



Grundlagen der Multimediaetechnik

Videokompression

10.12.2021, Prof. Dr. Enkelejda Kasneci

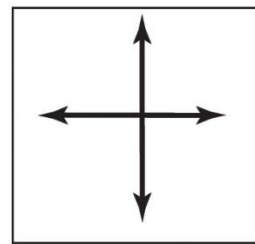


Termine und Themen

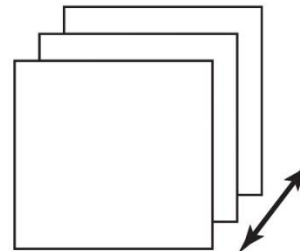
22.10.2021	Einführung
29.10.2021	Menschliche Wahrnehmung – visuell, akustisch, haptisch, ...
05.11.2021	Informationstheorie, Textcodierung und -komprimierung
12.11.2021	Bildverbesserung
19.11.2021	Bildanalyse
26.11.2021	Grundlagen der Signalverarbeitung
03.12.2021	Bildkomprimierung
10.12.2021	Bildkomprimierung
17.12.2022	Videokomprimierung Teil I
14.01.2022	Videokomprimierung Teil 2
21.01.2022	Videoanalyse
28.01.2022	Dynamic Time Warping
04.02.2022	Gestenanalyse
11.02.2022	FAQ mit den Tutoren
17.02.2022	Klausur, 14-16 Uhr, N10+N11



- **Videodaten haben vier Dimensionen:**
 - 2 Bilddimensionen
 - Eigenschaften der Pixel (Helligkeit, Farbe)
 - Zeitachse
- **Kompressionsansätze:**
 - **„räumliche“ Intra-Bild-Codierung:** „Redundanz“ aus einem Bild entfernen
 - DCT, Vektorquantisierung, Konturbasierte Kodierung
 - **„zeitliche“ Inter-Bild-Codierung:** „Redundanz“ zwischen Bildern entfernen
 - Differenzcodierung, Bewegungskompensation



räumlich



zeitlich



- **Idee:**

- **Bild aufteilen in Blöcke**, z.B. 4 x 4 Pixel
- **Suche nach Ähnlichkeiten** zwischen den Blöcken
- Ähnliche Blöcke durch einen „**Durchschnittsblock**“ ersetzen
- **Palette für Bildblöcke**, d.h. Kodierung durch Index

- Verwendung in Codecs (Codierern/Decodierern):

- Indeo, Cinepak
- Langsame Codierung (Spezial-Hardware)
- Schnelle Decodierung
- In Kompression und Bildqualität nicht besser als DCT und DWT

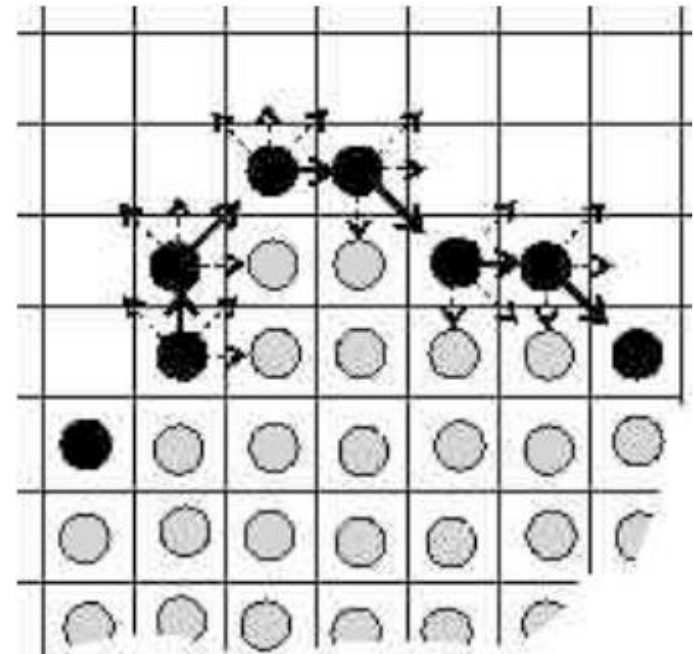


- **Idee:**

- **Bild trennen** in **Konturen** und **Texturen**
- **Konturen** z.B. durch Beziér-Kurven **beschreiben**
- **Texturen** z.B. **DCT kodieren**

- **Verwendung:**

- Ansatzweise in MPEG-4
- Vermeidet Darstellungsprobleme an Kanten
- Problem: Finden der Konturen in gegebenem Bild
 - Forschungsthema





Konzept: Differenzkodierung (frame differencing)

- Aufeinander folgende Bilder unterscheiden sich oft nur wenig
- **Idee:**
 - **Startbild** (und weitere Key-Frames) **intracodiert übertragen**
 - **Differenz** zum **nächsten Bild** als Bild auffassen und **komprimieren**
 - Z.B. mit DCT und anschließender Entropiecodierung
 - Viele niedrige Werte, also hoher Kompressionsfaktor möglich



Bildsequenz



Differenzbilder



Bewegungsanalyse: Was hat sich verändert?

a



b



c



d



Bewegungsanalyse: Differenzbild

Differenzbild von a und b



Differenzbild von c und d





Konzept: Bewegungskompensation (motion compensation)

- **Idee:**

- **Bewegungen** von **Objekten** zwischen Bildern **identifizieren**
- Für **Teilbilder** übertragen:
 - **Differenzbild** plus
 - **Verschiebungsvektor**
- Verwendung u.a.:
 - MPEG-1, -2 und -4, H.261-H.264

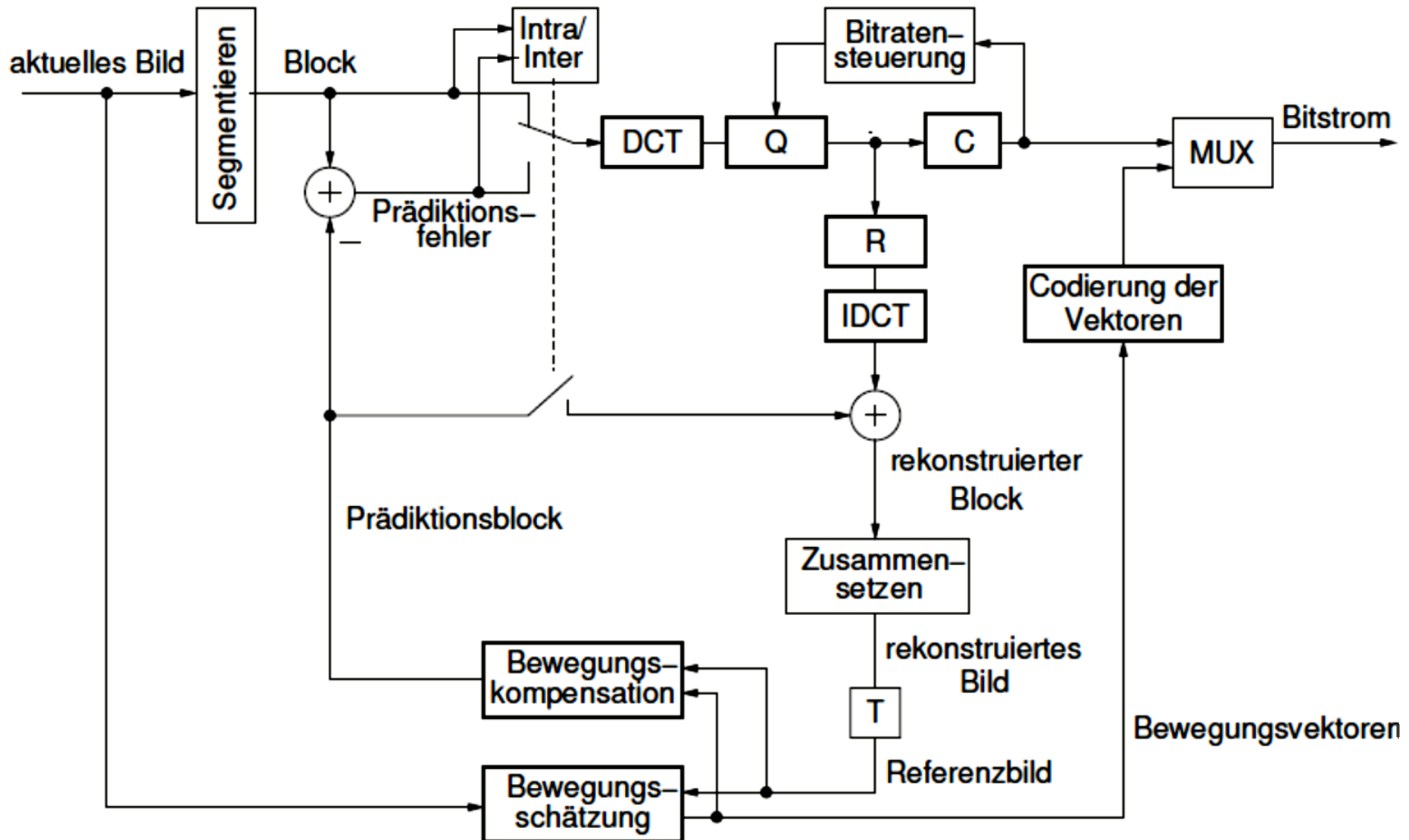
- **Problem:** Algorithmen zur Bewegungsschätzung

(motion estimation)

- Block-Matching
- Gradient-Matching
- Phase-Correlation



Aufbau eines Video-Codierers



Bildquelle: Tilo Strutz: Bilddatenkompression; Vieweg+Teubner, 2009.



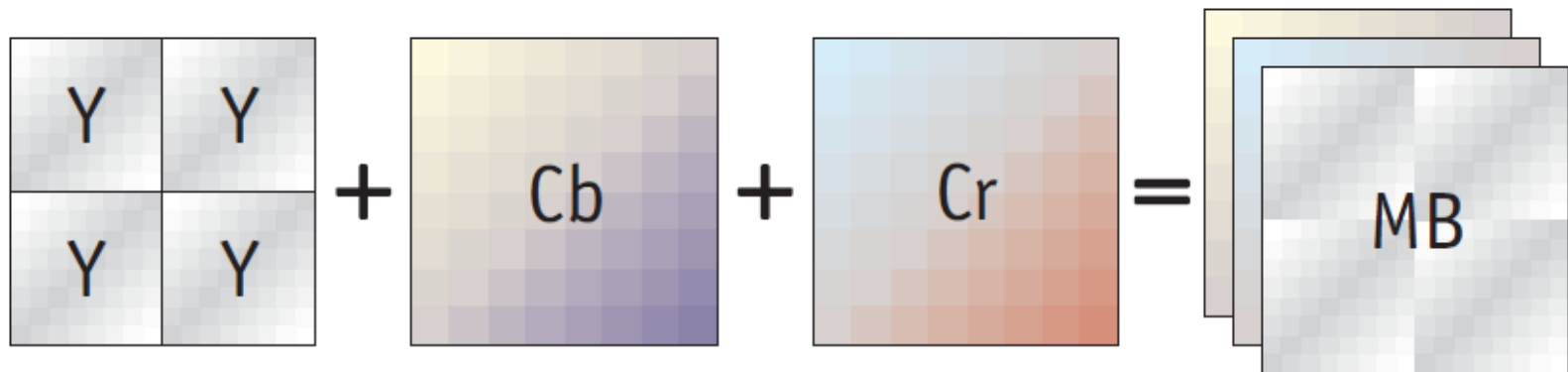
Vorgehensweise eines Video-Codierers

- Aktuelles Bild wird **blockweise DCT-transformiert** und **quantisiert (Q)**
 - **Kodierung** der **quantisierten Transformationskoeffizienten (C)**
 - **Bitratensteuerung** überwacht Datenaufkommen und **beeinflusst** durch **Verändern der Quantisierungsstärke** die **Datenrate**
 - **Ähnlichkeitsdetektion** aufeinanderfolgender Bilder (Inter-Frames)
 - Quantisierte Koeffizienten werden **rekonstruiert (R)** und **rücktransformiert (IDCT)**
 - **Rücktransformiertes Bild wird als Referenz** für einen oder mehrere Zeitschritte (T) gespeichert aufgehoben und steht als Referenz für später zu codierende Bilder zur Verfügung.
 - **Bewegungsschätzung** vergleicht neues Bild mit Referenzbild
→ **Menge von Bewegungsvektoren**
 - **Bewegungskompensation erzeugt** mittels Bewegungsvektoren aus dem Referenzbild blockweise ein **Prädiktionsbild**
 - **Differenz** zwischen neuem Bild und vorausgesagten Bild ergibt einen **Prädiktionsfehler**
 - **Intra/Inter-Modul** entscheidet, ob ein Block als **Intra-Block** ohne Prädiktion oder als **Inter-Block** mit Prädiktionsfehler **codiert** wird
-



Makroblöcke

- DCT arbeitet nur mit Graustufen
 - alle drei Farbkanäle werden getrennt DCT-codiert
- Chrominanzblöcke erscheinen im Ausgabebild wegen Subsampling in 16x16 Pixel Größe
- Zusammenfassung von 4 Y, 1 Cb und 1 Cr-Block zu einem
- Grundeinheit des Bilddatenstroms sind die Makroblöcke
- **Aufbau Makroblock:**

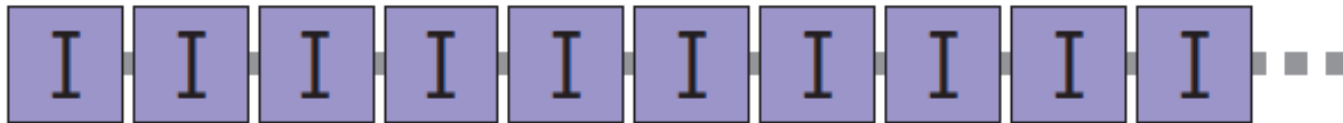




Frametypen: I-Frames

- **Videodarstellung** durch **einfaches Aneinanderreihen von Einzelbildern** (Intra-coded Frames, oder Keyframes)
 - Motion-JPEG (MJPEG), DV

- **Intra-Frames**

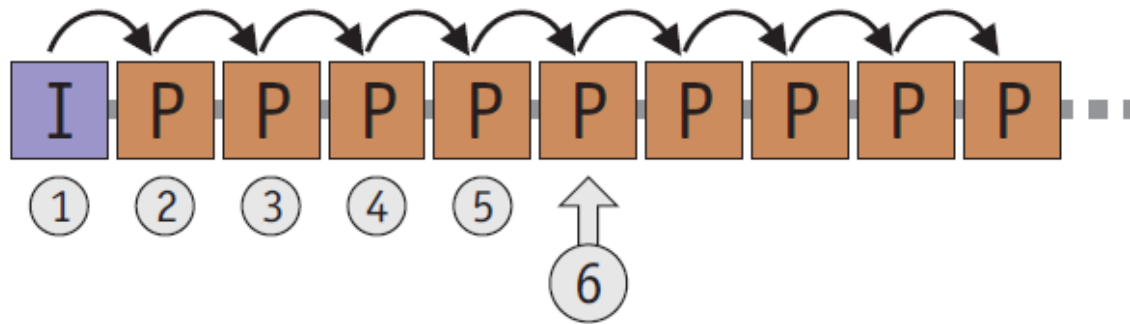


- **Vorteile:**
 - ermöglicht direkten Zugriff auf jedes Einzelbild (ideal zum Editieren)
 - kann ein Frame nicht decodiert werden, wird er einfach ausgelassen
- **Nachteil:**
 - niedrige Kodiereffizienz, da Abhängigkeiten zwischen Bildern nicht berücksichtigt werden



Frametypen: P-Frames

- Es wird zunächst ein I-Frame codiert
- danach folgen Frames, welche durch Bewegungsschätzung und Differenzbildung aus dem letzten Frame erzeugt werden
- **Prädiktiv-kodierte Frames (P-Frames)**



- **Vorteil:**
 - höhere Kodiereffizienz
- **Nachteil:**
 - soll ein bestimmter Frame dekodiert werden, so müssen zunächst alle vorherigen Frames dekodiert werden



Frametypen: P-Frames

- **Weiterer Nachteil:**

- kann ein P-Frame nicht dekodiert werden, so kann auch der Rest des Videos nicht dekodiert werden



- (Teil-)Lösung für die Probleme:

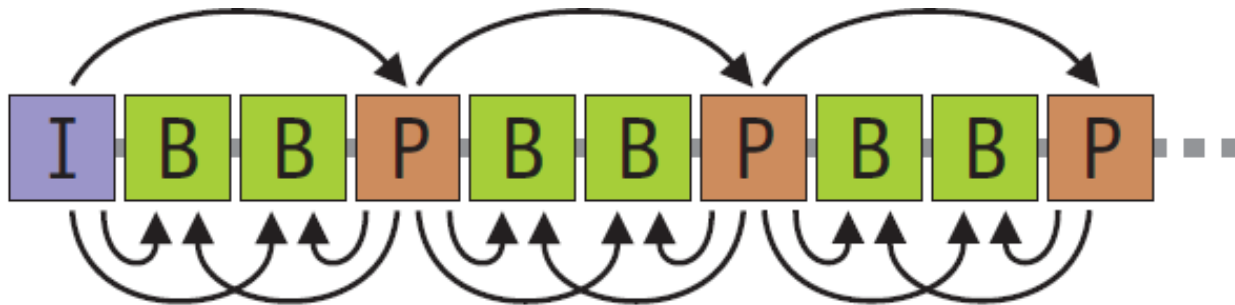
- periodisches Einstreuen von I-Frames
- bei großen Perioden trotzdem nicht hilfreich

- **Vorteil überwiegt Nachteile bei weitem!**



Frametypen: B-Frames

- Zusätzlich zu I- und P-Frames existiert noch ein dritter Frame-Typ:
Bidirektionale Prädiktiv-kodierte Frames (B-Frames)
- B-Frames kodieren den **Unterschied** zum **vorigen** und **nächsten I- oder P-Frame**
 - B-Frames werden jedoch ihrerseits als Grundlage für andere Frames verwendet
 - typischerweise 2-3 B-Frames zwischen zwei P-Frames

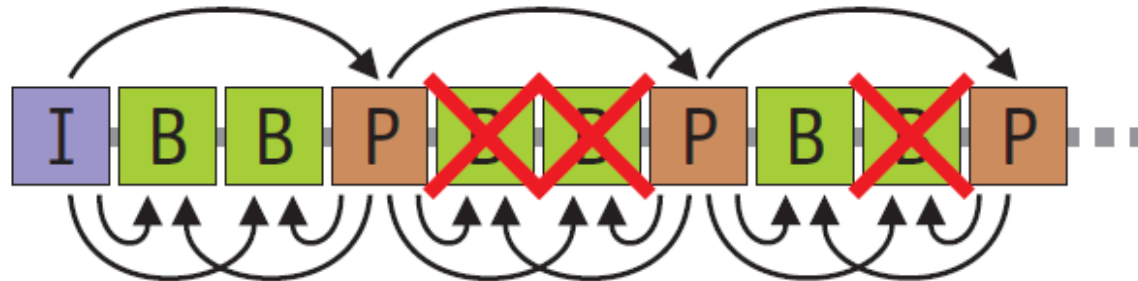




Frametypen: B-Frames

• Vorteile:

- höhere Kodiereffizienz
- schnellerer Zugriff auf beliebige Bilder
- B-Frames können bei Problemen ohne Folgen ausgelassen werden



• Nachteile:

- Kodierreihenfolge entspricht nicht mehr der Anzeigereihenfolge
- B-Frames leisten keinen Beitrag zu anderen Bildern („verschwendete Bits“)
- höhere algorithmische Komplexität, Speicherplatz und
- Bandbreitenanforderungen in Encoder und Decoder



Frametypen (Zusammenfassung)

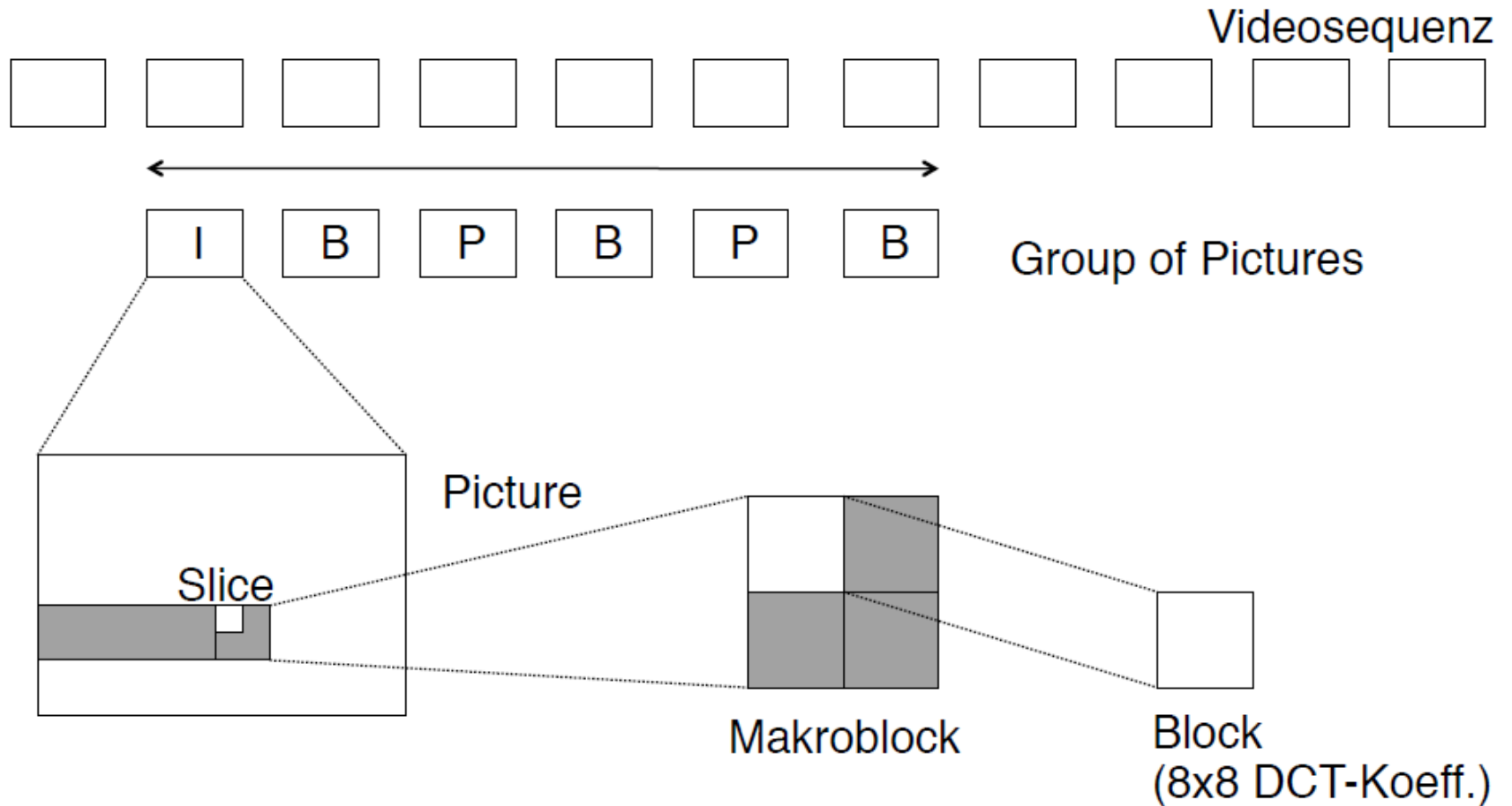
- **Intra-coded Frames** (I-Frames)
 - Entspricht JPEG-Bild, in Echtzeit dekodierbar
- **Predictive-coded Frames** (P-Frames)
 - Verweis auf vorheriger I- oder P-Frames
 - Bewegungsvektor nicht festgelegt
 - Differenz ähnlicher Makroblöcke DCT-kodiert
 - DC- und AC-Koeffizienten RLE-kodiert
- **Bidirectional predictive-coded Frames** (B-Frames)
 - Verweis auf vorherige und folgende I- und P-Frames
 - Interpolation zwischen Makroblöcken
- **DC-coded Frames** (D-Frames)
 - Nur DC-Komponenten DCT-kodiert
 - schnelles Vorwärts- und Rückwärtsspulen



- **MPEG-Sequenz** besteht aus **GOPs** (Group of Picture, Bildgruppen)
 - eine GOP ist eine Sequenz von typischerweise ca. 0,5 bis 1 Sekunde Länge
 - jede GOP hat ein I-Frame und eine Anzahl von P/B-Frames
 - typische GOP-Struktur: IBBP...
 - Encoder kann (z.B. bei Szenenwechseln) GOP auch kürzen
 - Schnitt von MPEG-Video ist nur an GOP-Grenzen möglich!
- Frames sind in **Slices** unterteilt
 - Z.B. durch **Zusammenfassung benachbarter Blöcke** mit **gleichen/ähnlichen Grauwerten**
 - erhöhen die Robustheit gegen Fehler
 - Resynchronisierung nach Dekodierfehlern



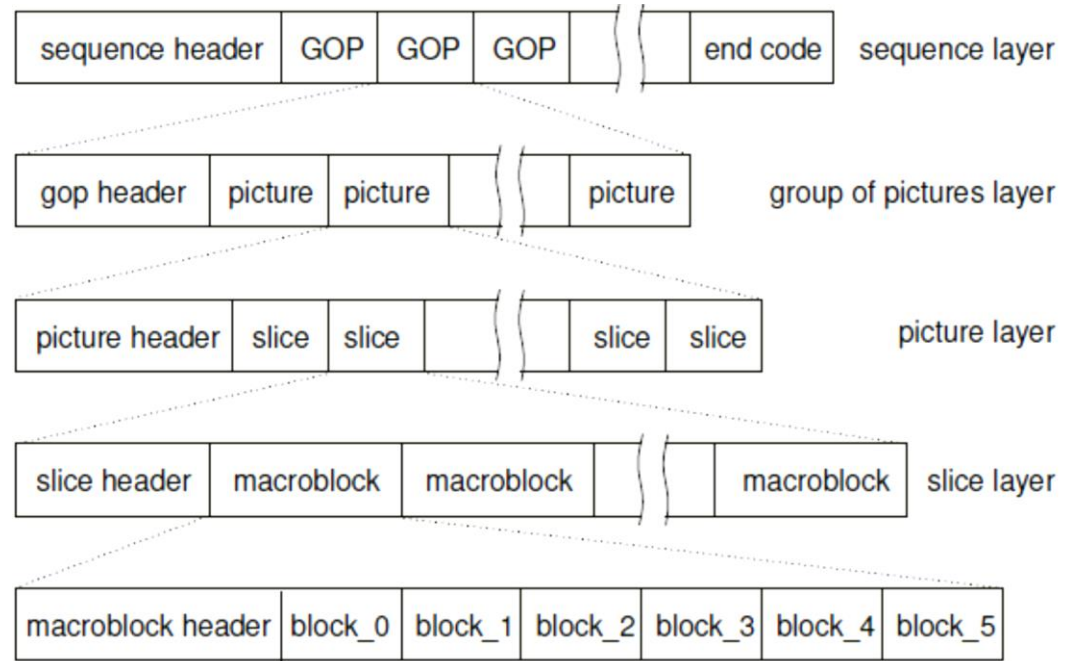
Struktur des MPEG-2 Videodatenstroms





Group of Pictures (GOP): Bildgruppe

- **Bildgruppe muss mind. 1 I-Frame enthalten**
- Bildgruppe wird oft zur Übertragung umsortiert
 - Vermeidung von Vorwärtsverweisen
 - Kleinere Zwischenspeicher



I B B P B B P B B P B B (I)
1 2 3 4 5 6 7 8 9 10 11 12 1

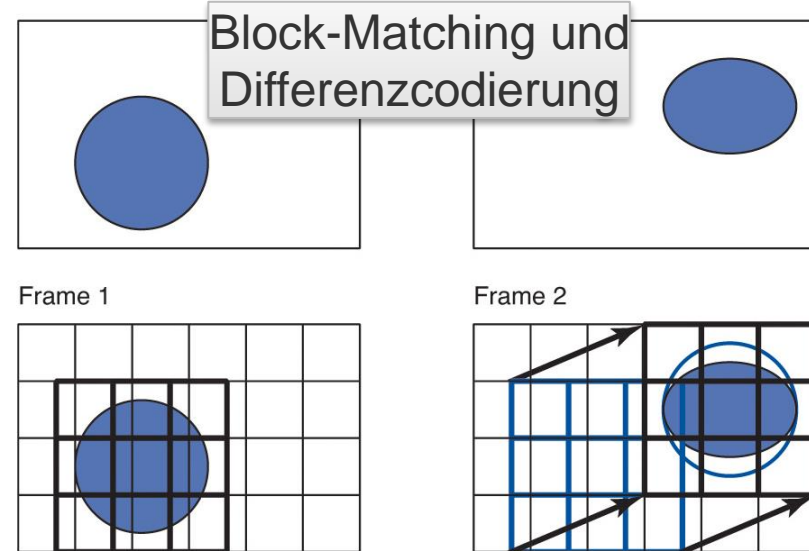
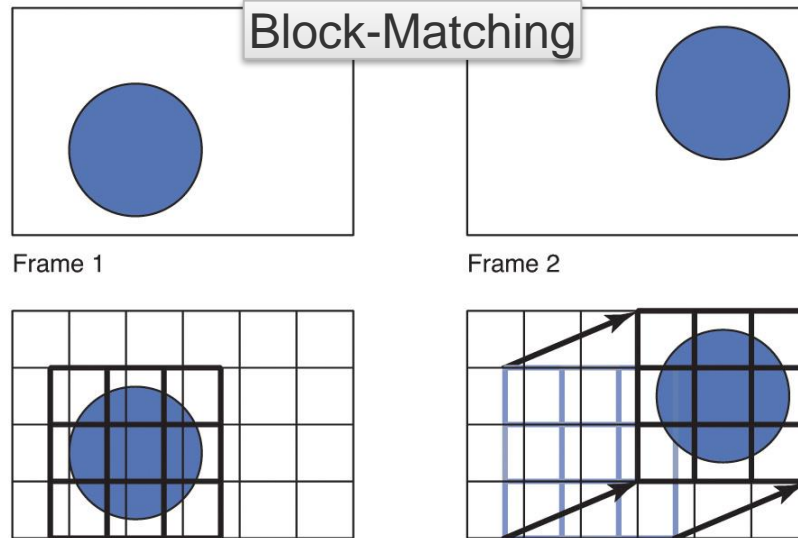
Darstellungsreihenfolge



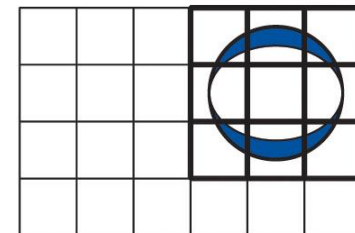
I P B B P B B P B B (I) B B
1 4 2 3 7 5 6 10 8 9 1 11 12

Übertragungsreihenfolge

Bewegungskompensation: Block-Matching



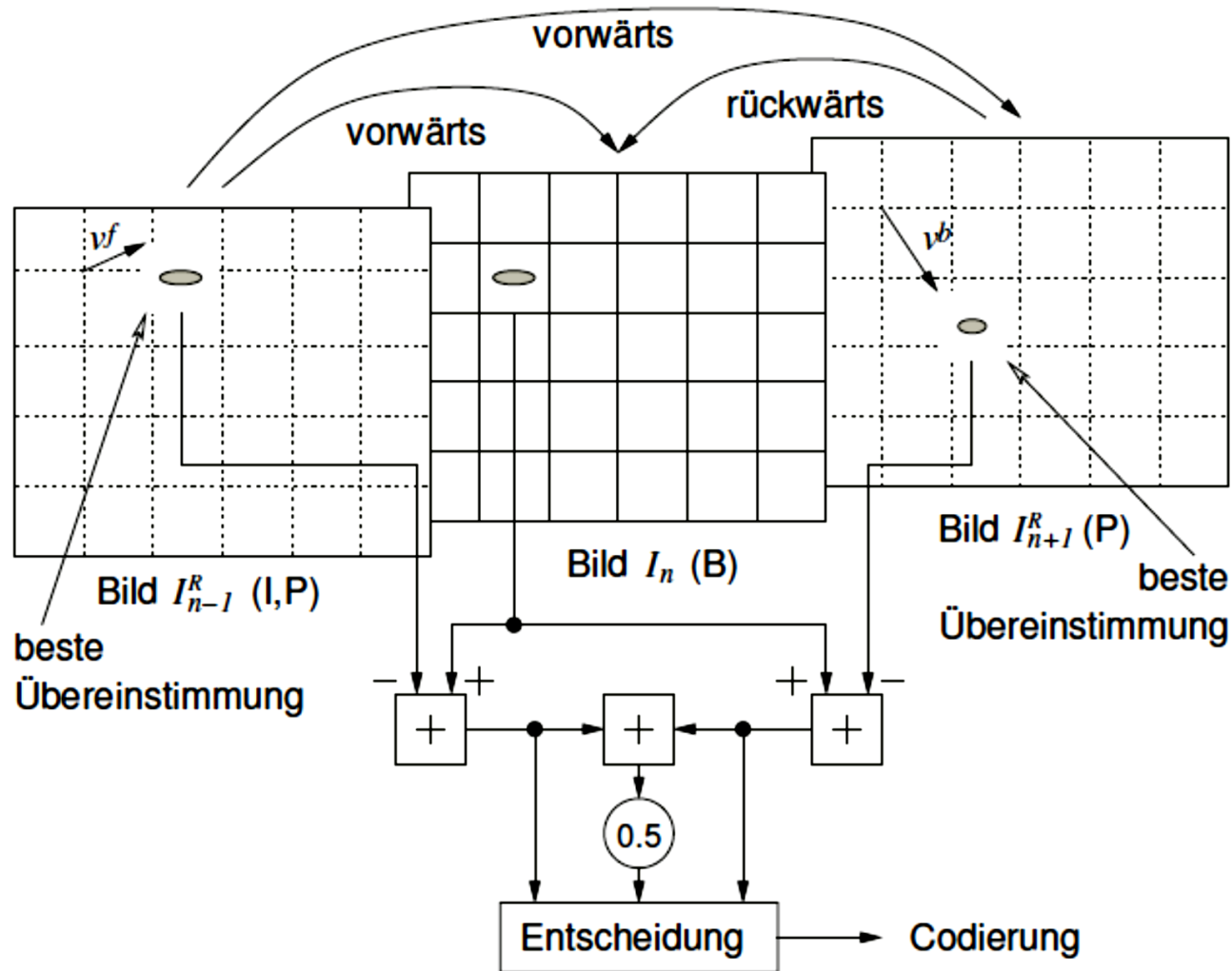
Differenz



- Referenzframe und Zielframe (aktueller Frame)
 - Referenzframe = vorheriges Bild
 - Einteilung des Bildes in Blöcke (z.B. 16x16)
 - Für jeden Block des Zielframes:
 - Suche nach „best match“ im Referenzframe
 - z.B. mittlere quadratische Abweichung oder mittlere Differenz
 - Speichern des Verschiebungsvektors in Differenzcodierung
 - Beschleunigung des Algorithmus
 - Hierarchische Suche zunächst auf vergrößertem Bild



Bidirektionale Prädiktion von Blöcken mit Block-Matching (B-Frames)

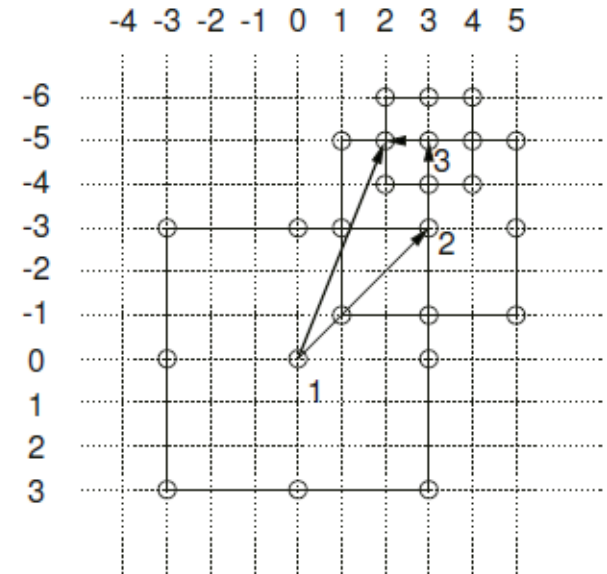


Bildquelle: Tilo Strutz: Bilddatenkompression; Vieweg+Teubner, 2009.



Suchstrategien

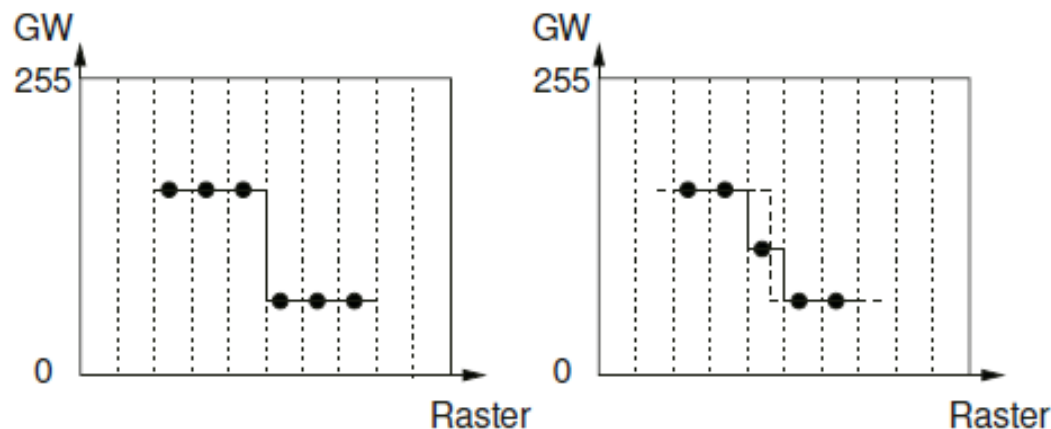
- **Suche** nach passenden **Referenzblock** ist **sehr rechenintensiv**
- Maximal zu erwartende Bewegungsdistanz
(Zeitaufwand + Flächenaufwand)
- Keine systematische Untersuchung aller Verschiebungsvarianten
- **Effizienter: Mehrschritt-Suche**
 - Untersuche 9 Positionen in
8er-Nachbarschaft
 - Verringere Schrittweite
 - Fahre mit bestem Vektor fort und
bilde neue Nachbarschaft bis
minimale Schrittweite erreicht





Suchstrategien: Subpixelschätzung

- Liegt eine Grauwertkante durch eine Verschiebung nicht mehr genau auf dem Bildpunktraster, so erscheint die Kante nicht mehr vollständig scharf und ist nicht mehr im Referenzbild zu finden
- **Idee:**
 - **Verschiebungen um halbe oder Viertelpixel**
 - **Subpixelschätzung interpoliert** zwischen den **vorhandenen Bildpunkten** wenn Verschiebung nicht mehr genau auf dem Bildpunktraster liegt





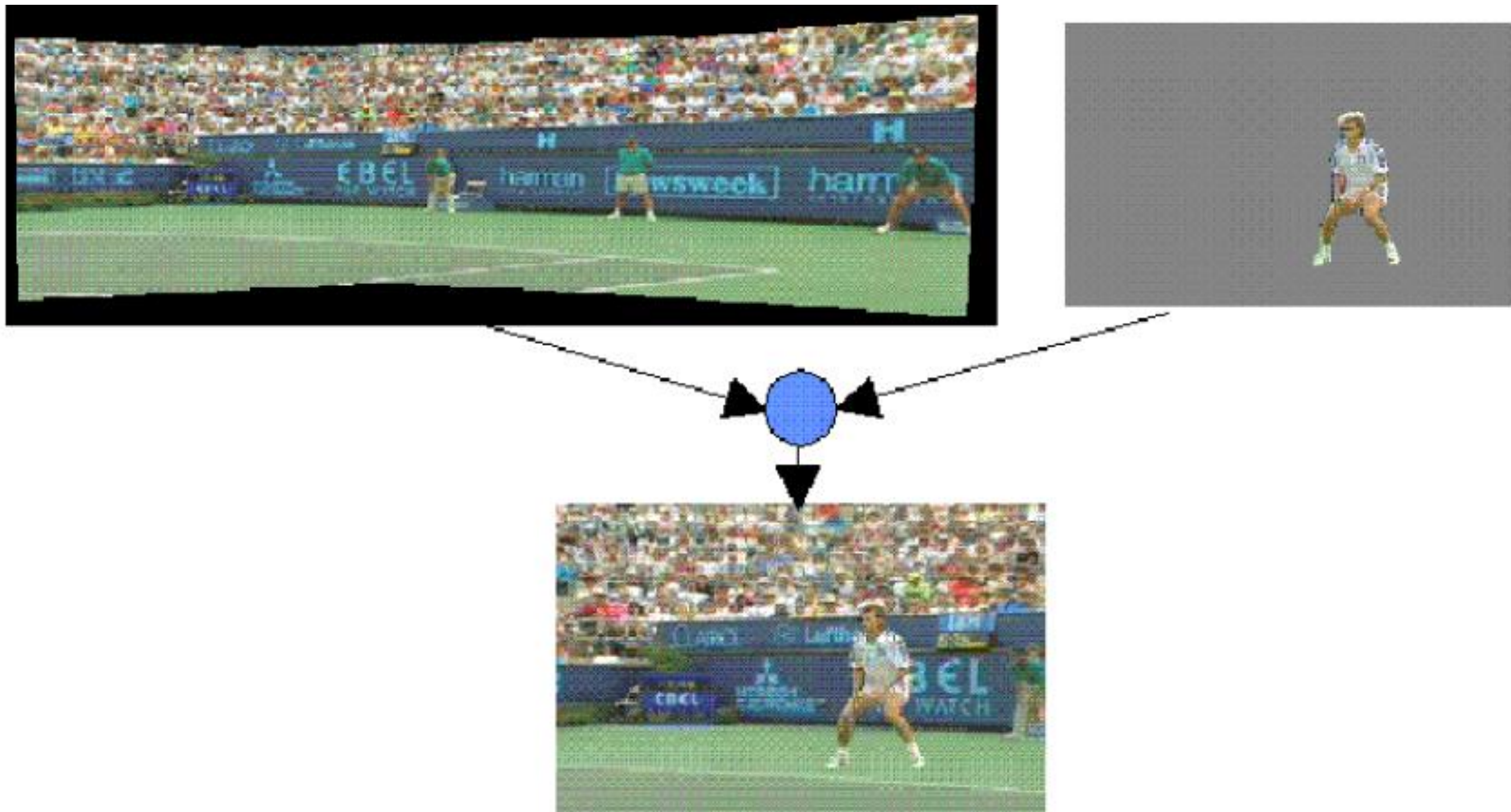
Videokompression: MPEG4

- **Fokus sowohl auf Systeme mit geringen Ressourcen als auch auf Studioanwendungen**
 - mobile Kommunikation
 - Bildtelefon und MMS
- **Datenraten und Dimensionen**
 - ca. 4,8-64 kBit/s
 - 176*144 Pixel, 10 Frames/s
 - Studioanwendungen: Auflösungen bis zu 4096x4096 Pixel
- **Verbesserung von Bildanalyse und -manipulation**
 - **Inhaltsorientierung**
 - **Kodierung und Kapselung von Einzelobjekten**
 - Nicht „rechteckige“ Video- und Bildobjekte
→ **Audiovisuelle Objekte (AVO)** basierend auf Konturinformationen und Hintergrundobjekte (z.B. Fußballspieler auf Fußballfeld)
 - Audio getrennt von Video
 - **Komposition von Objekten**
 - Erzeugung zusammengestellter Objekte
 - Resynthese der Szene
 - **Multiplexing und Synchronisation von AVOs**
 - **Interaktion** mit auf Empfängerseite generierten Szene



Videokompression: MPEG4 – Objektkodierung

- **Konturabhängige Videoobjekte** mittels Alphamasken (Konturinfo.)
- **Sprite Coding** / Arbitrary Shape Coding



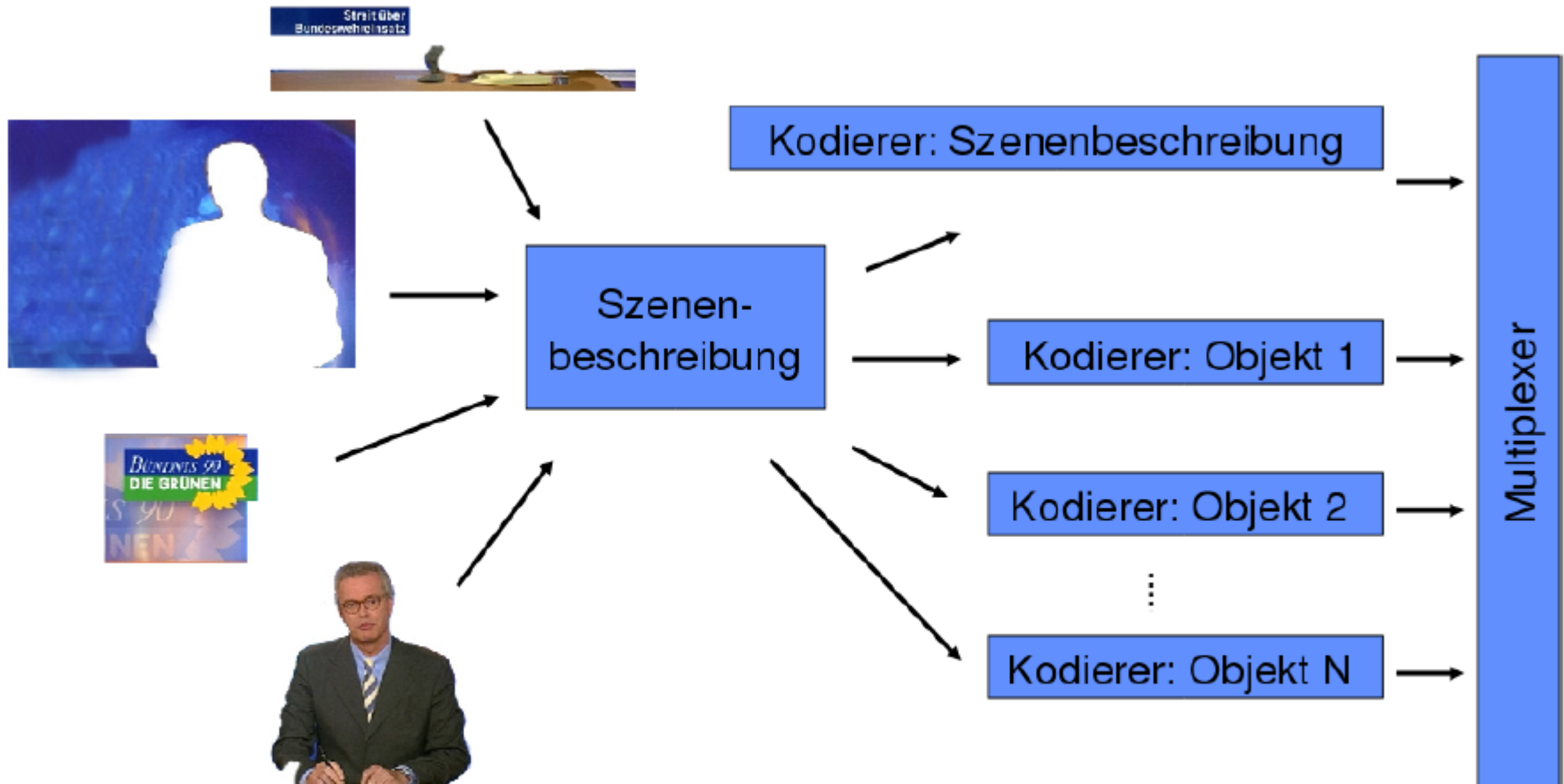


MPEG-4: Komposition von Szenen

- Beschreibung **hierarchischer Beziehungen zwischen AVOs** durch Baum
- **Position der Objekte** in Raum und Zeit
 - Konvertierung von lokalen in globale Koordinatensysteme
- **Attributwerte** wie Tonhöhe, Farbe, Textur, Animationsparameter
- Beschreibung auf Basis von **VRML-Konzepten**
 - Virtual Reality Modeling Language
- **Interaktion mit Szenen**, z.B.
 - Perspektivenwechsel
 - Start/Stopp von Videoströmen
 - Sprachwahl



MPEG-4: Komposition von Szenen



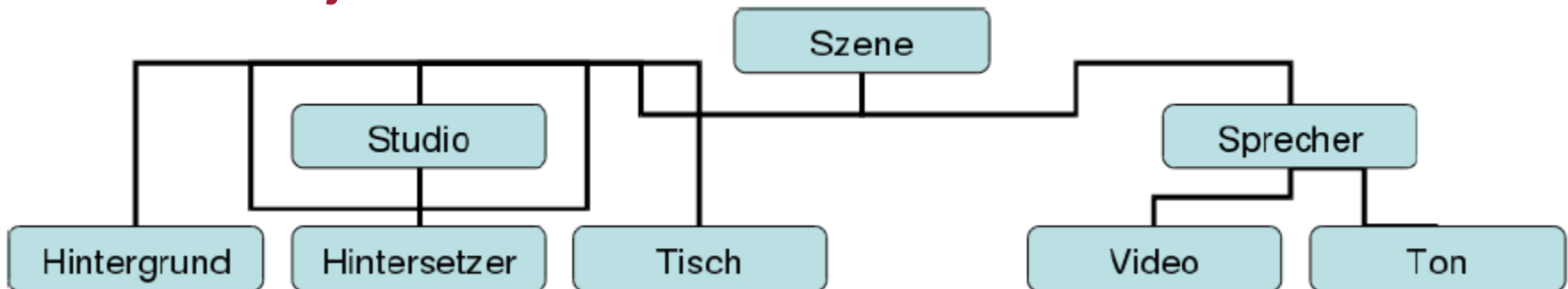
Objektkodierung und Szenensynthese



MPEG-4: Szenengraph

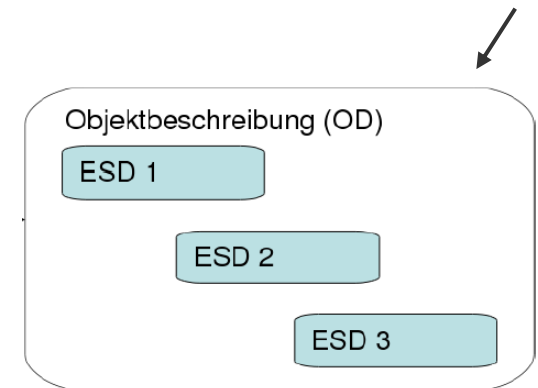
- **Binary Format for Scenes (BIFS)**

- abgeleitet aus Virtual Reality Modeling Language (VRML)
- **Spezifikation** von **räumlicher Positionierung** und **Interaktion von Objekten**



- **Objektbeschreibung**

- Deskriptoren für Elementarströme
- Kodierung von Eigenschaften, Abhängigkeiten und Prioritäten
- Spezifikation von Alternativen



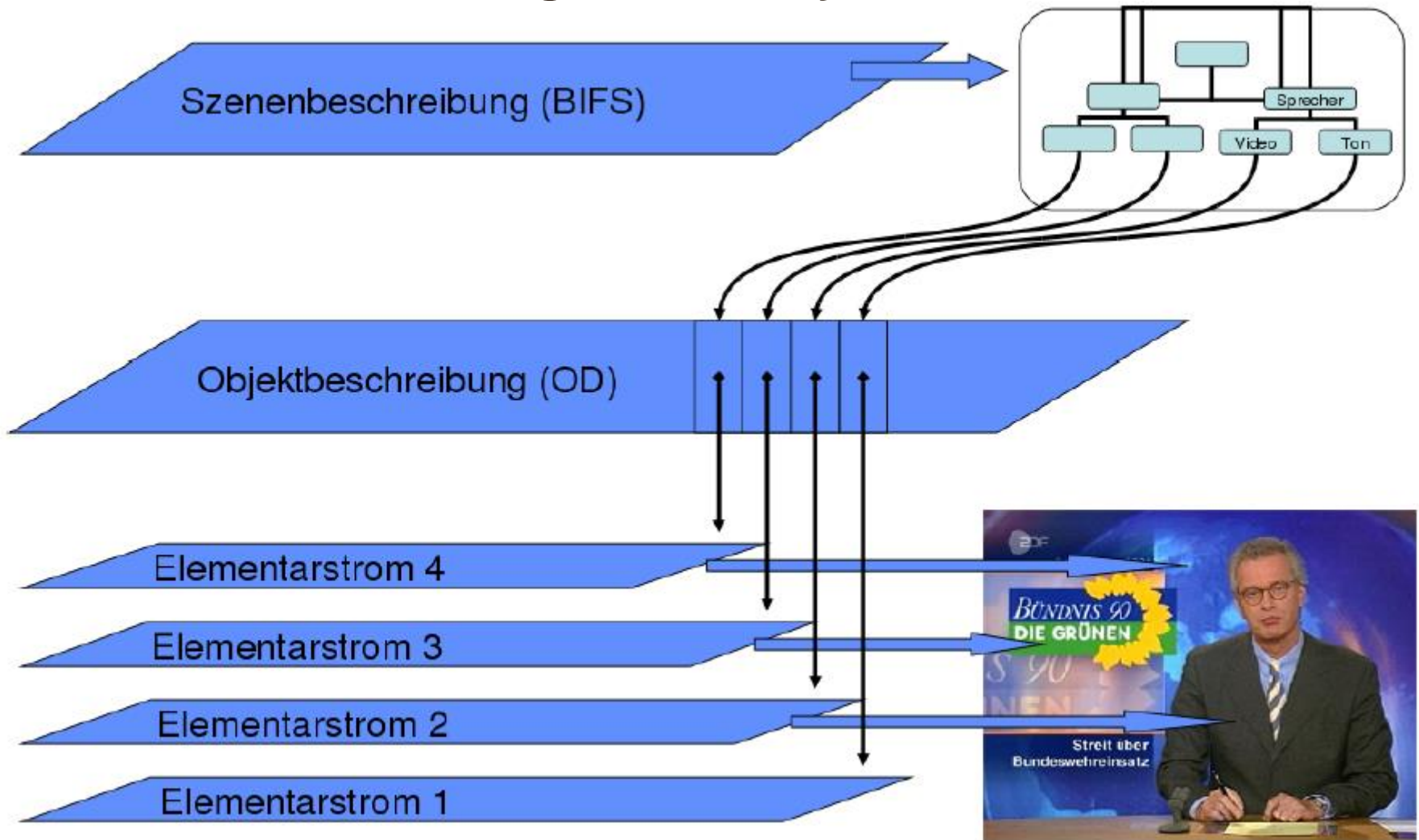


MPEG-4: Synthetische Objekte

- **Visuelle Objekte**
 - Virtuelle Teile der Szene, z.B. Hintergrund
 - Animation, z.B. Gesichter
- **Akustische Objekte**
 - Text-to-Speech, d.h. Spracherzeugung ausgehend von gegebenem
 - Text und prosodischen Parametern
 - Score-driven Synthesis, d.h. Musikerzeugung aus Partitur (allgemeiner als MIDI)
 - Spezialeffekte



MPEG-4: Vollständige Szenensynthese





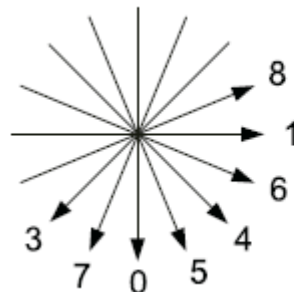
H.264 ist ähnlich zu MPEG-1/2 mit den folgenden Erweiterungen:

- **Makroblöcke können** bei Bedarf bis zu **4x4-Blöcken weiter unterteilt werden** → DCT auf 4x4-Blöcken
- **Bewegungsvektoren** mit einer **Genauigkeit** von **Viertel-Bildpunkten**
- **Bewegungsvektoren** dürfen **über Bildgrenzen** hinausragen
- **Mehrfache Referenzbilder** (I-Frames)
- **Loslösen** der **Referenzierung** von der **Reihenfolge** der Bilder
- **Gewichtete Mittelung** von Referenzblöcken bei bidirektionaler Prädiktion (B-Frames)
- **Örtliche Prädiktion** von **Blöcken**
- **Einsatz** einer **Integer-Transformation** mit 16-Bit-Arithmetik
- Kaskadierung der Transformation ähnlich dem Multiauflösungskonzept der Wavelet-Transformation
- **Logarithmische Abstufung** der **globalen Quantisierung**
- **Deblocking-Filter** in der zeitlichen Prädiktionsschleife
- **Verbesserte Entropiecodierung**
→ CABAC (Context-Adaptive Binary Arithmetic Coding)



H.264: Örtliche Prädiktion

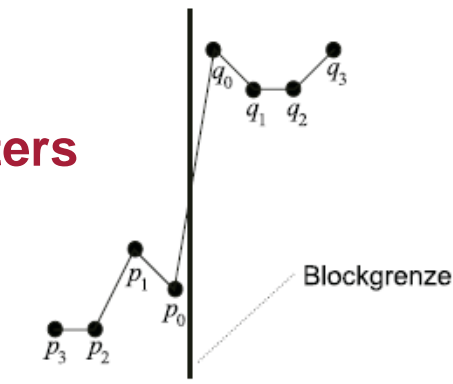
- **Prädiktion unter Zuhilfenahme benachbarter Blöcke**
- Wesentliche Eigenschaft: **Richtungsabhängigkeit**
 - Modus 0: beschreibt eine vertikale Prädiktion.
 - Modus 1: verwendet eine horizontale Voraussage
 - Modus 2: DC-Prädiktion (Mittelwert der direkten Nachbarpunkte)
 - Modus 3: (diagonal von rechts-oben nach links-unten)
 - Modus 4: (diagonal von links-oben nach rechts-unten)
 - Modus 5: (schräg vertikal nach rechts)
 - Modus 6: (schräg horizontal nach unten)
 - Modus 7: (schräg vertikal nach links)
 - Modus 8: (schräg horizontal nach oben)



	Q	A	B	C	D	E	F	G	H
I		a	b	c	d				
J		e	f	g	h				
K		i	j	k	l				
L		m	n	o	p				



- **Verallgemeinerung der zeitlichen Prädiktion**
 - **B-Frames** dürfen **nun** als **Referenz verwendet werden**
 - Voraussage von B-Frames aus zwei zeitlich vorangegangenen Referenzbildern (I-Frames)
 - **Prädiktion** basierend auf einer **gewichteten linearen Kombination zweier Referenzblöcke**
- **Prädiktion der Bewegungsvektoren**
 - Benachbarte Blöcke führen in der R. ähnliche Bewegung durch
 - Vorhersage der Bewegungsvektoren anhand von drei Nachbarvektoren
 - Übertragung des Differenzvektors, Rekonstruktion durch Differenz und Prädiktionsvektor
- **Filter zur Unterdrückung von Blockartefakten (Deblockingfilter)**
 - Künstliche Blockkante vorhanden, wenn Differenz der Bildpunkte im Bereich der Blockgrenze beobachtet wird → **Anwendung des Deblockingfilters**
 - Kleine Differenz oder sehr große Differenz, die nicht durch Quantisierungseffekte erklärt werden kann
→ **Bildpunkte bleiben unverändert**





Zusammenfassung

- Standards zur Bild- und Videokomprimierung
 - **Vielzahl unterschiedlicher Standards** zur Bild- und Videokomprimierung
 - Neuentwicklungen zielen auf eine **Verdopplung der Kompressionseffizienz** im Vergleich zum letzten Standard ab
 - Bsp.: MPEG-2 → H.264 / MPEG AVC → H.265