

# Bildmerkmale am Beispiel SIFT

## Medieninformatik IL

Andreas Unterweger

Vertiefung Medieninformatik und Bildverarbeitung  
Studiengang ITS  
FH Salzburg

Wintersemester 2023/24

# Bildmerkmale (engl. *features*) I

- Beschreiben markante Stellen eines Bildes (so genannte Schlüsselpunkte) und deren Umgebung
- Anwendungen (Auswahl):
  - Bildinhalte beschreiben
  - Objekte in Bildern (wieder-)finden
  - Bilder zusammenfügen
  - Ähnliche Bilder suchen
- Anforderungen:
  - **Robustheit** (Stabilität)
  - Kompaktheit (geringer Speicherbedarf)
  - Effizienz (geringer Zeitbedarf für Berechnung)

- Robustheitsgrade:
  - Invarianz
  - Quasi-Invarianz
  - Toleranz
  - Intoleranz
- Robustheitsdimensionen:
  - Translation
  - Rotation
  - Skalierung
  - Allgemeine affine Transformationen
  - Perspektivische Transformationen
  - Beleuchtung

- SIFT: Scale-invariant Feature Transform
- Bestandteile
  - Merkmalerkennung
  - Merkmalvergleich
- Quasi-invariant gegenüber
  - Translation
  - Rotation
  - Isotroper Skalierung
- Tolerant gegenüber
  - Perspektivischen Transformationen
  - Beleuchtung

- Ablauf Merkmalerkennung
  - Schlüsselpunktkandidatensuche
  - Schlüsselpunktkandidatenfilterung
  - Schlüsselpunktorientierungszuordnung
  - Schlüsselpunktdeskriptorerstellung
- Ablauf Merkmalvergleich (von erkannten Merkmalen zweier Bilder)
  - Merkmalübereinstimmungen finden
  - Falschzuordnungen erkennen und entfernen
  - Verifikation und Korrektur der Zuordnung (iterativ)
- Begriffsdefinitionen
  - Orientierung: Ausrichtung des Schlüsselpunktes (ermöglicht Neuordnung zur Erreichung von Quasi-Rotationsinvarianz)
  - Deskriptor: Merkmalvektor mit Informationen zum Schlüsselpunkt

- Motivation: Skalierungsinvarianz durch Modellierung des Scale (zu deutsch etwa Maßstab) erreichen
- Idee: Ein Bild  $L(x, y)$  in einem Scale  $\sigma$  entspricht einem weichgezeichneten (mit  $G$  gefalteten) Ausgangsbild  $I(x, y)$  mit Weichzeichnungsstärke  $\sigma$

$$L(x, y, \sigma) = I(x, y) * G(x, y, \sigma)$$

- $G$  ist zweidimensionales Gauß-Filter mit variablem  $\sigma$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

# Einschub: Scale Space II



$\sigma = 0$



$\sigma = 1$



$\sigma = 4$



$\sigma = 16$

Quelle: Wikipedia contributors: Scale space. *Wikipedia, The Free Encyclopedia* [http://en.wikipedia.org/wiki/Scale\\_space](http://en.wikipedia.org/wiki/Scale_space) (18.7.2014), 2014

# Schlüsselpunkt-kandidatensuche I

- Idee: Menschliche Wahrnehmung nachahmen
- Überlegung: Details/hochfrequente Anteile enthalten viel Information
- Umsetzung: Difference of Gaussians (DoG)  $D$  mit  $k \in \mathbb{R}$ :

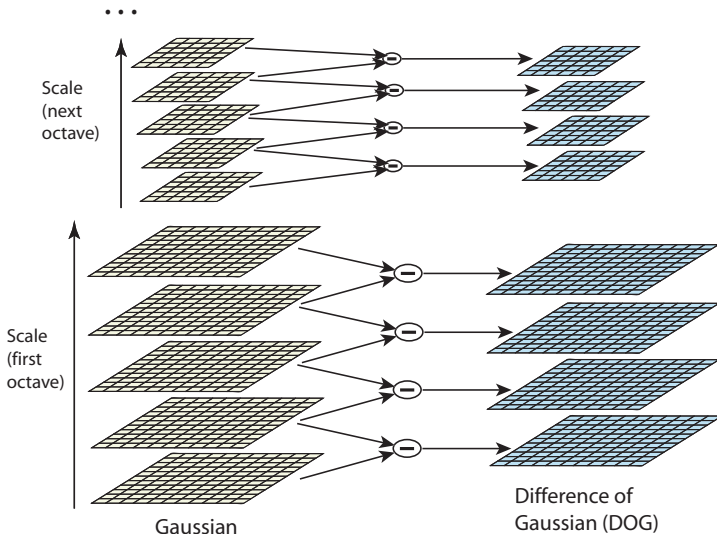
$$D(x, y, k) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$



Quellen: [http://en.wikipedia.org/wiki/File:Flowers\\_before\\_difference\\_of\\_gaussians.jpg](http://en.wikipedia.org/wiki/File:Flowers_before_difference_of_gaussians.jpg);  
[http://commons.wikimedia.org/wiki/File:Flowers\\_after\\_difference\\_of\\_gaussians\\_grayscale.jpg](http://commons.wikimedia.org/wiki/File:Flowers_after_difference_of_gaussians_grayscale.jpg)



# Schlüsselpunktkandidatensuche II



Quelle: Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

- Ohne Beweis: Scale-Invarianz durch Laplacian of Gaussians (LoG)  
 $\nabla^2 G$  mit zusätzlichem Faktor  $\sigma^2$  möglich:

$$\nabla^2 G = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

→ Approximation von LoG durch DoG

- Scale Space erfüllt Diffusionsgleichung:

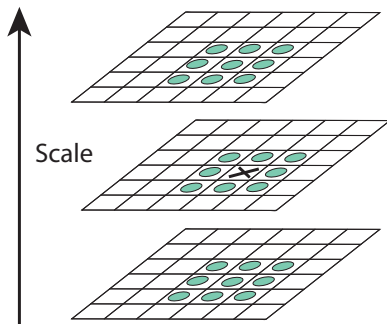
$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma}$$

- Diskretisierung erlaubt vereinfachte numerische Ableitung:

$$\rightarrow \sigma \nabla^2 G \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \rightarrow D(x, y, k) \approx (k-1)\sigma^2 \nabla^2 G * I(x, y)$$

# Schlüsselpunktkandidatensuche IV

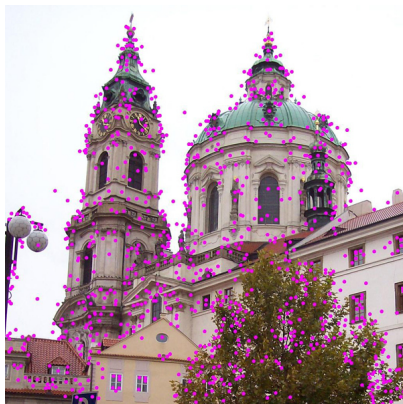
- Ohne Beweis: Scale-Space-Extrema (Minima und Maxima von  $D$ ) sind gute Kandidaten für Schlüsselpunkte
- Scale-Invarianz als Nebeneffekt von LoG-Approximation
- Diskreter Suchraum erlaubt Extremafindung durch Vergleich eines Pixels von  $D$  mit dessen direkten Nachbarn in allen drei Dimensionen



Quelle: Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

# Schlüsselpunktkandidatenfilterung

- Kandidatenfilterungsprozess (ohne Details):
  - Kandidaten mit niedrigem Kontrast entfernen (anfällig für Rauschen)
  - Kandidaten entlang von Kanten entfernen (Lokalisation schwierig)
  - Interpolation rund um Extremum erlaubt exaktere Positionsermittlung



Adaptiert von [http://commons.wikimedia.org/wiki/File:Sift\\_keypoints\\_filtering.jpg](http://commons.wikimedia.org/wiki/File:Sift_keypoints_filtering.jpg)

# Schlüsselpunktorientierungszuordnung I

- Beträge  $A$  und Winkel  $\varphi$  von Gradienten in Region um Schlüsselpunkt werden durch numerische Ableitung genähert
- Verwendung von  $L'(x, y) = L(x, y, \sigma_P)$  zur Beibehaltung von Scale-Invarianz ( $\sigma_P$  ist Scale von Schlüsselpunkt  $P$ )

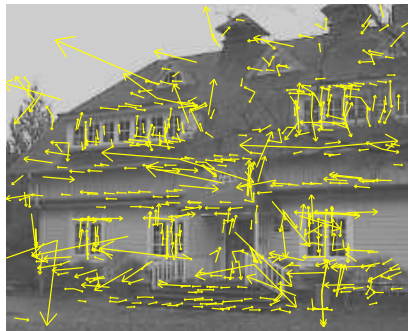
$$A(x, y) \approx \sqrt{(L'(x+1, y) - L'(x-1, y))^2 + (L'(x, y+1) - L'(x, y-1))^2}$$

$$\varphi(x, y) \approx \arctan \left( \frac{L'(x, y+1) - L'(x, y-1)}{L'(x+1, y) - L'(x-1, y)} \right)$$

- Winkel-Histogramm mit 10-Grad-Unterteilung wird erstellt
- Gradientenbeträge fließen entfernungsgewichtet ein (ohne Details)
- Maximum ist dominante Orientierung  $\rightarrow$  Rotations(-quasi-)invarianz
- Für Werte bis 80% des Maximums: Zusätzliche Schlüsselpunkte

# Schlüsselpunktorientierungszuordnung II

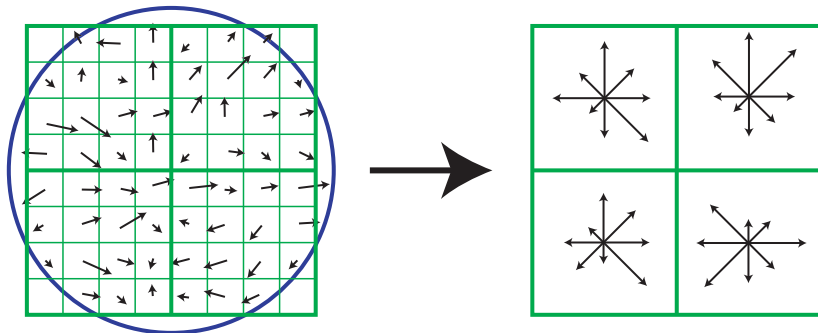
- Bisher bestimmte Schlüsselpunkteigenschaften:
  - Position (für Translationsinvarianz)
  - Scale (für Scale-Invarianz)
  - Orientierung (für Rotationsinvarianz)



Quelle: Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

# Schlüsselpunktdeskriptorerstellung

- 4 · 4 Orientierungshistogramme mit 45-Grad-Unterteilung aus Gradienten in 16 · 16-Region berechnen (vereinfachte Skizze)
- Vektor aus Histogrammwerten erstellen und (vereinfacht) Vektor normalisieren (für Beleuchtungsinvarianz)



Quelle: Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.





- Ausschluss von unpassenden Zuordnungen
  - Bestimmung der Pose (relative Position, Scale und Orientierung) jedes zugeordneten Schlüsselpunktes (mit Fehlertoleranz)
  - Gruppierung von Schlüsselpunkten, die in ihrer Pose übereinstimmen → Höhere Wahrscheinlichkeit für korrekte Zuordnung (ohne Details)
  - Verwerfen von Zuordnungsgruppen mit weniger als drei Zuordnungen  
→ Menge von Zuordnungsgruppen
- Geometrische Verifikation (für jede Zuordnungsgruppe)
  - Vereinfachte Annahme: Zuordnungen werden durch affine Transformation (mit Fehlertoleranz) hinreichend genau beschrieben
  - Unbekannte affine Transformationsmatrix aus linearem Gleichungssystem (mit Fehlertoleranz) bestimmen (ohne Details)
  - Zuordnungen mit zu großem Fehler verwerfen und Zuordnungsgruppe bei weniger als drei übrigen Zuordnungen verwerfen
  - Schritte bei Bedarf wiederholen (bis nichts mehr verworfen wird)

# Merkmalvergleich III

- Abschließende Zuordnung: Aus verbleibenden Gruppen anhand der Anzahl der verbleibenden Zuordnungen pro Gruppe und der Größe des Fehlers die wahrscheinlichste Zuordnung auswählen (ohne Details)



Quelle: Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

- Nachteile von SIFT
  - Nicht robust gegenüber einigen Verzerrungsarten
  - Relativ langsam (trotz vieler Optimierungen)
  - Lange Zeit (bis ca. März 2020) patentrechtlich geschützt
- Alternative Verfahren (Auswahl):
  - SURF (Speeded Up Robust Features) – patentrechtlich geschützt
  - ORB (Oriented FAST<sup>1</sup> and Rotated BRIEF<sup>2</sup>)
  - BRISK (Binary Robust Invariant Scalable Keypoints)
  - FREAK (Fast Retina Keypoint)
  - HOG (Histogram of Oriented Gradients)
  - Diverse Eckendetektoren (z.B. FAST und Harris)

---

<sup>1</sup>Features from Accelerated Segment Test

<sup>2</sup>Binary Robust Independent Elementary Features

Fragen?