



3 Kompression

3.6 Videokompression

Videokompression



- Die hohen Datenraten erzwingen Kompression
 - zur Speicherung
 - zur Übertragung über Netze
- Verlustfreie vs. Verlustbehaftete Kompression
 - Video verkraftet verlustbehaftete Kompression



Geschwindigkeit der Videokompression



- Symmetrische vs. Asymmetrische Kompression
 - Verhältnis zwischen Kompression und Dekompression
 - Gleiche Dauer → symmetrisch
 - Unterschiedliche Dauer → asymmetrisch
- Echtzeitkompression
 - Verfahren ist mindestens so schnell wie die Bildwiederholrate
 - zumindest die Dekompression sollte in Echtzeit erfolgen können



Intraframe vs. Interframe



- Intraframe Kompression
 - Einzelbild wird unabhängig von anderen komprimiert
 - Räumliche Kompression
 - Verfahren wie bei (Stand-)Bildern
- Interframe Kompression
 - Differenzen unter „benachbarten“ Bildern werden berücksichtigt
 - Zeitliche Kompression
 - Keyframes als Stützbilder und Differenzframes zwischen den Keyframes



Motion JPEG, MJPEG, MJPEG-A



- Einzelbilder werden mit JPEG kodiert
- Anwendung direkt auf $Y'CbCr$ Samples
- Einstellbare Qualität bzw. Kompression
- Meist hardware-unterstützt auf Capturekarten
 - Verfahren leicht unterschiedlich von Hersteller zu Hersteller
 - MJPEG-A wird von QuickTime unterstützt



Vektorquantisierungsverfahren



- Oft vorkommende Bestandteile eines Bildes werden als Vektor in einem Codebook abgelegt
- Tatsächliches Teilbild wird durch Index ins Codebook ersetzt
- Sehr asymmetrische Verfahren
- Ziel: schneller Software Decoder
- Beispiele
 - Cinepak, Intel Indeo, Sorenson (QuickTime)



MPEG



- Standardisierung der Motion Picture Experts Group (ISO/IEC 11172)
- Sequenz von MPEG-Normen
 - MPEG-1, MPEG-2, MPEG-4, MPEG-7
- MPEG behandelt Videoströme inklusive zugehöriger Audioströme
- allgemeiner Standard, der viele Anwendungen unterstützt
- Das exakte Kompressionsverfahren ist nicht standardisiert, lediglich die Datenstromsyntax und der Dekompressor



MPEG-1



- Ziele
 - Kompression eines Videosignals inklusive Audio auf 1,5 Mbit/s (T1 Datenrate)
 - Für progressives Rendering, also Computer
 - Qualität
 - 4:2:0 Subsampling
 - Auflösung SIF (source input format, 352 x 240) bei 25 fps
 - Audio in CD-Qualität
 - Wahlfreier Zugriff innerhalb von 0,5 s ohne wesentlichen Qualitätsverlust
 - schnelle Vorwärts-/Rückwärtssuche
 - Video rückwärts abspielbar
 - Video gut editierbar



MPEG-1 Eigenschaften



- Räumliche Kompression der Einzelbilder
 - JPEG – Variante
- Temporale Kompression
 - Differenzbilder zwischen Stützbildern (difference frame; key frame)
 - Idee:
nur die Veränderung gegenüber vorhergehenden Frames ist zu kodieren



Temporale Kompression



- Naiver Ansatz
 - Subtrahiere Frame vom Vorgängerframe
 - Es werden viele Pixel 0 sein
 - → dies lässt sich stärker komprimieren
- Gewählter Ansatz
 - Bewegungskompensation (motion compensation)
 - Veränderte Pixel stellen eine Bewegung dar
 - Es genügt den Bewegungsvektor zu jeder Region bewegter Pixel anzugeben



Temporale Kompression



- MPEG erkennt keine Objekte
- Bewegungskompensation erfolgt auf der Ebene von Makroblöcken (16 x 16 Pixel)
- Fragestellung:
 - An welcher Position kommt ein Makroblock X im neuen Frame vor?
- Aus Effizienzgründen wird diese Suche auf eine eingeschränkte Umgebung angewendet



Frametypen



- I-Frame (Intra Frame)
 - Intra-kodiertes Vollbild
 - ähnlich JPEG-Standbild
 - DCT, Quantisierung, Lauflängenkodierung, Huffman-Kodierung
- P-Frame (Predicted Frame)
 - Vorwärts-Vorhersage unter Bezugnahme auf vorhergehende I- und P-Bilder
 - Bewegungsvektoren zu bekannten Makroblöcken
 - JPEG-ähnlich kodierte „neue“ Makroblöcke



Frametypen



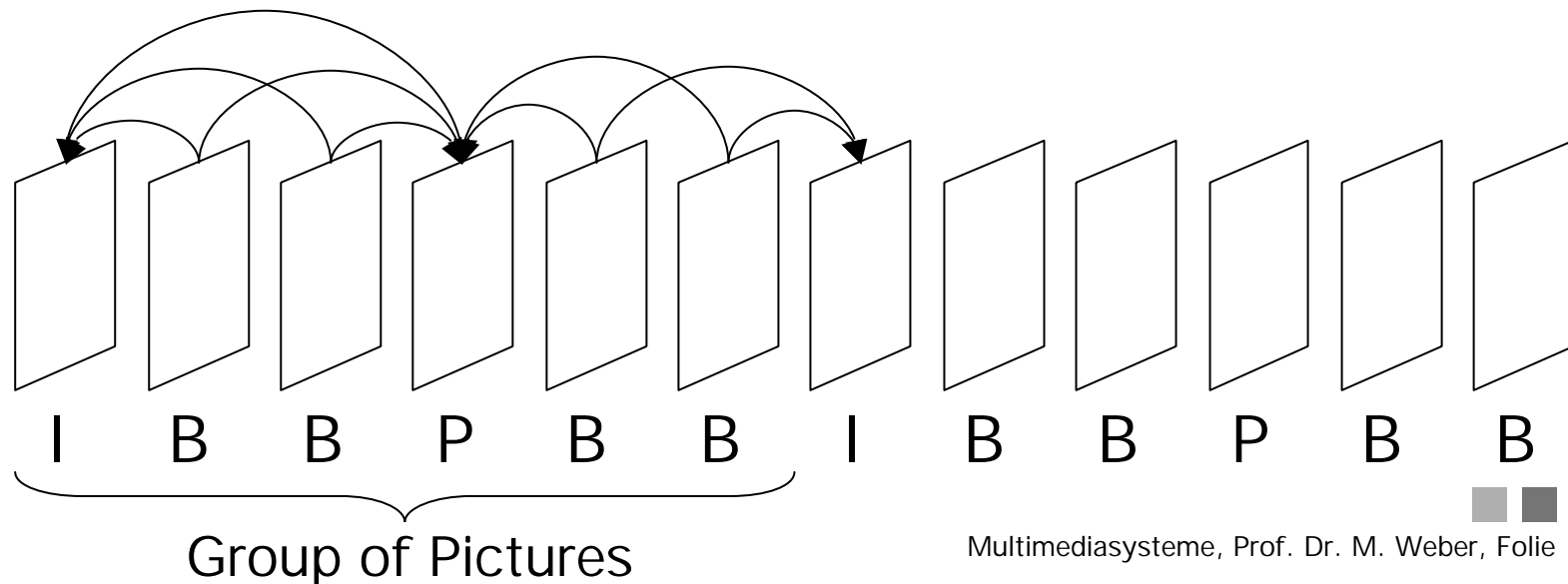
- B-Frame (Interpolated Frame)
 - Bidirektional „vorhergesagte“ Bilder unter Bezugnahme auf vorhergehende und nachfolgende Bilder (nur I- und P-Frames)
 - Interpolation zwischen Makroblöcken möglich
- D-Frame
 - nur DC-Koeffizienten der JPEG-Blöcke
 - Für schnellen Vor- und Rücklauf



Group of Pictures - GOP



- Folge von I, P und B-Bildern ist nicht standardisiert
 - kann von der Anwendung bestimmt werden, auch adaptiv während der Laufzeit
 - Bildqualität und Kompressionsrate lassen sich dadurch anpassen



Vorwärts-Prädiktion bei P-Frames



- wohin haben sich Makroblöcke bewegt?

$$\min(x, y) \sum_{i=0}^{15} \sum_{j=0}^{15} |I_1(i, j) - I_2(i + x, j + y)|$$

- Kodierung der Makroblöcke
 - Bewegungsvektor + Fehlerinformation (JPEG-kodiert)
 - voller JPEG-Block für „neue“ Makroblöcke
 - Bewegungsvektoren werden differenzkodiert



Bidirektionale Prädiktion bei B-Frames



- zwei Typen Pixelinformation
 - Objekt + auftauchender Hintergrund
- bei der Kodierung
 - Zwischenframes blockweise per Interpolation ermitteln
 - Unterschied zum tatsächlichen Block ermitteln (error block)
 - Korrekturterm mit JPEG behandeln
 - Bidirektional für Behandlung von auftauchendem Hintergrund



Bidirektionale Prädiktion bei B-Frames



- Drei Kodierungsoptionen pro Makroblock
 - Fehlerblock zur Korrektur der Interpolation
 - Bewegungsvektor
 - vorwärts, rückwärts, beide
 - Kodiertes Original
- Probleme der B-Frame Kodierung
 - Erhöhte Speicheranforderungen (3 Bilder)
 - Verzögerungen, da auf spätere Frames gewartet werden muss



MPEG Wiedergabe

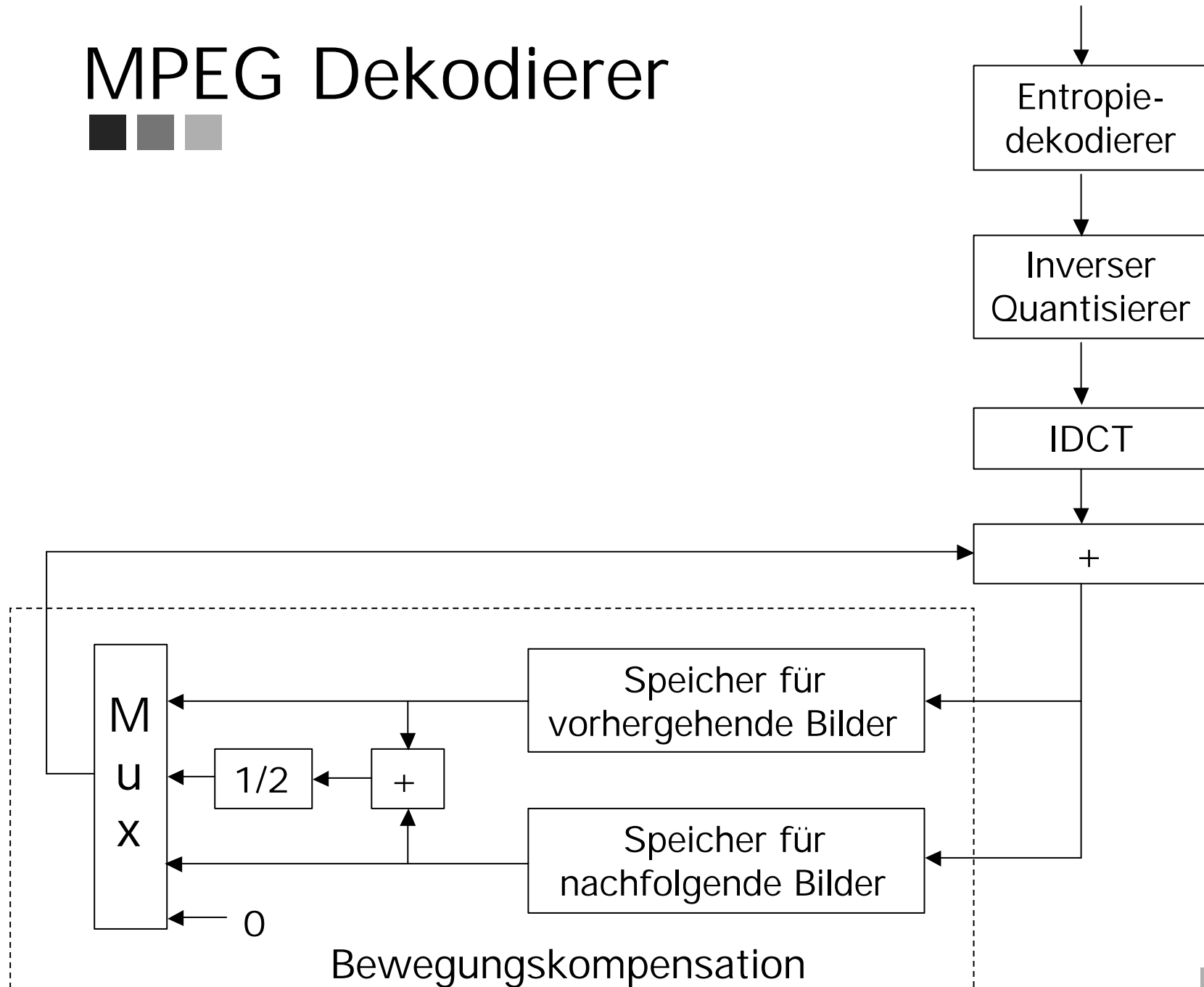


- Group of Pictures ist in Display Order angegeben
- Wegen bidirektionaler Prädiktion wird die Sequenz in Bitstream Order gespeichert

Display Order	I	B	B	P	B	B	I	B	B	P	B	B	I	B	B	P	...
Bitstream Order			I	P	B	B	I	B	B	P	B	B	I	B	B	P	...
Präsentation				I	B	B	P	B	B	I	B	B	P	B	B	I	...



MPEG Dekodierer



MPEG-Datenstrom



- Synchronisationspunkte auf Teilbildebene
 - Aufsetzen der Dekompression nach einem Übertragungsfehler
- sechs Ebenen
 - 1. Ebene: Sequence
 - Videostrom
 - 2. Ebene: Bildgruppe, GOP
 - IBBPBB, IBBBPBBB, IBBPBBPBB, IPPPPP, IIIIIII, ...
 - 3. Ebene: Bild
 - I, P, B



MPEG-Datenstrom



- 4. Ebene: Scheibe (slice)
 - Bildausschnitt mit wählbarer Größe zur Resynchronisation nach einem Fehler
- 5. Ebene: Makroblock
 - 16 x 16 Pixel, auf der die Bewegungskompensation aufbaut
- 6. Ebene: Block
 - 8 x 8 Pixel für DCT, Quantisierung, Lauflängenkodierung, Huffman



Variable Bitrate



- I, P und B Frames haben unterschiedliches Datenvolumen
- Szenenwechsel
 - freiwerdender Hintergrund führt zum Anstieg des Volumens
- Datenrate ist begrenzt durch z.B. CD-ROM-Leserate bzw. Übertragungskanal
 - Bei Source Input Format (352 x 240) und 30 fps → 1,86 Mbit/s
 - = Datenrate von CD Video



MPEG-2



- MPEG-2 erweitert MPEG-1 für höhere Bandbreiten und bessere Bildqualität
 - 4MBit/s – 100 MBit/s
 - Für digitale Broadcastmedien
 - Interlacing-Unterstützung
 - Auflösung von CCIR 601 bis HDTV
- gemeinsam von ISO (13818) und ITU entwickelt (H.262)
- skalierbare Datenströme
 - Bedienung von Empfängern verschiedener Leistungsklassen



Skalierbarkeit in MPEG-2



- SNR scalability
(SNR = signal to noise ratio)
 - Kodierung jeden Bildes in mehreren Schichten
 - base, middle, enhanced
 - wer die Basis-Schicht dekodiert erhält mäßige Bildqualität
 - wer auch höhere Schichten dekodiert, erhält Schicht für Schicht eine bessere Bildqualität
 - Beispiel:
Übertragung der U-/V-Komponenten in der Basis-Schicht mit einem Viertel der Auflösung von Y, in höheren Schichten höhere Auflösung



Skalierbarkeit in MPEG-2



- Spatial scalability
 - Kodierung des Bildes mit verschiedenen Pixel-Auflösungen
 - Beispiel:
Standard-Fernsehen und HDTV wird in gemeinsamem Datenstrom übertragen
- Temporal scalability
 - Übertragung von wenigen Frames/s in der Basis-Schicht und mehr Frames/s in höheren Schichten



Skalierbarkeit in MPEG-2



- Data partitioning
 - ermöglicht eine Skalierbarkeit bezüglich der Fehler-Resistenz
 - die wichtigsten Bestandteile des MPEG-Stromes werden in der Basis-Schicht übertragen, weniger wichtige Teile in den höheren Schichten
 - Beispiel:
 - niedrige Koeffizienten der DCT und Bewegungsvektoren in der Basis-Schicht
 - Basis-Schicht kann dann mit mehr Redundanz zur Fehlersicherung versehen werden
 - wird damit fehlerärmer empfangen als die höheren Schichten



MPEG-2 Videoprofile



	Simple profile	Main profile	SNR scalable profile	Spatially scalable profile	High profile
	No B frames No scalable	B frames Not scalable	B frames SNR scaling	B frames spatial scaling	B frames spatial or SNR scaling
High level 1920x1152x60		= 80 Mbit/s			= 100 Mbit/s
High level 1440x1152x60		= 60 Mbit/s		= 60 Mbit/s	= 80 Mbit/s
Main level 720x576x30	= 15 Mbit/s	= 15 Mbit/s	= 15 Mbit/s		= 20 Mbit/s
Low level 352x288x30		= 4 Mbit/s	= 4 Mbit/s		



MPEG-4



- Wesentliche Quelle
 - Rob Koenen (Ed.): MPEG-4 Overview, V.21, März 2002.
<http://mpeg.telecomitalia.com/standards/mpeg-4/mpeg-4.htm>
- Ziele von MPEG-4
 - zunächst für Video mit wenigen Ressourcen gedacht
 - z.B. im Mobilfunk, Videophone
 - Bandbreiten: 4,8 – 64 Kbit/s
 - Bildparameter: 176 x 144 x 10 Frames/s
 - Funktionalität zur Analyse und Manipulation von Bildinhalten
- Normung 1999
 - ISO/IEC 14496



MPEG-4 Ziele



- Standardisierte Verfahren zu:
 - Repräsentation von Audio/Visual Objects, AVO
 - für jedes Objekt kann die Kodierungstechnik unabhängig von anderen AVOs, Umgebung und Hintergrund gewählt werden
 - AVOs können natürliche und synthetische Objekte sein
 - Komposition der Objekte
 - Zusammengesetzte AVOs bilden audiovisuelle Szenen
 - Multiplexing und Synchronisation zwischen AVOs
 - Quality-of-Service zur Netzwerkübertragung
 - Interaktion mit der audiovisuellen Szene wird beim Dekodierer generiert



MPEG-4 Umfang



- MPEG-4 definiert
 - ein „System Decoder Model“
 - Spezifikation der Dekodiererimplementierung
 - eine Beschreibungssprache
 - Binärsyntax der AVO Bitstromrepräsentation
 - Informationen über die Szene
 - Konzepte, Werkzeuge und Algorithmen für
 - Inhaltsbasierte Kompression von AVOs und zusammengesetzten AVOs
 - Manipulation von Objekten
 - Übertragung von Objekten
 - Wahlfreier Zugriff auf Objekte
 - Animation
 - Skalierung
 - Robustheit gegenüber Fehlern



MPEG-4 Bitraten



- Very Low Bitrate Video Core, VLBV
 - 5 - 64 Kbit/s
 - CIF Auflösung mit bis zu 15 Frames/s
- Higher-quality Video
 - 64 Kbit/s - 4 Mbit/s
 - Qualität wie Digitales Fernsehen
- Natürliche Audiokodierung
 - 2 - 64 Kbit/s



MPEG-4 Videokodierung



- Kodierung von rechtwinkligen Bildern und Videos
 - Ähnlich wie MPEG-1/2
 - Bewegungsvorhersage
 - Texturkodierung
- Kodierung von Bildern und Videos beliebiger Form
 - 8x8 DCT oder formadaptive DCT
 - plus Kodierung der Form und von Transparenzinformation
- Kodierer
 - Generiert Timing Information
 - Die Geschwindigkeit der Uhr des Kodierers bestimmt die Zeitbasis
 - Kodiert erwartete Dekodierzeiten und –dauern durch Zeitstempel im Datenstrom
 - Spezifiziert untere Schranken der benötigten Pufferressourcen für den Dekodierer



Szenekomposition



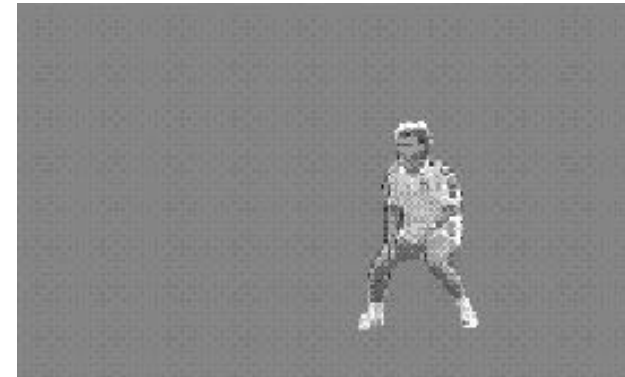
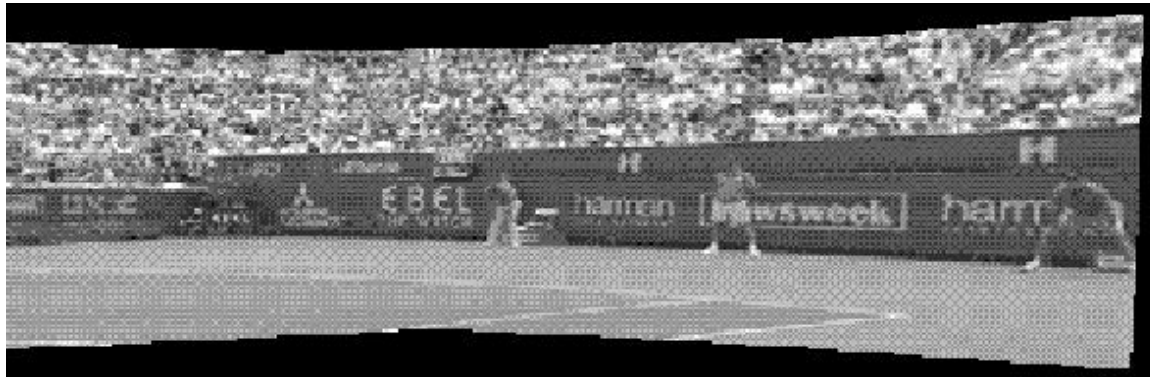
- Beschreibung enthält
 - Baum der hierarchischen Anordnung von AVOs
 - Objektpositionen in Raum und Zeit
 - Konvertierung des lokalen AVO-Koordinatensystems in globale Koordinaten
 - Wahl der Attributwerte
 - Z.B. Soundpitch, Farbe, Textur, Animationsparameter
- Beschreibung basiert auf VRML Konzepten
 - VRML = Virtual Reality Modelling Language
- Interaktionsmöglichkeiten mit der Szene
 - Z.B. Betrachterstandpunkt wechseln, Objektdragging, Start/Stop von Strömen, Auswahl der Sprache (Ton)



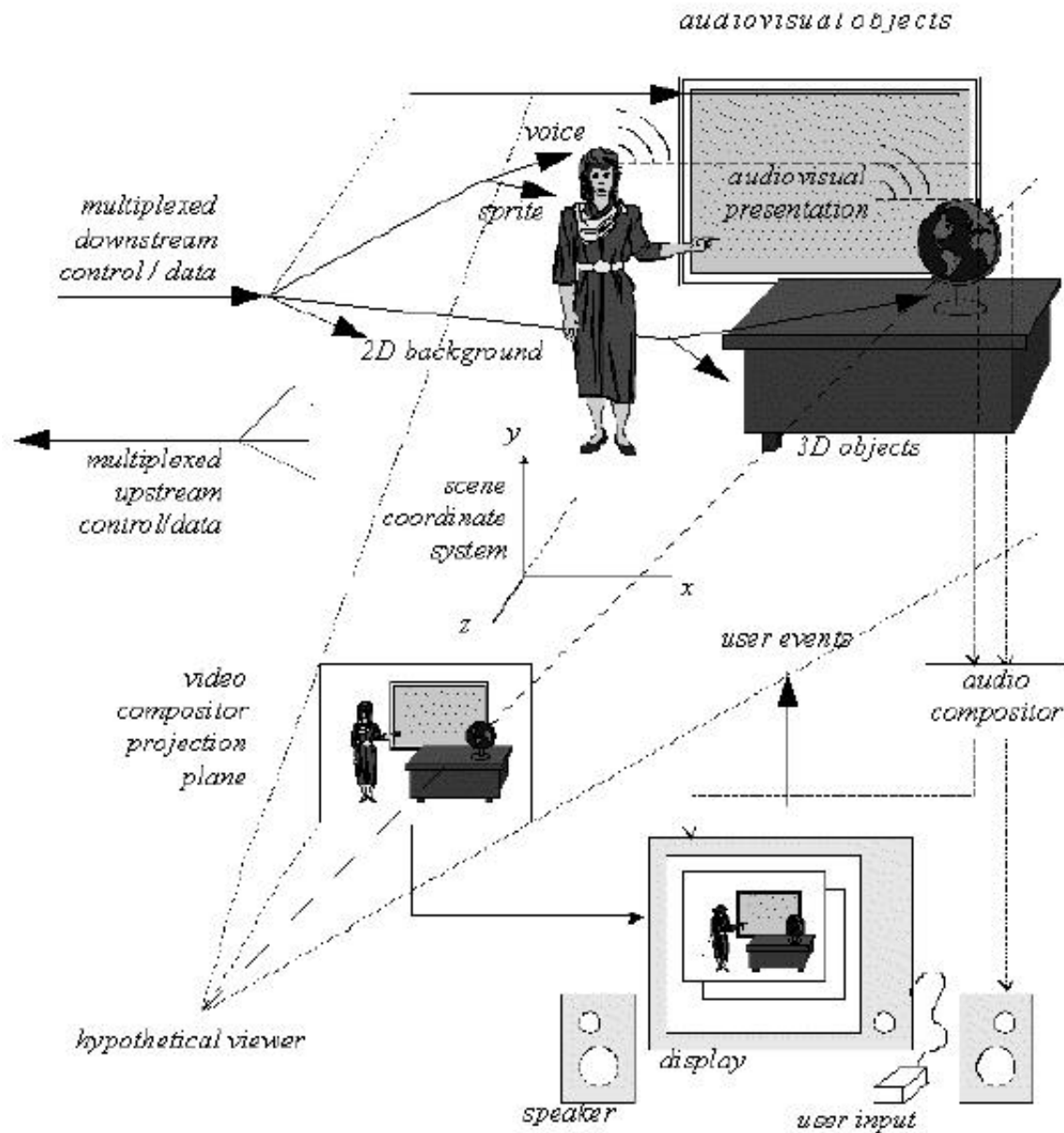
Szenekomposition



- Beispiel



MPEG-4 Szene



Skalierung



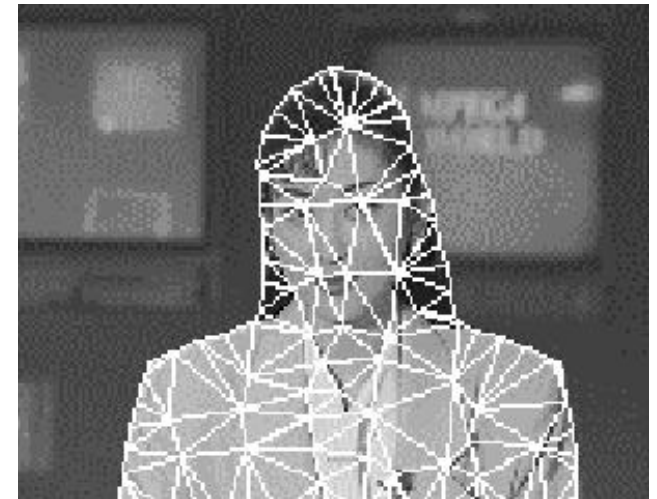
- Räumliche Skalierung
 - Dekodieren nur einer Teilmenge des verfügbaren Bitstroms zeigt Texturen und AVOs mit reduzierter räumlicher Auflösung
 - Maximal 32 solcher Ebenen sind für Texturen und Standbilder möglich
 - Maximal 3 Ebenen für Video
- Temporale Skalierung
 - Dekodierer spielt mit reduzierter zeitlicher Auflösung durch Dekodieren einer Teilmenge des verfügbaren Bitstroms
 - Maximal 3 Ebenen
- Qualitätsskalierung
 - Bitstrom wird durch Parsing in Ströme unterschiedlicher Bitraten aufgeteilt
 - Entweder bei der Versendung oder im Dekodierer



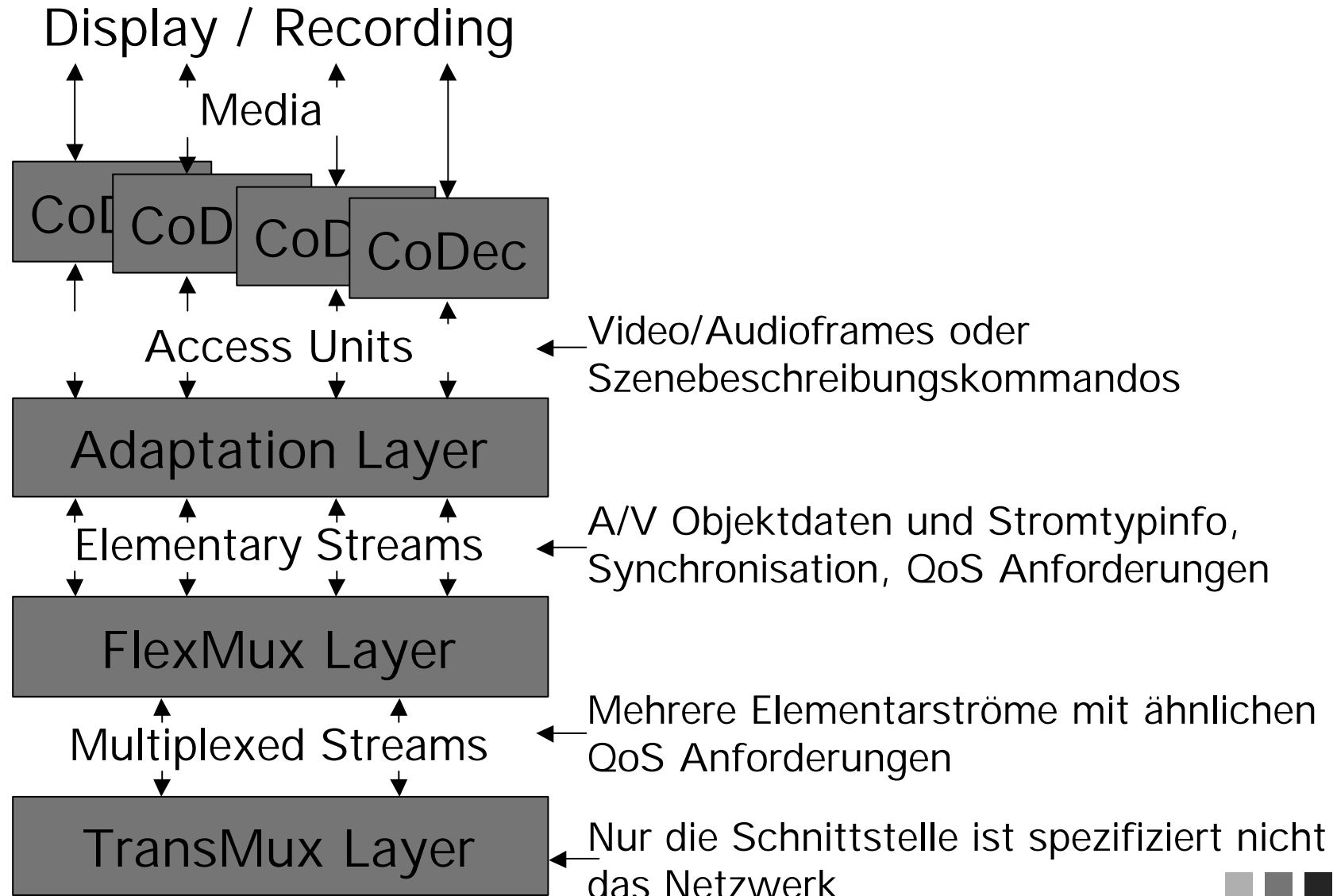
Synthetische Objekte



- Synthetische visuelle Objekte
 - Virtuelle Teile einer Szene
 - Z.B. virtueller Hintergrund
 - Einfügen von Animation
 - Z.B. animierte Gesichter
- Synthetische Audioobjekte
 - „Text-to-speech“
 - Sprachgenerierung aus gegebenem Text und prosodischen Parametern (Prosodie = Silbenlehre)
 - Kontrolle der Gesichtsanimation
 - „Score driven synthesis“
 - Generierung von Musik aus einer Szene



MPEG-4 Netzwerkaritektur



Delivery Multimedia Integration Framework, DMIF



- Sessions mit mehreren Teilnehmern
 - Interaktion mit
 - Entfernten interaktiven Partnern
 - Braodcast Systeme
 - Speichersysteme
 - remote interactive peers
 - Aufbau von Kanälen mit unterschiedlichen QoS und Bandbreiten
- DMIF kontrolliert die FlexMux und TransMux Layer



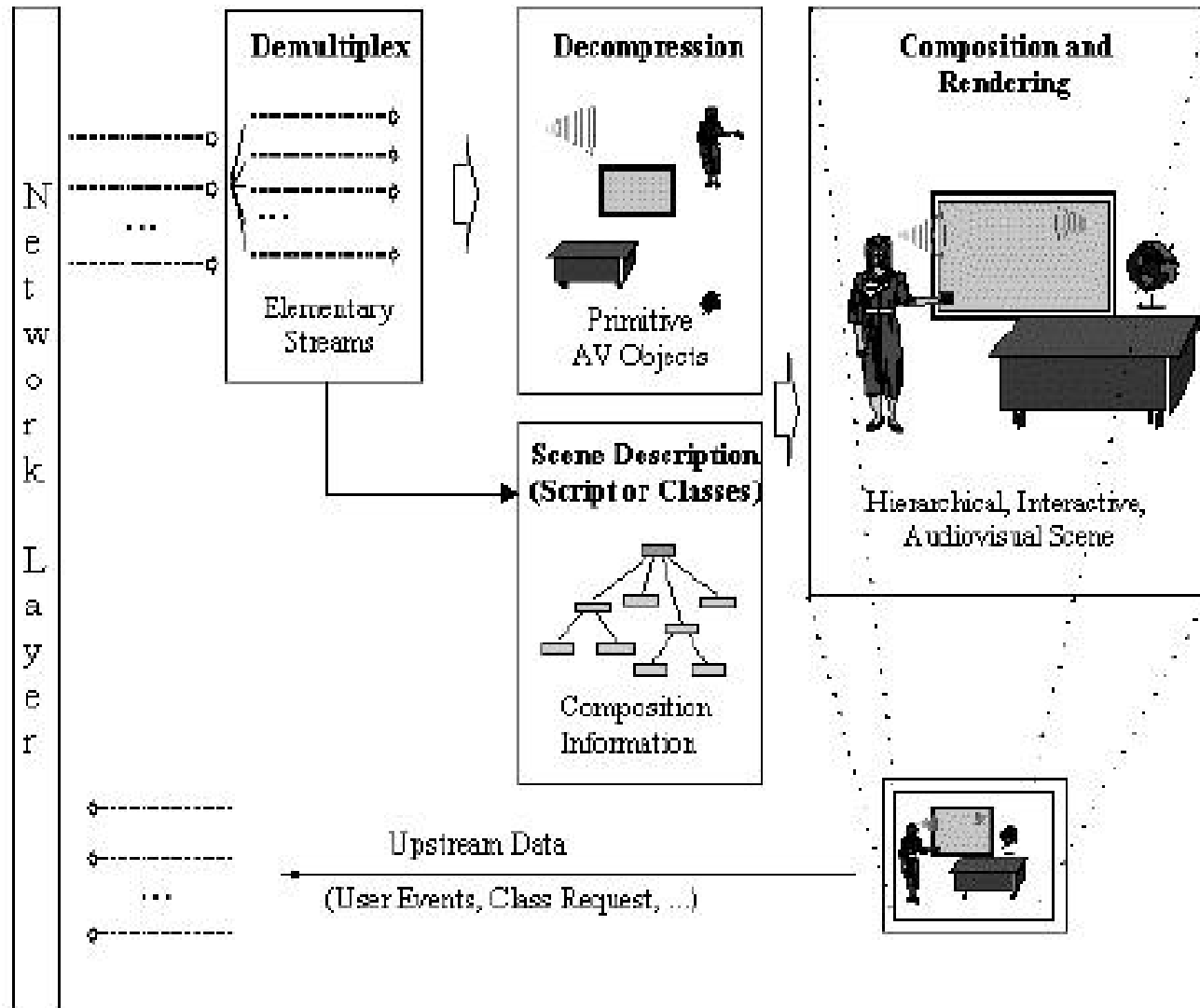
MPEG-4 Fehlerbehandlung



- Mobilkommunikation hat niedrige Bitraten und ist fehleranfällig
- Deswegen beinhaltet MPEG-4 Konzepte zur Fehlerbehandlung
 - Resynchronisierung
 - durch Marker im Bitstrom
 - Data recovery
 - Empfänger kann verlorene Daten rekonstruieren
 - Datenkodierung erfolgt fehlerrobust
 - Error concealment
 - Empfänger kann Lücken in den Daten überbrücken, z.B. durch Wiederholen alter Frames



MPEG-4 Player



MPEG-7



- Quelle: José M. Martínez, Overview of the MPEG-7 Standard (version 6.0), Dezember 2001
- ISO/IEC 15398, Herbst 2001
- Metadaten zur Ergänzung anderer MPEG-kodierter Quellen als eigene Spur
- Ziele
 - Verbesserung von Suchverfahren und Anzeigestrategien
 - Unterstützung von Konsistenzüberprüfung
 - Priorisierungsinformationen
 - Schnittstellen zu Werkzeugen zur Inhaltsanalyse



MPEG-7 Elemente



- Descriptor
 - Definition der Syntax und Semantik einer Featurerepräsentation
- Description Scheme, DS
 - Struktur und Semantik der Beziehungen zwischen Komponenten
 - Komponenten sind Descriptors bzw. Description Schemes
- Description Definition Language, DDL
 - Sprache zur Definition von Description Schemes



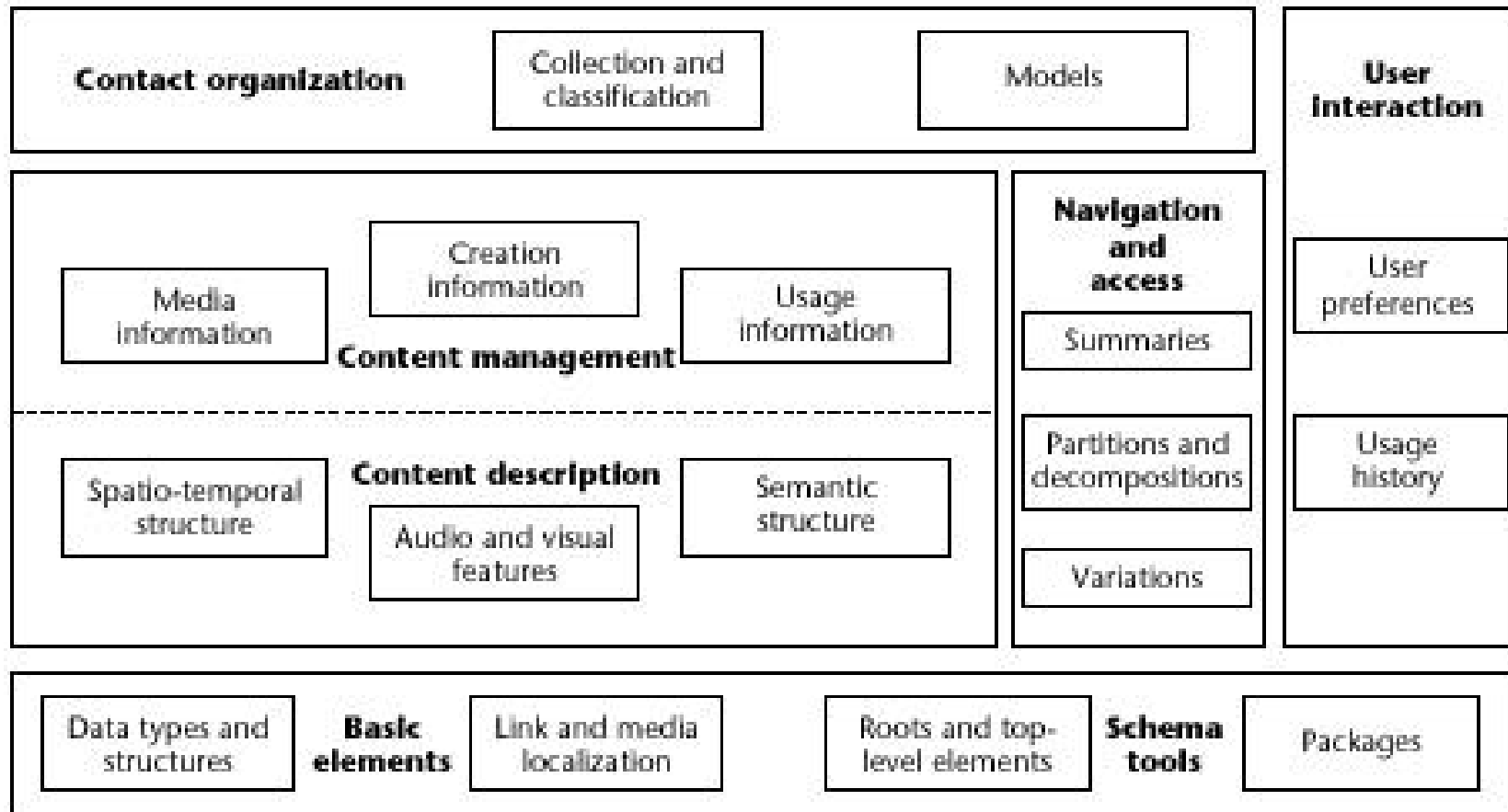
MPEG-7 Elemente



- Description
 - Instanz, d.h. tatsächliche Beschreibung als Menge von Instanzen einzelner Description Schemes
- System Werkzeuge
 - Multiplexing von Beschreibungen
 - Synchronisation von Beschreibung und Inhalt
 - Speicherungs- und Übertragungswerkzeuge
 - Management und Schutz von Intellectual Property



MPEG-7 Beschreibungswerkzeuge



MPEG-7 Beispiel



```
< Mpeg7 xmlns= "http://www.mpeg7.org/2001/MPEG-7_Schema "xml:lang= "en " type= "complete " >
  < ContentDescription xsi:type= "ContentEntityType " >
    < MultimediaContent xsi:type= "ImageType " >
      < Image >
        < MediaLocator >
          < MediaUri > http://www.tilab.org/mpeg/mpeg_logo-anim_1.gif < /MediaUri >
        < /MediaLocator >
        < CreationInformation >
          < Creation >
            < Title xml:lang= "en " > The animated MPEG Logo < /Title >
            < Creator >
              < Role href= "urn:mpeg:mpeg7:cs:RoleCS:AUTHOR " >
                < Name xml:lang= "en " > Author < /Name >
              < /Role >
              < Agent xsi:type= "OrganizationType " > < Name > MPEG < /Name > < /Agent >
            < /Creator >
          < /Creation >
          < RelatedMaterial >
            < MediaLocator > < MediaUri > http://www.tilab.com/mpeg/ < /MediaUri > < /MediaLocator >
          < /RelatedMaterial >
        < /CreationInformation >
      < /Image >
    < /MultimediaContent >
  < /ContentDescription >
< /Mpeg7 >
```



MPEG Audio



- Kompression auf
 - Bei MPEG-1: 32, 64, 96, 128 oder 192 KBit/s
 - Bei MPEG-2 auch 256 – 448 KBit/s
- Audiokanäle
 - Mono oder 2 unabhängige Kanäle oder Joint Stereo
 - Bei MPEG-2
 - Bis zu 5 Kanäle voller Bandbreite (Surround Sound)
 - plus 7 multilinguale Kommentarkanäle
- Sampling
 - Abtastrate 32 kHz, 44,1 kHz oder 48 kHz
 - 16 Bits pro Abtastwert
 - Verzögerung durch Kodieren und Dekodieren höchstens 80 ms bei 128 kbit/s
 - psychoakustisches Modell steuert die Quantisierung



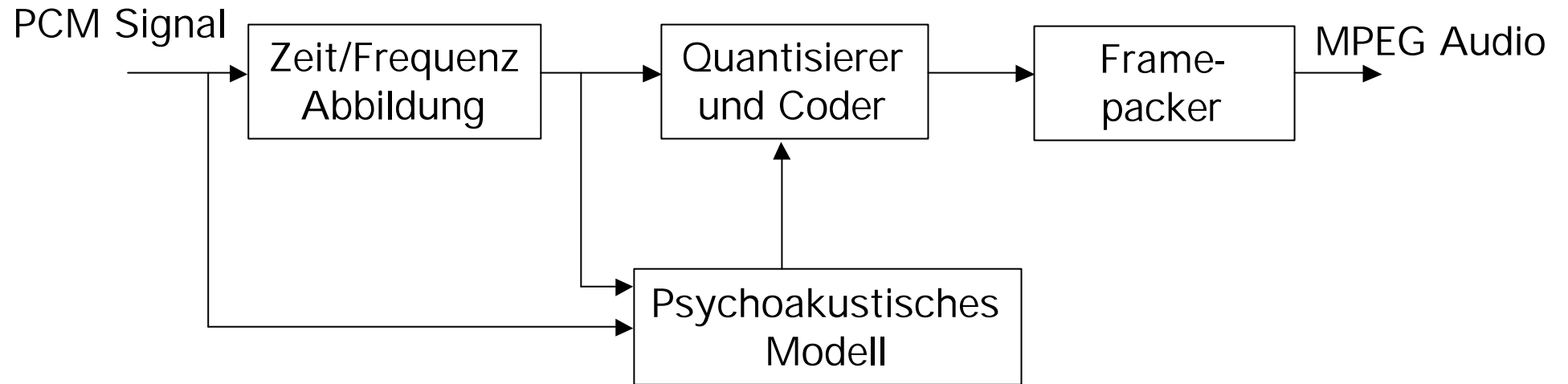
Verfahren



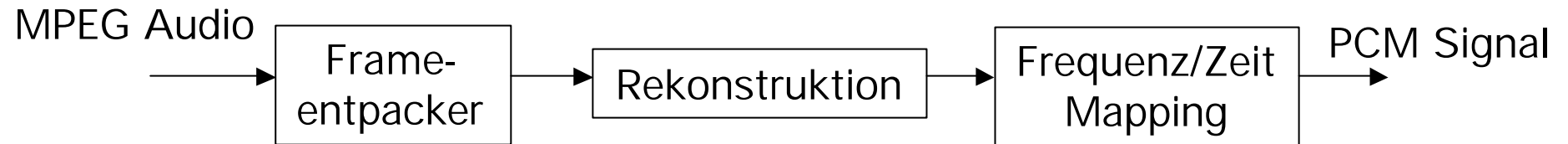
- MUSICAM
 - Masking Pattern Universal Subband Integrated Coding And Multiplexing
 - Institut für Rundfunktechnik, München
 - Basiert auf Teilbandkodierung
 - einfacher Aufbau
- ASPEC (MP3)
 - Advanced Spectral Entropy Coding
 - FhG Erlangen
 - Basiert auf Teilbandkodierung mit dynamisch überlappenden Frequenzbändern
 - sehr gute Ergebnisse bei niedrigen Bitraten



MPEG Audio Kodierer



MPEG Audio Dekodierer



MPEG Audio Layers



- Layer I
 - Teilbandkodierung mit 32 Bändern nach MUSICAM
 - hohe Datenrate
 - mono, stereo, 48 kHz, 44.1 kHz, 32 kHz
- Layer II
 - Teilbandkodierung nach MUSICAM, aufwendigeres psycho-akustisches Modell, bessere Bestimmung der Skalierungsfaktoren
 - mittlere Datenrate
 - bessere Klangqualität bei niedrigen Bitraten
 - multikanalfähig



MPEG Audio Layers



- Layer III
 - Transformationskodierung mit max. 512 dynamischen Fenstern und Entropie-Kodierung nach ASPEC (MP3)
 - niedrigste Datenrate
 - Stereo Audio in CD-Qualität: 128 KBit/s (Verhältnis 12:1)
 - Mono Audio in Telefonqualität: 8 KBit/s



H.261 - $p * 64\text{kb/s}$



- Recommendation der ITU
- Videokodierung für AV-Dienste bei Bitübertragungsraten von $p * 64\text{kb/s}$
- Ursprünglich gedacht für ISDN
 - p aus $\{1, \dots, 30\}$
 - $p = 1, p = 2$ (1 bzw. 2 ISDN Kanäle): Bildtelefon (schlechte Qualität)
 - $p = 6$: Videokonferenz (mittlere Qualität)



H.261 - p*64kbit/s



- Intraframe-Coding
 - DCT-basiert (wie JPEG)
- Interframe-Coding
 - ähnlich P-Frames in MPEG-1



H.261 Verfahren



- Grundideen sind gleich zu JPEG und MPEG
- H.261 ist weniger flexibel, aber leichter standardkonform zu implementieren
- Kodiert werden Y , C_b und C_r ,
in 8×8 Pixel-Blöcken
 - Luminanz mit voller Auflösung
 - Chrominanz jeweils mit halber Auflösung in
horizontaler und vertikaler Richtung
 - 4:2:0-Modell
- Bildformate
 - CIF (Common Intermediate Format) 352x288
 - QCIF (Quarter CIF) 176x144



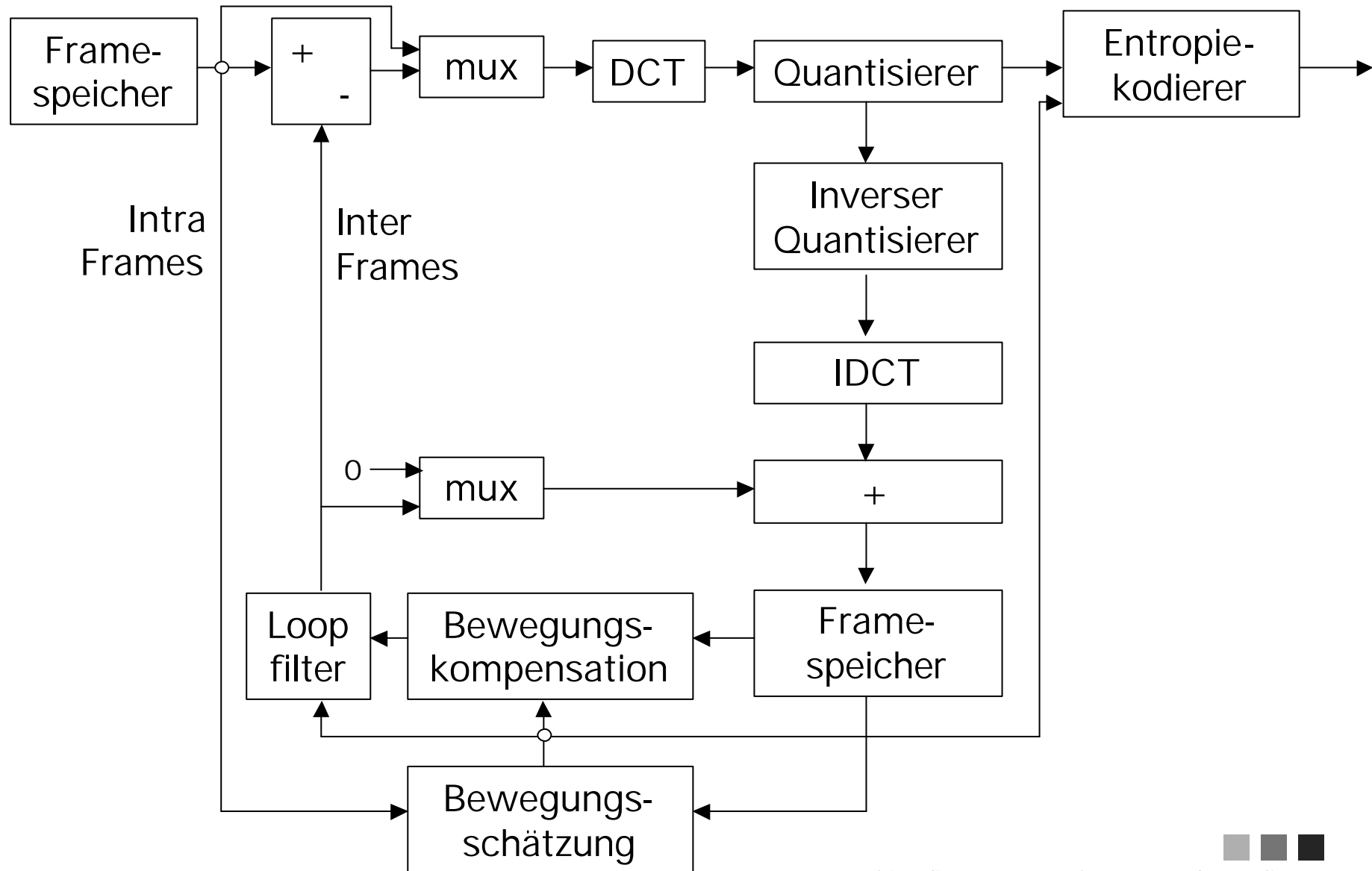
H.261 Strukturelemente



- picture ein Videobild (frame)
- group of blocks 33 Makroblöcke
- macro block 16 x 16 Y, 8 x 8 Cb, Cr
- block 8 x 8 Pixel
(Kodiereinheit für DCT)



H.261 Kodierer



H.263



- Weiterentwicklung von H.261
 - Einbeziehung von Erfahrungen aus MPEG
- Unterschiede zu H.261
 - fünf Bildgrößen
 - bidirektionale Interpolation
 - Genau ein B-Frame wird immer nach einem P-Frame platziert
 - verhandelbare Optionen zur Optimierung für bestimmte Anwendungen
 - z.B. arithmetische Kodierung in der Entropie-Kodierung statt Lauflängen- und Huffman-Kodierung

