Lehrstuhl für Datenverarbeitung
Fakultät für Elektrotechnik und Informationstechnik
Technische Universität München

Martin Kleinsteuber: Computer Vision

Ch. 1 – Characteristics of Images

1. Image Representation

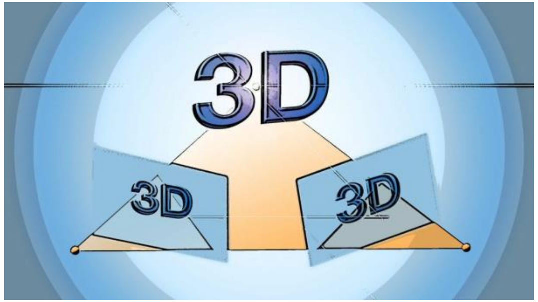
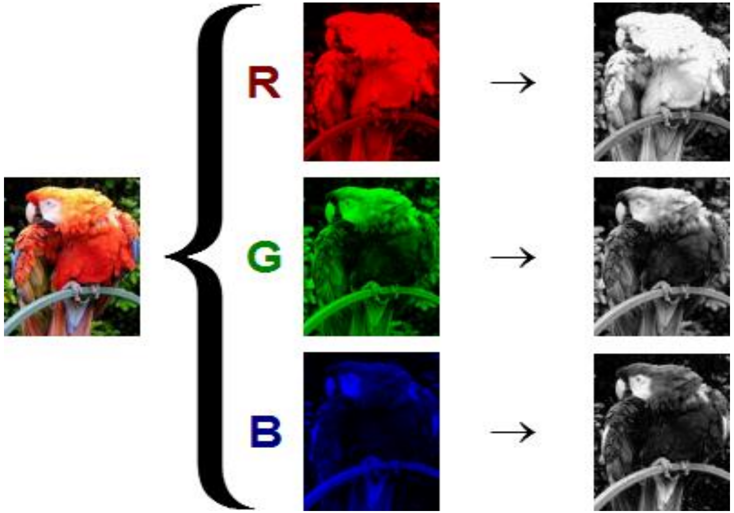


Image Representation

From Color Image to Intensity Image



- Color images consist of several channels
- This lecture deals exclusively with grayscale Images

Image Representation

Continuous and Discrete Representations

- **Continuous** Representation as a function of two variables (to derive algorithms)

  $$I : \mathbb{R}^2 \supset \Omega \to \mathbb{R}, \quad (x, y) \mapsto I(x, y)$$

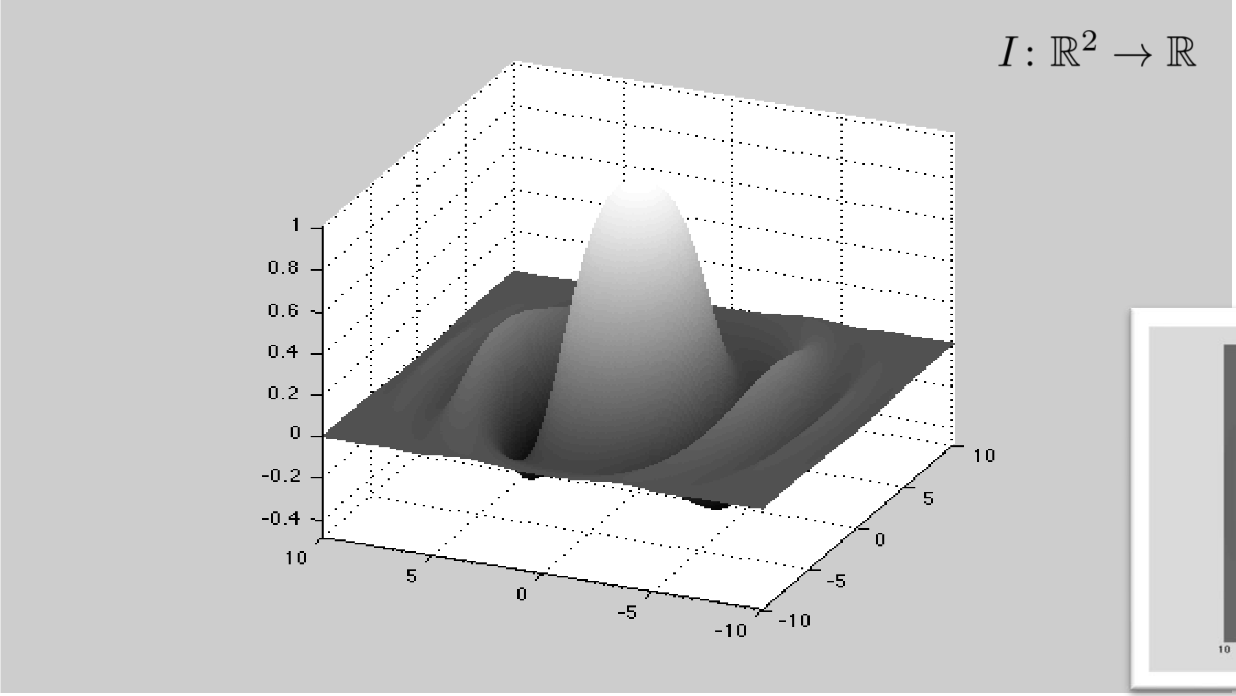
- Frequent Assumptions
  - I is differentiable
  - Ω is simply connected and bounded

- **Discrete** Representation as a Matrix $I \in \mathbb{R}^{m \times n}$ the item $I_{k,l}$ corresponds to the Intensity Value
- The scale is typically between [0, 255] or [0, 1]

VGA: 480 x 640 Pixel (ca. 0.3 Megapixel)
HD: 720 x 1280 Pixel (ca. 1.0 Megapixel)
FHD: 1080 x 1920 Pixel (ca. 2.1 Megapixel)

Image Representation

Graph of a function



Top view

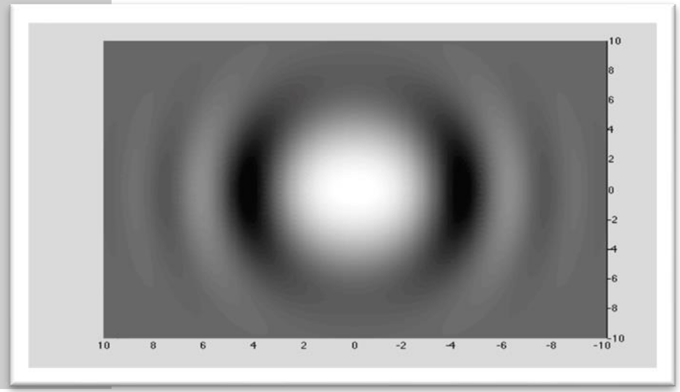
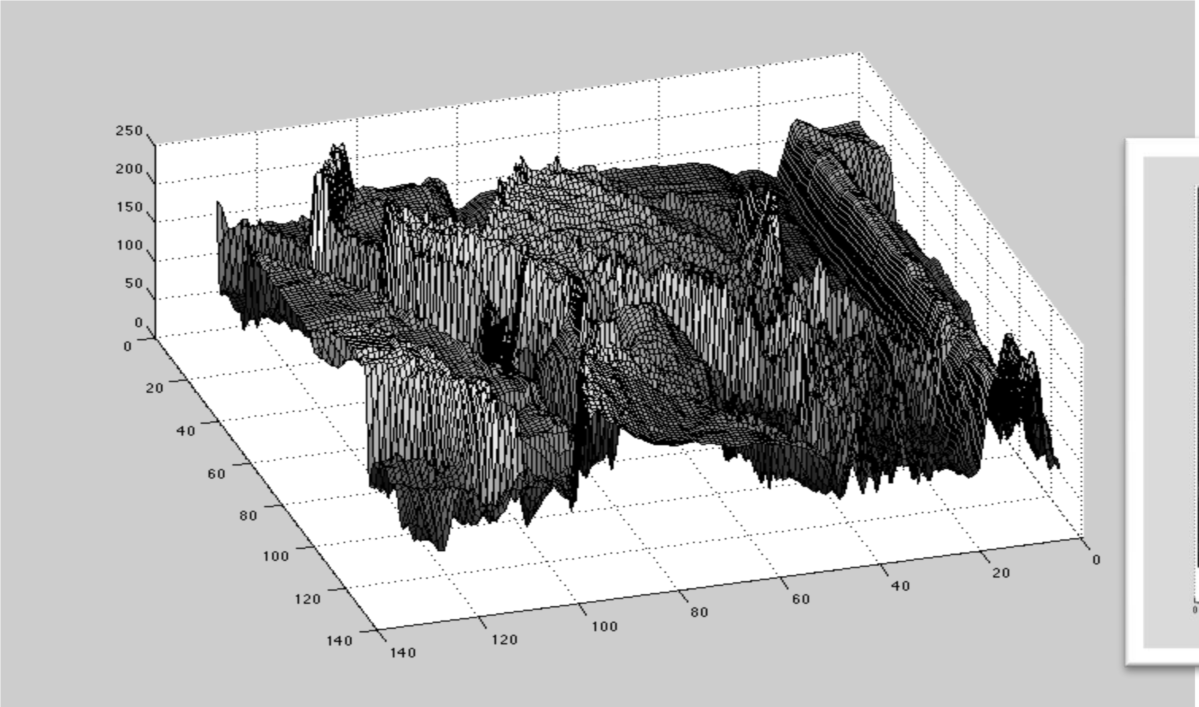


Image Representation

Graph of a photo



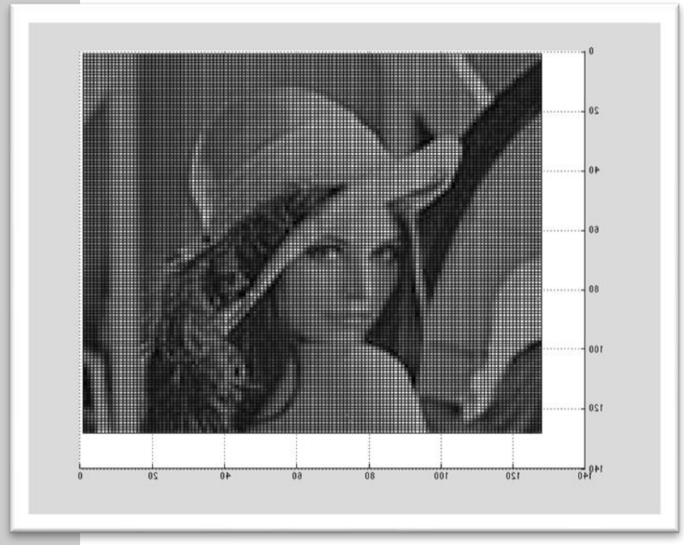


Image Representation

Discrete sampling

- Sampling of a one-dimensional signal

$$S\{f(x)\} = (\dots, f(x-1), f(x), f(x+1), \dots)$$

- Sampling of an image

$$S\{I(x,y)\} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \\ \dots & I(x-1,y-1) & I(x-1,y) & I(x-1,y+1) & \dots \\ \dots & I(x,y-1) & I(x,y) & I(x,y+1) & \dots \\ \dots & I(x+1,y-1) & I(x+1,y) & I(x+1,y+1) & \dots \\ & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Image Representation

Discrete representation / matrix representation

- Assumption: origin is in the top-left corner
- Items of the matrix are $I_{k,l} = S\{I(0,0)\}_{kl}$

Columns 1 through 41

163	163	162	161	160	160	159	159	159	165	169	173	176	170	148	126	103
162	162	161	161	160	159	159	159	160	166	171	173	174	167	147	126	102
161	161	160	160	159	159	159	158	162	168	173	173	171	163	145	126	101
160	160	160	159	159	159	159	159	164	170	174	172	169	161	143	126	99
160	160	160	160	160	160	160	160	167	171	172	169	167	160	143	125	99
160	160	161	161	161	162	162	162	170	170	168	166	166	162	144	124	99
161	161	162	162	163	164	164	165	172	169	163	162	167	164	145	123	100
162	162	162	163	164	165	166	166	174	168	161	160	167	166	146	123	100
158	160	162	164	167	170	171	170	164	167	164	160	164	166	147	121	97
162	162	162	163	167	170	171	170	163	166	164	160	164	165	146	121	97
166	163	162	165	168	169	168	168	161	165	164	160	163	164	145	120	97
167	163	164	169	170	165	161	160	159	165	165	161	163	163	144	120	96
166	164	167	173	170	157	150	151	158	165	167	163	164	163	144	120	94
166	165	170	174	164	145	138	143	157	166	169	165	165	163	145	122	91
169	167	171	171	154	132	129	140	158	168	172	168	167	164	146	123	89
172	169	171	167	146	123	124	140	158	169	173	169	168	165	147	124	87
175	173	171	156	124	101	115	143	163	172	173	167	165	161	143	120	86
171	172	166	144	110	93	113	144	163	171	174	168	165	161	143	120	86
169	167	155	126	94	86	113	144	162	171	174	168	166	162	142	119	87
171	159	137	107	86	89	116	142	161	171	174	169	166	162	142	118	87
170	146	116	95	88	97	120	140	161	171	174	170	167	162	142	117	88
159	129	101	93	95	102	120	140	160	170	175	170	168	163	141	117	88
139	111	93	97	98	97	116	144	159	170	175	171	169	163	141	116	89
123	99	91	101	99	91	111	148	159	170	175	171	169	163	141	116	89
96	99	97	99	101	97	111	141	157	167	171	167	166	162	143	119	90
99	101	100	102	103	99	113	143	157	167	171	167	166	162	143	119	90
101	103	102	103	105	100	114	144	157	167	171	167	166	162	143	119	89
101	103	101	103	104	99	112	142	157	167	171	167	166	162	143	119	89
99	102	99	100	101	96	109	139	157	167	171	167	166	162	143	119	88
99	101	99	100	100	95	108	137	157	167	171	167	166	162	143	119	88
102	104	101	101	102	96	109	138	157	167	171	167	166	162	143	119	87
104	106	103	104	104	98	111	140	157	167	171	167	166	162	143	119	87
103	108	105	103	106	105	115	137	156	166	171	170	172	168	145	118	87
104	109	106	104	107	106	115	138	157	166	172	171	172	168	145	118	87
105	110	107	105	108	107	117	139	157	167	172	171	173	169	146	119	88
106	111	107	106	109	107	117	140	158	168	173	172	174	170	147	120	88
106	111	107	106	109	107	117	140	159	169	174	174	175	171	148	121	89
105	110	107	105	108	107	117	139	160	170	175	175	176	172	149	122	90
104	109	106	104	107	106	115	138	161	171	176	175	177	173	150	123	91
103	108	105	103	106	105	115	137	162	172	177	176	177	173	150	123	91
103	108	104	103	107	107	119	143	162	171	175	173	175	172	150	124	92
104	108	105	103	107	108	120	143	162	171	175	173	175	172	150	124	91
104	109	105	104	108	109	120	144	162	171	176	174	175	172	150	124	91
105	110	106	105	109	110	121	145	163	172	176	174	176	173	151	125	90
106	111	108	106	110	111	122	146	164	173	177	175	176	173	152	125	89
107	112	109	107	111	112	123	147	164	173	177	175	177	174	152	126	88
108	113	109	108	112	112	124	148	164	173	178	176	177	174	152	126	88
109	113	110	108	112	113	125	148	165	174	178	176	178	175	153	126	87
117	114	107	110	115	111	120	144	159	170	177	175	176	173	153	129	91
118	115	108	110	114	109	118	143	159	170	177	175	176	173	153	129	91
119	116	109	110	112	107	115	140	159	170	177	175	176	173	153	129	89
119	117	111	110	111	103	112	138	159	170	177	175	176	173	153	129	89
118	117	111	109	109	101	110	137	159	170	177	175	176	173	153	129	89
116	116	111	109	107	99	109	137	159	170	177	175	176	173	153	129	89
113	114	110	108	106	98	109	138	159	170	177	175	176	173	153	129	91

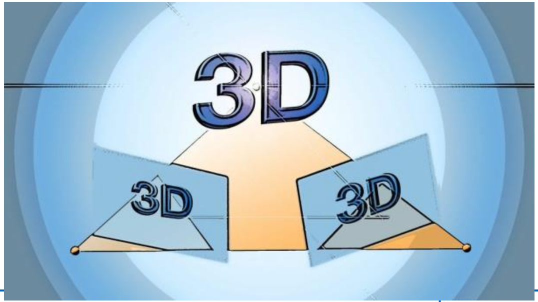
Summary

- Grayscale images
- Images as Matrices
- Images as smooth functions

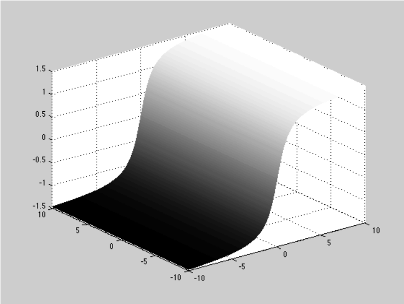
Martin Kleinsteuber: Computer Vision

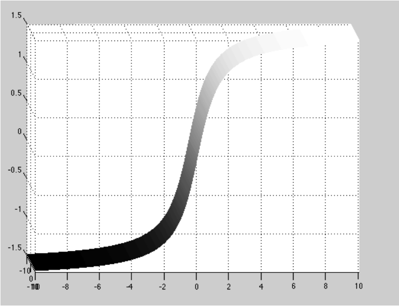
Ch. 1 – Characteristics of Images

2. Image Gradient



The Image Gradient





Edges correspond to stark local changes of the intensity

Local Changes are described by a gradient

$$\nabla I(x,y) = \begin{bmatrix} \frac{d}{dx} I(x,y) \\ \frac{d}{dy} I(x,y) \end{bmatrix}$$

The Gradient of an Image

How to estimate the gradient?

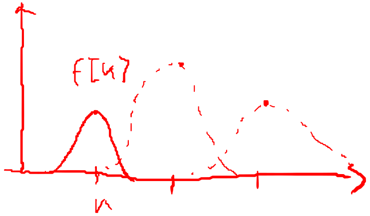
- The discrete Form of the Image is known $I \in \mathbb{R}^{m \times n}$
- Naive approach:

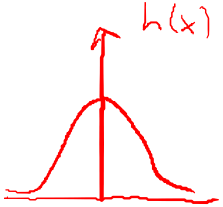
$$\frac{d}{dx} I(x, y) \approx I(x+1, y) - I(x, y)$$
$$\frac{d}{dy} I(x, y) \approx I(x, y+1) - I(x, y)$$

Discrete and Continuous Signals

Interpolation

- From the discrete signal $f[x] = S\{f(x)\}$ to the continuous signal $f(x)$



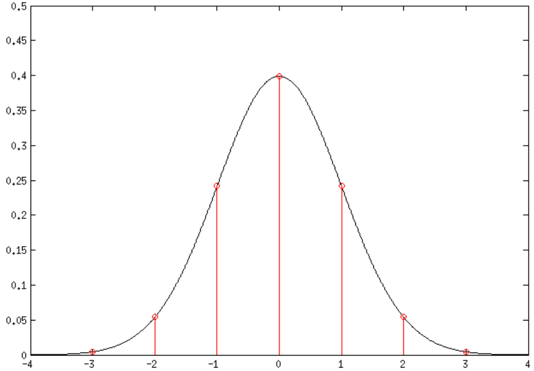


$$f(x) \approx \sum_{k=-\infty}^{\infty} f[k]h(x-k) =: f[x] * h(x)$$

Discrete and Continuous Signals

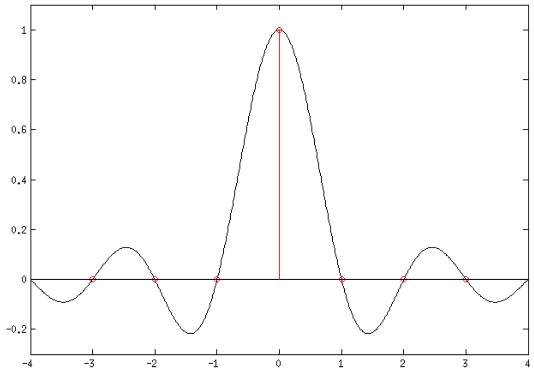
Interpolation filter

- Discrete signal: $f[x] = S\{f(x)\}$
- Continuous signal: $f(x) \approx f[x] * h(x)$
- Gaussian filter: $h(x) = g(x)$



$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}}$$

- Ideal interpolation filter: $h(x) = \mathrm{sinc}(x)$
- Therefore it is valid that $f[x] * h(x) = f(x)$



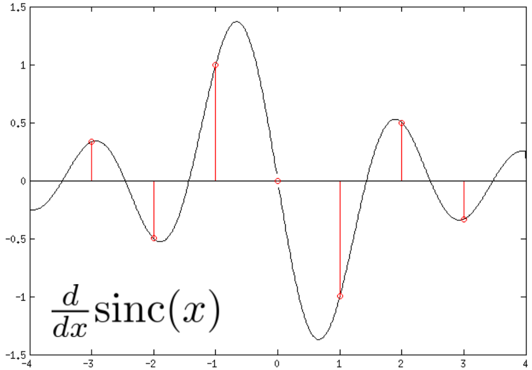
$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

The Discrete Derivative

With the Helpof the reconstructed signal

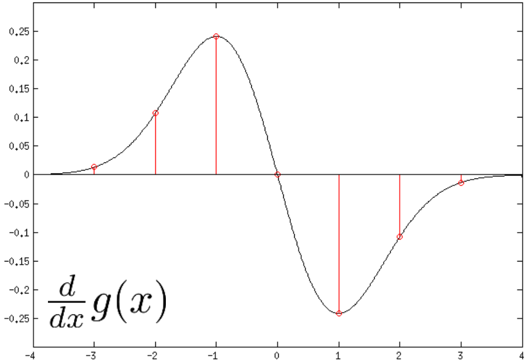
- Algorithmically
  1. Reconstruction of the continuous signal
  2. Differentiation of the continuous signal
  3. Sampling of the Derivative

- Sinc function
  - Slow decay



- Derivation:
  - $$f'(x) \approx \frac{d}{dx}(f[x] * h(x)) = f[x] * h'(x)$$
  - $$f'[x] = f[x] * h'[x] = \sum_k f[x-k]h'[k]$$

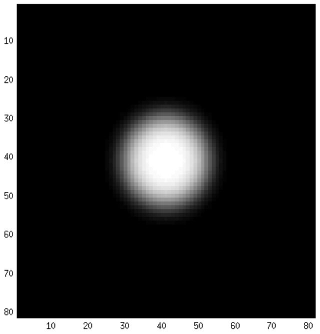
- Gaussian filter
  - Fast decay



2-Dimensional Reconstruction

Separable 2D Gaussian filter

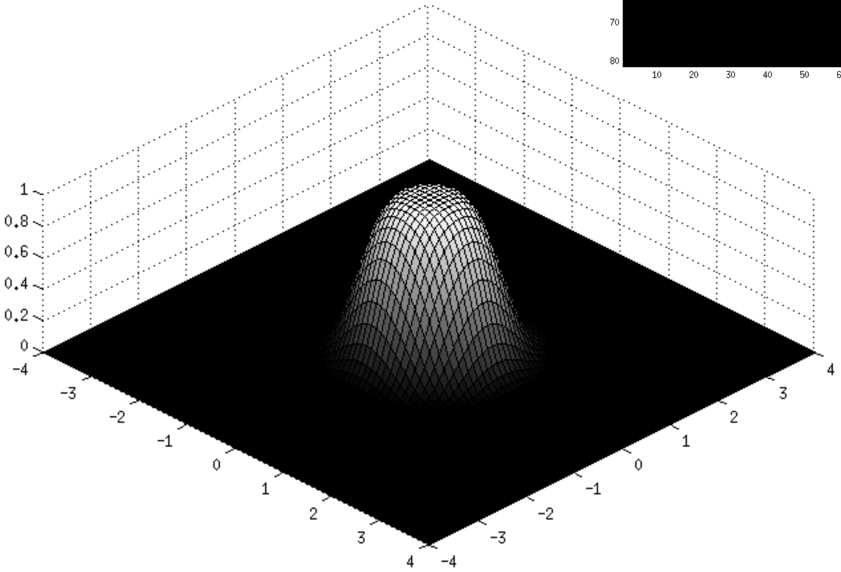
$$h(x,y) := g(x)g(y)$$



- 2D reconstruction:

$$I(x,y) \approx I[x,y] * h(x,y) =$$

$$= \sum_{k,l} I[x,y]g(x-k)g(y-l)$$

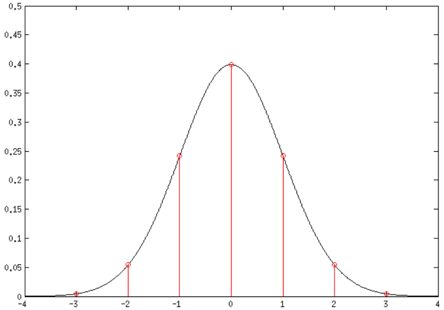


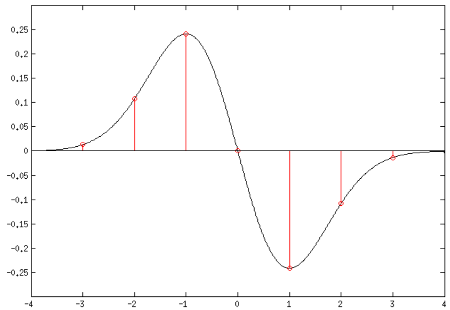
2-Dimenasional Derivative

Exploitation of the Separability

- Derivative in x direction

$$\frac{d}{dx} I(x,y) \approx I[x,y] * \left(\frac{d}{dx} h(x,y)\right)$$
$$= \sum_{k,l} I[k,l] g'(x-k) g(y-l)$$
$$S\{\tfrac{d}{dx} I(x,y)\} = I[x,y] * g'[x] * g[y]$$
$$= \sum_{k,l} I[x-k, y-l] g'[k] g[l]$$





- Finite approximation of the gaussian filter:

$$k, l \in \{-1, 0, 1\}$$
$$g'[0] = 0;\ g'[-1] \approx \tfrac{1}{4};\ g'[1] \approx -\tfrac{1}{4}$$
$$g[0] \approx \tfrac{1}{2};\ g[-1] \approx \tfrac{1}{4};\ g[1] \approx \tfrac{1}{4}$$

- Optain the discrete filter:

$$\Big(g'[k]g[l]\Big)_{k=-1,0,1;l=-1,0,1} = 0.6 \begin{bmatrix} 0.6 & 0 & -0.6 \\ 1 & 0 & -1 \\ 0.6 & 0 & -0.6 \end{bmatrix}$$

Example

Sobel Filtering

- Due to practical reasons, integer coefficients are desirable
- A multiple of the gradient is sufficient to detect the intensity differences

$$0.6 \begin{bmatrix} 0.6 & 0 & -0.6 \\ 1 & 0 & -1 \\ 0.6 & 0 & -0.6 \end{bmatrix}$$

Integer approximation of the double gradient

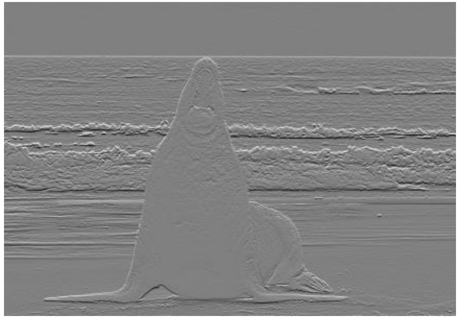
1	0	-1
2	0	-2
1	0	-1

Horizontales Sobel-Filter

Example

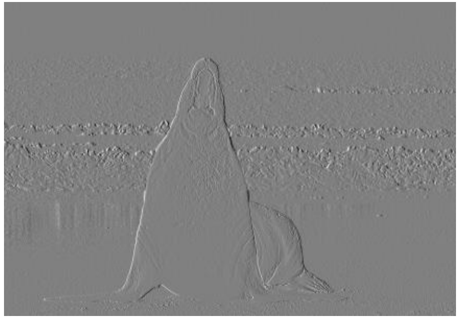
Sobel filtering

vertical



1	2	1
0	0	0
-1	-2	-1

horizontal



1	0	-1
2	0	-2
1	0	-1

Summary

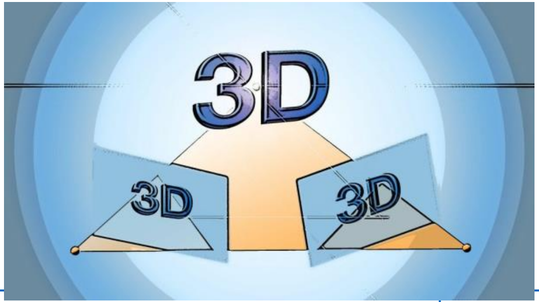
Image Gradient

- The image gradient is an important tool to ddetermine local intensity changes
- The discrete derivative is calculated by differentiating the interpolated signal
- Separable 2D filters allow for efficient calculation of the image gradient
- Sobel filters are integer approximations of the double gradient

Martin Kleinsteuber: Computer Vision

Ch. 1 – Characteristics of Images

3. Feature points – Corners and Edges



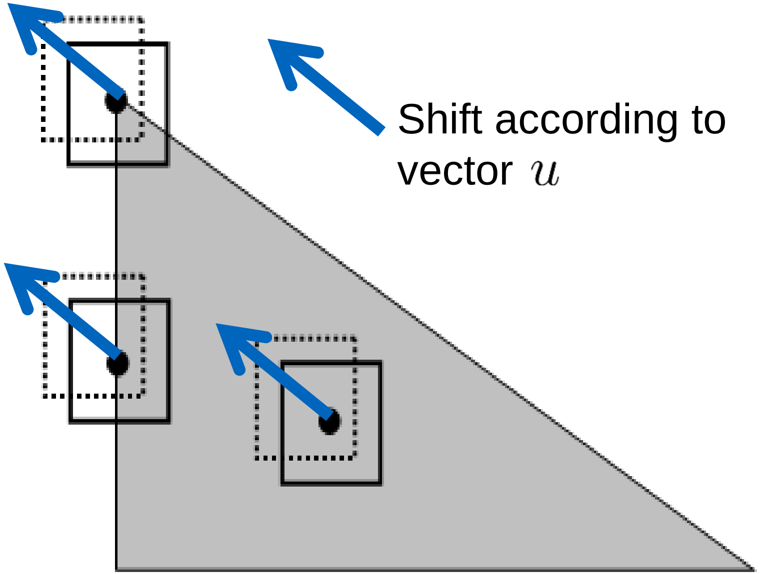
Corners and Edges...

...deliver prominent image features

- Determination of outlines
- Calculation of movements in sequences of images
- Estimation of camera movements
- Image registration
- 3D reconstruction

Harris Corner and Edge Detector

Change in an image section as a dependency of the shift



- Corner: shifts in all directions cause a change
- Edge: shifts in all directions except exactly one cause a change
- Homogenous surface: no change, independently of the direction

Harris Corner and Edge detector

Formal Definition of Change

- Position in the image: $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad I(x) = I(x_1, x_2)$

- Shift direction: $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$

- Change in the image segment:

$$S(u) = \int_W \Big(I(x+u) - I(x) \Big)^2 dx$$

- Differentiability of I:

$$\lim_{u \to 0} \frac{I(x+u) - I(x) - \nabla I(x)^\top u}{\|u\|} = 0$$

Harris Corner and Edge Detector

Approximation of the Change

- Because of differentiability: $I(x+u) - I(x) = \nabla I(x)^\top u + o(\|u\|)$
- Residual term $o(\|u\|)$ with the characteristic $\lim\limits_{u \to 0} o(\|u\|)/\|u\| = 0$

- Approximation for small shifts: $I(x+u) - I(x) \approx \nabla I(x)^\top u$

- Approximation of the change in the image section:

$$S(u) = \int_W \Big(I(x+u) - I(x)\Big)^2 dx \approx \int_W \Big(\nabla I(x)^\top u\Big)^2 dx$$

Harris Corner and Edge Detector

The Harris Matrix

- Expanding the integral:

$$\int_W \left(\nabla I(x)^\top u\right)^2 dx = u^\top \Big(\int_W \nabla I(x)\nabla I(x)^\top dx\Big) u$$

- Harris matrix: $G(x) = \int_W \nabla I(x)\nabla I(x)^\top dx$

$$\nabla I(x)\nabla I(x)^\top = \begin{bmatrix} (\frac{\partial}{\partial x_1} I(x))^2 & \frac{\partial}{\partial x_1} I(x) \frac{\partial}{\partial x_2} I(x) \\ \frac{\partial}{\partial x_2} I(x) \frac{\partial}{\partial x_1} I(x) & (\frac{\partial}{\partial x_2} I(x))^2 \end{bmatrix}$$

- Approximative change of the image section:

$$S(u) \approx u^\top G(x) u$$

Excursus: Linear Algebra

Positive definite and positiv semi-definite Matrices

- Definition. A real-valued symmetrical Matrix $A = A^\top$ is
  - positive definite, if $x^\top A x > 0, \quad x \neq 0$
  - positive semi-definite, if $x^\top A x \geq 0$
- Examples
  - The zero-matrix is positive semi-definite, although not positive definite.
  - The identity matrix is positive definite.
  - All positive definite matrices are also positive semi-definite.
  - $G(x)$ is positive semi-definite. Why?

Excursus: Linear Algebra

Positive definite and positiv semi-definite Matrices

- Theorem. For real-valued symmetrical matrices $A = A^\top$ it is equal to say:
  - A positive (semi-) definite
  - All eigenvalues of A are greater than zero (greater than or equal to zero)

- Eigenvalue decomposition of real-valued symmetrical Matrices.

  All real-valued symmetrical matrices $A = A^\top$ can be decomposed into a product $A = V \Lambda V^\top$ with $VV^\top = I$ and a diagonal matrix Λ, wherein the eigenvalues of A lie on the diagonal. The columns of V are the corresponding eigenvectors.

Harris Corner and Edge Detector

Eigenvalue decomposition

- Eigenvalue decomposition of the Harris matrix:

$$G(x) = \int_W \nabla I(x) \nabla I(x)^\top dx = V \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} V^\top$$

with $VV^\top = I_2$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq 0.$

- Change is dependent on the eigenvectors: $V = [v_1, v_2]$

$$S(u) \approx u^\top G(x) u = \lambda_1 (u^\top v_1)^2 + \lambda_2 (u^\top v_2)^2$$

Harris Corner and Edge Detector

Type of Feature based on eigenvalues

$$S(u) \approx u^\top G(x)u = \lambda_1(u^\top v_1)^2 + \lambda_2(u^\top v_2)^2$$

- Both eigenvalues positive
  - $S(u) > 0$ for all u (change in every direction)
  - Examinated image section contains a corner
- One eigenvalue positive, one eigenvalue equal to zero
  - $S(u) \begin{cases} = 0, & \text{falls } u = rv_2 \\ > 0, & \text{sonst} \end{cases}$ (no change, only in the direction of the eigenvector with eigenvalue 0)
  - Examinated image section contains an edge
- Both eigenvalues equal to 0
  - $S(u) = 0$ for all u (no change in any direction)
  - Examinated image section contains a homogenous surface

Falls = if
Sonst = else

Practical Implementation of the Harris detector

Computation of the Harris matrix

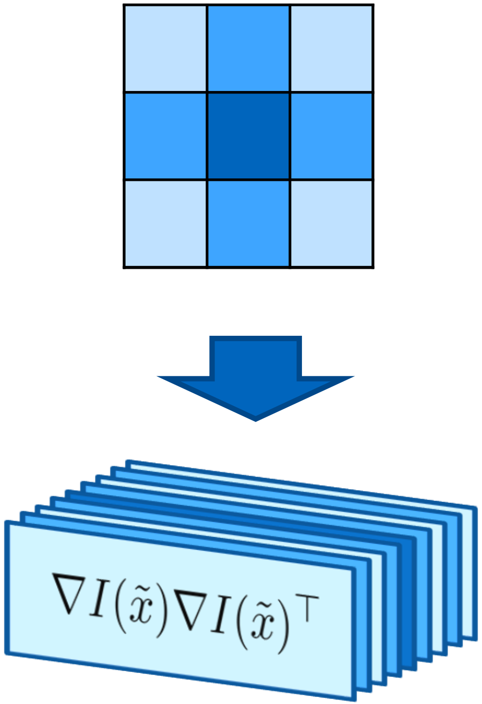
- Approximate $G(x)$ with a finite sum

$$G(x) = \int_W \nabla I(x) \nabla I(x)^\top dx \approx \sum_{\tilde{x} \in W(x)} \nabla I(\tilde{x}) \nabla I(\tilde{x})^\top$$

- Weighted Sum dependent on the position of $\tilde{x}$

$$G(x) \approx \sum_{\tilde{x} \in W(x)} w(\tilde{x}) \nabla I(\tilde{x}) \nabla I(\tilde{x})^\top$$

- Weights $\omega(\tilde{x}) > 0$ emphasize the influence of the central pixel



Practical Implementation of the Harris detector

Eigenvalues

- In the real world, eigenvalues are never exactly 0, e.g. due to noise, discrete sampling or numerical imprecisions.
- Practical Characteristics
  - Corner: two large eigenvalues
  - Edge, a large eigenvalue and a small eigenvalue
  - Homogeneous surface: two small eigenvalues
- Decision through empirical thresholds

Excursus: Linear Algebra

Eigenvalues, Determinant, Trace

- Correlation of eigenvalues, determinant and trace of a matrix.

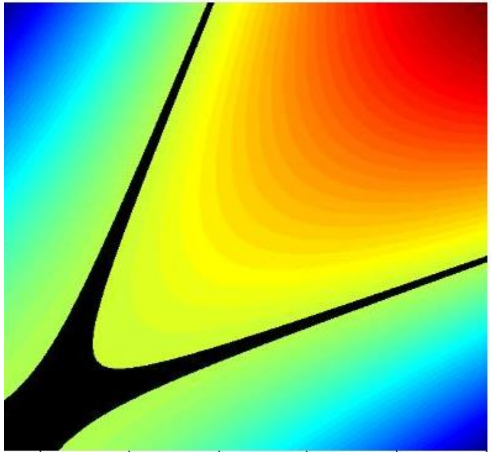
  Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of a matrix G .

  It follows,

  - $\det G = \prod_{i=1}^{n} \lambda_i$ (The determinant is the product of the eigenvalues)
  - $\operatorname{tr} G = \sum_{i=1}^{n} \lambda_i$ (The trace is the sum of the eigenvalues)

Practical Implementation of the Harris Detector

A simple criterium for corners and edges



- Analyze the quantity $H := \det(G) - k(\mathrm{tr}(G))^2$

$$H = (1-2k)\lambda_1\lambda_2 - k(\lambda_1^2 + \lambda_2^2)$$

- Corner (both eigenvalues large)
  - H larger than a positive threshold value
- Edge (one large and one small eigenvalue)
  - H smaller than a negative threshold value
- Homogeneous surface (both eigenvalues small)
  - H small (absolute value)

Summary

Harris Detector foridentifying feature points

- Evaluation of the (approximated) Harris matrix

$$G(x) \approx \sum_{\tilde{x} \in W(x)} w(\tilde{x}) \nabla I(\tilde{x}) \nabla I(\tilde{x})^\top$$

- The eigenvalue decomposition of $G(x)$ delivers informations on the corners and the edges
- Efficient implementation with the help of:

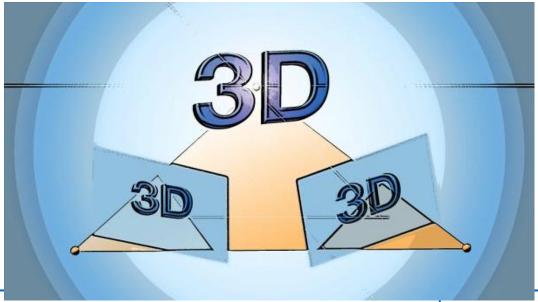
$$H := \det(G) - k(\mathrm{tr}(G))^2$$

- Decision through threshold values:
  - Corner: $0 < \tau_+ < H$
  - Edge: $H < \tau_- < 0$
  - Homogeneous surface: $\tau_- < H < \tau_+$

Martin Kleinsteuber: Computer Vision

Ch. 1 – Characteristics of Images

4. Correspondence Estimation of Feature Points



Correspondence Estimation

Problem

- Two images $I_1 : \Omega_1 \to \mathbb{R}, I_2 : \Omega_2 \to \mathbb{R}$ from the same 3D scene are known
- Find pairs of Image points $(x^{(i)}, y^{(i)}) \in \Omega_1 \times \Omega_2$ which correspond to the same 3D points.

Correspondence Estimation

Problem

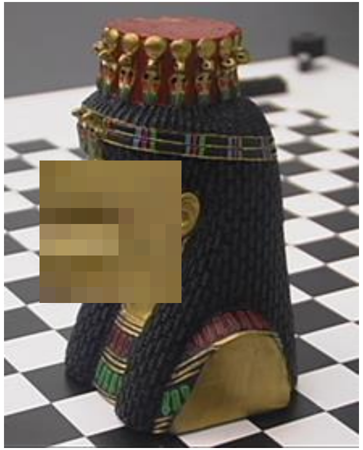
- In this session: correspondence of feature points in I_1 and I_2
- Feature points $\{x_1, \ldots x_n\} \subset \Omega_1$ and $\{y_1, \ldots y_n\} \subset \Omega_2$ are known
- Find compatible pairs of feature points

Naive Solution to the Problem

Sum of squared differences (SSD)

- Examine image sections V_i around x_i and W_i around y_i in matrix representation and compare the respective intensity values

d(,)



Sum of Squared Differences (SSD)

Formal Definition

- A criterion: $d(V,W) = \|V - W\|_F^2$
- Whereby $\|A\|_F^2 = \sum_{kl} A_{kl}^2$ describes the quadratic Frobenius norm

- Find a suitable W_j to a V_i with $j = \arg\min_{k=1,\ldots,n} d(V_i, W_k)$
- Assumption: If W_j suits V_i, then the opposite is also true

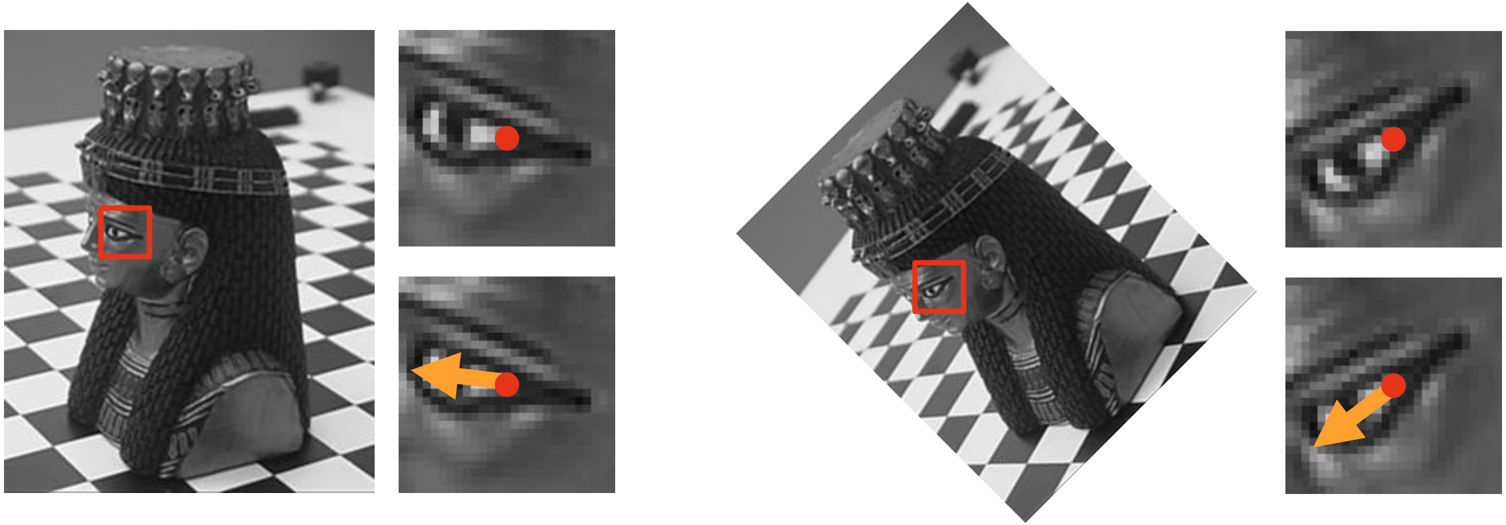
Weak Points of the SSD Method

Change in Illumination or Rotation

- The normalization of intensity and rotation is necessary!

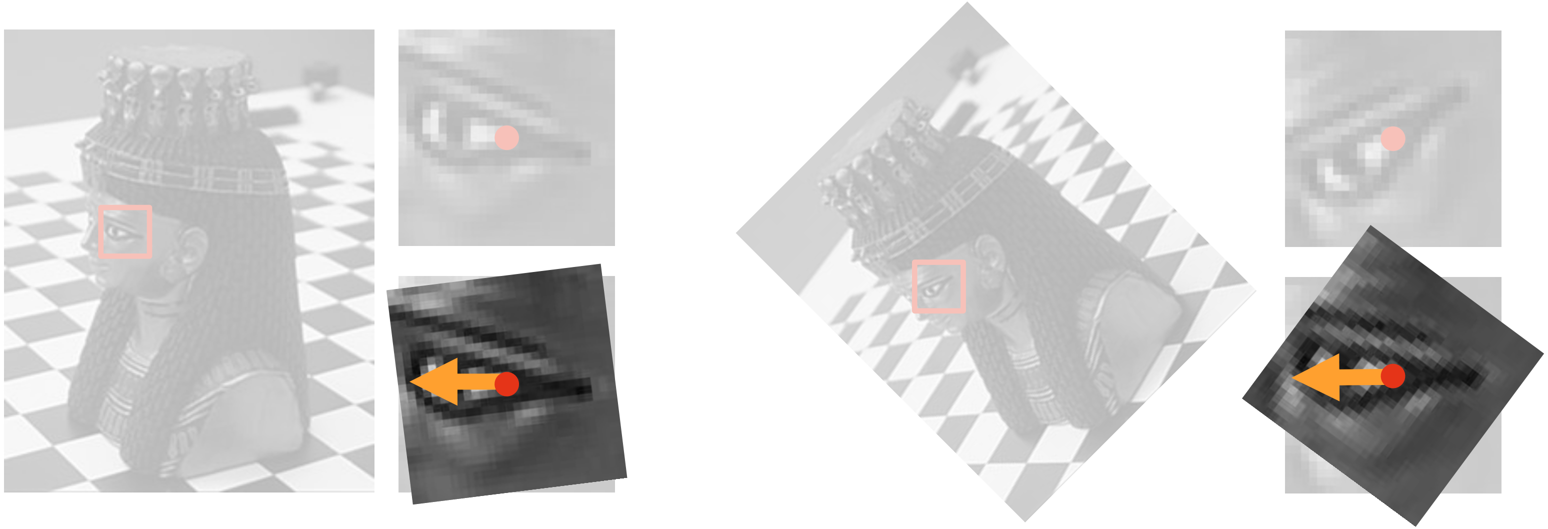
Rotation Normalization

By means of the Gradient Direction



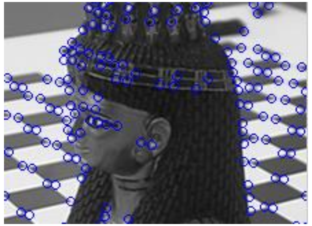
Rotation Normalization

By means of the Gradient Direction



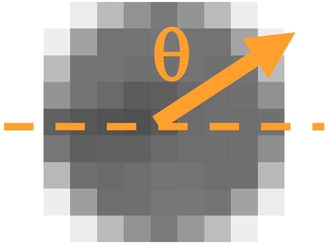
Rotation Normalization

By means of the Gradient Direction

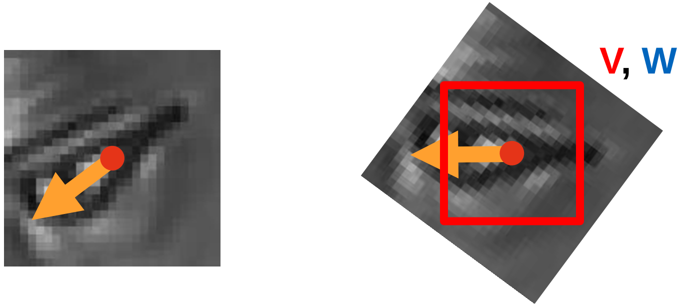


- Pre-processing:

  1. Determine the gradient in all feature points
  2. Rotate the regions around feature points such that the gradient points in one direction.
  3. Extrapolate V, W from the rotated regions



$$\theta = \arctan\left(\frac{\frac{\partial I}{\partial x_2}}{\frac{\partial I}{\partial x_1}}\right)$$



Bias-and-Gain-Model

Modelling of Contrast and Brightness

- Scaling of the intensity values (Gain) with α
- Shift of the intensity values (Bias) with β

Skalierung bewirkt Kontraständerung

Verschiebung bewirkt Helligkeitsänderung

- Gain-Model: $W \approx \alpha V$
- Bias-Model: $W \approx V + \beta \mathbb{1}\mathbb{1}^\top$

$$\mathbb{1} = (1, \ldots 1)^\top$$

- Bias-and-Gain Model: $W \approx \alpha V + \beta \, \mathbb{1}\mathbb{1}^\top$

Bias-and-Gain-Model

Calculating the Mean

- Calculate the intensity mean

$$\begin{aligned}\overline{W} &= \tfrac{1}{N}\left(\mathbb{1}\mathbb{1}^\top W \mathbb{1}\mathbb{1}^\top\right)\\ &\approx \tfrac{1}{N}\left(\mathbb{1}\mathbb{1}^\top\left(\alpha V + \beta\mathbb{1}\mathbb{1}^\top\right)\mathbb{1}\mathbb{1}^\top\right)\\ &= \alpha\tfrac{1}{N}\left(\mathbb{1}\mathbb{1}^\top V \mathbb{1}\mathbb{1}^\top\right) + \beta\mathbb{1}\mathbb{1}^\top\\ &= \alpha\overline{V} + \beta\mathbb{1}\mathbb{1}^\top\end{aligned}$$

- Subtract the mean-matrix

$$\begin{aligned}W - \overline{W} &\approx \alpha V + \beta\mathbb{1}\mathbb{1}^\top - \left(\alpha\overline{V} + \beta\mathbb{1}\mathbb{1}^\top\right)\\ &= \alpha\left(V - \overline{V}\right)\end{aligned}$$

Bias-and-Gain-Model

Calculating the Standard Deviation

- Standard deviation of the intensity

$$
\begin{aligned}
\sigma(W) &= \sqrt{\tfrac{1}{N-1}\|W - \overline{W}\|_F^2} \\
&= \sqrt{\tfrac{1}{N-1}\ \operatorname{tr}\left(\left(W - \overline{W}\right)^\top \left(W - \overline{W}\right)\right)} \\
&\approx \sqrt{\tfrac{1}{N-1}\ \operatorname{tr}\left(\alpha\left(V - \overline{V}\right)^\top \alpha\left(V - \overline{V}\right)\right)} \\
&= \alpha\ \sigma(V)
\end{aligned}
$$

Bias-and-Gain-Model

Compensation of Bias and Gain

- Normalization of the image sections through
  1. Subtraction of the mean
  2. Division by the standard deviation

$$
\begin{aligned}
W_n :=& \frac{1}{\sigma(W)}(W - \overline{W}) \\
\approx& \frac{1}{\alpha\ \sigma(V)}\left(\alpha\left(V - \overline{V}\right)\right) \\
=& \frac{1}{\sigma(V)}(V - \overline{V}) \\
=:& V_n
\end{aligned}
$$

Normalized Cross Correlation (NCC)

Derivation from SSD

- SSD of two normalized image sections

  $$\|V_n - W_n\|_F^2 = 2(N-1) - 2\mathrm{tr}(W_n^\top V_n)$$

- The Normalized Cross Correlation of th two image sections is defined as $\frac{1}{N-1}\mathrm{tr}(W_n^\top V_n)$
- It holds: $-1 \leq \mathrm{NCC} \leq 1$
- Two normalized image sections are similar, if
  - SSD small (few differences)
  - NCC close to +1 (high correlation)

$$\|V_n - W_n\|_F^2 = \underbrace{\overset{N-1}{tr(V_n^T V_n)}} + \overset{N-1}{tr(W_n^T W_n)} - 2 \cdot tr(W_n^T V_n)$$

$$V_n = \frac{V - \bar{V}}{\sqrt{\frac{1}{N-1} \cdot \|V - \bar{V}\|_F^2}} \Rightarrow tr(V_n^T V_n) = \frac{\|V - \bar{V}\|^2}{\frac{1}{N-1} \cdot \|V - \bar{V}\|^2} = N-1$$

$$\frac{1}{N-1} \cdot \|V_n - W_n\|_F^2 = 2 - 2 \cdot \underbrace{\frac{1}{N-1} \cdot tr(W_n^T V_n)}_{NCC}$$

Summary

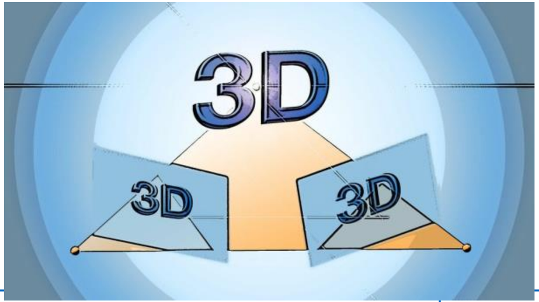
Correspondence Estimation of Feature Points

- Find feature points in image 1 and image 2
- Compensate the rotation by orienting the gradient.
- Extraxt image sections arounf each feature point
- Compensate the illumination by normalizing the image sections
- Compare the normalized image sections
  by SSD or NCC

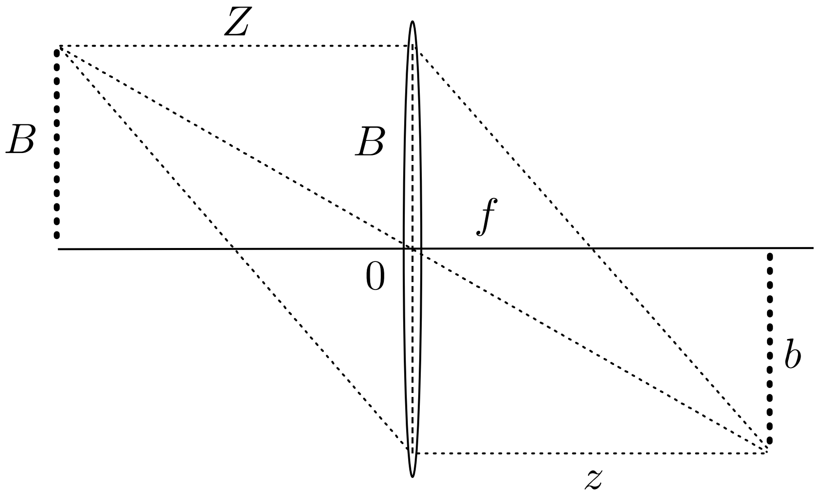
Martin Kleinsteuber: Computer Vision

Ch. 2 – Image Formation

1. The Pinhole Camera Model



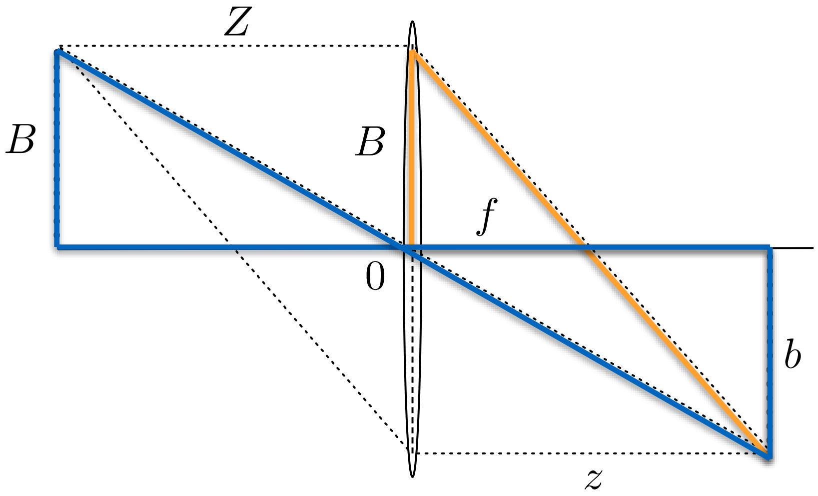
Projection through a Thin Lens



- Light rays parallel to the optical axis are refracted in such a way, that they pass through the focal plane
- Light rays passing through the optical centre are not refracted

Projection through a Thin Lens

Derivation of the Lens Equation

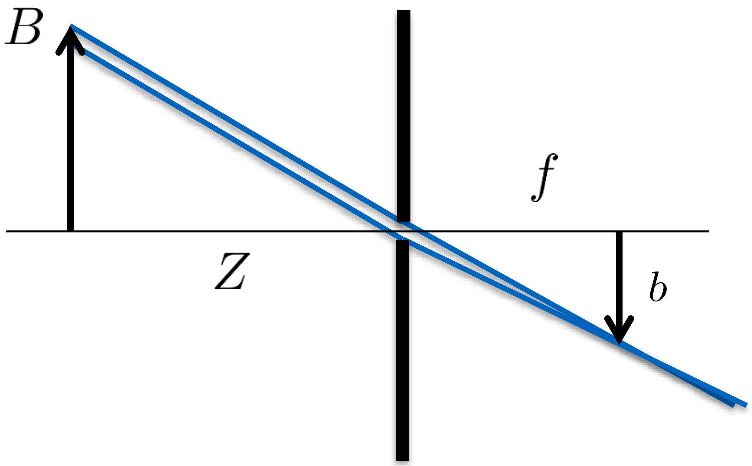


- $$\left|\frac{b}{B}\right| = \frac{z-f}{f}.$$
- $$\left|\frac{z}{Z}\right| = \left|\frac{b}{B}\right|$$
- Equation for thin lenses

$$\frac{1}{|Z|} + \frac{1}{z} = \frac{1}{f}$$

Projection through a Pinhole Camera

Ideal Assumptions

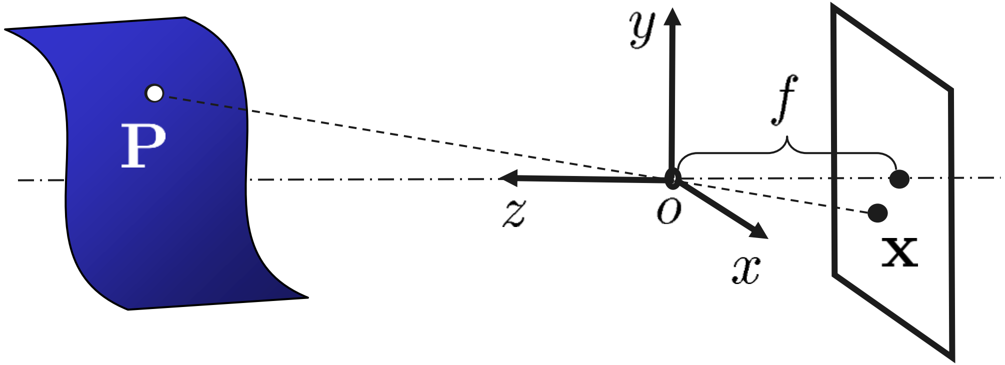


- Very small opening in front of the lens
- Arbitrarily large viewing angle
- Image projects sharply onto the focal plane
- It holds:

$$b = -\frac{fB}{Z}$$

Projection through a Pinhole Camera

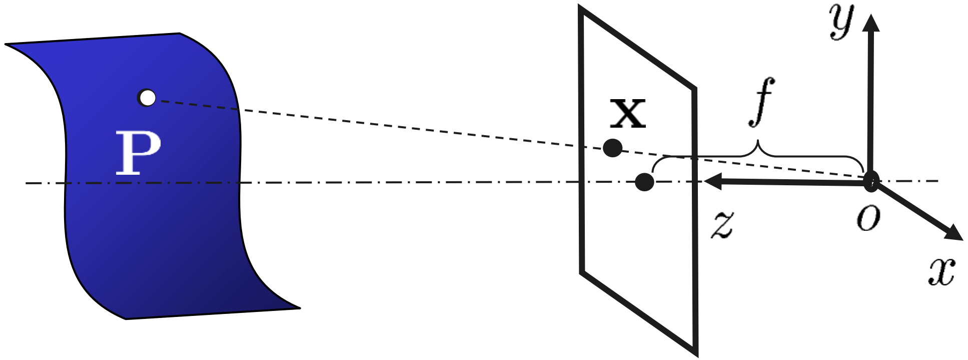
Projection of a Point in Space onto the Focal Plane



- Let the coordinates of P w.r.t the optical centre be (X, Y, Z)
- It follows that the coordinates of the image point are $\left(-\frac{fX}{Z}, -\frac{fY}{Z}, -f\right)$

Frontal Pinhole Camera Model

Projection of a point



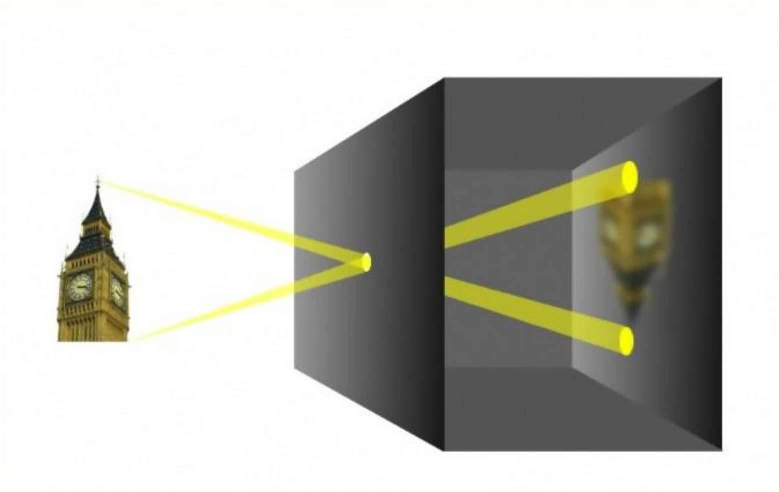
- The coordinates of the image point in the frontal pinhole camera model are $\left(\frac{fX}{Z}, \frac{fY}{Z}, f\right)$
- Ideal perspective projection:

$$\pi : \mathbb{R}^3 \setminus \{(X,Y)\text{-Ebene}\} \to \mathbb{R}^2, \quad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \mapsto \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

Ebene = Plane

Projection through a Pinhole Camera

Ideal Assumptions



Source: Wikipedia.de, 2013

- Thin Lens
- Small aperture
- Arbitrarily big viewing angle
- Perspective Projection:

$$\pi : \mathbb{R}^3 \setminus \{(X,Y)\text{-Ebene}\} \to \mathbb{R}^2, \qquad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \mapsto \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

Ebene = Plane

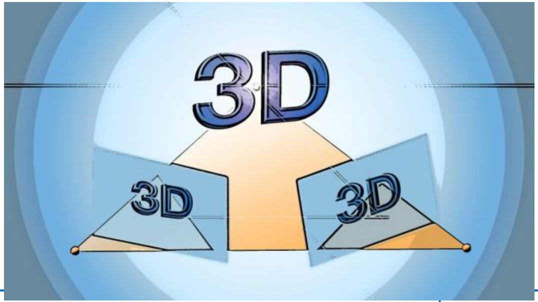
Summary

- Pinhole camera model
- Ideal perspective projection: Mapping of the coordinates of the 3D point onto the 2D coordinates in the focal plane

Martin Kleinsteuber: Computer Vision

Ch. 2 – Image Formation

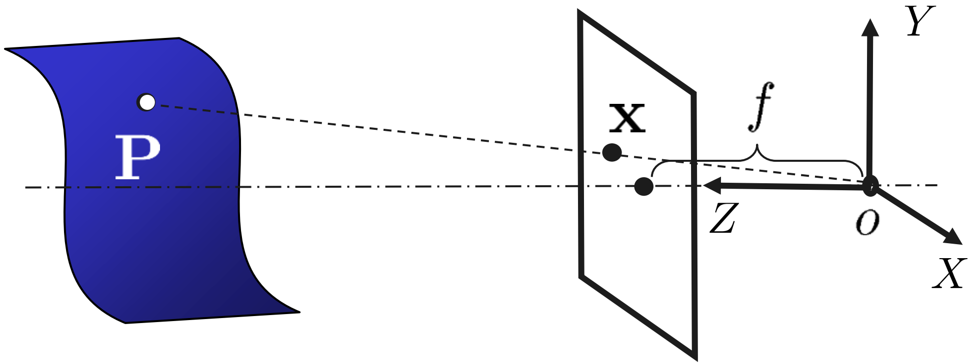
2. Homogeneous Coordinates



Recap: The Pinhole Camera System

Bildpunkte und Geraden

- All Points on a line going through the optical centre are projected on the same image point
- Inversely, each image point corresponds to exactly one line



The Projective Space

- Two Vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are called **equivalent** to one another, if there exists a $\lambda \neq 0$ with $\mathbf{x} = \lambda \mathbf{y}$. In this case we write

$$\mathbf{x} \sim \mathbf{y}$$

- A line through $\mathbf{x} \neq 0$ can only be described as an **equivalence class**

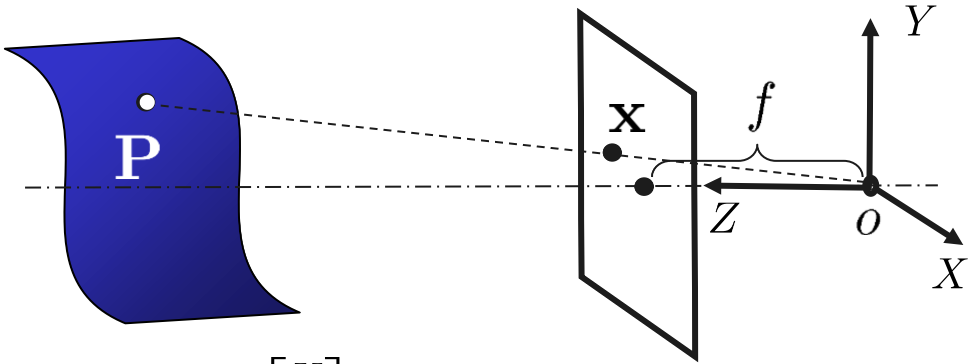
$$[\mathbf{x}] := \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} \sim \mathbf{x}\}$$

- The set of all lines in $\mathbb{R}^{n+1}$ is called **projective space**

$$\mathbb{P}_n = \{[\mathbf{x}] \mid \mathbf{x} \in \mathbb{R}^{n+1} \setminus \{0\}\}$$

Homogeneous Coordinates

- If we set the focal length as the length unit, the image plane can be defined as



$$BE = \left\{ \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \in \mathbb{R}^3 \mid Z = 1 \right\}$$

- The vector $\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$ describes the **homogeneous coordinates** of $\begin{bmatrix} X \\ Y \end{bmatrix}$

Homogeneous Coordinates

- General definition: $\mathbf{x} := (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$

  Then

  $$\mathbf{x}^{(\mathrm{hom})} := (X_1, \ldots, X_n, 1)^\top \in \mathbb{R}^{n+1}$$

  is the homogeneous coordinates of $\mathbf{x}$

Summary

- Equivalent vectors differ from one another only by means of the multiplication with a non-zero value.

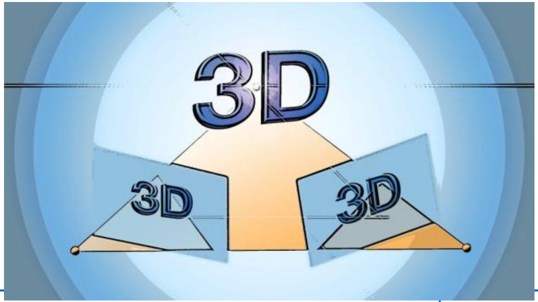
- The homogeneous coordinates of a vector can be obtained by expanding it with a further coordinate containing the value „1“ .

Lehrstuhl für Datenverarbeitung
Fakultät für Elektrotechnik und Informationstechnik
Technische Universität München

Martin Kleinsteuber: Computer Vision

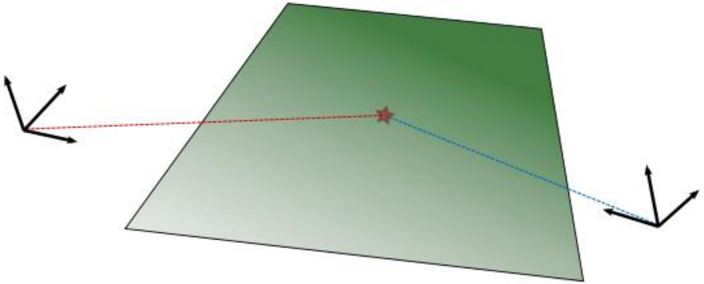
Ch. 2 – Image Formation

3. Euclidian Motion



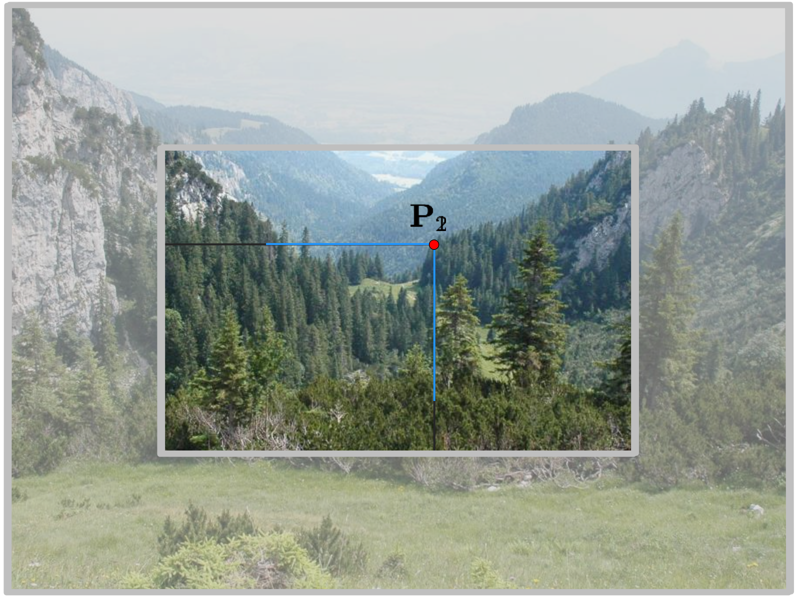
Motivation

- Two images, belonging to the same 3D scene, taken from different positions.
- Position changes comprise a rotation and a subsequent translation of the camera
- The movement of the camera is described by a coordinate change of a fix point in space



Video

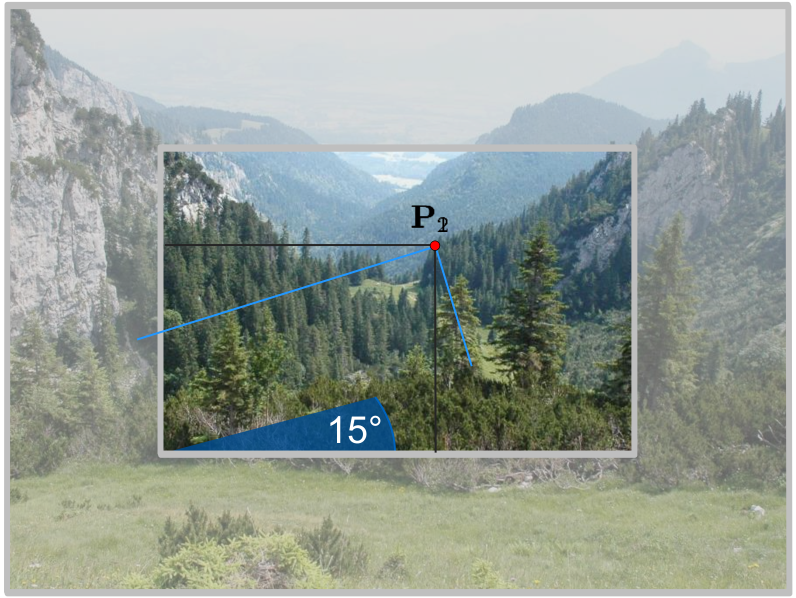
Visualization
Translation



Video

Visualization

Rotation



Video

Euclidian Motion

In General

- The Matrices $O(n) := \{O \in \mathbb{R}^{n \times n} \mid O^\top O = I_n\}$ are called orthogonal matrices
- Rotations in $\mathbb{R}^n$ are described by the special orthogonal matrices

$$SO(n) := \{R \in \mathbb{R}^{n \times n} \mid R^\top R = I_n, \det(R) = 1\}$$

- Euclidian motion by change in coordinates

$$g_{R,T} \colon \mathbb{R}^n \to \mathbb{R}^n, \qquad \mathbf{P} \mapsto R\mathbf{P} + T$$

- Let $\mathbf{P}_1$ be the coordinates of a point w.r.t CF1 and $\mathbf{P}_2$ be the coordinates of the same point w.r.t the moved CF , then it holds that $\mathbf{P}_2 = R\mathbf{P}_1 + T$

Euclidian Motion

In homogeneous coordinates

- In homogeneous coordinates, as matrix-vector-multiplication

$$M = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \quad , R \in SO(n)\,, T \in \mathbb{R}^n$$

$$\mathbf{P}_2^{(\mathrm{hom})} = M\,\mathbf{P}_1^{(\mathrm{hom})}$$

- Special euclidian group

$$SE(n) = \left\{ M = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \middle| R \in SO(n)\,, T \in \mathbb{R}^n \right\} \subset \mathbb{R}^{(n+1)\times(n+1)}$$

Video

Euclidian Motion

Characteristics

- The combination of two euclidian motions is itself an euclidian motion

$$M_1 M_2 = \begin{bmatrix} R_1 & T_1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_2 & T_2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_1 R_2 & R_1 T_2 + T_1 \\ 0 & 1 \end{bmatrix}$$

- Euclidian motions are invertible:

$$M^{-1} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} R^\top & -R^\top T \\ 0 & 1 \end{bmatrix}$$

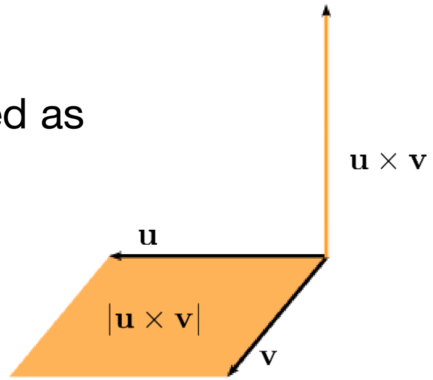
Video

The Cross Product

Definition

- The cross product between vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ is defined as

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} \in \mathbb{R}^3$$



- It holds: $\langle \mathbf{u} \times \mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{u} \times \mathbf{v}, \mathbf{v} \rangle = 0$ and $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$

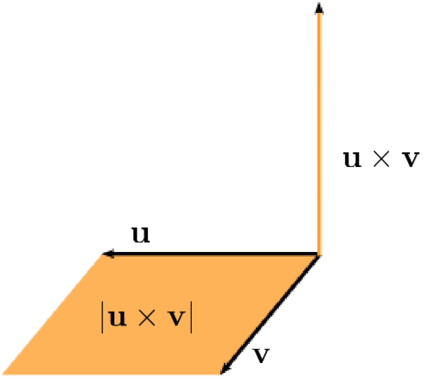
Video

The Cross Product

Matrix-Vector-Multiplication

- The cross product can be written as

$$\mathbf{u} \times \mathbf{v} = \hat{\mathbf{u}}\mathbf{v}, \quad \text{mit} \quad \hat{\mathbf{u}} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$



- It holds (for invertible A):

$$\widehat{A\mathbf{v}} = \det(A) A^{-\top} \hat{\mathbf{v}} A^{-1}$$

Video

Euclidian Motion

Characteristics of Orthogonal Transformations

- Orthogonal transformations preserve the properties of the scalar product $\langle \mathbf{u}, \mathbf{v} \rangle = \langle O\mathbf{u}, O\mathbf{v} \rangle$
- Euclidian transformations preserve the distance
- Special orthogonal transformations preserve the properties of the cross product

$$R\mathbf{u} \times R\mathbf{v} = R\left(\mathbf{u} \times \mathbf{v}\right)$$

Video

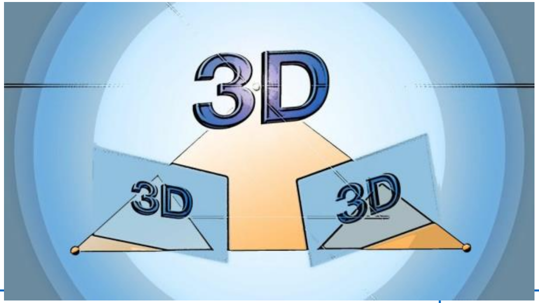
Summary

- Camera movements are described as a coordinate change of a fix point in space
- Coordinate changes are described as euclidian motions
- Rotationen by means of special orthogonal matrices
- Euclidian motion in homogeneous coordinates by means of matrix-vector-multiplication

Martin Kleinsteuber: Computer Vision

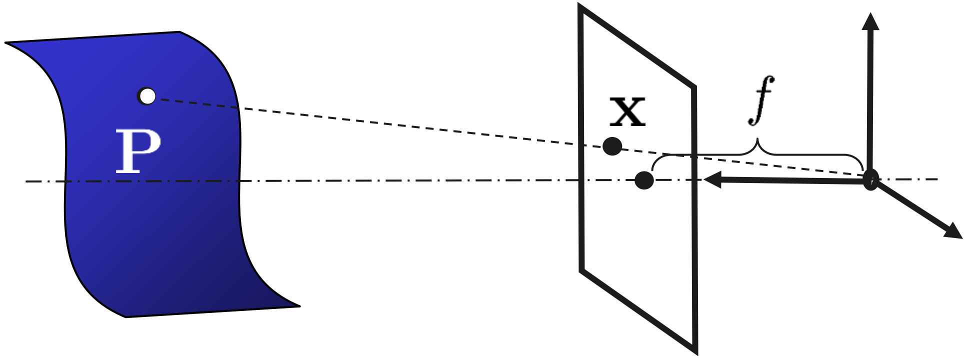
Ch. 2 – Image Formation

4. Perspective Projection with Calibrated Camera



Perspective Projection

Pinhole Camera Model



- Projection of a point onto the image point $\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$

Video

Perspective Projection

- The dependence of homogeneous coordinates is obtained by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{Z} \begin{bmatrix} X \\ Y \end{bmatrix}$$

$$Z \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f/Z \, X \\ f/Z \, Y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix}$$

$$= \frac{1}{Z} \cdot \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$= \frac{1}{Z} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Video

Perspective Projection

By the Focal Length Matrix and the Generic Projection Matrix

$x^{(hom)} = \frac{1}{z} \cdot K_f \cdot \Pi_0 \cdot p^{(hom)}$

- Define $K_f := \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Pi_0 := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
- Consider the 3D points $\mathbf{P}^{(\mathrm{hom})} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$ and $\mathbf{x}^{(\mathrm{hom})} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$
- The perspective projection is therefore

$$\mathbf{x}^{(\mathrm{hom})} \sim K_f \Pi_0 \mathbf{P}^{(\mathrm{hom})}$$

Video

The Ideal Camera

Mapping of World Coordinates onto Image Coordinates

- Transformation of the homogeneous coordinates $\mathbf{P}^{(\mathrm{hom})}$ through euclidian motion of the camera

$$\mathbf{P}^{(\mathrm{hom})} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{P_0}^{(\mathrm{hom})} \quad R \in SO(3), \quad T \in \mathbb{R}^3$$

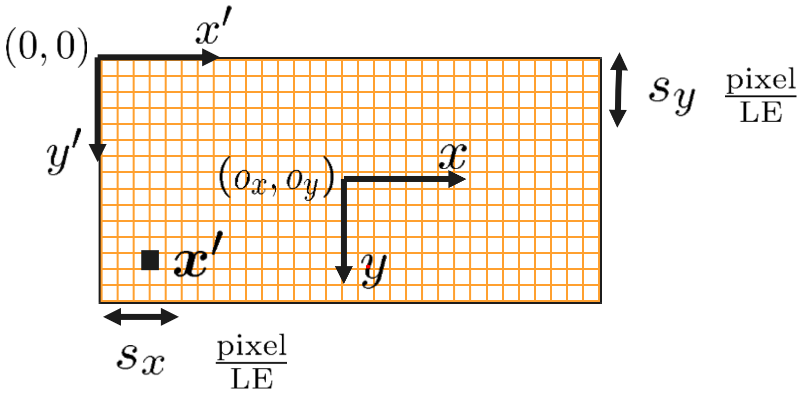
- Perspective projection by euclidian transformation

$$\mathbf{x}^{(\mathrm{hom})} \sim K_f \Pi_0 \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{P_0}^{(\mathrm{hom})}$$

Video

Sensor Parameter

Transformation of Image Coordinates to Pixel Coordinates



1. Specify the length unit (LE)

- $x_s = s_x x$, analogous to $y_s = s_y y$
- $s_x = s_y$ means quadratic pixel

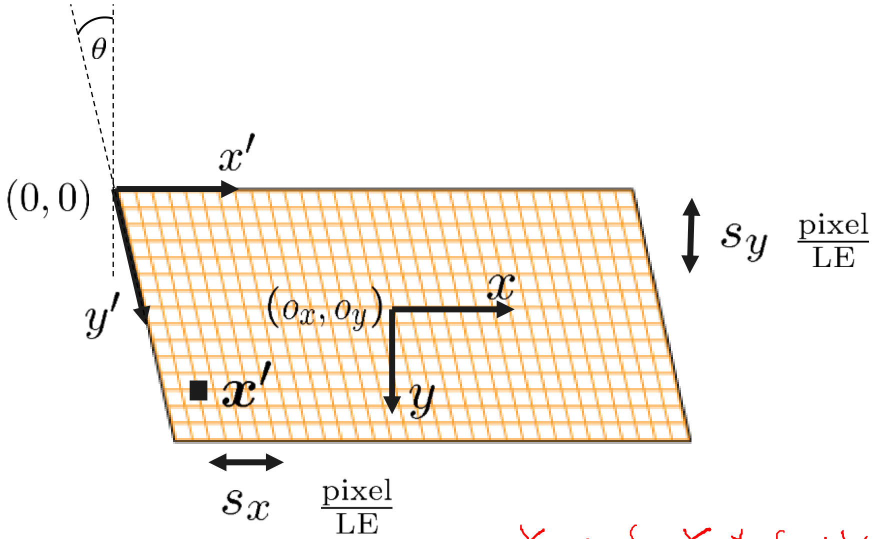
2. Adjustment of the origin

- Pixel coordinates $x' = x_s + o_x$, analogous to $y' = y_s + o_y$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Sensor Parameter

Linear Transformation of Image Coordinates to Pixel Coordinates



3. Introduction of a sheer factor s_θ

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{:=K_s} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$x_s = s_x \cdot x + s_\theta \cdot y$

$y_s = s_y \cdot y$

$$\mathbf{x}' = K_s \mathbf{x}$$

$$\mathbf{x} = K_s^{-1} \mathbf{x}'$$

Video

Calibration Matrix

Merging of Focal Length and Sensor Parameters

- Pixel coordinates

$$\mathbf{x}' = K_s \mathbf{x}$$

- Perspective projection

$$\mathbf{x} \sim K_f \Pi_0 \mathbf{P}$$

$$\mathbf{x}' \sim K_s K_f \Pi_0 \mathbf{P} = \underbrace{\begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{:=K} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Pi_0} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Calibration matrix (intrinsic camera parameters)

$$K = K_s K_f = \begin{bmatrix} f s_x & f s_\theta & o_x \\ 0 & f s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Video

Summary

- Mapping of world coordinates onto pixel coordinates

$$\mathbf{x}' \sim K\Pi_0 \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{P}_0$$

- Intrinsic parameter : K
- Extrinsic parameter: position of the camera
- Conversion of ideal image coordinates and pixel coordinates

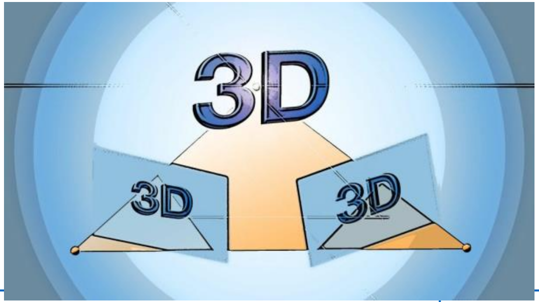
$$\mathbf{x}' = K_s \mathbf{x}$$

$$\mathbf{x} = K_s^{-1} \mathbf{x}'$$

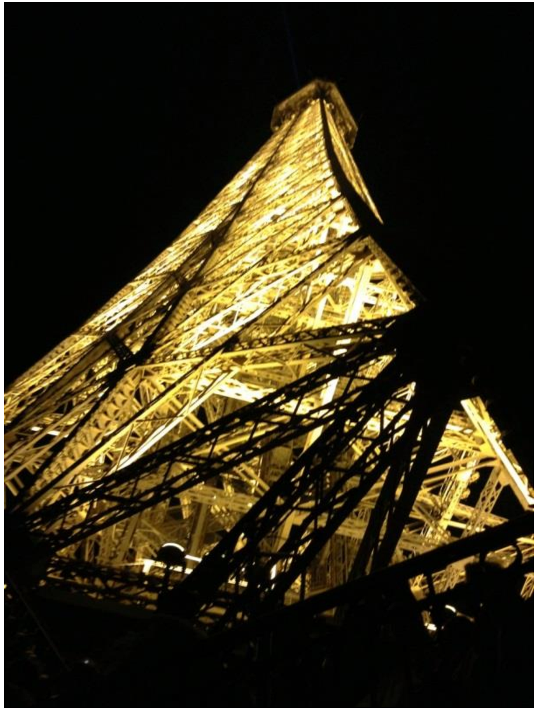
Martin Kleinsteuber: Computer Vision

Ch. 2 – Image Formation

5. Image, Preimage and Coimage



Motivation



Video

Recap

Projection of Lines

- Line in space in homogeneous coordinates

$$L^{(\mathrm{hom})} = \left\{ \mathbf{P}_0^{(\mathrm{hom})} + \lambda \left[v_1, v_2, v_3, 0 \right]^\top \mid \lambda \in \mathbb{R} \right\}$$

- Perspective projection of a line – example

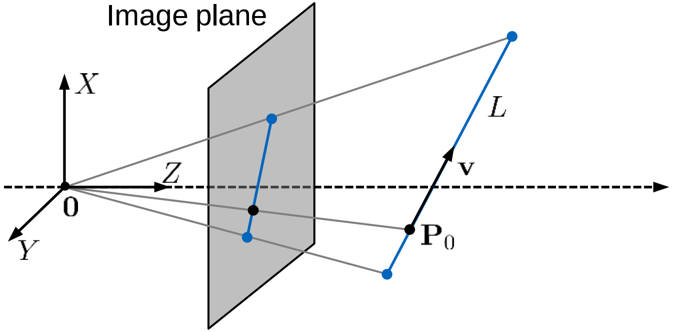


Image and Preimage

Definitions

- The image of a point and of a line, respectively, is their perspective projection
  $\Pi_0 \mathbf{P}^{(\mathrm{hom})}$ and $\Pi_0 L^{(\mathrm{hom})}$

- The Preimage of a point $\mathbf{P}$ is all the points in space which project onto a single image point in the image plane. The Preimage of a line L is all the points which project onto a single line in the image plane.

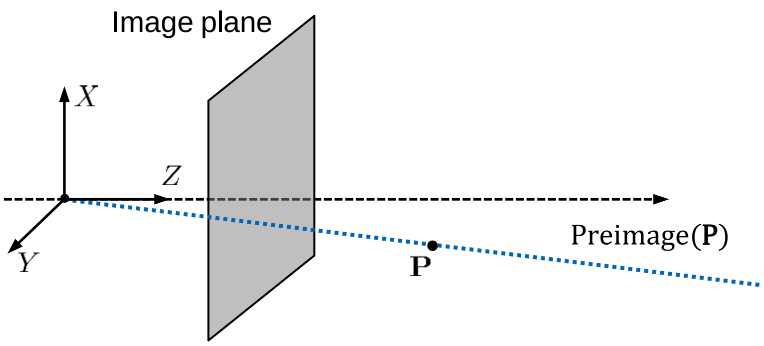
$$\mathrm{Urbild}(\mathbf{P}) = \{\mathbf{Q} \in \mathbb{R}^3 \mid \Pi_0 \mathbf{Q}^{(\mathrm{hom})} \sim \Pi_0 \mathbf{P}^{(\mathrm{hom})}\}$$

$$\mathrm{Urbild}(L) = \bigcup_{\mathbf{P} \in L} \mathrm{Urbild}(\mathbf{P})$$

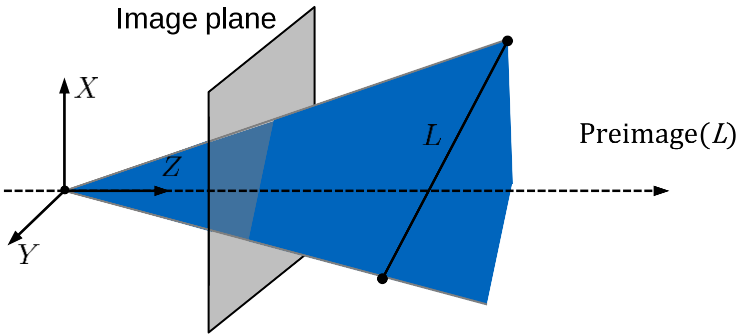
Urbild = Preimage

Image and Preimage
Properties

- Preimages of points are lines passing through the origin



- Preimages of lines are planes passing through the origin



Excursus: Linear Algebra

Span and Orthogonal Complement

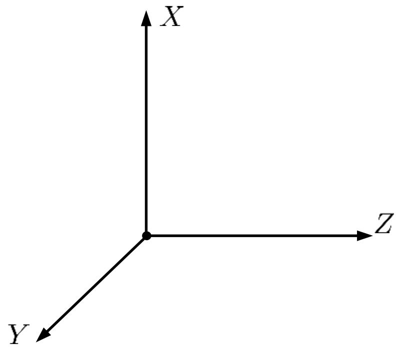
- Span of column vectors $\mathbf{a}_i$ of a matrix $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_m] \in \mathbb{R}^{n \times m}$

$$\operatorname{span}(\mathbf{A}) = \left\{ \sum_{i=1}^{m} \lambda_i \mathbf{a}_i \mid \lambda_i \in \mathbb{R} \right\}$$

- The span is a vector subspace of $\mathbb{R}^n$

- The orthogonal complement of a vector subspace

$$\operatorname{span}(\mathbf{A})^{\perp} = \{ \mathbf{v} \in \mathbb{R}^n \mid \langle \mathbf{a}_i, \mathbf{v} \rangle = 0, i = 1, \dots, m \}$$



Video

Excursus: Linear Algebra

Dimension, Kernel and Rank

- Dimension of a vector subspace:
  number of elements of a minimal generating set
- The rank of a matrix is the dimension of the span of the columns
- Kernel of a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$

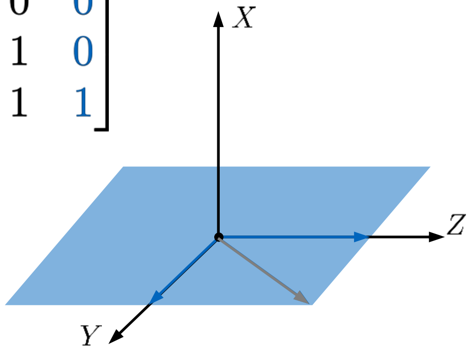
$$\mathrm{Ker}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^m \,|\, \mathbf{A}\mathbf{x} = 0\}$$

- $\mathrm{Ker}(\mathbf{A})$ is a vector subspace in $\mathbb{R}^m$
- Dimension theorem:

$$\mathbf{A} \in \mathbb{R}^{n \times m}$$

$$\mathrm{Rang}(\mathbf{A}) + \dim\,\mathrm{Ker}(\mathbf{A}) = m$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

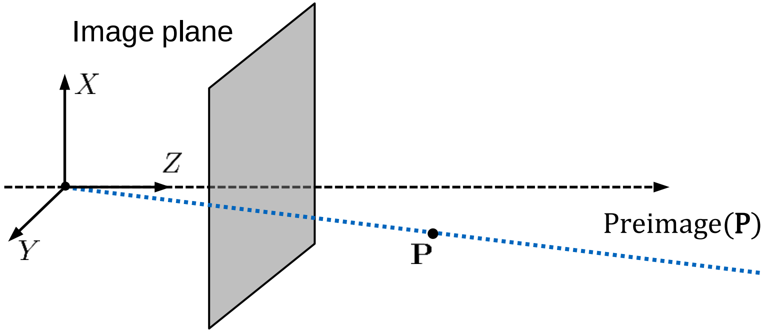


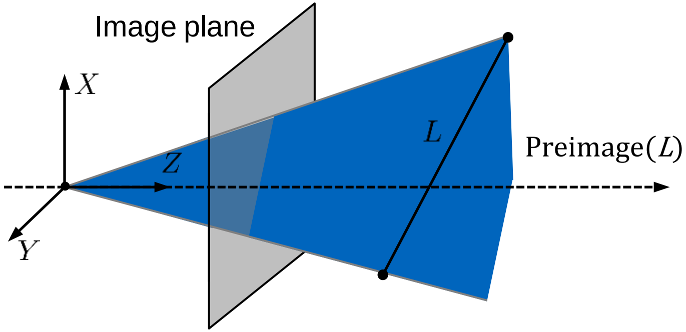
Rang = Rank

Coimage
Definition

Video

- The Coimage of points or lines is the orthogonal complement of the preimage





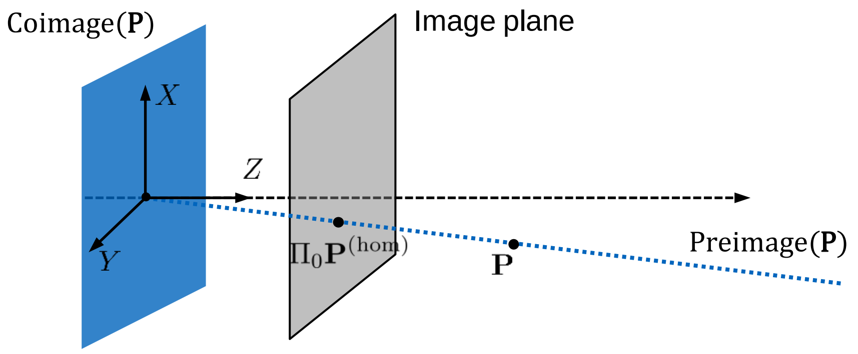
Image, Preimage and Coimage

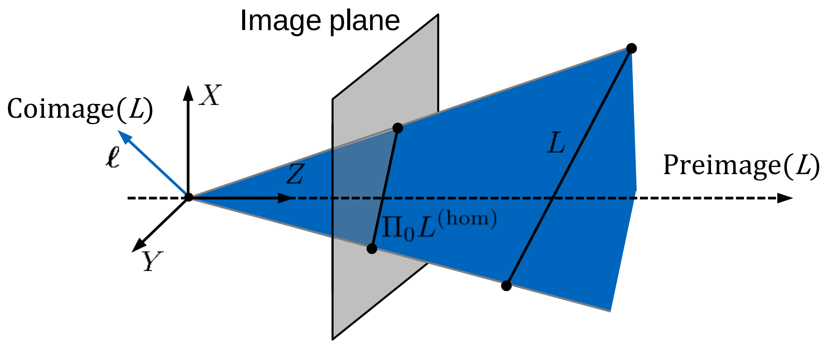
Correlations

Video

- Equivalent descriptions of image, preimage and coimage

	Bild	Urbild	Cobild
Punkt	$\mathrm{span}(\mathbf{P}) \cap \mathrm{BE}$	$\mathrm{span}(\mathbf{P})$	$\mathrm{span}(\hat{\mathbf{P}})$
Linie	$\mathrm{span}(\hat{\boldsymbol{\ell}}) \cap \mathrm{BE}$	$\mathrm{span}(\hat{\boldsymbol{\ell}})$	$\mathrm{span}(\boldsymbol{\ell})$





Coimage

Useful Properties

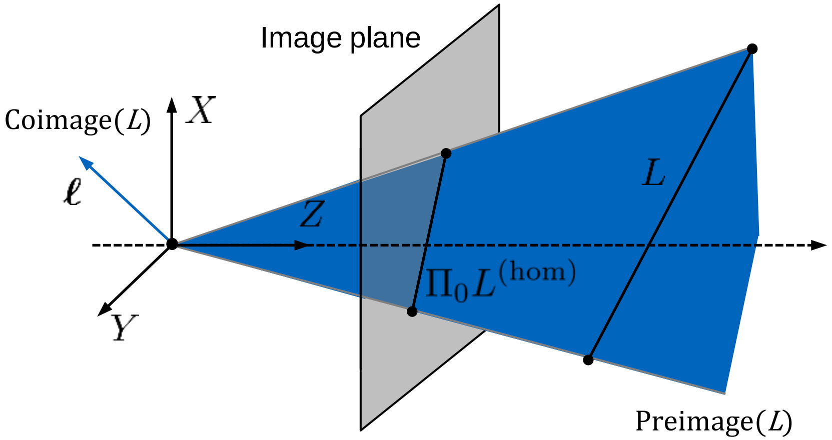
- Let L be a line in space with $\ell \in Preimage(L)$,
  let $\mathbf{x}$ be the image of a point on this line.
  It follows:

$$\mathbf{x}^\top \boldsymbol{\ell} = \boldsymbol{\ell}^\top \mathbf{x} = 0$$

- Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be the images of two points in space.
  For the Preimage ℓ of the connecting lines, it holds:

$$\boldsymbol{\ell} \sim \mathbf{x}_1 \times \mathbf{x}_2$$

Video



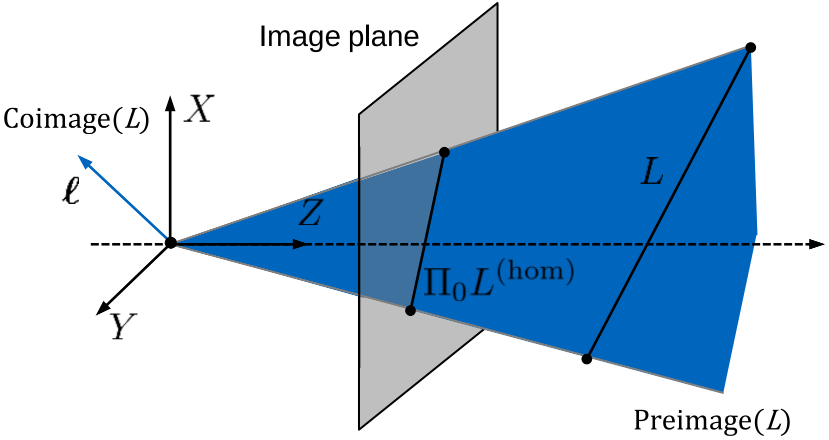
Cobild

Useful Properties

Video

- Let $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ be the preimages of two lines. For the point of intersection $\mathbf{x}$ of the images of these two lines, it follows:

$$\mathbf{x} \sim \boldsymbol{\ell}_1 \times \boldsymbol{\ell}_2$$



Collinearity of Image Points

Analysis with the Help of the Rank

- The image points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ lie exactly on a line (= they are collinear), if

$$\mathrm{Rang}([\mathbf{x}_1, \ldots, \mathbf{x}_n]) \leq 2$$

Generally, for $\mathbf{A} \in \mathbb{R}^{n \times n}$ the following are equivalent:

- $\det \mathbf{A} = 0$
- $\mathrm{Rang}(\mathbf{A}) < n$

- Three image points are exactly collinear, if

$$\det [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = 0$$

Video

Eigenvalues and Eigenvectors

Diagonalizability of Matrices

- Eigenvalues and eigenvectors of a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \qquad \text{falls} \quad \exists\, \mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

- Eigenvectors are only defined up to scaling

$$\alpha \mathbf{A}\mathbf{v} = \alpha \lambda \mathbf{v}$$

- Some matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ are diagonalizable, i.e.

$$\exists\, \mathbf{S} \in \mathbb{R}^{n \times n}, \mathbf{S} \text{ invertierbar} \qquad \mathbf{D}_A = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$$

- Real symmetrical matrices are diagonalizable and posess eigenvectors, which are orthogonal to one another, i.e.

$$\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]\,, \ \mathbf{v}_i \text{ EV von } \mathbf{A} \qquad \mathbf{D}_A = \mathbf{V}^\top \mathbf{A}\mathbf{V}$$

falls = if
invertierbar = invertible
von = of

Positive (semi-)definite Matrices

Orthogonal Diagonlizability

- A symmetrical matrix $\mathbf{A} = \mathbf{A}^\top$ is called positive (semi-) definite, if $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0,\ \mathbf{x} \neq \mathbf{0}$
- The eigenvalues of positive (semi-)definite matrices are positive (non-negative)
- positive (semi-)definite matrices can also be orthogonally diagonalized, i.e.

$$\lambda_i \geq 0,\ \forall\, \lambda_i \,\mathrm{EW}\ \text{ von } \mathbf{A}$$

$$\mathbf{D}_A = \mathbf{V}^\top \mathbf{A}\mathbf{V} \qquad \mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]\,,\ \mathbf{v}_i\,\mathrm{EV}\ \text{ von } \mathbf{A}$$

Video

Collinearity of Image Points

Analysis with the Help of Eigenvalues

- The image points $\mathbf{x}_1, \dots, \mathbf{x}_n$ lie on a line (= they are collinear), if and only if

$$\mathrm{Rang}([\mathbf{x}_1, \dots, \mathbf{x}_n]) \leq 2$$

- Three image points are collinear, if and only if

$$\det\left[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\right] = 0$$

- The image points $\mathbf{x}_1, \dots, \mathbf{x}_n$ are collinear, if the smallest eigenvalue of

$$\mathbf{M} = \sum_{i=1}^{n} \omega_i \mathbf{x}_i \mathbf{x}_i^\top$$

is equal to 0 $\quad \forall \omega_i > 0$

Video

Aspects of the Practical Implementation

Inaccuracies becaus of Discretization / Noise

- In Practice, the conditions
  - $\det[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = 0$
  - Smallest Eigenvalue $\mathbf{M} = \sum_{i=1}^{n} \omega_i \mathbf{x}_i \mathbf{x}_i^\top$ is equal to zero

  Never met

- Utilization of thresholds

Video

Summary

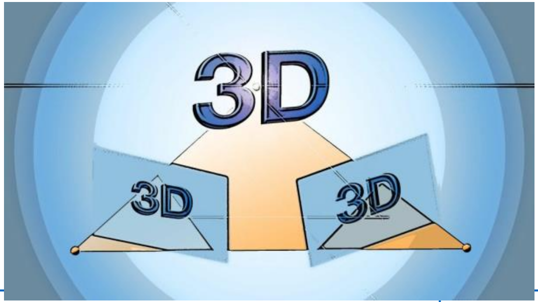
- Coimages of points and lines are vector subspaces
- The Coimage is the orthogonal complement of the preimage
- Definition of lines in the image by the Coimage
- Criteria for the collinearity of points
- Image, preimage and coimage are a useful formalism for the explanation of simple geometrical correlations of points and lines in space on the image plane

Video

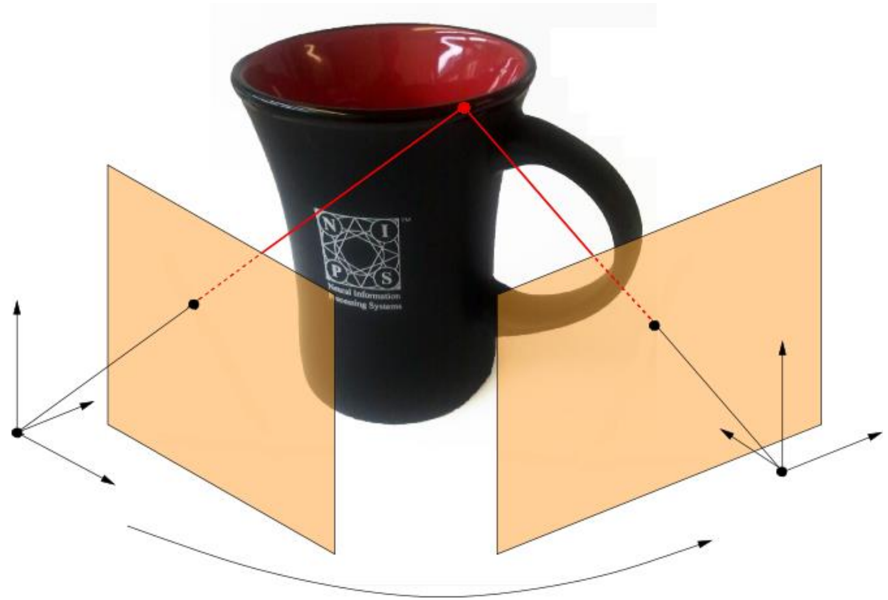
Martin Kleinsteuber: Computer Vision

Ch. 3 – Epipolar Geometry

1. Epipolar Equation



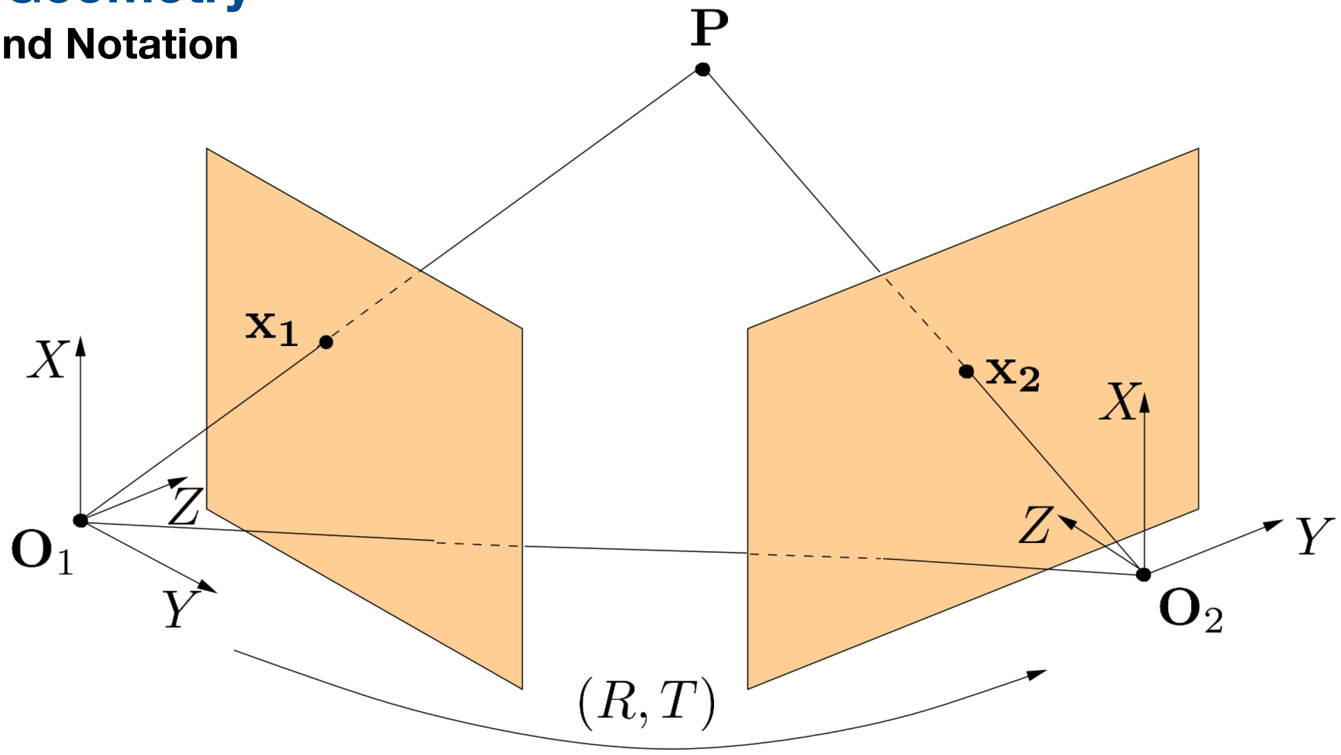
Motivation



- Aim: describing the correlation between corresponding image points with respect to the euclidian motion of the camera

Epipolar Geometry

Overview and Notation



Euclidian Motion

Conversion from Camera 1 into Camera 2

- $\mathbf{P_1}$ are the coordinates of the point $\mathbf{P}$ in camera system 1
- $\mathbf{P_2}$ are the coordinates of the point $\mathbf{P}$ in camera system 2
- Euclidian motion of the cameras described by the coordinate transformation

$$\mathbf{P}_2 = R\mathbf{P}_1 + T$$

with $R \in SO(3), T \in \mathbb{R}^3$

Video

Perspective Projection

Correlation between Points and Image Points

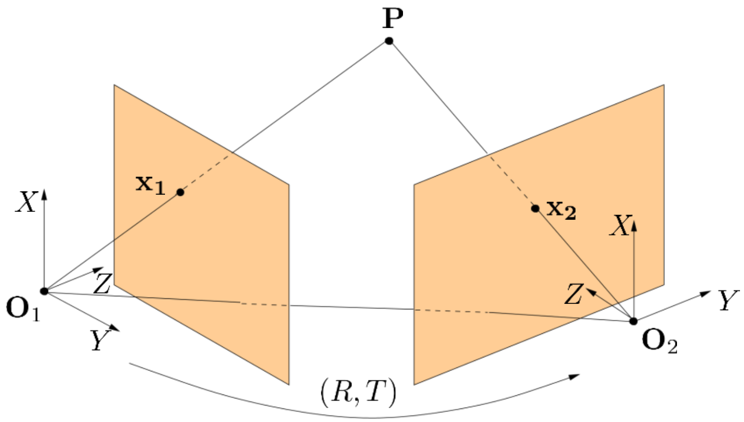
- Assuming an ideal camera, it holds

$$\lambda_i \mathbf{x_i} = \mathbf{P_i}, \quad i = 1,2\ , \quad \lambda \in \mathbb{R}$$

- Insrting it into euclidian motion

$$\lambda_2 \mathbf{x_2} = R\lambda_1 \mathbf{x_1} + T$$

- Problem: the scaling factors λ_i are generally unknown



Video

The Epipolar Equation

Formal Correlation between Corresponding Image Points

$$\lambda_2 \mathbf{x_2} = R\lambda_1 \mathbf{x_1} + T \quad \Big| \to \hat{T}$$

$$\lambda_2 \cdot \hat{T} \cdot x_2 = \lambda_1 \hat{T} R x_1 + \underbrace{\hat{T} \cdot T}_{=0} \quad \Big| \to x_2^T$$

$$\lambda_2 \cdot x_2^T \cdot (T \times x_2) = \lambda_1 x_2^T \hat{T} R x_1$$

- The matrix $E = \hat{T}R$ is called essential matrix to the euclidian motion (R, T)
- Epipolar equation $\mathbf{x_2}^\top E \mathbf{x_1} = 0$

Video

Properties of the Essential Matrix

Singular Value Decomposition (SVD)

- Every matrix $A \in \mathbb{R}^{n \times m}$ can be described as the product

$$A = U\tilde{\Sigma}V^\top$$

  whereby $U \in O(n), V \in O(m)$ and
- $n \leq m$: $\tilde{\Sigma} = \left[\ \Sigma \ \middle|\ 0 \right]$
- $m < n$: $\tilde{\Sigma} = \left[\begin{array}{c} \Sigma \\ \hline 0 \end{array}\right]$
- $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{\min\{n,m\}} \end{bmatrix} \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq 0$

Video

Properties of the Essential Matrix

Singular Value Decomposition (SVD)

- The singular values of a matrix are uniquely defined. U and V generally not.

Bsp.: $A = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & -1 \end{bmatrix}}_{\hat{=}\, U} \cdot \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}}_{\hat{=}\, \Sigma} \cdot \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}}_{\hat{=}\, V^T} = \underbrace{\begin{bmatrix} 1 & & \\ & -1 & \\ & & 1 \end{bmatrix}}_{\hat{=}\, U} \cdot \underbrace{\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}}_{\hat{=}\, \Sigma} \cdot \underbrace{\begin{bmatrix} 1 & & \\ & -1 & \\ & & -1 \end{bmatrix}}_{\hat{=}\, V^T}$

- Correlation with the eigenvalue decomposition of $AA^\top$ and $A^\top A$:

$A = U \Sigma V^T \Rightarrow AA^T = U \Sigma \Sigma^T U^T = U \cdot \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \cdot U^T$

$A^T \cdot A = V \cdot \Sigma^2 \cdot V^T$

Video

Properties of the Essential Matrix
Characterization through SVD

- A matrix E is an essential matrix (i.e. it has the form $E = \hat{T}R$, with skew-symmetric matrix $\hat{T}$ and rotation matrix R if and only if the singular value decomposition of E, produces the following

$$E = U \begin{bmatrix} \sigma & & \\ & \sigma & \\ & & 0 \end{bmatrix} V^\top$$

Video

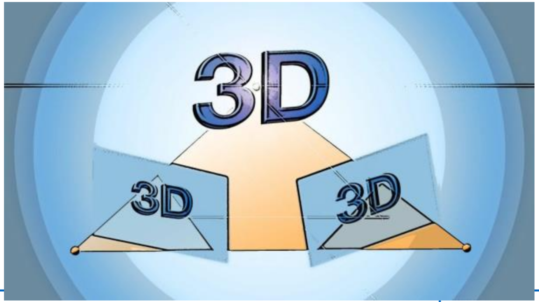
Summary

- The epipolar equation describes the correlation between corresponding image points from different images of the same scene
- The essential matrix contains the information of the euclidian motion
- Essential matrices are the ones, whose singular value decomposition produces two identical singular values and one singular value equal to zero

Martin Kleinsteuber: Computer Vision

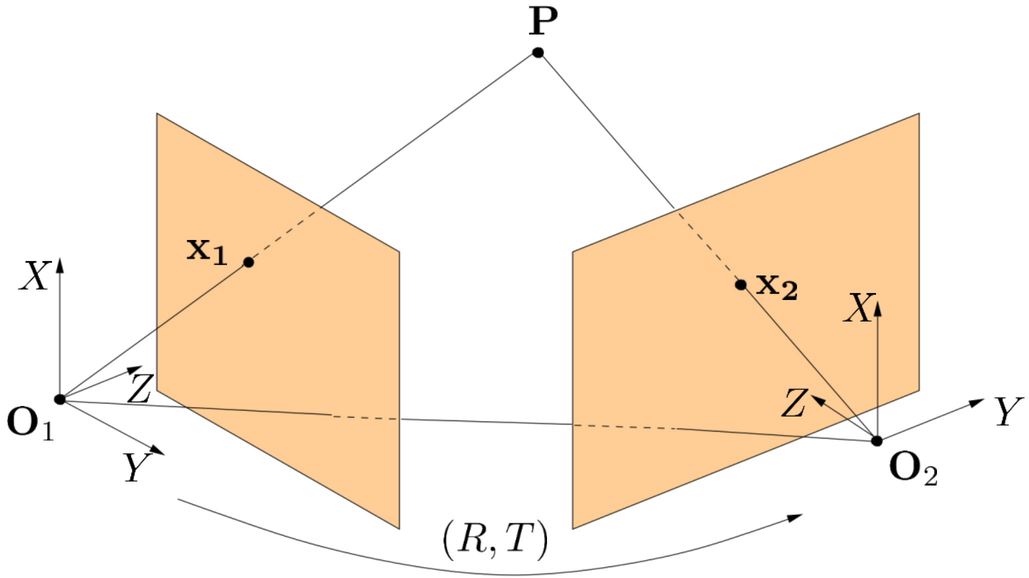
Ch. 3 – Epipolar Geometry

2. Epipoles and Epipolar Lines



Epipolar Geometry

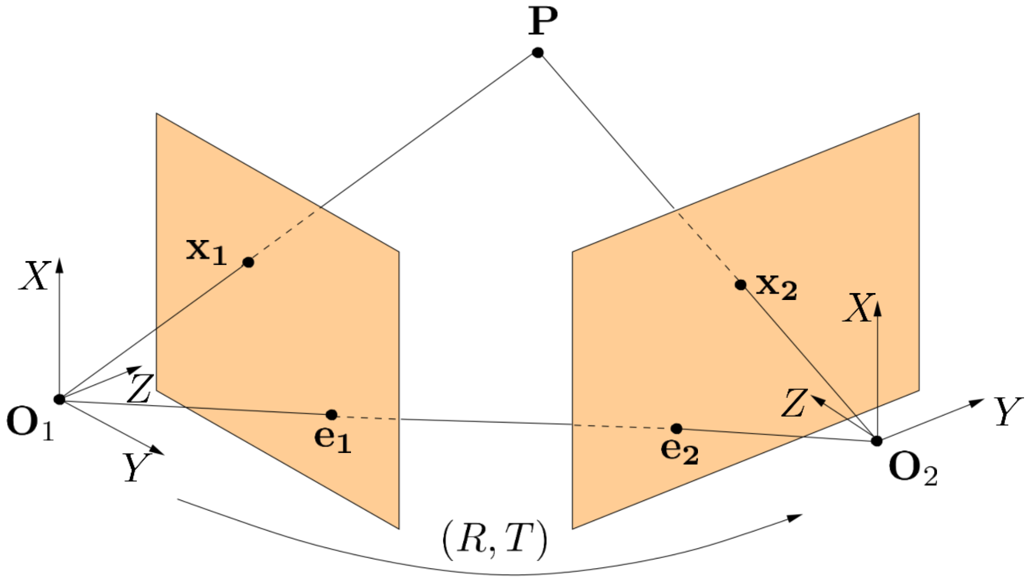
Recap Epipolar Equation



Epipolar Geometry

Definition of Epipoles

- The perspective projection of the respective optical centres in the other camera systems are called **Epipoles**



Epipoles

Properties

- From the geometry of the euclidian motion: $\mathbf{e_1} \sim R^\top \mathbf{T}, \quad \mathbf{e_2} \sim \mathbf{T}$

- $\mathbf{e_1}$ lies in the kernel of E : $E\mathbf{e_1} = 0$
- $\mathbf{e_2}$ lies in the kernel of $E^\top$: $E^\top \mathbf{e_2} = 0$

Video

The Zero Objects of E

From The Singular Value Decomposition

- $E = \begin{bmatrix} \mathbf{u_1} & \mathbf{u_2} & \mathbf{u_3} \end{bmatrix} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v_1}^\top \\ \mathbf{v_2}^\top \\ \mathbf{v_3}^\top \end{bmatrix}$

- $\mathbf{e_1} \sim \mathbf{v_3} \Rightarrow E\mathbf{e_1} = 0$

- The Coimage of $\mathbf{e_1}$ is equivalent to the third singular value of E from the right

Video

The Zero Objects of E$^\top$

From The Singular Value Decomposition

- $E^\top = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} \end{bmatrix} \begin{bmatrix} \sigma & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u_1}^\top \\ \mathbf{u_2}^\top \\ \mathbf{u_3}^\top \end{bmatrix}$

- $\mathbf{e_2} \sim \mathbf{u_3} \Rightarrow E^\top \mathbf{e_2} = 0$

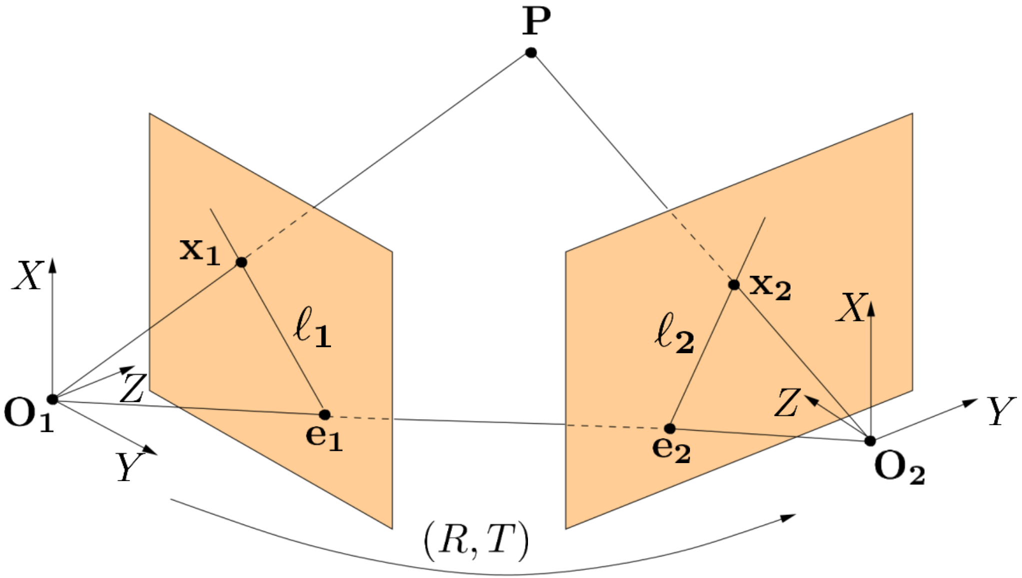
- The Coimage of $\mathbf{e_2}$ is equivalent to the third singular value of E from the left

Video

Epipolar Geometry

Definition Epipolarebene und Epipolarlinie

- The plane spanned by $\mathbf{O_1}, \mathbf{O_2}$ and **P**, is called the **epipolar plane of P**
- The intersection of the epipolar plane and the image plane is called **epipolar line**



Video

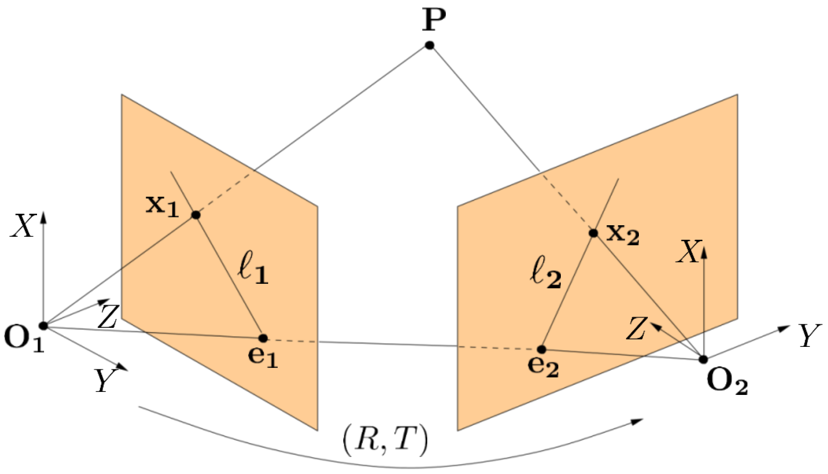
Epipolar Lines

Geometric Interpretation

- The epipolar line is the image that is created from the Preimage in the other camera system.

- The epipolar plane is spanned by the position vectors of the image point and of the epipole: $\mathrm{span}\left(\mathbf{x_i}, \mathbf{e_i}\right)$

- The epipolar line is identified by means of the Coimage $\ell_\mathbf{i} \sim \mathbf{e_i} \times \mathbf{x_i}$

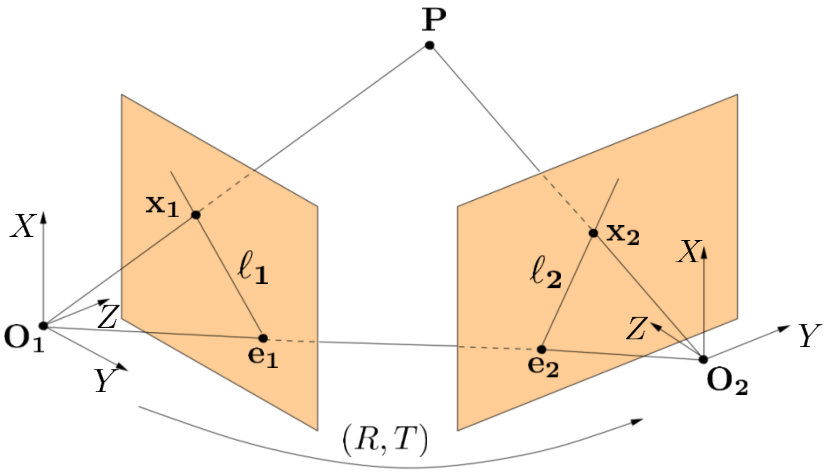


Video

Epipolar Lines

Properties

- $\boldsymbol{\ell_i}^\top \mathbf{x_i} = 0, \quad \boldsymbol{\ell_i}^\top \mathbf{e_i} = 0$
- $\boldsymbol{\ell_1} \sim E^T \mathbf{x_2}, \quad \boldsymbol{\ell_2} \sim E \mathbf{x_1}$

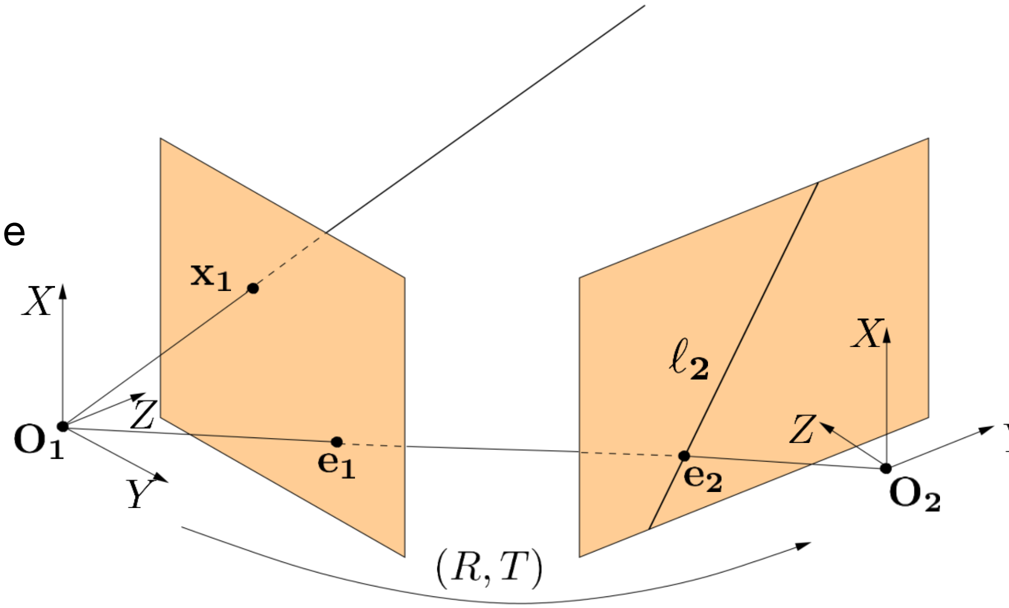


Video

Correspondence Search

By Means of the Essential Matrix

- Known: E and $\mathbf{x_1}$
- Calculate $\ell_2 \sim E\mathbf{x_1}$
- Determine the image of the epipolar line
- Search $\mathbf{x_2}$ (e.g. with NCC) along the image



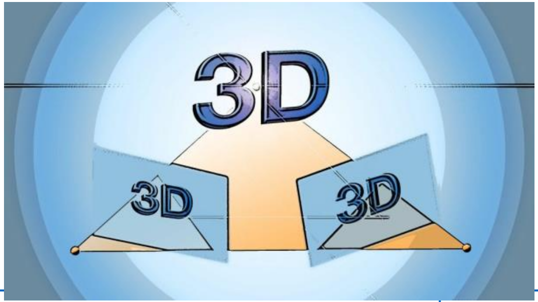
Summary

- Calculation of the epipoles and the epipolar line by means of the essential matrix
- Coimages of the epipols from the singular value decomposition of the essential matrix
- Epipolar lines through $\ell_2 \sim E\mathbf{x}_1$
- Simplified search of correspondences by means of the epipolar line

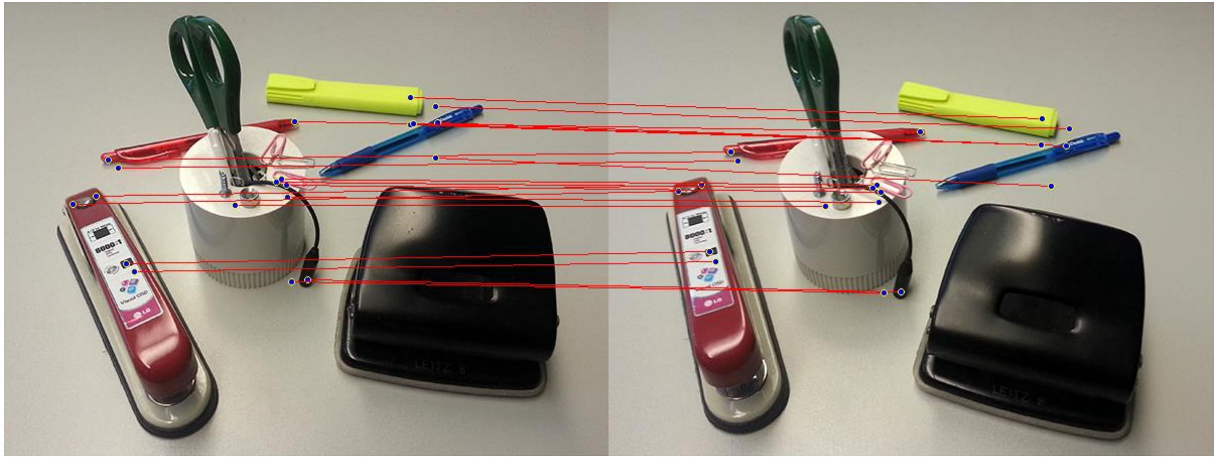
Martin Kleinsteuber: Computer Vision

Ch. 3 – Epipolar Geometry

3. The Eight-Point Algorithm



Motivation



$$(R, T)$$

- Euclidian Motion is unknown
- Feature points correspondence is known
- How to estimate the essential matrix?

Video

Eight-Point Algorithm for the estimation of the essential matrix

Motivation / Requirementd

- Given: n pairs of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j)$
- Ideally, all pair of points fulfill the epipolar equation

  $$\mathbf{x}_2^{j\top} E \mathbf{x}_1^j = 0$$

- Aim: Calculate the essential matrix E with the estimated pairs of corresponding points

  $E \in \mathbb{R}^{3\times 3}$,therefore 9 unknowns

- Scaling invariance: If E is the solution, then

  $\lambda E, \quad \lambda \in \mathbb{R}$ is one, too

- 8 independent equations are needed

Video

Vectorized Epipolar Equation

- Hiterto: $\mathbf{x}_2^{j\top} E \mathbf{x}_1^j = 0$

$$x_2^T \cdot [h_1, h_2, h_3] \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = x_2^T \cdot [x_1 h_1 + y_1 h_2 + z_1 h_3]$$

$$= x_1 \cdot x_2^T \cdot h_1 + y_1 x_2^T \cdot h_2 + z_1 \cdot x_2^T h_2$$

$$= [x_1 \cdot x_2^T, y_1 x_2^T, z_1 x_2^T] \cdot \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}$$

- For an ideal pair of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j)$ it holds therefore

$$\mathbf{a}^{j\top} \mathbf{E}^s = 0$$

Kronecker product $\otimes$

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$$

$$\mathbf{a} := \mathbf{x}_1 \otimes \mathbf{x}_2 = \begin{bmatrix} x_1x_2 \\ x_1y_2 \\ x_1z_2 \\ y_1x_2 \\ y_1y_2 \\ y_1z_2 \\ z_1x_2 \\ z_1y_2 \\ z_1z_2 \end{bmatrix} \in \mathbb{R}^9$$

Vectorization („stacking“)

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \quad \text{yields} \quad \mathbf{E}^s = \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{12} \\ e_{22} \\ e_{32} \\ e_{13} \\ e_{23} \\ e_{33} \end{bmatrix}$$

The Eight-Point Algorithm

System of Equations for the Determination of E

- n pairs of corresponding points produce the homogeneous system of linear equations $A\mathbf{E}^s = 0$, whereby

$$A := \begin{bmatrix} \mathbf{a}^{1\top} \\ \mathbf{a}^{2\top} \\ \vdots \\ \mathbf{a}^{n\top} \end{bmatrix} \in \mathbb{R}^{n\times 9}$$

- For generically distributed ideal pairs of corresponding points: $\dim(\mathrm{kern}(A)) = 1$ and therefore $\mathrm{rk}(A) = 8$

- Because of discretization errors, in reality it holds:

$$\mathrm{rk}(A) = 9$$

- For $n > 8$, the homogeneous system of linear equations has non-trivial solutions

Video

Excursus: Linear Algebra

Homogeneous Systems of Equations

- Idea: instead of $A\mathbf{x} = 0$, we solve the minimization problem

$$\min_{\mathbf{x}} \|A\mathbf{x}\|_2^2$$

- Because the solutions are scaling invariant, the search can be limited to

$$\mathbf{x} : \ \|\mathbf{x}\|_2 = 1$$

- I.e. find $\mathbf{x}$ whereby $\|\mathbf{x}\|_2 = 1$, that minimizes $\|A\mathbf{x}\|_2^2$.

Video

Excursus: Linear Algebra

Solution of the Minimization Problem with SVD

x^* mit $\|x^*\| = 1$ minimiert $\|Ax\|_2^2$

$\Leftrightarrow$ $w^* := V^\top x^*$ minimiert $\|AV \cdot w\|_2^2$

$\Leftrightarrow$ w^* minimiert $w^\top \cdot \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} w$

- Singular Value Decomposition $A = U\Sigma V^\top$

$\|Ax\|_2^2 = x^\top A^\top A x$

$$= \mathbf{x}^\top V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^\top \mathbf{x}$$

$\Rightarrow$ $w^* = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$

$\Rightarrow$ $x^* = V \cdot \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = v_n$

- $\mathbf{v}_n = \arg \min_{\|\mathbf{x}\|_2 = 1} \|A\mathbf{x}\|_2^2$

minimiert = minimizes

The Eight-Point Algorithm

Derivation until now

- Construct matrix A from $n \geq 8$ generically placed pairs of corresponding points

- Solve minimization problem $\mathbf{G}^s = \arg \min_{\|\mathbf{E}^s\|_2 = 1} \|A\mathbf{E}^s\|_2^2$:

  Singular value decomposition $A = U_A \Sigma_A V_A^\top$ yields the solution

  $\mathbf{G}^s = \mathbf{v}_9$ (9th column of V_A)

- Resorting the items of $\mathbf{G}^s$ yields $G \in \mathbb{R}^{3 \times 3}$

Video

The Eight-Point Algorithm

From the Solution of the Minimization Problem to the Essential Matrix

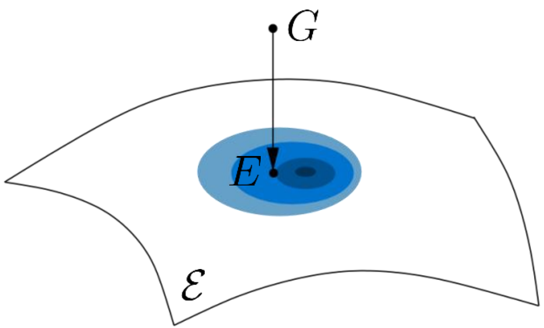
- G is generally not an essential matrix

$$G = U_G \Sigma_G V_G^\top \qquad \Sigma_G = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \sigma_3 \end{bmatrix}$$

- Find the „next“ essential matrix w.r.t. G

$$E = \arg\min_{E \in \mathcal{E}} \|E - G\|_F^2$$

- Projection onto the space of the essential matrices $\mathcal{E}$



The Eight-Point Algorithm

Projection onto the next Essential Matrix

- $E = \arg\min_{E \in \mathcal{E}} \|E - G\|_F^2$
- $E = U_G \begin{bmatrix} \sigma & & \\ & \sigma & \\ & & 0 \end{bmatrix} V_G^\top$
- $\sigma = \dfrac{\sigma_1 + \sigma_2}{2}$

- E can only be estimated without identifying ist scaling
- In practice therefore:

$$E = U_G \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} V_G^\top$$

Video

Summary

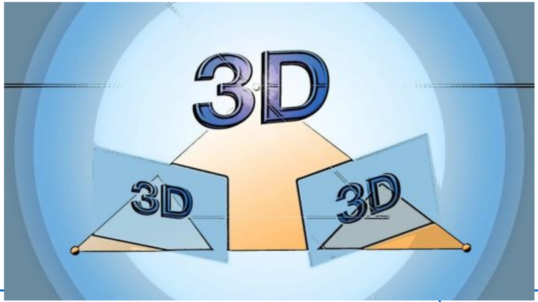
Eight-Point Algorithm

- System of linear equations from epipolar conditions
- Matrix of coefficients from Kronecker product of pairs of corresponding points
- The Solution is the 9th singular value from the right
- Projection onto normalized essential matrices
  - Singular value decomposition of the resorted solution
  - Set the first two singular values to one
  - Set the smallest singular value to zero

Martin Kleinsteuber: Computer Vision

Ch. 3 – Epipolar Geometry

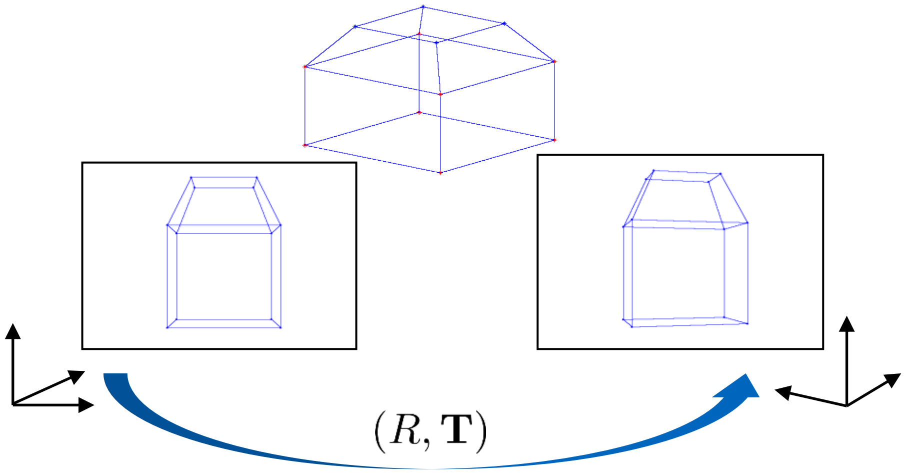
4. 3D Reconstruction



3D Reconstruction

Motivation and Assumptions

- Aim: Estimate the euclidian motion $(R, \mathbf{T})$ from the essential matrix and reconstruct 3D points from point correspondences $(\mathbf{x}_1^j, \mathbf{x}_2^j)$



Reconstruction of Rotation and Translation

Through the Singular Value Decomposition of the Essential Matrix

- Needed: $E = U\Sigma V^\top, \quad U, V \in \mathrm{SO}(3)$ (rotation matrices with determinant equal to 1)
- SVD is not unambiguous, U and V are not necessarily rotation matrices
- Example: $det(U) = -1$

$$E = \underbrace{U \begin{bmatrix}1&&\\&1&\\&&-1\end{bmatrix}}_{=:\hat{U}} \underbrace{\begin{bmatrix}1&&\\&1&\\&&-1\end{bmatrix} \begin{bmatrix}\sigma&&\\&\sigma&\\&&0\end{bmatrix}}_{=\Sigma} V^\top$$

$\Rightarrow \det(\hat{U}) = -\det U$

Video

Reconstruction of Rotation and Translation

Through Singular Value Decomposition of Essential Matrix

- Needed: $E = U\Sigma V^\top, \quad U, V \in \mathrm{SO}(3)$

- It can be shown, that with matrices

$$R_Z(\pm\tfrac{\pi}{2}) = \begin{bmatrix} 0 & \mp 1 & 0 \\ \pm 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

it holds:

$$R = U R_Z^\top(\pm\tfrac{\pi}{2}) V^\top$$
$$\hat{T} = U R_Z(\pm\tfrac{\pi}{2}) \Sigma U^\top$$

$$\hat{T}R = U \cdot R_Z(\tfrac{\pi}{2}) \cdot \Sigma U^T U \cdot R_Z^T(\tfrac{\pi}{2}) V^T$$

$$= U \cdot \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma & & \\ & \sigma & \\ & & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot V^T$$

vertauschen

$$= U \cdot \Sigma \cdot V^T = \underline{\underline{E}}$$

vertauschen = to swap

Reconstruction of Rotation and Translation

Ambiguity

$\lambda_2 \cdot x_2 = \lambda_1 \cdot \underline{R} x_1 + \underline{T}$ so dass

$\lambda_2 > 0,$

$\lambda_1 > 0.$

- Eight-Point algorithm yields $\pm E$
- Respectively two solutions for $(R, \hat{T})$
- Overall 4 euclidian transformations, which explain the correspondences
- Only one transformation is geometrically plausible
- $\mathbf{T}$ from $\hat{T}$ determinable, but without the scaling

Video

Reconstruction of the 3D Coordinates

Through Estimated Euclidian Transformation

$\rightarrow \hat{x}_2$

$\underbrace{\lambda_2 \hat{x}_2 \cdot x_2}_{=0} = \lambda_1 \cdot \hat{x}_2 R x_1 + \gamma \cdot \hat{x}_2 T$

$0 = \begin{bmatrix} \hat{x}_2 R x_1 & \hat{x}_2 T \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \gamma \end{bmatrix}$

- Aim: estimate the depth of points in space, e.g. λ_1^j in camera system 1
- $\lambda_2^j \mathbf{x}_2^j = \lambda_1^j R \mathbf{x}_1^j + \gamma \mathbf{T}, \quad j = 1, 2, \ldots, n$
- Each pair of corresponding points yields

$$\underbrace{\begin{bmatrix} \widehat{\mathbf{x}_2^j} R \mathbf{x}_1^j & \widehat{\mathbf{x}_2^j} \mathbf{T} \end{bmatrix}}_{:= M^j} \begin{bmatrix} \lambda_1^j \\ \gamma \end{bmatrix} = 0$$

Video

Reconstruction of the 3D Coordinates

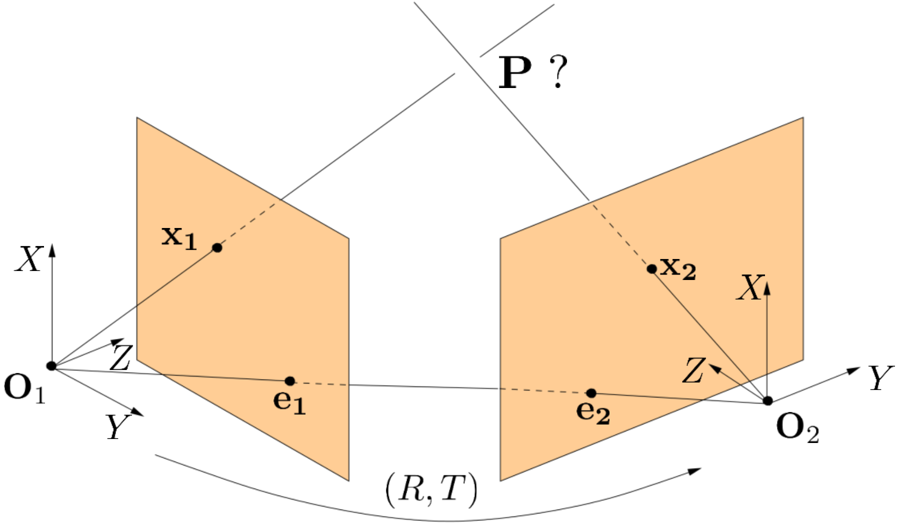
Possible, except Scaling

- Ideal pairs of corresponding points fulfill $M\mathbf{d} := \begin{bmatrix} \widehat{\mathbf{x}_2^1} R\mathbf{x}_1^1 & & & & \widehat{\mathbf{x}_2^1}\mathbf{T} \\ & \widehat{\mathbf{x}_2^2} R\mathbf{x}_1^2 & & & \widehat{\mathbf{x}_2^2}\mathbf{T} \\ & & \ddots & & \vdots \\ & & & \widehat{\mathbf{x}_2^n} R\mathbf{x}_1^n & \widehat{\mathbf{x}_2^n}\mathbf{T} \end{bmatrix} \begin{bmatrix} \lambda_1^1 \\ \lambda_1^2 \\ \vdots \\ \lambda_1^n \\ \gamma \end{bmatrix} = 0$

- In Practice: solve minimization problem $\min_{\|\mathbf{d}\|_2=1} \|M\mathbf{d}\|_2^2$ with the help of the SVD of M
- Inheren scaling invariance: without further knowledge of the scene, there is no difference between the scaling of the object and the distance between it and the camera
- Estimated 3D coordinates of $\mathbf{P}$ w.r.t. camera system 1 are $\mathbf{P}_1^j = \lambda_1^j \mathbf{x}_1^j$

Rekonstruktion der 3D-Koordinaten

Schwierigkeiten und Lösungsansätze



- Fehlerhafte Korrespondenzschätzung
  - Lösung z.B. über RanSaC-Methode
- Fehlerhafte 3D-Rekonstruktion durch Diskretisierungsfehler
  - Zusätzliche Schätzung von λ_2^j
  - Robuste Triangulationsverfahren, die 8PA als Initialisierung verwenden.

Video

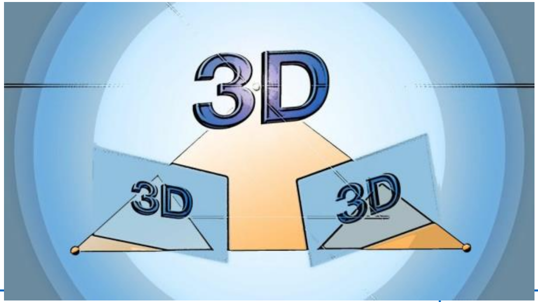
Summary

- Two euclidian transformations for one essential matrix
- Four euclidian transformations yielded by the Eight-Point algorithm
- Selection of the physically correct one by checking if the depth is positive
- Reconstruction of the 3D space coordinates through euclidian transformation

Martin Kleinsteuber: Computer Vision

Ch. 3 – Epipolar Geometry

5. The Fundamental Matrix



Motivation

- Epipolar equation for calibrated camera $\mathbf{x_2}^\top E \mathbf{x_1} = 0$
- Is it possible to find a similar relationship between pixel coordinates and the uncalibrated case ?
- The relationship between calibrated and uncalibrated coordinates is given by matrix K

$$\mathbf{x}' \sim K \Pi_0 \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{P}$$

$$\mathbf{x} = K^{-1}\mathbf{x}' \sim \Pi_0 \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \mathbf{P}$$

$$K = K_s K_f = \begin{bmatrix} f s_x & f s_\theta & o_x \\ 0 & f s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Epipolar Equation

Uncalibrated Case

- Epipolar equation for calibrated cameras $\mathbf{x_2}^\top E \mathbf{x_1} = 0$

$x_2 = K^{-1} x_2' \,,\quad x_1 = K^{-1} x_1' \quad \Rightarrow \quad x_2'^\top K^{-\top} E K^{-1} x_1' = 0$

- Uncalibrated version of the epipolar equation:

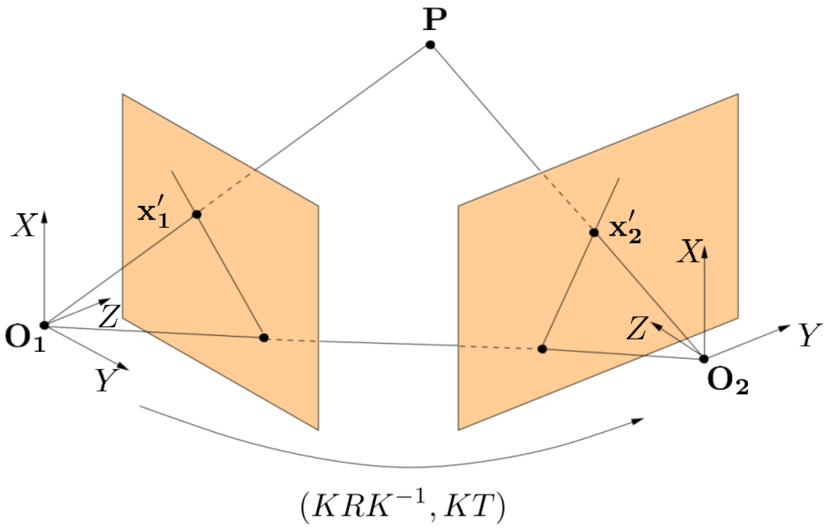
$$\mathbf{x_2'}^\top \underbrace{K^{-\top} E K^{-1}}_{=:\,F} \mathbf{x_1'} = 0$$

- The matrix $F := K^{-\top} E K^{-1}$ is called fundamental matrix.

Video

Epipolar Geometry

Uncalibrated case



- Euclidian Transformation $\lambda_2 \mathbf{x_2} = R\lambda_1 \mathbf{x_1} + T$

$$\lambda_2 \cdot K \cdot x_2 = \lambda_1 K \cdot R K^{-1} K x_1 + K \cdot T$$

$$\lambda_2 x_2' = \lambda_1 \cdot KRK^{-1} x_1' + T' \quad \Big| \to \hat{T}'$$

$$\lambda_2 \cdot \hat{T}' x_2' = \lambda_1 \cdot \hat{T}' KRK^{-1} \cdot x_1' \quad \Big| \to x_2'^{T}$$

$$0 = x_2'^{T} \cdot \hat{T}' KRK^{-1} \cdot x_1'$$

$$\widehat{KT} = \det(K) \cdot K^{-T} \hat{T} K^{-1}$$

$$\hat{T}' \sim K^{-T} \hat{T} K^{-1}$$

- It holds $\mathbf{x_2'}^\top \hat{T}' KRK^{-1} \mathbf{x_1'} = 0$
- With $\hat{T}' \sim K^{-\top} \hat{T} K^{-1}$ it follows $F \sim \hat{T}' KRK^{-1}$

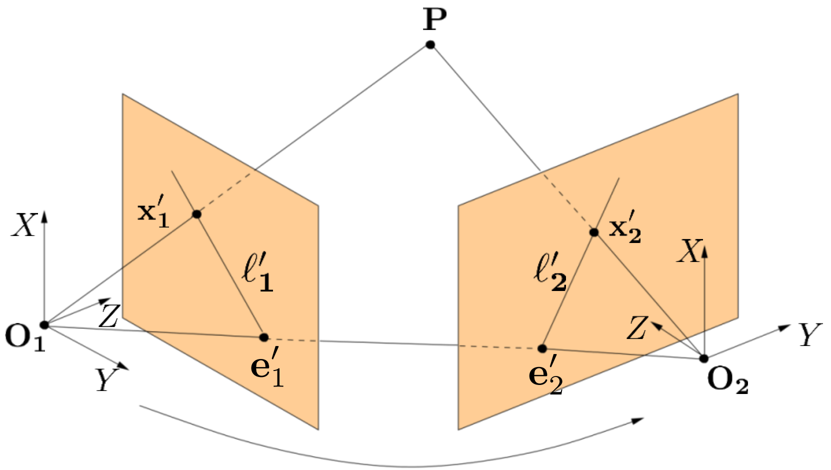
Video

Properties of the Fundamental Matrix

Epipoles and Epipolar Lines

$x_2'^\top \underbrace{F x_1'}_{} = 0$

- The correspondence point $\mathbf{x}_2'$ lies on th line $\ell_2' \sim F\mathbf{x}_1'$
  and inversely: $\mathbf{x}_1'$ lies on $\ell_1' \sim F^\top \mathbf{x}_2'$



- The epipoles in pixel coordinates:

$$\mathbf{e}_2'^\top F = 0, \quad F\mathbf{e}_1' = 0$$

$E e_1 = 0 \Leftrightarrow E \cdot K^{-1} K e_1 = 0 \Leftrightarrow E \cdot K^{-1} e_1' = 0 \Leftrightarrow \overbrace{K^{-\top} E K^{-1}}^{=F} \cdot e_1' = 0$

Video

Properties of the Fundamental Matrix

Singular Value Decomposition of the Fundamental Matrix

- A matrix is a fundamental matrix if and only if for ist singular value decomposition it holds that:

$$F = U\Sigma V^\top \qquad \text{mit} \quad \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & 0 \end{bmatrix}$$

- The fundamental matrix can be estimated by the Eight-Point Algorithm

Video

Properties of the Fundamental Matrix

Eight-Point Algorithm for the Estimation of the Fundamental matrix

- Known: n pairs of corresponding points in pixel coordinates
  $(\mathbf{x}_1'^{\,j}, \mathbf{x}_2'^{\,j}),\ j = 1, \ldots, n \quad (n \geq 8)$
- Construct coefficient matrix $A := \begin{bmatrix} \mathbf{a}^{1\,\top} \\ \vdots \\ \mathbf{a}^{n\,\top} \end{bmatrix} \in \mathbb{R}^{n \times 9}$ with $\mathbf{a}^{\mathbf{j}} = \mathbf{x}_1'^{\,j} \otimes \mathbf{x}_2'^{\,j}$
- Extract the singular vector corresponding to the smallest singular value of A and construct a matrix from it $G \in \mathbb{R}^{3 \times 3}$
- Project onto the set of fundamental matrices:
  - SVD of $G = U_G \mathrm{diag}\{\sigma_1, \sigma_2, \sigma_3\} V_G^\top$
  - Estimation of the fundamental matrix $F = U_G \mathrm{diag}\{\sigma_1, \sigma_2, 0\} V_G^\top$

Video

Kalibrierter vs. Unkalibrierter Fall

Überblick & Diskussion

Kalibrierte Kamera

- Epipolargleichung $\mathbf{x}_2^\top E \mathbf{x}_1 = 0$
- Essentielle Matrix $E = \hat{T} R$
- Epipole und Epipolarlinien

$$E\mathbf{e_1} = 0, \quad E^\top \mathbf{e_2} = 0$$

$$\ell_\mathbf{2} \sim E\mathbf{x_1}, \quad \ell_\mathbf{1} \sim E^\top \mathbf{x_2}$$

- 3D-Rekonstruktion möglich

Unkalibrierte Kamera

- Epipolargleichung $\mathbf{x}_2'^\top F \mathbf{x}_1' = 0$
- Fundamentalmatrix $F = K^{-\top} \hat{T} R K^{-1}$
- Epipole und Epipolarlinien

$$F\mathbf{e}_\mathbf{1}' = 0, \quad F^\top \mathbf{e}_\mathbf{2}' = 0$$

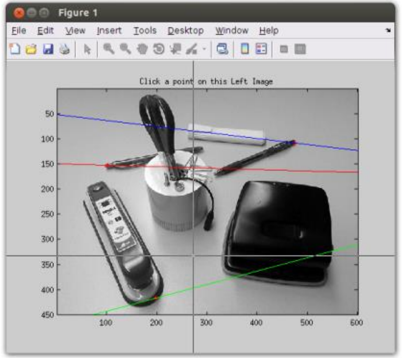
$$\ell_\mathbf{2}' \sim F\mathbf{x}_\mathbf{1}', \quad \ell_\mathbf{1}' \sim F^\top \mathbf{x}_\mathbf{2}'$$

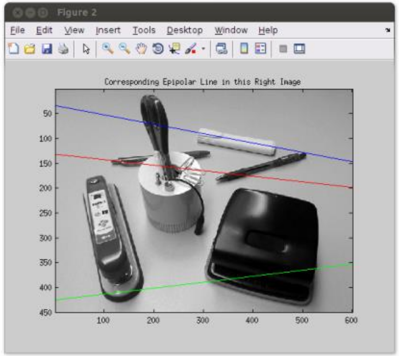
- 3D-Rekonstruktion nicht ohne Weiteres möglich

MATLAB-Demonstration

Eight-Point Algorithm and Epipolar Lines

- Mitul Saha & Rohit Singh, Stanford University
- http://ai.stanford.edu/~mitul/cs223b/fm.html





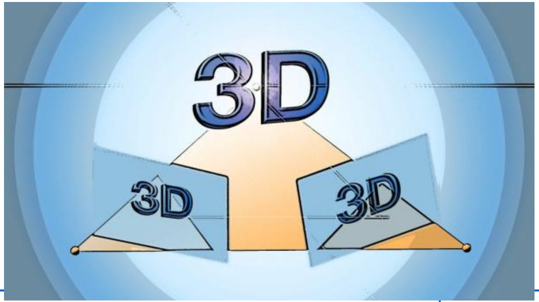
Summary

- Epipolar equation for uncalibrated cameras
- The fundamental matrix contains the information of the intrinsic camera parameters and of the euclidian motion
- Estimation of the fundamental matrix with the help of the Eight-Point algorithm and pairs of corresponding points in pixel coordinates

Martin Kleinsteuber: Computer Vision

Ch. 4 – Planar Scenes

1. The Planar Epipolar Equation



Motivation

Planar scene (aerial image)

- Aim 1: Formal correlation between correspondences
- Aim 2: Estimate the euclidian motion of the camera through correspondences in a planar scene

Video

Planar Scene

Plane Equation

- Plane in camera system 1

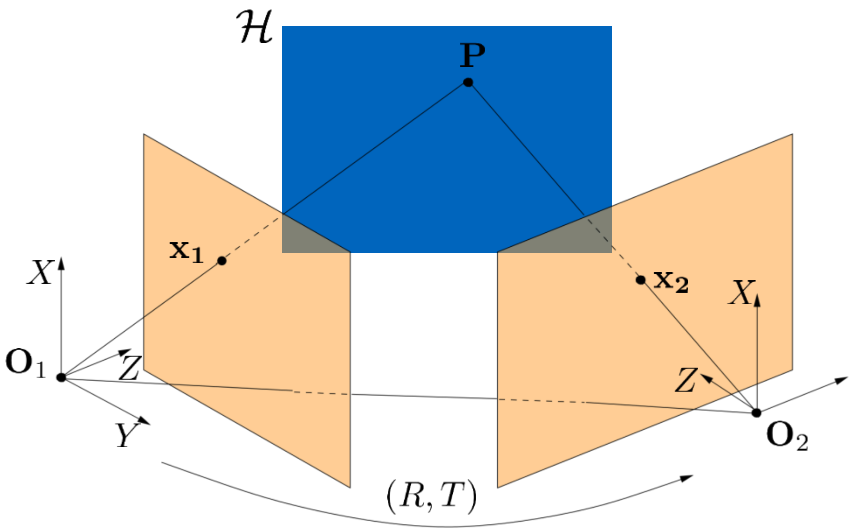
  $\mathcal{H} =: \{\mathbf{X} \in \mathbb{R}^3 \,|\, \mathbf{n}^\top \mathbf{X} = d\}$

  $\mathbf{n}$: Normalized normal vector

  d : Distance from $\mathcal{H}$ to $\mathbf{O}_1$

- Plane equation with the help of $\mathbf{P}_1$

  $$\mathbf{n}^\top \mathbf{P}_1 = n_1 X + n_2 Y + n_3 Z = d$$



Video

Planar Epipolar Equation

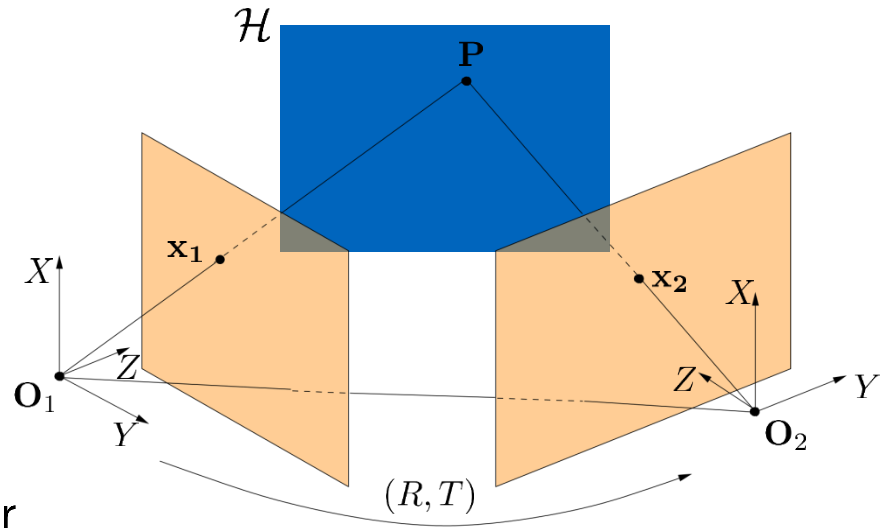
Correlation between Correspondence Point in a Planar Scene

- Euclidian Motion $\mathbf{P}_2 = R\mathbf{P}_1 + T$
- Plane Equation $\mathbf{n}^\top \mathbf{P}_1 = d$ $\Leftrightarrow \frac{1}{d} n^T \cdot P_1 = 1$

$$P_2 = R \cdot P_1 + T \cdot \left(\frac{1}{d} n^T P_1\right)$$

$$= \left(R + \frac{1}{d} T n^T\right) \cdot P_1$$

$$\Rightarrow x_2 \sim \underbrace{\left(R + \frac{1}{d} T n^T\right)}_{=: H} \cdot x_1$$



- H is the homography matrix for plane $\mathcal{H}$ and for euclidian motion (R, T)
- The equation $\mathbf{x}_2 \sim H\mathbf{x}_1$ is the planar epipolar equation

Properties of the Homography Matrix

Characterization through Singular Value Decomposition

- A matrix H a is a homography matrix, if and only if for its singular value decomposition it holds that

$$H = U \begin{bmatrix} \sigma_1 & & \\ & 1 & \\ & & \sigma_3 \end{bmatrix} V^\top$$

Video

Failure of the Eight-Point Algorithm

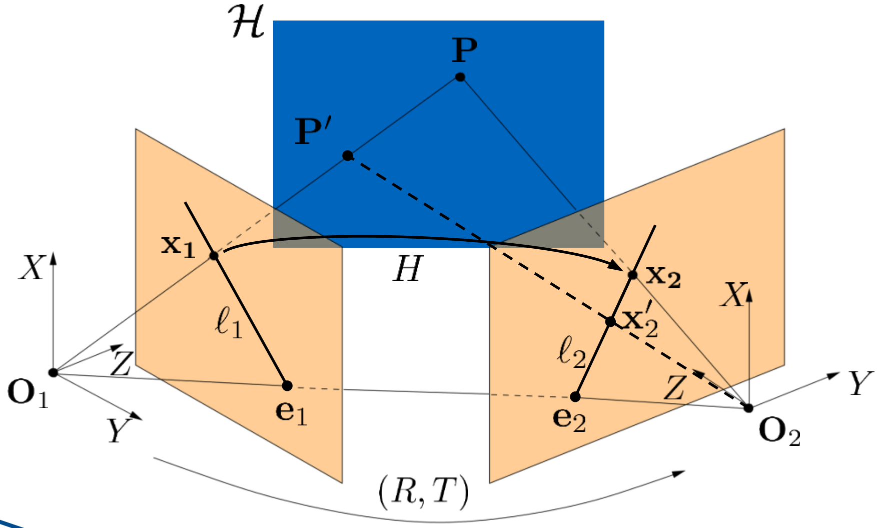
In the Planar Case

$A \cdot (\hat{u} H)^S = 0$ for all $u \in \mathbb{R}^3$

- Planar epipolar equation $\mathbf{x}_2 \sim H\mathbf{x}_1$
- $\mathbf{x}_2$ orthogonal to $\mathbf{u} \times H\mathbf{x}_1$ for all $\mathbf{u} \in \mathbb{R}^3$
- For ideal correspondence points, the coefficient matrix has $$A := \begin{bmatrix} \mathbf{a}^{1\top} \\ \mathbf{a}^{2\top} \\ \vdots \\ \mathbf{a}^{n\top} \end{bmatrix} \in \mathbb{R}^{n\times 9}, \quad \mathbf{a}^j := \mathbf{x}_1^j \otimes \mathbf{x}_2^j$$

$\dim(\mathrm{Ker}(A)) \geq 3$ and $\mathrm{Rang}(A) \leq 6$

Planar Epipolar Geometry
Conclusion



- For every image $\mathbf{x}_1$ of a point in the plane, it exists a unique correspondence point $\mathbf{x}_2$

Falls x_1 Bild eines Punkts ausserhalb der Ebene ist, gilt

$H \cdot x_1 \in \ell_2 \quad \ell_2 \sim E x_1$

$\ell_2^T H x_1 = 0 \quad \ell_2^T x_2 = 0 \quad \Rightarrow \ell_2 \sim x_2 \times H x_1$

$(H^T \ell_2)^T x_1 = 0 \quad \ell_1^T x_1 = 0 \quad \Rightarrow H^T \ell_2 \sim \ell_1$

- Epipolar lines defined by homography

$\boldsymbol{\ell}_2 \sim \hat{\mathbf{x}}_2 H \mathbf{x}_1, \quad \boldsymbol{\ell}_1 \sim H^\top \boldsymbol{\ell}_2$

„If x_1 is the image of a point outside of the plane, then it holds:“

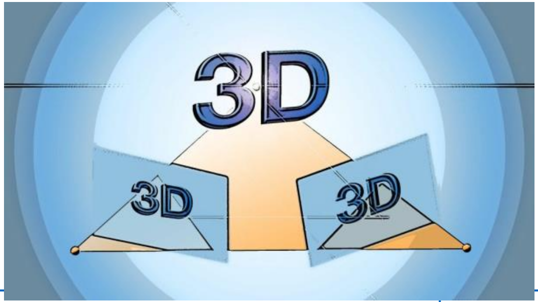
Summary

- In the planar case, points in 3D space lie on a plane in the scene
- The planar epipolar equation describes the formal correlation between image points
- Unique determination of correspondences with the help of the homography matrix
- The homography matrix allows for the calculation of epipolar lines outside of the plane, too

Martin Kleinsteuber: Computer Vision

Ch. 4 – Planar Scenes

2. The 4-Point Algorithm



Motivation

planar scene


- How to estimate the homography matrix from planar correspondence points?
- How to extract the euclidian motion from that ?
- How do homography matrix and essential matrix correlate ?

Video

4-Point Algorithm for the Estimation of the homography matrix

Motivation / Requirements

$x_2 \sim H x_1 \Leftrightarrow 0 = \hat{x}_2 H x_1$

- Known: n pairs of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j)$ of points in a plane
- Ideally, all correspondence points fulfill the planar epipolar equation

  $$\hat{\mathbf{x}}_2^j H \mathbf{x}_1^j = 0$$

- Aim: calculate the homography matrix H from the estimated correspondence points

  $H \in \mathbb{R}^{3\times 3}$, therefore 9 unknowns

- Scaling invariance: if H is a solution, then

  $\lambda H, \quad \lambda \in \mathbb{R}$ is too

Video

Vektorized Planar Epipolar Equation

- Hiterto: $\hat{\mathbf{x}}_2^j H \mathbf{x}_1^j = 0$

$$\hat{x}_2 \cdot [h_1, h_2, h_3] \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = [\hat{x}_2 h_1, \hat{x}_2 h_2, \hat{x}_2 h_3] \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$$

$$= x_1 \hat{x}_2 h_1 + y_1 \hat{x}_2 h_2 + z_1 \cdot \hat{x}_2 h_3 = 0$$

$$\Leftrightarrow \quad B^T \cdot H^s = 0$$

- $\mathrm{Rang}(\hat{\mathbf{x}}_2) = 2$, therefore generally also $\mathrm{Rang}(B) = 2$
- For each ideal pair of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j)$ the homogeneous system of linear equations $B^{j\top} \mathbf{H}^s = 0$ contains generally two independent equations
- A minimum of 4 generally distributed pairs of corresponding points are needed

Kronecker product $\otimes$

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \qquad \hat{\mathbf{x}}_2 = \begin{bmatrix} 0 & -z_2 & y_2 \\ z_2 & 0 & -x_2 \\ -y_2 & x_2 & 0 \end{bmatrix}$$

$$B := \mathbf{x}_1 \otimes \hat{\mathbf{x}}_2 = \begin{bmatrix} x_1 \hat{\mathbf{x}}_2 \\ y_1 \hat{\mathbf{x}}_2 \\ z_1 \hat{\mathbf{x}}_2 \end{bmatrix} \in \mathbb{R}^{9\times 3}$$

The 4-Point Algorithm

- n pairs cof corresponding points generate the homgeneous system of linear equations $A\mathbf{H}^s = 0\,,\quad A := \begin{bmatrix} B^{1\,\top} \\ B^{2\,\top} \\ \vdots \\ B^{n\,\top} \end{bmatrix} \in \mathbb{R}^{3n\times 9}$
- Solve minimization problem $\mathbf{H}^{\mathbf{s}}_{\mathbf{L}} = \arg \min_{\|\mathbf{H}^{\mathbf{s}}\|_2 = 1} \|A\mathbf{H}^{\mathbf{s}}\|_2^2$ with the SVD of A
- $H_L := \lambda H$ are resorted items of the righthand-side singular vector corresponding to the smallest singular value, Estimation of H except scaling
- From the properties of the homography matrix $|\lambda| = \sigma_2(H_L)$ and therefore $H = \frac{1}{\sigma_2(H_L)} H_L$

Video

The 4-Point Algorithm

Sign of the Estimated Homography Matrix

- Estimation of H except sign, $\pm H$ fulfill planar epipolar equation
- Exploitation of the positivity requirement of the depth yields the criterion $\mathbf{x}_2^\top H \mathbf{x}_1 > 0$

$$\lambda_2 x_2 = \lambda_1 R x_1 + T$$

$$\lambda_2 x_2 = \lambda_1 H x_1 \quad , \quad H = R + \frac{1}{d} T n^T$$

$$0 < \lambda_2 x_2^T \cdot \lambda_1 H x_1 \quad \Rightarrow \quad x_2^T H x_1 > 0.$$

- Choose the sign of H according to this criterion

Video

3D Reconstruction from Homography Matrix

Derivation

- A homography matrix $H = (R + \frac{1}{d}\mathbf{T}\mathbf{n}^\top)$ has at most two physically possible deompositions into the parameters $\{R, \frac{1}{d}\mathbf{T}, \mathbf{n}\}$
- Eigenvalue decomposition $H^\top H = V\Sigma^2 V^\top, \quad V = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} \in \mathrm{SO}(3)$
- Define

$$\mathbf{u}_1 := \frac{\sqrt{1-\sigma_3^2}\mathbf{v}_1 + \sqrt{\sigma_1^2 - 1}\mathbf{v}_3}{\sqrt{\sigma_1^2 - \sigma_3^2}}, \quad \mathbf{u}_2 := \frac{\sqrt{1-\sigma_3^2}\mathbf{v}_1 - \sqrt{\sigma_1^2 - 1}\mathbf{v}_3}{\sqrt{\sigma_1^2 - \sigma_3^2}}$$

- Define

$$U_1 := \begin{bmatrix}\mathbf{v}_2 & \mathbf{u}_1 & \hat{\mathbf{v}}_2\mathbf{u}_1\end{bmatrix}, \quad W_1 := \begin{bmatrix}H\mathbf{v}_2 & H\mathbf{u}_1 & \widehat{H\mathbf{v}_2}H\mathbf{u}_1\end{bmatrix}$$

$$U_2 := \begin{bmatrix}\mathbf{v}_2 & \mathbf{u}_2 & \hat{\mathbf{v}}_2\mathbf{u}_2\end{bmatrix}, \quad W_2 := \begin{bmatrix}H\mathbf{v}_2 & H\mathbf{u}_2 & \widehat{H\mathbf{v}_2}H\mathbf{u}_2\end{bmatrix}$$

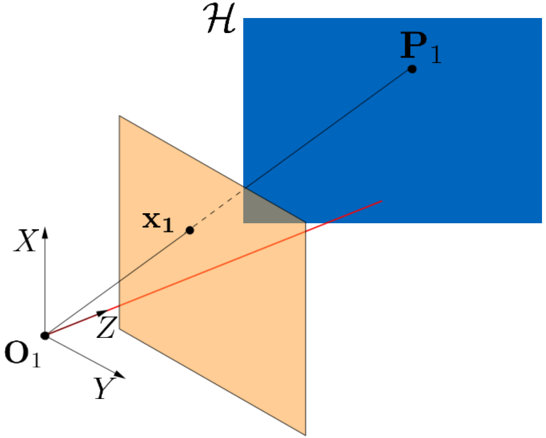
Video

3D Reconstruction from Homography Matrix

4 Solutions

Solution 1: $\begin{aligned} R_1 &= W_1 U_1^\top \\ \mathbf{n}_1 &= \mathbf{v}_2 \mathbf{u}_1 \\ \frac{1}{d}\mathbf{T}_1 &= (H - R_1)\mathbf{n}_1 \end{aligned}$	Solution 3: $\begin{aligned} R_3 &= R_1 \\ \mathbf{n}_3 &= -\mathbf{n}_1 \\ \frac{1}{d}\mathbf{T}_3 &= -\frac{1}{d}\mathbf{T}_1 \end{aligned}$
Solution 2: $\begin{aligned} R_2 &= W_2 U_2^\top \\ \mathbf{n}_2 &= \mathbf{v}_2 \mathbf{u}_2 \\ \frac{1}{d}\mathbf{T}_2 &= (H - R_2)\mathbf{n}_2 \end{aligned}$	Solution 4: $\begin{aligned} R_4 &= R_2 \\ \mathbf{n}_4 &= -\mathbf{n}_2 \\ \frac{1}{d}\mathbf{T}_4 &= -\frac{1}{d}\mathbf{T}_2 \end{aligned}$

$n^T \lambda \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = d > 0 \Rightarrow n^T \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} > 0$



Video

Homography Matrix and Essential Matrix

Correlations

- Consider $E = \hat{\mathbf{T}}R$ and $H = R + \mathbf{T}\mathbf{u}^\top, \quad R \in \mathbb{R}^{3\times 3},\ \mathbf{T}, \mathbf{u} \in \mathbb{R}^3, \quad \|\mathbf{T}\| = 1$
- Then it holds
  - $E = \hat{\mathbf{T}}H$
  - $H^\top E + E^\top H = 0$
  - $H = \hat{\mathbf{T}}^\top E + \mathbf{T}\mathbf{v}^\top$ for a specific $\mathbf{v} \in \mathbb{R}^3$

① $\hat{T} \cdot \hat{T}^T \cdot v = (I - T \cdot T^T) \cdot v$ orth. Projektion auf $\langle T \rangle^\perp$

② $\hat{T} \cdot H = \underbrace{\hat{T} \cdot R}_{\in \langle T \rangle^\perp} = \hat{T} \cdot \hat{T}^T \cdot \hat{T} \cdot R = \hat{T} \cdot \hat{T}^T \cdot E$

$\Rightarrow \hat{T} \cdot (H - \hat{T}^T E) = 0$ $\Rightarrow$ jede Spalte von $(H - \hat{T}^T E) \sim T$

$\Rightarrow H - \hat{T}^T E = [v_1 T, v_2 T, v_3 T] = T \cdot v^T$

Auf = onto
Jede = each
Spalt = column

Homography Matrix and Essential Matrix

Calculation of the Essential Matrix from the Homography Matrix

- Homography Matrix H is known
- Moreover, two pairs of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j),\ j = 1, 2$ are known, whose 3D-Points do not lie on the homography matrix's plane $\ell_2^1 \times \ell_2^2 \sim$
- Epipolar lines $\ell_2^j \sim \hat{\mathbf{x}}_2^j H \mathbf{x}_1^j$ intersect in the epipol $\mathbf{e}_2 \sim \mathbf{T}$
- Then: $E = \hat{\mathbf{T}} H$ whereby $\mathbf{T} \sim \hat{\ell}_2^1 \ell_2^2$ and $\|\mathbf{T}\|_2 = 1$

Video

Homography Matrix and Essential Matrix

Calculation of the Homography Matrix from the Essential Matrix

- Essential matrix E is known
- Moreover, three pairs of corresponding points $(\mathbf{x}_1^j, \mathbf{x}_2^j),\ j = 1, 2, 3$ are known, whose 3D points spane a plane in space
- Then: $H = \hat{\mathbf{T}}^\top E + \mathbf{T}\mathbf{v}^\top$ whereby $\mathbf{v} \in \mathbb{R}^3$, which is a solution to the system of equations $\hat{\mathbf{x}}_2^j \left(\hat{\mathbf{T}}^\top E + \mathbf{T}\mathbf{v}^\top\right) \mathbf{x}_1^j = 0, \quad j = 1, 2, 3$

Video

Summary

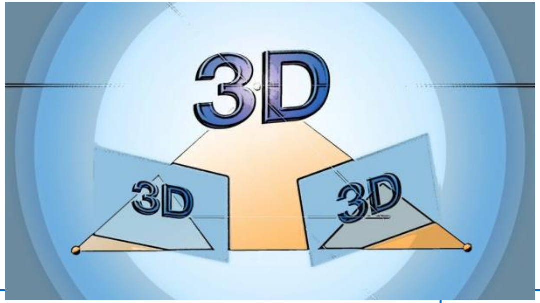
- 4-Point algorithm for the determination of the homography matrix
- Two physically plausible solutions for the parameters of the 3D reconstruction
- With the help of further correspondence points we can determine the essential matrix from the homography matrix (and the latter from the first)

Lehrstuhl für Datenverarbeitung
Fakultät für Elektrotechnik und Informationstechnik
Technische Universität München

Martin Kleinsteuber: Computer Vision

Ch. 4 – Planar Scenes

3. Camera Calibration



Motivation

- Determine calibration matrix K from several views of a chessboard
- To this end, use knowledge of the geometry of the chessboard

Recap

Perspective Projection

- Pixel coordinates

$$\mathbf{x}' = K_s \mathbf{x}$$

- Perspektive Projection

$$\mathbf{x} \sim K_f \Pi_0 \mathbf{P}$$

$$\mathbf{x}' \sim K_s K_f \Pi_0 \mathbf{P} = \underbrace{\begin{bmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{:=K} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\Pi_0} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

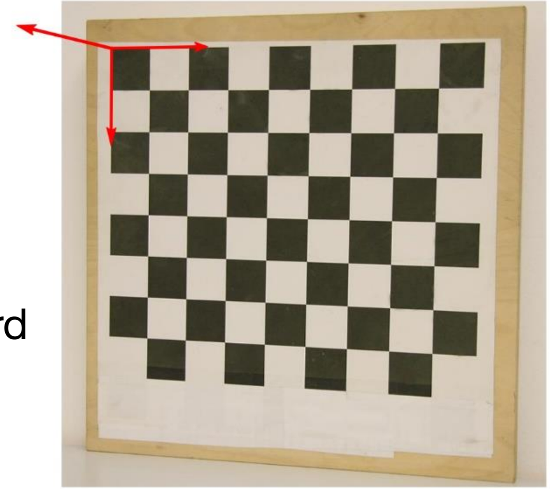
- With the euclidian transformation of $\mathbf{P}$

$$\mathbf{x}' \sim K \Pi_0 \begin{bmatrix} R & \mathbf{T} \\ 0 & 1 \end{bmatrix} \mathbf{P}$$

Video

Calibration

Approach



- The chessboard lies on the origin of the world coordinates
- Z-axis of the world coordinates perpendicular to the chessboard
- For a point $\mathbf{P} = \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix}$ on the chessboard, it holds

$$\mathbf{x}' \sim K\Pi_0 \begin{bmatrix} R & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

Video

Calibration

Estimation of the Homography

- The matrix $H := K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix}$ is a homography matrix, which projects the homogeneous coordinates of the chessboard onto homogeneous, uncalibrated coordinates in the image plane (except scaling)

$$\hat{\mathbf{x}}' H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = 0$$

- Estimate H with the 4-Point algorithm (except scaling)

with the help of known pairs of corresponding points $\left(\mathbf{x}'^{\,j}, \begin{bmatrix} X^j \\ Y^j \\ 1 \end{bmatrix} \right)$

Calibration

Constraints on the Calibration Matrix

$H \sim K[r_1\ r_2\ T]$

- It holds $K^{-1}\begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2\end{bmatrix} \sim \begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2\end{bmatrix}$
- With the orthogonality and the normalization of $\mathbf{r}_1$ and $\mathbf{r}_2$ we obtain the constraints:
  - $\mathbf{h}_1^\top K^{-\top} K^{-1} \mathbf{h}_2 = 0$
  - $\mathbf{h}_1^\top K^{-\top} K^{-1} \mathbf{h}_1 = \mathbf{h}_2^\top K^{-\top} K^{-1} \mathbf{h}_2$
- The matrix $K^{-\top}K^{-1}$ is symmetrical:

$$B := K^{-\top}K^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix}$$

Video

Calibration

Estimation of the Calibration Matrix

- The product $\mathbf{a}^\top B\mathbf{c}$ with $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$ can be

  reformulated as $\mathbf{a}^\top B\mathbf{c} = \mathbf{v}_{(\mathbf{a},\mathbf{c})}^\top \mathbf{b}$, $\quad \mathbf{v}_{(\mathbf{a},\mathbf{c})} = \begin{bmatrix} a_1c_1 \\ a_1c_2 + a_2c_1 \\ a_2c_2 \\ a_3c_1 + a_1c_3 \\ a_3c_2 + a_2c_3 \\ a_3c_3 \end{bmatrix}$

  and $\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^\top \in \mathbb{R}^6$

Video

Calibration

Estimation of the Calibration Matrix

- For each view, the constraints
  - $\mathbf{h}_1^\top B\,\mathbf{h}_2 = 0$
  - $\mathbf{h}_1^\top B\,\mathbf{h}_1 = \mathbf{h}_2^\top B\,\mathbf{h}_2$

  yield two equations

$$V^j\mathbf{b} = \begin{bmatrix} \mathbf{v}^\top_{(\mathbf{h}_1,\mathbf{h}_2)} \\ (\mathbf{v}_{(\mathbf{h}_1,\mathbf{h}_1)} - \mathbf{v}_{(\mathbf{h}_2,\mathbf{h}_2)})^\top \end{bmatrix}\mathbf{b} = \mathbf{0}$$

- n views of the chessboard yield the system of equations $V\mathbf{b} = \mathbf{0}$ whereby $V \in \mathbb{R}^{2n\times 6}$
- Therefore, we need at least 3 views to estimate $\mathbf{b}$ (except its scaling)

Video

Calibration

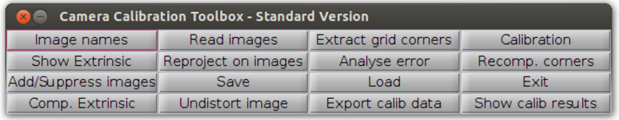
Steps to Calculate the Calibration Matrix

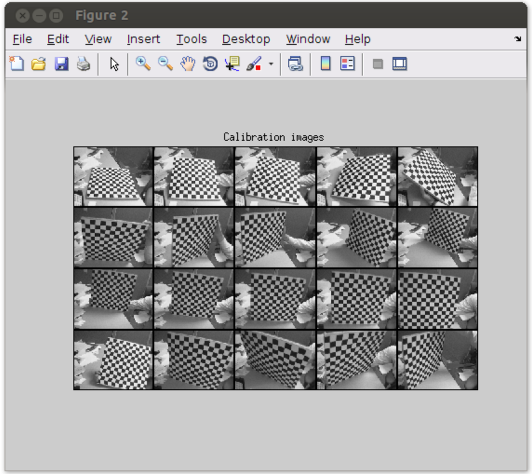
- Solve system of linear equations $V\mathbf{b} = \mathbf{0}$ with the SVD of V
- Generate symmetrical matrix $\tilde{B}$ from the right handside singular vector corresponding to the smallest singular value of V
- Choose positive definite $B = \pm\tilde{B}$
- Decompose B with the Cholesky factorization into the product $B = \tilde{K}^\top \tilde{K}$, whereby $\tilde{K}$ is an upper triangular matrix
- Then it holds: $K \sim \tilde{K}^{-1}$
- Normalize the calibration matrix to its (3,3)-item

MATLAB-Demonstration

Camera Calibration Toolbox

- Jean-Yves Bouget, CalTech
- http://www.vision.caltech.edu/bouguetj/calib_doc/index.html





Video

Summary

- At least 3 views of a chessboard are needed for the determination of the calibration camera
- Each view yields a homograhy between 3D points on the chessboard, as well as the corresponding uncalibrated image points
- Each view yields two equations for:

$$B = K^{-\top} K^{-1}$$

- Extract the calibration matrix with the help of the Cholesky factorization