

# Multimedia-Datenbanken

## Kapitel 6: Multimedia-Daten – Video

---

Friedrich-Alexander-Universität Erlangen-Nürnberg  
Technische Fakultät, Institut für Informatik  
Lehrstuhl für Informatik 6 (Datenbanksysteme)

**Prof. Dr. Klaus Meyer-Wegener**

Wintersemester 2002 / 2003

Technische Universität Kaiserslautern  
Fachbereich Informatik  
AG Datenbanken und Informationssysteme

**Dr. Ulrich Marder**

Wintersemester 2003 / 2004

---

# 6.1 Video-Daten

- ❑ **Kombination von Bild (Rasterbild oder Graphik) und Ton**
- ❑ **Rohdaten:**
  - enormes Datenvolumen:  
25 Bilder pro Sekunde à 250 KB,  
Tonaufzeichnung mit 11 Bit bei 16 kHz:
  - 6250 KB + 22 KB  $\approx$  6,3 MB pro Sekunde
  - zunächst spezielle Speichergeräte:  
Videorecorder (VCR), analoge Bildplatte ("LaserDisc")
- ❑ **Registrierungsdaten:**
  - Aufzeichnungsformat (VHS, Betacam, MII, 1-Zoll-B, U-matic u.v.a.)  
bzw. zu benutzender Recorder
  - Time-Codes
  - Dateiformat (MPEG, ... )
- ❑ **Beschreibungsdaten:**
  - Szenenstruktur: die einzelnen Schnitte (erstes Bild, Länge);  
Typ der Aufnahme: Totale, Halbtotale, Schwenk, Zoom

## ❑ **"Joint Photographic Expert Group"**

- gemeinsame Aktivität von ISO/IEC JTC1/SC2/WG10 und Kommission Q.16 der CCITT SGVIII
- seit 1992 ISO-Norm (international standard)

## ❑ **genormtes Format für Rasterbild**

- hohe Komprimierungsrate möglich (einstellbar), allerdings verlustbehaftet
- als "Motion-JPEG" auch für Video verwendet und Grundlage für MPEG (s. unten), daher hier behandelt

## ❑ **parametrisierbar:**

- Anwender kann entscheiden über Qualität des Bildes, Dauer der Kompression, Größe des komprimierten Bildes

## ❑ **vier Modi:**

- verlustbehaftet, sequenziell, DCT-basiert: "baseline process", Basis-Modus; muss von jedem JPEG-Decoder unterstützt werden
- verlustbehaftet, erweitert, DCT-basiert: Menge von Alternativen zum Basis-Modus

## □ **Modi (Forts.):**

- verlustfrei:  
geringerer Kompressionsfaktor  
(selten benutzt, da nicht besser als andere Formate)
- hierarchisch:  
Bild in verschiedenen Auflösungen,  
jeweils in einem der anderen drei Modi

## □ **Verfahren**

- siehe Literatur
- Schritte: Bildung von 8x8-Blöcken, diskrete Cosinus-Transformation (DCT), Quantisierung, Entropiecodierung

## □ **im erweiterten Modus progressive Codierung möglich:**

- Bild wird nicht zeilenweise aufgebaut (sequenziell),  
sondern von unscharf zu scharf

- ❑ **Norm für Übertragung von Bewegtbildern über ISDN**
  - Schmalband-ISDN-Anschluss: zwei B-Kanäle (à 64 kbit/s), ISDN-Hierarchie (für Nebenstellenanlagen): bis zu 30
  - 1984 Study Group XV der CCITT eingerichtet, 1990 CCITT-Empfehlung H.261 "Video Codec for Audiovisual Services at p x 64 Kbit/s" verabschiedet
- ❑ **Bildgröße**
  - 288 Zeilen mit jeweils 352 Pixel (3 : 4) für Luminanz, 144 x 176 für Chrominanz (Common Intermediate Format – CIF; für Videokonferenz)
  - daneben auch halbe Auflösung – obligatorisch (QCIF; für Bildtelefonie)
  - Kompressionsrate von 47 : 1 (für QCIF, 10 fps und 64 kbit/s)
- ❑ **Intraframe:**
  - nur Daten aus dem jeweiligen Bild; 8x8-Pixel-Blöcke, DCT, Entropiecodierung (wie JPEG)
- ❑ **Interframe:**
  - Daten aus anderen Bildern verwenden; ähnliche Makroblöcke suchen, Bewegungsvektor und Differenz speichern
- ❑ **Datenstrom:**
  - komprimierte Bilder, Angaben zur Fehlerkorrektur, Bildnummern (5 Bit), Kommandos zum "Einfrieren" des zuletzt angezeigten Bildes

- ❑ **"Moving Picture Expert Group"**
  - zunächst Untergruppe von ISO/IEC JTC1/SC2/WG8, inzwischen eigene WG11 in SC29
- ❑ **Bewegtbild und Audio**
  - konstante Datenrate von maximal 1.856.000 bit/s (auch für CD-ROM geeignet)
  - 1993 ISO-Norm
  - JPEG übernommen, Folge von Einzelbildern möglich
- ❑ **asymmetrische Kompression:**
  - Aufwand zum Codieren darf viel höher sein als der zum Decodieren
  - max. 768 x 576 Pixel
- ❑ **I-Bilder (intra coded pictures):**
  - unabhängig von anderen Bildern, wie JPEG
- ❑ **P-Bilder (predictive coded pictures):**
  - benötigt vorhergehende (I- oder P-) Bilder
- ❑ **B-Bilder (bidirectionally predictive coded pictures):**
  - benötigt vorhergehende und nachfolgende (I- oder P-) Bilder

# MPEG (2)

- ❑ **D-Bilder (DC coded pictures):**
  - Einzelbilder, nur grob dargestellt, für schnellen Vorlauf
- ❑ **Speicherungsreihenfolge**
  - kann wegen der B-Bilder von Präsentationsreihenfolge abweichen!
- ❑ **Entscheidung über I-, P- und B-Bilder**
  - anwendungsabhängig
  - Heuristik: IBBPBBPBBIBBPBBPBBBI .....
  - Auflösung für wahlfreien Zugriff dann 9 Bilder (330 ms), sehr gute Kompressionsrate
- ❑ **Audio:**
  - wie Audio-CD oder DAT
- ❑ **MPEG-2:**
  - 4–10 Mbit/s,  
Skalierung (verschiedenen Qualitäten für Auflösung, Bildrate usw.)

## □ ISO/IEC international standard 14496

- definiert ein Multimedia-System für die interoperable Kommunikation von komplexen Szenen, die Audio, Video, synthetisches Audio (MIDI) und Graphik enthalten
- begonnen Juli 1993
- Committee Draft November 1997
- IS April 1999
- für Autoren: mehr Wiederverwendung und Flexibilität
- für Netzbetreiber: generische Dienstgüte-Deskriptoren
- für Nutzer: mehr Interaktionen
- Normierung der
  - Codierung von Medienobjekten (aufgenommen oder synthetisch)
  - Komposition von Medienobjekten zu Szenen
  - Multiplexer und Synchronisierer für Transfer
  - Interaktion

## □ Teile der Norm

- Systeme, Visuell, Audio, Conformance-Tests, Referenz-Software und Delivery Multimedia Integration Framework (DMIF)

## □ System

- Rahmen für die Integration der Teile zu Szenen
- hierarchische Struktur (Graph)
- Nutzung des Ansatzes der Virtual Reality Modeling Language (VRML)

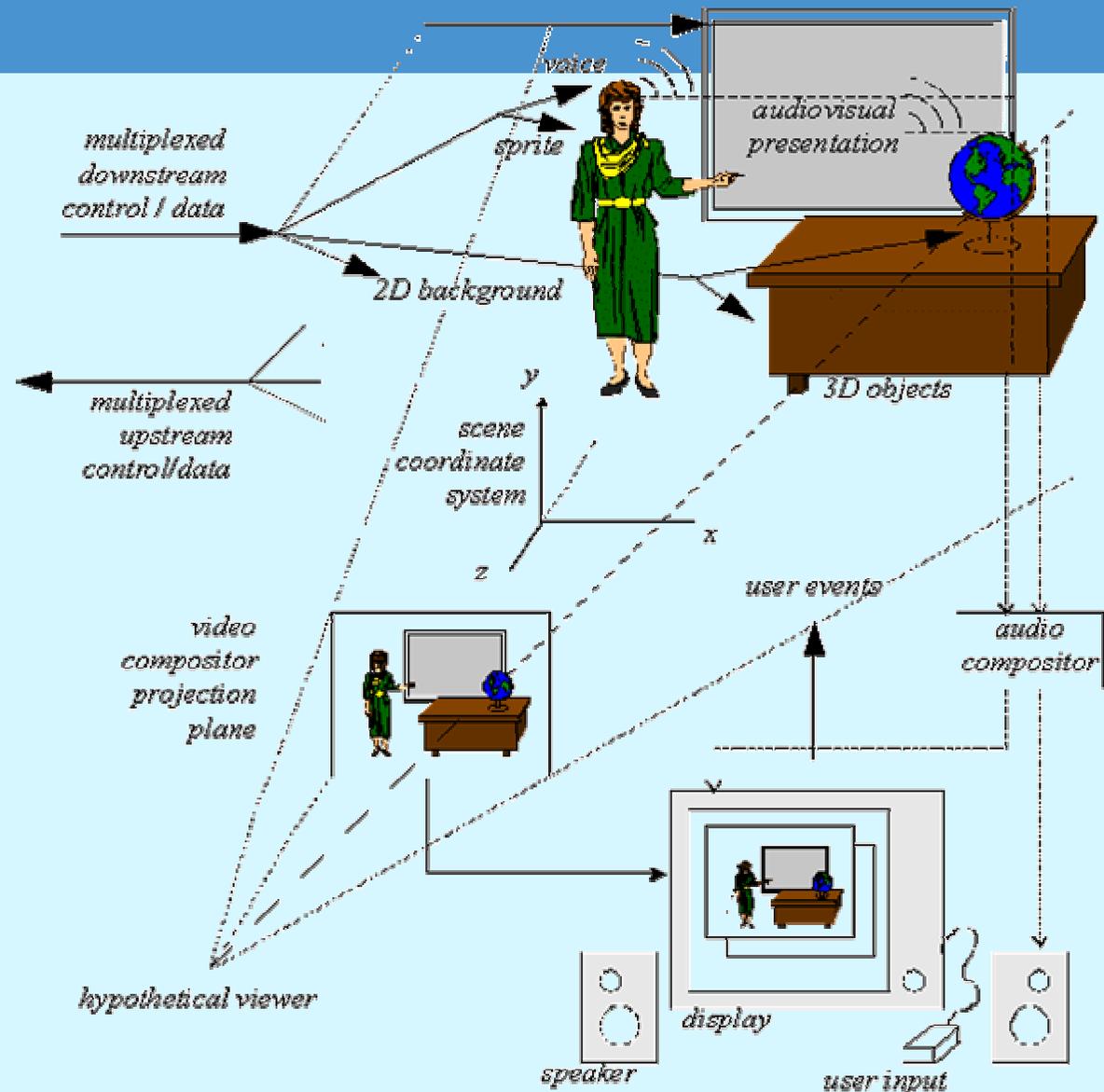
## □ Komposition

- Frames für Audio und Video
- außerdem noch **Objekte**, die eine Szene bilden
  - Videoobjekte in verschiedenen 2D-Formen (shapes)
  - Audioobjekte, möglicherweise Videoobjekten zugeordnet
- Szenenbeschreibung
  - Text, editierbar, oder binär (Binary Format for Scene Description, BIFS)

# MPEG-4 (3)

## □ Komposition einer Szene

- beliebige Platzierung im Koordinatensystem
- Gruppierung (z. B. *voice/sprite*)
- Interaktiver Wechsel des Betrachterstandpunkts möglich
- Information bleibt bei Codierung erhalten



## □ Familie von Codecs zu MPEG-4

- zuerst Microsoft MPEG-4 v3
- "nachgebaut" in DivX ;- ) bzw. DivX 3.x
  - nicht verwechseln mit Digital Video Express (DIVX/DVE) von Circuit City
- DivX 4 und nun DivX 5 von DivXnetworks ([www.divx.com](http://www.divx.com))
  - kommerziell
- OpenDivX von Project Mayo ([www.projectmayo.com](http://www.projectmayo.com))
  - open source
  - höhere Kompressionsrate als DivX
- XviD ([www.xvid.org](http://www.xvid.org))
  - open source
  - schnell
- sehr dynamische Entwicklung ... ([www.doom9.org](http://www.doom9.org))

## 6.2 Video-Operationen

### □ **Abspielen**

- auf einem separaten Monitor oder in einem Fenster der Workstation
- Abspielvorgang ist Prozess!  
Beeinflussung durch Betrachter muss möglich sein (Stop, Pause, Resume, ... )
- Standbild (evtl. ins Programm holen als Rasterbild)
- Zeitlupe, Zeitraffer
- zahlreiche Möglichkeiten elektronischer Manipulation (z. B. Überlagerung, Bluebox, Chroma-Key)

### □ **Schneiden, Kopieren, Zusammensetzen**

- große Probleme bei verlustbehafteter Komprimierung: vorher dekomprimieren und nachher wieder komprimieren, erhöht den Verlust!

### □ **Tonspur ersetzen (resynchronisieren)**

## 6.3 Video-Suche

- ❑ **metadatenbasiert:**
  - Titel, Autor/Produzent/Regisseur, Produktionsdatum, Typ usw.
- ❑ **textbasiert:**
  - Untertitel
- ❑ **audiobasiert:**
  - Tonspur
  - Abschnitte mit Sprache oder Musik
- ❑ **inhaltsbasiert:**
  - Bilder (frames)
  - alle oder in Gruppen (Schnitte, siehe unten)
- ❑ **kombiniert:**
  - mehrere der genannten Techniken zusammen
- ❑ **Ziel:**
  - Video *und* Teil davon (Szene)

## ❑ Voraussetzung: Segmentierung

## ❑ Struktur

- Schnitte (shots)
  - eine Kameraeinstellung
  - Problem: Überblendungen
- Szenen
  - mehrere Schnitte zusammen
  - mit gleicher Situation, d. h. kein Zeitsprung
  - z. B. Dialog
  - schwieriger zu identifizieren
  - Hilfsmittel (falls digital verfügbar) z. B. Storyboards, Drehbuch

## ❑ Schlüsselbilder (key frames)

- repräsentieren Szene
- suchbar wie Bild

## □ Differenz zwischen zwei unmittelbar aufeinander folgenden Bildern

- quantitativ erfassen: Metrik
- Schwellenwert

## □ einfachstes Maß: Summe der Pixeldifferenzen zwischen benachbarten Bildern

- nicht effektiv; zu viele fehlerhafte Erkennungen
- schnelle Bewegungen größerer Objekte – große Differenzen

## □ Summe der Histogramm-Differenzen

- Verteilung bleibt auch bei Bewegung noch gleich

$$SD_i = \sum_j |H_i(j) - H_{i+1}(j)|$$

- bei Farbe Komponenten einzeln betrachten

# Segmentierung (2)

## ❑ **Schwellenwert**

- kritisch!
- Ansatz: mittlere Differenz von benachbarten Bildern plus kleine Toleranz

## ❑ **nicht anwendbar bei graduellen Übergängen**

- dissolve, wipe, fade-in, fade-out
- Differenzen dabei größer als innerhalb von Schnitten, aber kleiner als bei harten Wechseln

## ❑ **Idee: zwei Schwellenwerte**

- Differenz größer als  $T_b$ : harter Wechsel
- Differenz kleiner als  $T_b$ , aber größer als  $T_s$ : evtl. gradueller Wechsel
- dann alle aufeinander folgenden Differenzen  $> T_s$  addieren und erneut mit  $T_b$  vergleichen: wenn größer, gradueller Wechsel
- trotzdem schlechte Erkennung:  $< 16\%$

# Segmentierung (3)

## ❑ fehlerhafte Erkennung

- Panning (Kameraschwenk) und Zooming
  - Bewegungserkennung nutzen
- Beleuchtungswechsel (Scheinwerfer, Wolken, Reflexionen)
  - Normalisierung vor Differenzenbildung

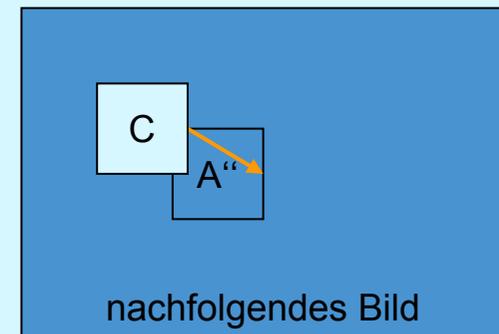
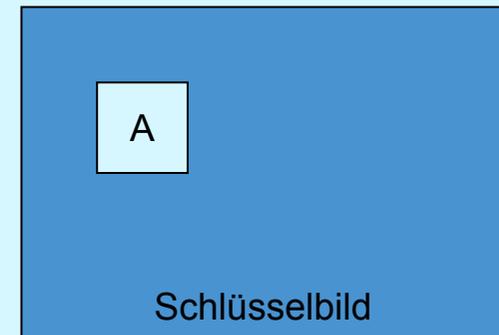
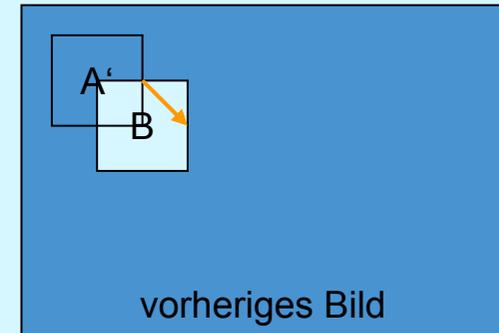
## ❑ weitere Verfahren

- Bewegungsfiler vor Differenzenbildung
- Kantenerkennung (edge detection)
  - zählen, wie viele Kanten verschwinden oder auftauchen
  - Schwellenwert
- moderne Kameras können die Information mit aufzeichnen
  - Position, Zeit, Orientierung

- ❑ **key frames oder representative frames (r frames)**
- ❑ **Wie viele pro Schnitt?**
  - genau eins
  - proportional zur Länge, z. B. für jede Sekunde eins
  - inhaltsabhängig (Bewegung etc., Unterstruktur bilden)
- ❑ **Welche?**
  - je nach Anzahl; "Segment" ist entweder ganzer Schnitt, eine Sekunde oder Teilstück
  - das erste Bild
  - "Durchschnittsbild": jedes Pixel aus dem Durchschnitt der entsprechenden Pixel aller Bilder des Segments, dann das Bild, das diesem Durchschnittsbild am ähnlichsten ist
  - statt der Pixel Histogramm verwenden
  - Trennung Vorder- und Hintergrund; zu künstlichem Bild zusammenfassen

# Bewegungsinformation

- ❑ **Ergänzung der Schlüsselbilder**
- ❑ **ableiten aus Bewegungsvektoren**
- ❑ **Parameter:**
  - Bewegungsinhalt
    - Gesamtmenge von Bewegung im Schnitt
  - Bewegungsgleichförmigkeit
  - Schwenk horizontal
  - Schwenk vertikal
- ❑ **zum ganzen Video, jedem Schnitt oder jedem Schlüsselbild**



## □ Filme werden "mit System" gedreht

- Bildsprache
- berühmtes Buch: Daniel Arijon: *Grammar of the film language*. Hastings House : New York, 1976
- z. B. Dialog:
  - man sieht immer den, der spricht
  - Kamera "springt" hin und her

## □ Idee:

- Schlüsselbilder von Schnitten einer Szene
- Ähnlichkeit zu Schlüsselbildern anderer Schnitte
- ergibt Muster: ABABAB...
- jeweils auch mit Zeit: Intervall
- zuordnen zu Stereotypen der Produktion, hier: Dialog

# Weitere Möglichkeiten

- ❑ **Suche über Objekte**
  - MPEG-4
- ❑ **Suche über Metadaten**
- ❑ **Suche über Annotation**
  - MPEG-7
- ❑ **Kombination**
  
- ❑ **für das Suchresultat:  
Browsing**

