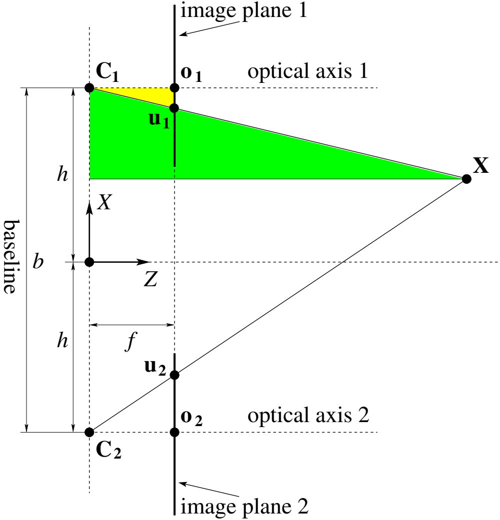
Geometry of standard stereo

$$\frac{u_1}{f} = \frac{h - X}{Z}$$
$$-\frac{u_2}{f} = \frac{h + X}{Z}$$
$$v_1 = v_2$$

$$Z = \frac{2hf}{u_1 - u_2} = \frac{bf}{d}$$
$$X = -\frac{b(u_1 + u_2)}{2d}$$
$$Y = \frac{bv_1}{d} = \frac{bv_2}{d}$$

$d \doteq u_1 - u_2$ disparity



Precision of depth estimation

- If $d \to 0$, and $Z \to \infty$
  - Disparity of distant points are small.
- Relation between disparity and precision of depth estimation

$$\frac{|\Delta Z|}{Z} = \frac{|\Delta d|}{|d|}$$

  - larger the disparity, smaller the relative depth error

$\to$ precision is increasing

- Influence of base length

$$d = \frac{bf}{Z}$$

  - For larger b, same depth value yields larger disparity

$\to$ Precision of depth estimation increasing

$\to$ more pixels $\longrightarrow$ precision of diparity increasing

Types of stereo reconstruction

- **Fully calibrated** reconstruction
  - Known **intrinsic and extrinsic** camera parameters
  - reconstruction by triangulation
  - known baseline $\longrightarrow$ known scale
- **Metric (Euclidean)** reconstruction
  - knonw **intrinsic** camera parameters, $n \geq 8$ point correspondences given
  - Extrinsic camera parameters obtained from **essential matrix**
  - Reconstruction up to a similarity transformation

  $\rightarrow$ up to a scale
- **Projective reconstruction**
  - **unknown** camera parameters, $n \geq 8$ point correspondences are given
  - Composition of projective matrices from a **fundamental matrix**
  - reconstruction can be computed up to a projective transformation

Overview

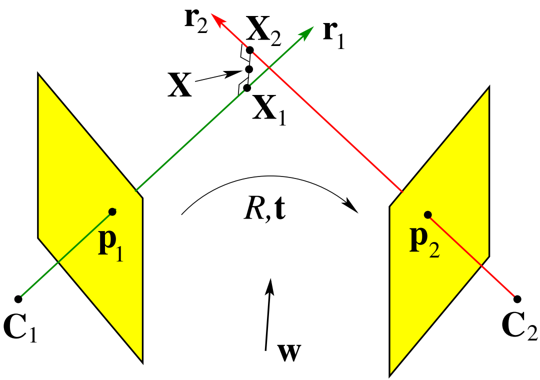
Triangulation

- **Task:**
  - Two calibrated cameras are given, including both intrinsic and extrinsic parameters, and
  - Locations $\mathbf{u}_1, \mathbf{u}_2$ of the projection of spatial point $\mathbf{X}$ are given
  - Goal is to estimate spatial location $\mathbf{X}$.
- Two calibration matrices are known, therefore
  - for a projection matrix: $\mathbf{K}^{-1}\mathbf{P} = [\mathbf{R}| - \mathbf{t}]$ and
  - for calibrated (aka. normalized) coordinates: $\mathbf{p} = \mathbf{K}^{-1}\mathbf{u}$.
- For the sake of simplicity, the first camera gives the world coordinate system
  - **non-homogeneous** coordinates are used

  $\rightarrow$ $\mathbf{p}_2 = \mathbf{R}(\mathbf{p}_1 - \mathbf{t}), \mathbf{p}_1 = \mathbf{t} + \mathbf{R}^T\mathbf{p}_2$
- Image points are bask-projected to 3D space
  - two rays obtained, they usually do not intersect each other due to noise/calibration error

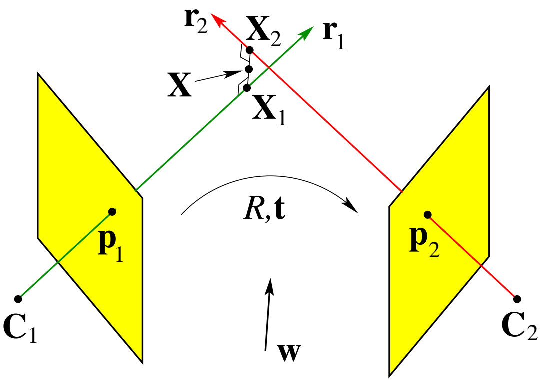
  $\rightarrow$ task is to give an estimate for spatial point $\mathbf{X}$.

Linear triangulation: geometry



- Line $\mathbf{X}_1\mathbf{X}_2$ perpendicular to both $\mathbf{r}_1$ and $\mathbf{r}_2$.
- Estimate $\mathbf{X}$ is the middle point of section $\mathbf{X}_1\mathbf{X}_2$
- Vector $\mathbf{w}$ is parallel to $\mathbf{X}_1\mathbf{X}_2$.

Linear triangulation: notations



- $\alpha \mathbf{p}_1$ is a point on ray $\mathbf{r}_1$ ($\alpha \in \Re$)
- $\mathbf{t} + \beta R^T \mathbf{p}_2$ a point on other ray $\mathbf{r}_2$ ($\beta \in \Re$)
  - $\rightarrow$ coordinate system fixed to the first camera
- Let $\mathbf{X}_1 = \alpha_0 \mathbf{p}_1$, $\mathbf{X}_2 = \mathbf{t} + \mathbf{R}^T(\beta_0 \mathbf{p}_2 - \mathbf{t})$

Linear triangulation: solution

- Task is to determine
  - the middle point of the line section $\mathbf{X}_1\mathbf{X}_2$

  $\rightarrow$ determination of α_0 and β_0 required
- Remark that
  - Vector $\mathbf{w} = \mathbf{p}_1 \times \mathbf{R}^T(\mathbf{p}_2 - \mathbf{t})$ perpendicular to both $\mathbf{r}_1$ and $\mathbf{r}_2$.
  - Line $\alpha\mathbf{p}_1 + \gamma\mathbf{w}$ parallel to $\mathbf{w}$ and contain the point $\alpha\mathbf{p}_1$ ($\gamma \in \Re$).

$\rightarrow$ α_0, β_0 (as well as γ_0) are given by the solution of the following linear system: :

$$\alpha\mathbf{p}_1 + \mathbf{t} + \beta\mathbf{R}^T(\mathbf{p}_2 - \mathbf{t}) + \gamma[\mathbf{p}_1 \times \mathbf{R}^T(\mathbf{p}_2 - \mathbf{t})] = 0 \tag{7}$$

- Triangulated point is obtained, e.g by $\alpha_0\mathbf{p}_1$
- There is no solution if $\mathbf{r}_1$ and $\mathbf{r}_2$ are parallel

Linear triangulation: an algebraic solution

- Two projected locations of spatial point $\mathbf{X}$ are given:

$$\lambda_1 \mathbf{u}_1 = \mathbf{P}_1 \mathbf{X}$$
$$\lambda_2 \mathbf{u}_2 = \mathbf{P}_2 \mathbf{X}$$

- λ_1 and λ_2 can be eliminated. 2 + 2 equations are obtained:

$$u \mathbf{p}_3^T \mathbf{X} = \mathbf{p}_1^T \mathbf{X}$$
$$v \mathbf{p}_3^T \mathbf{X} = \mathbf{p}_2^T \mathbf{X}$$

- where $\mathbf{p}_i^T$ is the i-th row of projection matrix $\mathbf{P}$.
- Both projections yield 2 equations. Only vector $\mathbf{X}$ is unknown.
- Solution for $\mathbf{X}$ is calculated by solving the homogeneous linear system of equations.
- Important remark: solution is obtained in homogeneous coordinates.

Refinement by minimizing the reprojection error

- Linear algorithm yield points $\mathbf{X}_i$, $i = 1, 2, \ldots, n$ if n point pairs are given
- The solution should be refined
  - minimization of **reprojection error** yields **more accurate** estimate
- For minimizing the reprojection error, the following parameters have to be refined:
  - Spatial points $\mathbf{X}_i$
  - Rotation matrix $\mathbf{R}$ and baseline vector $\mathbf{t}$

  $\rightarrow$ intrinsic camera parameters are usually fixed as cameras are pre-calibrated
- Initial values for numerical optimization
  - Spatial points $\mathbf{X}_i$ from linear triangulation
  - Initial rotation matrix $\mathbf{R}$ and baseline vector $\mathbf{t}$ by decomposing the essential matrix

Metric reconstruction by decomposing the essential matrix

- Intrinsic camera matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ given, fundamental matrix computed from $n \geq 8$ point correpondences
  - $\mathbf{E}$ can be retrieved from $\mathbf{F}$, $\mathbf{K}_1$ and $\mathbf{K}_2$.
  - from $\mathbf{E}$, extrinsic parameters can be obtained by decomposition
- Unknown baseline $\longrightarrow$ unknown scale
  - baseline normalized to 1
  - $\rightarrow$ Euclidean reconstruction possible up to a similarity transformation
- It is assumed that world coordinate is fixed to the first camera
  - $\rightarrow$ Therefore, $P_1 = [\mathrm{I}|\mathbf{0}]$, where I is the identity matrix
- Position of second camera computed from essential matrix $\mathbf{E}$ by SVD.
  - Four solutions obtained,
  - only one is correct.

Camera pose estimation by SVD

- The Singular Value Decompoisition of $\mathbf{E}$ is $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where $\mathbf{D} = \text{diag}(\delta, \delta, 0)$
  $\rightarrow$ $\mathbf{E}$ has two equal singuar values
- Four solutions can be obtained as follows:

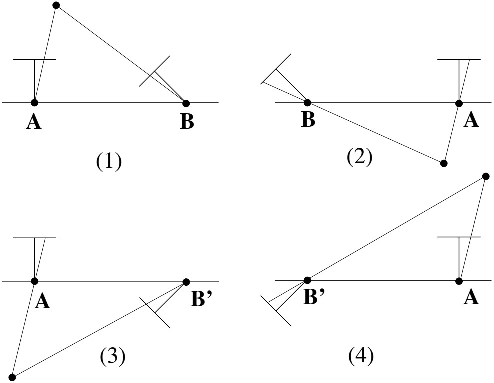
$$\mathbf{R}_1 = \mathbf{U}\mathbf{W}\mathbf{V}^T \qquad \mathbf{R}_2 = \mathbf{U}\mathbf{W}^T\mathbf{V}^T$$
$$[\mathbf{t}_1]_\times = \delta\mathbf{U}\mathbf{Z}\mathbf{U}^T \qquad [\mathbf{t}_2]_\times = -\delta\mathbf{U}\mathbf{Z}\mathbf{U}^T$$

- where

$$\mathbf{W} \doteq \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{Z} \doteq \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Combination of 2-2 candidates for translation and rotation yield 4 solutions.
- Determinants of $\mathbf{R}_1$ and $\mathbf{R}_2$ have to be positive, otherwise matrices should be multiplied by -1.

Visualization of the four solutions



- Left and right: camera locations replaces
- Top and bottom: mirror to base lane
- 3D point is in front of the cameras only in the top-left case.