

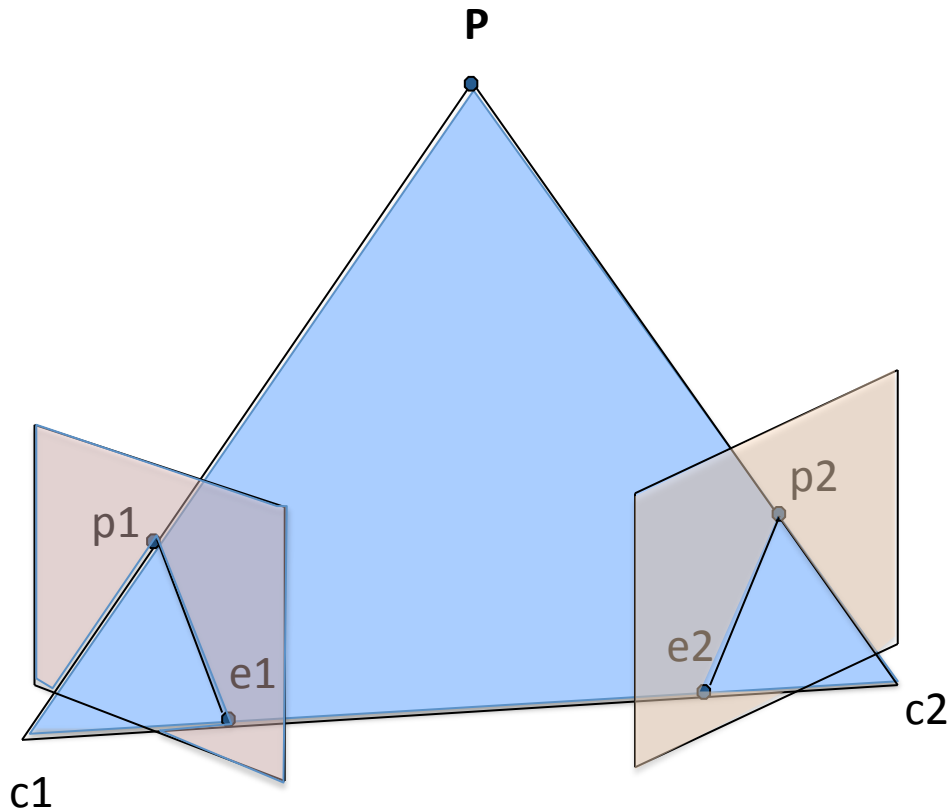
CS4243
Computer Vision
&
Pattern Recognition

Geometry
&
Deriving 3D from 2D Images

Deriving 3D Structure from Images

- Epipolar geometry
- Binocular stereo
- Depth from focus
- Focus of expansion
- Vanishing Points
- Structure from Motion

Epipolar Geometry (2 views)



P 3D point

c_1 optical center 1

c_2 optical center 2

p_1 image of **P** on image 1

p_2 image of **P** on image 2

e_1 epipole on image 1

e_2 epipole on image 2

Epipolar Geometry (2 views)

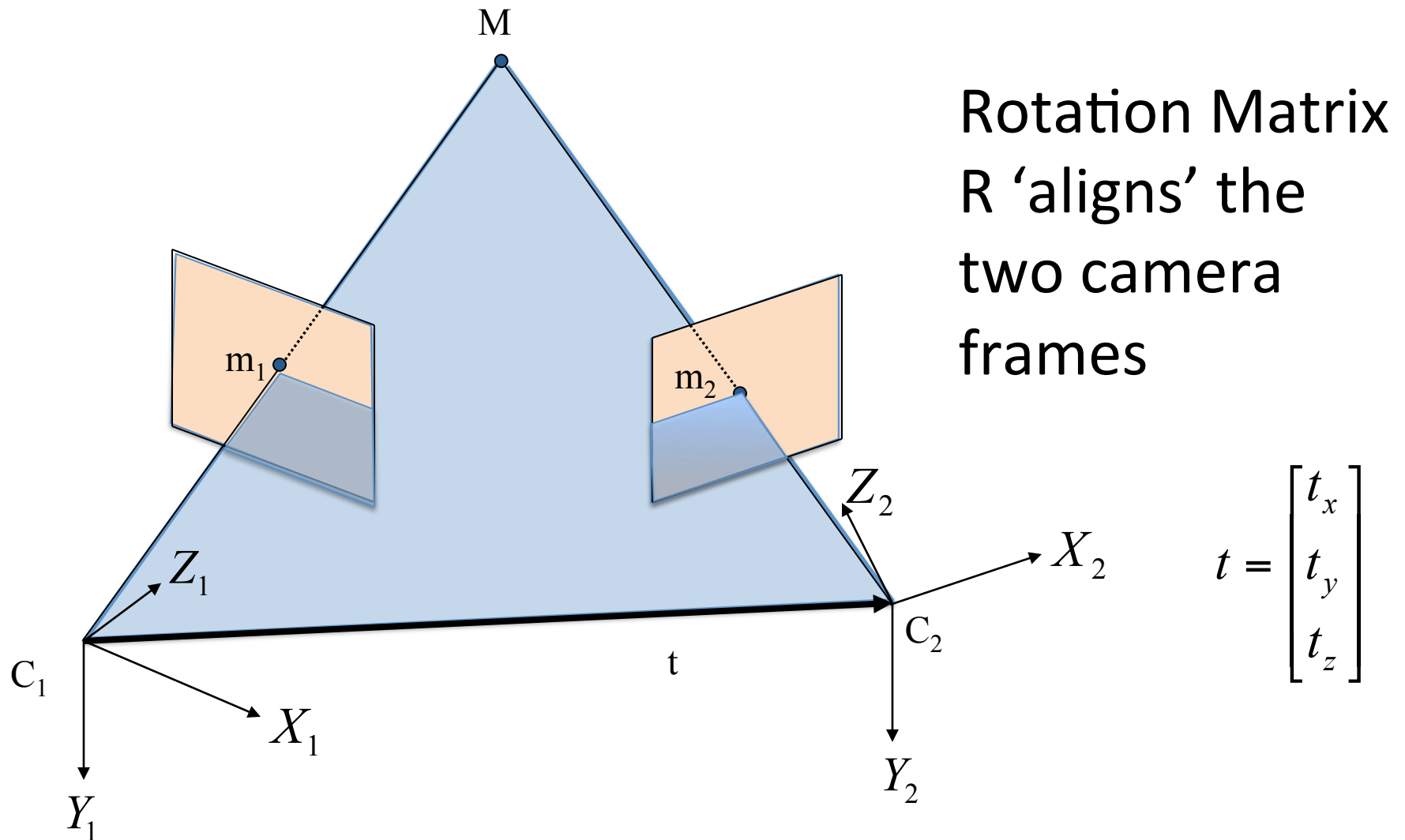
- The plane formed by c_1 , c_2 , and \mathbf{P} is called the epipolar plane
- The line passing through e_1 and p_1 is known as the epipolar line corresponding to image point p_2
- The line passing through e_2 and p_2 is known as the epipolar line corresponding to image point p_1
- Epipolar line e_1p_1 is the image of the 3D line $c_2\mathbf{P}$
- Epipolar line e_2p_2 is the image of the 3D line $c_1\mathbf{P}$
- All epipolar lines must pass through the epipole

Epipolar Geometry (2 views)

- Described by the Essential Matrix
- If the camera intrinsic parameters are not known, then a more general result is Fundamental Matrix.

Essential Matrix

Reference: “Three-Dimensional Computer Vision” by Faugeras.



$$m_1^T (t \wedge Rm_2) = 0$$

Cross product can be represented using a anti-symmetric matrix

$$\text{Let } T = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

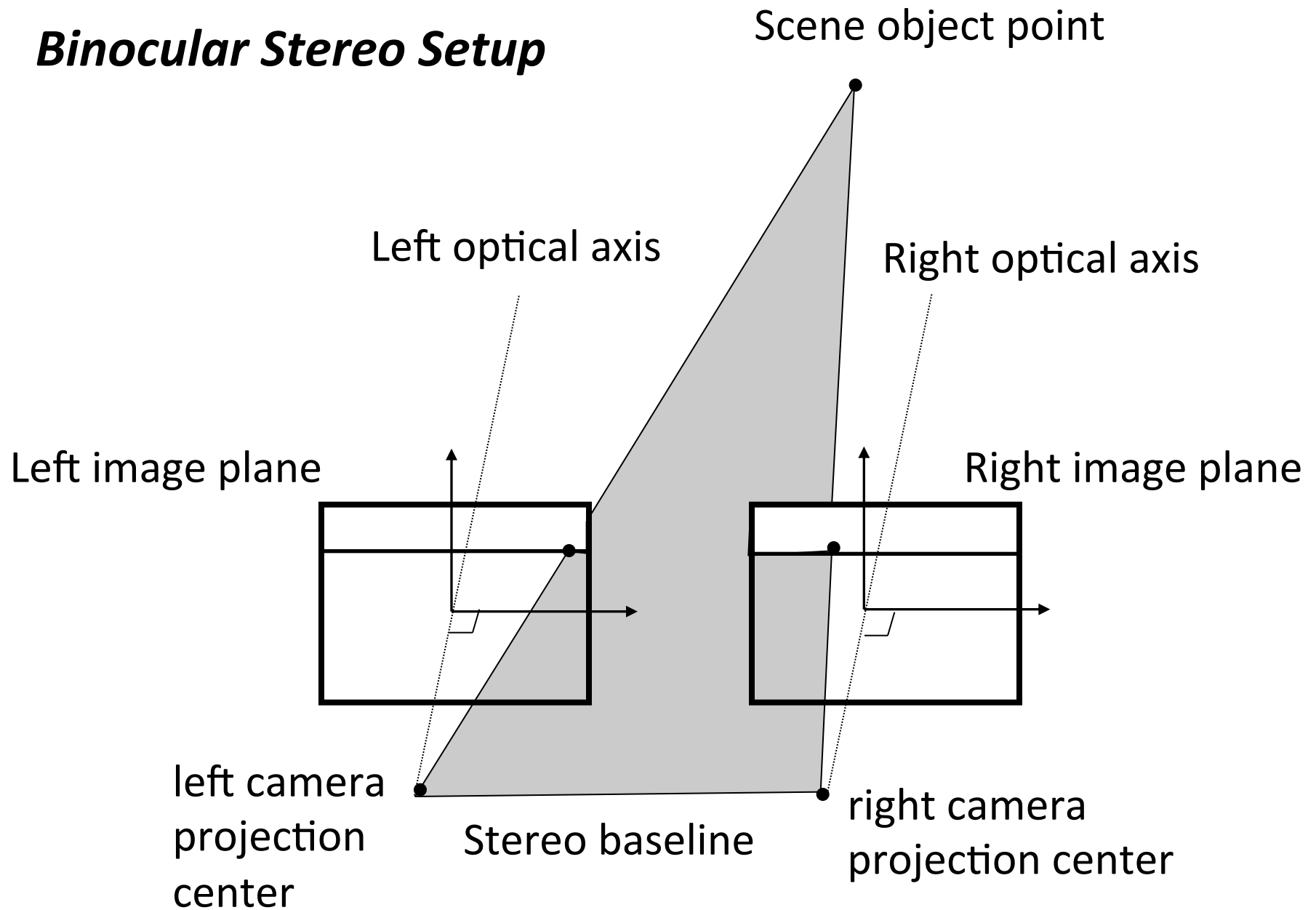
$$\text{then } m_1^T TR m_2 = 0$$

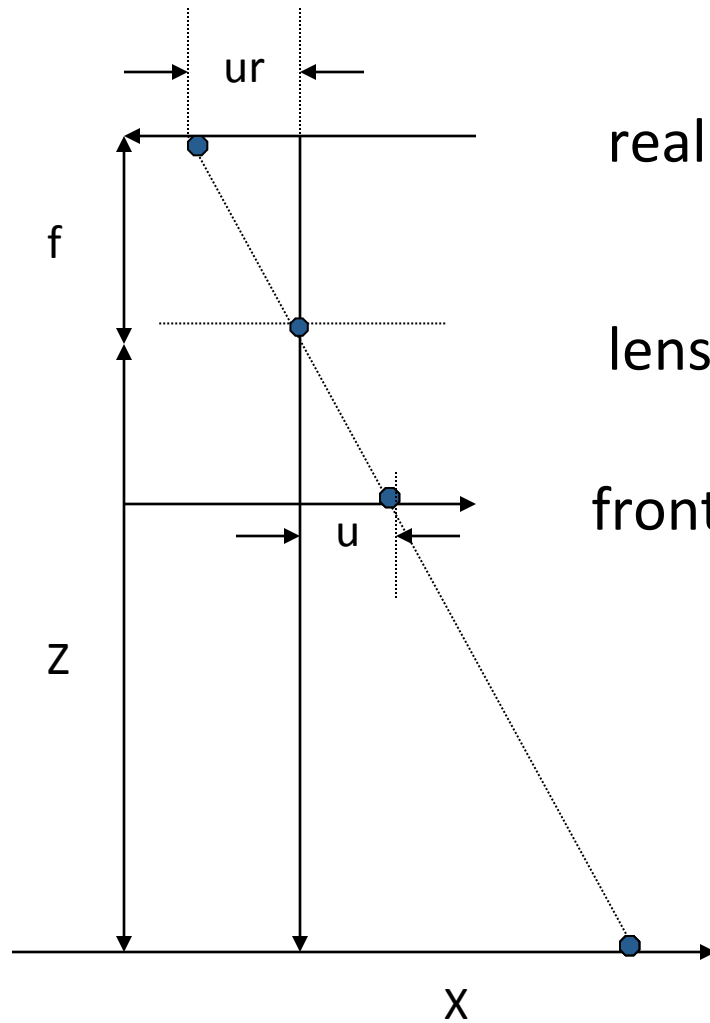
$$\text{Let } E = TR$$

$$\text{then } m_1^T E m_2 = 0$$

E is known as the Essential Matrix. $\det(E) = 0$.

Binocular Stereo Setup



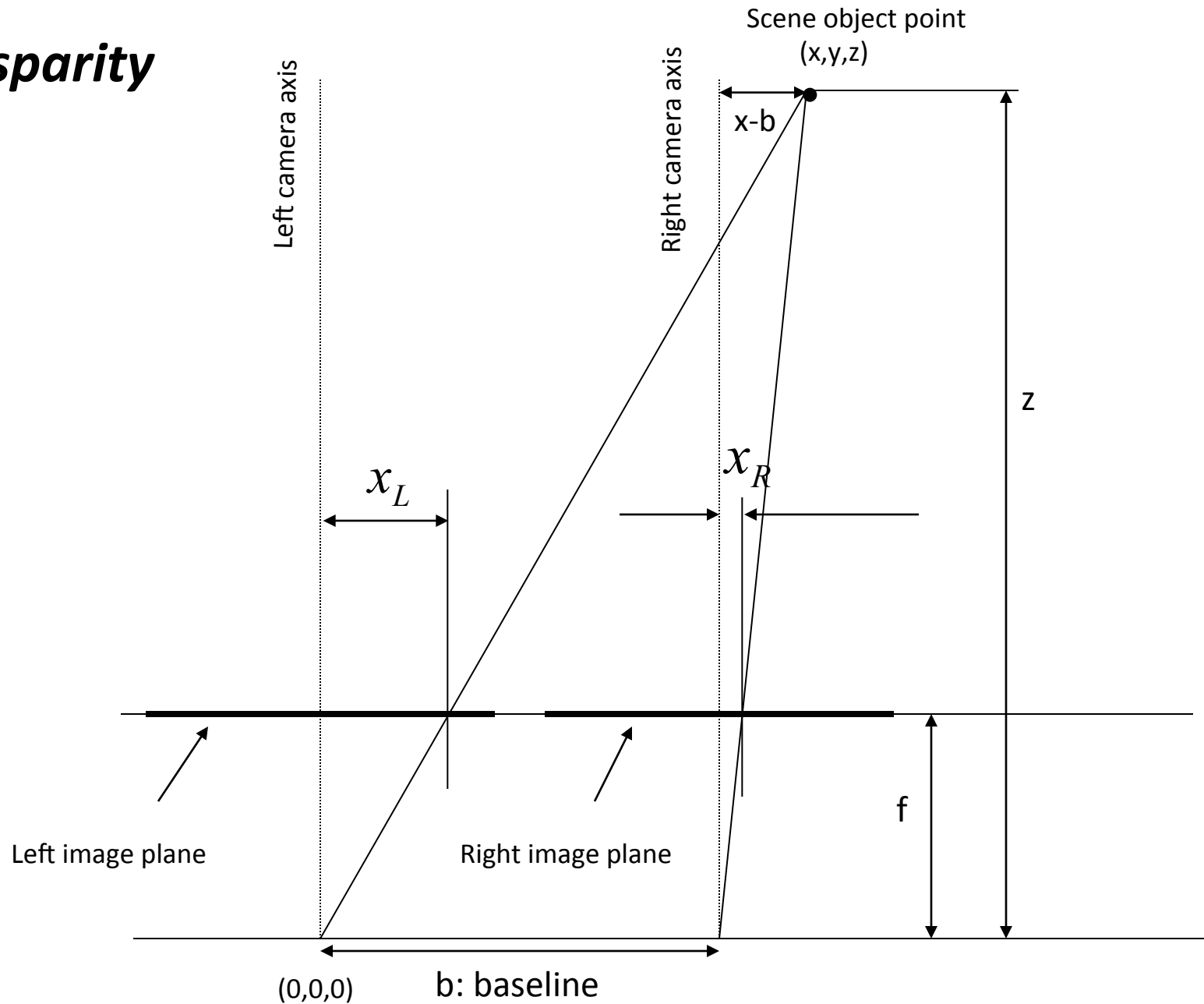


real image plane (image is inverted)

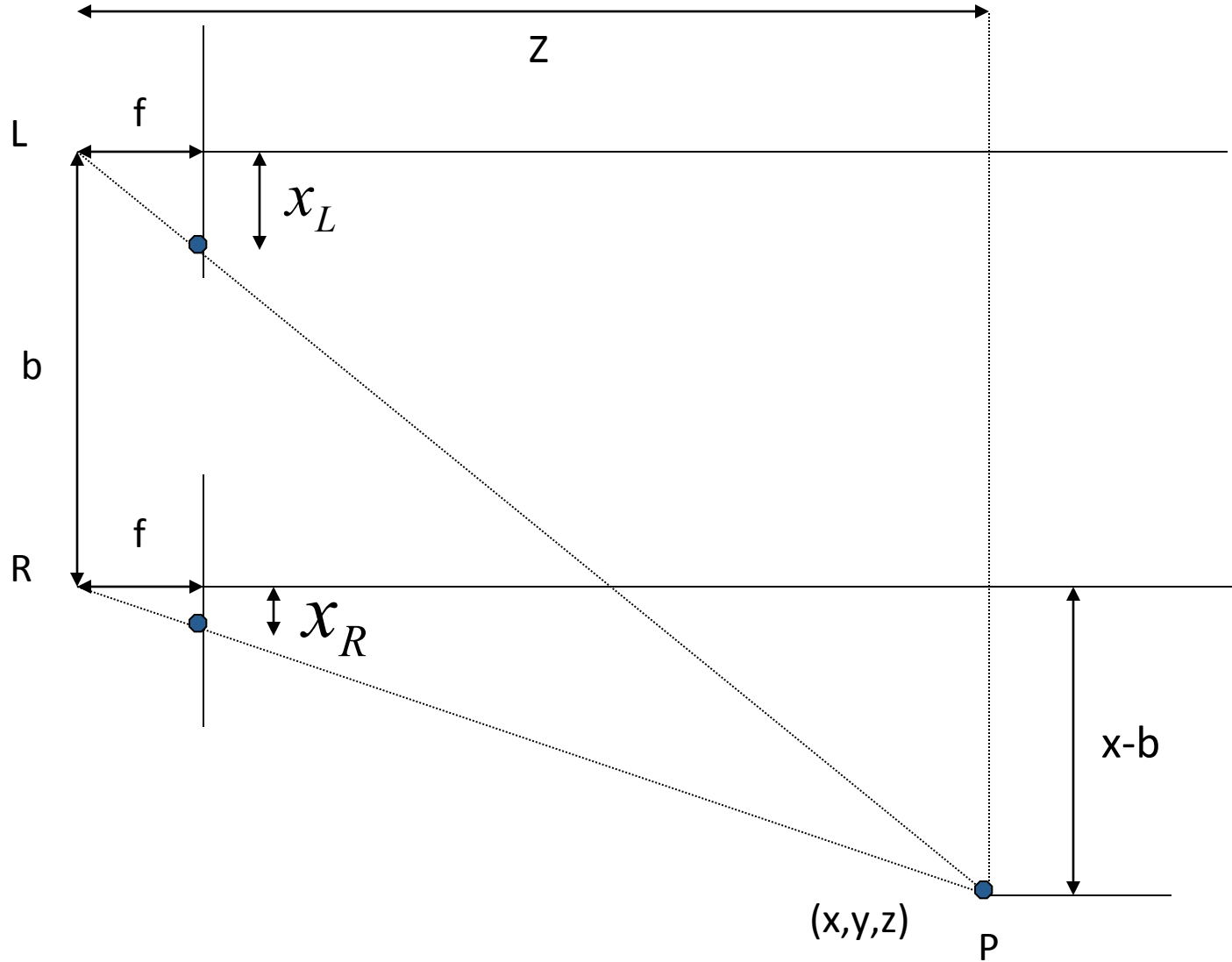
lens centre

front image plane (image is upright)

Disparity



Binocular Stereo Setup



Binocular Stereo (cont' d)

- disparity: $d = x_L - x_R$
- since camera is “aligned”, correspondence search is constrained to the same row in the left and right images i.e. The corresponding points have the same v-coordinates
- many existing systems, some of them real-time

Solving for depth

- Assume both image planes are parallel and aligned
- Given
 1. Focal length f
 2. Baseline distance b
- Let the global coordinates origin $(0,0,0)$ at left pinhole

TASK: Determine the z location of the scene point

Note: Once z is computed, the x and y locations of the scene point can be determined by considering perspective projection.

Geometry using left camera: $\frac{X}{Z} = \frac{x_L}{f}$

Geometry using right camera: $\frac{X - b}{Z} = \frac{x_R}{f}$

Eliminate x from the equations, get $Z = \frac{bf}{x_L - x_R}$

$(x_L - x_R)$ is known as the disparity

Binocular Stereo (cont' d)

- Error versus coverage

short baseline \rightarrow small error in image point location
result in large errors in computed
depth

but short baseline has more overlap regions (between the two images) and so enables a wider scene coverage in depth estimation.

Depth from Focus

Reference: “Computer Vision” section 12.4.3

- idea: change focal length, see which pixel has sharpest edge at which focal length
- Lenses with short focal length has good depth of field. Therefore, for depth from focus, should use long focal length

Focus Of Expansion

Reference: “Computer Vision” section 9.3

The images collected by a purely translating camera (i.e. camera does not rotate) exhibit optical flow vectors of various magnitude. When extended, these optical flow vectors will intersect at a single point, called the focus of expansion (FOE).

At the FOE, the optical flow is zero.

The FOE is the image of the ray along which the camera moves.

The concept of vanishing points

Reference: “Computer Vision” section 12.4.2

- vanishing point: the point where a 3D line appears to vanish
- parallel lines appear to intersect at the same vanishing point
- different groups of lines parallel to the **same plane** forms vanishing points that lie on a line, the vanishing line
- vanishing points can be used to calibrate cameras !
- can also be used to reconstruct 3D models !

v_1 , v_2 and v_3 are vanishing points. As all the lines are parallel to the same plane, these vanishing points form a straight line called the vanishing line.

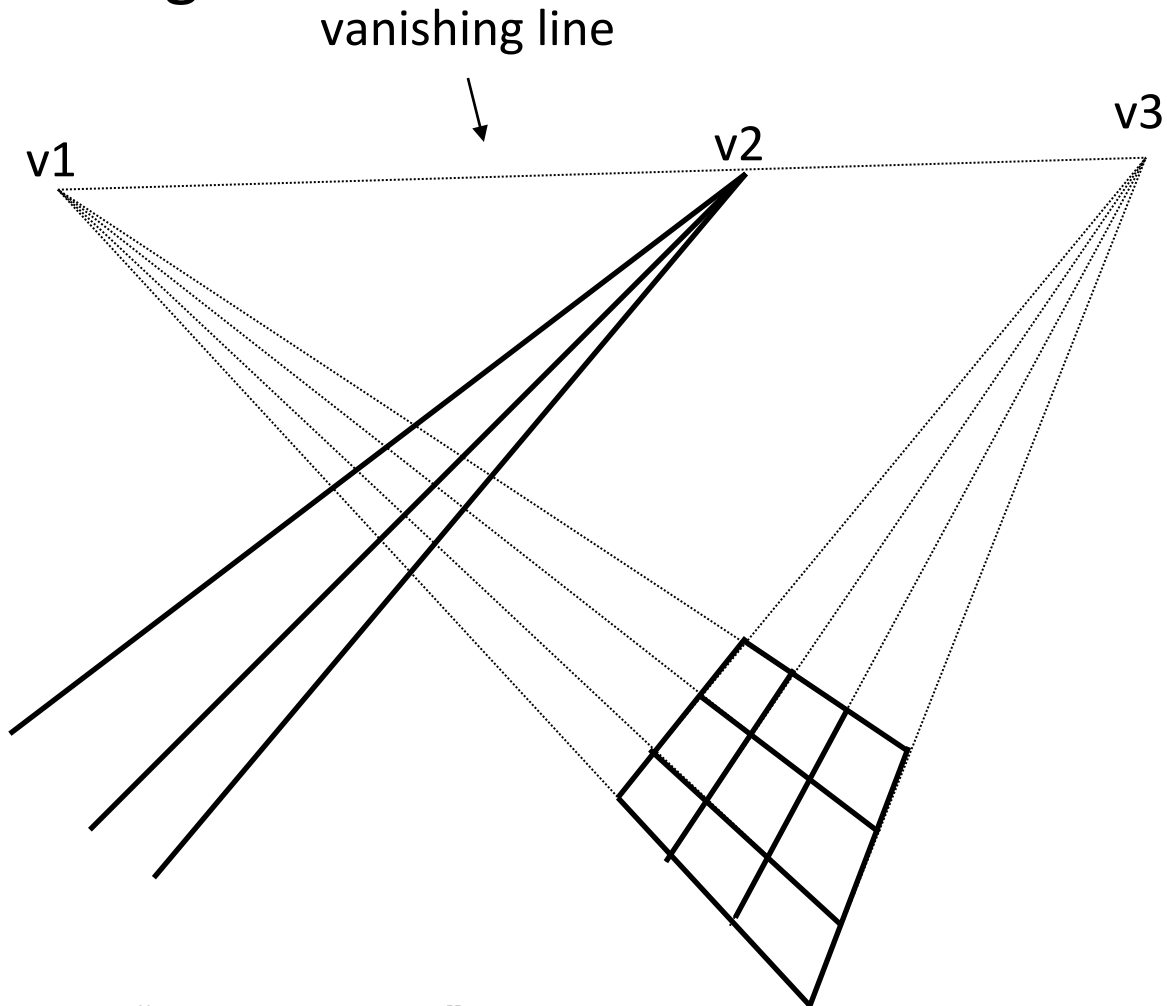
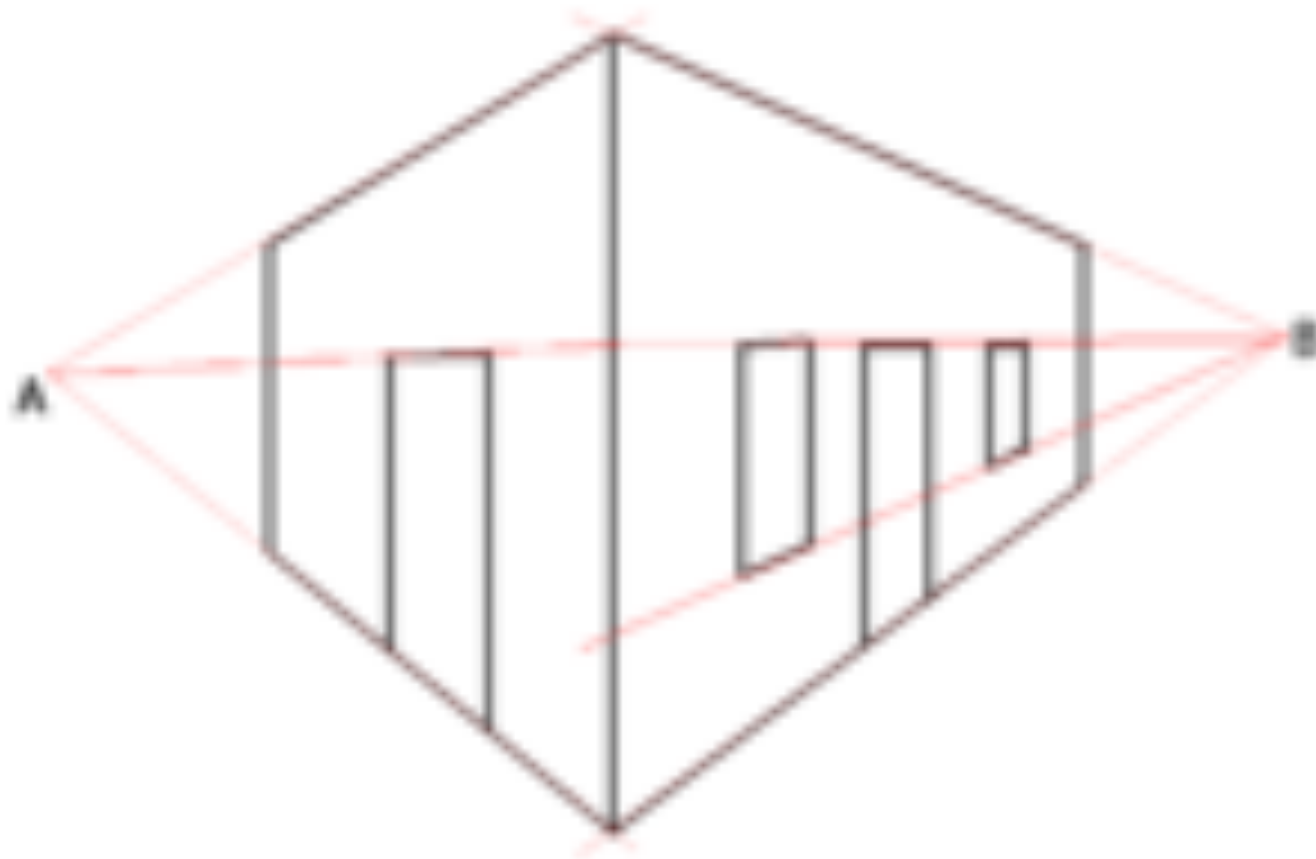


Figure 12.20 in “Computer Vision”



from http://en.wikipedia.org/wiki/Vanishing_point



from http://en.wikipedia.org/wiki/Vanishing_point



from http://en.wikipedia.org/wiki/Vanishing_point

From 2D Images to 3D Scenes

- In general, it is not possible to compute 3D shape from a single image.

Exception is when some a priori knowledge of scene is available eg. if we can tell certain lines are 3D parallel lines, certain planes are 3D vertical planes.

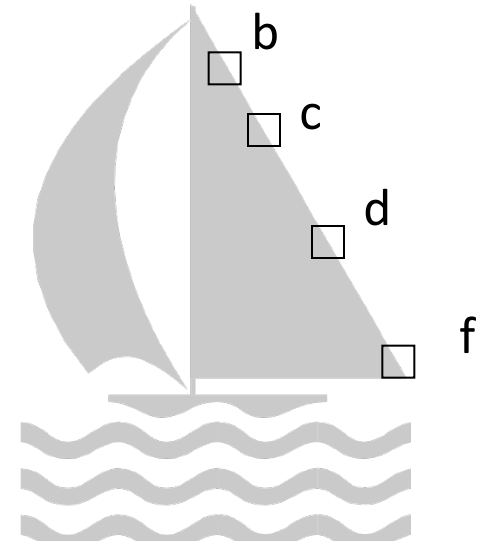
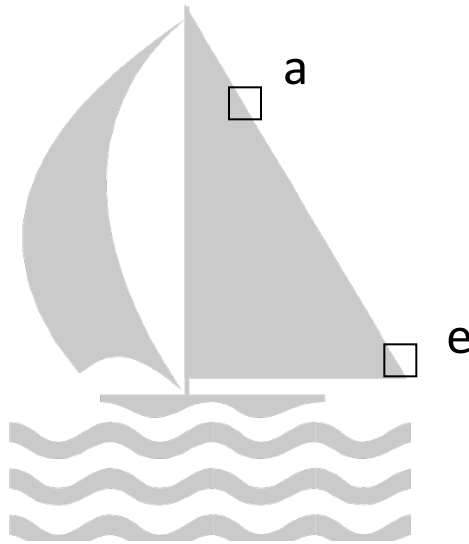
- Need at least two views to compute shape (eg. Binocular Stereo)
- Multiple views allow the recovery of both shape and camera motion (translation and rotation)

Structure from Motion

The Correspondence Problem

- In general, a single image cannot provide 3D information
- From a set of images taken with varying camera positions, we can extract 3D information of the scene. This requires us to match (associate) features in one image with the same feature in another image. The matching problem is called the correspondence problem.

Correspondence Problem: Good and bad features to Track



When epipolar line is unknown, straight edges are bad features for establishing feature correspondence. As in the diagram, feature 'a' can have multiple matches in the right image (eg. 'b', 'c', 'd'). On the other hand, feature 'e' is a good feature because it is unique.

Structure from Motion: The Factorization Algorithm

Reference:

Carlo Tomasi and Takeo Kanade, “Shape and Motion from Image Streams under Orthography: a Factorization Method”, International Journal of Computer Vision, 9:2, pages 137-154 (1992)

Orthographic model

$$u_{fp} = (s_p - t_f)^T i_f$$

$$v_{fp} = (s_p - t_f)^T j_f$$

F is total number of frames, N is total number of 3D points. (u,v) are the image coordinates of the 3D points.

Eg. (u_{53}, v_{53}) are the image coordinates of point 3 in frame 5.

Collect these images coordinates by feature tracking, then form the following matrix:

$$W = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ u_{21} & u_{22} & \cdots & u_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ u_{F1} & u_{F2} & \cdots & u_{FN} \\ v_{11} & v_{12} & \cdots & v_{1N} \\ v_{21} & v_{22} & \cdots & v_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ v_{F1} & v_{F2} & \cdots & v_{FN} \end{bmatrix}$$

Let
$$\bar{u}_f = \frac{1}{N} \sum_{p=1}^N u_{fp}$$

$$\bar{v}_f = \frac{1}{N} \sum_{p=1}^N v_{fp}$$

so

$$\begin{aligned}\bar{u}_f &= \frac{1}{N} \sum_{p=1}^N (s_p - t_f)^T i_f \\ &= \frac{1}{N} \sum_{p=1}^N (s_p^T i_f - t_f^T i_f) \\ &= \frac{1}{N} \sum_{p=1}^N s_p^T i_f - \frac{1}{N} \sum_{p=1}^N t_f^T i_f\end{aligned}$$

Without loss of generality, let

$$\sum_{p=1}^N S_p = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\text{so} \quad \bar{u}_f = - t_f^T i_f \quad \bar{v}_f = - t_f^T j_f$$

$$\begin{aligned} \hat{u}_{fp} &= u_{fp} - \bar{u}_f & \hat{v}_{fp} &= v_{fp} - \bar{v}_f \\ &= S_p^T i_f & &= S_p^T j_f \end{aligned}$$

form the following matrix:

$$\widehat{\mathcal{W}} = \begin{bmatrix} \widehat{u}_{11} & \widehat{u}_{12} & \cdots & \widehat{u}_{1N} \\ \widehat{u}_{21} & \widehat{u}_{22} & \cdots & \widehat{u}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{u}_{F1} & \widehat{u}_{F2} & \cdots & \widehat{u}_{FN} \\ \widehat{v}_{11} & \widehat{v}_{12} & \cdots & \widehat{v}_{1N} \\ \widehat{v}_{21} & \widehat{v}_{22} & \cdots & \widehat{v}_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{v}_{F1} & \widehat{v}_{F2} & \cdots & \widehat{v}_{FN} \end{bmatrix}$$

Observation:

$$\hat{W} = \begin{bmatrix} i_1^T \\ i_2^T \\ \vdots \\ i_F^T \\ j_1^T \\ j_2^T \\ \vdots \\ j_F^T \end{bmatrix} \begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_N \end{bmatrix}$$

← Camera orientation

3D shape points

$$2F \text{ by } 3 \rightarrow \begin{matrix} = M S \\ \leftarrow 3 \text{ by } N \end{matrix}$$

Observation:

\hat{W} has a rank less than or equal to 3

Using singular value decomposition (SVD), get

$$\hat{W} = U \Sigma V^T$$

note that U is $2F \times 2F$

Σ is $2F \times N$

V is $N \times N$

Σ is diagonal, with the first three diagonal entries non-zero, and the other diagonal entries close to zero.

Let Σ' be a 3x3 sub-matrix of Σ (taken from the top left hand corner of Σ).

Let U' be the first three columns of U ,
 V' be the first three columns of V

We have

$$\begin{aligned}\hat{w} &\approx U' \Sigma' V'^T \\ &= U' \Sigma'^{\frac{1}{2}} \Sigma'^{\frac{1}{2}} V'^T \\ &= \tilde{M} \tilde{S}\end{aligned}$$

where

$$\tilde{M} = U' \Sigma'^{\frac{1}{2}} \quad \tilde{S} = \Sigma'^{\frac{1}{2}} V'^T$$

There is ambiguity, because

$$\tilde{M} \tilde{S} = \tilde{M} A A^{-1} \tilde{S}$$

for any non-singular A

So need to find a matrix A (3X3) that will give a geometrically “correct” (i.e. Euclidean) solution. This is achieved by satisfying some constraints on the camera orientation vectors:

Constraints:

$$i_f^T i_f = 1 \quad j_f^T j_f = 1 \quad i_f^T j_f = 0$$

let

$$\tilde{M} = \begin{bmatrix} \tilde{m}_1^T \\ \tilde{m}_2^T \\ \vdots \\ \tilde{m}_F^T \\ \tilde{n}_1^T \\ \tilde{n}_2^T \\ \vdots \\ \tilde{n}_F^T \end{bmatrix}$$

$$(\tilde{m}_i^T A) (\tilde{m}_i^T A)^T = 1$$

$$\Rightarrow \tilde{m}_i^T Q \tilde{m}_i = 1 \quad \longleftarrow \text{This is first equation}$$

where $Q = A A^T$

$$= \begin{bmatrix} q_1 & q_2 & q_3 \\ q_2 & q_4 & q_5 \\ q_3 & q_5 & q_6 \end{bmatrix}$$

similarly

$j_i^T j_i = 1$ requires

$$(\tilde{m}_i^T A)(\tilde{m}_i^T A)^T = 1$$

$$\Rightarrow \tilde{n}_i^T Q \tilde{n}_i = 1$$



This is second equation

$i_i^T j_i = 0$ requires

$$(\tilde{m}_i^T A)(\tilde{n}_i^T A)^T = 0$$

$$\Rightarrow \tilde{m}_i^T Q \tilde{n}_i = 0$$



This is third equation

Use the three equations to solve for Q, then solve for A, then compute

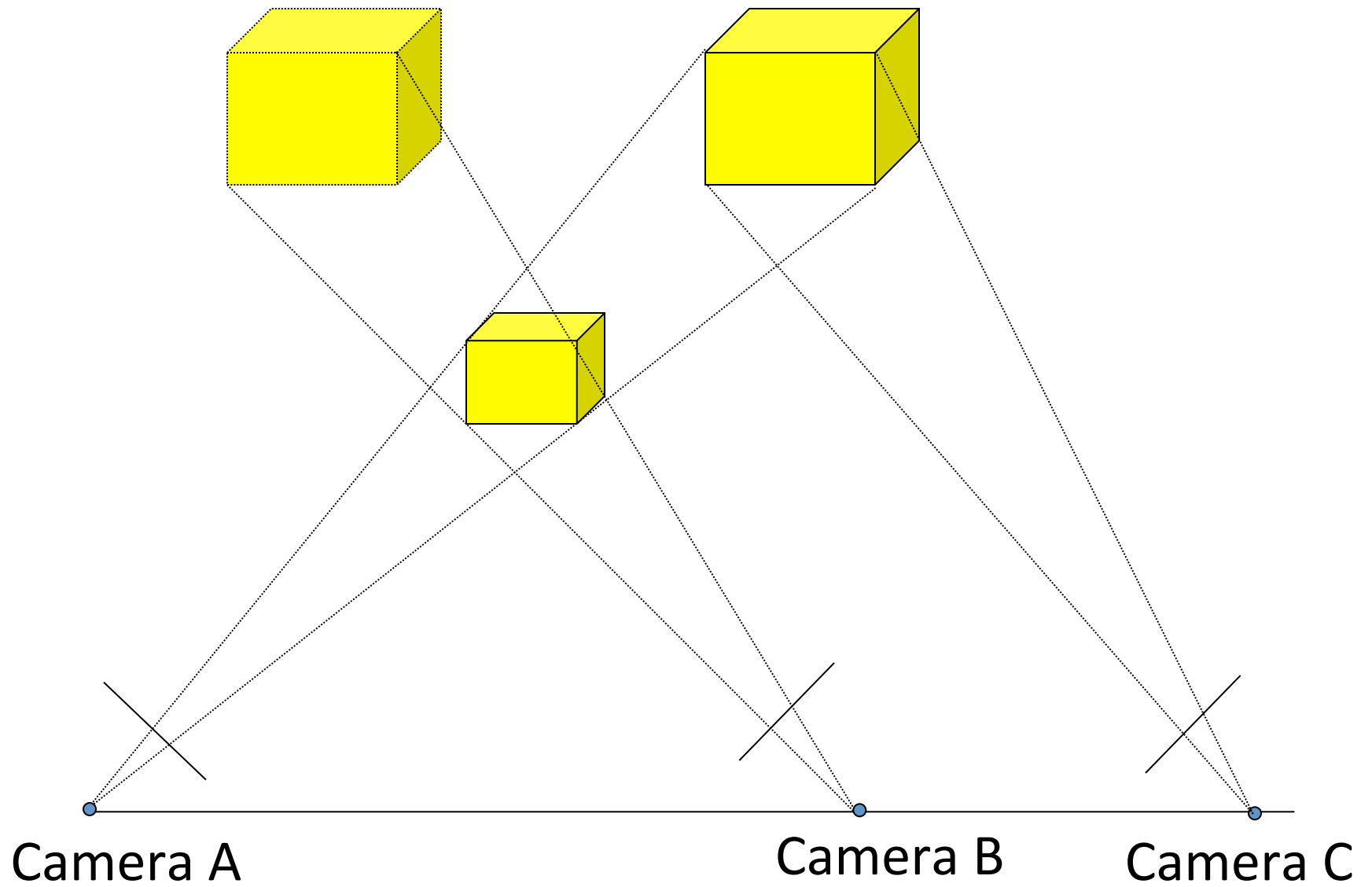
$$\dot{i}_i^T = \tilde{m}_i^T A$$

$$\dot{j}_i^T = \tilde{n}_i^T A$$

Recovered orientation
of camera

Recoverd 3D shape \longrightarrow $S_p = A^{-1} \tilde{S}_p$

Note that the recovered shape has scale ambiguity.
See the next two slides.



Camera B and camera C see exactly the same image. Therefore if we are only given images from Camera A and Camera B, there is no way we can calculate the absolute distance between camera A and camera B from the images alone. This ambiguity in distance also translates to an ambiguity in the scale of the reconstructed object – if camera B is the actual camera used, then the object is the smaller box; if camera C is used, then the object should be the bigger box.

Structure from Motion (5pt algorithm)

General Idea:

- solve for Essential Matrix E
- solve for T and R using $E = TR$
- solve for 3D points by triangulation
- nonlinear refinement

Method 1a for Solving for Essential Matrix E

8 point Algorithm

use Epipolar constraint to solve for
Fundamental Matrix F

$$m_2^T F m_1 = 0$$

then solve for Essential Matrix E using

$$E = K_2^T F K_1$$

where

$$K_1 = \begin{bmatrix} f_1 & s_1 & u_1 \\ 0 & f_1 & v_1 \\ 0 & 0 & 1 \end{bmatrix}$$

camera-1 internal
calibration matrix

$$K_2 = \begin{bmatrix} f_2 & s_2 & u_2 \\ 0 & f_2 & v_2 \\ 0 & 0 & 1 \end{bmatrix}$$

camera-2 internal
calibration matrix

Method 1b for Solving for Essential Matrix E

use Epipolar constraint to solve for Essential Matrix E directly

$$\hat{m}_2^T E \hat{m}_1 = 0$$

where

$$\hat{m}_1 = K_1^{-1} m_1 \quad \hat{m}_2 = K_2^{-1} m_2$$

Method 2 for Solving for Essential Matrix E

5 point Algorithm

$$\hat{m}_2^T E \hat{m}_1 = 0$$

epipolar constraint. 5 points give 5 eqns

$$\det(E) = 0$$

Singularity constraint

$$EE^T E - \frac{1}{2} \text{trace}(EE^T) E = 0$$

this constraint gives 9 eqns, but only 2 are independent

Proof of Why E has 2 identical singular values

$$E = \hat{T} R$$

there exist R_0 such that

$$R_0 T = \begin{bmatrix} 0 \\ 0 \\ \|T\| \end{bmatrix} = A$$

$$T = R_0^T A$$

$$TR_0 = R_0^T A R_0$$

Let $\hat{T} = TR_0$

$$\hat{A} = A R_0$$

$$\hat{T} = R_0^T \hat{A}$$

$$E E^T = \hat{T} R R^T \hat{T}^T$$

$$E E^T = \hat{T} \hat{T}^T$$

$$E E^T = R_0^T \hat{A} \hat{A}^T R_0$$

$$= R_0^T \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \|T\| & 0 \\ -\|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} R_0$$


$$= R_0^T \begin{bmatrix} \|T\|^2 & 0 & 0 \\ 0 & \|T\|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} R_0$$

Therefore,

Non-zero eigenvalues of $EE^T = \|T\|^2, \|T\|^2$

Non-zero singular values of $E = \|T\|, \|T\|$

Therefore,


$$EE^T E - \frac{1}{2} \text{trace}(EE^T) E = 0$$

A 3×3 matrix is an Essential Matrix if and only if 2 of its singular values are equal and the third is zero.

To get T and R from E , there are 4 possible solutions, all with scale ambiguity.

Let

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{rotation matrix}$$

$$Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{array}{l} \text{skew symmetric} \\ \text{matrix} \end{array}$$

$$E = U D V^T = T R$$

Translation: $T = U Z U^T$

Rotation: $R = U W V^T$

or

$$R = U W^T V^T$$

$$TR = UZU^T UWV^T = UZWV^T$$

$$TR = UZU^T UW^T V^T = UZW^T V^T$$

$$Z W = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$Z W^T = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that F , the fundamental matrix, has ambiguity in scale, so it only has 7 independent parameters (matrix elements).

Therefore, need only 7 points to solve for elements in F .

In reality, we use 8 points as 8 points give a more convenient solution.

Note that E , the essential matrix, also has ambiguity in scale and it only has 5 independent parameters (matrix elements).

Therefore, need only 5 points to solve for elements in E .

We use other constraints to enforce the properties of E .