Università Ca' Foscari Venezia

Computer Vision

Epipolar geometry

Filippo Bergamasco (filippo.bergamasco@unive.it)
http://www.dais.unive.it/~bergamasco
DAIS, Ca' Foscari University of Venice
Academic year 2018/2019

The imaging process

Supposing to know:

- The coordinates of a set of 3D points in our scene, according to some world reference system
- The camera extrinsic parameters **R**,**T** (ie. the pose of the camera in the scene)
- The camera intrinsic parameters **K**

We can compute the position of those point on the image plane after projection

Applications? We can blend **virtual** objects on the image plane (Augmented Reality)

3D reconstruction

But we are interested in the inverse process!

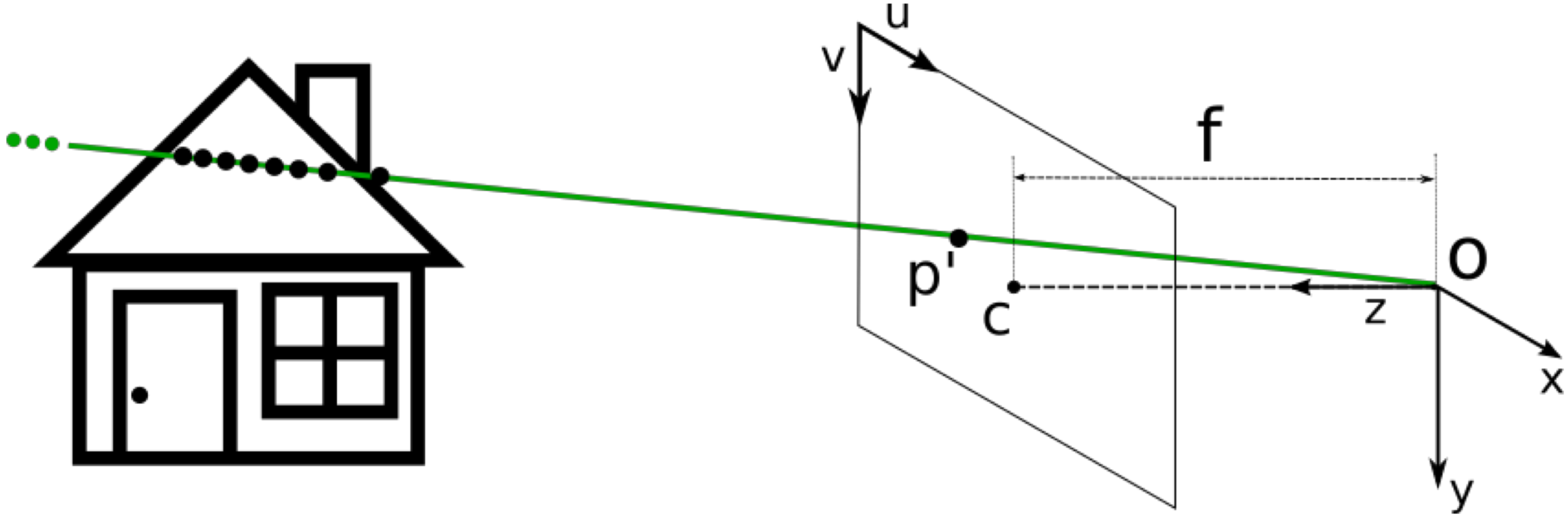
If we know:

- The intrinsic and extrinsic parameters of our imaging system
- The location of the object points in our image plane

Can we reconstruct their 3D structure?

3D reconstruction

Unfortunately, for each point on the image plane there is an infinite set of 3D points that would solve the inverse imaging problem

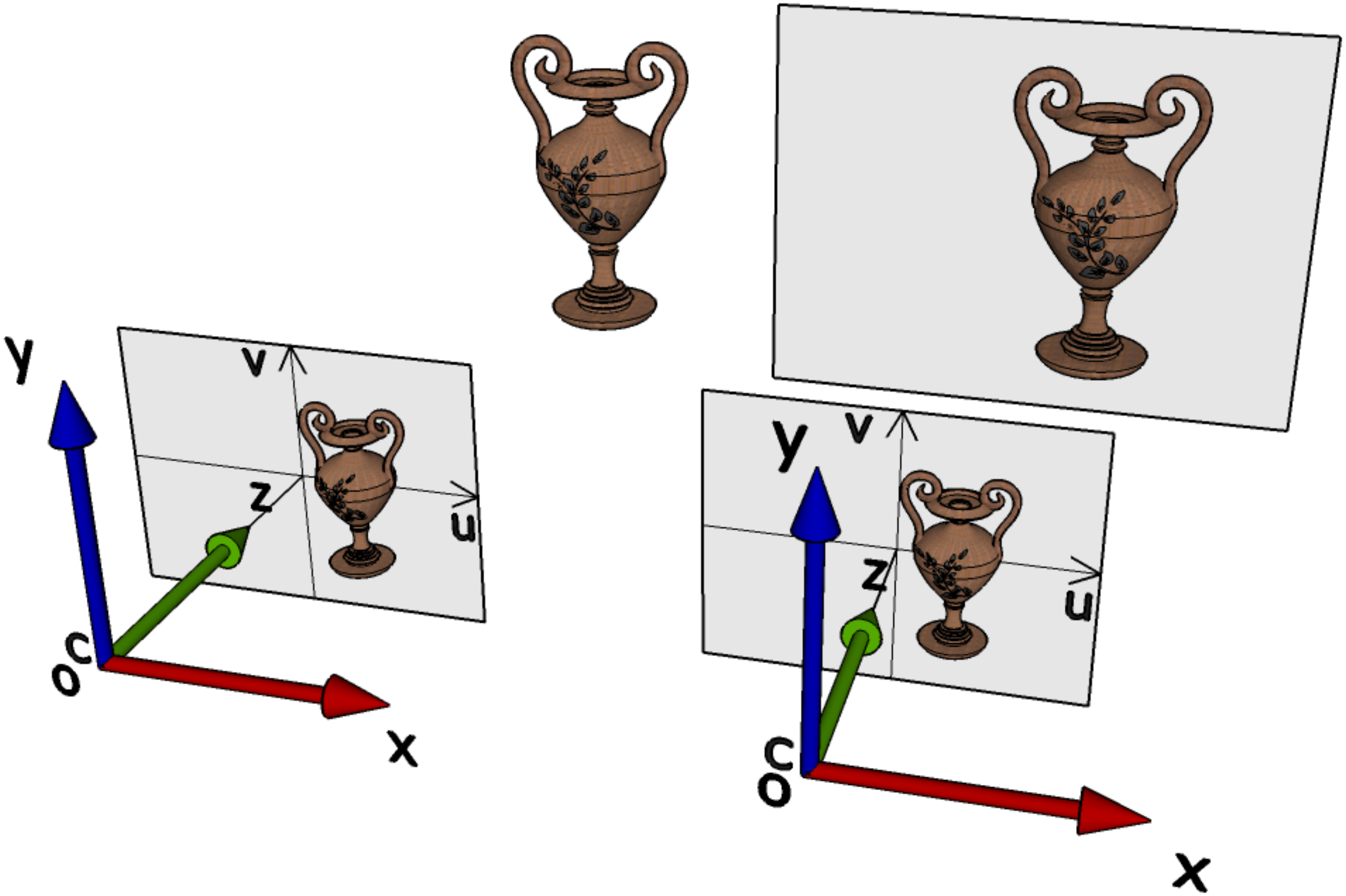


In other words, we do not have enough information from the 2D points alone

The reconstruction paradox

Università Ca' Foscari Venezia

The reconstruction paradox



Are we observing a vase or a picture of a vase?

… no way to know!

Università
Ca' Foscari
Venezia

Ambiguities…

Università Ca' Foscari Venezia

Ambiguities…

Ambiguities…

How to solve the problem?

- **Single image** alternatives:
  - Consider additional information, like the time it takes from each light beam to travel from the scene to the camera
  - Project controlled light onto the scene (ex. laser scanner)
  - Use properties of the optics (ex. depth from focus)
  - Assume a shading model (ex. shape from shading)

- **Multiple images** of the same scene

  By comparing images of the same scene from different points of view we can solve the ambiguity

  - Stereo Vision
  - Structure from motion
  - Etc.

Università Ca' Foscari Venezia

Epipolar geometry

The epipolar geometry is the intrinsic projective geometry between **two views** of the same scene

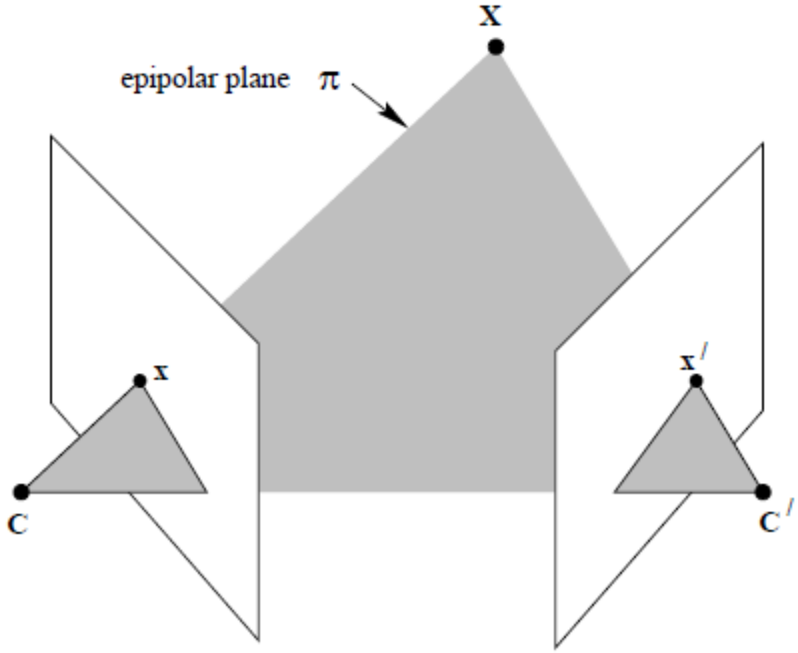
- It is independent of the scene structure
- It depends on the intrinsic parameters of the two cameras and their relative pose

Suppose that a 3D point $X \in \mathbf{P}^3$ is imaged in two views, $x \in \mathbf{P}^2$ in the first and $x' \in \mathbf{P}^2$ in the second.

What is the relation between the **corresponding points** x and x' ?

Università
Ca' Foscari
Venezia

Epipolar geometry

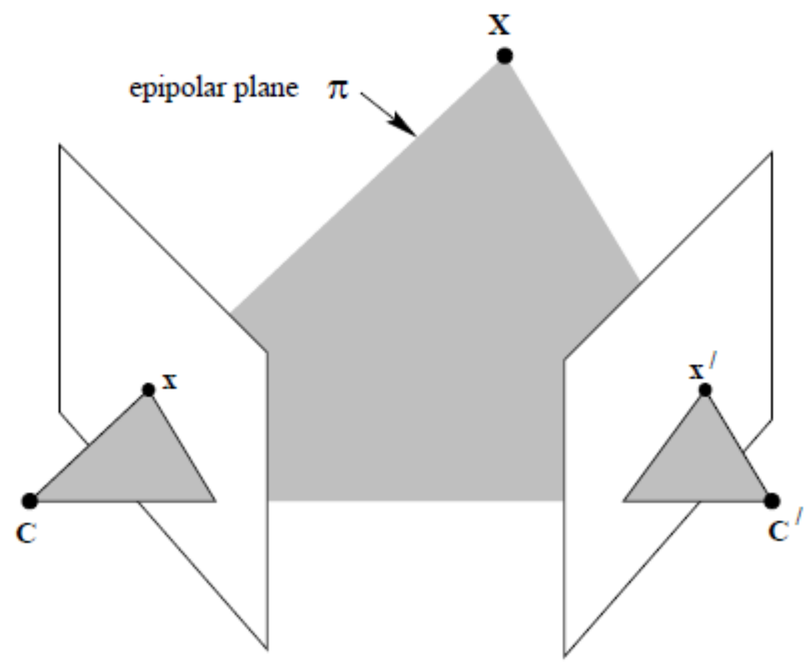


The point in space **X** and the two camera centres **C** and **C'** are coplanar.

We call this plane the **epipolar plane** π .

Università Ca' Foscari Venezia

Epipolar geometry

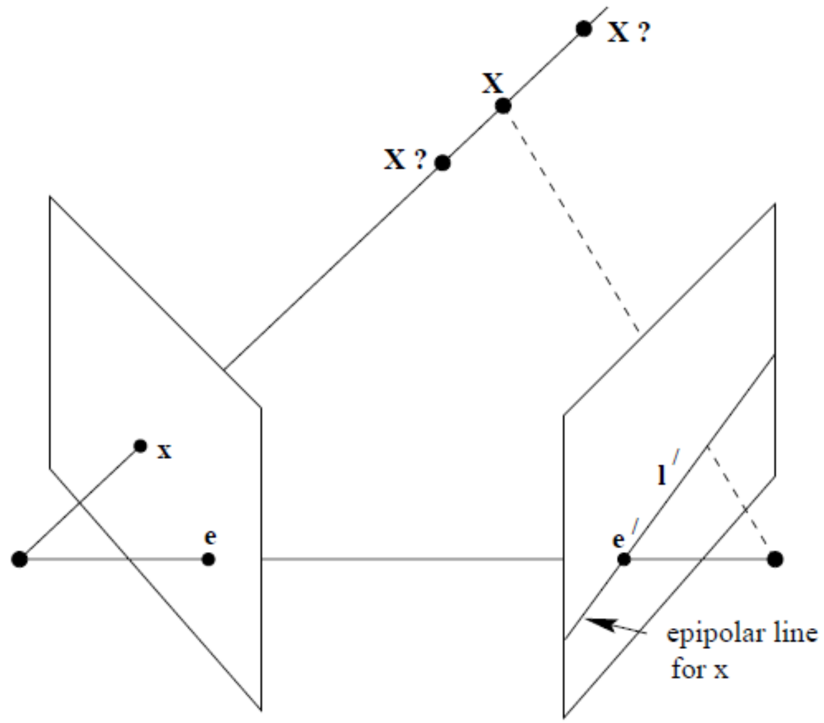


The rays back-projected from x and x' intersect at **X**. The rays are coplanar, lying in π.

Suppose that we only know x. Where do we expect to find the corresponding point x' in the second view?

Università Ca' Foscari Venezia

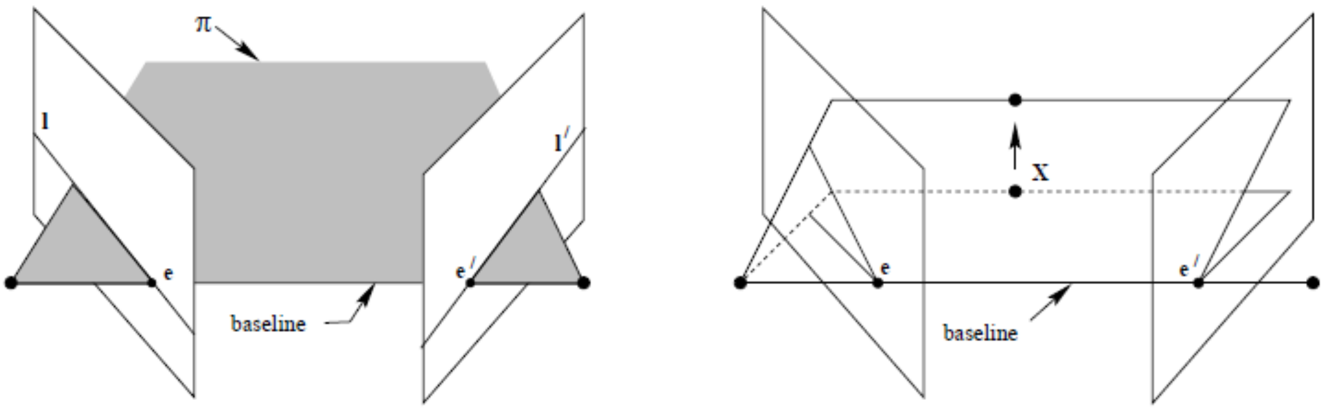
Epipolar geometry



The ray corresponding to the (unknown) point x' lies in π . Therefore, the point x' lies on the intersection between the plane π and the second image plane. Such intersection is a line l' called **the epipolar line corresponding to** x

Università Ca' Foscari Venezia

Epipolar geometry



<u>We define the following concepts:</u>

- The **baseline** is the line segment joining the two camera centers
- The **epipole** is the point of intersection of the baseline with the image plane (note: it can be at infinite).
- An **epipolar plane** is any plane containing the baseline. There is a one-parameter family of epipolar planes
- An **epipolar line** is the intersection of an epipolar plane with an image plane. All the epipolar lines intersect at the epipole.

The Fundamental matrix

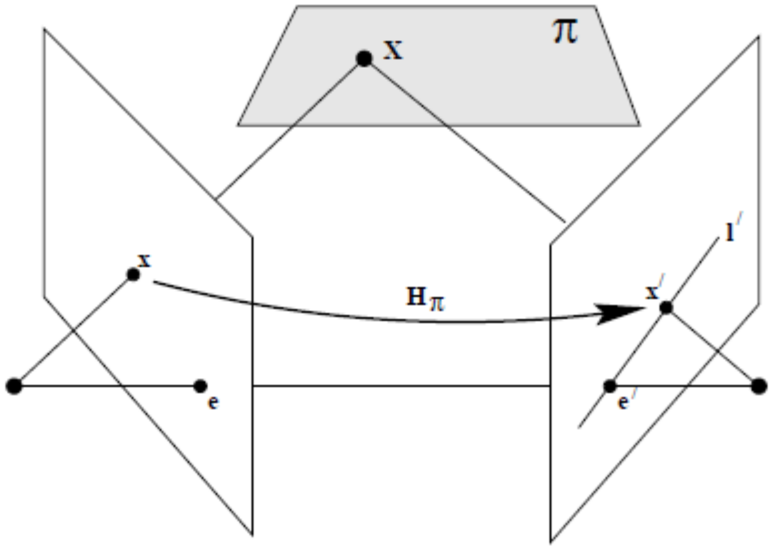
The fundamental matrix is the algebraic representation of epipolar geometry. It describes the following two facts:

1. For each point x in one image there exists a corresponding epipolar line l' in the other image.
2. Any point x' in the second image matching the point x must lie on the epipolar line l'

We will derive the fundamental matrix both geometrically and algebraically

Università Ca' Foscari Venezia

Deriving F geometrically

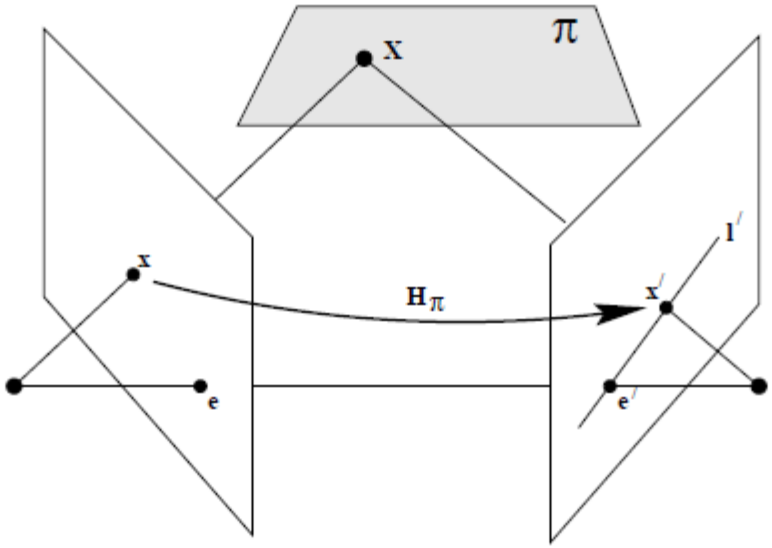


Let π be **any** plane in space and X a 3D point on that plane. The point x and x' are the image of X in the two images respectively. We got:

- x' must line on the epipolar line l' corresponding to the image of the ray passing through x and X.
- An homograpy H_π exists mapping each x_i to its corresponding point x'_i through π

Università
Ca' Foscari
Venezia

Deriving F geometrically



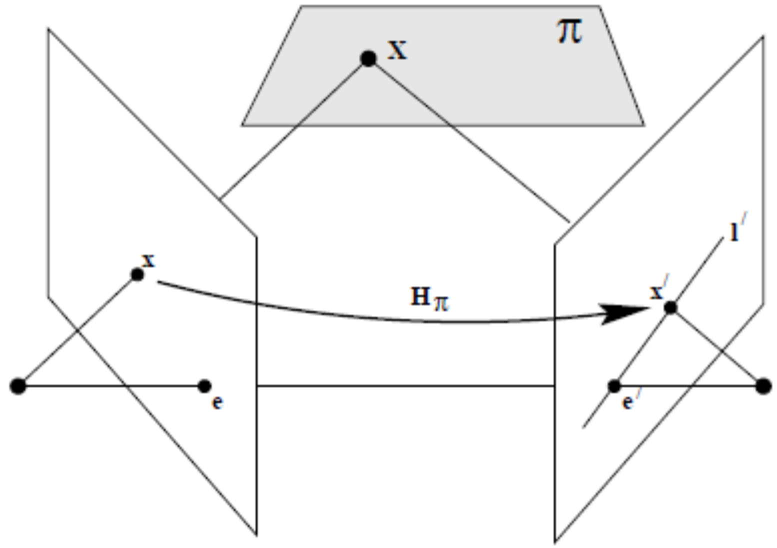
Given x', the epipolar line l' is the line passing through x' and the epipole e':

$$l' = e' \times x'$$

$$l' = [e']_{\times} x'$$

where $$[e']_{\times} = \begin{pmatrix} 0 & -e'_z & e'_y \\ e'_z & 0 & -e'_x \\ -e'_y & e'_x & 0 \end{pmatrix}$$

Deriving F geometrically



Since $l' = [e']_\times x'$ and $x' = H_\pi x$ we have:

$$l' = [e']_\times H_\pi x = Fx$$

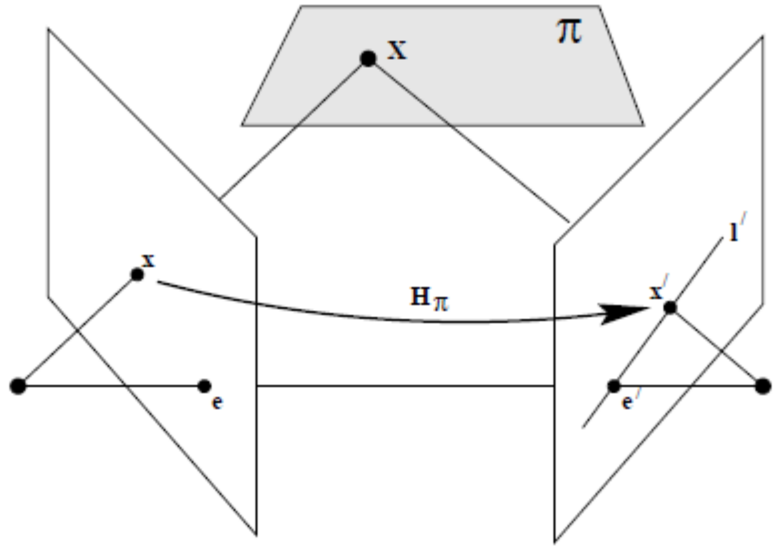
The fundamental matrix F may be defined as:

$$F = [e']_\times H_\pi$$

where H_π is the mapping from one image to the other via any plane π. Since $[e']_\times$ has rank 2, F is a matrix of rank 2.

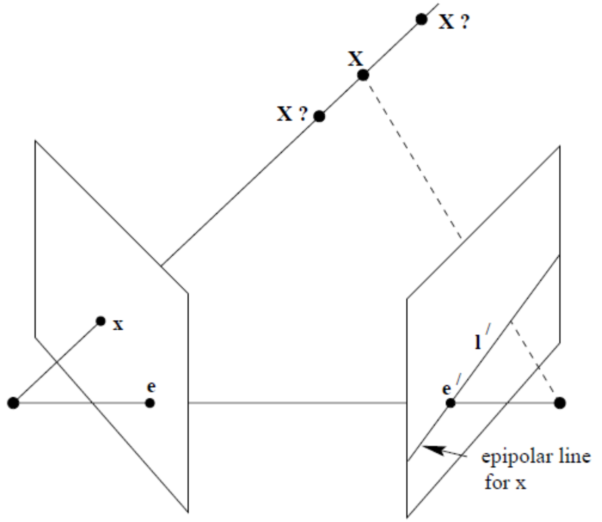
Università Ca' Foscari Venezia

Deriving F geometrically



Note that the geometric derivation involves a scene plane but **such plane is not required in order for F to exist**. The plane is used to define any point mapping between the first image and the second by means of an homography.

Deriving F algebraically

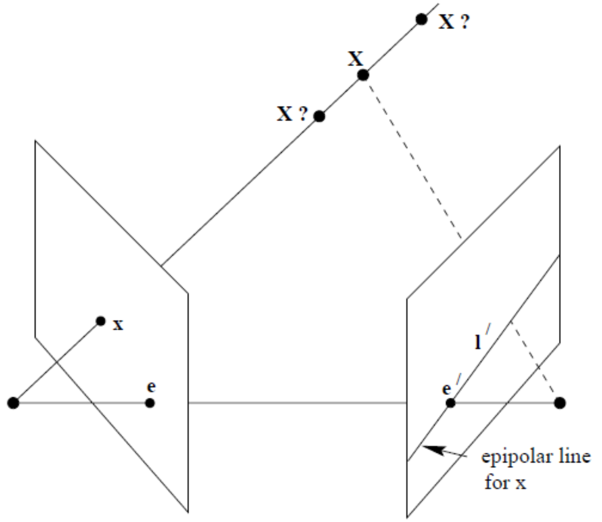


The ray back-projected from x by the first camera is given by:

$$X(\lambda) = P^{+}x + \lambda C$$

Where P^{+} is the pseudo-inverse of the projection matrix of the first camera and C is the null-vector $PC = 0$

Deriving F algebraically



The ray back-projected from x by the first camera is given by:

$$X(\lambda) = P^{+}x + \lambda C$$

In particular, the points $X(0) = P^{+}x$ and $X(\infty) = C$

When projected onto the second image P' they produce the epipolar line

$$l' = (P'C) \times (P'P^{+}x)$$

Università Ca' Foscari Venezia

Deriving F algebraically

$$l' = (P'C) \times (P'P^{+}x)$$

The point $P'C$ is the projection of the first camera center into the second image, ie. the epipole e'.

Therefore, we can write:

$$l' = [e']_{\times}(P'P^{+})x = Fx$$

Where $F = [e']_{\times}(P'P^{+})$ and $(P'P^{+}) = H_{\pi}$ is the explicit form of the point transfer homography between the two image planes.

Università Ca' Foscari Venezia

Relating F to K R T

Suppose that the projection matrices of the two cameras are:

$$P = K\left(I \mid 0\right) \qquad P' = K'\left(R \mid T\right)$$

Where the 3D points are defined in the reference frame of the first camera and **R T** is the rigid motion mapping 3D points from the first to the second camera reference frames.

Università Ca' Foscari Venezia

Relating F to K R T

$$P = K\,(I \mid 0) \qquad P' = K'\,(R \mid T)$$

We got:

$$P^{+} = \begin{pmatrix} K^{-1} \\ 0^{T} \end{pmatrix} \qquad C = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix}$$

Hence:

$$\begin{aligned} F &= [e']_{\times}(P'P^{+}) \\ &= [P'C]_{\times}P'P^{+} \\ &= [K'T]_{\times}K'RK^{-1} \\ &= K'^{-T}\,[T]_{\times}R\,K^{-1} \end{aligned}$$

Since for any vector t and non-singular matrix M one has:

$[\mathrm{t}]_{\times}\mathrm{M} = \mathrm{M}^{*}[\mathrm{M}^{-1}\mathrm{t}]_{\times} = \mathrm{M}^{-\mathrm{T}}[\mathrm{M}^{-1}\mathrm{t}]_{\times}$ *(up to scale)*.

Università Ca' Foscari Venezia

Correspondence condition

The fundamental matrix satisfies the condition that for any pair of corresponding points $x \leftrightarrow x'$ in the two images the following equation holds:

$$x'^T F x = 0$$

If $x \leftrightarrow x'$ correspond, then x' lies on the epipolar line $l' = Fx$ and therefore $0 = x'^T l' = x' F x$.

This is important because **we can obtain the fundamental matrix just by point correspondences** without knowing the projection matrices

Università Ca' Foscari Venezia

Properties

1. If F is the fundamental matrix of the pair of cameras (P,P'), F^T is the fundamental matrix of the same pair in the opposite order (P',P)
2. For any point correspondence $x \leftrightarrow x'$
   1. $l' = Fx$ is the epipolar line in the second image.
   2. $l = F^T x'$ is the epipolar line in the first image
3. For any point x, the epipolar line $l' = Fx$ contains the epipole e'. Thus:

$$e'^T(Fx) = (e'^T F)x = 0 \;\; \forall x \quad \Rightarrow e'^T F = 0$$

Hence e' is the left null vector of F. Similarly e is the right null vector of F.

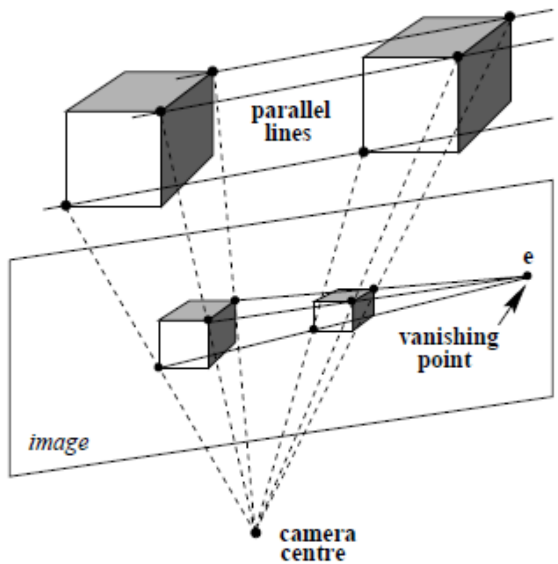
Properties

4. F has 7 degrees of freedom: 8 from the 3x3 homography matrix minus 1 due to the constratint det(F)=0
5. F is a projective map taking point to lines, and viceversa. If l and l' are corresponding epipolar lines, then any point x is mapped to the same line l'. (ie. THE MAPPING IS NOT INVERTIBLE)

F produced by a pure translation

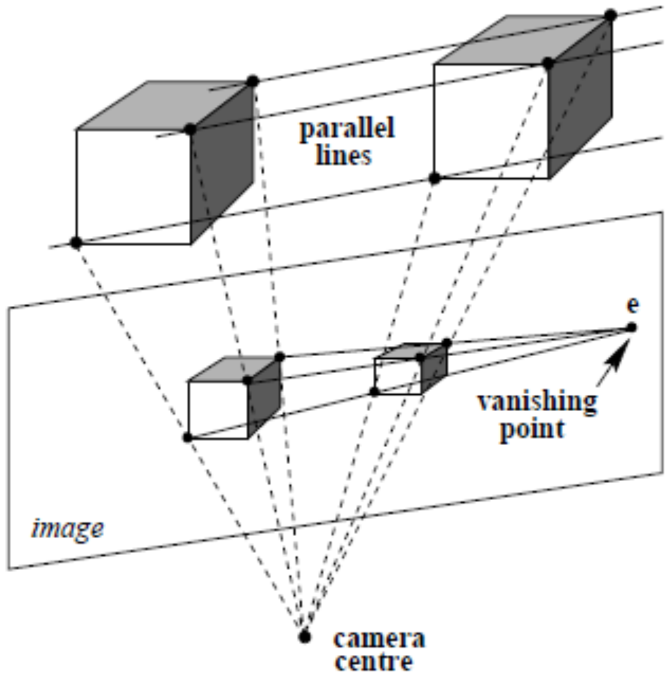
If the camera moves according to a pure translation **t** (ie. Intrinsic parameters remain fixed and rotation **R**=**I**) the points in $\mathbb{P}^3$ "move" on straight lines parallel to t with respect to the camera.

The **imaged intersection** of all those parallel lines is the vanishing point **v** in the direction of **t**.

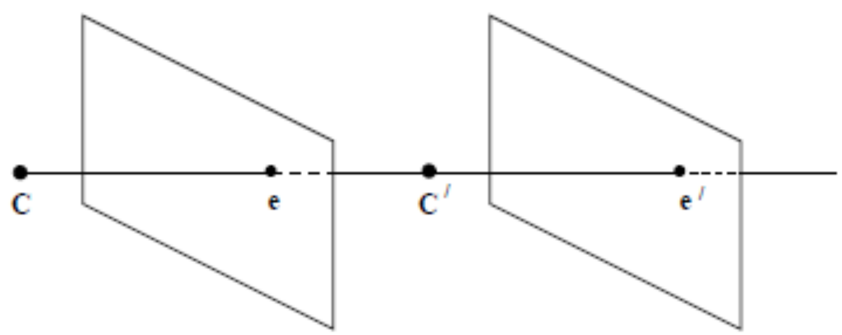


Università
Ca' Foscari
Venezia

F produced by a pure translation



v is the epopole of both views and the imaged parallel lines are the epipolar lines



Università Ca' Foscari Venezia

F produced by a pure translation

Algebraically we have:

$$P = K\left(I \mid 0\right) \qquad P' = K\left(I \mid \mathbf{t}\right)$$

$$F = [e']_\times (P'P^+)$$

$$F = [e']_\times KK^{-1} = [e']_\times$$

If an image point is normalized as $\mathbf{x} = (x, y, 1)^T$, then the (inhomogeneous) 3D coordinates of the originating point $\mathbf{X} = (X, Y, Z)^T$ are $ZK^{-1}x$. After the translation, the point is imaged to:

$$x' = P'X = K(\mathbf{X} + t) = x + \frac{K\mathbf{t}}{Z}$$

F produced by a pure translation

$$x' = x + \frac{K\mathbf{t}}{Z}$$

The extent of the motion in the image plane is:

- Proportional to the magnitude of the translation
- Inversely proportional to the depth Z (point distance)

Points close to the camera move faster than those further away

Projective ambiguity

If H is a 4x4 matrix representing a projective transformation in $\mathbb{P}^3$, then the fundamental matrices corresponding to the pair of camera matrices (P,P') and (PH, P'H) are the same.

This means that a pair of camera matrices (P,P') uniquely determine a fundamental matrix F **but the converse is not true**.

This mapping is not one-to-one but the only ambiguity is a right multiplication by a projective transformation

The essential matrix

Consider a camera matrix $P = K\left(R \mid T\right)$. If K is known we may apply K^{-1} to the image point to obtain a point in retinal plane expressed in normalized coordinates.

So, we may reason about normalized camera matrices $P = \left(I \mid 0\right)$ and $P' = \left(R \mid T\right)$ for which the corresponding fundamental matrix:

$$E = [T]_{\times} R$$

Is called the **essential matrix**. The relationship between the fundamental and essential matrices is

$$E = K'^{T} F K$$

Università
Ca' Foscari
Venezia

The essential matrix

The essential matrix has only 5 degrees of freedom (3 from rotation + 3 from translation -1 from the scale ambiguity).

A 3x3 matrix is an essential matrix if and only if two of its singular values are equal and the third is zero. This property is important to estimate the essential matrix given a set of point correspondences in retinal plane.

Università Ca' Foscari Venezia

RT from E

If the essential matrix E is known, the two normalized camera matrices can be recovered up to four possible choices.

Let E = U diag(1,1,0) V^T be the SVD decomposition of E and $P = (I \mid 0)$ the first (normalized) camera matrix. The four choices for the second camera matrix are:

$$P' = (UWV^T \mid +u_3) \quad P' = (UW^TV^T \mid +u_3)$$
$$P' = (UWV^T \mid -u_3) \quad P' = (UW^TV^T \mid -u_3)$$

with: $$W = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Computing F from points

Since the fundamental matrix defines the constraint

$$x'^T F x = 0$$

between corresponding points $x \leftrightarrow x'$ in the two images, we can collect multiple correspondences to estimate F.

Each point match gives rise to one linear equation in the unknown entries of F:

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0.$$

Università Ca' Foscari Venezia

Computing F from points

From a set of N point matches, we obtain a set of linear equations of the form:

$$\mathbf{Af} = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}$$

- To have exactly one solution, A must have rank 8 and hence we require 8 point matches (8-point algorithm)
- If more point are provided, the least-squares solution is given by the singular vector corresponding to the smallest singular value of A.

Università
Ca' Foscari
Venezia

The singularity constraint

Differently to other methods we have seen to estimate homographies or camera matrices, F must have rank 2 (hence be singular).

If we estimate it via least-squares, this constraint is not satisfied in general.

A convenient way to solve that is to enforce the rank 2 constraint a-posteriori by replacing F with F' according to the following minimization:

$$\text{minimize}_{F'} \|F - F'\| \quad \text{s.t.}\ \det(F') = 0$$

Università Ca' Foscari Venezia

The singularity constraint

$$\text{minimize}_{F'} \|F - F'\| \quad \text{s.t.}\, \det(F') = 0$$

The optimal solution can be obtained via SVD:

1. Let F =USV^T be the SVD decomposition of F
2. Let S = diag(r,s,t) be the diagonal matrix of the singular values of F, in descending order
3. F' = U diag(r,s,0) V^T

Università Ca' Foscari Venezia

Normalized 8-points

The 8-points algorithm works well in practice only if we "normalize" the input data such that:

- The centroid of the reference points is at the origin
- The RMS distance of the points from the origin is $\sqrt{2}$

Objective

Given $n \geq 8$ image point correspondences $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$, determine the fundamental matrix F such that $\mathbf{x}'^{\mathsf{T}}_i \mathrm{F} \mathbf{x}_i = 0$.

Algorithm

(i) **Normalization:** Transform the image coordinates according to $\hat{\mathbf{x}}_i = \mathrm{T}\mathbf{x}_i$ and $\hat{\mathbf{x}}'_i = \mathrm{T}'\mathbf{x}'_i$, where T and T′ are normalizing transformations consisting of a translation and scaling.

(ii) Find the fundamental matrix $\hat{\mathrm{F}}'$ corresponding to the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ by

(a) **Linear solution:** Determine $\hat{\mathrm{F}}$ from the singular vector corresponding to the smallest singular value of $\hat{\mathrm{A}}$, where $\hat{\mathrm{A}}$ is composed from the matches $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ as defined in (11.3).

(b) **Constraint enforcement:** Replace $\hat{\mathrm{F}}$ by $\hat{\mathrm{F}}'$ such that $\det \hat{\mathrm{F}}' = 0$ using the SVD (see section 11.1.1).

(iii) **Denormalization:** Set $\mathrm{F} = \mathrm{T}'^{\mathsf{T}} \hat{\mathrm{F}}' \mathrm{T}$. Matrix F is the fundamental matrix corresponding to the original data $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$.

Triangulation

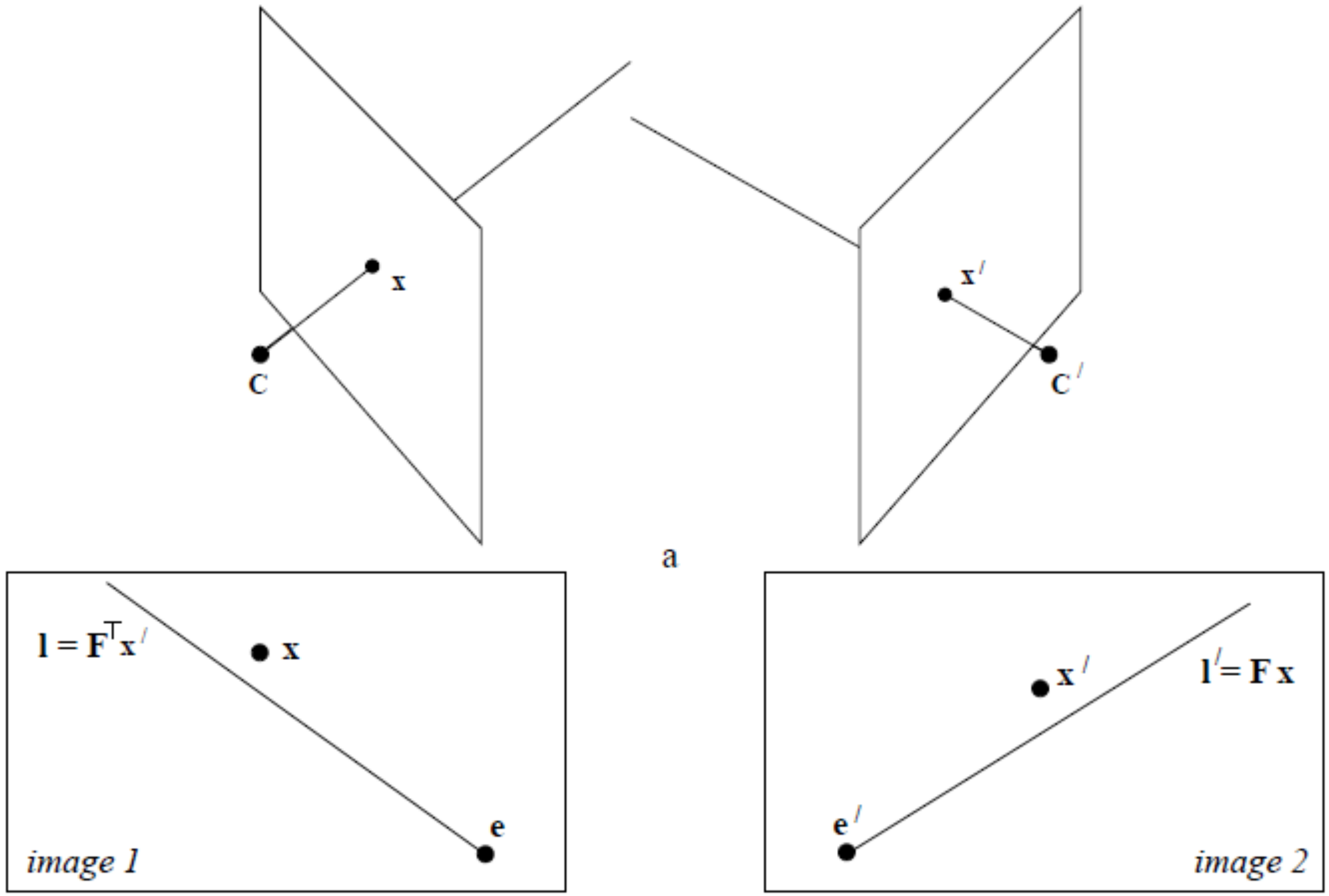
Triangulation refers to the problem of computing the 3D position of a point X given its image x and x' in two views.

We assume that errors are present only in the measured image coordinates (x,x') and not on the camera matrices P and P'.

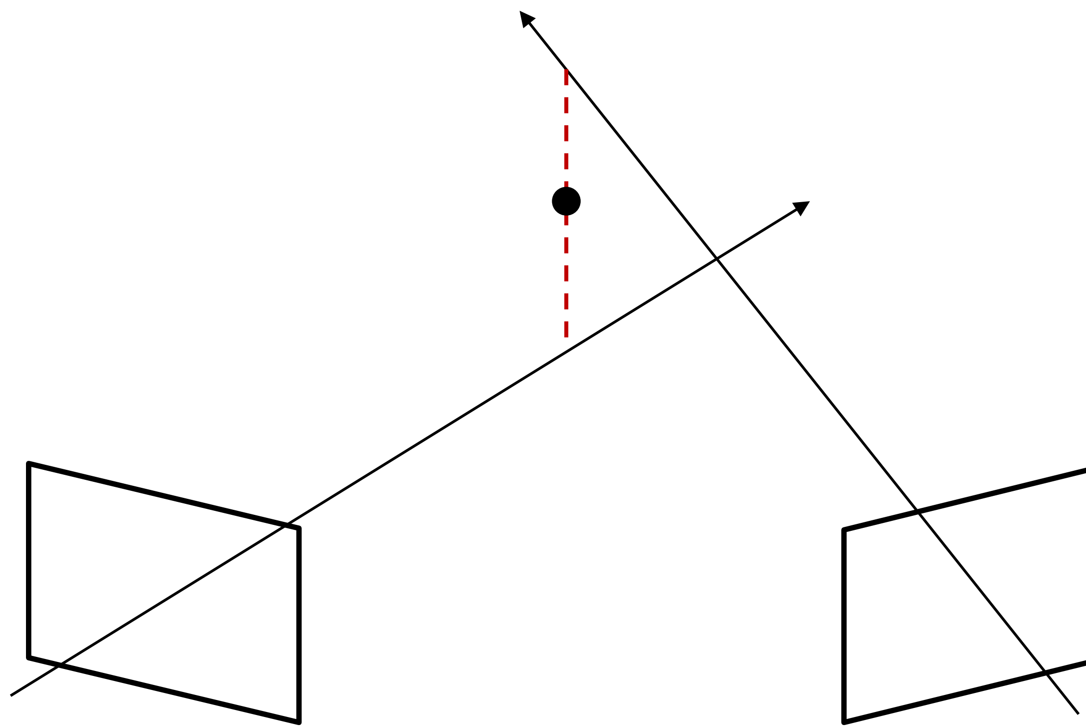
Naïve triangulation by "back-projecting" rays from x and x' fails because in general the two rays will not intersect exactly in space…

… we need to "estimate" a best solution for X

Triangulation

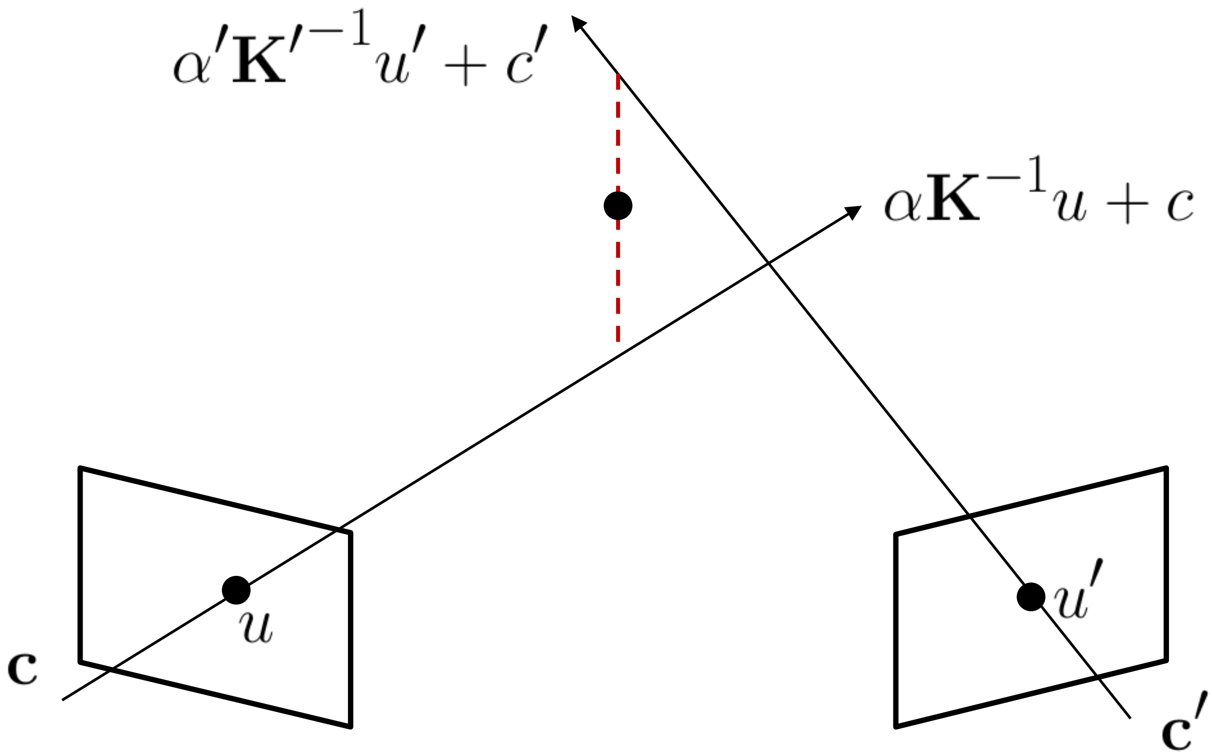


Mid-point triangulation



Choose the mid-point of the common perpendicular to the two rays. This method in general do not give optimal results, because of various approximations (evident if the two cameras are not equally distant to the midpoint)

Mid-point triangulation



To find the midpoint:

$$\alpha \mathbf{K}^{-1} u - \alpha' \mathbf{K}'^{-1} u' = c' - c$$

3 equations, two unknown (α, α') . Then, the midpoint is found as $(c + \alpha \mathbf{K}^{-1} u + c' + \alpha' \mathbf{K}'^{-1} u')/2$

Linear Triangulation methods

Is the most commonly used representing a good tradeoff between simplicity and accuracy.

Consider a projection of a point X:

$$\mathbf{u} = w \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = PX$$

Each component of u can be written as:

$$wu = \mathbf{p}_1^T X \qquad wv = \mathbf{p}_2^T X \qquad w = \mathbf{p}_3^T X$$

Università Ca' Foscari Venezia

Linear Triangulation methods

$$wu = \mathbf{p}_1^T X \qquad wv = \mathbf{p}_2^T X \qquad w = \mathbf{p}_3^T X$$

Substituting w into the first two, and considering two projections u, u' we obtain a system of 4 equations in the form $\mathbf{A}X = 0$:

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

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

$$u'\mathbf{p'}_3^T X = \mathbf{p'}_1^T X$$

$$v'\mathbf{p'}_3^T X = \mathbf{p'}_2^T X$$

A is a 4x4 matrix and we seek a non-zero solution for X (defined up to scale)

Linear Triangulation methods

Two ways to solve the system:

1. If working in **homogeneous** coordinates, find the X that minimize $||\mathbf{A}X||$ subject to $||X|| = 1$. The solution is the unit singular vector corresponding to the smallest singular value
2. By setting $X = (x, y, z, 1)^T$, the set of equations $\mathbf{A}X = 0$ is reduced to a set of 4 **non-homogeneous** equations in 3 unknowns that can be solved as Linear Least Squares.

Differences?

Università
Ca' Foscari
Venezia

Linear Triangulation methods

The second method assumes that the solution is not at infinity. This can cause numerical instability when we are seeking to reconstruct points that may lie (or are close to) the plane at infinity

The second method is **affine invariant**. In other words, if two cameras are transformed by a general affine transformation, the resulting 3D point will be the same point X transformed by the same transformation.

The triangulation is independent by any affine transformations of the reference frame.

Università Ca' Foscari Venezia

Iterative linear methods

The problem of the linear methods is that $\|\mathbf{A}X\|$ has no geometric meaning.

From each equation in the form

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

We'll get an error

$$\epsilon = u\mathbf{p}_3^T X - \mathbf{p}_1^T X$$

The geometric error would instead be given as:

$$\epsilon' = u - \boxed{\frac{\mathbf{p}_1^T X}{\mathbf{p}_3^T X}}$$

x coordinate of the projection of X

Iterative linear methods

$$\epsilon = u\mathbf{p}_3^T X - \mathbf{p}_1^T X \qquad \epsilon' = u - \frac{\mathbf{p}_1^T X}{\mathbf{p}_3^T X}$$

ϵ' can be obtained by weighting ϵ by the factor

$$\frac{1}{w} = \frac{1}{\mathbf{p}_3^T X}$$

Problem: the weights depends on the 3D point X to be estimated

Solution: Iteratively estimate X and update the weights (initial weight = 1)

Università Ca' Foscari Venezia

Iterative linear methods

Algorithm:

1. Set $w = 1, w' = 1$
2. Solve:

$$1/w \; u\mathbf{p}_3^T X = \mathbf{p}_1^T X \, 1/w$$

$$1/w \; v\mathbf{p}_3^T X = \mathbf{p}_2^T X \; 1/w$$

$$1/w' u' \mathbf{p'}_3^T X = \mathbf{p'}_1^T X \; 1/w$$

$$1/w \; v' \mathbf{p'}_3^T X = \mathbf{p'}_2^T X \, 1/w$$

3. Let:

$$w = \mathbf{p}_3^T X \qquad w' = \mathbf{p'}_3^T X$$

4. Return to step 2 until convergence (ie. the change in weights is small)