Open World Anwendungen der Bildsuche

Vorlesung Bildverarbeitung 2, Teil 4
Wintersemester 2001/2002
Ullrich Köthe, FB Informatik, Uni Hamburg

Gliederung

- **Farbsuche**
  - Wiederholung: Abstandsmaße für Farben
  - Erstellung einer Farbsignatur: Farbquantisierung
- **Textursuche**
  - Texturmerkmale (Laws- und Gabor-Filter, co-occurence Matrizen, autoregressive Modelle, WOLD-Merkmale)
  - Abstandsmaße für Texturen
  - Texturklassifikation
- **Suche nach Formmerkmalen**
  - Formmerkmale (Momente, Konturpolygon, shape scale-space)
  - Abstandsmaße: Hausdorff-Distanz, affine Distanz, Biegeenergie
- **Suche mit Hilfe von semantischen Konzepten**

Farbsuche

- **Transformation aus RGB in L*a*b* / L*u*v* / Y´PbPr**
  - RGB oder R´G´B´ ?
  - Euklidische Distanz näherungsweise wahrnehmungskonform
- **Erstellung eines Histogramms mit fester Quantisierung des Farbraums (typisch ~200 Farben): über 100 verschiedene Algorithmen(-Varianten)**
- **oder Erstellung einer Signatur durch adaptive Farbquantisierung (~ 10 Einträge)**
- **Vergleich mit Hilfe der *earth mover distance* oder der Mahalanobis-Distanz (nur Histogramme)**
  - Verwendung der Euklidischen Distanz zwischen den Clustern bzw. Bins als Basisdistanz

Farb-Clusterung

- **einfacher Algorithmus: *median cut* (Heckbert 82)**
  1. bestimme 3D Farbhistogramm mit typisch 32^3 Einträgen (RGB)
  2. platziere alle Pixel in ersten Cluster und berechne den umgebenden Quader
  3. solange maximale Clusteranzahl noch nicht erreicht ist
     1. wähle den größten Cluster in L_∞-Norm
     2. projiziere die Farben auf die längste Achse des aktuellen Clusters
     3. teile den Cluster beim Median-Wert, senkrecht dieser Achse
     4. berechne die neuen umgebenden Rechtecke
  4. Bestimme Repräsentanten jedes Clusters (Mittelwert der Pixel)
- **schnell, recht gute Ergebnisse für natürliche Bilder (für Anzeige: ca. 200 Farben, CBR auch weniger)**

Varianten

- **anderer Farbraum**
- **andere Bestimmung des größten Clusters:**

  Cluster mit den meisten Pixeln oder größte L_2-Norm (Berechnung der längsten Hauptachse)
- **andere Teilungsregeln:**

  teile entlang der längsten Hauptachse oder teile in der Mitte der Quaderseite (bessere Berücksichtigung von kleinen, aber deutlich abweichenden Regionen)
- **andere Bestimmung des Repräsentanten:**

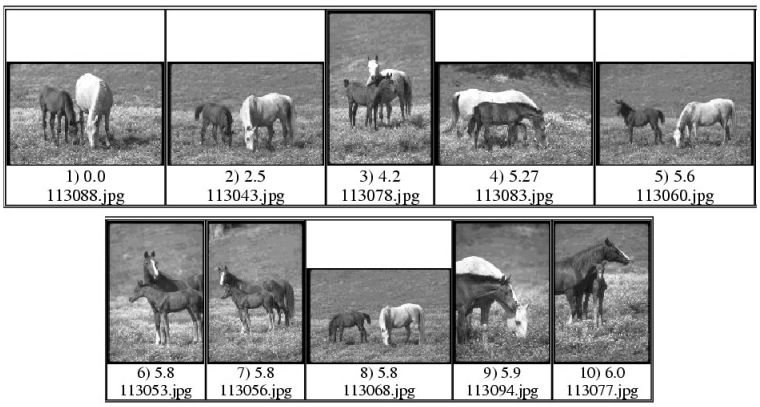
  Mittelpunkt des Quaders oder Mittelwert der Farben (statt Pixel)
- **Verbesserung durch Abwechseln von verschiedenen Kriterien während der Rekursion**
- **Abbruch, wenn Cluster genügend klein**
- **Beginn mit vielen Clustern und Verschmelzung**

Verbesserungen

- **erneute Quantisierung der Clusterrepräsentanten: kleinere Signaturen für CBR**
- **Nachbearbeitung mit k-Means Algorithmus: bessere Verteilung der Daten auf die Cluster**
  1. ordne jedem Pixel den nächsten Repräsentanten zu (kann aus anderem Cluster stammen!)
  2. berechne neue Repräsentanten aufgrund der neuen Zuordnung
  3. iteriere solange sich Repräsentanten signifikant ändern
- **Bestimmung der optimalen Clusterzahl oft aufwendig (bes. bei k-means), aber zu viele Cluster sind unkritisch, da dies die Abstandsmaße nur wenig beeinflußt**

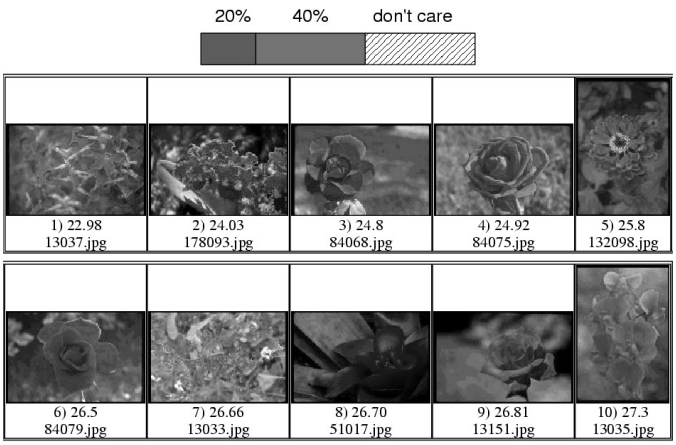
Suchergebnisse bei Farbsuche (1)

- **Suche nach Beispielbild (mit EMD):**



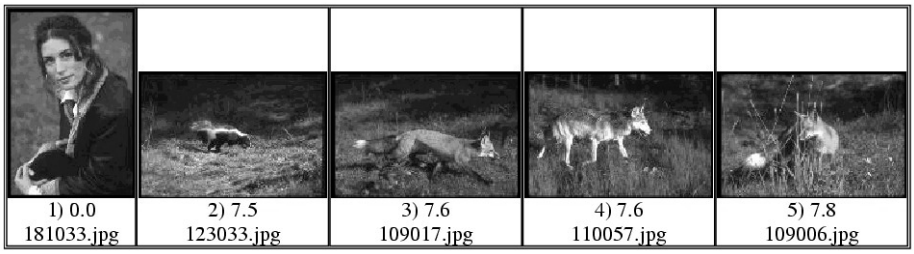
Suchergebnisse bei Farbsuche (2)

Suche nach Farbsignatur (mit EMD):



Suchergebnisse bei Farbsuche (3)

- **Diskrepanz zwischen semantischen und syntaktischen Merkmalen**
  - Unähnlichkeit der Bedeutung (Frau vs. Tiere) überdeckt Ähnlichkeit der Farben



Farbkonstanz (1)

- **Problem:**
  - Farben in Bildern derselben Szene ändern sich je nach Beleuchtung (Intensität, Farbe, Richtung), aber Wahrnehmung der Objektfarben durch Menschen ändert sich kaum

  „Farbkonstanz“

Farbkonstanz (2)

- **Lösungsansätze: Verwendung invarianter Farbräume oder Normierung der Bilder**
- **Invariante Farbräume für diffuse bzw. spiegelnde Reflexion:**

$$\left(\frac{R-G}{R+G}, \frac{R-B}{R+B}, \frac{G-B}{G+B}\right)$$

$$\frac{1}{|R-G|+|R-B|+|G-B|}\left(|R-G|, |R-B|, |G-B|\right)$$

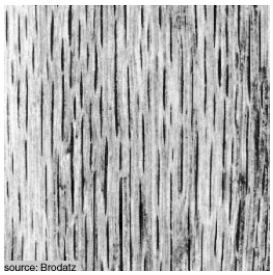
- **Nachteile:**
  - gilt nur für weißes Licht unterschiedlicher Intensität & Richtung
  - zu großer Verlust an Unterscheidungskraft - viele *false positives*
  - keine wahrnehmungskonforme Abstandsfunktion

Farbkonstanz (3)

- **Prinzipielles Vorgehen zur Normierung der Farben:**
  - Schätzung der Farbe der Lichtquelle, z.B. „Annahme einer grauen Welt“ – mittlere Farbe ist Lichtfarbe
  - Umrechnung der Farbwerte für eine Normlichtquelle
- **noch nicht vollständig gelöst**

Textur

- **Was ist Textur? – keine formale Definition**
- **informal: sich wiederholende, mehr oder weniger regelmäßige Muster, die als zusammenhängende Region interpretiert werden**

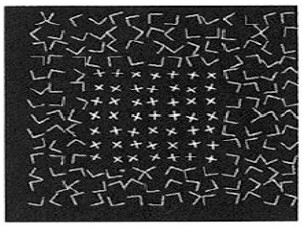

source: Brodatz

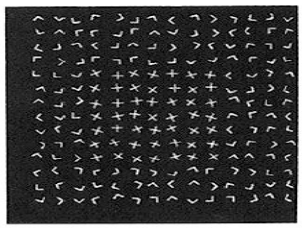
source: Simoncelli

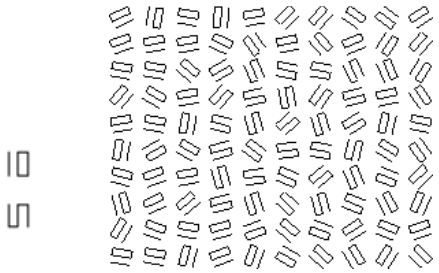
source: Simoncelli

Texturwahrnehmung (1)

- **Texturerkennung ist präattentiv (ansonsten scheinbar dieselbe Textur)**





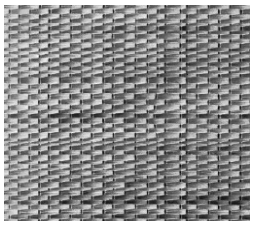


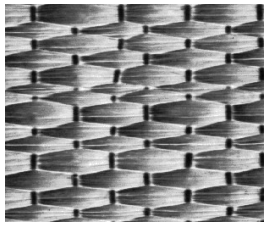
Texturwahrnehmung (2)

- **alte Vermutung (Julesz `66) „gleiche Statistik 2. Ordnung bedeutet gleiche Textur“ widerlegt**
- **grundlegende Texturcharakteristika aus neueren psychologischen Experimenten (z.B. Tamura `78):**
  - Grobheit (Skala): Kies vs. Schotter
  - Periodizität: Schotter vs. Ziegelmauer
  - Kontrast Differenz zwischen Hell und Dunkel bzw. extremen Farben
  - Gerichtetheit: Gras vs. Kies
- **Überlagerung der Phänomene auf mehreren Skalen möglich, z.B. unterschiedliche Grobheit in x- und y-Richtung**

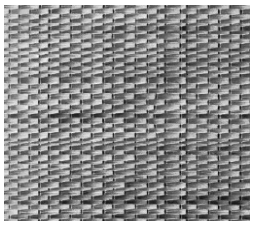
Texturwahrnehmung (3)

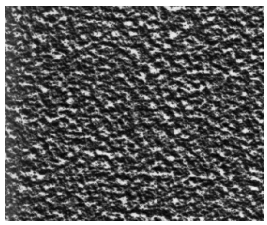
- **Grobheit**





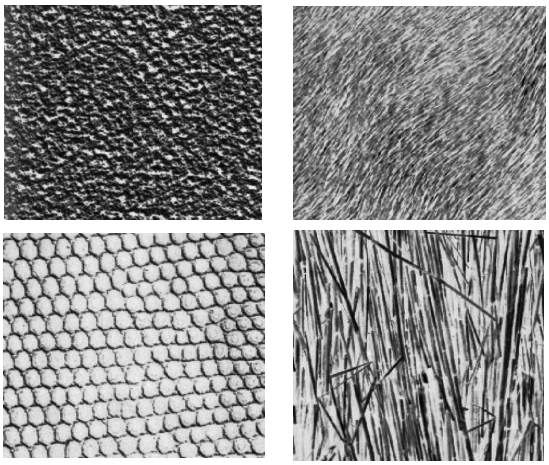
- **Regularität/ Priodizität**





Texturwahrnehmung (4)

- **Gerichtetheit**



Texturbeschreibung

- **Texturanalyse muß die Regelmäßigkeit erkennen und die Abweichungen ignorieren**
- **Methodenklassen:**
  - statistische Methoden: *co-occurence* Matrix, Markov Random Fields – besonders für unregelmäßige Texturen
  - Beschreibung durch Fourierkoeffizienten – regelmäßige Texturen
  - Wold-Merkmale: Kombination dieser Methoden
  - Filter: Laws-Filter, Gaborfilter, Wavelets
  - und andere: Neuronale Netze, *deterministic annealing*
- **Multiskalenverfahren**
- **Definition geeigneter Texturabstände:**
  - meist gewichtete Euklidische Distanz
  - existiert noch keine wirklich wahrnehmungskonforme Distanz

Co-occurence Matrizen (1)

- **Beschreiben Statistik 2. Ordnung einer Textur**
- **betrachte alle Dipole mit Länge *l* und Orientierung β**
- **bestimme 2D-Histogramm über die Grauwerte bzw. Farben der Dipole**
  - z.B. 4 Grauwerte, Dipol (l = 1, β = 0°)

	0	2	1	3	
	1	2	3	3	
	2	2	3	0	

	0	1	2	3	
0	0	0	1	1	
1		0	2	1	
2			1	2	
3				1	

Co-occurence Matrizen (2)

- **viele aus co-occurence Matrix abgeleitete Maße, z.B:**

$$\sum_{i,j} p_{ij}^2 \quad \text{- Uniformität}$$

$$\sum_{i,j} |i-j| p_{ij} \quad \text{- Kontrast}$$

$$\sum_{i,j} \frac{p_{ij}}{1+|i-j|} \quad \text{- Homogenität}$$

- **Abstandsfunktion zwischen heterogenen Maßen?**
- **Kombination der Maße von verschiedenen Dipolen?**
- **Maße für CBR weniger geeignet**

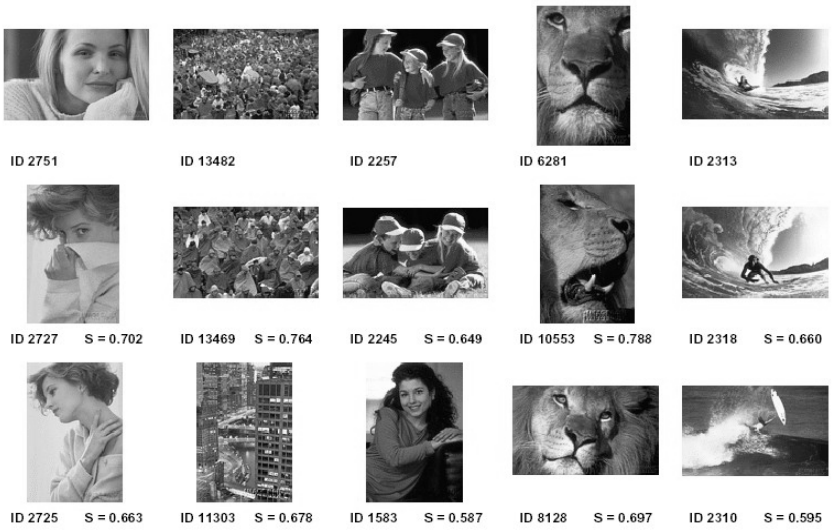
Co-occurence Matrizen (3)

- **rotationsinvariante co-occurence Matrizen:**
  - Histogramm über alle Orientierungen bei konstanter Länge der Dipole, z.B. $l = 1$
  - direkter Vergleich der Histogramme (Kovalev/Volmer)

$$D(k,l) = \frac{\sum_{i \in k \cap l} |k_i - l_i|}{\sum_{i \in k} k_i + \sum_{i \in l} l_i}, \qquad \text{falsche Formel aus dem Paper}$$

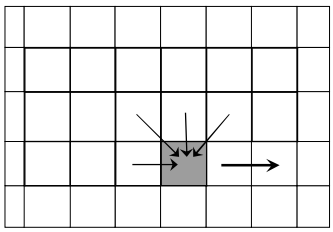
$$D(k,l) = \frac{\sum_{i \in k \cap l} |k_i - l_i| + \sum_{i \in k \setminus l} k_i + \sum_{i \in l \setminus k} l_i}{\sum_{i \in k} k_i + \sum_{i \in l} l_i} \qquad \text{verbessert}$$

Co-occurence Matrizen: Suchergebnisse



Markov Random Fields (1)

- **Betrachte Textur als Zufallsprozess (Mao & Jain ´92):**
  - erzeuge bzw. analysiere Textur in der Abtastreihenfolge
  - Farbe des nächsten Pixels hängt von den bisherigen Pixeln ab
  - Textur wird durch Mittelwert μ, Standardabweichung σ und Übergangskoeffizienten $a(\Delta x, \Delta y)$ charakterisiert



$$I(x,y) = \mu + \varepsilon(x,y) + \sum_{(i,j)<(x,y)} a(x-i, y-j) I(i,j)$$

$\varepsilon(x,y) = N(0,\sigma)$: Gaußsches Rauschen

Markov Random Fields (2)

- **Annahme stationärer Textur mit endlicher Wechselwirkungsweite: betrachte Nachbarschaft fester Größe**
  - einfachster Fall: 4 direkte kausale Nachbarn
  - subtrahiere Mittelwert der Textur
  - Minimiere

$$E = \sum_{(x,y)} \left(\sum_i a_i I(x+\Delta x_i, y+\Delta y_i) - I(x,y) \right)^2$$

	a_3: -1,-1	**a_2: 0,-1**	**a_1: 1,-1**	
	a_4: -1,0			

d.h. Lösung des linearen Gleichungssystems

$$\left[\sum_{(x,y)} \vec{I}_4(x,y)^T \vec{I}_4(x,y) \right] \vec{a} = \sum_{(x,y)} I(x,y) \vec{I}_4(x,y)$$

$$\vec{I}_4(x,y) = \left(I(x+1,y-1), I(x,y-1), I(x-1,y-1), I(x-1,y) \right)^T$$

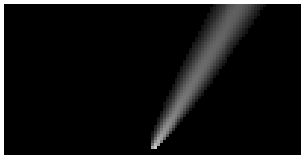
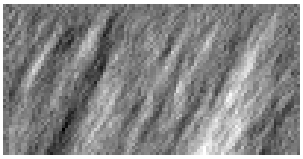
Markov Random Fields (3)

– bestimme Standardabweichung σ der Gaußverteilung aus Residuum: $\sigma = \sqrt{E}$

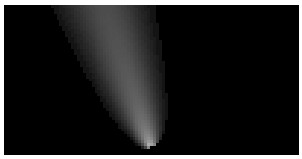
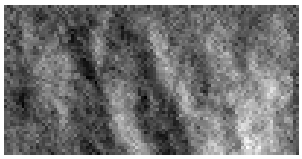
⇨ Modellierung der Textur als rekursive Filterung + Rauschen

- **Texturanalyse und –generierung möglich**

a_1= 0.6; a_2= 0.4
a_3= 0.0; a_4= 0.0
σ = 1.0

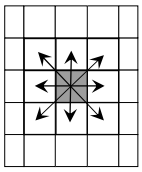



a_1= 0.25; a_2= 0.25
a_3= 0.25; a_4= 0.25
σ = 1.0

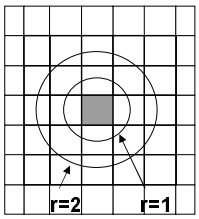
Markov Random Fields (4)

- **symmetrisches MRF:** Mittelung der korrespondierenden Pixel in der kausalen und akausalen Halbebene



$$I(x,y) = \mu + \varepsilon(x,y) + \sum_i a_i (I(x+\Delta x_i, y+\Delta y_i) + I(x-\Delta x_i, y-\Delta y_i))$$

- **rotationsinvariante MRF für ungerichtete Texturen:**
  – Mittelung der Grauwerte in Ringen um (*x*,*y*)
  – ein Rekursions-koeffizient pro Ring



$$I(x,y) = \mu + \varepsilon(x,y) + \sum_r a_r \bar{I}_r(x,y)$$

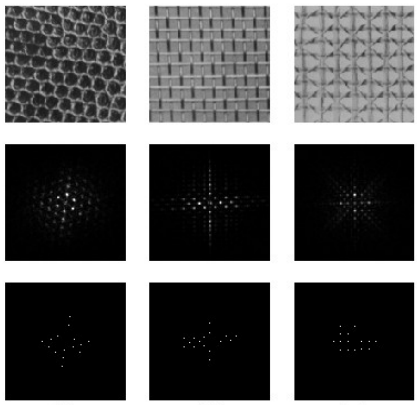
$\bar{I}_r(x,y)$: mittlerer Grauwert im Ring r um (x,y)

Markov Random Fields (5)

- **Verwendung größerer Nachbarschaften – genauere Texturbeschreinung, aber numerisch instabil**
- **Bestimmung der Koeffizienten auf mehreren Pyramidenstufen – *Multiresolution* MRF**
- **Texturabstand bei symmetrischem Multiresolution Markov Random Fields zweiter Ordnung**
  - berechne 4 Koeffizienten + Standardabweichung pro Ebene
  - verwende 3 Pyramidenebenen (Ebene 2 – 4)
  - bestimme Kovarianzmatrix der 15 Parameter über alle Texturen
  - verwende Mahalanobisdistanz zur Abstandsbestimmung
- **gilt für 1 Textur pro Bild, ansonsten Fensterung oder Textursegmentierung notwendig ⇨ Signatur**

Fourierkoeffizienten (1)

- **Fouriertransformierte einer regelmäßigen Textur hat diskrete, regelmäßig verteilte Maxima**
  - detektiere die 10 wichtigsten Maxima (Fundamentalwellen und deren harmonische)
  - drehe das Koordinaten-system in Richtung der 1. Fundamentalen
  - Lage und Höhe der 10 Peaks
  - „*harmonic peak signature*“



Fourierkoeffizienten (2)

- **Abstandsbestimmung:**
  - Vergleich der „*harmonic peak signatures*“

$\{m_d(\vec{\omega})\}, \{m_q(\vec{\omega})\}$: Signaturen Datenbank-/Anfragebild

Ähnlichkeitsfunktion wird maximal wenn die Signaturen identisch sind:

$$D = \sum_{\vec{\omega}_d} m_d(\vec{\omega}_d) \sum_{\vec{\omega}_q} w(\vec{\omega}_q - \vec{\omega}_d) \frac{m_d(\vec{\omega}_d) m_q(\vec{\omega}_q)}{\left[m_d(\vec{\omega}_d) + m_q(\vec{\omega}_q)\right]^2}$$

$w(\vec{\omega}_q - \vec{\omega}_d)$: Gaußsche Gewichtsfunktion für den Abstand der Peaks

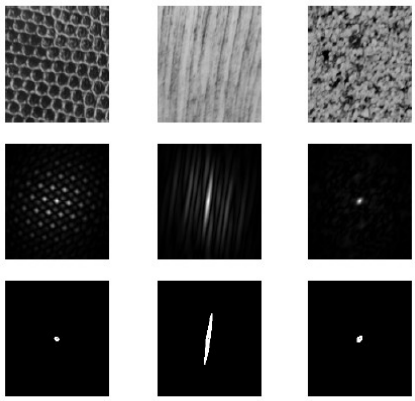
Wold-Merkmale (1)

- **jede Textur läßt sich als Überlagerung von drei Grundtypen darstellen:**
  - periodisch (deterministisch) in beiden Richtungen
  - periodisch (deterministisch) in einer Richtung, Zufallsprozeß (nicht-deterministisch) in der anderen
  - Zufallsprozeß (nicht-deterministisch) in beiden Richtungen
- **Wold-Merkmale (Liu und Picard ´94):**
  - generiere zwei Beschreibungen für jede Textur:
    - als periodische Funktion (Fourierbeschreibung)
    - als Zufallsprozeß (Markov Random Field)
  - berechne Konfidenz für jede dieser Beschreibungen
  - Texturdistanz: mit Konfidenz gewichtetes Mittel der Distanzen aus beiden Beschreibungen

Wold-Merkmale (2)

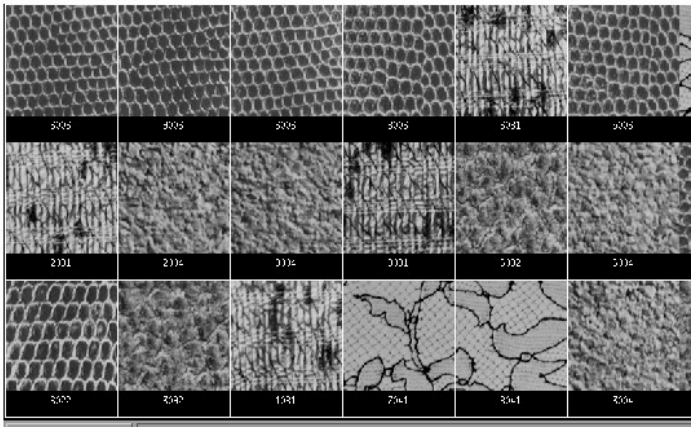
Konfidenzmaß für Periodizität:

- berechne Autokorrelation der Textur: hat bei periodischen Texturen mehrere Maxima, bei zufälligen nur zentrales Maximum
- Verhältnis der Energie im zentralen Maximum zur Gesamtenergie:
  - bimodales Histogramm über alle Texturen ⇨ Maß für die Regularität der Textur



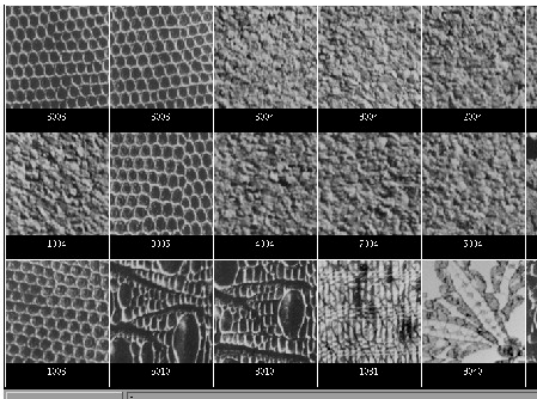
Wold-Merkmale (3)

- **Suchergebnisse: nur periodische Komponenten**



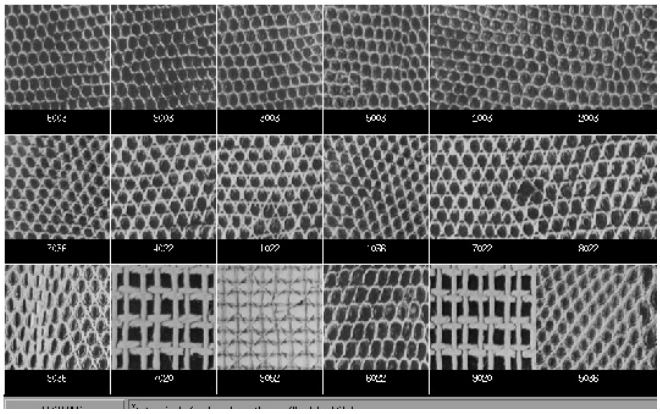
Wold-Merkmale (4)

- **Suchergebnisse: nur random field**



Wold-Merkmale (5)

- **Suchergebnisse: Wold (periodisch und random field)**



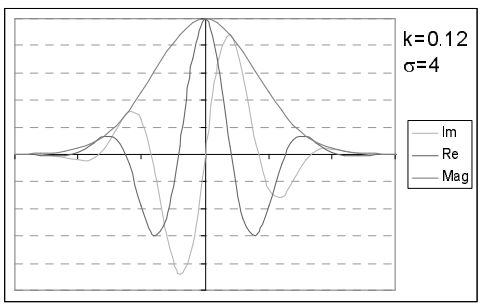
Gabor-Filter (1)

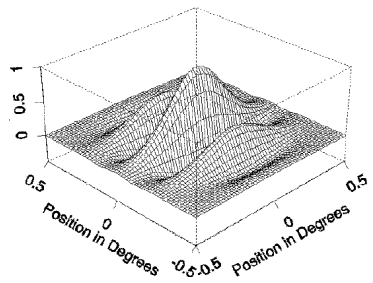
- **Nachteil der Wold-Merkmale:**
  - globales Texturmerkmal: falsche Ergebnisse, wenn mehrere Texturen
  - Fensterung oder Segmentierung möglich, aber kompliziert
- **Filter liefern lokale Beschreibung der Bildstruktur**
  - einfache Erzeugung eines Histogramms oder einer Signatur
- **Gaborfilter**
  - gute spektrale Eigenschaften: optimale Lokalisierung im Orts- und Fourierraum gemäß Unschärferelation
  - Betragsquadrat der Filterantwort phasenunabhängig: relative Lage der Textur zum Gitter verändert Texturmerkmale nicht
  - Aufspannen einer Filterbank von rotierten und skalierten Filtern

Gabor-Filter (2)

- **Form der Gaborfilter: rotierte und skalierte Versionen des Basisfilters:**

$$\mathrm{Gabor}_{k,\sigma}(x,y) = \left(\cos(2\pi\, k\, x) + i\sin(2\pi\, k\, x)\right) e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}$$

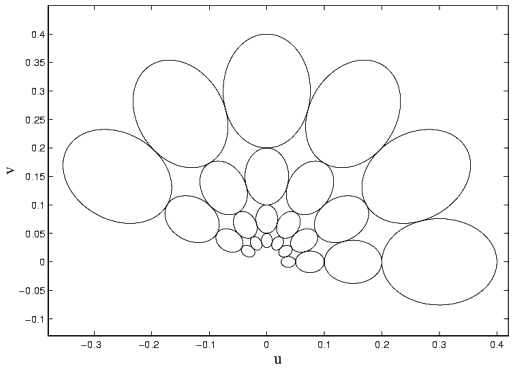




Gabor-Filter (3)

Berechnung der Parameter so, daß Fourierraum gleichmäßig abgedeckt

– 4 Skalen,
6 Richtungen



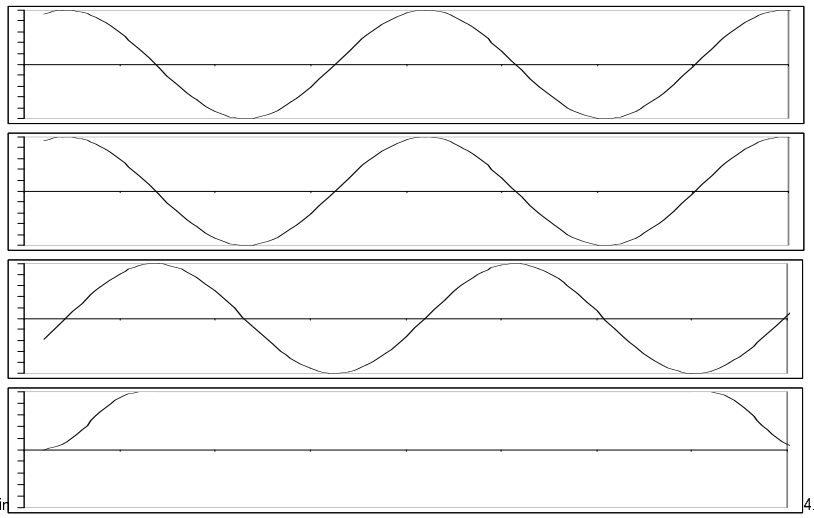
Gabor-Filter (4)

- **Energie der Gaborantwort – Phasenunabhängigkeit**

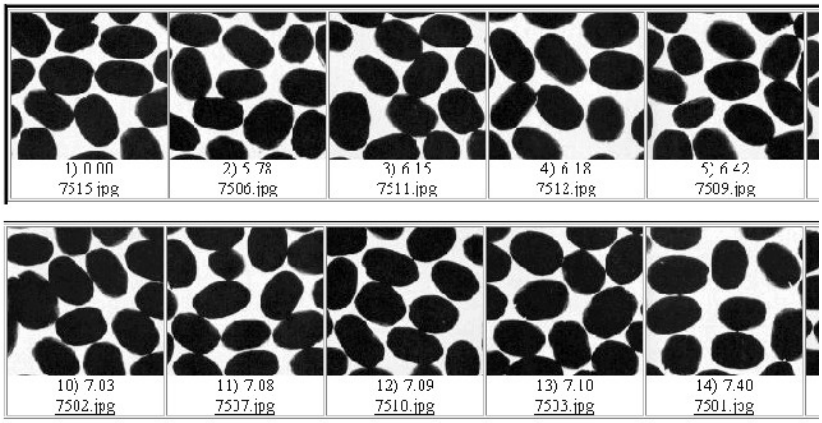
$$\text{Energie}(x, y) = \overline{\text{Gabor}(x, y)}\ \ \text{Gabor}(x, y)$$

(Betragsquadrat der komplexen Zahl $\text{Gabor}(x, y)$

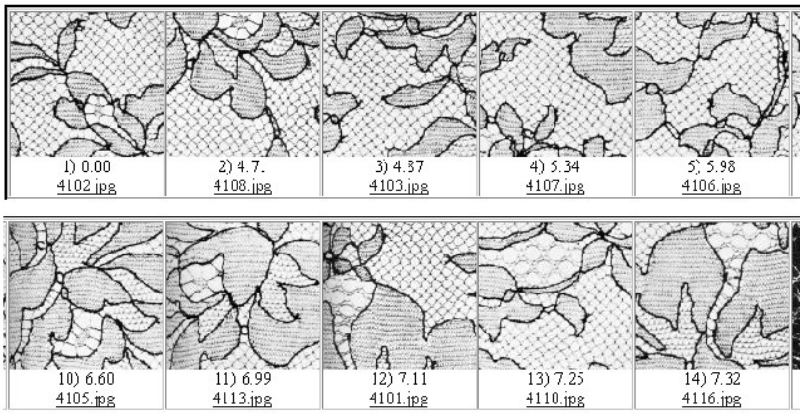
Gabor-Filter (5)



Gabor-Filter: Suchergebnisse (1)

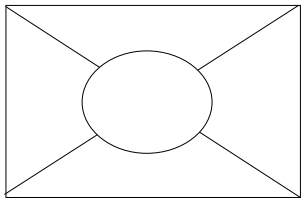


Gabor-Filter: Suchergebnisse (2)



Örtliche Zuordnung von Merkmalen (1)

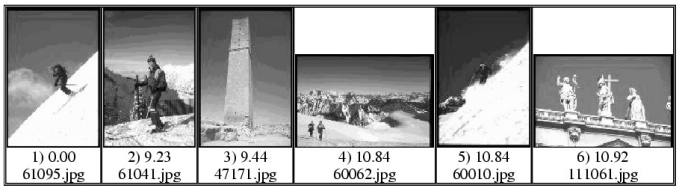
- **bisher nur globale Merkmale**
- **bessere Unterscheidung durch lokale Merkmale**
  - Farbe und Textur: Bildunterscheidung, falls keine Translationsinvarianz erwünscht
  - Textur: Vermeidung der Mischung verschiedener Texturen in einer Signatur
- **einfach: feste Zerlegung jedes Bildes**
  - 64x64 Raster
  - Zentrum und 4 Himmelsrichtungen



Örtliche Zuordnung von Merkmalen (2)

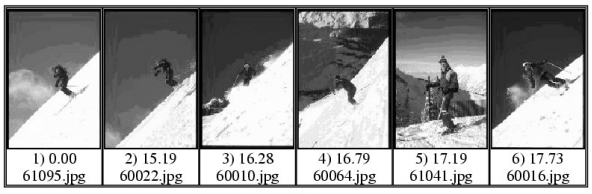
gemeinsame Signatur aus Farbe und Ort:

$Descriptor = (L, a, b, \lambda x, \lambda y),$
λ : Wichtung



(a)

Clusterung wie zuvor



Örtliche Zuordnung von Merkmalen (3)

- **Segmentierung (siehe später)**
  - 1 Signatur pro Segment, Kombination zu Globalsignatur

1) 130066 2) 130070 3) 130068

4)130051 5) 130038 6) 130039

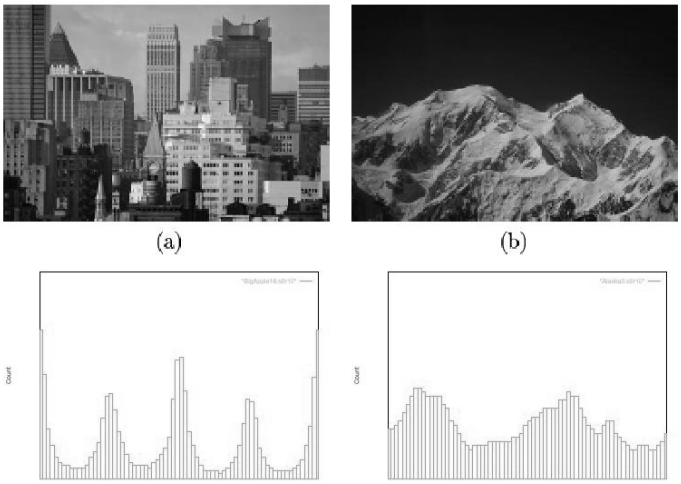
Suche nach Formmerkmalen

- **Formen spielen große Rolle für menschliches Ähnlichkeitsempfinden**
- **Ansätze:**
  - Formmerkmale ohne Segmentierung: z.B. Richtungshistogramm, Größenhistogramm
  - Template-Matching mit Hausdorff-Distanz
  - vollständige Segmentierung: Flächenmomente, Konturskalenraum
  - vollständige oder teilweise Segmentierung: Konturmorphing

Richtungshistogramm (1)

- **bestimme Gradienten des Bildes**
- **erstelle Histogramm der Gradientenrichtungen, wobei jedes Pixel mit seinem Gradientenbetrag gewichtet wird**
- **verwende *quadratic form distance* oder *earth mover distance*, wobei die Basisdistanz eine monoton fallende Funktion der Winkeldifferenz ist**

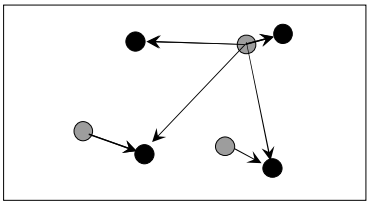
Richtungshistogramm (2)



Template-Matching mit Hausdorff-Distanz (1)

- **Bestimmung der Formähnlichkeit:**
  - finde für jeden (Kontur-)Punkt der einen Kontur den nächstgelegenen Punkt der anderen Kontur
  - wähle unter diesen NN den größten Abstand
  - symmetrisiere durch Wiederholung mit vertauschten Konturen

$$d(x,y) = \max\left(\max_{i \in x} \min_{j \in y} \|i - j\|, \max_{j \in y} \min_{i \in x} \|j - i\| \right)$$



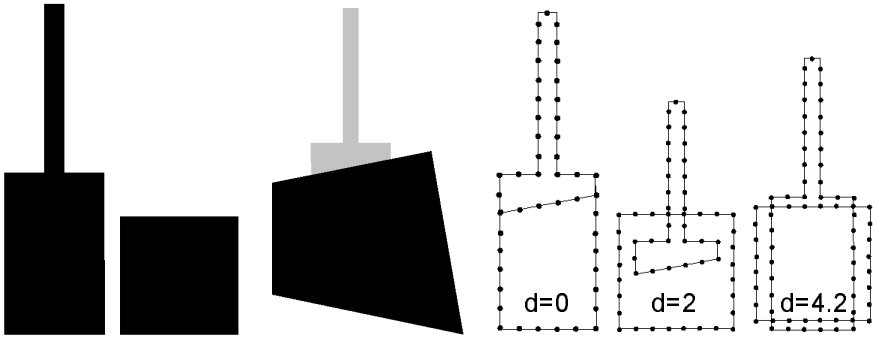
Template-Matching mit Hausdorff-Distanz (2)

- **Hausdorff-Distanz eignet sich zur konturbasierten Objektfindung:**
  - schiebe Template über das Bild
  - miß an jeder Position die Hausdorff-Distanz
  - Objekt ist erkannt, wo die Distanz ein Minimum und genügend klein ist
  - effiziente Berechnung mit Distanztransformation der Bildes
- **robuste Hausdorff-Distanz (gegen Ausreißer und Verdeckung): Ersetzen des max durch 60%-Quantil**

$$d(x,y) = \max\left(\underset{i \in x}{\text{med}_{60}} \min_{j \in y} \|i-j\|, \underset{j \in y}{\text{med}_{60}} \min_{i \in x} \|j-i\| \right)$$

Template-Matching mit Hausdorff-Distanz (3)

- **Verwendung zu Erkennung von verdeckten Objekten**
- **Dreiecksungleichung gilt nicht**



Flächenmomente (1)

- **Voraussetzung: binäre Indikatorfunktion für die von einem Objekt bedeckte Fläche**
- **allgemeine Formel:**

$$M_{ij} = \sum_{x,y \in R} x^i y^j$$

- **z.B.:** M_{00} **– Fläche**

  $x_s = M_{10}/M_{00}$, $y_s = M_{01}/M_{00}$ **– Schwerpunkt**
- **translationsinvariante Momente: Schwerpunkt in den Ursprung verschieben**

$$TM_{ij} = \sum_{x,y \in R} (x - x_s)^i (y - x_s)^j$$

Flächenmomente (2)

- **Trägheitstensor**

$$\begin{pmatrix} TM_{20} & TM_{11} \\ TM_{11} & TM_{02} \end{pmatrix}$$

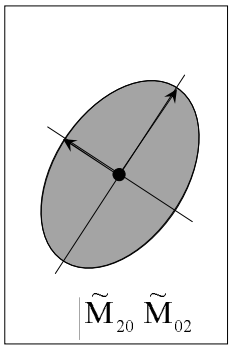
Eigenwerte: Hauptradien

$$\begin{pmatrix} \tilde{M}_{20} & 0 \\ 0 & \tilde{M}_{02} \end{pmatrix}$$

Eigenvektoren = Haupt(trägheits)achsen

Exzentrizität:

$$\frac{\tilde{M}_{20}}{\tilde{M}_{02}}$$



- **rotationsinvariante Momente: Drehung des Koordinatensystems in Richtung der Hauptachsen**

Flächenmomente (3)

- **direkte Berechnung rotationsinvarianter Momente:**

$$RM_1 = (TM_{20} + TM_{02})$$
$$RM_2 = (TM_{20} - TM_{02})^2 + 4TM_{11}^2$$
$$RM_3 = (TM_{30} - 3TM_{12})^2 + (3TM_{21} - TM_{03})^2$$
$$RM_4 = (TM_{30} + TM_{12})^2 + (TM_{21} + TM_{03})^2$$

- **Skaleninvarianz: Normierung mit Potenz von *RM$_1$***

$$SRM_2 = RM_2 / RM_1^2$$
$$SRM_3 = RM_3 / RM_1^3, \quad SIM_4 = RM_4 / RM_1^3$$

Flächenmomente (4)

- **auch affin invariante Momente möglich**
- **Ähnlichkeit: gewichtete Euklidische Distanz**
- **Leistung mäßig, da keine gute Übereinstimmung mit der menschlichen Wahrnehmung**

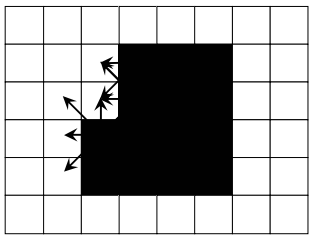
d = 0.033 **d = 0.85**

Konturskalenraum (1)

- **Bestimme, wie sich die Wendepunkte der Kontur verhalten, wenn die Kontur immer stärker geglättet wird (Mokhtarian)**
- **Algorithmus:**
  - Konturextraktion
  - natürliche (gleichabständige) Parametrisierung der Kontur
  - Faltung der Kontur mit Ableitungen von Gauß verschiedener Skalen
  - Bestimmung der Punkte, wo sich das Vorzeichen der Krümmung ändert ⇨ *"scale-space fingerprint"* der Kontur
  - Bestimmung der Lage der „Katastrophenpunkte" (Auslöschung von zwei Wendepunkten = einer Konkavität) ⇨ Kontursignatur
  - rotationsinvariantes Matching anhand der Signaturen

Konturskalenraum (2)

- **Konturverfolgung:**
  - beginne mit dem oberen linken Punkt der Kontur und initialisiere dort einen Nachbarschaftszirkulator, der nach links zeigt
  - solange die Kontur nicht vollständig:
    - inkrementiere den NZ solange, bis sein Ende auf einem Konturpunkt steht ⇨ Registriere den neuen Punkt
    - kehre die Richtung des NZ um und setze ihn auf den neuen Punkt



Konturskalenraum (3)

- **Ergebnis: Rohkontur (x_i, y_i)**
- **gleichabständige Abtastung der Kontur: Umfangsfunktion *s* an 100-1000 Stellen abtasten**

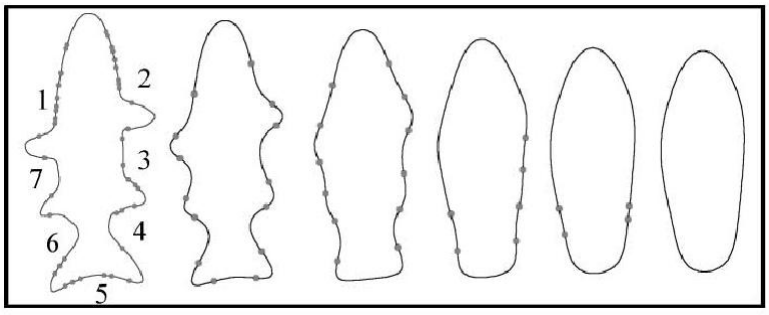
$$s = \frac{\int_0^u \sqrt{x'^2 + y'^2}\, du}{\int_0^{umax} \sqrt{x'^2 + y'^2}\, du}, \qquad \hat{P}_j = \left(x(s), y(s)\right)\Big|_{s=\frac{j}{N}}$$

- **Faltung der Kontur mit Ableitungen der Gaußfunktion bei Skala σ:**

$$\hat{x}' = \hat{x} * g'_\sigma \quad \hat{x}'' = \hat{x} * g''_\sigma, \quad \hat{y}' = \hat{y} * g'_\sigma \quad \hat{y}'' = \hat{y} * g''_\sigma$$

Konturskalenraum (4)

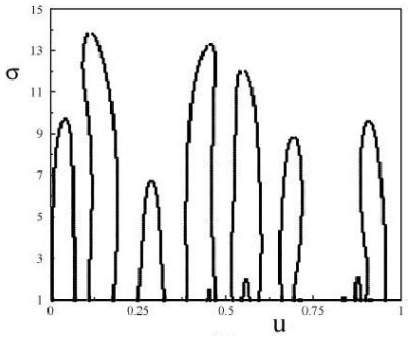
- **Glättung einer Kontur**



Konturskalenraum (5)

- **Berechnung der Krümmung:** $$\kappa(s,\sigma) = \frac{\hat{x}'(s,\sigma)\hat{y}''(s,\sigma) - \hat{x}''(s,\sigma)\hat{y}'(s,\sigma)}{\left(\hat{x}'^2(s,\sigma) + \hat{y}'^2(s,\sigma)\right)^{3/2}}$$

- **Eintrag der Nullstellen in ein s-σ Diagramm („Skalenraum-Fingerabdruck“, $\triangle\sigma$ = 0.1):**



Konturskalenraum (6)

- **Bestimmung der signifikanten Katastrophenpunkte (Maxima der Kurven = Auslöschung einer Konkavität):**
  - ⇨ Signatur: (s, σ)-Paare aller Maxima, die mindestens 1/6 der Höhe des höchsten Maximums haben
  - – unempfindlich gegen Rauschen
  - – translations- und skaleninvariant
  - – Rotation der Figur (oder Änderung des Startpunkts der Konturverfolgung) ⇨ zirkulare Verschiebung der Signatur
  - – affine Invarianz: Wendepunkte bleiben erhalten ⇨ affine Parametrisierung der Kontur

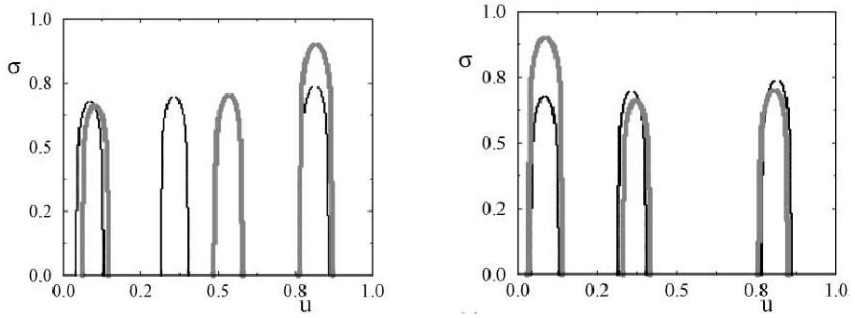
$$s = \frac{\int_0^u \sqrt[3]{x'y'' - x''y'}\,du}{\int_0^{umax} \sqrt[3]{x'y'' - x''y'}\,du}, \qquad \hat{P}_j = \left(x(s), y(s)\right)\Big|_{s=\frac{j}{N}}$$

Konturskalenraum (7)

- **Bestimmung der besten Übereinstimmung zweier Signaturen**
- **bestimme die optimale zirkulare Translation**
  - generiere Hypothesen
    - lege 2 Maxima aufeinander (größtes/größtes, größtes/2. usw.)
    - solange nicht alle Maxima zugeordnet sind:
      - bestimme die Distanz des größten noch freien zu allen Matchkandidaten
      - wähle die Zuordnung geringster Distanz, falls Schwellwert $\triangle s_{max}$=0.2 nicht überschritten wird

Konturskalenraum (8)

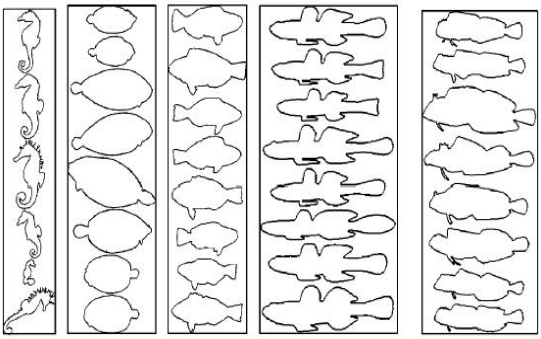
- **Beispiele für Korrespondenzbestimmung**



Konturskalenraum (9)

Tests für MPEG-7 Standardisierung

- Ähnlichkeitssuche für Konturen soll integriert werden (aber nicht das Segmentierungsverfahren)
- Datenbank binärer Muster zur Evaluierung: 70 Kategorien, 20 Bilder pro Kategorie
- Messung des *recall* bei 40 Antworten
- Konturskalenraum Sieger mit ca. 75%

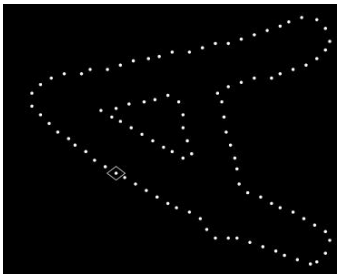


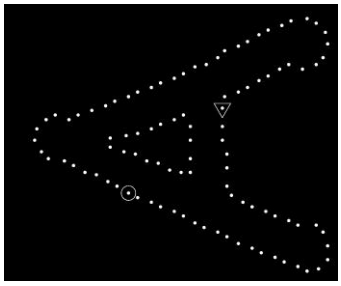
Formvergleich durch Verformung (1)

- **beschreibe Form durch Menge von Punkten**
- **charakterisiere Punkte durch „*shape context*“ Belongie et al. `01) und evtl. weitere Maße**
- **bestimme Korrespondenz und verforme eine Menge in die andere (z.B. mit *thin plate splines*)**
- **miß die Stärke der Verformung, die notwendig ist, um die Landmarken zur Deckung zu bringen**
- **sehr zuverlässiges Verfahren**
  - unempfindlich gegen Rauschen, da keine exakte Übereinstimmung notwendig
  - entspricht sehr gut der menschlichen Wahrnehmung

Formvergleich durch Verformung (2)

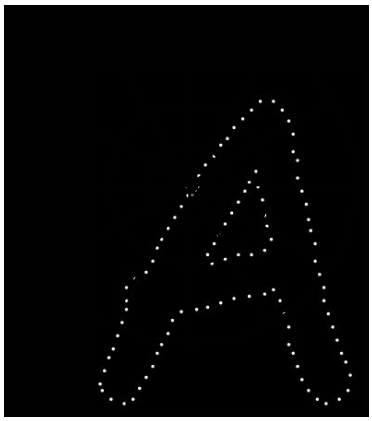
- **Auswahl der Punkte zur Beschreibung der Form**
  - (einfache) Kantendetektion
  - Auswahl von 100-300 möglichst gleichmäßig verteilten Punkten
  - müssen keine speziellen Punkte (Ecken, Kreuzungen, Landmarken) sein



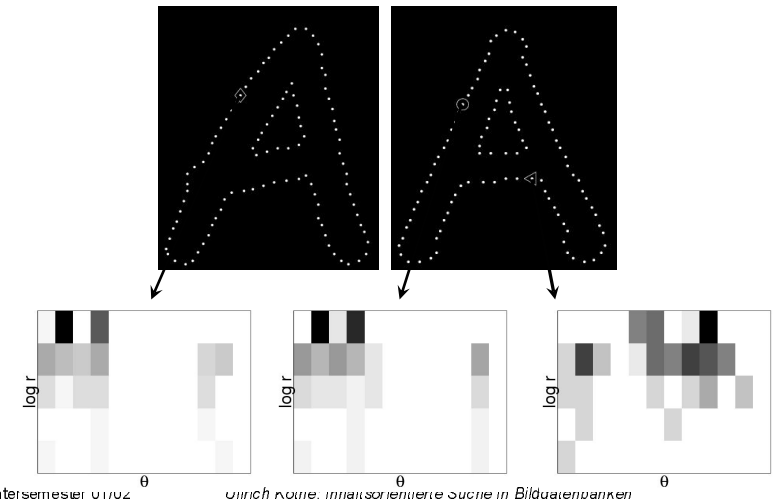


Formvergleich durch Verformung (3)

- **Bestimmung des *shape context:***
  - bestimme für jeden Punkt ein Histogramm über die Lage der anderen Punkte
  - logarithmischer Abstand (5 bins) und Winkel (12 bins)
  - falls Rotationsinvarianz benötigt: Winkel relativ zur Tangente
- **robust gegen Rauschen, teilweise Verdeckung**



Formvergleich durch Verformung (4)



Formvergleich durch Verformung (5)

- **Initiale Korrespondenz durch Matching der *shape context* Histogramme, z.B. χ^2-Test:**

$$D(k,l) = \frac{1}{2}\sum_i \frac{(k_i - l_i)^2}{k_i + l_i}$$

- **evtl. Einbeziehung weiterer Merkmale (z.B. Grauwertumgebung, Texturmerkmale)**
- **Berechnung der optimalen *thin plate spline* Transformation**
- **Neuberechnung der *shape contexts* des verformten Objekts**
- **Bestimmung verbesserter Korrespondenzen**
- **3mal wiederholen (oder bis Konvergenz)**

Formvergleich durch Verformung (6)

- **Methode der *thin plate splines* (TPS):**
  - Bestimmung eines optimalen Verschiebungsvektorfeldes, das die Punkte in Übereinstimmung bringt

$$\Delta x_i = (x_i^I - x_i^{II}) = f_A(x_i^I, y_i^I), \quad \Delta y_i = (y_i^I - y_i^{II}) = f_B(x_i^I, y_i^{II})$$

$$f_A(x,y) = a_1 + a_2 x + a_3 y + \sum_j w_j U\left(\left\|(x_j^{II}, y_j^{II}) - (x,y)\right\|\right)$$

$$f_B(x,y) = b_1 + b_2 x + b_3 y + \sum_j v_j U\left(\left\|(x_j^{II}, y_j^{II}) - (x,y)\right\|\right)$$

(a_k, b_k) - affine Transf.; $U(r) = r^2 \log r$ RBF des TPS

Formvergleich durch Verformung (7)

- **TPS minimiert die Biegeenergie (Bookstein `91):**

$$E[f] = \iint \left(f_{xx}^2 + 2 f_{xy}^2 + f_{yy}^2\right) dx\, dy \rightarrow \min$$

- **Lösung des linearen Gleichungssystems**

$$\begin{pmatrix} \mathbf{K} & \mathbf{P} \\ \mathbf{P}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \vec{w} \\ \vec{a} \end{pmatrix} = \begin{pmatrix} \Delta x \\ \vec{0} \end{pmatrix} \quad \text{mit}$$

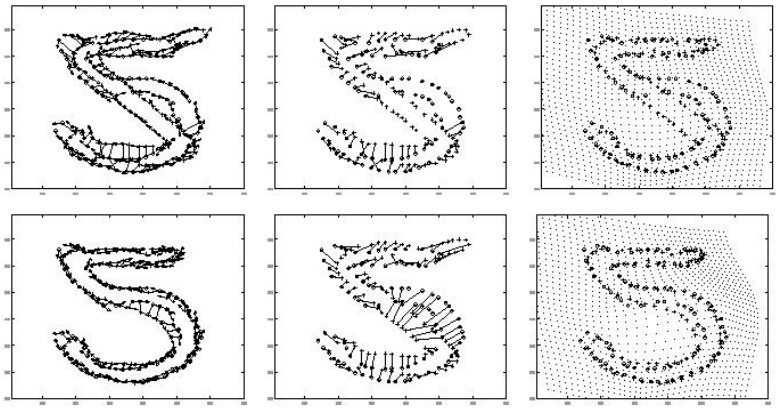
Interpolationsbedingung: $f_A(x_i^I, y_i^I) = x_i^I - x_i^{II} = \Delta x_i$

Nebenbedingungen: $\sum w_i = \sum w_i x_i^{II} = \sum w_i y_i^{II} = 0$

$K_{ij} = U\left(\left\|(x_i^I, y_i^I) - (x_j^{II}, y_j^{II})\right\|\right), \qquad P_i = (1, x_j^{II}, y_j^{II})$

Formvergleich durch Verformung (8)

- **Beispiel für TSP Morphing (1. und 5. Iteration)**



Formvergleich durch Verformung (9)

- **Abschwächung der Interpolationsbedingung:**
  - Punkte liegen nicht mehr exakt übereinander ⇨ geringere Rauschempfindlichkeit
  - Einführen der Biegeenergie als Regularisierungsterm

$$E_R[f_A] = \sum_i \left(\Delta x_i - \left|x_i^I - x_i^{II}\right|\right)^2 + \lambda E[f_A] \rightarrow \min$$

  - mit $\lambda \rightarrow 0$: exakte Interpolation, $\lambda \rightarrow \infty$: rein affines Modell (Biegeenergie verschwindet)
  - Lösung: ersetze K durch $K+\lambda I$ im Gleichungssystem
  - Skalierungsabhängigkeit:
    wenn Bild mit α skaliert wird, gilt $\lambda \rightarrow \alpha^2 \lambda$
    ⇨ setze λ_0**=1,** α ~ mittlerer NN-Abstand der Punkte

Formvergleich durch Verformung (10)

- **Vergleich:**
  - Bestimme optimale Transformation *q*
  - symmetrischer Abstand

$$D(I,II) = \frac{1}{n^I}\sum_{\{p^I\}} C(p_i^I, \text{match}(p_i^I)) + \frac{1}{n^{II}}\sum_{\{p^{II}\}} C(p_i^{II}, \text{match}(p_i^{II})) + D_{bending}(I,II)$$

$$\text{match}(p_i^I) = \arg\min_{\{p^{II}\}} C(p_i^I, q(p_j^{II})) \text{ - korrsepondierender Punkt}$$

C : "appearencecost" (Grauwert- Texturähnlichkeit)

$D_{bending}(I,II) = w^T \mathbf{K} w + v^T \mathbf{K} v$: Biegeenergie

Formvergleich durch Verformung (11)

- **MPEG-7 Datenbank: 76% (etwas besser als *contour scale space* Verfahren)**
- **Trademark-Datenbank: Ähnlichekeitssuche zur Feststellung potentieller Konflikte**

Gewinnung symbolischer Beschreibungen

- **Vorgehen in 2 Schritten**
  - Klassifikation aufgrund von low-level Merkmalen
  - high-level Suche aufgrund der Klassenzugehörigkeit oder Klassenhäufigkeit
- **typische Klassen:**
  - einfache globale Unterscheidungen: Photo/Graphik, Innen/Außen, Stadt/Landschaft
  - komplexere Klassen für Regionen: Himmel, Gras, Gesicht,... (ca. 10-20)
  - hierarchische Verknüpfung zu komplexeren Klassen, z.B. „Himmel + Schnee + Steine + Gras ⇨ Gebirge“
- **Einbeziehung von Text (z.B. Bildunterschriften)**

Globale binäre Klassifikation (1)

- **Graphik vs. Photo: Kombination einfacher Merkmale**
- **Fehlerrate (im WWW) < 10%**

	Graphik	**Photo**
Farbverteilung	wenig Farben	viele Farben
Sättigung	eher gesättigt	eher ungesättigt
Kantensteilheit	hoch	geringer
Bildgröße	kleiner	größer
Aspektverhältnis	groß	< 5/3

Globale binäre Klassifikation (2)

- **Landschaft vs. Stadt**
  - Richtungshistogramm mit ca. 10% Fehlerrate
  - Problem: Parks in Städten; Gebäude/Zäune, Bäume in Landschaften können zu Fehlklassifikation führen
  - insgesamt: feinere Aufteilung besser

	Landschaft	**Stadt**
Farbverteilung	wenig Aussagekraft	
Richtungshistogramm	kaum ausgeprägte Maxima	wenige ausgeprägte Maxima
Textur	unregelmäßig	regelmäßig

Globale binäre Klassifikation (3)

- **Innen- vs. Außenaufnahmen**
  - Kombination von Farbhistogramm, Richtungshistogramm und Schlüsselworthäufigkeit
  - Fehlerrate ca. 15%
- **Kenndaten für Merkmale:**
  - h_{kl}: Häufigkeit des Merkmals k in Bild l
  - h_{kC}: Häufigkeit des Merkmals k in Klasse C
  - h_k: Häufigkeit des Merkmals k in der gesamten Datenbank
  - $S_k = \log N / N_k$: Spezifität des Merkmals k
    (N: Anzahl der Bilder, N_k: Anzahl der Bilder, denen k zugeordnet ist)
  - $S_{kC} = \log N_C / N_{kC}$: Spezifität des Merkmals k bzgl. Klasse C

$$\text{score}(l, C) = \sum_k h_{kl} S_k * h_{kC} S_{kC}, \quad (l : \text{Bild}, C : \text{Klasse})$$

Globale binäre Klassifikation (4)

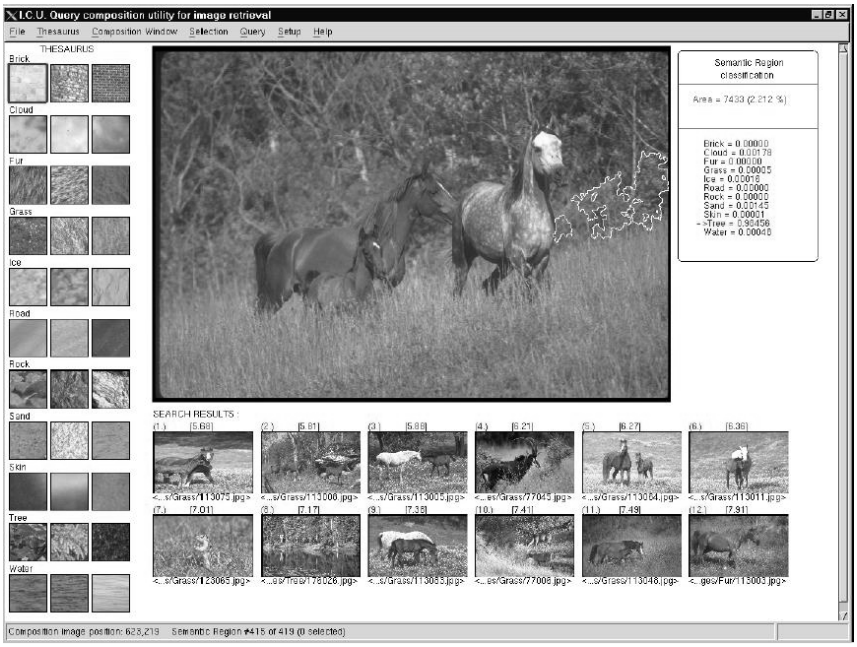
- **Optimale Auswahl der Worte:**
  - verwende 1. Satz der Bildunterschrift
  - verwende Substantive, Verben, Adjektive, Adverbien, Präpositionen, d.h. angepaßte Liste von *stop words*
- **wichtigste Wörter:**
  - Innen: Konferenz, Treffen, Hände, links, rechts
  - Außen: Soldaten, Polizei, Demonstration, von
- **Kombination der Merkmale durch gewichtetes Mittel**

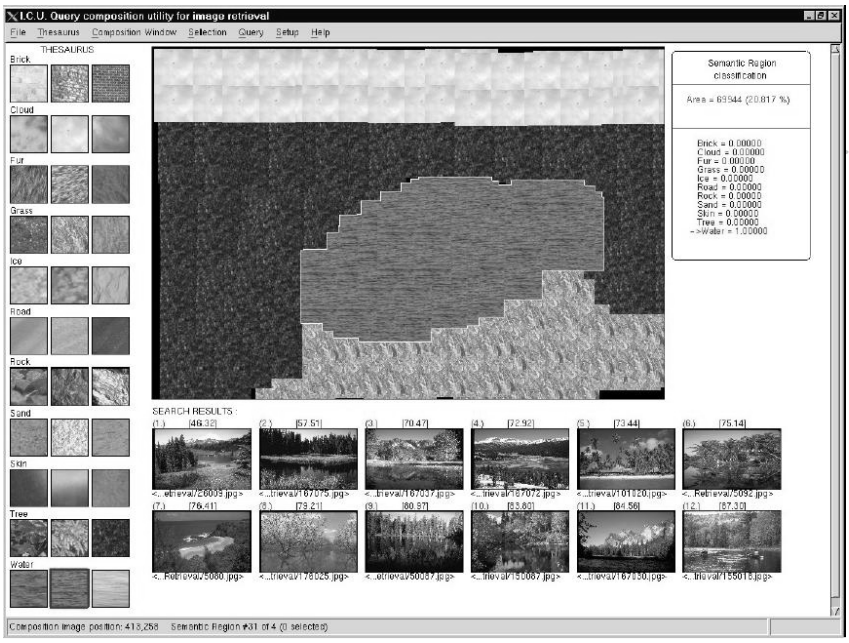
Suche mit Konzepten (1)

- **Typische Konzepte:**
  - Himmel, Wolken, Schnee/Eis, Wasser, Sand, Fels, Gras, Wald
  - Mauerwerk, Straße, Gebäude, Fahrzeug
  - Gesicht, Mensch, Menschenansammlung, Haut, Tier, Fell
- **Kachelung des Bildes oder Regionen-Segmentierung**
  - Berechnung von Merkmalen für jede Region (Farbe, Textur, Richtungen,...)
- **Erlernen der Klassifikation („Visual Thesaurus“):**
  - Nächster-Nachbar-Klassifikation
  - neuronale Netze
  - Entscheidungsbäume
  - Bayes-Klassifikator

Suche mit Konzepten (2)

- **Bildsignatur:**
  - Vorkommen (ja/nein) der Konzepte im Bild: wie Schlüsselwortsuche
  - Häufigkeit der Konzepte im Bild: Histogrammvergleich
  - Stärke der Ausprägung jedes Konzepts in einer Region: Euklidischer Abstand der „Klassifikationsvektoren“
- **einfache Berücksichtigung der räumlichen Verteilung**
- **Hierarchische Verknüpfung zu Konzepten höherer Ordnung möglich (aber welche? ⇨ anwendungsspezifisch)**
  - „Himmel + Schnee + Steine + Gras ⇨ Gebirge“





Verknüpfung von Text und Bild

Erweiterung des Ansatzes zur Unterscheidung Innen/Außen

Erstellung eines Histogramms der Worthäufigkeiten

- Hauptkomponentenanalyse
- ⇨ 256 dimensionaler Vektor (Linearkombination von Worthäufigkeiten) *„latent semantic indexing“*
- Kombination des LSI-Vektors mit Vektoren für ikonische Merkmale
- Verbesserung des *recall* von 25% auf 35%

Position	Gewicht
ALT-Feld von IMG	6
Titel/Bildunterschrift	5
Überschriften	2.2-4
Hervorhebungen	2.5-3
sonst ≤ 1 (je nach Entfernung)	