Advanced Robot Perception

Fortgeschrittene Konzepte der Wahrnehmung für Robotersysteme

Georg von Wichert, Siemens Corporate Technology

Letzte Woche

- Kamerageometrie und Bildentstehung

- Grundlegendes zur Bildverarbeitung
  - Nur einige Basics als Basis für den Rest der Vorlesung

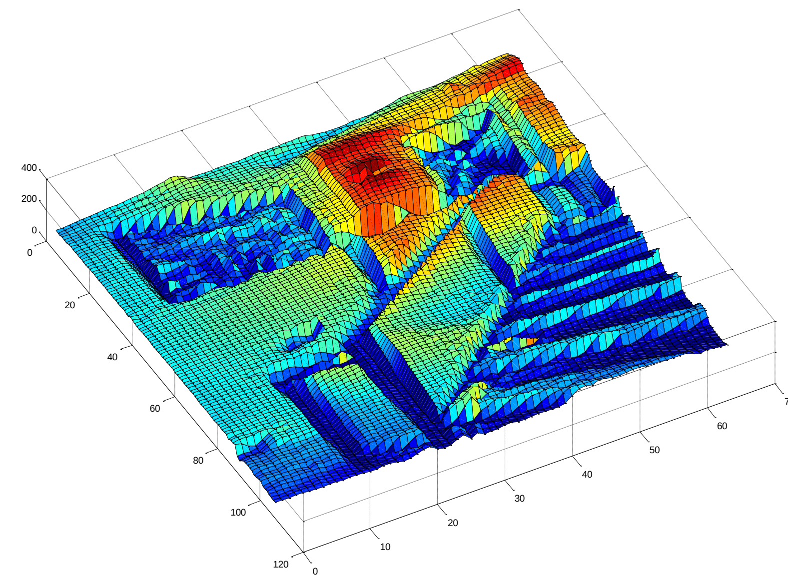
- Grundlegendes zur Mustererkennung
  - Bayes‘sche Statistik: Probability & Belief
  - „Gute“ Entscheidungen

Bild als Funktion des Ortes

- Grauwert des einzelnen Pixels als Funktion $f(u, v) \in [0, 255]$

$u \in [0, \mathrm{Width}]$

$v \in [0, \mathrm{Height}]$



- Charakterisierung lokaler Eigenschaften des Bildes über Analyse des Funktionsverlaufs von f(u,v)
  - z.B. Kanten als lokale Grauwertänderungen

Bildverarbeitung mittels linearer Filter

- Charakterisierung linearer Filter über (vgl. Systemtheorie/Regelungstechnik)
  - Impulsantwort *h(u,v)*
  - Übertragungsfunktion *H*

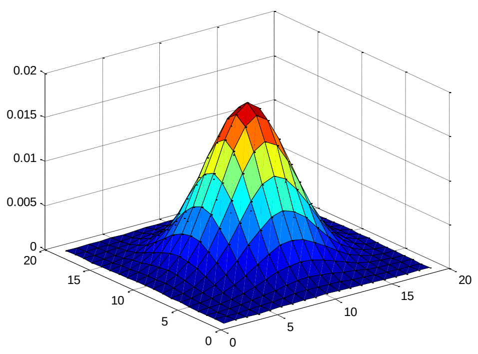
- In der digitalen Signalverarbeitung in der Regel Filter mit endlicher Impulsantwort
  - Filterung durch Faltungsoperation

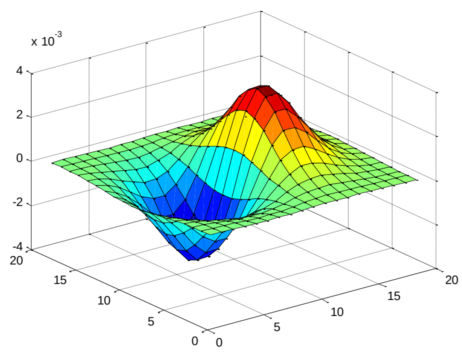
$$f(u,v) \; * \; h(u,v) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} h(\tau_1,\tau_2) \cdot f(x-\tau_1, y-\tau_2)$$

Gauß‘sche Richtungsableitungen

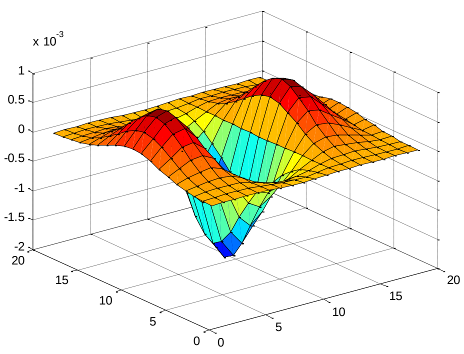
- Erst glätten, dann ableiten
  - Entspricht Faltung mit einen abgeleiteten Glättungskern

$$\frac{\partial}{\partial x}(f(x,y) * h(x,y)) = f(x,y) * \frac{\partial}{\partial x}h(x,y)$$





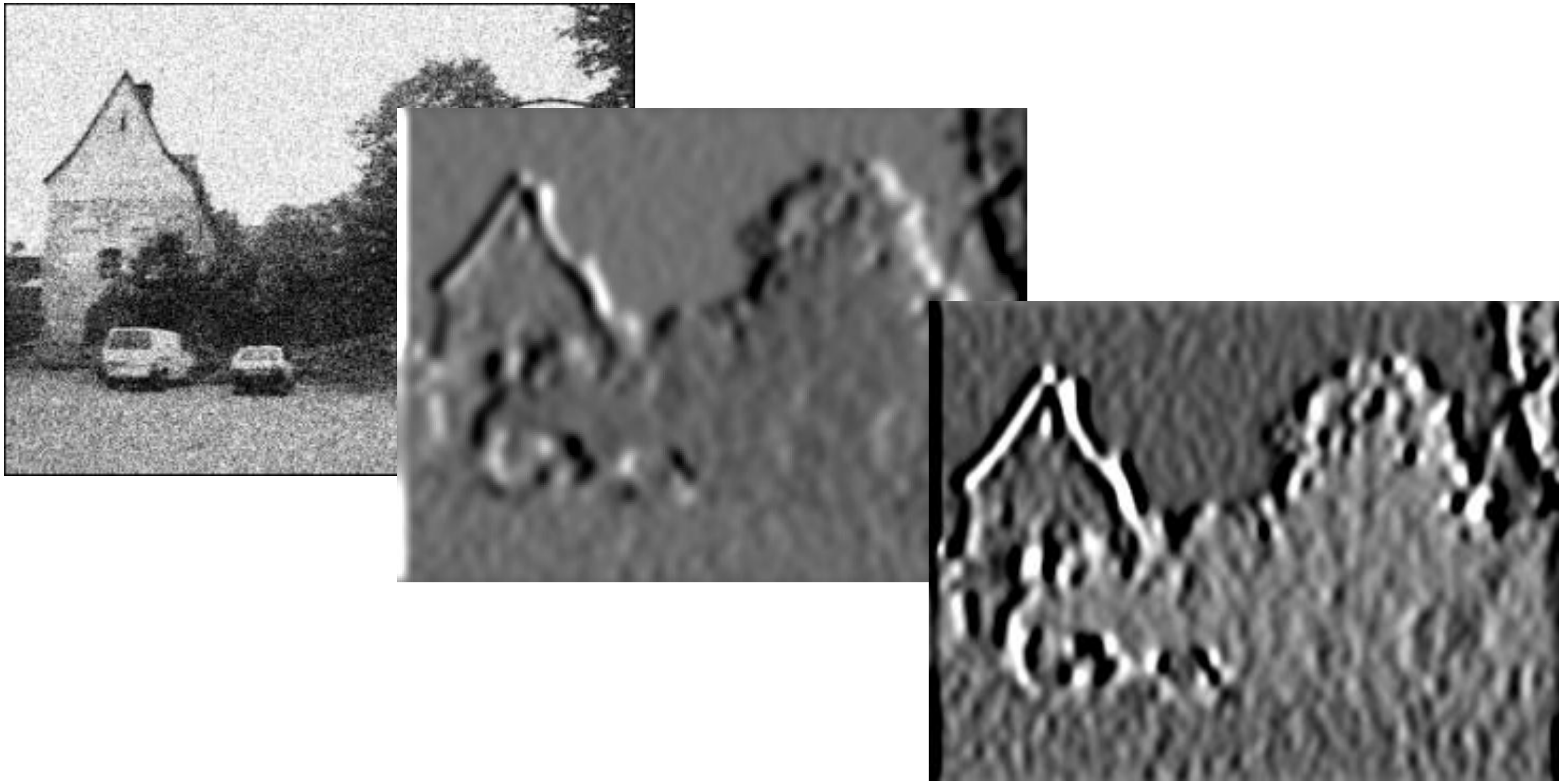
1. Ableitung



2. Ableitung

Gauß‘sche Richtungsableitungen

- Glättungsanteil des Filters dämpft Rauschen



Verbundverteilung , statistische Unabhängigkeit und Randverteilungen

- Verbundverteilung $$P(X,Y) = P(X|Y)\ P(Y)$$
- Verbunddichte $$p(x,y) = p(x|y)\ p(y)$$
- Statistische Unabhängigkeit $$P(X,Y) = P(X)\ P(Y)$$ $$p(x,y) = p(x)\ p(y)$$
- Randverteilung $$P(x) = \sum_y P(x,y)$$
- Randdichte $$p(x) = \int p(x,y)\mathrm{d}y$$

Randverteilung

	x_1	x_2	x_3	x_4	$p_y(Y)\downarrow$
y_1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
y_2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{4}$
y_3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
Y_4	$\frac{1}{4}$	0	0	0	$\frac{1}{4}$
$p_x(X)\rightarrow$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	1

Bedingte Unabhängigkeit

- Definition der bedingten Unabhängigkeit

$$P(x, y \mid z) = P(x \mid z)P(y \mid z)$$

- Dies entspricht $P(x \mid z) = P(x \mid y, z)$

$$P(y \mid z) = P(x \mid x, z)$$

- **Achtung**, dies impliziert nicht $P(x, y) = P(x)P(y)$

Merkmalsextraktion

- Merkmale (engl. Features) sind für die Erkennungsaufgabe hilfreiche Zahlenwerte
  - z.B. Dicke und Durchmesser der Brötchen

6,63027181282151	3,53867504726895
6,88686173034205	3,92131040308234
6,88430332022434	3,54949976124462
6,36520879057940	4,02253643322049
6,44322620306702	3,66220939079799
6,57727566517259	4,24876258223907
7,02477397855060	3,38155395317036
6,78444022562949	2,55339370408865
5,98173295133781	3,93233196957361
6,26804193710415	3,30271261332149
6,14414738063608	3,52581774912581

Merkmalsvektoren ***x*** = *[Durchmesser, Dicke]*

	Dicke	Durchmesser	Varianz Dicke	Varianz Durchmesser
Backer 1	3.5	6.5	0.1	0.3
Bäcker 2	2.5	5	0.2	0.5

Entscheidungstheorie

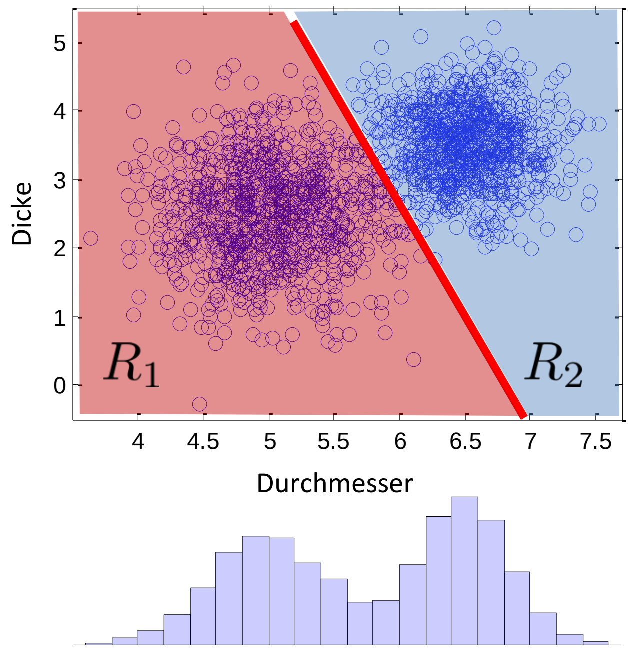
- Was ist die optimale Entscheidungsregel?
  - Minimierung der Wahrscheinlichkeit falscher Klassifikation

$$p(falsch) = p(\mathbf{x} \in R_1, \mathcal{C}_2) + p(\mathbf{x} \in R_2, \mathcal{C}_1)$$

$$= \int_{R_1} \underbrace{p(\mathbf{x}, \mathcal{C}_2)}\ d\mathbf{x} \ + \ \int_{R_2} \underbrace{p(\mathbf{x}, \mathcal{C}_1)}\ d\mathbf{x}$$

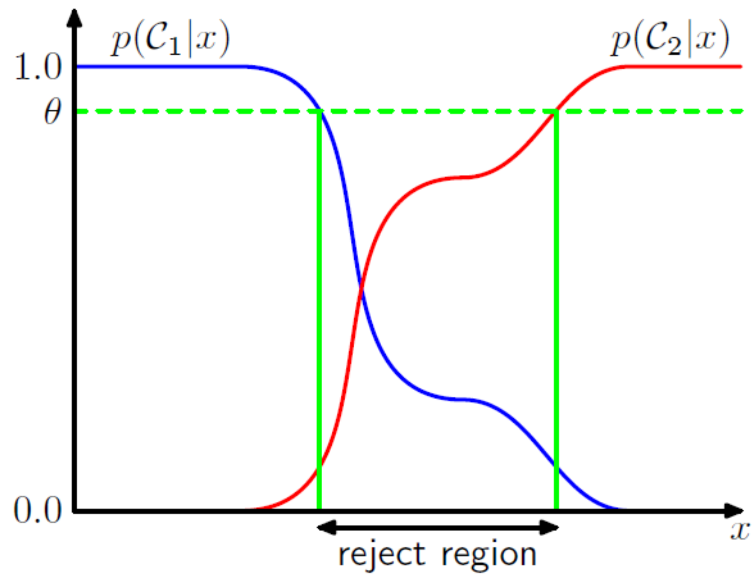
Minimierung der Wahrscheinlichkeit dass das Objekt aus der anderen Klasse kommt

$$p(\mathbf{x}, \mathcal{C}_x) = \underbrace{p(\mathcal{C}_x | \mathbf{x})}_{\text{Optimale Entscheidung: Klasse mit der maximalen Aposteriori-Wahrscheinlichkeit}} \ \overbrace{p(\mathbf{x})}^{\text{Für alle Klassen gleich}}$$



Rückweisung

- Vermeiden von Klassifikationsfehlern in mehrdeutigen Situationen
  - Nur entscheiden, wenn die maximale Aposteriori-Wahrscheinlichkeit größer ist, als eine vordefinierte Schwelle θ

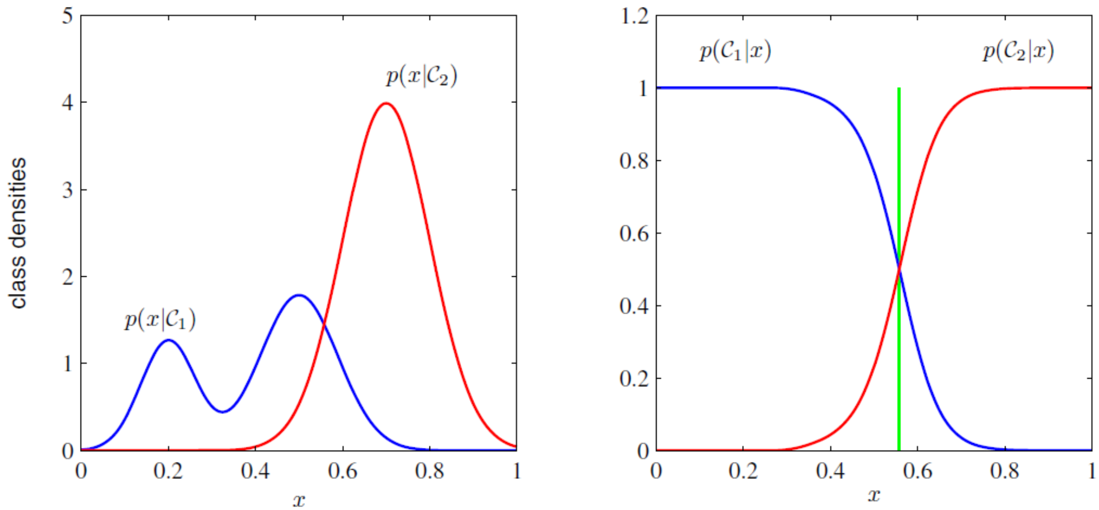


Inferenz und Entscheidung

- Klassifikation besteht aus zwei Schritten
  - Inferenz: Bestimmen von $p(\mathcal{C}_k|\mathbf{x})$ aus Trainingsdaten
  - Entscheidung: $\hat{\mathcal{C}} = \underset{k}{\operatorname{argmax}}\ p(\mathcal{C}_k|\mathbf{x})$

$$p(\mathbf{x}, \mathcal{C}_k) = p(\mathcal{C}_k|\mathbf{x})\ p(\mathbf{x}) \qquad p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)\ p(\mathcal{C}_k)}{p(\mathbf{x})}$$

Satz von Bayes!!



Generativ vs. Diskriminativ

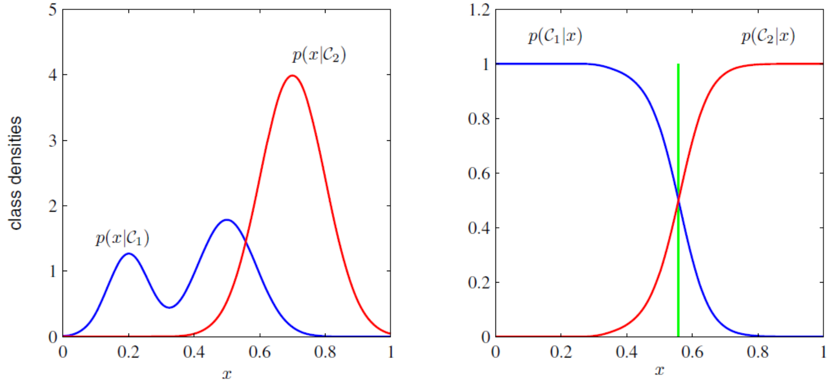
Generative Modelle

- Berechnung der Aposteriori-Wahrscheinlichkeiten $P(C_k|\mathbf{x})$ aus den klassenspezifischen Merkmalsverteilungen $p(\mathbf{x}|C_k)$ und den Apriori-Verteilungen $P(C_k)$

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)\ p(\mathcal{C}_k)}{p(\mathbf{x})}$$

Diskriminative Modelle

- Direkte Modellierung der Aposteriori-Wahrscheinlichkeit $P(C_k|\mathbf{x})$ oder einer Funktion $\hat{C}=f(\mathbf{x})$



Vorteile generativer Modelle

Kenntnis von $P(C_k|\mathbf{x})$

- Risikominimierung
- Rückweisung

$$p(\mathcal{C}_k|\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{C}_k)\ p(\mathcal{C}_k)}{p(\mathbf{x})}$$

Kenntnis von $p(\mathbf{x}|C_k)$ und $P(C_k)$

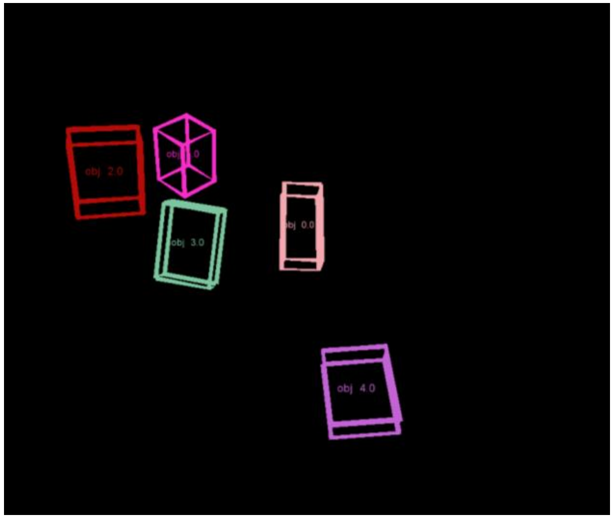
- Kompensation unterschiedlicher apriori-Verteilungen der Klassen in Trainings und Testdaten
- Modularisierung komplexer Probleme durch Zerlegen entsprechend der bedingte Unabhängigkeit

$$\begin{aligned} p(\mathcal{C}_k|\mathbf{x}_\mathrm{I}, \mathbf{x}_\mathrm{B}) &\propto p(\mathbf{x}_\mathrm{I}, \mathbf{x}_\mathrm{B}|\mathcal{C}_k)p(\mathcal{C}_k) \\ &\propto p(\mathbf{x}_\mathrm{I}|\mathcal{C}_k)p(\mathbf{x}_\mathrm{B}|\mathcal{C}_k)p(\mathcal{C}_k) \\ &\propto \frac{p(\mathcal{C}_k|\mathbf{x}_\mathrm{I})p(\mathcal{C}_k|\mathbf{x}_\mathrm{B})}{p(\mathcal{C}_k)} \end{aligned}$$

Für die Objekterkennung (und viele andere Zwecke):

„LOCAL FEATURES“

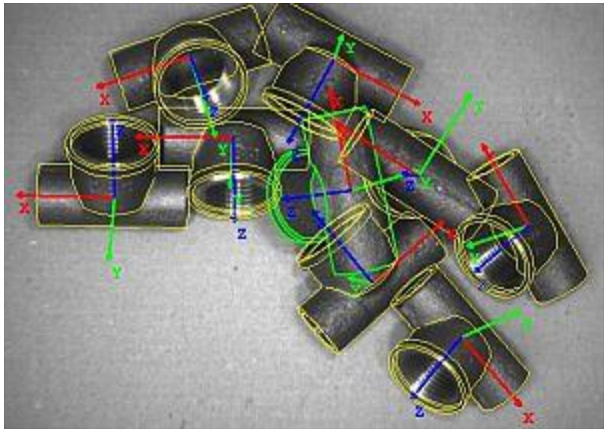
Szenenanalyse



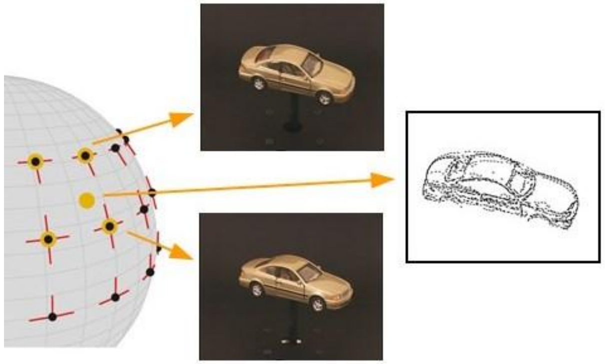
- Wahrnehmung des Umgebungszustands
- z.B.: Wo sind welche Objekte?
  - Segmentierung
  - Erkennung
  - Lokalisierung/ Lagebestimmung

Ansätze für die Objekterkennung

- Geometrie
  - CAD-Modell
  - Kantenextraktion



- Ansicht
  - Aus allen möglichen Richtungen
  - Numerische Beschreibung der Ansicht als Merkmalsvektor
    - z.B. Momenteninvarianten
  - Klassifikation (siehe oben)



Ansätze für die Objekterkennung

- Momenteninvarianten als Formmerkmale

$$M_{ij} = \sum_x \sum_y x^i y^j I(x,y)$$

  - Verschiebungsinvarianz

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x,y)$$

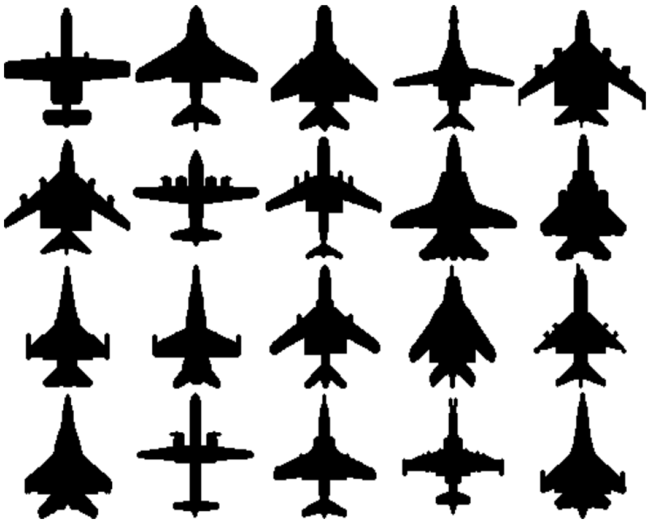
$$\bar{x} = \frac{M_{10}}{M_{00}} \qquad \bar{y} = \frac{M_{01}}{M_{00}}$$

  - Skalierungsinvarianz

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{\left(1+\frac{i+j}{2}\right)}}$$

  - Rotationsinvarianz ⇨

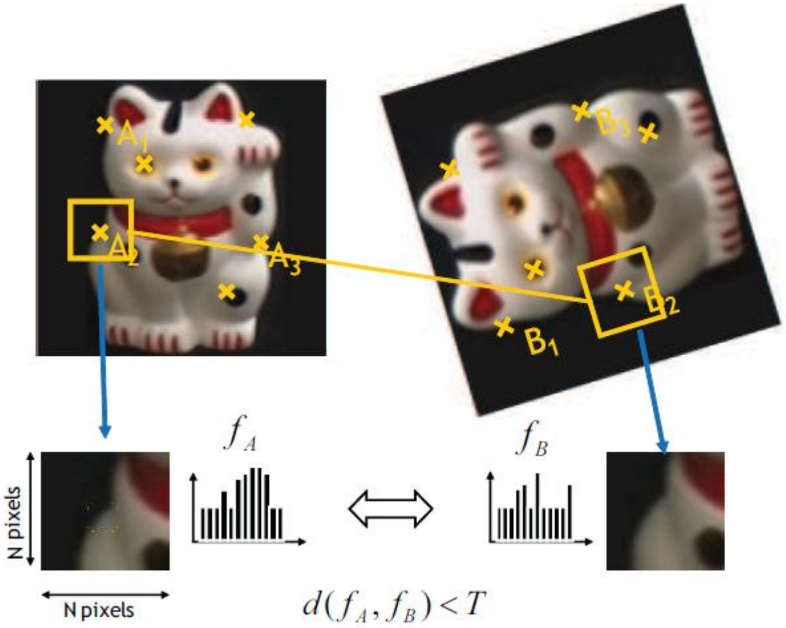
$$\Theta = \frac{1}{2} \arctan\left(\frac{2\mu'_{11}}{\mu'_{20} - \mu'_{02}}\right)$$

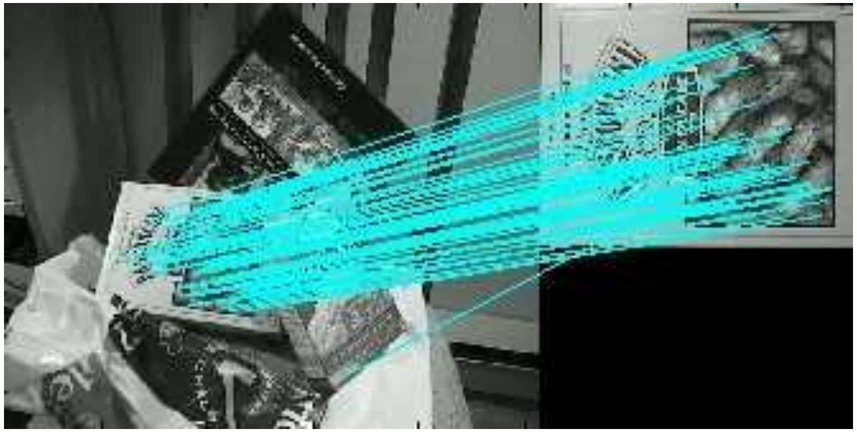


$$\begin{aligned}
I_1 &= \eta_{20} + \eta_{02} \\
I_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
I_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\
I_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\
I_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
I_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
I_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]. \\
I_8 &= \eta_{11}[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2] - (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})
\end{aligned}$$

Local Features: Das Grundprinzip

1. Bestimmen charakteristischer Punkte (key points / interest points)
2. Definition einer charakteristischen Region um den jeweiligen Keypoint
   - skaleninvariant
   - rotationsinvariant
   - …
3. Extraktion und Normalisierung der charakteristischen Region
4. Berechnung eines Deskriptors (Merkmalsvektors) für die Region
5. Erkennung der Region durch Deskriptorvergleich
6. Erkennung von Objekten durch erkennen vieler Regionen



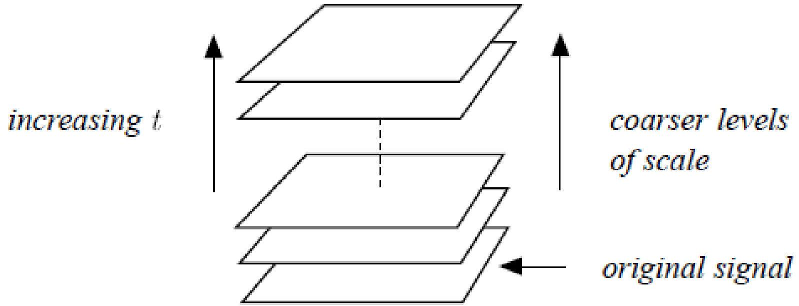


Scale space (Skalenraum)

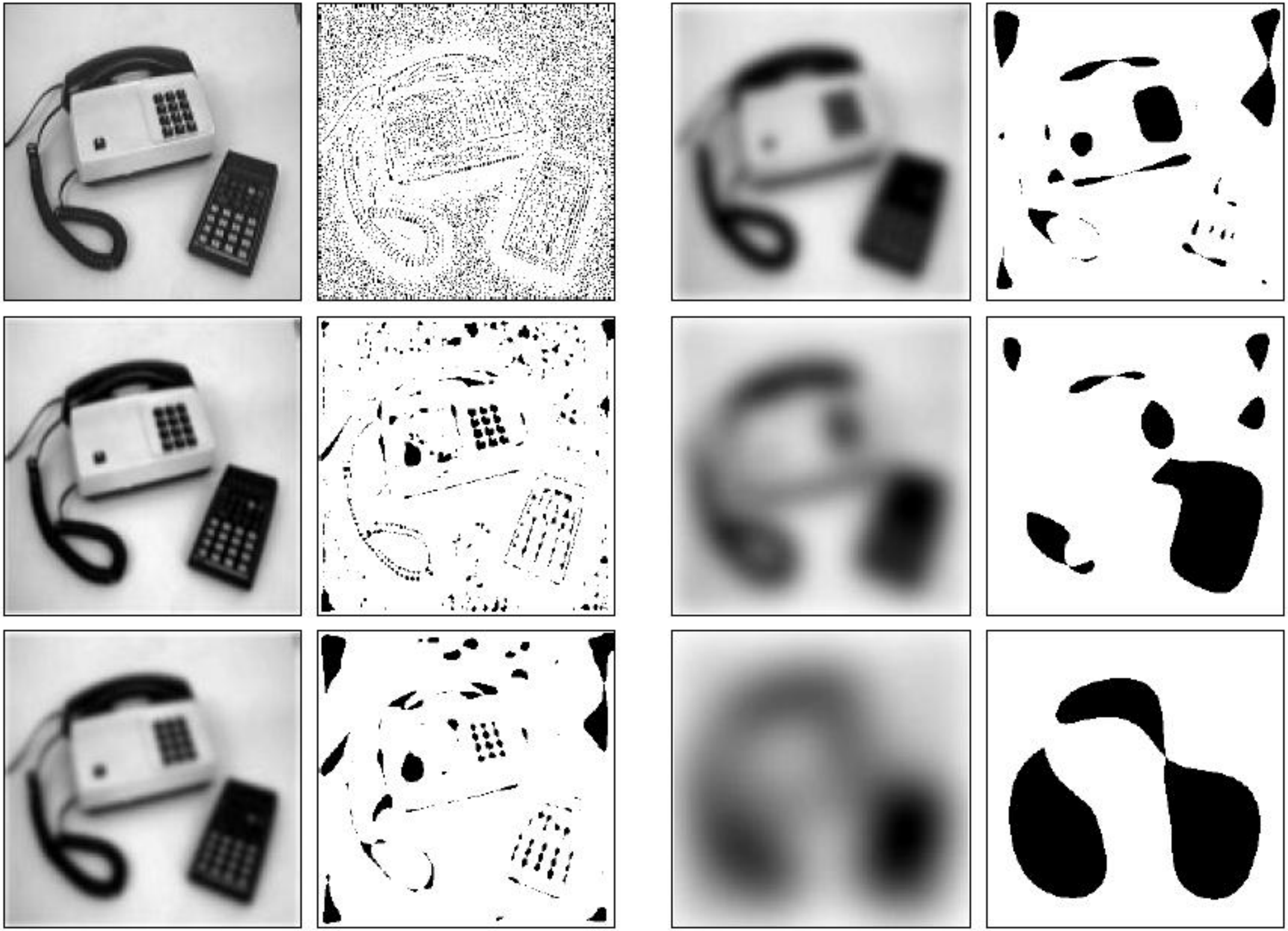
- Faltung des Bildes mit einem Gauß-Kern variabler „Breite“ *t*

$$g(x, y; t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}$$

$$L(\cdot, \cdot; t) \ = g(\cdot, \cdot; t) * f(\cdot, \cdot),$$



T. Lindeberg (1994). "Scale-space theory: A basic tool for analysing structures at different scales". *Journal of Applied Statistics (Supplement on Advances in Applied Statistics: Statistics and Images: 2)* **21** (2). pp. 224–270.



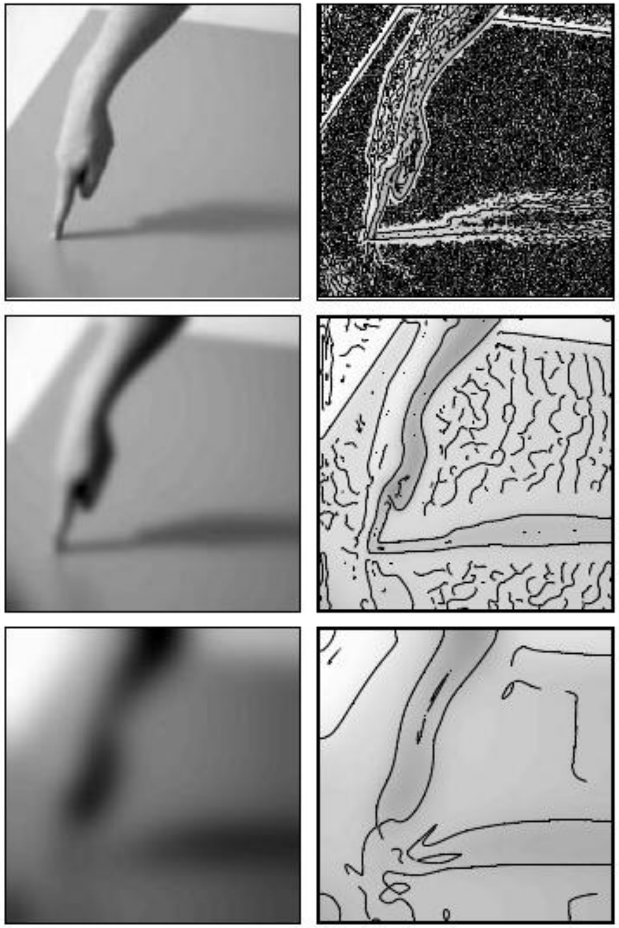
Scale space (Skalenraum)

- Differentielle Maße im Skalenraum über Gauß'sche Richtungsableitungen

$$L_{x^m y^n}(x, y; t) = (\partial_{x^m y^n} L)(x, y; t).$$

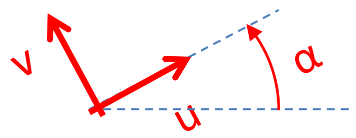
$$L_{x^m y^n}(\cdot, \cdot; t) = \partial_{x^m y^n} g(\cdot, \cdot; t) * f(\cdot, \cdot).$$

- L_{yy}: zweite Ableitung nach y



Skalenraum-Kanten

- Lokale (u,v)-Koordinaten

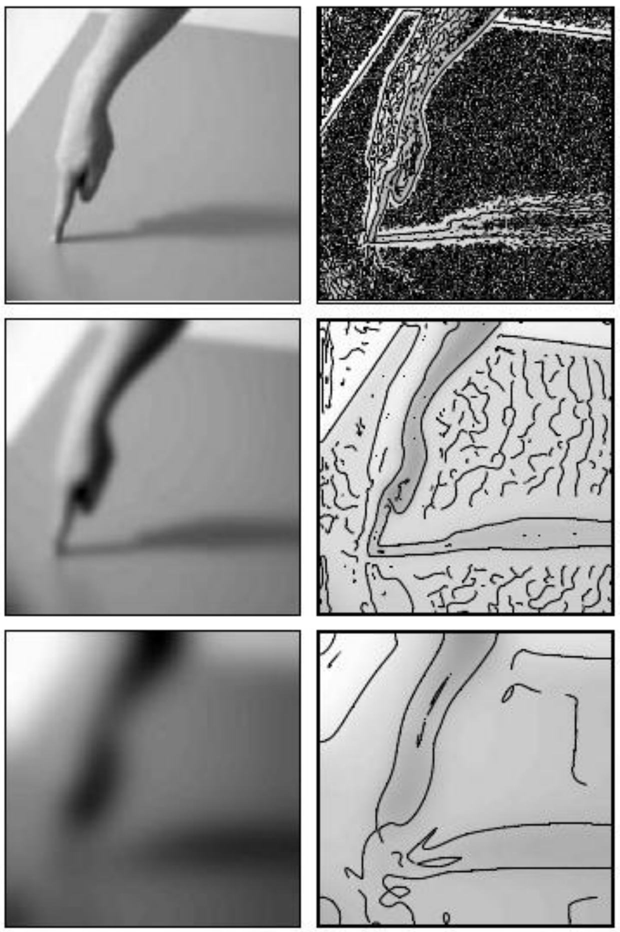


$$(\cos\alpha, \sin\alpha) = (L_x, L_y)/\sqrt{L_x^2 + L_y^2}$$

$$\partial_{\bar{u}} = \sin\alpha\,\partial_x - \cos\alpha\,\partial_y$$

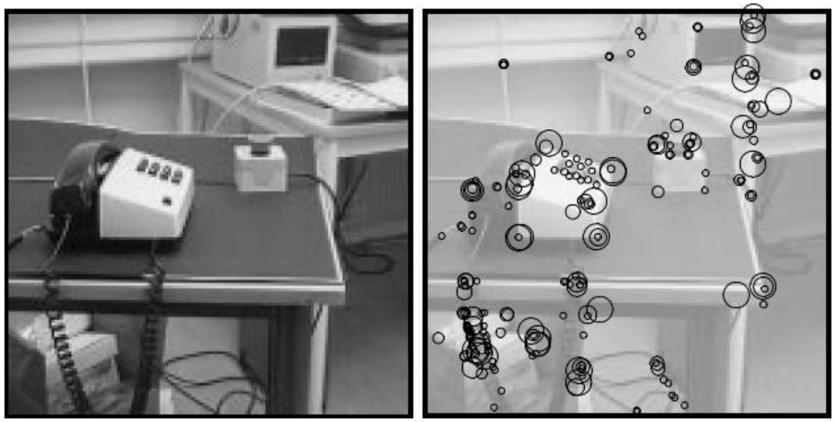
$$\partial_{\bar{v}} = \cos\alpha\,\partial_x + \sin\alpha\,\partial_y$$

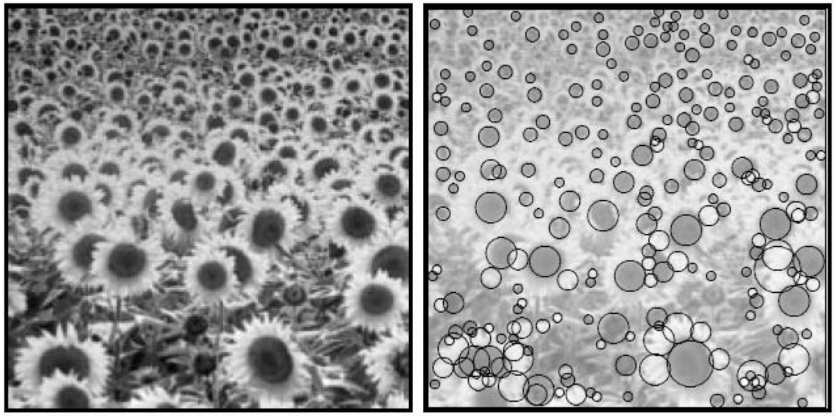
- Kantendetektion: $\begin{cases} L_{vv} = 0, \\ L_{vvv} < 0, \end{cases}$



Skalenraum-Merkmalspunkte

- Markante Punkte im Bild (Interest-Points)
- Lokalisierbar in x und y
  - Kanten ungeeignet
- Starke Änderung des bildes in mehren Richtungen
  - Ecken
  - Blobs



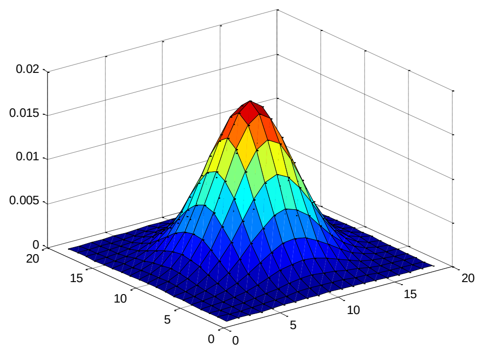


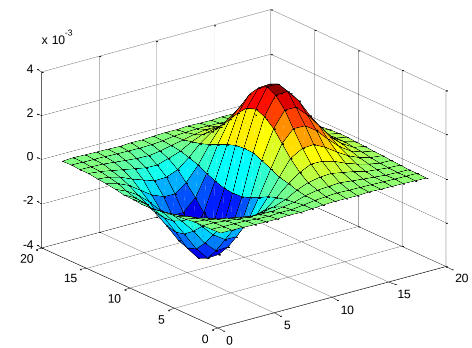
Gauß'sche Richtungsableitungen

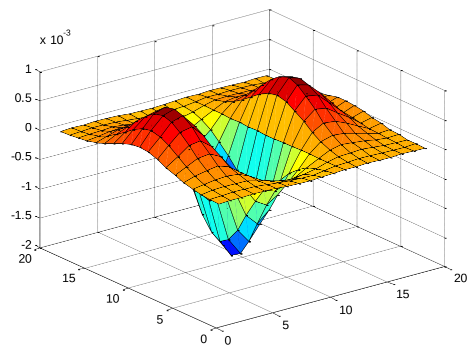
Erinnerung!!!

- Erst glätten, dann ableiten
  - Entspricht Faltung mit einen abgeleiteten Glättungskern

$$\frac{\partial}{\partial x}(f(x,y) * h(x,y)) = f(x,y) * \frac{\partial}{\partial x}h(x,y)$$





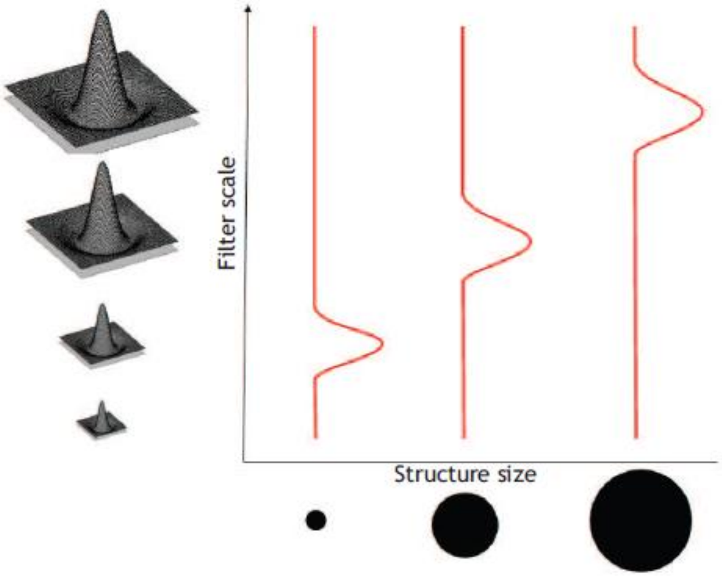


1. Ableitung

2. Ableitung

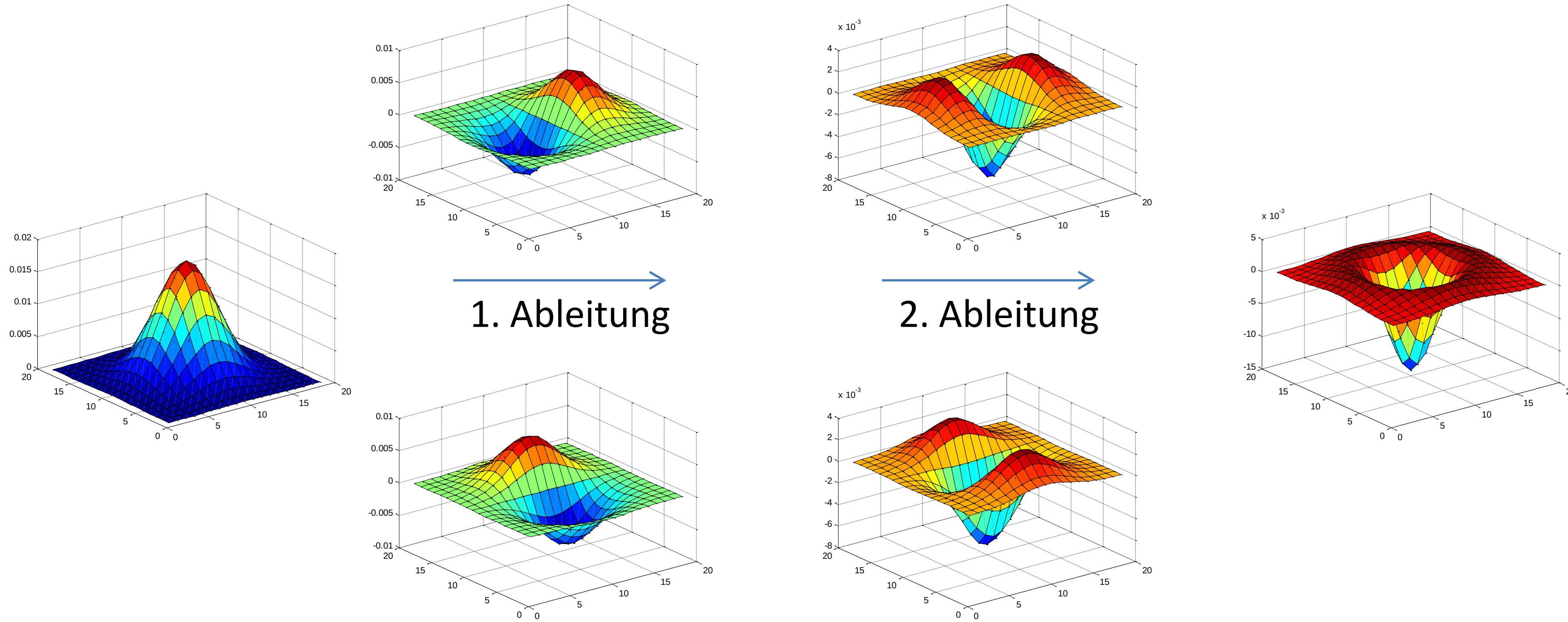
Skalenraum-Blobs

- Filterung mit einem punktförmigen Filter
- Antwort ist Skalenabhängig
  - Bei geeigneter Normierung
  - „Laplacian“



$$\nabla^2 L(\mathbf{x}, t) = t(L_{xx}(\mathbf{x}, t) + L_{yy}(\mathbf{x}, t))$$

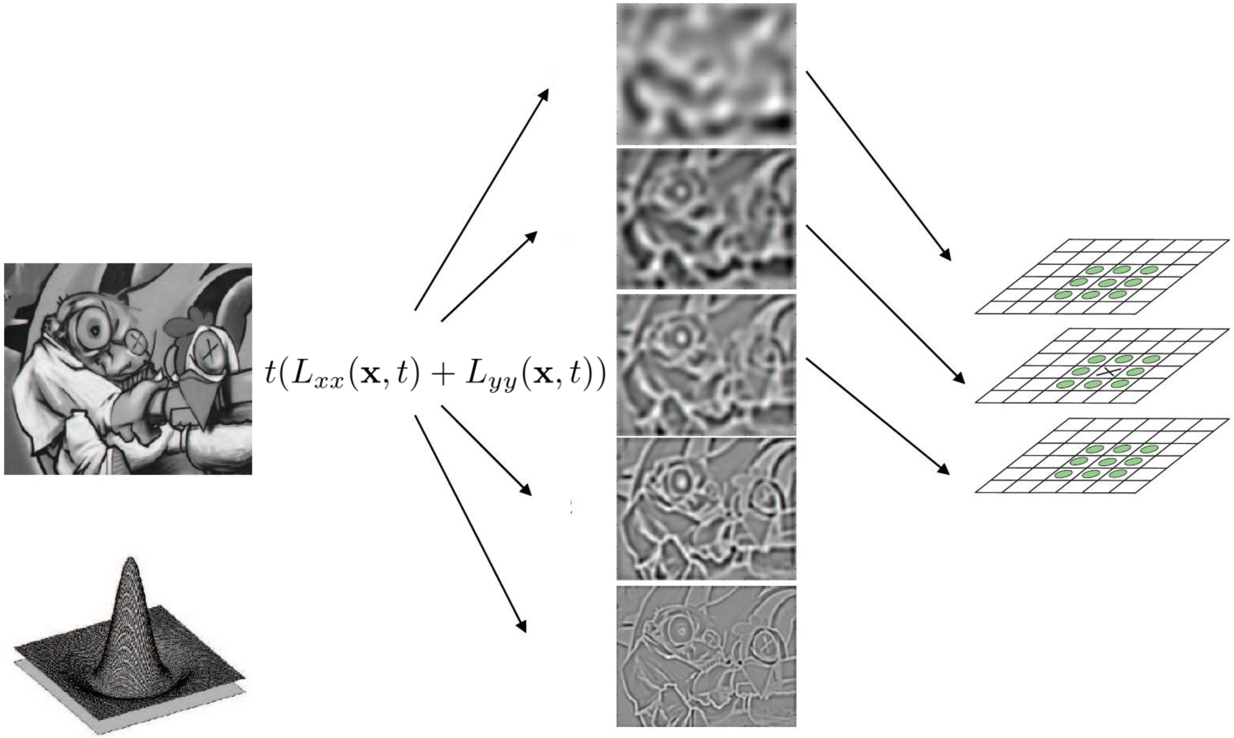
„Laplacian of the Gaussian“



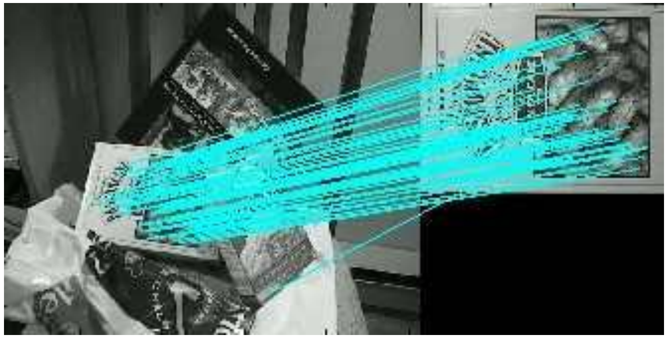
$$\nabla^2 L(\mathbf{x}, t) = t(L_{xx}(\mathbf{x}, t) + L_{yy}(\mathbf{x}, t))$$

Blobsuche im Skalenraum

- Automatische Skalenauswahl
  - Maximum von $\nabla^2 L(\mathbf{x}, t)$ in x,y, und *t*



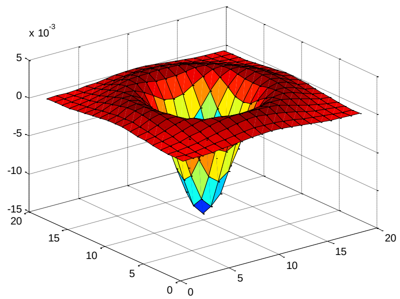
Scale Invariant Feature Transform (SIFT)



- **Sehr beliebter Algorithmus**
  - **Funktioniert hervorragend!**
- Zwei Schritte
  - Blobdetektion
  - Berechnung spezieller Deskriptoren für die Umgebung der Blobs
- Approximation des Laplacian of Gaussian (LoG) für die Blobdetektion

D. G. Lowe, *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, vol. 2, no. 60, pp. 91-110, 2004

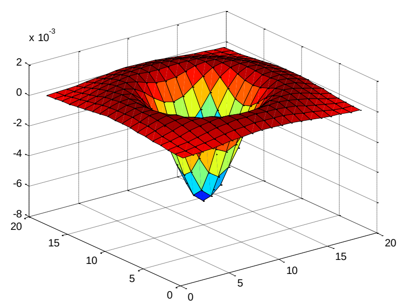
LoG-Approximation



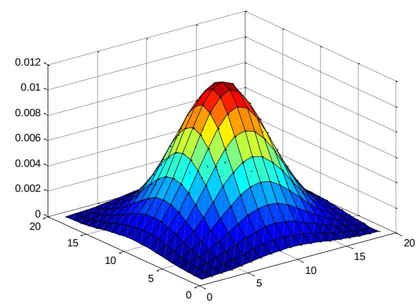
$$\nabla^2 L(\mathbf{x}, t) = t(L_{xx}(\mathbf{x}, t) + L_{yy}(\mathbf{x}, t))$$

Difference of Gaussians (DoG)

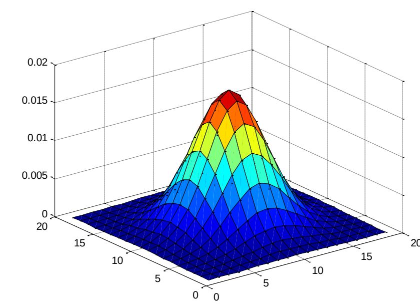
$$D(\mathbf{x}, t) = G(\mathbf{x}, kt) - G(\mathbf{x}, t)$$



$=$

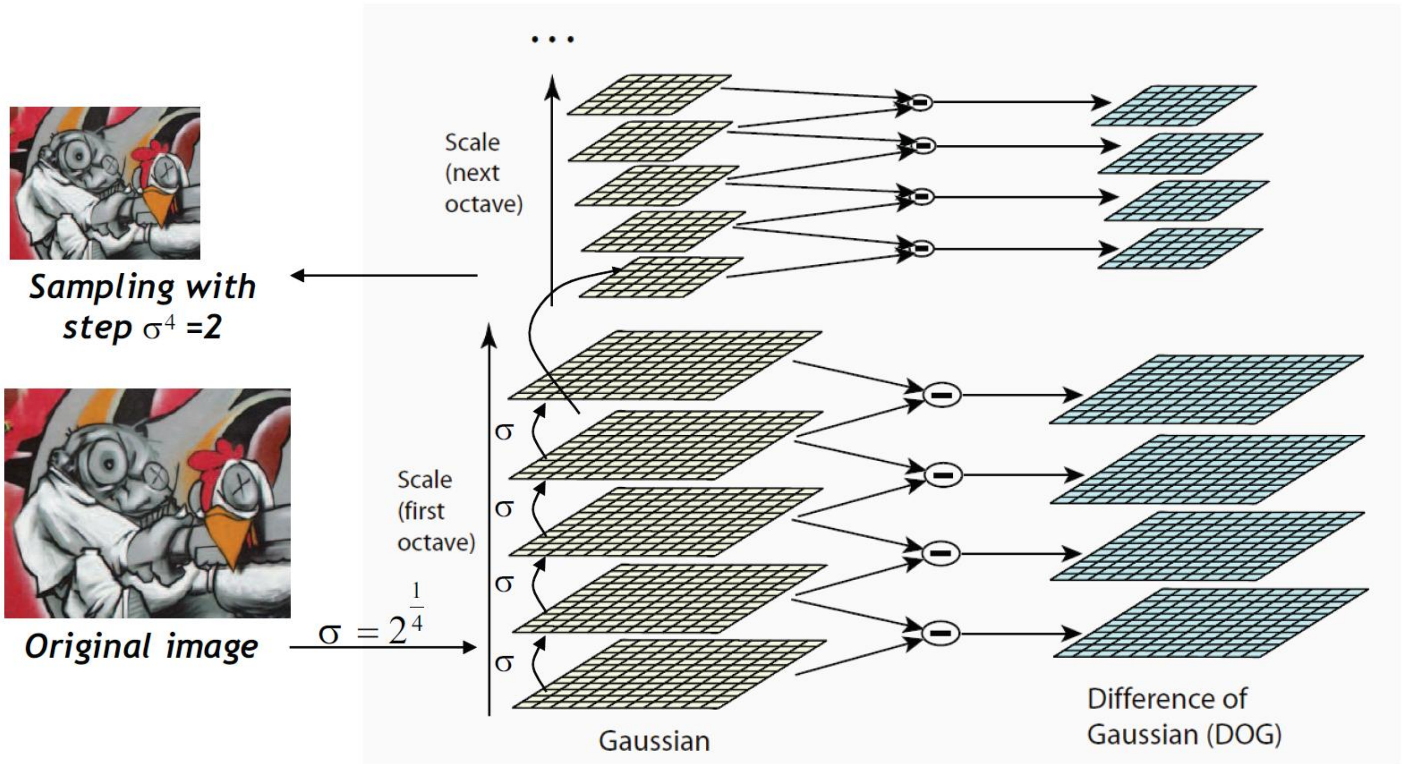


$-$



Blobdetektion

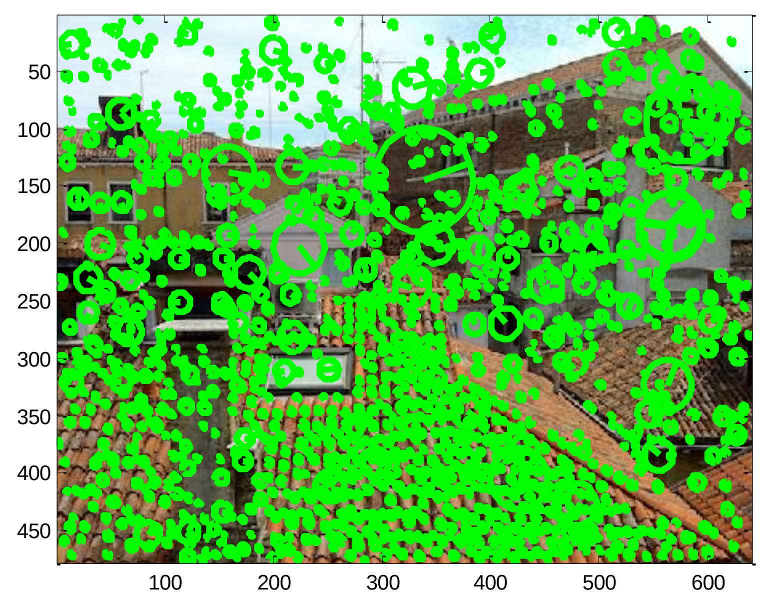
- Automatische Skalenauswahl
  - Maxima von $D(\mathbf{x}, t)$ in ***x***, und *t*
  - DoG innerhalb der Oktave, dann Unterabtastung (Bild-Pyramide zur Beschleunigung)



Blobdetektion

- Für gute Lokalisierung der Features muß zwischen den diskreten Pixeln und den diskretisierten Skalen interpoliert werden
  - Lokaler Fit einer quadratischen Funktion (Quadrik)
  - Finden des Maximums dieser Funktion

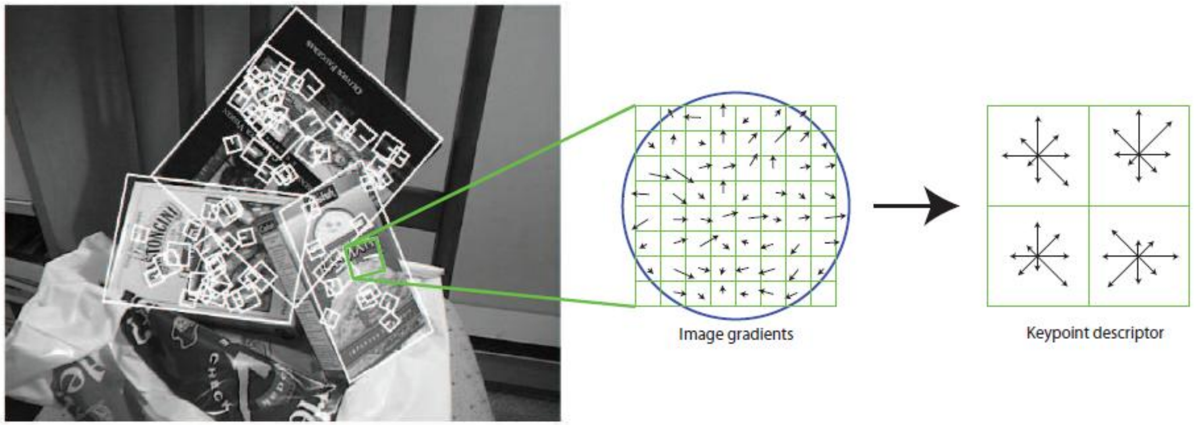
N = 1816

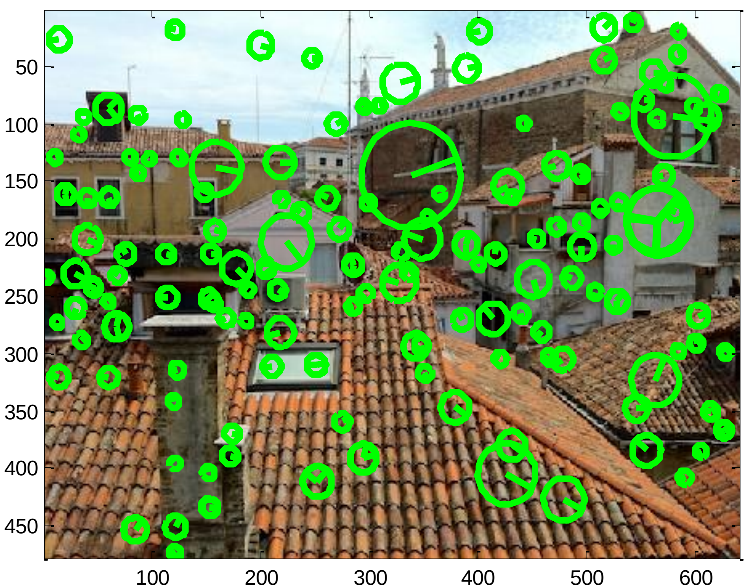


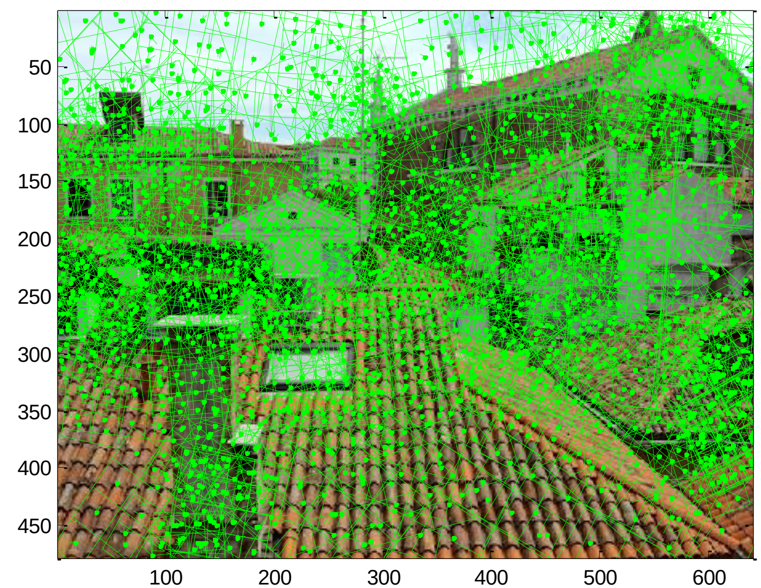
Orientierung

- Bestimmung der dominanten Gradienrenrichtung
- Für alle Pixel in der extrahierten Region
  - Berechnung der Gradientenrichtung auf der jeweiligen Skala
  - Bestimmung eines Histogramms mit 10 Grad Auflösung
  - Finden des Maximums (Parabelfit in 3 benachbarten Bins)
  - Falls mehrere gleich starke Orientierungen gefunden werden, wird der Punkt mehrfach zurückgegeben

Berechnung der Deskriptoren

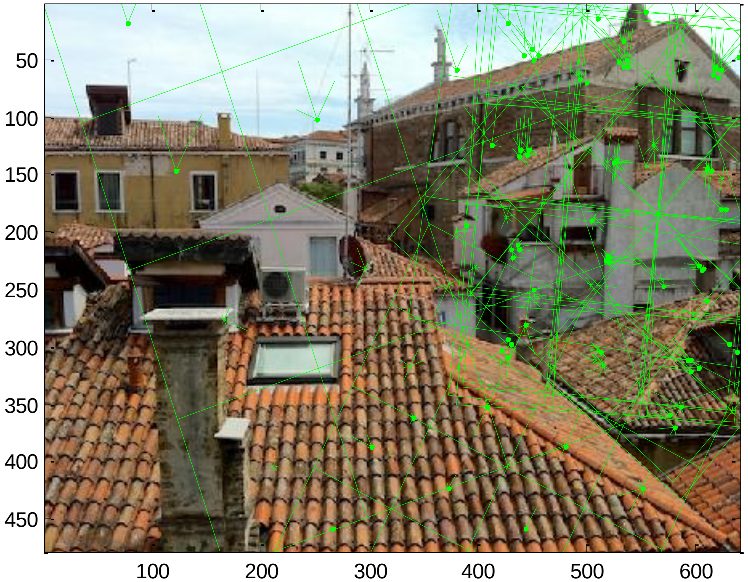






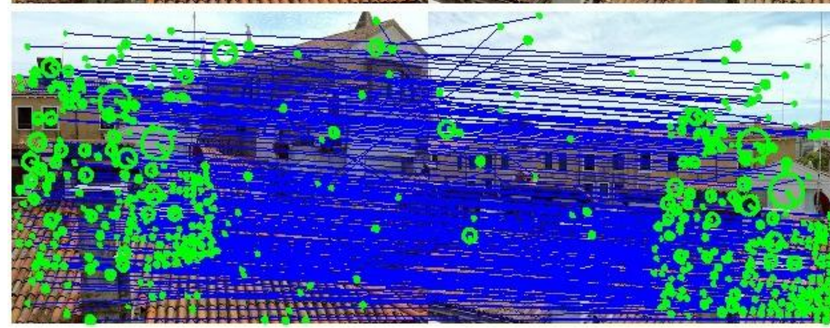
Berechnung der Deskriptoren

- In der skalierten und rotierten Umgebung des Featurepunkts werden Bildgradientenrichtungen berechnet und in einem 4x4-Fenster in Histogramme eingetragen (8 Bins)
- Damit ergibt sich ein 128-dimensionaler Deskriptor-vektor für jeden Featurepunkt

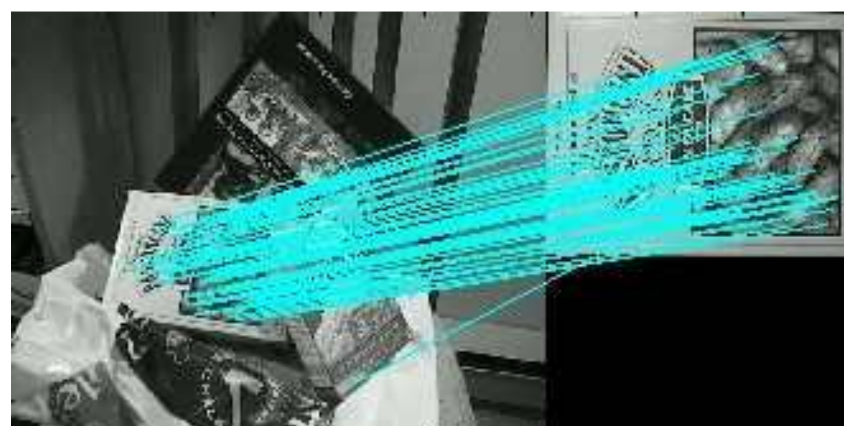


Korrespondenzfindung in verschiedenen Ansichten

- Vergleich der Deskriptoren über euklidischen Abstand



- Skaleninvarianz
  - Größenunterschiede der Bilder
- Rotationsinvarianz
  - In der Bildebene beliebig
  - sonst ca. 10-20°



Warum ist funktioniert das?

- Vergleich **vornormierter** Bildbereiche (Größe, Rotation)

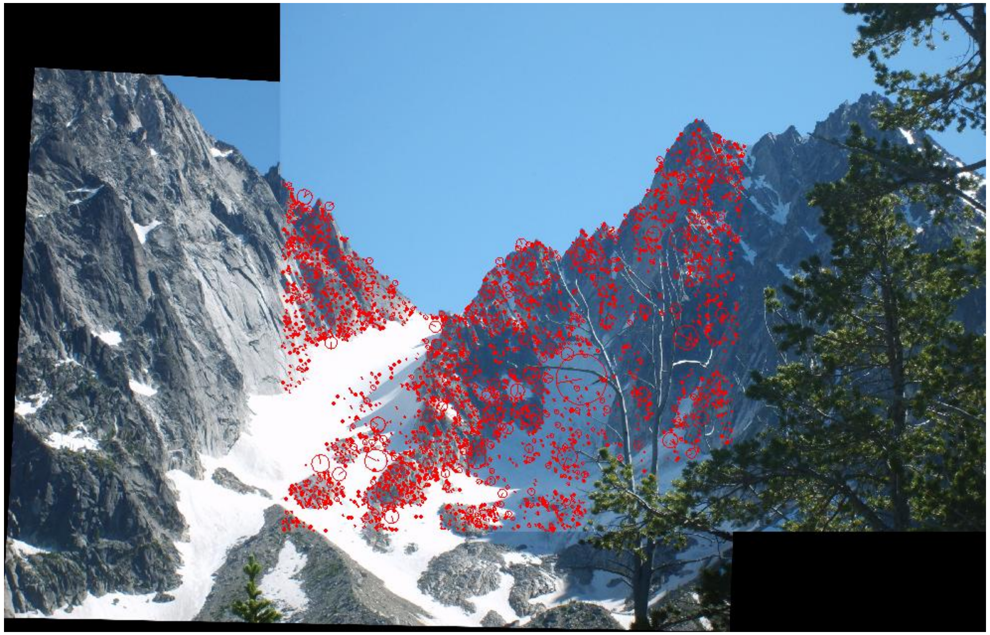
- Offensichtlich ist der SIFT-Deskriptor auf natürlichen Bildern sehr spezifisch
  - Deskriptor: Verteilung der Gradienterichtungen in der normierten Umgebung

- Statt das gesamte Bild zu betrachten konzentrieren sich Verarbeitung und Analyse auf die Bereiche, wo auch etwas zusehen ist

SIFT-Anwendungen

- Objekterkennung
  - 2D und 3D

- Szenenvergleich
  - Ansichstbasierte Navigation

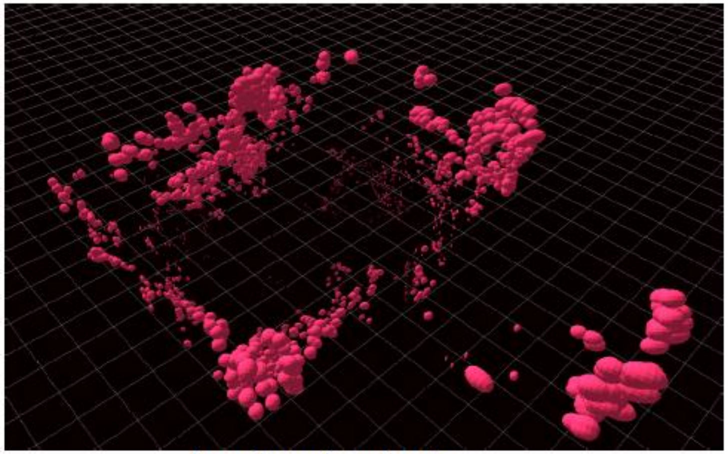


- Image Stitching
  - vollautomatisch

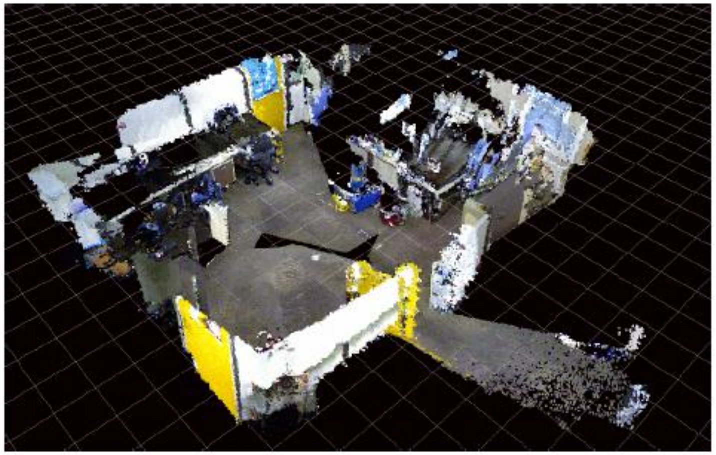
SIFT-Anwendungen

- SIFT-Punkte als Landmarken für die Navigation

- SIFT extrem beliebt, wie immer sind zahllose Varianten verfügbar



Landmarkenkarte



Punktwolke