COMPSCI 773 S1C

3D Reconstruction

A/Prof. Patrice Delmas and Prof. Georgy Gimel'farb

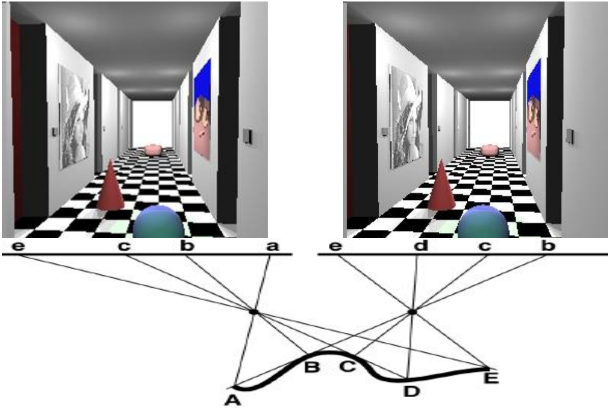
Computational Stereo Vision

Very diverse applications: from ophthalmology, biometrics, and architecture to virtual reality, autonomous navigation, robotics, cartography, reverse engineering, etc. . .

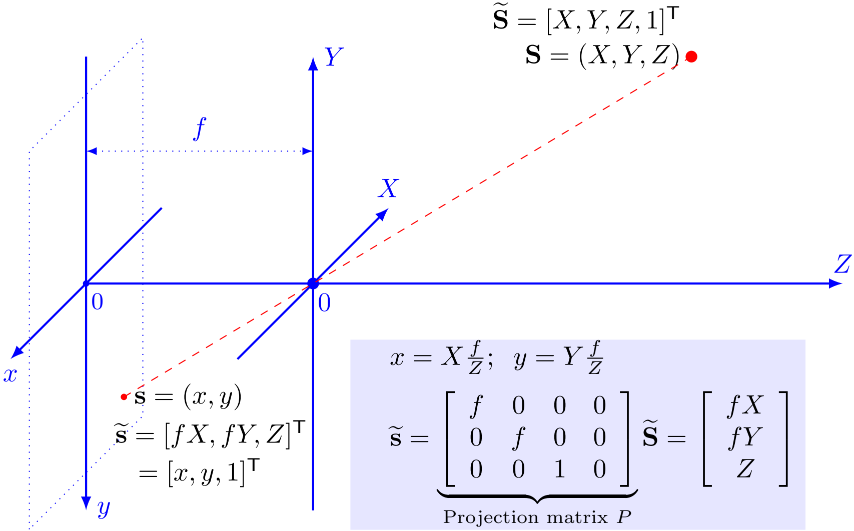
Computational Stereo: 3D Reconstruction

Reconstructing 3D points of an optical (visible) surface of a 3D scene by finding corresponding 2D points in two or more images

- **Stereo matching**: search for correspondences in a stereo pair or more images of a 3D scene



Imaging: 3D-to-2D Projection (the canonical coordinate frame)



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

$$\widetilde{\mathbf{s}} = \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{Projection matrix } P} \widetilde{\mathbf{S}} = \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix}$$

Imaging: 3D-to-2D Projection (the canonical coordinate frame)

Example of the projection: the focal distance $f = 10$

- (X, Y, Z)-coordinates of the 3D point: $\mathbf{S} = (30, 40, 20)$
- Homogeneous coordinates of this point: $\widetilde{\mathbf{S}} = [30, 40, 20, 1]^\mathsf{T}$

- (x, y)-coordinates of the projected point: $\mathbf{s} = (15, 20)$:

$$x = X\frac{f}{Z} = 30\frac{10}{20} = 15; \qquad y = Y\frac{f}{Z} = 40\frac{10}{20} = 20$$

- In homogeneous coordinates: $\widetilde{\mathbf{s}} = P\widetilde{\mathbf{S}}$

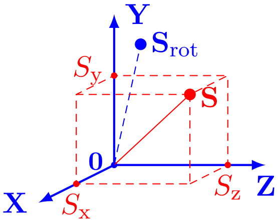
$$\widetilde{\mathbf{s}} = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 30 \\ 40 \\ 20 \\ 1 \end{bmatrix} = \begin{bmatrix} 300 \\ 400 \\ 20 \end{bmatrix} \equiv \begin{bmatrix} 15 \\ 20 \\ 1 \end{bmatrix}$$

Imaging: 3D-to-2D Projection (a general-case coordinate frame)

Canonical camera (XYZ)-coordinate frame with ortonormnal vectors $\mathbf{X} = [1, 0, 0]^\mathsf{T}$; $\mathbf{Y} = [0, 1, 0]^\mathsf{T}$, and $\mathbf{Z} = [0, 0, 1]^\mathsf{T}$:

- 3D point $\mathbf{S} = [S_\mathrm{x}, S_\mathrm{y}, S_\mathrm{z}]^\mathsf{T}$: the vector $\mathbf{S} = S_\mathrm{x}\mathbf{X} + S_\mathrm{y}\mathbf{Y} + S_\mathrm{z}\mathbf{Z}$
- Rotated point ($R = [r_{ij}]_{i,j=1,1}^{3,3}$ – the rotation matrix):

$$\mathbf{S}_\mathrm{rot} = R\mathbf{S} = \begin{bmatrix} S_\mathrm{rot:x} = r_{11}S_x + r_{12}S_y + r_{13}S_z \\ S_\mathrm{rot:y} = r_{21}S_x + r_{22}S_y + r_{23}S_z \\ S_\mathrm{rot:z} = r_{31}S_x + r_{32}S_y + r_{13}S_z \end{bmatrix}$$



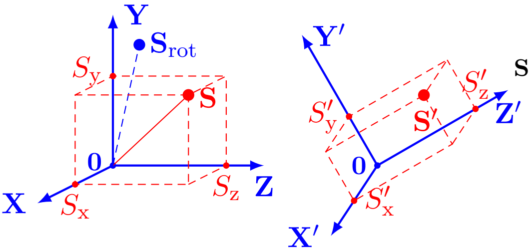
Imaging: 3D-to-2D Projection (a general-case coordinate frame)

Rotated $(X'Y'Z')$-frame with the orthonormal vectors:

$$\begin{aligned}\mathbf{X}' &= R\mathbf{X} = [r_{11}, r_{21}, r_{31}]^\mathsf{T}\\ \mathbf{Y}' &= R\mathbf{Y} = [r_{12}, r_{22}, r_{32}]^\mathsf{T}\\ \mathbf{Z}' &= R\mathbf{Z} = [r_{13}, r_{23}, r_{33}]^\mathsf{T}\end{aligned}$$

The same point $\mathbf{S}' = \left[S'_\mathrm{x}, S'_\mathrm{y}, S'_\mathrm{z}\right]^\mathsf{T} \rightarrow$

the vector $\mathbf{S}' = S'_\mathrm{x}\mathbf{X}' + S'_\mathrm{y}\mathbf{Y}' + S'_\mathrm{z}\mathbf{Z}'$



$$\mathbf{S}' = R^\mathsf{T}\mathbf{S} = \begin{bmatrix} S'_\mathrm{x} = (\mathbf{X}')^\mathsf{T}\mathbf{S} \\ S'_\mathrm{y} = (\mathbf{Y}')^\mathsf{T}\mathbf{S} \\ S'_\mathrm{z} = (\mathbf{Z}')^\mathsf{T}\mathbf{S} \end{bmatrix}$$

$$R = \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}}_{\text{The rotation matrix}}$$

Imaging: 3D-to-2D Projection (a general-case coordinate frame)

W.r.t. the canonical frame, let a new frame be rotated by rotation matrix R and its origin $\mathbf{0}'$ be translated by $\mathbf{T} = [t_\mathrm{x}, t_\mathrm{y}, t_\mathrm{z}]^\mathsf{T}$:

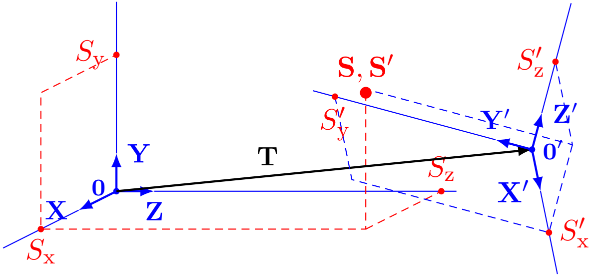
$$\overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}^{I_{3\times3}=[\mathbf{XYZ}]} \Rightarrow \overbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}}^{R=[\mathbf{X}'\mathbf{Y}'\mathbf{Z}']}$$

$$\Rightarrow \begin{cases} \mathbf{X}' = R\mathbf{X} & = [r_{11}, r_{21}, r_{31}]^\mathsf{T} \text{ in } \mathbf{0XYZ} \\ \mathbf{Y}' = R\mathbf{Y} & = [r_{12}, r_{22}, r_{32}]^\mathsf{T} \text{ in } \mathbf{0XYZ} \\ \mathbf{Z}' = R\mathbf{Z} & = [r_{13}, r_{23}, r_{33}]^\mathsf{T} \text{ in } \mathbf{0XYZ} \end{cases}$$

Orthonormal matrix $R = \| r_{ij} \|_{i,j=1}^{3}$: $\sum_{i=1}^{3} r_{ij} r_{ik} = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ otherwise} \end{cases}$

so that $R^{-1} = R^\mathsf{T}$ (inversion = transposition of R).

Imaging: 3D-to-2D Projection (a general-case coordinate frame)



$$\mathbf{S} = \begin{bmatrix} S_x \\ S_y \\ S_z \end{bmatrix} \text{ in } \mathbf{0XYZ}$$

$$\mathbf{S}' = \begin{bmatrix} S'_x \\ S'_y \\ S'_z \end{bmatrix} \text{ in } \mathbf{0'X'Y'Z'} = R^{\mathsf{T}}(\mathbf{S} - \mathbf{T})$$

Imaging: 3D-to-2D Projection (an example)

$\mathbf{S} = [5, 4, 3]^\mathsf{T}$; $\mathbf{T} = [1, 2, 3]$; $\mathbf{X} = [1, 0, 0]^\mathsf{T}$; $\mathbf{Y} = [0, 1, 0]^\mathsf{T}$; $\mathbf{Z} = [0, 0, 1]^\mathsf{T}$

$$R = \begin{bmatrix} 0.80 & -0.36 & -0.48 \\ 0.60 & 0.48 & 0.64 \\ 0.00 & -0.80 & 0.60 \end{bmatrix} \Rightarrow \begin{cases} [\mathbf{X}'\mathbf{Y}'\mathbf{Z}'] = R\,[\mathbf{XYZ}] \\ [\mathbf{XYZ}] = R^\mathsf{T}\,[\mathbf{X}'\mathbf{Y}'\mathbf{Z}'] \end{cases}$$

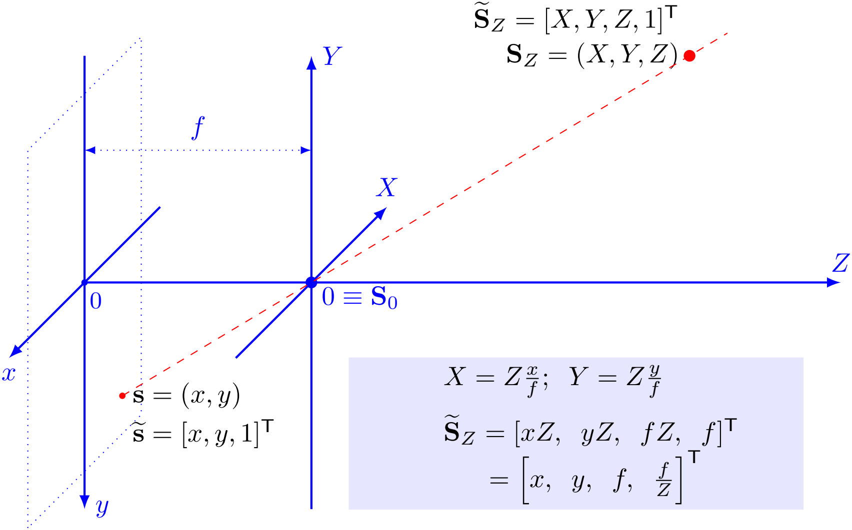
Vector $\mathbf{S}$ in the translated and rotated coordinate frame $\mathbf{0}'\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$:

- W.r.t. $\mathbf{0}'$ in $\mathbf{0XYZ}$: $\mathbf{U} = \mathbf{S} - \mathbf{T} = [5-1, 4-2, 3-3]^\mathsf{T} = [4, 2, 0]^\mathsf{T}$
- Components of $\mathbf{U} = \mathbf{S} - \mathbf{T}$ in the frame $\mathbf{0}'\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$:

$$\begin{array}{lclclcr} U'_\mathrm{x} & = & (\mathbf{X}')^\mathsf{T}\mathbf{U} & = & 4\cdot 0.80 + 2\cdot 0.60 + 0\cdot 0.00 & = & 4.40 \\ U'_\mathrm{y} & = & (\mathbf{Y}')^\mathsf{T}\mathbf{U} & = & -4\cdot 0.36 + 2\cdot 0.48 - 0\cdot 0.80 & = & -0.48 \\ U'_\mathrm{z} & = & (\mathbf{Z}')^\mathsf{T}\mathbf{U} & = & -4\cdot 0.48 + 2\cdot 0.64 + 0\cdot 0.60 & = & -0.64 \end{array}$$

- $\mathbf{U}' = R^\mathsf{T}\mathbf{U} \equiv 4.40\mathbf{X} - 0.48\mathbf{Y} - 0.64\mathbf{Z}$ in the frame $\mathbf{0}'\mathbf{X}'\mathbf{Y}'\mathbf{Z}'$
- $\mathbf{U} = R\mathbf{U}' \equiv 4.40(\mathbf{X}') - 0.48(\mathbf{Y}') - 0.64(\mathbf{Z}')$ in $\mathbf{0XYZ}$
- Vector $\mathbf{S} = \mathbf{T} + R\mathbf{U}' = [5, 4, 3]^\mathsf{T}$

Imaging: 2D-to-3D Inverse Projecting Ray



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

$$\widetilde{\mathbf{S}}_Z = [xZ, \ yZ, \ fZ, \ f]^\mathsf{T} = \left[x, \ y, \ f, \ \frac{f}{Z}\right]^\mathsf{T}$$

Imaging: 2D-to-3D Inverse Projecting Ray

2D image point: $\mathbf{s} = (15, 20)$; $\widetilde{\mathbf{s}} = [15, 20, 1]^\mathsf{T}$

Example of the projecting ray with the focal distance $f = 10$

(X, Y, Z)-points of the inverse ray: $\mathbf{S}_Z = (1.5Z, 2.0Z, Z)$:

$$X = Z\tfrac{x}{f} = Z\tfrac{15}{10} = 1.5Z;$$

$$Y = Z\tfrac{y}{f} = Z\tfrac{20}{10} = 2.0Z$$

In homogeneous coordinates:
$\widetilde{\mathbf{S}}_Z = [15Z,\ \ 20Z,\ \ 10Z,\ \ 10]^\mathsf{T} \equiv \left[15,\ \ 20,\ \ 10,\ \ \tfrac{10}{Z}\right]^\mathsf{T}$

Imaging: 2D-to-3D Inverse Projecting Ray

In homogeneous coordinates:

$$\widetilde{\mathbf{S}}_Z = [15Z,\ \ 20Z,\ \ 10Z,\ \ 10]^\mathsf{T} \equiv \left[15,\ \ 20,\ \ 10,\ \ \tfrac{10}{Z}\right]^\mathsf{T}$$

$$Z \to 0: \quad \widetilde{\mathbf{S}}_0 = \left[15, 20, 10, \tfrac{10}{0} = \infty\right]^\mathsf{T} \quad \to$$

$$\mathbf{S}_0 = \lim_{t\to\infty}\left(\tfrac{15}{t}, \tfrac{20}{t}, \tfrac{10}{t}\right) \quad = \quad (0, 0, 0)$$

$$Z \to \infty: \quad \widetilde{\mathbf{S}}_\infty = \left[15, 20, 10, \tfrac{10}{\infty} = 0\right]^\mathsf{T} \quad \to$$

$$\mathbf{S}_\infty = \lim_{t\to 0}\left(\tfrac{15}{t}, \tfrac{20}{t}, \tfrac{10}{t}\right) \quad = \quad (\infty, \infty, \infty)$$

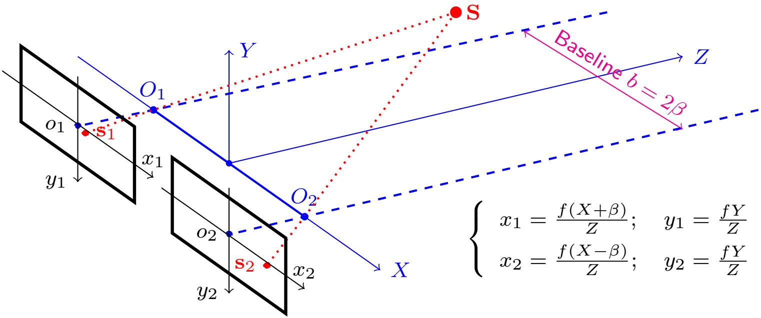
Note that homogeneous coordinates can discriminate between infinitely far 3D points of different rays!

3D Vision: Visible Surface from Two or More Images

Restoring a 3D visible surface by inverse projections of its images:

- Cameras **calibration** to find projective parameters.
  - *Extrinsic*: cameras poses w.r.t. a world coordinate frame.
    - Optical centre + optical axis + image coordinate axes.
  - *Intrinsic*: focal distance; principal point; optical distortions
- **Rectification** to simplify searching for corresponding points.
  - Epipolar geometry; canonical stereo geometry.
- **Stereo matching** to establish correspondences.
  - Similarity of image signals corresponding to a visible 3D point.
  - Prior constraints for partially occluded points.
- **3D surface reconstruction**.
  - Triangulation (optical rays' intersections) for visible 3D points.
  - Prior constraints for only monocularly visible points.
  - Prior constraints for surface continuity and smoothness

3D Vision: Basic Notions



$$\begin{cases} x_1 = \frac{f(X+\beta)}{Z}; & y_1 = \frac{fY}{Z} \\ x_2 = \frac{f(X-\beta)}{Z}; & y_2 = \frac{fY}{Z} \end{cases}$$

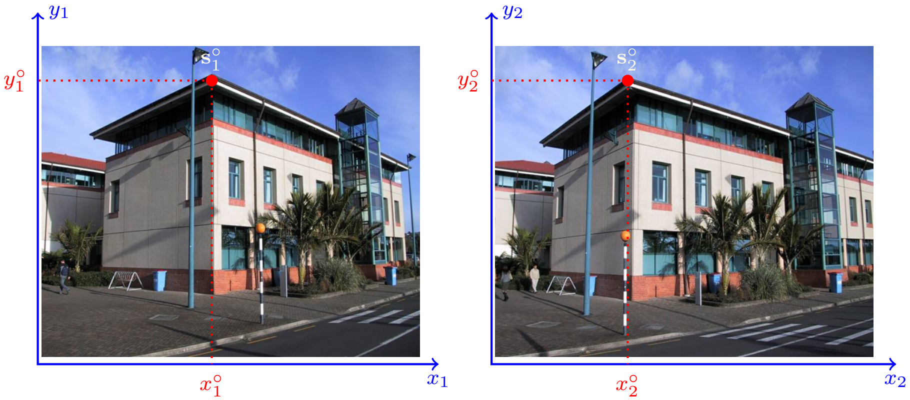
3D point coordinates $\mathbf{S} = (X, Y, Z)$ are linked to **disparities** (parallaxes), or differences between coordinates of corresponding 2D image points $\mathbf{s}_1$ and $\mathbf{s}_2$

- Horizontal, or x-disparity: $d = x_1 - x_2 \equiv \frac{bf}{Z}$
- Vertical, or y-disparity, $\delta = y_1 - y_2$, if images are not co-registered

Typical setup: horizontal (x-) baseline of size b; small or zero y-disparities

- **Canonical** stereo geometry (an *epipolar* pair): zero y-disparities

x- and y-Disparities on a Stereo Pair



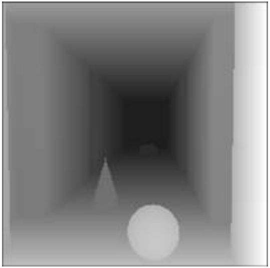
- Corresponding points: $\mathbf{s}_1^\circ = (x_1^\circ, y_1^\circ)$ and $\mathbf{s}_2^\circ = (x_2^\circ, y_2^\circ)$
- x-disparity: $d^\circ = x_1^\circ - x_2^\circ$
- y-disparity: $\delta^\circ = y_1^\circ - y_2^\circ$ (for the canonical geometry: $\delta^\circ = 0$)

Disparity Maps Formed by Stereo Matching

Generally, a vector-valued map $\mathbf{d}(x,y) = [d_{x,y}, \delta_{x,y}]$

- Mapping coordinates of binocularly visible points in one image to corresponding coordinates in the other image:
$$(x_1 = x, y_1 = y) \leftrightarrow (x_2 = x - d_{x,y}, y_2 = y - \delta_{x,y})$$
- Corresponding signals (intensities, colours) in the images: $g_1(x,y) \leftrightarrow g_2(x - d_{x,y}, y - \delta_{x,y})$

- **Partial occlusion**: if a 3D point is depicted only in one image
- For partially occluded points the mapping is undefined: no stereo correspondence!
- **Canonical epipolar geometry**: a scalar map of x-disparities $d_{x,y}$



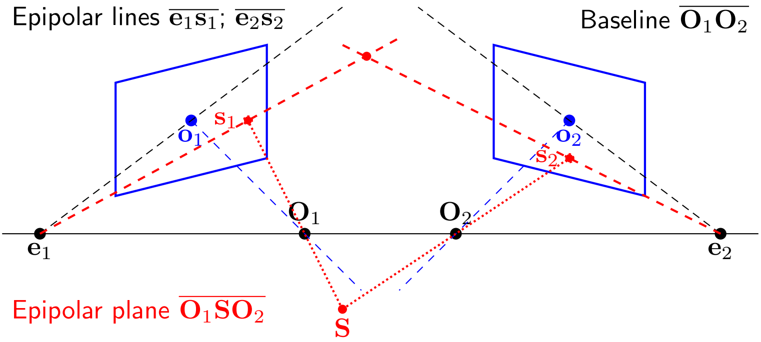
Gray-coded x-disparity map

Stereo Matching: Searching for Correspondences

- **Search region** for a position (x, y) in the left image: a set of candidate correspondences in the right image

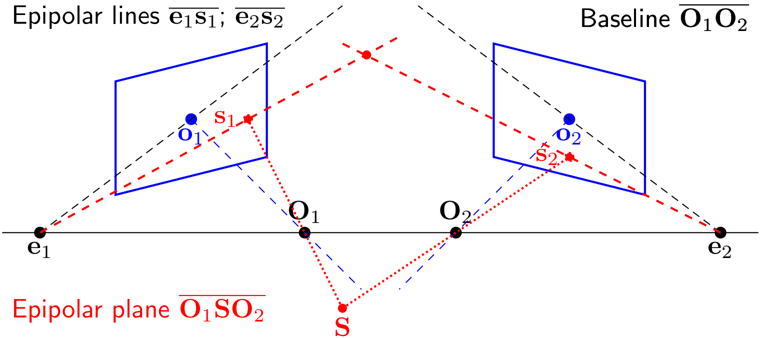
$$\{(x_2, y_2) : x - d_{\max} \leq x_2 \leq x - d_{\min}; y - \delta_{\max} \leq y_2 \leq y - \delta_{\min}\}$$

Corresponding (Conjugate) Epipolar Lines



- Image point $\mathbf{s}_i$ – the projection of a 3D point $\mathbf{S}$ onto image i
- Epipolar line through $\mathbf{s}_i$ – the trace of intersection of the image plane by the epipolar plane containing $\mathbf{S}$ and the baseline $\overline{\mathbf{O}_1\mathbf{O}_2}$
  - All epipolar planes contain both the optical centres $\mathbf{O}_1$ and $\mathbf{O}_2$
  - All epipolar planes contains both the epipoles $\mathbf{e}_1$ and $\mathbf{e}_2$

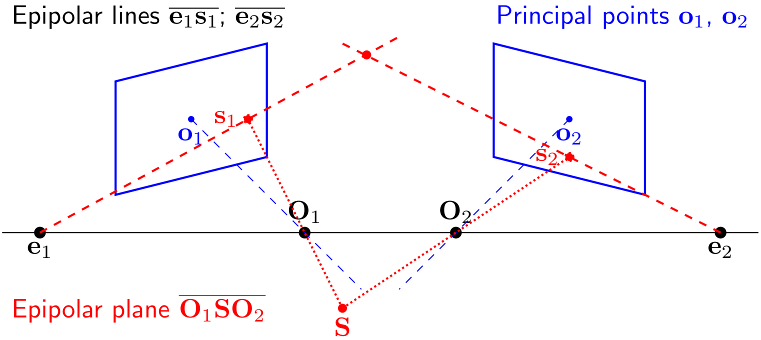
Corresponding (Conjugate) Epipolar Lines



Epipolar line through the projection of $\mathbf{S}$ on another image is the trace of intersection with the same epipolar plane containing $\mathbf{S}$ and the baseline

- Epipolar line $\overline{\mathbf{e}_j\mathbf{s}_j}$ conjugate on image j to the point $\mathbf{s}_i$ in image i is the projection onto the plane j of the optical ray producing $\mathbf{s}_i$
- Any 3D point on the epipolar plane $\overline{\mathbf{O}_1\mathbf{S}\mathbf{O}_2}$ is projected onto the conjugate epipolar lines $\overline{\mathbf{e}_1\mathbf{s}_1}$ and $\overline{\mathbf{e}_2\mathbf{s}_2}$

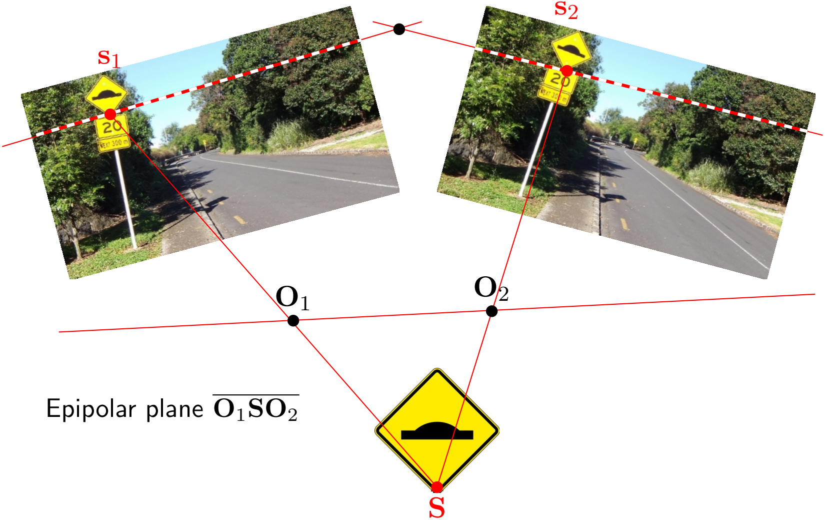
Epipolar Geometry



Conjugate epipolar lines $\overline{\mathbf{e}_1\mathbf{s}_1}$ and $\overline{\mathbf{e}_2\mathbf{s}_2}$ depict a visible 2D profile of a 3D scene in the intersecting epipolar plane $\overline{\mathbf{O}_1\mathbf{S}\mathbf{O}_2}$

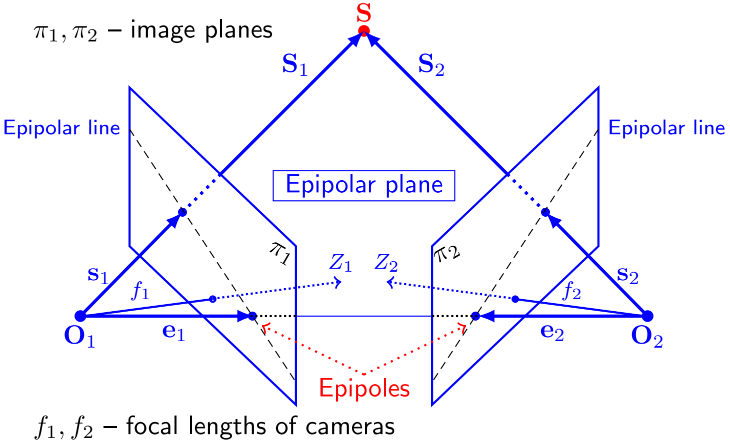
- $\mathbf{s}_1, \mathbf{s}_2$ – the projections of a 3D point $\mathbf{S}$
- $\mathbf{e}_1$, $\mathbf{e}_2$ – the **epipoles**, or projections of each optical centre (called a "pole"), $\mathbf{O}_2$ and $\mathbf{O}_1$, onto the other, i.e. opposite image plane

Epipolar Geometry



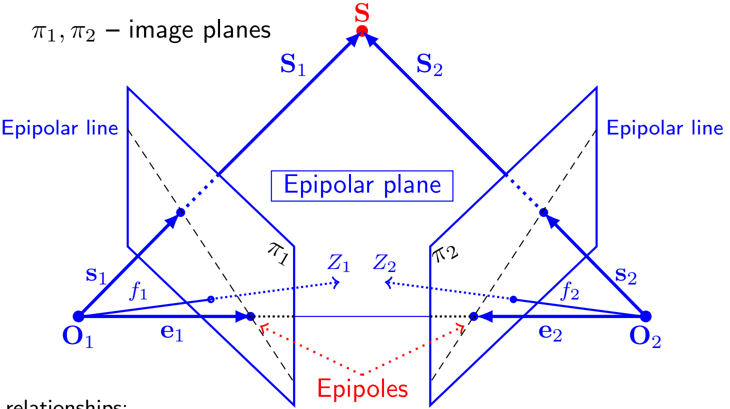
Epipolar plane $\overline{\mathbf{O}_1\mathbf{S}\mathbf{O}_2}$

Epipolar Geometry (image 1 – left; image 2 – right)



- $\mathbf{O}_1, \mathbf{O}_2$ – projection centres: the origins of the reference frames
- Z-axis of the 3D reference frame for each camera coincides with the optical axis of the camera

Epipolar Geometry

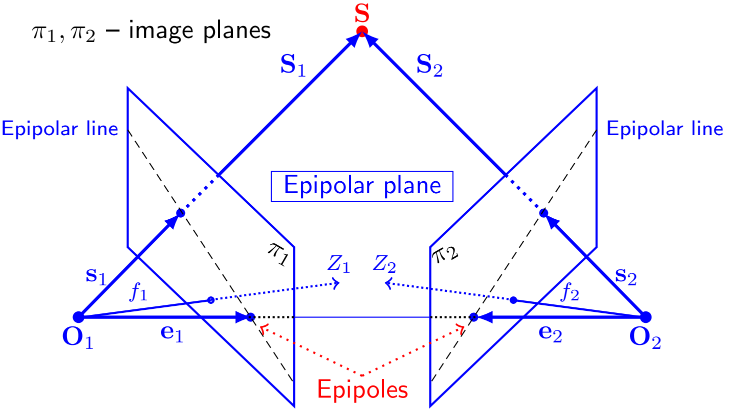


Basic relationships:

$$\mathbf{S}_1 = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}; \mathbf{S}_2 = \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \end{bmatrix}$$ – the same 3D point $\mathbf{S}$ in the reference frames

$$\mathbf{s}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 = f_1 \end{bmatrix}; \mathbf{s}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 = f_2 \end{bmatrix}$$ – projections of $\mathbf{S}$ onto the image planes

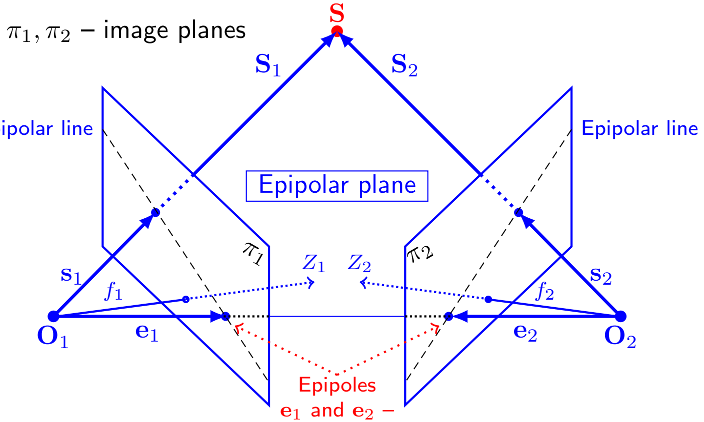
Reference Frames of the Left and Right Cameras



Related via **extrinsic parameters** of the stereo setup:
Given a 3-D point $\mathbf{S}$, a **rigid transformation** in 3D space between the coordinate vectors $\mathbf{S}_1$ and $\mathbf{S}_2$ in the reference frames is defined as $\mathbf{S}_2 = R(\mathbf{S}_1 - \mathbf{T})$

- $\mathbf{T} = \mathbf{O}_2 - \mathbf{O}_1$ – a translation vector
- R – 3×3 rotation matrix of the relative right-frame rotation

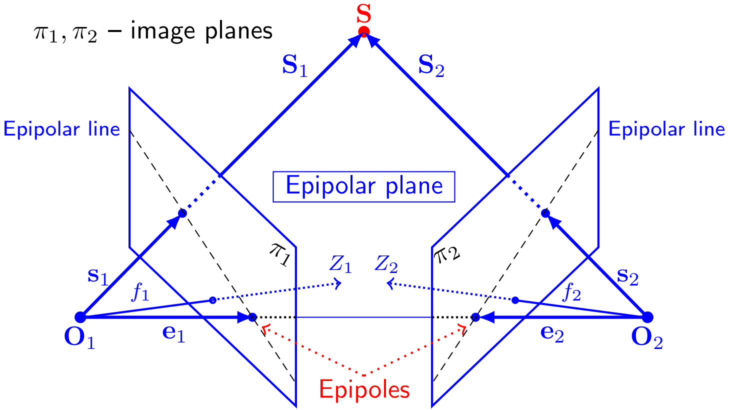
Reference Frames of the Left and Right Cameras



points at which the line through the projection centres intersects the images

- The left (right) epipole – the image of the right (left) projection centre
- **Canonical epipolar geometry**: the epipoles at infinity of the baseline
  - Projections $\mathbf{s}_1 = \frac{f_1}{Z_1}\mathbf{S}_1$ and $\mathbf{s}_2 = \frac{f_2}{Z_2}\mathbf{S}_2$ of a 3D point $\mathbf{S} = [X, Y, Z]^\mathsf{T}$

Epipolar Plane



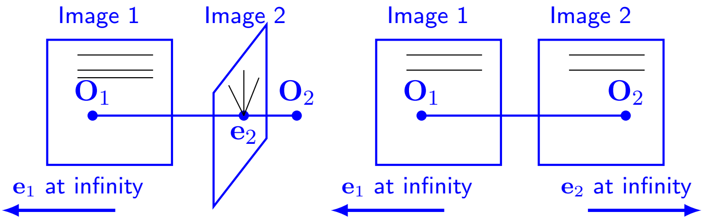
The plane through three 3D points: $\mathbf{S}$, $\mathbf{O}_1$, and $\mathbf{O}_2$

- **Epipolar line**: its intersection with each image plane
- **Conjugate lines**: both the epipolar lines for an epipolar plane

Epipolar Constraint

Given a point $\mathbf{s}_1$ (resp., $\mathbf{s}_2$) of the stereo image 1 (resp., 2), all the possible matches in the another image 2 (resp., 1) are sitting on the epipolar line through the epipole $\mathbf{e}_2$ (resp., $\mathbf{e}_1$)

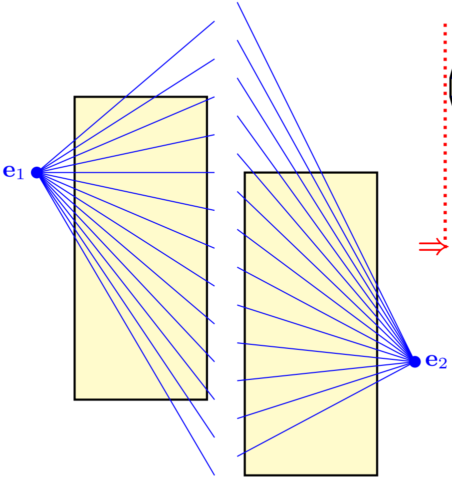
Parallel epipolar lines – in a special case of horizontal stereo pair:



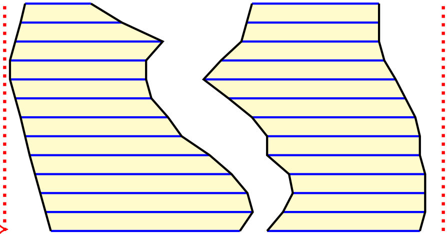
$\overline{\mathbf{O}_1\mathbf{O}_2}$ is parallel to Image 1

$\overline{\mathbf{O}_1\mathbf{O}_2}$ is parallel to both images

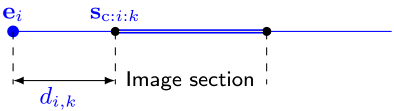
Epipolar Relations for Rectifying a Stereo Pair



Original stereo pair $(\mathbf{g}_1, \mathbf{g}_2)$

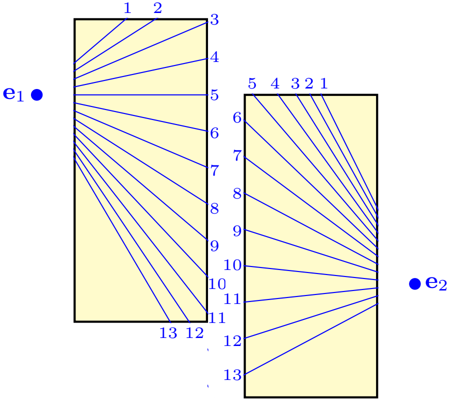


Rectified epipolar pair $(\widehat{\mathbf{g}}_1, \widehat{\mathbf{g}}_2)$

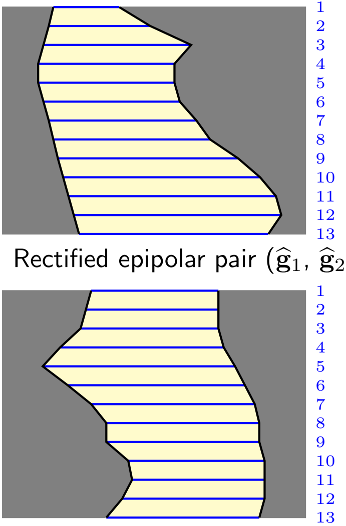


$d_{i:k}$ – the distance to the closest intersection point $\mathbf{s}_{\mathrm{c}:i:k}$ of an epipolar ray k with the frame of the image $\mathbf{g}_i$

Epipolar Relations for Rectifying a Stereo Pair



Original stereo pair $(\mathbf{g}_1, \mathbf{g}_2)$



Rectified epipolar pair $(\widehat{\mathbf{g}}_1, \widehat{\mathbf{g}}_2)$

Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

Given the 3×4 projection matrices $P_i = [Q_i \;\; \mathbf{q}_i]$; $i = 1, 2$, for 2 sensors, a 3D point $\mathbf{S}$ is projected to the corresponding image points $\widetilde{\mathbf{s}}_i = P_i \widetilde{\mathbf{S}}$

1. Find optical centres: $P_i \begin{bmatrix} \mathbf{O}_i \\ 1 \end{bmatrix} \equiv [Q_i \;\; \mathbf{q}_i] \begin{bmatrix} \mathbf{O}_i \\ 1 \end{bmatrix} = \widetilde{\mathbf{0}}$

   (i.e. the projected point with indefinite Cartesian coordinates $x, y, z = \frac{0}{0}$):
   $$Q_i \mathbf{O}_i + \mathbf{q}_i \cdot 1 = \mathbf{0} \;\Rightarrow\; \mathbf{O}_i = -Q_i^{-1} \mathbf{q}_i$$

2. Compute the epipoles (intersections of all epipolar lines) by projecting the optical centres $j \in \{1, 2\}$ ($i \in \{1, 2\}$; $j \neq i$):
   $$\widetilde{\mathbf{e}}_j = P_j \begin{bmatrix} \mathbf{O}_i \\ 1 \end{bmatrix} = [Q_j \;\; \mathbf{q}_j] \begin{bmatrix} -Q_i^{-1} \mathbf{q}_i \\ 1 \end{bmatrix} = -Q_j Q_i^{-1} \mathbf{q}_i + \mathbf{q}_j$$

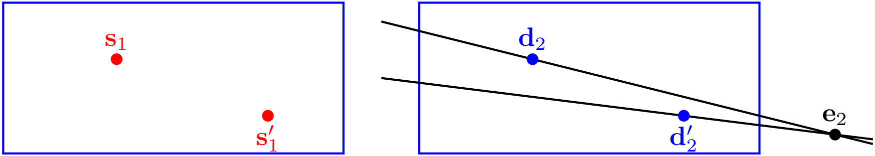
3. Find the point $\mathbf{D}_i$ at the infinity of the projecting ray $\overline{\mathbf{O}_i \mathbf{s}_i}$:
   $$P_i \begin{bmatrix} \mathbf{D}_i \\ 0 \end{bmatrix} = Q_i \mathbf{D}_i + \mathbf{q}_i \cdot 0 = \widetilde{\mathbf{s}}_i \;\Rightarrow\; \mathbf{D}_i = Q_i^{-1} \widetilde{\mathbf{s}}_i$$

4. Project $\mathbf{D}_i$ to the other image j: $\widetilde{\mathbf{d}}_j = P_j \begin{bmatrix} \mathbf{D}_i \\ 0 \end{bmatrix} = Q_j Q_i^{-1} \widetilde{\mathbf{s}}_i$

Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

As shown in Slide 32, a point $\mathbf{s}_i$ on one image, i, of a stereo pair and the projection matrices, $P_i = [Q_i \;\; \mathbf{q}_i]$; $i = 1, 2$, determines on the other image j; $j \neq i$, an epipolar line containing the point corresponding to $\mathbf{s}_i$

- This epipolar line is the projection of the optical ray producing $\mathbf{s}_i$
- This epipolar line is drawn through the epipole $\mathbf{e}_j$ and the projection $\mathbf{d}_j$ of the point $\mathbf{D}_i$ at infinity of the inverse optical ray $\overline{\mathbf{s}_i\mathbf{O}_i}$
  - The epipole $\widetilde{\mathbf{e}}_j = P_j \begin{bmatrix} \mathbf{O}_i \\ 1 \end{bmatrix} = [Q_j \;\; \mathbf{q}_j] \begin{bmatrix} -Q_i^{-1}\mathbf{q}_i \\ 1 \end{bmatrix}$
  - The projection $\widetilde{\mathbf{d}}_j = Q_j Q_i^{-1} \widetilde{\mathbf{s}}_i$
- The 2D epipolar line: $\mathbf{e}_j + \lambda\,(\mathbf{d}_j - \mathbf{e}_j)$; $\lambda \in (-\infty, \infty)$



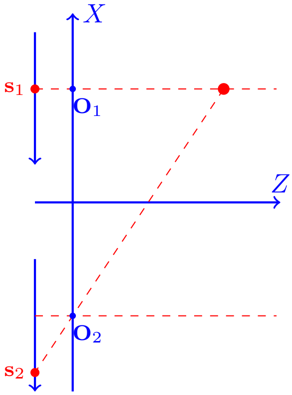
Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

Example: 2 cameras with the projection matrices $P_i = [Q_i \;\; \mathbf{q}_i]$; $i = 1, 2$

$$\begin{aligned}\widetilde{\mathbf{s}}_1 &= P_1\widetilde{\mathbf{S}} \\ &= \begin{bmatrix} 0.5 & 0 & 0 & -1.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}\widetilde{\mathbf{s}}_2 &= P_2\widetilde{\mathbf{S}} \\ &= \begin{bmatrix} 0.5 & 0 & 0 & 1.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}\end{aligned}$$

Image points $\begin{bmatrix} \frac{0}{2} \\ \frac{0}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} \frac{3}{2} \\ \frac{0}{2} \end{bmatrix} = \begin{bmatrix} 1.5 \\ 0 \end{bmatrix}$

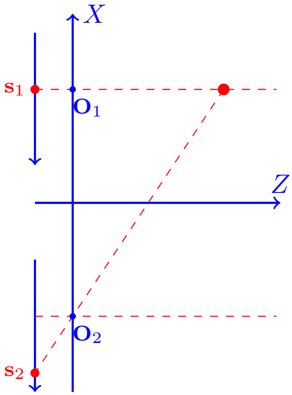


Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

Optical centres: $P_i \begin{bmatrix} \mathbf{O}_1 \\ 1 \end{bmatrix} = Q_i\mathbf{O}_i + \mathbf{q}_i \cdot 1 = \mathbf{0} \Rightarrow \mathbf{O}_i = -Q_i^{-1}\mathbf{q}_i$

$$\mathbf{O}_1 = -\begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1.5 \\ 0 \\ 0 \end{bmatrix} = -\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1.5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{O}_2 = -\begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1.5 \\ 0 \\ 0 \end{bmatrix} = -\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}$$



Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

- Epipoles $\widetilde{\mathbf{e}}_j = P_j \begin{bmatrix} \mathbf{O}_i \\ 1 \end{bmatrix} = P_j \begin{bmatrix} -Q_i\mathbf{q}_i \\ 1 \end{bmatrix}; \; j \neq i$
- For the example in Slides 34 and 35:

$$\widetilde{\mathbf{e}}_1 = P_1 \begin{bmatrix} -3 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 & -1.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} -3 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}$$

$$\widetilde{\mathbf{e}}_2 = P_2 \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 & 0 & 0 & 1.5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

Points at $\frac{-3}{0} = -\infty$ and $\frac{3}{0} = \infty$ along the x-axis

Epipolar Relations: Calibrated Cameras ($\widetilde{\ldots}$ – homogeneous coordinates)

- The point $\widetilde{\mathbf{D}}_i$ at infinity of the optical ray $\overline{\mathbf{O}_i\mathbf{s}_i}$ projecting to $\mathbf{s}_i$:

$$P_i\widetilde{\mathbf{D}}_i \equiv P_i \begin{bmatrix} \mathbf{D}_i \\ 0 \end{bmatrix} = Q_i\mathbf{D}_i + \mathbf{q}_i \cdot 0 = \mathbf{s}_i \;\Rightarrow\; \mathbf{D}_i = Q_i^{-1}\widetilde{\mathbf{s}}_i$$

- For the example in Slides 34 and 35:

$$\mathbf{D}_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

$$\mathbf{D}_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 2 \end{bmatrix}$$

- The projection $\widetilde{\mathbf{d}}_j$ of $\widetilde{\mathbf{D}}_i$; $j \neq i$, on the other image:

$$\widetilde{\mathbf{d}}_j = P_j \begin{bmatrix} \mathbf{D}_i \\ 0 \end{bmatrix} = Q_j Q_i^{-1}\widetilde{\mathbf{s}}_i$$

- For the example in Slides 34 and 35, $Q_1 = Q_2$ implying $\mathbf{d}_j = \mathbf{s}_i$

Epipolar Lines: A Summary

Given a single image point $\mathbf{s}_i$, the original 3D point $\mathbf{S}$ can sit anywhere on the inverse projecting ray $\overline{\mathbf{s}_i\mathbf{O}_i}$

- On the other image, this ray is projected onto the epipolar line, which goes through the corresponding image point $\mathbf{s}_j$; $j \neq i$
- **Epipolar constraint**: the true match sits on the epipolar line
- All the epipolar lines in an image go through the epipole
- With the exception of the epipole, only one epipolar line goes through any image point

Only 1D search region due to mapping between the points in one image and corresponding epipolar lines in the other image

- Corresponding points are on the conjugate epipolar lines
- Helps in rejecting false matches due to occlusions

An obvious goal: Estimate the epipolar geometry, i.e. determine the point-to-line mapping, for a stereo pair from uncalibrated cameras.

Vector Cross Product: A Math Prompt

Cross product $\mathbf{x} \times \mathbf{y}$ of 3D vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$:

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_2y_3 - x_3y_2 \\ -x_1y_3 + x_3y_1 \\ x_1y_2 - x_2y_1 \end{bmatrix} \equiv \overbrace{\begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}}^{\text{matrix } S_{\mathbf{x}} \text{ of rank 2}} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$\equiv [x_1 \;\; x_2 \;\; x_3] \underbrace{\begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix}}_{\text{matrix } S_{\mathbf{y}} \text{ of rank 2}}$$

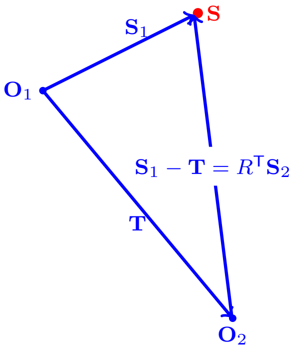
Vector $\mathbf{z} = \mathbf{x} \times \mathbf{y}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$: $\mathbf{x}^\mathsf{T}\mathbf{z} = \mathbf{y}^\mathsf{T}\mathbf{z} = 0$

See also: http://en.wikipedia.org/wiki/Cross_product

The Essential Matrix, E

Determines the mapping between the points in one image and epipolar lines in the other image:

- Equation of the epipolar plane through a 3D point $\mathbf{S}$ from the **co-planarity** of the vectors $\mathbf{S}_1$, $\mathbf{T}$, and $\mathbf{S}_1 - \mathbf{T} = R^\mathsf{T}\mathbf{S}_2$:



$$(\mathbf{S}_1 - \mathbf{T})^\mathsf{T}(\mathbf{T}\times\mathbf{S}_1) = 0$$

$$\Rightarrow \quad \left(R^\mathsf{T}\mathbf{S}_2\right)^\mathsf{T}(\mathbf{T}\times\mathbf{S}_1) = 0$$

$$\Rightarrow \quad \mathbf{S}_2^\mathsf{T}R\,(\mathbf{T}\times\mathbf{S}_1) = 0$$

$$\Rightarrow \quad \mathbf{S}_2^\mathsf{T}\underbrace{R\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}}_{S\mathbf{S}_1 \equiv \mathbf{T}\times\mathbf{S}_1}\mathbf{S}_1 = 0$$

$$\Rightarrow \quad \mathbf{S}_2^\mathsf{T}\underbrace{(RS)}_{E}\mathbf{S}_1 = 0 \Rightarrow \mathbf{S}_2^\mathsf{T}E\mathbf{S}_1 = 0$$

Rank 2 matrix S · Full-rank R $\rightarrow$ Essential matrix $E = RS$ of rank 2

The Essential Matrix, E

- By construction, the matrix S (and thus E) is of rank 2

$$\mathbf{T} \times \mathbf{S}_1 \equiv \overbrace{\begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}}^{\text{matrix } S} \mathbf{S}_1$$

- Essential matrix $E = RS$ links naturally the epipolar constraint, $\mathbf{S}_2^\mathsf{T} E \mathbf{S}_1 = 0$, and the extrinsic parameters of the stereo cameras:

$$\mathbf{S}_1 = \frac{Z_1}{f_1}\mathbf{s}_1;\ \mathbf{S}_2 = \frac{Z_2}{f_2}\mathbf{s}_2 \Rightarrow \frac{Z_1 Z_2}{f_1 f_2}\mathbf{s}_2^\mathsf{T} E \mathbf{s}_1 = 0 \Rightarrow \mathbf{s}_2^\mathsf{T} E \mathbf{s}_1 = 0$$

- Mapping between the points and epipolar lines
  - Vector $\mathbf{a}_2 = E\mathbf{s}_1$ – parameters of the epipolar line $\mathbf{s}_2^\mathsf{T}\mathbf{a}_2 = 0$ in the right image corresponding to the point $\mathbf{s}_1$ in the left image
  - Vector $\mathbf{a}_1^\mathsf{T} = \mathbf{s}_2^\mathsf{T} E$ – parameters of the epipolar line $\mathbf{a}_1^\mathsf{T}\mathbf{s}_1 = 0$ in the left image corresponding to the point $\mathbf{s}_2$ in the right image

The Fundamental Matrix, F

- No prior information about the stereo system: the unknown matrices M_1 and M_2 of the intrinsic camera parameters
- Mapping of points to epipolar lines can be obtained from the corresponding points only

Points $\widetilde{\mathbf{s}}_1$ and $\widetilde{\mathbf{s}}_2$ in pixel coordinates vs. the same points $\mathbf{s}_1$ and $\mathbf{s}_2$ in camera coordinates:

$$\widetilde{\mathbf{s}}_1 \equiv \begin{bmatrix} \widetilde{x}_1 \\ \widetilde{y}_1 \\ 1 \end{bmatrix} = M_1\mathbf{s}_1;\ \widetilde{\mathbf{s}}_2 \equiv \begin{bmatrix} \widetilde{x}_2 \\ \widetilde{y}_2 \\ 1 \end{bmatrix} = M_2\mathbf{s}_2$$

$$\Leftrightarrow \quad \mathbf{s}_1 = M_1^{-1}\widetilde{\mathbf{s}}_1;\ \mathbf{s}_2 = M_2^{-1}\widetilde{\mathbf{s}}_2$$

$$\Leftrightarrow \quad \widetilde{\mathbf{s}}_2^{\mathsf{T}} \underbrace{M_2^{-1} E M_1^{-1}}_{\text{Fundamental matrix } F} \widetilde{\mathbf{s}}_1 = \widetilde{\mathbf{s}}_2^{\mathsf{T}} F \widetilde{\mathbf{s}}_1$$

The Fundamental Matrix, F

Mapping pixels to the epipolar lines:

- Vector $\mathbf{a}_2 = F\widetilde{\mathbf{s}}_1$ – parameters of the epipolar line $\widetilde{\mathbf{s}}_2^\mathsf{T}\mathbf{a}_2 = 0$ in the right image corresponding to the pixel $\widetilde{\mathbf{s}}_1$ in the left image
- Vector $\mathbf{a}_1^\mathsf{T} = \widetilde{\mathbf{s}}_2^\mathsf{T}F$ – parameters of the epipolar line $\mathbf{a}_1^\mathsf{T}\widetilde{\mathbf{s}}_1 = 0$ in the left image corresponding to the pixel $\widetilde{\mathbf{s}}_2$ in the right image
- Just as the essential matrix E, the fundamental matrix F has rank 2
- The fundamental matrix F takes account of both the intrinsic and extrinsic parameters of the stereo system

The epipolar constraint can be established without prior knowledge of the stereo cameras parameters!

The Fundamental Matrix (here, $\widetilde{\ldots}$ – homogeneous coordinates)

An alternative derivation from optical rays projected onto image planes:

- Any point $\mathbf{s}_i$ of the image $i = 1$ or $i = 2$ is produced by an own optical ray
- Projecting that optical ray onto the other image j; $j \neq i$, forms the epipolar line containing the point, which corresponds to $\mathbf{s}_j$
- The epipolar line goes via the relevant epipole and the projection of the infinitely far 3D point of the ray
- Epipoles: $\widetilde{\mathbf{e}}_2 = -Q_2 Q_1^{-1}\mathbf{q}_1 + \mathbf{q}_2$ and $\widetilde{\mathbf{e}}_1 = -Q_1 Q_2^{-1}\mathbf{q}_2 + \mathbf{q}_1$
- Image points obtained by projecting the infinitely far 3D poinrs of the optical rays: $\widetilde{\mathbf{d}}_2 = Q_2 Q_1^{-1}\widetilde{\mathbf{s}}_1$ and $\widetilde{\mathbf{d}}_1 = Q_1 Q_2^{-1}\widetilde{\mathbf{s}}_2$
- A line via an arbitrary point $\widetilde{\mathbf{s}}$ in an image plane:

$$[\alpha_1 \;\; \alpha_2 \;\; \alpha_3] \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} \equiv \alpha_1 x_s + \alpha_2 y_s + \alpha_3 = 0, \text{ or } \boldsymbol{\alpha}^\mathsf{T}\widetilde{\mathbf{s}} = 0$$

The Fundamental Matrix (here, $\widetilde{\ldots}$ – homogeneous coordinates)

The epipolar line, $\boldsymbol{\alpha}_i^\mathsf{T}\widetilde{\mathbf{s}} = 0$, in the plane of the image i crosses the points $\widetilde{\mathbf{e}}_i$ and $\widetilde{\mathbf{d}}_i$, so that the relations $\boldsymbol{\alpha}_i^\mathsf{T}\widetilde{\mathbf{e}}_i = \boldsymbol{\alpha}_i^\mathsf{T}\widetilde{\mathbf{d}}_i = 0$ hold

- The line parameters, $\boldsymbol{\alpha}_i$, are specified by the cross product

$$\widetilde{\mathbf{e}}_i \times \widetilde{\mathbf{d}}_i = \underbrace{\widetilde{\mathbf{e}}_i \times Q_i Q_j^{-1}\widetilde{\mathbf{s}}_j}_{U_{\mathrm{ep}:i} Q_i Q_j^{-1}\widetilde{\mathbf{s}}_j}; i \in \{1,2\};\quad j \neq i$$

  - The 3×3 rank-2 matrix $U_{\mathrm{ep}:i}$ is built from $\mathbf{e}_i$ as in Slide 39

It can be shown that the epipolar line parameters for the points $\widetilde{\mathbf{s}}_1$ and $\widetilde{\mathbf{s}}_2$ are given by the vectors

$$\begin{cases} \boldsymbol{\alpha}_1 & = & \widetilde{\mathbf{e}}_1 \times \widetilde{\mathbf{d}}_1 & \equiv & U_{\mathrm{ep}:1} Q_1 Q_2^{-1}\widetilde{\mathbf{s}}_2 & = & F^\mathsf{T}\widetilde{\mathbf{s}}_2 \\ \boldsymbol{\alpha}_2 & = & \widetilde{\mathbf{e}}_2 \times \widetilde{\mathbf{d}}_2 & \equiv & U_{\mathrm{ep}:2} Q_2 Q_1^{-1}\widetilde{\mathbf{s}}_1 & = & F\widetilde{\mathbf{s}}_1 \end{cases}$$

where F is a fixed 3×3 *fundamental matrix* of rank 2

- Any pixel $\widetilde{\mathbf{s}}_1$ (resp., $\widetilde{\mathbf{s}}_2$) on the epipolar line for $\widetilde{\mathbf{s}}_2$ (resp., $\widetilde{\mathbf{s}}_1$) satisfies the **Longuet-Higgins equation**: $\widetilde{\mathbf{s}}_2^\mathsf{T} F\widetilde{\mathbf{s}}_1 = 0$

The Fundamental Matrix, F

The fundamental matrix relationship $\widetilde{\mathbf{s}}_{k,2}^{\mathsf{T}} F \widetilde{\mathbf{s}}_{k,1} = 0$ holds for any pair of corresponding points:

$$\left\{ \widetilde{\mathbf{s}}_{k,j} = \begin{bmatrix} x_{k,j} \\ y_{k,j} \\ 1 \end{bmatrix};\ j = 1, 2 \right\};\ k = 1, \ldots, n$$

Meaning of the relationship

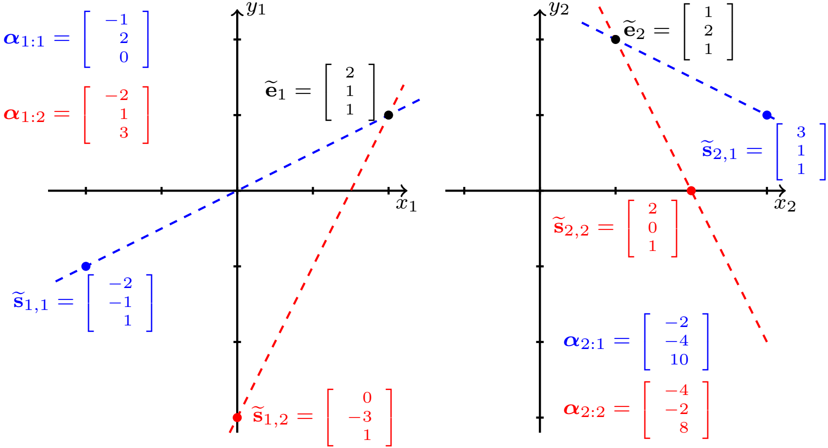
Any point $\mathbf{s}_{k,2}$ of the right image specifies in the left image a unique epipolar line with parameters $\widetilde{\mathbf{s}}_{k,2}^{\mathsf{T}} F$, such that it goes through the corresponding point $\mathbf{s}_{k,1}$

Alternatively, the point $\mathbf{s}_{k,1}$ of the left image specifies in the right image a unique corresponding epipolar line with parameters $F\widetilde{\mathbf{s}}_{k,1}$, such that it goes through the corresponding point $\mathbf{s}_{k,2}$

Conjugate Epipolar Lines for Corresponding Pixels

$$F = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -2 \\ -2 & -1 & 5 \end{bmatrix}: \text{ Lines } \begin{cases} \boldsymbol{\alpha}^\mathsf{T}_{1:k} = [y_{2:k} - 2,\ x_{2:k} - 1,\ -x_{2:k} - 2y_{2:k} + 5] \\ \boldsymbol{\alpha}^\mathsf{T}_{2:k} = [y_{1:k} - 1,\ x_{1:k} - 2,\ -2x_{1:k} - y_{1:k} + 5] \end{cases}$$

for the corresponding pixels $\mathbf{s}_{j:k}$, $j = 1, 2$; $k = 1, 2$:



Normalising the Fundamental Matrix

- Components of F are to be normalised to exclude the singular solution $F = 0$
- Canonical epipolar geometry for a stereopair (the epipolar lines $y_1 = y_2 = y$ are parallel to the x-axis of the images) has the fundamental matrix:

$$F \equiv \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} = \left[\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & 0 & \frac{1}{\sqrt{2}} \\ \hline 0 & -\frac{1}{\sqrt{2}} & 0 \end{array}\right)$$

and the epipoles $\mathbf{e}_1 = [\infty, c]$ and $\mathbf{e}_2 = [-\infty, c]$ with an arbitrary constant c

  - The normalization has to account for all the components f_{kl} of F; $k, l \in \{1, 2, 3\}$ (except of f_{33}), which cannot be equal to zero simultaneously

Estimating the Matrix F: The Eight-Point Algorithm

Given $n \geq 8$ known corresponding points' pairs in stereo images, **find** F

Each i-th correspondence results in the equation:

$$\tilde{\mathbf{s}}_{2:i}^{\mathsf{T}} F \tilde{\mathbf{s}}_{1:i} = 0 \;\Rightarrow\; [\tilde{x}_{2:i}\ \tilde{y}_{2:i}\ 1] \underbrace{\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}}_{\text{Fundamental matrix } F} \begin{bmatrix} \tilde{x}_{1:i} \\ \tilde{y}_{1:i} \\ 1 \end{bmatrix} = 0$$

i.e. the homogeneous linear equation $\mathbf{a}_i^{\mathsf{T}}\mathbf{f} = 0$ where

$$\begin{aligned} \mathbf{a}_i &= \left[\; \tilde{x}_{2:i}\tilde{x}_{1:i} \quad \tilde{x}_{2:i}\tilde{y}_{1:i} \quad \tilde{x}_{2:i} \quad \tilde{y}_{2:i}\tilde{x}_{1:i} \quad \tilde{y}_{2:i}\tilde{y}_{1:i} \quad \tilde{y}_{2:i} \quad \tilde{x}_{1:i} \quad \tilde{y}_{1:i} \quad 1 \;\right]^{\mathsf{T}} \\ \mathbf{f} &= \left[\; f_{11} \quad f_{12} \quad f_{13} \quad f_{21} \quad f_{22} \quad f_{23} \quad f_{31} \quad f_{32} \quad f_{33} \;\right]^{\mathsf{T}} \end{aligned}$$

If the n points do not form a degenerate configuration, the 9 entries of F are given by the non-trivial solution of the over-determined homogeneous linear system $A\mathbf{f} \equiv \begin{bmatrix} \mathbf{a}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{a}_n^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} f_{11} \\ \vdots \\ f_{33} \end{bmatrix} = \mathbf{0}$ with the $n \times 9$ matrix A

The Eight-Point Algorithm

Since the system $A\mathbf{f} = 0$ is homogeneous, the solution is unique up to a signed scaling factor

- Typically, $n > 8$, i.e. the system is over-determined
- The solution: the column of the matrix $V = [\mathbf{v}_1 \ \ldots \ \mathbf{v}_9]$ corresponding to the only null singular value in the SVD (singular value decomposition) $A = UDV^\mathsf{T}$
  - The columns $\mathbf{v}_1, \ldots, \mathbf{v}_9$ are the eigenvectors of the 9×9 matrix $A^\mathsf{T}A$
  - Due to noise, the solution is the column of V associated with the **least singular value**

Estimated fundamental matrix F_{est} is almost always non-singular (i.e. it is of the full rank 3 rather than the expected rank 2)

The SVD: Math Prompt (for those who forgot the SVD)

Any rectangular $n \times k$ matrix A ($n \geq k$) can be decomposed into the product of three matrices: $A = UDV^\mathsf{T}$

U – the column-orthonormal $n \times k$ matrix with the columns being mutually orthogonal unit vectors

- The columns of U are the top k eigenvectors $\mathbf{u}_j$; $j = 1, \ldots, k$, of the $n \times n$ matrix AA^T

V – the column-orthonormal $k \times k$ matrix with the columns being mutually orthogonal unit vectors

- The columns of V are the eigenvectors $\mathbf{v}_i$; $i = 1, \ldots, k$, of the $k \times k$ matrix $A^\mathsf{T}A$

D – the diagonal $k \times k$ matrix with non-negative diagonal elements σ_i called singular values: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k \geq 0$

SVD: Math Prompt

- The matrices U and V are not unique, but the singular values are fully determined by the matrix A
- The number of non-zero singular values equals the rank of A (if $n \geq k$, the rank of A is equal to or less than k)
- Basic properties of the SVD:
    - $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ and $A^\mathsf{T}\mathbf{u}_i = \sigma_i \mathbf{v}_i$
    - $A^\mathsf{T} A\mathbf{v}_i = \sigma_i A^\mathsf{T}\mathbf{u}_i = \sigma_i^2 \mathbf{v}_i$
    - $AA^\mathsf{T}\mathbf{u}_i = \sigma_i A\mathbf{v}_i = \sigma_i^2 \mathbf{u}_i$
    - The squared singular values are the eigen-values of both the $n \times n$ matrix AA^T and $k \times k$ matrix $A^\mathsf{T} A$

- An alternative definition of the SVD: the $n \times n$ matrix U, $n \times k$ matrix D, and $k \times k$ matrix V
    - A larger memory space for the matrices: $n^2 + k + k^2$ rather than $nk + k + k^2$ for the original definition as typically $n \gg k$

The 9×9 Matrix $A^\mathsf{T} A$

Simplifying notation: $X_2^\alpha Y_2^\gamma X_1^\beta Y_1^\delta = \sum_{i=1}^{n} x_{2:i}^\alpha y_{2:i}^\gamma x_{1:i}^\beta y_{1:i}^\delta$

$A^\mathsf{T} A =$

$$\begin{bmatrix}
X_2^2X_1^2 & X_2^2X_1Y_1 & X_2^2X_1 & X_2Y_2X_1^2 & X_2Y_2X_1Y_1 & X_2Y_2X_1 & X_2X_1^2 & X_2X_1Y_1 & X_2X_1 \\
X_2^2X_1Y_1 & X_2^2Y_1^2 & X_2^2Y_1 & X_2Y_2X_1Y_1 & X_2Y_2Y_1^2 & X_2Y_2Y_1 & X_2X_1Y_1 & X_2Y_1^2 & X_2Y_1 \\
X_2^2X_1 & X_2^2Y_1 & X_2^2 & X_2Y_2X_1 & X_2Y_2Y_1 & X_2Y_2 & X_2X_1 & X_2Y_1 & X_2 \\
X_2Y_2X_1^2 & X_2Y_2X_1Y_1 & X_2Y_2X_1 & Y_2^2X_1^2 & Y_2^2X_1Y_1 & Y_2^2X_1 & Y_2X_1^2 & Y_2X_1Y_1 & Y_2X_1 \\
X_2Y_2X_1Y_1 & X_2Y_2Y_1^2 & X_2Y_2Y_1 & Y_2^2X_1Y_1 & Y_2^2Y_1^2 & Y_2^2Y_1 & Y_2X_1Y_1 & Y_2Y_1^2 & Y_2Y_1 \\
X_2Y_2X_1 & X_2Y_2Y_1 & X_2Y_2 & Y_2^2X_1 & Y_2^2Y_1 & Y_2^2 & Y_2X_1 & Y_2Y_1 & Y_2 \\
X_2X_1^2 & X_2X_1Y_1 & X_2X_1 & Y_2^2Y_1 & Y_2X_1Y_1 & Y_2X_1 & X_1^2 & X_1Y_1 & X_1 \\
X_2X_1Y_1 & X_2Y_1^2 & X_2Y_1 & Y_2X_1Y_1 & Y_2Y_1^2 & Y_2Y_1 & X_1Y_1 & Y_1^2 & Y_1 \\
X_2X_1 & X_2Y_1 & X_2 & Y_2X_1 & Y_2Y_1 & Y_2 & X_1 & Y_1 & n
\end{bmatrix}$$

The Eight-Point Algorithm

The singularity of F is enforced by adjusting the entries of F_{est}:

1. The SVD $F_{\text{est}} = U_\circ D_\circ V_\circ^\mathsf{T}$
2. Set the smallest singular value in the diagonal matrix $D_\circ$ to zero in order to obtain the corrected matrix $D_\circ^+$
3. The corrected estimate: $F_{\text{est}}^+ = U_\circ D_\circ^+ V_\circ^\mathsf{T}$

To escape numerical instabilities: make comparable values of entries of A by normalising coordinates of the corresponding points

- Translate both the coordinates of each point to the centroid of each data set: $m_\text{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$; $m_\text{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
- Scale the norm of each point to make the unit average norm over the data set: $c = \frac{n\sqrt{2}}{\sum_{i=1}^{n}\sqrt{(x_i - m_\text{x})^2 + (y_i - m_\text{y})^2}}$:

$$\mathbf{s}_i' = H\mathbf{s}_i \equiv \begin{bmatrix} c & 0 & -m_\text{x}c \\ 0 & c & -m_\text{y}c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} c(x_i - m_\text{x}) \\ c(y_i - m_\text{y}) \\ 1 \end{bmatrix}$$

The Stable Eight-Point Algorithm

Input: n pixel-to-pixel correspondences:

$$\left\{\left(\widetilde{\mathbf{s}}_{1:i} = [x_{1:i}, y_{1:i}, 1]^\mathsf{T};\ \widetilde{\mathbf{s}}_{2:i} = [x_{2:i}, y_{2:i}, 1]^\mathsf{T}\right) : i = 1, \ldots, n\right\}$$

Data normalisation: $\{\widetilde{\mathbf{s}}'_{\alpha:i} = H_\alpha \widetilde{\mathbf{s}}_{\alpha:i} :\ \alpha \in \{1,2\};\ i = 1, \ldots, n\}$ where

$$H_1 = \begin{bmatrix} c_1 & 0 & -c_1 m_{\mathrm{x}:1} \\ 0 & c_1 & -c_1 m_{\mathrm{y}:1} \\ 0 & 0 & 1 \end{bmatrix};\ H_2 = \begin{bmatrix} c_2 & 0 & -c_2 m_{\mathrm{x}:2} \\ 0 & c_2 & -c_2 m_{\mathrm{y}:2} \\ 0 & 0 & 1 \end{bmatrix}$$

Data restoration $\{\widetilde{\mathbf{s}}_{\alpha:i} = H_\alpha^{-1} \widetilde{\mathbf{s}}'_{\alpha:i} :\ \alpha \in \{1,2\};\ i = 1, \ldots, n\}$

$$H_1^{-1} = \begin{bmatrix} \frac{1}{c_1} & 0 & m_{\mathrm{x}:1} \\ 0 & \frac{1}{c_1} & m_{\mathrm{y}:1} \\ 0 & 0 & 1 \end{bmatrix};\ H_2^{-1} = \begin{bmatrix} \frac{1}{c_2} & 0 & m_{\mathrm{x}:2} \\ 0 & \frac{1}{c_2} & m_{\mathrm{y}:2} \\ 0 & 0 & 1 \end{bmatrix}$$

SVD $A = UDV^\mathsf{T}$ of the $n \times 9$ matrix A for the homogeneous system of n linear equations $A\mathbf{F} = 0$; $n \geq 8$ (overdetermined system for $n > 8$)

The Stable Eight-Point Algorithm

- The SVD solution $\mathbf{f}^\circ = \begin{bmatrix} f^\circ_{11} \\ f^\circ_{12} \\ \vdots \\ f^\circ_{33} \end{bmatrix}$ (up to an unknown signed scale factor) by using the column of V corresponding to the least singular value of A
- The SVD of the 3×3 matrix $F^\circ_{\text{est}} = U_\circ D_\circ V_\circ^\mathsf{T}$ in order to enforce the singularity (i.e. rank 2) constraint of the fundamental matrix
  - Correct the matrix $D_\circ$ by setting the least singular value in its main diagonal equal to 0
  - Use the corrected matrix $D^+_\circ$ to compute the corrected estimate $F^+_{\text{est}} = U_\circ D^+_\circ V_\circ^\mathsf{T}$ of the fundamental matrix
- **Renormalised** output estimate $F_{\text{est}} = \left(H_2^{-1}\right)^\mathsf{T} F^+_{\text{est}} H_1^{-1}$

Locating the Epipoles

Accurate localisation of the epipoles:

- To refine the locations of the conjugate epipolar lines
- To simplify stereo geometry
- To recover 3D structure in the case of uncalibrated stereo

The left epipole $\mathbf{e}_1$ sits on all epipolar lines in the left image:

- The relationship $\mathbf{s}_2^\mathsf{T} F \mathbf{e}_1 = 0$ holds for every $\mathbf{s}_2$
- F is not identically zero, so it follows that $F\mathbf{e}_1 = 0$
- F has rank 2, so the epipole is the null space of F
  - Null space is the set of all solutions $\mathbf{s}$ to the equation $F\mathbf{s} = 0$
- Similarly, $\mathbf{e}_2$ is the null space of F^T

Algorithm: SVD $F = UDV^\mathsf{T}$

- $\mathbf{e}_1$ – the column of V corresponding to the null singular value
- $\mathbf{e}_2$ – the column of U corresponding to the null singular value

Locating the Epipoles: An Example

Canonical stereo geometry:

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \Rightarrow \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{FF^\top (= F^\top F) \Rightarrow |FF^\top - \lambda I| = -\lambda(1-\lambda)^2 = 0}$$

$$\Rightarrow \lambda_{1,2} = 1;\ \mathbf{e}_{1,2}^\top = [0\ \pm 1\ 0];\ \mathbf{e}_{2,1}^\top = [0\ 0\ \pm 1];\ \lambda_3 = 0;\ \mathbf{e}_3^\top = [\pm 1\ 0\ 0]$$

$$F = \underbrace{\begin{bmatrix} 0 & 0 & \mathbf{1} \\ 0 & 1 & \mathbf{0} \\ 1 & 0 & \mathbf{0} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \mathbf{0} \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -\mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{V^\top}$$

$$\Rightarrow \mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix};\ \mathbf{e}_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}$$

i.e. x-coordinates of both the epipoles are at $\pm$infinity and y-coordinates are indefinite

Distance to an Epipolar Line

Squared Cartesian distance between a point $\widetilde{\mathbf{s}}^\circ = \begin{bmatrix} x^\circ \\ y^\circ \\ 1 \end{bmatrix}$ and an epipolar line $\alpha x + \beta y + \gamma = 0$ with coefficients $\mathbf{c} = (\alpha, \beta, \gamma)$:

$$d^2 = \min_{x,y} \left\{ (x^\circ - x)^2 + (y^\circ - y)^2 \mid \alpha x + \beta y + \gamma = 0 \right\}$$

Constrained Lagrange optimisation:

$$d^2 = \frac{(\alpha x^\circ + \beta y^\circ + \gamma)^2}{\alpha^2 + \beta^2}$$

Distance: $d = \frac{1}{\sqrt{\alpha^2+\beta^2}} \left| \alpha x^\circ + \beta y^\circ + \gamma \right|$

Signed distance: $d = \frac{1}{\sqrt{\alpha^2+\beta^2}} \left(\alpha x^\circ + \beta y^\circ + \gamma \right)$

Distance to an Epipolar Line

Coefficients of the epipolar lines from the relation $\widetilde{\mathbf{s}}_{2:i}^{\mathsf{T}} F \widetilde{\mathbf{s}}_{1:i} = 0$:

1. Epipolar line on image 1:
   $\mathbf{c}_1 = (f_{11}x_{2:i} + f_{21}y_{2:i} + f_{31},\ f_{12}x_{2:i} + f_{22}y_{2:i} + f_{32},\ f_{13}x_{2:i} + f_{23}y_{2:i} + f_{33})$
2. Epipolar line on image 2:
   $\mathbf{c}_2 = (f_{11}x_{1:i} + f_{12}y_{1:i} + f_{13},\ f_{21}x_{1:i} + f_{22}y_{2:i} + f_{23},\ f_{31}x_{2:i} + f_{32}y_{2:i} + f_{33})$

Squared distances between the corresponding points and related conjugate epipolar lines:

$$d_{1:i}^2 = \frac{\left(\widetilde{\mathbf{s}}_{2:i}^{\mathsf{T}} F \widetilde{\mathbf{s}}_{1:i}\right)^2}{(f_{11}x_{2:i}+f_{21}y_{2:i}+f_{31})^2+(f_{12}x_{2:i}+f_{22}y_{2:i}+f_{32})^2}$$

$$d_{2:i}^2 = \frac{\left(\widetilde{\mathbf{s}}_{2:i}^{\mathsf{T}} F \widetilde{\mathbf{s}}_{1:i}\right)^2}{(f_{11}x_{1:i}+f_{12}y_{1:i}+f_{13})^2+(f_{21}x_{1:i}+f_{22}y_{1:i}+f_{23})^2}$$

where $\widetilde{\mathbf{s}}_{2:i}^{\mathsf{T}} F \widetilde{\mathbf{s}}_{1:i} = \mathbf{a}_i^{\mathsf{T}} \mathbf{f}$ (see Slide 49).

Distance to an Epipolar Line

- The 8-point algorithm accounts for only the nominators of the squared distances between the points and epipolar lines.
- The nominators are quadratic forms with the 9×9 matrices $A_i = \mathbf{a}_i \mathbf{a}_i^\mathsf{T}$ of the corresponding coordinates and their products: $\left(\mathbf{a}_i^\mathsf{T}\mathbf{f}\right)^2 = \mathbf{f}^\mathsf{T} A_i \mathbf{f}$.
- The denominators are also the quadratic forms:

$$
\begin{aligned}
d_{1:i}^2 &= \frac{\mathbf{f}^\mathsf{T} A_i \mathbf{f}}{(f_{11}x_{2:i}+f_{21}y_{2:i}+f_{31})^2+(f_{12}x_{2:i}+f_{22}y_{2:i}+f_{32})^2} = \frac{\mathbf{f} A_i \mathbf{f}}{\mathbf{f} B_{1:i} \mathbf{f}} \\
d_{2:i}^2 &= \frac{\mathbf{f}^\mathsf{T} A_i \mathbf{f}}{(f_{11}x_{1:i}+f_{12}y_{1:i}+f_{13})^2+(f_{21}x_{1:i}+f_{22}y_{1:i}+f_{23})^2} = \frac{\mathbf{f} A_i \mathbf{f}}{\mathbf{f} B_{2:i} \mathbf{f}}
\end{aligned}
$$

- Components of the 9×9 matrices $B_{1:i}$ and $B_{2:i}$ depend on the coordinates $\widetilde{\mathbf{s}}_2$ and $\widetilde{\mathbf{s}}_1$, respectively.

Distance to an Epipolar Line

- Minimising the total sum of the nominators:

$$\mathbf{f}^* = \arg\min_{|\mathbf{f}|=1} \left\{\mathbf{f}^\mathsf{T} A\mathbf{f}\right\}$$

  with $A = \sum_{i=1}^{n} A_i$ does not guarantee the minimal mean distance of the corresponding points from their epipolar lines.

- The constrained minimisation, making the total sum of the denominators, $\mathbf{f}^\mathsf{T} B\mathbf{f}$, equal to the unit value may result in a more accurate and noise-resistant 8-point algorithm:

$$\mathbf{f}^* = \arg\min_{|\mathbf{f}|=1} \left\{\mathbf{f}^\mathsf{T} A\mathbf{f} :\; \mathbf{f}^\mathsf{T} B\mathbf{f} = 1\right\}$$

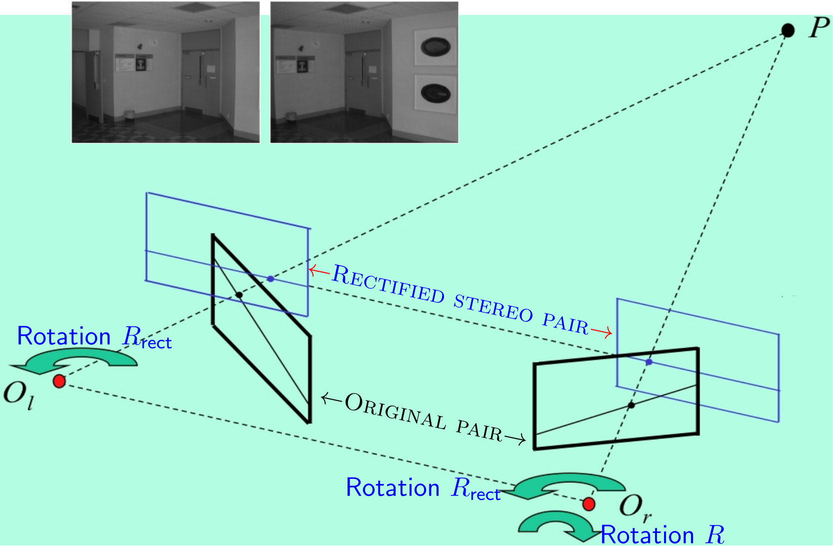
  where $B = \sum\limits_{i=1}^{n} (B_{1:i} + B_{2:i})$.

Rectification of Stereo Images

Rectification of a stereo pair – a transformation (**warping**) of each image such that pairs of conjugate epipolar lines become collinear and parallel to one of the image axes (typically, the x-axis)

- **Goal**: 1D search on scan-lines for point-to-point correspondences after rectification
- **Computation**: from the known intrinsic and extrinsic parameters of stereo cameras
  - Rectified images are thought of as acquired by a new stereo rig obtained by rotating the original cameras around their optical centres
  - Epipolar lines associated to a 3D point P in the original cameras become collinear in the rectified cameras
  - Original cameras can be in any position, and their optical axes may not intersect

Rectification of a Stereo Pair



Rectification Algorithm

Assumptions for the two cameras (without losing generality):

- Origin of the image reference frame – in the principal point (the trace of the optical axis)
- The same focal lengths f of both cameras
- Coordinate frames of the left and right cameras are related by the translation vector $\mathbf{T} = \mathbf{O}_2 - \mathbf{O}_1$ and the relative rotation matrix R

Basic steps of rectification:

1. Rotate the left camera to make its image plane parallel to the baseline of the system (the epipole goes to infinity along the x-axis)
2. Apply the same rotation R_{rect} to the right camera to recover the original geometry
3. Rotate the right camera by R to make its image plane parallel to the baseline
4. Adjust the scale in both the camera reference frames

Rectification Algorithm: Rotation Matrix R_{rect} for Step 1

From a triple of mutually orthogonal unit vectors: $R_{\text{rect}} = \begin{bmatrix} \mathbf{e}_1^\mathsf{T} \\ \mathbf{e}_2^\mathsf{T} \\ \mathbf{e}_3^\mathsf{T} \end{bmatrix}$

- An arbitrary choice due to the under-constrained problem
- Vector $\mathbf{e}_1$ – along the direction of translation (as the image centre is in the origin): $\mathbf{e}_1 = \frac{\mathbf{T}}{|\mathbf{T}|} = \frac{1}{\sqrt{T_x^2+T_y^2+T_z^2}} \left[T_x \;\; T_y \;\; T_z\right]^\mathsf{T}$
- Vector $\mathbf{e}_2$ – orthogonal to the plane containing both optical centres (i.e. the translation vector $\mathbf{T}$) and the optical axis Z of the left camera with the directional vector $\mathbf{d}_z = [0\ 0\ 1]^\mathsf{T}$:
$$\mathbf{e}_2 = \frac{\mathbf{e}_1 \times \mathbf{d}_z}{|\mathbf{e}_1 \times \mathbf{d}_z|} = \frac{1}{\sqrt{T_x^2+T_y^2}} \left[-T_y \;\; T_x \;\; 0\right]^\mathsf{T}$$
- Vector $\mathbf{e}_3$ – orthogonal to the first two vectors:

$$\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2 = \frac{1}{\sqrt{(T_x^2+T_y^2)(T_x^2+T_y^2+T_z^2)}} \left[-T_xT_z \quad -T_yT_z \quad T_x^2+T_y^2\right]^\mathsf{T}$$

Rectification Algorithm

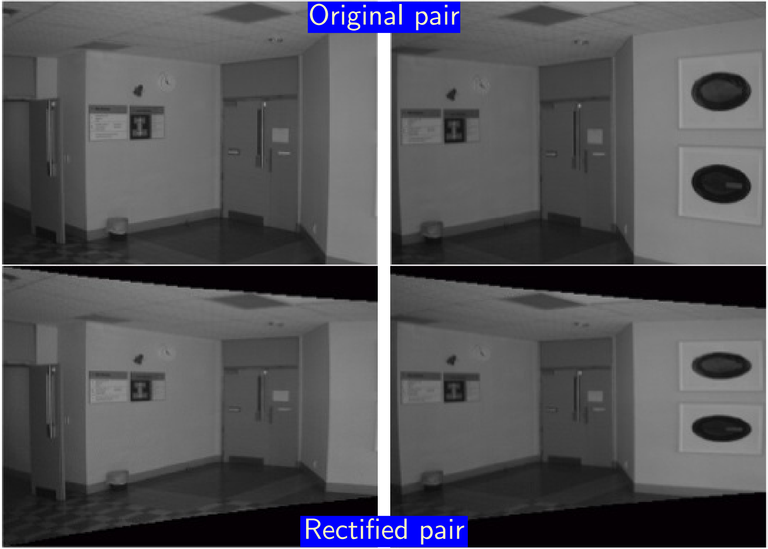
Input:

- Intrinsic and extrinsic parameters (in particular, the right camera translation $\mathbf{T}$ and rotation R w.r.t. the left camera)
- Images (or sets of their points) to be rectified
- Assumptions in Slide 65 about the origins and focal distances hold

Basic steps:

1. Build the matrix R_{rect} (see Slide 66)
2. Set $R_1 = R_{\text{rect}}$ and $R_2 = RR_{\text{rect}}$
3. For each left-camera point, $\mathbf{s}_1^\mathsf{T} = [x \quad y \quad f]$, compute
   - $\mathbf{s}'_1 \equiv [x' \quad y' \quad z']^\mathsf{T} = R_1\mathbf{s}_1$ and then
   - the rectified coordinates $\widehat{\mathbf{s}}_1^\mathsf{T} = \left[\frac{fx'}{z'} \quad \frac{fy'}{z'} \quad f\right]$
4. Repeat Step 3 for the right camera using R_2 and $\mathbf{s}_2$

Rectification of a Stereo Pair



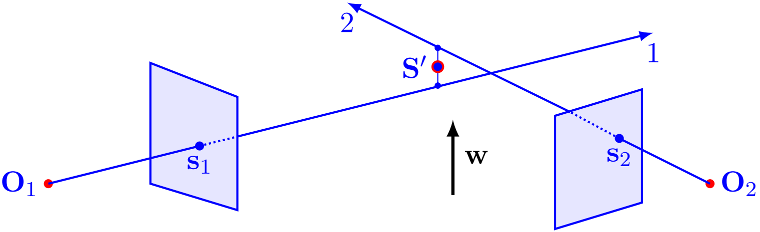
Three Basic Cases for a Stereo Pair

Reconstruction of an optical 3D surface depending on the amount of *a priori* knowledge about parameters of stereo cameras:

- **Known both intrinsic and extrinsic parameters**: a 3-D scene is uniquely reconstructed by triangulation of the corresponding image points
- **Known only intrinsic parameters**: a 3-D scene is still reconstructed and also the extrinsic parameters are estimated, but up to an unknown scaling factor
- **Unknown intrinsic and extrinsic parameters**: a 3-D scene is still reconstructed, but up to an unknown global projective transformation

Stereo matching: determining the corresponding points in a pair or multiple images of the same 3D scene

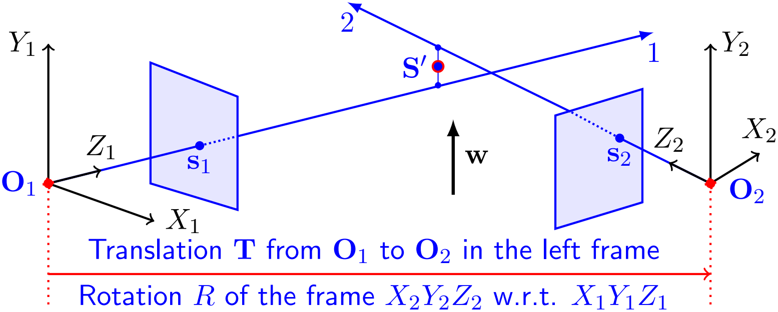
Triangulation from Projections



A 3D point $\mathbf{S}$, projected onto a pair of corresponding points $\mathbf{s}_1$ and $\mathbf{s}_2$, lies at the intersection of two inverse optical rays, 1 and 2, from $\mathbf{O}_1$ through $\mathbf{s}_1$ and from $\mathbf{O}_2$ through $\mathbf{s}_2$, respectively

- Due to approximate camera parameters and image locations, the two rays may not actually intersect in the 3D space
- Least-squares estimation of the intersection: the point $\mathbf{S}'$ at the minimum distance from both the rays along $\mathbf{w}$
- $\mathbf{w}$ – a vector being orthogonal to both the optical rays

Triangulation (in the Left Reference Frame)



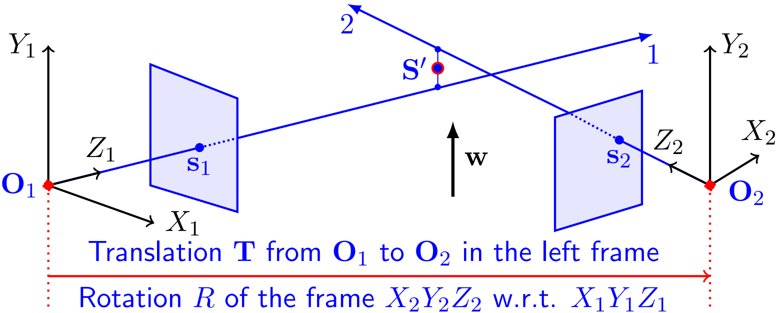
$\mathbf{w} = \mathbf{s}_1 \times R^{\mathsf{T}}\mathbf{s}_2$ – a vector orthogonal to both 1 and 2

- Cross product (see Slide 39 and http://en.wikipedia.org/wiki/Cross_product):

$$\begin{bmatrix} \alpha_x \\ \alpha_y \\ \alpha_z \end{bmatrix} \times \begin{bmatrix} \beta_x \\ \beta_y \\ \beta_z \end{bmatrix} = \begin{bmatrix} \gamma_x \\ \gamma_y \\ \gamma_z \end{bmatrix} \equiv \begin{bmatrix} \alpha_y\beta_z - \alpha_z\beta_y \\ \alpha_z\beta_x - \alpha_x\beta_z \\ \alpha_x\beta_y - \alpha_y\beta_x \end{bmatrix}$$

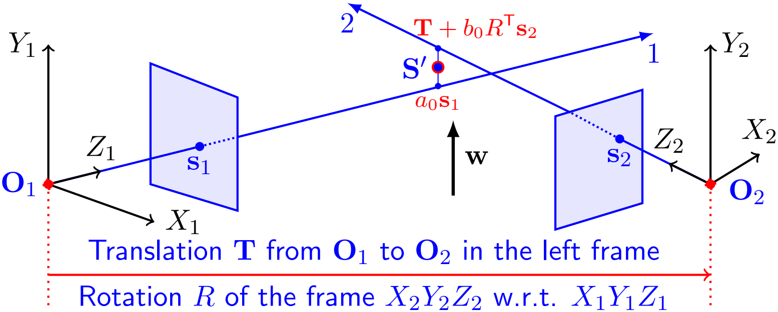
- $\mathbf{S}'$ – midpoint of the segment joining 1 and 2 and parallel to $\mathbf{w}$

Triangulation (in the Left Reference Frame)



- $a\mathbf{s}_1$; $a \in \mathbb{R} = (-\infty, \infty)$: ray 1 through the left origin $\mathbf{O}_1$ and point $\mathbf{s}_1$
- $\mathbf{T} + bR^{\mathsf{T}}\mathbf{s}_2$; $b \in \mathbb{R}$: ray 2 through the right origin $\mathbf{O}_2$ and point $\mathbf{s}_2$
  - R – orthonormal rotation matrix (3D rotation of the right frame w.r.t. the left frame)
  - R^{T} – transposed (i.e. inverted) matrix R (3D rotation of the right frame vectors back to the left frame)

Triangulation (in the Left Reference Frame)



- Endpoints, $a_0\mathbf{s}_1$, and $\mathbf{T} + b_0 R^\mathsf{T} \mathbf{s}_2$, of the segment joining 1 and 2 and parallel to $\mathbf{w}$ are computed by solving the linear system of equations for a_0, b_0, and c_0:

$$a\mathbf{s}_1 + c\left(\mathbf{s}_1 \times R^\mathsf{T}\mathbf{s}_2\right) = \mathbf{T} + bR^\mathsf{T}\mathbf{s}_2$$

$$\Rightarrow \quad a\mathbf{s}_1 - bR^\mathsf{T}\mathbf{s}_2 + c\left(\mathbf{s}_1 \times R^\mathsf{T}\mathbf{s}_2\right) = \mathbf{T}$$

Reconstruction up to a Scale

- Reconstruction by using the **essential matrix**, E
  - If $\mathbf{q}_1$ and $\mathbf{q}_2$ are *normalised* homogeneous coordinates of the corresponding points in images, then $\mathbf{q}_2^\mathsf{T} E \mathbf{q}_1 = 0$
    - Normalised coordinates are measured in a coordinate system with the origin at the trace of 3D axis Z on the image plane
    - See `http://en.wikipedia.org/wiki/Essential_matrix` and Slides 40 and 41
  - Known data: only the intrinsic parameters and n pairs of the corresponding points, $n \geq 8$
  - Since the baseline is unknown, the true scale of the viewed scene cannot be recovered

The estimated essential matrix, E, can only be specified up to an arbitrary scale factor

- Convenient normalisation of E - by normalising the length of the translation vector $\mathbf{T}$ to unit

Normalising the Essential Matrix E

From the definition: $E = RS$

$$
\begin{aligned}
E^\top E &= (RS)^\top RS = S^\top R^\top RS = S^\top S \\
&= \begin{bmatrix} 0 & T_z & -T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \begin{bmatrix} 0 & -T_z & T_y \\ -T_z & 0 & T_x \\ T_y & -T_x & 0 \end{bmatrix} \\
&= \begin{bmatrix} T_y^2 + T_z^2 & -T_x T_y & -T_x T_z \\ -T_y T_x & T_z^2 + T_x^2 & -T_y T_z \\ -T_z T_x & -T_z T_y & T_x^2 + T_y^2 \end{bmatrix} \Rightarrow Tr(E^\top E) = 2|\mathbf{T}|^2
\end{aligned}
$$

- Normalised translation: $\widehat{\mathbf{T}} = \frac{\mathbf{T}}{|\mathbf{T}|} \equiv \frac{[T_x \; T_y \; T_z]^\top}{\sqrt{T_x^2 + T_y^2 + T_z^2}} = \begin{bmatrix} \widehat{T}_x & \widehat{T}_y & \widehat{T}_z \end{bmatrix}$
- Normalised essential matrix: $\widehat{E} = \frac{E}{\sqrt{Tr(E^\top E)/2}}$
- $\widehat{E}^\top \widehat{E} = \begin{bmatrix} 1 - \widehat{T}_x^2 & -\widehat{T}_x \widehat{T}_y & -\widehat{T}_x \widehat{T}_z \\ -\widehat{T}_y \widehat{T}_x & 1 - \widehat{T}_y^2 & -\widehat{T}_y \widehat{T}_z \\ -\widehat{T}_z \widehat{T}_x & -\widehat{T}_z \widehat{T}_y & 1 - \widehat{T}_z^2 \end{bmatrix}$

Finding the Pair $(\widehat{\mathbf{T}}, R)$

Components of $\widehat{\mathbf{T}}$ – from any row or column of $G = \widehat{E}^{\mathsf{T}}\widehat{E}$

- The estimates $T_x = \pm\sqrt{1 - G_{11}}$; $T_y = -\frac{G_{12}}{T_x}$, and $T_z = -\frac{G_{13}}{T_x}$ may differ from the true components by a global sign change (due to quadratic entries of G)
- Rotation matrix R – from $\widehat{E}$ and $\widehat{T}$:

$$\widehat{E} = \begin{bmatrix} \widehat{\mathbf{E}}_1^{\mathsf{T}} \\ \widehat{\mathbf{E}}_2^{\mathsf{T}} \\ \widehat{\mathbf{E}}_3^{\mathsf{T}} \end{bmatrix} \Rightarrow \underset{i=1,2,3}{\mathbf{w}_i = \widehat{\mathbf{E}}_i \times \widehat{\mathbf{T}}} \Rightarrow R = \begin{bmatrix} \mathbf{R}_1^{\mathsf{T}} = (\mathbf{w}_1 + \mathbf{w}_2 \times \mathbf{w}_3)^{\mathsf{T}} \\ \mathbf{R}_2^{\mathsf{T}} = (\mathbf{w}_2 + \mathbf{w}_3 \times \mathbf{w}_1)^{\mathsf{T}} \\ \mathbf{R}_3^{\mathsf{T}} = (\mathbf{w}_3 + \mathbf{w}_1 \times \mathbf{w}_2)^{\mathsf{T}} \end{bmatrix}$$

- Due to the twofold ambiguity in the sign of $\widehat{E}$ and $\widehat{\mathbf{T}}$, there are four different estimates for the goal pair $(\widehat{\mathbf{T}}, R)$
- 3D reconstruction of the viewed points resolves the ambiguity and finds the only correct estimate

Finding the Pair $(\widehat{\mathbf{T}}, R)$

To resolve the ambiguity, the third component of every point in the left reference frame is computed for each of the four pairs $(\widehat{\mathbf{T}}, R)$

$$\mathbf{S}_2 = R(\mathbf{S}_1 - \widehat{\mathbf{T}}) \;\rightarrow\; Z_2 = \mathbf{R}_3^\top(\mathbf{S}_1 - \widehat{\mathbf{T}}) \;\rightarrow\; \mathbf{s}_2 = \frac{f_2}{Z_2}\mathbf{S}_2$$

$$\Rightarrow \quad \mathbf{s}_2 = \frac{f_2 R(\mathbf{S}_1 - \widehat{\mathbf{T}})}{\mathbf{R}_3^\top(\mathbf{S}_1 - \widehat{\mathbf{T}})} \;\rightarrow\; x_2 = \frac{f_2 \mathbf{R}_1^\top(\mathbf{S}_1 - \widehat{\mathbf{T}})}{\mathbf{R}_3^\top(\mathbf{S}_1 - \widehat{\mathbf{T}})}$$

$$\Rightarrow \quad \mathbf{s}_1 = \frac{f_1}{Z_1}\mathbf{S}_1 \;\rightarrow\; Z_1 = f_1 \frac{(f_2\mathbf{R}_1 - x_2\mathbf{R}_3)^\top\widehat{\mathbf{T}}}{(f_2\mathbf{R}_1 - x_2\mathbf{R}_3)^\top\mathbf{s}_1}$$

$$\Rightarrow \quad \mathbf{S}_1 = \frac{(f_2\mathbf{R}_1 - x_2\mathbf{R}_3)^\top\widehat{\mathbf{T}}}{(f_2\mathbf{R}_1 - x_2\mathbf{R}_3)^\top\mathbf{s}_1}\mathbf{s}_1; \;\; \mathbf{S}_2 = R(\mathbf{S}_1 - \widehat{\mathbf{T}})$$

Reconstruction Algorithm

Input: a set of corresponding points and an estimated E

1. Recover the normalised translation vector $\widehat{\mathbf{T}}$
2. Recover the rotation matrix R
3. Reconstruct the 3D coordinates Z_1 and Z_2 of each point
    1. If the signs of Z_1 and Z_2 of the reconstructed points are both negative for some point, change the sign of $\widehat{\mathbf{T}}$ and go to 3
    2. Otherwise if the signs of Z_1 and Z_2 of the reconstructed points are one negative and one positive for some point, change the sign of each entry of E and go to 2
    3. Otherwise if the signs of Z_1 and Z_2 of the reconstructed points are both positive for all points, exit

Uncalibrated Reconstruction

No information on the intrinsic and extrinsic parameters

- Only n point-to-point correspondences, $n > 8$, are given
  - Location of the epipoles is thus known
  - Reconstruction accuracy is affected by accuracy of disparities, not calibration

Reconstruction is unique only up to an unknown projective transformation of the world

- Projection matrix of each camera is recovered from 5 arbitrary scene points and the epipoles up to this transformation
  - No three of these 3D points should be collinear
  - No four of these 3D points should be coplanar
- Then 3D location of any point can be found by triangulation
  - Five 3D points $\mathbf{S}_1, \ldots, \mathbf{S}_5$ to be recovered from their 2D locations, $\mathbf{s}_{1:1}, \ldots, \mathbf{s}_{1:5}$ and $\mathbf{s}_{2:1}, \ldots, \mathbf{s}_{2:5}$ on the images

Uncalibrated Reconstruction

- **Planar projective transformation** is fixed if destinations of 4 spatial points in an image are known
- **Spatial projective transformation** is fixed if the destinations of 5 points are known: $M\mathbf{S}_i = \rho_i \mathbf{s}_i$; $\rho_i \neq 0$; $i = 1, \ldots, 5$, where M is the projection 3×4 matrix
- Without losing generality, a projective transformation is set up to associate the 5 image points with the following 3D points:
  1. $\mathbf{S}_1 = [1, 0, 0, 0]^\mathsf{T}$ – an infinitely far 3D point along the X-axis
  2. $\mathbf{S}_2 = [0, 1, 0, 0]^\mathsf{T}$ – an infinitely far 3Dpoint along the Y-axis
  3. $\mathbf{S}_3 = [0, 0, 1, 0]^\mathsf{T}$ – an infinitely far 3D point along the Z-axis
  4. $\mathbf{S}_4 = [0, 0, 0, 1]^\mathsf{T}$ – the 3D coordinate frame origin $[0\ 0\ 0]$
  5. $\mathbf{S}_5 = [1, 1, 1, 1]^\mathsf{T}$ – the "unit" 3D point $[1\ 1\ 1]^\mathsf{T}$

  in homogeneous coordinates: $[X, Y, Z]^\mathsf{T} \leftrightarrow [X, Y, Z, 1]^\mathsf{T}$ and $[x, y]^\mathsf{T} \leftrightarrow [x, y, 1]^\mathsf{T}$

Uncalibrated Reconstruction

Spatial projective transformation is set up to associate the image points $\mathbf{s}_i$ with the standard projection basis $\mathbf{S}_i$ on Slide 80:

1. $\mathbf{s}_1 = [1, 0, 0]^\mathsf{T}$ – an infinitely far 2D point along the x-axis
2. $\mathbf{s}_2 = [0, 1, 0]^\mathsf{T}$ – an infinitely far 2D point along the y-axis
3. $\mathbf{s}_3 = [0, 0, 1]^\mathsf{T}$ – the 2D coordinate frame origin $[0\ 0]^\mathsf{T}$
4. $\mathbf{s}_4 = [1, 1, 1]^\mathsf{T}$ – the "unit" 2D point $[1\ 1]^\mathsf{T}$
5. $\mathbf{s}_5 = [\alpha, \beta, \gamma]^\mathsf{T}$ – an arbitrary other point $\left[\frac{\alpha}{\gamma}\ \frac{\beta}{\gamma}\right]$ in the image

The projection matrix M is found from $M\mathbf{S}_i = \rho_i \mathbf{s}_i;\ i = 1, \ldots, 5$: e.g.

$$\left[\begin{array}{cccc} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \end{array}\right] \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array}\right] = \rho_1 \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right] \Rightarrow M_{11} = \rho_1;\ M_{21} = 0;\ M_{31} = 0$$

Uncalibrated Reconstruction

The chosen standard basis simplifies the expression of M:

$$
\begin{array}{rcl}
M\mathbf{S}_1 = \rho_1 \mathbf{s}_1 & \Rightarrow & M_{11} = \rho_1;\ M_{21} = 0;\ M_{31} = 0 \\
M\mathbf{S}_2 = \rho_1 \mathbf{s}_2 & \Rightarrow & M_{12} = 0;\ M_{22} = \rho_2;\ M_{32} = 0 \\
M\mathbf{S}_3 = \rho_3 \mathbf{s}_3 & \Rightarrow & M_{13} = 0;\ M_{23} = 0;\ M_{33} = \rho_3 \\
M\mathbf{S}_4 = \rho_4 \mathbf{s}_4 & \Rightarrow & M_{14} = \rho_4;\ M_{24} = \rho_4;\ M_{34} = \rho_4 \\
M\mathbf{S}_5 = \rho_5 \mathbf{s}_5 & \Rightarrow & M_{11} + M_{12} + M_{13} + M_{14} = \alpha\rho_5 \\
 & & M_{21} + M_{22} + M_{23} + M_{24} = \beta\rho_5 \\
 & & M_{31} + M_{32} + M_{33} + M_{34} = \gamma\rho_5 \\
\hline
 & & \rho_1 = \alpha\rho_5 - \rho_4 \\
 & & \rho_2 = \beta\rho_5 - \rho_4 \\
 & & \rho_3 = \gamma\rho_5 - \rho_4 \\
\hline
\end{array}
$$

$$
M = \begin{bmatrix} \alpha\rho_5 - \rho_4 & 0 & 0 & \rho_4 \\ 0 & \beta\rho_5 - \rho_4 & 0 & \rho_4 \\ 0 & 0 & \gamma\rho_5 - \rho_4 & \rho_4 \end{bmatrix}
$$

Uncalibrated Reconstruction

$$\left.\begin{array}{r} M = \begin{bmatrix} \rho_1 & 0 & 0 & \rho_4 \\ 0 & \rho_2 & 0 & \rho_4 \\ 0 & 0 & \rho_3 & \rho_4 \end{bmatrix} \\ \\ M\mathbf{S}_5 = \rho_5 \mathbf{s}_5 \end{array}\right\} \quad \underset{i=1,\dots,5}{M\mathbf{S}_i = \rho_i \mathbf{s}_i} \Rightarrow$$

$$\Rightarrow \quad M = \begin{bmatrix} \alpha\rho_5 - \rho_4 & 0 & 0 & \rho_4 \\ 0 & \beta\rho_5 - \rho_4 & 0 & \rho_4 \\ 0 & 0 & \gamma\rho_5 - \rho_4 & \rho_4 \end{bmatrix}$$

$$\Rightarrow \quad M = \begin{bmatrix} \alpha\kappa - 1 & 0 & 0 & 1 \\ 0 & \beta\kappa - 1 & 0 & 1 \\ 0 & 0 & \gamma\kappa - 1 & 1 \end{bmatrix} ; \ \kappa = \frac{\rho_5}{\rho_4}$$

The projection matrix M depends on the parameter κ and the chosen image point $[\alpha, \beta, \gamma]^\mathsf{T} \rightarrow \left[\frac{\alpha}{\gamma} \ \frac{\beta}{\gamma}\right]^\mathsf{T}$

Uncalibrated Reconstruction

Projection matrices of the left (M_1) and right (M_2) cameras are found up to unknown parameters κ_1 and κ_2:

$$M_c = \begin{bmatrix} \alpha_c\kappa_c - 1 & 0 & 0 & 1 \\ 0 & \beta_c\kappa_c - 1 & 0 & 1 \\ 0 & 0 & \gamma_c\kappa_c - 1 & 1 \end{bmatrix} \quad \text{where} \quad c \in \{1, 2\}$$

The parameters are computed using known projection centres $\mathbf{O}_1$, $\mathbf{O}_2$ and locations of the epipoles $\mathbf{e}_1$, $\mathbf{e}_2$ (see Slide 32).

- Projection centres – the null spaces of M_1 and M_2, i.e. $M_1\mathbf{O}_1 = \mathbf{0}$ and $M_2\mathbf{O}_2 = \mathbf{0}$
- Epipoles: $M_1\mathbf{O}_2 = \sigma_1\mathbf{e}_1$ and $M_2\mathbf{O}_1 = \sigma_2\mathbf{e}_2$ with $\sigma_1 \neq 0$, $\sigma_2 \neq 0$

Then any 3D point is reconstructed using the inverse projective rays through $\mathbf{O}_1$ and $\mathbf{O}_2$

Uncalibrated Reconstruction

Parametric projection matrices for $c = 1, 2$:

$$M_c = \underbrace{\begin{bmatrix} \alpha_c\kappa_c - 1 & 0 & 0 \\ 0 & \beta_c\kappa_c - 1 & 0 \\ 0 & 0 & \gamma_c\kappa_c - 1 \end{bmatrix}}_{3\times3 \text{ matrix } Q_c} \underbrace{\begin{matrix} 1 \\ 1 \\ 1 \end{matrix}}_{\mathbf{q}_c} \Rightarrow [Q_c \;\; \mathbf{q}_c]$$

- Optical centres: $\mathbf{O}_c = -Q_c^{-1}\mathbf{q}_c = -\begin{bmatrix} (1/(\alpha_c\kappa_c - 1) \\ 1/(\beta_c\kappa_c - 1) \\ 1/(\gamma_c\kappa_c - 1) \end{bmatrix}$
- Epipoles: $\widetilde{\mathbf{e}}_j = -Q_j Q_i^{-1}\mathbf{q}_i + \mathbf{q}_j =$

$$\begin{bmatrix} 1 - (\alpha_j\kappa_j - 1)/(\alpha_i\kappa_i - 1) \\ 1 - (\beta_j\kappa_j - 1)/(\beta_i\kappa_i - 1) \\ 1 - (\gamma_j\kappa_j - 1)/(\gamma_i\kappa_i - 1) \end{bmatrix} \equiv \begin{bmatrix} (\alpha_i\kappa_i - \alpha_j\kappa_j)/(\alpha_i\kappa_i - 1) \\ (\beta_i\kappa_i - \beta_j\kappa_j)/(\beta_i\kappa_i - 1) \\ (\gamma_i\kappa_i - \gamma_j\kappa_j)/(\gamma_i\kappa_i - 1) \end{bmatrix}$$