An Introduction to Multimedia Data Indexing and Retrieval

Michael Kropfberger

michael.kropfberger@gmx.net

November 28, 1999

Traum & Realität

Traum:

- komplett automatisierte Indizierung
- Erkennen von Objekten, Handlung, Szenarien
- medienübergreifende KI (Bild und Ton)

Realität:

- nichtsemantische Indizierung
- semantische Indizierung

Nichtsemantische Indizierung

erkennt Farbbereiche, Kamerabewegungen, Objektumrisse …

+ (semi-) automatisch
+ beliebig kompliziert/langwierig
– "weiss" nichts über Objeke und deren Sinn
– eingeschränkte Abfragemöglichkeiten

Implementationsbeispiel: QBIC

Semantische Indizierung

textuelle Beschreibungen von Objekten, Aktivitäten ...

+ "menschlichere" Suche möglich
+ beliebig abstrakte Beschreibungssprachen für Szenen
− manuelle Indizierung
− eingeschränkte Abfragemöglichkeiten

Implementationsbeispiel: AVIS, OVID

Query by Image Content (QBIC)

entwickelt 1995 vom IBM Almaden Research Center

erlaubt Suche nach

- Beispielbildern
- interaktiv gezeichneten Bilder und Skizzen
- ausgewählten Farb- und Texturmustern
- Kamera- und Objektbewegungen

Objekterkennung mit QBIC

Objektscans entlang Umrissen

+ Bildkataloge mit wenigen Gegenständen pro Bild, separierbaren Hintergründen
+ textuelle Betitelung möglich
− möglicherweise Benutzereingriffe nötig
  - Schwellenwerte (Thresholds) ändern
  - Gummiband-Kurven zeichnen
  - mehrere Teile zu einem Objekt zusammensetzen

Videos mit QBIC

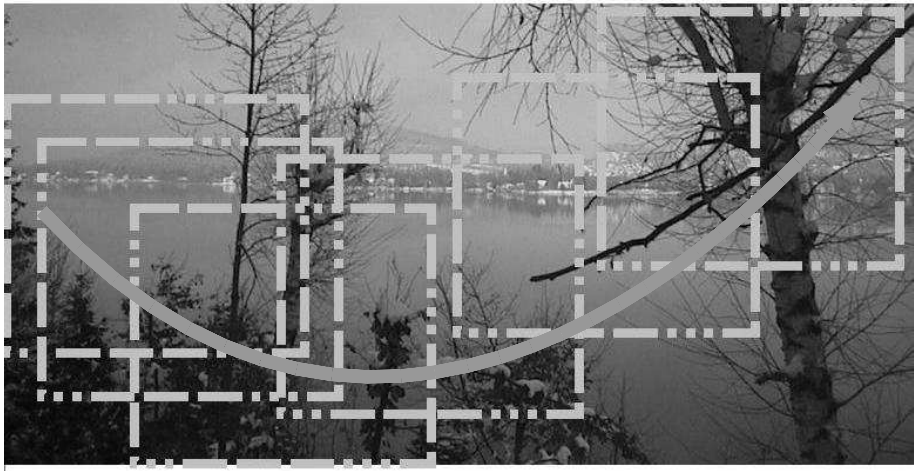
besondere Merkmale:

- Shot Detection
- repräsentatives Standbild pro Shot
- Erkennung von sich bewegenden Gebäuden und Objekten

neue Möglichkeiten:

- Zusammenkleben (Stitching)
- Objektfilterung
- Layerfilterung

Stitching



- Bildgröße $w \times h$, N Bilder im Shot
- linker oberer Bildeckpunkt $P_i(X_i, Y_i)$ mit $1 \leq i \leq N$

$$Groesse_{ges} = \Big(|\max(X_i) - \min(X_i)| + w\Big) \times \Big(|\max(Y_i) - \min(Y_i)| + h\Big)$$

Unser fiktiver Film: "The Killing Game"

Billy the Kid steht an der Bar und bricht einen Streit vom Zaun. Und zwar mit dem Bartender. Marshall DoJ sitzt im Hintergrund an einem Tisch und trinkt seinen Whiskey. Natürlich eilt er bald dem Bartender zu Hilfe und – da sich Billy nicht zur Räson bringen lässt – erschießt diesen in einem Duell. Daraufhin bläst er den Rauch von seinem Colt, verlässt den Saloon und reitet aus der Stadt, um andere Bösewichte zu fangen.

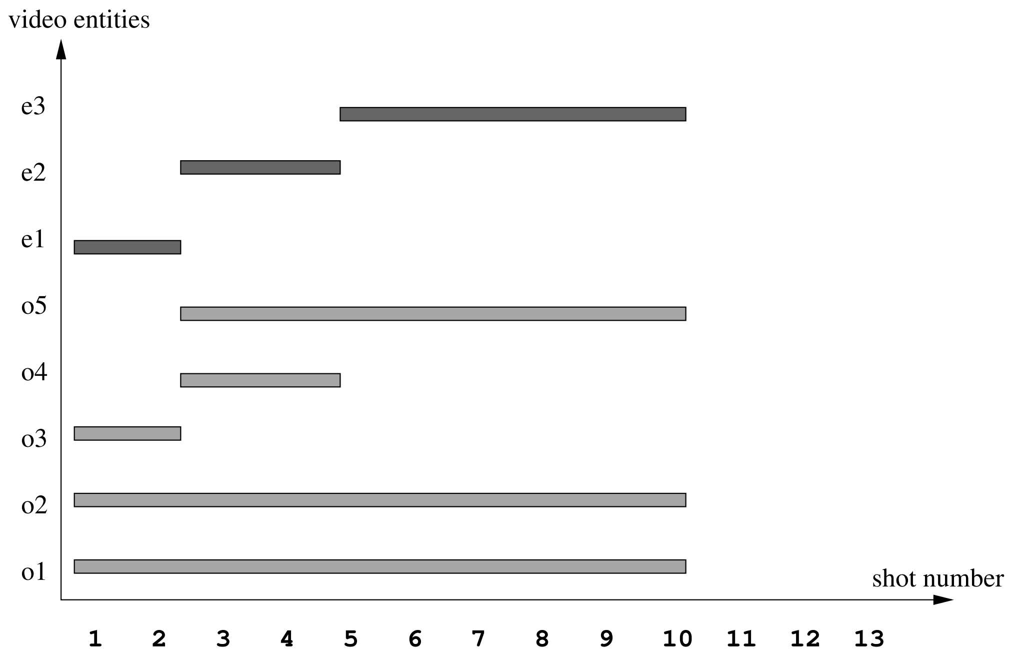
Advanced Video Information System (AVIS)

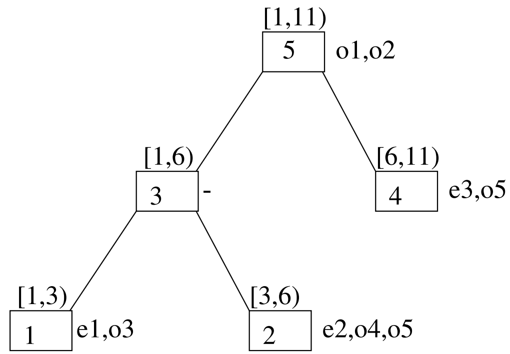
einfach, aber effizient zu implementieren

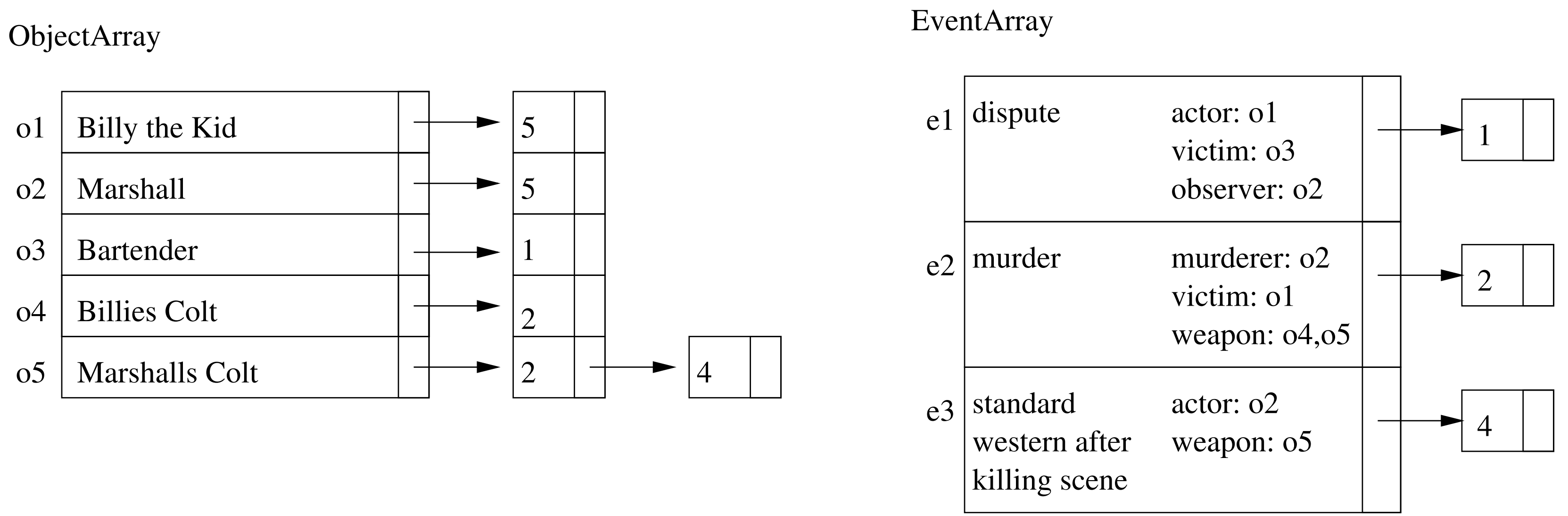
die meisten Abfragen sind in linearer Zeit lösbar

- Objekte
- Aktivitäten
- Ereignisse (Objekte+Aktivitäten)
- Rollen

AVIS Beziehungsgraph und Segmentbaum



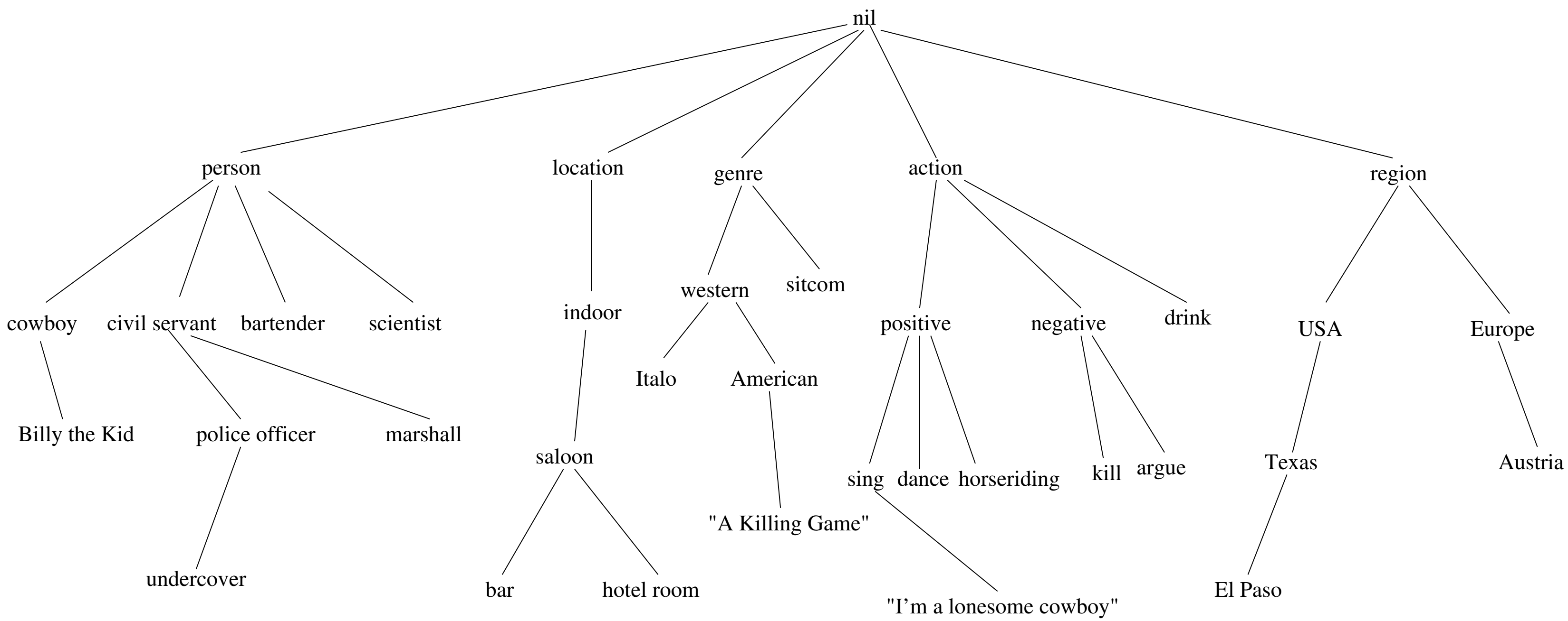


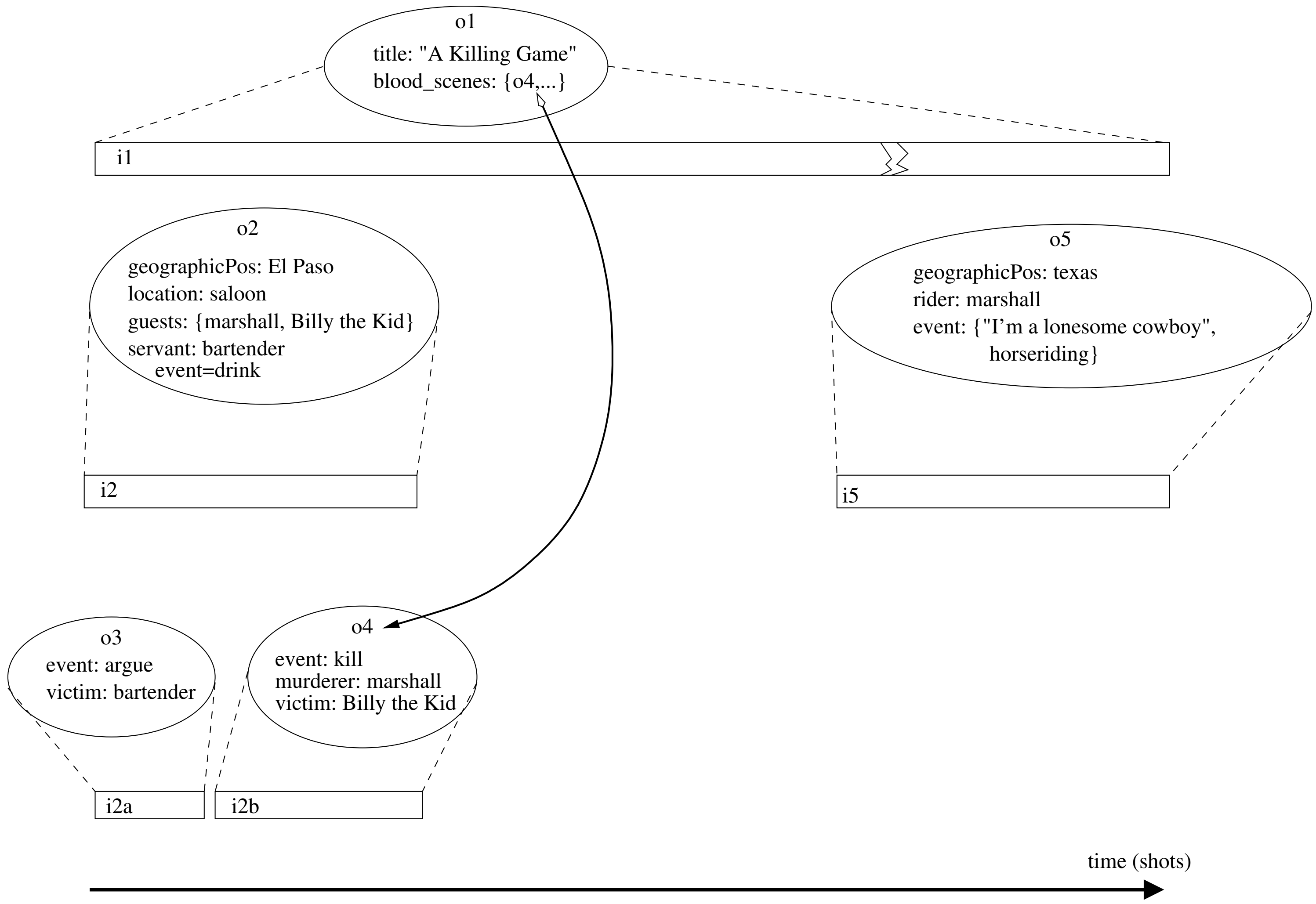


AVIS Objeke, Ereignisse, Rollen mit Segmentbaum-Zeigern

Object-Oriented Video Information Database (OVID)

- hierarchischer Generalisierungsbaum
- Video-Objekt: Tripel (oid, I, v) mit Attribut/Wert-Paaren $[a_1 : v1, \ldots, a_n : v_n]$
- Vererbung von speziell gekennzeichneten Attributen





OVID: überlappende Video-Objekte

OVID Abfragen

1. Gib mir alle Bilder, wo eine Person in Texas ist und singt. (ignoriert Attributnamen)

   formal: finde alle *oid*s | $\exists\, v_i \succeq person$ und $\exists\, v_j \succeq Texas$ und $\exists\, v_k \succeq sing$; Rückgabe von allen Frameintervallen.

2. Gib mir alle Bilder mit *blood_scenes* aus verschiedenen Genres.

   formal: finde alle *oid*s | $\exists\, v_i \succeq genre$ und $\exists\, a_j = blood\_scenes$; Rückgabe von allen Frameintervallen aller Objekte, welche in a_j gespeichert sind (für jede *oid*).

OVID Kompositionsoperationen: Merge

- einfache Vererbung
- **Merge** von zwei Objekten o_i und o_j
  - Vereinigung der Intervallmengen
  - Übernahme gemeinsamer Attribute
  - gleichnamige Wertattribute: erster gemeinsamer Wert aus dem hierarchischen Generalisierungsbaum
  - gleichnamige Objektattribute: rekursives Merge

Bsp: Merge $o2$ und $o5 \leadsto o6(oid, I, v)$ mit
$I = \{i2, i5\}$ $v = \{geographicPos = Texas, event = action\}$

OVID Kompositionsoperationen: Overlap

- **Overlap** von zwei Objekten o_i und o_j
  - echt überschneidende Frameintervalle
  - Übernahme aller Attribute
  - gleichnamige Wertattribute: alle Werte in die Mengenliste
  - gleichnamige Objektattribute: rekursives Overlap

Bsp: Overlap $o6$ und $o4 \leadsto o7(oid, I, v)$ mit
$I = \{i2b\}$ $v = \{geographicPos = Texas, event = \{action, kill\}, murderer = marshall, victim = Billy\ the\ Kid\}$

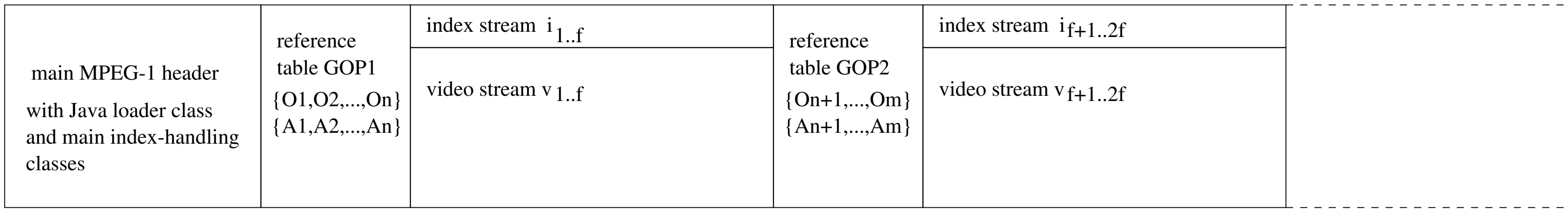
Intelligente Multimediadatenströme

- basierend auf MPEG-1 und Java
- Direkt in den Datenstrom inkludiert man
  - Indizes
  - Software (Klassen mit Wissen über Indexaufbau)
- mögliche Positionen sind spezielle Userbereiche im Haupt-Header, in jedem GOP und Frame

– Notwendigkeit von modifiziertem Encoder und Decoder!

+ Standarddecoder ignorieren Daten in Userbereichen

Erweiterter MPEG-1 Datenstrom



- Objekt/Aktions-Differenzlisten möglich
- Suche nach Objekten $\leadsto$ sequentiell über alle Daten
- zusätzliche Datenbank möglich

Authentifizierung für Videoinhalte

- Passwortlisten im Haupt-Header
- Frameintervalle nur für spezielle Benutzer(gruppen)
- Integration des Authentifizierungsprogrammes
- Verschlüsselung authentifizierungspflichtiger Intervalle

Abschluss und Ausblick

- Nichtsemantische Indizierung mit QBIC
- Semantische Indizierung mit AVIS und OVID
- Ideal ist Integration von beidem
- Integration von Indizes und intelligenter SW im Datenstrom

Standards:

MPEG-4 Layers, Objekte, Keywords und Interaktivität

MPEG-7 Definitionen für in-line Indizes und Informationen über deren Interpretation