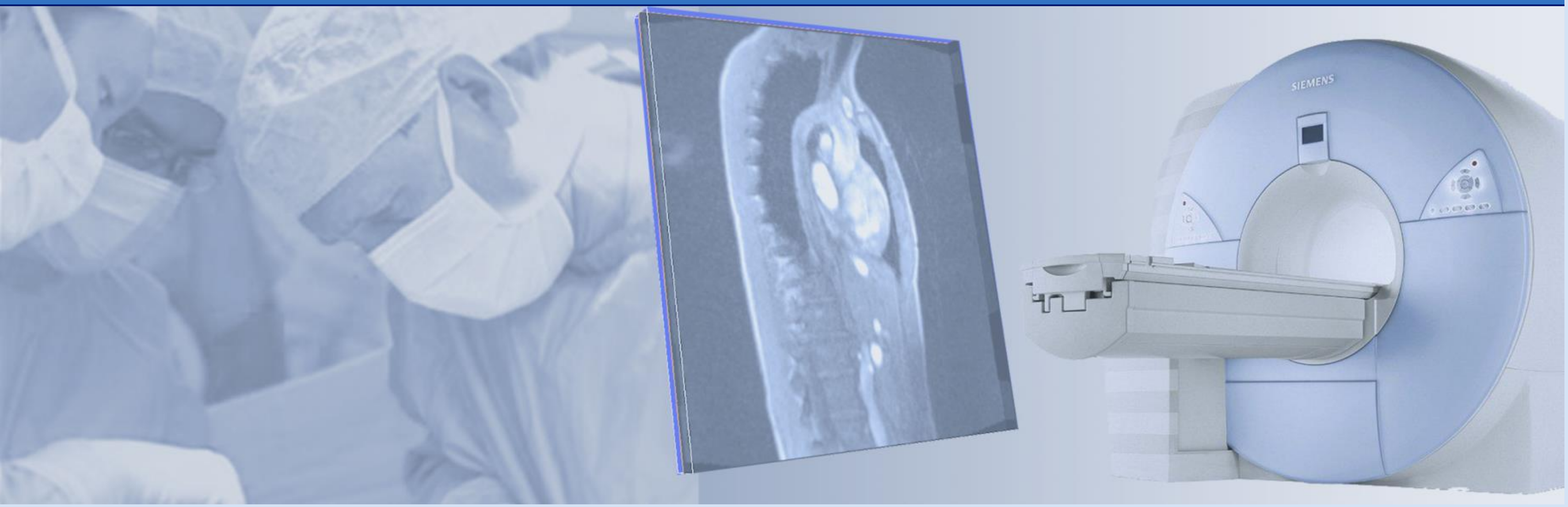
Computer- and robot-assisted Surgery



NCT
NATIONALES CENTRUM
FÜR TUMORERKRANKUNGEN
PARTNERSTANDORT DRESDEN
UNIVERSITÄTS KREBSCENTRUM UCC

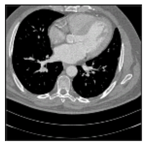
getragen von:
Deutsches Krebsforschungszentrum
Universitätsklinikum Carl Gustav Carus Dresden
Medizinische Fakultät Carl Gustav Carus, TU Dresden
Helmholtz-Zentrum Dresden-Rossendorf

Lecture 7

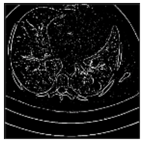
Basics of Computer Vision – Part 2

Process chain computer-assisted surgery

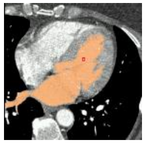
Image acquisition



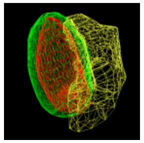
Pre-processing



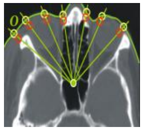
Segmentation



Modeling



Registration



Visualization/ Assistance

Interaction and Feedback

- https://pingo.coactum.de -> 392473

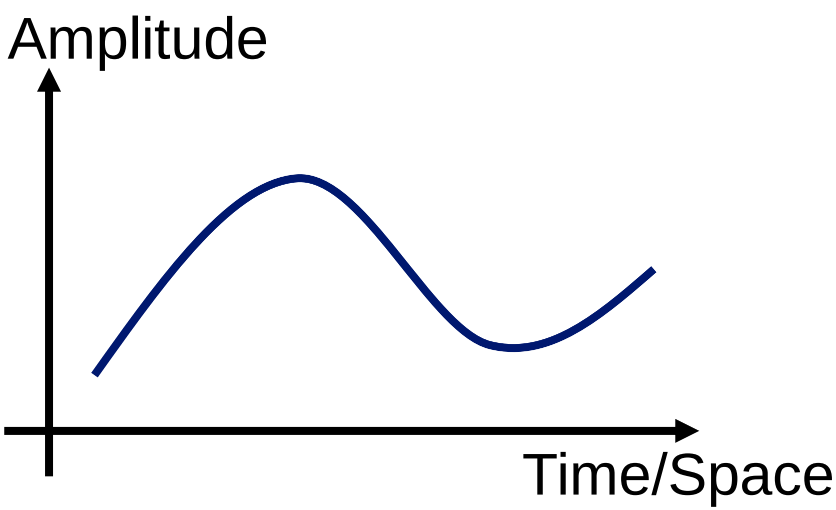
Image representation

- Context:

Signal capture → Processing (Amplification) → Digitalization

analog → digital

- Digitalization: Discretization + Quantization



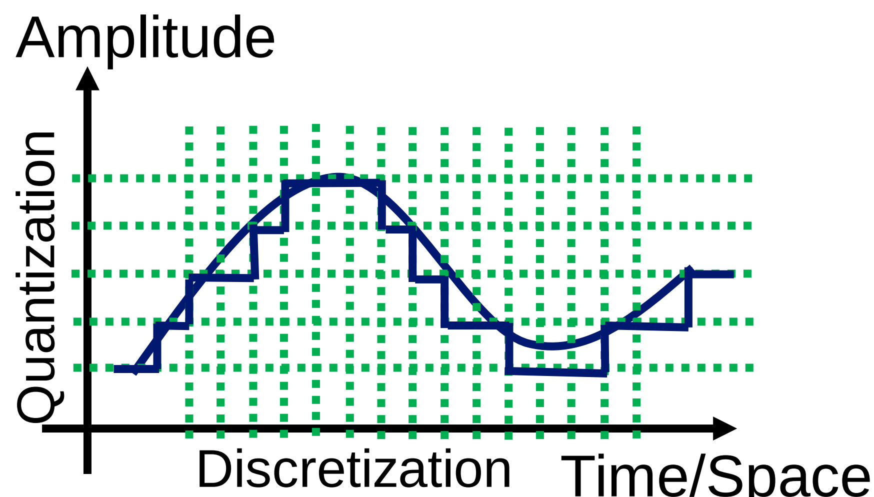
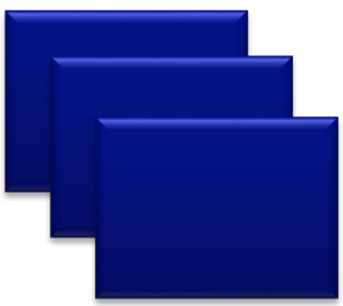


Image representation

2D grayscale image: Discrete function

$$\mathrm{Img}\ :\ [0..n]\times[0..m] \rightarrow [0..q]$$
$$(x,\ y) \mapsto \mathrm{G}(x,\ y) = g$$



Rows \ Column	0			y		m-1
0	80	0	100	70	0	80
	0	20	30	20	110	30
	25	79	136	100	30	0
x	20	20	30	50	90	85
	22	46	0	5	36	87
n-1	112	0	44	50	50	0

4-Neighbors

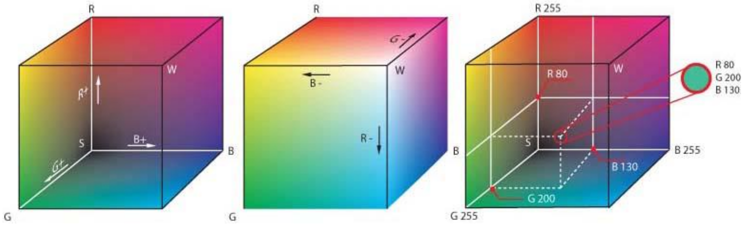
8-Neighbors

Image point (x, y)

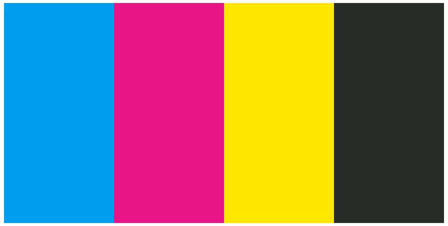
NCT

Image representation

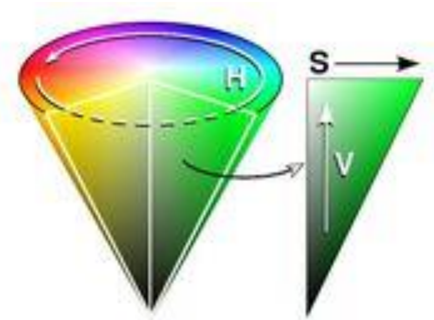
- Color image: different models for different applications
  - B/W: Grayscale
  - RGB-model: specific for screens (Phosphor-crystals), very common
  - CMYK-model: Color printer (subtractive color mix)
  - YCbCr: Breakdown into luminescence Y und two color components Cb, Cr
  - HSV (Hue, Saturation, Value): specific for color segmentation



RGB



CMYK



HSV

Quelle: Wikipedia

Image representation

- RGB-model:

$$\text{Img} : [0..n] \times [0..m] \rightarrow [0..R] \times [0..G] \times [0..B]$$
$$(x, y) \mapsto \text{G}(x, y) = (r, g, b)$$

3 components: red, green and blue

usually 256 x 256 x 256 nuances = 16,8 Mio. colors

Image representation

Conversion between different models

- RGB to B/W: $I = \frac{R+G+B}{3}$
- RGB to HSV: $V = MAX(R,G,B)$ or $V = \frac{R+G+B}{3}$

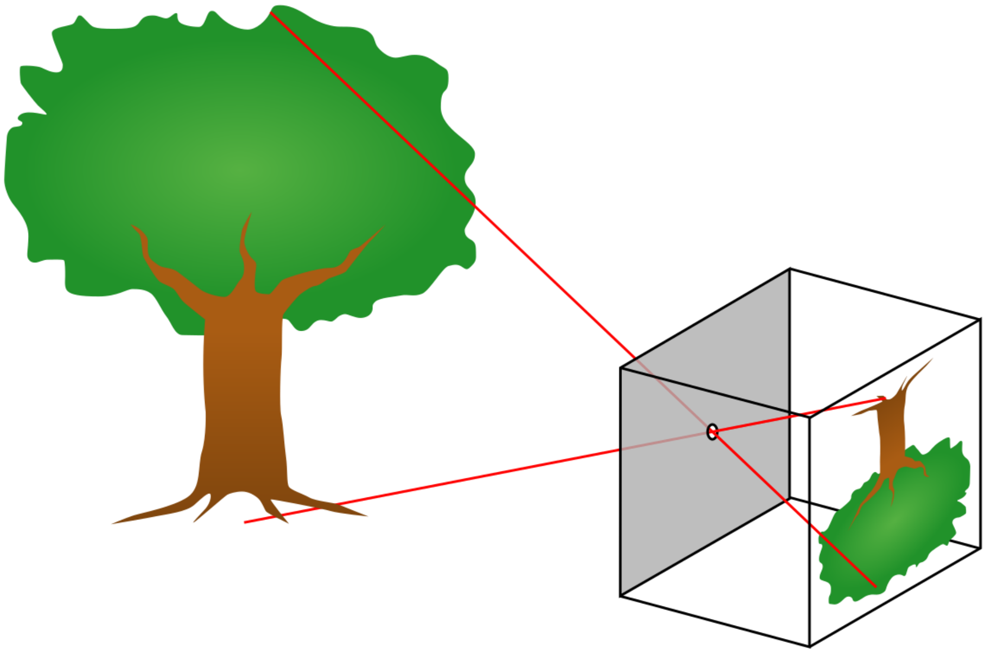
$$S = 1 - \frac{MIN(R,G,B)}{V} \text{ or } S = \begin{cases} \frac{3}{2}(R-V), & B+R \geq 2G \\ \frac{3}{2}(V-B), & B+R < 2G \end{cases}$$

$$H = \begin{cases} 60° \left(0 + \dfrac{G-B}{MAX(R,G,B) - MIN(R,G,B)}\right), & R = MAX(R,G,B) \\ 60° \left(2 + \dfrac{B-R}{MAX(R,G,B) - MIN(R,G,B)}\right), & G = MAX(R,G,B) \\ 60° \left(4 + \dfrac{R-G}{MAX(R,G,B) - MIN(R,G,B)}\right), & B = MAX(R,G,B) \end{cases}$$

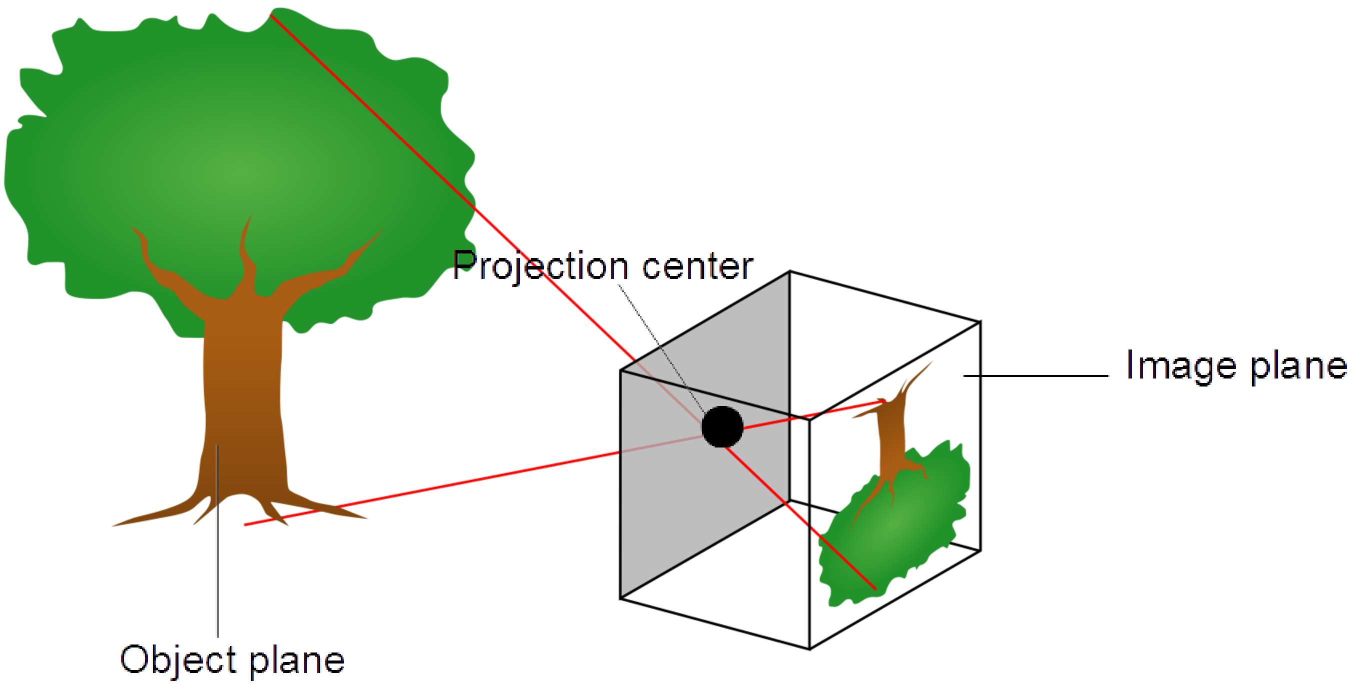
$$H < 0° \Rightarrow H = H + 360°$$

Pinhole camera model

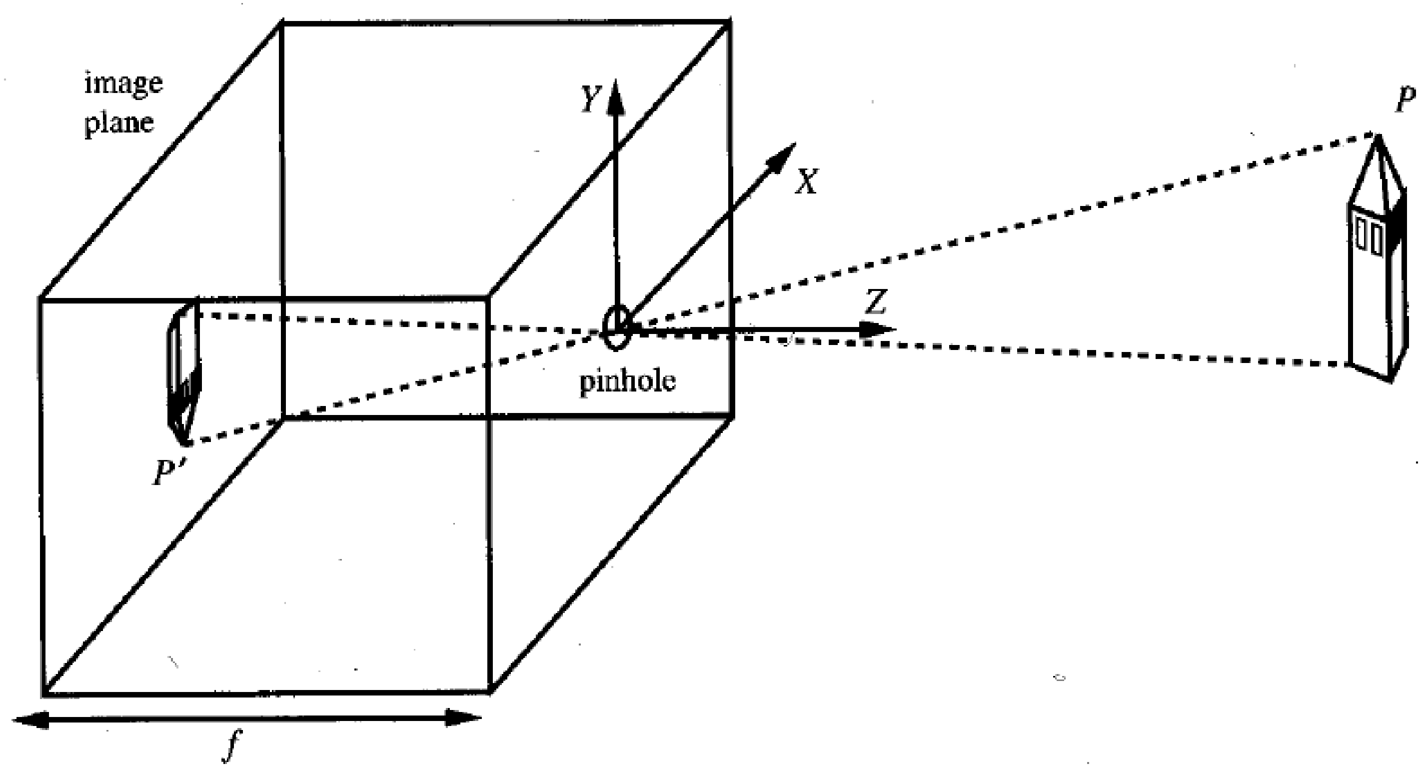
- Describes mathematical relationship between coordinates in 3D and their projection on a 2D plane
  - Simple model
  - No modeling of lens
  - No world coordinate system



Pinhole camera model



Pinhole camera model



Projection of 3D point $P = (x, y, z)$ onto an image point $p = (u, v, w)$ with focal length f:

$$-\frac{u}{f} = \frac{x}{z} \qquad -\frac{v}{f} = \frac{y}{z} \qquad w = -f$$

$$\mathrm{x} = -\frac{uz}{f} \qquad \mathrm{y} = -\frac{vz}{f}$$

$$p = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} u \\ v \\ -f \end{pmatrix} = -\frac{f}{z}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = -\frac{f}{z}P$$

Back projection

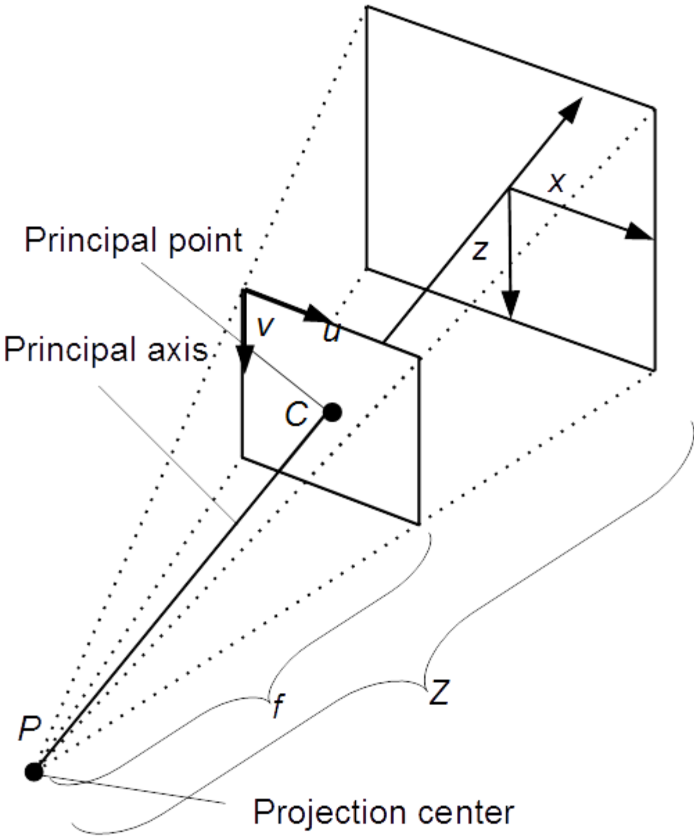
Perspective projection

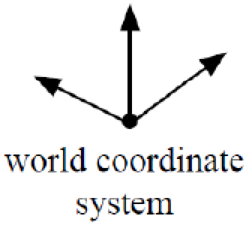
Extended camera model

- Pinhole camera model strongly simplifies reality
  - Image origin identical with principle point
  - Pixels are square
  - No lens distortion
  - No world coordinate system
- Therefore in practice, extensions are used
- A few definitions:
  - Principal axis: Line through the projection center, orthogonal to image plane
  - Principal point C: Point of intersection of principal axis and image plane
  - Image coordinate system: 2D, unit [pixels]. Origin in the upper left corner u-axis to the right, v-axis to the bottom
  - Camera coordinate system: 3D, unit [mm]. Origin in the projection center, axis parallel to those of the image coordinate system (x to u, y to v and z away from the projection center)
  - World coordinate system: 3D, unit [mm]. Origin arbitrary, anywhere in space possible

Extended camera model

- Common variant: Pinhole camera model in positive position





$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = -\frac{f}{z}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \Rightarrow \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \frac{f}{z}\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

- Projection center behind image plane
- No mirroring (negative signs are omitted)

Extended camera model

- Intrinsic parameters
  - Focal length *f*
    - In practice, the conversion from [mm] to [pixel] is incorporated into the focal length
    - As we assume non-quadratic but rectangular pixels, there is a parameter for each direction: f_x, f_y
    - Since product of actual focal length [mm] and conversion factor [pixel/mm] they have the unit [pixel]
  - Principal point $c(c_x, c_y)$
    - Point of intersection of principle axis and image plane
    - Has to be taken into consideration when moving origin of image plane

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \frac{f}{z} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{z} \begin{pmatrix} f_x \cdot x \\ f_y \cdot y \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}$$

Which of the following coordinate systems is NOT in 3D?

- A: Image coordinate system
- B: Camera coordinate system
- C: World coordinate system
- D: None of the above

Homogenous coordinates

$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{z}\begin{pmatrix} f_x \cdot x \\ f_y \cdot y \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}$ can be expressed better

- Homogenous coordinates
  - Add a new dimension with value of 1 to vector, e.g.: $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \rightarrow \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad \begin{pmatrix} u \\ v \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$
- Allows expression of certain operations with matrix multiplication

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{z}\begin{pmatrix} f_x \cdot x \\ f_y \cdot y \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} \rightarrow \begin{pmatrix} u \cdot w \\ v \cdot w \\ w \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

- Afterwards, normalize so the “additional” dimension becomes 1:

$$\frac{1}{w} \cdot \begin{pmatrix} u \cdot w \\ v \cdot w \\ w \end{pmatrix} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \end{pmatrix}$$

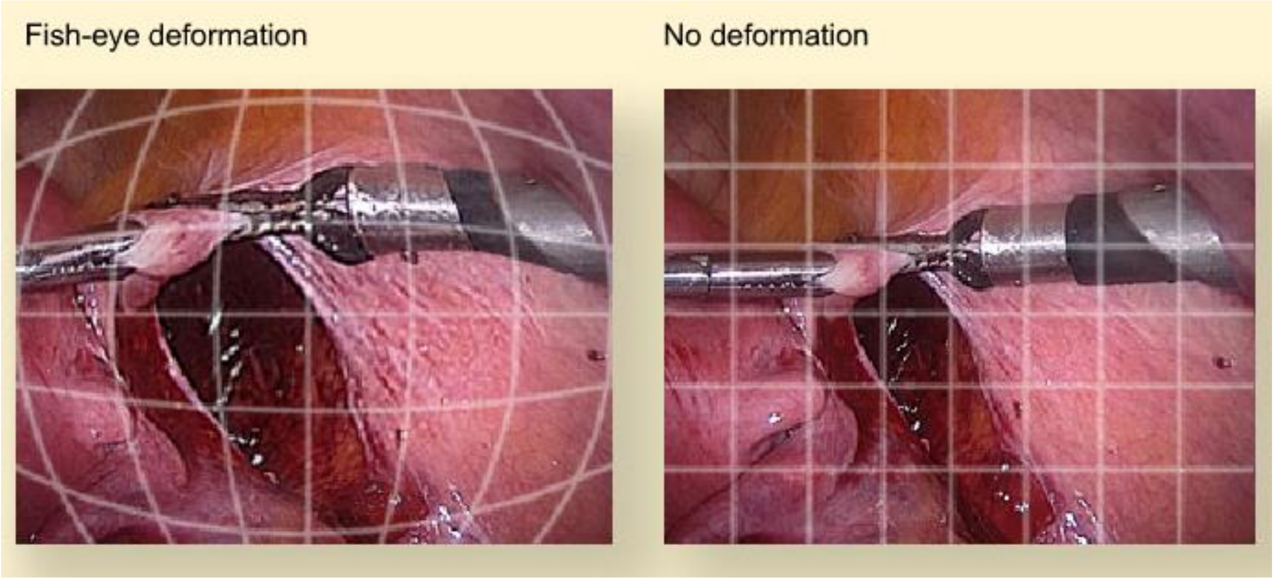
Extended camera model

- Intrinsic parameters
  - Focal length *f*
    - In practice, the conversion from [mm] to [pixel] is incorporated into the focal length
    - As we assume non-quadratic but rectangular pixels, there is a parameter for each direction: $\boldsymbol{f_x}, \boldsymbol{f_y}$
    - Since product of actual focal length [mm] and conversion factor [pixel/mm] they have the unit [pixel]
  - Principal point $\boldsymbol{c}(\boldsymbol{c_x}, \boldsymbol{c_y})$
    - Point of intersection of principal axis and image plane
    - Has to be taken into consideration when moving origin of image plane
  - Contained in the camera matrix K

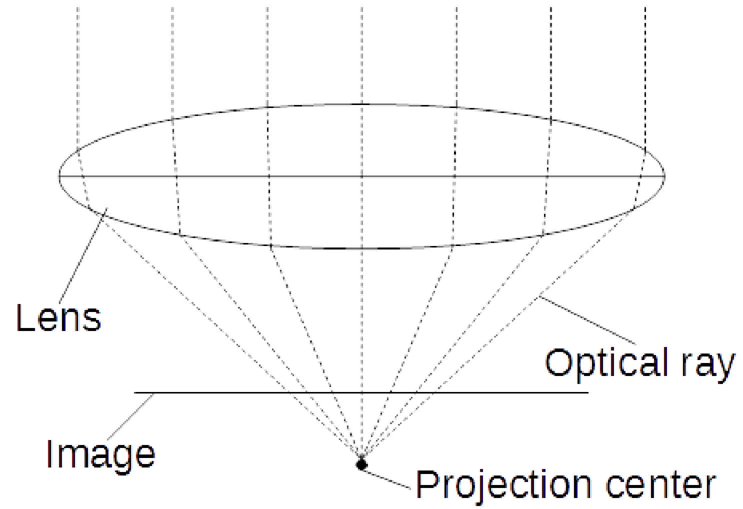
$$K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

Lens distortion

- Wide angle lenses (often encountered in endoscope) can significantly distort the image
  - Radial distortion
    - Symmetric from principle point
  - Other types of distortion are possible



Extended camera model



- Intrinsic parameters
  - Radial lens distortion
    - Project points "back onto lens":

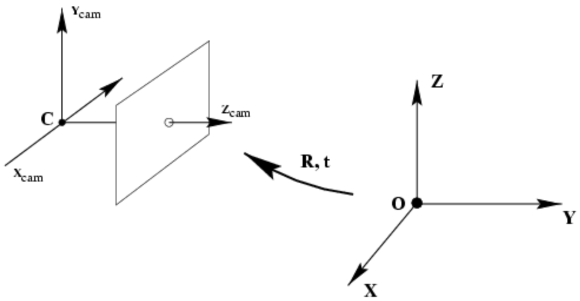
$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} := \begin{pmatrix} \frac{u - c_x}{f_x} \\ \frac{v - c_y}{f_y} \end{pmatrix}$$

    - Distortion is proportional to distance from principle point $r := \sqrt{x_n^2 + y_n^2}$
    - The distorted coordinate can then be computed from a distortion model
    - Often approximated using first two or three terms of a Taylor polynomial

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = (1 + \boldsymbol{d_1} r^2 + \boldsymbol{d_2} r^4 + \cdots) \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

    - Images can be "undistorted" by using a lookup table and interpolation

Extended camera model



- Extrinsic parameters
  - Offset camera to world coordinate, e.g. when using multiple cameras or a robot
  - Transformation from world to camera coordinate system
  - Defined through a coordinate transform consisting of
    - Rotation matrix ***R***

$$R = R_z(\gamma)R_y(\beta)R_x(\alpha)$$

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix} \quad R_y(\beta) = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix}$$

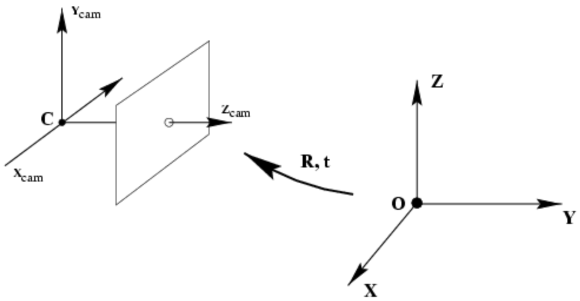
$$R_z(\gamma) = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

    - Translation vector ***t***

$$t = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

NCT

Extended camera model



- Extrinsic parameters
  - Offset camera to world coordinate, e.g. when using multiple cameras or a robot
  - Transformation from world to camera coordinate system
  - Defined through a coordinate transform consisting of
    - Rotation matrix $\boldsymbol{R}$
    - Translation vector $\boldsymbol{t}$

$$x_c = R \cdot x_w + t$$

  - In homogenous coordinates:

$$\begin{pmatrix} x_c \\ 1 \end{pmatrix} = \left(\begin{array}{ccc|c} & R & & t \\ 0 & 0 & 0 & 1 \end{array}\right) \begin{pmatrix} x_w \\ 1 \end{pmatrix}$$

- Projection matrix $\boldsymbol{P}$: 3x4 matrix containing both intrinsic and extrinsic parameters

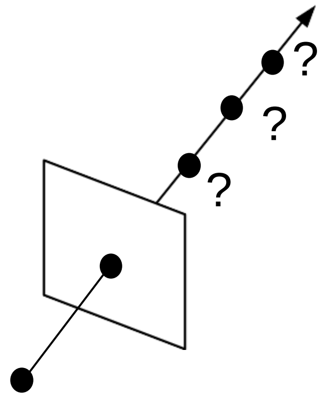
$$\begin{pmatrix} u \cdot w \\ v \cdot w \\ w \end{pmatrix} = P \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \qquad P = (KR|Kt)$$

Which of the following is NOT contained in the projection matrix?

- A: Principal point
- B: Focal length
- C: Lens distortion parameters
- D: Translation to world coordinate system

Camera calibration

- Process of determining intrinsic and extrinsic parameters
- Intrinsic parameters should remain constant for different setups unless zoom or focus of a camera changes
- Extrinsic parameters are dependent on the selection of world coordinate system and change depending on setup
- Once calibrated, a function f is know that maps points in world coordinate system onto the image coordinate system $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$
- f is defined through the projection matrix P and normalizing of the homogenous coordinates
- The inverse function maps a point of the image coordinate system onto a straight line in world coordinate system that runs through the projection center

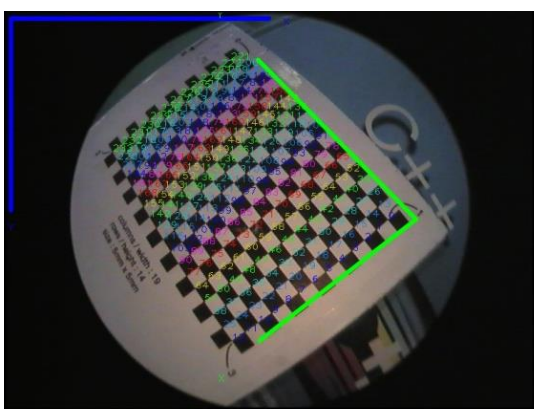


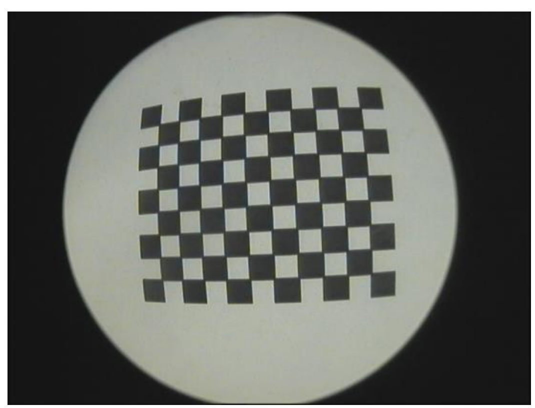
Camera calibration

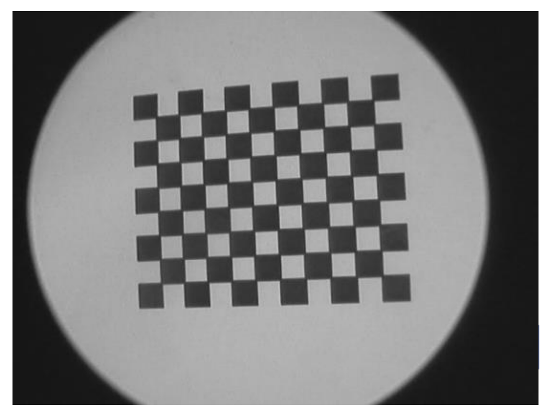
Wanted:

P is a 3×4-Matrix => 12 unknown variables

- Process *calibration*:
  1. Locating a number of 3D/2D point correspondences
  2. 3D points are known from usage of an appropriate calibration object or pattern
  3. 2D points are located through computer vision methods
  4. Estimation of P
  5. Estimation of distortion parameters from backprojection error
  6. Undistort 2D points and repeat from 4.







Direct Linear Transformation

- Standard method for computation of projection matrix P is the Direct Linear Transformation (DLT)

$$\begin{pmatrix} x \cdot w \\ y \cdot w \\ w \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad P = \left(K\,R \mid K\,\mathbf{t}\right) = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{pmatrix}$$

$$\Rightarrow \quad x = \frac{p_1 X + p_2 Y + p_3 Z + p_4}{p_9 X + p_{10} Y + p_{11} Z + p_{12}}$$

One parameter can be normalized. Usually $p_{12} = 1$.

$$y = \frac{p_5 X + p_6 Y + p_7 Z + p_8}{p_9 X + p_{10} Y + p_{11} Z + p_{12}}$$

Direct Linear Transformation

$$\Rightarrow \begin{aligned} p_1 X + p_2 Y + p_3 Z + p_4 &= x p_9 X + x p_{10} Y + x p_{11} Z + x \\ p_5 X + p_6 Y + p_7 Z + p_8 &= y p_9 X + y p_{10} Y + y p_{11} Z + y \end{aligned}$$

- Formulation as a linear system $A\boldsymbol{x} = \boldsymbol{b}$ with $n \geq 6$ point correspondences

$$A = \begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 X_1 & -x_1 Y_1 & -x_1 Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -y_1 X_1 & -y_1 Y_1 & -y_1 Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -x_n X_n & -x_n Y_n & -x_n Z_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -y_n X_n & -y_n Y_n & -y_n Z_n \end{pmatrix} \quad \boldsymbol{x} = \begin{pmatrix} p_1 \\ \vdots \\ p_{11} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} x_1 \\ y_1 \\ \vdots \\ x_n \\ y_n \end{pmatrix}$$

3D cameras

Until now: 2D vision

- Spatial information only secondary (experience)

Goal: Computer-assistance via 3D model

- Navigation
- Augmented reality

?

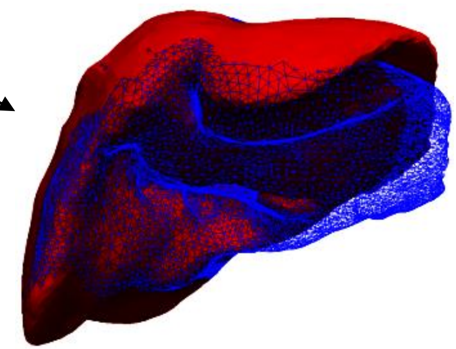
3D endoscopy

- Goal: Navigation, augmented reality
  - Create intraoperative model with the endoscope
  - Registration with pre-operative model

Intraoperative

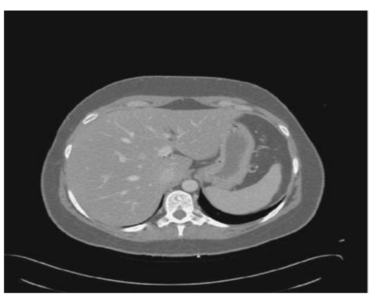
Endoscope image

3D model

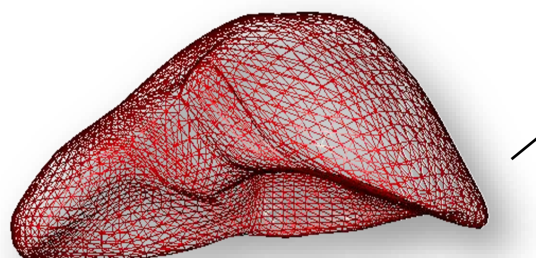


Registration

Pre-operative



CT image



Volume model

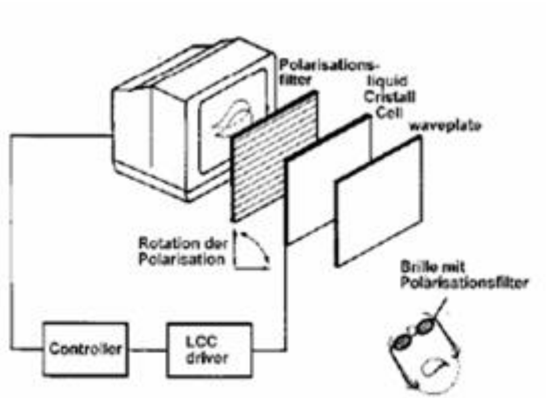
3D endoscopy

- **Applications without 3D reconstruction**
  - E.g. Shutter glasses
  - Accessories required (Glasses, ...)

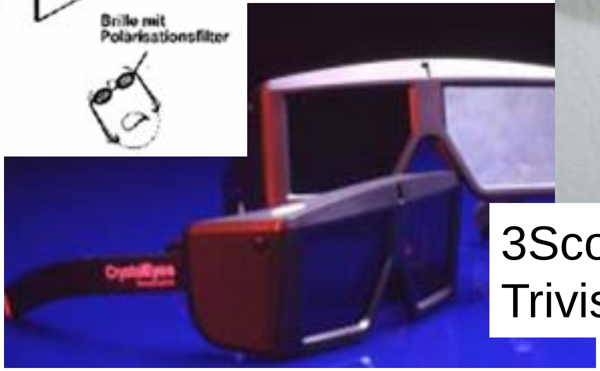
- **Applications with 3D reconstruction**
  - Active or passive methods:
    - Stereo endoscopy
    - Structure from Motion
    - Time-of-Flight
    - Structured Light
  - Different endoscope types

Methods without 3D reconstruction

- No computer-support
- Depth perception is a result from the natural stereo vision of the viewer



Shutter glasser



3Scope HMD,
Trivisio Prototyping GmbH

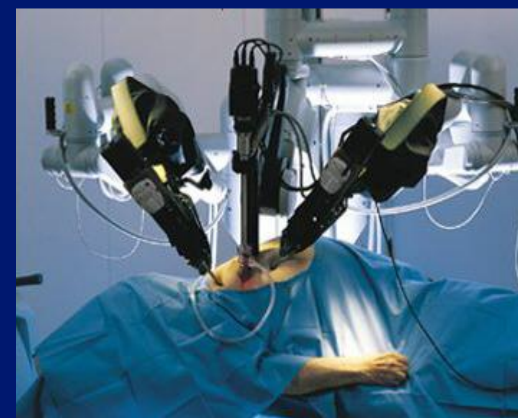
Quelle: Intuitive Surgical

da Vinci®
Surgical robot,
Intuitive Surgical, Inc.

DaVinci

Console

Manipulator

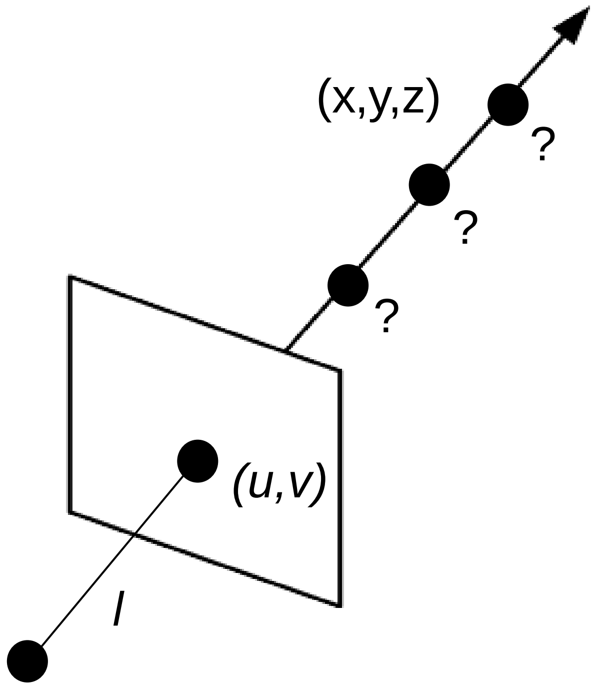


Quelle: Intuitive Surgical

DaVinci

2D to 3D

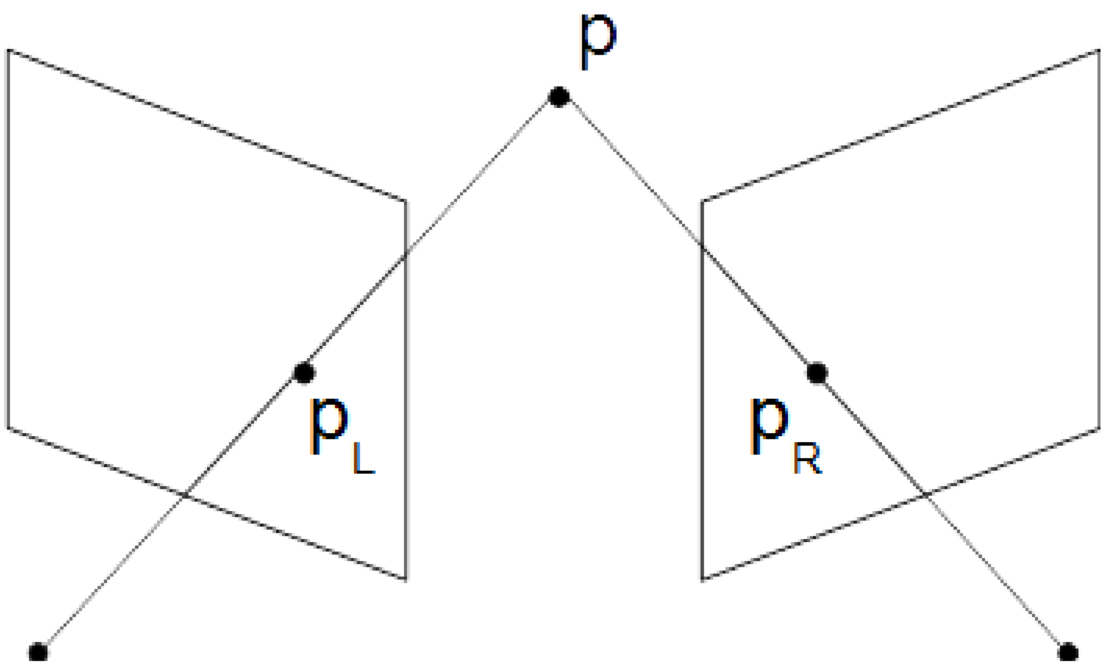
Given a 2D point, how do we reconstruct the original 3D point?



$$\begin{pmatrix} u \\ v \end{pmatrix} = R \cdot K \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + t \quad \Rightarrow \quad l: \lambda \cdot \left(R^{-1} K^{-1} \begin{pmatrix} u \\ v \end{pmatrix}\right) - R^{-1} t$$

λ variable describing position on line l

Stereo camera system - Triangulation



Given two calibrated cameras, each containing a projection (p_1, p_2) of point p, two lines can be computed:

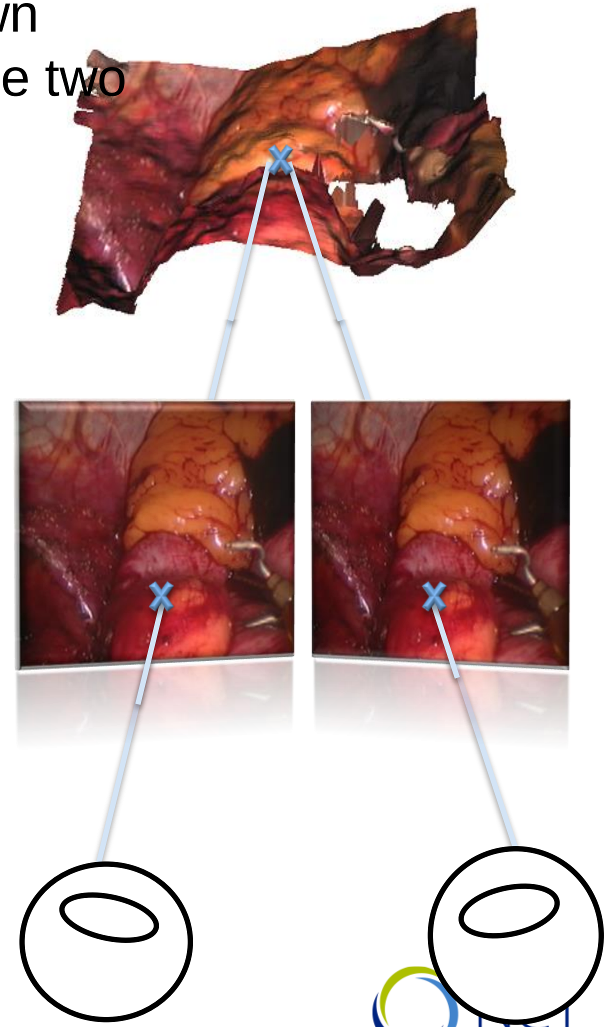
$$l_L(\lambda_L) = \lambda_L \cdot (R_L^{-1} K_L^{-1}) - R_L^{-1} t_L$$
$$l_R(\lambda_R) = \lambda_R \cdot (R_R^{-1} K_R^{-1}) - R_R^{-1} t_R$$

Solve for λ_L, λ_R so that $l_L(\lambda_L) = l_R(\lambda_R)$, reconstructing point p

Stereo endoscopy

Used endoscope: Stereo endoscope (two channels)

Reconstruction: Triangulation with known relationship between the two cameras



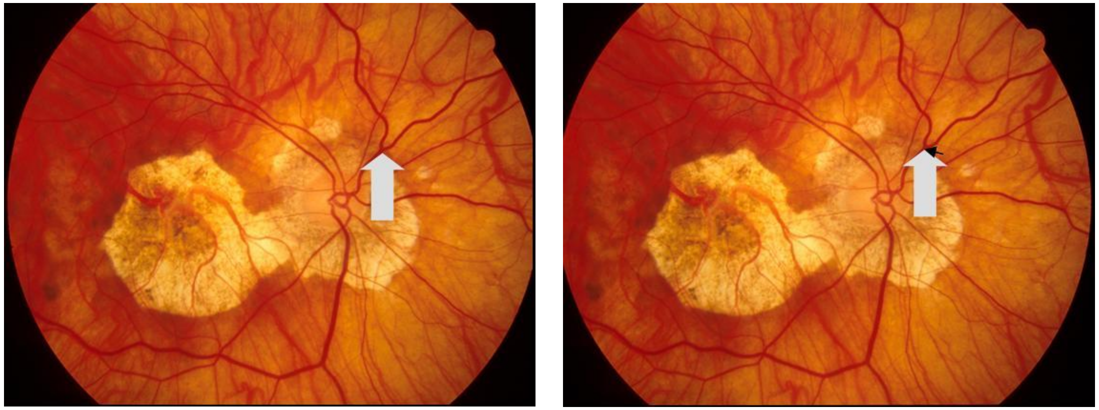
Pro: known stereoscopic basis

Cons: greater diameter
more expensive endoscope

Problem Stereo

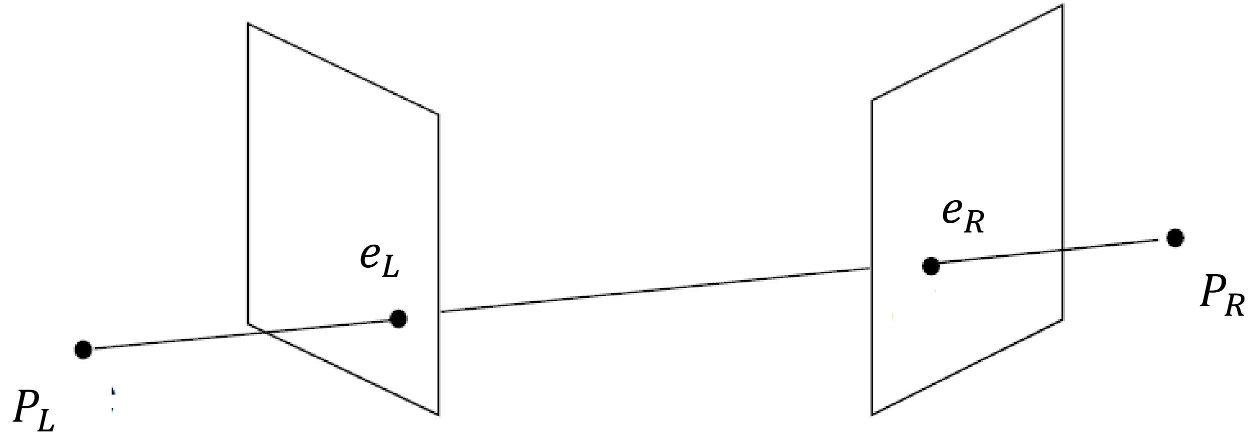
- Correspondence:

  Which point in the left image belongs to which point in the right image?



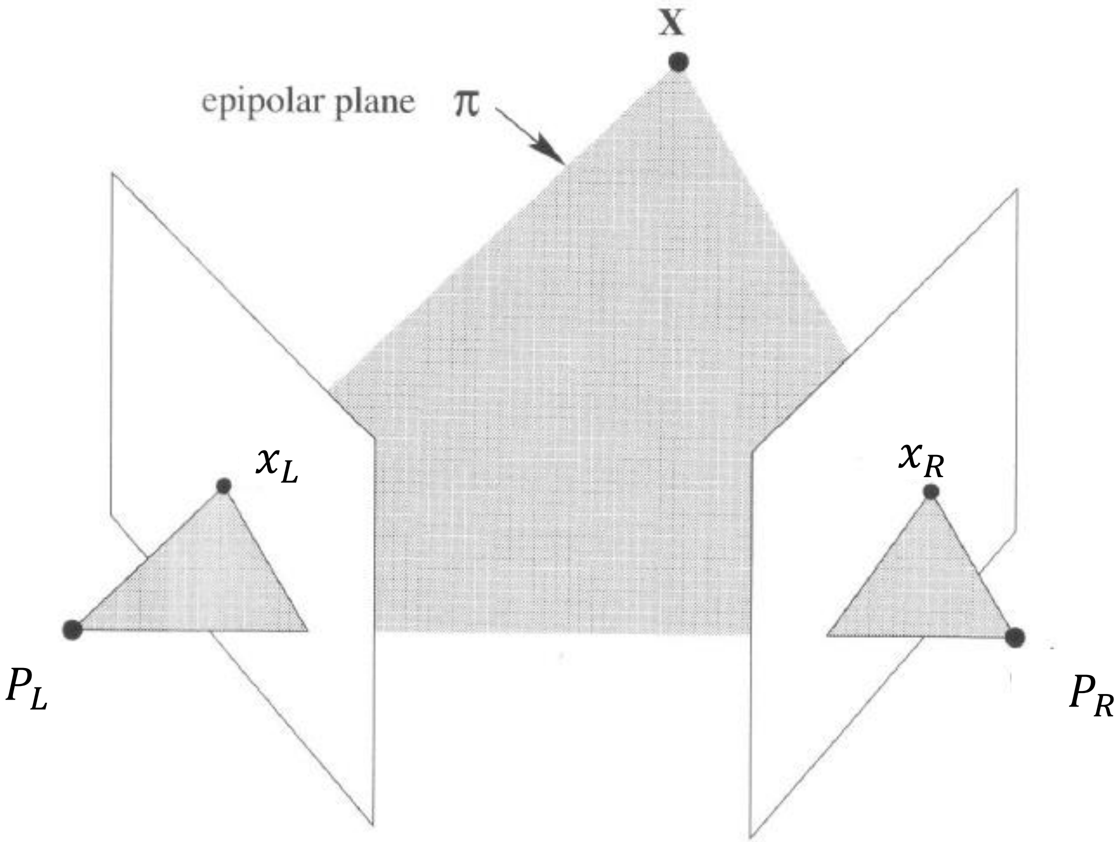
Epipolar geometry

- Relationship between two cameras is described through *Epipolar geometry*
- The points of intersection, e_L, e_R, of the line between the projection centers, P_L, P_R, are called *epipoles*



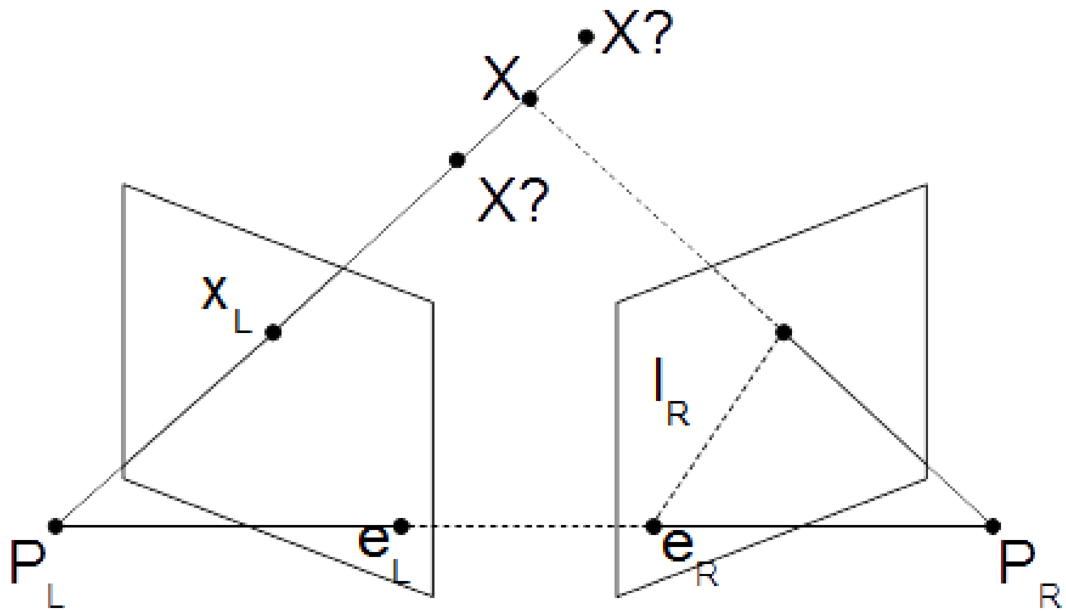
Epipolar geometry

- *Epipolar plane* $\pi(X)$:
  - Plane created through a 3D point X in the scene and the two projection centers P_L, P_R



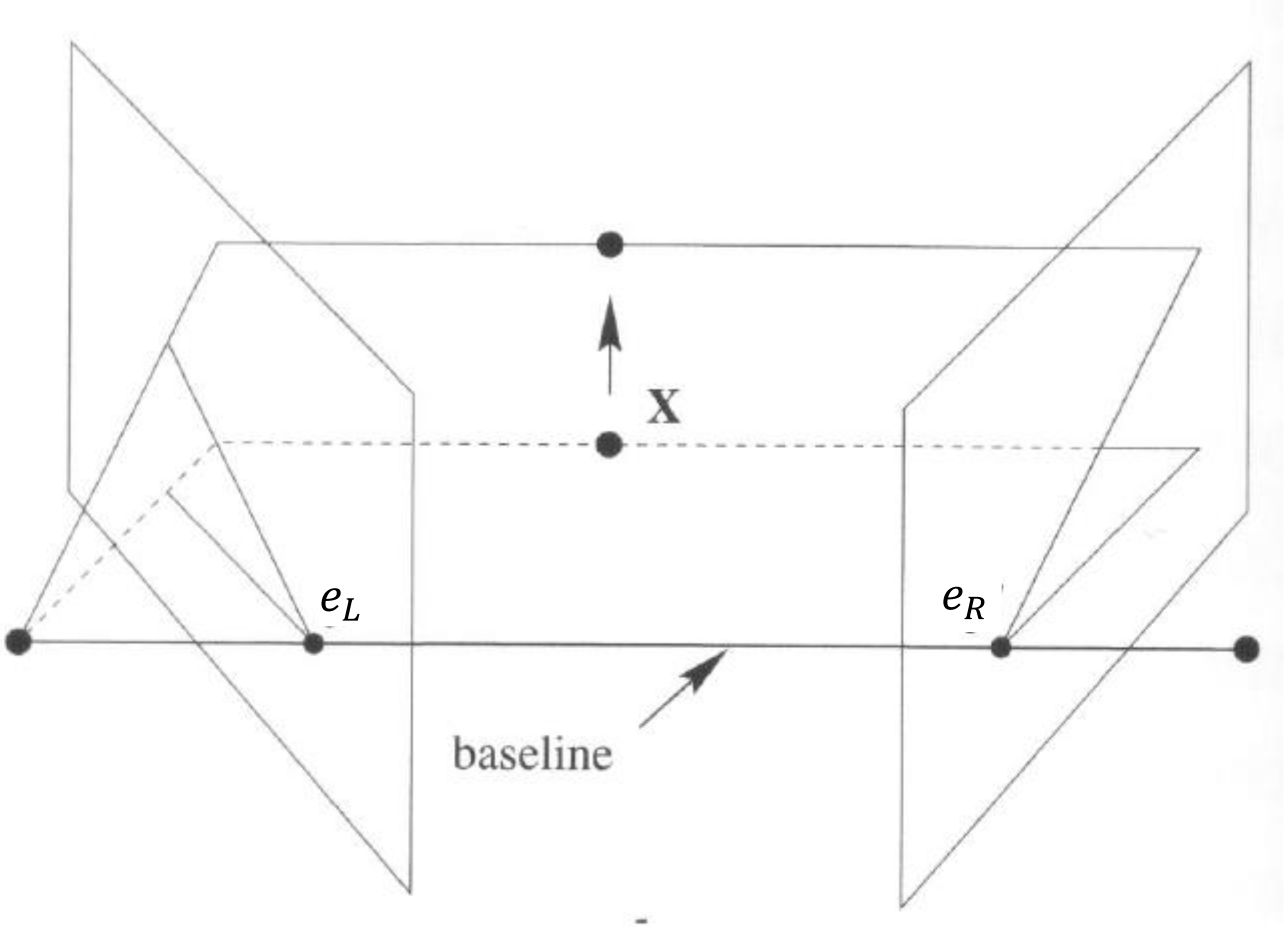
Epipolar geometry

- *Epipolar line* $l_R(x_L)$: line of intersection of $\pi(X)$ and image plane
  - All 3D points X that could be projected onto x_L in the left image, are mapped onto a line $l_R(x_L)$ in the right image



Epipolar geometry

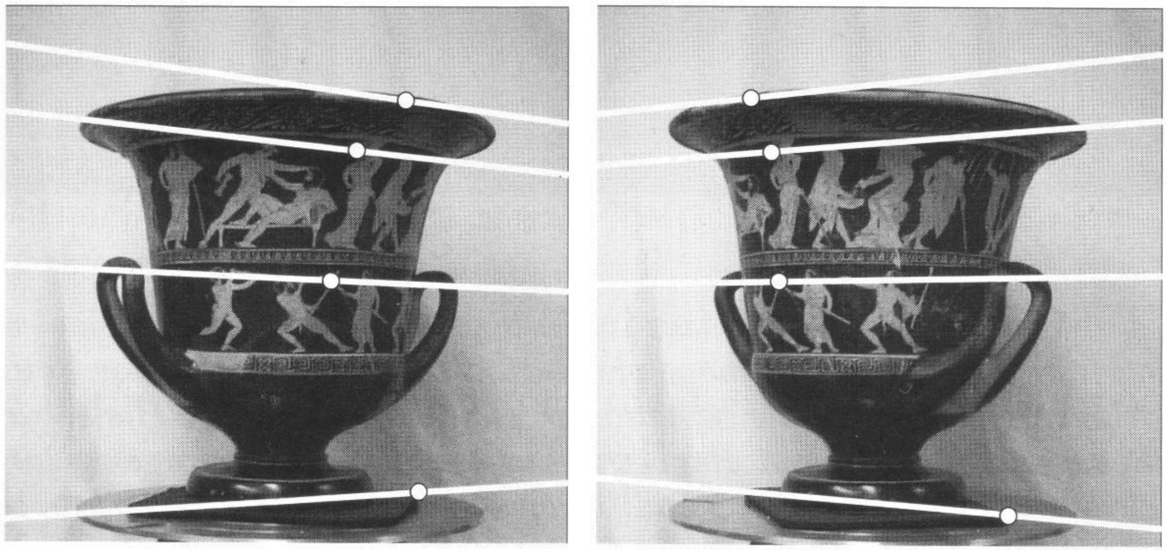
- All epipolar lines of a stereo camera system intersect in the epipoles e_L, e_R



Epipolar geometry

Usage:

- Reduction of the correspondence problem from two dimensions onto one dimension, as only points on an epipolar line have to be considered:
  - Higher robustness (less wrong correspondences)
  - Higher efficiency



Quelle: Multiple View Geometry

Fundamental matrix

- Mathematical description of the epipolar geometrie
- Properties of the Fundamental matrix F
  - 3x3 matrix
  - Has rank of 2
  - For all corresponding points x_L, x_R:
    - $x_L^T F x_R = 0$ (x_L, x_R are image points in homogenous form with $\mathrm{w} = 1$

Fundamental matrix

- Can be used to compute epipolar lines:
  - $l_L(x_R) = F^T x_R$
  - $l_R(x_L) = F x_L$
- For the epipoles:
  - $F^T e_R = 0$
  - $F e_L = 0$
- l_L (l_R analog) describes a 2D line in the following manner:
  - $l_L x = 0$ for all x (in homogenous form with $w = 1$) that lie on this line

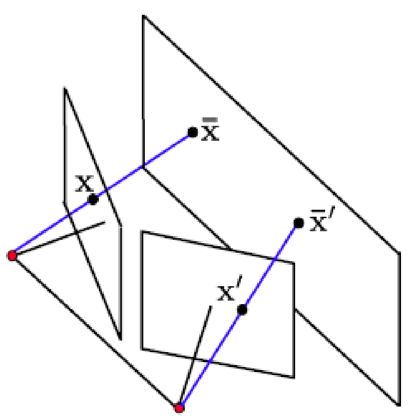
- Fundamental matrix can be compute in multiple ways
  - Using known image correspondences in the left and right images
  - When intrinsic and extrinsic parameters are known, directly using K_L and K_R and the Essential matrix E, which contains the extrinsic parameters

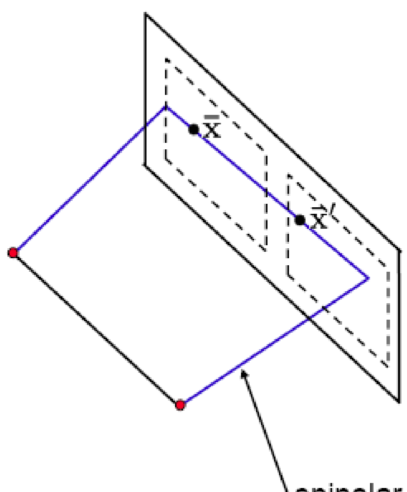
Fundamental matrix

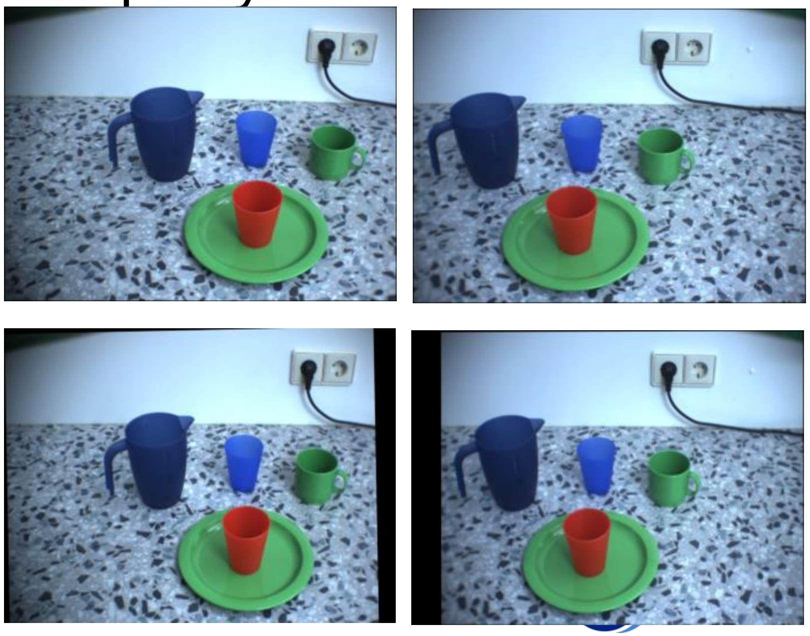
- Computation with know intrinsic and extrinsic parameters
  - Assumption extrinsic parameters
    - Left camera $(I|0)$ as transformation, i.e. identity
    - Right camera $(R|t)$ as transformation
  - Essential matrix E can be computed in the following manner:
    - $E = [t]_x \cdot R = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} R$
  - Fundamental matrix F can then be composed:
    - $F = K_R^{-T} \cdot E \cdot K_L^{-1}$
- If the Fundamental matrix has been computed from point correspondences and the intrinsic parameters are known, the Essential matrix can be computed:
  - $E = K_R^T \cdot F \cdot K_L$

Rectification

- If the epipolar geometry is know, images can be rectified:
  - Epipolar lines are parallel to the horizontal axis in a rectified image pair
  - Search for correspondences is restricted on a horizontal direction
  - Corresponding points share same y-coordinate, difference in x-coordinate is called disparity d







Rectification

- Rectified images have the benefit that optimized algorithms for correlation can be used to find correspondences

- Cons:
  - Interpolation necessary for rectifying images

    ⇨ Loss in quality
  - Depending on setup, images can be highly distorted

Rectification
Which statement is incorrect?

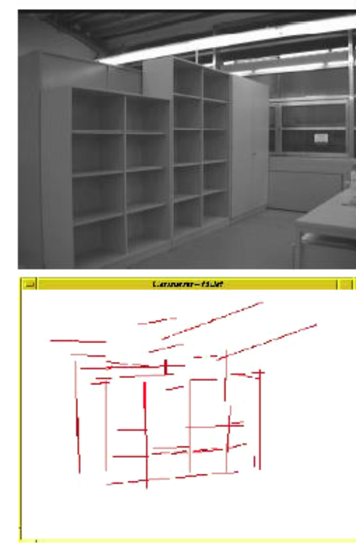
- A: only possible with calibration
- B: reduces image quality
- C: improves runtime
- D: reduces dimensionalty during correspondence analysis

Correspondence analysis

- Classification into 2 types:

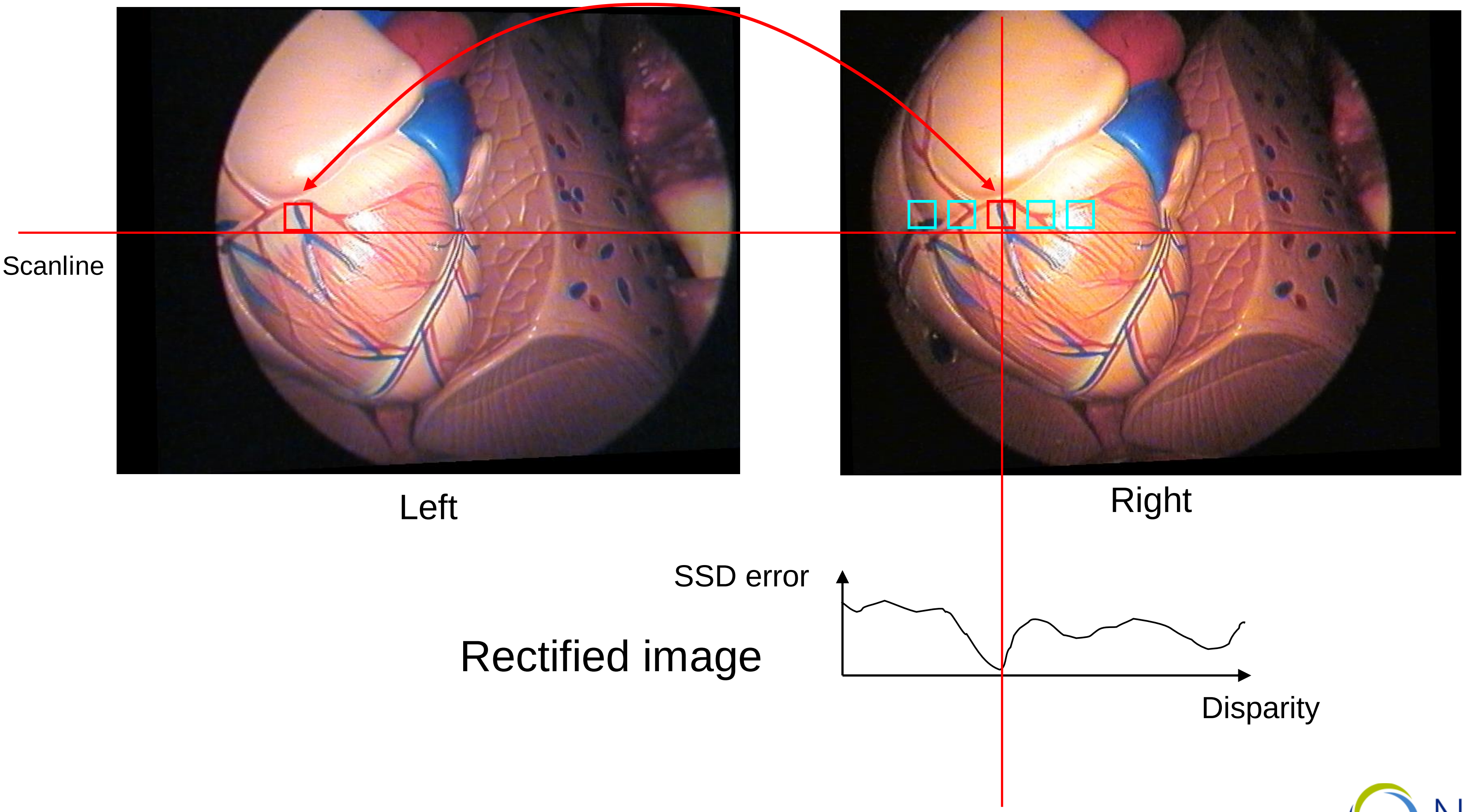
  - Correlation-based approaches
    - Correspondence for each pixel
    - Dense depth map
    - Application: Textured scene

  - Feature-based approaches:
    - Correspondence only for certain features
    - Sparse depth map
    - Application: Structured scene (Indoor)



Correspondences via correlation

- Corresponding elements are image windows
- Correlation as similarity measure



Rectified image

Normalization

- Images from different cameras can very due to varying lighting conditions

- Differing sensitivity of the sensors

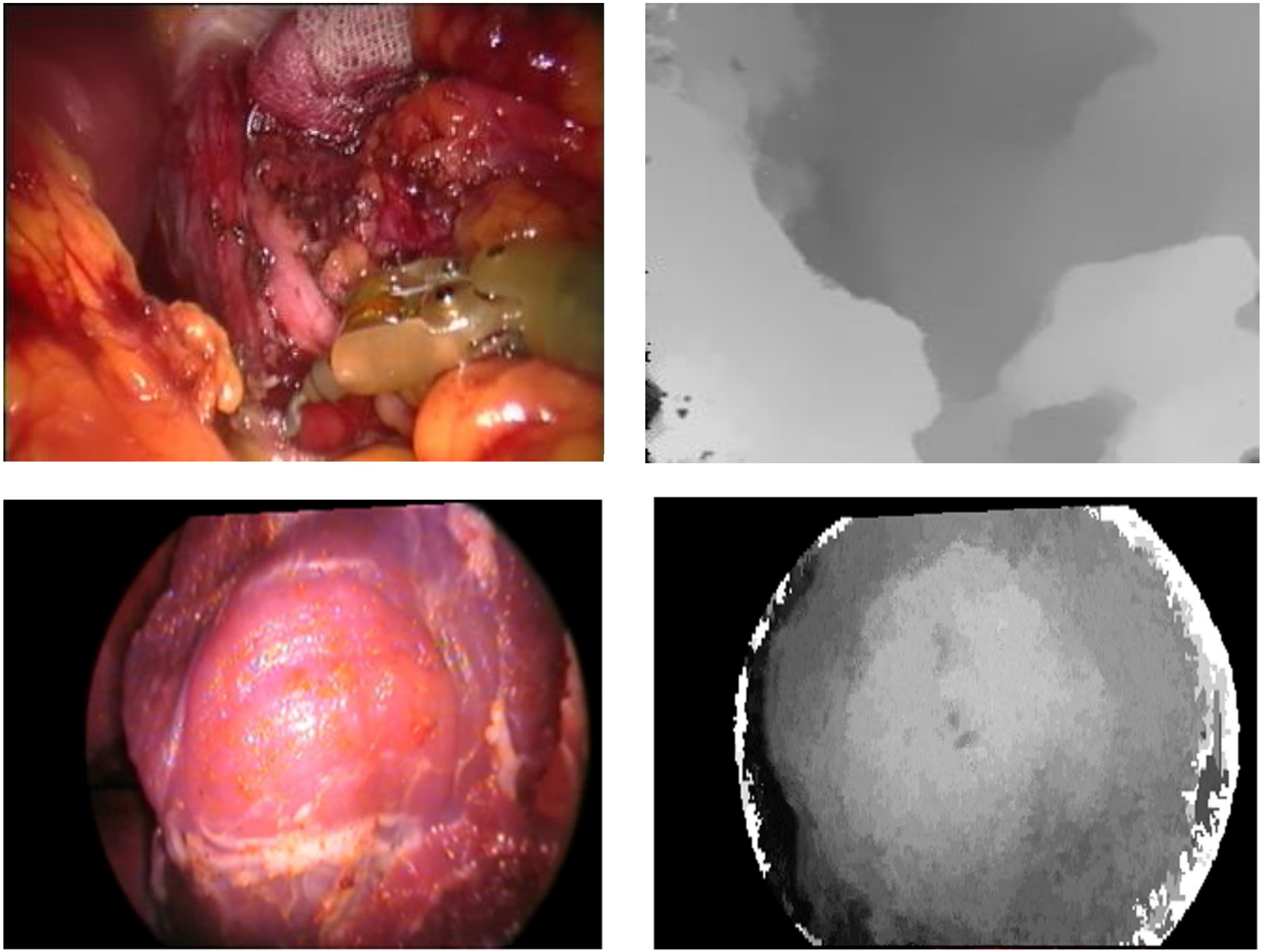
→ Normalization of pixels in each search window

$$\bar{I} = \frac{1}{|W_m(x,y)|} \sum_{(u,v) \in W_m(x,y)} I(u, v)$$ Average

$$\|I\|_{W_m(x,y)} = \sqrt{\sum_{(u,v) \in W_m(x,y)} [I(u, v)]^2}$$ Magnitude

$$\hat{I}(x, y) = \frac{I(x, y) - \bar{I}}{\|I - \bar{I}\|_{W_m(x,y)}}$$ Normalization

Disparity maps



Disparity map
A dark pixel in the disparity map implies:

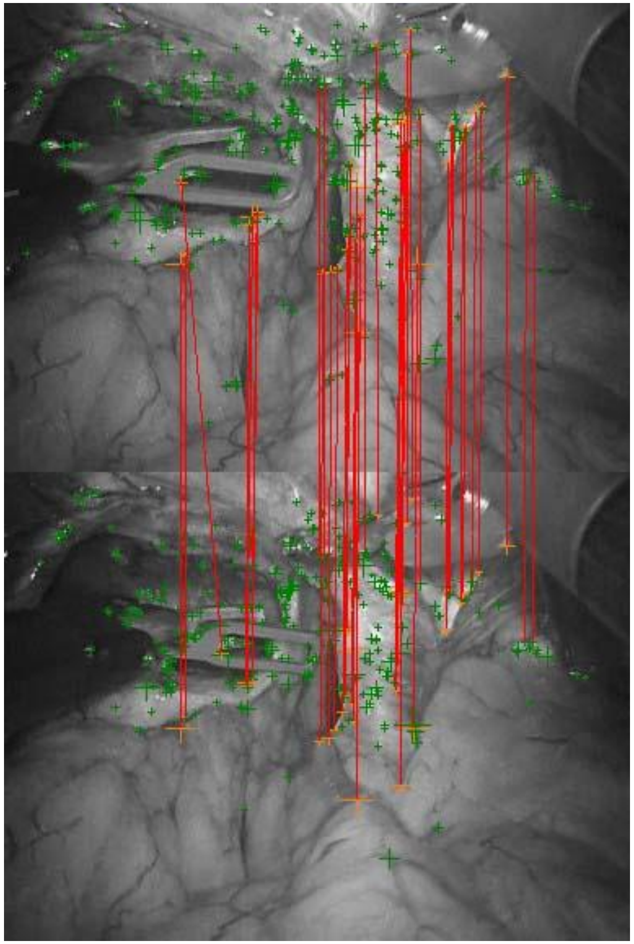
- A: Point close to the camera
- B: Point far from the camera
- C: Point not defined
- D: Point has many correspondences

Disparity map
A dark pixel in the disparity map implies:

- A: Point close to the camera
- **B: Point far from the camera**
- *C: Point not defined (when completely black)*
- D: Point has many correspondences

Correspondences via features

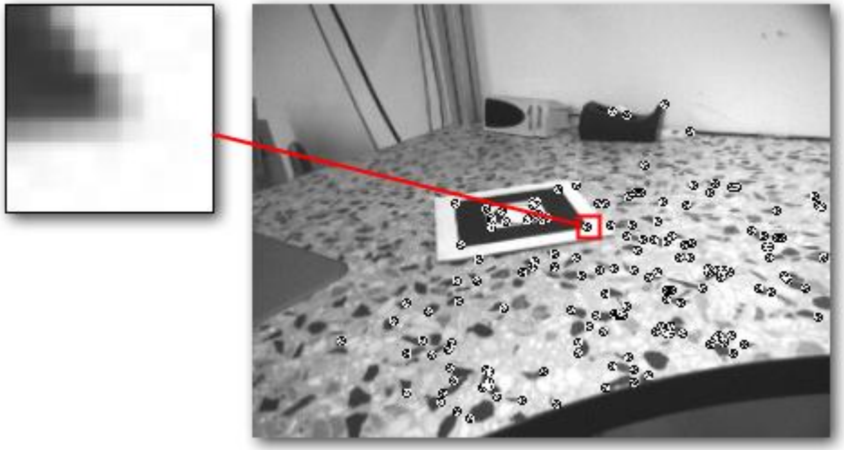
- Search restricted to few local point features
- Properties:
  - No occlusions, reproducible in different views, can be re-detected
  - Invariant against: Scaling, rotation, lighting
  - Neighborhood contains information
- Pixel feature: $(2n+1)\times(2n+1)$-Pixel-Block around Pixel p
- Computation is divided into detection and descriptor



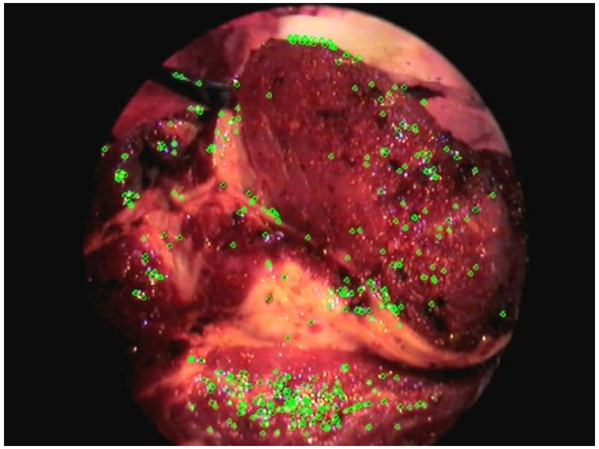
Correspondences via features

- Detection: Locating stable, transformation invariant key points
- Depending on the algorithm due points different in localization, scale and structure
- Descriptor: Robust, unique characterization of the local neighborhood, e.g. through gradient information or frequency spectrum
- Examples:

Harris Corner Detector



SIFT-Detector



Correspondences via features

Example: **Harris-Corner-Detector** :

If the Eigenvalues of the matrix

$$A = \begin{pmatrix} \left(\frac{\partial \mathrm{Img}(x,y)}{\partial x}\right)^2 & \frac{\partial \mathrm{Img}(x,y)}{\partial x}\frac{\partial \mathrm{Img}(x,y)}{\partial y} \\ \frac{\partial \mathrm{Img}(x,y)}{\partial x}\frac{\partial \mathrm{Img}(x,y)}{\partial y} & \left(\frac{\partial \mathrm{Img}(x,y)}{\partial y}\right)^2 \end{pmatrix}$$

are large, a small step in any direction will cause a large change in gray value.

Finding corner through looking for local maxima in:

$$R = \det A - k \cdot \mathrm{trace}(A)^2, k \approx 0.04$$

Problems/Comparison

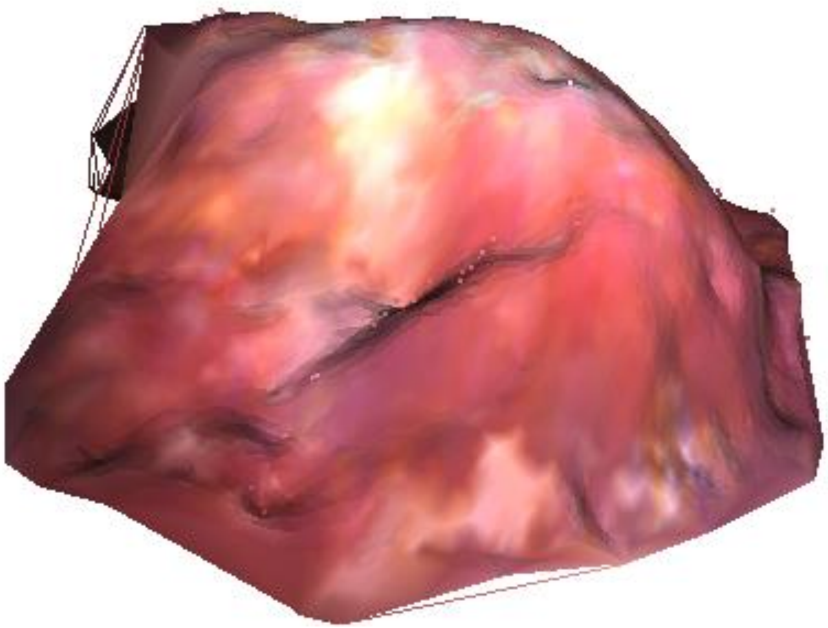
- Problems:
  - Occlusion
  - Limited field of view
  - Specularities, changes in light conditions
  - Surface structure: Sparse texture / repeating texture

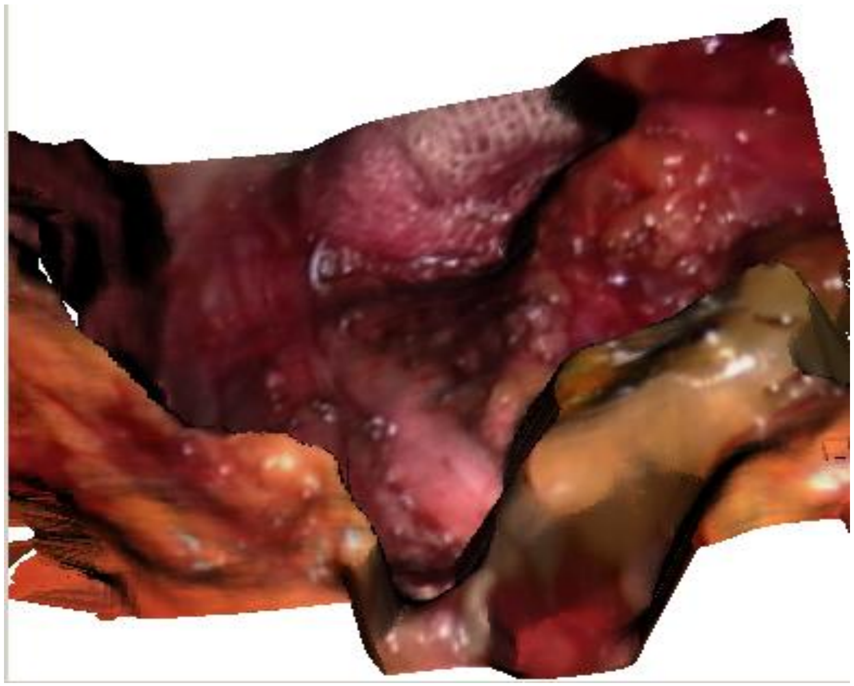
- Comparison:

Correlation-based	**Feature-based**
Dense depth map	Sparse depth map
Only for textured scenes	
Prone to errors from changes in direction	Prone to errors from wrong correspondences

Specific pros and cons have to be weighted for each use case

3D-model





Stereo endoscopy

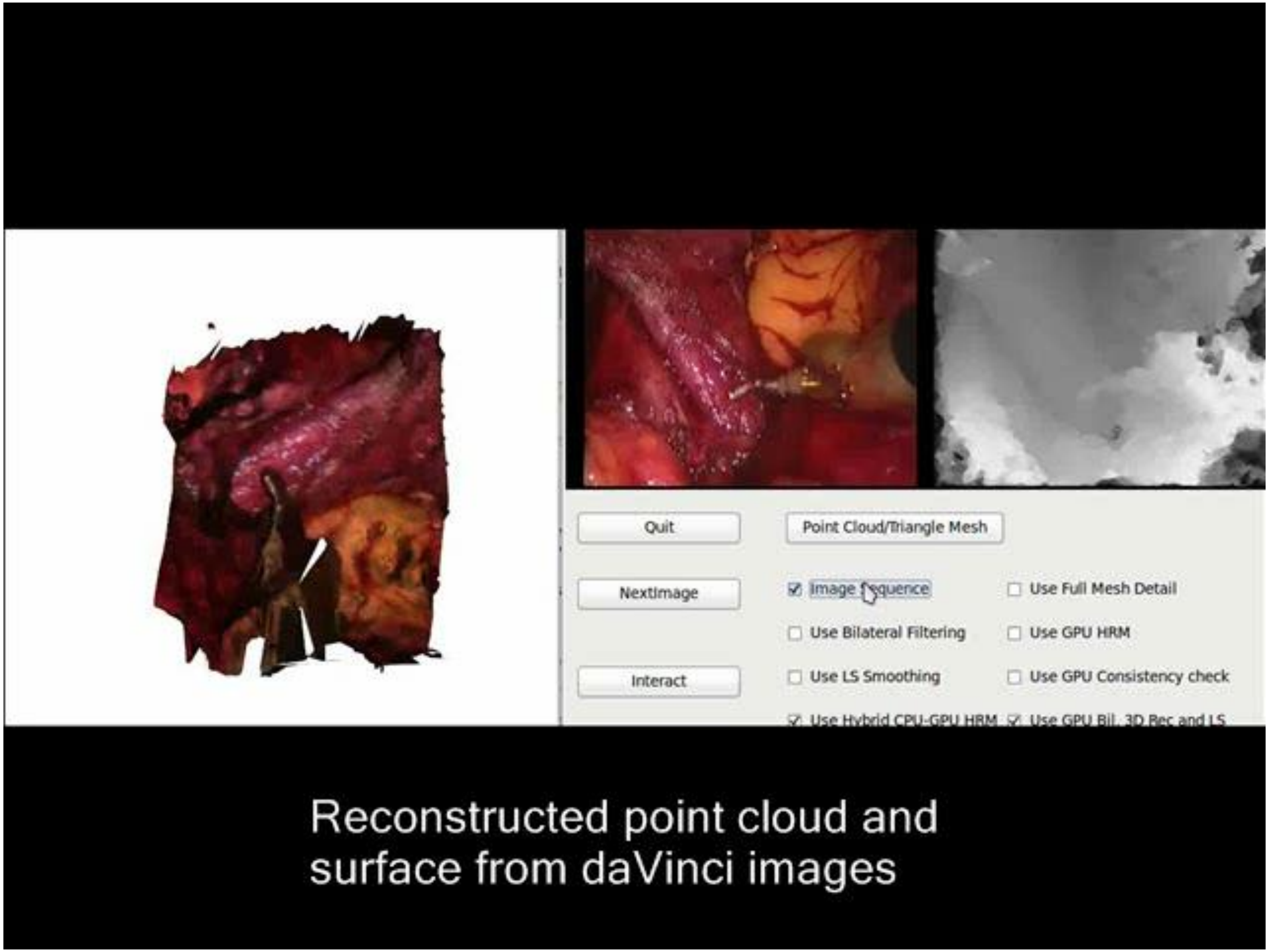
- Given:

  two calibrated cameras

  (Projection matrices given)
- 3D-Reconstruction:
  - Rectification
  - Correspondence search:

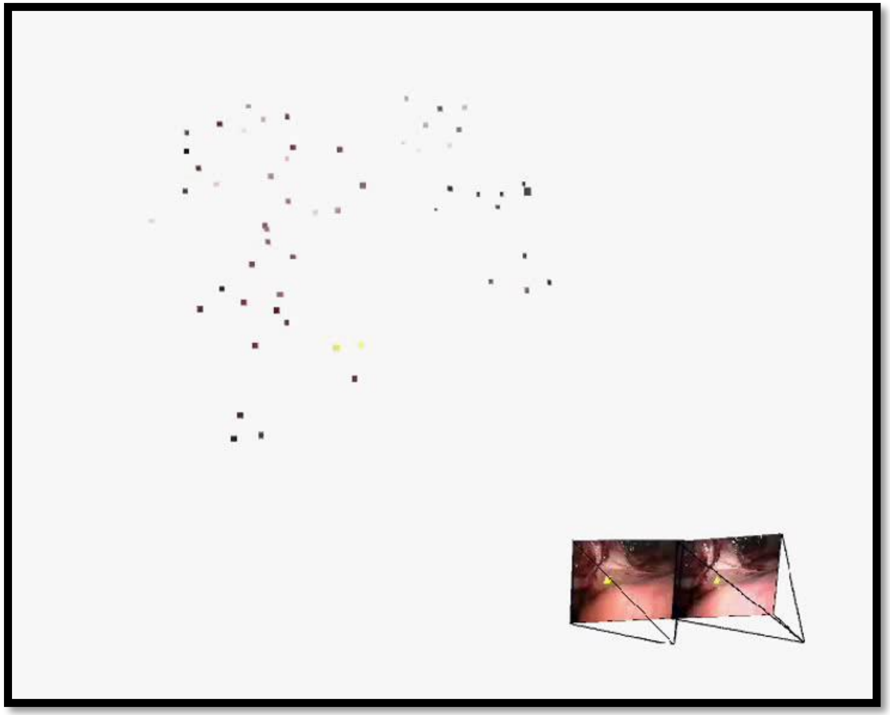
    Correlation-based or feature-based
    - Optional: Left/Right check
    - Optional: Detection of wrong correspondences
  - Triangulation
- Net generation
  - Texturizing of net

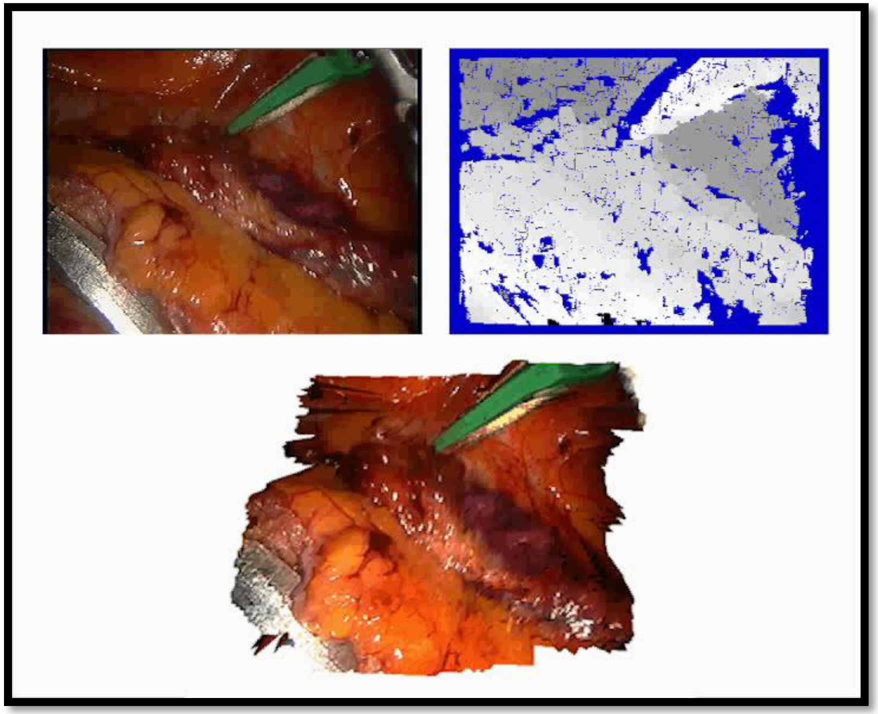
Stereo endoscopy



NCT

Stereo endoscopy





Stoyanov *et al.*: ”Real-time Stereo Reconstruction in Robotic Assisted Minimally Invasive Surgery”, MICCAI 2010

Stereo endoscopy - Evaluation

- Pros
  - High accuracy with good correspondences
  - No additional hardware (e.g. Tracking system, special light source) besides camera necessary
  - Dense depth map with correlation-based approaches

- Cons
  - Expensive hardware
  - Accuracy decreases with lower distances between cameras
  - Very accurate calibration necessary
  - For feature-based approaches:
    - Potentially fewer points on surface
    - Prone to wrong correspondences
  - Occlusions, shadows
  - Problems through weakly textured surface, smoke, blood, etc.

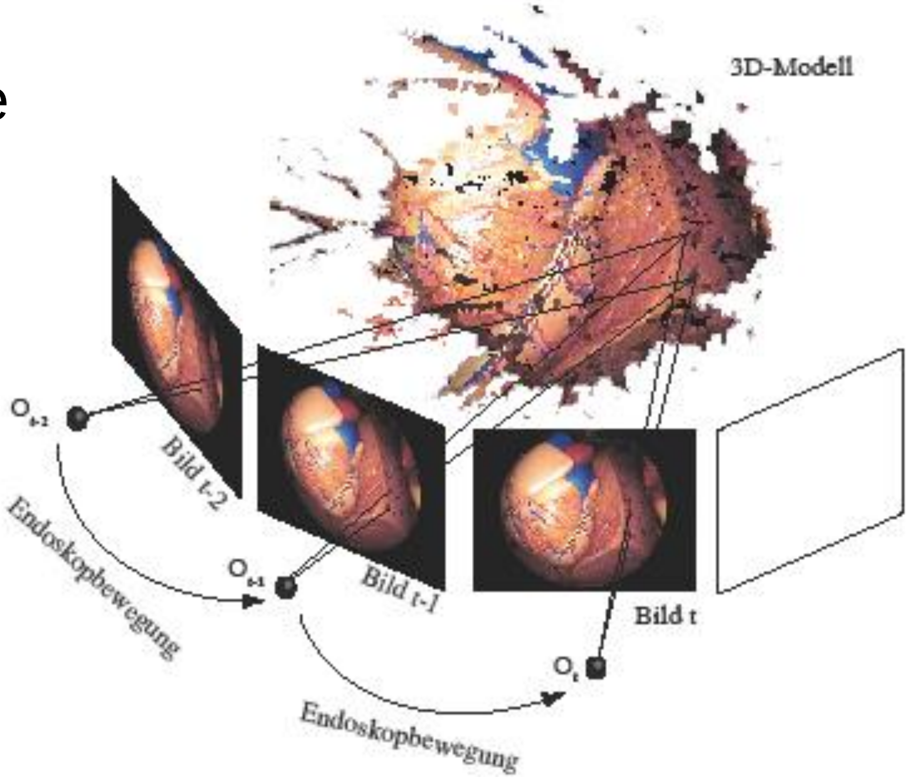
Structure-from-Motion

Problem:

- One channel endoscope

- Computation of scene structure and camera movement from images

- Image either simultaneously or sequential, scenes are geometrically equivalent
- Position of camera not know: has to be estimated from correspondences



NCT

Quelle:Mountney, Imperial College

Structure from Motion - Evaluation

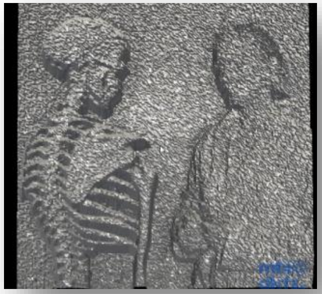
- Pros
  - High accuracy with good correspondences
  - Relative low hardware cost in comparison to other methods

- Cons
  - Potentially few features
  - Prone to correspondence errors
  - Occlusions, shadows
  - Problems from sparsely textured surfaces, specularities, smoke etc.
  - Often requires tracking
  - Difficult with fast moving objects

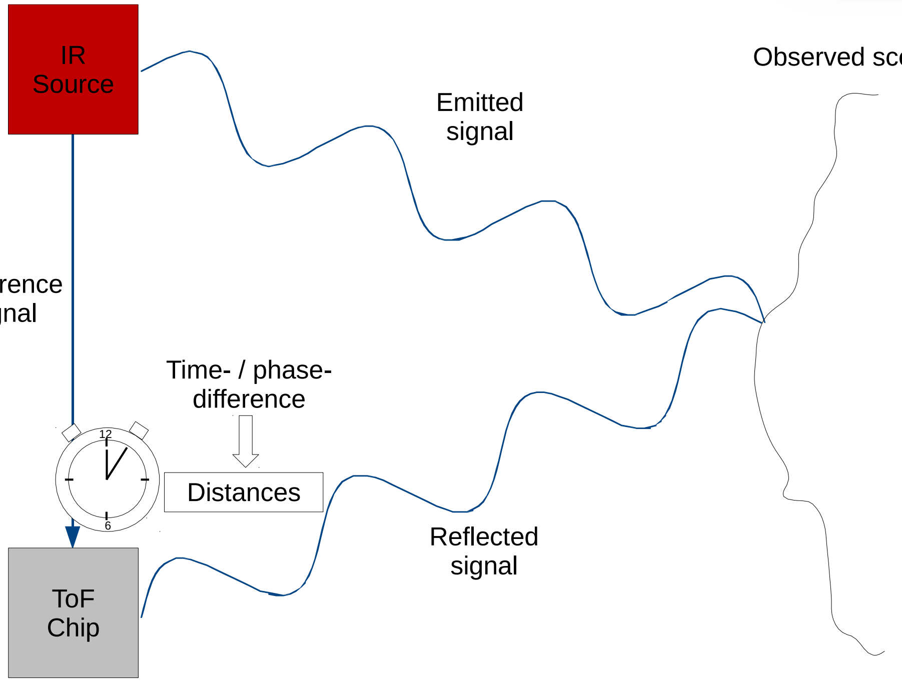
Further methods: Time-of-Flight

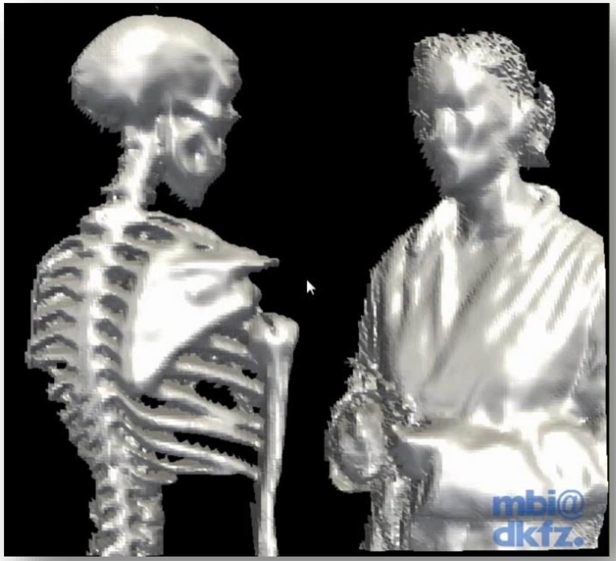
Time-of-Flight (ToF) Camera

Range image



Surface





Quelle:Seitel, Maier-Hein, DKFZ

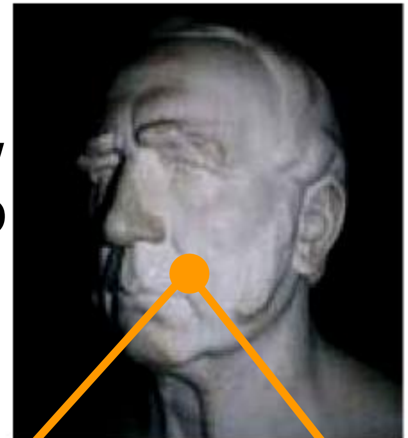
ToF - Evaluation

- Pros
  - No features required
  - Dense depth map
  - No shadow effects
  - No image processing necessary
- Cons
  - No color image
  - Low resolution
  - Systematic errors
  - Can’t deal with transparent structures

Quelle: Maier-Hein, DKFZ

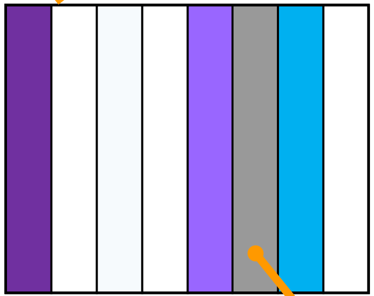
Further methods: Structured light

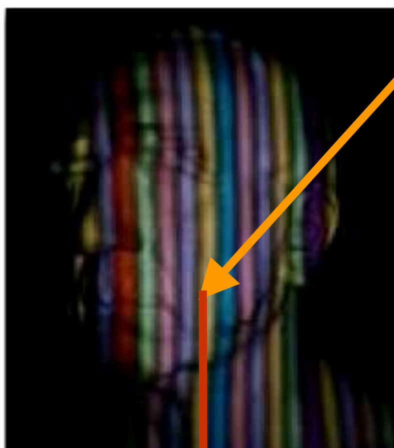
- Through projection of a known pattern, we can draw conclusion on an objects 3D shape



Example frequency coding

Pattern





Color identifies the stripe

Structured light - Evaluation

- Pro
  - No need to rely on features
  - Density of correspondence selectable
  - Less complex correspondence search
  - Works well for homogenous surfaces
- Cons
  - Additional hardware
  - Projected light can be bothersome
  - Sensitive against reflections and transparencies
  - Difficult: Segmentation of symbols/correct detection of color values

Quelle: Maier-Hein, DKFZ

Literature

- Trucco,Verri: “Introductory Techniques for 3D Computer Vision”
- Hartley, Zisserman: “Multiple View Geometry”
- Vogt. et al.: “Bildverarbeitung in der Endoskopie des Bauchraums”. BVM 2001
- Zimmerman et al.: “Automatic Detection of Specular Reflections in Uterine Cervix Images“. SPIE Medical Imaging 2006
- Wengert et al.: „Markerless Endoscopic Registration and Referencing“. MICCAI 2006
- Stoyanov et al.: “Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures”. MICCAI 2005
- Mountney et al.: Motion Compensated SLAM (MC-SLAM) for Image Guided Surgery
- Dillmann et al.: Lecture Robotik III, KIT

NCT