

# Semantic Based Image Retrieval

Frieder Hofmann

Seminar CBIR – 11.02.2005

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
  
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

## Die Nachteile des Traditionellen CBIR

*Obwohl viele Fortschritte erzielt wurden, sind die Probleme der Bildsuche noch lange nicht gelöst!*

- Die Wahrnehmung des Menschen ist schwer erklär- und nachvollziehbar
- Häufig wird Hintergrundwissen benötigt um Objekte korrekt zu klassifizieren
- Unter Umständen große semantische Unterschiede (semantic gap) zwischen Low-Levels features und Benutzer-Interessen
- Suchergebnisse für Anwender oft unbefriedigend/schwer zu interpretieren

## Die Nachteile des Traditionellen CBIR

*Obwohl viele Fortschritte erzielt wurden, sind die Probleme der Bildsuche noch lange nicht gelöst!*

- Die Wahrnehmung des Menschen ist schwer erklär- und nachvollziehbar
- Häufig wird Hintergrundwissen benötigt um Objekte korrekt zu klassifizieren
- Unter Umständen große semantische Unterschiede (semantic gap) zwischen Low-Levels features und Benutzer-Interessen
- Suchergebnisse für Anwender oft unbefriedigend/schwer zu interpretieren

## Die Nachteile des Traditionellen CBIR

*Obwohl viele Fortschritte erzielt wurden, sind die Probleme der Bildsuche noch lange nicht gelöst!*

- Die Wahrnehmung des Menschen ist schwer erklär- und nachvollziehbar
- Häufig wird Hintergrundwissen benötigt um Objekte korrekt zu klassifizieren
- Unter Umständen große semantische Unterschiede (semantic gap) zwischen Low-Levels features und Benutzer-Interessen
- Suchergebnisse für Anwender oft unbefriedigend/schwer zu interpretieren

## Die Nachteile des Traditionellen CBIR

*Obwohl viele Fortschritte erzielt wurden, sind die Probleme der Bildsuche noch lange nicht gelöst!*

- Die Wahrnehmung des Menschen ist schwer erklär- und nachvollziehbar
- Häufig wird Hintergrundwissen benötigt um Objekte korrekt zu klassifizieren
- Unter Umständen große semantische Unterschiede (semantic gap) zwischen Low-Levels features und Benutzer-Interessen
- Suchergebnisse für Anwender oft unbefriedigend/schwer zu interpretieren

das Gesuchte sind nicht unbedingt „ähnliche Bilder“ ...



Abbildung: Gleiche Klasse, aber sehr unterschiedliche Bilder

das Gesuchte sind nicht unbedingt „ähnliche Bilder“ ...



Abbildung: Gleiche Klasse, aber sehr unterschiedliche Bilder

das Gesuchte sind nicht unbedingt „ähnliche Bilder“ ...



Abbildung: Gleiche Klasse, aber sehr unterschiedliche Bilder

das Gesuchte sind nicht unbedingt „ähnliche Bilder“ ...

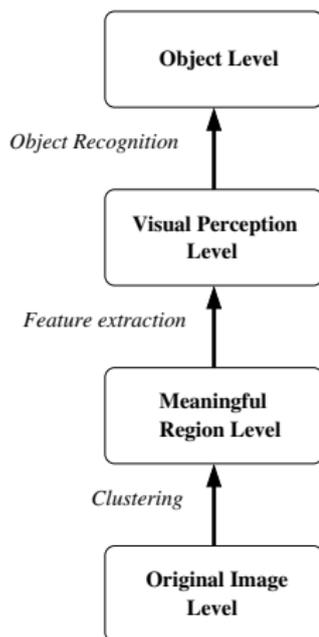


Abbildung: Gleiche Klasse?

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

# Was ist Semantic Based Image Retrieval



- Beschreibe das Bild in einem Mehrschichtenmodell von dem Originalbild über *Meaningful Regions* bis zum Object Level
- Es wird versucht die Semantik des Bildes zu erkennen bzw. abzubilden.
- Keine Benennung der Objekte notwendig.
- *Meaningful Regions* spielen hierbei eine zentrale Rolle

# Meaningful Regions

- Objekte extrahieren ist oft immer noch nicht möglich
- Deshalb suche nach „meaningful regions“
  - stellen eine Zwischenschicht zwischen dem ursprünglichen Bild und den Objekten des Bildes dar
  - nur grobe Segmentierung nötig
  - semantische Bedeutung kann später relativ leicht automatisch extrahiert werden
  - Objekte können dann mithilfe einer Knowledgebase zugeordnet werden
  - *später genauer*

## Visual Perception Level & Object Level

- Berechne Features für jede Meaningful Region
  - Auswahl an Features eingeschränkt, da nicht notwendigerweise rechteckige Gebiete
- mithilfe der berechneten Features für jede Region kann in einer Feature-Datenbank das zugehörige Objekt gesucht werden
  - Objekte sollten bekannt sein (bzw. es sollten Objekte von diesem Typ in der Datenbank enthalten sein)!

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

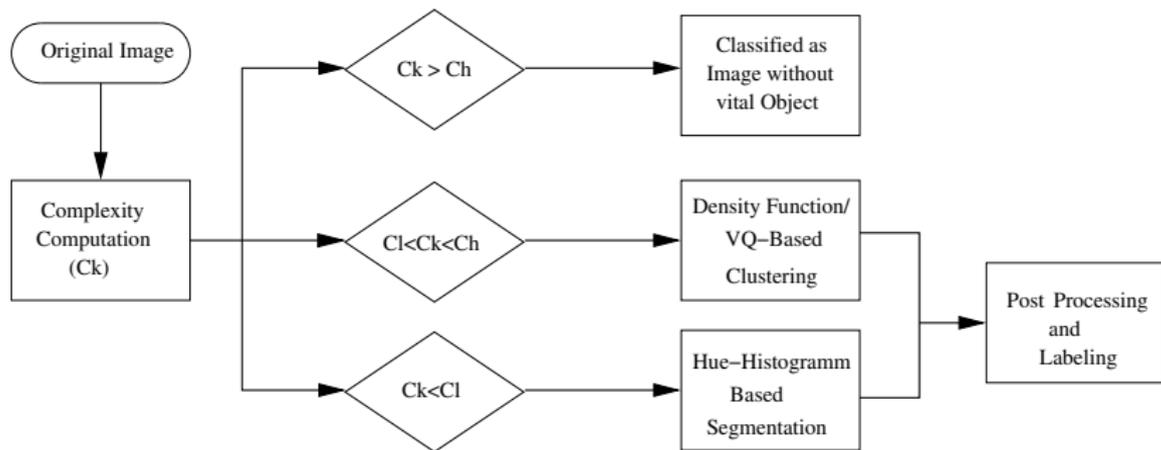


Abbildung: Flowchart: Meaningful Region Extraction

## Bestimmen des Komplexitätsindex $C^k$

Bestimme für jedes Bild den Wert:

$$C^k := \frac{1}{N} \sum R_k(i, j)$$

wobei

$R_k(i, j)$  = Anteil der Pixel mit größerem Unterschied von Pixel  $(i, j)$  als ein gegebenes  $\epsilon > 0$  .

Fall:  $C^k > C_h$  - Nicht separierbar, zu komplex

⇒ *Ganzes Bild eine Meaningful Region/ein Objekt*



Abbildung: Beispiele für zu komplexe Bilder

## Fall: $C^k < C_l$ |

### Hue-Histogramm Based Multi Threshold (1)

- 1 Betrachte das Histogramm des Hue-Kanals im HSV-Farbraum
- 2 sortiere die Bins des Histogramms nach Größe
- 3 Die Position des Maximums ( $P_1$ ) gibt den ersten Objekt-Mittelpunkt
- 4 Suche jetzt die Position ( $P_2$ ) des größten Wertes, für den gilt:  
 $|P_1 - P_2| > PEAK\_SPACE$   
(siehe Beispiel...)
- 5 setze alle Pixel entsprechend ihrem Abstand zu den  $P_i$  (nearest Neighbour) auf einen die Klasse repräsentierenden Wert

## Fall: $C^k < C_l$ II

- Falls innerhalb einer Region noch zu große (Farb-) Unterschiede vorhanden sind (z.B. zu große Varianz): Teile diese Regionen nach dem gleichen Prinzip noch einmal auf

# Beispiel: Hue-Histogramm Based Multi Threshold

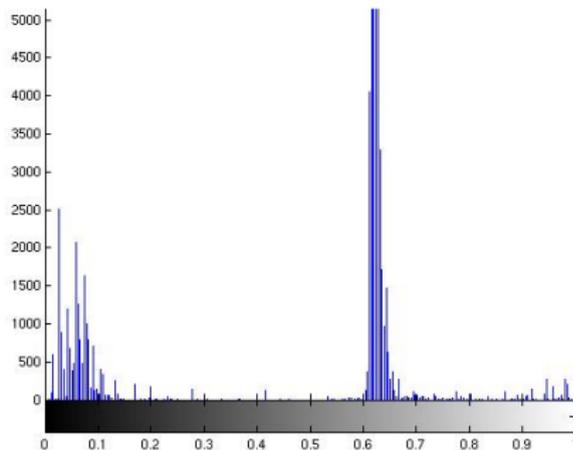


Abbildung: Histogramm des Hue-Wertes (HSV-Farbraum) Bild 220

## Beispiel: Hue-Histogramm Based Multi Threshold (2)



Abbildung: Histogramm des Hue-Wertes (HSV-Farbraum) Bild 220

## Weitere Beispiele I

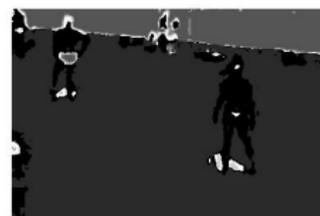


Abbildung: Bild 123 mit PEAK\_SPACE = 10, DIST = 0.01, mit PEAK\_SPACE = 20, und DIST = 0.1 bzw DIST = 0.001

## Weitere Beispiele II



Abbildung: Bild 126 mit  $PEAK\_SPACE = 10$ ,  $DIST = 0.01$ , mit  $PEAK\_SPACE = 20$ , und  $DIST = 0.1$  bzw  $DIST = 0.001$

## Schwierigkeiten bei Hue Histogramm Based Multi Thresh.

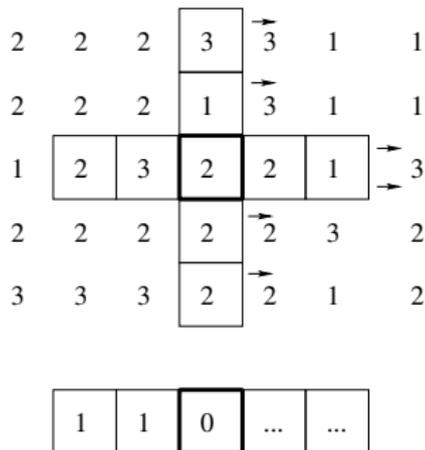
- *Nur Farbinformation enthalten*  
(ok, da dieses Verfahren nur auf wenig Komplexe Bilder angewandt wird)
- Der Wert PEAK\_SPACE muss bestimmt werden
- Der Threshold für die rekursive Segmentierung muss experimentell bestimmt werden

## Fall: $C_l \leq C^k \leq C_h$

*Verschiedene Möglichkeiten zur Clusterung:*

- VQ-Based Segmentation
- Density Function Based (finden von Regionen mit homogenen Feature-Vektoren...)

# VQ-Based Segmentation (1)



Berechne für jedes Bild den Wert  $T_n^k$  („Texture Coarseness“)

$$T_n^k = \sum_{i,j} E_{ij}$$

wobei  $E_{ij} = 1$  falls Pixel  $i,j$  ein lokales Zeilen- oder Spalten-Extremum des Value-Kanals (HSV-Farbraum) ist, 0 sonst (betrachte Nachbarschaften der Größe  $k$ ).

## VQ-Based Segmentation (2)

Berechne für jedes Pixel einen Feature-Vektor  $\vec{\chi}$  abhängig von der Textur-Information  $T^k$  des Bildes

- Fall  $T_k > T_{high}$  (Viel Information liegt in der Textur):  
setze  $\vec{\chi} = (H_{ij}, E_{ij}, V_{ij})$
- Fall  $T_k \leq T_{high}$  (weniger Information in Textur):  
setze  $\vec{\chi} = (H_{ij}, S_{ij}, V_{ij})$

(wobei  $H$  = Hue (HSV),  $E$  wie vorher,  $V$  = Value (HSV))

Hier können auch andere/höherdimensionale Feature Vektoren verwendet werden (z.B. Mittelwert/Varianz aller Pixel in einer Umgebung...) auch Gewichtung ist möglich.

# VQ-Based Segmentation (Algorithmus) I

Eigentliche Segmentierung:

Sei  $F^n$  die Matrix mit den Feature Vektoren  $\vec{x}$  von Bild  $n$

- 1 Berechne den Mittelwert  $\vec{m}$  aller Vektoren in  $F^n$
- 2 Berechne einen zufälligen „kleinen“ Vektor  $\vec{r}$  und wähle  $\vec{c}_1 = \vec{m} + \vec{r}$  und  $\vec{c}_2 = \vec{m} - \vec{r}$  als erstes *Codebook*
- 3 Klassifiziere alle Vektoren gemäß ihrem Abstand zu den Einträgen im Codebook
- 4 Erzeuge ein neues Codebook aus den Mittelwerten aller Klassen und klassifiziere erneut
- 5 Falls in einer Klasse  $\vec{c}_i$  zu große Unterschiede vorliegen, splitte diese wie in Schritt 2 auf und ersetze  $\vec{c}_i$  im Codebook durch die beiden neuen Vektoren. Gehe dann zurück zu Schritt 3

## VQ-Based Segmentation (Algorithmus) II

- 6 Stoppe falls entweder eine bestimmte Anzahl an Regionen erreicht, oder der Unterschied innerhalb der Regionen nur noch gering ist.

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
  
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - **Klassifizieren von Regionen und Bildern**
  - Möglichkeiten zur Verbesserung der Ergebnisse

# Klassifizieren der Regionen

Wieder VQ-Based

- berechne einen Feature-Vektor für jede Region (Meaningful Region)
- benutze VQ-Based Segmentation für diese Vektoren und bilde so *Cluster von Regionen* (Sky, People,...)

## Das eigentliche Finden der Bilder

- Finde Bilder, die möglichst viele Regionen der gleichen Klassen enthalten
- Es können dann noch Low-Level Features (Hu-Momente, Wavelet-Package, ...) verwendet werden um die zurückgelieferten Bilder zu sortieren und so „schönere“ Ergebnisse zu bekommen.

# Übersicht

- 1 Einleitung
  - Warum „Semantic Based CBIR“
- 2 Semantic Based Image Retrieval
  - Kurzüberblick
  - Extrahieren von Meaningful Regions
  - Klassifizieren von Regionen und Bildern
  - Möglichkeiten zur Verbesserung der Ergebnisse

## Verbesserungen in allen Einzel-Schritten möglich...

- Bessere Segmentierungsmethoden
- Komplexere (bessere) Features
- Gewichtete Features (je nachdem wie wichtig Farbe/Textur/... ist)
- Scene Level oberhalb des Object Levels einführen  
Dann Suchanfragen wie (Sky + Sand + Water) möglich
- *Relevance Feedback zum Verfeinern der Suchergebnisse*  
(Gegenstand aktueller Forschungen)

# Vielen Dank für die Aufmerksamkeit — Fragen???

?

?

?

# Literatur

-  Y. Luo, Y. J. Zhang, Y. Y. Gao, W. P. Yang  
Extracting meaningful regions for content-based retrieval of image and video.
-  Luo Y, Y. J. Zhang, Gao  
Object-Based Techniques for Image Retrieval.
-  Y. J. Zhang, Y. Y. Gao, N. S. Merzlykov  
Semantic-Based Image Description Model and Its Implementation for Image Retrieval.