

Offline-Kodierungstechniken am Beispiel H.264

Medientechnologie IL

Andreas Unterweger

Vertiefung Medieninformatik
Studiengang ITS
FH Salzburg

Sommersemester 2017

- Offlinekodierung
 - Kodierung ohne Echtzeitvoraussetzungen
 - Zu kodierende Daten stehen fest
 - Verzögerungen quasi irrelevant
- Ausgewählte Offline-Kodierungstechniken
 - RDO (Rate-Distortion-Optimierung)
 - Profile und Level
- Nicht behandelte Aspekte (Auswahl)
 - Videoqualitätsoptimierung
 - Exakte Bitratensteuerung
 - Puffermanagement
 - Vorverarbeitung (Farbraum- und -tiefenumwandlung etc.)
 - Nachbearbeitung (künstliches Rauschen, anamorphe Abtastung etc.)
 - Überschneidungen mit nicht behandelten Online-Kodierungstechniken (z.B. Skalierbarkeit und Optimierung)

- Videospuren auf Blu-rays
 - Sehr hohe Qualität
 - Sehr große Speicherkapazität
 - Harte Dekodierungsbeschränkungen durch Hardware
 - Kodierdauer quasi beliebig (terminbeschränkt)
- Video on Demand (vorkodiert)
 - Mittlere Qualität
 - Beschränkte Übertragungsbandbreite
 - Weiche Dekodierungsbeschränkungen
 - Kodierdauer abhängig von Kodier-Vorlauf
- Videoplattform (vorkodiert, z.B. *YouTube*)
 - Variable Qualität (Mehrfachkodierung)
 - Beschränkte Übertragungsbandbreite
 - Weiche Dekodierungsbeschränkungen
 - Möglichst kurze Kodierdauer (quasi-sofortige Verfügbarkeit)

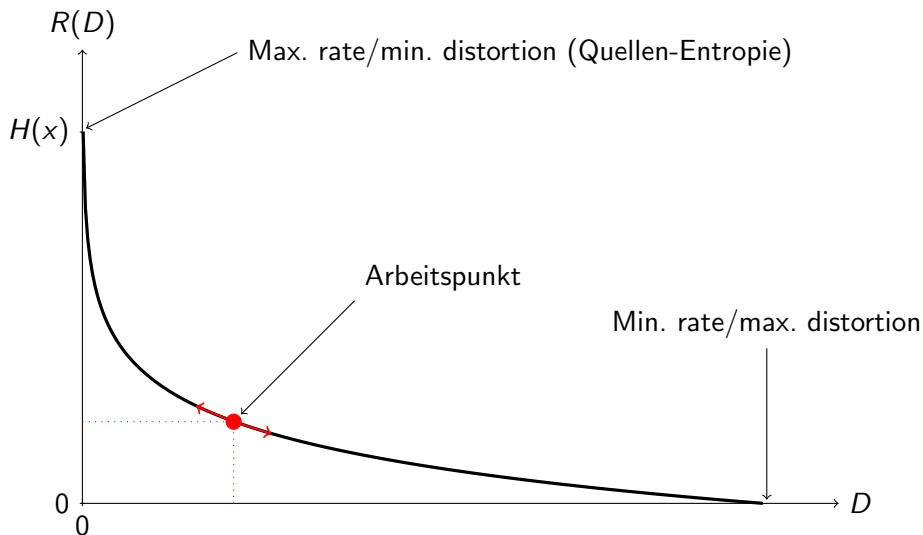
→ Verschiedene Anforderungen je nach Anwendungsgebiet

Motivation Rate-Distortion-Optimierung (RDO) I

- Datenrate (englisch *rate*) und quantisierungsbedingte Verzerrung (englisch *distortion*) hängen zusammen
- Hohe Qualität (niedrige Verzerrung) benötigt hohe Datenrate
- Niedrige Qualität (hohe Verzerrung) benötigt niedrige Datenrate
- Minimalqualität ist durch Minimaldatenrate beschränkt
- Maximalqualität ist durch Maximaldatenrate (bei redundanzfreier Kompression) durch Quellen-Entropie beschränkt

- Praktische Forderung: Minimale Verzerrung bei gegebener Datenrate
- Problem: Verschiedene Datenraten **und** verschiedene Verzerrungen → Verzerrungen nicht direkt vergleichbar
- Lösung: RDO (englisch *Rate-Distortion Optimization*)

Motivation Rate-Distortion-Optimierung (RDO) II



- Vereinfachende Annahme: Kodierte Makroblöcke (inkl. Datenrate und Verzerrung) sind unabhängig voneinander (nicht korrekt, aber globale Optimierung wäre völlig unpraktikabel, da viel zu aufwändig)
- Für jeden Makroblock
 - Alle Modi (Intra, Inter mit verschiedenen MV etc.) ausprobieren, d.h. voll kodieren (inkl. Entropiekodierung!) → **sehr** aufwändig
 - Modi, die Anforderungen nicht entsprechen, ausschließen
 - Datenrate und Verzerrung jedes gültigen Modus aufzeichnen
 - Modus mit dem besten Rate-Distortion-Verhältnis¹ wählen
- Problem: Bewertung des Rate-Distortion-Verhältnisses
- Lösung als Optimierungsproblem mit Nebenbedingungen bzw. Einschränkungen (englisch *constrained optimization problem*)

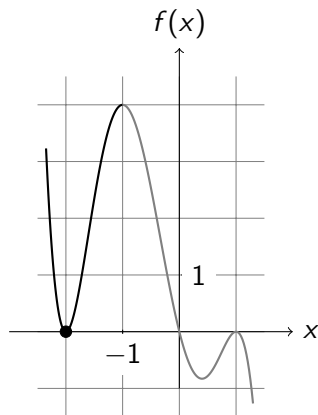
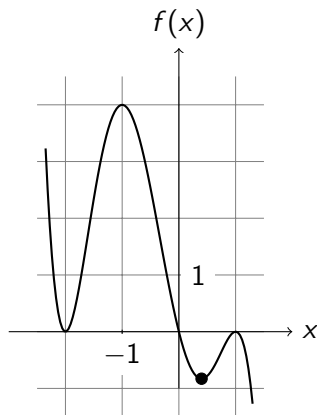
¹Verhältnis meint hier **nicht** $\frac{R}{D}$!

- $R_{max} \in \mathbb{R}$: Maximale Datenrate
- Makroblock besteht aus $p_{max} \in \mathbb{N}$ Partitionen \rightarrow Partitionsindex $p \in \{x \in \mathbb{N} \mid 1 \leq x \leq p_{max}\}$
- \mathbb{M}_p : Menge aller möglichen Modi für Partition mit Index p
- $R_p : \mathbb{M}_p \rightarrow \mathbb{R}$: Datenrate in verwendetem Modus
- $D_p : \mathbb{M}_p \rightarrow \mathbb{R}$: Verzerrung in verwendetem Modus
- Optimierungsziel pro Makroblockpartition (mit Nebenbedingung):

$$\min_{m_p \in \mathbb{M}_p} D_p(m_p)$$
$$R_p(m_p) \leq R_{max}$$

Einschub: Optimierungsbeispiel mit Nebenbedingung

- Minimiere $f(x) = -x^5 - 2x^4 + 3x^3 + 4x^2 - 4x$, wobei $x \leq -1$



Einschub: Lagrange-Multiplikatoren

- Optimierungsziel: Minimiere $f(x)$ unter Einhaltung von $x \leq K \in \mathbb{R}$
- $f : \mathbb{R} \rightarrow \mathbb{R}$: Zu optimierende Funktion
- $g : \mathbb{R} \rightarrow \mathbb{R}$: Nebenbedingungsfunktion (ohne K)
- Methode von Lagrange (vereinfacht): Minimiere stattdessen

$$\Lambda(x) = f(x) + \lambda g(x), \lambda \in \mathbb{R}^+$$

→ Optimierungsproblem ohne Nebenbedingungen

- λ : Lagrange-Multiplikator (Hilfsgröße)
- $\lambda \geq 0 \rightarrow$ Minimum (ansonsten Maximum)
- Wird λ (konstant) gewählt und $\Lambda(x)$ minimiert, wird $f(x)$ für einen bestimmten Wert von g minimal \rightarrow Kosten werden vergleichbar

→ Minimiere Kostenfunktion $f(x) + \lambda g(x)$ (trivial)

$$\min_{m_p \in \mathbb{M}_p} D_p(m_p), R_p(m_p) \leq R_{max}$$

- Mit Lagrange-Multiplikator (Kostenfunktionsminimierung):

$$\min_{m_p \in \mathbb{M}_p} D_p(m_p) + \lambda R_p(m_p), \lambda \in \mathbb{R}^+$$

- Wenn D additiv, einfache Optimierung für gesamten Makroblock (alle Partitionen unter Vernachlässigung von deren Abhängigkeiten):

$$\sum_{p=1}^{p_{max}} \min_{m_p \in \mathbb{M}_p} D_p(m_p) + \lambda R_p(m_p), \lambda \in \mathbb{R}^+$$

$$\min_{m_p \in \mathbb{M}_p} D_p(m_p) + \lambda R_p(m_p), \lambda \in \mathbb{R}^+$$

- Nachfolgende Formeln adaptiert aus²
- Modientscheidung für Partition p über RDO:

$$m_{optimal} = \arg \min_{m_p \in \mathbb{M}_p} (SSD(m_p) + \lambda_{Modus} R_p(m_p))$$

$$\lambda_{Modus} = \alpha Q_{step}^2, \alpha \in \mathbb{R}^+ (\text{empirisch})$$

²Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T. K. und Wiegand, T.: Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC). In *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 2012.

$$m_{optimal} = \arg \min_{m_p \in \mathbb{M}_p} (SSD(m_p) + \lambda_{Modus} R_p(m_p))$$

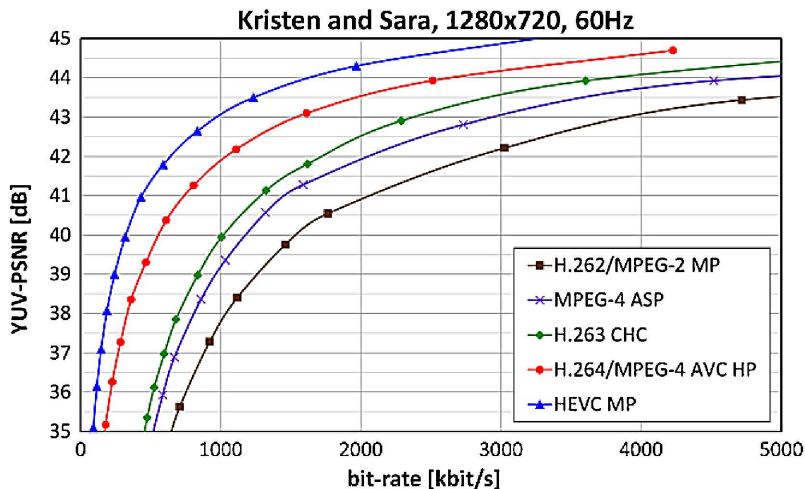
- Analog MV-Entscheidung für Partition p (vereinfacht nur für Vollpixel-ME, d.h. ohne Subpixel-MV)
- Zusätzliche Variablen, „dimensionen“ (hier als in \mathbb{M}_p inkludiert angenommen): Referenzbildindex, x- und y-Komponente des MV

$$m_{optimal} = \arg \min_{m_p \in \mathbb{M}_p} (SAD(m_p) + \lambda_{MV} R_p(m_p))$$

$$\lambda_{MV} = \sqrt{\lambda_{Modus}} = \sqrt{\alpha} Q_{step}$$

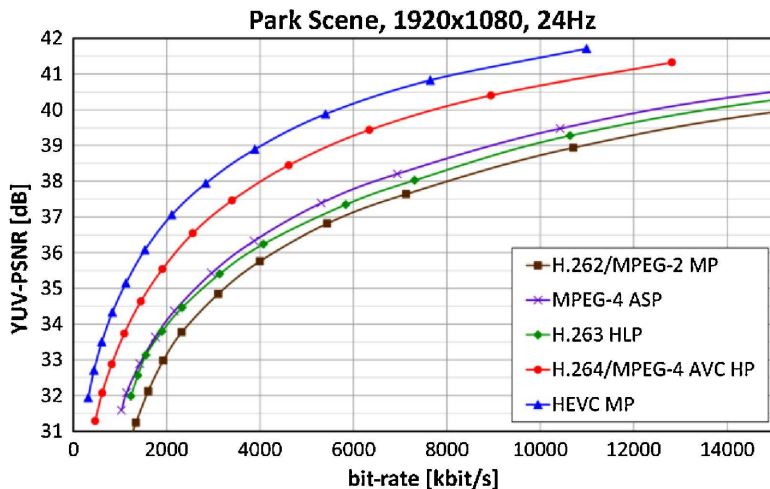
- Bidirektionale MV werden getrennt betrachtet und ihre Kosten addiert
- Weitere Möglichkeiten (ohne Details): Koeffizienten (z.B. *Trellis*)

RDO-Beispiel (Effizienzvergleich) I



Quelle: Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T. K. und Wiegand, T.: Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC). In *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 2012.

RDO-Beispiel (Effizienzvergleich) II



Quelle: Ohm, J.-R., Sullivan, G. J., Schwarz, H., Tan, T. K. und Wiegand, T.: Comparison of the Coding Efficiency of Video Coding Standards – Including High Efficiency Video Coding (HEVC). In *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 2012.

- Unterschiede zwischen Abspielgeräten
 - Dekodierleistung (prozessorabhängig)
 - Maximaler Datendurchsatz (busabhängig)
 - Verfügbarer Speicher
- Dekodierleistung abschätzbar machen
- Zwei unabhängige Komplexitätsdimensionen
 - Profil: Verwendbarer Satz von Coding Tools
 - Level: Schranke für Datendurchsatz und Speicherbedarf
- Kombination aus Profil und Level ergibt Komplexitätsabschätzung
- Encoder kommuniziert Kombination an Decoder
- Decoder entscheidet anhand seiner maximalen Profil-Level-Kombination, ob Abspielen möglich ist

- Definieren verwendbare Coding Tools
 - Schließen bestimmte Coding Tools aus
 - Encoder **muss nicht** alle Coding Tools verwenden
 - Decoder **muss** alle Coding Tools unterstützen
- Verschiedene Profile für verschiedene Anwendungsfälle (Auswahl):
 - Baseline: Telekonferenzen
 - Main: Fernsehübertragungen
 - High: Videospeicherung (z.B. für Blu-rays)
 - High 4:4:4 Predictive: Professionelles Authoring sowie verlustfreie Speicherung mit erweiterter Farbtiefe und -unterabtastung
- Weitere Intra-Profile für Schnittanwendungen u.ä.
- Zusätzliche Profile für skalierbare Videokodierung (SVC)
- Zusätzliche Profile für Multi-view Video Coding (MVC)

Profile II

Feature	CBP	BP	XP	MP	ProHiP	HiP	Hi10P	Hi422P	Hi444PP
Bit depth (per sample)	8	8	8	8	8	8	8 to 10	8 to 10	8 to 14
Chroma formats	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0	4:2:0/4:2:2	4:2:0/4:2:2/4:4:4
Flexible macroblock ordering (FMO)	No	Yes	Yes	No	No	No	No	No	No
Arbitrary slice ordering (ASO)	No	Yes	Yes	No	No	No	No	No	No
Redundant slices (RS)	No	Yes	Yes	No	No	No	No	No	No
Data Partitioning	No	No	Yes	No	No	No	No	No	No
SI and SP slices	No	No	Yes	No	No	No	No	No	No
Interlaced coding (PicAFF, MBAFF)	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes
B slices	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CABAC entropy coding	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
4:0:0 (Monochrome)	No	No	No	No	Yes	Yes	Yes	Yes	Yes
8x8 vs. 4x4 transform adaptivity	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Quantization scaling matrices	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Separate C _b and C _r QP control	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Separate color plane coding	No	No	No	No	No	No	No	No	Yes
Predictive lossless coding	No	No	No	No	No	No	No	No	Yes

Quelle: Wikipedia: H.264/MPEG-4 AVC. http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC#Profiles (1.10.2013), 2013.

CBP = Constrained Baseline Profile, BP = Baseline Profile, XP = Extended Profile, MP = Main Profile, ProHiP = Progressive High Profile, HiP = High Profile, Hi10P = High 10 Profile, Hi422P = High 4:2:2 Profile, Hi444PP = High 4:4:4 Predictive Profile

- Beschränken die Speicher- und Rechenanforderungen
 - Liste quantitativer Beschränkungen (Maximalwerte)
 - Encoder **muss** Beschränkungen **nicht** ausreizen
 - Decoder **muss** Ausreizung der Beschränkungen unterstützen
- Verschiedene Level für typische Anwendungsfälle (Auswahl):
 - 1: Telekonferenzen (176x144@15fps)
 - 3: Fernsehübertragungen (720x576@25fps)
 - 4.1: Videospeicherung (1920x1080@30fps)
 - 5.1: Großprojektionen (4096x2048@30fps)
- Wenn Decoder ein Level unterstützt, muss er auch alle darunter liegenden (niedrigeren) unterstützen
- Encoder muss Level angeben (kann Benutzervorgabe sein)

Level	Max decoding speed		Max frame size		Max video bit rate for video coding layer (VCL) kbit/s		
	Luma samples/s	Macroblocks/s	Luma samples	Macroblocks	Baseline, Extended and Main Profiles	High Profile	High 10 Profile
1	380,160	1,485	25,344	99	64	80	192
1b	380,160	1,485	25,344	99	128	160	384
1.1	768,000	3,000	101,376	396	192	240	576
1.2	1,536,000	6,000	101,376	396	384	480	1,152
1.3	3,041,280	11,880	101,376	396	768	960	2,304
2	3,041,280	11,880	101,376	396	2,000	2,500	6,000
2.1	5,068,800	19,800	202,752	792	4,000	5,000	12,000
2.2	5,184,000	20,250	414,720	1,620	4,000	5,000	12,000
3	10,368,000	40,500	414,720	1,620	10,000	12,500	30,000
3.1	27,648,000	108,000	921,600	3,600	14,000	17,500	42,000
3.2	55,296,000	216,000	1,310,720	5,120	20,000	25,000	60,000
4	62,914,560	245,760	2,097,152	8,192	20,000	25,000	60,000
4.1	62,914,560	245,760	2,097,152	8,192	50,000	62,500	150,000
4.2	133,693,440	522,240	2,228,224	8,704	50,000	62,500	150,000
5	150,994,944	589,824	5,652,480	22,080	135,000	168,750	405,000
5.1	251,658,240	983,040	9,437,184	36,864	240,000	300,000	720,000
5.2	530,841,600	2,073,600	9,437,184	36,864	240,000	300,000	720,000

Quelle: Wikipedia: H.264/MPEG-4 AVC. http://en.wikipedia.org/wiki/H.264/MPEG-4_AVC#Levels (1.10.2013), 2013.

Fragen?