Simple stereo

We now have a calibrated camera. Let's look to a simple method for recovering the 3-dimensional structure of a scene from two images.
If we have a fully calibrated camera, we can say that each point in the image lies on a outgoing ray. We know the equation of these rays.

3D to 2D (point)

$$u = f_x \frac{x_c}{z_c} + o_x$$

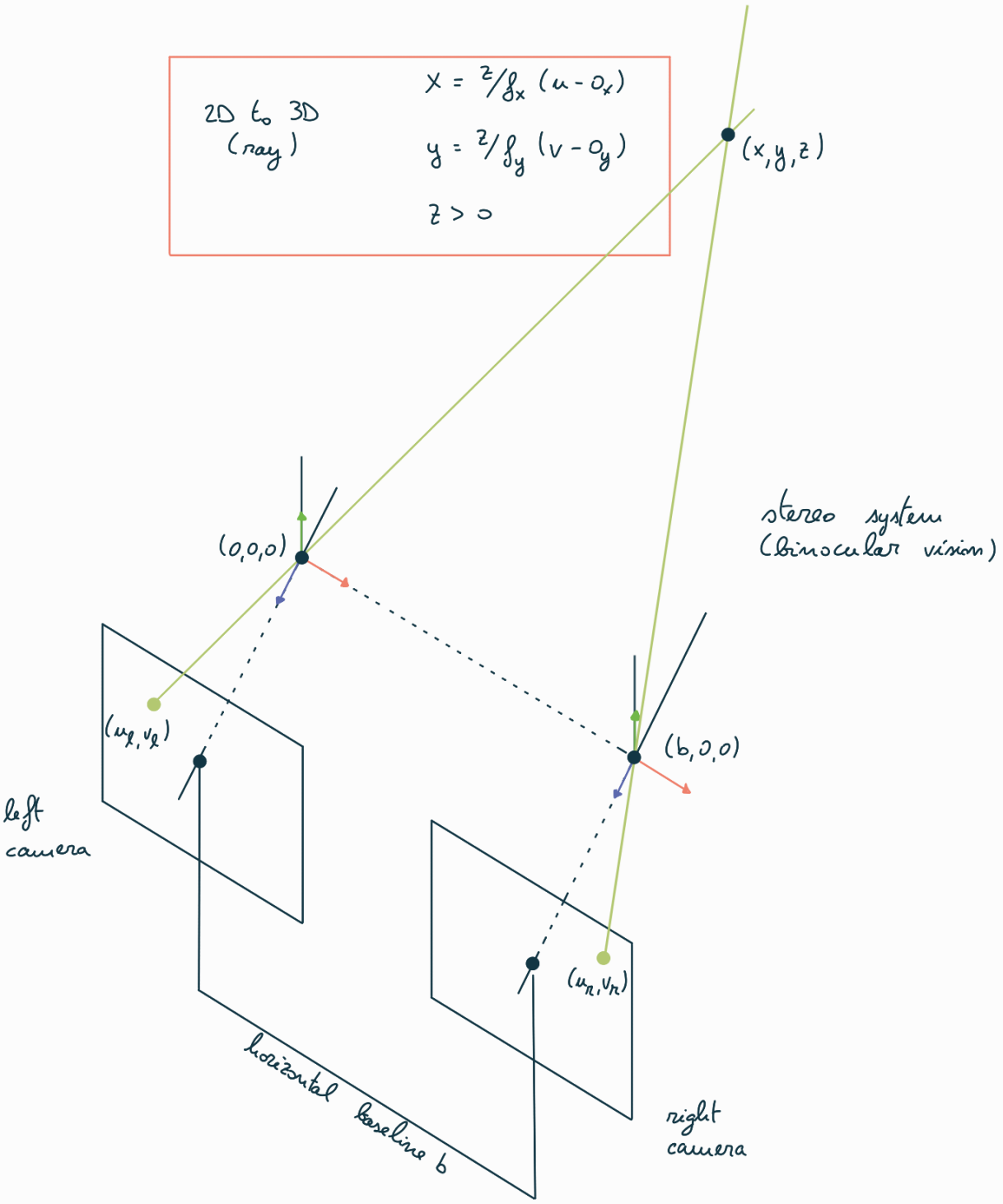
$$v = f_y \frac{y_c}{z_c} + o_y$$

2D to 3D (ray)

$$x = z/f_x \,(u - o_x)$$

$$y = z/f_y \,(v - o_y)$$

$$z > 0$$



If we have the two corresponding points we are left with 4 equations:

$$u_l = f_x \frac{x}{z} + o_x$$

$$v_l = f_y \frac{y}{z} + o_y$$

left point

$$u_r = f_x \frac{x - b}{z} + o_x$$

$$v_r = f_y \frac{y}{z} + o_y$$

right point

We have everything (f_x, f_y, o_x, o_y from the calibration procedure, u_l, v_l, u_r, v_r are supposed as given) except x, y, z.

Solving for x, y, z:

$$x = \frac{b(u_l - o_x)}{(u_l - u_r)}$$

$$y = \frac{b f_x (v_l - o_y)}{f_y (u_l - u_r)}$$

$$\text{depth of the point } (z) = \frac{b f_x}{(u_l - u_r)}$$

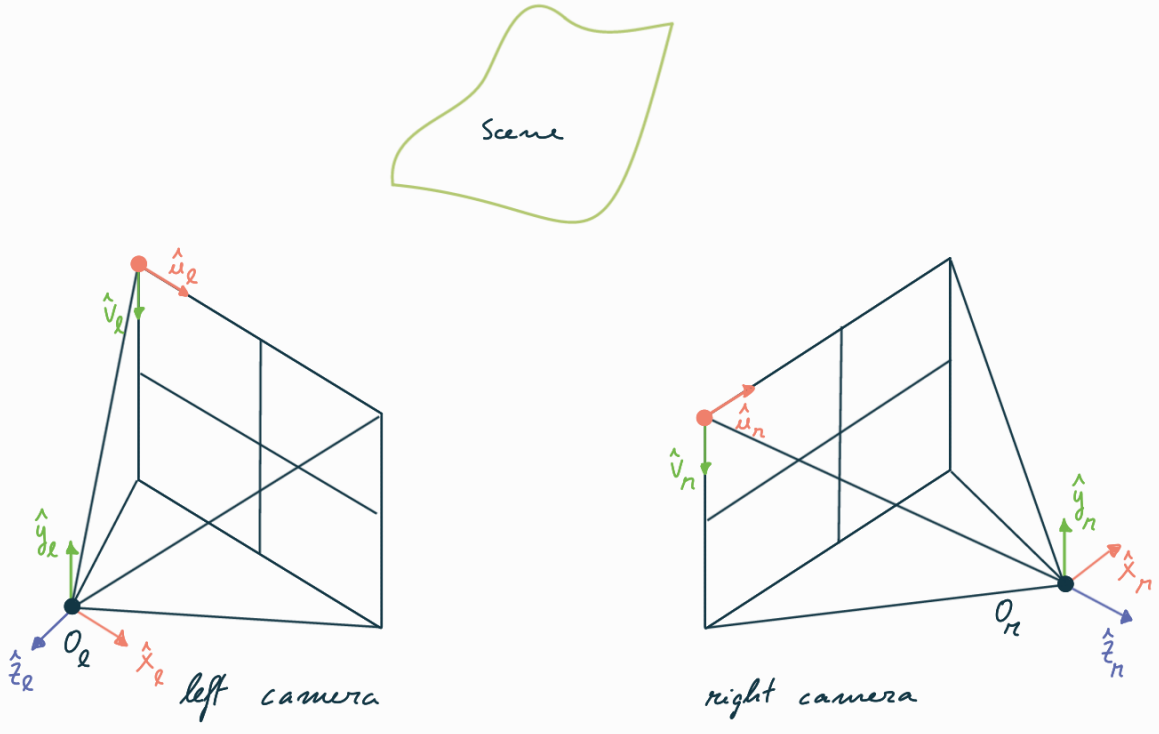
if the point in the second image is not given we need to find it. This is called stereo matching.

where $u_l - u_r$ is called disparity. The disparity is inversely proportional to the depth and proportional to the baseline.

Uncalibrated stereo

If you know the internal parameters of the two cameras it is possible to compute the translation and rotation of one camera with respect to the other one from two arbitrary views: we can then estimate the 3D structure of a static scene from two arbitrary views.
Often these days the internal parameters of an image are stored in its metadata; what we don't know is the translation and rotation among the points where two images of the same 3D scene are taken (two different cameras or same one).

To reconstruct the scene we need to formulate the relationship between left and right cameras. This is very concisely described in what's called epipolar geometry. Epipolar geometry tells us that points in the left and right image are related to each other through a sing 3x3 matrix called a fundamental matrix. If we can find the fundamental matrix we can find rotation and translation of one camera with respect to the other one.
We can develop a method to estimate the fundamental matrix F by using a small number of corresponding points in these two arbitrary views.
Once we have F we can find rotation and translation and the uncalibrated stereo is now fully calibrated.
Now we can recover the structure of the scene.



① We assume that the camera matrix K (intrinsic parameters) is known for each camera

② Find a small number of Reliable Corresponding Points between those two arbitrary views (e.g. SIFT). We will get a set of points:

$$(u_l^{(1)}, v_l^{(1)}) \qquad (u_r^{(1)}, v_r^{(1)})$$
$$\vdots \qquad\qquad \vdots$$
$$(u_l^{(m)}, v_l^{(m)}) \qquad (u_r^{(m)}, v_r^{(m)})$$

Turns out we don't need much points, we need a minimum of 8.

③ Find Relative Camera Position t and Orientation R

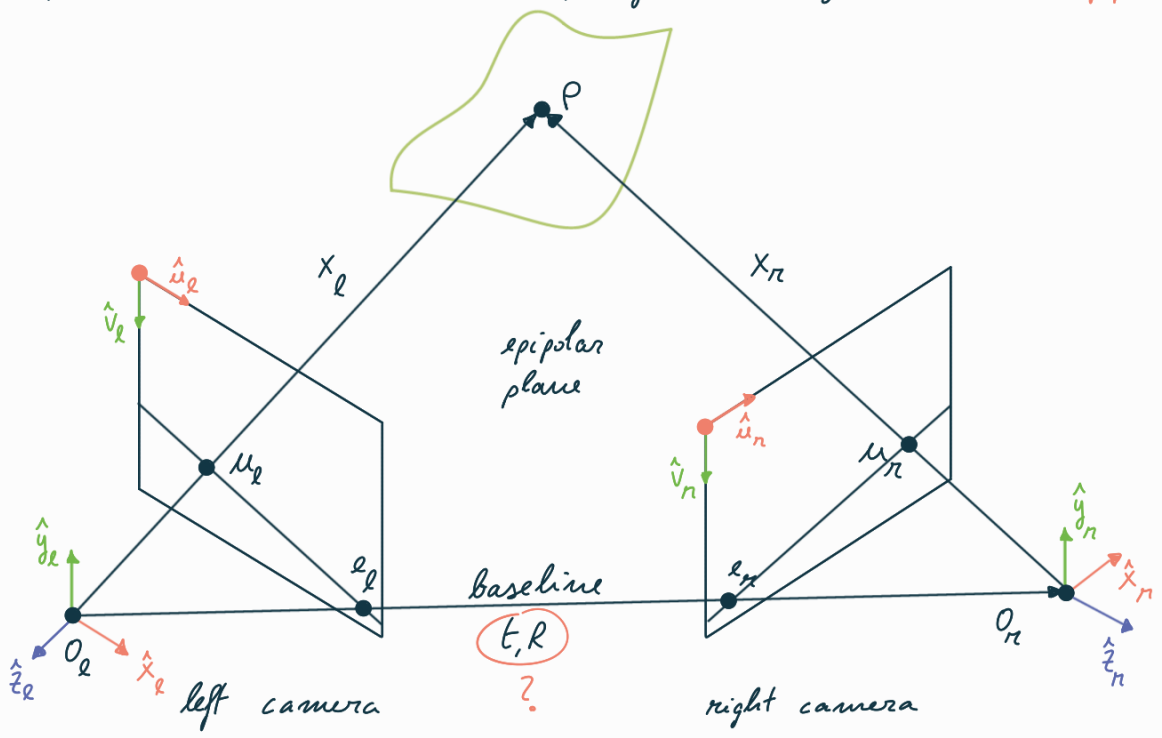
now the uncalibrated stereo system is calibrated

④ Find Dense Correspondance (1D search)

⑤ Compute Depth using Triangulation

Epipolar geometry

Goal: find the relative position and orientation of one camera with respect to the other; this is the process of calibrating an uncalibrated stereo system.
Relative position and orientation are completely described by what's called epipolar geometry.

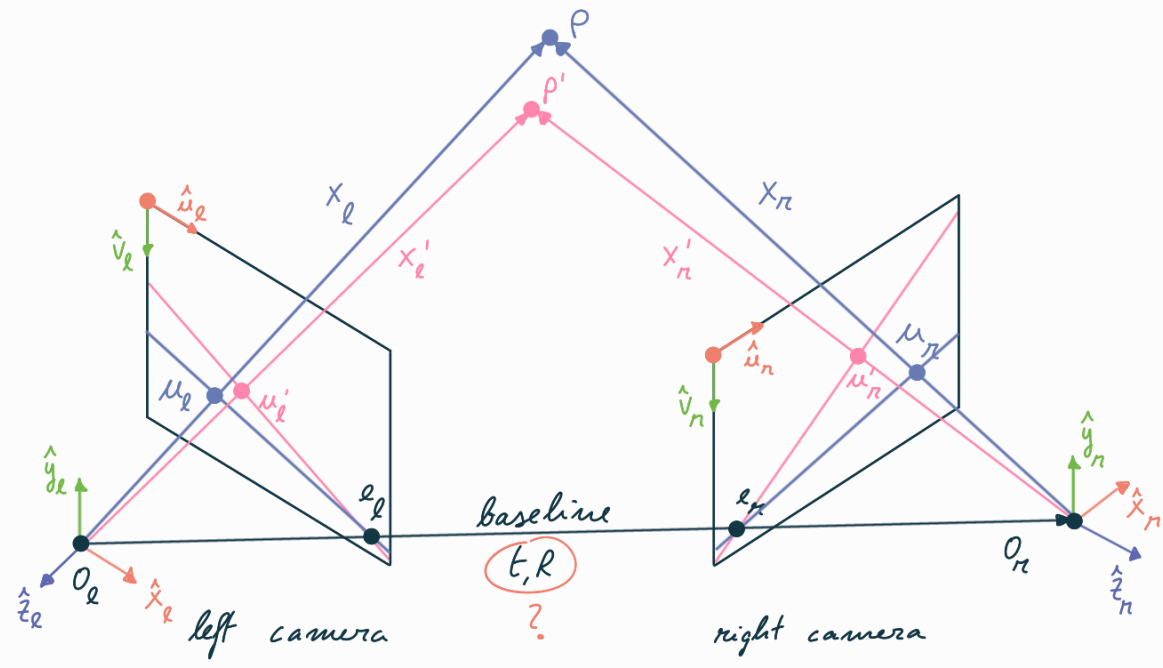


Baseline : line connecting the two camera centers

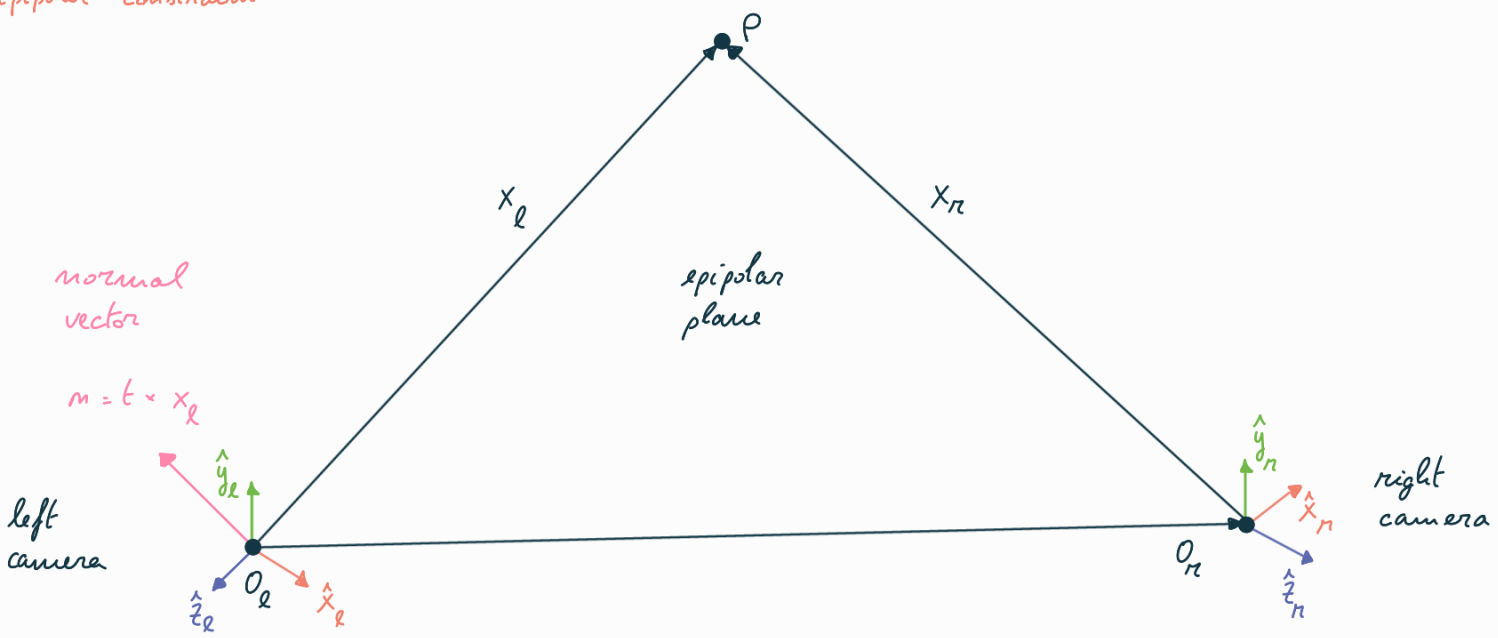
Epipole : image point of origin/pinhole of one camera as viewed by the other camera; every stereo system has a unique pair of epipoles. Equivalently, each epipole is the intersection point between the baseline and the image frame.

Epipolar plane : plane formed by camera origins (O_l and O_r), epipoles (e_l and e_r) and scene point P. Each point in the scene lies on a unique epipolar plane. There is a one parameter family, or a pencil, of epipolar planes.

Epipolar line : intersection of an epipolar plane with the image plane. All epipolar lines intersect at the epipole



Epipolar constraint



Vector normal to the epipolar plane : $n = t \times x_l$

The normal vector n should be perpendicular to x_l : $n \cdot l_l = 0$

Epipolar constraint : dot product of n and x_l (perpendicular vectors) is zero :

$$x_l \cdot (t \times x_l) = 0$$

writing it in vector form :

$$\begin{vmatrix} x_l & y_l & z_l \end{vmatrix} \begin{vmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{vmatrix} = 0 \qquad \text{(cross product definition)}$$

writing it in matrix form :

$$\begin{vmatrix} x_l & y_l & z_l \end{vmatrix} \begin{vmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ t_y & t_x & 0 \end{vmatrix} \begin{vmatrix} x_l \\ y_l \\ z_l \end{vmatrix} = 0 \qquad \text{(matrix vector form)}$$

epipolar constraint

We have another constraint :

$t_{3\times1}$: Position of Right Camera in left Camera's Frame

$R_{3\times3}$: Orientation of left Camera in Right Camera's Frame

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

relationship among x_l and x_r

Putting these two constraints together we have:

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

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

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

$t \times t = 0$

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

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

Essential Matrix E

The essential matrix expresses the relationship between x_l and x_r

$$E = T_x R$$

It is possible to decompose E back into T_x and R

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

Given that T_x is a Skew-Symmetric matrix ($a_{ij} = -a_{ji}$) and R is an Orthonormal matrix, it is possible to decompose E back into T_x and R from their product using Single Value Decomposition (SVD).
Once you have E you can compute t and R and calibrate the uncalibrated stereo.

$$x_l^T E x_r = 0$$

- x_l: 3D position in left camera coordinates
- E: essential matrix
- x_r: 3D position in right camera coordinates

We don't have x_l and x_r but we have their projections

We will try to use the projections of x_l and x_r to estimate the essential matrix.

Back to perspective projection equations

$$u_l = f_x^{(l)} \frac{x_l}{z_l} + o_x^{(l)} \longrightarrow z_l u_l = f_x^{(l)} x_l + z_l o_x^{(l)}$$

$$v_l = f_y^{(l)} \frac{x_l}{z_l} + o_y^{(l)} \longrightarrow z_l v_l = f_y^{(l)} x_l + z_l o_y^{(l)}$$

Representing in matrix form:

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

Known camera matrix K_l → we have the internal parameters

left camera

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

right camera

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

$$x_l^T = \begin{vmatrix} u_l & v_l & 1 \end{vmatrix} z_l K_l^{-1\,T} \qquad x_r = K_r^{-1} z_r \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix}$$

$$x_l^T E x_r = 0$$

$$\boxed{\begin{vmatrix} u_l & v_l & 1 \end{vmatrix} z_l K_l^{-1\,T} E K_r^{-1} z_r \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix} = 0}$$

We don't have z_l. We know z_l (depth of the 3D point in the scene) $\neq 0$: if it is 0, it is on the same plane of the camera center

$$\begin{vmatrix} u_l & v_l & 1 \end{vmatrix} \cancel{z_l} \underbrace{K_l^{-1\,T}}_{3\times3} E \underbrace{K_r^{-1}}_{3\times3} \cancel{z_r} \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix} = 0$$

$K_l^{-1\,T} E K_r^{-1}$ is a 3x3 matrix and it's called fundamental matrix F.

The fundamental matrix is an expression of the epipolar constraints.

$$\begin{vmatrix} x_l & y_l & 1 \end{vmatrix} \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix} = 0$$

Once we have F :

$E = K_R^T F K_\pi$

Once we have E :

$E = T_x R$ (with Single Value Decomposition)

We only need to find a single 3x3 matrix, the fundamental matrix, to calibrate the uncalibrated stereo system

Stereo Calibration Procedure

The epipolar constraint tells us that there is a single 3×3 matrix, the fundamental matrix, that we need to find to calibrate our uncalibrated stereo system.
Our goal is to estimate the fundamental matrix.

① Find a set of corresponding features in left and right image (e.g. using SIFT)

$$(u_l^{(1)}, v_l^{(1)}) \qquad (u_r^{(1)}, v_r^{(1)})$$
$$\vdots \qquad\qquad \vdots$$
$$(u_l^{(m)}, v_l^{(m)}) \qquad (u_r^{(m)}, v_r^{(m)})$$

② For each correspondance i, write the epipolar constraint

$$\begin{vmatrix} x_l^{(i)} & y_l^{(i)} & 1 \end{vmatrix} \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} \begin{vmatrix} u_r^{(i)} \\ v_r^{(i)} \\ 1 \end{vmatrix} = 0$$

Known — unKnown — Known

Expanding the expression we get a linear equation:

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

we get one such linear equation for each same point.

③ Stack them to form a linear system

$$\begin{vmatrix} u_r^{(1)}u_l^{(1)} & v_r^{(1)}u_l^{(1)} & u_l^{(1)} & u_r^{(1)}v_l^{(1)} & v_r^{(1)}v_l^{(1)} & v_l^{(1)} & u_r^{(1)} & v_r^{(1)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_r^{(i)}u_l^{(i)} & v_r^{(i)}u_l^{(i)} & u_l^{(i)} & u_r^{(i)}v_l^{(i)} & v_r^{(i)}v_l^{(i)} & v_l^{(i)} & u_r^{(i)} & v_r^{(i)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_r^{(m)}u_l^{(m)} & v_r^{(m)}u_l^{(m)} & u_l^{(m)} & u_r^{(m)}v_l^{(m)} & v_r^{(m)}v_l^{(m)} & v_l^{(m)} & u_r^{(m)} & v_r^{(m)} & 1 \end{vmatrix} \begin{vmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{vmatrix} = \begin{vmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{vmatrix}$$

Known — unKnown

$$Af = 0$$

Fundamental matrix acts on homogeneous coordinates

$$\begin{vmatrix} x_l & y_l & 1 \end{vmatrix} \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix} = 0 = \begin{vmatrix} x_l & y_l & 1 \end{vmatrix} \begin{vmatrix} kf_{11} & kf_{12} & kf_{13} \\ kf_{21} & kf_{22} & kf_{23} \\ kf_{31} & kf_{32} & kf_{33} \end{vmatrix} \begin{vmatrix} u_r \\ v_r \\ 1 \end{vmatrix}$$

Fundamental matrix is only defined up to a scale ($F = kF$).
We can set it to some arbitrary scale

$$\|f\|^2 = 1 \quad \text{(for convenience)}$$

④ Find least squares solution for fundamental matrix F.

The 8 point algorithm

We want Af as close as possible to 0 and $\|f\|^2 = 1$:

$$\min_f \|Af\|^2 \text{ such that } \|f\|^2 = 1$$

constrained linear least squares problem

⑤ Rearrang f to get matrix F.

⑥ Comput essential matrix from fundamental matrix

$$E = K_l^T F K_r$$

⑦ Extract R and t from E using Singular Value Decomposition

$$E = T_x R$$

Fully calibrated stere system