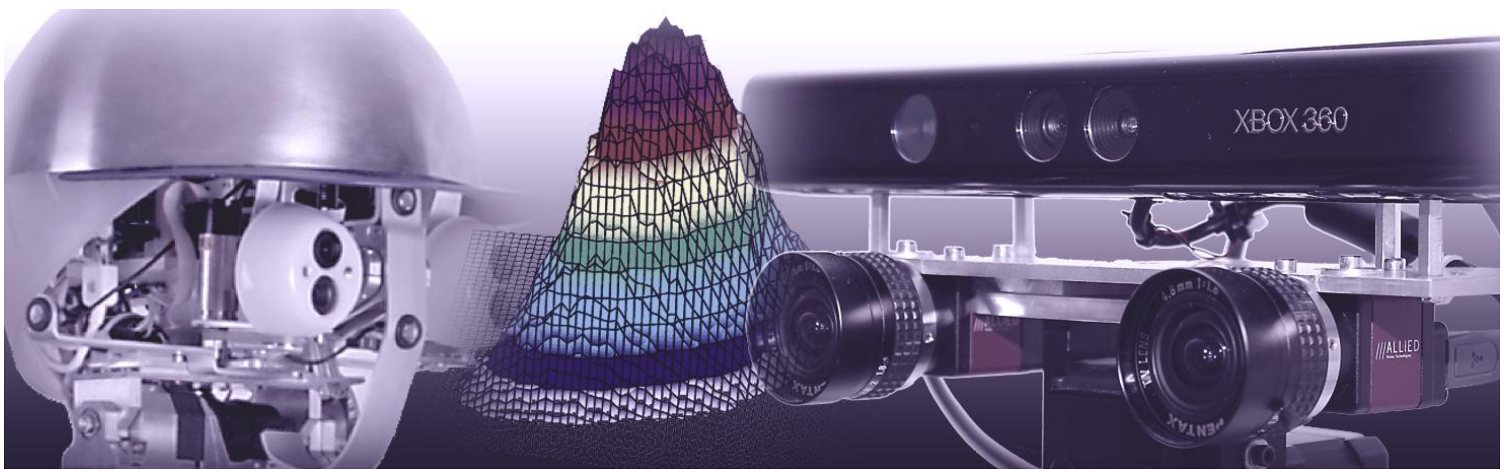
Robotics III: Sensors

Chapter 7: Optical 3D Sensors

Tamim Asfour

Department of Informatics, Institute for Anthropomatics and Robotics (IAR)
High Performance Humanoid Technologies Lab (H^2T)



http://www.humanoids.kit.edu

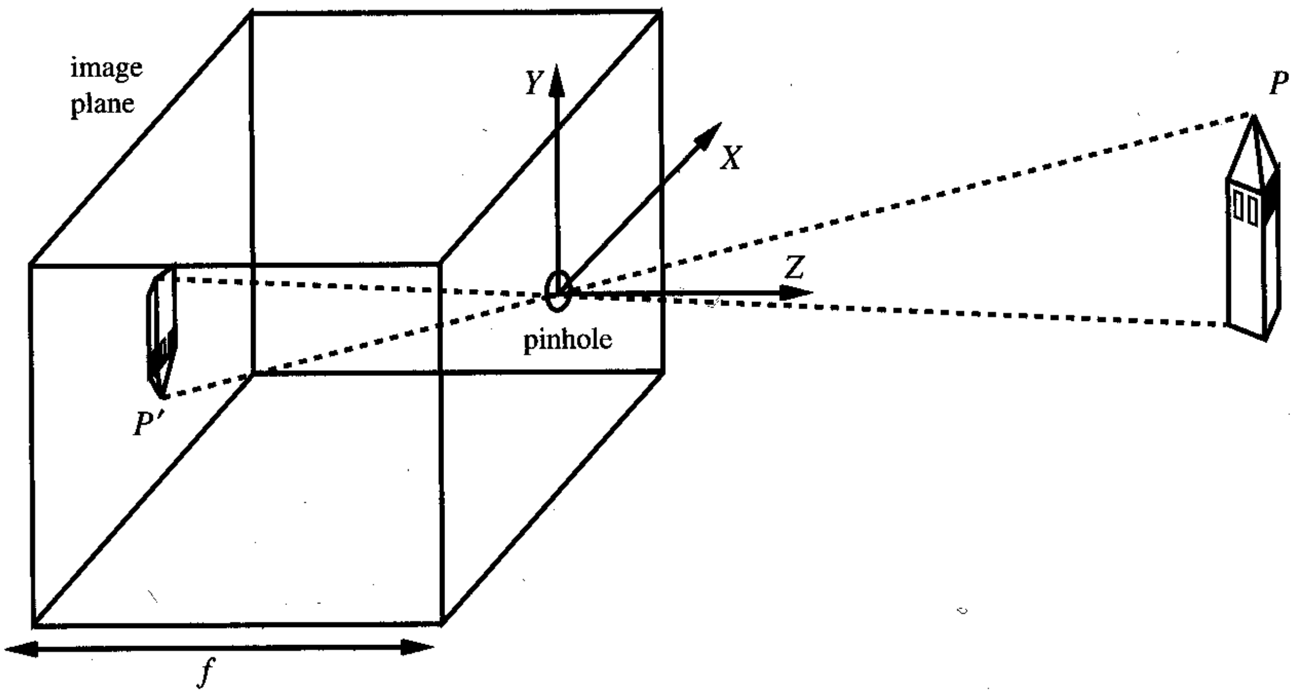
http://h2t.anthropomatik.kit.edu

Inhalt

- Pinhole camera model
- Extended camera model
  - Projection
  - World coordinate system
  - Consideration of lens distortions
- Stereo geometry (Epipolar geometry)
- Optical 3D sensors
  - Passive methods
  - Active methods

Pinhole I

- The simplest model: Pinhole camera model



Internal parameters: focal length f ("focal distance")

Pinhole II

- Projection of a scene point $P = (X,Y,Z)$ on to a pixel $p = (u,v,w)$:

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

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

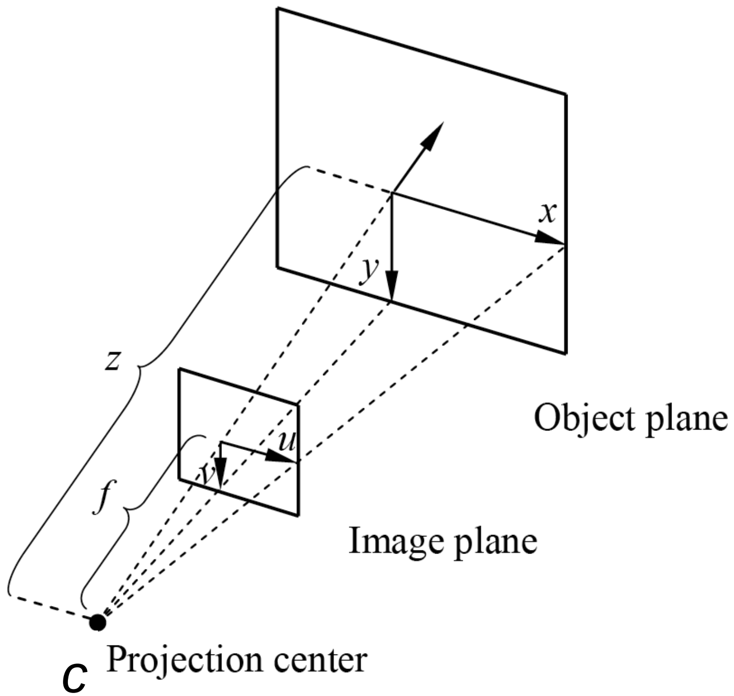
Perspective projection

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

Projecting back

Pinhole III

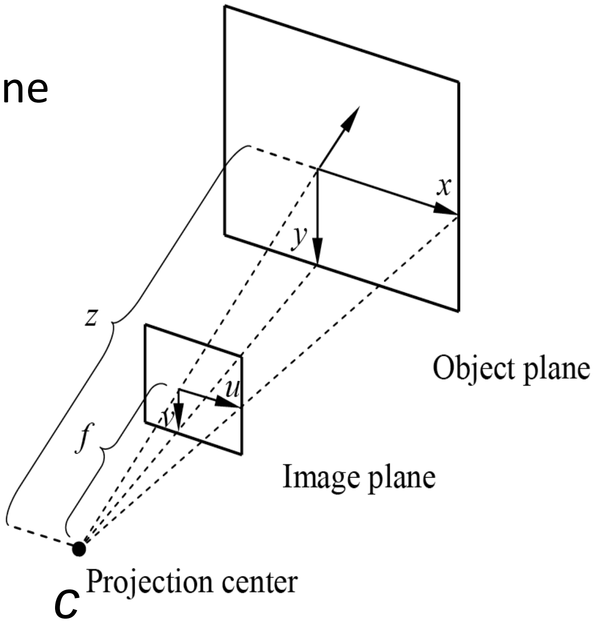
- Often used version:
  Pinhole camera model in *Positive Location* :
  - Projection center C is located behind the image plane
  - This means: no mirroring (minus signs are omitted)



Extended Camera Model I

- Pinhole camera model simplifies the real conditions strongly. Therefore, this model needs to be extended to be used also in practice.

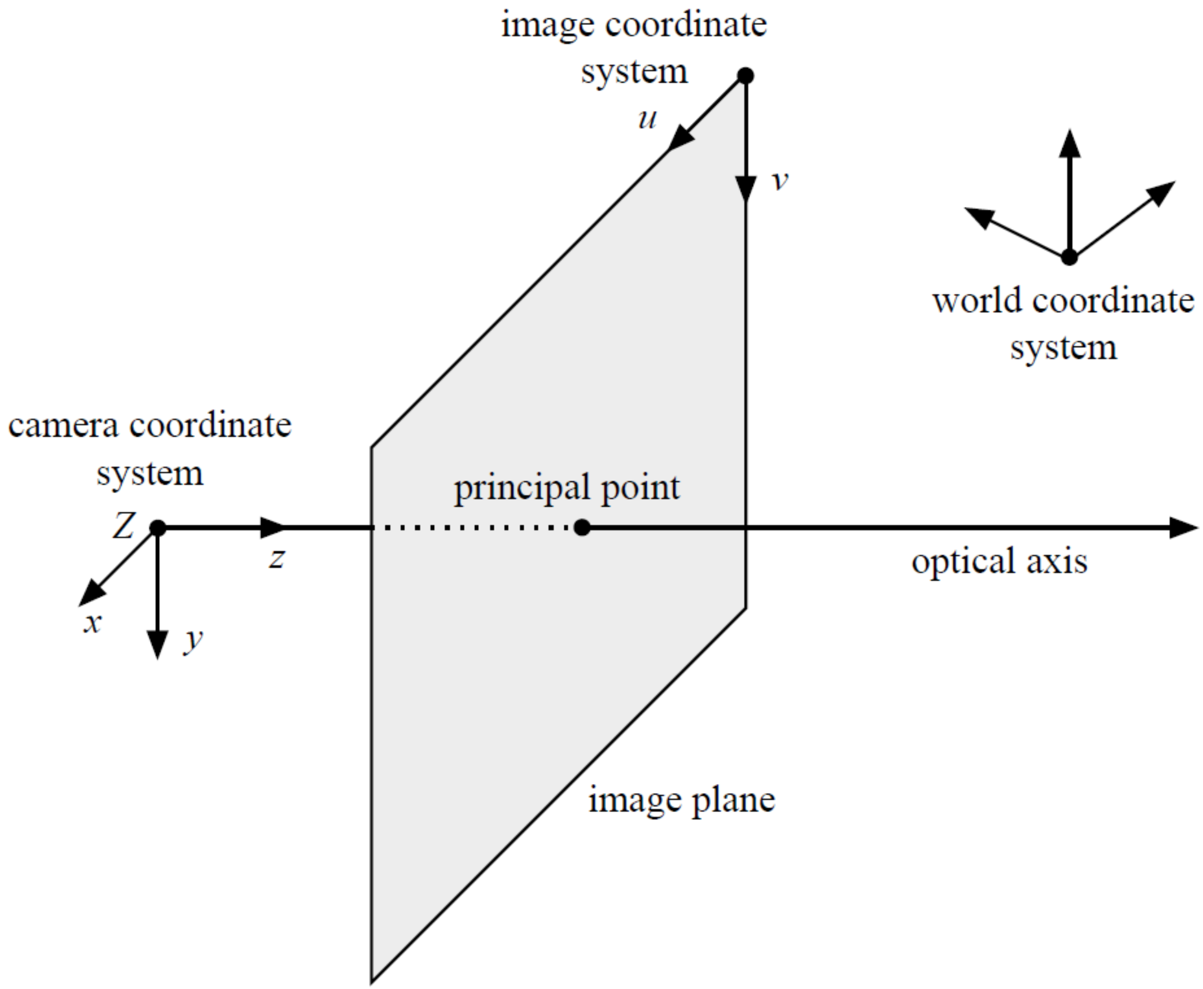
- First, some definitions:
  - Optical axis:
    Straight through the projection center, perpendicular to the image plane
  - Principal point $C(c_x, c_y)$:
    Intersection of the optical axis with the image plane



Extended Camera Model II

- Coordinate Systems:
  - Image coordinate system:
    - 2D coordinate system
    - Unit [pixels]
    - Agreement for the Lecture (applies to most camera drivers): origin in the upper left corner of the image, u axis points to the right, v Axis points downwards
  - Camera coordinate system:
    - 3D coordinate system
    - Unit [mm]
    - Origin is in the Projection center, axes parallel to the axes of the Image coordinate system, i.e. x axis to the right, y axis downwards, and the z axis in accordance with the three-finger rule for a Right-handed coordinate system to the front
  - World coordinate system:
    - 3D coordinate system
    - Unit [mm]
    - Basic coordinate system that can be anywhere in the room

Extended Camera Model III



Extended Camera Model IV

- Terms:
  - Intrinsic camera parameters:
    - Focal length, image point
    - Parameters for the description radial / tangential Lens distortion
    - Define the non (unambiguous) reversible illustration from camera coordinate system into the Image coordinate system
  - Extrinsic camera parameters:
    - Define the relationship between the camera and the World Coordinate System
    - Generally described by a rotation ***R*** and a Translation ***t***

Extended Camera Model V

- Simplifications of the Pinhole camera model:
  - Principle point is in the center of the image plane
  - Pixels are assumed to be square
  - No modeling of lens distortion
  - There is no world coordinate system or it is identical with the camera coordinate system, i.e., no extrinsic camera parameters

Extended Camera Model VI

- Focal length:
  - Focal length is the distance between projection center and image plane
  - Since pixels are not like square but rather like rectangular, there is one parameter for each direction, i.e.: f_x, f_y
  - The parameters f_x, f_y are the products from the actual Focal length with unit [mm] and the respective conversion factor with unit [Pixel / mm]
  - The unit for the parameter f_x, f_y is thus [Pixel]

Extended Camera Model VII

- The imaging of the camera coordinate system in the Image coordinate system, exclusively with the Intrinsic parameters is then defined by:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \frac{1}{Z} \begin{pmatrix} f_x \cdot X \\ f_y \cdot Y \end{pmatrix}$$

- Or, as a matrix multiplication by **calibration matrix K** in Homogeneous coordinates:

$$\begin{pmatrix} u \cdot w \\ v \cdot w \\ w \end{pmatrix} = K \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \qquad K = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

Extended Camera Model VIII

- Extrinsic camera calibration
  - Is defined by a coordinate transformation from rotation ***R*** and translation ***t***
  - Coordinate transformation from the world coordinate system to the Camera coordinate system:

$$\boldsymbol{x}_c = R\boldsymbol{x}_w + \boldsymbol{t}$$

  - The final output is a 3×4 **projection matrix *P*** (involving both intrinsic and extrinsic parameters) in homogeneous coordinates:

$$\begin{pmatrix} u \cdot w \\ v \cdot w \\ w \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \qquad P = (KR \mid K\boldsymbol{t})$$

Lens Distortions I

- The imaging by real lenses is not perfectly linear
- In particular, lenses with a small focal length form the (Radial) distortion

A sample ***distorted*** camera image!

Lens Distortions II

- Models are generally used
  - Radial lens distortions
  - Tangential lens distortions

- The output is the projection of the undistorted Coordinates on the plane z = 1:

- For the distorted image coordinates: $$\begin{pmatrix} x_n \\ y_n \end{pmatrix} := \begin{pmatrix} \frac{u - c_x}{f_x} \\ \frac{v - c_y}{f_y} \end{pmatrix}$$

- Radius: $$r := \sqrt{x_n^2 + y_n^2}$$

Lens Distortions III

- From the coordinates x_n, y_n, the distorted coordinates are computed according to the distortion model

- Radial lens distortion

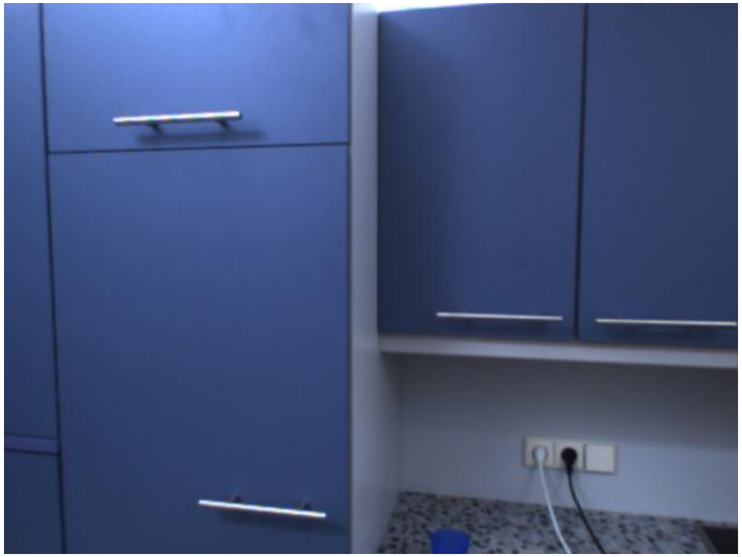
$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = (1 + d_1 r^2 + d_2 r^4) \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

- Tangential lens distortion

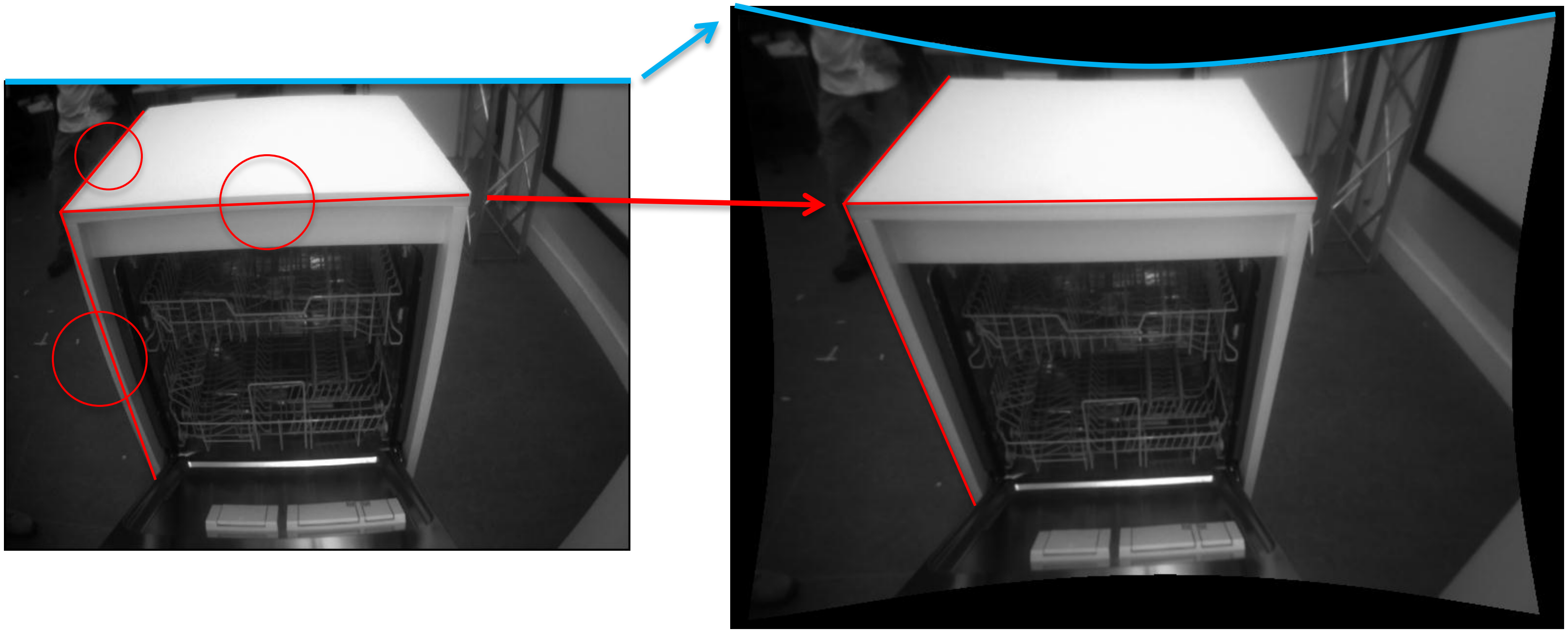
$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} + \begin{pmatrix} d_3 (2 x_n y_n) + d_4 (r^2 + 2 x_n^2) \\ d_3 (r^2 + 2 y_n^2) + d_4 (2 x_n y_n) \end{pmatrix} \quad \begin{pmatrix} u_d \\ v_d \end{pmatrix} = \begin{pmatrix} f_x x_d + c_x \\ f_y y_d + c_y \end{pmatrix}$$

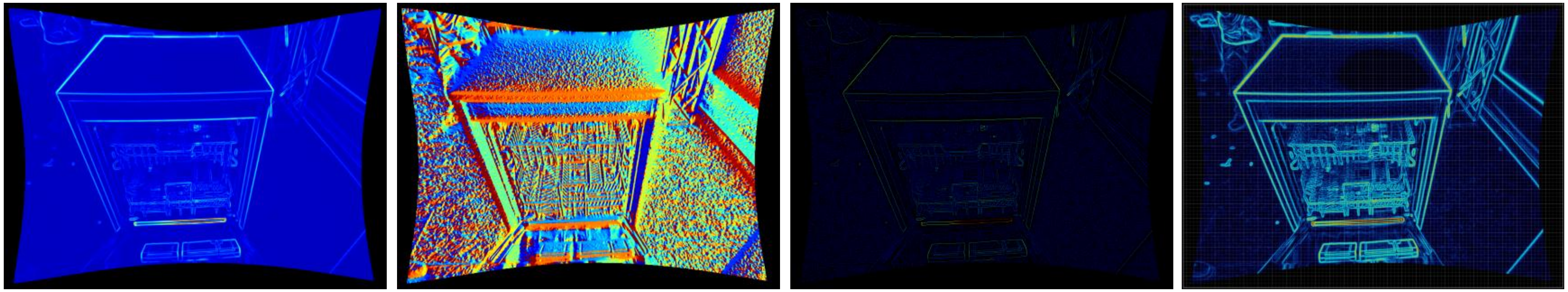
Lens Distortions IV

- Example of an undistorted image
  - For each pixel in the rectified image, the intensity or color value is determined by "lookup" in the distorted original image and interpolation (e.g., bilinear)



Lens Distortions V





Camera Calibration I

- The calibration of a camera means the determination of the parameters with respect to a selected one camera model

- The determination of the intrinsic parameters is independent of the structure; As long as the zoom and focus of the camera remain the same, these parameters do not change

- The determination of the extrinsic parameters depends on the selection of the world coordinate system and changes depending on the structure

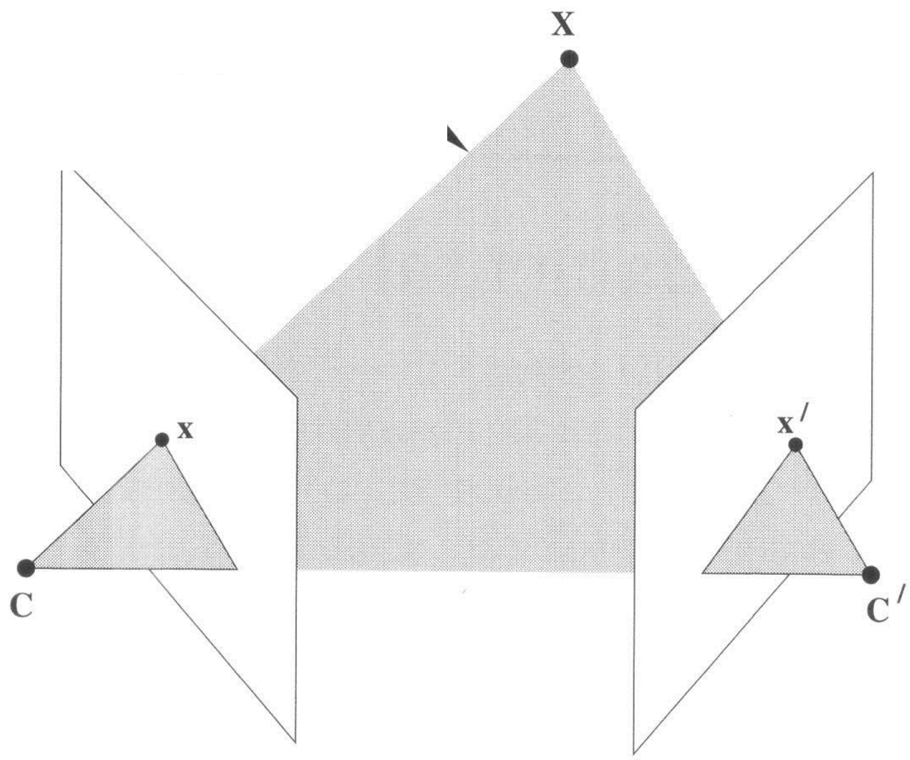
Camera Calibration II

- If the camera is calibrated, then the imaging function f maps a point from the world coordinate system unambiguously into the image coordinate system:
  - $f : R^3 \rightarrow R^2$
- f is defined by the projection matrix P and subsequent transformation of the homogeneous coordinates by division of w
- The inverse image maps a point in the image coordinate system to a straight line in the world coordinate system that passes through the projection center

Stereo Reconstruction I

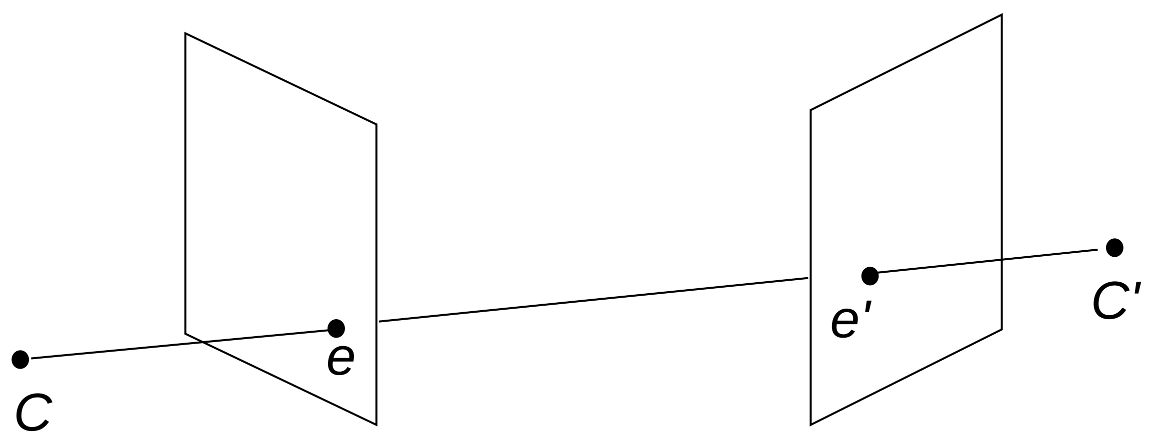
Given:

- Two cameras with projection matrices C and C'
- Two images $\mathbf{x}$ und $\mathbf{x}'$ of the point X
- Then X can be reconstructed



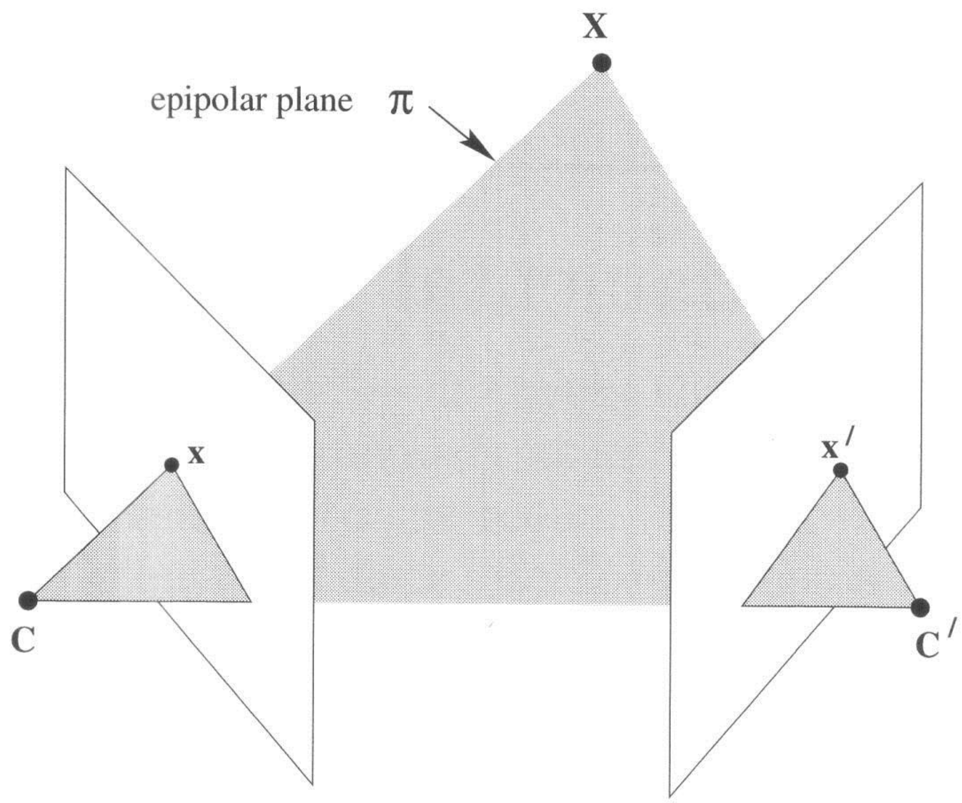
Epipolar Geometry I

- Connection between two cameras is given by the epipolar geometry
- *The intersections* e and e' of *the straight line through the projection centers with the image planes are called Epipole*



Epipolar Geometry II

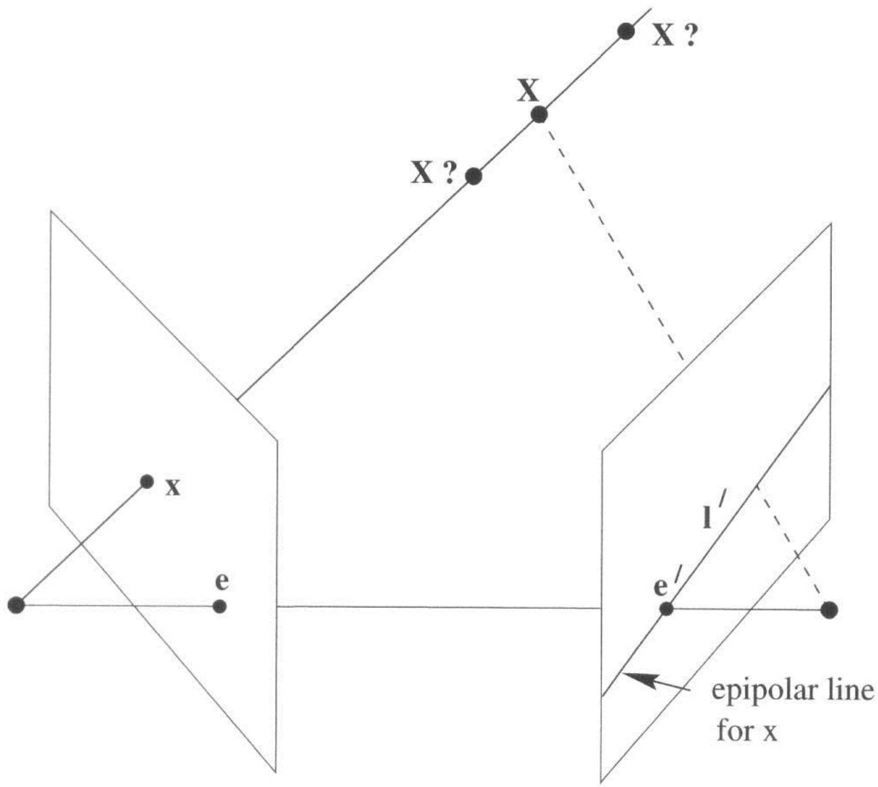
- *Epipolar plane $\pi(X)$:*
  A plane passes by C, C' and scene point X



Epipolar Geometry III

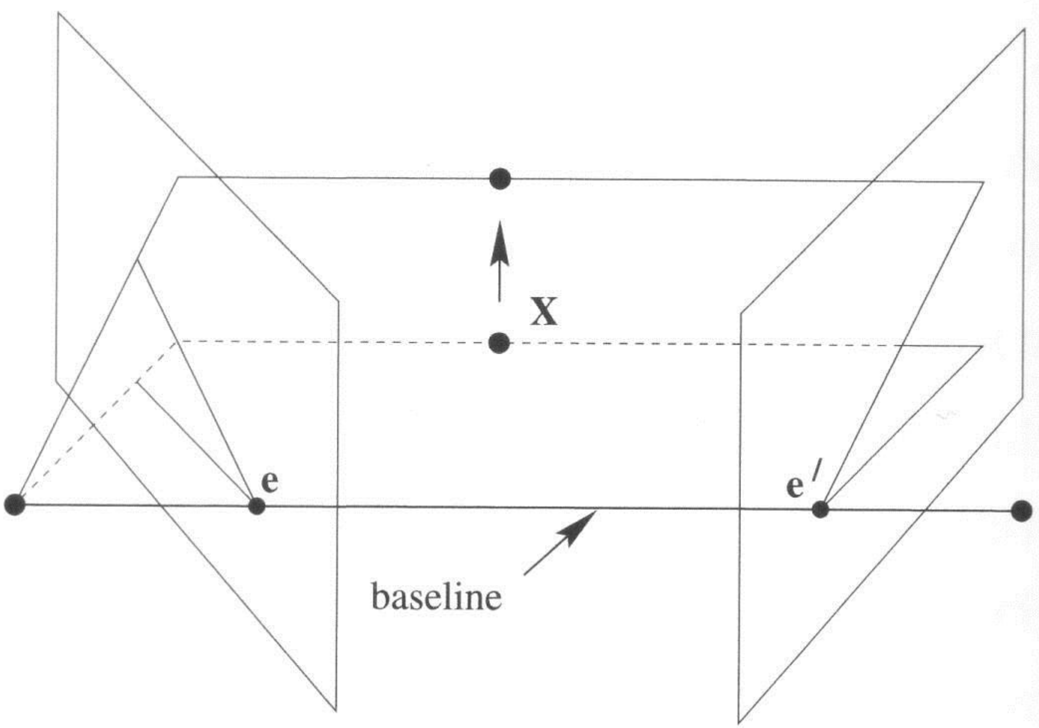
- Epipolar line $l'(x)$: Line of intersection of $\pi(X)$ with image plane

- All points X, which are imaged on x in camera image 1, are mapped to a Point of the line $l'(x)$ in camera image 2.



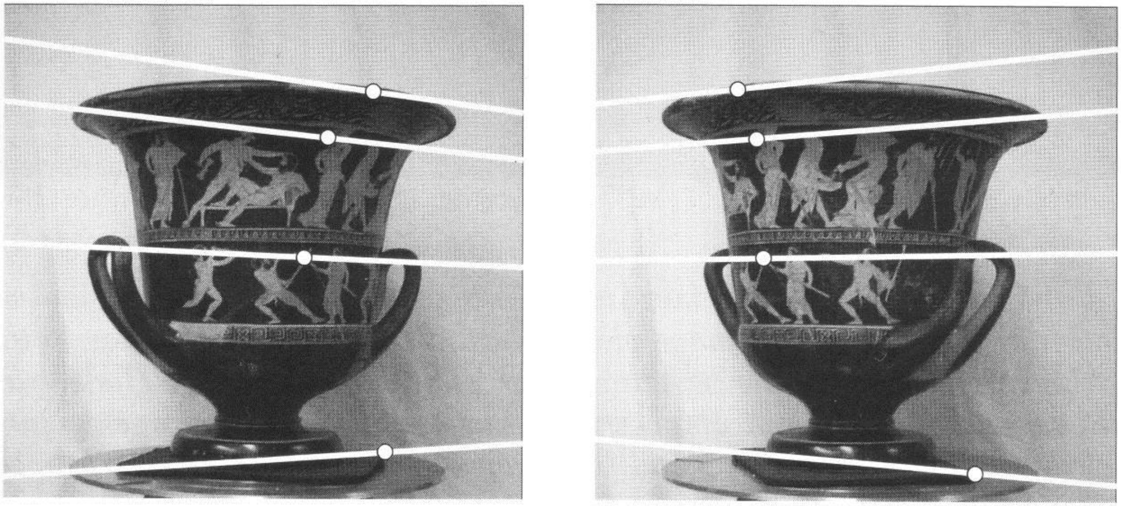
Epipolar Geometry IV

- All epipolar lines of a camera system intersect in the epipoles e and e'



Epipolar Geometry V

- Use
  - Restriction of the correspondence problem from two dimensions to one dimension since, according to corresponding features, only the epipolar line has to be searched, therefore:
  - Higher robustness (less false correspondences)
  - Higher efficiency



Fundamental Matrix I

- Mathematical description of epipolar geometry is performed by the fundamental matrix
- Properties of the **fundamental matrix** $\boldsymbol{F}$:
  - Is a 3×3-Matrix
  - Has Rank 2
  - For all correspondences $\boldsymbol{x}, \boldsymbol{x}'$:
    $\boldsymbol{x}'^{T} F \boldsymbol{x} = 0$
    ($\boldsymbol{x}$ and $\boldsymbol{x}'$ are pixels in homogenous coordinates with $w = 1$)

Fundamental Matrix II

- The epipolar lines can be calculated with the fundamental matrix
- Epipolar lines:
  - $\boldsymbol{l} = F^T\boldsymbol{x}'$
  - $\boldsymbol{l}' = F\boldsymbol{x}$
- The following applies to the epipoles:
  - $F\boldsymbol{e} = 0$
  - $F^T\boldsymbol{e}' = 0$
- Note: $\boldsymbol{l}$ (or $\boldsymbol{l}'$) defines a 2D straight line as follows: $\boldsymbol{l}\cdot\boldsymbol{x} = 0$ for all pixels $\boldsymbol{x}$ (in homogenous coordinates with $w = 1$), which lies on this straight line

Fundamental Matrix III

- The fundamental matrix can be calculated in several ways:
  - About image point correspondences in the left and right camera
  - For known intrinsic and extrinsic calibration of the cameras directly via the calibration matrices K, K' and the essential matrix E, which is defined by the extrinsic parameters

Fundamental Matrix IV

- Calculation of the fundamental matrix via Essential matrix is possible

- Essential matrix can be calculated by the extrinsic parameters:
  - Given:
    - Camera 1 with $(I \mid \boldsymbol{0})$ as Transformation (Identical)
    - Camera 2 with $(R \mid \boldsymbol{t})$ as Transformations
  - Essential matrix E can be calculated as:

$$E = [t]_\times R = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} R$$

The following applies to the epipoles:

$\boldsymbol{e} = -KR^T\boldsymbol{t}$

$\boldsymbol{e}' = K'\boldsymbol{t}$

Fundamental Matrix V

- Having computed the essential matrix (e.g., calculated via the extrinsic parameters) and the intrinsic parameters, i.e. the calibration matrices K, K', the fundamental matrix can be calculated as:

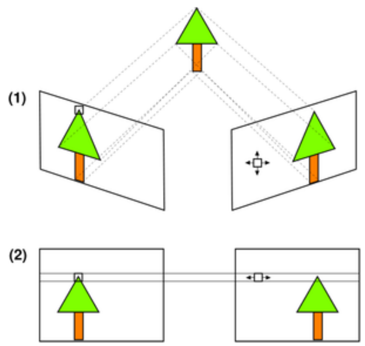
$$F = K'^{-T} E K^{-1}$$

- Conversely, if the fundamental matrix has been determined (e.g., via pixel correspondences) and the intrinsic parameters, i.e. the calibration matrices K, K', the essential matrix can be calculated as:

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

Stereoscopy: Depth Maps I

- Benefits of the Fundamental Matrix:
  - By using the fundamental matrix, the input images can be *rectified*
    - After rectification, all epipolar lines run horizontally with the same v-coordinate as the image point in the other camera image
    - After correspondences only horizontal (in one direction) has to be searched



Stereoscopy: Depth Maps II

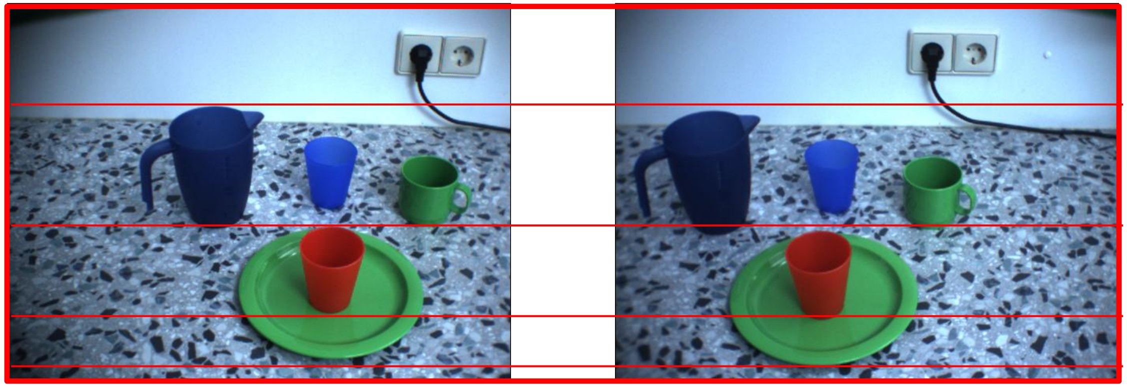
- Rectified images have the advantage that optimized correlation algorithms can be used for solving the correspondence problem
  ⇨ 30 Hz (and higher) for calculating the disparity card at 640 × 480 8-bit gray scale

- Disadvantage:
  - Interpolation necessary for the calculation of the rectified images ⇨ Quality loss
  - Images strongly distorted depending on the structure

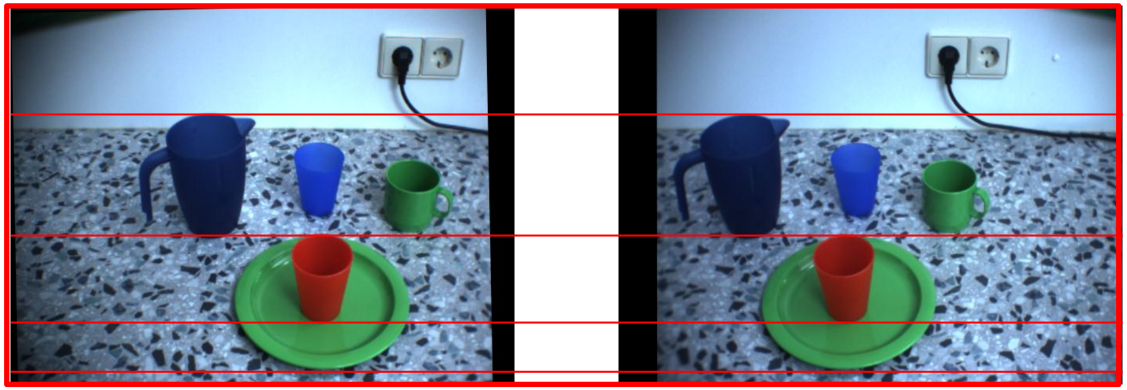
Stereoscopy: Depth Maps III

- Example of rectification with a standard stereo setup ⇨ relatively low distortion

Original Images
Left / Right



Rectified Images
Left / Rigth

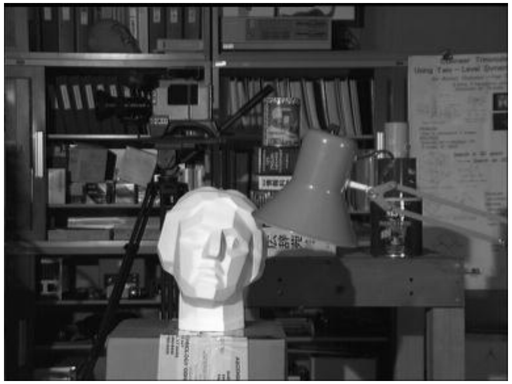


Stereoscopy: Depth Maps IV

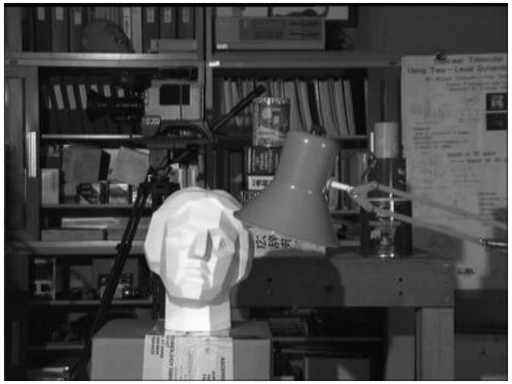
- After solving the correspondence problem:
  - Point clouds can be calculated by triangulation, as explained before
  - Depth images are generated by recording the disparities (Difference of u-coordinates for correspondence found in the rectified images) into a gray scale image: ⇨ The higher the gray value, the closer the corresponding 3D point to the camera is

Stereoscopy: Depth Maps V

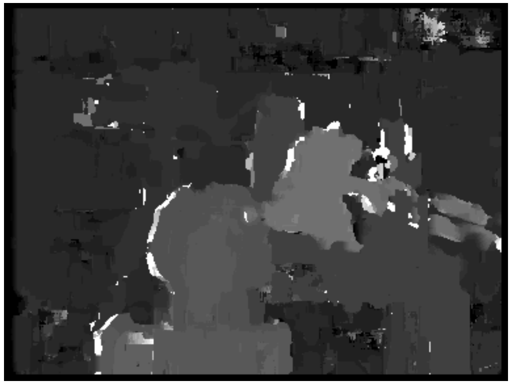
- Example of standard benchmark image pair “Tsukuba”



Left Image



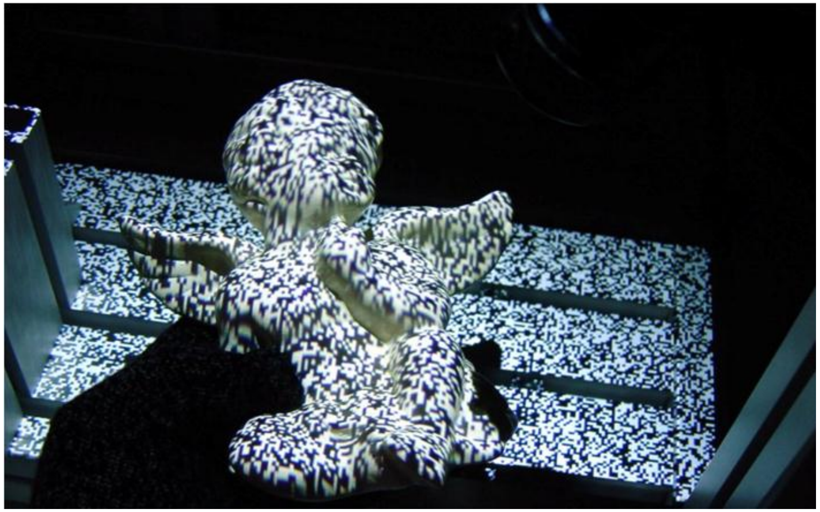
Right Image



Depth Image

Passive Pattern Projection

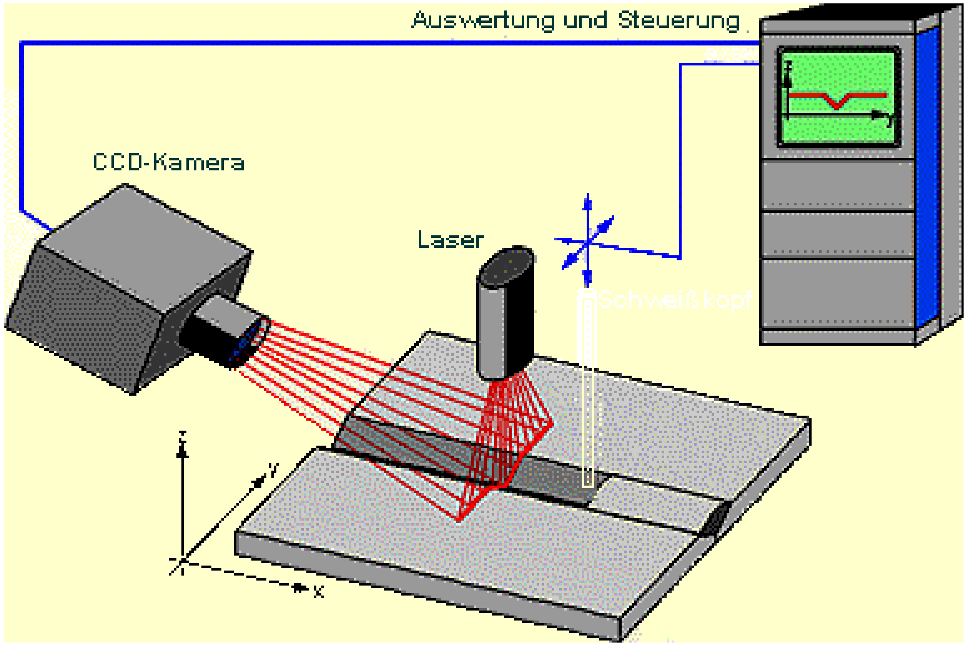
- A pattern is projected to make homogeneous surfaces structured
- Knowledge of the pattern is not necessary
- Projector does not need to be calibrated
- Correspondence problem for stereo camera systems can be solved more effectively



Active Pattern Projection

- Idea:
  Geometrical structure coded in projected light can be read back from the image

- Principle: Triangulation
  - Projection of a light pattern on object
  - Observation of the projected pattern by camera
  - Calculation of the selected 3D point



Structured light: Faster recording

- Projection of two dimensional patterns
- Problem: Correspondence problem

- Which point in the camera image corresponds to which ray of the projector?

Solution

- Types of patterns for solving correspondence problems

  - Time coded methods

  - Phase shift method

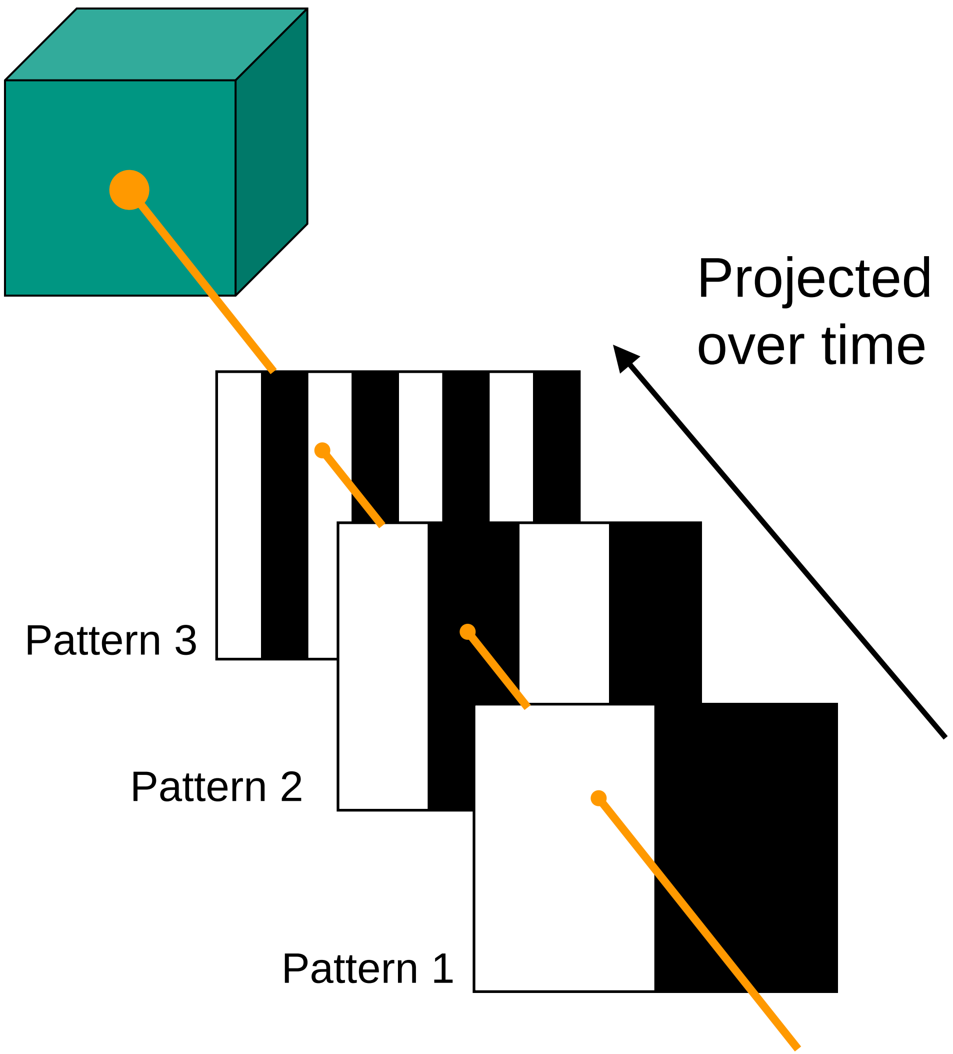
  - Frequency encoding

  - Locally coding methods
    - Color coding
    - Binary coded black and white pattern

Binary Coding / Time Coded Process

Example:

3 binary-encoded patterns which allows the measuring surface to be divided in 8 sub-regions

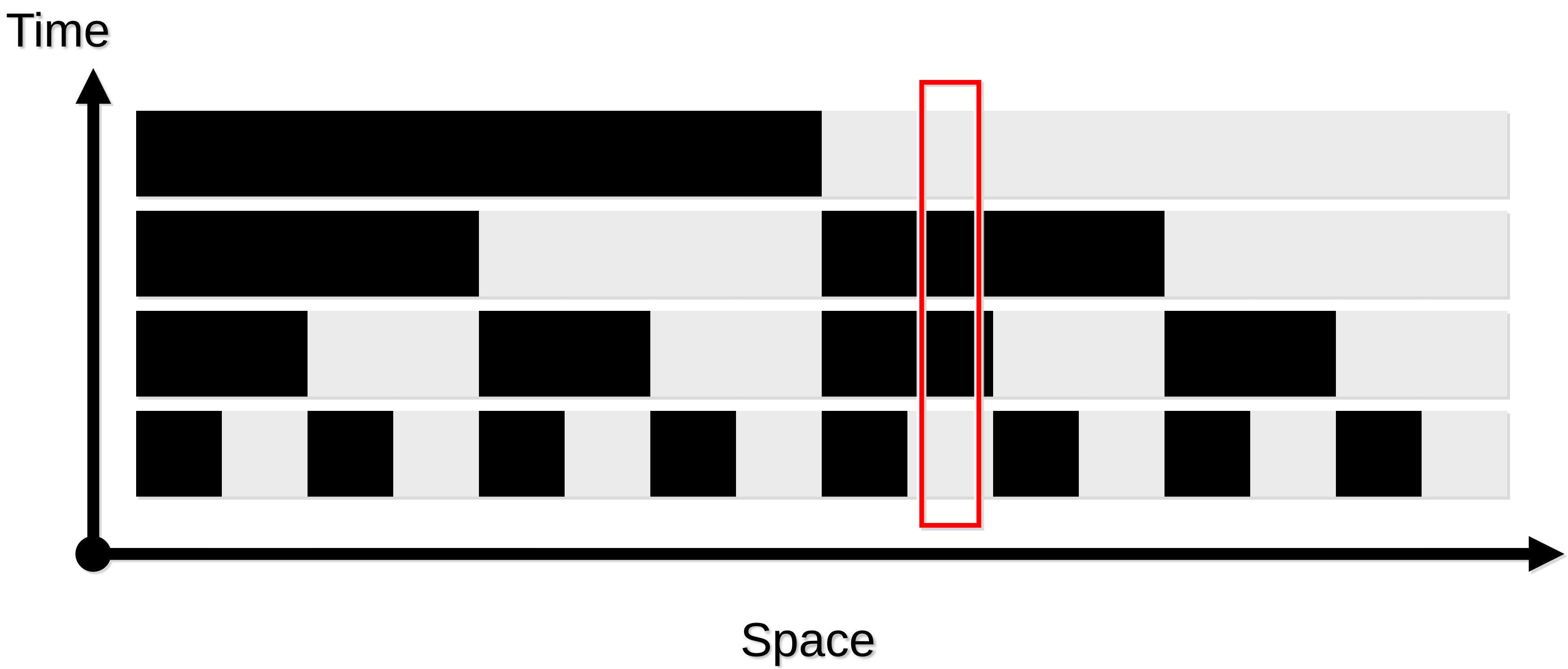


Temporal Coding I

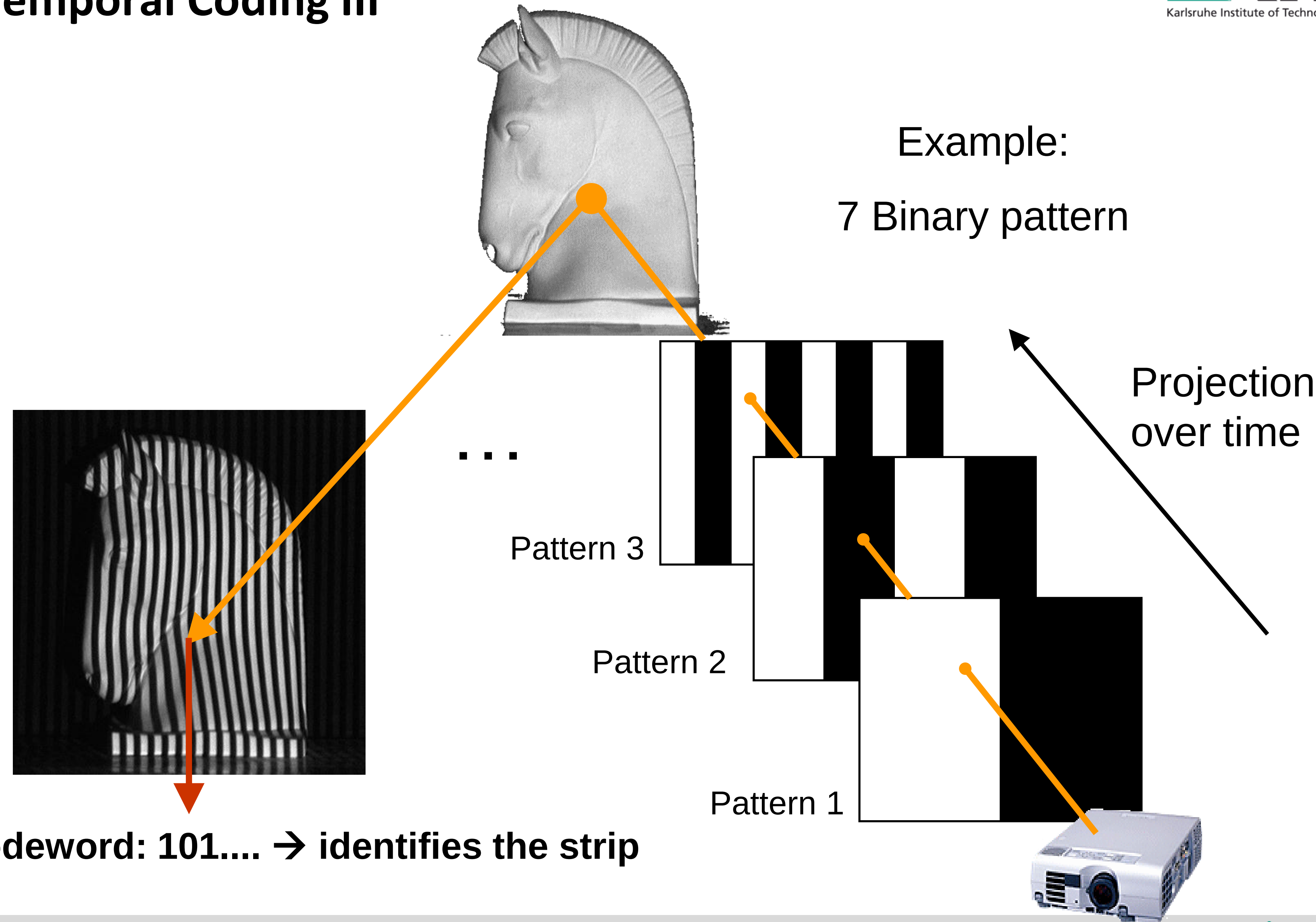
- Projecting many strips one after the other
  - Binary coding of stripe pattern → smaller number of projections
  - When n projections with different patterns n → 2^n strips
  - In the event of a faulty evaluation of a pixel code value, max. Error: 2^{n-1}
  - Using the GrayCode → max. Error: 1

Temporal Coding II

- Each strip is made by projecting several patterns each of which has a unique code [Posdamer 82]



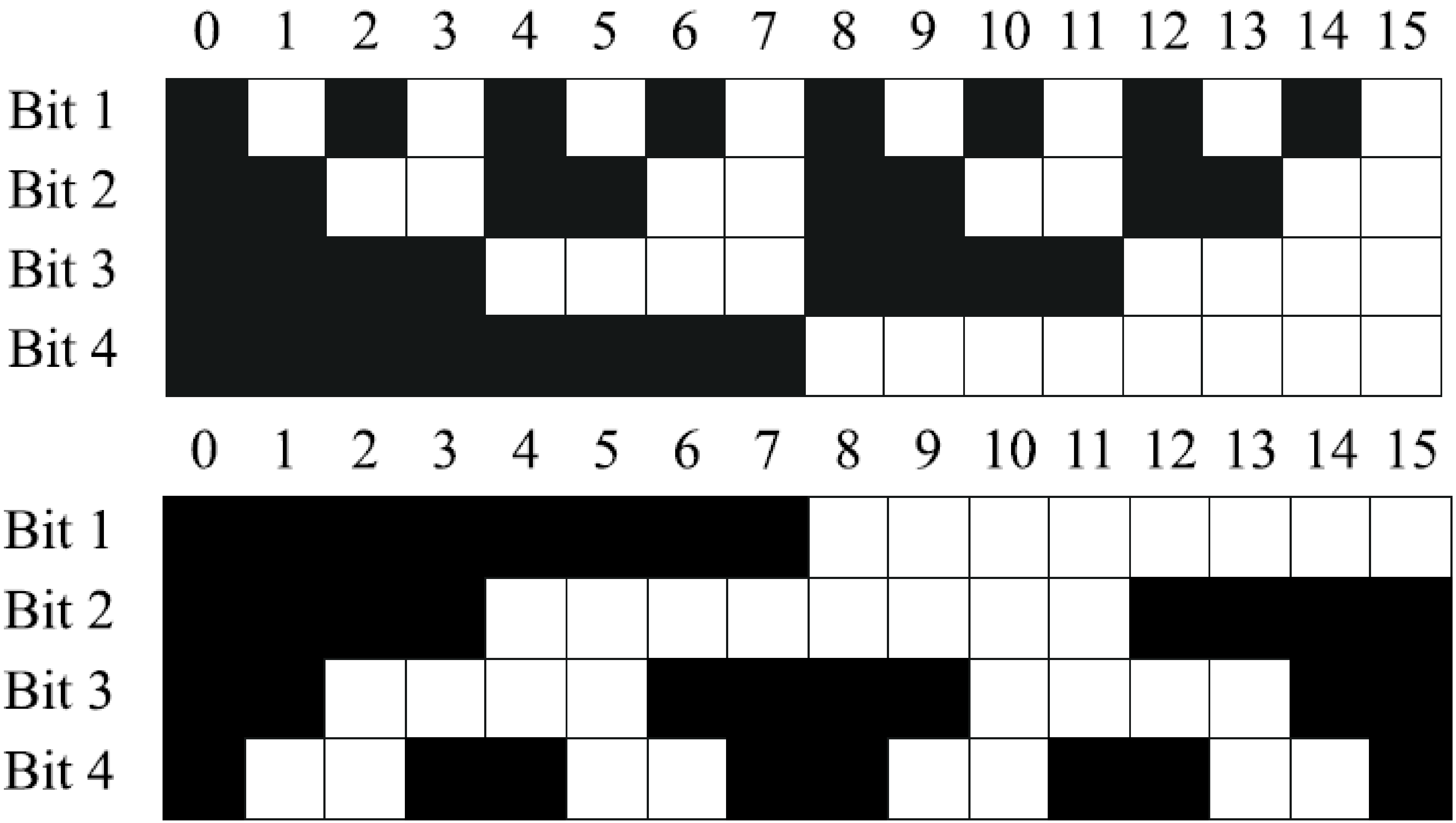
Temporal Coding III



Codeword: 101.... → identifies the strip

Temporal Coding IV

Top: Binary code, Bottom: Gray code

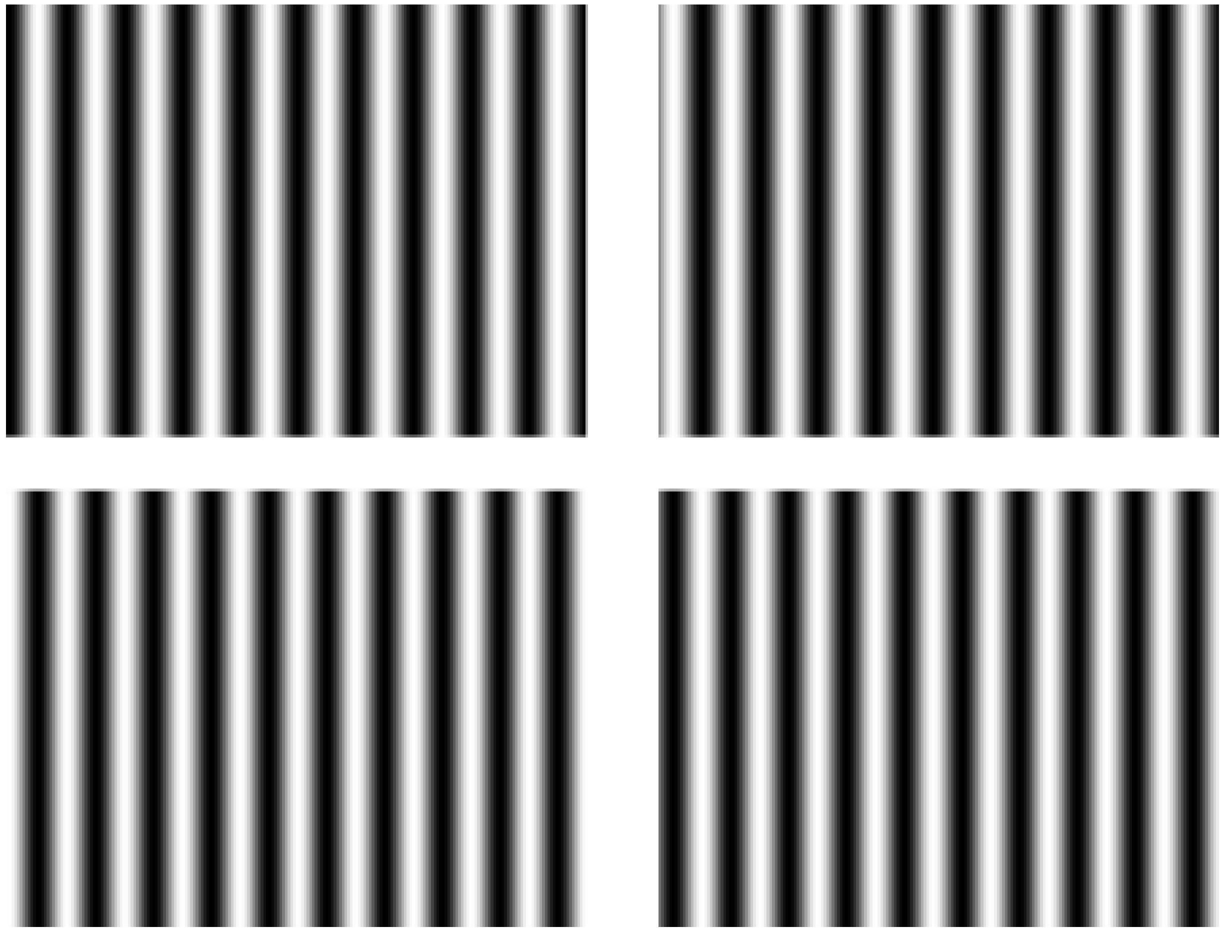


Temporal Coding V

- For multiple projection of binary patterns (or Gray Code), the achievable resolution is limited by the resolution of the projector

- Therefore: Combination with phase shifting
  - Phase only uniquely in the interval [$-\pi/2$, $+\pi/2$]
  - Combination solves ambiguity
  - Sub pixel resolution (regarding projector) is achieved

Temporal Coding VI



- Four different phases in the phase shifting process

Phase Coded Methods I

- Sinusoidal gray scale is projected onto the scene
- Intensity value $I_i(x,y)$ in the i-th phase pattern

$$I_i(x,y) = I_0 + A(x,y) \cdot \sin(\varphi(x,y) + i \cdot \Delta\varphi)$$

I_0:	Intensity offset
$A(x,y)$:	Amplitude
$\varphi(x,y)$:	Searched phase value
$\Delta\varphi$:	Phase shift per stage

Phase Coded Methods II

- Ex. One case with 4 measurements and $\Delta\varphi = \pi/2$

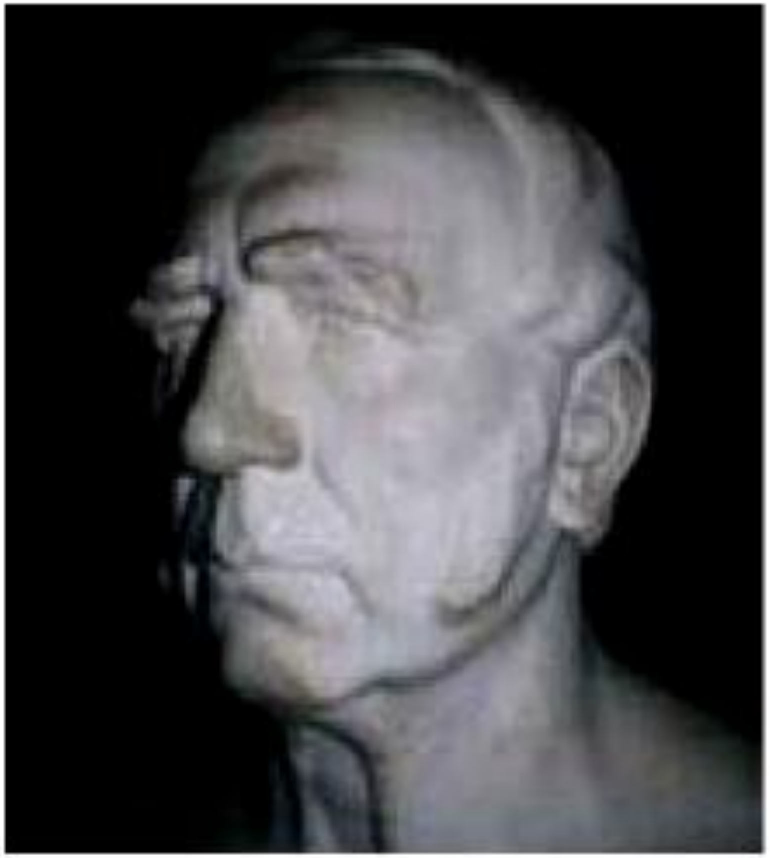
$$\varphi(x, y) = \arctan \frac{I_3(x, y) - I_1(x, y)}{I_2(x, y) - I_0(x, y)}$$

- Uniqueness of the phase value only within one period guaranteed
  → Combine with Graycode method to increase the resolution

Frequency Coding I

- Coding the stripes over color
  - RGB-Image → Hue, Saturation, Intensity – HSI-Colorspace
  - → Use the Hue value
- Hue value indexed in lookup table on stripe number
- Requirement:
  - Maximum of many color values, however, in the picture can be clearly distinguished → no rainbow pattern
  - Projection via flash light source → High contrast, influences of the object texture
  - Example: Minolta 3D 1500

Frequency Coding II

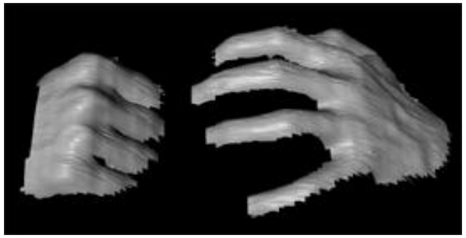


Frequency Coding III

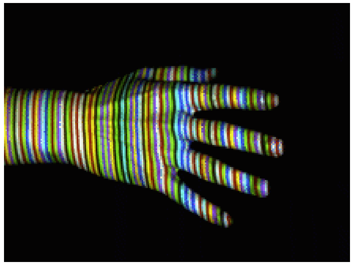
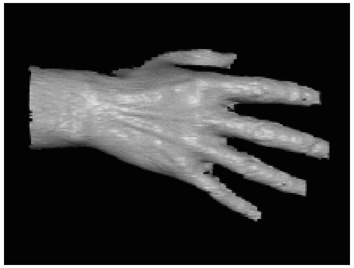
- Advantages
  - A single image is taken
  - Therefore suitable for dynamic scenes
  - Fast

- Disadvantage
  - Prefers homogeneous surface
  - White or color calibration with respect to known material
  - Resolution limited by virtually distinguishable colors

More Complex Procedures

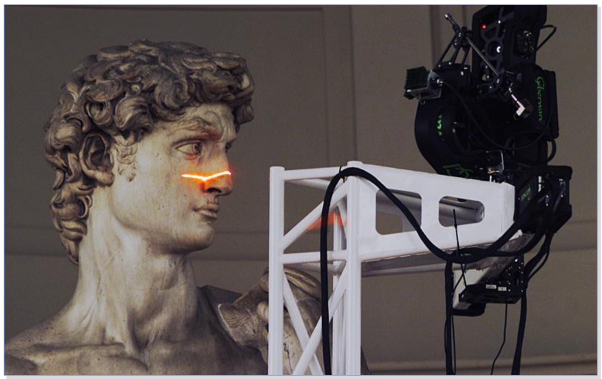
Works despite complex appearances

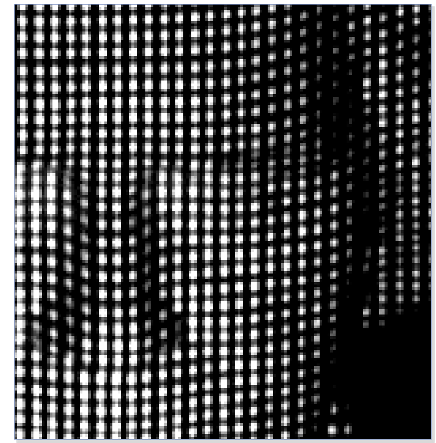
 

Works in real-time and on dynamic scenes

- Need only a few pictures (1 oder 2)
- But requires a more complex correspondence algorithm

Summary: Active Pattern Projection





Several beams
Multiple frames

Single beam

Single frame

Slow, robust

Fast, fragile

Literature

- Camera Modeling
  - Book of Pedram Azad chap. 2.2
- Stereo Vision
  - Book of Pedram Azad chap. 2.10
- Pattern Projection
  - Dissertation by Tilo Gockel – Kap. 2.2