Vision and Image Processing: Camera Models, Homographies

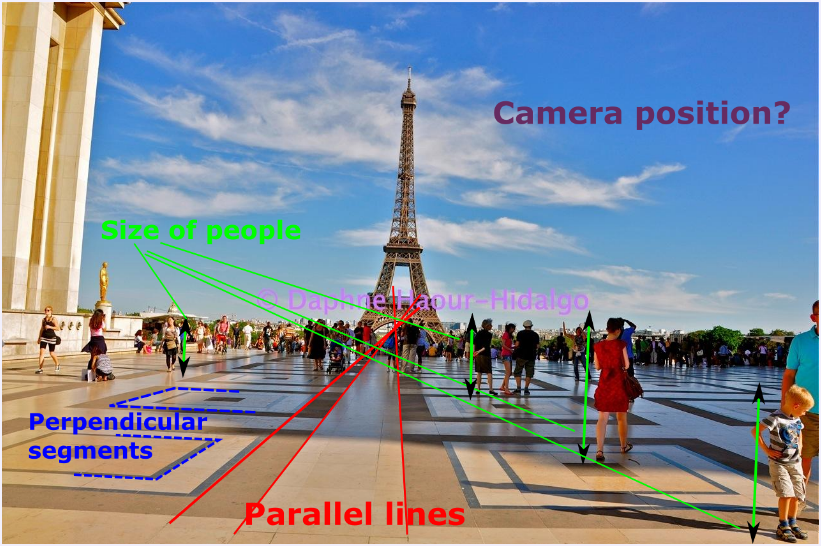
Søren Olsen

Department of Computer Science
University of Copenhagen

Plan for today

- Introduction to Camera Models, specifically the pinhole model and the perspective transformation
- Vanishing points and vanishing lines
- Homogeneous coordinates, Homographies
- Camera matrices and parameters
- Camera calibration
- Triangulation
- Questions for ssignment 4

Cameras and projections

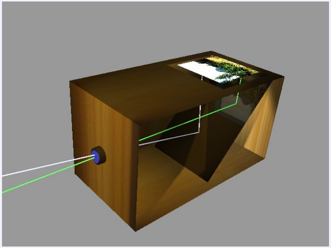


Questions

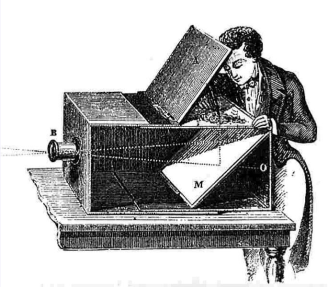
Previous picture raises some questions about:

- Lines?
- Parallelism?
- Angles / orthogonality?
- Sizes?
- Camera position / Horizon?

Camera Obscura



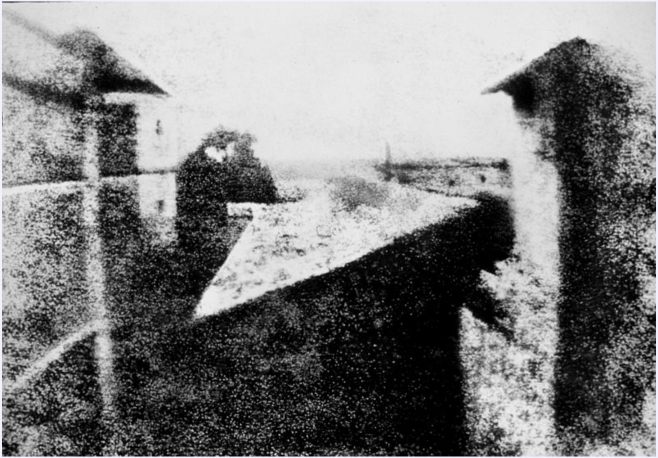
Principle of Camera Obscura



18th Century Camera Obscura

- Known from old chines writings
- Mentioned by Aristotle
- Plaque with photosensitive material: Photographic camera!

The Very First Photography, 1826

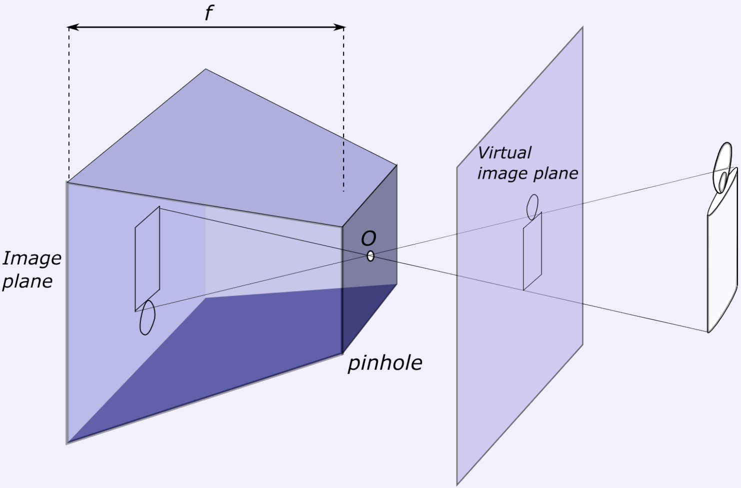


J.N. Niépce, View from the window at Le Gras, Saint Loup de Varennes, France – Now at University of Texas at Austin.

The pin-hole camera

Magnesium light was used to make light enough enter the pin-hole box.

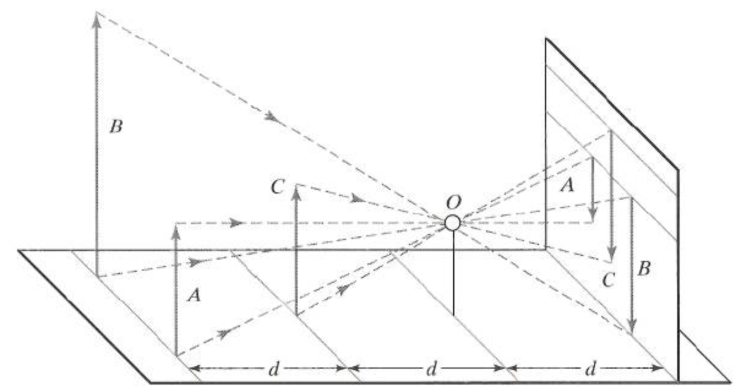
The Pinhole Camera Model



- *f* is the focal length,
- *O* is the camera center.

Perspective Effects

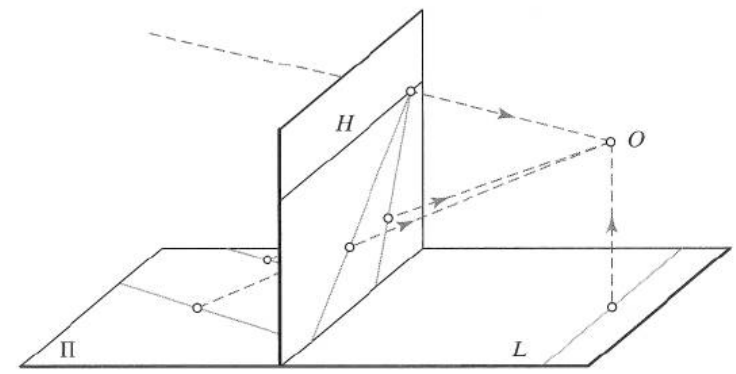
Remember from first lecture



Far objects appear smaller that close ones.

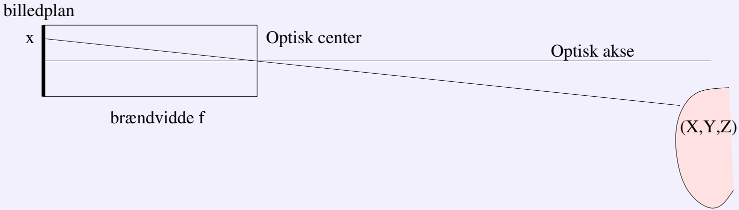
Perspective Effects

Remember from first lecture again



Images of parallel lines intersect at the horizon (virtual image plane).

Projection Equations



- Project $P(X, Y, Z)$ onto $(x, y, -f)$. Remember: similar triangles:

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}$$

- To get non-negative pixel indexes:

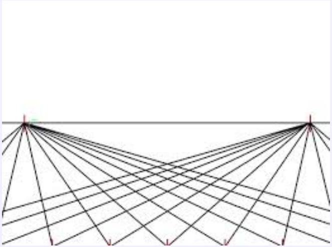
$$x \; = \; f\frac{X}{Z} \; + \; c_x, \quad y \; = \; f\frac{Y}{Z} \; + \; c_y$$

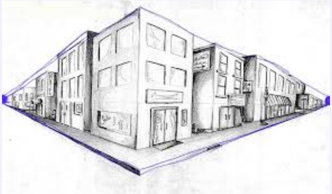
where (c_x, c_y) is called the principal point.

Vanishing Points

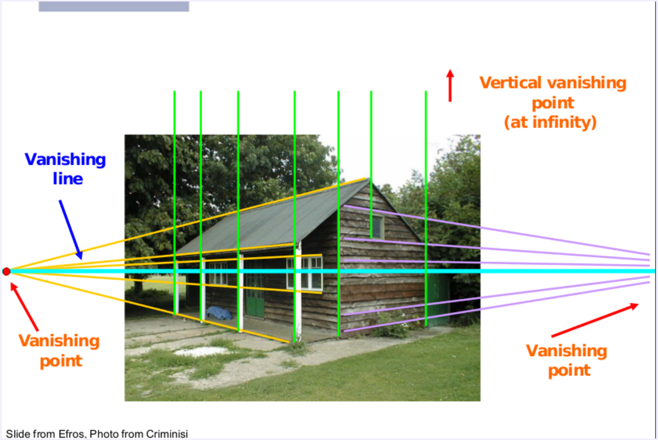
Projections of parallel lines intersect at common points.

Vanishing line





Vanishing line

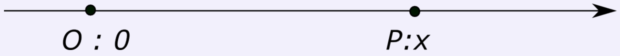


Shape (VP's) from texture

If we can measure a texture density, then we may estimate the directions of minimal and maximal density change. Theses points at the VPs. thus giving the VL and the 3D surface normal.

Homogeneous coordinates 1D

- In 1D coordinate is just 1 number.



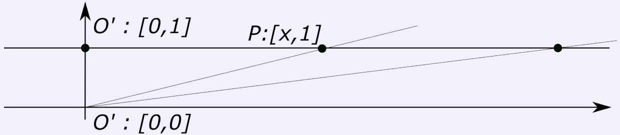
- 1D coordinate to 1D Homogeneous coordinates:

$$x \sim \begin{bmatrix} x \\ 1 \end{bmatrix}$$

- 1D homogeneous coordinate to 1 D coordinate

$$\begin{bmatrix} x \\ w \end{bmatrix} \sim x/w$$

- What can we do with that? we can "tame" infinity!
- A point with homogeneous coordinate $[x, 0]^T$ has "normal" coordinate $x/0 = \infty$ as if we took homogeneous coordinate $[\infty, 1]^T$



Homogeneous coordinates, 2D

"Natural Coordinates" for projective geometry

- From 2D point coordinate to 2D Homogeneous coordinate

$$(x, y) \Longrightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- From 2D homogeneous coordinates to 2D coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Longrightarrow (x/w, y/w)$$

Homogeneous coordinates, 3D

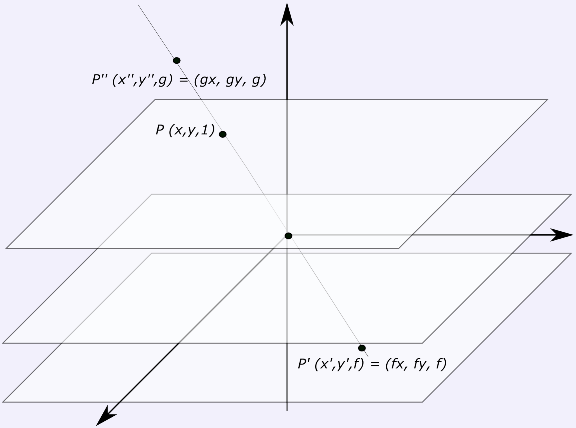
- From 3D point coordinate to 3D Homogeneous coordinate

$$(x, y, z) \Longrightarrow \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- From 3D homogeneous coordinates to 3D coordinates

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Longrightarrow (x/w, y/w, z/w)$$

Homogeneous coordinates



Homogeneous coordinates in 2D correspond to points in plane $z = 1$ but also to lines through the origin and this point.

Cameras and homogeneous coordinates

- Projection to image plane in standard coordinates:

$$P : (x, y, z) \mapsto P' : (f\frac{x}{z}, f\frac{y}{z})$$

- How does the mapping look like in homogeneous coordinates:

$$P : \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \mapsto P' : \begin{bmatrix} fx \\ fy \\ z \end{bmatrix}$$

- Matrix notation

$$\begin{bmatrix} fx \\ fy \\ z \end{bmatrix} = \underbrace{\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{K} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

- K is (a simple version of) the Camera Calibration Matrix and contains the intrinsic calibration parameters (here f).

World, Camera and Image Coordinates

In the previous slides we were not precise on the many coordinate systems are implicitly used:

- 3D World Coordinates: Coordinate system of the 3D world.
- Camera Coordinates: 3D coordinate system attached to the camera.
- Image Coordinates: 2D Coordinate system attached to the image plane.
- The coordinate system for the sampled and digitised image

Intrinsic vs Extrinsic Camera Parameters

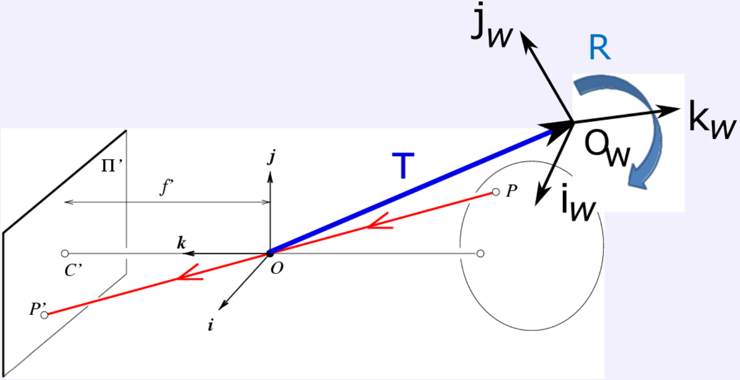
Intrinsic parameters refer to internal parameters:

- The principal point: 2 parameters
- Scale factors (or sampling frequencies) for the pixels sizes in both x and y directions: 2 parameters, multiplied by the focal length, resulting in the effective focal length $(f_x, f_y) = f \cdot (s_x, s_y)$.
- Skewness of pixels: 1 parameter. (Often assumed zero)

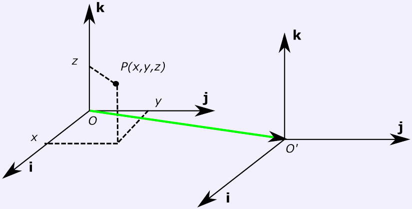
Extrinsic Camera parameters:

- Position of the camera coordinate system in the world coordinates system: translation: 3 parameters,
- Orientation of the camera coordinate system in the world coordinates system: rotation: 3 parameters.

Oriented and Translated Camera



Translating Coordinate System



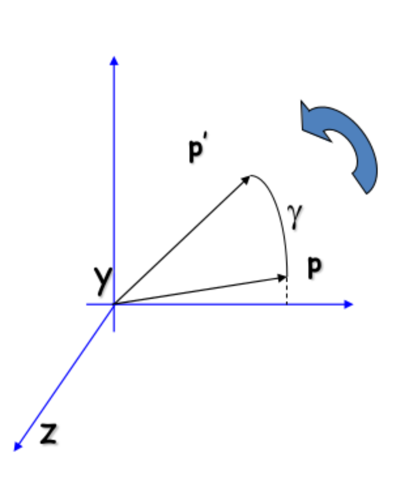
- Subtract the translation vector:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} x_{O'} \\ y_{O'} \\ z_{O'} \end{pmatrix} = \begin{pmatrix} x - x_{O'} \\ y - y_{O'} \\ z - z_{O'} \end{pmatrix}$$

- Transformation in homogeneous coordinates:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 & -x_{O'} \\ 0 & 1 & 0 & -y_{O'} \\ 0 & 0 & 1 & -z_{O'} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Rotating Coordinate System



Rotations along coordinate axes

$$R_{x\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

$$R_{y\beta} = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix}$$

$$R_{z\gamma} = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Can also be written in homogeneous coordinates

3D rotations

To obtain a 3D rotation apply all 3 single-axis rotations:

$$
\begin{aligned}
R &= R_{x\alpha} R_{y\beta} R_{z\gamma} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

Please notice that the order of the 3 rotation matrices do matter. However, no matter which order (α, β, γ) may be found such that the resulting 3D rotation is correct.

Camera Matrix

- Camera calibration matrix, now extended with Image plane transformation (axis scalings, shear, translation)

$$\mathbf{C} = \mathbf{K}\,[\mathbf{R}\ \ \mathbf{t}]$$

- $\mathbf{K}3 \times 3$ matrix encoding the homogeneous transformations inside the camera. $\mathbf{K}$ specifies the Intrinsic parameters.
- $[\mathbf{R}\ \ \mathbf{t}]$ Concatenation of world coordinates rotation and origin translation to align camera and world coordinates.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} = \underbrace{\begin{pmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathbf{K}} \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{pmatrix}}_{[\mathbf{R}\ \ \mathbf{t}]} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Exercise

- How many independent parameters are there to estimate in a camera calibration ?

Exercise

- How many independent parameters are there to estimate in a camera calibration ?

  11: 5 intrinsic and 6 extrinsic (sometimes simplified to 3+6)

- May all calibration parameters be found using Linear algebra ?

Exercise

- How many independent parameters are there to estimate in a camera calibration ?

  11: 5 intrinsic and 6 extrinsic (sometimes simplified to 3+6)

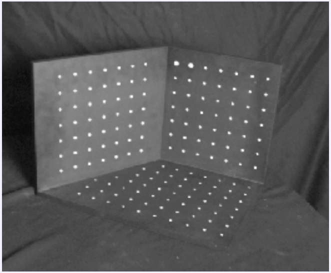
- May all calibration parameters be found using Linear algebra ?

  No, they don’t appear in a linear combinations. Some are multiplied together.

QUESTIONS ?

Geometric Calibration

- Computing the camera matrix is called geometric calibration.
- Extrinsic parameters: (R,t): Usually easy, but requires metric knowledge on scene features.
- Intrinsic parameters: (**K**): Some parameters (f) are easy and some ((u_0, v_0)) are difficult to estimate correctly.



- Use an object with known geometry
- Use vanishing points / lines
- Use other cues...

Repetition: The camera matrix

Please recall the definition of the camera matrix:

$$M = K\,[R\ \mathbf{t}]$$

and the projection written out:

$$w\begin{bmatrix}x\\y\\1\end{bmatrix} = \underbrace{\underbrace{\begin{pmatrix}f_x & 0 & u_0\\ 0 & f_y & v_0\\ 0 & 0 & 1\end{pmatrix}}_{\mathbf{K}}\underbrace{\begin{pmatrix}r_{11} & r_{12} & r_{13} & t_x\\ r_{21} & r_{22} & r_{23} & t_y\\ r_{31} & r_{32} & r_{33} & t_z\end{pmatrix}}_{[\mathbf{R}\ \mathbf{t}]}}_{\mathbf{M}}\begin{pmatrix}X\\Y\\Z\\1\end{pmatrix}$$

or

$$w\begin{bmatrix}x\\y\\1\end{bmatrix} = \begin{pmatrix}\mathbf{m}^1\\ \mathbf{m}^2\\ \mathbf{m}^3\end{pmatrix}\begin{pmatrix}X\\Y\\Z\\1\end{pmatrix} = \begin{pmatrix}\mathbf{m}^1 U\\ \mathbf{m}^2 U\\ \mathbf{m}^3 U\end{pmatrix}$$

where $\mathbf{m}^i$ is the i'th row of the camera matrix.

Camera calibration

In Camera calibration the task is to recover the 12 unknowns $\mathbf{m}_{ij}$ in the camera calibration matrix. Converting the result before to image coordinates we have:

$$\begin{aligned} x\mathbf{m}_3 U &= \mathbf{m}_1 U \\ y\mathbf{m}_3 U &= \mathbf{m}_2 U \end{aligned}$$

Let $\mathbf{m} = (m_{11}, \ldots, m_{14}, \ldots, m_{34})^\top$ be the vector of the 12 unknowns. Isolating these in the equations above lead to: $A\mathbf{m} = 0$, where:

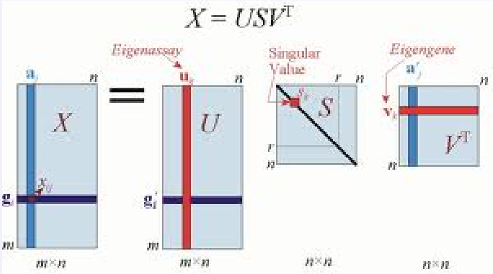
$$A = \begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 Z_1 & -x_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 & -y_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -x_2 X_2 & -x_2 Y_2 & -x_2 Z_2 & -x_2 \\ 0 & 0 & 0 & 0 & X_2 & Y_2 & Z_2 & 1 & -y_2 X_2 & -y_2 Y_2 & -y_2 Z_2 & -y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -x_N X_N & -x_N Y_N & -x_N Z_N & -x_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -y_N X_N & -y_N Y_N & -y_N Z_N & -y_N \end{bmatrix}$$

Solving $Ax = \mathbf{0}$

Many different numerical approaches. Singular Value Decomposition (SVD) is based on decomposition of the matrix A into a product of 3 matrices:

$$A = UDV^\top$$

where A and U is $m \times n$ and $U^\top U = I$,
D is a diagonal $n \times n$ matrix with non-negative singular values,
and V is a $n \times n$ unitary matrix $VV^\top = V^\top V = I$.



Singular value decomposition (svd)

$$A\,\mathbf{x} \;=\; UDV^\top\,\mathbf{x} \;=\; 0$$

We are not interested in the trivial solution $\mathbf{x} = \mathbf{0}$ but rather all solution lying in the Null-space of the matrix A.

The singular values in the diagonal matrix D specifies the energy contained in each of the dimensions. Usually, these are sorted in decreasing order.

Often, due to noise all singular values $\sigma_i > 0$. If we know that A should be singular may find the closest such matrix by zeroing out the smallest singular value.

We get the solution to $A\mathbf{x} = \mathbf{0}$ as the last column of V. In MATLAB:

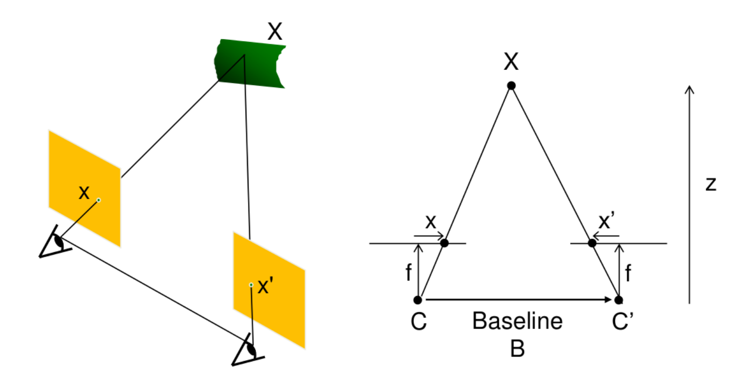
```
[U,D,V] = svd(A); x = V(:,end);
```

Notice that $||x|| = 1$

Camera calibration continued

- Given 6 points (x, y) projected from 6 known 3D points (X, Y, Z) we can solve for the camera matrix elements m_{ij}.
- In practise many more than 6 points are preferred to cancel noise.
- From the 12 elements m_{ij} it is possible to recover all 11 parameters f_x, f_y, α, β, γ, etc.
- There exist more advanced calibration methods than the shown linear calibration.

Simple depth recovery



If we can recover x' from x we can recover depth: $z = -\frac{fB}{x' - x}$.

Linear triangulation for 2 calibrated cameras

Let $U = (X, Y, Z, 1)^\top$. For the first coordinate in the left camera we have:

$$x^L = \frac{\mathbf{m}_L^1 U}{\mathbf{m}_L^3 U}$$

and similar for the y-coordinate and the right camera. Multiplying by the denominator we get:

$$\begin{aligned}
\mathbf{m}_L^3 x_L U &= \mathbf{m}_L^1 U \\
\mathbf{m}_L^3 y_L U &= \mathbf{m}_L^2 U \\
\mathbf{m}_R^3 x_R U &= \mathbf{m}_R^1 U \\
\mathbf{m}_R^3 y_R U &= \mathbf{m}_R^2 U
\end{aligned}$$

Subtracting the right side and putting U outside a parentesis we get the homogeneous equation $AU = 0$, where:

$$A = \begin{pmatrix} \mathbf{m}_L^3 x_L - \mathbf{m}_L^1 \\ \mathbf{m}_L^3 y_L - \mathbf{m}_L^2 \\ \mathbf{m}_R^3 x_R - \mathbf{m}_R^1 \\ \mathbf{m}_R^3 y_R - \mathbf{m}_R^2 \end{pmatrix}$$

More triangulation

- If the calibration matrices $\mathbf{M}_{L/R}$ is not available, a projective reconstruction still is possible. This will be relative to say the left camera and valid up to a projective transformation only.
- Notice that the dertived triangulation is for points. No concepts of surfaces etc. is imposed. Analysis of the derived point cloud data is a topic in itself.
- Remember that no (true) reconstructions are possible for partially occluded areas (often related to surface discontinuities).
- There is no evidence that humans do triangulations. Stereo analysis is a significant 3D inforrmation source, but seem to be based on disparity information alone.

QUESTIONS ?

QUESTIONS ?

QUESTIONS to assignment 4 ?

Next time

Next time, our first meeting in 2022, we will continue with stereo analysis, multi view geometry and may be a bit about image stitching.