

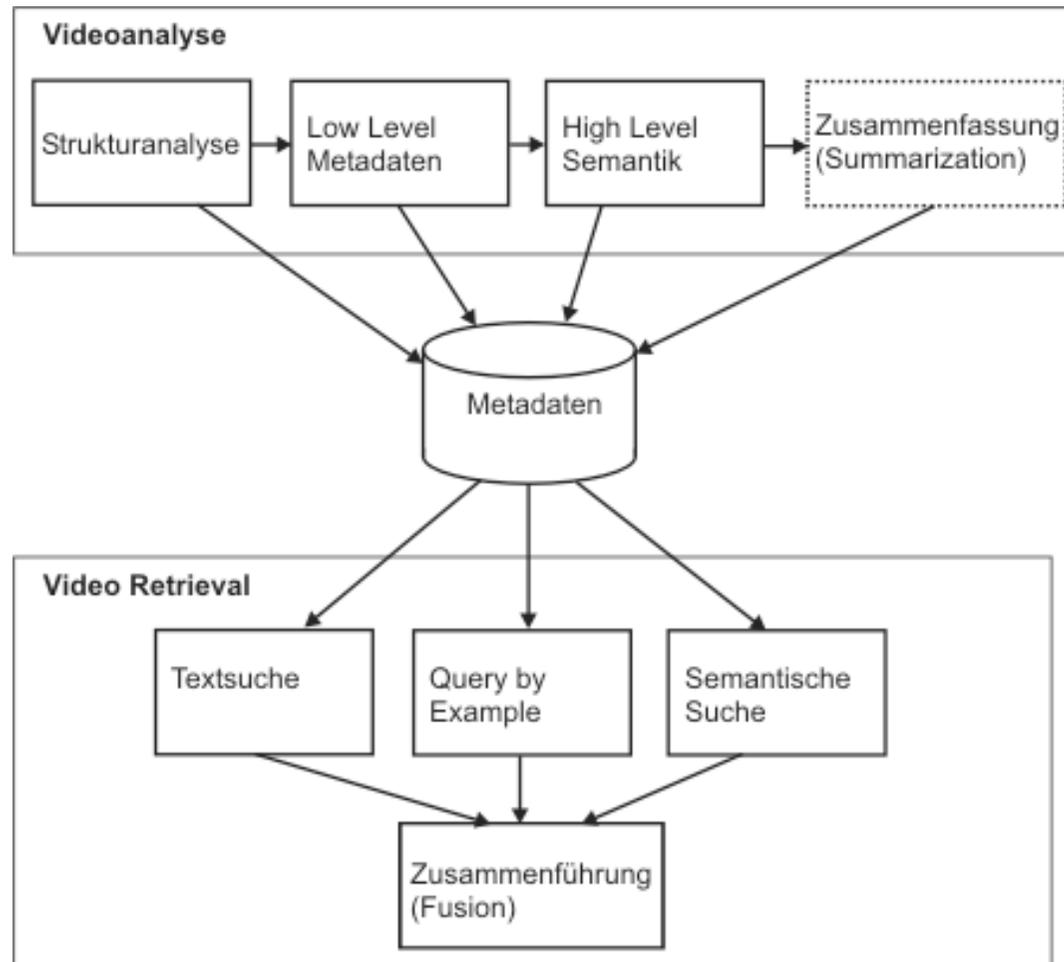
# Multimedia-Metadaten und ihre Anwendung

14.02.2006

Video Retrieval und Video  
Summarization

Maria Wagner

# Aspekte des Video Retrieval



# Videoanalyse

- Analyse nötig, um Struktur und Metadaten zu erkennen
- Video ist ein strukturiertes Medium, keine unstrukturierte Folge von Frames
- Es beinhaltet Bilder, Audio und Texte

## →Strukturanalyse nötig

- Erkennung von Schnitten
- Erkennung von Szenen
- Extraktion von Metadaten (High Level und Low Level)
- Zusammenfassung wird meist ebenfalls in Zuge der Videoanalyse generiert und als Metainformation gespeichert

# Shot Boundary Detection

- „Shot“: Kontinuierliche Aufnahme ohne Schnitt
- Vorgehensweise:
  - Extrahiere ein oder mehrere Merkmale (z.B. Histogramm, Kontrast) aus jedem n-ten Frame des Videos
  - Berechne die Differenzen für aufeinander folgende Frames
  - Vergleiche die Differenzen mit einem vorgegebenen Grenzwert
  - Wird der Grenzwert überschritten, so wurde ein Schnitt festgestellt

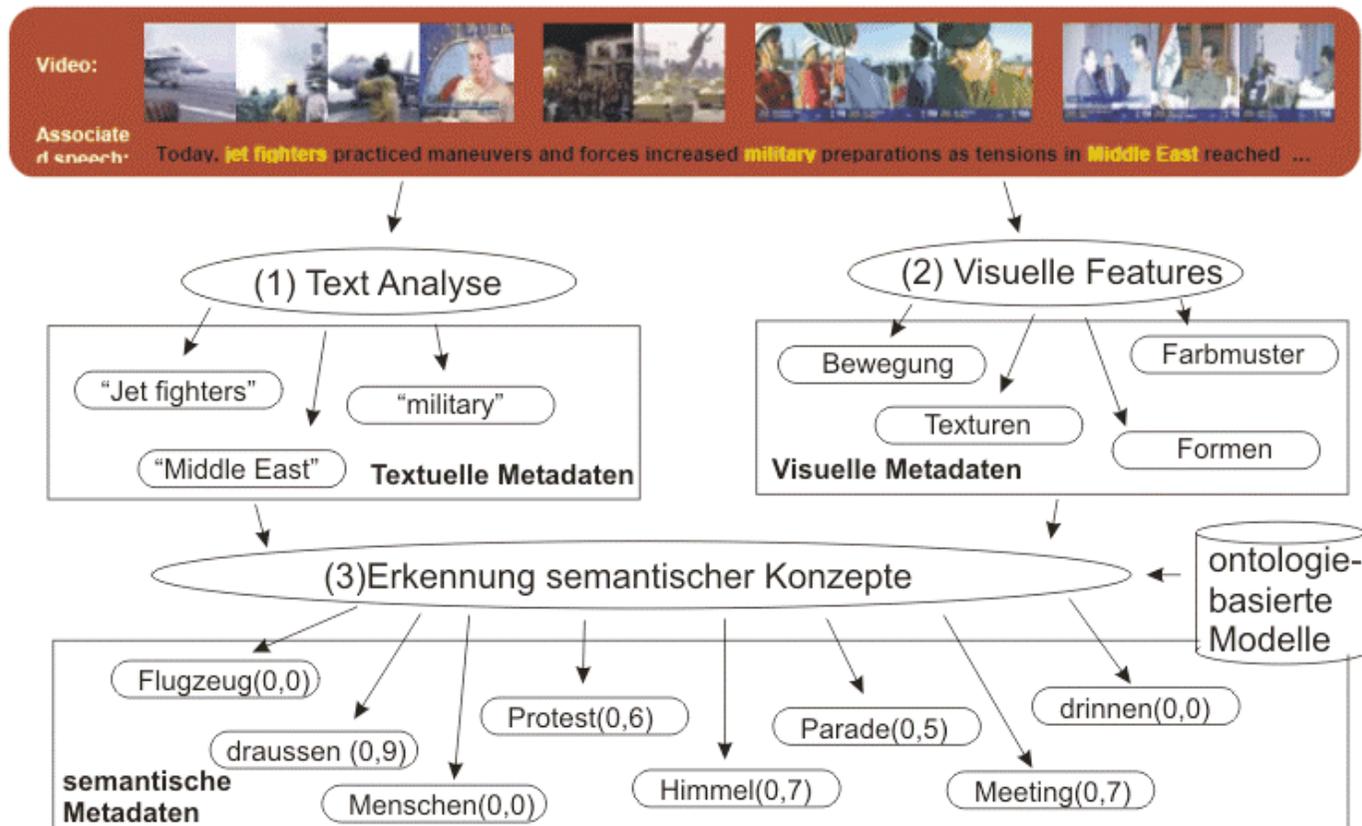
# Low Level Metadaten

- Visuelle Features:
  - Texturen, Farben, Formen
  - Bewegung
- Extraktion von Text
  - Spracherkennung
  - Video OCR
- Auswahl der sinnvollen Metadaten abhängig von der Art des Videos!
- Zusätzlich zur Spracherkennung kann Audio bei der Strukturierung des Videos helfen
- Low Level Metadaten bilden die Basis zur automatischen Erkennung der Semantik

# High Level Metadaten

- Ähnlichkeit von Videos liegt oft auf semantischer Ebene!
- Aber: Computer erkennt nur Low Level Merkmale automatisch! → „semantische Lücke“
- Lösungsansätze:
  - Manuelle Annotation
  - Ontologien
  - Maschinelles Lernen
- Problem: semantische Modelle sind meist nur für eine eingeschränkte Wissensdomäne einsetzbar

# Automatische Gewinnung von High Level Metadaten

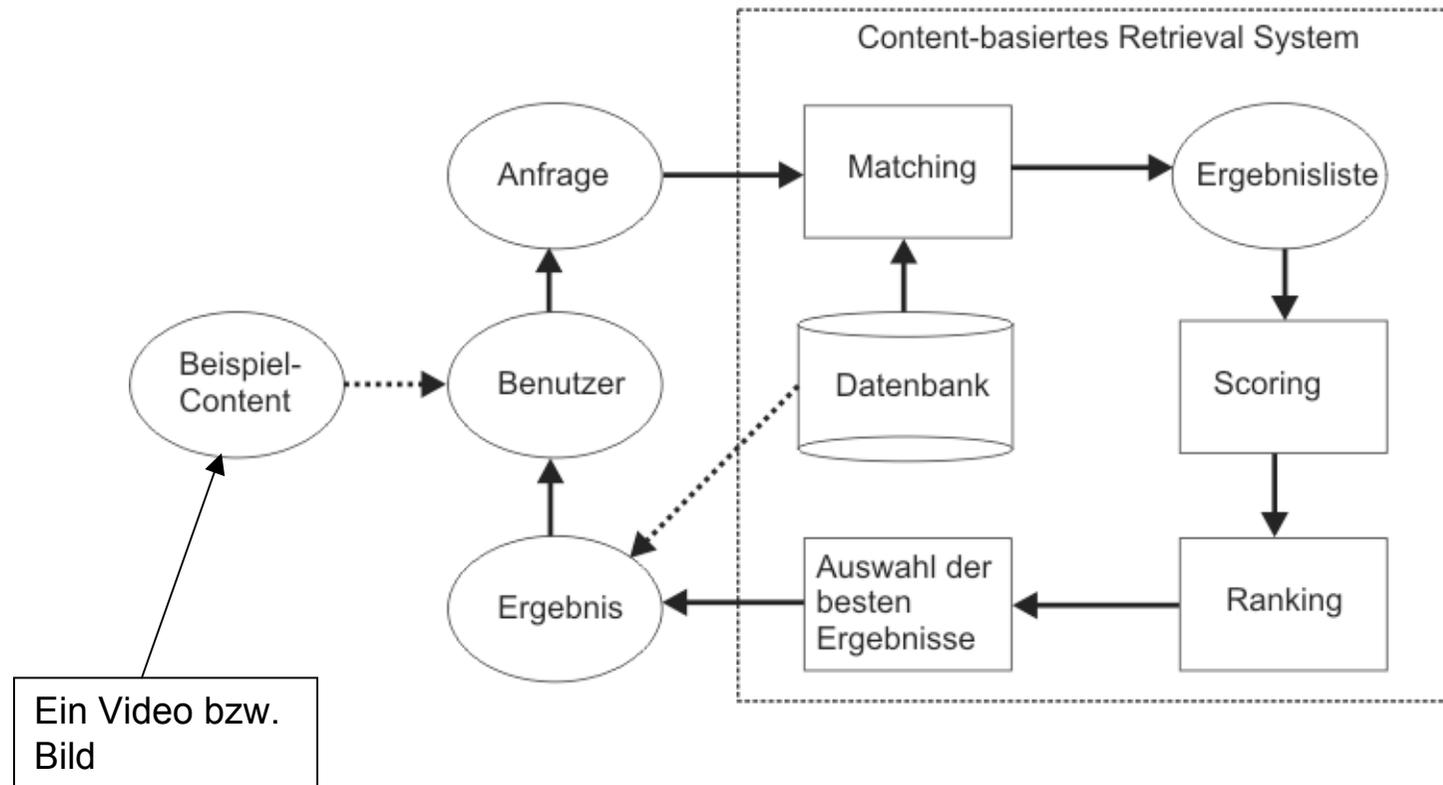


Quelle: Smith, J.R., Naphade, N., Natsev, A., Tesic, J.: Multimedia Research Challenges for Industry.

# Anfrage und Suche

- Von Retrievalsystemen unterstützte Anfragemöglichkeiten:
  - Textbasierte Suche
  - Query by Example
  - Semantische Suche
- Meist Kombination verschiedener Anfragen möglich  
→ Zusammenführung der Ergebnisse („Fusion“) nötig

# Query-by-Example



# Zusammenführung/Fusion

- Kombination verschiedener Anfragen hat höchste Effizienz
- Zusammenführung der verschiedenen Ergebnisse nötig
- Durchführung meist nach Ranking oder nach Scoring
- Normalisierung nötig, da Ergebnisse der Suchmethoden heterogen

# Video Summarization

- Problem bei der Präsentation von Suchergebnissen: Video ist zeitabhängiges Medium
- Ziel: schneller Überblick über den Inhalt eines Videos - Zusammenfassung
- Videosummary aus Keyframes
  - „optimaler“ Keyframe rechenintensiv zu bestimmen (z.B. durch Clustering)
  - Einfachster Ansatz: Erster Frame eines Shots
  - Erweiterte Ansätze: Wahl von Keyframes anhand von Metadaten, zum Beispiel MPEG-7 Motion Activity
- Videosummary als Video
  - Schnelleres Abspielen bzw. Subsampling

# Beispiel eines Videosummary



- Den Keyframes wurden „Importance Scores“ zugewiesen
- „Wichtigere“ Shots sind durch größere Keyframes repräsentiert
- Packalgorithmus baut daraus ein Layout

# Fazit

- Video beinhaltet Bilder, Audio und Text (bzw. Sprache)  
→ Indexierung und Retrieval sehr komplex
- Low Level Metadaten allein ermöglichen keine effiziente Suche
- High Level Metadaten können meist nur für bestimmte Domänen extrahiert werden
- Semantische Indexierung aufwendig  
→ noch ein weites Forschungsfeld

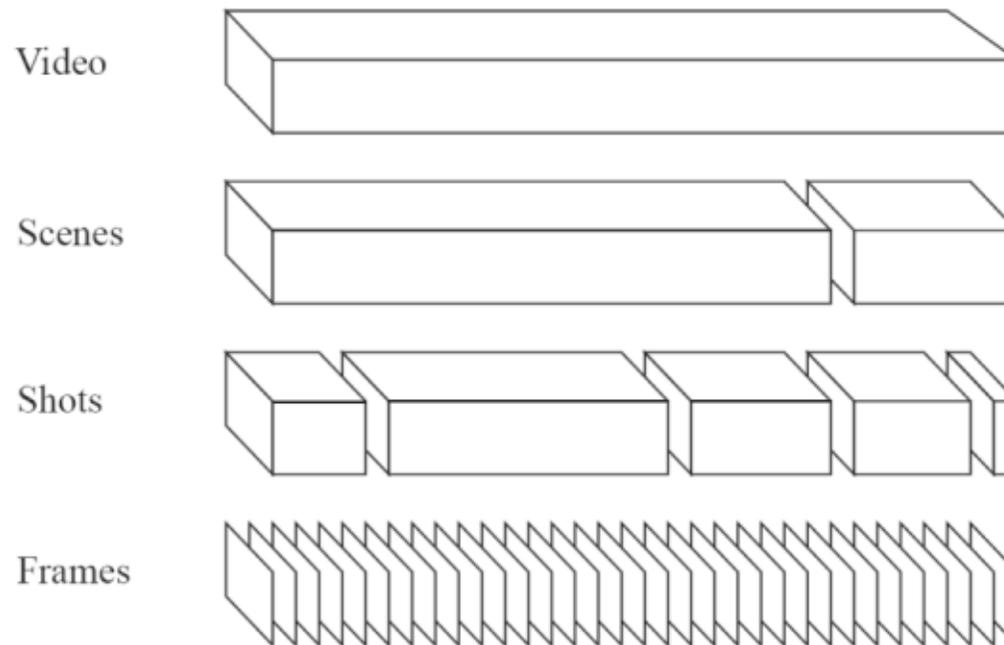
# Diskussion



# Literatur

- Adams, B., Amir, A., Dorai, C., Ghosal, S., Iyengar, G., Jaimes, A., Lang, C., Lin, C., Natsev, A., Naphade, M., Neti, C., Nock, H.J., Permuter, H.H., Singh, R., Smith, J.R., Srinivasan, S., Treng, B.L., Ashwin, T.V., Zhang, D.: (IBM Research TREC-2002 Video Retrieval System)  
<http://www.itl.nist.gov/iaui/894.02/projects/t2002v/results/notebook.papers/ibm.smith.pdf>
- Divakaran, A., Peker, K.A., Radhakrishnan, R., Xiong, Z., Cabasson, R.: Video Summarization using MPEG-7 Motion Activity and Audio Descriptors. Technical report  
<http://www.merl.com/reports/docs/TR2003-34.pdf>
- IBM T. J. Watson Research Center: (MARVEL: Multimedia Analysis and Retrieval System)  
<http://www.research.ibm.com/marvel/Marvel%20Whitepaper.pdf>
- Lienhart, R.: Comparison of Automatic Shot Boundary Detection Algorithms.  
<http://www.lienhart.de/spie99.pdf>
- Mc Donald, K., Smeaton, A.F.: A Comparison of Score, Rank and Propability-Based Fusion Methods for Video Shot Retrieval. In Leow, W.K., Lew, M., Chua, T.S., Ma, W.Y., Chaisorn, L., Bakker, E., eds.: Image and Video Retrieval. Springer-Verlag Berlin- Heidelberg (2005) 61-70
- Uchihashi, S., Foote, J., Girgensohn, A., Boreczky, J.: Video Manga: Generating Semantically Meaningful Video Summaries. <http://www.fxpall.com/publications/FXPAL-PR-99-136.pdf>

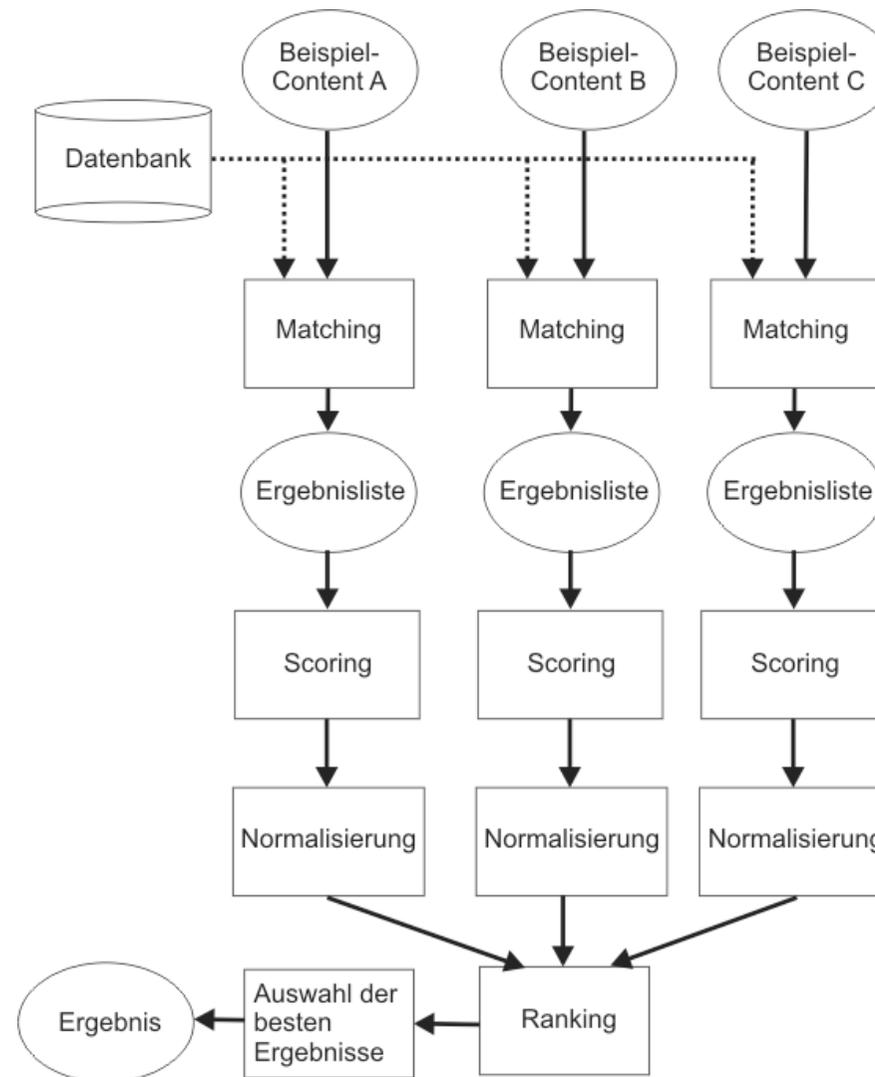
# Struktur eines Videos



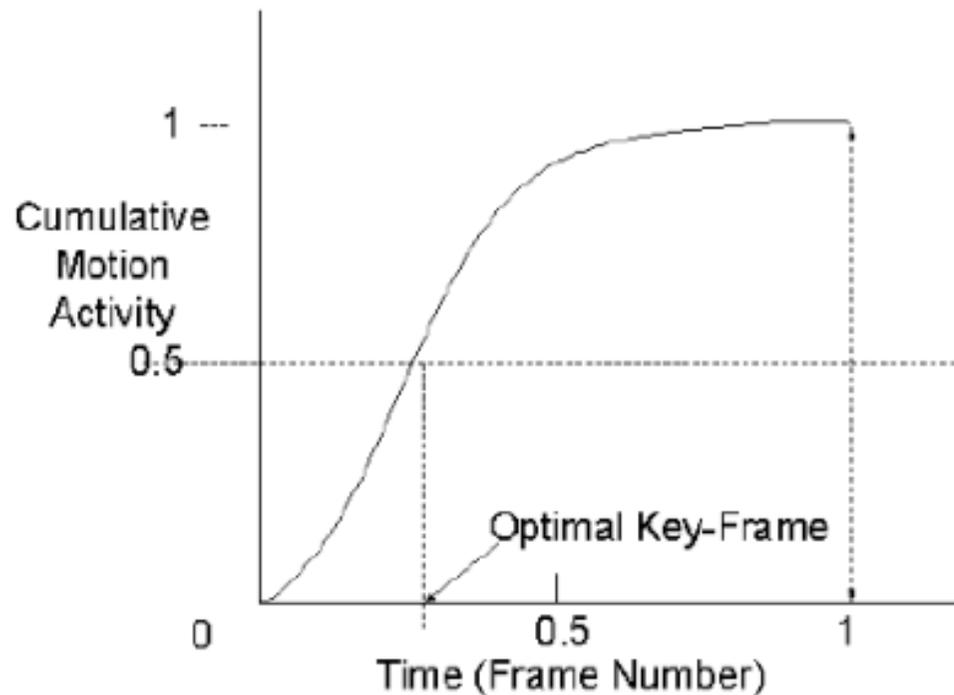
**Quelle:** Lienhart, R., Pfeier, S., Eelsberg, W.: Video Abstracting.

<http://www.lienhart.de/cacm.pdf>

# Score-basierte Fusion



# Kumulative Bewegungsintensität



**Quelle:** Divakaran, A., Peker, K.A., Radhakrishnan, R., Xiong, Z., Cabasson, R.: Video Summarization using MPEG-7 Motion Activity and Audio Descriptors.

<http://www.merl.com/reports/docs/TR2003-34.pdf>