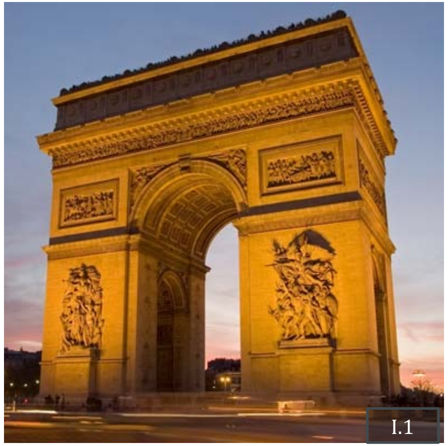
Johns Hopkins Engineering

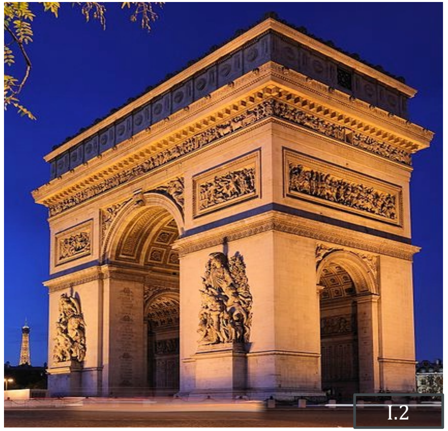
Computer Vision

Uncalibrated Binocular Stereo and Multi-View Geometry

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Uncalibrated Binocular Stereo





Left Image

Right Image

Uncalibrated Binocular Stereo

Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may be unknown.

Topics:

- Epipolar Geometry
- Essential and Fundamental Matrix
- Stereo Self-Calibration
- Stereopsis and Multiview Reconstruction

Uncalibrated Binocular Stereo

Method to estimate 3D structure from two arbitrary images of a scene captured with cameras whose intrinsic parameters are known.

Topics:

- Epipolar Geometry

Review: Linear Camera Model

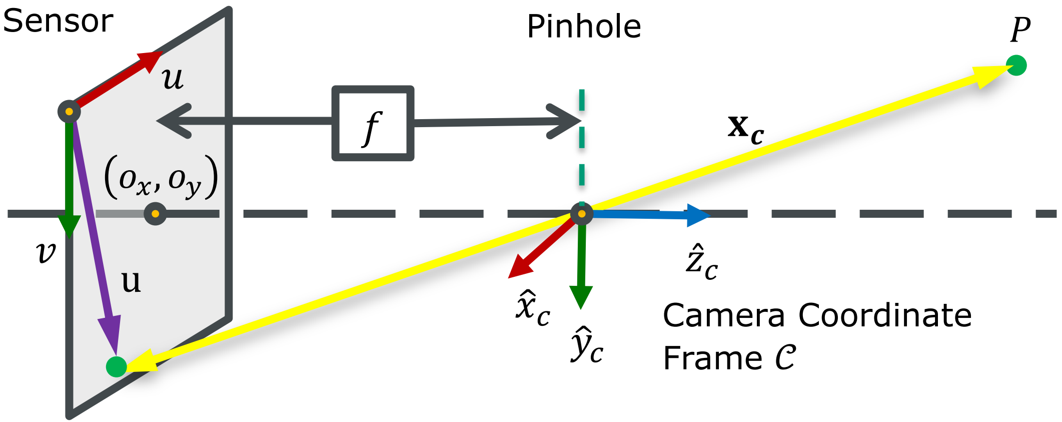


Image Coordinates

Camera Coordinates

$$\mathbf{u} = \begin{bmatrix} u \\ v \end{bmatrix} \longleftarrow \mathbf{x}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}$$

Perspective Projection

Review: Linear Camera Model

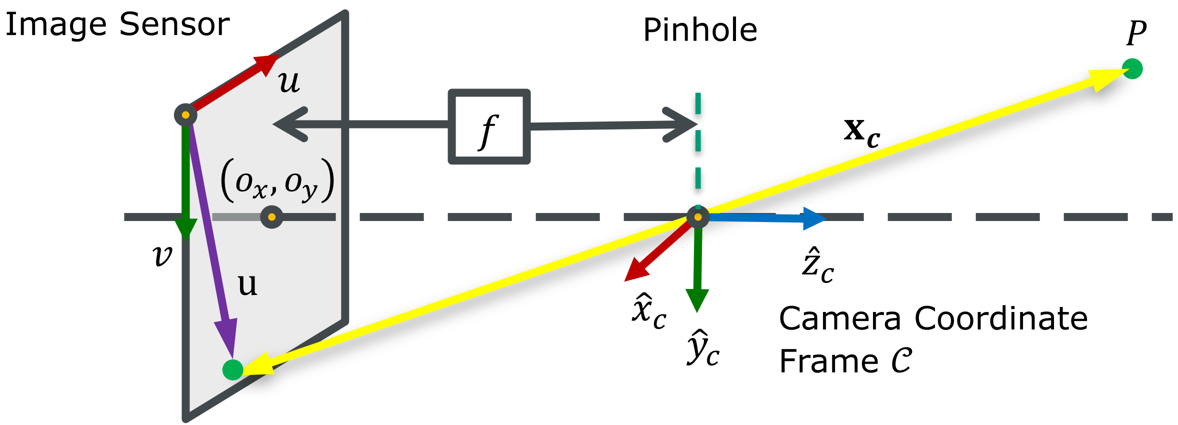


Image Homogenous Coordinates

Camera Homogenous Coordinates

$$\tilde{\mathbf{u}} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \xleftarrow{M_{int}} \tilde{\mathbf{x}}_c = \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}$$

Perspective Projection

Review: Linear Camera Model

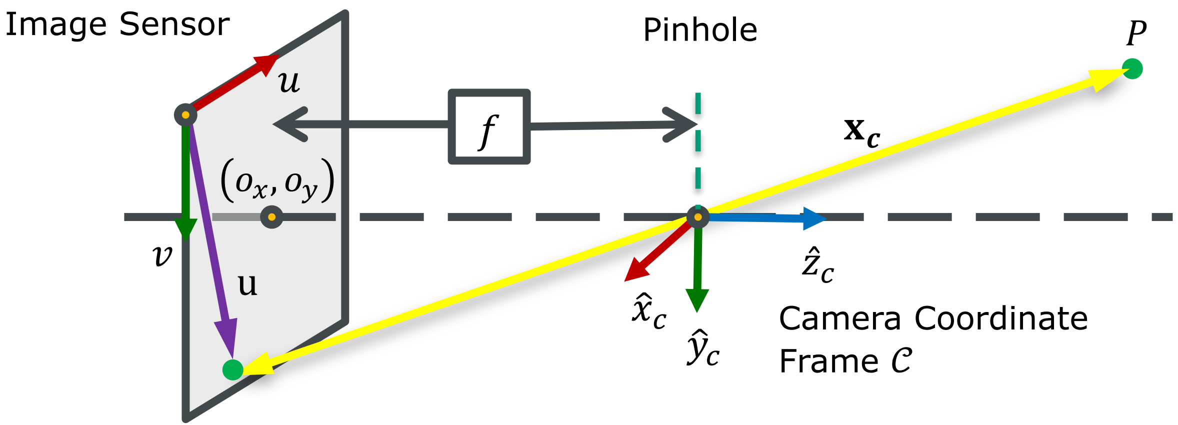
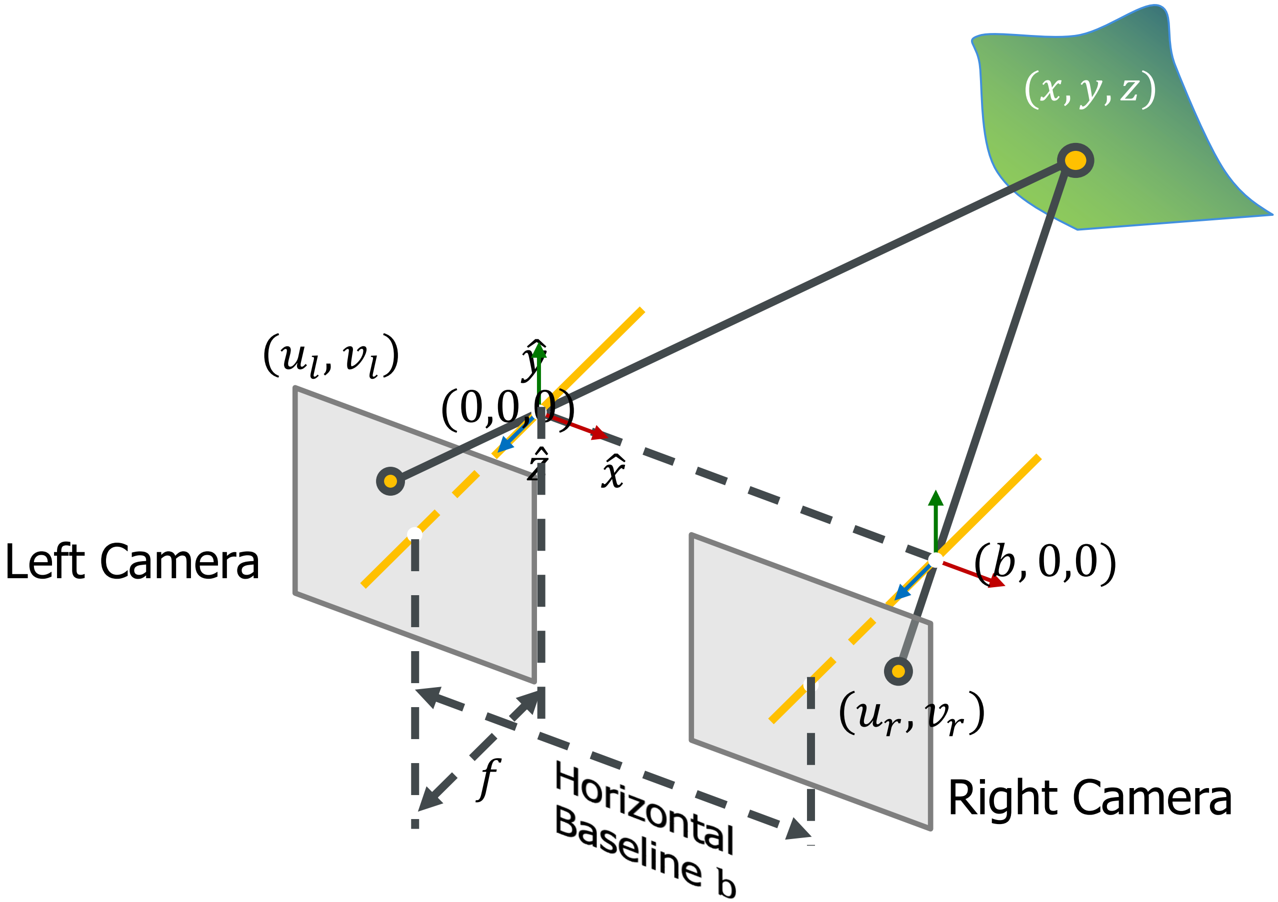


Image Homogenous Coordinates

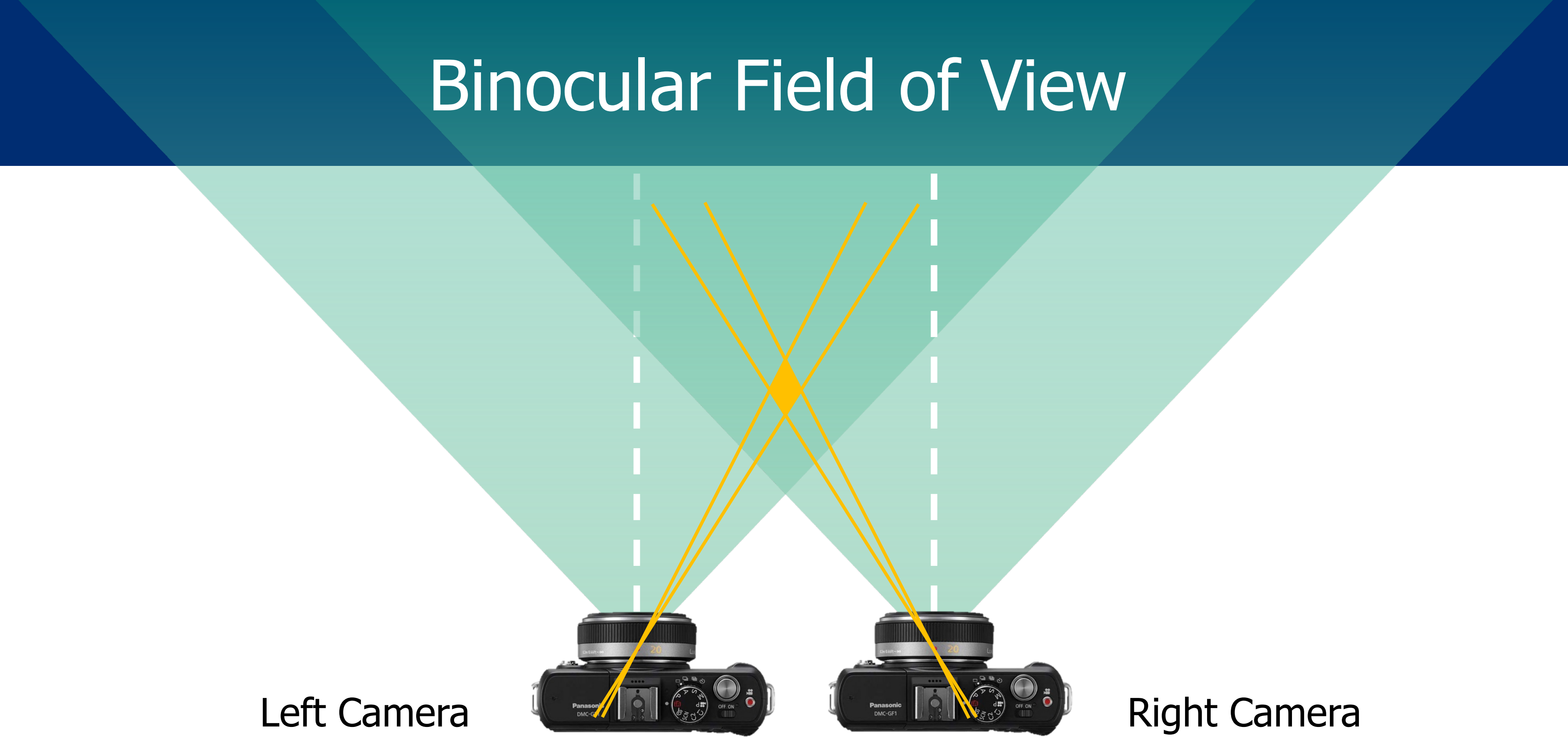
Camera Homogenous Coordinates

$$\underbrace{\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix}}_{\tilde{\mathbf{u}}} = \underbrace{\begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M_{int}} \underbrace{\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}}_{\tilde{\mathbf{x}}_c}$$

Review: Simple Stereo

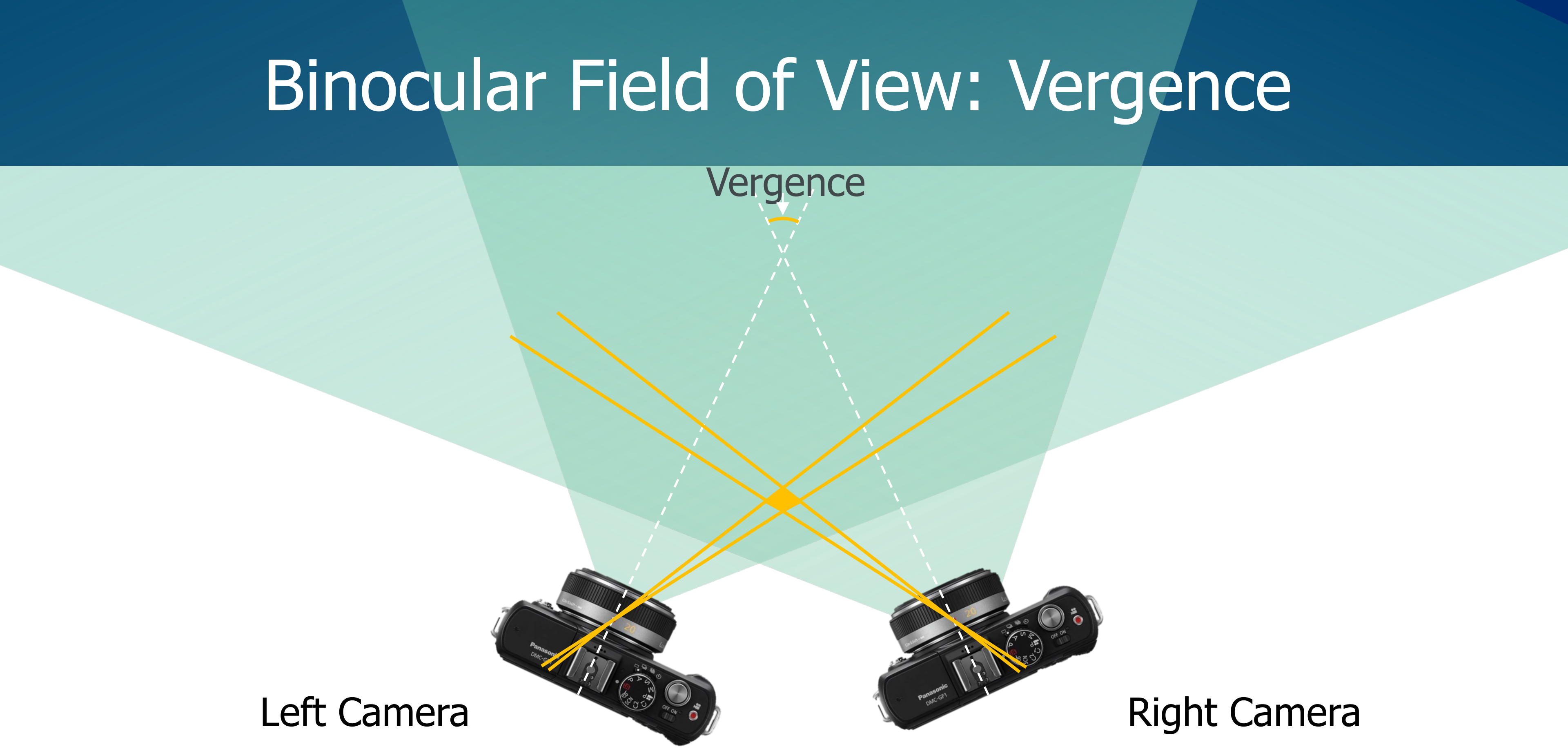


Binocular Field of View



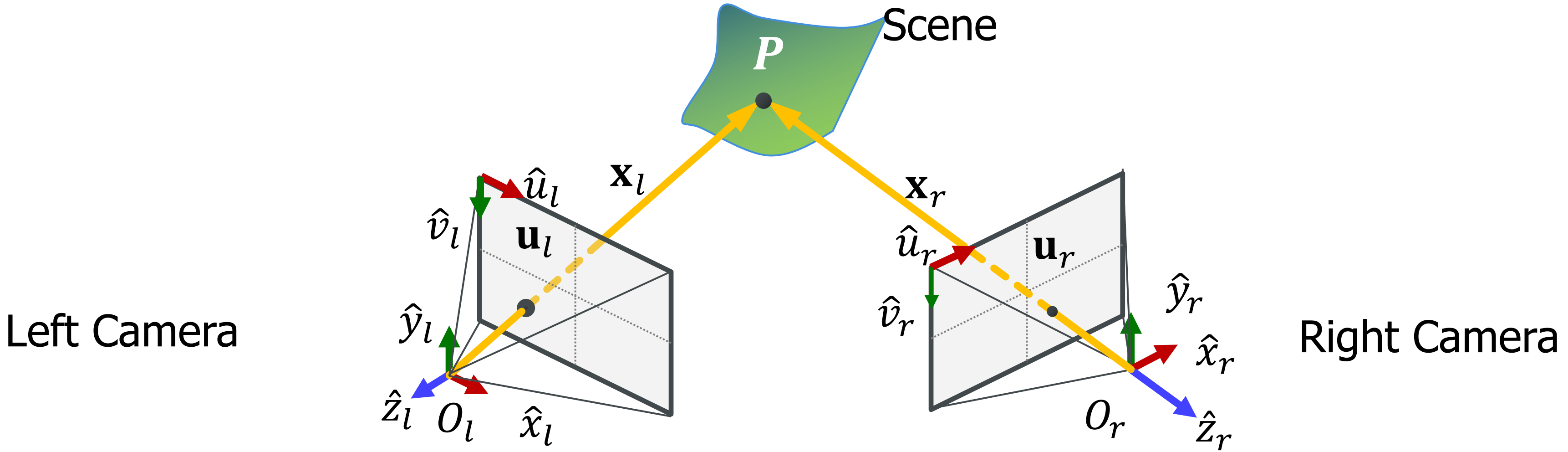
Binocular Field of View is the overlapping field of view.

Binocular Field of View: Vergence



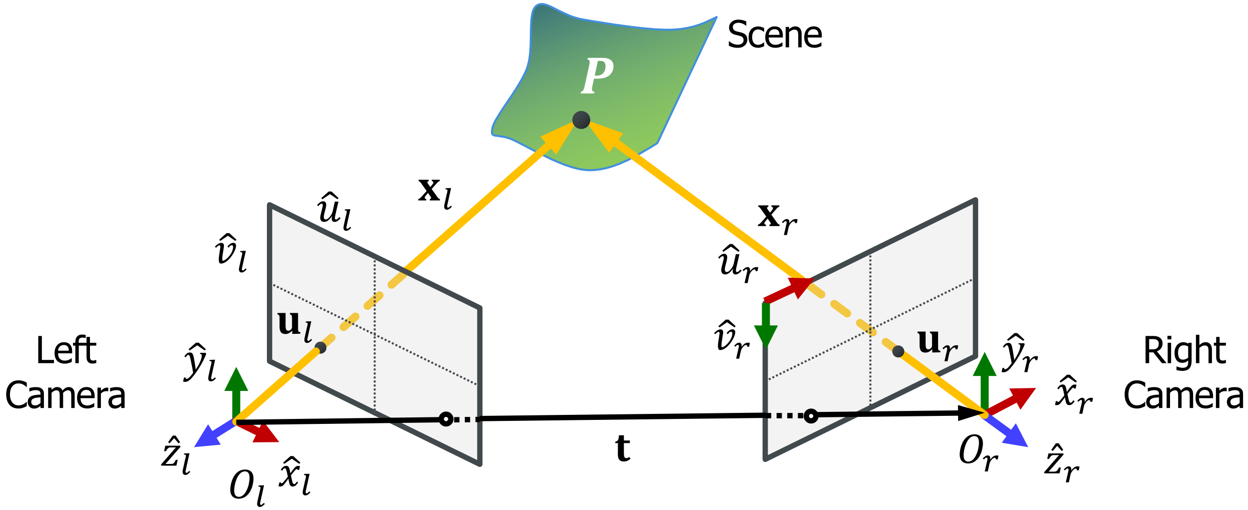
Field of view Decreases; Accuracy Increases with Vergence

Uncalibrated Binocular Stereo



Compute depth using two cameras (whose intrinsics are known) with arbitrary position and orientation.

Relative Position and Orientation



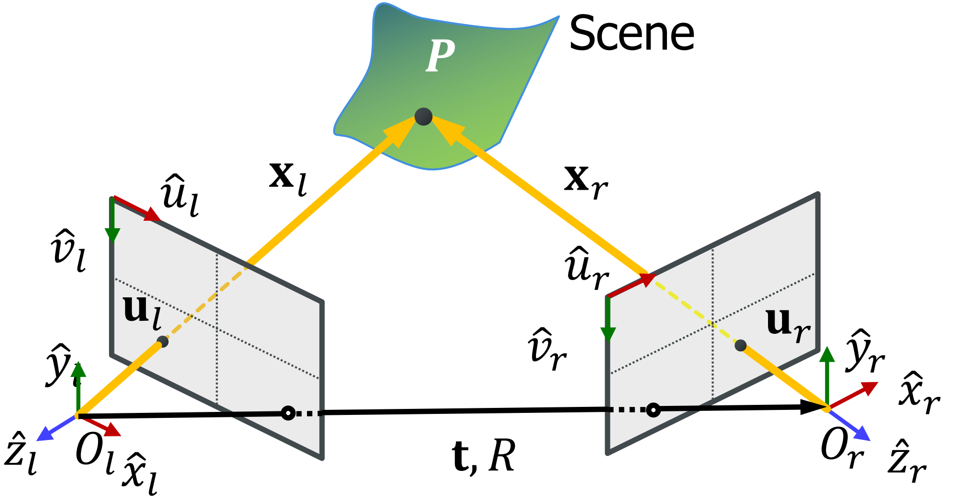
$\mathbf{t}_{3\times1}$: Position of Right Camera in Left Camera Coordinate Frame($\overrightarrow{O_l O_r}$)

$R_{3\times3}$: Rotation from Right to Left Camera Coordinate Frame

$$\boxed{\mathbf{x}_l = R\mathbf{x}_r + \mathbf{t}}$$

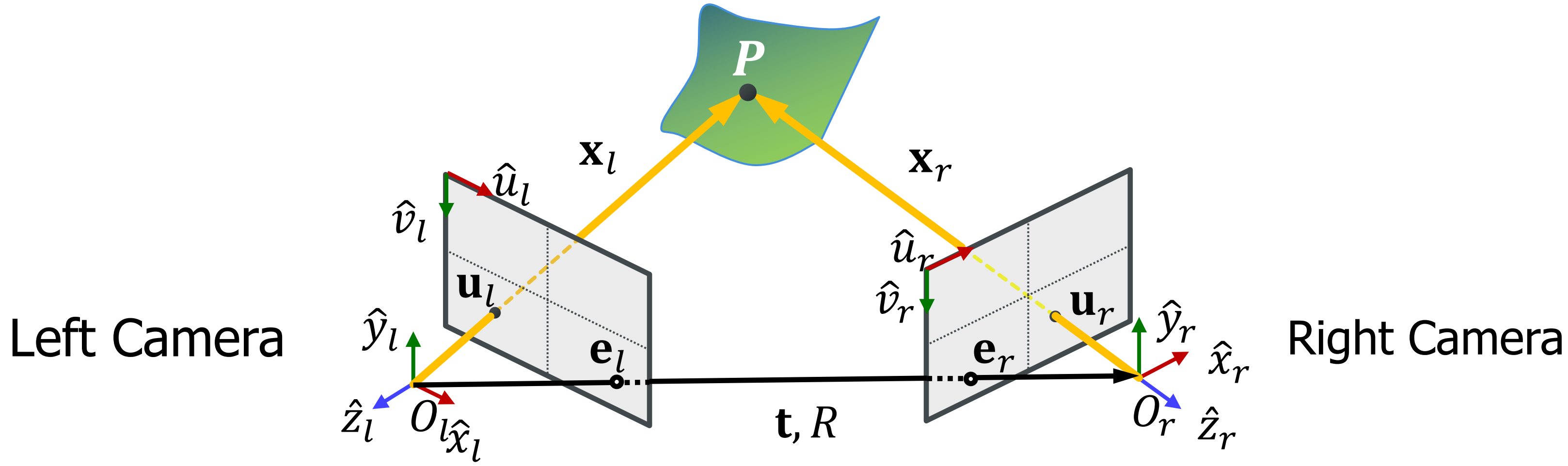
$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

Binocular Stereo



1. Assume Camera Intrinsic Parameters f_x, f_y, o_x, o_y are known.
2. Find Relative Camera Position $\mathbf{t}$ and Orientation R from the two images.
3. Find Correspondence for each pixel in the two images.
4. Compute Depth for each pixel using Triangulation.

Epipolar Geometry: Epipole

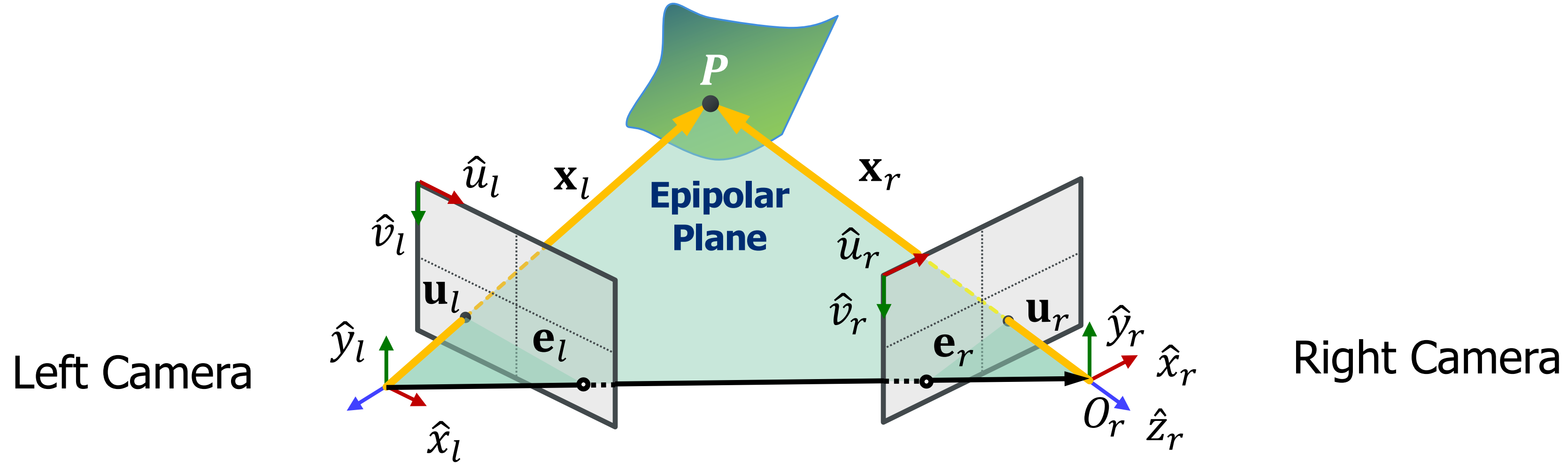


The image point of the origin/pinhole of one camera as viewed by the other camera is called the epipole.

$\mathbf{e}_l$ and $\mathbf{e}_r$ are the epipoles.

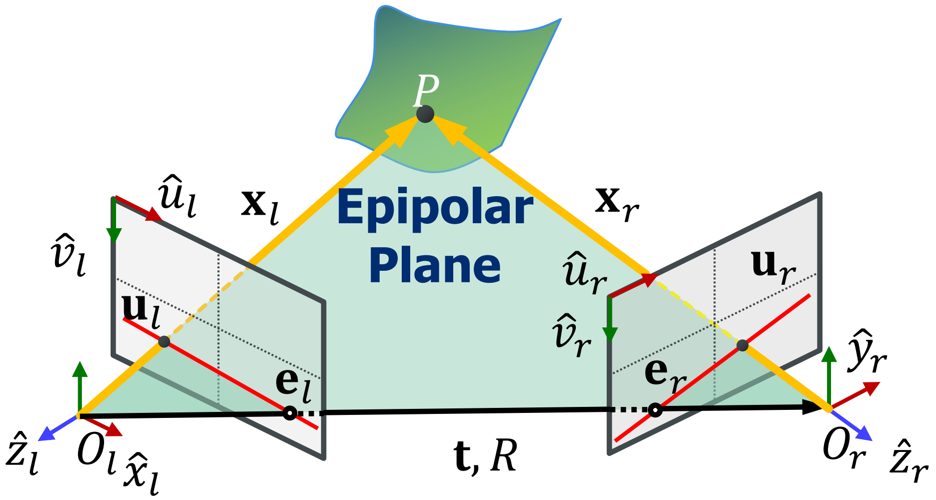
$\mathbf{e}_l$ and $\mathbf{e}_r$ are unique for a given stereo pair.

Epipolar Geometry: Epipolar Plane



The camera origins (O_l and O_r), the epipoles ($\mathbf{e}_l$ and $\mathbf{e}_r$) and any given scene point all lie on a plane called the Epipolar Plane.

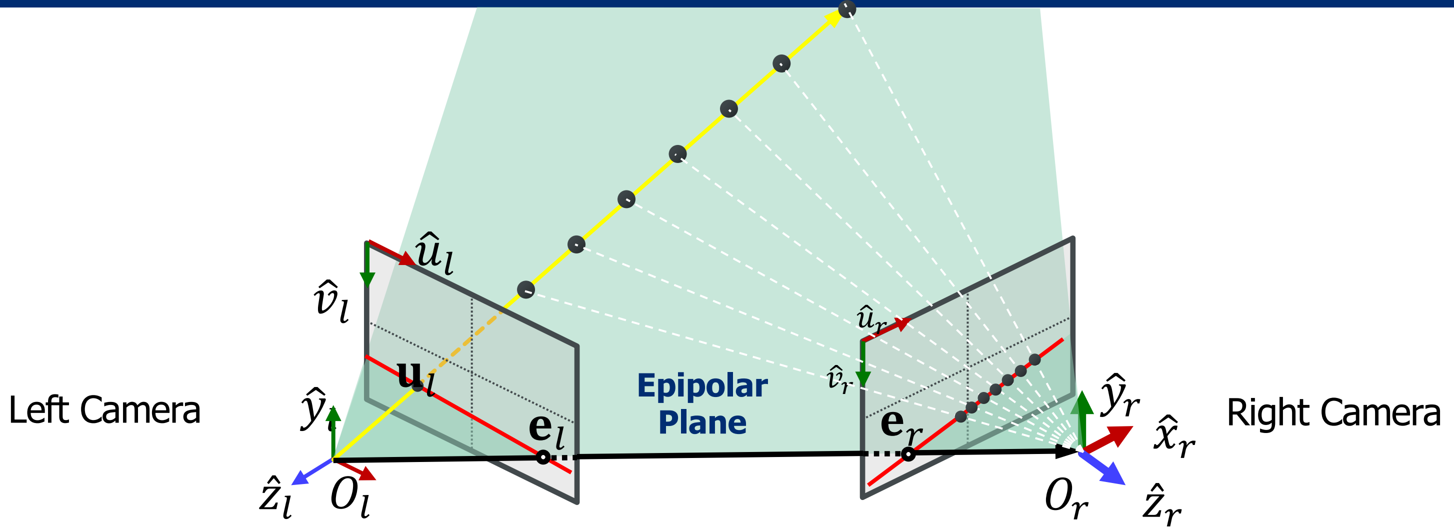
Epipolar Geometry: Epipolar Line



Intersection of the image plane and epipolar plane is the **Epipolar Line**.

Each scene point corresponds to two Epipolar Lines, one each on the two image planes.

Epipolar Geometry: Epipolar Constraint



Given a point in one image, the corresponding point in the other image must lie on the epipolar line.

Epipolar constraint reduces the problem of finding correspondence to a **1D search.**

Summary

Uncalibrated Multiview Reconstruction: Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may be unknown.

- Essential concepts in this lecture:
  - Epipolar geometry and how it relates to stereo
  - Epipoles
  - Epipolar lines

Johns Hopkins Engineering

Computer Vision

Uncalibrated Binocular Stereo

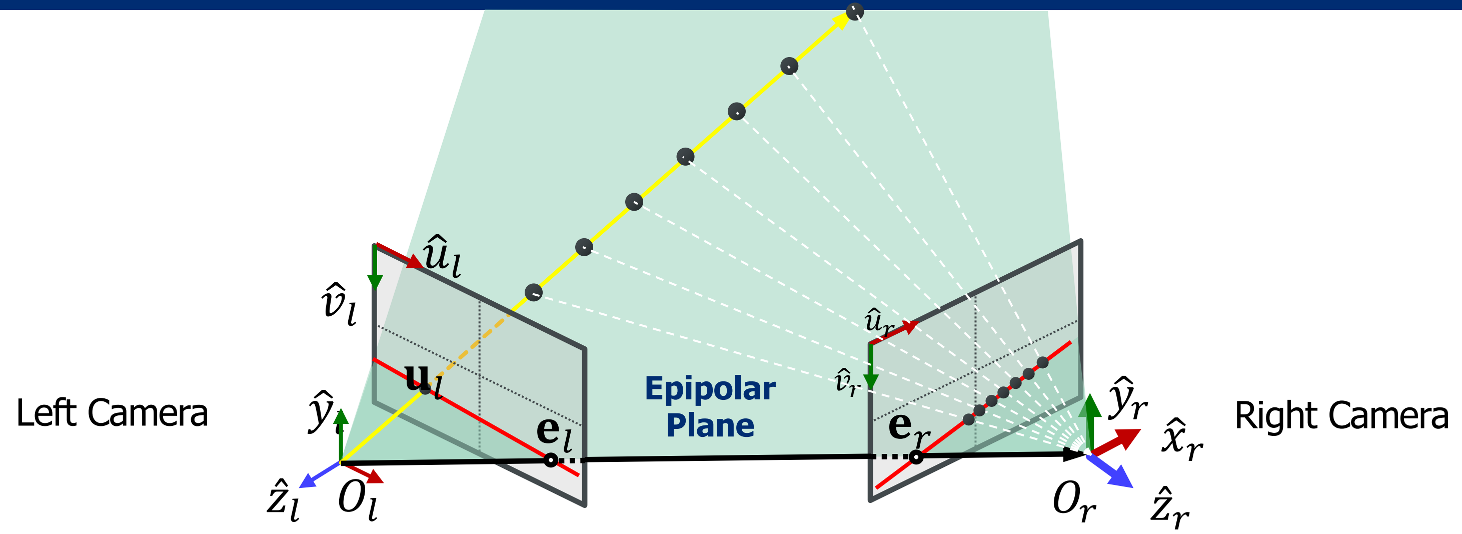
Uncalibrated Binocular Stereo

Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may not be known.

Topics:

- Essential Matrix and Fundamental Matrix

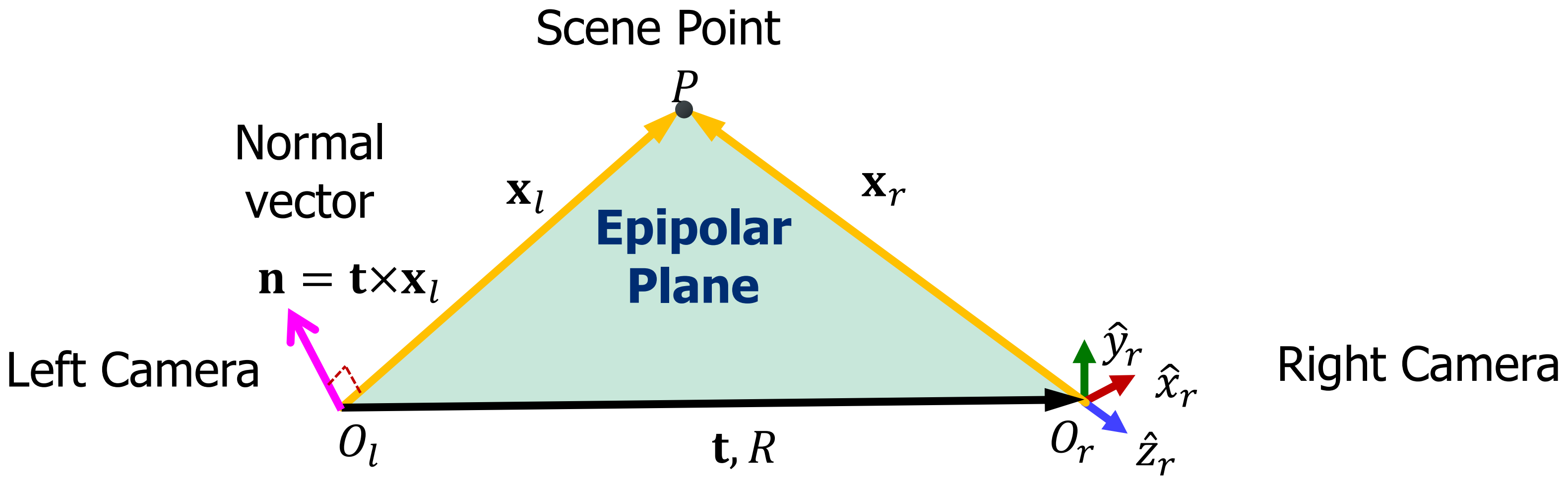
Epipolar Geometry: Epipolar Constraint



Given a point in one image, the corresponding point in the other image must lie on the epipolar line.

Epipolar constraint reduces the problem of finding correspondence to a **1D search.**

Epipolar Constraint



Vector normal to the epipolar plane: $\mathbf{n} = \mathbf{t} \times \mathbf{x}_l$

Dot product of $\mathbf{n}$ and $\mathbf{x}_l$ (perpendicular vectors) is zero.

$$\boxed{\mathbf{x}_l \cdot (\mathbf{t} \times \mathbf{x}_l) = 0}$$

Epipolar Constraint in Matrix Form

$$\mathbf{x}_l \cdot (\mathbf{t} \times \mathbf{x}_l) = 0$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{bmatrix} = 0$$

Cross-product definition

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \underbrace{\boxed{\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}}}_{T_\times} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0$$

Matrix-vector form

But we know that:

$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

The Epipolar Constraint

Substituting into the epipolar constraint gives:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \left(\boxed{\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \boxed{\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}} \right) = 0$$

$$\boldsymbol{t} \times \boldsymbol{t} = \boldsymbol{0}$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \boxed{\begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix}} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Essential Matrix E

$$\boxed{E = T_{\times} R}$$

The Essential Matrix E

Essential Matrix E: Relates position of scene point in left camera coordinate (x_l, y_l, z_l) to position in right camera coordinates(x_r, y_r, z_r)

$$\boxed{\mathbf{x}_l \cdot E\mathbf{x}_r \; = 0}$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

3D position in left camera coordinates

3x3 Essential Matrix

3D position in right camera coordinates

[Longuet-Higgins 1981]

Epipolar Constraint in Image Coordinates

Forward imaging equations:

$$\begin{bmatrix}\tilde{u}_l\\ \tilde{v}_l\\ \tilde{w}_l\end{bmatrix} = \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)}\\ 0 & f_y^{(l)} & o_y^{(l)}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_l\\ y_l\\ z_l\end{bmatrix}$$

$$\boxed{\tilde{\mathbf{u}}_l = K_l\,\mathbf{x}_l}$$

$$\begin{bmatrix}\tilde{u}_r\\ \tilde{v}_r\\ \tilde{w}_r\end{bmatrix} = \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)}\\ 0 & f_y^{(r)} & o_y^{(r)}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_r\\ y_r\\ z_r\end{bmatrix}$$

$$\boxed{\tilde{\mathbf{u}}_r = K_r\,\mathbf{x}_r}$$

Inverse imaging equations:

$$\begin{bmatrix}x_l\\ y_l\\ z_l\end{bmatrix} = \begin{bmatrix} \frac{1}{f_x^{(l)}} & 0 & -\frac{o_x^{(l)}}{f_x^{(l)}}\\ 0 & \frac{1}{f_y^{(l)}} & -\frac{o_y^{(l)}}{f_y^{(l)}}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\tilde{u}_l\\ \tilde{v}_l\\ \tilde{w}_l\end{bmatrix}$$

$$\boxed{\mathbf{x}_l = K_l^{-1}\tilde{\mathbf{u}}_l}$$

$$\begin{bmatrix}x_r\\ y_r\\ z_r\end{bmatrix} = \begin{bmatrix} \frac{1}{f_x^{(r)}} & 0 & -\frac{o_x^{(r)}}{f_x^{(r)}}\\ 0 & \frac{1}{f_y^{(r)}} & -\frac{o_y^{(r)}}{f_y^{(r)}}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\tilde{u}_r\\ \tilde{v}_r\\ \tilde{w}_r\end{bmatrix}$$

$$\boxed{\mathbf{x}_r = K_r^{-1}\tilde{\mathbf{u}}_r}$$

Epipolar Constraint in Image Coordinates

Rewriting the epipolar constraint:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Substituting with the inverse imaging equations gives:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \boxed{\underbrace{\begin{bmatrix} \frac{1}{f_x^{(l)}} & 0 & 0 \\ -\frac{o_x^{(l)}}{f_x^{(l)}} & \frac{1}{f_y^{(l)}} & 0 \\ 0 & -\frac{o_y^{(l)}}{f_y^{(l)}} & 1 \end{bmatrix}}_{\boxed{(K_l{}^{-1})^T}} \underbrace{\begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix}}_{\boxed{E}} \underbrace{\begin{bmatrix} \frac{1}{f_x^{(r)}} & 0 & -\frac{o_x^{(r)}}{f_x^{(r)}} \\ 0 & \frac{1}{f_y^{(r)}} & -\frac{o_y^{(r)}}{f_y^{(r)}} \\ 0 & 0 & 1 \end{bmatrix}}_{\boxed{K_r{}^{-1}}}} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

Epipolar Constraint in Image Coordinates

Rewriting the epipolar constraint:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0$$

Substituting with the inverse imaging equations gives:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \boxed{\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

Fundamental Matrix F

$$\boxed{F = (K_l^{-1})^T \, E \, K_r^{-1}}$$

The Fundamental Matrix F

Fundamental Matrix F: Relates position of scene point in left image $(u_l, v_l, 1)$ to position of the same scene point in the right image $(u_r, v_r, 1)$

$$\boxed{\tilde{\mathbf{u}}_l \cdot F\tilde{\mathbf{u}}_r = 0}$$

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0$$

Homogeneous 2D vector in left image coordinates

3x3 Fundamental Matrix

Homogeneous 2D vector in right image coordinates

[Longuet-Higgins 1981]

Scale of Fundamental Matrix F

Fundamental matrix acts on homogenous coordinates.

We know that: $\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix} \equiv k \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{bmatrix}$ ($k \neq 0$ is any constant)

That is: $$[u_l \quad v_l \quad 1] \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = [u_l \quad v_l \quad 1] k \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix}$$

Therefore, Fundamental Matrices F and kF produce the same epipolar constraint.

Fundamental Matrix F needs to be determined only up to a scale factor.

Epipolar Lines

If we know the Fundamental matrix F then,

- given a point (u_l, v_l) in the left image, we can find the line in the right image that the corresponding point must lie on,
- and, given a point (u_r, v_r) in the right image, we can find the line in the left image that the corresponding point must lie on.

Epipolar Lines from F Matrix

Given F and (u_r, v_r), the Epipolar Constraint Equation:

$$\underbrace{\begin{bmatrix} u_l & v_l & 1 \end{bmatrix}}_{\text{Unknown}} \underbrace{\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix}}_{\text{Known}} = 0$$

We can expand the matrix equation as:

$$(\underline{f_{11}u_r + f_{12}v_r + f_{13}})u_l + (\underline{f_{21}u_r + f_{22}v_r + f_{23}})v_l + (\underline{f_{31}u_r + f_{32}v_r + f_{33}}) = 0$$

$$\boxed{au_l + bv_l + c = 0}$$

Equation for left epipolar line

Epipolar Lines from F Matrix

Given F and (u_l, v_l), the Epipolar Constraint Equation:

$$\underbrace{\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}}_{\text{Known}} \underbrace{\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix}}_{\text{Unknown}} = 0$$

We can expand the matrix equation as:

$$\underline{(f_{11}u_l + f_{21}v_l + f_{31})}u_r + \underline{(f_{12}u_l + f_{22}v_l + f_{32})}v_r + \underline{(f_{13}u_l + f_{23}v_l + f_{33})} = 0$$

$$\boxed{a'u_r + b'v_r + c' = 0}$$

Equation for right epipolar line

Fundamental Matrix Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

and the left image point

$$\tilde{\boldsymbol{u}}_l = \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix}$$

Left Image

Right Image

The equation for the epipolar line in the right image is

$$\begin{bmatrix} u_r & v_r & 1 \end{bmatrix} \begin{bmatrix} -.003 & -.003 & 2.97 \\ -.028 & -.008 & 56.38 \\ 13.19 & -29.2 & -9999 \end{bmatrix} \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix} = 0$$

Fundamental Matrix Example

Given the Fundamental matrix,

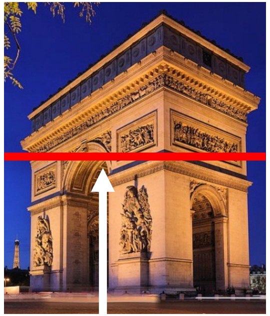
$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

and the left image point

$$\tilde{\boldsymbol{u}}_l = \begin{bmatrix} 343 \\ 221 \\ 1 \end{bmatrix}$$

Left Image

Right Image



Epipolar Line

The equation for the epipolar line in the right image is

$$.03u_r + .99v_r - 265 = 0$$

Fundamental Matrix Example

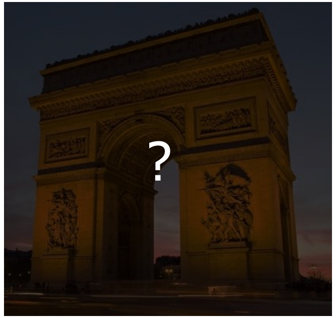
Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

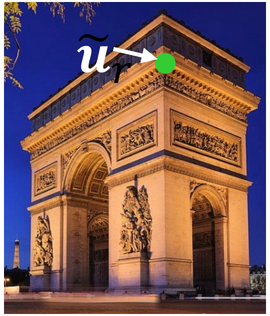
and the right image point

$$\tilde{\boldsymbol{u}}_r = \begin{bmatrix} 205 \\ 80 \\ 1 \end{bmatrix}$$

Left Image



Right Image



The equation for the epipolar line in the left image is

$$[u_l \quad v_l \quad 1] \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix} \begin{bmatrix} 205 \\ 80 \\ 1 \end{bmatrix} = 0$$

Fundamental Matrix Example

Given the Fundamental matrix,

$$F = \begin{bmatrix} -.003 & -.028 & 13.19 \\ -.003 & -.008 & -29.2 \\ 2.97 & 56.38 & -9999 \end{bmatrix}$$

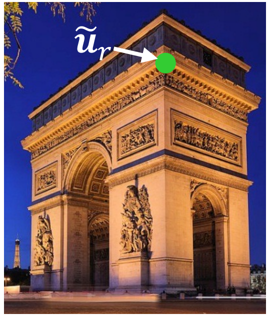
and the right image point

$$\tilde{\boldsymbol{u}}_r = \begin{bmatrix} 205 \\ 80 \\ 1 \end{bmatrix}$$

Left Image

Epipolar Line

Right Image



The equation for the epipolar line in the left image is

$$.32u_l - .95v_l - 151 = 0$$

Summary

Uncalibrated Multiview Reconstruction: Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may be unknown.

- Essential concepts in this lecture:
  - E matrix derivation
  - F matrix derivation
  - Computing epipolar lines

Johns Hopkins Engineering

Computer Vision

Uncalibrated Binocular Stereo

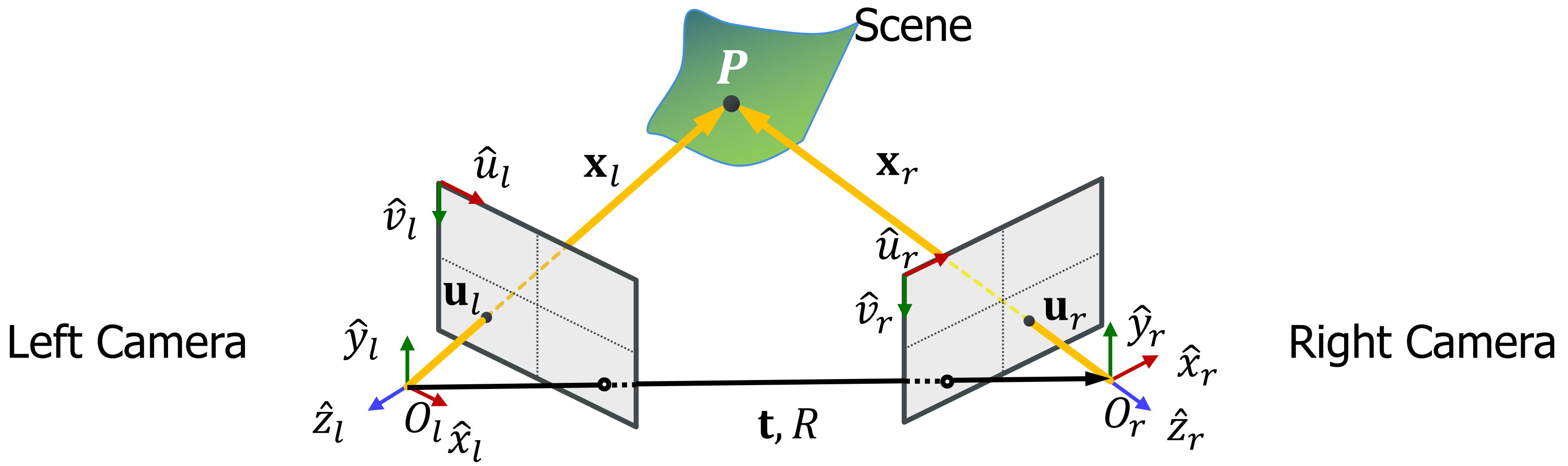
Uncalibrated Stereo

Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may not be known.

Topics:

- Stereo Self-Calibration
- Stereopsis and Multicamera reconstruction

Binocular Stereo



1. Assume Camera Intrinsic Parameters f_x, f_y, o_x, o_y are known.
2. Find Relative Camera Position $\mathbf{t}$ and Orientation R from the two images.
3. Find Correspondence for each pixel in the two images.
4. Compute Depth for each pixel using Triangulation.

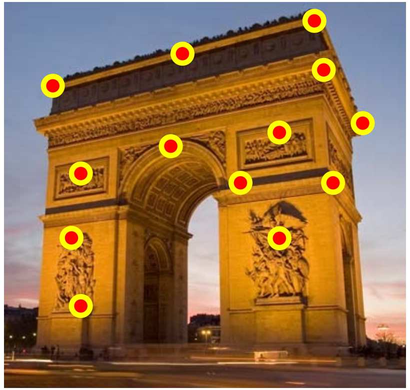
Stereo Calibration Using Fundamental Matrix

We use epipolar geometry to "**Calibrate**" the cameras to determine the relative camera position $\mathbf{t}$ and orientation R.

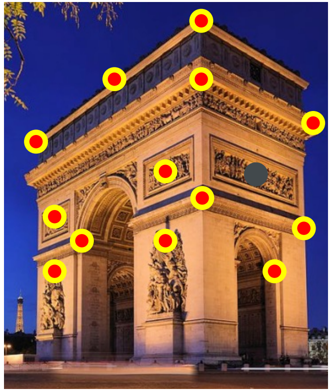
Stereo Calibration Procedure

Step 1: Find a set of features in left and right images (e.g. using SIFT)

Left image



Right image



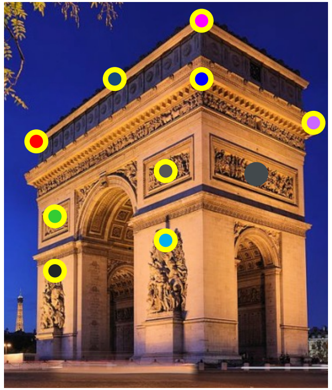
[Longuet-Higgins 1981]

Stereo Calibration Procedure

Step 2: Find correspondences by matching features.

Left image

Right image



$(\boldsymbol{u}_l^{(1)}, \boldsymbol{v}_l^{(1)})$

$\vdots$

$(\boldsymbol{u}_l^{(m)}, \boldsymbol{v}_l^{(m)})$

$(\boldsymbol{u}_r^{(1)}, \boldsymbol{v}_r^{(1)})$

$\vdots$

$(\boldsymbol{u}_r^{(m)}, \boldsymbol{v}_r^{(m)})$

[Longuet-Higgins 1981]

Stereo Calibration Procedure

Step 3: For each correspondence i, write out epipolar constraint

$$\underbrace{\begin{bmatrix} u_l^{(i)} & v_l^{(i)} & 1 \end{bmatrix}}_{\text{Known}} \underbrace{\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}}_{\text{Unknown}} \underbrace{\begin{bmatrix} u_r^{(i)} \\ v_r^{(i)} \\ 1 \end{bmatrix}}_{\text{Known}} = 0$$

Expand the matrix as linear equations

$$\left(f_{11}u_r^{(i)} + f_{12}v_r^{(i)} + f_{13}\right)u_l^{(i)} + \left(f_{21}u_r^{(i)} + f_{22}v_r^{(i)} + f_{23}\right)v_l^{(i)} + f_{31}u_r^{(i)} + f_{32}v_r^{(i)} + f_{33} = 0$$

[Longuet-Higgins 1981]

Stereo Calibration Procedure

Rearranging the terms:

$$\begin{bmatrix} u_l^{(1)}u_r^{(1)} & u_l^{(1)}v_r^{(1)} & u_l^{(1)} & v_l^{(1)}u_r^{(1)} & v_l^{(1)}v_r^{(1)} & v_l^{(1)} & u_r^{(1)} & v_r^{(1)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_l^{(i)}u_r^{(i)} & u_l^{(i)}v_r^{(i)} & u_l^{(i)} & v_l^{(i)}u_r^{(i)} & v_l^{(i)}v_r^{(i)} & v_l^{(i)} & u_l^{(i)} & u_r^{(i)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_l^{(m)}u_r^{(m)} & u_l^{(m)}v_r^{(m)} & u_l^{(m)} & v_l^{(m)}u_r^{(m)} & v_l^{(m)}v_r^{(m)} & v_l^{(m)} & u_l^{(m)} & u_r^{(m)} & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

A

(Known)

$\mathbf{f}$

(Unknown)

$$\boxed{A\,\mathbf{f} = \mathbf{0}}$$

Note F is rank deficient ➔ two zero singular vectors

[Longuet-Higgins 1981]

Extracting Essential Matrix

Step 5: Given the intrinsic parameters of the two cameras, compute essential matrix E from the fundamental matrix F.

From definition: $\boxed{F = (K_l^{-1})^T \; E \; K_r^{-1}}$

Therefore: $\boxed{E = K_l^T \; F K_r}$

$$\boxed{E = \begin{bmatrix} f_x^{(l)} & 0 & 0 \\ 0 & f_y^{(l)} & 0 \\ o_x^{(l)} & o_y^{(l)} & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} \\ 0 & f_y^{(r)} & o_y^{(r)} \\ 0 & 0 & 1 \end{bmatrix}}$$

Extracting Rotation and Translation

Step 6: Extract R and $\mathbf{t}$ from E

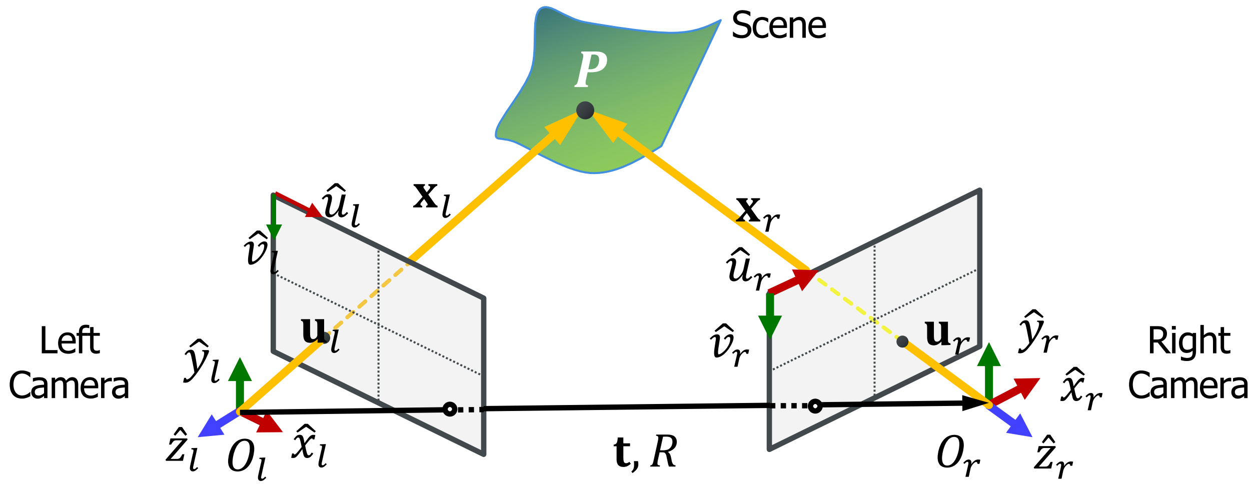
From definition, we know that:

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Given that $T_\times$ is a skew-symmetric matrix and R is an orthonormal matrix, it is possible to "decouple" $T_\times$ and R from their product using SVD factorization (see Szelinski, Chap. 7).

Note that translation is only known up to scale(!)

Binocular Stereo

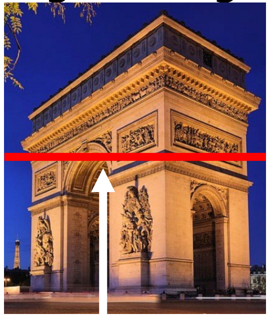


1. Assume Camera Intrinsic Parameters f_x, f_y, o_x, o_y are known.
2. Find Relative Camera Position $\mathbf{t}$ and Orientation R from the two images.
3. Find Correspondence for each pixel in the two images.
4. Compute Depth for each pixel using Triangulation.

Correspondence using Fundamental Matrix

Left Image

Right Image

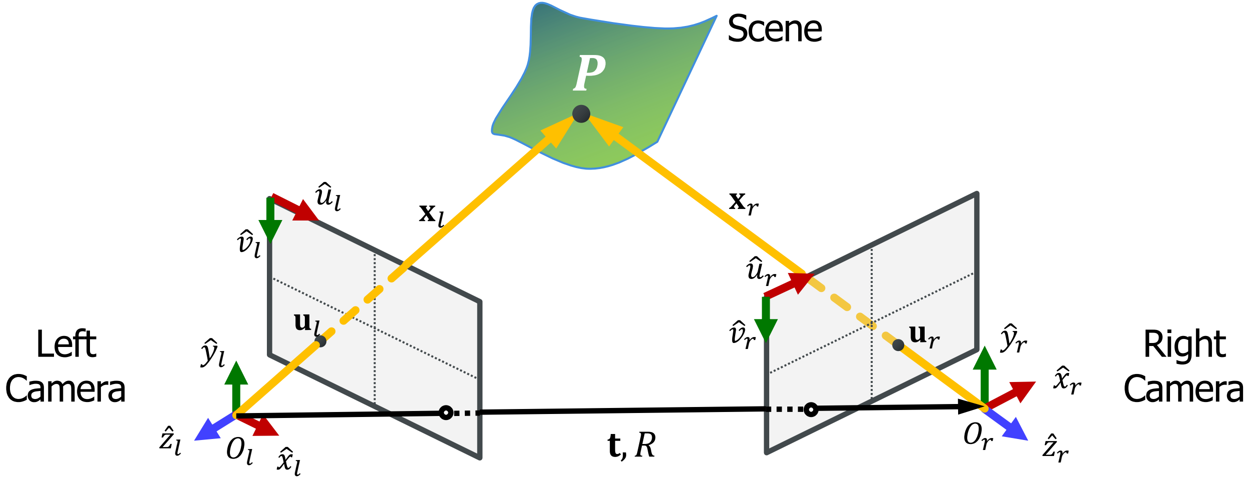


Epipolar Line

Given the Fundamental matrix and the left image point we can find the epipolar line in the right image and vice versa.

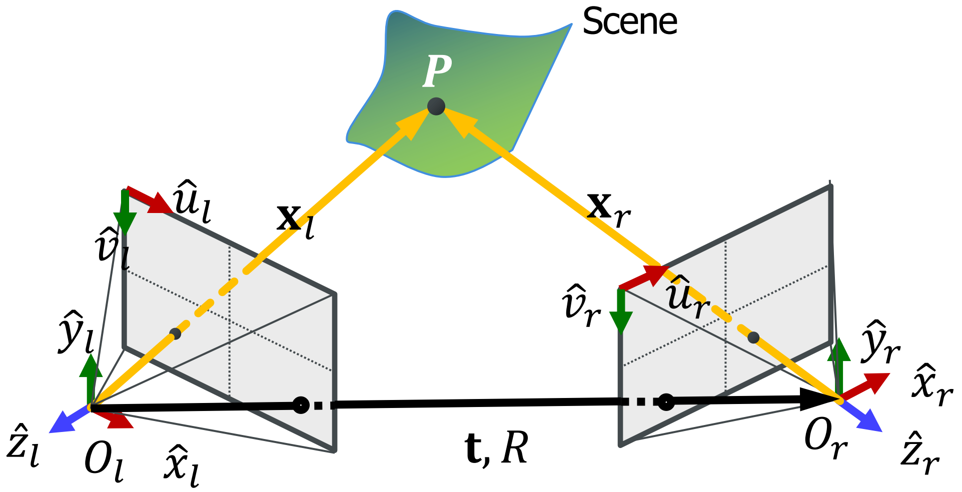
Perform template matching only along the epipolar line. (See Simple Stereo/Image Processing I lectures)

Binocular Stereo



1. Assume Camera Intrinsic Parameters f_x, f_y, o_x, o_y are known.
2. Find Relative Camera Position $\mathbf{t}$ and Orientation R from the two images.
3. Find Correspondence for each pixel in the two images.
4. Compute Depth for each pixel using Triangulation.

Computing Depth



Given the intrinsic parameters, the projection of scene points on to the image sensor is given by:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \qquad \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

Computing Depth

Left Camera Imaging Equation

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \boxed{\begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix}}$$

Right Camera Imaging Equation

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

We also know that relative position and orientation between the two cameras.

$$\boxed{\begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix}} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

Computing Depth

Left Camera Imaging Equation:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(l)} & 0 & o_x^{(l)} & 0 \\ 0 & f_y^{(l)} & o_y^{(l)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

$$\boxed{\tilde{\mathbf{u}}_{\boldsymbol{l}} = M_l \, \tilde{\mathbf{x}}_{\boldsymbol{r}}}$$

Right Camera Imaging Equation:

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} f_x^{(r)} & 0 & o_x^{(r)} & 0 \\ 0 & f_y^{(r)} & o_y^{(r)} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}$$

$$\boxed{\tilde{\mathbf{u}}_{\boldsymbol{r}} = P_r \, \tilde{\mathbf{x}}_{\boldsymbol{r}}}$$

Computing Depth

Expanding the imaging equations:

$$\tilde{\mathbf{u}}_{\boldsymbol{r}} = P_r\,\tilde{\mathbf{x}}_{\boldsymbol{r}} \qquad\qquad \tilde{\mathbf{u}}_{\boldsymbol{l}} = M_l\,\tilde{\mathbf{x}}_{\boldsymbol{r}}$$

$$\underbrace{\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\text{Known}} \underbrace{\begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}}_{\text{Unknown}} \qquad \underbrace{\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \equiv \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix}}_{\text{Known}} \underbrace{\begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix}}_{\text{Unknown}}$$

Rearranging the terms:

$$\begin{bmatrix} u_r p_{31} - p_{11} & u_r p_{32} - p_{12} & u_r p_{33} - p_{13} \\ v_r p_{31} - p_{21} & v_r p_{32} - p_{22} & v_r p_{33} - p_{23} \\ u_l m_{31} - m_{11} & u_l m_{32} - m_{12} & u_l m_{33} - m_{13} \\ v_l m_{31} - m_{21} & v_l m_{32} - m_{22} & v_l m_{33} - m_{23} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} p_{14} - p_{34} \\ p_{24} - p_{34} \\ m_{14} - m_{34} \\ m_{24} - m_{34} \end{bmatrix}$$

Computing Depth: Least Squares Solution

$$\underbrace{\begin{bmatrix} u_r p_{31} - p_{11} & u_r p_{32} - p_{12} & u_r p_{33} - p_{13} \\ v_r p_{31} - p_{21} & v_r p_{32} - p_{22} & v_r p_{33} - p_{23} \\ u_l m_{31} - m_{11} & u_l m_{32} - m_{12} & u_l m_{33} - m_{13} \\ v_l m_{31} - m_{21} & v_l m_{32} - m_{22} & v_l m_{33} - m_{23} \end{bmatrix}}_{A_{4\times 3}} \underbrace{\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix}}_{\mathbf{x}_r} = \underbrace{\begin{bmatrix} p_{14} - p_{34} \\ p_{24} - p_{34} \\ m_{14} - m_{34} \\ m_{24} - m_{34} \end{bmatrix}}_{\mathbf{b}_{4\times 1}}$$

$A_{4\times 3}$ (Known)

$\mathbf{x}_r$ (Unknown)

$\mathbf{b}_{4\times 1}$ (Known)

Find least squares solution using pseudo-inverse:

$$A\mathbf{x}_r = \mathbf{b}$$

$$A^T A\mathbf{x}_r = A^T\mathbf{b}$$

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

Note that reconstruction is only known up to scale!

Computing Depth: Bundle Adjustment

- What if I have lots of images?

$$\min_{\{\Sigma,D\}} \sum_{i} \sum_{p_j \in D} \mathrm{v}_{i,j} \left|\left|\Pi\left(H_i, p_j\right) - q_{i,j}\right|\right|^2$$

- Initialize this using pairwise estimates of camera poses, compute depths, and then optimize jointly with nonlinear optimization

- Szelinski, Chap 7.4

- Paper: "Structure-from-Motion Revisited, CVPR 2016https://demuc.de/papers/schoenberger2016sfm.pdf
  - Website: https://colmap.github.io/

Computing Depth: Bundle Adjustment

- Feature matching
- Triangulation and bundle adjustment

→ Reconstruction from acquired images

Snavely, N., Seitz, S. M., & Szeliski, R. (2006, July). Photo tourism: exploring photo collections in 3D. In ACM transactions on graphics (TOG) (Vol. 25, No. 3, pp. 835-846). ACM.

Results

Multiple views of the object

3D Structure

https://grail.cs.washington.edu/projects/mvscpc

Summary

Uncalibrated Multiview Reconstruction: Method to estimate 3D structure from two or more arbitrary images of a scene captured with cameras whose intrinsic parameters may be unknown.

- Essential concepts in this lecture:
  - Estimating F from correspondences
  - Using F and intrinsics to compute depths
  - Extending from two views to multiple views

Appendix A: Least Squares Solution for F

$$\boxed{\min_{\mathbf{f}} \|A\mathbf{f}\|^2 \quad \text{such that } \|\mathbf{f}\|^2 = 1}$$

We know that: $\|A\mathbf{f}\|^2 = (A\mathbf{f})^T(A\mathbf{f}) = \mathbf{f}^T A^T A\mathbf{f}$ and $\|\mathbf{f}\|^2 = \mathbf{f}^T\mathbf{f} = 1$

Create a Loss function $L(\mathbf{f})$ and find $\mathbf{p}$ that minimizes it.

$$\min_{\mathbf{f}}\{L(\mathbf{f}) = \mathbf{f}^T A^T A\mathbf{f} + \lambda(\mathbf{f}^T\mathbf{f} - 1)\}$$

Taking derivatives w.r.t $\mathbf{f}$ and λ: $\boxed{A^T A\mathbf{f} + \lambda\mathbf{f} = 0}$ Eigenvalue Problem

Clearly, eigenvector $\mathbf{f}$ with smallest eigenvalue λ of matrix $A^T A$ minimizes the loss function $L(\mathbf{f})$.

Appendix B: Using the SVD to Extract *R* and **t**

The SVD factorization of the Essential Matrix is

$$E = U\Sigma V^T$$
$$= \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}^T$$

Where U and V are orthonormal matrices, and $(\sigma_1, \sigma_2, \sigma_3)$ are the singular values of the matrix E

The stereo calibration parameters R and $\mathbf{t}$ can be calculated from the SVD of E using the equations

$$R = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}^T \qquad \mathbf{t} = \begin{bmatrix} u_{13} \\ u_{23} \\ u_{33} \end{bmatrix}$$

References: Textbooks

Robot Vision (Chapter 13)
Horn, B. K. P., MIT Press

Computer Vision: A Modern Approach (Chapter 10)
Forsyth, D and Ponce, J., Prentice Hall

Multiple View Geometry (Chapters 8-10)
Hartley, R. and Zisserman, A., Cambridge University Press

Computer Vision: Algorithms and Applications (Chapter 7)
Szeliski, R., Springer

An introduction to 3D Computer Vision (Chapter 3)
Cyganek, B., Siebert, J. P., Wiley Pub

References: Papers

[Longuet-Higgins 1981] H.C. Longuet-Higgins. "A computer algorithm for reconstructing a scene from two projections." Nature, 1981.

[Fagueras 1992] O. Fagueras. "What can be seen in three dimensions with an uncalibrated stereo rig?." European Conference on Computer Vision, 1992.

Image Credits

I.1 http://www.lasplash.com/publish/International_151/Hilton_Arc_de_Triomphe_Review-A_Parisian_Gem.php

I.2 http://nelietatravellingadventures.blogspot.com/2011/01/arc-de-triomphe-paris-france.html

I.3 http://vision.middlebury.edu/mview/eval/

I.3 http://grail.cs.washington.edu/projects/stfaces/

I.4-I.12 Adapted from Gregory, *Eye and Brain.*

I.13 http://xkcd.com/941/

JOHNS HOPKINS
WHITING SCHOOL *of* ENGINEERING