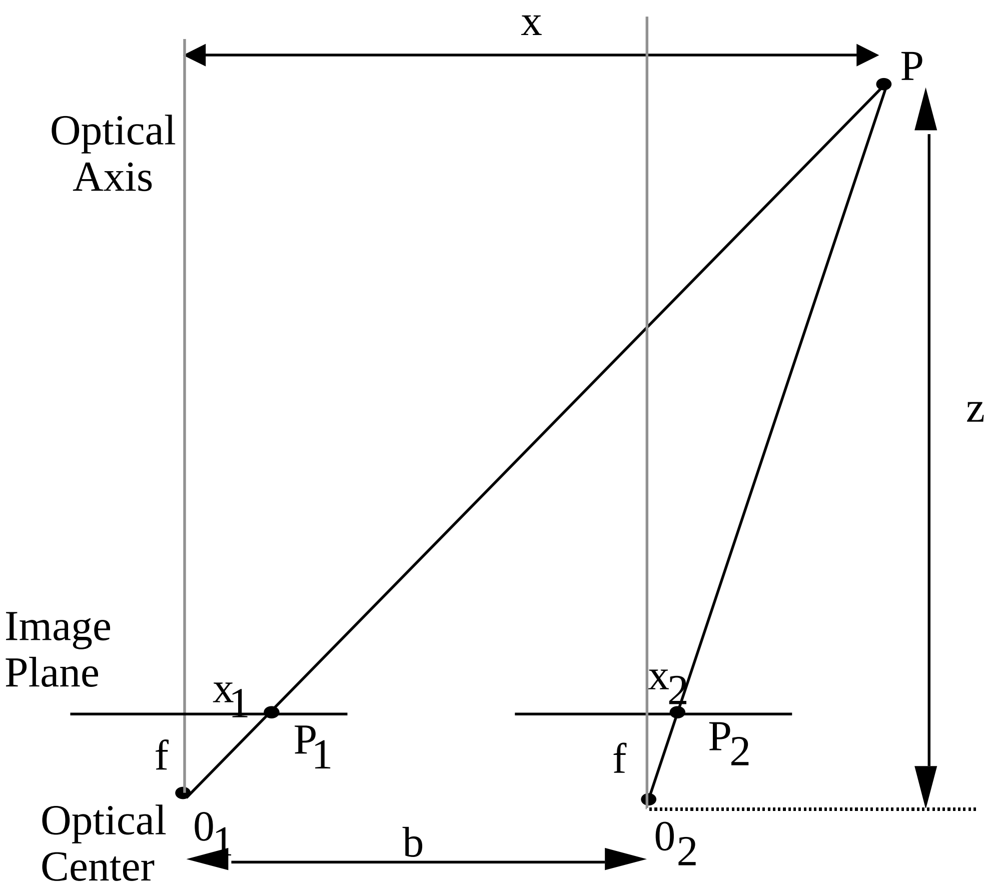
$$\mathbf{A}\mathbf{x}_w + \mathbf{b} = 0$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \dots \\ \mathbf{A}_n \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \dots \\ \mathbf{b}_n \end{pmatrix}$$

  - Least Squares Solution:

$$\mathbf{x}_w = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$

Example: parallel cameras



$$R_{c_1 w} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad t_{c_1 w} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$R_{c_2 w} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad t_{c_2 w} = \begin{pmatrix} -b \\ 0 \\ 0 \end{pmatrix}$$

- Choosing as absolute reference $W = C_1$

Example: parallel cameras

- Linear system:

$$\begin{pmatrix} 1 & 0 & -u_{n1} \\ 0 & 1 & -v_{n1} \\ 1 & 0 & -u_{n2} \\ 0 & 1 & -v_{n2} \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -b \\ 0 \end{pmatrix} = 0$$

- Analytical solution:

$$z_w = \frac{b}{u_{n1} - u_{n2}}$$

$$x_w = u_{n1} \frac{b}{u_{n1} - u_{n2}}$$

$$y_w = v_{n1} \frac{b}{u_{n1} - u_{n2}}$$

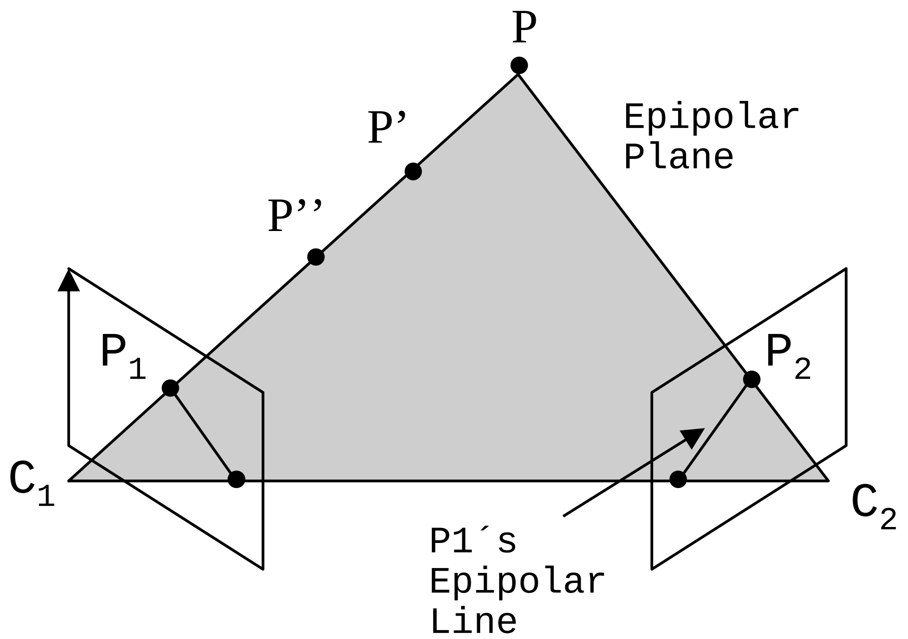
$$v_{n1} = v_{n2}$$

Epipolar Constraint

- Epipolar constraint: The corresponding points must be on the same row in both images

5. Correspondence Search

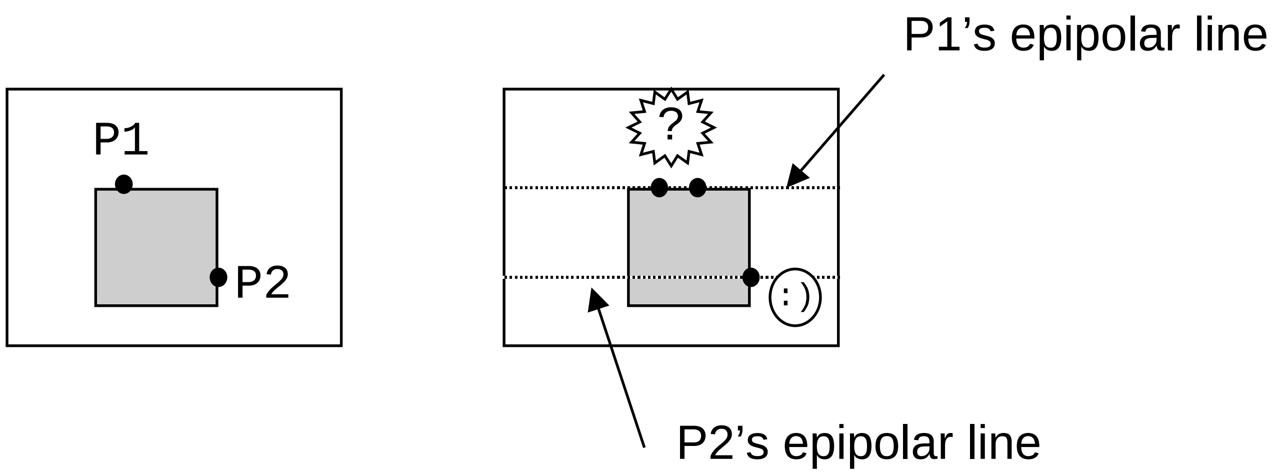
- Give a point P, its epipolar plane is PC_1C_2
- Its projections P_1 and P_2 belong to P's epipolar plane



- Epipolar constraint for P_1:
  - Given C_1 , C_2 y P_1, compute the epipolar plane
  - Intersect the epipolar plane with the image plane 2 to obtain P_1's epipolar line in image 2
  - The point corresponding to P_1 must be on the epipolar line
- They can be several candidates P_2 , P_2',...

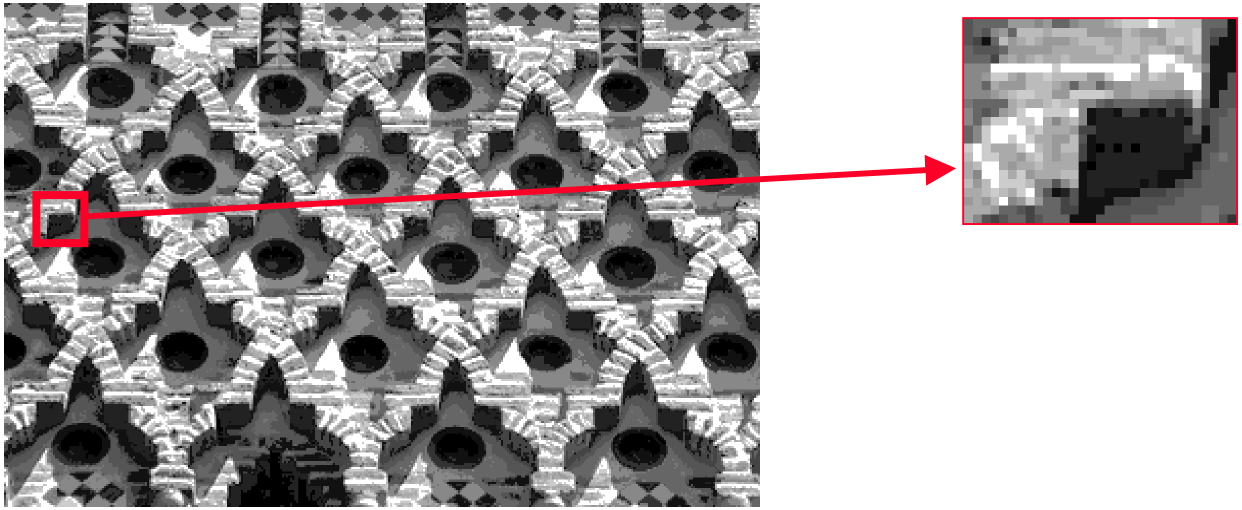
Edge Correspondences

- Detect edge points in both images
- For each edge in image 1, search for an edge along its epipolar line having:
  - Similar gradient magnitude
  - Similar gradient orientation
- Contours oriented parallel to its epipolar line cannot be matched:



Matching using Correlation

- For each pixel, take a local patch and search for a similar patch along the epipolar line on the second image



Correlation

- Sum of Squared Differences SSD:

$$\sum_{i,j}\left(W_{ij}^{a}-W_{ij}^{b}\right)^{2}$$

Searched nxm patch from the first image

Each posible nxm patch on the second image

» Problematic if illumination changes

- Normalized Correlation:

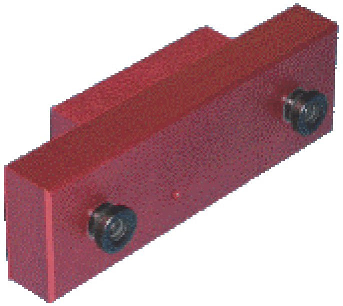
$$\rho=\frac{\sum_{ij}\left(W_{ij}^{a}-\overline{W_{ij}}^{a}\right)\left(W_{ij}^{b}-\overline{W_{ij}}^{b}\right)}{\sqrt{\sum_{ij}\left(W_{ij}^{b}-\overline{W_{ij}}^{b}\right)^{2}}\sqrt{\sum_{ij}\left(W_{ij}^{a}-\overline{W_{ij}}^{a}\right)^{2}}} \qquad \rho\in[-1,1]$$

» ρ close to 1: linear gray level relation → correspondence

Stereo using Correlation

- SRI Stereo Engine algorithm
  by Kurt Konolige
  http://www.ai.sri.com/~konolige/

  1. Rectification - remove distortion from images
  2. Feature Extraction - Laplacian of Gaussian
  3. Correlation - compute disparities by matching
  4. Filtering - remove uncertain matches
  5. 3D reconstruction - convert disparities to 3D points



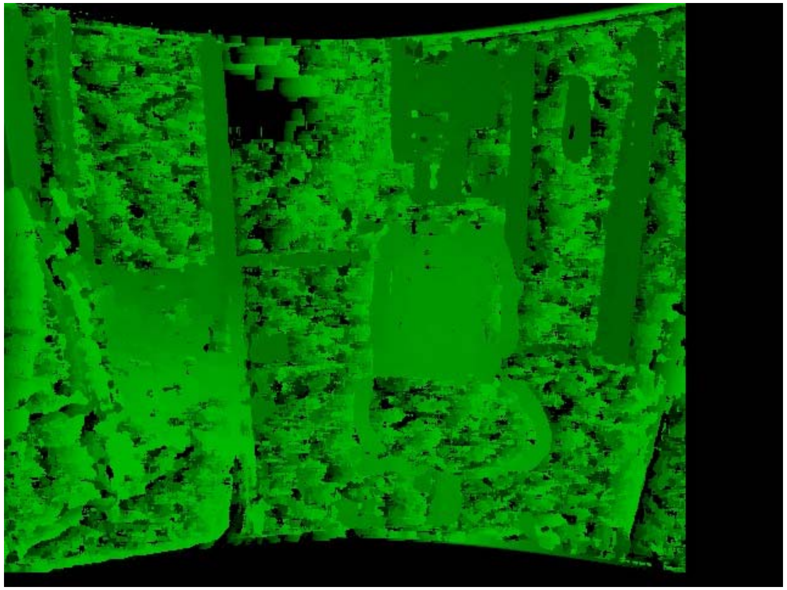
Stereo using Correlation

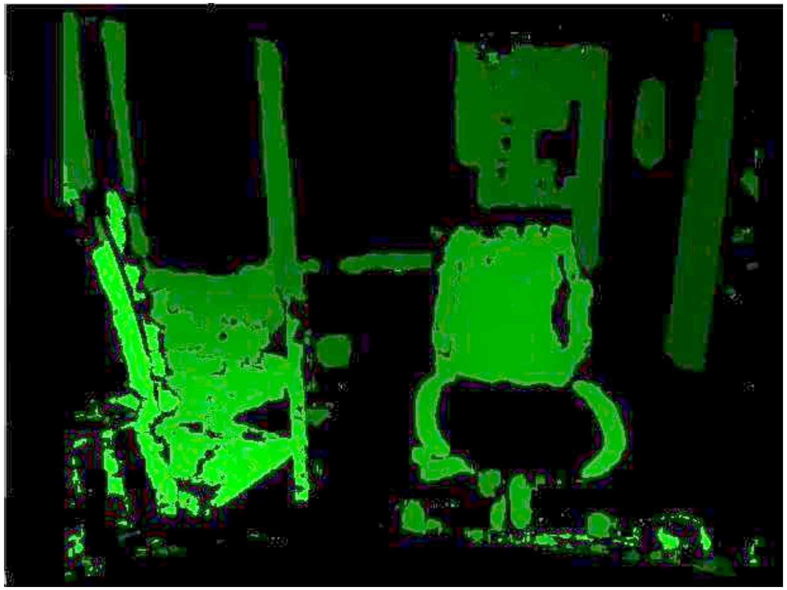
- SRI Stereo Engine algorithm

  by Kurt Konolige

  http://www.ai.sri.com/~konolige/

  1. Rectification - remove distortion from images
  2. Feature Extraction - Laplacian of Gaussian
  3. Correlation - compute disparities by matching
  4. Filtering - remove uncertain matches
  5. 3D reconstruction - convert disparities to 3D points





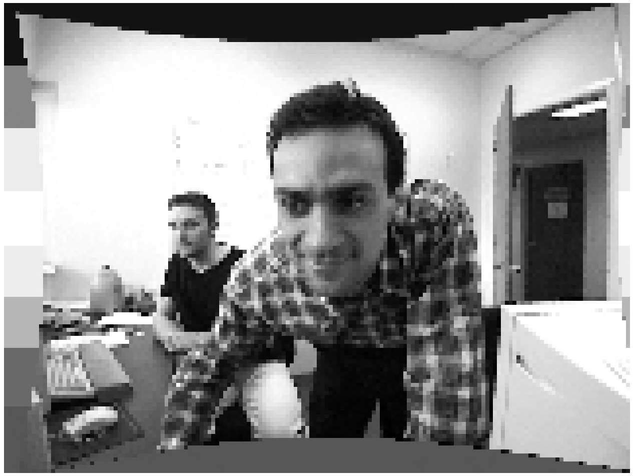
Stereo using Correlation

- SRI Stereo Engine algorithm
  by Kurt Konolige
  http://www.ai.sri.com/~konolige/

  1. Rectification - remove distortion from images
  2. Feature Extraction - Laplacian of Gaussian
  3. Correlation - compute disparities by matching
  4. Filtering - remove uncertain matches
  5. 3D reconstruction - convert disparities to 3D points

Stereo using Correlation

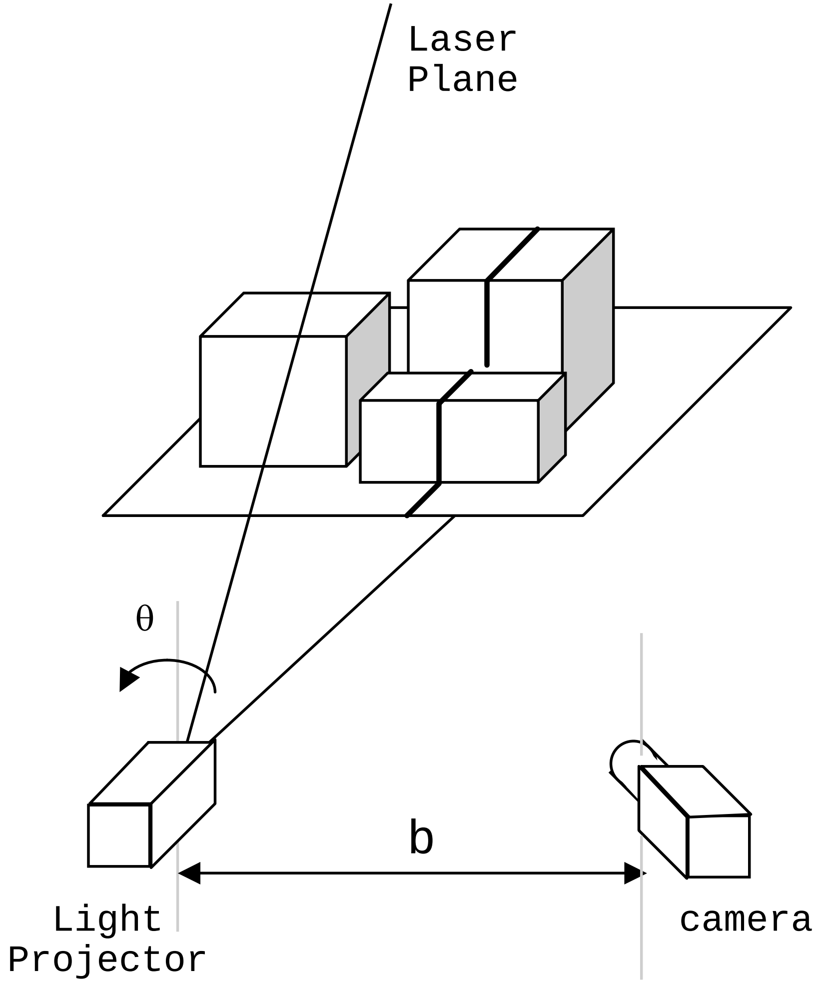
- SRI Stereo http://www.ai.sri.com/~konolige/



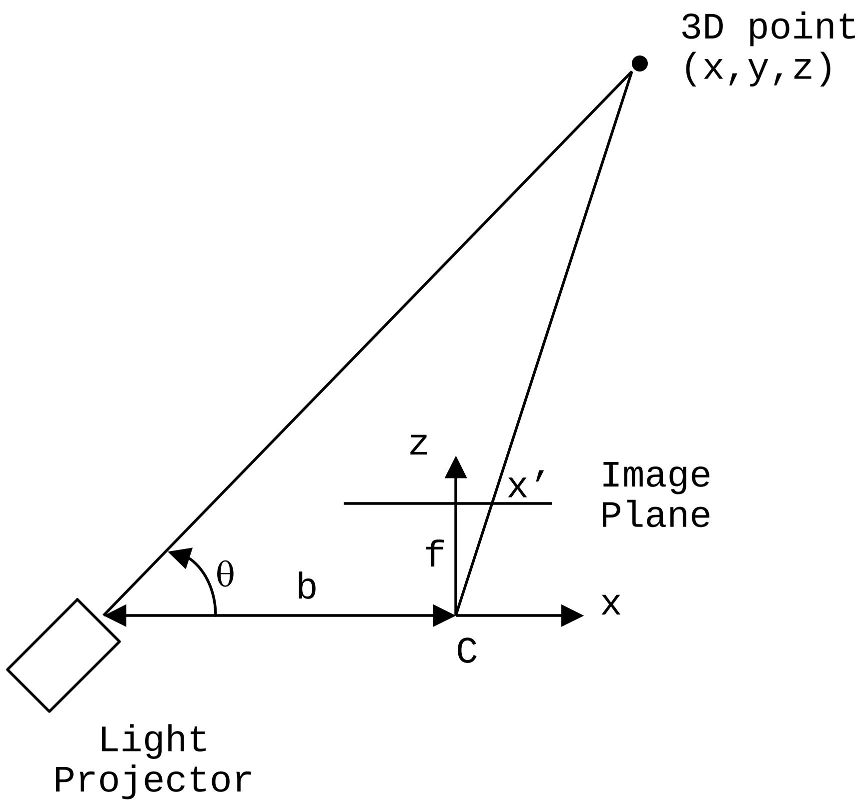
6. Other 3D techniques: Structured Light

- Sweep the scene with a laser point or a laser plane



Structured Light

- 3D information from triangulation

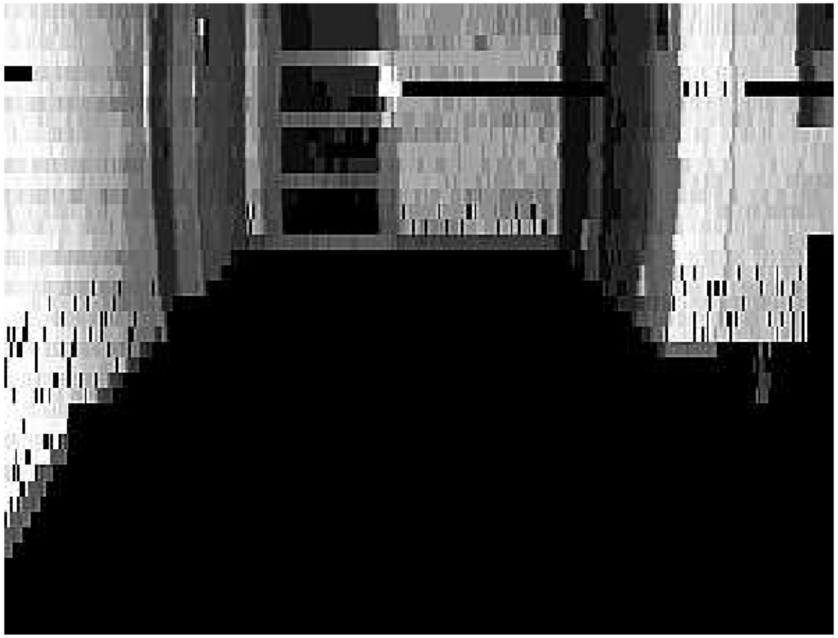


$$x' = f\,\frac{x}{z} \qquad y' = f\,\frac{y}{z}$$

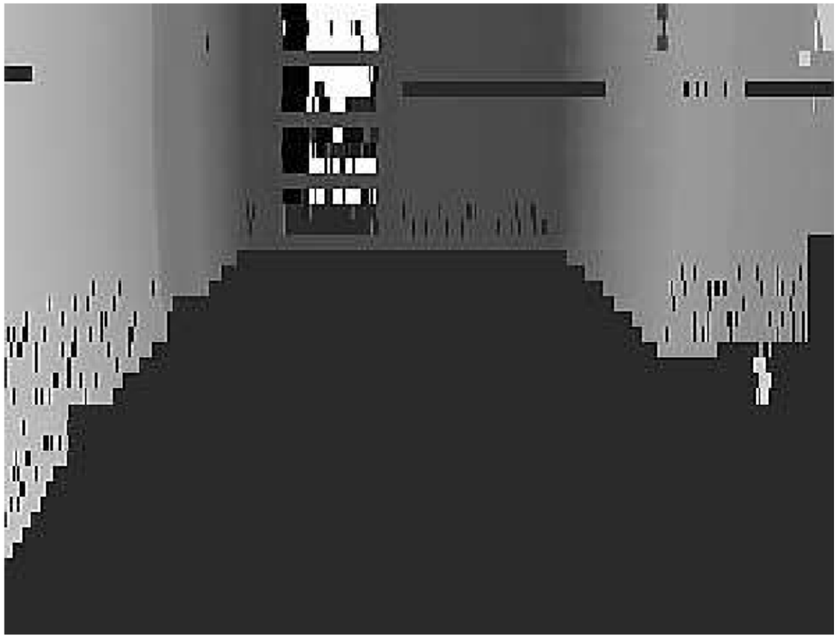
$$\tan\theta = \frac{z}{x+b}$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{b}{f\cot\theta - x'} \begin{pmatrix} x' \\ y' \\ f \end{pmatrix}$$

Other 3D techniques: Range Camera

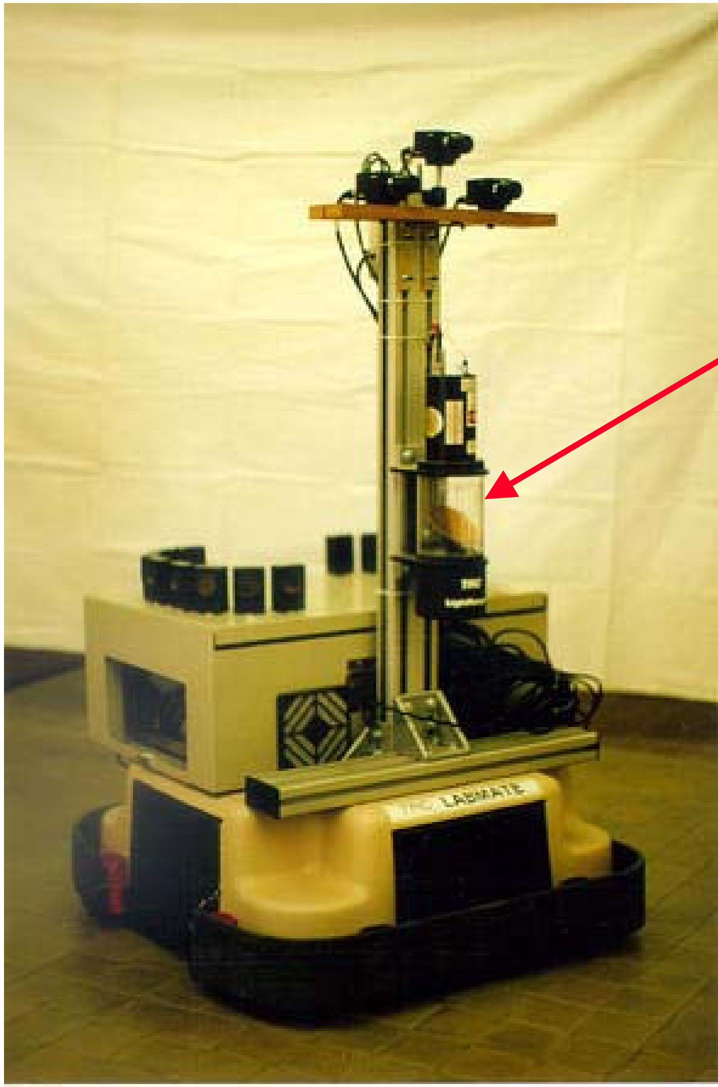


Intensity Image

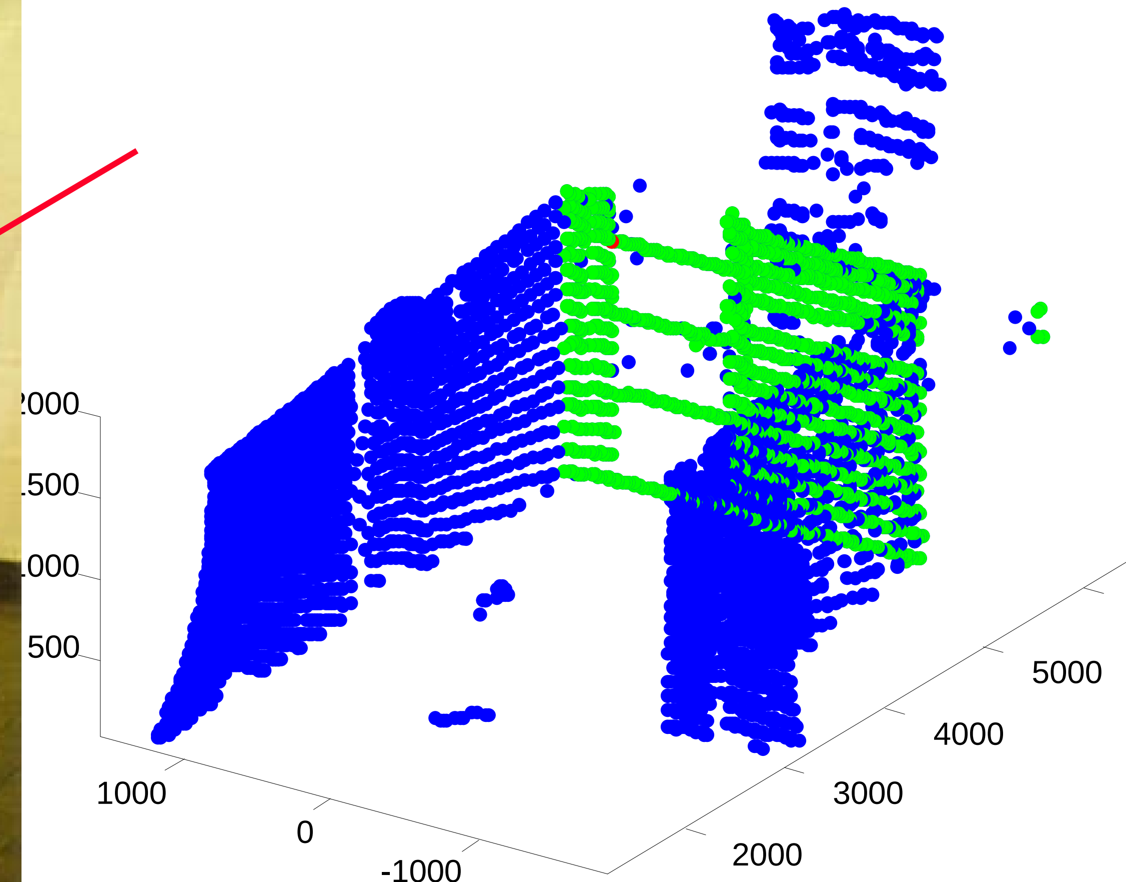


Depth Image

Other 3D techniques: 3D Laser Scanner

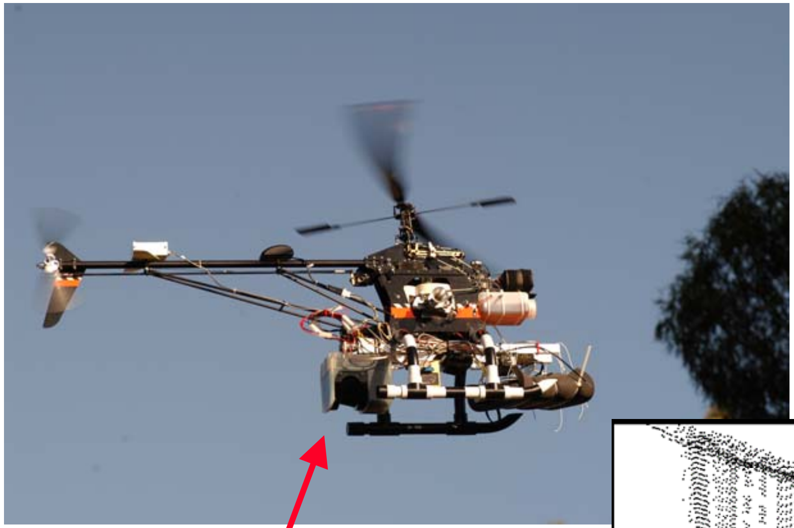


Otilio



2D Laser Scanner + Motion

- Scan Alignment and 3-D Surface Modeling with a Helicopter Platform by Sebastian Thrun, Mark Diel, and Dirk Haehnel, 2003 International Conference of Field and Service Robotics.



2D Laser Scanner