Geometry and Physics (at a run)

D.A. Forsyth, University of Illinois at Urbana Champaign

SOME APPLICATION SLIDES MISSING

Main points

- Geometry
  - camera model
  - multi camera relations
- Fitting:
  - correspondence is reliable
  - yields robust fundamental matrix estimation
- Physics:
  - relegated to notes
- Yields
  - Mosaics
  - 3D reconstructions
  - Odometry
  - SLAM

Homogeneous coordinates

- Add an extra coordinate and use an equivalence relation
- for 2D
  - equivalence relation
    k*(X,Y,Z) is the same as (X,Y,Z)
- for 3D
  - equivalence relation
    k*(X,Y,Z,T) is the same as (X,Y,Z,T)
- “Ordinary” or “non-homogeneous” coordinates
  - properly called affine coordinates
  - in 3D, affine -> homogeneous $(x, y, z) \to k * (x, y, z, 1)$
  - in 3D, homogeneous to affine $(X, Y, Z, T) \to \left(\frac{X}{T}, \frac{Y}{T}, \frac{Z}{T}\right)$

Homogeneous coordinates

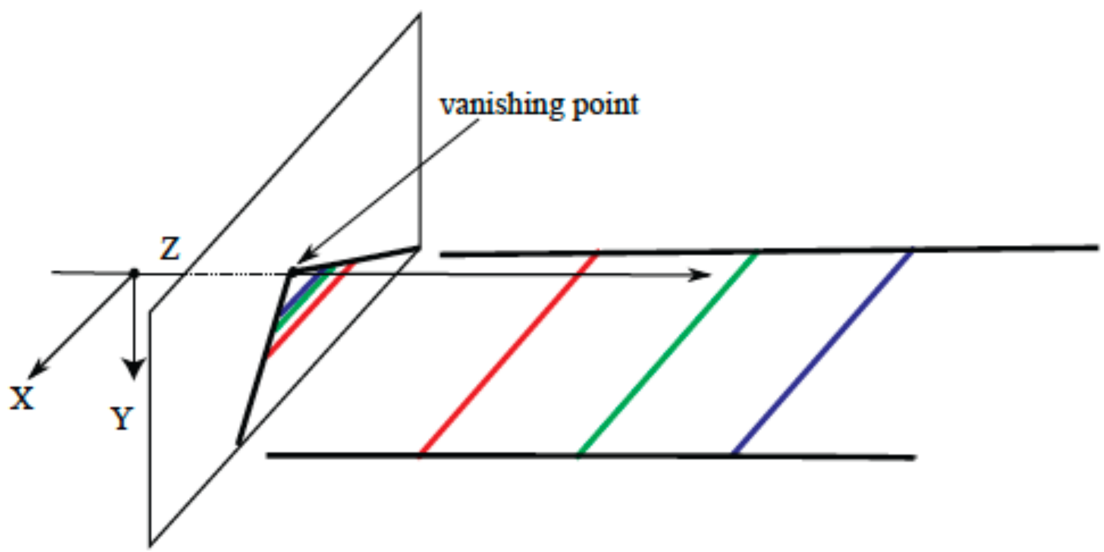
- Notice (0, 0, 0, 0) is meaningless (HC's for 3D)
  - also (0, 0, 0) in 2D
- Basic notion
  - Possible to represent points "at infinity" by careful use of zero
  - Where parallel lines intersect
    - eg

$$(tX, tY, tZ, 1) \quad \text{and} \quad (tX + a, tY + b, tZ + c, 1)$$

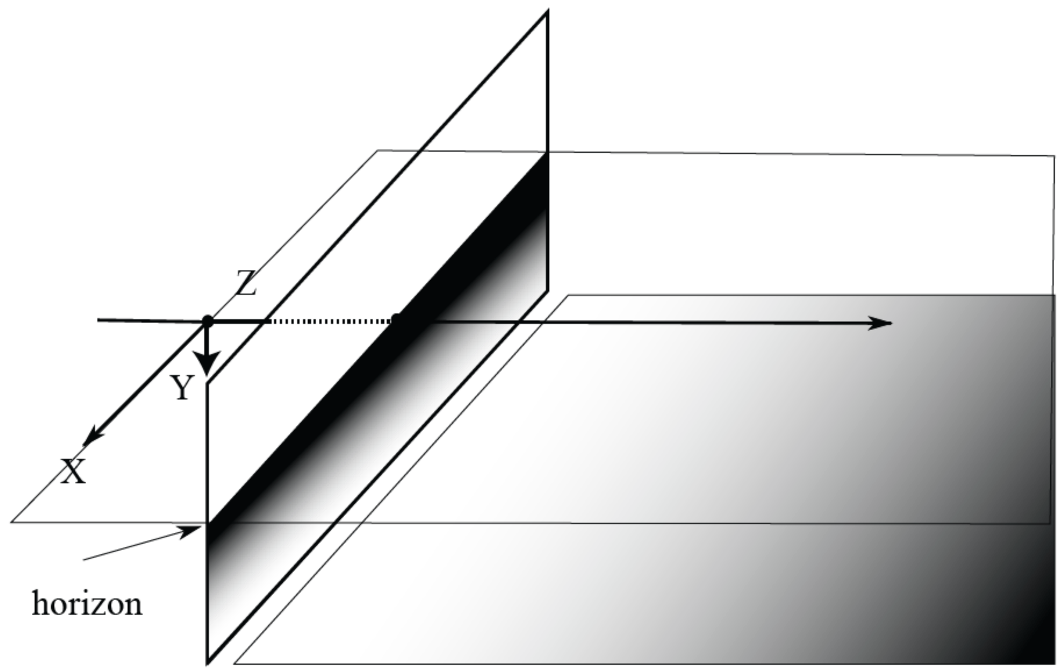
$$\text{intersect at} \quad (X, Y, Z, 0)$$

  - Where parallel planes intersect (etc)
- Can write the action of a perspective camera as a matrix

You can see the point at infinity



You can see the line at infinity, too!



Recall

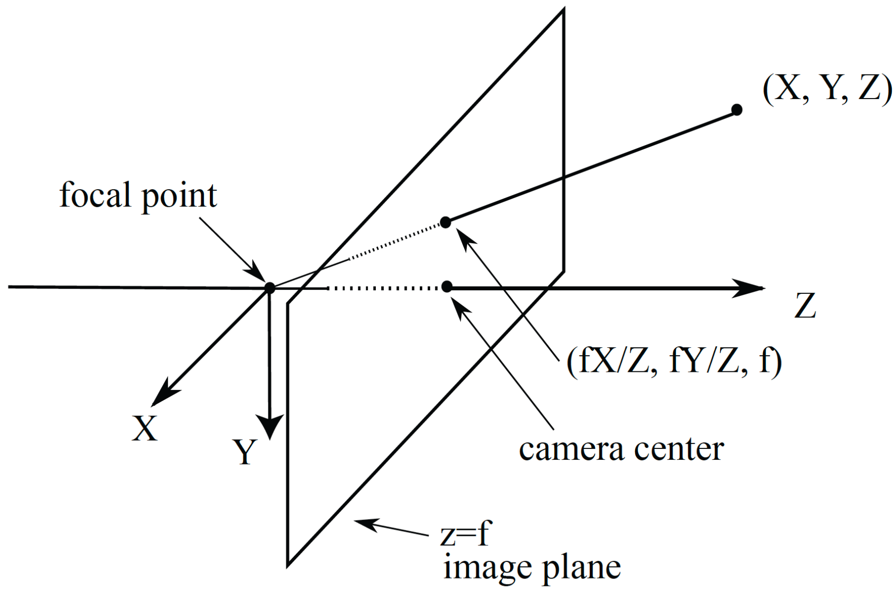


FIGURE 2.4: *The usual geometric abstraction of the pinhole model. The box doesn't affect the geometry, and is omitted. The pinhole has been moved to the back of the box, so that the image is no longer inverted. The image is formed on the plane* $z = f$*, by convention. Notice the y-axis goes down in the image. This allows me to use a* right handed *coordinate system and also have* z increase *as one moves into the image.*

Notice:

- Change in focal length is just a scaling
  - deal with this later
  - currently, f=1
- Now turn into homogenous coordinates

In 3D

Affine | Homogenous

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cong \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

and

Camera

Affine | Homogenous

$$\begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \cong \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

The perspective camera matrix

In affine coordinates, the camera mapping is $(X, Y, Z) \to (X/Z, Y/Z)$, and account for f later. Now write the 3D point in homogeneous coordinates as

$$\mathbf{X} = (X_1, X_2, X_3, X_4)$$

and the point in the image plane in homogeneous coordinates as

$$\mathbf{I} = (I_1, I_2, I_3).$$

Now we have

$$\mathbf{I} = (I_1, I_2, I_3) \equiv (X/Z, Y/Z, 1) \equiv (X, Y, Z) \equiv (X_1/X_4, X_2/X_4, X_3/X_4) \equiv (X_1, X_2, X_3).$$

This means that, in homogeneous coordinates, we can represent perspective projection as

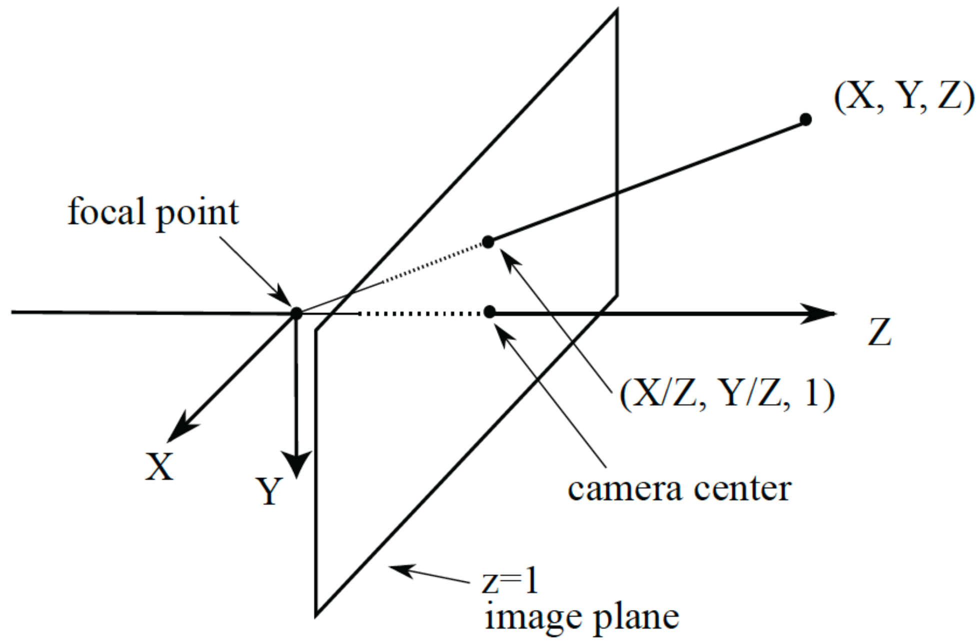
$$(X_1, X_2, X_3, X_4) \to (X_1, X_2, X_3).$$

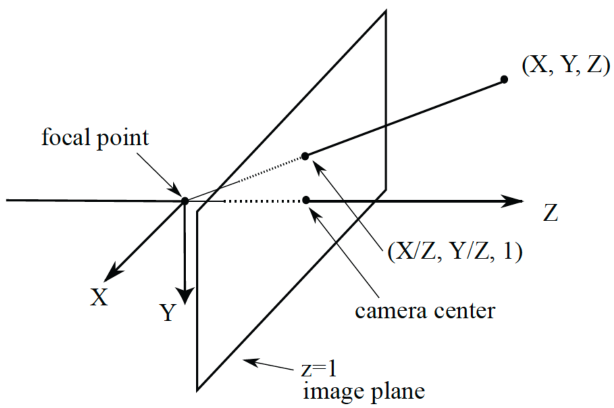
or

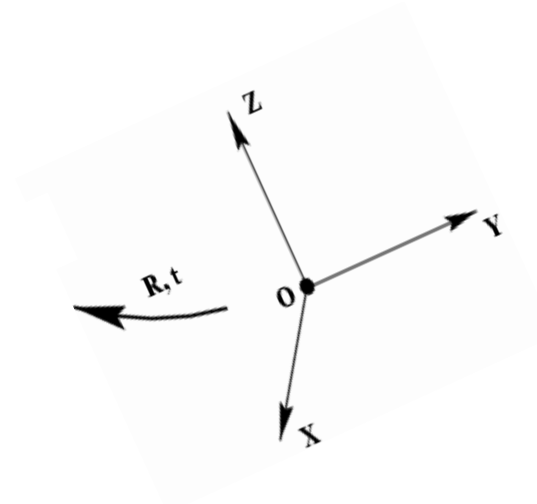
$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

where the matrix is known as the *perspective camera matrix* (write $\mathcal{C}_p$).

Canonical camera

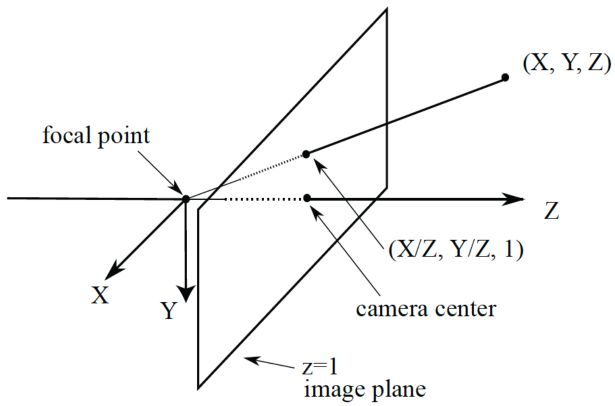


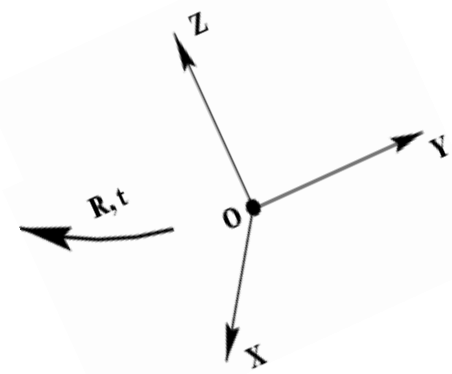




Actual camera

$$\begin{pmatrix} c_{11} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & c_{34} \end{pmatrix}$$





A general perspective camera transformation can be written as:

$$
\begin{aligned}
\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} &= \begin{bmatrix} \text{Transformation} \\ \text{mapping image} \\ \text{plane coords to} \\ \text{pixel coords} \end{bmatrix} \mathcal{C}_p \begin{bmatrix} \text{Transformation} \\ \text{mapping world} \\ \text{coords to camera} \\ \text{coords} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \\
&= \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}
\end{aligned}
$$

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
$$

The parameters of $\mathcal{T}_i$ are known as *camera intrinsic parameters* or *camera intrinsics*, because they are part of the camera, and typically cannot be changed. The parameters of $\mathcal{T}_e$ are known as *camera extrinsic parameters* or *camera extrinsics*, because they can be changed.

Extrinsics

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \text{Transformation} \\ \text{mapping image} \\ \text{plane coords to} \\ \text{pixel coords} \end{bmatrix} \mathcal{C}_p \underbrace{\begin{bmatrix} \text{Transformation} \\ \text{mapping world} \\ \text{coords to camera} \\ \text{coords} \end{bmatrix}}_{\text{Extrinsics}} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

$$= \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

any Euclidean transformation maps the vector $\mathbf{x}$ to

$$\mathcal{R}\mathbf{x} + \mathbf{t}$$

where $\mathcal{R}$ is an appropriately chosen 3D rotation matrix (check the endnotes if you can't recall) and $\mathbf{t}$ is the translation. Any map of this form is a Euclidean transformation. You should confirm the transformation that maps the vector $\mathbf{X}$ representing a point in 3D in homogeneous coordinates to

$$\lambda \begin{bmatrix} \mathcal{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{X}$$

represents a Euclidean transformation, but in homogeneous coordinates. It follows that any map of this form is a Euclidean transformation. Because $\mathcal{T}_e$ represents a Euclidean transformation, it must have this form **exercises** .

Intrinsics

Intrinsics

$$\begin{aligned}\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} &= \begin{bmatrix} \text{Transformation} \\ \text{mapping image} \\ \text{plane coords to} \\ \text{pixel coords} \end{bmatrix} \mathcal{C}_p \begin{bmatrix} \text{Transformation} \\ \text{mapping world} \\ \text{coords to camera} \\ \text{coords} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \\ &= \mathcal{T}_i \mathcal{C}_p \mathcal{T}_e \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}\end{aligned}$$

Notice that

$$\mathcal{T}_i \mathcal{C}_p \mathcal{T}_e = \mathcal{T}_i \left[\mathcal{R} \mid \mathbf{t} \right]$$

which has a significant effect on the form of the intrinsic parameters. Any square matrix of full rank can be factored into a product of an upper triangular term and a rotation (**exercises**). Now assume you are *given* a general camera matrix

$$\mathcal{C} = \left[\mathcal{M} \mid \mathbf{v} \right].$$

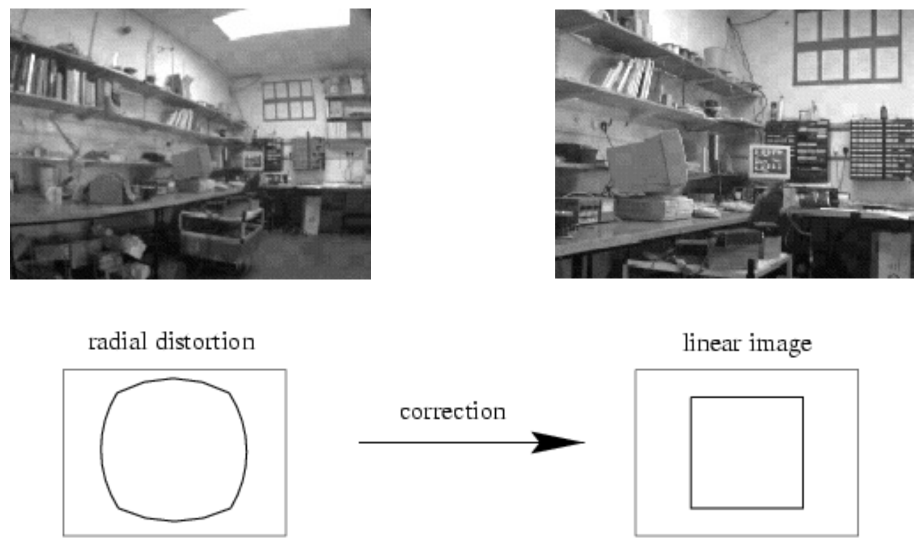
Intrinsics

This could only be a camera matrix if $\mathcal{M}$ had full rank **exercises** . You can factor $\mathcal{M}$ into $\mathcal{U}\mathcal{Q}$, where $\mathcal{U}$ is upper triangular and $\mathcal{Q}$ is a rotation. The only ambiguities have to do with signs **exercises** . This means that, if $\mathcal{T}_i$ is *not* upper triangular, an appropriate choice of rotation would make it upper triangular **exercises** .

It is usual to work with an upper triangular $\mathcal{T}_i$. There are easy physical interpretations for the elements of $\mathcal{T}_i$. Write

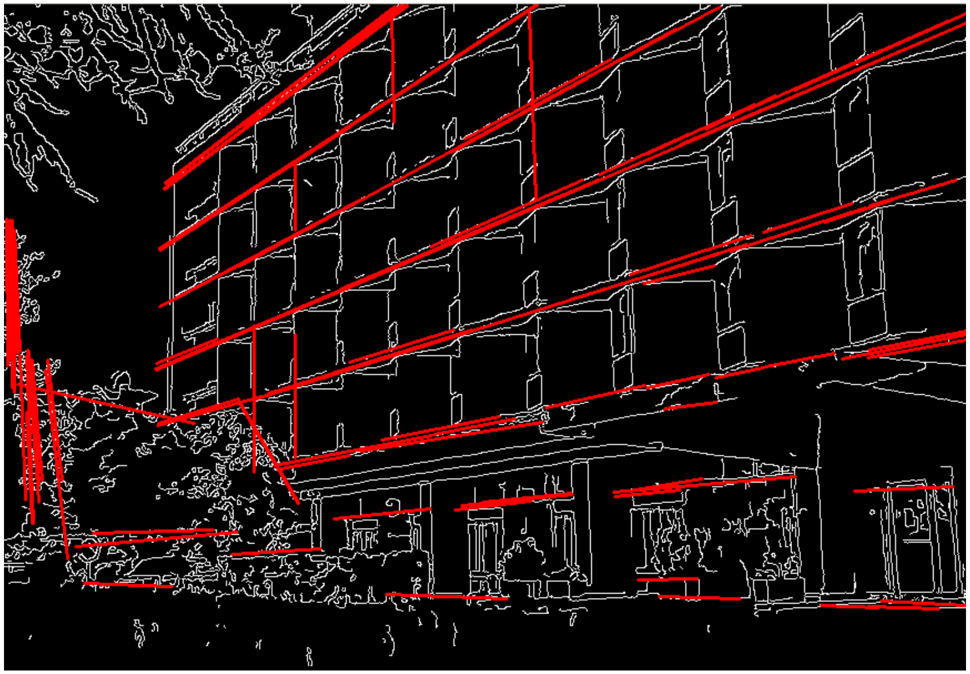
$$\mathcal{T}_i = \begin{bmatrix} s_x & k & c_x \\ 0 & s_y & c_y \\ 0 & 0 & 1 \end{bmatrix}.$$

I did not discuss...



Fitting

- Form a higher level, more compact representation of the image features
- Model case:
    - fit a line



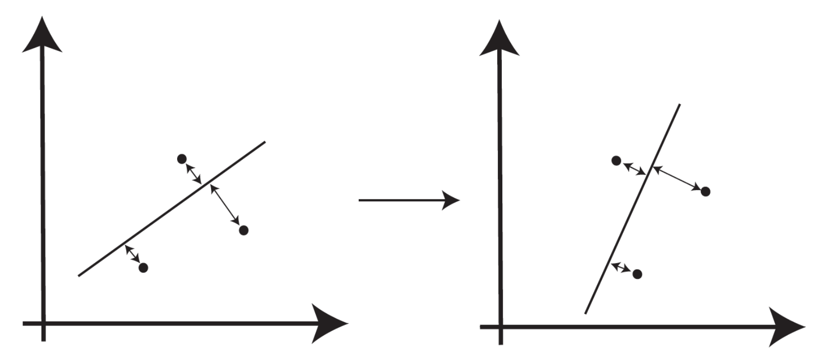
Fitting: Challenges

Case study: Line detection

- **Noise** in the measured feature locations
- **Extraneous data:** clutter (outliers), multiple lines
- **Missing data:** occlusions

Fitting – some crucial facts of life

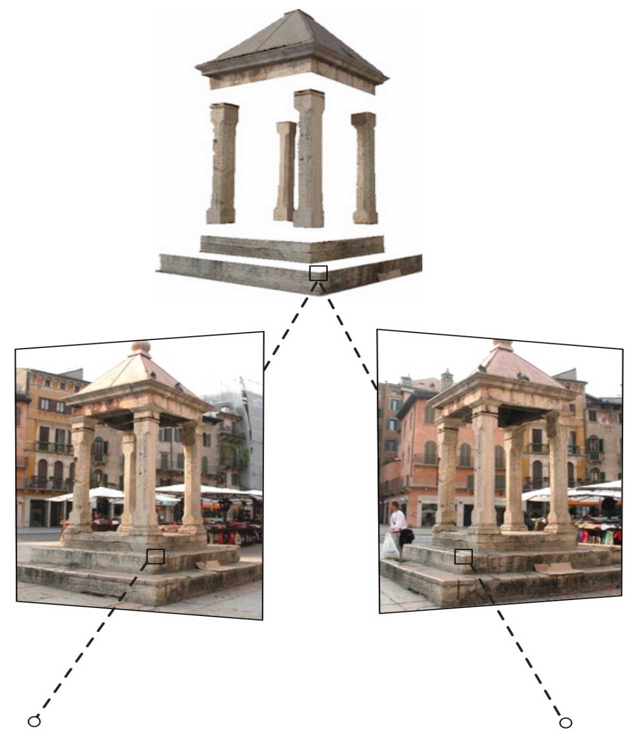
- Some parametric model explains data
  - example: line fitting, explain points by line
  - an error/loss measures how well
- Fitting: find the parameters
  - noise – total least squares
  - outliers – RANSAC
- These methods are highly effective
  - but there isn't time for detail

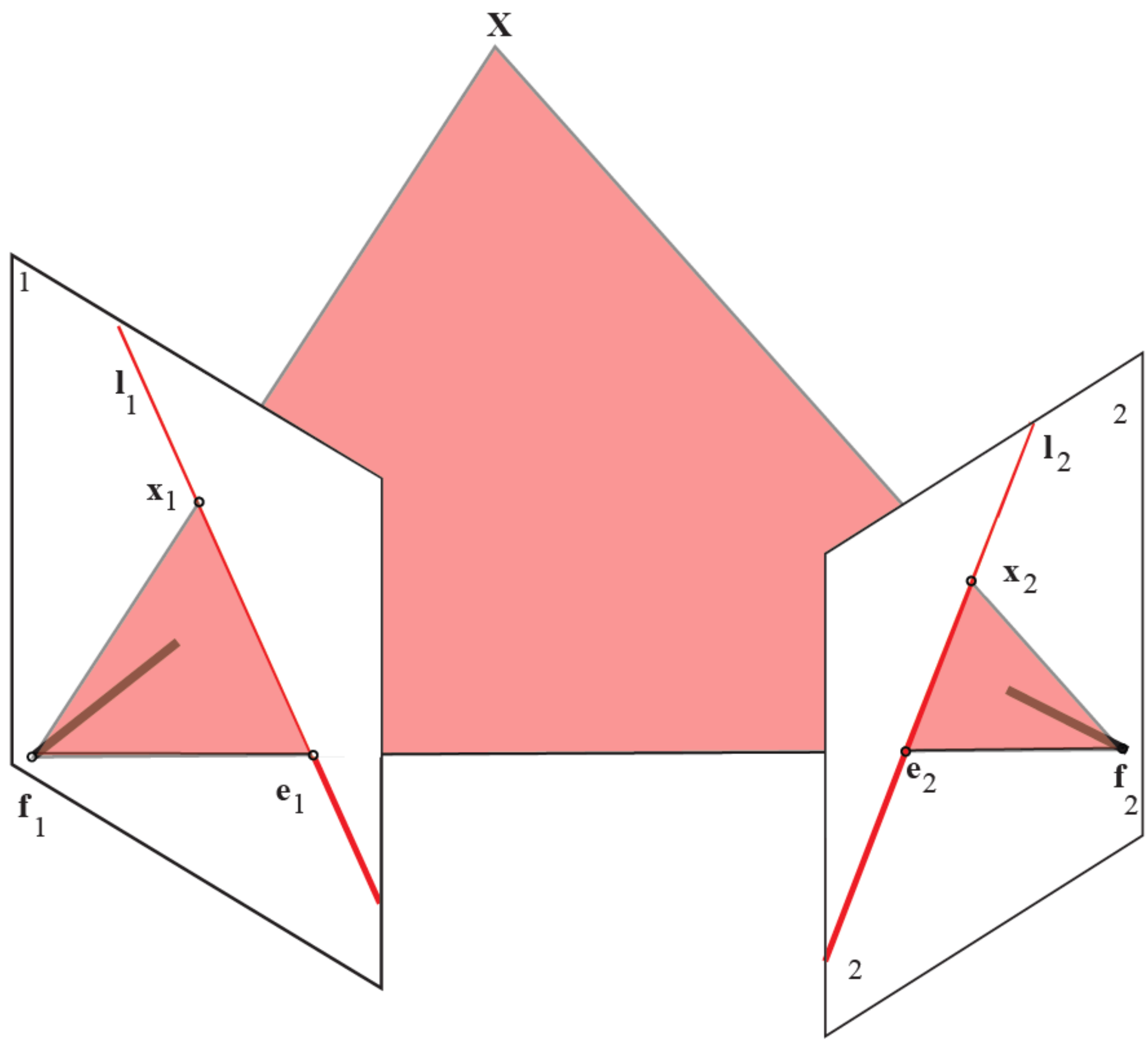


Total least squares

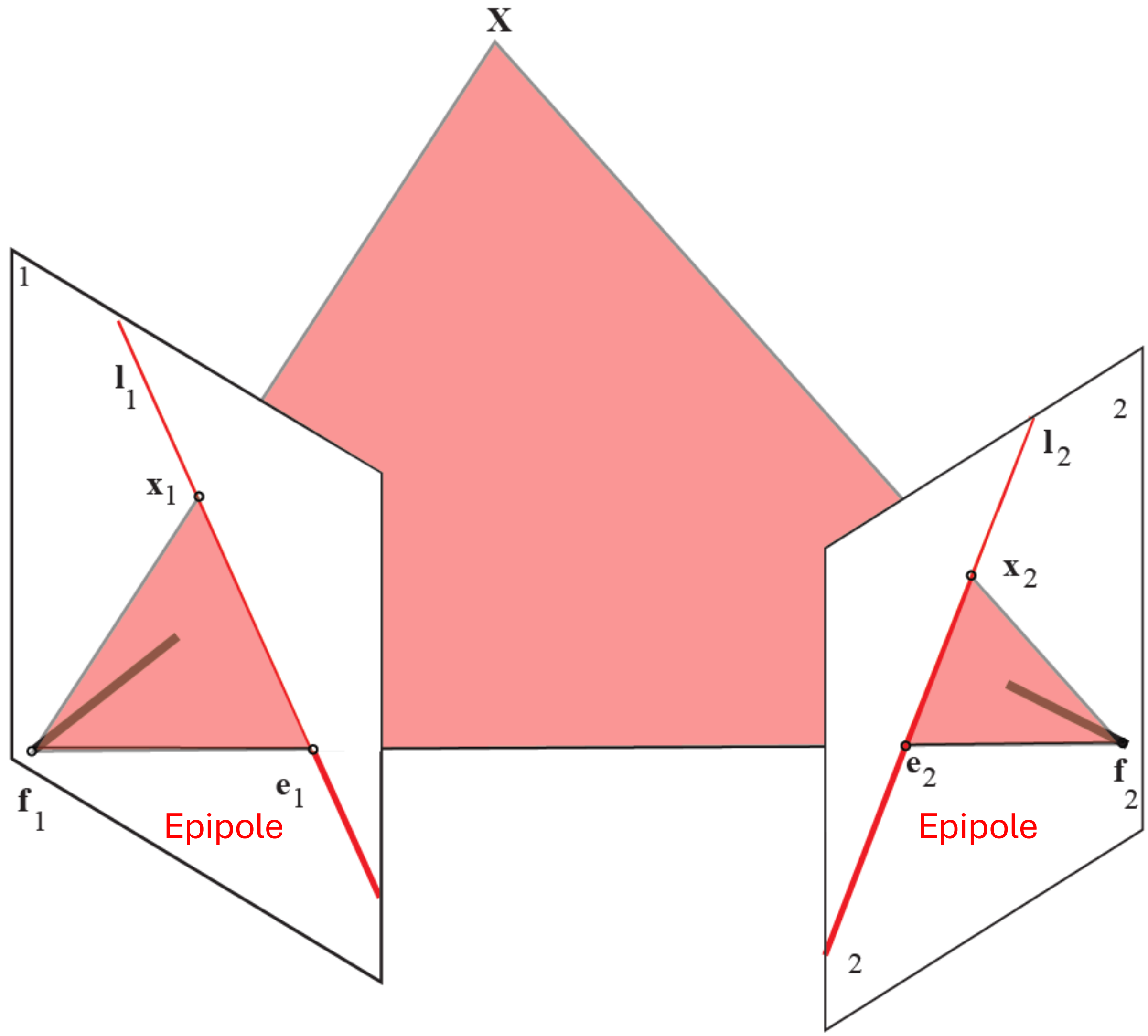
Two views of the same 3D scene

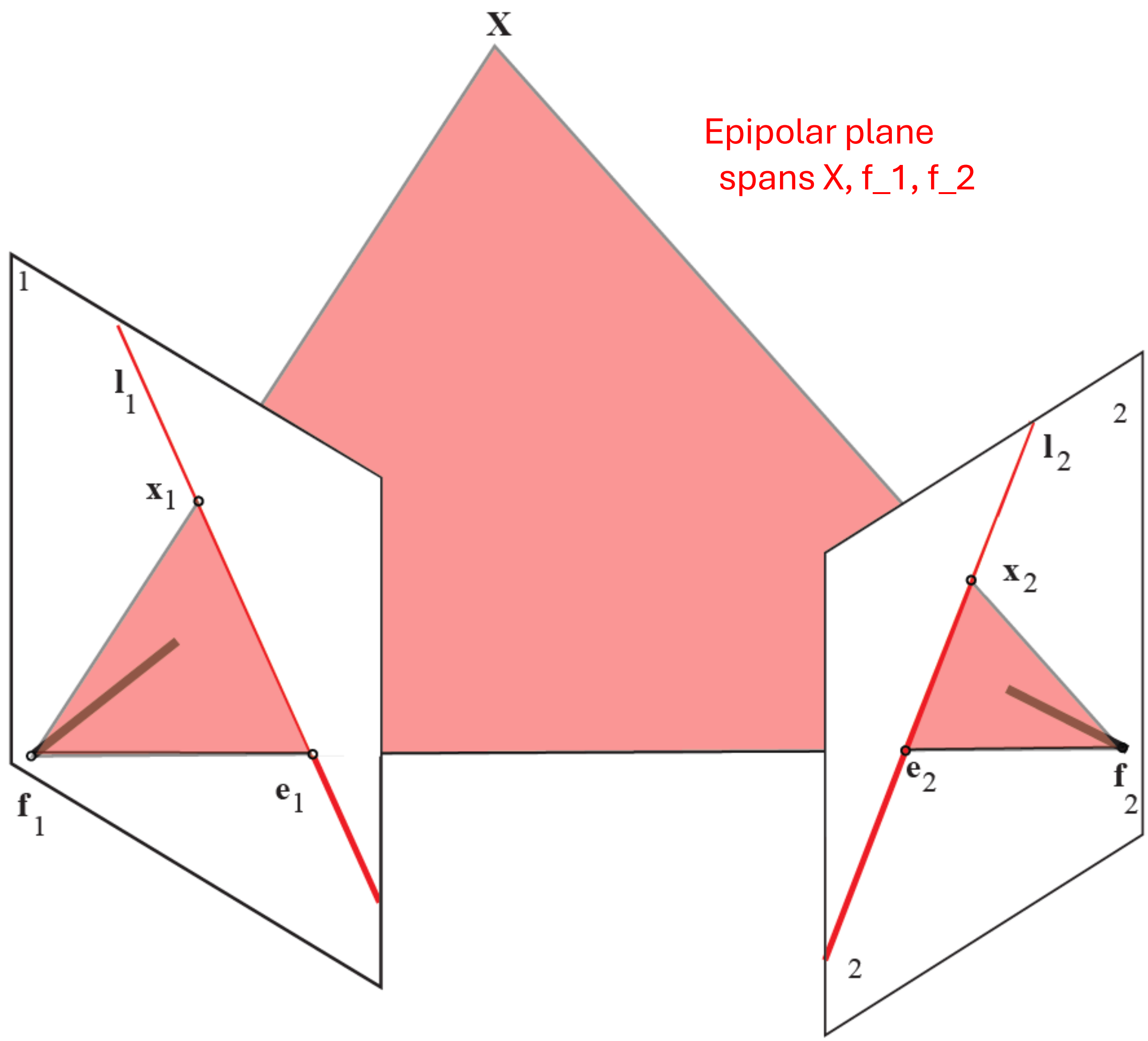
- Because:
  - You have two
    - Eyes
    - Cameras
  - OR the camera moves





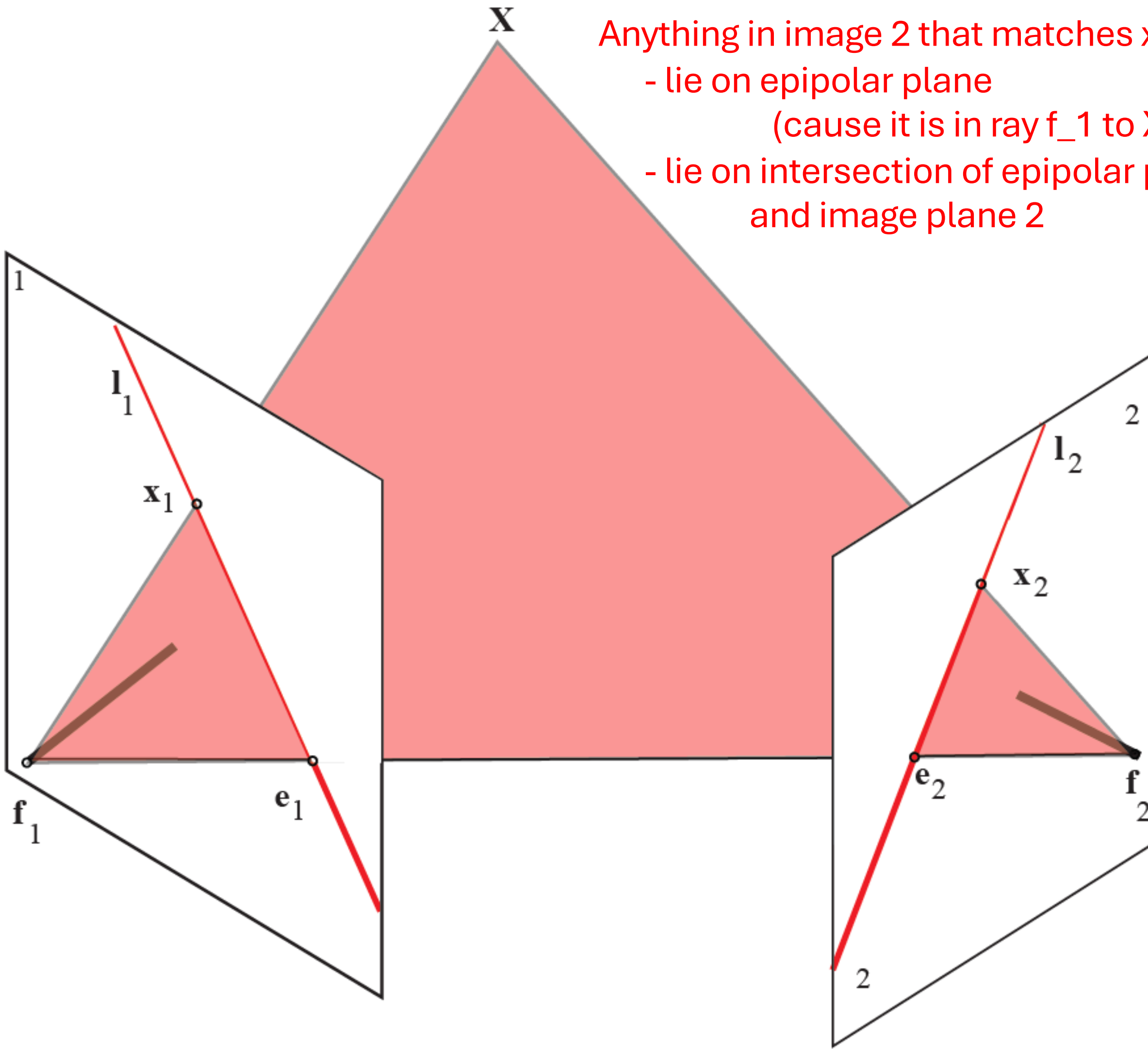
X
1
l1
x1
f1
e1
2
l2
x2
e2
f2
2
2

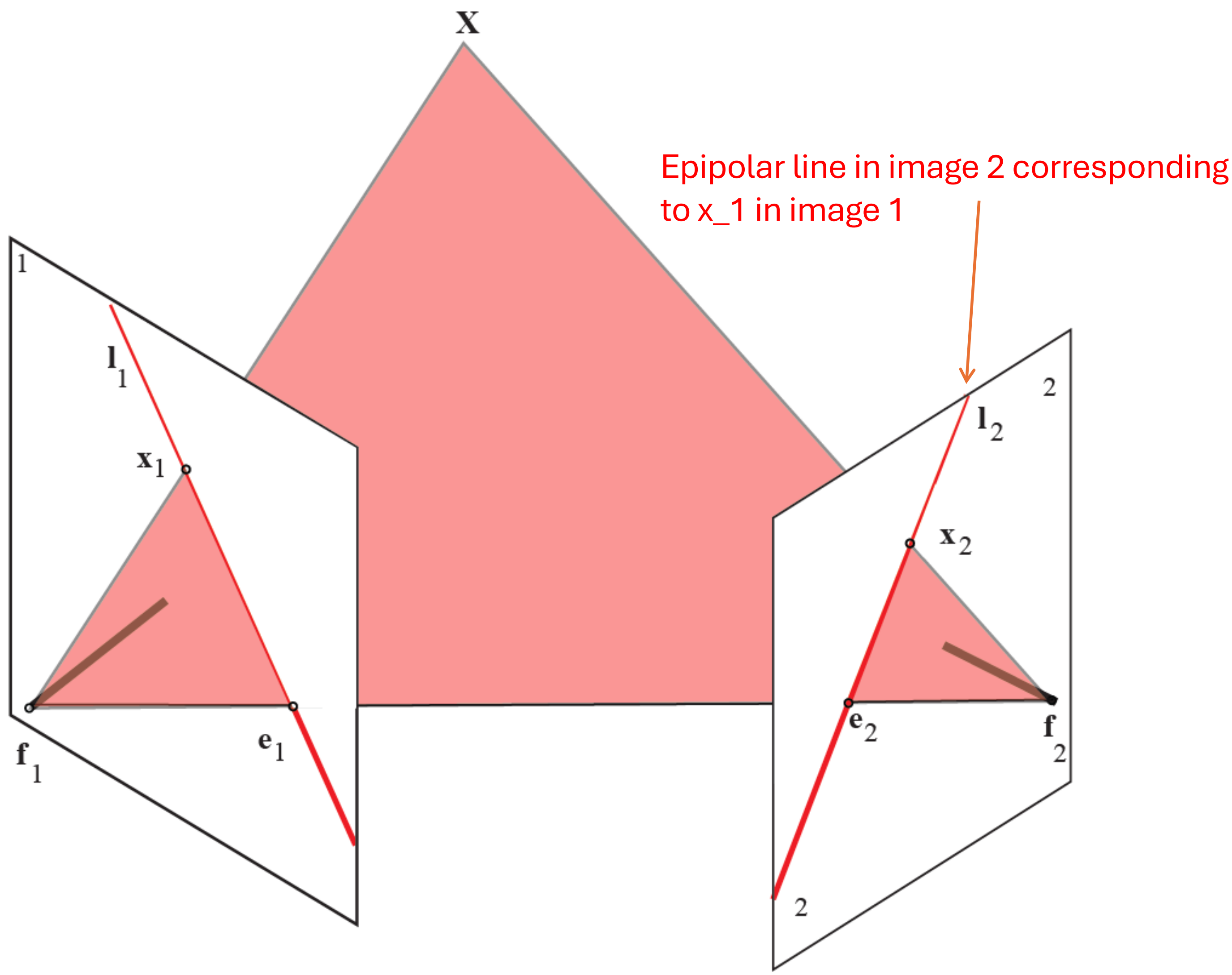

X
1
$\mathbf{l}_1$
$\mathbf{x}_1$
$\mathbf{f}_1$
$\mathbf{e}_1$
Epipole
2
$\mathbf{l}_2$
$\mathbf{x}_2$
$\mathbf{e}_2$
$\mathbf{f}_2$
Epipole
2

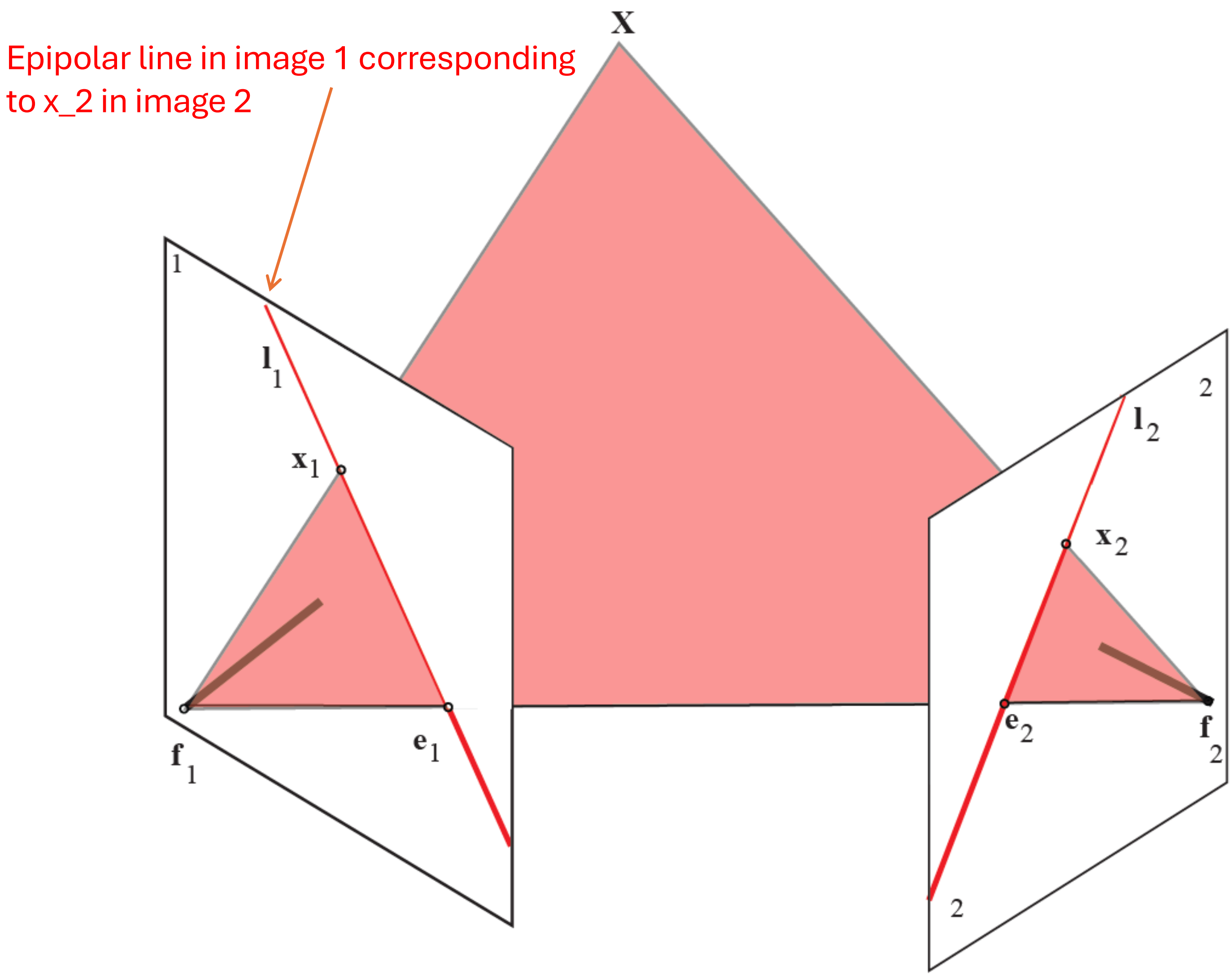


X
Epipolar plane
spans X, f_1, f_2
1
l 1
x 1
e 1
f 1
2
l 2
x 2
e 2
f 2
2

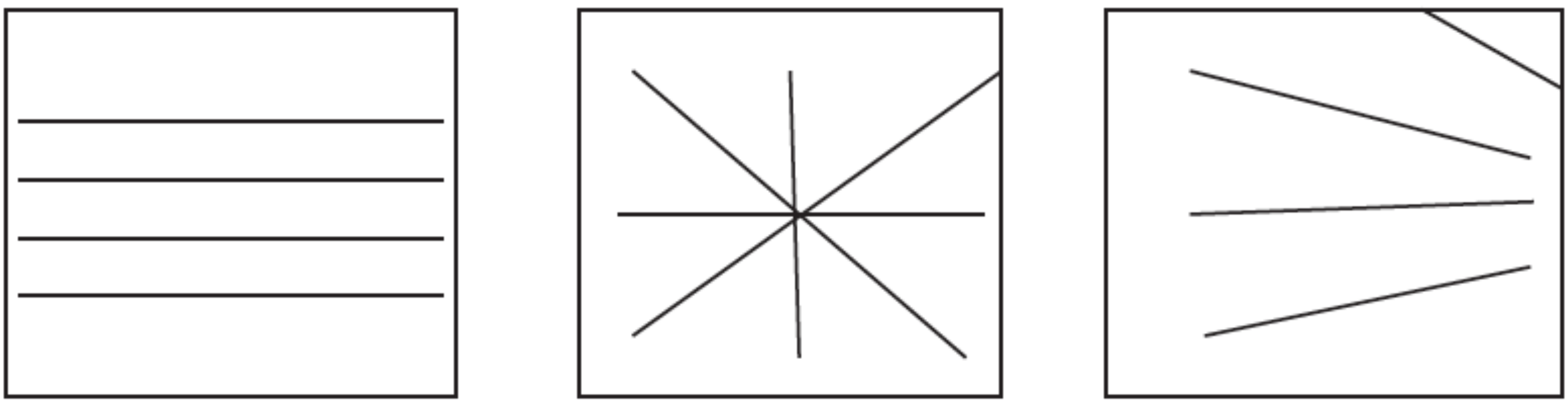
Anything in image 2 that matches x_1 MUST
- lie on epipolar plane
  (cause it is in ray f_1 to X)
- lie on intersection of epipolar plane
  and image plane 2

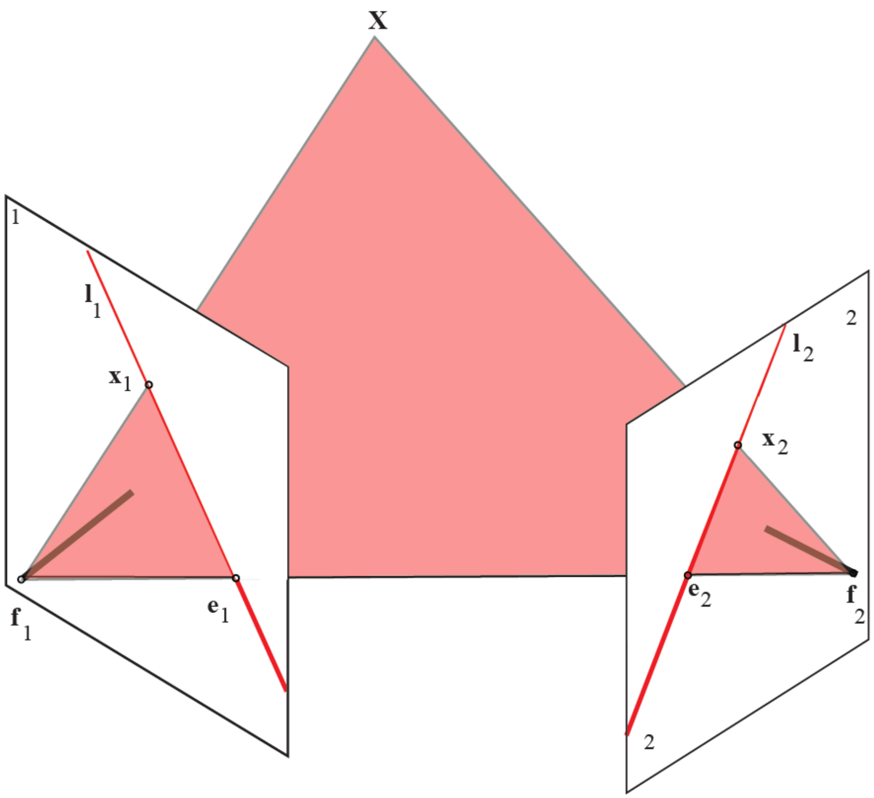



X
Epipolar line in image 2 corresponding
to x_1 in image 1
1
l 1
x 1
e 1
f 1
2
l 2
x 2
e 2
f 2
2
2


Epipolar line in image 1 corresponding to x_2 in image 2
X
1
l1
x1
e1
f1
2
l2
x2
e2
f2
2
2

Interpreting epipoles and epipolar lines





Conclude:

x_1, x_2 and focal points are coplanar

It follows there is a matrix F such that

x_1^T F x_2 =0

The fundamental matrix

Remember this: *For any pair of cameras which do not share a focal point, there is a fundamental matrix* $\mathcal{F}$ *with the property that for any pair* $\mathbf{x}_1$, $\mathbf{x}_2'$, *where* $\mathbf{x}_1$ *is the image of a 3D point in the first camera and* $\mathbf{x}_2'$ *is the image of that point in the second camera,*

$$\mathbf{x}_1^T \mathcal{F} \mathbf{x}_2' = 0$$

Procedure: 32.1 *Obtaining an epipolar line from a fundamental matrix*

The epipolar line in camera 2 corresponding to $\mathbf{x}_1$ in camera 1 consists of the set of points $\mathbf{x}'$ in camera 2 which satisfy the equation

$$\left(\mathbf{x}_1^T \mathcal{F}\right) \mathbf{x}' = 0$$

and the coefficients of the line are

$$\left(\mathcal{F}^T \mathbf{x}_1\right).$$

It follows that the coefficients of the epipolar line in camera 1 corresponding to $\mathbf{x}_2'$ in camera 2 are

$$(\mathcal{F}\mathbf{x}_2').$$

The fundamental matrix

Procedure: 32.2 *Obtaining epipoles from a fundamental matrix*

The epipole in camera 2 is the point $\mathbf{e}_2'$ such that

$$\mathcal{F}\mathbf{e}_2' = \mathbf{0}.$$

It follows that the epipole in camera 1 is the point $\mathbf{e}_1$ such that

$$\mathcal{F}^T\mathbf{e}_1 = \mathbf{0}.$$