

Technische Universität München  
Lehrstuhl für Integrierte Systeme

# **Echtzeitfähige Datenratenkontrolle für die Videokommunikation**

**Markus Zeller**

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der  
Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktor-Ingenieurs**

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Klaus Diepold

Prüfer der Dissertation:

1. Univ.-Prof. Dr.-Ing. Ingolf Ruge, em.
2. Univ.-Prof. Dr.-Ing. Gert Hauske

Die Dissertation wurde am 01.06.2004 bei der Technischen Universität München eingereicht und  
durch die Fakultät für Elektrotechnik und Informationstechnik am 25.10.2004 angenommen.



## **Vorwort**

Mein erster und besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr.-Ing. Ingolf Ruge für die Ermöglichung und Förderung dieser Arbeit. Seine aktive Unterstützung und wertvollen Tipps haben wesentlich zum Gelingen dieser Arbeit beigetragen.

Herrn Prof. Dr.-Ing. Gert Hauske danke ich für die Übernahme des Zweitberichts und das damit bekundete Interesse an der Arbeit. Seine zahlreichen und wertvollen Hinweise und Anregungen während des gesamten Entstehungsprozesses haben mir sehr geholfen.

Ein besonderer Dank geht an Herrn Dr.-Ing. Rudi Knorr und Herrn Dr.-Ing. Walter Stechele für die Unterstützung der Arbeit. Ihre stets hilfreichen Diskussionen und die vielen guten Anregungen haben mich bei der Entstehung der Arbeit wesentlich unterstützt und motiviert. Nur durch die Unterstützung von Herrn Dr.-Ing. Rudi Knorr und auch Herrn Dr.-Ing. Helmut Steckenbiller ist diese Arbeit im Rahmen meiner Anstellung als wissenschaftlicher Mitarbeiter bei der Fraunhofer-Einrichtung für Systeme der Kommunikationstechnik ESK ermöglicht worden.

Herrn Mike Heidrich, Herrn Holger Hutzelmann, Herrn Dr.-Ing. Rudi Knorr, Herrn Dr.-Ing. Walter Stechele und Herrn Rainer Steffen danke ich für die kritische Durchsicht des Manuskripts und den intensiven, fachlich anregenden Austausch.

Ein spezieller Dank gilt meiner ganzen Familie für das entgegengebrachte Verständnis und die Geduld während der gesamten Arbeit. Insbesondere möchte ich mich bei meiner Frau Ivonne Zeller herzlich bedanken, die sich zusätzlich noch die Mühe des Korrekturlesens gemacht hat.

Diedorf im Mai 2004.



## Kurzfassung

Diese Arbeit untersucht und entwickelt Lösungen zur Problematik der echtzeitfähigen Datenratenkontrolle für aktuelle Videocodecs mit dem Schwerpunkt Kommunikationsanwendungen.

Für eine verzögerungs- und bandbreitenbegrenzte Übertragung eines codierten Videobitstroms ist die Regelung der Ausgangsbitrate notwendig. Eine Codierung mit konstanter Verzerrung (Distortion) resultiert in einer von der Verteilungsdichtefunktion des Bildsignals abhängigen variablen Bitrate. Die Festlegung der Verzerrung des Bildsignals erfolgt durch die Quantisierung nach der Dekorrelation. Eine Übertragung mit festgelegter Bitrate erfordert somit, dass die Verzerrung des Bildes angepasst wird. Hierzu werden Puffer im Encoder und Decoder benutzt, in denen der Videobitstrom bei der Übertragung zwischengespeichert wird. Droht ein Puffer über oder leer zu laufen, muss gegengesteuert werden, indem die Verzerrung erhöht bzw. verringert wird.

Der funktionale Zusammenhang zwischen Bitrate und Verzerrung des Encoders in Verbindung mit dem aktuellen Bildsignal kann anhand statistischer Modelle, wie z.B. Gauß- oder Laplacequellen näherungsweise beschrieben werden. Für den Einsatz unter Echtzeitbedingungen werden die statistischen Modelle durch eine vereinfachte Modellierung approximiert. Für die Vereinfachungen dienen Annahmen über die Signalstatistik, die sich aus dem jeweils interessanten Bereich der Komplexität des Bildsignals oder der erwarteten Ausgangsbitrate ableiten. Zudem ist eine Anpassung an das reale Verhalten des Encoders in Verbindung mit dem aktuellen Bildsignal notwendig.

Seit über 10 Jahren bildet bei allen Standardisierungen ein Hybrid-DCT-Videocodierverfahren (Diskrete Cosinus Transformation) mit Bewegungskompensation die Basis. Aktuelle Videostandards, wie MPEG-4 (Moving Picture Experts Group) oder H.264, entsprechen dabei in ihrer Grundfunktion H.263. Ein neuer, wichtiger Aspekt aktueller Videostandards ist, dass sie nicht mehr fest auf einzelne Zielapplikation ausgerichtet sind. Durch die Unterstützung unterschiedlicher Anwendungen und Netze ergeben sich neue Anforderungen an die im Standard nicht definierten, aber qualitätsbestimmenden Module des Videocodecs, wie die Datenratenkontrolle.

Die Analyse bekannter, echtzeitfähiger Verfahren für die Datenratenkontrolle zeigt, dass sie bisher auf einen eng begrenzten Anwendungsbereich oder für die Unterstützung spezieller Netze konzipiert worden sind. Sobald das breite Anwendungsspektrum neuer Videocodecs unterstützt werden muss, kann die Datenratenkontrolle nicht mehr fest auf bestimmte Anwendungsbereiche mit weitgehend fixen Randbedingungen, wie z.B. Zielbitrate oder zu erwartende Komplexität des Bildmaterials, ausgerichtet werden.

Diese Problematik begründet die Zielsetzung dieser Arbeit. Es soll ein neues Verfahren für die Datenratenkontrolle konzipiert werden, welches das breite Anwendungsspektrum aktueller Hybrid-DCT-Videoencoder im Kommunikationsbereich unterstützt. Wesentliches Kriterium für das adaptive Verfahren ist dabei eine stabile und effiziente Encodierung unter Einhaltung der Vorgaben der Standardisierung, des jeweiligen Übertragungsnetzes und der Zielanwendung. Zusätzlich ist eine Echtzeitfähigkeit verbunden mit einer geringen Komplexität gefordert.

Im ersten Teil dieser Arbeit wird auf Basis der Analyse der Anforderungen und der Untersuchung des Rate-Distortion-Verhaltens ein funktionales Systemmodell für ein echtzeitfähiges Verfahren für die Datenratenkontrolle erstellt. Als Anforderungskriterien sind zu nennen: die Unterstützung eines breiten Spektrums von Ausgangsbitraten (von einigen kBit/s bis hin zu mehreren MBit/s), eine effiziente Codierung von Bildmaterial unterschiedlicher Komplexität, die Unterstützung einer im Voraus wählbaren maximalen Ende-zu-Ende-Verzögerung und die Gewährleistung der Regelung einer variablen Ausgangsbitrate.

Für die Festlegung des Quantisierungsparameters ist eine prädiktive Rate-Distortion-Schätzung des aktuellen Signals in Verbindung mit dem Encoder notwendig. Die Wahl der jeweiligen Quantisierungsparameter beeinflusst die Effizienz der Videoencodierung. Die Untersuchung bekannter Ansätze für die Rate-Distortion-Modellierung anhand von unterschiedlichen Testsequenzen zeigt, dass dabei im geforderten Anwendungsbereich hohe Schätzfehler mit einer mittleren Abweichung von über 25% auftreten. Dies erfordert eine verbesserte Rate-Distortion-Modellierung beim Entwurf neuer Verfahren für die Datenratenkontrolle. Zudem ist es notwendig, dass eine rasche Adaption auf variierende Randbedingungen und an das aktuelle Eingangssignal stattfindet.

Im zweiten Teil dieser Arbeit wird aufbauend auf den vorherigen Untersuchungsergebnissen ein Ansatz für die echtzeitfähige Rate-Distortion-Modellierung hergeleitet, der die Grundlage des neuen Verfahrens für die Datenratenkontrolle bildet. Die Modellierung beruht sowohl auf einer genaueren Rate-Distortion-Schätzung als auch auf einem verallgemeinerten Ansatz zur Beschreibung von nieder- bis hochbitratigen Videosignalen. Zudem wird ein neuer Schätzparameter auf Basis der DCT-Koeffizienten definiert, der gegenüber der Varianz als Schätzparameter im Bitratenbereich unter 0,3 Bit/Pixel eine genauere Schätzung ermöglicht. Die Kombination der hergeleiteten Rate-Distortion-Modellierung und des neuen Schätzparameters ermöglicht, innerhalb eines Bildes auf sich ändernde Bedingungen zu reagieren. Sie bildet damit die Basis für eine stabile und effiziente Encodierung.

Für die Unterstützung zusätzlicher Leistungsmerkmale von aktuellen Videostandards ist eine Erweiterung des neuen Verfahrens notwendig. Folgende Ansätze sind hierfür zusätzlich konzipiert worden: Unterstützung von variierenden Ausgangsdatenraten, Modellierung von unterschiedlich codierten Bildern bzw. Makroblöcken, die Unterstützung mehrerer Objekte in einer Szene und die Möglichkeit, psychooptisch wichtige Bildbereiche gezielt zu betonen.

Im Vergleich zu den Verfahren aus der Standardisierung von MPEG-4 und H.263 kann eine Verbesserung des *PSNR*-Wertes (Peak Signal to Noise Ratio) im Schnitt um ca. 0,7 dB erreicht werden. In einigen Fällen ist eine Steigerung bis über 2 dB möglich.

Im dritten Teil dieser Arbeit wird auf Implementierungsaspekte eingegangen. Aktuelle Standardprozessoren sind aufgrund spezieller SIMD-Funktionseinheiten (Single Instruction Multiple Data) für die Unterstützung der Signalverarbeitung in der Lage, Videosignale in Fernsehqualität in Echtzeit zu codieren. Eine Analyse des zur Verfügung stehenden Befehlssatzes und der erstellten Implementierung zeigt, dass damit insbesondere die Berechnung statistischer Schätzparameter auf Basis einzelner Pixel oder Koeffizienten beschleunigt werden kann. Der zur Verfügung stehende H.263-Videoencoder kann durch den Einsatz der SIMD-Funktionseinheiten im Mittel um den Faktor 13 beschleunigt werden. Einzelne Funktionen werden sogar bis zum Faktor 20 gegenüber der teiloptimierten C-Implementierung beschleunigt. Insbesondere die zur Verfügung stehende Speichertransferleistung, das Fehlen von Registern und Einschränkungen im Befehlssatz begrenzen eine weitere Beschleunigung.

Neben der Bildqualität spielt die Ende-zu-Ende-Verzögerung bei Kommunikationsanwendungen eine wichtige Rolle. Die Verzögerung bei der Videokommunikation wird vor allem durch die Verarbeitungsgeschwindigkeit im Videocodec und durch den Einsatz von Puffern bestimmt. Eine Reduzierung der Verzögerung der Videokommunikation kann damit sowohl mit der Beschleunigung der Videosignalverarbeitung, z.B. durch den Einsatz von SIMD-Prozessorerweiterungen, als auch durch eine Reduzierung der notwendigen Puffer erreicht werden. Die Untersuchung des neuen Verfahrens für die Datenratenkontrolle zeigt, dass es durch geringe Modifikationen möglich ist, auf die bisher für die Datenratenkontrolle notwendige Zwischenpufferung von Makroblöcken zu verzichten. Ohne eine nennenswerte Verringerung der Codiereffizienz kann damit die Verzögerung der Videocodierung in Abhängigkeit von der Framerate um 50 - 150 ms verringert werden.

Die Ergebnisse dieser Arbeit zeigen, dass es für einen breiten Anwendungsbereich möglich ist, ein adaptives, echtzeitfähiges Verfahren für die Datenratenkontrolle mit einem im Vergleich zu anderen Verfahren vergleichbaren Aufwand zu implementieren.





# Inhalt

	<b>Verzeichnis der Abbildungen</b> .....	<b>XIII</b>
	<b>Verzeichnis der Tabellen</b> .....	<b>XVII</b>
<b>1</b>	<b>Einleitung</b> .....	<b>1</b>
1.1	Entwicklung der Videocodierverfahren .....	4
1.2	Anwendungsszenarios und deren Anforderungen .....	5
1.3	Struktur und Aufbau der Arbeit .....	6
<b>2</b>	<b>Grundlagen der Datenratenkontrolle in Videocodecs</b> .....	<b>9</b>
2.1	Einführung .....	9
2.2	Prinzipien der Videocodierung .....	9
2.2.1	Beurteilung von Videocodierverfahren .....	10
2.2.2	Videocodiermethoden .....	12
2.2.3	Hybrid-DCT-Videoencoder mit Bewegungskompensation .....	13
2.2.4	Standardisierung im Bereich Videocodierung .....	16
2.3	Datenratenkontrolle in Videocodecs .....	21
2.3.1	Rate-Distortion-Theorie .....	22
2.3.2	Funktionsprinzip der Datenratenkontrolle bei der Videocodierung .....	25
2.3.3	Realisierungsansätze für die Datenratenkontrolle .....	33
2.4	Zusammenfassung .....	40

---

<b>3</b>	<b>Systemmodellierung von neuen Verfahren zur Datenratenkontrolle . . . . .</b>	<b>41</b>
3.1	Einführung . . . . .	41
3.2	Anforderungen und Randbedingungen . . . . .	42
3.3	Signalcharakteristika und deren Modellierung . . . . .	44
3.3.1	Analyse der Signalcharakteristik . . . . .	44
3.3.2	Interpretation und Zusammenfassung der Auswertung . . . . .	51
3.4	Systemansatz eines neuen Verfahrens für die Datenratenkontrolle . . . . .	52
3.5	Zusammenfassung . . . . .	56
<b>4</b>	<b>Vorschlag eines neuen Verfahrens für die Datenratenkontrolle . . . . .</b>	<b>57</b>
4.1	Einführung . . . . .	57
4.2	Lösungsansatz für eine adaptive Datenratenkontrolle . . . . .	58
4.2.1	Grundlegende Rate-Distortion-Modellierung . . . . .	58
4.2.2	Verallgemeinerung der Rate-Distortion-Modellierung . . . . .	59
4.2.3	Parametrisierung der Rate-Distortion-Modellierung . . . . .	63
4.2.4	Untersuchung statistischer Parameter im Spektralbereich . . . . .	64
4.3	Datenratenkontrolle für einen Hybrid-DCT Videoencoder . . . . .	68
4.3.1	Eingangsparameter und globale Initialisierung . . . . .	70
4.3.2	Datenratenkontrolle auf der GOP-Ebene . . . . .	72
4.3.3	Datenratenkontrolle auf der Bild-Ebene . . . . .	73
4.3.4	Datenratenkontrolle auf der Makroblock-Ebene . . . . .	73
4.4	Erweiterungsmöglichkeiten . . . . .	77
4.4.1	Analyse der DCT-Koeffizienten . . . . .	77
4.4.2	Variable Ausgangsdatenraten . . . . .	78
4.4.3	Modellierung von I- und B-Bildern bzw. Makroblöcken . . . . .	81
4.4.4	Multi-Objektfähigkeit . . . . .	84
4.4.5	Subjektive Bildqualität . . . . .	85
4.5	Zusammenfassung . . . . .	89

---

<b>5</b>	<b>Verifikation des neuen Verfahrens für die Datenratenkontrolle</b> . . . . .	<b>91</b>
5.1	Einführung . . . . .	91
5.2	Testbedingungen . . . . .	92
5.3	Untersuchungsergebnisse . . . . .	94
5.3.1	Leistungsfähigkeit von FLEX . . . . .	94
5.3.2	Zusammenfassung und Interpretation der Ergebnisse . . . . .	102
5.4	Untersuchung der Teilaspekte von FLEX . . . . .	103
5.4.1	Datenratenkontrolle auf Bild-Ebene . . . . .	104
5.4.2	Aktualisierung der Modellparameter . . . . .	106
5.4.3	Analyse der DCT-Koeffizienten . . . . .	107
5.5	Zusammenfassung . . . . .	108
<b>6</b>	<b>Realisierungsuntersuchungen</b> . . . . .	<b>111</b>
6.1	Einführung . . . . .	111
6.2	Multimedia-Prozessorerweiterungen . . . . .	112
6.3	Einsatz von SIMD-Befehlserweiterungen in Videocodecs . . . . .	114
6.3.1	Beschleunigung wichtiger Funktionen im Videocodec . . . . .	115
6.3.2	Auswertung der SIMD-Implementierung . . . . .	123
6.3.3	Zusammenfassung und Ausblick . . . . .	127
6.4	Minimierung der Ende-zu-Ende-Verzögerung . . . . .	128
6.4.1	Ermittlung der Verarbeitungsverzögerung . . . . .	129
6.4.2	Lösungsansätze zur Minimierung der Verzögerung . . . . .	133
6.5	Zusammenfassung . . . . .	136
<b>7</b>	<b>Zusammenfassung, Bewertung und Ausblick</b> . . . . .	<b>139</b>
<b>Anhang A</b>	<b>Beschreibung der Testsequenzen</b> . . . . .	<b>147</b>
<b>Anhang B</b>	<b>Ergänzende Auswertungen zu Kapitel 5</b> . . . . .	<b>153</b>
<b>Anhang C</b>	<b>Ergänzungen zu SIMD-Prozessorerweiterungen in Kapitel 6</b> . . . . .	<b>159</b>
	<b>Verzeichnis der Abkürzungen, Definitionen und Nomenklatur</b> . . . . .	<b>163</b>
	<b>Literaturverzeichnis</b> . . . . .	<b>177</b>



# Verzeichnis der Abbildungen

## 1 Einleitung

<b>Abb. 1.1:</b>	Ende-zu-Ende-Videokommunikation zwischen den Endgeräten A und B. . . . .	1
<b>Abb. 1.2:</b>	Klassifikation der verschiedenen Netze. . . . .	3
<b>Abb. 1.3:</b>	Die Videostandardisierung definiert das Bitstromformat und die Decoderstruktur. . .	4
<b>Abb. 1.4:</b>	Kommunikationsmöglichkeiten eines mobilen Endgerätes. Jedes Netz stellt unterschiedliche Anforderungen an die Videocodierung. . . . .	5
<b>Abb. 1.5:</b>	Der Videocodec steht als Basisdienst des Betriebssystems für unterschiedliche Anwendungen zur Verfügung. Die Übertragung der Daten erfolgt über verschiedene Netze. . . . .	6
<b>Abb. 1.6:</b>	Einordnung der Datenratenkontrolle im Videoencoder in ein Ende-zu-Ende-Videokommunikationssystem. . . . .	7

## 2 Grundlagen der Datenratenkontrolle in Videocodecs

<b>Abb. 2.1:</b>	Schematischer Aufbau eines Hybrid-DCT-Videoencoders mit zeitlich prädiktiver Bewegungskompensation und Transformationscodierung des Prädiktionsfehlers. . .	13
<b>Abb. 2.2:</b>	Grundprinzip der blockbasierten Bewegungsschätzung. . . . .	14
<b>Abb. 2.3:</b>	Zeitliche Abfolge der P- und B-Bilder mit einem I-Bild als Ausgangsbasis für die Bewegungskompensation. . . . .	15
<b>Abb. 2.4:</b>	Vergleich verschiedener Videocodecs hinsichtlich ihrer Codiereffizienz [126]. . . .	20
<b>Abb. 2.5:</b>	Distortion-Rate-Kennlinien von Einzelbildern der Sequenz Foreman. . . . .	21
<b>Abb. 2.6:</b>	Beispielhafter Verlauf einer Rate-Distortion-Funktion bei wertkontinuierlichen (a) und wertdiskreten (b) Signalen [96]. . . . .	24
<b>Abb. 2.7:</b>	Resultierende Bitmenge pro Bild bei der Intracodierung der Sequenz Foreman mit const. Quantisierungsparameter. . . . .	25
<b>Abb. 2.8:</b>	Füllstandverlauf des Eingangspuffers des Decoders in Abhängigkeit von der Zeit [73]. . . . .	28
<b>Abb. 2.9:</b>	Entropie einer Quelle, deren quantisiertes Ausgangssignal einer Laplace-Verteilung entspricht (durchgezogene Linien) und der durch (2.30) approximierten Bitrate (gestrichelte Linien) [106]. . . . .	32

### 3 Systemmodellierung von neuen Verfahren zur Datenratenkontrolle

<b>Abb. 3.1:</b>	Einordnung der Videocodierung in ein Kommunikationssystem. Die Adaptionsschicht passt einen allgemein gültigen Videocodec an verschiedene Übertragungsmedien an [96]. . . . .	42
<b>Abb. 3.2:</b>	Rate-Distortion-Kennlinien der Sequenz Mobile&Calendar. . . . .	46
<b>Abb. 3.3:</b>	Rate-Distortion-Kennlinien der Sequenz Mother&Daughter. . . . .	47
<b>Abb. 3.4:</b>	Rate-Distortion-Kennlinien der Sequenz Weather. . . . .	48
<b>Abb. 3.5:</b>	Bitraten der einzelnen Makroblöcke pro Bild (Sequenz Mobile&Calendar). . . . .	49
<b>Abb. 3.6:</b>	Verzerrung der einzelnen Makroblöcke pro Bild (Sequenz Mobile&Calendar). . . . .	50
<b>Abb. 3.7:</b>	Schnittstellenkonzept eines Videocodecs in Anlehnung an MPEG-4 mit einer zweistufigen Adaption an verschiedene Übertragungsnetze. . . . .	53
<b>Abb. 3.8:</b>	Prinzipieller Aufbau der Datenratenkontrolle. . . . .	54
<b>Abb. 3.9:</b>	Funktionsmodule im Encoder mit Zwischenpuffer für "vorcodierte" Makroblöcke. . . . .	55

### 4 Vorschlag eines neuen Verfahrens für die Datenratenkontrolle

<b>Abb. 4.1:</b>	Verhalten von $\hat{Q}_H^F / \hat{Q}^F$ für die Sequenzen Children, Mobile&Calendar und Stephan. . . . .	62
<b>Abb. 4.2:</b>	Die Parameter für die Rate-Distortion-Schätzung des Videosignals können im Hybrid-DCT-Videoencoder vor und nach der DCT-Transformation gewonnen werden. . . . .	65
<b>Abb. 4.3:</b>	Vergleich von Rate-Distortion-Kennlinien der Sequenz Mobile&Calendar in Abhängigkeit der Schätzparameter Standardabweichung und $m_{DCT}$ . . . . .	67
<b>Abb. 4.4:</b>	Funktionsblöcke des Basisverfahrens für die Datenratenkontrolle. . . . .	69
<b>Abb. 4.5:</b>	Zielbefüllungshöhe des Ausgangspuffers in Abhängigkeit der vorgegebenen Puffergröße. . . . .	71
<b>Abb. 4.6:</b>	Reaktionszeit auf sich ändernde Randbedingungen bei Kontrolle nach jeder Makroblock-Zeile. . . . .	79
<b>Abb. 4.7:</b>	Verhalten des vorgeschlagenen Verfahrens für die Datenratenkontrolle bei schwankenden Ausgangsdatenraten. . . . .	81
<b>Abb. 4.8:</b>	GOP-Struktur. . . . .	83
<b>Abb. 4.9:</b>	Bewertungsmodell für die Gewichtung einzelner Bildbereiche hinsichtlich der subjektiven Bildwahrnehmung. . . . .	86
<b>Abb. 4.10:</b>	Gesichtsfarbenbereich in Abhängigkeit der Farbdifferenzkomponenten U und V, der aus verschiedenen Bildern und Videosequenzen ermittelt worden ist. . . . .	87
<b>Abb. 4.11:</b>	Einzelergebnisse der Gewichtung einzelner Bildbereiche hinsichtlich der subjektiven Bildwahrnehmung. . . . .	88

## 5 Verifikation des neuen Verfahrens für die Datenratenkontrolle

<b>Abb. 5.1:</b>	Vergleich der Codiereffizienz des neuen Verfahrens FLEX mit den Referenzverfahren beim Bildformat QCIF. . . . .	97
<b>Abb. 5.2:</b>	Vergleich der Codiereffizienz des neuen Verfahrens FLEX mit den Referenzverfahren beim Bildformat CIF. . . . .	98
<b>Abb. 5.3:</b>	Qualitätsvergleich des neuen Verfahrens FLEX mit den Referenzverfahren. . . . .	99
<b>Abb. 5.4:</b>	Vergleich der Qualitätsschwankung innerhalb von Einzelbildern für verschiedene Verfahren für die Datenratenkontrolle. . . . .	101
<b>Abb. 5.5:</b>	Qualitätsvergleich von unterschiedlichen Verfahren für die Bitallocation. . . . .	105
<b>Abb. 5.6:</b>	Qualitätsvergleich unterschiedlicher Strategien für die Parameteraktualisierung. . . . .	106
<b>Abb. 5.7:</b>	Qualitätsvergleich von FLEX mit und ohne Analyse der DCT-Koeffizienten. . . . .	107

## 6 Realisierungsuntersuchungen

<b>Abb. 6.1:</b>	Prozessorarchitekturen für die Multimedia-Signalverarbeitung. . . . .	112
<b>Abb. 6.2:</b>	Vereinfachtes Prinzip einer SIMD-Verarbeitungseinheit. . . . .	113
<b>Abb. 6.3:</b>	Beschleunigung der SIMD-Implementierung gegenüber dem nicht optimierten H.263-Videoencoder aufgeschlüsselt nach Funktionseinheiten. . . . .	124
<b>Abb. 6.4:</b>	Aufteilung der zur Encodierung notwendigen Rechenleistung auf einzelne Funktionseinheiten. . . . .	125
<b>Abb. 6.5:</b>	Komponenten einer Ende-zu-Ende-Kommunikation. . . . .	128
<b>Abb. 6.6:</b>	Verzögerungbestimmende Funktionsblöcke eines Hybrid-DCT Videoencoders. . . . .	130
<b>Abb. 6.7:</b>	Verzögerungbestimmende Funktionsblöcke eines Hybrid-DCT Videodecoders. . . . .	131
<b>Abb. 6.8:</b>	Verarbeitungsverzögerung mit (a) und ohne (b) Zwischenpufferung von Codiererergebnissen im Encoder und Decoder. . . . .	132
<b>Abb. 6.9:</b>	Qualitätsvergleich unterschiedlicher Varianten der Datenratenkontrolle FLEX. . . . .	135

## 7 Zusammenfassung, Bewertung und Ausblick

### Anhang A Beschreibung der Testsequenzen

<b>Abb. A.1:</b>	Sequenz "Container Ship" der Klasse A. . . . .	149
<b>Abb. A.2:</b>	Sequenz "Mother&Daughter" der Klasse A. . . . .	149
<b>Abb. A.3:</b>	Sequenz "Bowling" der Klasse B. . . . .	149
<b>Abb. A.4:</b>	Sequenz "Coastguard" der Klasse B. . . . .	150
<b>Abb. A.5:</b>	Sequenz "Foreman" der Klasse B. . . . .	150
<b>Abb. A.6:</b>	Sequenz "Irene" der Klasse B. . . . .	150
<b>Abb. A.7:</b>	Sequenz "News" der Klasse B. . . . .	151
<b>Abb. A.8:</b>	Sequenz "Mobile&Calendar" der Klasse C. . . . .	151
<b>Abb. A.9:</b>	Sequenz "Stephan" der Klasse C. . . . .	151
<b>Abb. A.10:</b>	Sequenz "Table Tennis" der Klasse C. . . . .	152
<b>Abb. A.11:</b>	Sequenz "Children" der Klasse E. . . . .	152
<b>Abb. A.12:</b>	Sequenz "Weather" der Klasse E. . . . .	152

**Anhang B Ergänzende Auswertungen zu Kapitel 5**

- Abb. B.1:** Qualitätsvergleich der Einzelbilder der CIF-Sequenz Mobile&Calendar für FLEX und den Referenzverfahren. ....157
- Abb. B.2:** Qualitätsvergleich der Einzelbilder der QCIF-Sequenz Foreman für FLEX und den Referenzverfahren. ....158

**Anhang C Ergänzungen zu SIMD-Prozessorerweiterungen in Kapitel 6**

- Abb. C.1:** SIMD-Funktionsbeispiel für die Berechnung des *SAD*-Wertes zweier 64-Bit Register mit 8 Bit breiten Einzeldaten. ....159
- Abb. C.2:** Ports und Verarbeitungseinheiten der Pentium 4 Architektur [41]. ....161



# Verzeichnis der Tabellen

## 1 Einleitung

## 2 Grundlagen der Datenratenkontrolle in Videocodecs

**Tabelle 2.1:** Eigenschaften der digitalen Bildformate QCIF und CIF (europäische Werte) [96]. . 10

**Tabelle 2.2:** Minimales  $BPP_{maxKb}$  für unterschiedliche Bildformate [73]. . . . . 27

## 3 Systemmodellierung von neuen Verfahren zur Datenratenkontrolle

**Tabelle 3.1:** Anforderungen von Kommunikationsanwendungen an den Videocodec. . . . . 43

## 4 Vorschlag eines neuen Verfahrens für die Datenratenkontrolle

## 5 Verifikation des neuen Verfahrens für die Datenratenkontrolle

**Tabelle 5.1:** Einordnung der Testsequenzen in unterschiedliche Klassen [59], [65]. . . . . 92

**Tabelle 5.2:** Randbedingungen im definierten Testset. . . . . 93

**Tabelle 5.3:** Zusammenfassung des Qualitätsvergleichs der QCIF- und CIF-Sequenzen. . . . . 96

## 6 Realisierungsuntersuchungen

**Tabelle 6.1:** Mathematische Basisoperationen eines Videocodecs [20]. . . . . 114

**Tabelle 6.2:** Charakteristik von ausgewählten schnellen DCT-Algorithmen [39]. . . . . 118

**Tabelle 6.3:** Initialisierung und Berechnung der Bitmasken für die Quantisierung und Dequantisierung. . . . . 120

**Tabelle 6.4:** Rechenzeit der verschiedenen Verfahren für die Datenratenkontrolle pro Sequenz. 120

**Tabelle 6.5:** Initialisierung der für FLEX notwendigen Bitmasken. . . . . 122

**Tabelle 6.6:** Vergleich der Rechenzeit der nicht optimierten C und der SIMD-optimierten Funktionsmodule. . . . . 123

**Tabelle 6.7:** Effizienzanalyse der SIMD-Funktionsmodule. . . . . 126

## 7 Zusammenfassung, Bewertung und Ausblick

**Anhang A Beschreibung der Testsequenzen**

<b>Tabelle A.1:</b> Kurzbeschreibung der Sequenzen. ....	147
--	-----

**Anhang B Ergänzende Auswertungen zu Kapitel 5**

<b>Tabelle B.1:</b> Auflistung der einzelnen Tests der QCIF-Sequenzen. ....	154
---	-----

<b>Tabelle B.2:</b> Auflistung der einzelnen Tests der CIF-Sequenzen. ....	154
--	-----

<b>Tabelle B.3:</b> Zusammenfassung des Qualitätsvergleichs der QCIF-Sequenzen. ....	155
--	-----

<b>Tabelle B.4:</b> Zusammenfassung des Qualitätsvergleichs der CIF-Sequenzen.. ....	156
--	-----

**Anhang C Ergänzungen zu SIMD-Prozessorerweiterungen in Kapitel 6**

<b>Tabelle C.1:</b> Beispiele für Bearbeitungszeiten verschiedener Pentium 4 Befehle [41], [46]. ...	162
--	-----

# 1 Einleitung

Als treibende Kräfte für die wachsende Marktrelevanz der Multimedia-Kommunikation können drei Faktoren herausgestellt werden: die Forderung nach ständiger Erreichbarkeit in Verbindung mit einer hohen Mobilität der Anwender und die Möglichkeit sich überall und jederzeit Informationen zu beschaffen. In der Vergangenheit ist eine qualitativ ansprechende Übertragung von Multimediainhalten oftmals an technischen, aber auch wirtschaftlichen Problemen gescheitert.

Die wesentlichen technischen Hindernisse waren dabei die zur Verfügung stehende Bandbreite der Kommunikationsnetze, die Leistungsfähigkeit der Endgeräte und nicht zuletzt die unzureichende Effizienz der Multimediatechnologiestandards. Durch die Einführung von neuen Übertragungstechnologien und der stetig wachsenden Anzahl von multimediafähigen Endgeräten, wie Mobiltelefonen, Handhelds, Notebooks und PCs (Personal Computer, PC), sind die Grundlagen geschaffen, qualitativ hochwertige Videokommunikation zu ermöglichen [4], [5], [26], [137].

Eine Ende-zu-Ende-Kommunikation von Endgerät A zu Endgerät B, wie in Abbildung 1.1 dargestellt, kann dabei über ein einziges oder auch mehrere verschiedene Netze erfolgen. Die folgenden Ausführungen beziehen sich, ebenso wie die Darstellung in Abbildung 1.1, im Schwerpunkt auf die Videocodierung.

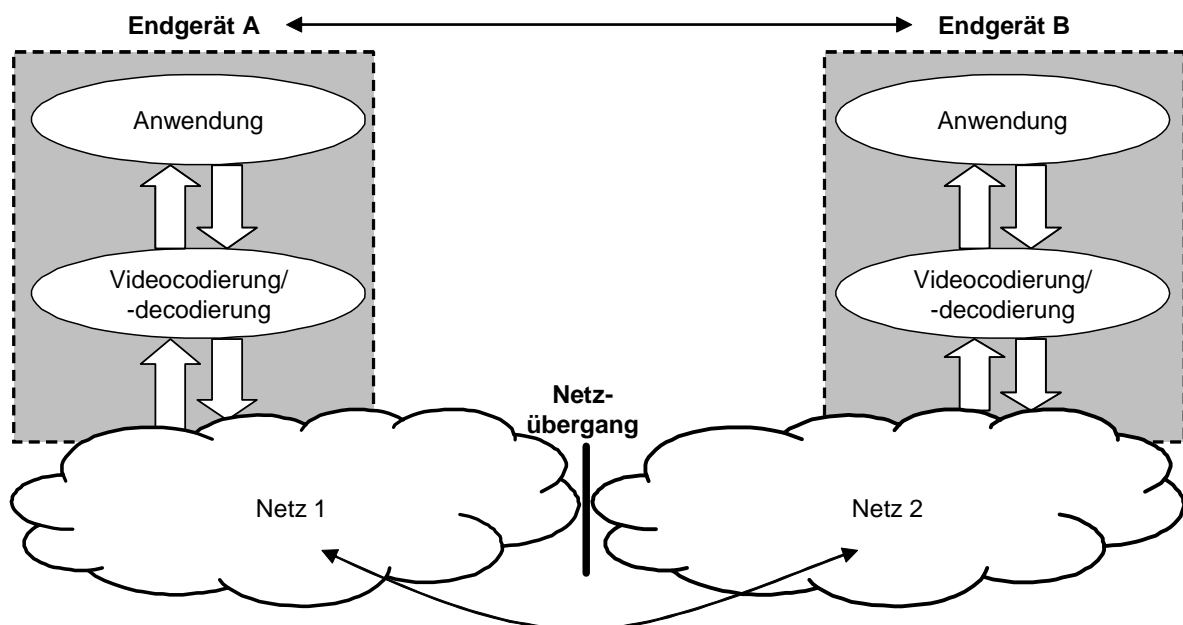


Abbildung 1.1: Ende-zu-Ende-Videokommunikation zwischen den Endgeräten A und B.

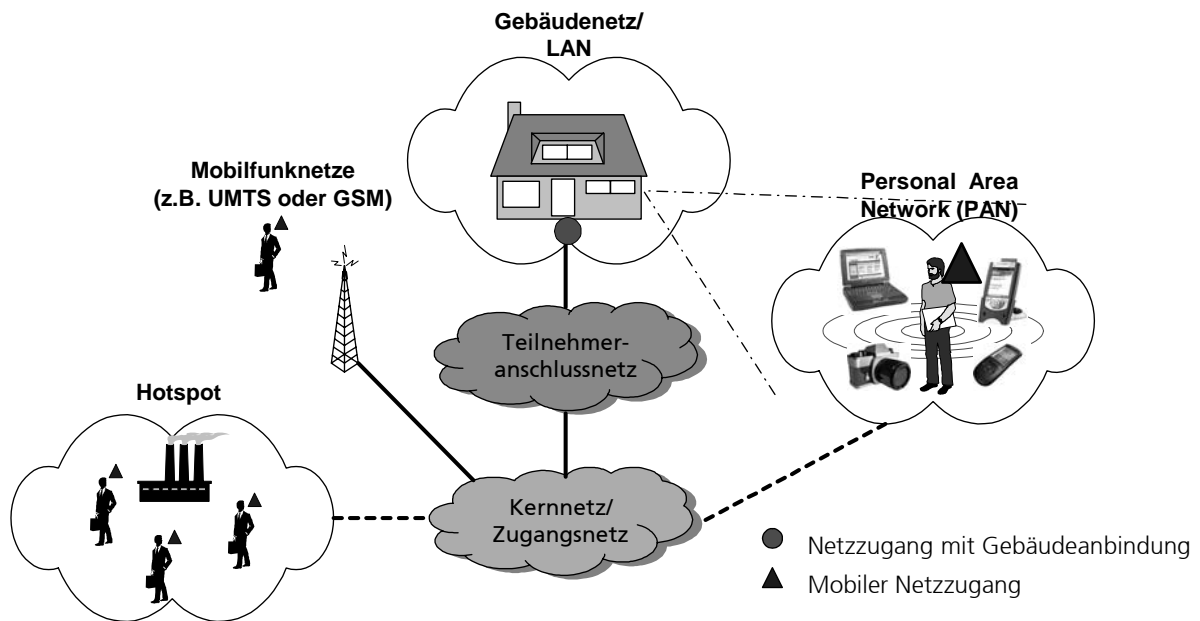
Ein Videokommunikationssystem setzt sich im Wesentlichen aus den Einzelkomponenten: Anwendung, Videocodec und Übertragungsnetz zusammen. Die drei Teilsysteme stehen dabei zueinander in folgender Beziehung:

- Die Anwendung definiert sowohl die Anforderungen an die Videocodierung als auch an das Transportnetz. Kriterien sind z.B. Qualitätsanforderungen, Vorgaben für die Ende-zu-Ende-Verzögerung aber auch die Charakteristik des Eingangssignals. Als Kriterien dienen dabei z.B. Anzahl und Geschwindigkeit von bewegten Bildinhalten, Detailreichtum oder Häufigkeit von Szenenwechseln.
- Der Videoencoder hat die Aufgabe, das Eingangssignal effizient zu codieren. Der Videocodec - der Encoder und der Decoder - stellt der Anwendung dabei zahlreiche Funktionalitäten, wie z.B. Zugriffsmöglichkeit auf einzelne Bildobjekte oder Streamingfähigkeit zur Verfügung. Dabei sind die Vorgaben, wie z.B. Anforderungen an die Verzögerung oder Vorgaben für die Datenrate, einzuhalten. Für die Videoübertragung ist es notwendig, dass der Videocodec an die Eigenschaften der Übertragung angepasst wird. Eine "netzfreundliche" Videocodierung unterstützt z.B. vereinfachtes Paketieren, Schnittstellen für priorisierende bzw. zeitvariante Netze, Skalierbarkeit und effektiven Fehlerschutz.
- Das Netz ist für die Übertragung der codierten Videosignale zuständig. Hierfür werden die Videodaten mittels Kommunikationsprotokollen über die jeweiligen Netze transportiert. Für die eigentliche Übertragung über den physikalischen Kanal muss der resultierende Protokollbitstrom kanalcodiert und anschließend moduliert werden.

Wie in Abbildung 1.1 dargestellt, kann eine Ende-zu-Ende-Kommunikation über ein einziges Netz oder aber über mehrere verschiedene Netze stattfinden. Die Eigenschaften des Übertragungsnetzes, wie Häufigkeit von Bitfehlern oder effektive Datenrate in Abhängigkeit von der Zeit, sind von entscheidender Bedeutung für die Auswahl und die Optimierung geeigneter Videocodierverfahren und eventuell notwendiger Fehlerschutzstrategien. So erfordern insbesondere drahtlose Übertragungsnetze eine hohe Robustheit der Videocodierung bezüglich Übertragungsfehlern und können meist keine konstante Datenrate für die Übertragung zur Verfügung stellen.

Eine Klassifikation der verschiedenen Übertragungsnetze ist anhand ihrer Aufgaben bzw. ihrer Lokalität in unterschiedliche Ebenen möglich. Eine beispielhafte Einordnung ist in Abbildung 1.2 dargestellt. Darin stellt das Kernnetz den Backbone eines Netzwerks dar, über den die globale Kommunikation abgewickelt wird. Das Zugangnetz bildet die Schnittstelle zwischen dem Kernnetz und dem Teilnehmeranschlussnetz ("letzte Meile"). Es stellt das Bindeglied zwischen dem Zugangnetz und dem Teilnehmer dar. Die einzelnen Endgeräte können direkt an das Teilnehmeranschlussnetz oder über ein Gateway an ein Gebäudenetz, ein LAN (Local Area Network) bzw. ein PAN (Personal Area Network) angebunden werden.

Des Weiteren ist auch eine mobile Anbindung über öffentliche Netze möglich, z.B. über GSM (Global System for Mobile Communication), einen Hotspot oder UMTS (Universal Mobile Telecommunication System). Für die Verteilung von Audio- oder Multimedialinhalten stehen zudem die mobilen Systeme DAB (Digital Audio Broadcast) und DVB (Digital Video Broadcast) zur Verfügung.



**Abbildung 1.2:** Klassifikation der verschiedenen Netze.

Der technologische Fortschritt ermöglicht die Nutzung verschiedener Übertragungsmedien und die dazu notwendigen leistungsfähigen Übertragungstechnologien. Aktuell stehen folgende Übertragungsmedien mit jeweils spezifischen Eigenschaften zur Verfügung:

- drahtlos (Satellitenkommunikation, Funk),
- drahtgebunden (Stromnetz, Koaxialkabel, Kupferdoppelader) und
- optisch (Glas- und Polymerfasern).

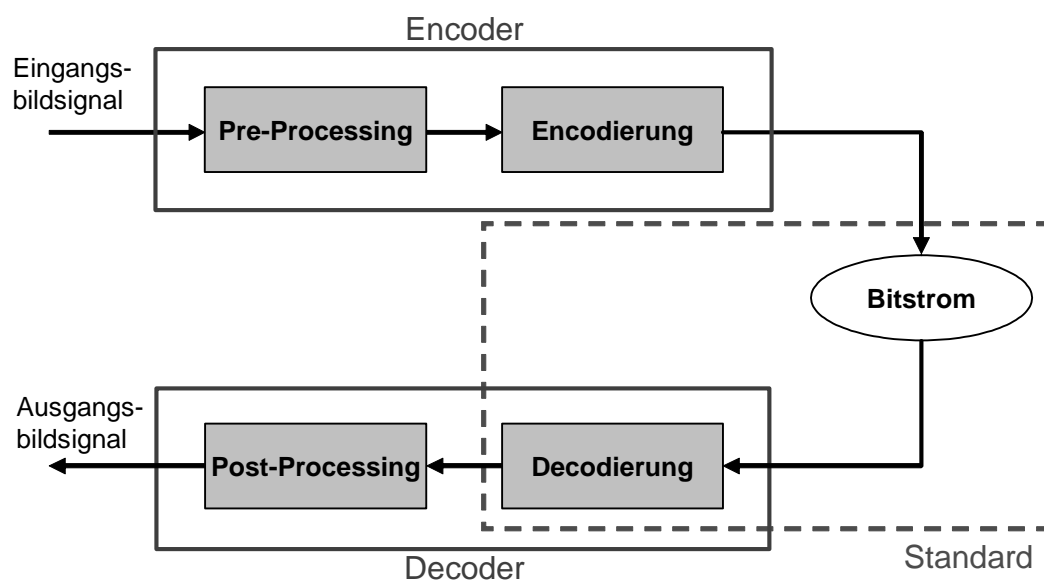
Jedes einzelne Netz stellt dabei in Verbindung mit dem jeweiligen physikalischen Kanal für den Datentransport eine unterschiedliche Dienstgüte (Quality of Service, QoS) zur Verfügung, so z.B. Datenrate, Fehlerrate, Häufigkeit von Paketverlusten, Verzögerung oder Jitter. Während klassische verbindungsorientierte Sprachnetze, wie z.B. ISDN (Integrated Services Digital Network), Datenraten bis maximal 64 kBit/s pro Kanal und eine hohe Dienstgüte zur Verfügung stellen, können drahtlose Netzwerke für den LAN-Bereich bis über 100 MBit/s zur Verfügung stellen. Je nach Qualität der Verbindung und der Anzahl der Kommunikationsteilnehmer kann aber die Dienstgüte erheblich schwanken.

In einigen Netzen, wie z.B. in zukünftigen IP-Netzen (Internetprotokoll) [37], [81] oder in speziellen Gebäudenetzwerken [136], ist es möglich QoS-Garantien auszuhandeln. Die Übertragungscharakteristik des Netzes kann damit an die Anforderungen der Anwendung und der Videocodierung angepasst werden.

## 1.1 Entwicklung der Videocodierverfahren

In der Vergangenheit sind für jeden Einsatzbereich eigene Videocodierverfahren standardisiert worden, so z.B. H.261 für die Videoübertragung in Konferenzen und MPEG-1 (Moving Pictures Expert Group) für Video-CDs. Aufgrund der Konvergenz der Telekommunikation, der Unterhaltungselektronik und dem allgegenwärtigen Einsatz von Computern haben aktuelle und vor allem zukünftige Standardisierungsbestrebungen, wie MPEG-4 [57], [66], [68], H.263 [73], [75], [76], [78] oder H.264 [71], [74], [128], die Schaffung flexibler Standards zum Ziel. In [82] treffend bezeichnet als "Visual Communications of Tomorrow: Natural, Efficient and Flexible". Die Zielsetzung ist, für unterschiedliche Anwendungsbereiche und der Übertragung über verschiedene Netze stets eine effektive Codierung auf Basis einheitlicher, adaptiver Verfahren sicherzustellen.

Die Videostandardisierung definiert die Struktur des Videodecoders und das Format des zu übertragenden Bitstroms. Der Decoder interpretiert den Bitstrom des Encoders, um damit das codierte Videosignal zu rekonstruieren (siehe Abbildung 1.3). Damit kann durch technische Neuerungen oder auch durch gezielte Optimierungen eine Qualitätsverbesserung erreicht werden. Zur Reduzierung der Komplexität können aber auch Vereinfachungen vorgesehen werden. Damit bestimmt überwiegend die jeweilige Realisierung des Videocodecs und nur bedingt der Videostandard selbst die unter bestimmten Bedingungen erzielbare Videoqualität.



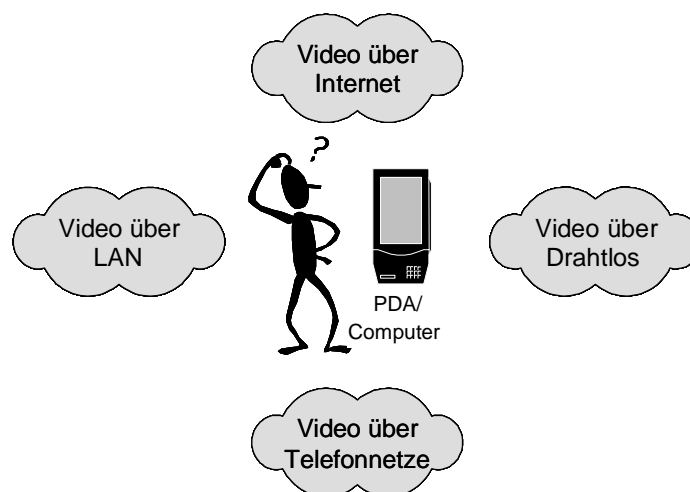
**Abbildung 1.3:** Die Videostandardisierung definiert das Bitstromformat und die Decoderstruktur.

Bei der Standardisierung von MPEG-4, aber auch in den Weiterentwicklungen von H.263 sind in umfangreichen Untersuchungen verschiedene Videocodiermethoden bewertet worden. Es hat sich gezeigt, dass das bisher eingesetzte Hybrid-DCT-Videocodierverfahren mit Bewegungskompensation (Diskrete Cosinus Transformation, DCT) auf Basis von H.263 einen sehr guten Kompromiss zwischen Aufwand und effektiven Nutzen darstellt [79], [96]. In aktuellen Videocodecs kann das Basisverfahren durch neue Funktionalitäten an unterschiedliche Einsatzbereiche angepasst werden, so z.B. durch zusätzliche Bewegungsmodelle.

Ebenso spielt die Vorverarbeitung des Videosignals (Pre-Processing) vor der eigentlichen Codierung eine wichtige Rolle. So z.B. die Erkennung von Szenenänderungen oder die getrennte Codierung und Übertragung des Szenenhintergrundes und des Vordergrundes. Der Videoencoder muss für eine effiziente Codierung - teilweise sogar im laufenden Betrieb - das entsprechende Vorgehen anhand der aktuellen Bildsignalcharakteristik bestimmen. Gleichzeitig muss die Codierung an die Übertragung über die unterschiedlichen, teilweise zeitvarianten Netze angepasst werden.

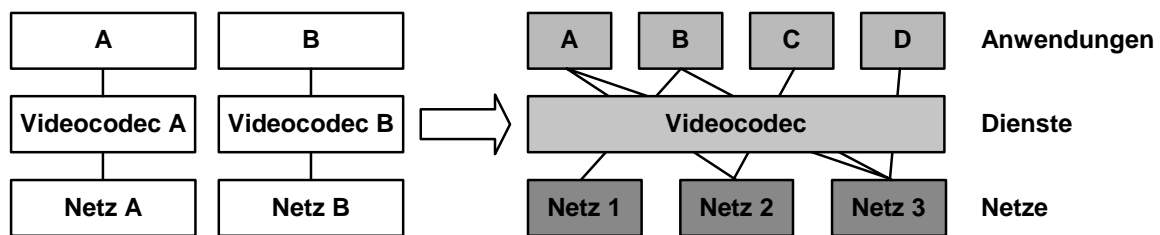
## 1.2 Anwendungsszenarios und deren Anforderungen

Die Einsatzmöglichkeiten eines flexiblen Videostandards können beispielhaft an drei Anwendungsszenarios dargestellt werden. Im ersten Fall sollen multimediafähige Endgeräte, wie z.B. ein Handheld oder WebPad, über jeweils unterschiedliche Netze eine Videokommunikation realisieren (siehe Abbildung 1.4). Im mobilen Betrieb ist die Anbindung z.B. über UMTS oder im lokalen Netz über IEEE 802.11 (Institute of Electrical and Electronics Engineers, Inc.) vorgesehen. Im stationären Betrieb dagegen über LAN. Bei jeder Anbindung muss der Videocodec in Abhängigkeit von der aktuellen Situation stets eine effiziente Codierung und Übertragung sicherstellen. Dazu muss er sich flexibel auf verschiedene - bei mobiler Anbindung auch auf stark variierende - Randbedingungen einstellen.



**Abbildung 1.4:** Kommunikationsmöglichkeiten eines mobilen Endgerätes. Jedes Netz stellt unterschiedliche Anforderungen an die Videocodierung.

Im zweiten Szenario wird der Videocodec, wie es aktuell bereits häufig realisiert ist, als Basisdienst in ein Betriebssystem integriert (siehe Abbildung 1.5). Damit kann der Codec von unterschiedlichen Anwendungen verwendet werden, so z.B. für eine Videokommunikation über ISDN und LAN, für das Versenden von Videomails oder für die Aufzeichnung von TV-Programmen. Im Gegensatz zu bisher ist der Videocodec damit nicht mehr fest mit einer einzigen Anwendung verbunden. Er kann aber auch nicht mehr anhand des vorgegebenen Einsatzbereiches gezielt entwickelt und optimiert werden, sondern muss sich adaptiv der jeweiligen Situation anpassen.



**Abbildung 1.5:** Der Videocodex steht als Basisdienst des Betriebssystems für unterschiedliche Anwendungen zur Verfügung. Die Übertragung der Daten erfolgt über verschiedene Netze.

Das dritte Beispiel ist ein Videocodex, der für unterschiedliche Plattformen und Anwendungen realisiert werden soll. Die Implementierung des Codex soll soweit wie möglich unabhängig von dem jeweiligen Einsatzbereich realisiert werden. Erst in einem späteren Schritt soll mit minimalem Aufwand eine Anpassung an die Zielplattform und die jeweilige Anwendung erfolgen. Damit entfällt der Aufwand, den Videocodex jeweils neu optimieren zu müssen.

Anhand der Beispiele werden der Bedarf und die neuen Möglichkeiten von flexiblen Videostandards offensichtlich. Für Kommunikationsanwendungen, die z.B. auf H.263, H.264 oder MPEG-4 basieren, lassen sich anhand der definierten Einsatzbereiche und der Vorgaben aus der Standardisierung folgende Anforderungen definieren [59], [62], [64], [65], [70]: Einsatz von Bildmaterial in verschiedenen Formaten bzw. mit unterschiedlicher Bildstatistik (z.B. Detailreichtum und Anzahl bzw. Geschwindigkeit von bewegten Objekten) bei Datenraten von 10 kBit/s - 2 MBit/s und Bildwiederholraten von 7,5 - 30 fps (frames per second). Die je nach Anwendung geforderte maximale Ende-zu-Ende-Verzögerung liegt zwischen 0,2 - 2 s. Weiterhin müssen neben konstanten auch variable bzw. abschnittsweise konstante Datenraten berücksichtigt werden.

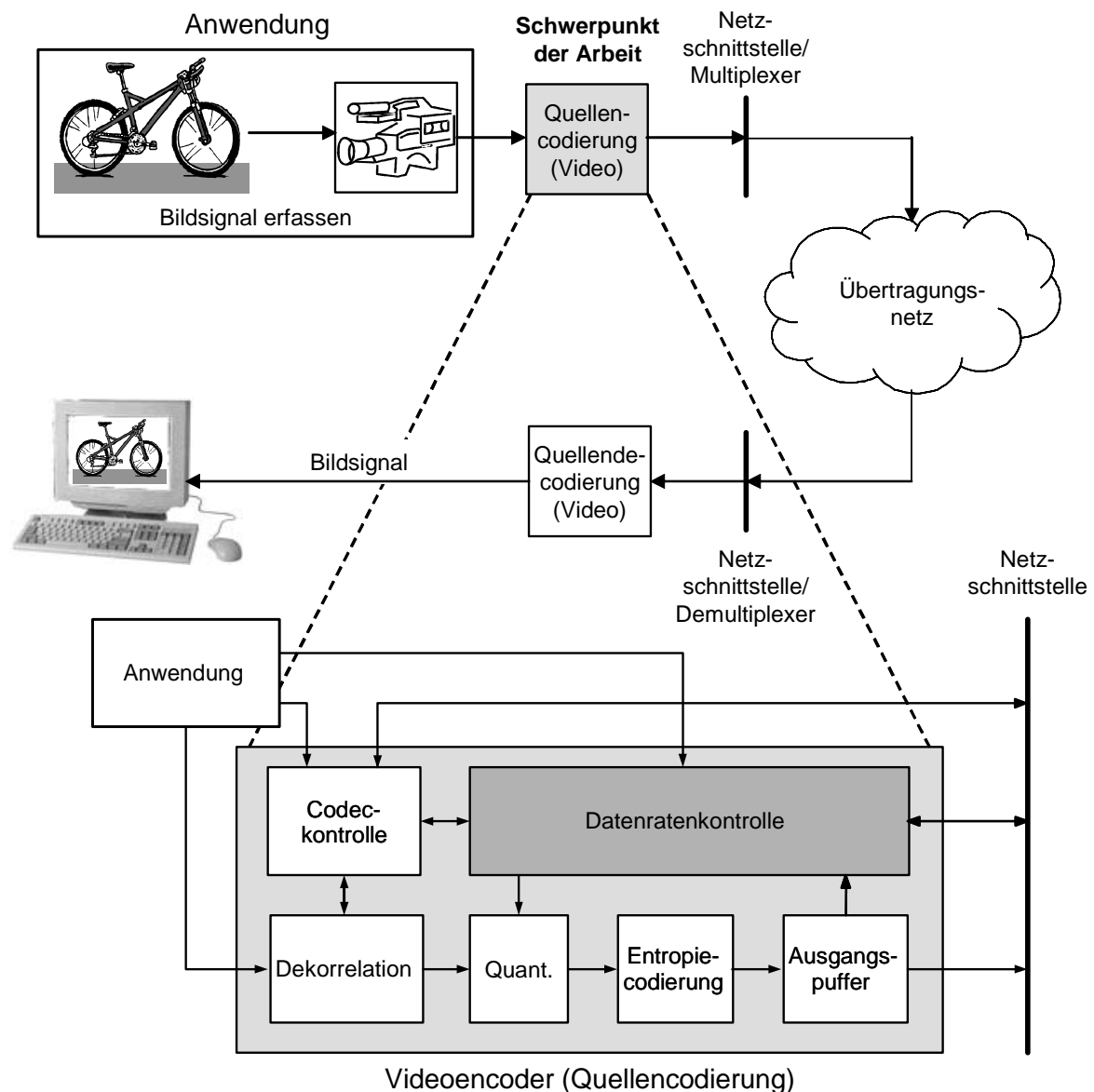
Für die Auslegung des Videoencoders, dessen Arbeitsweise die Codiereffizienz und die Ende-zu-Ende-Verzögerung festlegt, ergeben sich viele Einstellmöglichkeiten. Allerdings entfällt aufgrund der notwendigen Flexibilität die bisher praktizierte Möglichkeit, den Videocodex im Voraus auf feste Randbedingungen zu optimieren. Vielmehr müssen unterschiedliche Anforderungen - teilweise im laufenden Betrieb - mit einem einzigen Videocodex effizient verarbeitet werden können. Gleichzeitig ist dabei die Übertragung über mehrere, teilweise zeitvariante Netze sicherzustellen.

Die Datenratenkontrolle hat im Videoencoder die Aufgabe, die Ausgangsdatenrate an die verfügbare Datenrate des Netzes anzupassen, um eine verzögerungsbegrenzte Ende-zu-Ende-Kommunikation sicherzustellen. Dabei müssen die Vorgaben des Videostandards und der Anwendungen eingehalten werden. Durch die flexiblen Videocodexes ergeben sich für die Datenratenkontrolle als qualitätsbestimmendes Element im Encoder neue Anforderungen.

### 1.3 Struktur und Aufbau der Arbeit

Die Zielsetzung dieser Arbeit ist die Untersuchung und Optimierung von neuen Verfahren für die Datenratenkontrolle in Videocodexes mit dem Einsatzbereich Videokommunikation. Der Schwerpunkt liegt auf der Unterstützung von Hybrid-DCT-Codexes, wie sie H.263 und MPEG-4 zu Grunde liegen. Insbesondere die Anforderungen des breiten Einsatzbereichs und der Übertragung über verschiedene Netze stellen hohe Ansprüche an die Datenratenkontrolle. Die Einordnung der Thematik in ein Ende-zu-Ende-Kommunikationssystem ist in Abbildung 1.6 dargestellt.





**Abbildung 1.6:** Einordnung der Datenratenkontrolle im Videoencoder in ein Ende-zu-Ende-Videokommunikationssystem.

Im Kapitel 2 wird auf die Videocodierung und -standards eingegangen. Die Grundlagen der Videocodierung bilden die Basis für die weiteren Ausführungen. Im Anschluss wird die Datenratenkontrolle im Videoencoder beschrieben. Die Aufgabe der Datenratenkontrolle ist es, die Ausgangsdatenrate des Videoencoders effizient an die Vorgaben des Decoders, des Übertragungssystems und der Anwendung anzupassen. Eine vorgegebene Ende-zu-Ende-Verzögerung darf dabei nicht überschritten werden. Neben der Funktionsweise und den Anforderungen an die Verfahren für die Datenratenkontrolle wird auf die Rate-Distortion-Theorie [111] eingegangen. Daran schließt sich die Beschreibung bestehender Lösungsansätze an. Dabei wird beispielhaft auf existierende Verfahren aus der Standardisierung von H.263 und MPEG-4 eingegangen.

Der Schwerpunkt des Kapitels 3 ist - aufbauend auf dem aktuellen Stand der Technik - die Konzeption einer Datenratenkontrolle für flexible Videokommunikation. Zielsetzung ist die Unterstützung eines breiten Einsatzbereichs und der Übertragung über verschiedene Netze. Hierzu werden die Anforderungen und damit auch die Randbedingungen, die sich hieraus an die Datenratenkontrolle ergeben, ermittelt. Anhand der Rate-Distortion-Charakteristik unterschiedlicher Sequenzen wird die Eignung existierender Rate-Distortion-Modelle bewertet. Zugleich werden Verbesserungsansätze aufgezeigt. Anhand der Ergebnisse wird eine Funktionsspezifikation auf Systemebene für einen neuen Ansatz für die Datenratenkontrolle aufgestellt.

In Kapitel 4 wird ein neues Verfahren für die Datenratenkontrolle vorgeschlagen. Aufgrund seiner Flexibilität erfüllt es die Anforderungen verschiedenster Anwendungsbereiche und Übertragungsnetze im Bereich Echtzeitvideokommunikation. Das neue Verfahren beruht sowohl auf einer genaueren Schätzung der Verzerrung (Distortion) und der Bitrate als auch auf einem Ansatz zur verallgemeinerten Beschreibung von nieder- bis hochbitratigen Videosignalen. Zudem wird ein neuer Schätzparameter auf Basis der DCT-Koeffizienten definiert. Wichtig für die Leistungsfähigkeit des neuen Verfahrens ist die Anpassung an das aktuelle Bildsignal und eine automatische Systemparametrisierung, um in allen Bereichen eine stabile und hochwertige Bildqualität sicherstellen zu können. Im Anschluss wird auf wichtige Erweiterungsmöglichkeiten des neuen Verfahrens, wie sie bei aktuellen Videocodecs optional gefordert werden, eingegangen: so z.B. Modellierung von unterschiedlichen Methoden für die Codierung von Bildern bzw. Makroblöcken, Unterstützung von mehreren Bildobjekten und die Einbeziehung der subjektiven Bildqualität.

Den Schwerpunkt des Kapitels 5 bildet die Verifikation des im vorherigen Abschnitt vorgeschlagenen neuen Verfahrens für die Datenratenkontrolle. Nach der Einführung werden die Testbedingungen für die Untersuchungen beschrieben. Sie bilden die Basis für die Bestimmung der Leistungsfähigkeit des neuen Verfahrens im Vergleich zu den Referenzverfahren aus der Standardisierung von H.263 und MPEG-4. Neben der Ermittlung der Codiereffizienz für verschiedene Bitratenbereiche werden für weitere Auswertungen auch die Ergebnisse aus einem Testset untersucht und diskutiert. Für die Ermittlung des Verhaltens unter verschiedenen Bedingungen sind hierfür 38 Einzeltests zusammengefasst. Zusätzlich wird der Einfluss verschiedener Implementierungsvarianten auf die Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle analysiert. Anschließend werden die einzelnen Ergebnisse zusammengefasst und bewertet.

Im Kapitel 6 wird auf die Realisierungsaspekte von echtzeitfähigen Videocodecs eingegangen. In diesem Zusammenhang werden die Grundlagen von Multimedia-Prozessorerweiterungen beschrieben. Anschließend wird sowohl auf deren Einsatzmöglichkeiten im Videocodec und speziell in der Datenratenkontrolle eingegangen als auch das sich daraus ergebende Beschleunigungspotenzial diskutiert. In Kommunikationsanwendungen ist neben der Codiereffizienz die systembedingte Ende-zu-Ende-Verzögerung ein wichtiges Kriterium. Aus diesem Grund wird in Kapitel 6 der Aufbau des Videocodecs inklusive der für die Bearbeitung notwendigen Puffer analysiert. Darauf aufbauend werden Ansätze zur Minimierung der Verzögerung untersucht und hinsichtlich der erzielbaren Videocodiereffizienz bewertet.

Die Zusammenfassung und die Diskussion wichtiger Ergebnisse befinden sich zusammen mit dem Ausblick in Kapitel 7.

## **2 Grundlagen der Datenratenkontrolle in Videocodecs**

In diesem Kapitel wird auf Videostandards und die Datenratenkontrolle als Funktionsmodul im Videoencoder eingegangen. Hierzu werden im Abschnitt 2.2 die Grundlagen der Videocodierung und deren Umsetzung in standardisierte Videocodecs als Basis für die weiteren Ausführungen beschrieben. Anschließend wird im Abschnitt 2.3 das Prinzip der Datenratenkontrolle dargestellt. Wichtige Aspekte sind dabei die Anforderungen an die Datenratenkontrolle und deren Funktionsweise. Zum Schluss werden existierende Lösungsansätze für die Datenratenkontrolle analysiert und bewertet. Die wichtigsten Inhalte und Aussagen dieses Kapitels sind im Abschnitt 2.4 zusammengefasst.

### **2.1 Einführung**

Die Einführung digitaler Übertragungstechnologien, wie ISDN, DSL (Digital Subscriber Line), WLAN (Wireless Local Area Network) und UMTS, und die Verfügbarkeit von multimediafähigen Endgeräten haben die Einsatzmöglichkeiten der Videokommunikation wesentlich erweitert. In neuen Multimediastandards konnte sowohl die Videocodiereffizienz als auch die Funktionalität deutlich gesteigert werden. Ein weiteres Kennzeichen aktueller Standards ist die Unterstützung unterschiedlicher Anwendungsbereiche und die Übertragung über verschiedene Netze.

Mit der wachsenden Flexibilität der Videocodierung erhöht sich jedoch der Kontroll- und Steuerungsanfang insbesondere im Encoder. Damit werden an die Datenratenkontrolle hohe Anforderungen gestellt. Ihre Aufgabe ist es im Zusammenspiel mit der Codeckontrolle den Videocodierprozess an die Anforderungen unterschiedlicher Anwendungen, aber auch an die aktuelle Netzumgebung adaptiv anzupassen.

### **2.2 Prinzipien der Videocodierung**

Videocodierverfahren verarbeiten das durch Abtastung und Quantisierung digitalisierte analoge Bildsignal, das aus einer Abfolge von zeitlich sortierten Einzelbildern besteht. Die Eigenschaften der bei der Videokommunikation häufig eingesetzten digitalen Bildformate QCIF (Quarter Common Intermediate Format) und CIF (Common Intermediate Format) sind in Tabelle 2.1 aufgelistet. Zum Vergleich ist das CCIR 601 (Comité Consultatif International des Radiocommunications) als digitales Fernsehformat aufgeführt. Für die Farbkomponenten-Darstellung werden dabei zusätzlich zur Helligkeitskomponente Y (Luminanz) die zwei Farbdifferenzkomponenten U und V (Chrominanz) bereitgestellt (YUV-Format).

	QCIF	CIF	CCIR 601
<b>Abtastparameter</b>			
Bildpunkte/Zeile (Y)	176	352	720
Anzahl der Zeilen (Y)	144	288	576
Bildpunkte/Zeile (U, V)	88	176	360
Anzahl der Zeilen (U, V)	72	144	576
Bildseitenverhältnis	4:3	4:3	4:3
Typische Bildfolgefrequenz [Hz]	7,5 - 15	10 - 30	25
<b>Datenraten</b>			
Datenmenge eines Einzelbildes [kByte] bei 8 Bit-PCM (Pulse Code Modulation)	38,02	152,1	829,4
Datenraten Bildsequenz bei 8 Bit-PCM [MBit/s]	2,28 - 4,56	12,17 - 36,5	165,9

**Tabelle 2.1:** Eigenschaften der digitalen Bildformate QCIF und CIF (europäische Werte) [96].

Die Videocodierung hat die Aufgabe das digitale Videosignal möglichst effizient zu beschreiben. Ziel ist es bei vorgegebener Bildqualität eine möglichst geringe Datenrate für die Übertragung zu benötigen (Codiereffizienz). Für die Reduzierung der Datenmenge wird überflüssige Information (Redundanz) entfernt.

Die Verfahren der Videocodierung basieren dabei auf Modellen, die sich aus der Informationstheorie ableiten und meist die statistischen Eigenschaften der Bildsignale ausnutzen, siehe z.B. [24], [25], [27], [102] und [111]. Grundsätzlich kann man zwischen der statistischen und der subjektiven Redundanz, auch Irrelevanz genannt, unterscheiden. Die statistische Redundanz ergibt sich aus der Korrelation der Bilddaten, wie der Abhängigkeit von räumlich und zeitlich benachbarten Bildpunkten. Man kann sie entfernen, ohne dass Information verloren geht, sodass die Daten in ihrer ursprünglichen Form wieder zurückgewonnen werden können. Dagegen ist die subjektive Redundanz durch Bildinformationen gekennzeichnet, für die der menschliche Betrachter weniger empfindlich ist. Im Idealfall wird ein Fehlen dieser Information vom menschlichen Betrachter nicht wahrgenommen. Im Gegensatz zur statistischen Redundanz können die ursprünglichen Bilddaten nur mit Fehlern rekonstruiert werden.

### 2.2.1 Beurteilung von Videocodierverfahren

Für die Bewertung und die Optimierung von Videocodierverfahren ist es notwendig, die durch die Codierung verursachte Verzerrung zu messen. Für eine allgemein gültige Bewertung müssen eine Vielzahl von psychooptischen Kriterien berücksichtigt werden, so z.B. Verdeckungseffekte, Abhängigkeit von der Ortsfrequenz bzw. frequenzabhängige Fehlergewichtung und Lokalität von Fehlerhäufungen [28], [96]. Im Moment werden daher für Echtzeitanwendungen meist vereinfachte Modelle, die zudem auf bestimmte Einsatzbereiche optimiert sind, eingesetzt.

Oftmals werden daher vereinfachte (statistische) Verzerrungsmaße verwendet, so Sum of Absolut Differences (*SAD*) nach (2.1), Mean Absolut Differences (*MAD*) nach (2.2), Sum of Squared Differences (*SSD*) nach (2.3) oder Mean Squared Error (*MSE*) nach (2.4). Für Vergleiche verschiedener Videocodierverfahren wird oft Peak Signal to Noise Ratio (*PSNR*) nach (2.5) verwendet. Die Maße sind dabei für ein Rechteck  $R$  mit  $M \times N$  Pixel für das Originalbild  $x$  und das rekonstruierte Bild  $y$  mit einer maximalen Amplitude  $A$  definiert:

$$SAD(x_R, y_R) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |x_R(m, n) - y_R(m, n)|^1 \quad (2.1)$$

$$MAD(x_R, y_R) = \frac{1}{|M \cdot N|} SAD(x_R, y_R) \quad (2.2)$$

$$SSD(x_R, y_R) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |x_R(m, n) - y_R(m, n)|^2 \quad (2.3)$$

$$MSE(x_R, y_R) = \frac{1}{|M \cdot N|} SSD(x_R, y_R) \quad (2.4)$$

$$PSNR(x_R, y_R)[dB] = 10 \cdot \log_{10} \left( \frac{A^2}{MSE(x_R, y_R)} \right) \quad (2.5)$$

Für die Bewertung einer ganzen Videosequenz mit  $N$  Bildern (Frames) wird der Mittelwert der *PSNR*-Werte der einzelnen codierten Bilder mit Bezug zu den entsprechenden Originalbildern berechnet. Dabei werden allerdings die Qualitätsschwankungen zwischen aufeinander folgenden Bildern nicht berücksichtigt. In [139] wird hierzu beispielsweise eine modifizierte *PSNR*-Berechnung für eine Videosequenz mit  $N$  Bildern eingeführt, bei der die Differenz der *PSNR*-Werte des Luminanzanteils des Bildes  $j$  zum vorherigen Bild  $j - 1$  berücksichtigt wird.

Neben der Verzerrung sind für die Beurteilung von Videocodierverfahren folgende Aspekte wichtig:

- die für eine bestimmte Verzerrung notwendige Datenrate,
- die systembedingte Verzögerung des Coders,
- die Robustheit des Verfahrens und
- die Komplexität des Codierverfahrens.

Mit höherer Komplexität des Videocoders oder einer höheren Verzögerungszeit ist es im Allgemeinen möglich, die Videocodiereffizienz zu erhöhen, d.h. bei gleicher Datenrate<sup>2</sup> eine geringere Verzerrung zu erzielen oder aber bei gleicher Verzerrung eine geringere Datenrate.

---

1.  $R$ : Bezug auf einen rechteckigen Bildausschnitt.

2. Der Begriff "Bitrate" stellt den Bezug auf die einzelnen Bit-Werte in einem Bitstrom her. Dagegen bezieht sich der Begriff "Datenrate" auf aggregierte Bitwerte. Oftmals werden die beiden Begriffe aber auch in der gleichen Bedeutung benutzt.

## 2.2.2 Videocodiermethoden

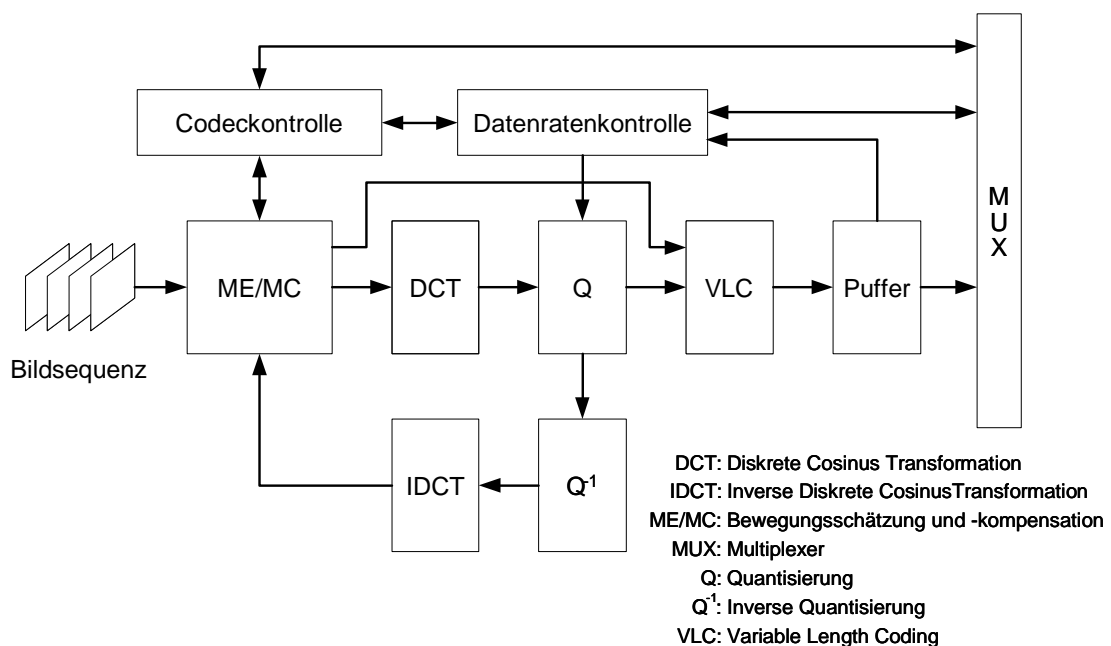
Das zentrale Ziel der Videocodierung ist die Reduktion der Datenmenge für die Übertragung oder Speicherung. Während der letzten 40 Jahre ist eine Vielzahl von Algorithmen basierend auf der Informationstheorie entwickelt worden. Das Funktionsprinzip dieser Algorithmen für die Videocodierung basiert dabei auf folgenden Methoden [96]:

- Entropiecodierung: Es ist eine verlustfreie Codierung, bei der die statistische Redundanz aus dem Videosignal entfernt wird, so z.B. die Lauflängencodierung (Variable Length Coding, VLC) oder die arithmetische Codierung.
- Skalare Quantisierung: Zur verlustbehafteten Reduzierung der Datenrate wird das Bildsignal oder seine dekorrelierte Äquivalente auf einen kleineren Codesymbolvorrat abgebildet.
- Vektorquantisierung: Anwendung eines Blockcodes auf wertkontinuierliche oder -diskrete Quellensignale, wobei der Blockcode verzerrungsbehaftet ist. Durch die blockbasierte Arbeitsweise ergibt sich im Vergleich zu der skalaren Quantisierung ein größeres Maß an Freiheitsgraden für die Auswahl von Rekonstruktionswerten.
- Prädiktive Differenzcodierung: Entfernt die zeitliche oder örtliche Redundanz des Bildsignals, indem die Differenzwerte zwischen Bildpunkten berechnet und codiert werden.
- Transformationscodierung: Aus dem Bildsignal wird durch eine lineare Transformation oder lineare Filterung eine dekorrelierte spektrale Repräsentation berechnet. Bei einer Transformationscodierung, wie z.B. der 2D-DCT-Transformation, erfolgt dies durch eine blockseparat arbeitende, zweidimensionale orthogonale Transformation.
- Bewegungskompensierte Prädiktion: Nutzt die zeitliche Redundanz zwischen benachbarten Bildern einer Videosequenz aus. Bei der bewegungskompensierten Prädiktion wird ein Bild durch Bewegungsvektoren auf ein oder mehrere vorherige Bilder und ein Fehlersignal beschrieben. Das Fehlersignal entspricht der Abweichung zwischen dem Original- und dem bewegungskompensierten Rekonstruktionsbild.
- Örtliche und/oder zeitliche Unterabtastung: Bei der örtlichen Unterabtastung wird das zu übertragende Bildsignal gegenüber dem Eingangssignal in der Auflösung reduziert. Nach der Übertragung wird es für die Darstellung gewöhnlich wieder auf die ursprüngliche Auflösung skaliert. Die zeitliche Unterabtastung reduziert die Bildwiederholrate des Eingangssignals vor der Codierung und Übertragung. Zur Darstellung werden die fehlenden Bilder durch Interpolation oder einfache Bildwiederholungen rekonstruiert.
- Fraktale Codierung: Nutzt die Selbstähnlichkeit der Bildsignale aus, wobei für die einzelnen Bildbereiche eine fraktale Transformation gefunden werden muss, die die Selbstähnlichkeit ausreichend beschreibt. Der Decoder wendet dann die fraktale Transformation iterativ auf ein beliebiges Ausgangsbild an, um das ursprüngliche Bild zu rekonstruieren.
- Inhaltsorientierte, objektorientierte und semantische Codierung: Die objektorientierte Codierung geht von einer Kontinuität einzelner Objekte im Bild aus und überträgt für die einzelnen Objekte zusätzlich Kontur und Umriss. Bei Codierverfahren mit räumlicher 3D-Modellierung und bei semantischer Codierung wird zusätzlich noch das Wissen über die Bildinhalte ausgenutzt, sodass sich hier die höchsten Kompressionsraten erzielen lassen.

Der kombinierte Einsatz von unterschiedlichen Codierverfahren wird in der Videocodierung als hybride Codierung bezeichnet. So bildet z.B. die Kombination aus einer zeitlich prädiktiven Bewegungskompensation mit einer Transformationscodierung des Prädiktionsfehlers die Basis für das Funktionsprinzip von H.263 und MPEG-4 [57], [66], [68], [73], [75], [76], [78]. Im Folgenden wird dieses Verfahren als Hybrid-DCT-Videoencoder mit Bewegungskompensation bezeichnet und bildet die Grundlage für die weitere Ausführung im Rahmen dieser Arbeit [79], [96], [138].

### 2.2.3 Hybrid-DCT-Videoencoder mit Bewegungskompensation

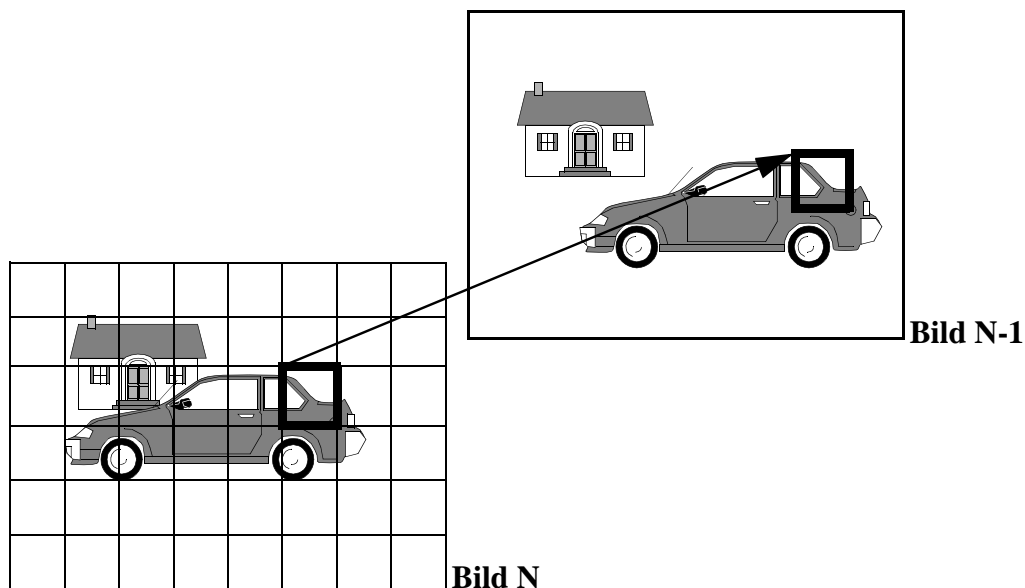
Das Grundprinzip für die zeitlich prädiktive Bewegungskompensation (Motion Compensation, MC) mit einer Transformationscodierung des Prädiktionsfehlers ist in Abbildung 2.1 dargestellt.



**Abbildung 2.1:** Schematischer Aufbau eines Hybrid-DCT-Videoencoders mit zeitlich prädiktiver Bewegungskompensation und Transformationscodierung des Prädiktionsfehlers.

Für die Bearbeitung unterteilt man das zu codierende Bild meist in 16 x 16 Pixel-Blöcke, die so genannten Makroblöcke (MB). Ein Makroblock besteht aus 4 Luminanz-Blöcken (Y-Blöcke) und aufgrund der Unterabtastung der Farbinformation je einem Block für die U- und V-Farbinformation (siehe Tabelle 2.1). Bei der Bewegungsschätzung (Motion Estimation, ME) wird für jeden Block des Bildes ein Bewegungsvektor (Motion Vector, MV) gesucht, der auf denjenigen Bereich im vorhergehenden Bild zeigt, der die kleinste Abweichung zum aktuellen Block aufweist.

Das Prinzip der zeitlich prädiktiven Bewegungsschätzung ist in Abbildung 2.2 dargestellt. Aktuelle Verfahren sind nicht fix auf die Makroblockgröße festgelegt, sondern arbeiten in Abhängigkeit vom Inhalt z.B. auch mit 4 x 4 Pixel-Blöcken. Zusätzlich kann die Genauigkeit erhöht werden, indem die Suche auch auf den Subpixelbereich erweitert wird, so z.B. 1/2-Pixel (half-pel) oder 1/4-Pixel (quarter-pel). Die Steuerung der Bewegungssuche ist ebenso die Aufgabe der Codecsteuerung bzw. -kontrolle, wie die Festlegung der restlichen freien Parameter des Encoders.



**Abbildung 2.2:** Grundprinzip der blockbasierten Bewegungsschätzung.

Im nächsten Arbeitsschritt, der Bewegungskompensation (MC), wird ein so genanntes Prädiktionsbild erstellt, das aus den Verweisen der Bewegungsvektoren zusammengesetzt wird. Die Bewegungskompensation arbeitet ebenso wie die Bewegungsschätzung entweder mit ganzzahligen Vielfachen des Pixel-Abstandes oder mit Zwischenpixelschritten.

Für die Abweichungen des Prädiktionsbildes zum aktuellen Bild werden mithilfe der 2D-DCT-Transformation die Spektralkoeffizienten für die einzelnen  $8 \times 8$  Blöcke berechnet. Durch diese Vorgehensweise ist es möglich, bereits übertragene Bildinformation des vorherigen Bildes durch Angabe von Bewegungsvektoren zu referenzieren und nur noch den Prädiktionsfehler zu codieren.

In einem weiteren Bearbeitungsschritt werden die DCT-Koeffizienten variabel quantisiert, wobei die Genauigkeit und damit auch der Informationsgehalt der Darstellung variiert werden kann. Der jeweilige Quantisierungsparameter kann von der Datenratenkontrolle (RateControl, RC) festgelegt oder fix vorgegeben werden. Zum Schluss folgt eine Lauflängencodierung.

Der Multiplexer (MUX) ist nicht mehr Bestandteil des Videoencoders, sondern bildet das Bindeglied zwischen Netz und Videocodec. Seine Aufgabe ist unter anderem die Synchronisation verschiedener Audio- und Videodatenströme untereinander.

Für die Bewegungskompensation des nächsten Bildes ist es notwendig, nach der Quantisierung alle Bildblöcke invers zu quantisieren ( $Q^{-1}$ ) und die Spektralkoeffizienten mittels der Inversen Diskreten Cosinus Transformation (IDCT) in den Ortsbereich zurück zu transformieren. Zusammen mit dem aktuellen Prädiktionsbild erhält man das rekonstruierte Ausgangsbild.



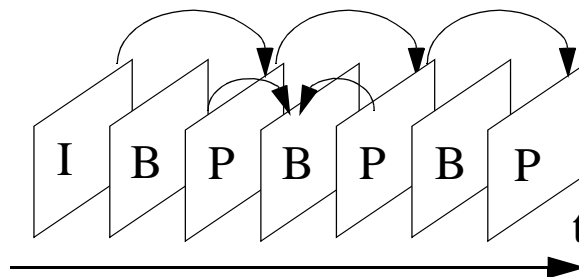
Im Decoder laufen im Wesentlichen für die Rekonstruktion der Bilder aus dem Bitstrom folgende Arbeitsschritte ab: Variable Length Decoding (VLD), inverse Quantisierung der DCT-Koeffizienten, IDCT und die Decodierung der Bewegungsvektoren. Auf Basis des vorherigen, rekonstruierten Bildes, der Bewegungsvektoren und des nach der IDCT vorliegenden Fehlerbildes kann das aktuelle Bild rekonstruiert werden.

Insgesamt können beim Hybrid-DCT-Codierverfahren drei verschiedene Verfahren angewendet werden, um die Einzelbilder einer Bildsequenz zu codieren (siehe Abbildung 2.3):

**I-Bild (Intra Frame):** Sämtliche Bildinformation wird mittels DCT-Koeffizienten gespeichert, wie es z.B. für das erste Bild einer Sequenz notwendig ist (Intracodierung).

**P-Bild (Predicted Frame):** Information wird durch Verweise auf das vorherige Bild und mittels DCT für die Fehlerkompensation der Prädiktion gespeichert (Intercodierung).

**B-Bild (Bidirectional Predicted Frame):** Ähnlich dem P-Bild, allerdings wird der Inhalt durch Interpolation zwischen vorherigem und zukünftigem Bild vorhergesagt. Die Abweichungen werden ebenfalls wieder mittels DCT codiert.



**Abbildung 2.3:** Zeitliche Abfolge der P- und B-Bilder mit einem I-Bild als Ausgangsbasis für die Bewegungskompensation.

In Kommunikationsanwendungen wird meist aufgrund der notwendigen Interaktivität eine geringe Verzögerung gefordert, sodass auf den Einsatz von B-Bildern verzichtet wird. Meist wird nur das erste Bild intracodiert (I-Bild) und die restlichen als P-Bilder übertragen (folgend als IPPP...-Codierung bezeichnet). In Abhängigkeit vom jeweiligen Bildsignal ist die Codiereffizienz in der Regel bei P-Bildern höher als bei I-Bildern.

Insbesondere bei höherbitratigen Anwendungen mit geringen Anforderungen bezüglich der Ende-zu-Ende-Verzögerung, wie z.B. bei der Fernsehübertragung, gibt es zwischen den P- und B-Bildern in regelmäßigen Abständen I-Bilder. Damit kann sich der Decoder beispielsweise bei Fehlern im Bitstrom einfach resynchronisieren.

In P- und B-Bildern kann von der Codeckontrolle pro Makroblock die genaue Codierungsart festgelegt werden. So kann z.B. ein Makroblock, dessen Bildinhalt nicht im vorherigen Bild enthalten ist, intracodiert werden. In diesem Fall wird also keine Bewegungsinformation übertragen. Ebenso können Makroblöcke, die in beiden Bildern identisch sind, übersprungen werden (skipped makroblocks).

## 2.2.4 Standardisierung im Bereich Videocodierung

Die Standardisierung eines Videocodecs gewährleistet, dass verschiedene Geräte auch von unterschiedlichen Herstellern miteinander kommunizieren können. Innerhalb der Standardisierung liegt der Schwerpunkt auf der Definition des Decoders und des Bitstroms, den der Decoder interpretiert, um das Videosignal zu rekonstruieren (siehe Kapitel 1). Dadurch wird ermöglicht, dass z.B. Verbesserungen im Encoder realisiert werden können.

Unter diesem Aspekt wird die softwarebasierte (Teil-)Realisierung von Videocodecs immer wichtiger. Im Gegensatz zu reinen Hardware-Lösungen können hierdurch sowohl die für aktuelle Standards notwendige Flexibilität als auch der Einsatz verbesserter Methoden für unterschiedliche Einsatzbereiche gewährleistet werden. Für die Auslegung des Videoencoders, dessen Arbeitsweise die Effizienz der Quellencodierung und die Ende-zu-Ende-Verzögerung festlegt, bleiben viele Freiheiten. Je nach technischem Aufwand und Know-how können sich einzelne Implementierungen hinsichtlich ihrer Codiereffizienz stark unterscheiden. Innerhalb der Standardisierung werden meist für die Realisierung des Videocodecs Empfehlungen gegeben. Bei der MPEG-Standardisierung werden sogar komplette softwarebasierte Referenzcodecs zur Verifikation der Funktionalität implementiert. Sie bilden die Basis für Vergleichstests von unterschiedlichen Algorithmen.

In der Vergangenheit sind während des Standardisierungsprozesses der ITU-T<sup>1</sup> bzw. der ISO/IEC<sup>2</sup> für den jeweiligen Einsatzzweck sehr konkrete Anforderungen aufgestellt und der neue Codec darauf ausgerichtet worden. So hat H.120 bzw. H.261 im Wesentlichen für Bildtelefonie- und Videokonferenz-Übertragungen gedient. MPEG-1 wiederum ist gezielt für die digitale Speicherung von Videosignalen entwickelt worden.

Vor dem Hintergrund des raschen Zusammenwachsens der Telekommunikation, der Computertechnologie und der Konsumerelektronik sind aktuelle Standards, wie MPEG-4 [57], [66], [68], H.263 [73], [75], [76], [78] oder H.264 [71], [74], [128], flexibel ausgelegt. Sie können für unterschiedlichste, teilweise auch interaktive Multimediaanwendungen eingesetzt werden und stellen die Übertragung der codierten Inhalte über verschiedene Netze sicher.

In der folgenden Übersicht sind verschiedene Standardisierungen der ITU-T und ISO/IEC zu finden, die im Wesentlichen nach dem gleichen Funktionsprinzip arbeiten [79], [80], [96], [126].

### 2.2.4.1 Videokonferenz-Standard: H.120

Die Recommendation CCITT H.120 "Codecs for Videoconferencing using Primary Digital Group Transmission" ist der erste Standard für die digitale Videocodierung aus dem Jahre 1984 mit einer Zieldatenrate für Videokonferenz-Übertragungen von 2 MBit/s. In der Version 1 (1984) arbeitete er nach dem Replenishment-Verfahren. Bildinhalte, die nicht an der gleichen Position im Vorgängerbild vorhanden sind, werden mit einem Intraframe-DPCM-Verfahren (Difference Pulse Code Modulation) übertragen. Zusätzlich werden eine skalare Quantisierung und eine VLC unterstützt. Die Version 2 (1988) bietet zusätzlich die Möglichkeiten einer Bewegungskompensation und einer Hintergrundprädiktion.

- 
1. ITU Telecommunication Standardization Sector ist einer der drei Sektoren der International Telecommunication Union (ITU).
  2. International Organization for Standardization (ISO) bzw. International Electrotechnical Commission (IEC).

### 2.2.4.2 Bildtelefonie-Standard: H.261

Zielanwendungen für die Entwicklung der ITU-T Recommendation H.261 "Video Codec for Audiovisual Services at p x 64 kBit/s" im Jahre 1990 ist die Bildtelefon- und Videokonferenzübertragung mit fester Datenrate von p x 64 kBit/s (p = 1 ... 30) gewesen [72]. Die Struktur des Standards basiert auf einem Hybrid-DCT-Codec mit Bewegungskompensation.

Die wichtigsten Eigenschaften des Standards sind: Bildgrößen QCIF und CIF, Frameraten bis 30 fps, optionale Bewegungsschätzung für Makroblöcke der Größe 16 x 16 mit  $\pm 15$  Pixel pro Richtung bei 1-Pixel Suchgenauigkeit, Umschaltung auf Intra- und Interframe-Codierung pro Makroblock, Lauflängencodierung der DCT-Koeffizienten, lineare ungewichtete Quantisierung der Koeffizienten und verlustlose Differenzcodierung der Bewegungsparameter. Der Aufbau des Encoders entspricht Abbildung 2.1.

### 2.2.4.3 Standard für die digitale Speicherung: MPEG-1

MPEG-1 ist als ISO/IEC 11172 "Coded Representation of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 MBit/s" im Jahre 1993 standardisiert [55] worden. Der Multimediastandard besteht aus 5 Teilen: Systems (11172-1), Video (11172-2), Audio (11172-3), Compliance testing (11172-4), Simulation Software (11172-5). Die folgenden Ausführungen beziehen sich ausschließlich auf den Videoteil.

Zielsetzung ist die Entwicklung eines Codecs gewesen, der zum einen höheren Qualitätsansprüchen gerecht wird und zum anderen zusätzlicher Funktionalitäten, wie z.B. wahlfreier Zugriff und schnellen Vor- und Rücklauf bereitstellt. Zudem hat die Notwendigkeit bestanden, dass der Codec mit komplexerem Bildmaterial als H.261 zurechtkommen musste. So treten bei Heim- oder Studio-Videosignalen vor allem oftmals schnellere Bewegungen, erhöhte Detailstrukturierung, bewegte Kameraeinstellungen und Szenenwechsel auf. Das Codiervorgehen von MPEG-1 basiert auf H.261, wobei es unter anderem um folgende Punkte erweitert worden ist:

- Zusammenfassung einer Anzahl von Bildern zu einer Gruppe (Group Of Pictures, GOP). Dabei wird mindestens ein Bild (I-Bild) intracodiert. Die Anzahl der Bilder in einem GOP ist frei wählbar.
- Verbesserung der Bewegungskompensation, indem der Suchbereich auf  $\pm 64$  Pixel erweitert und die Genauigkeit auf  $1/2$ -Pixel erhöht worden ist.
- Einführung einer bewegungskompensierten Interpolation (B-Bild). Dabei ist zu berücksichtigen, dass die zur Bewegungskompensation notwendigen Bilder vor dem B-Bild codiert und auch decodiert werden müssen. Damit erhöht sich die Ende-zu-Ende-Verzögerung und führt im Vergleich zum H.261 zu einem erhöhten Speicherbedarf im Encoder und Decoder.
- Möglichkeiten eigene Gewichtungstabellen für die Quantisierung festzulegen.

#### 2.2.4.4 Standard für universelle (hochbitratige) Anwendungen: MPEG-2

Aufgrund der Einschränkungen von MPEG-1 auf 1,5 MBit/s und der möglichen Bildgrößen ist der MPEG-2-Standard "Generic Coding of Moving Pictures and Associated Audio Information" unter der Bezeichnung ISO/IEC 13818 definiert worden [56]. Die ITU-T hat den Standard unter der Bezeichnung H.262 übernommen. Der Videoteil (13818-2) spezifiziert Datenraten bis ca. 40 MBit/s und Bildgrößen bis zu den HDTV-Formaten (High Definition Television). Der Einsatzbereich umfasst nicht nur die Videoaufzeichnung, sondern auch die Videoübertragung, z.B. über Breitbandnetze. Als Erweiterungen gegenüber MPEG-1 sind vor allem die Unterstützung von Zeilensprungabtastung und die Halbbild-Prädiktionsstrategie zu nennen. Zudem ist eine skalierbare und hierarchische Codierung zur Rückwärtskompatibilität zu MPEG-1 definiert worden.

Die Vielzahl der Anwendungen und die daraus folgende Kombinationsvielfalt der Standardeigenschaften machen es notwendig, so genannte Anwendungsprofile (Profiles) mit jeweils unterschiedlichen Auflösungsstufen (Levels) zu definieren. Dabei definiert das Profile normalerweise die Funktionalität und der Level die notwendige Performance. Die Gruppierung von Profiles und Levels ermöglicht die Vereinfachung von Encodern und Decodern für bestimmte Anwendungsbereiche. Aktuell wird der Standard im Wesentlichen für DVDs und digitales Fernsehen verwendet, wobei der Datenratenbereich normalerweise zwischen 2 - 20 MBit/s liegt.

#### 2.2.4.5 Standard für universelle Anwendungen: H.263

Die ITU-T Empfehlung H.263 "Video Coding for Low Bitrate Communication" ist ein Videokompressionsstandard, für Datenraten von 10 - 2048 kBit/s [73]. Beim H.263 Testmodell sind durch eine 1/2-Pixel genaue Bewegungskompensation die *PSNR*-Werte um ca. 3-4 dB gegenüber H.261 bei Datenraten unter 64 kBit/s erhöht worden. Der prinzipielle Aufbau des H.263-Encoders entspricht H.261 (siehe Abbildung 2.1). Zusätzlich zu den Bildformaten des H.261 sind die Bildformate Sub-QCIF, 4CIF und 16CIF definiert worden. Die wesentlichen Erweiterungen gegenüber H.261 sind:

- Erhöhung der Genauigkeit der Bewegungskompensation auf 1/2-Pixel.
- Effizientere Codierung der Bewegungsvektoren und eine verbesserte VLC.
- Erweiterte Bereiche der Bewegungsvektoren mit Extrapolation des Bildes (optional).
- Verbesserte Bewegungsschätzung mit variabler Blockgröße, überlappenden Blöcken und Extrapolation des Bildes (optional).
- Möglichkeit P- und B-Bilder als eine Einheit (PB-Bild) zu codieren (optional).
- Arithmetische Entropie Codierung (optional).

Ergänzend zur Version 1 (1996) sind die Versionen 2 (1998) und 3 (2001) definiert worden, die auch H.263+ bzw. H.263++ bezeichnet werden. Durch Erweiterungen sind sowohl eine Steigerung der Effizienz als auch der Funktionalität erreicht worden, sodass folgende Applikationsbereiche abgedeckt werden können: Übertragung über fehlerbehaftete Netze, flexible Datenraten und spezielle Effekte, wie z.B. die Kompensation globaler Bewegungen. Für eine Vereinfachung von En- und Decodern sind auch hier Profiles und Levels eingeführt worden [75], [76], [78].

#### 2.2.4.6 Standard für universelle Anwendungen: MPEG-4

Zielsetzung bei der Entwicklung von MPEG-4 mit dem offiziellen Namen ISO/IEC 14496 "Information Technology - Coding of audio-visual Objects -" ist vor allem die Schaffung eines universell einsetzbaren Standards gewesen, der aktuell aus 17 Teilen besteht [57]. Die Videocodierung wird im Teil 2 des Standards beschrieben. Sie deckt das gesamte Anwendungsspektrum von sehr geringen Datenraten von ca. 5 kBit/s bis zu mehreren hundert MBit/s für transparente Qualität ab. Zudem wird die Übertragung über unterschiedliche Netze berücksichtigt.

Als Videoeingangsformate sind Größen von QCIF bis HDTV vorgesehen, die sowohl progressiv, als auch im Interlaced-Format vorliegen können. Die wichtigsten Profiles (Simple und Advanced Real Time Simple) für Kommunikationsanwendungen sehen Datenraten bis 2 MBit/s vor. Das Codierprinzip folgt dabei H.263. Allerdings ist es zur Steigerung der Funktionalität und der Videocodiereffizienz um zusätzliche Funktionen ergänzt worden. Für MPEG-4 sind aktuell zwei Versionen bereits standardisiert: MPEG-4v1 (1999) und MPEG-4v2 (2000).

Das Simple Profile von MPEG-4v1 entspricht weitestgehend H.263. Im so genannten Core Profile ist als wichtigste Erweiterung eine binäre Shape-Codierung definiert worden. Im Main Profile sind zusätzlich folgende Funktionalitäten definiert: Graustufen Shape-Codierung, Unterstützung von Interlaced-Bildmaterial und statischen Sprites. Des Weiteren sind in MPEG-4v1 noch Skalierbarkeit, 12-Bit Farbauflösung und Gesichtsanimation definiert. In MPEG-4v2 ist unter anderem die Genauigkeit der Bewegungskompensation auf 1/4-Pixel erhöht, eine globale Bewegungskompensation definiert und die Resistenz gegenüber Bitfehlern (Error Resilience) verbessert worden. Zusätzliche Informationen zu MPEG-4 sind z.B. in [18], [31], [32], [63], [66]-[68] und [80] zu finden.

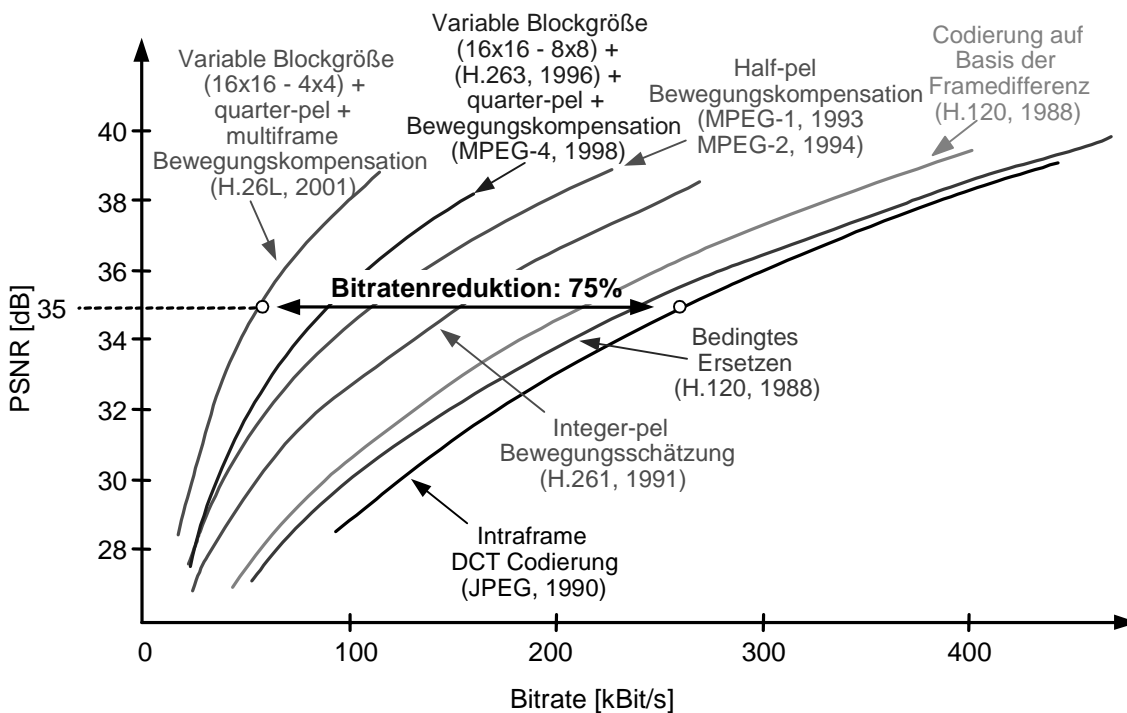
#### 2.2.4.7 Standard für universelle Anwendungen: H.264 / MPEG-4 AVC

Unter dem Namen H.264 "Advanced video coding for generic audiovisual services" standardisierte die ITU-T bzw. die ISO/IEC mit der Bezeichnung 14496-10 AVC (MPEG-4 AVC) im Jahre 2003 einen neuen Standard für die Videocodierung [74]. Der Einsatzbereich erstreckt sich von Videomails über Echtzeitkommunikation bis hin zu digitalem TV. Entsprechend sind als Datenraten von ca. 5 kBit/s bis hin zu mehreren hundert MBit/s definiert. Die Übertragung ist über unterschiedliche Netze vorgesehen, so z.B. paketorientiert, mobil oder über klassische Telekommunikationsnetze. Im Vergleich zu MPEG-4 ist vor allem die Videocodiereffizienz gesteigert worden.

Der Basiscodec ist ähnlich zu H.263, wobei insbesondere die Bewegungskompensation erweitert worden ist, so z.B. Erhöhung der Genauigkeit auf 1/4- oder 1/8-Pixel und zusätzliche Freiheiten für unterschiedliche Blockgrößen. Des Weiteren können mehrere Bilder für die Bewegungskompensation benutzt werden. Anstelle der 2D-DCT wird auf 4 x 4 Blöcke eine separierbare Ganzzahlen-Transformation angewendet, die ähnliche Ergebnisse wie die 2D-DCT liefert. Zusätzlich wird eine modifizierte Quantisierung eingesetzt. Weitergehende Informationen sind in [60], [63], [71], [74] und [117] zu finden.

### 2.2.4.8 Bewertung

Seit über 10 Jahren wird bei allen wichtigen Videostandards ein hybrides Verfahren eingesetzt. Während H.261, MPEG-1 und MPEG-2 speziell für die jeweiligen Einsatzbereiche konzipiert worden sind, haben aktuelle Standards wie H.263, MPEG-4 oder H.264 ein breites Anwendungsspektrum und können eine effiziente Übertragung über verschiedene Netze sicherstellen. Für die Steigerung der Videocodiereffizienz, auch unter schwierigen Bedingungen, werden hierzu zahlreiche optionale Funktionalitäten angeboten. Damit erhöht sich allerdings wesentlich der Aufwand für die Codecsteuerung, um die notwendige Adaptivität an die aktuelle Bildsignalcharakteristik sicherstellen zu können. In Abbildung 2.4 ist die Steigerung der Videocodiereffizienz beispielhaft dargestellt. Dabei basieren sowohl MPEG-4 als auch H.264/MPEG-4 AVC in ihrer Grundfunktionalität auf H.263.



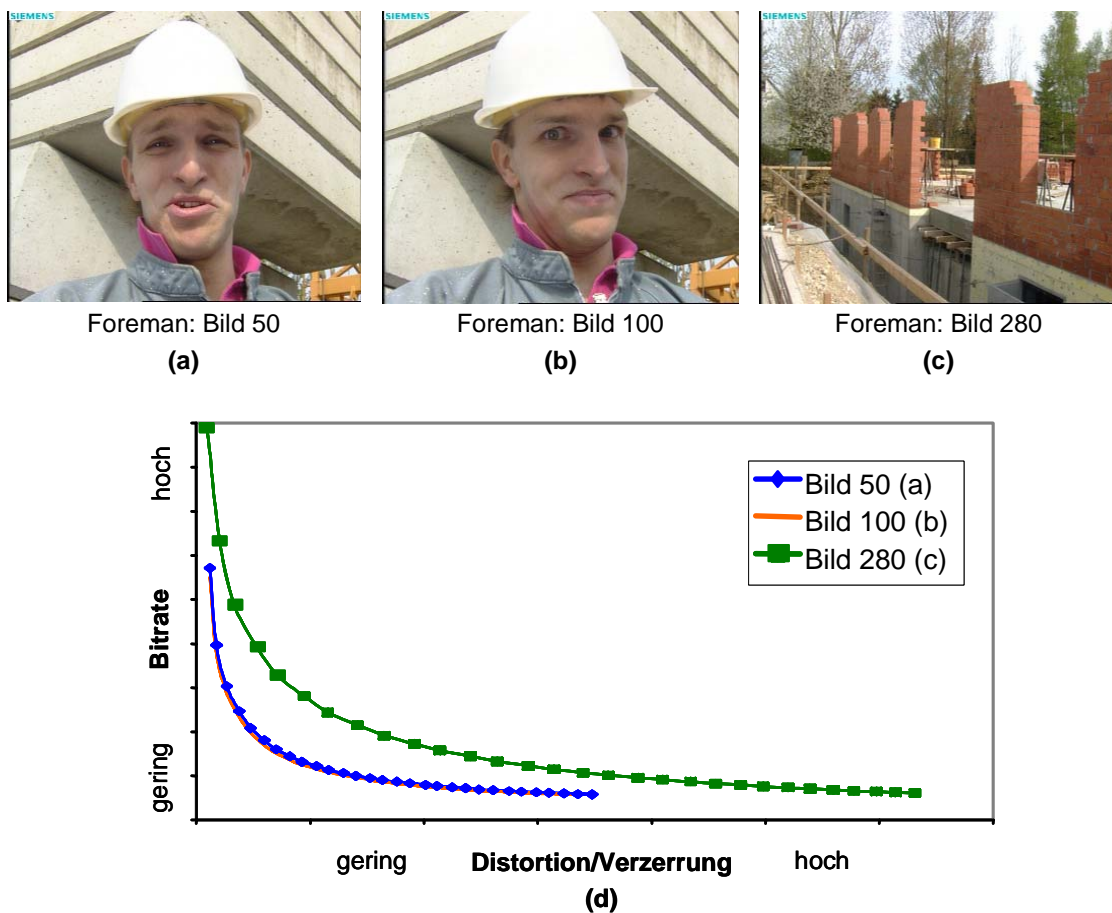
**Bedingungen:** Sequenz Foreman, QCIF, 10 fps und 100 codierte Bilder

**Abbildung 2.4:** Vergleich verschiedener Videocodecs hinsichtlich ihrer Codiereffizienz [126].

## 2.3 Datenratenkontrolle in Videocodecs

Die Aufgabe der Datenratenkontrolle ist die Anpassung der Ausgangsdatenrate des Videoencoders an die vorgegebene Datenrate, um eine verzögerungsbegrenzte Übertragung sicherzustellen. Dabei müssen die Vorgaben des Videostandards und der Anwendung eingehalten werden. Wird ein digitales Videosignal verlustlos komprimiert, so ist vor allem die resultierende Datenrate von Interesse. Werden allerdings verlustbehaftete Methoden eingesetzt, so stellt sich die Frage: Welche Bitrate ist notwendig, um eine vorgegebene Verzerrung zu erreichen bzw. wie hoch ist die Verzerrung bei vorgegebener Bitrate?

Neben der Auswahl und der Umsetzung des Videocodiervorgangs bestimmen im Wesentlichen zwei Faktoren die erzielbare Qualität: zum einen die Eigenschaften des Bildes und zum anderen die Bitrate. In Abbildung 2.5 ist für drei unterschiedliche Bilder der Sequenz Foreman der Zusammenhang zwischen der notwendigen Bitrate und der jeweiligen Verzerrung dargestellt<sup>1</sup>.



Bedingungen: H.263 ohne Optionen, I-Bilder, QP = 1, ..., 31

Abbildung 2.5: Distortion-Rate-Kennlinien von Einzelbildern der Sequenz Foreman.

- Die Kennlinien ergeben sich aus der Bestimmung der Bitrate und der Verzerrung in Abhängigkeit des Quantisierungsparameters  $Q$  für  $Q = 1, 2, \dots, 31$  mit anschließender Approximation der Kennlinien. Daher ist der für wertdiskrete Signale typische Schnittpunkt mit der y-Achse nicht vorhanden.

Deutlich zu erkennen ist im Diagramm (d), dass für die Codierung von Bild 280 der Sequenz Foreman (Teilbild (c)) eine deutlich größere Bitmenge notwendig ist, um die gleiche Bildqualität wie Bild 50 (Teilbild (a)) oder Bild 100 (Teilbild (b)) der gleichen Sequenz zu erreichen. Im Umkehrschluss gilt, dass Bild 280 der Sequenz Foreman im Vergleich zu den anderen dargestellten Bildern nur mit schlechterer Bildqualität codiert werden kann, wenn jeweils eine gleich große Bitmenge für alle Teilbilder zur Codierung zur Verfügung steht.

Der Zusammenhang zwischen Datenrate und erzielbarer Verzerrung bei der Codierung wird durch die Rate-Distortion-Theorie und das Quellencodierungstheorem beschrieben.

### 2.3.1 Rate-Distortion-Theorie

Die Rate-Distortion-Funktion gibt für die Codierung einer gegebenen Signalstatistik einen funktionalen Zusammenhang zwischen Bitrate und Verzerrung an. Im Folgenden wird auf die statistischen Grundlagen für die Beschreibung der Rate-Distortion-Funktion eingegangen [96].

#### 2.3.1.1 Statistische Grundlagen

Der Informationsgehalt  $i$  [bit] eines Ereignisses  $j$  mit der Wahrscheinlichkeit  $p(j)$  ist:

$$i(j) = \log_2\left(\frac{1}{p(j)}\right) = -\log_2(p(j)) \quad (2.6)$$

Aus (2.6) folgt, dass der Informationsgehalt hoch ist, wenn ein Ereignis selten auftritt ("Überraschungseffekt"). Die Entropie  $H$  ist der Mittelwert des Informationsgehalts, wenn  $j$  eines von  $J$  möglichen Ereignissen ist:

$$H(j) = -\sum_{j=1}^J p(j) \cdot \log_2(p(j)) \quad (2.7)$$

Ableitend aus der Definition der bedingten Wahrscheinlichkeit  $p(j_1|j_2)$  für die Wahrscheinlichkeiten  $p(j_1)$ ,  $p(j_2)$  und der Annahme über die Verbundwahrscheinlichkeit  $p(j_1, j_2) = p(j_2, j_1)$  der Ereignisse  $j_1$  und  $j_2$ :

$$p(j_1|j_2) = \frac{p(j_1, j_2)}{p(j_2)}; p(j_2|j_1) = \frac{p(j_2, j_1)}{p(j_1)} \quad (2.8)$$

lässt sich in ähnlicher Form auch der bedingte Informationsgehalt eines Ereignisses  $i(j_2|j_1)$  definieren. Er gibt den Anteil der Information über ein Ereignis  $j_2$  an, der noch unbekannt ist, wenn man über die Information eines Ereignisses  $j_1$  bereits verfügt:

$$i(j_2|j_1) = -\log_2(p(j_2|j_1)) = -\log_2\left(\frac{p(j_1, j_2)}{p(j_1)}\right) \quad (2.9)$$



Für statistisch unabhängige Ereignisse  $p(j_2|j_1) = p(j_2)$  folgt, dass der bedingte Informationsgehalt  $i(j_2|j_1)$  gleich  $i(j_2)$  wird. Als Transinformationsgehalt  $i(j_2;j_1)$  wird die Reduktion an Informationsgehalt bezeichnet, die erfolgt, wenn das Wissen über die statistischen Zusammenhänge zwischen  $j_1$  und  $j_2$  beim Auftreten von  $j_2$  ausgenutzt wird:

$$i(j_2;j_1) = i(j_2) - i(j_2|j_1) = \log_2\left(\frac{p(j_2|j_1)}{p(j_2)}\right) = \log_2\left(\frac{p(j_1|j_2)}{p(j_1)}\right) \quad (2.10)$$

Aus dem bedingten Informationsgehalt bestimmt sich der Mittelwert, die bedingte Entropie  $H(j_2|j_1)$  eines diskreten Quellensignals zu:

$$\begin{aligned} H(j_2|j_1) &= -\sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} p(j_1) \cdot p(j_2|j_1) \cdot \log_2(p(j_2|j_1)) = \\ &= -\sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} p(j_1, j_2) \cdot \log_2(p(j_2|j_1)) \end{aligned} \quad (2.11)$$

Damit ergibt sich aus (2.7), (2.10) und (2.11) der mittlere Transinformationsgehalt  $H(j_2;j_1)$  zu:

$$\begin{aligned} H(j_2;j_1) = H(j_1;j_2) &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} p(j_1, j_2) \cdot \log_2\left(\frac{p(j_1, j_2)}{p(j_1) \cdot p(j_2)}\right) = \\ &= H(j_2) - H(j_2|j_1) \end{aligned} \quad (2.12)$$

In der folgenden Beschreibung der Rate-Distortion-Funktion werden die Begriffe Informationsgehalt und Transinformationsgehalt verwendet. Ausführliche Informationen hierzu und zu statistischen Grundlagen befinden sich in [24], [25], [27], [96], [102], [111].

### 2.3.1.2 Rate-Distortion-Funktion

Werden  $K$  Bildpunkte  $x(n)$  zu einem Vektor  $\mathbf{X}$  zusammengefasst und mittels des Codebuchs  $\mathbf{C}$  auf die Rekonstruktionswerte  $y(n)$  des Vektors  $\mathbf{Y}$  abgebildet, so ergibt sich der mittlere Transinformationsgehalt zwischen den Vektoren  $\mathbf{X}$  und  $\mathbf{Y}$  bei Verwendung des Codebuches  $\mathbf{C}$  zu:

$$H(\mathbf{X};\mathbf{Y})|_{\mathbf{C}} = H(\mathbf{X}) - H(\mathbf{X}|\mathbf{Y})|_{\mathbf{C}} \quad (2.13)$$

Aus den Definitionen (2.10) und (2.12) folgt, dass der mittlere Transinformationsgehalt in (2.13) ein Maß für die notwendige Information zur Beschreibung des Vektors  $\mathbf{X}$  durch den Rekonstruktionsvektor  $\mathbf{Y}$  ist, wenn das Codebuch  $\mathbf{C}$  verwendet wird.  $\mathbf{C}_{\mathbf{D}}$ <sup>1</sup> stellt dabei die Menge aller Codes dar, die eine Codierung des Ausgangsvektors mit der gleichen Verzerrung  $D$  ermöglichen. Derjenige Code aus  $\mathbf{C}_{\mathbf{D}}$ , der die kleinste Bitrate  $R(D)$  bei der vorgegebenen Verzerrung  $D$  benötigt, hat auch den geringsten Transinformationsgehalt:

---

1.  $D$ : Bezug auf Verzerrung (Distortion).

$$R(D) = \min_{\mathbf{C} \in \mathbf{C}_D} (H(\mathbf{X};\mathbf{Y})|_{\mathbf{C}}) \quad (2.14)$$

Das Quellencodierungstheorem, das von Shannon [111] mit der Verallgemeinerung auf wertkontinuierliche und korrelierte Quellen formuliert worden ist, lautet [96]:

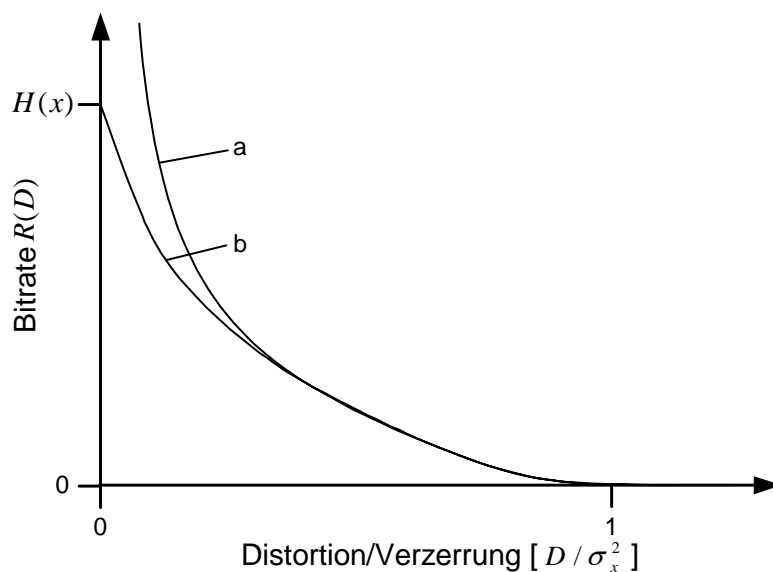
"Für die Codierung eines diskreten Quellensignals existiert, wenn eine Verzerrung kleiner oder gleich  $D$  zugelassen wird, ein Blockcode mit der Bitrate  $R = R(D) + \varepsilon$ ,  $\varepsilon > 0$ , wenn die Blocklänge  $K$  des Codes groß genug gewählt wird."

Aufgrund des konvexen Verlaufs von  $R(D)$  kann sowohl die inverse Funktion  $D(R)$  definiert werden als auch die Umkehrung des Quellentheorems [96]:

"Steht zur Codierung eines diskreten Quellensignals eine Bitrate  $R$  zur Verfügung, so kann eine Verzerrung  $D(R)$  nicht unterschritten werden."

Damit können aus dem Quellencodierungstheorem zwei Aussagen abgeleitet werden: Bei der Codierung eines abgetasteten Signals besteht ein Zusammenhang zwischen vorgegebener Verzerrung  $D$  und der dafür notwendigen minimalen Bitrate. Zudem kann diese minimal notwendige Bitrate  $R(D)$  beliebig nahe approximiert werden, wenn eine genügend hohe Anzahl von Abtastwerten nicht einzeln, sondern durch Blockcodes zusammengefasst als Vektoren codiert werden.

In Abbildung 2.6 ist ein beispielhafter Verlauf einer Rate-Distortion-Funktion  $R(D)$  dargestellt. Wird als Fehlermaß der Verzerrung  $D$  der quadratische Fehler angesetzt, so wird bei der Bitrate  $R = 0$  die Verzerrung  $D = \sigma^2$  und damit identisch mit der Varianz des Signals, das hier zur Vereinfachung mittelwertfrei angenommen wird. Für wertdiskrete Quellen ist eine verzerrungsfreie Codierung mit der Bitrate der Entropie  $H(\mathbf{X})$  möglich. Dies folgt aus (2.13) und (2.14). Hingegen wird die Bitrate für wertkontinuierliche Quellen prinzipiell unendlich, wenn eine verzerrungsfreie Codierung gefordert ist.



**Abbildung 2.6:** Beispielhafter Verlauf einer Rate-Distortion-Funktion bei wertkontinuierlichen (a) und wertdiskreten (b) Signalen [96].

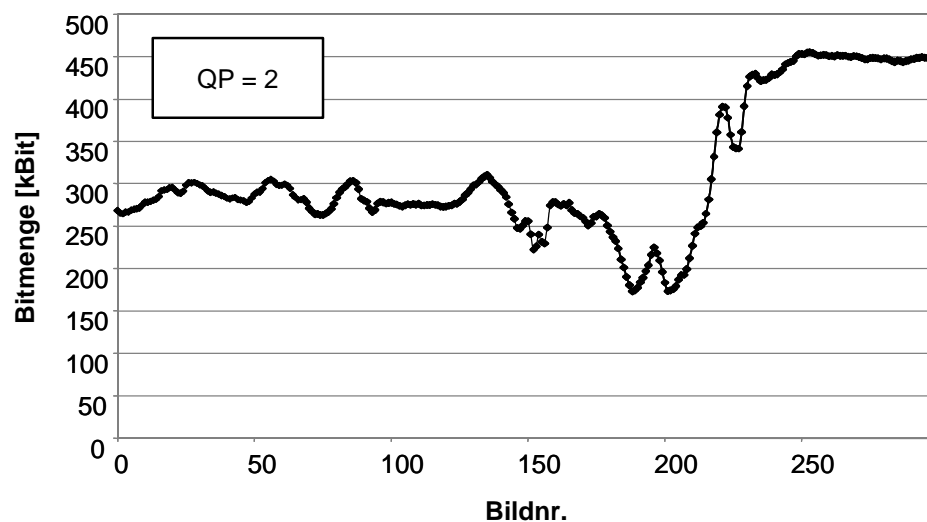
Eine analytische Bestimmung der Rate-Distortion-Funktion ist nur für Modellsignale mit stationärer Statistik und bei Verwendung des quadratischen Fehlerkriteriums  $D$  möglich. Für die wichtigen Fälle eines stationären, unkorrelierten Signals  $x(n)$  der Varianz  $\sigma_x^2$  mit Gauß-Verteilung gilt:

$$R(D) = \max\left(\frac{1}{2}\log_2\left(\frac{\sigma_x^2}{D}\right), 0\right) \quad (2.15)$$

Für alle unkorrelierten Signale derselben Varianz  $\sigma^2$  stellt (2.15) die Obergrenze der Bitrate dar. Mit Einschränkungen auf AR(1)-Prozesse (autoregressiv) mit der Varianz  $\sigma_x^2$  und einer kleinen Verzerrung  $D$  ist (2.15) auch für korrelierte Gauß-Prozesse gültig [96].

### 2.3.2 Funktionsprinzip der Datenratenkontrolle bei der Videocodierung

In Abbildung 2.1 ist das Grundkonzept von aktuellen Videocodierverfahren dargestellt. Dabei wird das Bildsignal nach der Dekorrelation quantisiert, wobei je nach Quantisierungsstufe bzw. Quantisierungsparameter vor der Entropiecodierung eine Verzerrung festgelegt wird. Am Ausgang der Entropiecodierung ist die resultierende Bitrate damit abhängig von der Verteilungsdichtefunktion des Bildes, oder seines dekorrelierten Äquivalents. Eine Codierung mit konstanter Verzerrung ergibt damit stets eine - vom Informationsgehalt des jeweiligen Bildes abhängige - variable Bitrate (siehe Abbildung 2.7).



**Abbildung 2.7:** Resultierende Bitmenge pro Bild bei der Intracodierung der Sequenz Foreman mit const. Quantisierungsparameter.

---

1.  $x$ : Bezug zum Signal  $x$ .

Soll jedoch eine verzögerungsbegrenzte Übertragung mit konstanter (Constant BitRate, CBR) bzw. an die Vorgaben des Netzes angepasster Bitrate erfolgen, so wird meist die Verzerrung in Abhängigkeit vom Detailgehalt des Bildes variiert. Hierzu wird ein Puffer im Encoder und Decoder benutzt, in dem der resultierende Bitstrom bei der Übertragung zwischengespeichert wird. Drohen die Puffer über oder leer zu laufen, muss gegengesteuert werden, indem die Verzerrung bzw. der Quantisierungsparameter im Encoder erhöht bzw. verringert wird.

Große Puffer sind für den Regelprozess weniger kritisch. Allerdings hat die Größe der Puffer maßgeblichen Einfluss auf die Ende-zu-Ende-Verzögerung eines Videokommunikationssystems. Wichtige Parameter für den Regelmechanismus sind zum einen die effektive Puffergröße und zum anderen die Zeitabstände für die Überprüfung des Pufferfüllstands [96]. Die Verifikation des Videopuffermanagements wird im Rahmen der Videostandards beschrieben (siehe Abschnitt 2.3.2.1).

Neben der Festlegung des Quantisierungsparameters kommen noch weitere Methoden für die Regelung der Ausgangsdatenrate in Frage:

- Regulierung der aktuellen Framerate: Wenn in den Puffern zu viele Daten zwischengespeichert sind, kann das nächste oder die nächsten Bilder übersprungen werden. Bewegungsabläufe erscheinen dadurch oftmals nicht mehr fließend. Häufige und ungleichmäßige Änderung der Framerate verstärkt diesen Eindruck.
- Dynamische Auflösungskonvertierung: Ist die aktuelle Ausgangsrate des Videoencoders höher als die Übertragungsbandbreite, können die Eingangsbilder vor der Encodierung verkleinert und nach der Decodierung im Decoder wieder aufskaliert werden. Bisher wird diese Möglichkeit nur in wenigen Videocodecs optional unterstützt.
- Vorverarbeitung des Bildsignals: Prinzipiell ist es möglich, die Details im Bild zu verringern. Dies kann sowohl durch Filterung als auch durch Segmentierung erreicht werden. Eine Filterung weist oftmals einen ähnlichen Effekt wie (ungleichmäßige) Quantisierung bzw. das Verwerfen von hochfrequenten DCT-Koeffizienten auf. Dagegen hat die Segmentierung meist nur einen sehr geringen Einfluss auf die Datenrate [103].

Im Gegensatz zur CBR-Codierung darf die Ausgangsdatenrate bei der Übertragung mit variabler Bitrate (Variable BitRate, VBR) - in Grenzen - variieren. Somit ist es möglich, dass die Verzerrung - weitgehend unabhängig vom jeweiligen Bildinhalt - konstant bleibt. Allerdings stellt diese Übertragungsweise neue Anforderungen an das Übertragungsnetz. Für die Einhaltung der vorgegebenen Netzparameter ist es jedoch meist notwendig, ein ähnliches Vorgehen wie bei der CBR-Codierung einzusetzen.

Im Anschluss an die Beschreibung der Verifikation des Videopuffermanagements wird im Abschnitt 2.3.2.2 auf die Rate-Distortion-Modellierung in der Datenratenkontrolle eingegangen.

### 2.3.2.1 Referenzmodell für die Verifikation des Videopuffermanagements

In der Standardisierung von Videocodecs wird prinzipiell nur das Bitstromformat und die Decoderstruktur bzw. dessen grobe Arbeitsweise festgeschrieben. Am Beispiel des Standards H.263 [73] werden im Folgenden die Kriterien für die Verifikation des Videopuffermanagements beschrieben<sup>1</sup>.

In der Definition des Hypothetical Reference Decoders (HRD) wird für die Verifikation des Videopuffermanagements davon ausgegangen, dass sowohl der Encoder als auch der Decoder synchron mit der gleichen Referenz-Framerate  $REF\_FR$  arbeiten. Solange keine anderen Werte zwischen Encoder und Decoder signalisiert werden, wird als Standardeinstellung die CIF-Referenz-Framerate bzw. CIF-Bild-Periode von 30 000/1001 (ungefähr 29.97) fps definiert.

Die Empfangspuffergröße  $BS_{max}^D$  im Decoder ist bei H.263 durch  $BS + BPPmaxKb \cdot 1024$  Bit definiert [73]. Der minimale Wert für  $BS$  ist definiert durch:

$$BS = 4 \frac{R_{max}}{REF\_FR}. \quad (2.16)$$

$BPPmaxKb \cdot 1024$  ist die maximale Anzahl der Bits pro Bild, die zwischen Encoder und Decoder auf Basis der Mindestwerte in Tabelle 2.2 signalisiert wird.

Pixelanzahl des Y-Bildes	Minimales $BPPmaxKb$ [kBit]
Bis zu 25 344 (oder QCIF)	64
25 360 bis zu 101 379 (oder CIF)	256
101 392 bis zu 405 504 (oder 4CIF)	512
405 520 und mehr	1024

**Tabelle 2.2:** Minimales  $BPPmaxKb$  für unterschiedliche Bildformate [73].

Dabei ist  $R_{max}$  die maximale Videobitrate während der Verbindung in Bits pro Sekunde.

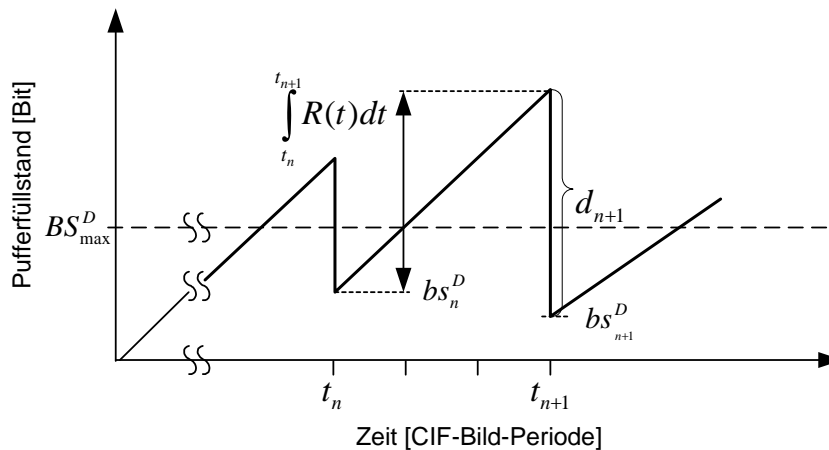
In Abbildung 2.9 ist ein beispielhafter Füllstandverlauf des spezifizierten Eingangspuffers des Decoders  $BS_{max}^D$  in Abhängigkeit von der Zeit dargestellt. Für die Initialisierung wird der Puffer vollständig geleert. Alle  $1000/REF\_FR$  ms wird der Pufferfüllstand kontrolliert. Sobald ein ganzes Bild im Puffer ist, werden alle Daten des ältesten Bildes als Ganzes sofort aus dem Puffer entfernt (z.B. zum Zeitpunkt  $t_n + 1$ <sup>3</sup>). Unmittelbar nach dem Entfernen der Daten muss der Füllstand per Definition kleiner  $BS_{max}^D$  sein.

1. Auch in der Standardisierung von MPEG-4 [66] ist diese Vorgehensweise zu finden. Allerdings werden hier aufgrund der erweiterten Fähigkeiten von MPEG-4, wie z.B. die Unterstützung mehrerer Objekte, zusätzlich noch die Referenzmodelle für das Management der Komplexität und des Speichers definiert.
2.  $D$ : Bezug auf den Decoder bzw.  
 $max$ : Bezug auf Maximal.
3.  $n$ : Bezug auf das  $n$ -te Bild.

Für das Erfüllen der Bedingung muss die Bitmenge  $d_{n+1}$  beim Codieren des  $(n+1)$ -ten Bildes folgender Ungleichung genügen:

$$d_{n+1} \geq bs_n^D + \int_{t_n}^{t_{n+1}} R(t) dt - BS_{max}^D \quad (2.17)$$

Dabei ist  $t_n$  der Zeitpunkt an dem das  $n$ -te Bild vom Eingangspuffer des Decoders entfernt wird,  $bs_n^D$  ist der Pufferfüllstand des Decoders zu dieser Zeit und  $R(t)$  ist die Bitrate zum Zeitpunkt  $t$ .



Anmerkung: Zeit  $(t_{n+1} - t_n)$  entsprechend dem ganzzahligen Vielfachen der CIF-Bild-Periode (1/29,97, 2/29,97, 3/29,97, ...)

**Abbildung 2.8:** Füllstandverlauf des Eingangspuffers des Decoders in Abhängigkeit von der Zeit [73].

Im Videostandard wird nur der Eingangspuffer des Decoders betrachtet, allerdings besteht zwischen dem Ausgangspuffer des Encoders und des Eingangspuffers im Decoder ein direkter Zusammenhang. Im Folgenden wird dieser Aspekt in Verbindung mit der Einführung einer effektiven Puffergröße  $BS_{EFF}^E$ <sup>1</sup> im Encoder untersucht. Für eine störungsfreie IPPP...-Videokommunikation und der Annahme, dass eine Encodierung und Decodierung gleichmäßig und ohne schwankende Verzögerung abläuft, muss die Verzögerung  $\Delta T$  zwischen dem Puffer im Encoder bzw. dem Puffer im Decoder konstant sein:

$$\Delta T = \Delta t^{EP} + \Delta t^N + \Delta t^{DP} = const.^2, \quad (2.18)$$

wobei  $\Delta t^{EP}$  bzw.  $\Delta t^{DP}$  die Pufferverzögerungen im Encoder bzw. Decoder und  $\Delta t^N$  die Transportverzögerung des Netzes sind.

- 
1.  $E$ : Bezug auf den Encoder bzw.  
 $EFF$ : Bezug auf den effektiven Wert.
  2.  $EP$ : Bezug auf den EncoderPuffer,  
 $DP$ : Bezug auf den DecoderPuffer bzw.  
 $N$ : Bezug auf das Transportnetz.

Die Verzögerung in Transportnetzen variiert. Bei der Übertragung über Netze, die zumindest prinzipiell den Anforderungen der Echtzeitkommunikation entsprechen, ist die durch Jitter und die Übertragung verursachte Verzögerung klein im Verhältnis zu der Pufferverzögerung im Encoder und Decoder und wird im Folgenden vernachlässigt. Mit der Annahme, dass der zeitliche Abstand bei der Referenz-Framerate  $REF\_FR$  zwischen zwei Bildern  $t_{REF}$ <sup>1</sup> ist, erhält man als Anzahl der Bilder  $N$  im Encoder- und Decoderpuffer:

$$N = \frac{\Delta T - \Delta t^N}{t_{REF}} \quad (2.19)$$

$N$  entspricht dabei der Anzahl der Bilder bezüglich der Referenz-Framerate  $REF\_FR$ . Bei  $t_{REF} \neq t_{FR}$ <sup>2</sup> entspricht dies nicht der physikalischen Anzahl der Bilder in den Puffern<sup>3</sup>.

Die Datenratenkontrolle muss sicherstellen, dass die Puffer weder im Encoder noch im Decoder über bzw. leer laufen. Im Encoder kann durch Bitstuffing verhindert werden, dass der Puffer geleert wird. Kritisch ist das Überlaufen des Encoderpuffers bzw. das damit verbundene Leerlaufen des Decoderpuffers. Es kommt dabei zu einer Verletzung der Ende-zu-Ende-Verzögerung. Entweder geht ein Bild verloren oder der Decoder kann nicht zur richtigen Zeit ein neues Bild decodieren. Um dies zu verhindern, wird eine effektive Puffergröße im Encoder definiert, die nicht überschritten werden darf. Die effektive Puffergröße  $BS_{EFF}^E$  stimmt dabei nicht mit der Puffergröße  $BS_{max}^E$  überein, sondern kann geringer sein  $BS_{EFF}^E \leq BS^E$ .

Angenommen die Füllstände im Encoder und Decoder zum Zeitpunkt  $t_n = n \cdot t_{REF}$ , definiert als der Zeitpunkt des  $n$ -ten Bildes bezogen auf die Referenz-Framerate, werden durch  $bs_n^E$  und  $bs_n^D$  ausgedrückt.  $C_n$  ist die Bitmenge, die im Intervall zwischen den Bildern  $n$  und  $n+1$  über den Übertragungskanal transportiert werden kann. Für die folgende Untersuchung ist  $C_n$  konstant ( $C_n = C$ ).  $d_n$  entspricht der Anzahl der Bits, die der Encoder für Bild  $n$  produziert. Wenn weder der Encoder- noch der Decoderpuffer über bzw. leer läuft, dann gilt zum Zeitpunkt  $t_n$ :

$$bs_n^E = \sum_{j=1}^n d_j - (n \cdot C) \quad (2.20)$$

$$bs_n^D = \begin{cases} n \cdot C - \sum_{j=1}^{n-N} d_j, & \text{wenn } n \geq N \\ n \cdot C, & \text{wenn } n < N \end{cases}, \quad (2.21)$$

wobei der Decoder nach Erhalt der ersten Bits die Zeit  $N \cdot t_{REF}$  wartet, bis er das erste Bild decodiert.

- 
1.  $REF$ : Bezug auf die Referenz-Framerate  $REF\_FR$ .
  2.  $FR$ : Bezug auf Framerate.
  3. Herleitung über  $\Delta t^{EP} + \Delta t^{DP} = N \cdot t_{REF}$ .

Fasst man die Gleichungen (2.20) und (2.21) zusammen, also den Pufferfüllstand des Encoders und des Decoders zum Zeitpunkt  $t_n$ , erhält man:

$$bs_{n+N}^D = N \cdot C - bs_n^E, \text{ wenn } n \geq N \quad (2.22)$$

Wenn man nun  $N \cdot C - bs_n^E$  zu jedem Zeitpunkt  $> 0$  halten kann, dann wird auch  $bs_{n+N}^D$  immer  $> 0$  sein. Der Puffer im Decoder wird damit zu keinem Zeitpunkt leer laufen. Aus diesem Grund wird die effektive Puffergröße im Encoder wie folgt definiert:

$$BS_{EFF}^E = N \cdot C \quad (2.23)$$

Damit kann man garantieren, dass der Puffer im Decoder niemals leer läuft, wenn der Füllstand im Puffer des Encoders  $bs_n^E$  immer kleiner als  $BS_{EFF}^E$  ist. Aus der Gleichung für  $bs_{n+N}^D$  kann man erkennen, dass der maximale Wert für den Pufferfüllstand im Encoder  $bs_n^E$  gleich  $N \cdot C$  ist, wenn  $bs_n^D = 0$ . Daher reicht es für den Encoder aus, wenn er eine Puffergröße von  $N \cdot C = BS_{EFF}^E$  aufweist. Er kann dann in jedem Fall alle codierten Daten fassen, wenn die Ende-zu-Ende-Verzögerung eingehalten wird. Ein ähnlicher Ansatz gilt für den Decoder.

In einem Szenario mit variablen Übertragungsdatenraten können ähnliche Annahmen getroffen werden, allerdings ist dann die Datenrate des Übertragungskanals nicht konstant, sondern abhängig von der Zeit. Damit ergibt sich für die effektive Puffergröße im Encoder  $BS_{EFF_n}^E$ :

$$BS_{EFF_n}^E = \sum_{j=n+1}^{n+N} C_j \quad (2.24)$$

Damit hängt die effektive Puffergröße für eine konstante Ende-zu-Ende-Verzögerung von der zukünftigen Kapazität des Übertragungskanals ab. Sobald statistische Aussagen über das Übertragungsnetz getroffen werden können, kann dies in der Encodierung berücksichtigt werden.

Oftmals wird auch die effektive Puffergröße  $BS_{EFF}^E$  äquivalent zu ihrer Verzögerung angegeben, wobei mit konstanter Übertragungsdatenrate  $R$  und damit  $C = R \cdot t_{REF}$  gilt:

$$BS_{EFF}^E = N \cdot C = N \cdot R \cdot t_{REF} \quad (2.25)$$

Damit erhält man für die Pufferverzögerung des Encoders und Decoders  $\Delta t^{EP} + \Delta t^{DP}$ :

$$\Delta t^{EP} + \Delta t^{DP} = N \cdot t_{REF} = \frac{BS_{EFF}^E}{R} \quad (2.26)$$



### 2.3.2.2 Rate-Distortion-Modellierung in der Datenratenkontrolle

Für eine stabile Funktionsweise der Datenratenkontrolle und die a priori Festlegung des Quantisierungsparameters vor der Entropiecodierung in Abhängigkeit vom Pufferstand, wird meist die Rate-Distortion-Charakteristik des Encoders in Verbindung mit dem aktuellen Signal modelliert. Die Anpassung der Modellierung erfolgt dabei anhand einfacher Messgrößen des optional bewegungskompensierten Bildsignals  $x$ , wie dem  $MAD_x$ -Wert oder der Varianz  $\sigma_x^2$ . Dabei wird für einen rechteckigen Ausschnitt  $R$  des Signals  $x_R$  mit  $M \times N$  Pixel der mittlere absolute Fehler nach (2.2) und die Varianz  $\sigma_x^2$  nach

$$\sigma_x^2 = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [x_R(m, n) - \mu_x]^2 = P_x - \mu_x^2 \quad (2.27)$$

berechnet.

Der Mittelwert  $\mu_x$  und die Leistung  $P_x$  ergeben sich dabei zu:

$$\mu_x = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_R(m, n) \quad (2.28)$$

$$P_x = \frac{1}{M \cdot N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_R^2(m, n) \quad (2.29)$$

Insbesondere bei Verfahren ohne Mehrfachcodierung (Single-Pass-Verfahren) wird bei der Datenratenkontrolle ein zweistufiges Verfahren eingesetzt, das je nach Anforderung mehr oder weniger abgewandelt wird:

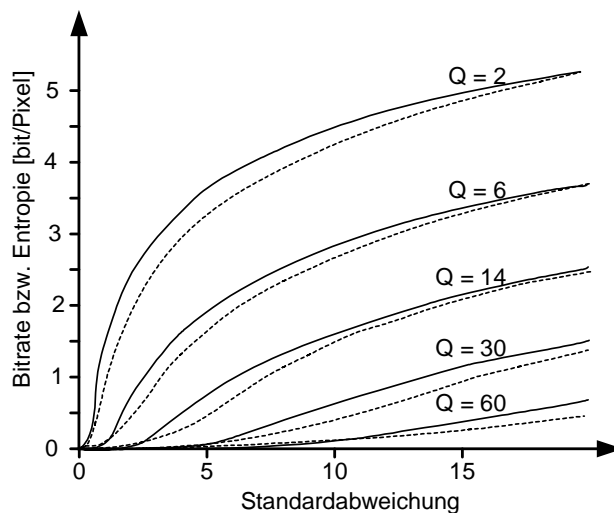
1. Bereitstellung einer Bitmenge (Bitallocation) für die Codierung eines oder auch mehrerer Bilder in Abhängigkeit des Pufferfüllstands für eine vorgegebene Datenrate. Dabei werden die Komplexität und die Codiermethode der Bilder geschätzt und berücksichtigt.
2. Festlegung des Quantisierungsparameters auf Bild- oder Makroblock-Ebene, um die Zielbitmenge zu erreichen. Für die Festlegung des Quantisierungsparameters auf Makroblock-Ebene werden die lokale Komplexität und die zur Verfügung stehende Bitmenge berücksichtigt.

Die analytischen Lösungen der Rate-Distortion-Funktion, die für Modellsignale mit stationärer Statistik und bei Verwendung des quadratischen Fehlerkriteriums existieren, werden für den Einsatz unter Echtzeitbedingungen vereinfacht, indem sie für den jeweils interessanten Bereich durch einfachere Kennlinien approximiert werden. In vielen Veröffentlichungen, wie [3], [21], [58], [100], [106] - [108], wird als Basis für die Abschätzung des Verhaltens der Bitrate des bewegungskompensierten Videosignals folgende Vereinfachungen angenommen: Die DCT-Koeffizienten des bewegungskompensierten Differenzbildes sind näherungsweise unkorreliert und entsprechen einer Laplace-Verteilung mit der Varianz  $\sigma^2$ .

Die Bitrate  $\hat{R}(Q)$  [Bit/Pixel] für den Luminanzanteil des bewegungskompensierten Differenzbildes, dessen Spektralkoeffizienten mit einem gleichförmigen Quantisierungsparameter der Höhe  $Q$  linear quantisiert werden, wird durch folgende Gleichungen approximiert:

$$\hat{R}(Q) = \begin{cases} \frac{1}{2} \log_2 \left( 2e^2 \frac{\sigma^2}{Q^2} \right), & \frac{\sigma^2}{Q^2} > \frac{1}{2e} \quad (1) \\ \frac{e}{\ln 2} \frac{\sigma^2}{Q^2}, & \frac{\sigma^2}{Q^2} \leq \frac{1}{2e} \quad (2) \end{cases} \quad (2.30)$$

In Abbildung 2.9 ist die Entropie  $H(Q)$  des quantisierten Signals als Funktion von  $\sigma$  für mehrere Quantisierungsparameter  $Q$  [93] und der nach (2.30) approximierten Bitrate  $\hat{R}(Q)$  dargestellt. Dabei gilt Modell (1) für kleine  $Q$ -Werte mit einer geringen Verzerrung oder für hohe Bitraten, sodass  $\sigma^2/Q^2$  größer als  $1/(2e)$  wird. Modell (2) gilt für große  $Q$ -Werte und einer damit verbundenen hohen Verzerrung oder entsprechend niedrige Bitraten, sodass  $\sigma^2/Q^2$  kleiner gleich  $1/(2e)$  wird. Das Modell (2) entsteht aus Modell (1) durch Linearisierung bezüglich  $\sigma^2$  und ist so definiert, dass die beiden Modelle im Punkt  $\sigma^2/Q^2 = 1/(2e)$  stetig ineinander übergehen.



**Abbildung 2.9:** Entropie einer Quelle, deren quantisiertes Ausgangssignal einer Laplace-Verteilung entspricht (durchgezogene Linien) und der durch (2.30) approximierten Bitrate (gestrichelte Linien) [106].

In [59], [66], [121] wird die Bitrate  $R(Q)$  für das bewegungskompensierte Quellsignal als näherungsweise laplaceverteilt angenommen. Als Approximation dient:

$$\hat{R}(Q) = \frac{(X_1 \cdot MAD)}{Q} + \frac{(X_2 \cdot MAD)}{Q^2} \quad (2.31)$$

Dabei wird die mittlere absolute Differenz  $MAD$  nach (2.2) berechnet und die Parameter  $X_1$  und  $X_2$  dienen zur Anpassung an die Charakteristik des Encoders in Verbindung mit dem aktuellen Bildsignal. Andere Ansätze, wie z.B. [103], schätzen die Bitrate  $\hat{R}(Q) \sim MAD$  für  $Q = const.$  oder  $\hat{R}(Q) \sim 1/Q$  bzw.  $\hat{R}(Q)$  in Abhängigkeit von  $1/Q$  und  $1/Q^2$ .

Als Abschätzung für die Verzerrung  $\hat{D}(Q)$  in Abhängigkeit des Quantisierungsparameters der Höhe  $Q$  dient oftmals die für höhere Bitraten bzw. kleine Verzerrungen und gleichförmige Quantisierung wohlbekannte Annahme [3], [22], [95]:

$$\hat{D}(Q) = \frac{Q^2}{12} \quad (2.32)$$

In [103] wird die Schätzung der Verzerrung, ebenso wie die Bitrate, in Abhängigkeit des mittleren absoluten Fehlers  $MAD$  nach (2.2) berechnet:  $\hat{D}(Q) \sim MAD$  für  $Q = const.$  oder ebenfalls  $\hat{D}(Q) \sim Q$  bzw.  $\hat{D}(Q)$  in Abhängigkeit von  $Q$  und  $Q^2$ .

### 2.3.3 Realisierungsansätze für die Datenratenkontrolle

In der Literatur sind bereits unterschiedliche Ansätze für die Datenratenkontrolle für Videocodern vorgestellt worden. Anhand ihrer Arbeitsweise und Einsatzbereiches ist eine Klassifizierung in echtzeitfähige und nicht echtzeitfähige Verfahren möglich. Neben dem Entwurf kompletter Ansätze für die Datenratenkontrolle sind auch zahlreiche Einzelaspekte oder auch Lösungen für spezielle Anwendungsbereiche, wie z.B. gezielte Optimierung auf mobile Netze, betrachtet worden. Diese Gruppe wird im Folgenden extra aufgeführt.

#### 2.3.3.1 Nicht echtzeitfähige Ansätze

Für Videoencodierung ohne Echtzeitanpruch, wie z.B. die Codierung von Inhalten für Video-CDs, Streaming-Video oder für die Archivierung sind bereits zahlreiche Ansätze beschrieben worden, so [16], [17], [97] - [99], [118] oder [125]. Meist hat der Encoder hier Zugriff auf die ganze Videosequenz. So ist es möglich, die Informationen über zukünftige Bilder zu berücksichtigen, um eine höhere Codiereffizienz zu erreichen. Szenenänderungen können erkannt werden und die Bit-allocation kann sich dynamisch an die Erfordernisse anpassen. Oftmals werden Bereiche auch mehrfach codiert, um gezielt die Rate-Distortion-Charakteristik des Encoders in Verbindung mit dem aktuellen Bild zu ermitteln, so z.B. [89] oder [90]. Im Anschluss wird die beste Methode ausgewählt. Für Kommunikationsanwendungen sind alle Verfahren, die eine Mehrfachcodierung eines Bildes erfordern, nur bedingt geeignet, da hier aufgrund der notwendigen Rechenleistung und der Einbeziehung zukünftiger Bilder eine zusätzliche Verzögerung entsteht.

### 2.3.3.2 Echtzeitfähige Verfahren

In den letzten Jahren sind zahlreiche echtzeitfähige Verfahren für die Datenratenkontrolle vorgestellt worden, so z.B. für Videokommunikation mit H.261 oder H.263 [3], [14], [29], [58], [75], [78], [105], [106], [112]-[114] für Echtzeitübertragungen mit MPEG-1 oder MPEG-2 [9], [15], [36], [139] bzw. MPEG-4 [11], [103], [104], [109], [120]. Dabei ist für jeden Anwendungsbereich das Videocodiervorgehen untersucht und die Verfahren für die Datenratenkontrolle hinsichtlich der sich daraus ergebenden Anforderungen, wie Datenrate, geforderte Ende-zu-Ende-Verzögerung oder die Komplexität des Bildmaterials, entworfen worden.

Insbesondere ein kleiner Ausgangspuffer mit der damit verbundenen geringen Ende-zu-Ende-Verzögerung stellt viele Verfahren vor Probleme. Häufig werden hier viele Bilder übersprungen oder die zur Verfügung stehende Übertragungsbandbreite kann nicht ausgenutzt werden, so z.B. [11], [19], [106], [119] und [131]. Zudem sind die Rate-Distortion-Modellierung und die Anpassung an das aktuelle Bildsignal auf den zu erwarteten Bitratenbereich und die Komplexität des Bildmaterials angepasst.

Ein genereller Nachteil aller echtzeitfähigen Single-Pass-Verfahren liegt in den hohen Modell- und Schätzfehlern. Bedingt ist dies dadurch, dass sie nur die Informationen, wie Komplexität der momentanen Szene oder der augenblicklichen Pufferbefüllung, aus vergangenen und dem aktuellen Bild für die Entscheidungsprozesse heranziehen können. Vor allem bei Szenenänderungen kommt es hier aufgrund der fehlenden Kenntnis über die Statistik der zukünftigen Bilder oftmals zu einer Verschlechterung der Bildqualität. Verfahren, die sich rasch auf die neuen Bedingungen einstellen können, haben hier Vorteile.

Für die Bitratenadaptation kann neben der Anpassung des Quantisierungsparameters auch die Framerate variiert werden. Allerdings erscheinen dadurch die Bewegungsabläufe oftmals nicht mehr fließend. Neben sehr einfachen Verfahren, wie z.B. [75], sind auch Lösungsansätze basierend auf der theoretischen Analyse des Bildsignals und des zugrunde liegenden Encoders in [10], [27] untersucht worden. Dabei zeigt sich, dass auch eine komplexe Rate-Distortion-Modellierung mindestens anhand von drei empirisch ermittelten Adaptionsparametern an einen realen Encoder angepasst werden muss. Ebenso wie bei einfacheren Modellen kann es damit auch hier aufgrund der fehlenden Kenntnis des Rate-Distortion-Charakters der zukünftigen Bilder zu hohen Schätzfehlern kommen.

Ein weiterer wichtiger Bewertungsfaktor für Echtzeitanwendungen ist die Komplexität des Verfahrens. So ist es z.B. bei [106] ausreichend als Aktivitätsmaß die Signalvarianz  $\sigma^2$  des bewegungskompensierten Bildsignals zu bestimmen. In [103] ist neben der Ermittlung des *MAD*-Wertes als Aktivität zudem noch die Berechnung der Verzerrung zwischen originalem und rekonstruiertem Bildsignal notwendig. Daraus ergibt sich für dieses Verfahren eine höhere algorithmische Komplexität.

### 2.3.3.3 Teilaspekte und spezielle Anwendungsbereiche

Neben dem Entwurf kompletter Verfahren für die Datenratenkontrolle sind auch zahlreiche Einzelaspekte betrachtet worden, so z.B. die Auswahl der Codierungsweise des nächsten Bildes [83], [85] und der Prädiktionmethode für MPEG [118] oder H.263 [127]. Der auf der Lagrange-Theorie basierende Ansatz in [116] bzw. [127] verspricht dabei durch die optimale Steuerung der Codeckontrolle eine hohe Bildqualität, der sowohl in H.263 als auch zusätzlich in H.264 implementiert worden ist. Ein weiterer Aspekt ist die Betrachtung des Bitallocation-Problems in Verbindung mit der Kontrolle des Füllstands des Ausgangspuffers, so z.B. [12], [13], [15], [36] und [87].

In vielen Veröffentlichungen sind spezielle Verfahren für die MPEG-2 Übertragung mit variabler Datenrate über ATM-Netze entworfen worden, so z.B. [13], [15], [36] oder [132]. Durch das Aushandeln von statistischen Kenngrößen zwischen Applikation und Netz, wie z.B. mittlere oder maximale Datenrate, kann die Bildqualität unabhängig von der aktuellen Szenenkomplexität weitgehend konstant bleiben, da die aktuelle Ausgangsdatenrate - in Grenzen - variieren kann. Zusätzlich kann auf den Ausgangspuffer verzichtet werden. Die Datenratenkontrolle muss dabei sicherstellen, dass die ausgehandelten Netzparameter von der Videocodierung eingehalten werden.

Vor allem für Videocodecs mit hoher Codiereffizienz sind Datenratenkontrollverfahren zur Übertragung über mobile Netze entworfen worden, so z.B. [26], [33]-[35], [91] und [113]. In [113] wird ein Verfahren vorgestellt, das durch die Anpassung der aktuellen Framerate, gezielt die Bildqualität auf Basis der Bewegungsinformation regelt. Das Verfahren in [34] nutzt die Information eines integrierten Kanalmodells mit Rückkanal und einer Rate-Distortion-Optimierung, um die beste Bitallocation für jedes Videobild anhand einer Variationsrechnung bzw. Lagrange-Optimierung herauszusuchen. Aufgrund der Komplexität ist es nicht für Echtzeitanwendungen geeignet.

In [91] wird ein Modell für die Datenratenkontrolle basierend auf einem ARQ<sup>1</sup>-Verfahren vorgeschlagen, das die Paketfehlerrate auswertet. Anhand dieser Informationen werden dann sowohl die Anzahl erzwungener Intra-Makroblöcke für die Fehlerbegrenzung als auch der Quantisierungsparameter für das nächste Bild festgelegt. Einen ähnlichen Ansatz verfolgt auch [3]: Hier werden über ein Feedback-Verfahren die Übertragungskanalinformationen ausgewertet, um die Anzahl der übersprungenen Bilder zu minimieren. Als Kanalmodell wird ein zweistufiges Markov-Modell benutzt.

---

1. Automatic Retransmission reQuest (ARQ).

### 2.3.3.4 Realisierungsbeispiele aus der Standardisierung

Innerhalb der Standardisierung von H.263 bzw. MPEG-4 sind verschiedene Verfahren für die Datenratenkontrolle miteinander verglichen worden. Die ausgewählten Verfahren sind in den jeweiligen Referenzcodecs implementiert worden und dienen im Weiteren als Bewertungsmaßstab für das in dieser Arbeit neu entwickelte Verfahren für die Datenratenkontrolle.

#### MPEG-4

Neben dem aktuellen Verification Model (VM) [66] wird das grundlegende Verfahren und dessen Implementierung z.B. in [59] und [121] beschrieben. Der Ansatz besteht aus den zwei Teilschritten: Bitallocation und Bestimmung des Quantisierungsparameters. In der Bitallocation wird anhand des aktuellen Pufferfüllstands und der Zieldatenrate die Bitmenge für das nächste Bild festgelegt. Zusätzlich wird noch die Bitrate des letzten Bildes berücksichtigt.

Die Bestimmung der Quantisierungsparameter basiert im Kern auf folgendem Ansatz:

$$\hat{B}_i = \frac{c_1 \cdot S_i}{Q_i} + \frac{c_2 \cdot S_i}{Q_i^2} \quad (2.33)$$

mit:

$\hat{B}_i$ : Bitmenge für das  $i$ -te Bild bzw. den  $i$ -ten Makroblock.

$c_1, c_2$ : Modellparameter, die aus der Bildstatistik ermittelt werden.

$Q_i$ : Quantisierungsparameter für das  $i$ -te Bild bzw. den  $i$ -ten Makroblock.

$S_i$ : *MAD* des  $i$ -ten Bildes bzw. des  $i$ -ten Makroblocks.

Je nach Anwendungsfall wird dieses Modell eingesetzt, um für eine bestimmte Bitmenge einen globalen Quantisierungsparameter für das gesamte Bild zu berechnen oder aber den Quantisierungsparameter pro Makroblock zu berechnen. Die Parameter  $c_1$  und  $c_2$  werden in einem zweistufigen Verfahren mittels Ausgleichsrechnung nach dem Prinzip des kleinsten quadratischen Fehlers auf Basis der Ergebnisse der bereits codierten Bilder bzw. Makroblöcke berechnet. Beide Verfahren greifen dabei auf jeweils bis zu 20 zurückliegende Codierergebnisse zurück. Beim ersten Durchlauf werden die Parameter  $c_1$  und  $c_2$  auf Basis aller Werte berechnet. Der zweite Durchlauf berücksichtigt dann für die Berechnung der Parameter nur noch die Werte, deren Abweichung zu (2.33) innerhalb einer bestimmten Schwelle liegen.

Für die Bestimmung des Quantisierungsparameters für den Makroblock  $i$  ist eine Aufteilung der Bitmenge des Bildes auf die einzelnen Makroblöcke notwendig. Hierzu dient folgender Ansatz:

$$\hat{B}_i^{MB} = \frac{S_i^{MB}}{S_{re_i}^F} \hat{B}_{re_i}^F \quad (2.34)$$

Dabei ist  $\hat{B}_i^{MB}$  die geschätzte Bitmenge und  $S_i^{MB}$  der  $MAD$ -Wert des Makroblocks  $i$ . Die Bitmenge  $\hat{B}_{re_i}^F$  und der  $MAD$ -Wert  $S_{re_i}^F$  werden nach der Codierung jedes Makroblocks ausgehend von der geschätzten Bitmenge für das aktuelle Bild  $\hat{B}^F$  bzw. dessen  $MAD$ -Wert  $S^F$  aktualisiert:

$$\hat{B}_{re_i}^F = \hat{B}^F - \sum_{j=0}^{i-1} d_j^{MB} \quad \text{bzw.} \quad (2.35)$$

$$S_{re_i}^F = S^F - \sum_{j=0}^{i-1} S_j^{MB} \quad (2.36)$$

$d_j^{MB}$  entspricht dabei der Bitmenge, die für die Codierung des Makroblocks  $j$  benötigt worden ist.

### H.263 (TMN8 bzw. TMN11)

In der aktuellen Version von H.263 bzw. in den Beschreibungen des Videocoderns TMN8 [75] und TMN11 [78] wird ein Verfahren eingesetzt, das speziell für niedrige Datenraten und eine geringe Ende-zu-Ende-Verzögerung entworfen worden ist. Eine Beschreibung des Algorithmus und dessen Implementierung befindet sich in [58], [59] und [106].

Das Verfahren teilt sich auf in die Bitallocation und die Bestimmung des Quantisierungsparameters auf Makroblockebene. Bei der Bitallocation wird die Bitmenge für das aktuelle Bild aus der aktuellen Bitrate und dem Pufferfüllstand ermittelt. Die Berechnung der Quantisierungsparameter auf Makroblock-Ebene basiert auf der Minimierung der prädiktiv geschätzten Verzerrung des aktuellen Bildes mittels der Lagrange-Theorie, unter Einhaltung der vorgegebenen Bitmenge.

Hierzu dienen folgende Annahmen: Auf Basis von (2.30) erhält man für die Bitratenschätzung  $\hat{B}_i^{MB}$  für den Makroblock  $i$  mit dem Quantisierungsparameter  $Q_i$  und der Standardabweichung  $\sigma_i$  für den Fall  $\sigma_i^2 / Q_i^2 \leq 1 / (2e)$ :

$$\hat{B}_i^{MB} = A \left( c \frac{\sigma_i^2}{Q_i^2} + \hat{B}_{over} \right)^2 \quad (2.37)$$

- 
1.  $MB$ : Bezug auf MakroBlock,  
 $F$ : Bezug auf ein Bild (Frame) bzw.  
 $re$ : Bezug auf restliche Werte im Bild (Frame).
  2.  $over$ : Bezug auf Overhead.

Der Adaptionsfaktor  $c$  kann aus den Ergebnissen der Codierung gewonnen werden und passt das Modell an die aktuelle Signalstatistik an. Der Faktor  $A$  entspricht der Anzahl der Luminanz-Pixel in einem Makroblock.  $\hat{B}_{over}$  entspricht dem aktuellen Overhead bei der Codierung pro Pixel, wie z.B. der Bitmenge für die Übertragung der Bewegungsvektoren oder der Steuerinformationen.

Die Abschätzung der Verzerrung  $\hat{D}^F$  für das aktuelle Bild mit  $N$  Makroblöcken erfolgt auf Basis der für höhere Bitraten bzw. kleinen Verzerrungen wohlbekannten Annahme [3], [22], [95]:

$$\hat{D}^F = \frac{1}{N} \sum_{i=1}^N \alpha_i^2 \frac{Q_i^2}{12} \quad (2.38)$$

Dabei wird der Makroblock  $i$  mit dem Quantisierungsparameter  $Q_i$  quantisiert und mittels des Faktors  $\alpha_i$  gewichtet. Die Verzerrung entspricht dem  $MSE$ , wenn gilt:  $\alpha_i^2 = 1; \forall i \in [1;N]$ .

### 2.3.3.5 Bewertung der Verfahren und Ausblick<sup>1</sup>

Für Kommunikationsanwendungen sind in der Regel aufgrund der geforderten geringen Ende-zu-Ende-Verzögerung nur echtzeitfähige Verfahren für die Datenratenkontrolle einsetzbar. Diese sind überwiegend auf einen begrenzten Anwendungsbereich oder für die Unterstützung spezieller Netze konzipiert. Die Untersuchung der Datenratenkontrollen aus der Standardisierung zeigt, dass die einzelnen Verfahren jeweils in unterschiedlichen Bereichen Vorteile aufweisen.

Die Berechnung der Quantisierungsparameter auf Makroblock-Ebene basiert beim Verfahren aus TMN8 auf einem analytisch lösbaeren Ansatz zur Minimierung der Verzerrung. Dabei wird die vorgegebene Bitmenge zur Codierung aller Makroblöcke im aktuellen Bild eingehalten. Dadurch kann der Füllstand des Ausgangspuffers gezielt festgelegt und somit die Zielframerate genau eingehalten werden. Die Rate-Distortion-Modellierung basiert auf Ansätzen aus der Literatur, so z.B. [3], [22], [95], [107] oder [108]. Allerdings führt die eingesetzte Rate-Distortion-Modellierung oftmals zu hohen Schätzfehlern, da ein Distortion-Modell gültig für hohe Bitraten und ein Bitraten-Modell gültig für niedrige Bitraten miteinander kombiniert werden.

Die gezielte Ausrichtung auf Kommunikationsanwendungen mit einer sehr kleinen Ende-zu-Ende-Verzögerung und dem hierzu notwendigen kleinen Ausgangspuffer spiegelt sich bei TMN8 auch bei den Ergebnissen wider. Untersuchungen auf Basis des in Kapitel 5 aufgestellten Testsets und der definierten Bedingungen zeigen im Vergleich zu den bei MPEG-4 eingesetzten bild- und makroblockbasierten Verfahren einen mittleren  $PSNR$ -Gewinn von bis zu 0,49 dB bei 100 ms bzw. 0,14 dB bei 250 ms Ausgangspuffergröße im Encoder.

Bei größerer Verzögerung führt sowohl die schlechte Nutzung des verfügbaren Ausgangspuffers als auch die Schätzfehler durch die verwendete Rate-Distortion-Modellierung zu einer Verschlechterung der mittleren  $PSNR$ -Werte von ca. 0,23 dB bei 1000 ms Ausgangspufferverzögerung. Dabei ist der Verlust nicht gleichmäßig über alle Sequenzen verteilt. In kritischen Bereichen kann das Verfahren auch bei großen Ausgangspuffergrößen oftmals eine hohe Codiereffizienz erzielen.

---

1. Die Bewertung der verschiedenen Verfahren für die Datenratenkontrolle basiert auf der Realisierung in einem H.263-Encoder und ist in Kapitel 5 unter verschiedenen Testbedingungen durchgeführt worden.



Für die Bestimmung eines globalen Quantisierungsparameters pro Bild wird bei MPEG-4 ein Verfahren beschrieben, das die Bitrate sowohl mit einer linearen als auch einer quadratischen Abhängigkeit vom Quantisierungsparameter modelliert. Auf Basis der bereits codierten Bilder wird dabei die Gewichtung für die linearen und quadratischen Komponenten berechnet. Dadurch weist die Modellierung der Bitrate eine hohe Flexibilität auf. Allerdings kann durch den globalen Quantisierungsparameter für ein ganzes Bild der Füllstand bei kleinen Ausgangspuffergrößen bzw. bei niedriger Framerate nicht genau festgelegt werden.

Die Ergebnisse des Testsets ergeben, dass das MPEG-4 Verfahren für die Bestimmung eines globalen Quantisierungsparameters für das gesamte Bild oftmals eine hohe und homogene Bildqualität erreicht. Allerdings ist das Verfahren für kleinere Ausgangspuffergrößen ( $< 500$  ms) nur bedingt geeignet, da hier sehr oft einzelne Bilder oder auch ganze Sequenzabschnitte übersprungen werden müssen, um ein Überlaufen des Puffers zu verhindern. Aufgrund der bildbasierten Arbeitsweise ist die algorithmische Komplexität im Vergleich zu den makroblockbasierten Verfahren gering.

Das MPEG-4 Verfahren für die Bestimmung der Quantisierungsparameter pro Makroblock ergänzt das bildbasierte MPEG-4 Verfahren um eine Verteilung der Bitmenge auf die einzelnen Makroblöcke. Dies wird auf Basis einer analytisch lösbaren Rate-Distortion-Modellierung analog zu dem Verfahren aus TMN8 realisiert. Für die anschließende Berechnung des Quantisierungsparameters dient das Bitratenmodell von MPEG-4. Daher kann es ebenso wie beim Verfahren aus TMN8 zu hohen Schätzfehlern bei der Aufteilung der Bitmenge auf die Makroblöcke kommen. Die Berechnung der Gewichtungskoeffizienten auf Basis weniger Makroblöcke für die linearen bzw. quadratischen Komponenten kann dabei zu un stetigen Modellverläufen und damit teilweise zu Fehlern bei der Bestimmung der Quantisierungsparameter führen.

Die Anwendung des Verfahrens für die Bestimmung des Quantisierungsparameters auf Makroblock-Ebene ergeben die Untersuchungen gute Ergebnisse, wenn die Ausgangspuffergröße mindestens 250 ms beträgt. Bei Puffergrößen über 500 ms liefert es allerdings eine geringfügig schlechtere Codiereffizienz (im Mittel ca. 0,1 dB schlechtere *PSNR*-Werte) als die bildbasierte Berechnung des Quantisierungsparameters. Aufgrund der notwendigen Berechnungen auf Makroblock-Ebene weist das Verfahren die höchste algorithmische Komplexität im Vergleich zu den anderen Verfahren auf. Zudem führt die Anpassung an die lokale Bildsignalcharakteristik zu größeren Qualitätsschwankungen, auch innerhalb von einzelnen Bildern.

Neue Videocodecs, wie z.B. MPEG-4, unterstützen ein breites Anforderungsprofil. Daher entfällt die Möglichkeit, die Datenratenkontrolle fest auf bestimmte Bereiche, mit weitgehend festen Randbedingungen ausrichten zu können. Die Kombination mehrerer Verfahren jeweils für spezielle Einsatzbereiche und die hierfür notwendige Rate-Distortion-Modellierung führt zu einer deutlich erhöhten algorithmischen Komplexität.

Unterschiedliche Rate-Distortion-Modelle müssen hierzu mit den jeweiligen Schätzparametern sowohl implementiert als auch parametrisiert werden. Zudem muss ein stufenloser Übergang zwischen den verschiedenen Modellen gewährleistet sein.

Aus diesem Grund ist es notwendig, flexible Verfahren für die Datenratenkontrolle zu untersuchen bzw. zu konzipieren. Gefordert ist ein Verfahren, das eine effiziente Codierung unter unterschiedlichen Randbedingungen, wie z.B. Zieldatenrate, Ende-zu-Ende-Verzögerung oder Komplexität des Bildmaterials, sicherstellt und zudem eine geringe Komplexität aufweist.

## 2.4 Zusammenfassung

Seit über 10 Jahren bildet bei allen Standardisierungen von Videocodecs ein Hybrid-DCT-Video-codierverfahren mit Bewegungskompensation die Basis. Sowohl MPEG-4 als auch H.264/MPEG-4 AVC basieren in ihrer Grundfunktionalität auf H.263. Durch Erweiterungen, insbesondere bei der Bewegungsschätzung, konnte die Codiereffizienz gesteigert werden. Zusätzlich ist die Funktionalität bei den aktuellen Standards erhöht worden. Während H.261, MPEG-1 und MPEG-2 speziell für die jeweiligen Anwendungsbereiche konzipiert worden sind, unterstützen aktuelle Standards ein breites Anwendungsspektrum und stellen eine effiziente Übertragung über mehrere Netze sicher.

Die hierfür notwendige Adaptivität und der Einsatz unterschiedlicher Codierverfahren erhöhen deutlich den Kontroll- und Steueraufwand im Encoder. Daher werden an die qualitätsbestimmenden Module des Videoencoders, wie die Codeckontrolle und Datenratenkontrolle, hohe Anforderungen gestellt.

Die Aufgabe der Datenratenkontrolle ist die Anpassung der Ausgangsdatenrate des Videoencoders an die Bandbreite des Übertragungsnetzes. Dadurch kann eine verzögerungsbegrenzte Ende-zu-Ende-Kommunikation gewährleistet werden. Der resultierende Bitstrom des Encoders wird bei der Übertragung zwischengespeichert. In Abhängigkeit vom aktuellen Pufferstand und dem Detailgehalt des Bildes wird die Verzerrung durch die Wahl des Quantisierungsparameter erhöht bzw. verringert. Die Größe der Puffer hat einen maßgeblichen Einfluss auf die Ende-zu-Ende-Verzögerung bei der Videokommunikation.

Durch eine Approximation der Rate-Distortion-Charakteristik des Videosignals in Verbindung mit dem Encoder für die Berechnung der Quantisierungsparameter kann sichergestellt werden, dass die Datenratenkontrolle die notwendige Stabilität aufweist. Die Rate-Distortion-Theorie ist seit Shannon bekannt. Eine analytische Lösung existiert allerdings nur für Sonderfälle. Für die Anwendung in der Datenratenkontrolle, speziell für echtzeitfähige Verfahren, ist es notwendig, die Modelle anhand von Annahmen zu vereinfachen und an die aktuelle Bildsignalcharakteristik anzupassen.

Bisher sind die Verfahren für die Datenratenkontrolle überwiegend auf einen eng begrenzten Anwendungsbereich oder für die Unterstützung spezieller Netze konzipiert worden. Dabei unterstützen neue Videocodecs ein breites Anforderungsprofil. Aus diesem Grund ist es notwendig, neue Verfahren für die Datenratenkontrolle zu entwickeln. Die Zielsetzung ist eine effiziente Datenratenkontrolle, die eine gleich bleibende Codiereffizienz unter unterschiedlichen Randbedingungen, wie z.B. Zieldatenrate, Ende-zu-Ende-Verzögerung oder Komplexität des Bildmaterials, sicherstellt.

## **3 Systemmodellierung von neuen Verfahren zur Datenratenkontrolle**

Der Schwerpunkt dieses Kapitels ist die Konzeption eines Systemmodells einer echtzeitfähigen Datenratenkontrolle für die Videokommunikation. Als Basis für die Beschreibung der Funktionalität des Systemmodells dient die Analyse der Anforderungen und die Untersuchung des Rate-Distortion-Verhaltens im Anwendungsbereich. Im Abschnitt 3.2 werden hierzu die Anforderungen und damit auch die Randbedingungen an die Datenratenkontrolle spezifiziert. Aufbauend auf der Untersuchung der Rate-Distortion-Charakteristik über einen breiten Datenratenbereich werden im Abschnitt 3.3 die Eignung existierender Rate-Distortion-Modelle bewertet und Verbesserungsansätze aufgezeigt. Auf Basis der Ergebnisse wird im Abschnitt 3.4 eine Funktionsspezifikation auf Systemebene für ein neues Verfahren für die Datenratenkontrolle aufgestellt und ein entsprechender Aufbau vorgeschlagen. Die Zusammenfassung der wesentlichen Ergebnisse des Kapitels befindet sich im Abschnitt 3.5.

### **3.1 Einführung**

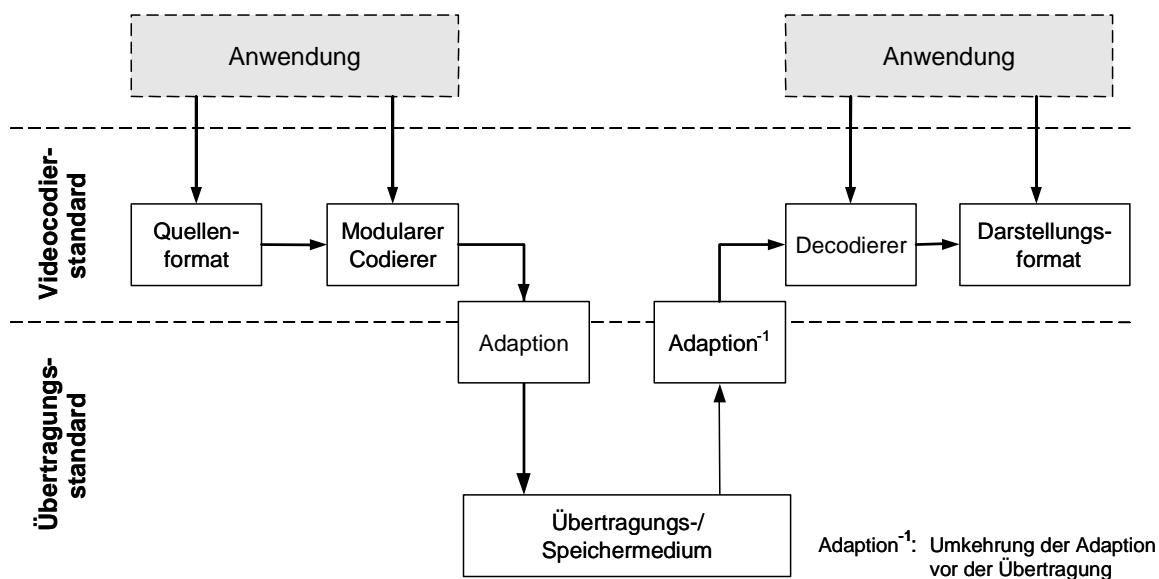
Besonderes Interesse bei der Entwicklung von Videostandards gilt der Erschließung neuer Einsatzbereiche. Für jede Anwendung sind bisher spezialisierte Verfahren für die Datenratenkontrolle entwickelt oder bestehende Modelle angepasst worden.

Ausgangspunkt für die Entwicklung der Datenratenkontrolle sind dabei oftmals statistische Modelle, wie z.B. Quellen, die einer Gauß- oder Laplace-Verteilung folgen. Insbesondere für Kommunikationsanwendungen mit Echtzeitanforderungen werden diese Modelle dann im jeweils interessanten Bereich durch einfachere Kennlinien approximiert, wie z.B. [11] bzw. [106], und an das reale Verhalten des Encoders angepasst. Zudem werden je nach Einsatzbereich Sonderfunktionen, wie z.B. spezielle Netzcharakteristika oder Paketfehlerraten, ergänzt, so z.B. [3], [26], [34], [91] oder [113]. Eine Adaption auf neue Anwendungsbereiche ist aufgrund der Spezialisierung teilweise nicht oder oftmals nur ineffizient möglich.

Ein wesentlicher Aspekt bei der Standardisierung von MPEG-4 ist die Unterstützung unterschiedlicher Anwendungen und die Übertragung über unterschiedliche Netze. Hierzu ist es notwendig, die Videocodecs flexibel an unterschiedlichste Anwendungsszenarien anzupassen. Außerdem muss für die Übertragung eine Adaption an die teilweise zeitvarianten Randbedingungen des jeweiligen Kommunikationsnetzes stattfinden. Hierzu sind neue Konzepte nötig. Für deren Entwicklung ist es erforderlich, die Anforderungen an einen flexiblen Videocodec zu untersuchen.

## 3.2 Anforderungen und Randbedingungen

Die Videocodierung ist innerhalb eines Videokommunikationssystems das Bindeglied zwischen der Anwendung und dem Übertragungsnetz. Stellt die Videocodierung ihre Dienste unterschiedlichen Anwendungen für die Übertragung über verschiedene Netze zur Verfügung, so ergeben sich neue Anforderungen an den Videocodec. Er muss unter verschiedensten Randbedingungen bestmögliche Bildqualität liefern. Dabei müssen die Vorgaben der verschiedenen Komponenten des Videokommunikationssystems - Anwendung, Videocodec und Netz - eingehalten werden. Gegenüber dem Übertragungsmedium kann diese Flexibilität durch die Einführung einer Adaptionsschicht zwischen Videocodec und Netz erreicht werden, wie in Abbildung 3.1 dargestellt. Sie ermöglicht, einen allgemein gültigen Kommunikationscodec an verschiedene Übertragungsmedien anzupassen [96].



**Abbildung 3.1:** Einordnung der Videocodierung in ein Kommunikationssystem. Die Adaptionsschicht passt einen allgemein gültigen Videocodec an verschiedene Übertragungsmedien an [96].

Für die Datenratenkontrolle im Encoder sind hierzu neue Ansätze notwendig. Zielsetzung ist es, unterschiedlichstes Bildmaterial mit jeweils passender Ende-zu-Ende-Verzögerung über zeitvariante mobile Netze mit mehreren Kilobits pro Sekunde oder aber über hochbitratige Festverbindungen effektiv übertragen zu können.

Neben den klassischen Aufgaben für die Datenratenkontrolle, wie der Sicherstellung von homogener, bestmöglicher Bildqualität und der Einhaltung der Vorgaben, wie die Ende-zu-Ende-Verzögerung und mittlere Ausgangsdatenrate werden zusätzlich folgende Aspekte wichtig:

- Effiziente Codierung von variablem Bildmaterial. Dies kann sowohl im Format als auch in der Komplexität variieren, wenn z.B. während einer Videokonferenz eine Sportsequenz übertragen wird oder die Anwendung wechselt.
- Unterstützung von unterschiedlicher Ende-zu-Ende-Verzögerung, die von der Anwendung vorgegeben wird. Damit verbunden auch die Begrenzung der Puffergröße bzw. der dadurch verursachten Verzögerung im Encoder.

- Gewährleistung einer variablen Ausgangsdatenratensteuerung, um bei mobilen Netzen mit teilweise schwankenden Datenraten eine effektive Übertragung sicher zu stellen.
- Unterstützung eines breiten Spektrums von Ausgangsdatenraten (von einigen kBit/s bis zu mehreren MBit/s).

Anhand der genannten Anforderungen von zukünftigen Kommunikationsszenarien bzw. der Vorschläge aus der Standardisierung von MPEG-4 und H.264, wie z.B. [59], [62], [64], [65], [70] und [77], können für die Analysen der verschiedenen Verfahren der Datenratenkontrolle folgende Einsatzbedingungen (siehe Tabelle 3.1) aufgestellt werden. Sie bilden zugleich die Basis für die Verifikation des neuen Verfahrens für die Datenratenkontrolle in Kapitel 5:

Kriterium	Wertebereich
Bereich der Ausgangsdatenrate:	10 kBit/s - 2 MBit/s
Zielbereich der Framerate:	7,5 - 30 fps
Verzögerung durch den Übertragungspuffer:	0,1 - 2 s (Kommunikationsanwendungen: bis ca. 0,5 s)
Videoeingangssignal:	Unterschiedlichste Charakteristik und Codeceignung (z.B. Detailreichtum oder Keraschwenks)
Ausgangsdatenrate:	Konstant oder variabel
Maximale Anzahl der Bits pro Bild:	Ausgangsdatenrate / 2 s (erstes Bild ist davon ausgenommen)
Minimale Framerate:	2 fps
Max. physikalischer Übertragungspuffer im Decoder:	In Abhängigkeit der max. Datenrate und des Profiles/Levels bei MPEG-4 zwischen $(10 \dots 80) * 16384$ Bits [65].

**Tabelle 3.1:** Anforderungen von Kommunikationsanwendungen an den Videocodec.

Neben der reinen Funktionsweise und der Stabilität sind im Besonderen folgende Kriterien beim Entwurf von neuen Verfahren zu berücksichtigen:

- Geringe Komplexität und einfache Implementierbarkeit.
- Effiziente Videocodierung über den gesamten Anwendungsbereich.
- Schnelle Adaptivität, um schnell auf variierende Bedingungen reagieren zu können.
- Geringe systembedingte Verzögerung.

Ein Videoencoder muss für die Übertragung mit fester Bitrate die Verzerrung in Abhängigkeit vom Bildinhalt variieren (siehe Kapitel 2). Für eine Bewertung der Leistungsfähigkeit von verschiedenen Verfahren für die Datenratenkontrolle ist es notwendig, Vergleichskriterien aufzustellen. Bei der Standardisierung von MPEG-4 sind folgende Kriterien als wichtig erachtet worden: gemittelter *PSNR*-Wert der codierten Bilder einer Sequenz, konstante Framerate, geringe Fluktuation der Bitmenge pro Bild, Vermeidung von Pufferüber- oder -leerlauf und das Erreichen der Zieldatenrate.

Zusätzlich soll die subjektive Bildqualität in die Bewertung mit einfließen, um eine homogene Gesamtqualität sicherzustellen. Ein weiteres wichtiges Kriterium ist die Verzögerung. Ihr Maximalwert wird im Wesentlichen durch die gewählte Ausgangspuffergröße bestimmt (siehe Abschnitt 2.3.2) und kann je nach Anwendung variieren.

Im Folgenden werden die Auswirkungen des breiten Anforderungsbereichs von Tabelle 3.1 hinsichtlich der Modellierung der Signalcharakteristika untersucht.

### 3.3 Signalcharakteristika und deren Modellierung

Die Analyse eines bewegungskompensierten Bildsignals zeigt, dass es sich aus drei unterschiedlichen Bestandteilen zusammensetzt [96]:

1. Neue Bildinhalte, für die keine Bewegungskompensation möglich ist. Meist überwiegen dabei tieffrequente Spektralanteile und es ist eine hohe örtliche Korrelation vorhanden.
2. Bestandteile, die durch die Rückkopplung des Codierungsfehlers entstehen.
3. Bestandteile, die durch ungenaue Bewegungskompensation verursacht werden.

Die beiden zuletzt genannten Komponenten weisen meist eine geringe örtliche Korrelation auf und die hochfrequenten Spektralanteile überwiegen. Aktuell ist noch kein Modell bekannt, das alle Bestandteile in ihrer Gesamtheit exakt beschreiben kann [96].

Für die Datenratenkontrolle muss der Quantisierungsparameter prädiktiv bestimmt werden. Dabei stellen die Unterstützung eines großen Bandbreitenspektrums, die Codierung von Bildmaterial mit variierender Signalcharakteristik und wechselnde Netzdatenraten hohe Anforderungen an die Datenratenkontrolle und damit an die zugrunde liegende Rate-Distortion-Modellierung.

#### 3.3.1 Analyse der Signalcharakteristik

Die Rate-Distortion-Charakteristik eines Videocodecs wird am Beispiel H.263 unter verschiedenen Bedingungen analysiert. Ziel ist die Eignung bekannter Rate-Distortion-Modelle anhand realistischer Messungen zu ermitteln. Interessant ist dabei die Rate-Distortion-Abhängigkeit vom Schätzparameter Varianz und die Häufigkeitsverteilung der Makroblöcke ebenfalls in Abhängigkeit der Varianz. Für die Bewertung existierender Rate-Distortion-Modelle sind die jeweiligen Schätzfehler und die Eignung für unterschiedliche Bitratenbereiche interessant.

In den Abbildungen 3.2, 3.3 und 3.4 sind zum Vergleich des Verhaltens bei verschiedener Signalcharakteristik die Rate-Distortion-Kennlinien der Sequenzen Mobile&Calendar, Mother&Daughter bzw. Weather dargestellt. Die Werte sind auf Basis eines H.263-Codecs ohne Optionen in einer IPPP...-Codierung mit jeweils konstantem Quantisierungsparameter für die P-Bilder ermittelt worden. Das erste I-Bild ist stets mit dem Quantisierungsparameter  $QP = 13$  entsprechend  $Q = 26$  codiert worden. Die Statistik der P-Bilder ist jeweils für die Quantisierungsparameter  $QP = 2, 6, 15, 22$  und  $30$  dargestellt worden, um ein breites Datenratenspektrum abzubilden. Dabei sind folgende Frameraten vorgegeben worden:  $30$  fps für  $QP = 2$  und  $6, 15$  fps für  $QP = 10$  und  $15$ ,  $10$  fps für  $QP = 22$  bzw.  $7,5$  fps für  $QP = 30$ .

In den Diagrammen (a) der Abbildungen 3.2, 3.3 und 3.4 ist jeweils die relative Häufigkeit der Makroblöcke in Abhängigkeit von der Standardabweichung dargestellt. Es ist für jeden Makroblock der Sequenz die Standardabweichung des Fehlersignals auf 0,5 gerundet und dem entsprechenden Wert zugeordnet worden. Die Diagramme (b) stellen jeweils den Verlauf des Adaptionparameters  $c$  für (2.37) gemittelt über ein Bild dar. Die Diagramme (c) und (e) visualisieren den Verlauf der Bitrate bzw. der Verzerrung in Abhängigkeit von der Standardabweichung. Analog zu den Diagrammen (a) ist hier die Standardabweichung pro Makroblock ermittelt und auf 0,5 gerundet worden. Allerdings sind hier pro Standardabweichungsstufe die Mittelwerte dargestellt.

Die mittleren Fehler zwischen der Schätzung der Bitrate bzw. der Verzerrung und den tatsächlichen Messwerten für die Makroblöcke des jeweiligen Bildes sind in den Diagrammen (d) bzw. (f) dargestellt. Die Fehlerberechnungen zwischen bestimmten und geschätzten Werten ist auf Basis der Schätzung der Bitrate nach (2.37) und der Verzerrung nach

$$\hat{D}^F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{Q_i^2}{a} \quad (3.1)$$

pro Bild mit  $N$  Makroblöcken ermittelt worden. Die Adaptionparameter  $a$  aus (3.1) bzw.  $c$  aus (2.37) sind mithilfe der Methode des kleinsten quadratischen Fehlers (z.B. [92], [110]) auf Basis der Werte der Makroblöcke des letzten Bildes berechnet worden.

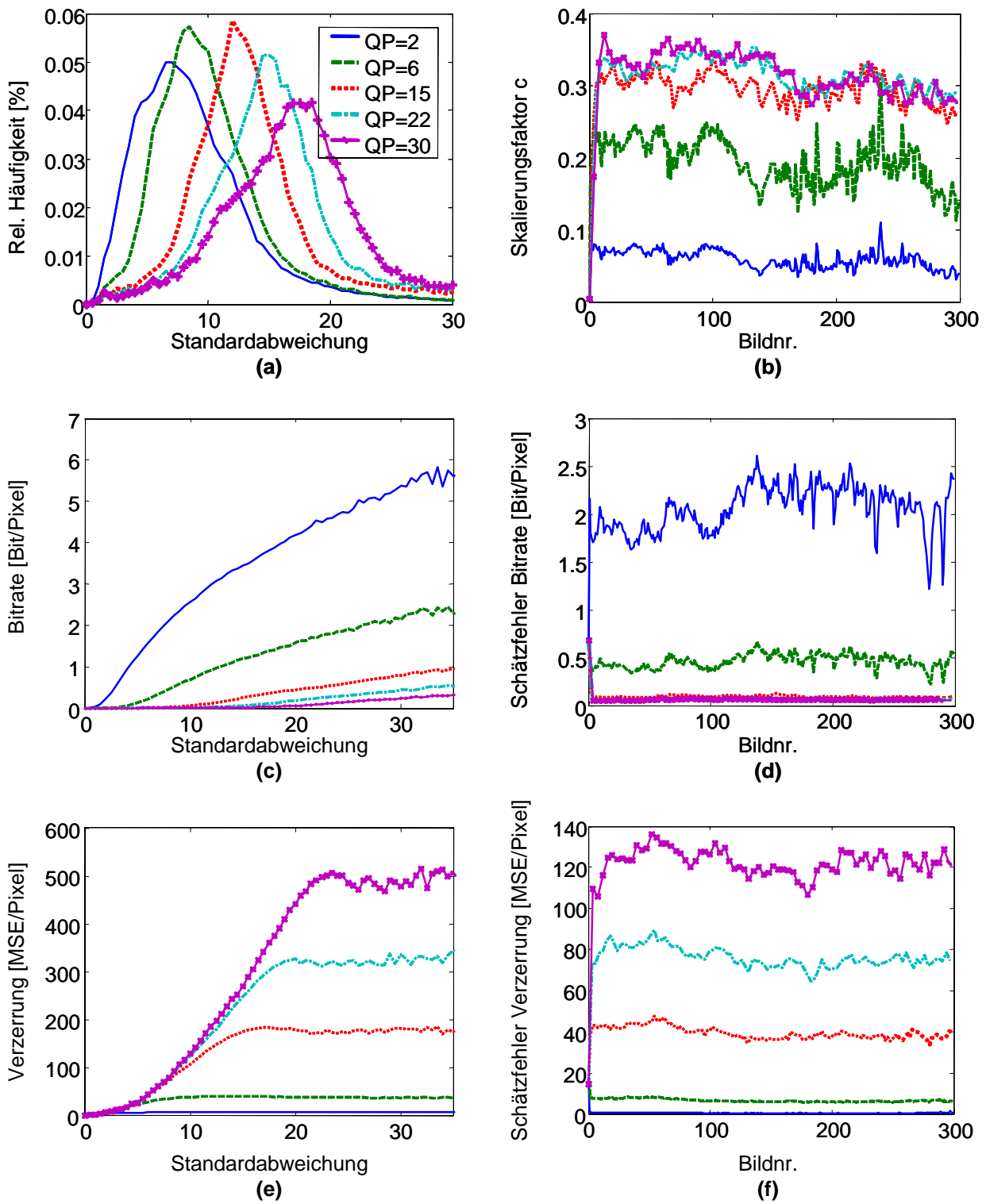
Die Auswertung der Abbildungen 3.2, 3.3 und 3.4 zeigt, dass sowohl für die Schätzung der Verzerrung als auch der Bitrate mittlere Fehlerraten von oftmals weit über 25% zu den gemessenen Werten auftreten (siehe jeweils die Diagramme (c) und (d) bzw. (e) und (f))<sup>1</sup>. Teilweise auch bis 100 %. Weiterhin zeigt sich, dass die im Abschnitt 2.3.2 angenommene Abhängigkeit zwischen Quantisierungsparameter und Verzerrung für die Mehrzahl der Makroblöcke nicht erfüllt ist (siehe Häufigkeitsverteilung in den Diagrammen (a) im Vergleich zum Verlauf der Kennlinie der Verzerrung in den Diagrammen (e)). In diesen Bereichen sind die Bedingungen für (2.32) nicht erfüllt. Hier ist die Schätzung der Verzerrung  $\hat{D}_{Qu}^2$  mit dem Adaptionparameter  $a$  in Abhängigkeit von der Varianz  $\sigma_i^2$  der einzelnen Makroblöcke  $i$  sinnvoll nach:

$$\hat{D}_{Qu} = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\sigma_i^2 \cdot Q_i^2}{a} \quad (3.2)$$

Der Verlauf des Adaptionparameters  $c$  in den Diagrammen (b) ist stark abhängig von der jeweiligen Bildsignalcharakteristik. Eine Prädiktion zwischen unterschiedlichen Quantisierungsparameter ist nur sehr bedingt möglich.

In den Bitratenverläufen (Diagramme (c)) ist sowohl der Bereich mit quadratischer Abhängigkeit von der Standardabweichung als auch der Bereich mit logarithmischer gut erkennbar (siehe Abbildung 2.9). Dabei ist der Bereich in Abhängigkeit des Quantisierungsparameters auffällig, der nahe der Nulllinie verläuft.

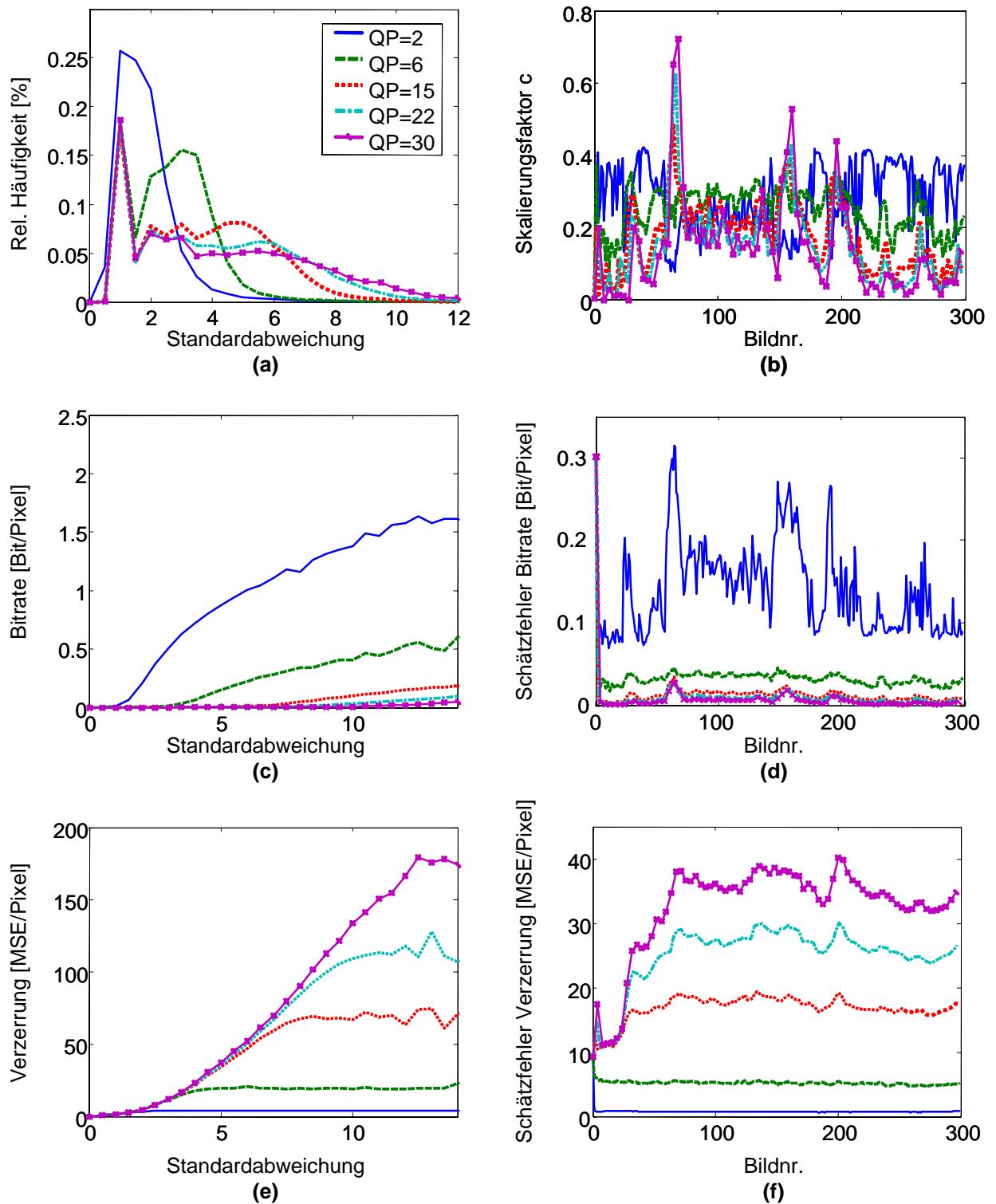
- 
1. Ermittlung der relativen Fehler bei der Bitratenschätzung in Abhängigkeit vom Quantisierungsparameter über den Vergleich der Diagramme (c) und (d) bzw. für die Schätzung der Distortion über den Vergleich der Diagramme (e) und (f) jeweils bei den Abbildungen 3.2, 3.3 und 3.4.
  2.  $Q_u$ : Bezug auf geringe Datenraten mit Quadratischem Verlauf der Verzerrungskennlinie.



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.  
 Die Legende für alle Diagramme befindet sich links oben.

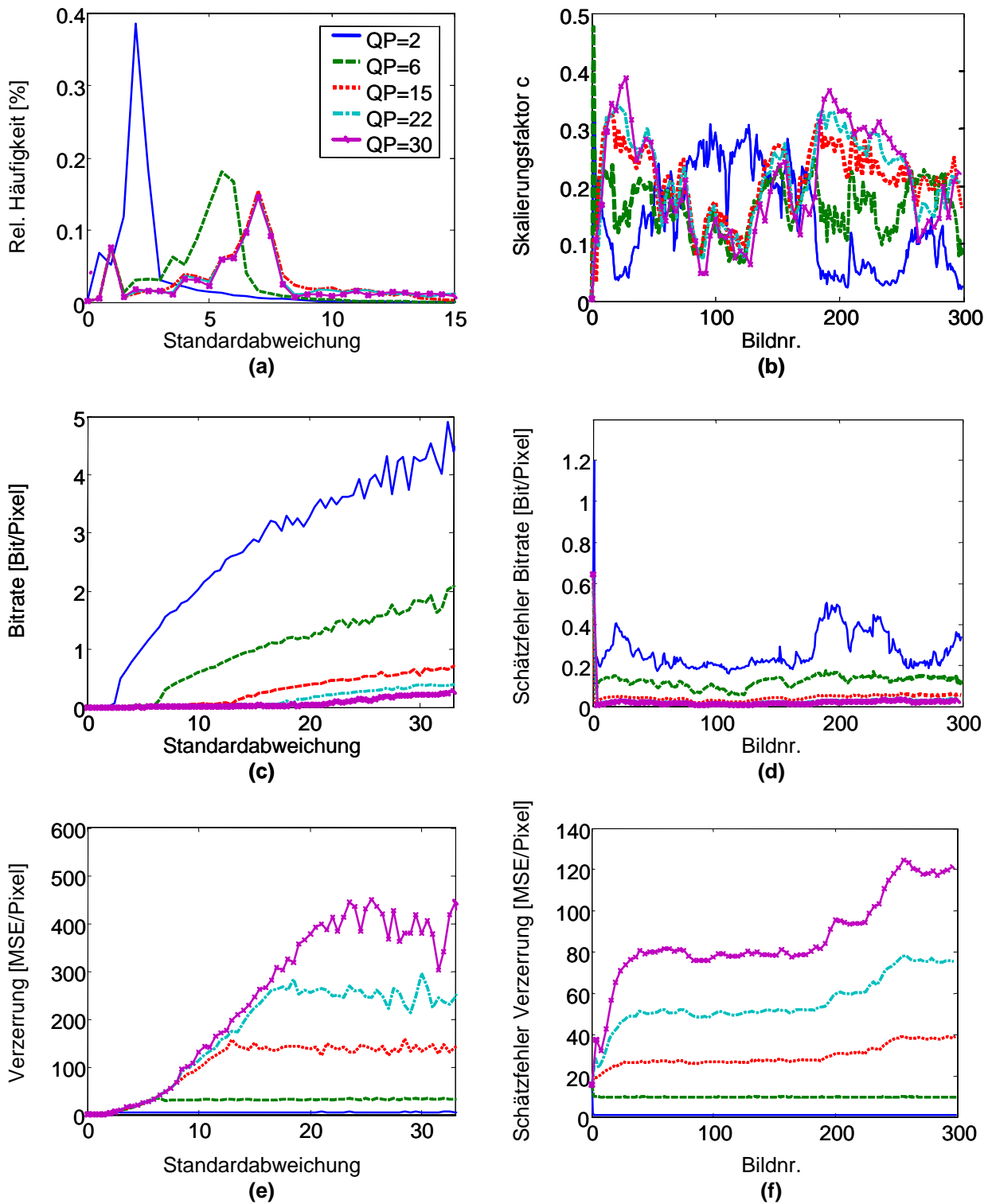
**Abbildung 3.2:** Rate-Distortion-Kennlinien der Sequenz Mobile&Calendar.





**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.  
Die Legende für alle Diagramme befindet sich links oben.

**Abbildung 3.3:** Rate-Distortion-Kennlinien der Sequenz Mother&Daughter.

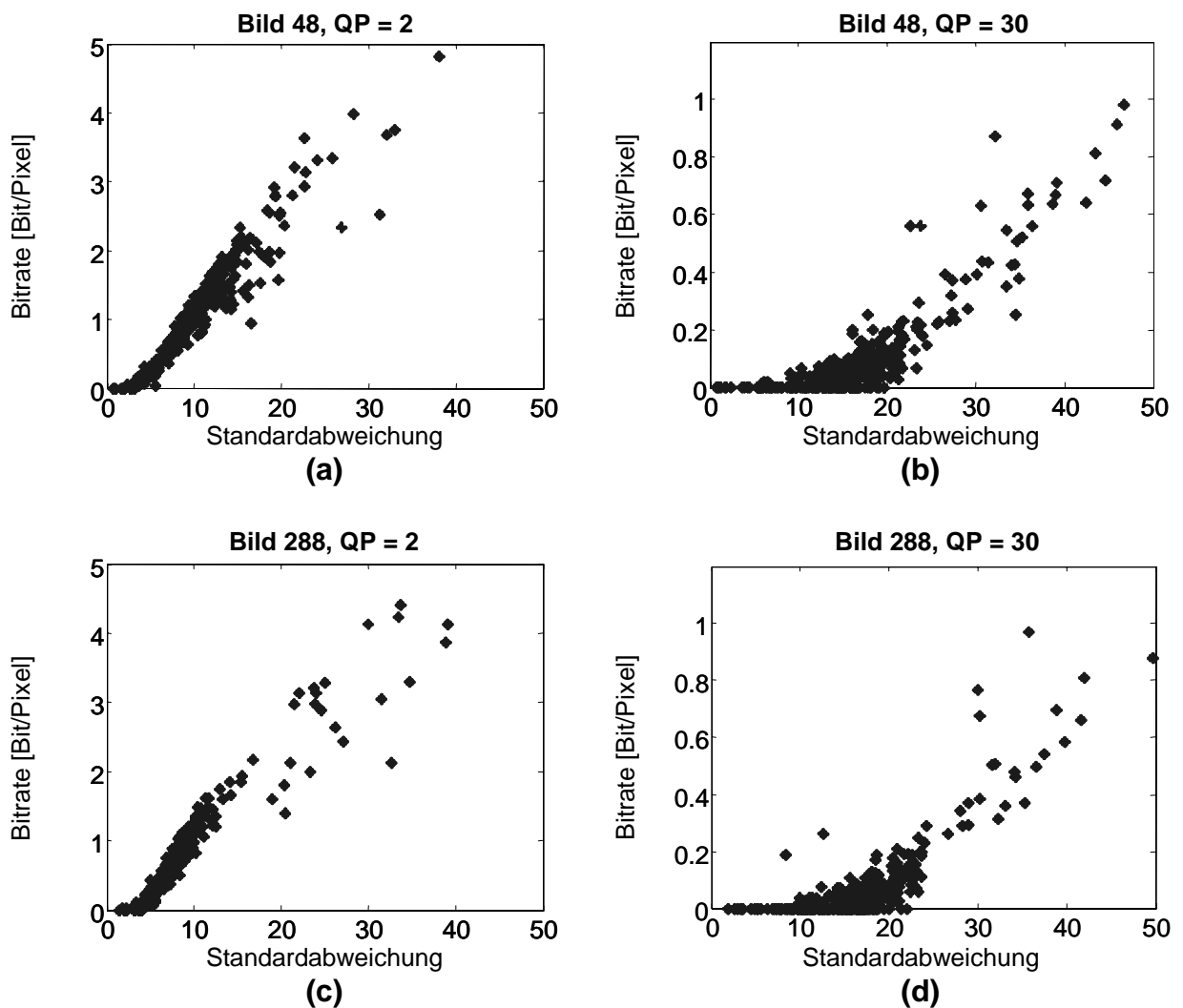


**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.  
 Die Legende für alle Diagramme befindet sich links oben.

**Abbildung 3.4:** Rate-Distortion-Kennlinien der Sequenz Weather.

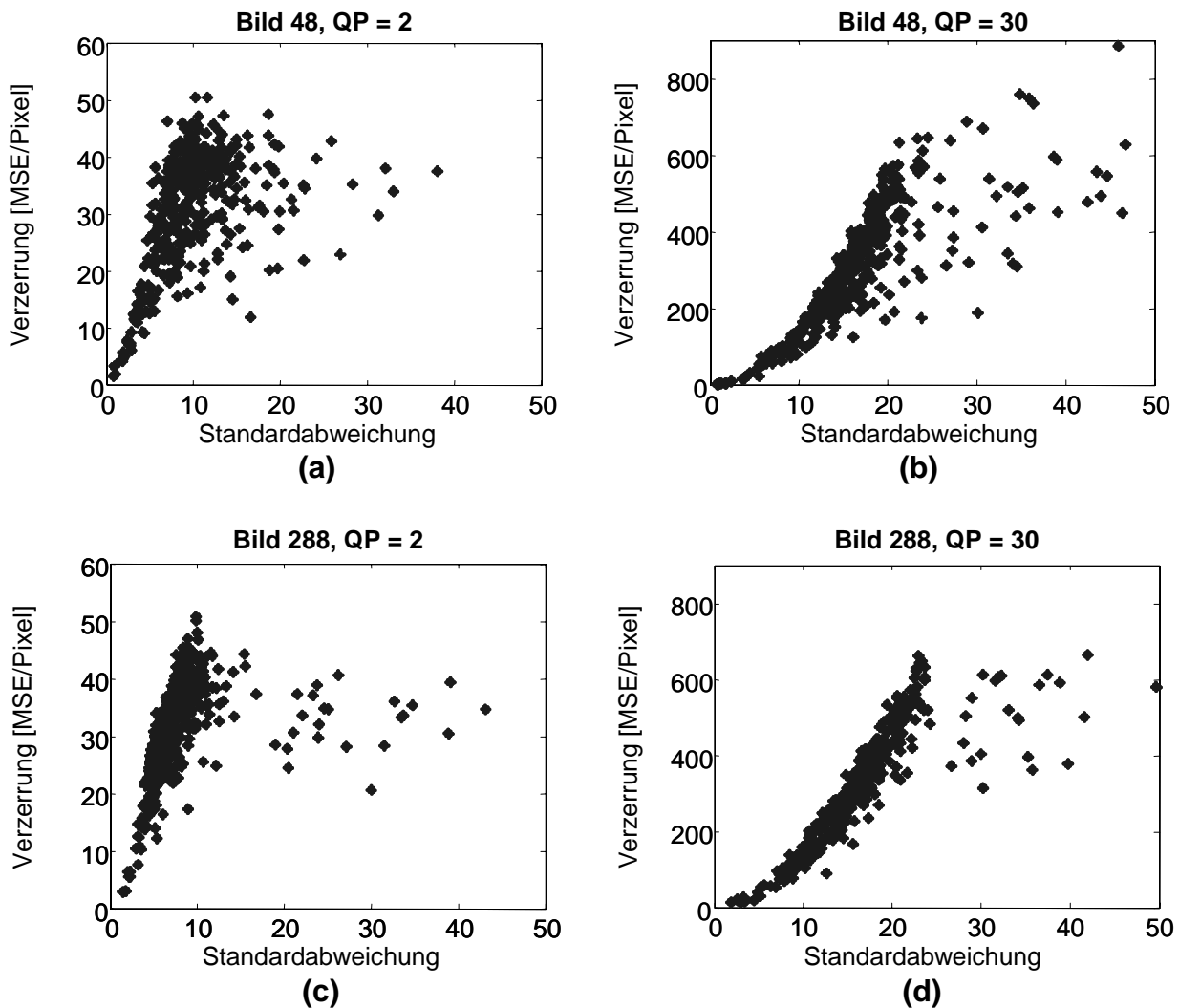
Für weitere Analysen sind die ermittelten Messwerte für die Bitrate bzw. die Verzerrung in Abhängigkeit von der Standardabweichung in den Abbildungen 3.5 bzw. 3.6 dargestellt. Die Werte pro Makroblock sind ebenfalls mittels eines H.263-Encoders ohne Optionen mit IPPP...-Codierung und konstanten Quantisierungsparametern ermittelt worden.

Im oberen Bereich der Abbildung 3.5 werden die Ergebnisse der Bitratenmessung für jeden Makroblock des Bildes 48 der Sequenz Mobile&Calendar für die Quantisierungsparameter  $QP = 2$  (Diagramm (a)) und  $QP = 30$  (Diagramm (b)) dargestellt. In den Diagrammen (c) und (d) sind die Werte entsprechend für Bild 288 der Sequenz Mobile&Calendar visualisiert. In der Abbildung 3.6 sind die Messergebnisse der Verzerrung analog dargestellt.



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.

**Abbildung 3.5:** Bitraten der einzelnen Makroblöcke pro Bild (Sequenz Mobile&Calendar).



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.

**Abbildung 3.6:** Verzerrung der einzelnen Makroblöcke pro Bild (Sequenz Mobile&Calendar).

Die Rate-Distortion-Charakteristik einzelner Bilder in den Abbildungen 3.5 und 3.6 entspricht dem gemittelten Verlauf der Gesamtsequenzen in den Abbildungen 3.2, 3.3 und 3.4. In den Diagrammen von Abbildung 3.6 ist deutlich zu erkennen, dass sich die Mehrzahl der Messwerte sehr gut mit der Schätzung der Verzerrung  $\hat{D}$  nach (3.2) bestimmen lassen.

Der Verlauf der Mehrheit der Messwerte in den Diagrammen von Abbildung 3.5 lässt sich näherungsweise mit der Bitratenschätzung in (2.37) beschreiben. Der Bereich der gemittelten Bitratenkennlinien in den Abbildungen 3.2, 3.3 und 3.4, der nahe der Nulllinie verläuft, ist durch die Abbildung 3.5 erklärbar. Mit größerem Quantisierungsparameter nimmt in Abhängigkeit vom Bildmaterial die Anzahl der Makroblöcke stark zu, bei denen alle DCT-Koeffizienten nach der Quantisierung null werden. Hier wird also kein Fehlersignal übertragen.

### 3.3.2 Interpretation und Zusammenfassung der Auswertung

Die Auswertung der Ergebnisse aus den Abbildungen 3.2 - 3.6 zeigt, dass die Beschreibung der Rate-Distortion-Kennlinien nach (2.37) und (2.32) hohe Fehlerraten bei der prädiktiven Schätzung der Bitrate und der Verzerrung gegenüber den ermittelten Werten liefert. Mittlere Abweichungen von 25% und mehr sind üblich. Eine Reduzierung des prädiktiven Schätzfehlers ist daher ein wichtiger Aspekt, um die Quantisierungsparameter in der Datenratenkontrolle genauer bestimmen zu können und damit die Codiereffizienz zu verbessern. Dies kann durch folgende Maßnahmen erreicht werden:

1. Einsatz von verbesserten Schätzmaßen für die prädiktive Beschreibung der Signalcharakteristik in Verbindung mit den Codeeigenschaften.
2. Verbesserte Modellierung der Rate-Distortion-Charakteristik des Encoders bzw. des Bildsignals.

Insbesondere für echtzeitfähige Kommunikationsanwendungen ist es notwendig, die Systemkomplexität der Datenratenkontrolle zu berücksichtigen. Die notwendige Rechen- und die Speichertransferleistung darf für die Bestimmung neuer Schätzmaße nicht deutlich über der Komplexität zur Bestimmung der Signalvarianz bzw. der *MAD*-Werte liegen. Für die Modellierung sind analytisch lösbare Ansätze ohne numerisch iterative Approximation und eine Beschränkung auf die Grundrechnungsarten notwendig.

Für die Schätzung der Verzerrung bietet sich in Abhängigkeit von der Bitrate und der Signalcharakteristik eine Kombination von Modell (3.1) und (3.2) an. Dabei muss ein stufenloser Wechsel zwischen den Modellen sichergestellt sein, um z.B. während der Encodierung auf Basis der Ergebnisse des letzten Bildes eine Modellauswahl für das nächste Bild vornehmen zu können. Einheitliche Modellparameter und Schätzparameter sind für eine Reduzierung der notwendigen Rechenleistung und Speicheranforderungen wünschenswert. Für eine Verringerung des Modellfehlers ist zudem eine schnelle Anpassung des Modells an die lokale Rate-Distortion-Charakteristik notwendig. Dabei zeigen z.B. die Ergebnisse in den Abbildungen 3.5 und 3.6, dass eine stabile Modelladaption auf Bildebene möglich ist.

Die prädiktive Erkennung der Makroblöcke, für die keine DCT-Koeffizienten übertragen werden, ist sowohl für die Erhöhung der Modellgenauigkeit als auch für eine Erhöhung der Codiereffizienz vorteilhaft und sollte berücksichtigt werden.

### 3.4 Systemansatz eines neuen Verfahrens für die Datenratenkontrolle

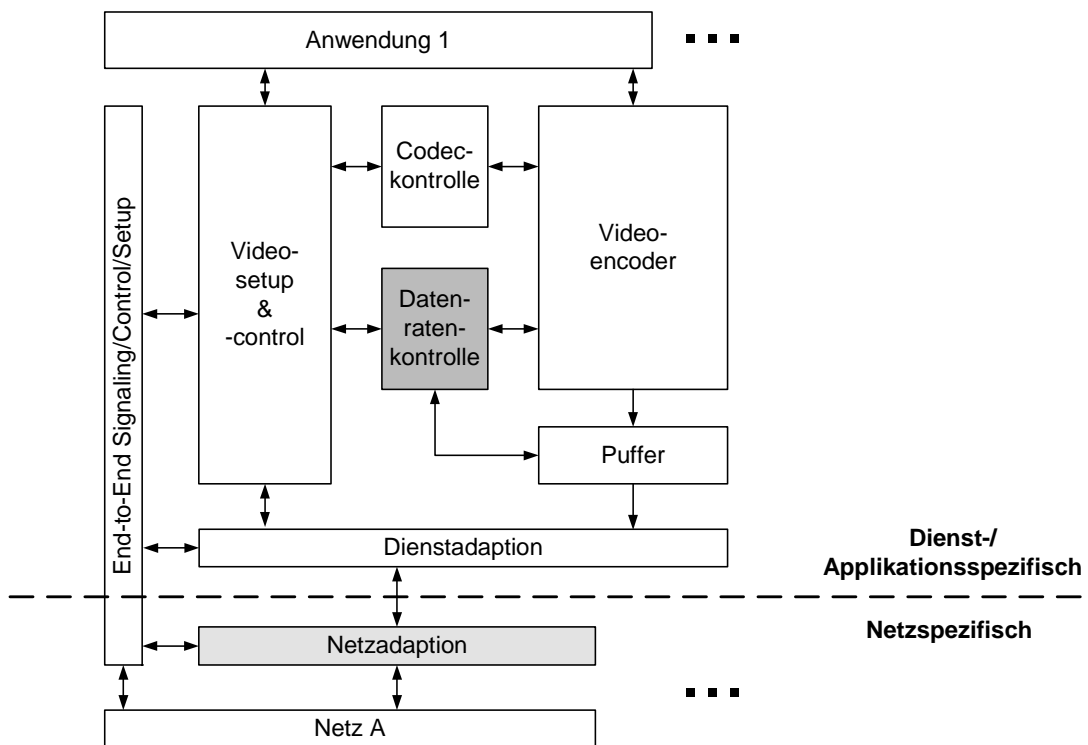
Die Zielsetzung für die Rate-Distortion-Modellierung des neuen Verfahrens für die Datenratenkontrolle ist die Minimierung der Modell- und Schätzfehler unter verschiedenen Randbedingungen, wie Komplexität des Bildmaterials oder unterschiedliche Ausgangsdatenraten. Dabei ist für die Anpassung an die lokalen Gegebenheiten eine hohe Adaptivität des Verfahrens und der zugrunde liegenden Rate-Distortion-Modellierung notwendig. Für die Bestimmung der Modellparameter direkt aus dem Encoder und die darauf aufbauende Berechnung der Quantisierungsparameter spielen folgende Aspekte eine wichtige Rolle:

- Der Einsatz von einfach zu berechnenden Schätz- und Modellparametern, um den notwendigen Rechenleistungsbedarf gering zu halten.
- Berücksichtigung von Vorgaben aus dem Videostandard, so kann z.B. bei H.263 der Quantisierungsparameter nur um max. 2 Stufen von Makroblock zu Makroblock variiert werden [73], sodass die Quantisierungsparameter nicht frei wählbar sind.
- Sicherstellung der Systemstabilität, so z.B.:
  - Bei Schätzfehlern, die sich aufgrund der prädiktiven Schätzung ergeben. Dies ist insbesondere dann problematisch, wenn sich die Bildstatistik in kurzer Zeit sehr stark ändert, wie z.B. bei einem Szenenwechsel oder bei einem Kameranachschwenk.
  - Bei Modellfehlern, die durch eine starke Abweichung des aktuellen Bildsignals von der angenommenen Statistik, aber auch durch eine gleichzeitige Veränderung der Bitrate und der Verzerrung auftreten können. In den Annahmen für die Rate-Distortion-Modellierung ist meist ein Parameter fix.
  - Bei variablen Datenraten, wie sie insbesondere bei mobilen Anwendungen auftreten.
- Wechselwirkung zwischen den verschiedenen Einflussfaktoren, sodass sich Einzeleffekte gegenseitig aufheben bzw. auch verstärken können.

Die Berücksichtigung aller genannten Effekte ist aufgrund deren Komplexität und Nichtlinearität ohne Mehrfachcodierung nicht möglich, sodass in echtzeitfähigen Videoencodern oftmals mit Näherungen oder empirischen Werten gearbeitet werden muss, so z.B. [27], [73], [75] oder [106].

Für eine einfache Adaption der Videocodierung auf unterschiedliche Netze und verschiedene Videoencoder ist ein modularer Ansatz für das neue Verfahren zur Datenratenkontrolle notwendig. In Abbildung 3.7 ist ein vereinfachtes Schnittstellenkonzept für ein Videoterminal dargestellt. Die Datenratenkontrolle kommuniziert dabei über die "Videosetup&-control"-Einheit mit der Anwendung und dem Netz.

Die Adaption der Videocodierung auf die verschiedenen Netze ist dabei zweigeteilt: in eine dienstspezifische und eine netzspezifische Anpassung mit entsprechender Aufgabenteilung [57]. Aufgrund der unterschiedlichen Eigenschaften der verschiedenen Netze, wie variierende Datenrate durch Paketwiederholungen, variable Fehlerkorrektur oder zeitvariantem Übertragungskanal, ist es sinnvoll, im netzspezifischen Teil die aktuell zur Verfügung stehende Datenrate und deren zukünftige Entwicklung zu ermitteln.



**Abbildung 3.7:** Schnittstellenkonzept eines Videocoders in Anlehnung an MPEG-4 mit einer zweistufigen Adaption an verschiedene Übertragungsnetze.

Wie im Kapitel 2 beschrieben, ist es insbesondere für echtzeitfähige Encoder - ohne Mehrfachcodierung - sinnvoll, die Datenratenkontrolle in Anlehnung an den Videocodierprozess auf die zwei folgenden Stufen aufzuteilen: Bitallocation mit der GOP- und Bild-Ebene bzw. Makroblock-Ebene. Hierdurch wird eine einfache Erweiterbarkeit der Datenratenkontrolle erreicht. Für das weitere Vorgehen wird daher folgender zweistufiger Ansatz gewählt, der in Abbildung 3.8 dargestellt ist:

- Bei der Bitallocation wird die Bitmenge für die folgende Codierung eines oder mehrerer Bilder reserviert. Hierzu wird die Aktivität des oder der Bildtypen und der Pufferfüllstand für eine vorgegebene Bitrate berücksichtigt. Ebenso wird die geschätzte Komplexität der zukünftigen Bilder einbezogen. Zusätzlich wird überprüft, ob es aufgrund eines drohenden Pufferüberlaufs notwendig ist, Bilder bei der Codierung zu überspringen.
- Auf Makroblock-Ebene wird der Quantisierungsparameter in Abhängigkeit von der lokalen Aktivität und der zur Verfügung stehenden Bitmenge festgelegt. Randbedingung ist das Einhalten der Bitmenge für das aktuelle Bild, die von der Bitallocation vorgegeben wird, sodass kein Überlauf des Ausgangspuffers erfolgt. Ein Leerlaufen des Puffers kann durch das Einfügen von zusätzlichen Bits im Bitstrom verhindert werden, die der Decoder verwirft. Hierdurch wird ein Erreichen der Zielbitrate sichergestellt.

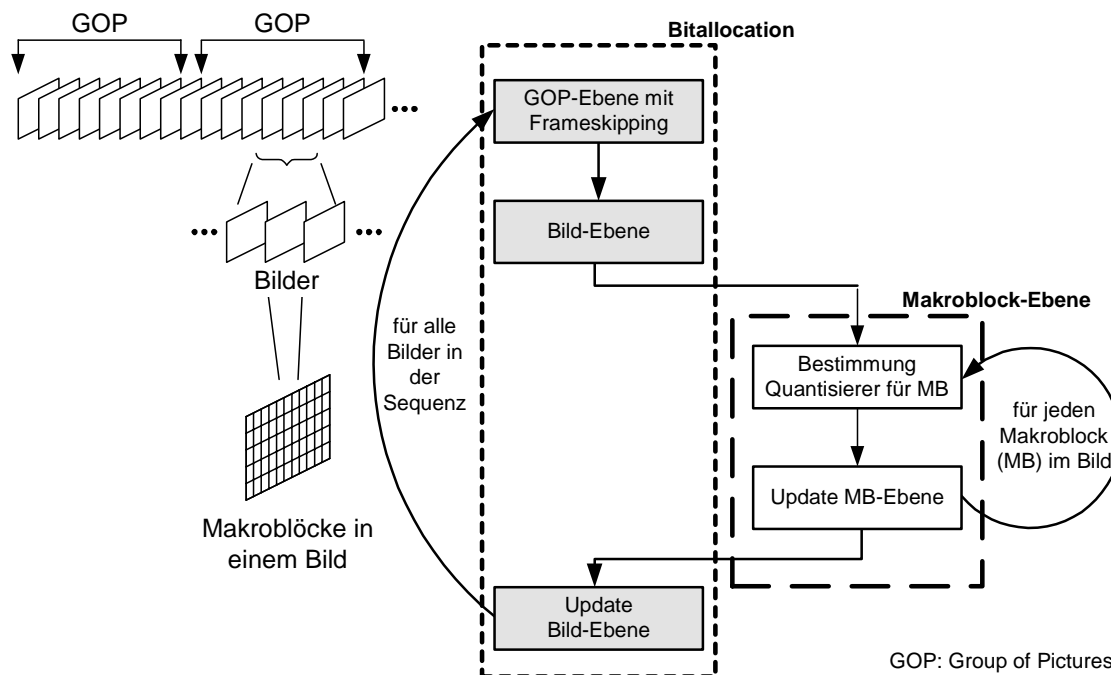


Abbildung 3.8: Prinzipieller Aufbau der Datenratenkontrolle.

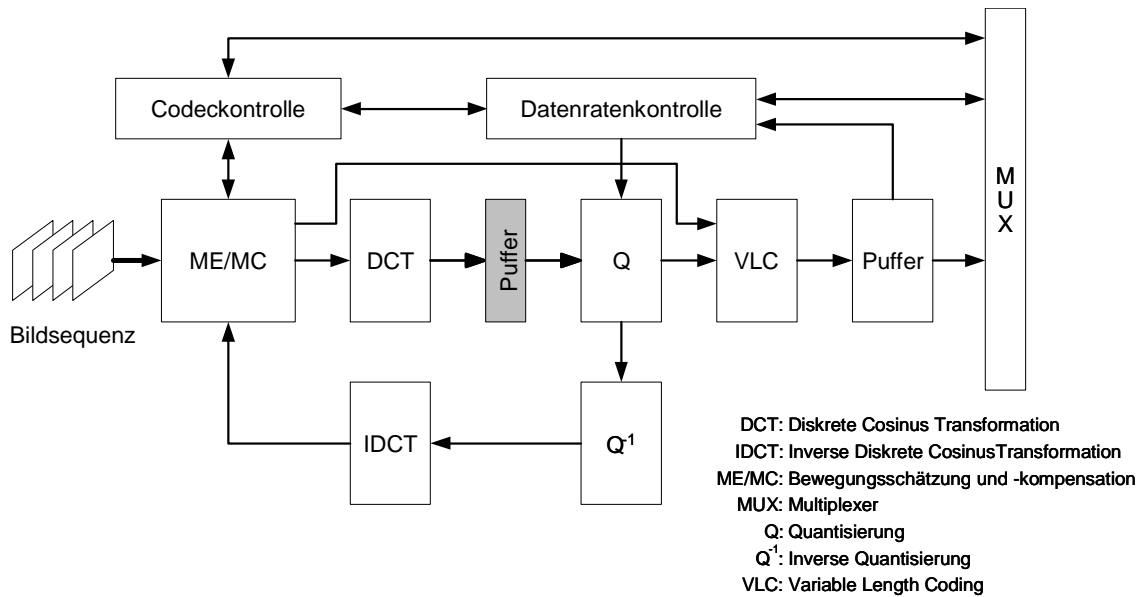
In Abbildung 3.9 sind die wichtigsten Funktionsmodule und der Datenfluss im Encoder mit Zwischenpufferung für eine Aktivitätsschätzung auf Bild-Ebene dargestellt. Nach der Bitallocation und der Festlegung des Typs für das nächste Bild wird für alle Makroblöcke im aktuellen Bild die Bewegungsschätzung und -kompensation durchgeführt. Im Anschluss werden auf Bild-Ebene die Schätzgrößen für die Datenratenkontrolle ermittelt. Optional kann dies auch nach der DCT-Transformation erfolgen. Nach Festlegung des jeweiligen Quantisierungsparameters auf Makroblock-Ebene wird die Codierung des aktuellen Makroblocks fertig gestellt. Darauf folgt die Codierung des nächsten Makroblocks, solange bis das Bild fertig codiert ist (siehe Abbildung 3.8).

Zur Sicherstellung einer effizienten Videocodierung ist eine stabile Arbeitsweise der Datenratenkontrolle notwendig. Hierzu muss gewährleistet sein, dass die Schätz- und Modellparameter nicht einseitig auf Basis spezieller Charakteristika von Bildteilen bestimmt werden. Zudem müssen die Ergebnisse der Datenratenkontrolle im laufenden Betrieb überwacht werden.

Alle in Kapitel 2 vorgestellten Verfahren für die Datenratenkontrolle bestimmten - für die Erhöhung der Genauigkeit - die Schätzparameter für alle Makroblöcke auf Bild-Ebene. Hierzu ist eine Zwischenspeicherung der teilweise codierten Makroblöcke notwendig, wodurch sich die Verzögerung des Videoencoders erhöht. Prinzipiell ist es sowohl bei H.263 als auch bei MPEG-4 möglich den aktuellen Makroblock ohne Informationen über zukünftige Makroblöcke zu codieren. Die Encodierung kann also rein makroblockbasiert ablaufen. Lediglich spezielle Optionen wie z.B. Global Motion Compensation erfordern prinzipbedingt eine Bearbeitung auf Bild-Ebene.

In Kapitel 4 und 5 werden für eine Vergleichbarkeit des neuen Verfahrens für die Datenratenkontrolle mit existierenden Verfahren auch die Schätzparameter auf Bild-Ebene bestimmt. In Kapitel 6 wird untersucht, inwieweit eine Verringerung der Verzögerung durch eine reine makroblockbasierte Bearbeitung möglich ist bzw. inwieweit sich die Codiereffizienz aufgrund der verschlechterten Rate-Distortion-Schätzung verändert.





**Abbildung 3.9:** Funktionsmodule im Encoder mit Zwischenpuffer für "vorcodierte" Makroblöcke.

Die funktionalen Anforderungen an das neue Verfahren für die Datenratenkontrolle ergeben sich direkt aus der erweiterten Funktionalität aktueller Videocodecs: effiziente Codierung von unterschiedlichem Bildmaterial mit jeweils passender Ende-zu-Ende-Verzögerung über zeitvariante mobile Netze mit mehreren Kilobit pro Sekunde oder aber über hochbitratige Festverbindungen. Das neue Verfahren für die Datenratenkontrolle muss unter diesen Randbedingungen effizient und stabil arbeiten. Zusätzlich ergeben sich folgende Anforderungen an den Systementwurf:

- Bestimmung der Quantisierungsparameter auf Makroblockebene, um die Verzerrung zu minimieren und der Bedingung die Zielbitmenge aus der Bitallocation einzuhalten.
- Eine geringe systembedingte Verzögerung, um die für Kommunikationsanwendungen notwendige niedrige Ende-zu-Ende-Verzögerung sicherstellen zu können.
- Echtzeitfähigkeit und geringe Systemkomplexität. Hierzu ist eine einfache Rate-Distortion-Modellierung notwendig, die unter den aufgestellten Bedingungen eine stabile und effiziente Videocodierung ohne iterative Näherungsverfahren ermöglicht. Aufgrund der Echtzeitanforderungen und der begrenzt zur Verfügung stehenden Rechenleistung kann keine Mehrfachcodierung durchgeführt werden.
- Ein möglichst breites Einsatzgebiet, das einen breiten Bitratenbereich, unterschiedliches Bildmaterial und auch die Codierung mit variablen Ausgangsdatenraten beinhaltet.
- Modulares Konzept für eine leichte Erweiterbarkeit und klare Schnittstellen zwischen Applikation, Videocodec mit Datenratenkontrolle und Netz.

Auf Basis der ermittelten Anforderungen wird in Kapitel 4 zunächst vertieft auf eine geeignete Rate-Distortion-Modellierung für den genannten Einsatzbereich eingegangen. Auf Basis dieser Modellierung und der funktionalen Systembeschreibung wird das neue grundlegende Verfahren für die Datenratenkontrolle entworfen. Im Anschluss werden Erweiterungsansätze für eine zusätzliche Steigerung der Effizienz bzw. der Funktionalität beschrieben.

## 3.5 Zusammenfassung

Der Schwerpunkt dieses Kapitels ist ein Systemkonzept für ein neues echtzeitfähiges Verfahren für die Datenratenkontrolle, das eine netz- und anwendungsunabhängige Videokommunikation ermöglicht. Als Basis für den Entwurf ist es notwendig, die Anforderungen und Randbedingungen, die sich aus den Vorgaben der Standardisierung flexibler Videocodecs ergeben, zu ermitteln. Als wichtigste Kriterien sind zu nennen: die Unterstützung eines breiten Spektrums von Ausgangsdatenraten (von einigen kBit/s bis hin zu mehreren MBit/s), eine effiziente Codierung von Bildmaterial unterschiedlicher Komplexität, die Unterstützung von variierender maximaler Ende-zu-Ende-Verzögerung und von zeitvarianten Ausgangsdatenraten.

In echtzeitfähigen Verfahren für die Datenratenkontrolle ist es notwendig, den Quantisierungsparameter in Abhängigkeit vom Bildinhalt a priori zu bestimmen. Hierzu muss die aktuelle Charakteristik des Bildmaterials in Verbindung mit dem Codec prädiktiv geschätzt werden. Durch eine Rate-Distortion-Modellierung kann dabei sichergestellt werden, dass die notwendige Stabilität gewährleistet ist. Dabei ist die Qualität der Modellierung ein wichtiger Faktor für die Effizienz der Videoencodierung. Die Verifikation bekannter prädiktiver Ansätze zeigt hohe Schätzfehler mit einer mittleren Abweichungen von 25% und mehr.

Dieser Aspekt wird bei der funktionalen Spezifikation des neuen Verfahrens berücksichtigt. Ebenso wie ein modularer Ansatz mit einer Funktionsaufteilung in die zwei Bereiche Bitallocation und Makroblockebene, eine geringe algorithmische Komplexität und systembedingte Verzögerung. Darüber hinaus ist es insbesondere für die Leistungsfähigkeit unter realen Bedingungen wichtig, dass eine rasche Adaption auf variierende Randbedingungen und an das aktuelle Eingangssignal stattfindet.

## **4 Vorschlag eines neuen Verfahrens für die Datenratenkontrolle**

In diesem Kapitel wird ein neues Verfahren für die Datenratenkontrolle eines Hybrid-DCT-Videoencoders vorgeschlagen. Aufgrund seiner Flexibilität erfüllt es die Anforderungen verschiedener Anwendungsbereiche und Übertragungsnetze im Bereich Echtzeitvideokommunikation. Nach einer Einführung wird im Abschnitt 4.2 ein Lösungsansatz für das neue Verfahren für die Datenratenkontrolle erläutert. Wichtig für die Leistungsfähigkeit des neuen Verfahrens ist neben der zugrunde liegenden Rate-Distortion-Modellierung die Anpassung an das aktuelle Eingangssignal. Das neue Verfahren für die Datenratenkontrolle wird im Abschnitt 4.3 beschrieben. Auf optionale Erweiterungsmöglichkeiten, wie sie bei aktuellen Videocodecs notwendig sind, wird im Abschnitt 4.4 eingegangen. Die wesentlichen Inhalte dieses Kapitels sind im Abschnitt 4.5 zusammengefasst.

### **4.1 Einführung**

Wie in Kapitel 2 beschrieben, ist die Rate-Distortion-Theorie für statistische Signale wohlbekannt [24], [25], [27], [102], [111]. Eine analytische Lösung existiert allerdings nur für Sonderfälle. Die Rate-Distortion-Charakteristik des Encoders in Verbindung mit dem aktuellen Bildsignal kann daher in echtzeitfähigen Verfahren für die Datenratenkontrolle nur prädiktiv geschätzt werden.

Der Schätzung liegt oftmals eine Rate-Distortion-Modellierung zugrunde. Damit kann sichergestellt werden, dass der Encoder die notwendige Stabilität aufweist. Ein wichtiger Aspekt ist die Anpassung der Modelle an die tatsächlichen Gegebenheiten im Encoder, da die zugrunde gelegten statistischen Annahmen für das Bildsignal nur näherungsweise zutreffen.

In Kapitel 3 ist die Eignung bekannter Rate-Distortion-Modelle [3], [58], [106] untersucht worden. Es zeigt sich, dass die Verbesserung der Genauigkeit der Rate-Distortion-Schätzung eine besondere Herausforderung darstellt. Sowohl eine Vereinfachung auf Basis unzutreffender Annahmen, als auch eine unzureichende Parametrisierung des Modells führen zu hohen Schätzfehlern und zu Qualitätsverlusten bei der Videocodierung.

## 4.2 Lösungsansatz für eine adaptive Datenratenkontrolle

Für Videokommunikation ist ein einfaches, echtzeitfähiges Verfahren für die Datenratenkontrolle notwendig. Die Rate-Distortion-Modellierung basiert daher beim neuen Lösungsansatz für die Datenratenkontrolle auf einem analytisch lösbaren Ansatz ohne numerisch iterative Approximation [134], [135]. Durch eine Parametrisierung wird die Anwendbarkeit der Modellierung auf unterschiedliche Bitratenbereiche und die jeweils aktuelle Signalcharakteristik sichergestellt.

### 4.2.1 Grundlegende Rate-Distortion-Modellierung

Der Ausgangspunkt für die Rate-Distortion-Modellierung des Videoencoders basiert auf (2.30), der bekannten Approximation der Entropie  $\hat{H}$  einer Quelle mit Laplace-Verteilung, deren Signal mit dem Quantisierungsparameter  $Q$  quantisiert wird. Dabei wird das Modell (2) aus (2.30) für niedrige bis mittlere Datenraten ( $\sigma^2/Q^2 \leq 1/(2e)$ ) durch Linearisierung vom Ansatz (1) aus (2.30) für hohe Datenraten hergeleitet. Der Vorfaktor  $e/\ln 2$  ergibt sich aus der notwendigen Stetigkeit. Für  $\sigma^2/Q^2 \leq 1/(2e)$  folgt die Schätzung der Bitmenge  $\hat{B}_{DCT_i}^{MB}$  für die Codierung der DCT-Koeffizienten des Makroblocks  $i$  mit  $A$  Pixel damit zu:

$$\hat{B}_{DCT_i}^{MB} = A \cdot c \cdot \frac{\sigma_i^2}{Q_i^2} \quad (4.1)$$

Der Adaptionfaktor  $c$  wird während der Codierung an die aktuelle Signalstatistik angepasst. Die prinzipielle Eignung des Ansatzes ist für niedrige Bitraten bereits in [3], [23], [58] und [106] nachgewiesen worden. In Kapitel 3 ist zusätzlich das Verhalten bei höheren Bitraten und mit komplexerem Bildmaterial analysiert worden.

Für die Schätzung der Verzerrung  $\hat{D}_{Li}^F$  pro Bild mit  $N$  Makroblöcken dient die für höhere Bitraten bzw. kleine Verzerrungen und gleichförmige Quantisierung wohlbekannte Annahme (4.2), wobei durch den Faktor  $\alpha_i^2$  die Verzerrung des Makroblocks  $i$  gezielt gewichtet werden kann [3], [22], [95], [106]:

$$\hat{D}_{Li}^F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\alpha_i^2 \cdot Q_i^2}{a} \quad (4.2)$$

- 
1.  $MB$ : Bezug auf einen Makroblock bzw.  
 $DCT$ : Bezug auf die DCT-Koeffizienten.
  2.  $Li$ : Bezug auf mittlere und höhere Bitraten mit Linearem Verlauf der Verzerrungskennlinie bzw.  
 $F$ : Bezug auf ein Bild (Frame).

Der Adaptionfaktor  $a$  wird während der Codierung an die aktuelle Signalstatistik angepasst.

Für alle Bereiche, in denen diese Bedingungen nicht erfüllt sind, kann die Verzerrung  $\hat{D}_{Qu}^F$  sehr gut in Abhängigkeit von der Varianz  $\sigma_i^2$  des Makroblocks  $i$  geschätzt werden (siehe Kapitel 3):

$$\hat{D}_{Qu}^F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\alpha_i^2 \cdot \sigma_i^2 \cdot Q_i^2}{a} \quad (4.3)$$

Der Adaptionfaktor  $a$  wird während der Codierung an die aktuelle Signalstatistik angepasst.

## 4.2.2 Verallgemeinerung der Rate-Distortion-Modellierung

Aufbauend auf der grundlegenden Rate-Distortion-Modellierung im vorherigen Abschnitt wird im Folgenden ein einheitlicher verallgemeinerter Modellansatz für die Datenratenkontrolle in Videokommunikationssystemen entworfen. Wichtige Aspekte sind dabei ein breiter Einsatzbereich und eine hohe Adaptionfähigkeit, um in allen Bereichen eine möglichst genaue Schätzung sicherzustellen.

Insbesondere bei Kommunikationsanwendungen ist eine Videocodierung mit geringer Verzögerung gefordert. Ist das zugrunde liegende Transportnetz in seiner Datenrate beschränkt, so muss der Videoencoder seine Ausgangsdatenrate an die Vorgaben anpassen. In diesem Fall hat es sich bewährt, die Datenratenkontrolle mehrstufig aufzubauen (siehe Kapitel 3). Die Funktionen der Datenratenkontrolle werden auf die Bitallocation mit GOP- und Bild-Ebene bzw. auf die Makroblock-Ebene aufgeteilt. Dadurch wird pro Bild eine Bitmenge  $B_{DCT}^F$  für die DCT-Koeffizienten vorgegeben, sodass als Lösungsansatz eine Extremwertaufgabe mit einer Nebenbedingung möglich ist.

Die Minimierung der jeweiligen Schätzung der Verzerrung  $\hat{D}_{Li}^F$  und  $\hat{D}_{Qu}^F$  unter Berücksichtigung der vorgegebenen Bitanzahl für die DCT-Codierung pro Bild  $B_{DCT}^F$  führt mit (4.1) auf folgende Ansätze:

$$\hat{D}_{Li}^F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\alpha_i^2 \cdot Q_i^2}{a} = \min!; \quad \text{NB. } B_{DCT}^F - \sum_{i=0}^{N-1} \hat{B}_{DCT_i}^{MB} = 0 \quad (4.4)$$

$$\hat{D}_{Qu}^F = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\alpha_i^2 \cdot \sigma_i^2 \cdot Q_i^2}{a} = \min!; \quad \text{NB. } B_{DCT}^F - \sum_{i=0}^{N-1} \hat{B}_{DCT_i}^{MB} = 0 \quad (4.5)$$

Eine Lösung von (4.4) und (4.5) ist mittels der Lagrange-Multiplikatorregel möglich [8], [92]. Die Einführung des Parameters  $t_i$  führt auf den folgenden Lösungsansatz für die Berechnung des Quantisierungsparameters  $\hat{Q}_i$ , der beide Ansätze vereinheitlicht:

$$\hat{Q}_i = \sqrt{\frac{A \cdot c \cdot \sigma_i \sum_{k=0}^{N-1} t_k \alpha_k \sigma_k}{t_i \cdot \alpha_i \cdot B_{DCT}^F}} \text{ mit } \begin{cases} t_i = 1 & \text{für } \hat{D}_{Li}^F \quad (1) \\ t_i = \sigma_i & \text{für } \hat{D}_{Qu}^F \quad (2) \end{cases} \quad (4.6)$$

Die Lösung in (4.6) basiert auf der Annahme  $\sigma^2/Q^2 \leq 1/(2e)$ . In Abbildung 2.9 ist ersichtlich, dass die Charakteristik der beiden in (2.30) abschnittsweise definierten Funktionen im Übergangsbereich nur gering voneinander abweicht. Zugleich zeigen die Untersuchungen in Kapitel 3 für die Bitratenschätzung in (4.1) Werte von  $c$ , die stark vom ermittelten Vorfaktor  $e/\ln 2$  abweichen. Es ist daher sinnvoll, zu untersuchen, inwieweit der Modellansatz in (4.6) für die Beschreibung der Rate-Distortion-Charakteristik mit  $\sigma^2/Q^2 > 1/(2e)$  geeignet ist.

Die Minimierung des Ansatzes (4.4) für die Schätzung der Bitmenge pro Makroblock bei höheren Bitraten ( $\sigma^2/Q^2 > 1/(2e)$ ):

$$\hat{B}_{DCT_i}^{MB} = A \frac{1}{2} \log_2 \left( 2e^2 \frac{\sigma_i^2}{Q_i^2} \right) \quad (4.7)$$

ist in [106] untersucht worden. Dabei zeigt sich, dass die Quantisierungsparameter für die einzelnen Makroblöcke in einem Bild in etwa konstant sind. Unter der Annahme, dass die Bitratenschätzung (4.1) näherungsweise gültig ist, erhält man mit dem Ansatz  $Q_0 = Q_1 = \dots = Q_{N-1}$  und (4.6) für die Schätzung der Verzerrung  $\hat{D}_{Li}^F$  mit  $\alpha_i^2 = 1; \forall i \in [0; N-1]$  für die Schätzung des Quantisierungsparameters  $\hat{Q}_{H_i}^1$  bei hohen Bitraten:

$$\hat{Q}_{H_i} = \sqrt{\frac{A \cdot c \cdot \sigma_i \sum_{k=0}^{N-1} \sigma_k}{B_{DCT}^F}} \approx const.; \forall i \in [0; N-1] \quad (4.8)$$

damit folgt

$$\sigma_0 = \sigma_1 = \dots = \sigma_{N-1} = \frac{1}{N} \sum_{k=0}^{N-1} \sigma_k \text{ und} \\ \hat{Q}_{H_0} = \hat{Q}_{H_1} = \dots = \hat{Q}_{H_{N-1}} = \sqrt{\frac{A \cdot c \cdot \frac{1}{N} \left( \sum_{k=0}^{N-1} \sigma_k \right)^2}{B_{DCT}^F}} \quad (4.9)$$

---

1.  $H$ : Bezug auf hohe (High) Bitraten.

Vergleicht man die Ansätze (4.6) für  $\hat{D}_{Qu}^F$  mit (4.9), so ergibt sich:

$$\frac{\hat{Q}_{H_i}}{\hat{Q}_i} = \frac{\hat{Q}_H^F}{\hat{Q}^F} = \frac{\sqrt{\frac{A \cdot c \cdot \frac{1}{N} \left( \sum_{k=0}^{N-1} \sigma_k \right)^2}{B_{DCT}^F}}}{\sqrt{\frac{A \cdot c \sum_{k=0}^{N-1} \sigma_k^2}{B_{DCT}^F}}} = \sqrt{\frac{\left( \sum_{k=0}^{N-1} \sigma_k \right) \left( \sum_{k=0}^{N-1} \sigma_k \right)}{N \left( \sum_{k=0}^{N-1} \sigma_k^2 \right)}} \quad (4.10)$$

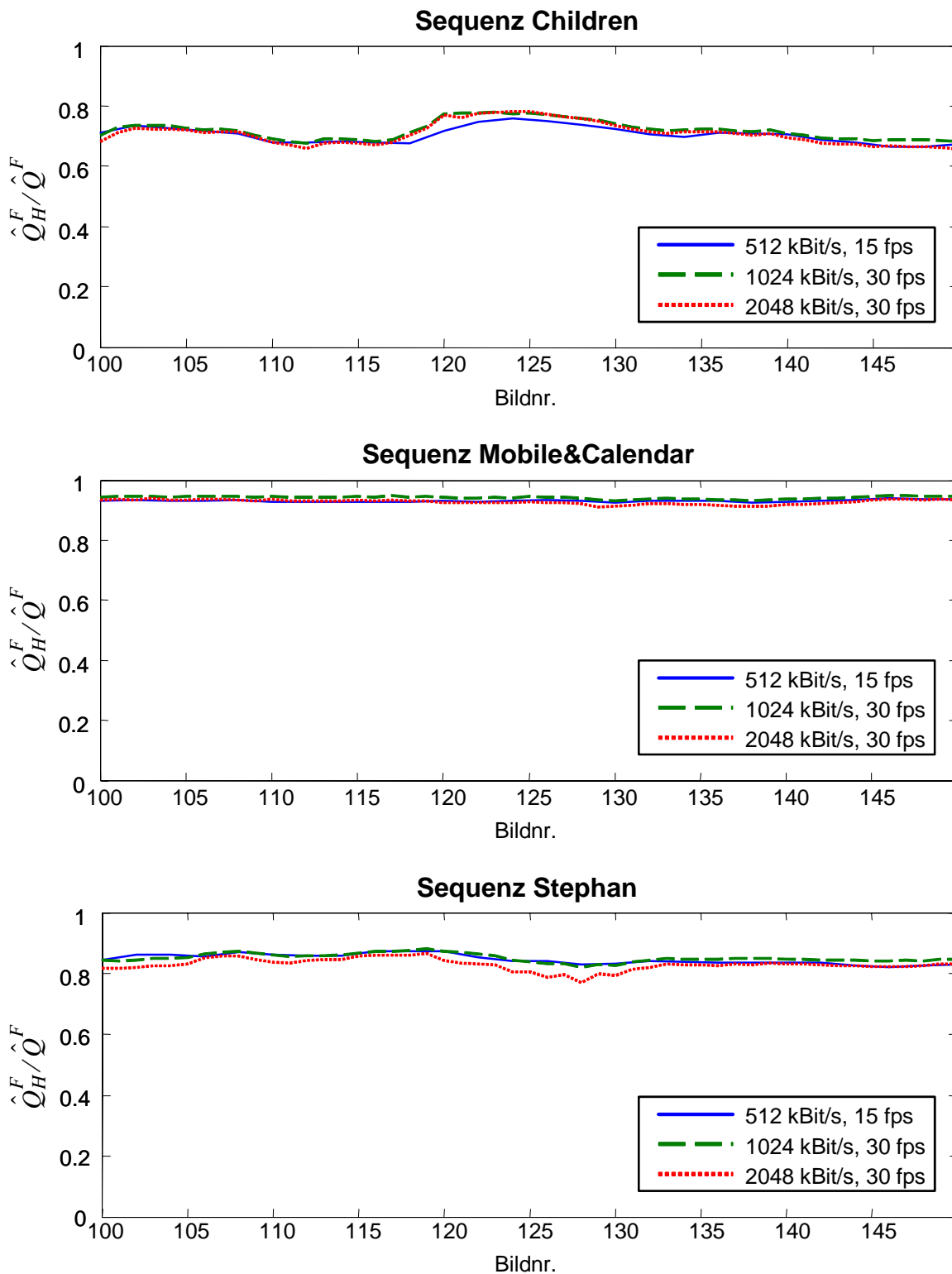
Der Index  $F$  drückt dabei aus, dass in beiden Ansätzen der Quantisierungsparameter konstant über ein Bild ist.

Unter der Annahme, dass in einem Bild nur eine geringe Schwankungsbreite von  $\sigma_i$  um  $\bar{\sigma}$  vorhanden ist (siehe Abschnitt 3.3) und damit oftmals  $\sigma_i \approx \sigma_c = const.; \forall i \in [0; N-1]$  als Abschätzung dienen kann, so kann (4.10) approximiert werden:

$$\frac{\hat{Q}_H^F}{\hat{Q}^F} \approx \sqrt{\frac{(N\sigma_c)^2}{N(N\sigma_c^2)}} = 1 \quad (4.11)$$

Daraus folgt, dass mit den genannten Annahmen die Ansätze (4.6) für  $\hat{D}_{Qu}^F$  und (4.9) näherungsweise gleich sind. In Abbildung 4.1 ist das Ergebnis der Untersuchung von (4.11) dargestellt. Für eine realistische Ermittlung der Werte für (4.10) dient das in einem H.263-Encoder implementierte neue Verfahren für die Datenratenkontrolle aus Abschnitt 4.3 mit folgenden Bedingungen: H.263 Baseline, IPPP...-Codierung, 1 s Ausgangspufferverzögerung, Bitraten jeweils 512 kBit/s bei 15 fps und 1024 bzw. 2048 kBit/s bei 30 fps.

Die Untersuchung bestätigt die Richtigkeit der Annahme über weite Bereiche. Weiter zeigt sich, dass das Verhältnis von  $\hat{Q}_H^F / \hat{Q}^F$  nur sehr gering von Bild zu Bild variiert. Damit können sowohl die Modellunterschiede als auch die zeitlichen Änderungen gut mittels des linearen Adaptionfaktors  $c$  in (4.6) berücksichtigt werden, solange für die Bestimmung von  $c$  nur wenige zurückliegende Bilder herangezogen werden.



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, 1 s Ausgangspufferverzögerung, Bitraten jeweils 512 kBit/s bei 15 fps und 1024 bzw. 2048 kBit/s bei 30 fps

**Abbildung 4.1:** Verhalten von  $\hat{Q}_H^F / \hat{Q}^F$  für die Sequenzen Children, Mobile&Calendar und Stephan.



Somit lassen sich mit einem einzigen parametrisierten Modell neben niedrigen und mittleren Bitraten auch Bereiche über  $\sigma^2/Q^2 > 1/(2e)$  beschreiben:

$$\hat{Q}_i = \sqrt{\frac{A \cdot c \cdot \sigma_i \sum_{k=0}^{N-1} t_k \cdot \alpha_k \cdot \sigma_k}{t_i \cdot \alpha_i \cdot B_{DCT}^F}} \quad \text{mit} \quad \begin{cases} t_i = 1 & \text{mittlere Bitraten} & (1) \\ t_i = \sigma_i & \text{niedrige bzw. hohe Bitraten} & (2) \end{cases} \quad (4.12)$$

### 4.2.3 Parametrisierung der Rate-Distortion-Modellierung

Die Parametrisierung von (4.12) wird von verschiedenen Faktoren bestimmt. Für eine Quelle mit Laplace-Verteilung, deren Signal quantisiert wird, ist die Grenze für hohe Bitraten nach (2.30) mit  $\sigma^2/Q^2 > 1/(2e)$  gegeben. Die Abgrenzung von niedrigen und mittleren Bitraten wird von der Auswahl der jeweiligen Schätzung der Verzerrung  $\hat{D}_{MH}^F$  für (4.2) und  $\hat{D}_L^F$  für (4.3) bestimmt. Untersuchungen zeigen zudem, dass nichtlineare Effekte im Videoencoder eine Rolle spielen, so z.B. die Eingrenzung des Wertebereichs von  $Q_i$  in Abhängigkeit von  $Q_{i-1}$ ,  $Q_{i-2}$ , ... oder der Signalisierungsaufwand für einen Wechsel des Quantisierungsparameters von Makroblock zu Makroblock.

Für die Veranschaulichung des Einflusses des Signalisierungsaufwands auf die Videocodierung werden die Quantisierungsparameter aller Makroblöcke in einem Bild mit der Gesamtbitmenge  $B_{NCH}^F$ <sup>1</sup> zur Codierung der DCT-Koeffizienten auf zwei unterschiedliche Weisen für den Bereich  $\hat{D}_{Li}^F$  berechnet. Für jeden Fall wird die jeweilige Verzerrung ermittelt.

Im ersten Fall ergibt sich die Verzerrung  $\hat{D}_{CH}^F$ <sup>2</sup>. Basierend auf (4.12) ist mit  $t_i = 1; \forall i \in [0; N-1]$  für jeden Makroblock  $i$  in Abhängigkeit von  $\sigma_i$  ein Quantisierungsparameter  $Q_{CH_i}$  berechnet worden. Die Verzerrung  $\hat{D}_{NCH}^F$  ergibt sich für den zweiten Fall, bei dem mit  $t_i = \sigma_i; \forall i \in [0; N-1]$  und (4.12) der Quantisierungsparameter  $Q_{NCH_i}$  für jeden Makroblock  $i$  für  $\forall i \in [0; N-1]$  konstant ist.

Aufgrund des Signalisierungsoverheads  $B_{SO}^F$ <sup>3</sup> pro Bild für die wechselnden  $Q_{CH_i}$  gilt:  $B_{NCH}^F = B_{CH}^F + B_{SO}^F$ , sodass für den ersten Fall für die Codierung der DCT-Koeffizienten nur die Bitmenge  $B_{CH}^F$  zur Verfügung steht.

- 
1. *NCH*: Bezug auf keinen Wechsel des Quantisierungsparameters pro Makroblock in Abhängigkeit seiner Standardabweichung (No CHange).
  2. *CH*: Bezug auf Wechsel des Quantisierungsparameters pro Makroblock möglich, abhängig von seiner Standardabweichung (CHange).
  3. *SO*: Bezug auf SignalisierungsOverhead, der für wechselnde Makroblöcke notwendig wird.

Damit ergibt sich für das Verhältnis von  $\hat{D}_{CH}^F/\hat{D}_{NCH}^F$ :

$$\frac{\hat{D}_{CH}^F}{\hat{D}_{NCH}^F} = \frac{\frac{1}{N} \sum_{i=0}^{N-1} \frac{Q_{CH_i}^2}{a}}{\frac{1}{N} \sum_{i=0}^{N-1} \frac{Q_{NCH_i}^2}{a}} = \frac{\sum_{i=1}^{N-1} \frac{A \cdot c \cdot \sigma_i \sum_{k=0}^{N-1} \sigma_k}{B_W^F}}{\sum_{i=1}^{N-1} \frac{A \cdot c \sum_{k=0}^{N-1} (\sigma_k)^2}{B_{NW}^F}} \quad (4.13)$$

Für  $\hat{D}_{CH}^F/\hat{D}_{NCH}^F < 1$  weist dabei der erste Fall mit der Berechnung des Quantisierungsparameters  $Q_{CH_i}$  für den Makroblock  $i$  in Abhängigkeit von  $\sigma_i$  eine höhere Codiereffizienz auf. Durch einfache Umformung von (4.13) erhält man:

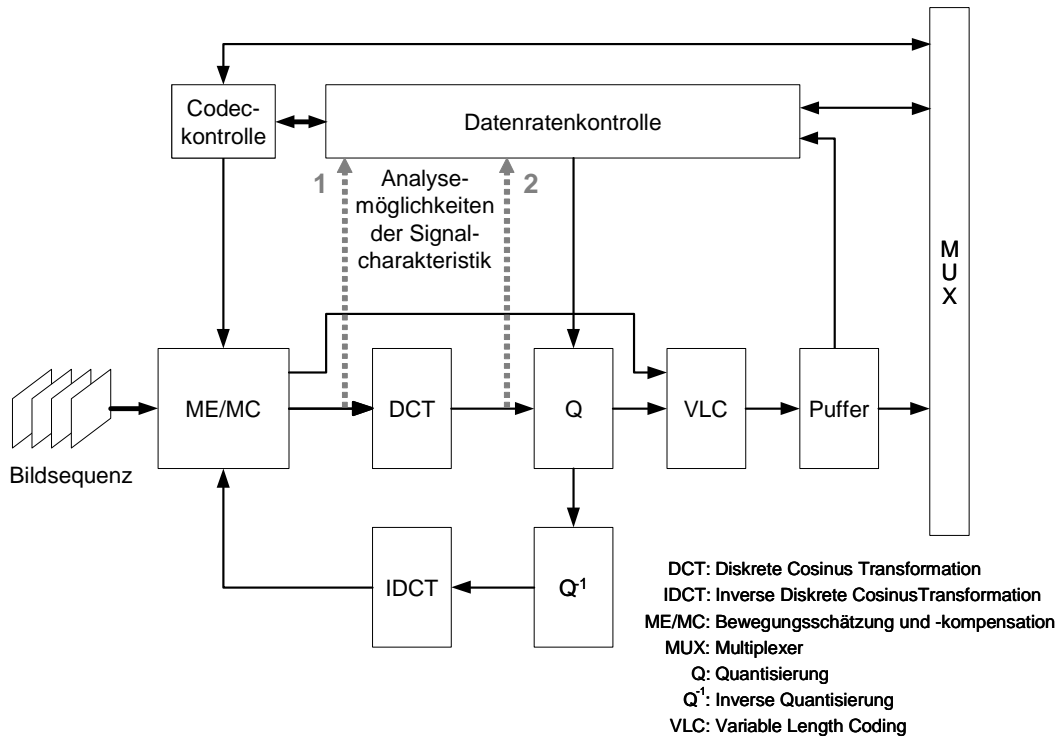
$$\left(1 + \frac{B_{SO}^F}{B_{CH}^F}\right) \frac{\left(\sum_{i=0}^{N-1} \sigma_i\right)^2}{N \sum_{i=0}^{N-1} \sigma_i^2} < 1 \quad (4.14)$$

Die Analyse mit z.B.  $\hat{B}_{SO}^F \approx 1000\text{Bit}$  (Bildformat: CIF, 5 Bit Overhead pro Änderung des Quantisierungsparameters von Makroblock zu Makroblock und Änderung bei 50 % der Makroblöcke) zeigt, dass es nur in bestimmten Bereichen sinnvoll ist, den Quantisierungsparameter  $Q_i$  für den Makroblock  $i$  in Abhängigkeit von  $\sigma_i$  zu berechnen (Fall (1) in (4.12)). Aufgrund von Modell-schätzfehlern und der Eingrenzung des Wertebereichs von  $Q_i$  in Abhängigkeit der Werte von  $Q_{i-1}, Q_{i-2}, \dots$  wird dieser Bereich zusätzlich beschränkt.

Auf die Festlegung der jeweiligen Grenzen und Entscheidungsschwellen für das neue Verfahren wird im Abschnitt 4.3 näher eingegangen.

#### 4.2.4 Untersuchung statistischer Parameter im Spektralbereich

Die Qualität der Schätzung des Rate-Distortion-Verhaltens wird neben der gewählten Modellierung auch durch den eingesetzten Schätzparameter beeinflusst. In Abbildung 4.2 sind die Funktionsmodule eines bewegungskompensierten DCT-Videoencoders vereinfacht dargestellt. Wie im Abschnitt 2.2 beschrieben, werden die Makroblöcke nach der optionalen Bewegungskompensation in  $8 \times 8$  große Blöcke unterteilt und mittels der 2D-DCT in einen Satz von  $8 \times 8$  zweidimensionalen Frequenzkomponenten transformiert, quantisiert und anschließend laulängencodiert. Die Parameter für die Schätzung der Rate-Distortion-Charakteristik des aktuellen Signals können auf Makroblockebene im Encoder, wie in Abbildung 4.2 dargestellt, sowohl aus dem örtlichen Bildsignal als auch nach der DCT-Transformation gewonnen werden. Die bisherigen Betrachtungen basierten auf der Varianz  $\sigma_i^2$  als Schätzparameter, die in vielen Verfahren eingesetzt wird [3], [22], [95], [106].



**Abbildung 4.2:** Die Parameter für die Rate-Distortion-Schätzung des Videosignals können im Hybrid-DCT-Videoencoder vor und nach der DCT-Transformation gewonnen werden.

Es bietet sich an, die DCT-Koeffizienten am Ausgang der DCT-Transformation zu analysieren, da diese nach der Quantisierung im Bitstrom codiert werden. Wichtige statistische Parameter, wie die Leistung  $P_x$  und Varianz  $\sigma_x^2$  eines digitalen Bildsignals  $x(m, n)$  können direkt aus den Koeffizienten  $c_{uv}$  der DCT-Transformation berechnet werden (allgemein:  $U = M, V = N$  und in diesem Fall:  $U = M = V = N = 8$ ) [96]:

$$P_x = \frac{1}{UV} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} c_{uv}^2 \quad (4.15)$$

$$\sigma_x^2 = \frac{1}{UV} \left( \left( \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} c_{uv}^2 \right) - c_{00}^2 \right) \quad (4.16)$$

Die Untersuchung des Rate-Distortion-Verhaltens im Spektralbereich von korrelierten und unkorrelierten Gaußprozessen liefert Ansatzpunkte für die Interpretation des erzielbaren Codiergewinns gegenüber der PCM (Pulse Code Modulation). Die Rate-Distortion-Funktion für korrelierte Gaußprozesse (siehe Abschnitt 2.3.1) legt die benötigte Bitanzahl  $R_{uv}$  für die Codierung des Spektralkoeffizienten  $c_{uv}$  für eine vorgegebene quadratische Verzerrung  $D$  wie folgt fest:

$$R_{uv} = \max \left( 0, \frac{1}{2} \log_2 \frac{E\{c_{uv}^2\}}{D} \right) \quad (4.17)$$

1. Der Operator  $E\{\dots\}$  berechnet den Erwartungswert des Arguments.

Daraus ist ersichtlich, dass für Koeffizienten geringer Energie die Bitrate null wird. Der Codiergewinn verringert sich, sobald einzelne Koeffizienten zu null gesetzt werden, wie es bei niedrigen und mittleren Bitraten gebräuchlich ist. Der Fehler wird dabei identisch mit dem Signal und der Codierungsfehler enthält - normalerweise hochfrequente - Komponenten, die mit dem Bildsignal korreliert sind. Aus (4.17) ist ersichtlich, dass die notwendige Bitrate logarithmisch von den Energien der Koeffizienten abhängt, dagegen ist die Energiesumme gleich der Energie des Signals. Daher ist der Gewinn umso größer, je ungleichmäßiger die Verteilung der Energie auf die einzelnen Koeffizienten ist [96].

Wird die Varianz als Parameter für die Rate-Distortion-Schätzung verwendet, folgt aus (4.15), dass zum einen die Energieverteilung auf die einzelnen Koeffizienten und zum anderen die Codierung des DCT-Koeffizienten  $c_{00}$  nicht berücksichtigt wird. Dies kann speziell in inhomogenen Bereichen mit wechselnden Eigenschaften zu großen Fehlern bei der Schätzung führen.

Eine Lösungsmöglichkeit bietet der Polynomische-Satz [8], bei dem die Energieverteilung im Gegensatz zur Definition der Varianz in (4.16) auf die einzelnen Koeffizienten berücksichtigt wird:

$$(a_1 + a_2 + \dots + a_k)^n = \sum_{t_1 + t_2 + \dots + t_k = n} \binom{n}{t_1, t_2, \dots, t_k} a_1^{t_1} a_2^{t_2} \dots a_k^{t_k} \quad (4.18)$$

Dabei ist die Summe über alle  $k$ -Tupel natürlicher Zahlen  $(t_1, t_2, \dots, t_k)$  zu erstrecken, für die

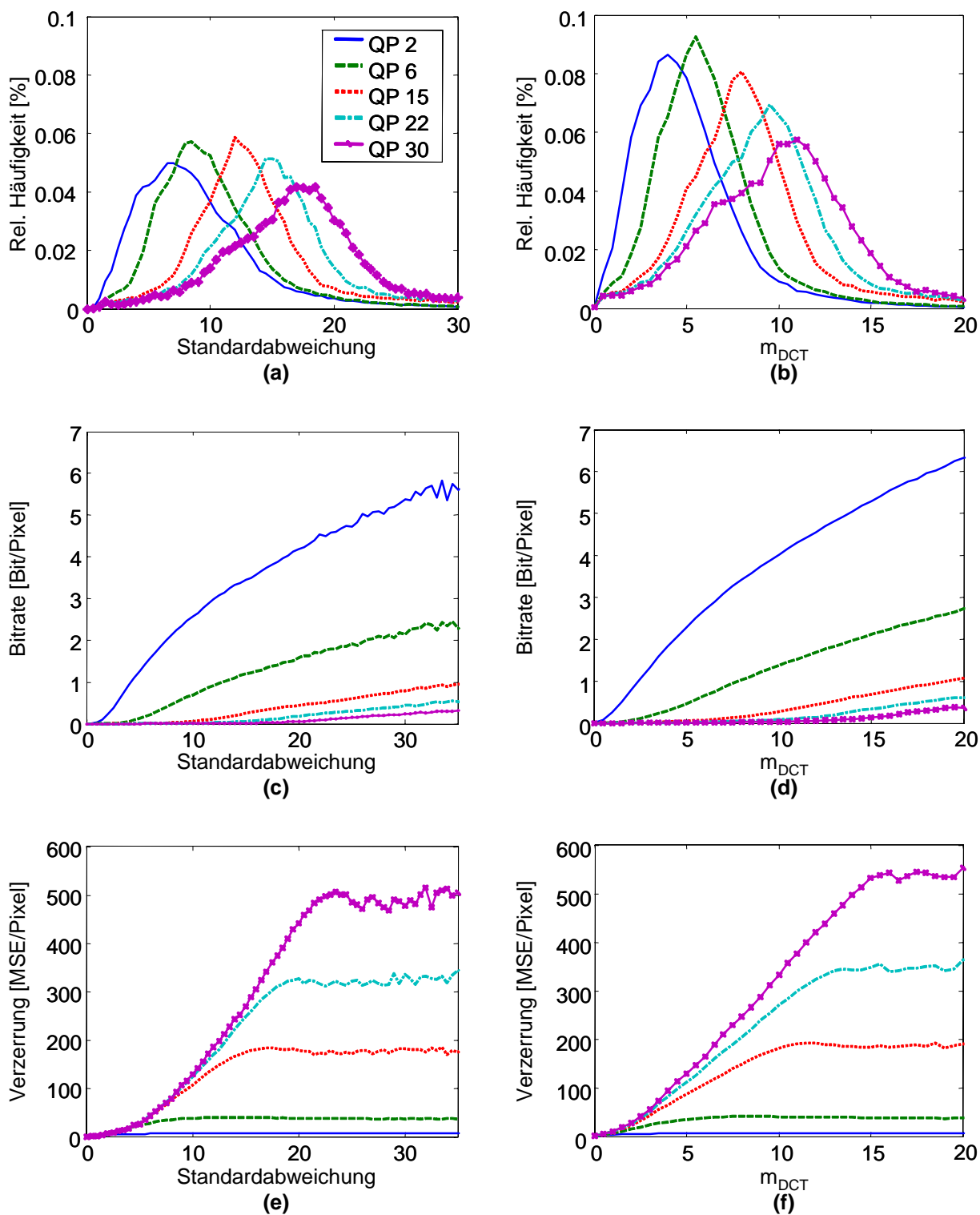
$$\sum_{i=1}^k t_i = n \text{ ist, wobei hier } n = 2 \text{ gilt.}$$

Daher wird für die folgenden Untersuchungen anstelle der Varianz  $\sigma_x^2$  ein neuer Schätzparameter  $m_{DCT_x}^2$  definiert, der diese Aspekte berücksichtigt:

$$m_{DCT_x}^2 = \left( \frac{1}{UV} \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} |c_{uv}| \right)^2 \quad (4.19)$$

Multipliziert man den rechten Term aus, ist ersichtlich, dass in den neuen Schätzparameter sowohl die Produkte der verschiedenen Koeffizienten und damit deren Energieverteilung als auch die Energie der Einzelkoeffizienten  $c_{uv}^2$  analog zur Definition der Varianz in (4.16) eingehen. Zudem wird der Koeffizient  $C_{00}$  berücksichtigt. Damit kann der neue Schätzparameter  $m_{DCT}^2$  insbesondere in Bereichen mit einer stark inhomogenen Energieverteilung auf die einzelnen Koeffizienten die Rate-Distortion-Modellierung verbessern.

Bedingt durch die Analogien der Definition der Varianz in (4.16) und des in (4.19) definierten Schätzparameters  $m_{DCT_x}^2$  zeigen die Verläufe der Rate-Distortion-Kennlinien in Abhängigkeit des jeweiligen Schätzparameters eine große Ähnlichkeit (siehe Abbildung 4.3). Dies bestätigen weitergehende Analysen, wie z.B. die Untersuchungen in [96], die einen  $1/x$ -ähnlichen Verlauf der Verteilungsdichte der DCT-Wechselanteil-Koeffizienten ergeben. Somit können die untersuchten Modelle übernommen werden, allerdings müssen aufgrund des veränderten Wertebereichs die Gültigkeitsgrenzen der Rate-Distortion-Modellierung im weiteren Entwurf neu bestimmt werden.



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, QP const.  
Die Legende für alle Diagramme befindet sich links oben.

**Abbildung 4.3:** Vergleich von Rate-Distortion-Kennlinien der Sequenz Mobile&Calendar in Abhängigkeit der Schätzparameter Standardabweichung und  $m_{DCT}$ .

Aufgrund der inhomogeneren Energieverteilung auf die einzelnen Koeffizienten bei niedrigen und mittleren Bitraten verspricht der neue Parameter  $m_{DCT}^2$  insbesondere hier eine verbesserte Schätzung. Eine Überprüfung der Eignung des neuen Schätzparameters  $m_{DCT}^2$  anhand des in Kapitel 5 aufgestellten Testsets und der darin definierten Bedingungen ergibt bis ca. 0,3 Bit pro Pixel eine Verbesserung des *PSNR*-Werts von durchschnittlich 0,12 dB im Vergleich zur Varianz.

Dabei wird der mittlere Fehler zwischen der Schätzung durch die Rate-Distortion-Modellierung und den aus der Codierung ermittelten Werten nur unwesentlich verringert. Allerdings werden die Maximalwerte des Schätzfehlers deutlich reduziert. Ein Qualitätsgewinn kann dabei vor allem bei variierenden Quantisierungsparametern und der kombinierten Rate-Distortion-Modellierung, wie im Abschnitt 4.2.2 beschrieben, erreicht werden.

Bei höheren Bitraten ergibt sich im Testset aus Kapitel 5 unter den gleichen Randbedingungen eine geringfügige Verschlechterung des mittleren *PSNR*-Werts von 0,01 dB. In diesem Bereich ist die Energieverteilung auf die einzelnen Koeffizienten homogener. Daher ist es besser, die Signalcharakteristik anhand der Varianz in Verbindung mit einem skalaren Adaptionsfaktor  $c$  in (4.12) zu beschreiben. Im Maß  $m_{DCT}^2$  wird die Energieverteilung auf die einzelnen Koeffizienten in diesem Bereich zu stark gewichtet.

### 4.3 Datenratenkontrolle für einen Hybrid-DCT Videoencoder

Auf Basis der Rate-Distortion-Modellierung im Abschnitt 4.2, wird im Folgenden das neue grundlegende Verfahren beschrieben. Dabei wird speziell auf den für Kommunikationsanwendungen sehr wichtigen Fall einer IPPP...-Codierung eingegangen.

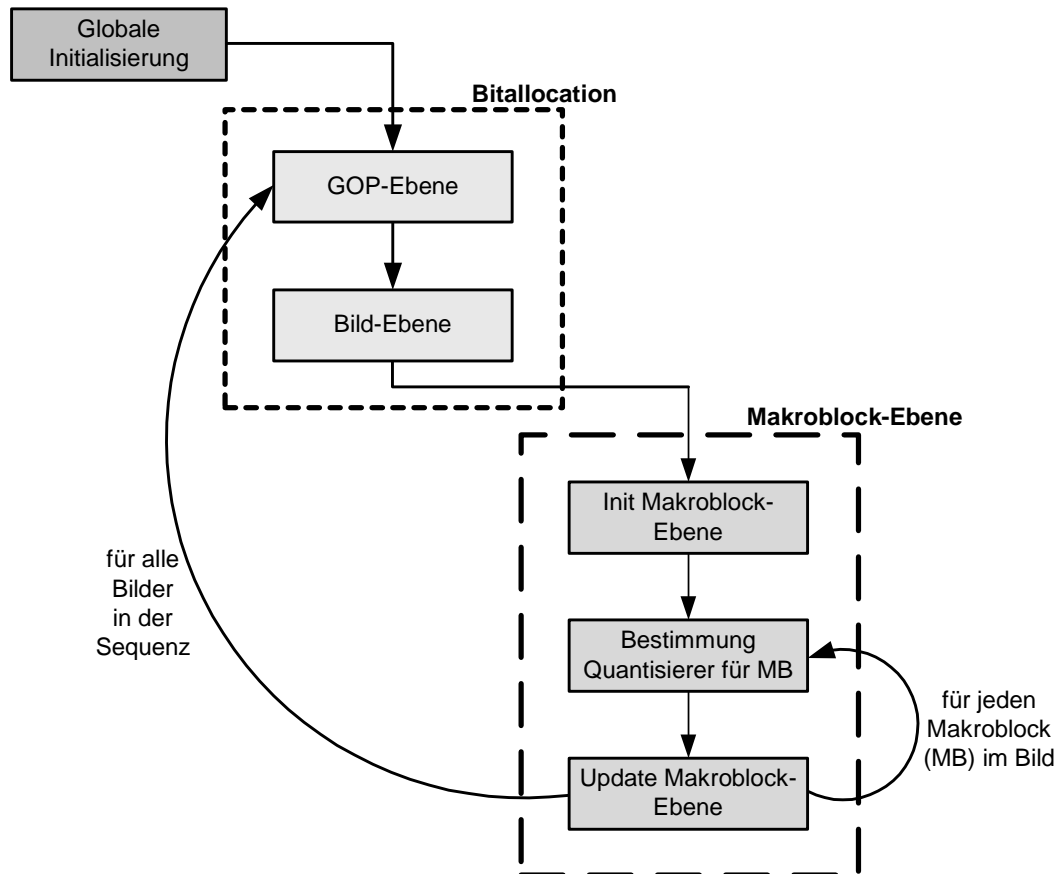
Der zugrunde liegende Aufbau des neuen Verfahrens mit der entsprechenden Aufgabenverteilung in GOP-, Bild- und Makroblock-Ebene ist für echtzeitfähige Kommunikationsanwendungen mit beschränkter Rechenleistung wohl erprobt, so z.B. [3], [22], [59], [106] oder [121].

Für die Sicherstellung einer möglichst hohen Videocodiereffizienz durch eine Bestimmung der Quantisierungsparameter basierend auf (4.12) sind folgende Aspekte berücksichtigt worden:

- Der Einsatz verbesserter Schätzparameter nach (4.19).
- Die Anpassung an die aktuelle Signalstatistik durch eine adaptive Modellparametrisierung, um den Bitratenbereich von 7,5 kBit/s bis zu 2 MBit/s abdecken zu können.
- Die Sicherstellung der Stabilität auch unter wechselnden oder kritischen Situationen, wie Szenenänderungen oder variablen Ausgangsdatenraten.
- Modularer Aufbau des Verfahrens, um gezielt neue Funktionalitäten, wie die Unterstützung variabler Ausgangsdatenraten, unterschiedlicher Bildtypen (I-, P- und B-Bilder) und mehrerer Objekte pro Bild gewährleisten zu können.

In Abbildung 4.4 ist der Systemaufbau der Datenratenkontrolle dargestellt. Auf Erweiterung des neuen Verfahrens bezüglich neuer Funktionalitäten wird im Abschnitt 4.4 eingegangen, so z.B. die Unterstützung variabler Ausgangsdatenraten, unterschiedlicher Bildtypen (I-, P- und B-Bilder) und mehrerer Objekte pro Bild.

Die tatsächliche Leistungsfähigkeit des neuen echtzeitfähigen Verfahrens für die Datenratenkontrolle wird in Kapitel 5 anhand von unterschiedlichsten Testsequenzen mit verschiedenen Videocodierparametern (z.B. Datenrate oder Framerate) im Vergleich zu wichtigen Verfahren aus der Standardisierung beurteilt.



**Abbildung 4.4:** Funktionsblöcke des Basisverfahrens für die Datenratenkontrolle.

### 4.3.1 Eingangparameter und globale Initialisierung

Für die Initialisierung sind folgende Parameter notwendig, die sich direkt aus dem Videocodec bzw. aus dem zur Verfügung stehenden Kommunikationsnetz ergeben:

$A$ : Anzahl der Pixel pro Makroblock (hier:  $A = 16 \cdot 16$ ).

$br$ : Größe des Ausgangspuffers in Abhängigkeit von der Bitrate  $R$

( $BS_{EFF}^E = br \cdot R$  ergibt die Größe des Puffers [Bit]) [s].

$b\_tmax$ : Vorgabe für die maximale Befüllungshöhe des Ausgangspuffers [%].

$FR$ : Framerate ( $REF\_FR/FR$  muss eine ganze Zahl ergeben) [fps].

$N$ : Anzahl der Makroblöcke pro Bild.

$R$ : Bitrate [Bit/s].

$REF\_FR$ : Referenz-Framerate [fps].

Auf Basis der Eingangswerte müssen für die Steuerung bzw. die Anpassung der Datenratenkontrolle an das zu erwartende Bildmaterial einige Steuerparameter bestimmt werden. Ab einer mittleren Bitrate pro Pixel von 0,3 Bit/Pixel wird von einer hochbitratigen Videocodierung ausgegangen und die Steuervariable  $hb$  gleich eins gesetzt, sonst  $hb = 0$ . Die Annahme basiert auf einer Abschätzung der mittleren Bitrate pro Pixel mithilfe der in (2.30) definierten Grenzen und den in Kapitel 3 ermittelten Werten von  $c$  für (4.1). Untersuchungen mit weiteren Testsequenzen haben diesen Wert bestätigt.

Prinzipiell gibt es zwei Möglichkeiten, den Quantisierungsparameter  $\hat{Q}_i$  aus (4.12) und (4.19) mit

$$\hat{Q}_i = \sqrt{\frac{A \cdot c \cdot m_{DCT_i} \sum_{j=0}^{N-1} t_j \alpha_j m_{DCT_j}}{t_i \cdot \alpha_i \cdot \hat{B}_{DCT}^F}} =$$

$$= \sqrt{\frac{A \cdot c \cdot m_{DCT_i} k_i}{t_i \cdot \alpha_i \cdot \hat{B}_{DCT_i}^F}} \text{ mit } \begin{cases} t_i = 1 & \text{für mittlere Bitraten} & (1) \\ t_i = m_{DCT_i} & \text{für niedrige und hohe Bitraten} & (2) \end{cases} \quad (4.20)$$

zu bestimmen.

Der erste Ansatz berechnet den Quantisierungsparameter  $\hat{Q}_i$  für den Makroblock  $i$  bezogen auf

alle Makroblöcke in einem Bild. Damit sind  $k_i = \sum_{j=0}^{N-1} t_j \cdot \alpha_j \cdot m_{DCT_j}$  und  $\hat{B}_{DCT_i}^F = \hat{B}_{DCT}^F$  aus

(4.20) konstant für  $\forall i \in [0; N-1]$ . Es findet also stets eine Minimierung der Verzerrung auf Bild-Ebene statt.

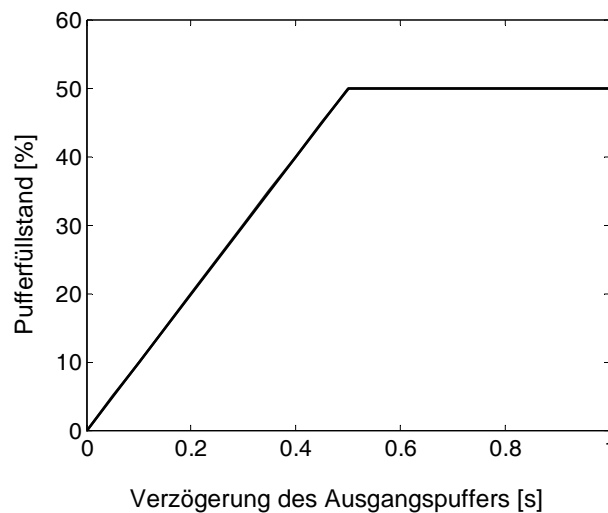


Im anderen Fall wird die Bestimmung des Quantisierungsparameters  $\hat{Q}_i$  für die restlichen Makroblöcke  $i, \dots, N-1$  des Bildes durchgeführt, indem für jeden Makroblock  $i$  die Werte  $k_i = \sum_{j=i}^{N-1} t_j \cdot \alpha_j \cdot m_{DCT_j}$  und  $\hat{B}_{DCT_i}^F = \hat{B}_{DCT}^F - d_{sf_i}^F$  verwendet werden.  $d_{sf_i}^F$  entspricht dabei der Bitmenge, die für die DCT-Codierung der Makroblöcke  $0, \dots, i-1$  benötigt worden ist.

Untersuchungen der Codiereffizienz bzw. der Arbeitsweise der Datenratenkontrolle anhand von Testsequenzen ergeben, dass die bildbasierte Optimierung leistungsfähiger ist. Allerdings zeigt sich, dass es teilweise zu starken Abweichungen zwischen der vorgegebenen Bitmenge  $\hat{B}^F$  und der tatsächlichen für die Codierung benötigten Bitmenge kommt.

Aus diesem Grund wird für niedrige und mittlere Bitraten ( $hb = 0$ ), solange der Ausgangspuffer  $br < 0,25$  s ist, die Bestimmung des Quantisierungsparameters  $\hat{Q}_i$  jeweils basierend auf den restlichen Makroblöcke durchgeführt (Steuervariable  $lb = 0$ , sonst  $lb = 1$ ). Durch die ständige Aktualisierung der Werte kann die Bitmenge  $\hat{B}_{DCT}^F$  sehr genau eingehalten werden, sodass ein Überlauf des Puffers vermieden wird.

Eine weitere Maßnahme für die Verhinderung von Pufferüberläufen, ist die empirische Bestimmung der Zielfüllhöhe des Ausgangspuffers  $b_t$  in Abhängigkeit von dessen Größe  $br$ . Hierzu dient die Zielfunktion in Abbildung 4.5. Der zur Verfügung stehende Ausgangspuffer wird dabei möglichst effizient genutzt.



**Abbildung 4.5:** Zielbefüllungshöhe des Ausgangspuffers in Abhängigkeit der vorgegebenen Puffergröße.

- 
1.  $sf$ : Bezug auf die Bitmenge für die teilweise abgeschlossene Codierung des aktuellen Bildes bis zum aktuellen Makroblock (so far).

Um zu vermeiden, dass sich Schwankungen bei der Pufferbefüllung zu stark auf das nächste Bild auswirken und es somit zu starken Qualitätsschwankungen von Bild zu Bild kommt, wird die notwendige Korrektur auf mehrere Bilder verteilt. In Abhängigkeit von der Zielframerate  $FR$  und der Referenz-Framerate  $REF\_FR$  wird die Anzahl der zukünftigen Bilder  $nff$ , die in der Datenratenkontrolle auf der Bild-Ebene benötigt werden (siehe Abschnitt 4.3.3), bestimmt zu  $nff = \max(0, 5 \cdot FR/REF\_FR, 1)$ .

### 4.3.2 Datenratenkontrolle auf der GOP-Ebene

Bevor das aktuelle Bild  $n$  codiert wird, wird ein Update des Füllstands des Ausgangspuffers  $bs_n^E$  durchgeführt:

$$bs_n^E = \max(bs_{n-1}^E + d_{n-1} - R'/FR, 0) \quad (4.21)$$

Dabei ist  $bs_{n-1}^E$  der Füllstand nach dem letzten Update,  $d_{n-1}$  die tatsächliche Bitmenge für die Codierung des letzten Bildes und  $R' = R$ , wenn die Bitrate konstant ist bzw. sonst gleich dem Mittelwert der Bitrate, seit dem letzten Update. Für die Initialisierung vor der Codierung des ersten P-Bildes, wird der Puffer auf  $b\_t = 50\%$  gesetzt [75].

Im Anschluss wird das nächste zu codierende Bild bestimmt. Hierzu wird überprüft, wie viele Bilder  $nsf$  bezogen auf die aktuelle Framerate  $FR$  übersprungen werden müssen, um folgende Ungleichung zu erfüllen:

$$\max(b\_t \cdot BS_{EFF}^E - (nsf + 1) \cdot R'/FR, 0) < b\_tmax \cdot BS_{EFF}^E \quad (4.22)$$

In der aktuellen Implementierung wird  $R' = R$  angenommen. Bei Kenntnis der zukünftigen Ausgangsdatenrate kann diese berücksichtigt werden.

Diese einfache Methode für die Bestimmung der Anzahl der zu überspringenden Bilder wird ebenfalls bei den Verfahren für die Datenratenkontrolle bei H.263 und MPEG-4 [59], [75], [78], [106], [121] eingesetzt. Eine Vergleichbarkeit der Ergebnisse ist damit gegeben. Wie in [75], [78] wird davon ausgegangen, dass die Encodierung unterbrochen wird, wenn die Puffergröße  $BS_{EFF}^E$  nahezu erreicht ist. Damit kann niemals ein Pufferüberlauf auftreten und die Pufferverzögerung ergibt sich bei konstanter Datenrate direkt aus  $br$  (siehe Kapitel 2).

Ist eine GOP-Struktur vorgegeben, dann wird für das nun aktuelle Bild anhand der Vorgabe ein Bildtyp vorausgewählt. Dieser Vorschlag wird dann innerhalb der folgenden Codierung berücksichtigt bzw. kann in speziellen Fällen, wie z.B. einem Szenenwechsel, modifiziert werden.

### 4.3.3 Datenratenkontrolle auf der Bild-Ebene

Die Datenratenkontrolle auf der Bild-Ebene bestimmt die Zielbitmenge in Abhängigkeit des aktuellen Pufferstandes und der im Standard vorgegebenen Maximalgrößen, die für die Codierung des aktuellen Bildes zur Verfügung stehen. In diesem Abschnitt wird von einer IPPP...-Codierung ausgegangen, die im Kapitel 4.4.3 auf weitere Bildtypen erweitert wird. Für die Bestimmung der geschätzten Zielbitmenge  $\hat{B}_{DCT}^F$ , die für die Codierung der DCT-Koeffizienten aller Makroblöcke im Bild zur Verfügung steht, hat sich folgender Ansatz in Anlehnung an die Verfahren in der Standardisierung von H.263 und MPEG-4 (z.B. [59], [75] oder [121]) bewährt:

$$\hat{B}_{DCT}^F = \frac{R'}{FR} - \frac{(bs_n^E - b_{-t} \cdot BS_{EFF}^E)}{nff} - d_{over}^F \quad (4.23)$$

Der Term  $R'/FR$  berechnet dabei für  $R' = R = const.$  die mittlere Bitmenge in Abhängigkeit von der Framerate  $FR$  und der Bitrate  $R$ , die pro Bild zur Verfügung steht. Die Abweichung von der Sollfüllhöhe des Ausgangspuffers wird durch  $(bs_n^E - b_{-t} \cdot BS_{EFF}^E)/nff$  berücksichtigt. Um starke Qualitätsschwankungen von Bild zu Bild zu verhindern, werden dabei  $nff$  Bilder berücksichtigt. Der Ausgangsdatenstrom des Encoders beinhaltet neben den codierten DCT-Koeffizienten unter anderem noch Steuerinformationen und die Bewegungsvektoren. Dieser Anteil wird nach der Begrenzung auf die Maximalvorgaben des jeweiligen Videostandards durch  $d_{over}^F = d_{LF}^F - d_{DCT_{LF}}^F$ <sup>1</sup> berücksichtigt.  $d_{LF}^F$  entspricht der Anzahl aller Bits, die für die Codierung der letzten Bilder notwendig waren bzw. die Bitmenge  $d_{DCT_{LF}}^F$  entsprechend für die Codierung der DCT-Koeffizienten. Nach der Codierung des ersten I-Bildes wird  $d_{over}^F = 0$  initialisiert.

In Kapitel 5 sind zusätzlich Untersuchungen mit aufwendigeren Verfahren basierend auf dem Rate-Distortion-Modell in (4.12) durchgeführt worden. Es zeigt sich, dass gegenüber dem hier beschriebenen Ansatz bei der IPPP...-Codierung mit Übertragung über einen fehlerfreien Kanal keine Verbesserungen der Leistungsfähigkeit erzielt werden konnte.

### 4.3.4 Datenratenkontrolle auf der Makroblock-Ebene

In dieser Ebene wird für alle Makroblöcke im aktuellen Bild der jeweilige Quantisierungsparameter auf Basis von (4.12) bestimmt. Ziel ist die Minimierung der Verzerrung eines ganzen Bildes bezüglich des  $MSE$ -Wertes ( $\alpha_i = 1; \forall i \in [0; N-1]$ ). Die Summe der Bitmengen für die DCT-Codierung der Makroblöcke soll der vorgegebenen Bitmenge  $\hat{B}_{DCT}^F$  entsprechen. In der folgenden Beschreibung wird davon ausgegangen, dass das erste Bild mit einem fest vorgegebenen Quantisierungsparameter intracodiert worden ist. Für alle Makroblöcke des aktuellen P-Bildes sind zudem die Bewegungskompensation und die 2D-DCT-Transformation bereits durchgeführt worden und der jeweilige Makroblock-Typ steht fest.

---

1.  $LF$ : Bezug auf das vorherige Bild (Last Frame) bzw.  
 $over$ : Bezug auf Overhead.

#### 4.3.4.1 Initialisierung

Vor der Berechnung der Quantisierungsparameter für jeden Makroblock  $i$  werden bildbasierte Parameter bestimmt, die passende Modellierung der Verzerrung ausgewählt und eine Initialisierung durchgeführt.

#### Berechnung bildbasierter Parameter

Zuerst wird für jeden Y-Block  $j$  des Makroblocks  $i$  der Mittelwert der Absolutbeträge der DCT-Koeffizienten aus dem transformierten Prädiktionsfehler bestimmt und im Anschluss der Mittelwert für den Makroblock  $i$ :

$$m_{DCT_{ij}} = \frac{1}{8 \cdot 8} \sum_{u=0}^7 \sum_{v=0}^7 |c_{uv}|; \quad \forall j \in [0, 3] \quad (4.24)$$

$$m_{DCT_i} = \frac{1}{4} \sum_{j=0}^3 m_{DCT_{ij}} \quad (4.25)$$

Daraus wird der Mittelwert für das ganze Bild berechnet:

$$m_{DCT}^F = \frac{1}{N} \sum_{i=0}^{N-1} m_{DCT_i} \quad (4.26)$$

Für die Berechnung von  $k_0 = \sum_{j=0}^{N-1} t_j \cdot \alpha_j \cdot m_{DCT_j}$  aus (4.20) ist die Festlegung des Parameters  $t_k$  in Abhängigkeit von der aktuellen Bitrate und der geschätzten Verzerrung notwendig. Hierzu wird  $t_j = m_{DCT_j}$ ;  $\forall j \in [0, N-1]$  gesetzt, wenn folgende drei Bedingungen erfüllt sind:

- $hb = 1$ . Es liegt eine hochbitratige Codierung vor (siehe Abschnitt 4.3.1).
- $m_{DCT}^F < \frac{1}{N} \sum_{i=0}^{N-1} QP_{LF_i}$ . Die Abschätzung für den Einsatz der Modellierung der Verzerrung  $\hat{D}_L$  basiert auf den Untersuchungen zu Abbildung 4.3 und konnte für weitere Fälle bestätigt werden.  $QP_{LF_i}$  ist der Quantisierungsparameter des Makroblocks  $i$  aus dem vorherigen Bild. Bei der Verifikation des neuen Verfahrens in Kapitel 5 hat sich gezeigt, dass die Bedingung  $(m_{DCT}^F)^2 < \frac{1}{N} \sum_{i=0}^{N-1} QP_{LF_i}$  geringfügig bessere Ergebnisse liefert.
- Die Bitrate pro Pixel im letzten Bild kleiner als 0,5 Bit/Pixel ist. Dieser empirische Ansatz stellt im Vergleich zu (4.14) eine deutliche Vereinfachung dar und hat sich in der Praxis bewährt.

In allen anderen Fällen wird  $t_j = 1$ ;  $\forall j \in [0, N-1]$  gesetzt. Für die Initialisierung der Werte des ersten P-Bildes wird auf die Ergebnisse des vorherigen I-Bildes zurückgegriffen.

### Initialisierung von Zählern und Modellparametern

Für die Berechnung des ersten Quantisierungsparameters wird  $\hat{B}_{DCT_0}^F = \hat{B}_{DCT}^F$  gesetzt. Der Makroblockzähler  $i$  wird mit  $i = 0$  initialisiert. Der Modellparameter  $c$  wird mithilfe der Methode des kleinsten quadratischen Fehlers (siehe z.B. [92], [110]) aus den ermittelten Werten der letzten  $N$  Makroblöcke basierend auf

$$\hat{B}_{DCT_i}^{MB} = 16^2 \cdot c \frac{m_{DCT_i}^2}{Q_i^2} \quad (4.27)$$

ermittelt. Auch hier wird für die Initialisierung vor dem ersten P-Bild auf die Ergebnisse des vorherigen I-Bildes zurückgegriffen.

#### 4.3.4.2 Berechnung des Quantisierungsparameters für den $i$ -ten Makroblock

Die Berechnung des Quantisierungsparameters  $\hat{QP}_i$  für den aktuellen Makroblock  $i$  nach (4.20) erfolgt durch:

$$\hat{QP}_i = \frac{1}{2} \hat{Q}_i = \frac{1}{2} \sqrt{\frac{16^2 \cdot c \cdot k_i \cdot m_{DCT_i}}{t_i \cdot \hat{B}_{DCT_i}^F}} \quad (4.28)$$

Im H.263-Codec ist der Wertebereich für den aktuellen Quantisierungsparameter  $QP_i$  begrenzt. Daher wird für alle  $QP_i < 1$  der Parameter auf  $QP_i = 1$  und für alle  $QP_i > 31$  auf  $QP_i = 31$  gesetzt. In allen anderen Fällen wird  $QP_i$  ganzzahlig gerundet. Zusätzlich muss aufgrund der Standardvorgaben bei H.263 für alle  $i > 0$  der Wertebereich von  $QP_i$  auf  $QP_{i-1} \pm 2$  begrenzt werden.

#### 4.3.4.3 Update der Datenratenkontrolle auf der Makroblock-Ebene

Nachdem der Makroblock  $i$  des aktuellen P-Bildes codiert ist, werden alle Werte für die Berechnung des nächsten Quantisierungsparameters aktualisiert.  $d_{DCT_j}^{MB}$  ist dabei die ermittelte Bitmenge für die Codierung der DCT-Koeffizienten des Makroblocks  $j$ .

In Abhängigkeit der Entscheidungsvariablen  $lb$  werden

$$\begin{aligned} \hat{B}_{DCT_{i+1}}^F &= \hat{B}_{DCT}^F - \sum_{j=0}^i d_{DCT_j}^{MB}, k_{i+1} = k_0 - \sum_{j=0}^i t_j \cdot m_{DCT_j} \quad \text{für } lb = 1 \\ \hat{B}_{DCT_{i+1}}^F &= mk_i \cdot \hat{B}_{DCT}^F, k_{i+1} = k_0 \quad \text{für } lb = 0 \end{aligned} \quad (4.29)$$

aktualisiert.

Für den Fall  $lb = 0$  wird für die Erhöhung der Systemstabilität und eine verbesserte Einhaltung der vorgegebenen Bitmenge ein Stabilisierungsfaktor  $mk_i$  berechnet. Basierend auf (4.1) und (4.12) mit  $\alpha_i = 1; \forall i \in [0, N-1]$ , dem Ersatz von  $\sigma_i$  durch  $m_{DCT_i}$  und der Verwendung der Bitmenge  $d_{DCT_i}^{MB}$  für die Codierung der DCT-Koeffizienten des Makroblocks  $i$  erhält man:

$$\frac{d_{DCT_i}^{MB}}{\hat{B}_{DCT}^F} = \frac{t_i \cdot m_{DCT_i}}{\sum_{j=0}^{N-1} t_j \cdot m_{DCT_j}} \quad (4.30)$$

Daraus kann, durch Einbeziehung der Ergebnisse aller  $i + 1$  bereits codierten Makroblöcke, die zu erwartende Bitmenge  $\hat{B}_{DCT_s}^F$ <sup>1</sup> für das aktuelle Bild abgeschätzt werden:

$$\hat{B}_{DCT_s}^F = \left( \sum_{j=0}^i d_{DCT_j}^{MB} \right) \frac{\sum_{l=0}^{N-1} t_l \cdot m_{DCT_l}}{\sum_{j=0}^i t_j \cdot m_{DCT_j}} \quad (4.31)$$

Die Anpassung der geschätzten Bitmenge  $\hat{B}_{DCT_s}^F$  an die Zielbitmenge  $\hat{B}_{DCT}^F$  basierend auf den Ergebnissen der codierten Makroblöcke  $0, \dots, i$  liefert der Ansatz  $\hat{B}_{DCT}^F = mk_i \cdot \hat{B}_{DCT_s}^F$ . Damit wird  $mk_i$  zu:

$$mk_i = \begin{cases} \frac{\hat{B}_{DCT}^F \sum_{j=0}^i t_j \cdot m_{DCT_j}}{\left( \sum_{l=0}^{N-1} t_l \cdot m_{DCT_l} \right) \left( \sum_{j=0}^i d_{DCT_j}^{MB} \right)} & \text{für } \sum_{j=0}^i d_{DCT_j}^{MB} > 0 \\ 1 & \text{sonst} \end{cases} \quad (4.32)$$

bestimmt. Alle Werte von  $mk_i > 1$  werden auf  $mk_i = 1$  beschränkt. Für den Fall  $hb = 1$  ist es vorteilhaft, nur bei starken Abweichungen von der Sollbitmenge ( $mk_i < 0,6$ ) die Modellkorrektur anzuwenden.

Nach der Codierung einer Makroblockzeile wird der Modellparameter  $c$  mithilfe der Methode des kleinsten quadratischen Fehlers aus den ermittelten Werten der letzten  $N$  Makroblöcke basierend auf (4.27) neu ermittelt. Die Aktualisierung pro Zeile (bei CIF nach jeweils 22 und bei QCIF nach 11 Makroblöcken) ist ein Kompromiss aus schneller Anpassung an die aktuelle Signalstatistik und einem reduzierten Berechnungsaufwand (siehe Kapitel 5). Die Analysen zeigen, dass sich aufgrund der Ähnlichkeit von zeitlich nahe liegenden Bildern die Statistik der jeweils gleichen Makroblockzeile ähnelt und dieser Ansatz damit praktikabel ist.

---

1.  $s$ : Bezug auf Schätzung.

Ebenso hat sich die Begrenzung der Berechnung auf  $N$  Makroblöcke als günstig erwiesen. Es zeigt sich, dass die Anzahl ausreichend groß ist, um Extremwerte durch statistische Ausreißer und der damit verbundenen Instabilität zu verhindern. Gerade bei Szenenänderungen erweist sich die Begrenzung auf  $N$  Makroblöcke bzw. die Berechnung auf Zeilenebene als sinnvoll, da rasch auf veränderte Bedingungen reagiert wird.

Zuletzt wird der Zähler für den aktuellen Makroblock  $i$  um eins erhöht.

## 4.4 Erweiterungsmöglichkeiten

Der modulare Aufbau erlaubt, das im Abschnitt 4.3 beschriebene Basisverfahren einfach um neue Funktionalitäten zu erweitern. So können z.B. durch die Unterstützung variabler Ausgangsdatenraten, unterschiedlicher Bildtypen (I-, P- und B-Bilder) oder mehrerer Objekte pro Bild auch optionale Anforderungen aktueller Videocodecs erfüllt werden.

In der Beschreibung der zusätzlichen Funktionalitäten in den folgenden Unterabschnitten wird Bezug auf den Systemaufbau des neuen Verfahrens für die Datenratenkontrolle in Abbildung 4.4 genommen.

### 4.4.1 Analyse der DCT-Koeffizienten

Die Untersuchungen in Kapitel 3 und die Analysen im Abschnitt 4.2.4 zeigen, dass insbesondere bei niedrigen und mittleren Bitraten aufgrund der Quantisierung häufig alle DCT-Koeffizienten eines Makroblocks zu null gesetzt werden. Der Fehler wird damit identisch mit dem Signal und ist daher mit ihm korreliert. Die Codiereffizienz wird verringert und die Adaption der Rate-Distortion-Modellierung erschwert.

Daher ist es auf Makroblock-Ebene sinnvoll, die Anzahl dieser Makroblöcke zu verringern, wie die Untersuchungen der Codiereffizienz in Kapitel 5 zeigen. Einen möglichen Ansatz liefert die Definition der Quantisierung von H.263 [73]. Bei der Bestimmung von  $m_{DCT_i}$  für die Berechnung des Quantisierungsparameters wird gleichzeitig der Maximalwert der absoluten DCT-Koeffizienten  $a_{max_i}$  aller vier Y-Blöcke des aktuellen Makroblocks  $i$  ermittelt:

$$a_{max_i} = \max(|c_{j_{uv}}|); \quad \forall j \in [0, 3], \forall u, v \in [0, 8] \quad (4.33)$$

Bei allen Makroblöcken  $i$  mit  $a_{max_i} \leq 2,5 \cdot QP_i$  sind aufgrund der Definition der Quantisierung [73] alle DCT-Koeffizienten der vier Y-Blöcke null. Sind zudem die Bewegungsvektoren Nullvektoren, so wird dieser Makroblock mit sehr hoher Wahrscheinlichkeit geskippt. Im Bitstrom wird damit nur signalisiert, dass dieser Makroblock übersprungen wird, ohne zusätzliche Informationen zu übertragen.

Wird der Quantisierungsparameter  $QP_i$  mit

$$QP_i = \frac{2 \cdot a_{max_i}}{5} - 1 \quad (4.34)$$

neu bestimmt, so ist mindestens ein DCT-Koeffizient nach der Quantisierung ungleich null, solange die Einschränkungen des Wertebereichs der Quantisierungsparameter erfüllt sind.

Die beschriebene Modifikation des Quantisierungsparameters ist allerdings nicht in allen Fällen sinnvoll. Befinden sich in einem Bild nur einzelne Makroblöcke, für die die Bedingung  $a_{max_i} \leq 2,5 \cdot QP_i$  zutrifft, verschlechtert sich aufgrund der Beschränkungen des Wertebereichs und dem erhöhten Signalisierungsoverhead für den neuen Quantisierungsparameter die Codiereffizienz.

In der Implementierung wird daher die Anzahl der potenziellen Makroblöcke pro Bild ermittelt. Ist die Anzahl im vorherigen Bild größer als  $N/2$ , so wird der Ansatz für das aktuelle Bild verwendet. Die Nachberechnung des Quantisierungsparameters beeinflusst die Einhaltung der Zielbitmenge für das aktuelle Bild. Daher wird das Verfahren aus Stabilitätsgründen nur dann eingesetzt, wenn der Modellkorrekturfaktor  $mk_i > 0,8$  ist. Eine große Abweichung von der vorgegebenen Zielbitmenge wird damit vermieden.

#### 4.4.2 Variable Ausgangsdatenraten

Wesentliche Aspekte für die Unterstützung variabler Ausgangsdatenraten durch die Datenratenkontrolle sind zum einen die Möglichkeit die Pufferbefüllung exakt festlegen zu können und zum anderen eine schnelle Reaktion auf sich ändernde Randbedingungen (siehe Kapitel 3). Beide Anforderungen sind bei dem im Abschnitt 4.3 beschriebenen Verfahren für die Datenratenkontrolle erfüllt.

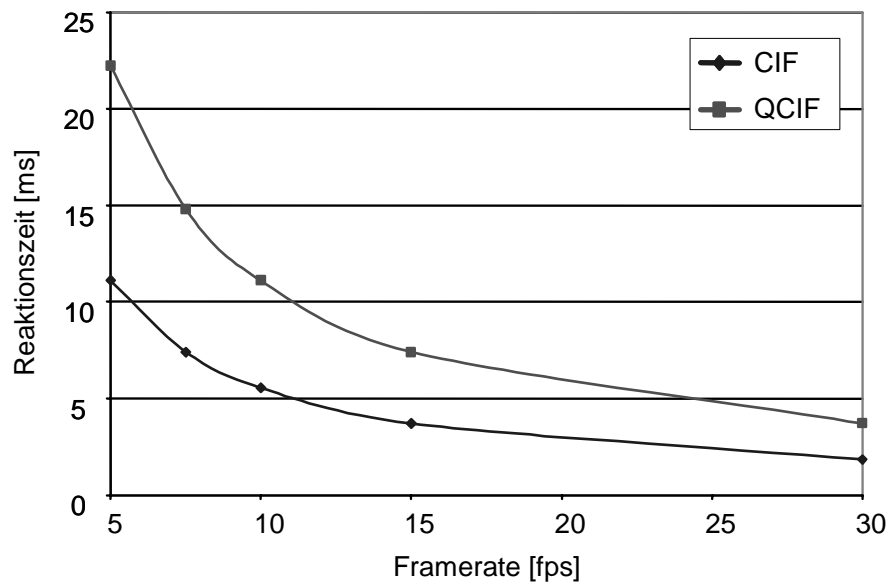
Durch die verwendete Rate-Distortion-Modellierung ist es möglich, die Zielrate und damit die Pufferbefüllung genauer einzustellen. Zudem können die notwendigen Parameter für die Modellierung bereits auf Basis eines einzigen codierten Bildes berechnet werden, sodass eine schnelle Adaption auf sich ändernde Signalcharakteristika möglich ist. Die Regelung auf Makroblock-Ebene erlaubt daher sogar Reaktionszeiten im unteren Millisekundenbereich. Für die Unterstützung variabler Ausgangsdatenraten sind im beschriebenen Verfahren für die Datenratenkontrolle allerdings Ergänzungen notwendig.

Wie bereits im Abschnitt 3.4 beschrieben, wird das jeweilige Verhalten, wie z.B. erwartete Fehlerrate, notwendige Paketwiederholungen oder die Prädiktion zukünftiger Bitraten, netzspezifisch ermittelt. Für die aktuelle Implementierung ist es ausreichend, wenn die mit hoher Wahrscheinlichkeit zu erwartende zukünftige Mindestbitrate an die Datenratenkontrolle signalisiert wird. Im idealen Fall reicht die Bitratenprädiktion entsprechend der aktuellen Ausgangspufferverzögerung in die Zukunft (siehe Abschnitt 2.3.2). Je größer der Ausgangspuffer ist, desto einfacher können Fehlprädiktionen kompensiert werden.



Die Datenratenkontrolle überprüft auf Makroblock-Ebene nach jeder Makroblock-Zeile die Ausgangsdatenrate auf Veränderungen. Bei positiver Prüfung wird die Steuervariable  $lb$  zu  $lb = 1$  modifiziert, um bereits während der Codierung des aktuellen Bildes ein möglichst exaktes Einhalten der vorgegebenen Bitmenge sicherzustellen. In Abbildung 4.6 sind die Reaktionszeiten auf veränderte Randbedingungen in Abhängigkeit von der aktuellen Framerate dargestellt, wenn nach jeder Makroblock-Zeile auf Änderungen überprüft wird.

Gleichzeitig wird die Zielbitmenge  $\hat{B}_{DCT_{i+1}}^F$  für das aktuelle Bild auf Basis der zukünftigen Bitratenschätzung angepasst. Zur Vereinfachung wird angenommen, dass für jeden Makroblock des aktuellen Bildes die gleiche Bitmenge für die Codierung zur Verfügung steht. Die Bestimmung der restlichen Quantisierungsparameter für das restliche Bild erfolgt unmittelbar nach deren Festlegung auf Basis der neuen Steuergrößen.



**Abbildung 4.6:** Reaktionszeit auf sich ändernde Randbedingungen bei Kontrolle nach jeder Makroblock-Zeile.

Auf der Bild-Ebene sind basierend auf den neuen Randbedingungen folgende Parameter zu bestimmen:

- Neuberechnung der Ausgangspuffergröße  $b_{max}^e$  mit  $b_{max}^e = br \cdot R'$  in Abhängigkeit der neuen Bitrate  $R'$  bzw. deren Neuschätzung und der relativen Ausgangspufferverzögerung  $br$ .
- Entscheidung, ob eine hochbitratige Videocodierung vorliegt ( $hb = 1$ , sonst  $hb = 0$ ).

- Begrenzung der Zielfüllhöhe  $b_t$  des Ausgangspuffers auf 10% der relativen Puffergröße  $br$ . Setzen der Steuervariablen  $lb = 1$  bzw.  $nff = 1$ , um ein möglichst exaktes Einhalten der Zielbitmenge sicherzustellen. Erfolgt die Schätzung der zukünftigen Bitrate mit einer hohen Genauigkeit, so kann die Zielfüllhöhe  $b_t$  des Ausgangspuffers unverändert bleiben. Die Pufferwirkung kann weiterhin gezielt für die Encodierung genutzt werden.
- Bei großen Änderungen der Bitrate ist zusätzlich eine Neubestimmung der Zielframerate  $FR$  notwendig. In der Verifikation hat sich eine einfache Anpassung der Framerate in Abhängigkeit von der Bitrate mittels einer quadratischen Funktion durch den Ursprung und die Punkte (10 kBit/s; 7,5 fps) und (768 kBit/s; 30 fps) bewährt. Weitergehende Ansätze zur Bestimmung der Zielframerate in Abhängigkeit der aktuellen Rate-Distortion-Charakteristik sind z.B. in [10] zu finden.

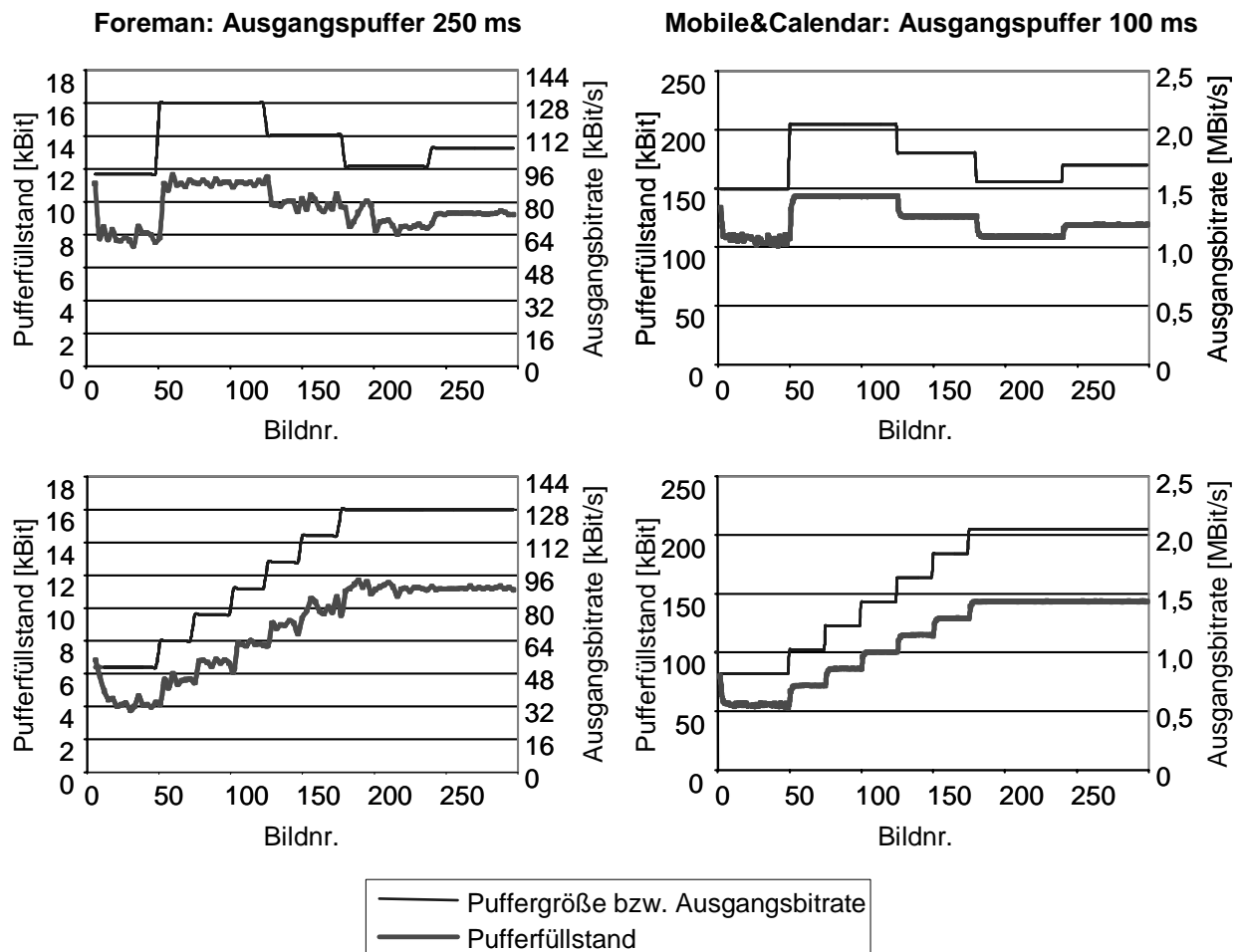
Für die weitere Codierung bilden die oben definierten Steuerparameter die Grundlage. In der aktuellen Implementierung werden nach zehn Bildern ohne Änderung alle Steuerparameter auf Basis der im Abschnitt 4.3 definierten Vorgaben neu bestimmt. Durch die Aufhebung der für variable Ausgangsdatenraten notwendigen Restriktionen wird sichergestellt, dass unter den neuen Randbedingungen eine effektive Datenratenkontrolle gewährleistet ist.

In Abbildung 4.7 sind beispielhaft die Ergebnisse für die zwei Sequenzen Foreman und Mobile&Calendar bei variablen Ausgangsdatenraten und damit variierender Größe des Ausgangspuffers dargestellt<sup>1</sup> (siehe Abschnitt 2.3.2). Dabei wird die Sequenz Foreman im QCIF-Format mit einer maximalen Bitrate von 128 kBit/s bei 10 fps codiert und die Sequenz Mobile&Calendar im CIF-Format mit einer maximalen Bitrate von 2 MBit/s bei 30 fps. In allen Fällen wird von einer IPPP...-Codierung ausgegangen, wobei die Zielpufferhöhe fest auf 70% der Maximalhöhe festgelegt wird.

Bei der Videocodierung mit niedrigen Datenraten bzw. Frameraten und bei kleinen Ausgangspuffergrößen sind die Schwankungen bei der Pufferbefüllung relativ hoch im Vergleich zu der Puffergröße. Insbesondere bei kleinen Puffergrößen ist eine schnelle und effektive Anpassung der Bitrate notwendig.

---

1. Ein direkter Leistungsvergleich mit den ausgewählten Vergleichsverfahren aus der Standardisierung von H.263++ bzw. MPEG-4 ist nicht möglich, da diese Verfahren keine variablen Ausgangsdatenraten berücksichtigen.



**Abbildung 4.7:** Verhalten des vorgeschlagenen Verfahrens für die Datenratenkontrolle bei schwankenden Ausgangsdatenraten.

#### 4.4.3 Modellierung von I- und B-Bildern bzw. Makroblöcken

In Kommunikationsanwendungen wird meist das erste Bild intracodiert und alle Folgenden als P-Bilder übertragen. Das notwendige Update der Bildinformation wird durch eingestreute I-Makroblöcke gewährleistet.

In Anwendungen, bei denen die Verzögerung höher sein kann, werden zur Steigerung der Bildqualität bzw. zur Verringerung der notwendigen Rechenleistung zusätzlich B-Bilder eingesetzt. Bei höherbitratigen Anwendungen werden zusätzliche I-Bilder im Bitstrom codiert, um z.B. den Wechsel zwischen verschiedenen laufenden Videoübertragungen bzw. ein vereinfachtes Bearbeiten der Datenströme zu ermöglichen.

Die Berücksichtigung der unterschiedlichen Rate-Distortion-Charakteristik der I-, P- und B-Codierung erfordert, dass das Verfahren aus Abschnitt 4.3 auf GOP- und Bild-Ebene erweitert wird.

Die Bestimmung der Bitmenge für die Codierung des jeweiligen Bildes in Abhängigkeit des Bildtyps, basiert im Wesentlichen auf (4.30) und der Erweiterung auf die GOP- bzw. Bild-Ebene. Für die Codierung von  $K$  Bildern gleichen Typs mit der Gesamtbitmenge  $\hat{B}_{DCT}^{GOP}$  für die Codierung der DCT-Koeffizienten ergibt sich damit für die Bestimmung der Bitmenge  $\hat{B}_{DCT_n}^F$  des Bildes  $n$ :

$$\hat{B}_{DCT_n}^F = \frac{\sum_{j=0}^{N-1} t_{nj} \cdot m_{DCT_{nj}}}{\sum_{l=0}^{K-1} \sum_{j=0}^{N-1} t_{lj} \cdot m_{DCT_{lj}}} \hat{B}_{DCT}^{GOP} \quad (4.35)$$

Der Doppelindex z.B. von  $m_{DCT_{nj}}$  adressiert dabei den jeweiligen Makroblock  $j$  im Bild  $n$  mit jeweils  $N$  Makroblöcken. Die Werte aller Bilder  $j \geq n$  der Doppelsumme  $\sum_{l=0}^{K-1} \sum_{j=0}^{N-1} t_{lj} \cdot m_{DCT_{lj}}$  sind zum Zeitpunkt der Codierung des Bildes  $n$  unbekannt. Als Näherung der Summe wird daher innerhalb des Codierprozesses der Mittelwert

$$\overline{tm} = \frac{1}{M} \sum_{l=0}^{M-1} \sum_{j=0}^{N-1} t_{lj} \cdot m_{DCT_{lj}} \quad (4.36)$$

aus den letzten  $M$  Bildern (z.B.  $M = 10$ ) berechnet. Die gewünschte Gesamtbitmenge  $\hat{B}_{DCT}^{GOP}$  wird in Abhängigkeit der aktuellen bzw. zukünftig geschätzten Übertragungsdatenrate, der Anzahl der zu codierenden Bilder, der Zielframerate und dem Füllstand des Ausgangspuffers bestimmt (siehe auch Abschnitt 4.3.3).

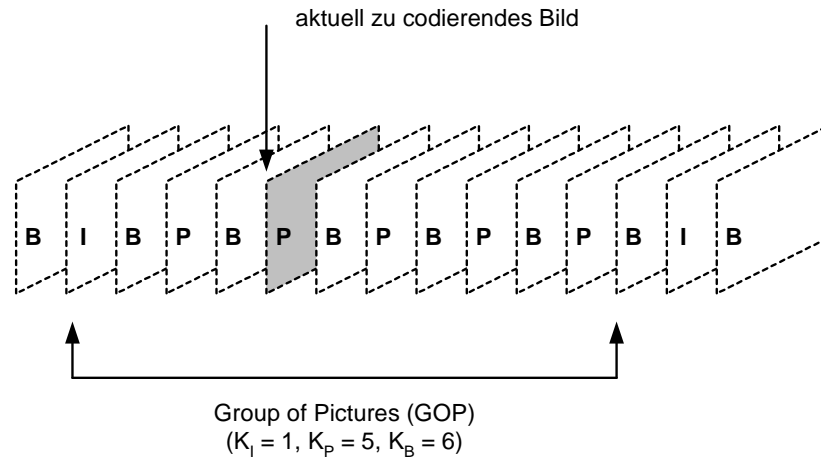
Erweitert man den Ansatz aus (4.35) auf I-, P- und B-Bilder so erhält man:

$$\hat{B}_{DCT}^{F_{I,P,B}} = \frac{\sum_{j=0}^{N-1} t_{ij} \cdot m_{DCT_{ij}}}{K_I \cdot \overline{tm}_I + K_P \cdot \overline{tm}_P + K_B \cdot \overline{tm}_B} \hat{B}_{DCT}^{GOP} \quad (4.37)$$

Die Mittelwerte  $\overline{tm}_I$ ,  $\overline{tm}_P$  und  $\overline{tm}_B$  werden dabei nach (4.36) für jeden Bild-Typ getrennt bestimmt. Wie in Abbildung 4.8 dargestellt sind pro GOP jeweils  $K_I$ ,  $K_P$  bzw.  $K_B$  Bilder des jeweiligen Typs enthalten. Das aktuell zu codierende Bild  $n$  in Abbildung 4.8 ist ein P-Bild, sodass in diesem Beispiel  $\hat{B}_{DCT_n}^{F_P}$  bestimmt werden muss. Für die Sicherstellung einer gleichmäßigen Qualität kann das Ergebnis noch auf Plausibilität anhand der bisherigen Ergebnisse überprüft und bei Bedarf modifiziert werden.

B-Bilder dienen in aktuellen Codecs nicht als Referenz für andere Bilder. Daher hat es sich in der Praxis bewährt, sie mit geringerer Qualität als die anderen Bilder zu codieren. In [139] wird z.B. ein Gewichtungsfaktor von 0,7 vorgeschlagen.

- 
1.  $GOP$ : Bezug auf die Bitmenge, die für den ganzen GOP zur Verfügung steht. Der Index  $DCT$  drückt den Bezug auf die DCT-Transformationskoeffizienten aus.
  2. Die Indizes  $I$ ,  $P$  und  $B$  beziehen sich jeweils auf die Codierart: I-, P- oder B-Bilder.



**Abbildung 4.8:** GOP-Struktur.

Je nach Anforderung werden nach der Codierung die jeweiligen Werte für den aktuellen Bild-Typ bzw. für  $\hat{B}_{DCT}^{GOP}$  aktualisiert. Für einen geschlossenen GOP wird die Bitmenge für die Codierung des Bildes  $n$  von  $\hat{B}_{DCT}^{GOP}$  abgezogen und  $K_I$ ,  $K_P$  bzw.  $K_B$  um eins verringert. Nach der Codierung aller  $K_I + K_P + K_B$  Bilder im GOP werden die Werte für den nächsten GOP neu initialisiert. Eine andere Methode, die Werte für den GOP nach Abschluss der Codierung des Bildes  $n$  neu zu bestimmen, besteht darin, das nächste zukünftige Bild außerhalb des GOPs mit in die Berechnung einzubeziehen und die Werte des GOPs neu zu initialisieren.

Auf Makroblock-Ebene muss die unterschiedliche Rate-Distortion-Charakteristik der verschiedenen Codierungsmethoden ebenfalls berücksichtigt werden. Zur Bestimmung des jeweiligen Quantisierungsparameters wird für (4.27) ein spezifischer Skalierungsfaktor  $c_I$ ,  $c_P$  bzw.  $c_B$  auf Basis bereits codierter Makroblöcke bestimmt und für die Berechnung nach (4.28) verwendet.

In einigen Fällen, z.B. bei häufigen Szenenwechseln, der Übertragung über fehlerbehaftete Netze oder sehr hohen und schnellen Bewegungsanteilen im Bild, werden auch in P- oder B-Bildern viele Makroblöcke intracodiert. Auch hier bietet es sich an, nach dem gleichen Prinzip die Bitmenge für das Gesamtbild aufzuteilen und spezifische Skalierungsfaktoren  $c_I$ ,  $c_P$  bzw.  $c_B$  für die Berechnung der Quantisierungsparameter nach (4.28) zu bestimmen.

#### 4.4.4 Multi-Objektfähigkeit

In MPEG-4 [57] ist es möglich, nicht nur ein ganzes Bild zu codieren, sondern die Szene in mehrere Objekte aufzuteilen, die jeweils getrennt codiert und übertragen werden. Erst nach der Decodierung werden die verschiedenen Objekte zu einer Gesamtszene zusammengesetzt (siehe Kapitel 2). Für die Codierung ist es notwendig, die Gesamtbitmenge auf die verschiedenen Objekte aufzuteilen.

Die Bestimmung der Bitmenge, die für die Codierung des jeweiligen Objekts zur Verfügung steht, basiert analog zu dem Vorgehen im Abschnitt 4.4.3 im Wesentlichen auf einer entsprechenden Erweiterung von (4.30). Für die Codierung von  $K$  Objekten gleichen Typs mit der Gesamtbitmenge  $\hat{B}_{DCT}^{MO}$ <sup>1</sup> ergibt sich damit für die Bestimmung der Bitmenge  $\hat{B}_{DCT_n}^O$ <sup>2</sup> des Objekts  $n$  mit  $N_n$  Makroblöcken:

$$\hat{B}_{DCT_n}^O = \frac{\sum_{j=0}^{N_n-1} t_{nj} \cdot m_{DCT_{nj}}}{\sum_{l=0}^{K-1} \sum_{j=0}^{N_l-1} t_{lj} \cdot m_{DCT_{lj}}} \hat{B}_{DCT}^{MO} \quad (4.38)$$

Der Doppelindex z.B. von  $m_{DCT_{nj}}$  adressiert dabei den jeweiligen Makroblock  $j$  im Objekt  $n$ . Die Gesamtbitmenge  $\hat{B}_{DCT}^{MO}$  wird in Abhängigkeit von der aktuellen bzw. zukünftig geschätzten Übertragungsdatenrate, der gewählten Frame- bzw. Objektrate, dem Füllstand des Ausgangspuffers und dem erwarteten Overhead für die Codierung der Bewegungsvektoren und der Steuerinformation analog zu Abschnitt 4.3.3. bestimmt.

Die unterschiedliche Rate-Distortion-Charakteristik der verschiedenen Objekte muss auf Makroblock-Ebene ebenfalls berücksichtigt werden. Die Objekte werden auf Makroblock-Ebene wie getrennte Bilder codiert. Die Statistik und die Parametrisierung werden daher getrennt für jedes Objekt bestimmt. Durch entsprechende Erweiterungen ist es auch möglich, Objekte unterschiedlicher Typen und Objektraten zu berücksichtigen.

---

1.  $MO$ : Bezug auf Mehrere BildObjekte.

2.  $O$ : Bezug auf ein BildObjekt.

### 4.4.5 Subjektive Bildqualität

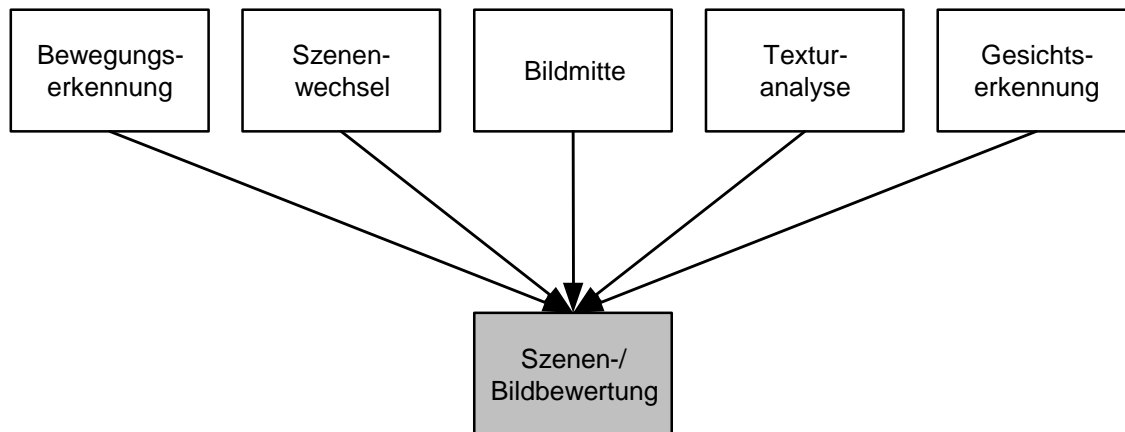
Die bisherige Zielsetzung ist die Minimierung der Verzerrung eines Bildes gewesen, wobei als Verzerrungsmaße *PSNR* bzw. *MSE* verwendet worden sind. Hierfür ist der Faktor  $\alpha_i$  im Abschnitt 4.3, für jeden einzelnen Makroblock  $i$  auf  $\alpha_i = 1$  gesetzt worden. Allerdings entsprechen die objektiven Verzerrungsmaße wie *PSNR* oftmals nicht der subjektiven Wahrnehmung eines menschlichen Betrachters.

Das bisherige Verfahren verteilt die zur Verfügung stehende Bitmenge anhand von Rate-Distortion-optimalen Gesichtspunkten auf die einzelnen Bildteile. Als Fehlermaß ist die Energie des Fehlers zwischen Original und rekonstruiertem Bild verwendet worden. Es findet also keine Bewertung der Bildinhalte statt. Eine Bewertung hinsichtlich psychooptischer Kriterien führt somit direkt zu einer subjektiv höheren Bildqualität. Ein menschlicher Betrachter nimmt somit eine bessere Bildqualität wahr. Aktuelle nur bedingt echtzeitfähige Realisierungen von MPEG-4 [18], [31] implementieren einfache Verfahren, um die subjektive Bildqualität zu erhöhen.

Die Modellierung der psychooptischen Wahrnehmung des Menschen ist bereits Gegenstand in zahlreichen Untersuchungen, so z.B. [6], [28], [96], [123], [129] und [130]. Dabei zeigt sich, dass eine realistische Modellierung im Vergleich zur Videocodierung rechenaufwendig ist. Unter anderem spielen folgende Aspekte bei der subjektiven Wahrnehmung des Menschen eine wichtige Rolle:

- Vordergrundobjekte, bewegte Objekte, Gesichter und das Bildzentrum erwecken die Aufmerksamkeit des Betrachters.
- Neue Objekte oder Szenenwechsel können erst nach einer bestimmten Zeit wahrgenommen werden.
- Farben werden mit unterschiedlicher Empfindlichkeit wahrgenommen.

Ein einfaches Modell für die ersten beiden Punkte ist in Abbildung 4.9 dargestellt. Dabei wird die Sequenz bzw. die Einzelbilder auf Bild-Ebene hinsichtlich folgender Einzelfaktoren bewertet: einer Gesichtserkennung, einer Texturanalyse, der Bildmitte, Szenenwechsel und einer Bewegungserkennung. Anschließend werden für die Steuerung der Datenratenkontrolle die Einzelergebnisse zusammengefasst.



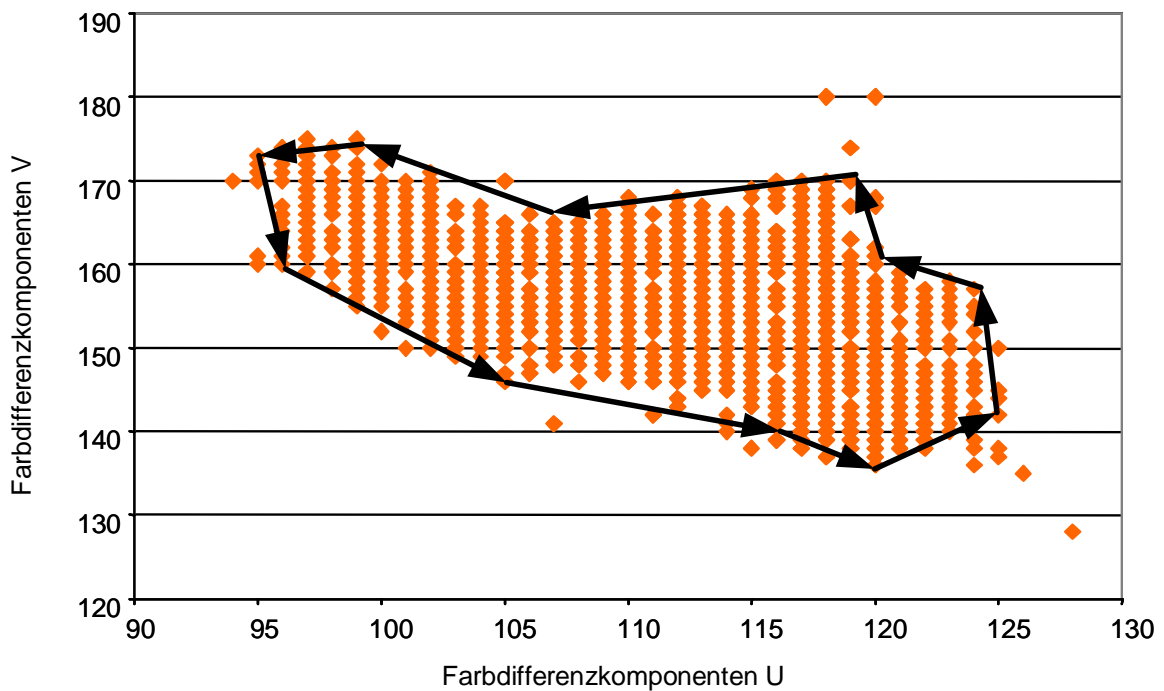
**Abbildung 4.9:** Bewertungsmodell für die Gewichtung einzelner Bildbereiche hinsichtlich der subjektiven Bildwahrnehmung.

Für eine Überprüfung des Ansatzes ist gegenüber Abbildung 4.9 ein vereinfachtes Modell zur Steuerung der Datenratenkontrolle implementiert worden. Auf Bild- und Makroblock-Ebene dient der Faktor  $\alpha_i$  in (4.2), (4.3) bzw. (4.12) als Gewichtungsfaktor, um die Verzerrung für jeden einzelnen Makroblock  $i$  zu gewichten. Dabei werden die Einzelergebnisse aus einer Gesichtsfarbenerkennung, einer Texturanalyse auf Basis der Varianz  $\sigma_i^2$  des Makroblocks  $i$  und die Berücksichtigung der relativen Lage der Makroblöcke im Bild, zu einer Gesamtbewertung zusammengefasst.

Für die Ermittlung der Bewertungsfaktoren  $\alpha_i$  für jeden Makroblock  $i$  im aktuellen Bild werden zunächst alle Werte der Einzelbewertungen berechnet und in Matrixform entsprechend der Makroblockanordnung im Bild aufgestellt:

- Die festen Werte der Matrix für die Auswertung der relativen Lage des jeweiligen Makroblocks sind in Abbildung 4.11 Teilbild (b) visualisiert.
- Für die Texturanalyse wird die Varianz  $\sigma_i^2$  für jeden Makroblock  $i$  des Originalbildes ermittelt und mit dem Maximalwert des aktuellen Bildes normiert. In die Matrix wird der Quotient  $1/\sigma_{norm_i}^2$  für jeden Makroblock eingetragen.
- Als vereinfachte Methode der Gesichtserkennung dient eine Analyse der Gesichtsfarben. Es wird überprüft, ob die Mehrzahl der Pixel im aktuellen Makroblock in einem fest vorgegebenen UV-Farbbereich enthalten ist. Für diese Makroblöcke wird der jeweilige Matrixwert gleich 1 gesetzt, andernfalls 0. Der in Abbildung 4.10 dargestellte Bereich ist anhand einer Farbanalyse der jeweiligen Gesichtsfarben aus verschiedenen Bildern und Videosequenzen ermittelt worden. Zusätzlich ist der Farbbereich so vereinfacht, dass eine Beschreibung mittels eines Polygons möglich ist (im Beispiel anhand von 11 Geraden). Dies reduziert den notwendigen Rechenaufwand bei der Farbanalyse.





**Abbildung 4.10:** Gesichtsfarbenbereich in Abhängigkeit der Farbdifferenzkomponenten U und V, der aus verschiedenen Bildern und Videosequenzen ermittelt worden ist.

Für die abschließende Gesamtauswertung werden die Matrixwerte der Einzelergebnisse aufsummiert und normiert, sodass der Maximalwert gleich 1 ist. In Abbildung 4.11 ist eine beispielhafte Auswertung für das Bild 45 der Sequenz Mother&Daughter dargestellt.

Die Ergebnisse verschiedener Testsequenzen zeigen, dass sich durch die Gewichtung auf Makroblock-Ebene bei vorgegebener Datenrate die mittleren *PSNR*-Werte um ca. 0,4 dB gegenüber der Minimierung der Verzerrung pro Bild anhand des *MSE*-Wertes verschlechtern. Allerdings erscheint dabei in vielen Fällen die subjektive Bildqualität höher.

In Kapitel 6 wird die notwendige Rechenleistung für die Realisierung des einfachen Modells im Vergleich zum eigentlichen Videocodec analysiert. Eine Erweiterung hinsichtlich der Erkennung und Bewertung von Szenenwechseln oder bewegten Objekten ist möglich, so z.B. durch die in [38] vorgestellte Methode. Allerdings ist dafür eine deutlich höhere Rechenleistung notwendig.

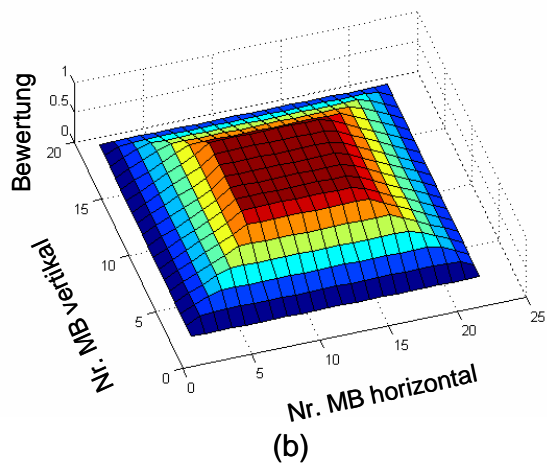
Die Verifikation des neuen Verfahrens für die Datenratenkontrolle erfolgt in Kapitel 5 auf Basis des gemittelten *PSNR*-Wertes. Dabei wird die subjektive Bildqualität nicht berücksichtigt, um einfach reproduzierbare (objektive) Vergleichsergebnisse der verschiedenen Verfahren zu erhalten.

**Bild 45 der Sequenz  
Mother&Daughter**



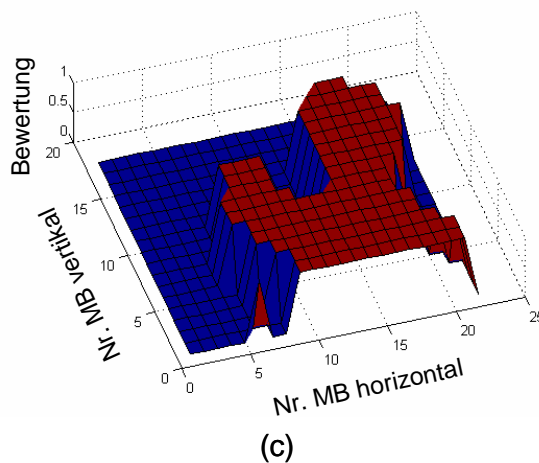
(a)

**Gewichtung der Bildlokalität**



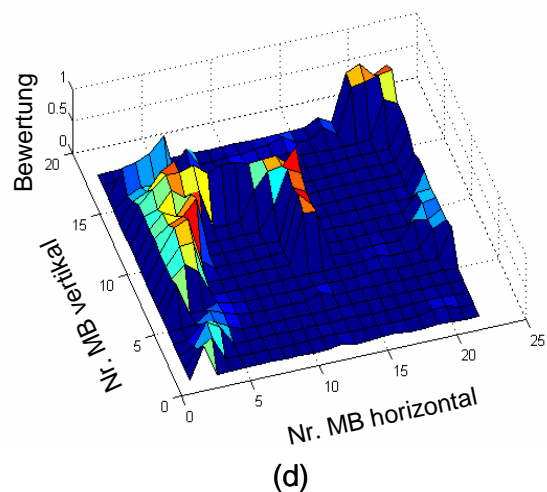
(b)

**Vereinfachte Gesichtsfarbenanalyse**



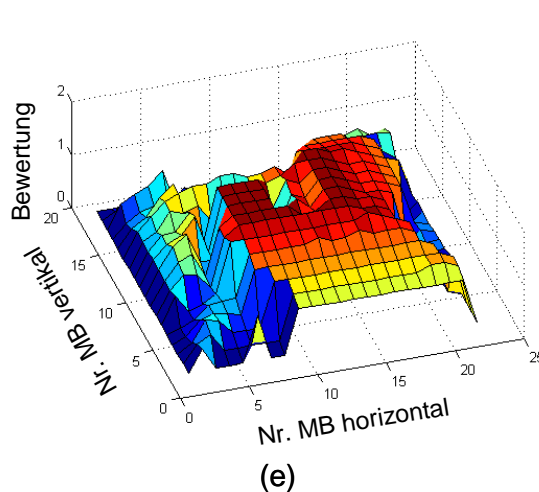
(c)

**Aktivitätsanalyse**



(d)

**Gesamtauswertung**



(e)

**Abbildung 4.11:** Einzelergebnisse der Gewichtung einzelner Bildbereiche hinsichtlich der subjektiven Bildwahrnehmung.

## 4.5 Zusammenfassung

In echtzeitfähigen Verfahren für die Datenratenkontrolle muss die aktuelle Rate-Distortion-Charakteristik des Bildmaterials in Verbindung mit dem Encoder prädiktiv geschätzt werden. Durch die Anwendung einer Rate-Distortion-Modellierung kann dabei gewährleistet werden, dass die prädiktive Regelung des Encoders die notwendige Stabilität aufweist.

Die Rate-Distortion-Theorie selbst ist wohlbekannt. Allerdings existieren analytische Lösungen nur für Sonderfälle. Für die Anwendung in realen Systemen, speziell für echtzeitfähige Lösungen, ist es notwendig, die Modelle anhand bestimmter Annahmen zu vereinfachen. Zur Sicherung einer hohen Codiereffizienz ist für die Bestimmung der Quantisierungsparameter sowohl die Genauigkeit der Modelle selbst als auch die adaptive Anpassung an die tatsächlichen Gegebenheiten im Encoder notwendig. Insbesondere die Anwendbarkeit über einen breiten Bitratenbereich auch für unterschiedliche Signalcharakteristika stellt hohe Anforderungen an die Modellierung.

Auf Basis der Anforderungen und die Untersuchung des Rate-Distortion-Verhaltens im Anwendungsbereich ist in diesem Kapitel ein neuer echtzeitfähiger Ansatz für die Rate-Distortion-Modellierung vorgeschlagen worden. Er beruht sowohl auf einer genauen Rate-Distortion-Schätzung als auch auf einem verallgemeinerten Ansatz zur vereinfachten Beschreibung von nieder- bis hochbitratigen Videosignalen. Zudem ist ein neuer Schätzparameter auf Basis der DCT-Koeffizienten definiert worden, der gegenüber der Varianz insbesondere im Bitratenbereich unter 0,3 Bit/Pixel überlegen ist.

Die neue Rate-Distortion-Modellierung bildet die Grundlage des neuen Verfahrens für die Datenratenkontrolle. Wichtig für die Leistungsfähigkeit unter realen Bedingungen ist die Anpassung an das aktuelle Eingangssignal und die automatische Systemparametrisierung. Dies ist beim Design des neuen Verfahrens für die Datenratenkontrolle berücksichtigt worden.

Für die Unterstützung optionaler Leistungsmerkmale von aktuellen Videostandards ist eine Erweiterung des neuen Verfahrens notwendig. Folgende Ansätze sind zusätzlich definiert worden: Unterstützung von variablen Ausgangsdatenraten, Modellierung von I- und B-Bildern bzw. Makroblöcken und die Unterstützung mehrerer Objekte in einer Szene.

Die objektive Bildqualität spiegelt oftmals nur ungenügend die Wahrnehmung des Menschen wider. Aus diesem Grund ist im Design des neuen Verfahrens berücksichtigt worden, andere Verzerrungsmaße für die Qualitätsbewertung heranzuziehen. Jeder Makroblock im Bild kann gewichtet werden, um das Codierergebnis besser an die Wahrnehmung des Menschen anzupassen. Die Gewichtung erfolgt beispielhaft auf Basis eines einfachen Modells für die Ermittlung subjektiv wichtiger Bildbereiche.



## 5 Verifikation des neuen Verfahrens für die Datenratenkontrolle

Der Schwerpunkt dieses Kapitels ist die Verifikation des neuen Verfahrens für die Datenratenkontrolle. Nach der Einführung im Abschnitt 5.1 werden die Testbedingungen für die anschließenden Untersuchungen im Abschnitt 5.2 spezifiziert. Die Leistungsfähigkeit des neuen Verfahrens wird unter verschiedenen Bedingungen im Abschnitt 5.3 im Vergleich zu Referenzverfahren aus der Standardisierung von H.263 und MPEG-4 untersucht und bewertet. Neben dem Vergleich der Codiereffizienz über verschiedene Bitratenbereiche werden für weitere Auswertungen die Ergebnisse aus 38 einzelnen Untersuchungen diskutiert, die in einem Testset zusammengefasst sind. Die Testbedingungen decken die Anforderungen an die Videocodierung ab, wie sie im Abschnitt 3.2 beschrieben sind. Die wichtigsten Ergebnisse des Kapitels werden im Abschnitt 5.5 zusammengefasst.

### 5.1 Einführung

Für eine Bewertung der Leistungsfähigkeit eines Videocodecs ist eine Verifikation anhand von realen Bildsequenzen notwendig. Bei der Standardisierung von MPEG-4, H.263 und H.264 sind für die Beurteilung der verschiedenen Verfahren sowohl die Testsequenzen als auch die genaue Vorgehensweise definiert worden [54], [59], [61], [62], [64], [65], [70], [77], [122]. Speziell für die Bewertung der Datenratenkontrolle sind bei MPEG-4 die folgenden zusätzlichen Bewertungskriterien definiert worden: subjektive Bildqualität, mittlere  $PSNR$ -Werte<sup>1</sup>, mittlere Framerate der codierten Sequenz und die Fluktuation der Bitmenge pro Bild. Des Weiteren sind die Testsequenzen und die Testbedingungen für die Beurteilung der verschiedenen Verfahren für die Datenratenkontrolle definiert worden [59].

Für eine Identifikation möglicher Schwachstellen ist es notwendig, dass die Testbedingungen möglichst umfassend die Anforderungen des jeweiligen Einsatzbereiches abdecken. So erfordern z.B. Sequenzen mit geringer zeitlicher Änderung, dass die Codierung der Informationen für die Bewegungskompensation eine möglichst geringe Anzahl von Bits benötigt. Bei Sequenzen mit schnellen Bewegungsanteilen ist es dagegen sinnvoll, einen großen Suchbereich abzudecken. Die Codierung der längeren Bewegungsvektoren erfordert allerdings eine größere Bitmenge.

Insbesondere unter diesem Aspekt ist es wichtig, die Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle unter verschiedensten Bedingungen und mit unterschiedlichem Bildmaterial zu verifizieren.

---

1. Aufgrund der festgelegten Ausgangsdatenrate des Encoders besteht eine direkte Abhängigkeit zwischen Codiereffizienz und erzielter Bildqualität, sodass beide Begriffe gleichbedeutend verwendet werden.

## 5.2 Testbedingungen

Für die Ermittlung der Leistungsfähigkeit dient die Implementierung des neuen Verfahrens für die Datenratenkontrolle in einem H.263-Encoder. Die Realisierung entspricht dem im Abschnitt 4.3 beschriebenen Verfahren, das um die Analyse der DCT-Koeffizienten nach Abschnitt 4.4.1 ergänzt worden ist. Im Folgenden wird dieses Verfahren als FLEX<sup>1</sup> bezeichnet.

Als Vergleichsmaßstab dienen die Referenzverfahren aus der MPEG-4- und H.263-Standardisierung, die jeweils im Abschnitt 2.3.3.4 beschrieben sind. Für eine einfache Referenzierung auf die einzelnen Verfahren werden im Weiteren folgende Kurzbezeichnungen gewählt:

- MPEG4: für das bildbasierte Verfahren aus der MPEG-4 Standardisierung, das jeweils einen Quantisierungsparameter pro Bild bestimmt (siehe Kapitel 2).
- MPEG4MB: für das makroblockbasierte Verfahren aus der MPEG-4 Standardisierung, das jeweils einen Quantisierungsparameter pro Makroblock in einem Bild berechnet (siehe Kapitel 2).
- TMN8: für das makroblockbasierte Verfahren aus der H.263-Standardisierung mit entsprechendem Bezug auf das Testmodell (hier TMN8), das die Implementierungsempfehlungen für die Umsetzung des H.263-Standards beschreibt (siehe Kapitel 2).

Die einzelnen Verfahren für die Datenratenkontrolle sind entsprechend dem Aufbau von Abbildung 4.4 in den H.263-Encoder implementiert worden. Auf GOP-Ebene wird für alle Verfahren der Algorithmus nach Abschnitt 4.3.2 verwendet. Bei jedem Verfahren wird das nächste zu codierende Bild nach den gleichen Vorgaben ausgewählt. Für die Untersuchungen gilt stets  $b_{tmax} = 80\%$  als Vorgabe für die maximale Befüllungshöhe des Ausgangspuffers.

Für die Verifikation sind zwölf Sequenzen entsprechend den Vorschlägen aus der Standardisierung von MPEG-4, H.263 bzw. H.264 mit unterschiedlicher Charakteristik jeweils in den Bildformaten QCIF und CIF ausgewählt worden [54], [59], [64], [65], [69], [122]. Die Einordnung der verwendeten Testsequenzen in verschiedene Klassen mit entsprechender Charakterisierung befindet sich in Tabelle 5.1 bzw. im Anhang A.

Klasse	Sequenzname	Charakterisierung
A	Container Ship, Mother&Daughter	Wenig Details und Bewegung
B	Bowing, Coastguard, Foreman, News, Irene	Mehr Details, wenig Bewegung oder wenig Details und mehr Bewegung
C	Mobile&Calendar, Stephan, Table Tennis	Viele Details, mehr Bewegung oder mehr Details und viel Bewegung
E	Children, Weather	Gemischt natürliches und synthetisches Bildmaterial

**Tabelle 5.1:** Einordnung der Testsequenzen in unterschiedliche Klassen [59], [65].

1. FLEX: Kennzeichnung für die Flexibilität des neuen Verfahrens für die Datenratenkontrolle.

Für die Analyse des Verhaltens der Datenratenkontrolle unter verschiedenen Bedingungen werden auf Basis der in Kapitel 3 definierten Anforderungen bzw. entsprechend den Vorschlägen aus der Standardisierung von MPEG-4 und H.264 zwei unterschiedliche Methoden für die Verifikation gewählt:

1. Bestimmung der Codiereffizienz von einzelnen Sequenzen über einen Bitratenbereich in Abhängigkeit vom Bildformat. Hierzu werden die Sequenzen aus Tabelle 5.1 bei QCIF im Bitratenbereich zwischen 10 - 128 kBit/s und bei CIF zwischen 128 - 2048 kBit/s jeweils bei unterschiedlichen Ausgangspuffergrößen untersucht.
2. Bestimmung der Codiereffizienz für 38 festgelegte Einzeluntersuchungen, die zu einem repräsentativen Testset zusammengestellt sind. Dabei werden folgende Anforderungen erfüllt: Bitraten von 10 - 2048 kBit/s, Frameraten von 7,5 - 30 fps und Bildmaterial mit unterschiedlicher Charakteristik. In Tabelle 5.2 sind die einzelnen Testbedingungen aufgeführt, die jeweils mit 100, 250, 500 und 1000 ms Ausgangspufferverzögerung untersucht werden.

Nr.	Kurzbeschreibung (Sequenz; Format; Datenrate; Framerate)	Nr.	Kurzbeschreibung (Sequenz; Format; Datenrate; Framerate)
1	Mother&Daughter; QCIF; 10 kBit/s; 7,5 fps	20	Irene; CIF; 384 kBit/s; 30 fps
2	Container Ship; QCIF; 15 kBit/s; 7,5 fps	21	Table Tennis; CIF; 384 kBit/s; 15 fps
3	Container Ship; QCIF; 24 kBit/s; 10 fps	22	Weather; CIF; 384 kBit/s; 15 fps
4	Mother&Daughter; QCIF; 24 kBit/s; 10 fps	23	Children; CIF; 512 kBit/s; 15 fps
5	News; QCIF; 24 kBit/s; 10 fps	24	Mobile&Calendar; CIF; 512 kBit/s; 15 fps
6	Coastguard; QCIF; 48 kBit/s; 10 fps	25	Stephan; CIF; 512 kBit/s; 15 fps
7	Foreman; QCIF; 48 kBit/s; 10 fps	26	Table Tennis; CIF; 512 kBit/s; 30 fps
8	News; CIF; 48 kBit/s; 7,5 fps	27	Children; CIF; 768 kBit/s; 30 fps
9	News; QCIF; 48 kBit/s; 10 fps	28	Mobile&Calendar; CIF; 768 kBit/s; 30 fps
10	Coastguard; QCIF; 112 kBit/s; 15 fps	29	Stephan; CIF; 768 kBit/s; 30 fps
11	Foreman; QCIF; 112 kBit/s; 10 fps	30	Children; CIF; 1024 kBit/s; 30 fps
12	News; CIF; 112 kBit/s; 15 fps	31	Mobile&Calendar; CIF; 1024 kBit/s; 30 fps
13	Bowing; CIF; 128 kBit/s; 10 fps	32	Stephan; CIF; 1024 kBit/s; 30 fps
14	Irene; CIF; 128 kBit/s; 10 fps	33	Children; CIF; 1536 kBit/s; 30 fps
15	Table Tennis; QCIF; 128 kBit/s; 10 fps	34	Mobile&Calendar; CIF; 1536 kBit/s; 30 fps
16	Weather; CIF; 128 kBit/s; 10 fps	35	Stephan; CIF; 1536 kBit/s; 30 fps
17	Bowing; CIF; 384 kBit/s; 15 fps	36	Children; CIF; 2048 kBit/s; 30 fps
18	Coastguard; CIF; 384 kBit/s; 30 fps	37	Mobile&Calendar; CIF; 2048 kBit/s; 30 fps
19	Foreman; CIF; 384 kBit/s; 15 fps	38	Stephan; CIF; 2048 kBit/s; 30 fps

**Tabelle 5.2:** Randbedingungen im definierten Testset.

Die Bestimmung der Codiereffizienz basiert auf der Berechnung der gemittelten *PSNR*-Werte der codierten Einzelbilder mit einem H.263-Encoder ohne optionale Modi. Die sonstigen Parameter wie Datenrate, Ausgangspuffergröße und Framerate werden vorgegeben. Dabei wird die für Kommunikationsanwendungen übliche IPPP...-Codierung gewählt. Für die Bestimmung des gemittelten *PSNR*-Wertes wird für jedes Bild entsprechend der Zielframerate der jeweilige *PSNR*-Wert nach (2.5) auf Basis des entsprechenden Originalbildes ermittelt und aufaddiert. Die Gesamtsumme wird anschließend durch die Bildanzahl entsprechend der gewählten Framerate dividiert<sup>1</sup>. Werden einzelne Bilder bei der Codierung übersprungen, so wird der *PSNR*-Wert zwischen dem letzten codierten Bild und dem übersprungenen Originalbild für die Berechnung verwendet. Ein Überspringen von Bildern wird damit bei der Berechnung des gemittelten *PSNR*-Wertes berücksichtigt.

Voruntersuchungen mit Ausgangspuffergrößen größer als 1000 ms zeigen, dass sich hier keine signifikanten Änderungen gegenüber 1000 ms ergeben, sodass im Folgenden nur Ergebnisse bis 1000 ms Ausgangspufferverzögerung aufgeführt werden.

## 5.3 Untersuchungsergebnisse

Auf Basis der in Abschnitt 5.2 definierten Testbedingungen wird die Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle FLEX ermittelt und bewertet. Hierzu wird in Abschnitt 5.3.1 die Codiereffizienz von FLEX im Vergleich zu den Verfahren aus der Standardisierung von MPEG-4 und H.263 untersucht. In Abschnitt 5.4 werden verschiedene Variationen von FLEX analysiert, die bereits in Kapitel 4 als Designoption diskutiert worden sind.

Während in diesem Kapitel der Schwerpunkt auf der Untersuchung und Bewertung der Codiereffizienz von FLEX liegt, wird in Kapitel 6 auf die Realisierungsaspekte und die Ansätze zur Minimierung der Ende-zu-Ende-Verzögerung eingegangen.

### 5.3.1 Leistungsfähigkeit von FLEX

Für eine Verifikation der Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle FLEX wird dessen Codiereffizienz unter praxisrelevanten Bedingungen im Vergleich zu den Verfahren MPEG4, MPEG4MB und TMN8 ermittelt. Hierzu wird jeweils der gemittelte *PSNR*-Wert einer codierten Sequenz bestimmt, wobei die Datenrate, die Framerate und die Größe des Ausgangspuffers vorgegeben werden.

Die Verifikation selbst gliedert sich in zwei Teilbereiche: Zuerst werden für unterschiedliche Sequenzen die gemittelten *PSNR*-Werte für alle vier Verfahren für die Datenratenkontrolle über definierte Bitratenbereiche in Abhängigkeit von der Bildgröße ermittelt. Im Anschluss werden die einzelnen *PSNR*-Werte der im Testset (siehe Tabelle 5.2) definierten Randbedingungen bestimmt. In allen zwei Untersuchungen werden jeweils unterschiedliche Ausgangspufferverzögerungen berücksichtigt.

---

1. Die Schwankung der Bildqualität von Bild zu Bild entsprechend Abschnitt 2.2.1 geht dabei nicht in die Bewertung ein, sondern wird gesondert ausgewertet.



Die Ermittlung des *PSNR*-Verlaufs einzelner Sequenzen über einen definierten Bitratenbereich basiert auf folgende Testbedingungen:

- QCIF-Sequenzen: Bitratenbereich von 10 - 128 kBit/s mit 8 Einzelmessungen, Framerate bis einschließlich 15 kBit/s gleich 7,5 fps bzw. für höhere Bitraten gleich 10 fps und Ausgangspufferverzögerungen jeweils 100, 250, 500 bzw. 1000 ms. In Abhängigkeit von den Ergebnissen wird der Bitratenbereich für einzelne Sequenzen so weit eingegrenzt, dass alle Verfahren für die Datenratenkontrolle interpretierbare Ergebnisse liefern. Der Anteil der übersprungenen Bilder darf dabei bei maximal 10 % der Gesamtsequenzlänge liegen. Eine detaillierte Aufstellung der jeweiligen Parameter befindet sich in Tabelle B.1.
- CIF-Sequenzen: Bitratenbereich von 128 - 2048 kBit/s mit jeweils 6 Einzelmessungen, Framerate bis 256 kBit/s gleich 10 fps, für 512 kBit/s gleich 15 fps bzw. für höhere Bitraten gleich 30 fps und Ausgangspufferverzögerungen jeweils 100, 250, 500 bzw. 1000 ms. In Abhängigkeit von den Ergebnissen wird der Bitratenbereich für einzelne Sequenzen so weit eingegrenzt, dass alle Verfahren für die Datenratenkontrolle interpretierbare Ergebnisse liefern (siehe Tabelle B.2). Der Anteil der übersprungenen Bilder darf dabei ebenfalls bei maximal 10 % der Gesamtsequenzlänge liegen.

In Tabelle 5.3 sind die Ergebnisse der QCIF-Sequenzen über den Bitratenbereich 10 - 128 kBit/s bzw. der CIF-Sequenzen über den Bitratenbereich 128 - 2048 kBit/s zusammengestellt. In einem ersten Schritt sind die einzelnen Untersuchungsergebnisse pro Sequenz, Bitrate, Framerate und Bildformat mit jeweils einer der definierten Ausgangspufferverzögerungen ermittelt worden (siehe Tabelle B.1 für die QCIF-Sequenzen bzw. Tabelle B.2 für die CIF-Sequenzen). Für die Auswertung ist dann jeweils die Differenz zwischen den gemittelten *PSNR*-Werten in dB von FLEX und den *PSNR*-Werten in dB des jeweiligen Vergleichsverfahrens gebildet worden. Bei positiven Werten weist FLEX eine höhere Codiereffizienz aus.

Die relativen Ergebnisse pro Vergleichsverfahren sind dann für jede Sequenzklasse getrennt zusammengefasst worden, indem der Mittelwert, der maximale und der minimale Wert der Einzelergebnisse bestimmt worden sind. Diese Ergebnisse sind für QCIF in der Tabelle B.3 und für CIF in der Tabelle B.4 aufgeführt.

Für die Auswertung in Tabelle 5.3 sind die Ergebnisse der Tabellen B.3 und B.4 über alle Ausgangspufferverzögerungen pro Vergleichsverfahren, Sequenzklasse und Bildformat zusammengefasst worden. Die Interpretation der Ergebnisse lautet damit z.B. für Tabelle 5.3 vierte Zeile für die Spalte TMN8: FLEX erzielt bei QCIF und dem vorgegebenen Bitratenbereich von 10 - 128 kBit/s bei Sequenzen der Klasse A gegenüber TMN8 über alle getesteten Ausgangspuffergrößen im Schnitt 0,23 dB höhere *PSNR*-Werte. Der maximale Gewinn beträgt 0,69 dB und der kleinste Gewinn beträgt -0,16 dB.

<b>Zusammenfassung des Qualitätsvergleichs der jeweiligen Bitratenbereiche, der Sequenzklasse und der Ausgangspufferverzögerungen 100, 250, 500 und 1000 ms:</b>				
Klasse	Sequenzen	Mid. / Max. / Min. relativ zu FLEX in PSNR [dB]		
		TMN8	MPEG4	MPEG4MB
Bildformat: QCIF				
A	Container Ship, Mother&Daughter	0,23 / 0,69 / -0,16	0,17 / 0,98 / -0,19	0,41 / 1,39 / -0,11
B	Bowing, Coastguard, Foreman, Irene, News	0,25 / 0,86 / -0,10	0,57 / 5,25 / -0,09	0,42 / 1,66 / 0,01
C	Stephan, Table Tennis	0,26 / 0,70 / -0,11	0,30 / 0,96 / -0,05	0,44 / 0,98 / 0,15
E	Children, Weather	0,45 / 1,63 / -0,36	0,66 / 1,75 / 0,17	0,44 / 1,70 / -0,24
<b>Gesamtergebnis für QCIF-Sequenzen</b>		<b>0,30 / 1,63 / -0,36</b>	<b>0,43 / 5,25 / -0,19</b>	<b>0,43 / 1,70 / -0,24</b>
Bildformat: CIF				
A	Container Ship, Mother&Daughter	0,24 / 0,57 / -0,02	0,33 / 1,41 / -0,12	0,42 / 1,12 / -0,02
B	Bowing, Coastguard, Foreman, Irene, News	0,20 / 0,73 / -0,47	0,51 / 4,96 / -0,11	0,39 / 0,95 / -0,03
C	Mobile&Calendar, Stephan, Table Tennis	0,22 / 0,58 / -0,02	0,25 / 2,57 / -0,10	0,28 / 0,65 / 0,10
E	Children, Weather	0,77 / 2,08 / 0,01	0,68 / 2,03 / -0,11	0,40 / 2,19 / -0,13
<b>Gesamtergebnis für CIF-Sequenzen</b>		<b>0,36 / 2,08 / -0,47</b>	<b>0,44 / 4,96 / -0,12</b>	<b>0,37 / 2,19 / -0,13</b>

**Tabelle 5.3:** Zusammenfassung des Qualitätsvergleichs der QCIF- und CIF-Sequenzen.

Die Ergebnisse zeigen, dass FLEX im Mittel bei jeder Sequenzklasse unter den oben genannten Bedingungen in Abhängigkeit vom Vergleichsverfahren eine Steigerung des *PSNR*-Werts von ca. 0,4 dB erreicht. Dabei schwankt die Höhe des *PSNR*-Gewinns in Abhängigkeit des Vergleichsverfahrens, der Höhe der Ausgangspufferverzögerung und der jeweiligen Sequenz bzw. Sequenzklasse. So kann z.B. bei QCIF bei 100 ms Ausgangspufferverzögerung im Vergleich zu TMN8 bei der Klasse E ca. 1 dB Gewinn erzielt werden, aber bei den restlichen Klassen (A, B und C) wird nur eine Verbesserung von jeweils ca. 0,1 dB erreicht (siehe Tabelle B.3).

Bei höheren Ausgangspufferverzögerungen (über 250 ms) zeigt der Performancegewinn gegenüber allen Vergleichsverfahren eine geringere Abweichung vom mittleren Gewinn. Vor allem MPEG4 liefert bei Ausgangspufferverzögerungen über 250 ms neben FLEX eine gute Performance.

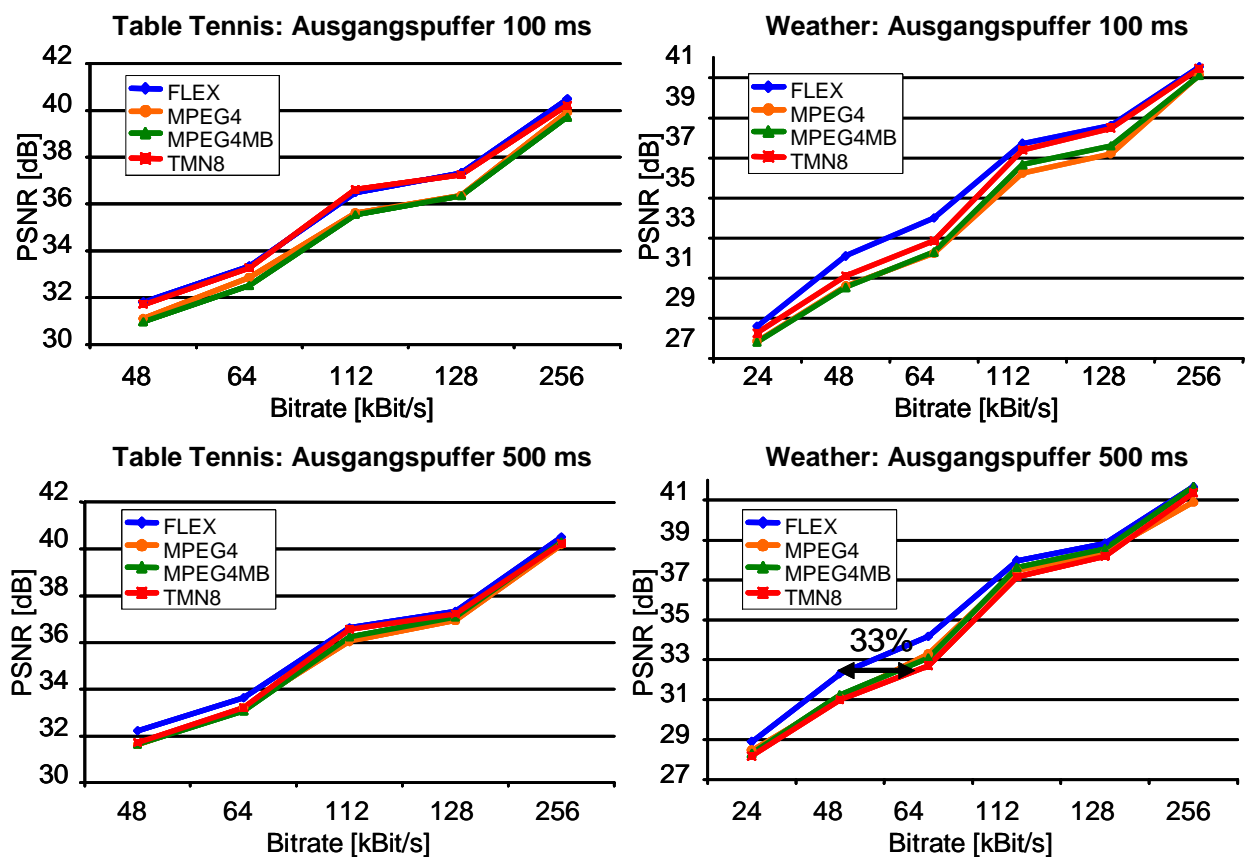
TMN8 zeigt insbesondere bei kleinen Ausgangspuffergrößen eine zu FLEX vergleichbare Codiereffizienz. Von größeren Ausgangspuffergrößen kann TMN8 nur wenig profitieren, sodass FLEX in diesem Bereich deutlich bessere Ergebnisse liefert.

Im Vergleich zu MPEG4MB sind die mittleren *PSNR*-Gewinne nahezu bei allen Ausgangspuffergrößen und Sequenzen ausgeglichen und bewegen sich meist zwischen 0,3 und 0,4 dB.

Insgesamt zeigt sich anhand der Mittel- bzw. der Minimalwerte, dass FLEX nur in Einzelfällen eine schlechtere Performance als die Vergleichsverfahren liefert. Der *PSNR*-Verlust beträgt maximal zwischen -0,1 und -0,5 dB. In den meisten Fällen kann aber eine deutlich gesteigerte Codiereffizienz erzielt werden. Der Gewinn bezüglich der *PSNR*-Werte liegt zwischen ca. 1,6 dB und 5,25 dB. Der teilweise sehr hohe Gewinn gegenüber MPEG4 entsteht vor allem dadurch, dass dieses Verfahren nicht auf Makroblock-Ebene den Quantisierungsparameter bestimmen kann. In kritischen Situationen bzw. bei Fehlschätzungen müssen deshalb Bilder übersprungen werden. Dies wirkt sich negativ auf die gemittelten *PSNR*-Werte aus.

Für verschiedene Sequenzen sind beispielhaft die *PSNR*-Kennlinien in Abhängigkeit von der Bitrate für QCIF in Abbildung 5.1 bzw. für CIF in Abbildung 5.2 dargestellt. Dabei sind jeweils die absoluten *PSNR*-Werte für das jeweilige Bitratenkontrollverfahren angegeben. Die Auswertung entspricht der Tabelle 5.3. Beispielhaft ist in den Abbildungen 5.1 und 5.2 zusätzlich eine Bitratenreduzierung von ca. 30 % eingezeichnet, die FLEX bei gleicher Bildqualität ermöglicht.

Deutlich sichtbare Einbrüche der Codiereffizienz, wie z.B. beim MPEG4-Verfahren in Abbildung 5.2 sichtbar (Sequenz Forman, 100 ms Ausgangspuffergröße), entstehen durch ein vermehrtes Überspringen von Bildern bei der Codierung. Die Anzahl der zu überspringenden Bilder wird auf GOP-Ebene nach (4.22) in Abhängigkeit des aktuellen Pufferfüllstandes ermittelt und bei der Bestimmung der mittleren *PSNR*-Werte berücksichtigt.



**Abbildung 5.1:** Vergleich der Codiereffizienz des neuen Verfahrens FLEX mit den Referenzverfahren beim Bildformat QCIF.

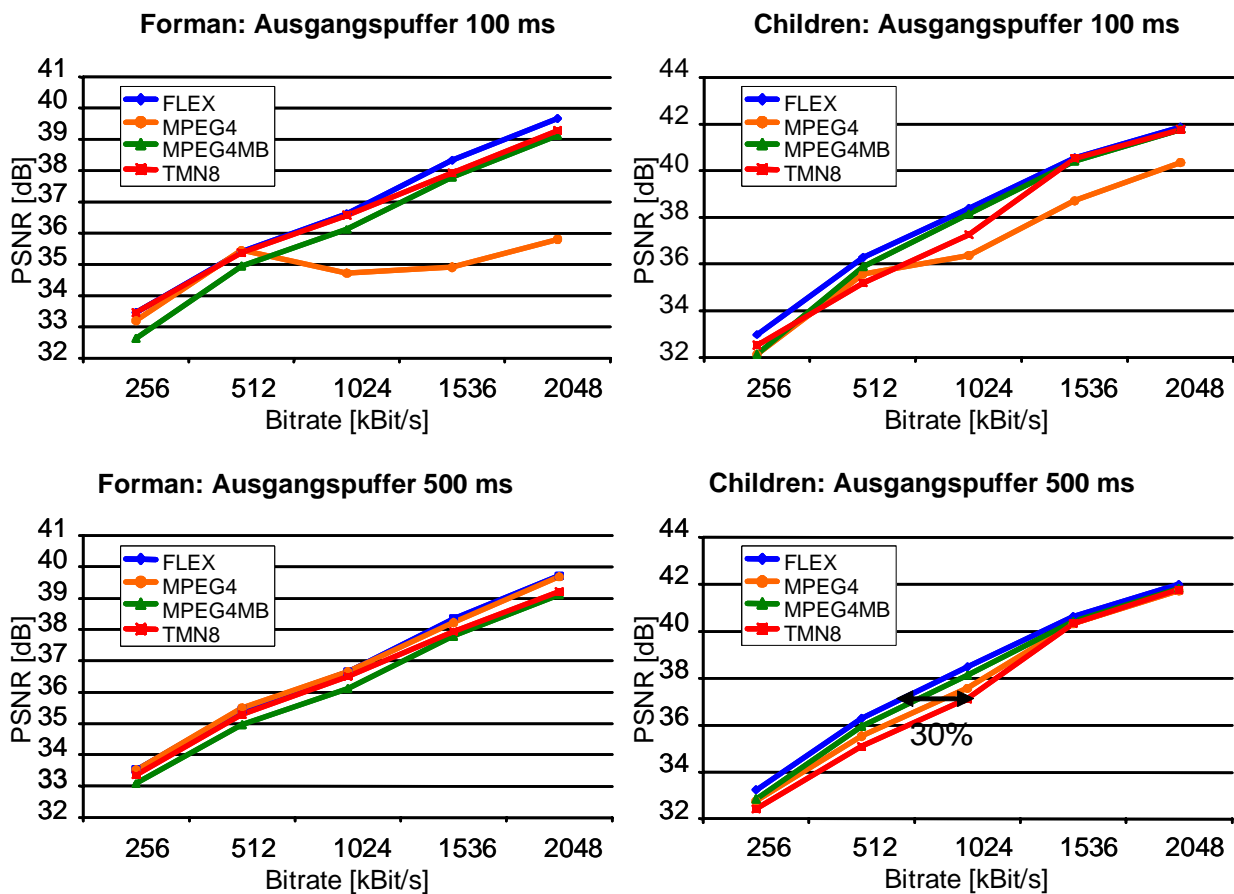


Abbildung 5.2: Vergleich der Codiereffizienz des neuen Verfahrens FLEX mit den Referenzverfahren beim Bildformat CIF.

Zusätzlich wird im Folgenden die Leistungsfähigkeit von FLEX im Vergleich zu den Referenzverfahren MPEG4, MPEG4MB und TMN8 auf Basis verschiedener, in einem Testset definierten Randbedingungen bestimmt. Auch hier wird zwischen verschiedenen Ausgangspuffergrößen unterschieden.

In Abbildung 5.3 sind drei Auswertungen des Testsets für die Ausgangspufferverzögerungen 100, 250 und 1000 ms dargestellt. Für jeden Einzeltest ist der mittlere *PSNR*-Wert der Sequenz unter Vorgabe der Datenrate, der Ausgangspuffergröße und der Framerate bestimmt worden und repräsentiert einen Messpunkt der dargestellten Kennlinien.

Dabei beziehen sich die oberen *PSNR*-Ergebnisse auf die linke y-Achse und geben die absoluten Werte der verschiedenen Einzeltests in dB an. Die rechte y-Achse gibt die Skalierung für die unteren Kurven an. Sie stellen den Vergleich der Verfahren jeweils zu FLEX dar und geben damit die *PSNR*-Differenz zwischen FLEX und den Vergleichsverfahren in dB an. Bei positiven Ergebnissen weist FLEX einen höheren *PSNR*-Wert als das jeweilige Vergleichsverfahren auf.

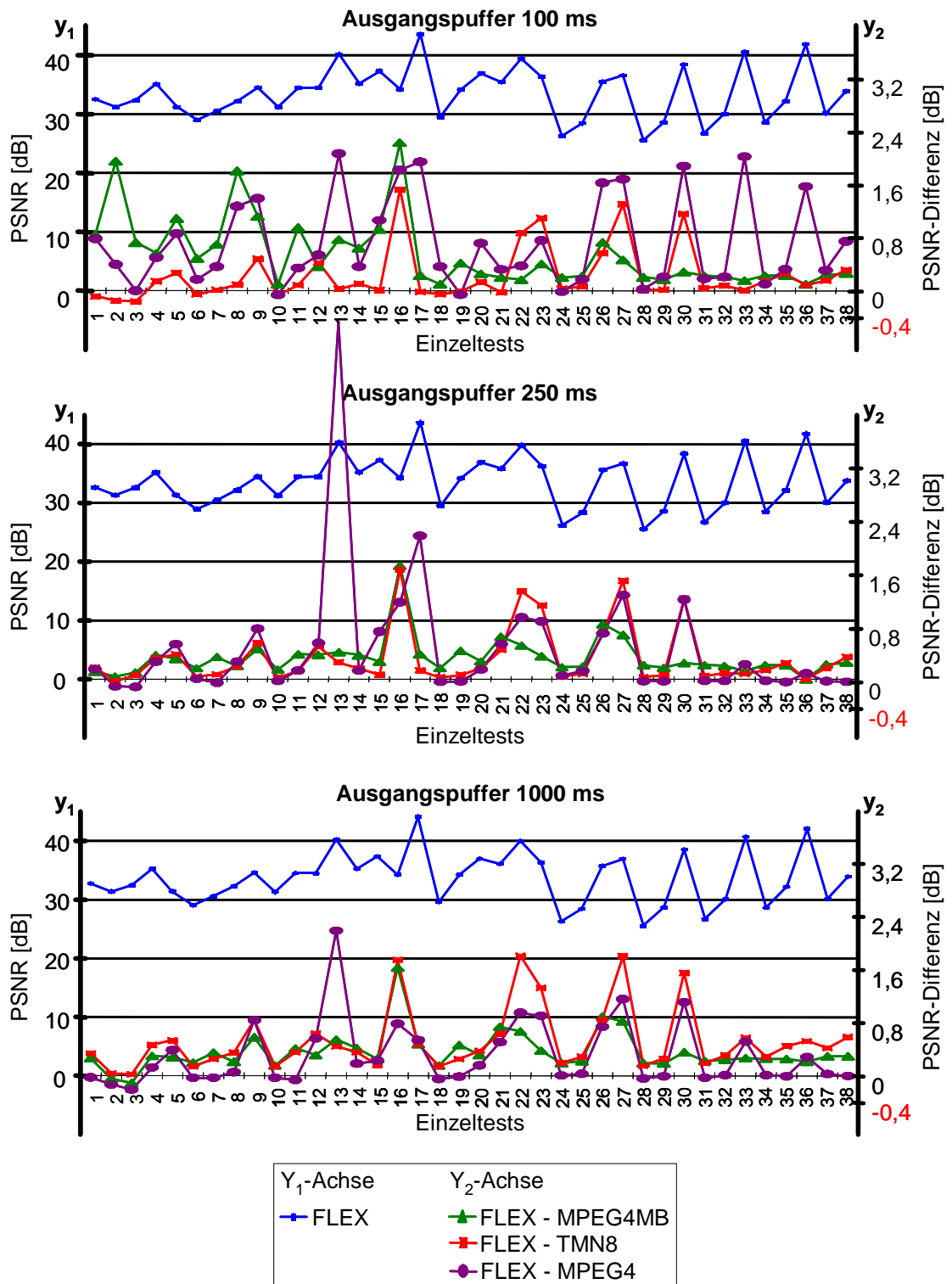


Abbildung 5.3: Qualitätsvergleich des neuen Verfahrens FLEX mit den Referenzverfahren.

Die Auswertung bestätigt die Ergebnisse der Tabelle 5.3. Zusammenfassend ergeben sich folgende Aussagen:

- FLEX kann in nahezu allen Untersuchungen eine Steigerung der *PSNR*-Werte gegenüber den Vergleichsverfahren erzielen. Der Gewinn ist abhängig von der Ausgangspuffergröße und den jeweils vorgegebenen Randbedingungen und schwankt zwischen ca. -0,5 und 2,2 dB mit einem mittleren Gewinn von ca. 0,7 dB. Nur in einzelnen Fällen sind die *PSNR*-Werte im Vergleich zu den Referenzverfahren minimal geringer (bis zu -0,5 dB).
- Bei Ausgangspuffergrößen über 250 ms kann MPEG4 im Vergleich zu TMN8 und MPEG4MB eine hohe Codiereffizienz erzielen. Speziell in Bereichen mit kleinen Ausgangspuffern und niedrigen Bitraten werden allerdings oftmals Bilder übersprungen, sodass hier teilweise nur geringe *PSNR*-Werte erreicht werden. Eine starke Änderung der Bildsignalstatistik wirkt sich zudem negativ auf die Ermittlung der Modellparameter aus. Für die Berechnung dieser Parameter wird auf bis zu 20 codierte Bilder zurückgegriffen.
- MPEG4MB zeigt bei allen Testsequenzen und bei jeder Ausgangspufferverzögerung meist eine mittlere Codiereffizienz. Bei Ausgangspuffergrößen bis 250 ms erreicht es im Vergleich zu TMN8 eine geringere Codiereffizienz. Für größere Puffergrößen zeigt MPEG4MB eine höhere Codiereffizienz als TMN8.
- TMN8 erzielt insbesondere bei kleinen Ausgangspuffergrößen (bis ca. 250 ms) oder niedrigen Bitraten sehr gute Ergebnisse. Außerhalb dieser Bereiche ist die Codiereffizienz im Vergleich zu den anderen Verfahren MPEG4 und MPEG4MB meist nur durchschnittlich bzw. oftmals auch unterdurchschnittlich.

Neben den gemittelten *PSNR*-Ergebnissen sind zusätzlich für einzelne Sequenzen der Verlauf der *PSNR*-Werte pro Bild und die Qualitätsschwankungen innerhalb von einzelnen Bildern untersucht worden. Im Anhang B sind in Abbildung B.1 die Verläufe für die CIF-Sequenz Mobile&Calendar mit 2048 kBit/s und 30 fps bzw. in Abbildung B.2 für die QCIF-Sequenz Foreman mit 48 kBit/s und 10 fps beispielhaft dargestellt.

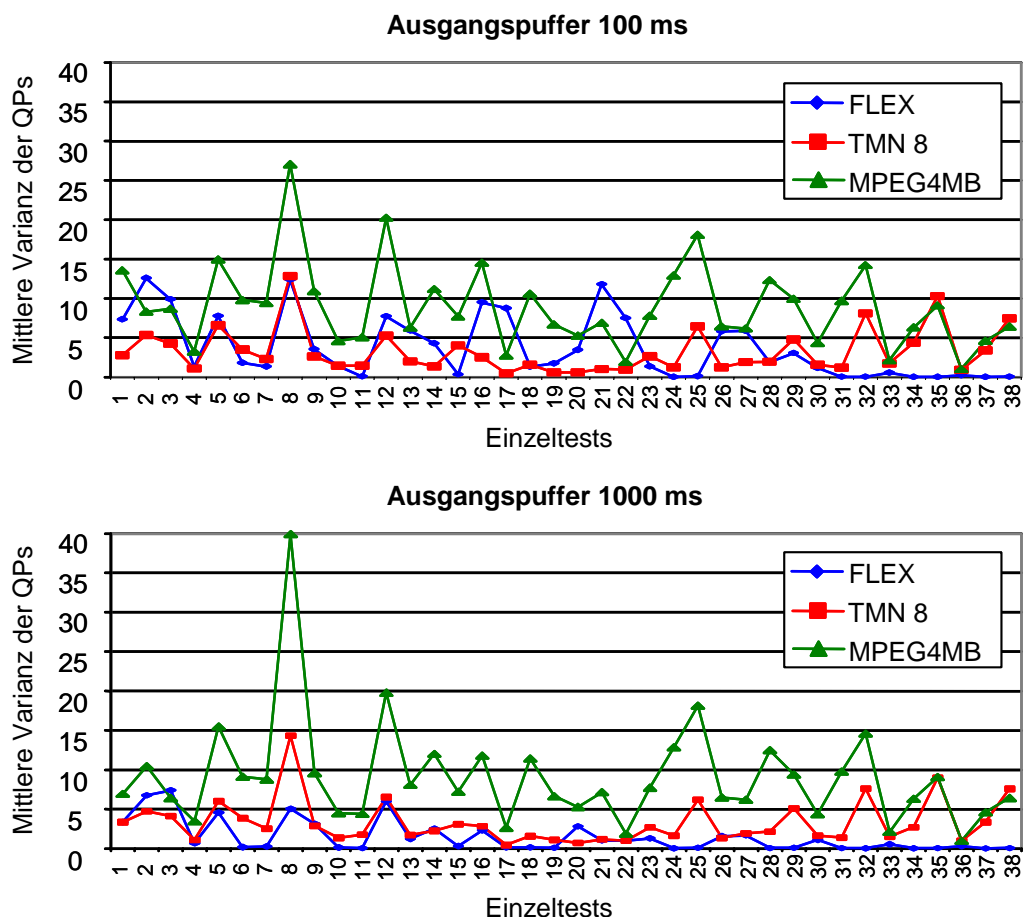
Es zeigt sich, dass nach einer Einschwingphase von bis zu 50 Bildern meist der generelle Verlauf der verschiedenen Verfahren für die Datenratenkontrollen ähnlich ist. Allerdings sind die Kennlinien teilweise zu höheren bzw. niedrigeren *PSNR*-Werten verschoben, so z.B. in Abbildung B.2 die Auswertung für 100 ms Ausgangspufferverzögerung. Die einzelnen Verfahren für die Datenratenkontrolle weisen allerdings teilweise einen von Bild zu Bild schwankenden Bildqualitätsverlauf auf. So sind z.B. in Abbildung B.1 bei FLEX mit 100 ms Ausgangspufferverzögerung *PSNR*-Schwankungen um ca. 0,5 dB innerhalb der ersten 50 Bildern zu sehen. Die anderen Verfahren für die Datenratenkontrolle weisen dagegen bei 1000 ms Ausgangspufferverzögerung eine ähnlich schwankende Bildqualität auf.

Insgesamt zeigen alle Verfahren für die Datenratenkontrolle bei Qualitätsschwankungen von Bild zu Bild ein ähnliches Verhalten, das vor allem durch unvorhersehbare Änderungen in der Signalstatistik verursacht wird. In Abhängigkeit von der verwendeten Rate-Distortion-Modellierung sind davon unterschiedliche Sequenzen oder Sequenzabschnitte betroffen.

Zusätzlich zu der Untersuchung der Qualitätsschwankungen von Bild zu Bild sind noch die Schwankungen innerhalb von einzelnen Bildern untersucht worden. Oftmals ist dieser Effekt nicht nur störend, sondern deckt Schwächen des jeweiligen Verfahrens auf. So wird z.B. durch Fehler in der Rate-Distortion-Modellierung die zur Verfügung stehende Bitmenge nicht den Anforderungen entsprechend auf das Bild verteilt. Die letzten Makroblöcke werden dann in einer schlechten Qualität codiert, um die aus der Bitallocation vorgegebene Bitmenge nicht zu überschreiten.

Die Ergebnisse für das definierte Testset sind in Abbildung 5.4 für unterschiedliche Ausgangspufferverzögerungen für die Verfahren FLEX, MPEG4MB und TMN8 dargestellt. Als Maß für die Qualitätsschwankung ist die Varianz des Quantisierungsparameters ausgewertet worden. Hierzu ist für jede codierte Sequenz im Testset die Varianz der Quantisierungsparameter einzeln pro codiertes Bild berechnet und über die Sequenz gemittelt worden.

Die Ergebnisse sind über alle Sequenzen im Testset für die Verfahren in Abbildung 5.4 dargestellt. Das bildbasierte Verfahren MPEG4 ist in Abbildung 5.4 nicht enthalten, da hier die Varianz der Quantisierungsparameter, der pro Bild immer konstant ist, stets gleich null ist.



**Abbildung 5.4:** Vergleich der Qualitätsschwankung innerhalb von Einzelbildern für verschiedene Verfahren für die Datenratenkontrolle.

Die Auswertung der Ergebnisse von Abbildung 5.4 zeigen deutlich, dass das Verfahren MPEG4MB in nahezu allen Bereichen die höchsten Qualitätsschwankungen innerhalb von einzelnen Bildern aufweist. Dieser Effekt tritt bei allen Ausgangspuffergrößen im Bereich von 100 - 1000 ms in ähnlicher Weise auf. Bei 100 ms Ausgangspuffergrößen weisen FLEX und TMN8 in etwa gleich große Schwankungen auf. Allerdings sind jeweils unterschiedliche Sequenzen betroffen. Von größeren Ausgangspuffern kann vor allem FLEX profitieren. Bei 1000 ms Ausgangspuffergröße weist es gegenüber den anderen Verfahren die kleinste Qualitätsschwankung innerhalb von einzelnen Bildern auf.

Bei den beiden Verfahren TMN8 und MPEG4MB wird oftmals der untere Bildbereich mit schlechterer Qualität codiert. Dies weist darauf hin, dass für diesen Bereich aufgrund von Schätz- bzw. Modellfehlern eine zu geringe Bitmenge für die Codierung zur Verfügung steht.

### 5.3.2 Zusammenfassung und Interpretation der Ergebnisse

Die Untersuchungsergebnisse zeigen, dass das in Kapitel 4 entworfene Verfahren für die Datenratenkontrolle die geforderte Flexibilität aufweist. FLEX stellt bei verschiedenen Sequenzen im Bitratenbereich von 10 kBit/s bis zu 2 MBit/s und unterschiedlichen Ausgangspuffergrößen unter den definierten Bedingungen eine effiziente Codierung sicher.

Im Vergleich zu den Referenzverfahren aus der Standardisierung TMN8, MPEG4 und MPEG4MB kann FLEX bei den Tests in den meisten Fällen eine Steigerung der Codiereffizienz erreichen. Der *PSNR*-Gewinn ist dabei abhängig von der Ausgangspuffergröße und den jeweiligen Randbedingungen und schwankt im Testset zwischen ca. -0,5 und 2,2 dB mit einem mittleren Gewinn von ca. 0,7 dB. Nur in einzelnen Fällen sind die *PSNR*-Werte minimal geringer (maximal -0,5 dB). Entsprechend ist dadurch bei gleicher Bildqualität in Abhängigkeit von der aktuellen Rate-Distortion-Charakteristik eine Reduzierung der Bitrate um bis zu 30 % und mehr möglich.

Die Qualitätsschwankungen von Bild zu Bild sind ähnlich den Ergebnissen der Vergleichsverfahren. Allerdings weist FLEX gegenüber den ebenfalls auf Makroblock-Ebene arbeitenden Verfahren TMN8 und MPEG4MB meist geringere Qualitätsschwankungen innerhalb von einzelnen Bildern auf. Aufgrund dieser Ergebnisse zeigt sich, dass das neue Verfahren unter den geforderten Bedingungen stabil arbeitet und eine effiziente Codierung sicherstellt. Dies bestätigt auch die Kontrolle der internen Modellparameter.

Die Leistungsfähigkeit der Rate-Distortion-Optimierung ist abhängig von der Genauigkeit der geschätzten Modellparameter. Im Vergleich zu den Referenzverfahren kann FLEX durch die Parametrisierung der Rate-Distortion-Modellierung von (4.20) besser an die Statistik des aktuellen Bildsignals und der jeweiligen Datenrate angepasst werden. Allerdings ist es hierzu notwendig, den Modellparameter  $t_i$  für den jeweiligen Makroblock  $i$  bzw. für das aktuelle Bild prädiktiv richtig zu schätzen, um hohe Modellfehler zu vermeiden. Die Steigerung der Codiereffizienz im Vergleich zu den Referenzverfahren zeigt, dass die für (4.20) notwendigen Parameter unter realistischen Testbedingungen mit der notwendigen Genauigkeit prädiktiv geschätzt werden können.



Die Analyse der Arbeitsweise von FLEX zeigt, dass der Modellparameter  $t_i$  für den Makroblock  $i$  in (4.20) entgegen der Theorie in vielen Fällen für  $t_i = m_{DCT_i}$  eine höhere Codiereffizienz als  $t_i = 1$  liefert. Dies ist bedingt durch die nichtlinearen Effekte im Videocodec, wie der Eingrenzung des Wertebereichs des Quantisierungsparameters  $Q_i$  in Abhängigkeit der Werte von  $Q_{i-1}, Q_{i-2}, \dots$ , dem erhöhten Signalisierungsaufwand für einen Wechsel des Quantisierungsparameters von Makroblock zu Makroblock und den damit verbundenen Schätzfehlern. Die Berücksichtigung dieses Ergebnisses vereinfacht die Bestimmung der Parametrisierung in Abschnitt 4.3.4.

## 5.4 Untersuchung der Teilaspekte von FLEX

Neben der Bewertung der Codiereffizienz von FLEX dient das im Abschnitt 5.2 definierte Testset auch für die Bewertung und Auswahl unterschiedlicher Algorithmen für das neue Verfahren für die Datenratenkontrolle. In Kapitel 4 sind beim Entwurf des neuen Verfahrens folgende Optionen diskutiert worden:

1. Auswahl des eingesetzten Verfahrens für die Datenratenkontrolle auf Bild-Ebene. Dabei stehen ein einfaches Verfahren, wie im Abschnitt 4.3.3 beschrieben, und ein Rate-Distortion-optimiertes Modell basierend auf (4.12) zur Auswahl.
2. Vergleich unterschiedlicher Strategien für die Aktualisierung der Modellparameter: pro Bild, Makroblock-Zeile oder Makroblock.
3. Einsatzmöglichkeiten und Parametrisierung der Analyse der DCT-Koeffizienten für die Erhöhung der Codiereffizienz nach Abschnitt 4.4.1.

Analog zu dem Vorgehen im Abschnitt 5.3.1 basiert die Bewertung der unterschiedlichen Algorithmen auf dem direkten Vergleich der gemittelten  $PSNR$ -Werte des definierten Testsets mit dem Verfahren für die Datenratenkontrolle FLEX. Dabei sind die Datenrate, die Ausgangspuffergröße und die Framerate jeweils vorgegeben. Jedes Einzelergebnis repräsentiert dabei einen Messpunkt der dargestellten Kennlinien. Die oberen  $PSNR$ -Ergebnisse entsprechen den absoluten Werten der verschiedenen Einzeltests in dB und beziehen sich auf die linke y-Achse. Die unteren Kurvenverläufe beziehen sich auf die rechte y-Achse und stellen den Vergleich der Optionen jeweils zu FLEX dar. Sie entsprechen der  $PSNR$ -Differenz zwischen FLEX und der jeweiligen Optionen in dB. Bei positiven Werten weist dabei FLEX eine höhere Codiereffizienz als die jeweilige Vergleichsoption auf.

### 5.4.1 Datenratenkontrolle auf Bild-Ebene

Die Aufgabe der Datenratenkontrolle auf Bild-Ebene ist die Bestimmung der Zielbitmenge für die Codierung des aktuellen Bildes in Abhängigkeit des aktuellen Pufferstandes und der Datenrate. Dabei müssen zusätzlich die im jeweiligen Standard vorgegebenen Restriktionen berücksichtigt werden. Wesentliches Kriterium für die Bewertung ist die Codiereffizienz, wobei größere Schwankungen zwischen aufeinander folgenden Bildern vermieden werden sollen.

In Kapitel 4 sind für eine IPPP...-Codierung zwei Verfahren für die Datenratenkontrolle auf Bild-Ebene diskutiert worden. Das im Abschnitt 4.3.3 beschriebene Verfahren basiert auf (4.23) und berechnet die Zielbitmenge für das aktuelle Bild in Abhängigkeit von der Framerate und der Bitmenge, die pro Bild zur Verfügung steht. Zusätzlich wird der Pufferfüllstand berücksichtigt.

Als Vergleich dient ein Verfahren basierend auf dem Rate-Distortion-Modell in (4.12) bzw. auf (4.30). Analog zu der Modellierung von I- und B-Bildern im Abschnitt 4.4.3 basiert die Berechnung der Zielbitmenge  $\hat{B}_{DCT}^{F_{opt}}$  auf den Modell- und Schätzparameter  $t_j$  und  $m_{DCT_j}$  für den Makroblock  $j$  mit  $j \in [0; N-1]$  des aktuellen Bildes:

$$\hat{B}_{DCT}^{F_{opt}} = \frac{\sum_{j=0}^{N-1} t_j \cdot m_{DCT_j}}{\overline{tm}_P} \hat{B}_{DCT} \quad (5.1)$$

Dabei wird die mittlere Zielbitmenge  $\hat{B}_{DCT}^F$  für das aktuelle Bild nach (4.23) bestimmt. Die Berechnung des Vergleichsparameters  $\overline{tm}_P$  wird auf Basis der letzten  $M$  codierten Bilder (hier:  $M = 10$ ) durchgeführt:

$$\overline{tm}_P = \frac{1}{M} \sum_{l=0}^{M-1} \sum_{j=0}^{N-1} t_{lj} \cdot m_{DCT_{lj}} \quad (5.2)$$

Der Doppelindex bei  $m_{DCT_{nj}}$  adressiert dabei den jeweiligen Makroblock  $j$  im Bild  $n$  mit jeweils  $N$  Makroblöcken.

In Abbildung 5.5 sind die *PSNR*-Ergebnisse aus dem Testset von FLEX mit dem im Abschnitt 4.3.3 beschriebenen Verfahren für die Bitalloation auf Bildebene für unterschiedliche Ausgangspuffergrößen dargestellt. Als Vergleich dient eine Modifikation von FLEX, die im Unterschied dazu die beschriebene Rate-Distortion-Optimierung der Datenratenkontrolle auf Bild-Ebene einsetzt (Bezeichnung: OPT\_BILD).

Der direkte Vergleich der beiden Verfahren bei der IPPP...-Codierung zeigt, dass der Einsatz der Rate-Distortion-Optimierung für die Datenratenkontrolle auf Bild-Ebene in den meisten Fällen zu einer Verschlechterung der Codiereffizienz führt. Dabei ist das Ergebnis, wie in Abbildung 5.5 deutlich sichtbar, unabhängig von der Ausgangspuffergröße.

---

1. *opt* : Rate-Distortion-Optimierung.

Das Prinzip der Rate-Distortion-Optimierung bei der Datenratenkontrolle auf Bild-Ebene nutzt den Ausgangspuffer, um die zur Verfügung stehende Bitmenge in Abhängigkeit von der relativen Komplexität des jeweiligen Bildes aufzuteilen. Hierzu muss die Komplexität der zukünftigen Bilder prädiktiv geschätzt werden. Die Ergebnisse zeigen, dass sich diese Vorgehensweise trotz höherer algorithmischer Komplexität in vielen Fällen bei einer IPPP...-Codierung zu einer Verschlechterung der Codiereffizienz führt. Die prädiktive Schätzung ist hierzu nicht genau genug.

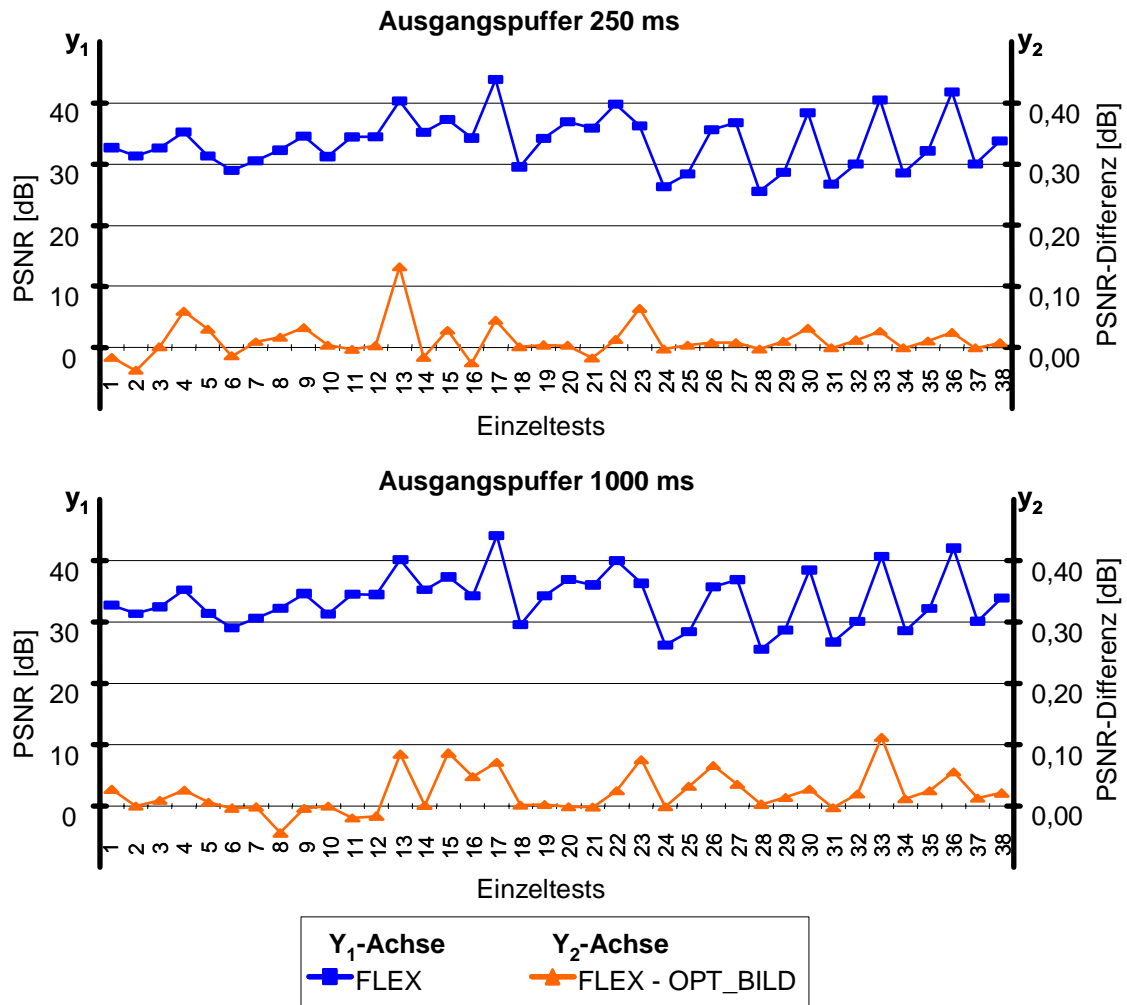


Abbildung 5.5: Qualitätsvergleich von unterschiedlichen Verfahren für die Bitalloation.

### 5.4.2 Aktualisierung der Modellparameter

In Kapitel 4 wird der Modellparameter  $c$  des neuen Verfahrens für die Datenratenkontrolle auf Basis von (4.27) pro Makroblock-Zeile neu berechnet. Die Aktualisierung pro Zeile (bei CIF nach jeweils 22 und bei QCIF nach 11 Makroblöcken) ist ein Kompromiss aus einer schnellen Anpassung an die aktuelle Signalstatistik und einer Reduzierung der notwendigen Rechenleistung.

Aufgrund der Arbeitsweise der Videocodierung bieten sich drei Möglichkeiten für ein Update des Modellparameters  $c$  an: pro Bild, Makroblock-Zeile oder Makroblock. Im Folgenden werden die drei Strategien für das Update des Modellparameters  $c$  hinsichtlich der Auswirkung auf die erzielbare Codiereffizienz untersucht.

Bei FLEX, wie im Abschnitt 5.2 definiert, wird  $c$  pro Makroblock-Zeile aktualisiert. Als Vergleich sind in Abbildung 5.6 die Ergebnisse aus dem Testset für modifizierte FLEX-Varianten mit einer Aktualisierung des Modellparameters  $c$  pro Makroblock (Bezeichnung: MB\_UPDATE) und pro Bild (Bezeichnung: BILD\_UPDATE) für unterschiedliche Ausgangspuffergrößen dargestellt.

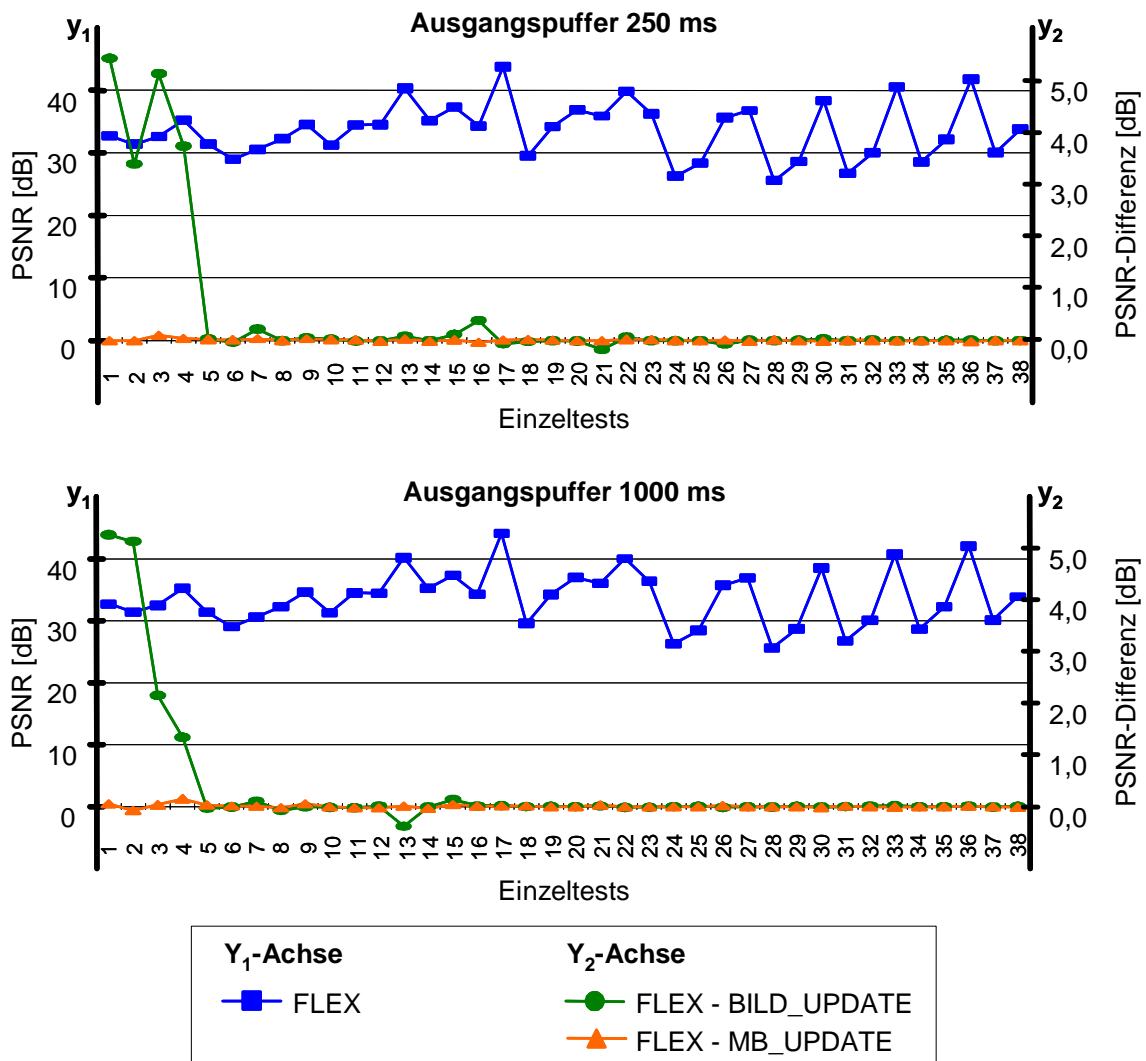


Abbildung 5.6: Qualitätsvergleich unterschiedlicher Strategien für die Parameteraktualisierung.

Die Ergebnisse zeigen, dass ein Update des Modellparameters  $c$  auf Bild-Ebene bei geringen Datenraten und niedrigen Frameraten unabhängig von der Ausgangspuffergröße zu einer Verschlechterung der Codiereffizienz gegenüber FLEX führt. Ein Update pro Makroblock führt im Vergleich zu FLEX zu keiner Erhöhung der Codiereffizienz. Daher wird für FLEX aufgrund der geringeren algorithmischen Komplexität eine Aktualisierung pro Makroblock-Zeile als Default-Einstellung gewählt. Die erzielbare Reaktionszeit in Abhängigkeit der Framerate ist in Abbildung 4.6 dargestellt. Ist aufgrund von stark schwankenden Ausgangsdatenraten eine schnellere Modellanpassung notwendig, kann jederzeit eine Aktualisierung pro Makroblock gewählt werden.

### 5.4.3 Analyse der DCT-Koeffizienten

In Abschnitt 4.4.1 ist ein Verfahren für die Analyse der DCT-Koeffizienten zur Modifikation des Quantisierungsparameters beschrieben. Das Ziel ist die Vermeidung von Makroblöcken, deren DCT-Koeffizienten alle zu null gesetzt werden. Dieses Verfahren wird in FLEX eingesetzt. Zusätzlich wird sichergestellt, dass die Quantisierungsparameter nicht zu häufig gewechselt werden. In Abbildung 5.7 sind die Ergebnisse des Testsets von FLEX und einer modifizierten Variante (Bezeichnung: NO\_DCT\_ANA), bei der das Verfahren für die Analyse der DCT-Koeffizienten deaktiviert ist, bei einer IPPP...-Codierung dargestellt.

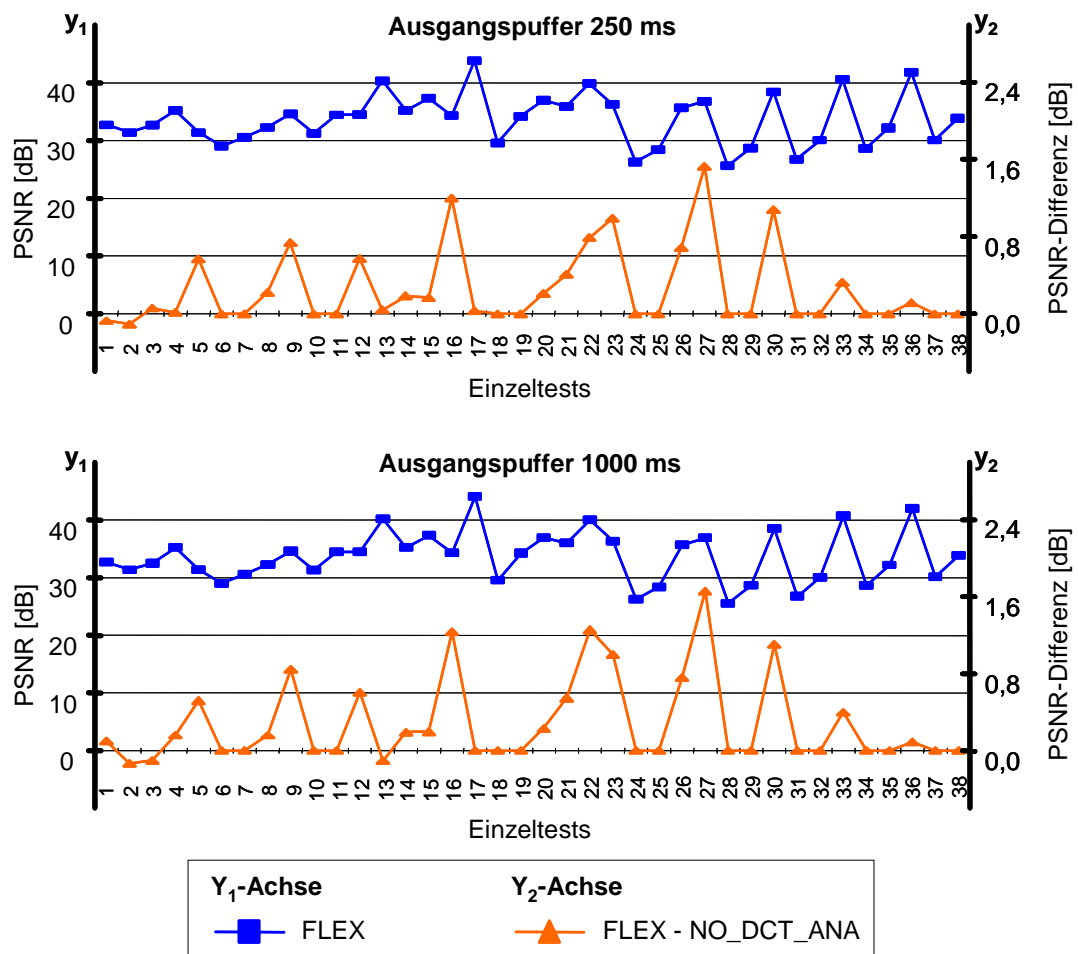


Abbildung 5.7: Qualitätsvergleich von FLEX mit und ohne Analyse der DCT-Koeffizienten.

Die Ergebnisse zeigen, dass das in Abschnitt 4.4.1 beschriebene Verfahren zur Analyse der DCT-Koeffizienten bei Sequenzen mit weitgehend festen Hintergrundbereichen die Codiereffizienz deutlich steigern kann. Bei den Sequenzen News, Weather, Children und Table Tennis wird bei einer IPPP...-Codierung eine Verbesserung des mittleren  $PSNR$ -Wertes von 0,6 und 1,6 dB erreicht. Allerdings erfordert die Modifikation der Quantisierungsparameter eine Ausgangspuffergröße von mindestens 250 ms, um die dadurch verursachten Schwankungen der Datenrate ausgleichen zu können. Über dieser Grenze ist der Gewinn der Codiereffizienz weitgehend unabhängig von der Ausgangspuffergröße.

## 5.5 Zusammenfassung

Das Ziel des neuen Verfahrens für die Datenratenkontrolle ist es, die Anforderungen aktueller Videocodecs, wie H.263 oder MPEG-4 zu erfüllen. Hierzu ist es notwendig, dass über den gesamten Anforderungsbereich eine stabile und effiziente Videocodierung im Bezug auf die Datenratenkontrolle sichergestellt wird.

Für die Verifikation des neuen Verfahrens dienen die ermittelten Anforderungen aus der Standardisierung aktueller Videocodecs für Kommunikationsanwendungen: die Unterstützung eines breiten Ausgangsdatenbereichs von 10 kBit/s bis zu 2 MBit/s, von Bildmaterial unterschiedlicher Komplexität und Ausgangspuffergrößen von 100 ms bis zu 2000 ms<sup>1</sup>.

Diese Anforderungen bilden die Grundlage für die Definition der Testbedingungen, anhand derer die Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle im Vergleich zu den Referenzverfahren aus der Standardisierung von H.263 und MPEG-4 ermittelt wird. Die zwölf verwendeten Testsequenzen unterscheiden sich dabei vor allem in den Kriterien Detailreichtum und dem Anteil schneller Bewegungen. Zusätzlich sind in einigen Sequenzen Kamerazooms bzw. -schwenks und eine Mischung aus natürlichem und synthetischem Bildmaterial enthalten.

Die Verifikation des neuen Verfahrens für die Datenratenkontrolle auf Basis dieser Anforderungen zeigt, dass es die geforderte Flexibilität aufweist, um die Anforderungen aktueller Videocodecs, wie H.263 oder MPEG-4, zu erfüllen. Aufgrund der Ergebnisse und der Untersuchung der Rate-Distortion-Modellierung zeigt sich, dass das neue Verfahren unter den geforderten Bedingungen stabil arbeitet und eine effiziente Codierung sicherstellt.

Im Vergleich zu den Referenzverfahren aus der H.263- und MPEG-4-Standardisierung kann das neue Verfahren dabei in vielen Fällen eine Steigerung der Codiereffizienz erreichen. Der  $PSNR$ -Gewinn pro codierte Sequenz ist dabei abhängig von den jeweiligen Randbedingungen und dem jeweiligen Vergleichsverfahren und beträgt bis zu 2 dB und mehr. Als mittlerer Gewinn über 38 einzelne Tests mit unterschiedlichen Randbedingungen ergibt sich in etwa 0,7 dB. In Einzelfällen ergibt sich auch eine Verschlechterung des mittleren  $PSNR$ -Werts von bis zu 0,4 dB.

---

1. Die Untersuchung und Bewertung bezüglich der Unterstützung von variablen Ausgangsdatenraten befindet sich im Abschnitt 4.4.2.

Insgesamt zeigt sich, dass für die Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle eine Kombination verschiedener Faktoren notwendig ist. Den Kern des Verfahrens stellt die parametrisierte Rate-Distortion-Modellierung aus Kapitel 4 dar. Der Einsatz des neuen Schätzparameters auf Basis der DCT-Koeffizienten - anstelle der Varianz - bildet in Verbindung mit der Modellparametrisierung die Grundlage für eine verbesserte Rate-Distortion-Modellierung.

Die Untersuchungen zeigen, dass es ausreicht, den Modellparameter  $c$  basierend auf (4.27) zur Anpassung an die aktuelle Bildsignalstatistik pro Makroblock-Zeile zu berechnen. Für die Bestimmung der Parameter ist es ausreichend, auf die Ergebnisse eines codierten Bildes zurückzugreifen. Bei der Datenratenkontrolle auf Bild-Ebene kann durch den Einsatz einer Rate-Distortion-Modellierung keine Steigerung der Codiereffizienz erreicht werden. Es zeigt sich, dass die Modellfehler aufgrund der prädiktiven Schätzparameter hierfür zu groß sind. Allerdings kann durch die zusätzliche Analyse der DCT-Koeffizienten zur Modifikation des Quantisierungsparameters eine deutliche Steigerung der Codiereffizienz erreicht werden. Insbesondere bei Sequenzen mit festem Hintergrund ist eine Steigerung des  $PSNR$ -Werts von bis zu 1,6 dB möglich.





## 6 Realisierungsuntersuchungen

In diesem Kapitel wird auf die Realisierungsaspekte von echtzeitfähigen, softwarebasierten Videocodecs auf Standardprozessoren eingegangen. In diesem Zusammenhang werden im Abschnitt 6.2 die Grundlagen von Multimedia-Prozessorerweiterungen dargestellt. Im Abschnitt 6.3 werden die Einsatzmöglichkeiten und die sich daraus ergebenden Beschleunigungsmöglichkeiten im Videocodec untersucht. Neben der erzielbaren Codiereffizienz spielt bei Kommunikationsanwendungen die Verzögerung eine wichtige Rolle. Aus diesem Grund wird im Abschnitt 6.4 der Videocodec in Bezug auf dessen Ende-zu-Ende-Verzögerung analysiert. Zudem werden Ansätze zur Minimierung der Verzögerung untersucht und bewertet. Die wesentlichen Inhalte dieses Kapitels sind im Abschnitt 6.5 zusammengefasst.

### 6.1 Einführung

Das in dieser Arbeit entwickelte neue Verfahren für die Datenratenkontrolle zeichnet sich insbesondere durch seine Einsatzmöglichkeiten und seine Videocodiereffizienz in Kommunikationsanwendungen aus. Die hierfür notwendige Videocodierung in Echtzeit stellt hohe Anforderungen an die Realisierung [2], [20]. Es müssen unterschiedliche, sich teilweise widersprechende Kriterien berücksichtigt werden, die als mehrdimensionales Optimierungsproblem angesehen werden können:

1. die algorithmische Komplexität des Codierverfahrens,
2. der Speicherbedarf und die notwendige Speichertransferleistung,
3. die notwendige Geschwindigkeit der Videocodierung,
4. die prinzipiell vom Verfahren vorgegebene und die zusätzliche durch die Implementierung verursachte Verzögerung,
5. die erzielbare (Bild-)Qualität und
6. die hierfür notwendige Bitrate.

In vielen Fällen steht für die Realisierung nur eine begrenzte Rechenleistung zur Verfügung, sodass die Komplexität der Encoderimplementierung - häufig auf Kosten der Codiereffizienz - reduziert werden muss. Dabei können unterschiedliche Lösungsansätze gewählt werden, so z.B. die Beschränkung der Bildgröße bzw. der Framerate oder der Einsatz von vereinfachten Methoden für die Bewegungsschätzung. Bei der Auswahl der Verfahren für die Realisierung des Codecs wird daher oftmals der Einfluss auf das Codierergebnis im Verhältnis zu der Komplexität gesetzt, um die beste Konstellation auszuwählen.

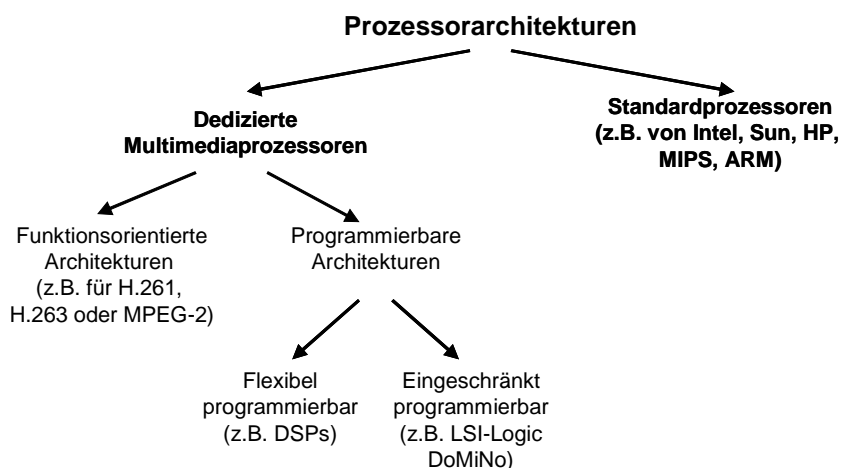
Unter diesem Aspekt wird das neu entwickelte Verfahren für die Datenratenkontrolle im Gesamtkontext des Videoencoders anhand eines echtzeitfähigen H.263-Softwarecodecs untersucht. Zudem werden Ansätze für eine optimierte Implementierung aufgezeigt.

Bei Kommunikationsanwendungen ist neben der erzielbaren Codiereffizienz eine geringe Verzögerung ein wichtiges Qualitätskriterium. Die vom Codec verursachte Verzögerung ist abhängig vom Algorithmus, der zur Verfügung stehenden Rechenleistung und von der Implementierung.

Auf die erzielbare Bildqualität und die hierzu notwendige Bitrate ist im Zusammenhang mit der Verifikation des neuen Verfahrens für die Datenratenkontrolle bereits ausführlich im Kapitel 5 eingegangen worden. Im Folgenden liegt der Schwerpunkt der Untersuchung damit auf den ersten vier Aspekten der obigen Aufstellung.

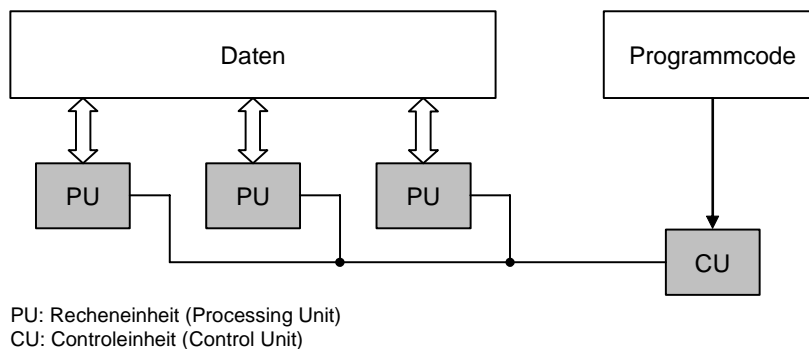
## 6.2 Multimedia-Prozessorerweiterungen

Bis vor kurzem waren dedizierte Multimediaprozessoren für eine qualitativ hochwertige Videokommunikation unter Echtzeitbedingungen notwendig. Wie in Abbildung 6.1 dargestellt, können die Prozessoren hinsichtlich ihrer Flexibilität unterschieden werden: Die funktionsorientierten Architekturen, z.B. für H.261 oder MPEG-2, sind speziell für den jeweiligen Anwendungsfall entwickelt und können meist nur im geringen Umfang innerhalb ihres Anwendungsbereiches konfiguriert werden. Dagegen können bei programmierbaren Architekturen - zumindest eingeschränkt - z.B. neue Standard-Erweiterungen oder auch Fehler bei der Codecimplementierung durch eine erweiterte Programmierbarkeit nachträglich berücksichtigt werden [101], [133].



**Abbildung 6.1:** Prozessorarchitekturen für die Multimedia-Signalverarbeitung.

Die zweite Klasse neben den dedizierten Multimediaprozessoren sind die Standardprozessoren, wie sie z.B. in PCs eingesetzt werden. In vielen Fällen erreichen inzwischen Vertreter dieser Klasse die notwendige Rechenleistung, um Multimediakommunikation in Echtzeit zu ermöglichen. Neben der Steigerung der Taktfrequenzen ist dies vor allem auf spezielle Multimedia-Architekturserweiterungen zurückzuführen, die von allen wichtigen Herstellern, wie z.B. Intel, AMD, ARM und Motorola, in ihre aktuellen Prozessoren integriert werden, so z.B. [3], [4], [7], [43], [45]-[47], [86] oder [94]. Das Funktionsprinzip der so genannten "Single Instruction stream, Multiple Data stream (SIMD)"-Architekturen ist in Abbildung 6.2 dargestellt.



**Abbildung 6.2:** Vereinfachtes Prinzip einer SIMD-Verarbeitungseinheit.

Durch das parallele Ausführen des jeweiligen Befehls auf mehreren Recheneinheiten können voneinander unabhängige Daten sehr effektiv bearbeitet werden. Der herstellerabhängige Befehlssatz ist dabei an die speziellen Belange der Audio- und Bildsignalverarbeitung angepasst. Meist werden auch weitere Anwendungsbereiche der digitalen Signalverarbeitung, wie Verschlüsselung, Spracherkennung und 3D-Grafikverarbeitung unterstützt. Bei der für den PC-Bereich sehr wichtigen IA-32-Architekturen (Intel Architecture, IA) der Firmen Intel und AMD stellen inzwischen die Streaming SIMD Extensions 2 (SSE2), die Intel mit dem Pentium 4 einführte, die gemeinsame Basis dar (siehe Anhang C).

Der mögliche Parallelisierungsgrad kann nur als erster Anhaltspunkt für die Abschätzung der potenziellen Beschleunigung dienen. Die tatsächliche Leistungssteigerung ist von vielen Einzelfaktoren abhängig, so z.B. der Zeit für den Speicher- bzw. Cachezugriff, der Busbandbreite, der IO-Leistungsfähigkeit und auch von möglichen Ressourcenkonflikten. In vielen Fällen können die Leistungsdaten nur durch mehrere Programmläufe bestimmt werden. Allein durch die schnelle Anbindung des Caches mit bis zu 100 GByte/s im Vergleich zu einer maximalen Speicherbandbreite von 6,4 GByte/s, kann ein im Cache ablaufender Algorithmus um den Faktor 16 schneller sein.

Aktuell ist die Compilerunterstützung für die automatische Parallelisierung von C-Code meist ungenügend, sodass rechenintensive Algorithmen direkt in SIMD-Assembler programmiert<sup>1</sup> werden müssen. Daher ist es notwendig, neben der möglichst guten Ausnutzung des Caches, auch den Aufbau der Ziel-Architektur zu kennen, um den Code darauf anzupassen. So kann z.B. durch eine Veränderung der Befehlsreihenfolge der Durchsatz deutlich erhöht werden. Werden z.B. die zwei Befehle `PADDW` bzw. `PMULHUW` (siehe Tabelle C.1) jeweils mehrfach hintereinander in einem Algorithmus benötigt, so ist der Durchsatz<sup>2</sup> jeweils zwei Taktzyklen. Eine abwechselnde Gruppierung der Befehle verringert dagegen den effektiven Durchsatz auf einen Taktzyklus. Die beiden Befehle benutzen zwar den gleichen Port, aber unterschiedliche Verarbeitungseinheiten [41], [46].

Eine ergänzende Beschreibung von SIMD-Prozessorerweiterungen befindet sich im Anhang C. Weitere Informationen zur SIMD-Programmierung ist z.B. in [39], [42]-[47] und [53] zu finden.

1. Der C++-Compiler von Intel unterstützt Ansätze für die automatische Parallelisierung von C-Code.
2. Die Verzögerung entspricht der Anzahl der Taktzyklen, die der Prozessor benötigt bis alle  $\mu$ -Operationen des Befehls komplett abgearbeitet sind. Dagegen entspricht der Durchsatz der Anzahl der Taktzyklen, die gewartet werden muss bevor die gleiche Verarbeitungseinheit erneut benutzt werden kann.

### 6.3 Einsatz von SIMD-Befehlserweiterungen in Videocodecs

Die Multimedia-Einheiten der Prozessoren erlauben durch das parallele Ausführen des jeweiligen Befehls auf mehrere Einzeldaten, dass voneinander unabhängige Daten sehr effektiv bearbeitet werden können. In Tabelle 6.1 sind die notwendigen mathematischen Basisoperationen der einzelnen Funktionsblöcke in einem Videocodec aufgeführt (siehe Abschnitt 2.2.3 bzw. Abbildung 2.1). Zusätzlich wird deren Komplexität und die Eignung für eine parallele Verarbeitung bewertet. Dabei stehen  $x_i$ ,  $x_j$  bzw.  $y_i$  für einzelne Pixel der Bilder  $x$  und  $y$ . Der Gewichtungsfaktor  $c$  ist ohne Index konstant bzw. mit Index  $i$  bzw.  $j$  konstant für ein Pixel  $x_{i/j}$  oder  $y_{i/j}$ . Die Variablen  $a$  und  $b$  stehen für beliebige Zahlenwerte.

Funktion	Operationen	Komplexität	Parallelität
<b>Bewegungsschätzung</b>	$\sum  x_i - y_i $ oder $\sum (x_i - y_i)^2$ , $\min(a, b)$	Abhängig von der Implementierung, oftmals konstant für jedes Pixel	Abhängig von der Implementierung, meist jedoch hoch parallel und zusätzlich noch hoher Daten- und Befehlsdurchsatz notwendig
<b>Bewegungskompensation</b>	$x_i + cx_j$ , kopieren von Datenblöcken, Pixel-Interpolation	Weitgehend konstant für jedes Pixel	Hoch parallel
<b>DCT und IDCT</b>	$\sum c_i x_i$ , $c \cdot x_i + b$	Implementierungsabhängig: entweder konstant oder eine Funktion von der mittleren Anzahl der DCT-Koeffizienten ungleich null	Abhängig von der Implementierung
<b>Quantisierung</b>	$x_i / c_i$	Konstant für jedes Pixel	Hoch parallel
<b>Dequantisierung</b>	$c_i \cdot x_i$	Abhängig von der mittleren Anzahl der DCT-Koeffizienten ungleich null	Hoch parallel
<b>VLC</b>	Schiebeoperationen und Vergleiche	Abhängig von der mittleren Anzahl der Symbole im Bitstrom	Vollständig sequenziell
<b>Sonstiges (z.B. Konvertieren von Farbräumen, Vor- und Nachbearbeitung)</b>	$\sum c_i x_i$ , Clipping, $(x_i + x_j) / 2$ , $\frac{1}{c} \sum x_i$ , (um)kopieren von Datenblöcken	Konstant für jedes Pixel	Hoch parallel

**Tabelle 6.1:** Mathematische Basisoperationen eines Videocodecs [20].

Mit Ausnahme der VLC eignen sich damit prinzipiell alle wichtigen Funktionseinheiten im Videocodec für eine Beschleunigung durch die parallele Bearbeitung in den SIMD-Prozessoreinheiten. Allerdings müssen die eingesetzten Algorithmen hinsichtlich ihrer Parallelisierungsmöglichkeiten ausgewählt werden. Zudem ist es für die Videoencodierung notwendig, dass der Prozessor eine hohe IO-Leistungsfähigkeit (Input Output, IO) und eine flexible Adressierung unterstützt.

Ein wichtiger Aspekt bei der Beschleunigung einzelner Funktionen ist, den Implementierungsaufwand mit dem konkreten Nutzen für die Anwendung abzuwägen. Dieser Punkt ist insbesondere dann von Bedeutung, wenn Verfahren eingesetzt werden, die sich negativ auf die Codiereffizienz auswirken. Für die Beurteilung ist das Gesetz von Amdahl (Amdahl's Law [30]) ein wesentliches Kriterium. Dieses Gesetz besagt, dass die Beschleunigung einer Teilkomponente nur maximal mit dem Faktor zur Verbesserung der Systemleistung beiträgt, der indirekt proportional zum Anteil der übrigen Komponenten ist. Die Beschleunigung unwichtiger Teile, auch um hohe Faktoren, trägt nur gering zur Gesamtbeschleunigung bei. Nur die Verbesserung wichtiger Komponenten trägt spürbar zur Leistungssteigerung bei, beeinflusst aber direkt den Anteil der übrigen Teile.

### 6.3.1 Beschleunigung wichtiger Funktionen im Videocodec

Vor diesem Hintergrund sind auf Basis des in Kapitel 5 beschriebenen H.263-Encoders<sup>1</sup> verschiedene Sequenzen (Mobile&Calendar, Foreman und Stephan) hinsichtlich der notwendigen Rechenzeit für die einzelnen Funktionsmodule untersucht worden. Jede Sequenz ist im Format CIF mit 30 fps jeweils mit 512 bzw. 2048 kBit/s ohne zusätzliche Optionen mit einer IPPP-Bildfolge auf einem PC<sup>2</sup> codiert worden. Für die Encodierung der 300 Bilder der jeweiligen CIF-Sequenzen mit 512 kBit/s und 2048 kBit/s werden durchschnittlich 88,24 s bzw. 88,6 s benötigt. Der geringe Unterschied zwischen 512 kBit/s und 2048 kBit/s entsteht vor allem durch die höhere Anzahl der DCT-Koeffizienten ungleich null, die jeweils einzeln im VLC-Modul bearbeitet werden müssen.

Insgesamt zeigt sich, dass die Bewegungsschätzung mit ca. 95 % die meiste Rechenzeit von allen Funktionsmodulen benötigt. Der Rest der Rechenzeit teilt sich in etwa gleichmäßig auf die sonstigen Module auf. Für die Beschleunigung des Encoders eignen sich mit absteigender Priorität folgende Module für eine optimierte SIMD-Umsetzung (siehe auch Tabelle 6.1):

- Bewegungsschätzung (ME) 1- und 1/2-Pixel genau,
- DCT,
- Bewegungskompensation (MC),
- Quantisierung/Dequantisierung,
- Datenratenkontrolle und
- IDCT.

Auf die jeweiligen Implementierungsaspekte der einzelnen Module wird im Folgenden eingegangen. Im Anschluss daran werden die Ergebnisse im Zusammenhang ausgewertet.

- 
1. Das Hauptziel für das Software-Design des untersuchten H.263-Codecs (ohne die im Folgenden beschriebenen SIMD-Module, die auf Geschwindigkeit optimiert sind) ist ein modularer Aufbau, um eine einfache Erweiterung sicherzustellen. In Abhängigkeit des verwendeten Compilers könnte daher ein optimiertes Design nach Schätzung des Autors eine um 25 - 35 % höhere Performance erreichen.
  2. Als Basis für die Untersuchungen hat folgende Hardware-Ausstattung gedient: Pentium 4 mit 3 GHz, 875P-Chipsatz, 1024 MByte DDR400 Speicher, Windows XP und einer RamDisk, um das Ergebnis nicht durch die Leistungsfähigkeit der Festplatten zu verfälschen. Als Entwicklungsumgebung hat Microsoft Visual C++ 6.0 ServicePack 5 und dem für die SIMD-Unterstützung notwendigen Prozessorpack gedient.

### 6.3.1.1 Bewegungsschätzung und -kompensation

Bei H.263 wird bei der Bewegungsschätzung für jeden Makroblock  $i$  des aktuellen Bildes der  $16 \times 16$  große Bereich im rekonstruierten Vorgängerbild gesucht, der den geringsten  $SAD$ -Wert aufweist. Dabei liegt der maximale Wertebereich der  $x$ - und  $y$ -Komponenten der Bewegungsvektoren jeweils zwischen  $-16$  und  $15,5$  relativ zur Position des jeweiligen Makroblocks  $i$  [73].

Die Schätzung kann auf unterschiedliche Weise durchgeführt werden, so können z.B. alle möglichen Positionen abgesucht werden (Fullsearch) oder es können die vielversprechendsten ausgewählt werden. Oftmals wird bei Fullsearch mit 1-Pixel Genauigkeit der Luminanzanteil des Makroblocks  $i$  mit dem korrespondierenden Startbereich im rekonstruierten Vorgängerbild Pixel für Pixel verglichen und die jeweilige absolute Differenz der einzelnen Pixel aufsummiert. Anschließend wird der Vergleichsbereich im Vorgängerbild um einen Pixel verschoben und erneut mit dem Makroblock  $i$  verglichen. Dies wird solange wiederholt, bis der gesamte Suchbereich bearbeitet ist. Die beste Übereinstimmung liefert dann der Bereich, der den kleinsten  $SAD$ -Wert aufweist.

Abhängig von der Implementierung kann die Bewegungsschätzung für den Makroblock  $i$  mit höherer Genauigkeit im ersten Durchgang bei allen möglichen Positionen berechnet werden. Meist wird aber aufgrund der notwendigen Rechenleistung ein zweistufiges Verfahren eingesetzt. Im ersten Schritt wird mit einer Genauigkeit von einem Pixel der Bereich mit der besten Übereinstimmung gesucht und entschieden, ob eine genauere Schätzung sinnvoll erscheint oder ob der Makroblock intracodiert werden soll. Für eine Inter-Codierung wird dann im zweiten Schritt in einem reduzierten Suchbereich (z.B. im Bereich von  $1/2$  oder  $1$  Pixel) eine Schätzung mit höherer Genauigkeit durchgeführt.

Für die SIMD-Umsetzung der Bewegungsschätzung mit 1-Pixel Genauigkeit bietet der Pentium 4 den Befehl PSADBW, der die Summe der absoluten Differenz zwischen einzelnen Pixeln (vorzeichenlose 8 Bit breite Daten) berechnet. In Abhängigkeit der Registerbreite können dabei entweder 8 oder 16 Pixel auf einmal berechnet werden. Somit kann eine Fullsearch mit 1-Pixel Genauigkeit für einen Makroblock  $i$  direkt in zwei ineinander geschachtelten Schleifen für die horizontale bzw. vertikale Bearbeitung umgesetzt werden. Der jeweils kleinste  $SAD$ -Wert wird zusammen mit den jeweiligen Koordinaten des Bewegungsvektors abgespeichert [40], [51].

Die Bewegungsschätzung mit  $1/2$ -Pixel Genauigkeit kann nach einem ähnlichen Prinzip erfolgen. Allerdings müssen hier vier Fälle bei der Schätzung im rekonstruierten Vorgängerbild unterschieden werden. Je nach Lage des Vergleichsbereichs ist eine Berechnung von Halbpixel-Werten durch Interpolation von Bildpixeln notwendig:

1. Horizontal liegt der Vergleichsbereich zwischen zwei Pixeln, sodass ein Zwischenwert berechnet werden muss. In vertikaler Richtung ist keine Interpolation notwendig.
2. Vertikal ist eine Interpolation notwendig, horizontal ist kein Zwischenwert notwendig.
3. Für die Ermittlung des Vergleichsbereichs müssen jeweils vier Pixel interpoliert werden.
4. Es ist keine Interpolation notwendig, da der Vergleichsbereich direkt den Pixeln des rekonstruierten Vorgängerbildes entspricht. Die  $SAD$ -Werte entsprechen damit den entsprechenden Werten der Bewegungsschätzung mit 1-Pixel Genauigkeit.

Die Interpolation von zwei Pixeln wird durch den SIMD-Befehl PAVGB unterstützt. In Abhängigkeit der Breite der beiden angegebenen Register werden entweder 8 oder 16 Pixel jeweils mit 8 Bit Auflösung parallel bearbeitet. Hierzu werden intern die jeweiligen Pixelwerte der Register aufsummiert, das Ergebnis für die Rundung um eins erhöht, durch zwei dividiert und zum Schluss im zuerst angegebenen Register abgespeichert.

Eine Interpolation von vier Pixeln kann über einen Zwischenschritt realisiert werden. Hierzu werden jeweils zwei Pixel der ersten und der zweiten Bildzeile interpoliert. Im Anschluss wird durch Interpolation der Zwischenergebnisse das Endergebnis berechnet. Allerdings tritt hier aufgrund von Rundungsfehlern in 37 % der Pixel ein Fehler von +1 auf. Durch Subtraktion von 1 pro Pixel bei einem der beiden Zwischenergebnisse kann der Fehler auf -1 in 17 % der Pixel reduziert werden. Mit einer aufwendigeren Berechnung der jeweils notwendigen Subtraktionswerte pro Pixel kann der Fehler auch vollständig kompensiert werden [48].

Die Bewegungskompensation kann nach dem gleichen Prinzip, wie die Bewegungsschätzung mit  $1/2$ -Pixel Genauigkeit realisiert werden. Allerdings müssen hierfür auch die Chrominanz-Blöcke berücksichtigt werden.

Im H.263-Encoder sind für die Bewegungsschätzung folgende Module mittels SIMD-Befehle parallelisiert worden: Fullsearch im Bereich  $[-16;15, 5]$  Pixel mit einer Genauigkeit von 1-Pixel und eine Schätzung mit  $1/2$ -Pixel Genauigkeit im Bereich  $\pm 0,5$  Pixel um den Ergebnisvektor der Fullsearch. Darüber hinaus ist die Varianz des Originalmakroblocks auf Basis einer Ganzzahlarithmetik abgeschätzt worden, um das Ergebnis der Bewegungsschätzung für den weiteren Ablauf zu beurteilen.

Für die Bewegungskompensation ist zum einen die Addition zweier Bilder implementiert worden, die für die Berechnung des rekonstruierten Bildes notwendig ist. Dabei werden gleichzeitig die Pixelwerte des Ergebnisbildes auf den Wertebereich  $[0;255]$  begrenzt. Zum anderen sind die Subtraktion des Prädiktionsbildes vom Originalbild, die für die Berechnung des Fehlerbildes notwendig ist, parallelisiert worden.

### 6.3.1.2 Transformationscodierung

In der Videocodierung wird meist eine  $8 \times 8$  2D-DCT-Transformation eingesetzt. Hierfür sind viele Algorithmen veröffentlicht worden, die eine effiziente Berechnung ermöglichen. Einige basieren davon auf einer 1D-DCT, andere nützen die 2D-Struktur der DCT für eine direkte Umsetzung. Andere kombinieren eine effiziente 1D-DCT Berechnung mit einer anschließenden 2D Skalierung [39], [49], [50].

Allen gemeinsam ist, dass sie für eine optimierte Berechnung die Anzahl der notwendigen arithmetischen Operationen reduzieren, insbesondere die Anzahl der Multiplikationen. So haben z.B. Feig und Winograd einen der schnellsten  $8 \times 8$  2D-DCT-Algorithmen entworfen. Der Loeffler-Ligtenberg-Moschytz (LLM) und der Arai-Agui-Nakajima (AAN) Algorithmus sind sehr schnelle Algorithmen für die Berechnung der 1D-DCT [39], [49], [50].

Alle drei Algorithmen können auch für die Berechnung der IDCT modifiziert werden. In Tabelle 6.2 sind die notwendigen arithmetischen Operationen der verschiedenen Algorithmen zusammengefasst.

Algorithmus	Anzahl der Operationen für 1D-DCT mit 8 Elementen		Anzahl der Operationen für 8 x 8 2D-DCT	
	Additionen	Multiplikationen	Additionen	Multiplikationen
<b>Feig &amp; Winograd</b>	- (direkter 2D Algorithmus)		454	94
<b>LLM</b>	28	11	448	176
<b>AAN</b>	29	5	464	144

**Tabelle 6.2:** Charakteristik von ausgewählten schnellen DCT-Algorithmen [39].

Für die Auswahl des Algorithmus, der auf einer SIMD-Recheneinheit eines Pentium 4 implementiert werden soll, sind aufgrund der schnellen Multiplikationsbefehle die zusätzlichen Kosten für die Multiplikationen gering. Wichtiger ist es, irreguläre Speicherzugriffsmuster zu vermeiden.

Aus diesem Grund scheidet der Algorithmus von Feig aus. Untersuchungen in [39] und [49] zeigen, dass die LLM- gegenüber der AAN-Implementierung einen geringeren Fehler im Vergleich zur IEEE-Standard-Spezifikation der DCT aufweist. Daher ist der LLM-Algorithmus auf Basis der Beschreibung von [39] und [49] mit 32-Bit-Elementen für die Implementierung der DCT und der IDCT in den H.263 ausgewählt worden.

### 6.3.1.3 Quantisierung und Dequantisierung

Nach der Transformationscodierung werden die Koeffizienten der Blöcke bzw. Makroblöcke verlustbehaftet quantisiert und mittels einer VLC im Bitstrom zum Decoder übertragen. Für die Bewegungskompensation ist im Encoder auch eine Dequantisierung der Koeffizienten notwendig, um das codierte Bild rekonstruieren zu können (siehe Kapitel 2). Das Prinzip der Quantisierung und der Dequantisierung wird in [73] und [96] beschrieben.

Für die Quantisierung auf Basis des SIMD-Prinzips sind folgende Bearbeitungsschritte und die Bitmasken *sign\_mask* und *reciprodiv\_cor\_mask* aus Tabelle 6.3 notwendig:

1. Initialisierung und Berechnung der notwendigen Bitmasken in Abhängigkeit des Quantisierungsparameters (siehe Tabelle 6.3).
2. Laden der Koeffizienten in ein SIMD-Register (z.B. 4 Koeffizienten mit je 16 Bit Breite in ein 64 Bit breites SIMD-Register), z.B. mit den Befehlen `MOVDQA` oder `MOVDQU` in Abhängigkeit von der Lage im Speicher (siehe Tabelle C.1).
3. Berechnung des Vorzeichens und der Absolutwerte pro Koeffizient mittels Schiebebefehle, Oder-Verknüpfung durch *sign\_mask* und Multiplikation mit dem Vorzeichen.
4. Durchführung der Quantisierung mit Ersetzen der Division durch die Multiplikation des Kehrwertes und der Subtraktion eines Korrekturwertes *reciprodiv\_cor\_mask*.
5. Berechnung und Speicherung des vorzeichenbehafteten, quantisierten Koeffizienten  $cqp_{uv}$ .
6. Weiter mit Punkt 2 bis alle Koeffizienten des Makroblocks bearbeitet sind.



Zusätzlich wird durch eine Oder-Verknüpfung über alle Koeffizienten überprüft, ob mindestens ein quantisierter Koeffizient ungleich null ist.

Für die Dequantisierung mit dem Quantisierungsparameter  $QP$  sind folgende Berechnungen für alle quantisierten Koeffizienten  $cqp_{uv}$  ungleich null notwendig:

$$cdqp_{uv} = \begin{cases} QP \times (2cqp_{uv} + 1) - !(QP \& 1) & \text{für } cqp_{uv} > 0 \\ -QP \times (2(-cqp_{uv}) + 1) + !(QP \& 1) & \text{für } cqp_{uv} < 0 \end{cases} \quad (6.1)$$

Für alle Werte  $cdqp_{uv}$  größer 4047 wird  $cdqp_{uv} = 4047$  und für alle Werte kleiner -4048 wird  $cdqp_{uv} = -4048$  gesetzt.

Die SIMD-Umsetzung von (6.1) basiert im Kern auf folgenden Operationen, wobei die Masken  $qp\_mul\_mask$ ,  $s\_add\_mask$  und  $s\_sub\_mask$  aus Tabelle 6.3 benötigt werden:

1. Initialisierung und Berechnung der notwendigen Bitmasken (siehe Tabelle 6.3).
2. Laden der Koeffizienten in ein SIMD-Register (z.B. 4 Koeffizienten mit je 16 Bit Breite in ein 64 Bit breites SIMD-Register).
3. Multiplikation des quantisierten Absolutbetrags des Koeffizienten mit  $qp\_mul\_mask$  und Speicherung der rechten 16 Bit des 32 Bit breiten Ergebnisses.
4. Vorzeichenbehaltete 16 Bit Addition mit Saturation des obigen Ergebnisses mit der Maske  $s\_add\_mask$  und vorzeichenlose 16 Bit Subtraktion mit Saturation des Additionsergebnisses mit der Maske  $s\_sub\_mask$ .
5. Multiplikation des dequantisierten Absolutbetrags mit dem Vorzeichen des ursprünglichen Koeffizienten  $c_{uv}$  und Abspeicherung des Dequantisierungsergebnisses  $cdqp_{uv}$ .
6. Weiter mit Punkt 2 bis alle Koeffizienten des Makroblockes bearbeitet sind.

Bei der Berechnung der quantisierten Koeffizienten  $cqp_{uv}$  ist es sinnvoll, die für die Bewegungskompensation notwendige Dequantisierung mit zu berechnen. So können Zwischenergebnisse und die Berechnung von notwendigen Bitmasken für beide Fälle verwendet werden. Die Koeffizienten  $c_{00}$  von Intra-Blöcken werden nach der Beschreibung in [73] gesondert ohne Einsatz von SIMD-Befehlen berechnet.

---

1.  $!(...)$ : Nicht-Operator mit  $!(a = 0) = 1$  und  $!(a \neq 0) = 0$ .  
 $a \& b$ : Und-Operator (bitweise).

Maskenname	Bitmaske für ein vorzeichenloses Word (16 Bit)
<i>qp_mul_mask</i> :	$QP \ll 1$
<i>s_sub_mask</i> :	$0xF800^a$
<i>s_add_mask</i> :	$0xF800 + \begin{cases} QP & \text{für } (QP \& 1) = 1 \\ QP - 1 & \text{sonst} \end{cases}$
<i>reciprodiv_cor_mask</i> :	$0x0000$ für Intra-Blöcke $QP \gg 1$ sonst
<i>sign_mask</i> :	$0x0001$
<i>reciprodiv_mask</i> <sup>b</sup> :	$reciprodiv[QP] = \{0, 32767, 16384, 10923, 8192, 6554, 5462, 4682, 4096, 3641, 3277, 2979, 2731, 2521, 2341, 2185, 2048, 1928, 1821, 1725, 1639, 1561, 1490, 1425, 1366, 1311, 1261, 1214, 1171, 1130, 1093, 1058\}$
a. 0x: Kennzeichnet Hexadezimaldarstellung von Zahlen. b. $QP = 1$ kann auf diese Weise nicht berechnet werden. Hierzu ist eine Schiebeoperation ( $\gg 1$ ) notwendig.	

**Tabelle 6.3:** Initialisierung und Berechnung der Bitmasken für die Quantisierung und Dequantisierung.

### 6.3.1.4 Datenratenkontrolle

In Tabelle 6.4 sind die gemittelten Rechenzeiten pro Sequenz der verschiedenen Verfahren für die Datenratenkontrolle aufgeführt. Die Testsequenzen entsprechen Abschnitt 6.3. Vergleicht man die verschiedenen Verfahren für die Datenratenkontrolle hinsichtlich der notwendigen Rechenleistung, so ergeben sich keine großen Unterschiede. Lediglich das Verfahren MPEG4 benötigt eine geringere Rechenleistung, da nur ein Quantisierungsparameter pro Bild berechnet wird.

Datenrate [kBit/s]	TMN8 (C-Code) [s]	MPEG4 (C-Code) [s]	MPEG4MB (C-Code) [s]	FLEX (C-Code) [s]
512	0,3	0,2	0,4	0,4
2048	0,3	0,2	0,4	0,4

**Tabelle 6.4:** Rechenzeit der verschiedenen Verfahren für die Datenratenkontrolle pro Sequenz.

Auf Basis des neuen Verfahrens für die Datenratenkontrolle FLEX werden für die Bestimmung des Quantisierungsparameters auf Makroblock-Ebene Beschleunigungsmöglichkeiten untersucht. Die Bearbeitung kann in zwei Schwerpunkte gegliedert werden: zum einen die Berechnung des Schätzparameters  $m_{DCT_i}$  für jeden Makroblock  $i$  des aktuellen Bildes und die eigentliche Bestimmung des Quantisierungsparameters.

Die Berechnungen des Schätzparameters auf Basis der DCT-Koeffizienten ist rechenintensiv. Die Berechnung lässt sich aber parallelisieren und weist reguläre Speicherzugriffsmuster auf. Somit sind die Voraussetzungen für eine effiziente SIMD-Umsetzung erfüllt. Zur Berechnung des Schätzparameters  $m_{DCT_i}$  auf Basis von (4.24) und (4.25) ist ein Durchlaufen aller DCT-Koeffizienten eines Makroblockes notwendig. Neben der Summation ist noch die Bestimmung des Absolutwertes notwendig.

Dagegen sind für die Berechnung der Varianz auf Basis der Pixelwerte der Y-Blöcke eines Makroblocks nach (2.27), (2.28) und (2.29) zwei Durchläufe erforderlich. Der Erste für die Bestimmung des Mittelwertes und der Zweite für die eigentliche Bestimmung der Varianz. Neben der Summation der Einzelwerte ist für diese Berechnung pro Pixel noch eine Subtraktion und eine Multiplikation notwendig. Für die Berechnung der Varianz auf Basis der DCT-Koeffizienten nach (4.16) ist ebenfalls nur ein Durchlauf notwendig, allerdings ist neben der Summation der Einzelwerte noch eine Multiplikation pro Koeffizient notwendig.

Die Implementierung der Berechnung des Schätzparameters  $m_{DCT_i}$  mittels der SIMD-Einheit des Pentium 4 basiert auf folgenden Operationen, wobei die Masken *sign\_mask*, *h\_mask*, *l\_mask* und *a\_mask* aus Tabelle 6.5 benötigt werden:

1. Initialisierung der notwendigen Bitmasken (siehe Tabelle 6.5).
2. Laden der Koeffizienten in ein Register, z.B. mit dem Befehl `MOVDQA` oder `MOVDQU` in ein `XXM`-Register mit Berücksichtigung der Lage im Speicher (siehe Tabelle C.1).
3. Berechnung des Vorzeichens und der Absolutwerte der Koeffizienten mittels Schiebebefehlen, Oder-Verknüpfung mit *sign\_mask* und Multiplikation mit dem Vorzeichen.
4. Aufsummieren der Absolutbeträge der einzelnen Koeffizienten im aktuellen SIMD-Register. Um diesen Vorgang zu vereinfachen, wird der Befehl `PSADBQ` zur Berechnung der *SAD* zwischen den jeweiligen Werten zweier Register verwendet. In diesem Fall ist das zweite Register mit Null initialisiert, sodass als Ergebnis die Summe der Werte des ersten Registers berechnet wird. Aufgrund der fehlenden Unterstützung von 16 Bit breiten Werten bei der Berechnung der *SAD*, ist es notwendig, die einzelnen 16 Bit breiten Absolutwerte in ein oberes und ein unteres Byte aufzuteilen. Hierzu ist es notwendig das Register zu kopieren. Im Anschluss wird eine Und-Verknüpfung des ursprünglichen Registers mit der Maske *h\_mask* und der Registerkopie mit der Maske *l\_mask* durchgeführt. Im Anschluss wird auf beide Register der Befehl `PSADBQ` zur Berechnung der *SAD* ausgeführt. Der zweite Befehlsoperand ist, wie oben beschrieben, jeweils ein mit Null initialisiertes Register. Für die Berechnung der Summe der Absolutbeträge der einzelnen Koeffizienten wird jetzt die Summe der oberen Bytes um 8 Bit nach links verschoben und zu der Summe der unteren Bytes hinzuaddiert.
5. Addition der Ergebnisse aus Punkt 4 zu den bereits ermittelten Summen der vorherigen DCT-Koeffizienten.
6. Weiter mit Punkt 2, bis alle DCT-Koeffizienten des Makroblockes bearbeitet sind und speichern der Gesamtsumme.

Für die Analyse der DCT-Koeffizienten zur Erhöhung der Codiereffizienz ist die Berechnung des Maximums der Absolutwerte der DCT-Koeffizienten notwendig (siehe Abschnitt 4.4.1). Die Bestimmung des Maximums kann parallel zu der oben beschriebenen Berechnung erfolgen. Hierzu werden die Absolutwerte mit einem Vergleichsregister  $y$ , das anfangs mit Null initialisiert worden ist, verglichen. Die jeweiligen Maximalwerte werden im Vergleichsregister  $y$  abgespeichert.

Nach Abschluss der Berechnungen für den jeweiligen Makroblock wird der maximale Wert im Vergleichsregister  $y$  berechnet. Hierzu wird das Vergleichsregister  $y$  in ein Register  $x$  kopiert. Im Anschluss wird mittels einer Bitverschiebung die obere Hälfte des Registers  $x$  auf die untere Hälfte verschoben. Die Maximalwerte der Register  $y$  und  $x$  werden wieder in  $y$  gespeichert, sodass jetzt alle möglichen Maximalwerte in der unteren Hälfte des Registers  $y$  stehen. Dieser Vorgang wird solange wiederholt, bis nur noch ein Wert, der Maximalwert des aktuellen Makroblocks im Vergleichsregister  $y$  steht und abgespeichert werden kann.

Maskenname	Bitmaske für ein vorzeichenloses Word (16 Bit)
<i>sign_mask</i> :	0x0001
<i>h_mask</i> :	0xFF00
<i>l_mask</i> :	0x00FF
<i>a_mask</i> :	0xFFFF

**Tabelle 6.5:** Initialisierung der für FLEX notwendigen Bitmasken.

Für den zweiten Schwerpunkt, die Bestimmung der Quantisierungsparameter auf Makroblockebene, wird aufgrund der nur geringen Parallelisierungsmöglichkeiten keine SIMD-Implementierung vorgenommen.

Die C-Implementierung des Bewertungsmodells für die Gewichtung von Bildbereichen hinsichtlich der subjektiven Bildwahrnehmung (siehe Abschnitt 4.4.5) benötigt eine ähnliche Rechenleistung, wie die Verfahren für die Datenratenkontrolle selbst (jeweils im Schnitt 0,4 s Rechenzeit sowohl für 512 kBit/s als auch für 2048 kBit/s).

Auch hier ist eine Beschleunigung möglich: zum einen kann für die Texturanalyse die bereits für die Bewegungsschätzung ermittelte Varianzschätzung verwendet werden und zum anderen können die notwendigen Vektoroperationen [53] für die Ermittlung der Gesichtsfarben effektiv beschleunigt werden. Durch diese Maßnahmen kann die notwendige Rechenzeit auf einen geschätzten Wert von ca. 0,1 s für die Bewertung von 300 CIF-Bildern verringert werden.

### 6.3.2 Auswertung der SIMD-Implementierung

Durch die Nutzung der SIMD-Erweiterung des Pentium 4 Prozessors im H.263-Encoder, wie im Abschnitt 6.3.1 beschrieben, kann eine Beschleunigung nahezu um den Faktor 13 (siehe Tabelle 6.6) im Vergleich zur reinen C-Implementierung erreicht werden. Die Beschleunigung ist ausreichend, alle Testsequenzen schneller zu codieren, als es für Echtzeitbedingungen notwendig ist. Im Einzelnen basieren die weiteren Auswertungen auf der Parallelisierung folgender Funktionen des H.263-Encoders:

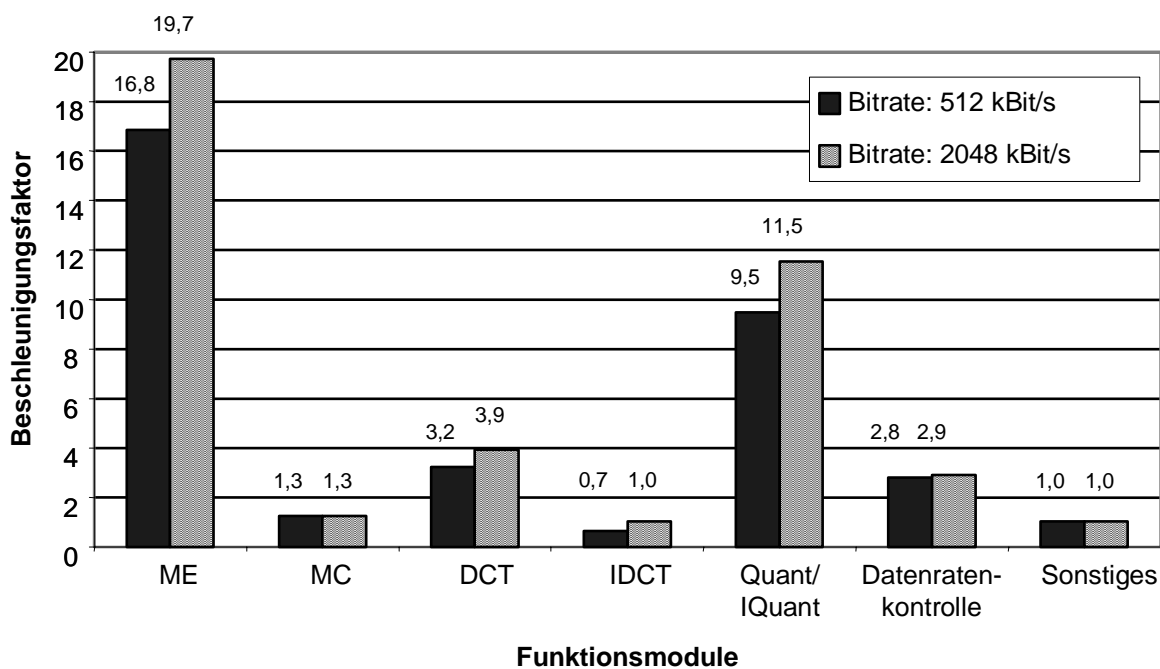
- Bewegungsschätzung (Fullsearch) mit 1-Pixel Genauigkeit im Bereich  $[-16;15, 5]$  um den jeweiligen Makroblock (ME, SSE2).
- Bewegungsschätzung mit  $1/2$ -Pixel Genauigkeit im Bereich von  $\pm 0, 5$  Pixel um den Ergebnisvektor der Fullsearch (ME, SSE2).
- Abschätzung der Varianz des Originalbildes auf Basis einer Ganzzahlenarithmetik (ME, SSE2).
- Addition zweier Bilder zur Berechnung des rekonstruierten Bildes mit Begrenzung der Pixelwerte des Ergebnisbildes auf den Bereich  $[0;255]$  (MC, SSE2).
- Subtraktion des Prädiktionsbildes vom Originalbild zur Berechnung des Fehlerbildes (MC, SSE2).
- Berechnung der 2D-DCT (DCT, MMX).
- Inverse DCT der dequantisierten Koeffizienten (IDCT, MMX).
- Quantisierung und Dequantisierung der Spektralkoeffizienten (Quant/IQuant, MMX).
- Berechnung des Schätzparameters  $m_{DCT_i}$  für die Datenratenkontrolle (Datenratenkontrolle, SSE2).

Sequenz	Gesamtzeit (C-Impl.) [s]	Gesamtzeit (SIMD-Opt.) [s]	Beschleunigungs- faktor
Foreman, CIF, 2048 kBit/s, 30 fps, 300 Bilder	81,33	7,06	11,52
Foreman, CIF, 512 kBit/s, 30 fps, 300 Bilder	80,98	6,67	12,14
Mobile&Calendar, CIF, 2048 kBit/s, 30 fps, 300 Bilder	93,78	7,06	13,28
Mobile&Calendar, CIF, 512 kBit/s, 30 fps, 300 Bilder	93,45	6,69	13,97
Stephan, CIF, 2048 kBit/s, 30 fps, 300 Bilder	90,72	7,06	12,85
Stephan, CIF, 512 kBit/s, 30 fps, 300 Bilder	90,28	6,67	13,54

**Tabelle 6.6:** Vergleich der Rechenzeit der nicht optimierten C und der SIMD-optimierten Funktionsmodule.

Eine Analyse auf Funktionsebene zeigt, dass nicht alle Berechnungen im gleichen Maße von der Parallelisierung durch die SIMD-Einheit profitieren. Wie in Abbildung 6.3 dargestellt, schwanken die jeweiligen mittleren Beschleunigungsfaktoren zwischen eins und nahezu zwanzig. Während insbesondere die Bewegungsschätzung (ME) und die Quantisierung/Dequantisierung (Quant/IQuant) stark von der Parallelisierung profitieren können, ergibt sich für die DCT bzw. IDCT nur eine geringe oder keine Beschleunigung.

Dabei ist zu berücksichtigen, dass bereits im C-Code die DCT- und IDCT-Implementierung auf Geschwindigkeit optimiert worden ist. Ein weiterer Grund ist, dass erst der SSE2-Befehlssatz durch seinen erweiterten Befehlssatz und die 128 Bit breiten Register die DCT- und IDCT-Algorithmen effektiv parallelisieren und damit auch beschleunigen kann. Aufgrund der exklusiven Verwendung der 128 Bit breiten Register ist es zudem nicht notwendig, den Registerstatus nach Abschluss der Berechnung wieder herzustellen. Im Gegensatz zur Verwendung der 64 Bit breiten Register, die gleichzeitig für Fließkommaberechnungen verwendet werden. Insbesondere bei der Berechnung von nur 64 Koeffizienten pro Block ergibt sich hierdurch ein hoher Overhead. Gegenüber der aktuellen Implementierung ist damit beim Einsatz von SSE2 eine Beschleunigung um den Faktor 2 realistisch.

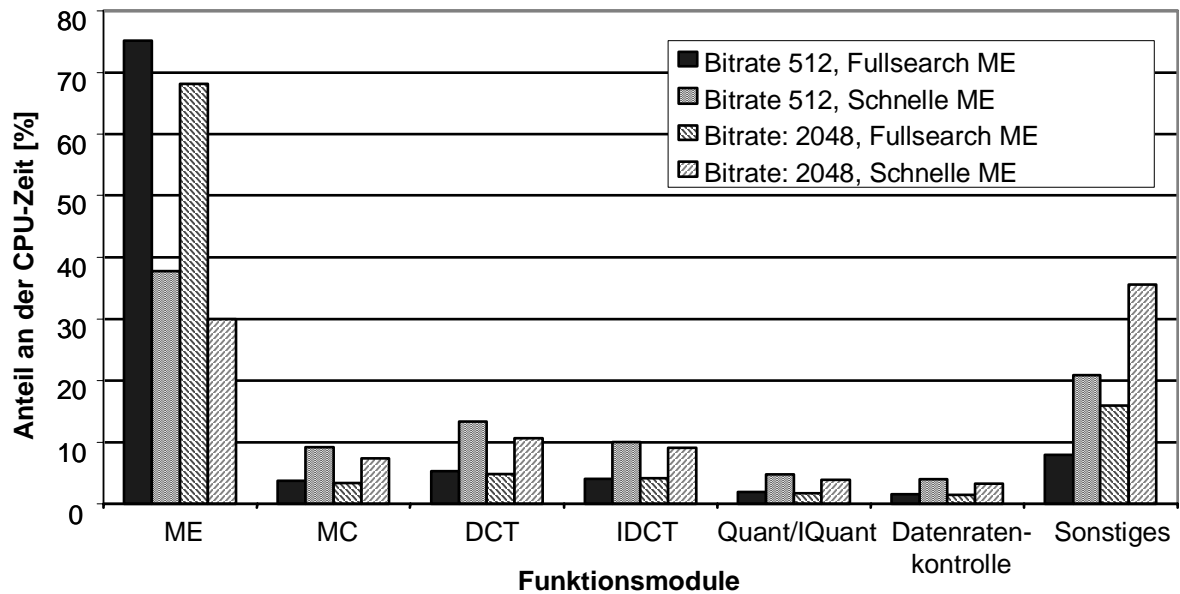


ME: Bewegungsschätzung, MC: Bewegungskompensation,  
 Quant/IQuant: Quantisierung und Inverse Quantisierung

**Abbildung 6.3:** Beschleunigung der SIMD-Implementierung gegenüber dem nicht optimierten H.263-Videoencoder aufgeschlüsselt nach Funktionseinheiten.

In Abbildung 6.4 ist die notwendige Rechenleistung für die einzelnen Funktionseinheiten des Videoencoders mit entsprechender SIMD-Optimierung dargestellt. Insbesondere für die Bewegungsschätzung muss - auch bei einer SIMD-Implementierung - ein hoher Anteil, der zur Verfügung stehenden Rechenleistung, aufgewendet werden.

Aus diesem Grund werden in kostenoptimierten Codecs oftmals vereinfachte Verfahren für die Bewegungsschätzung eingesetzt. Im Vergleich zur Fullsearch weisen sie eine wesentlich verringerte algorithmische Komplexität auf. In der Literatur werden zahlreiche Verfahren für eine schnellere Bewegungsschätzung vorgeschlagen. Teilweise kann dabei eine Beschleunigung um den Faktor 5 erreicht werden. Allerdings verschlechtert sich hierdurch in Abhängigkeit der Bildsignalstatistik die Codiereffizienz [67], [84]. Für eine Visualisierung der Auswirkungen auf die anderen Encodermodule, ist in Abbildung 6.4 noch eine um den Faktor 5 beschleunigte schnelle Bewegungsschätzung berücksichtigt worden [67], [115].



ME: Bewegungsschätzung, MC: Bewegungskompensation,  
Quant/IQuant: Quantisierung und Inverse Quantisierung

**Abbildung 6.4:** Aufteilung der zur Encodierung notwendigen Rechenleistung auf einzelne Funktionseinheiten.

Für eine Analyse der Charakteristik der Algorithmen und der begrenzenden Faktoren sind für die einzelnen Funktionsmodule folgende Kenngrößen ermittelt worden und in Tabelle 6.7 aufgeführt:

1. Speichertransferleistung der SIMD-Module: Hierzu sind im Assemblercode alle IO-Befehle für die Verwaltung und die Berechnungen ausgewertet worden. Auf Basis der gemessenen Rechenzeit und der Anzahl der Durchläufe und der vorhandenen Schleifen im Code wird auf die Transferleistung in GByte/s geschlossen.
2. Für die Berechnung notwendige Speichertransferleistung für die C- und SIMD-Implementierungen: analog dem ersten Punkt, allerdings sind hier nur die reinen Nutzdaten für die Berechnung berücksichtigt worden.
3. Verhältnis der Befehlsanzahl zwischen Berechnung und Verwaltung für die SIMD-Module: es sind jeweils auf Basis des Assemblercodes die Anzahl der Befehle für die Berechnung und für die Verwaltung ermittelt worden. Als Verwaltungsbefehle sind dabei die Initialisierung, die Berechnung von Speicheradressen, das Berechnen und Laden von Bitmasken, die Verwaltung von Schleifen bzw. Sprüngen und sonstiger Overhead gewertet worden.

4. Verhältnis des Speichertransfers für Berechnung und Verwaltung für die SIMD-Implementierung: analog dem dritten Punkt, allerdings sind hier alle IO-Operationen ausgewertet und entsprechend der Zuordnung ins Verhältnis gesetzt worden.
5. Takte pro Pixel für die C- und SIMD-Implementierung: auf Basis der ermittelten Rechenzeit, der Taktfrequenz des Prozessors und der bekannten Anzahl der berechneten Pixel kann der Takt pro Pixel ermittelt werden.

Einzelergebnisse	Bewegungsschätzung (1-Pixel genau)	Bewegungsschätzung (1/2 -Pixel genau) <sup>a</sup>	Erstellung Fehlerbild (Sub)	Bildrekonstruktion (Add)	Quant/IQuant	FLEX ( $m_{DCT_i}$ )
Speichertransferleistung gesamt: SIMD [GByte/s]	11,17	4,79 (4,05)	4,05	3,28	5,76	4,77
Speichertransferleistung Berechnungsdaten: SIMD [GByte/s]	10,05	4,22 (3,57)	3,41	2,50	2,66	1,75
Speichertransferleistung Berechnung: IA32 (C) [GByte/s]	0,64	0,28	0,96	0,96	0,26	0,27
Verhältnis der Befehlsanzahl zwischen Berechnung und Verwaltung: SIMD	2,16	2,41	2,76	3,10	3,17	4,09
Verhältnis Speichertransfer für Berechnung und Verwaltung: SIMD	8,96	8,55	5,33	3,20	0,86	0,58
Takte pro Pixel SIMD	483,13	1,72 (2,19)	3,28	4,48	12,61	4,86
Takte pro Pixel IA32 (C)	7609,39	28,40	11,63	11,63	126,96	31,50

a. Bei der Bewegungsschätzung mit 1/2 -Pixel Genauigkeit ist bei der SIMD-Implementierung eine Fallunterscheidung (siehe Abschnitt 6.3.1.1) notwendig. Für die Schätzung wird von einer übergeordneten Funktion das jeweils passende Verfahren ausgewählt und der weitere Ablauf bestimmt. Bei den Werten in Klammern, ist der Verwaltungsaufwand für die Auswahl des jeweils notwendigen Pixel-Interpolationsverfahrens berücksichtigt und bei den Werten ohne Klammern ist der Overhead herausgerechnet worden.

**Tabelle 6.7:** Effizienzanalyse der SIMD-Funktionsmodule.

Die Bewegungsschätzung stellt in Tabelle 6.7 einen Sonderfall dar. Aufgrund der vielen Vergleiche zwischen den Pixeln des jeweiligen Makroblocks mit den Pixeln im Referenzbild ist zum einen eine hohe Anzahl von Takten pro Pixel notwendig und zum anderen müssen durch die geringe Registeranzahl die jeweils notwendigen Pixel stets neu vom Speicher geladen werden. Dies führt trotz der direkten Unterstützung der Bewegungssuche mit speziellen Befehlen auf ein geringes Verhältnis zwischen der Anzahl der für die Berechnung und der für die Verwaltung notwendigen Befehle. Zudem führt das ständige Nachladen der Pixel zu einer hohen Speichertransfertrate.



Ein Vergleich zwischen einer theoretisch nahezu doppelt so schnellen SSE2- und einer MMX-Implementierung auf einem Pentium 4 im H.263-Encoder ergab nur eine ermittelte Beschleunigung unter 10 %. Dies zeigt, dass bei der Bewegungsschätzung mit 1-Pixel Genauigkeit die Speichertransferleistung der begrenzende Faktor ist.

Insbesondere bei der Implementierung der Bewegungsschätzung ist es daher wichtig, auf die Befehlsreihenfolge zu achten. Durch eine abwechselnde Gruppierung der Befehle können damit möglichst viele Lade-Operationen parallel zu einer davon unabhängigen Berechnung durchgeführt werden (siehe Tabelle C.1).

Bei der Quantisierung und der Dequantisierung führt das häufige Laden von Bitmasken ebenso wie bei der Datenratenkontrolle FLEX zu einem schlechten Wirkungsgrad zwischen Berechnung und dem notwendigen Overhead. Insbesondere das Verhältnis der Speicheroperationen für Berechnungs- und Verwaltungsaufgaben spiegelt dies wider. Die erhöhte Anzahl der Takte pro Pixel bei Quant/IQuant liegt zum einen an der zusammengefassten Berechnung der Quantisierung mit anschließender Dequantisierung und zum anderen an der MMX-Implementierung.

### 6.3.3 Zusammenfassung und Ausblick

Der Vergleich der C- mit der SIMD-Implementierung zeigt deutlich die hohen Beschleunigungsfaktoren, die durch die Parallelisierung möglich sind. Insbesondere die Berechnungen auf Pixel-Ebene können deutlich davon profitieren.

Wie die Untersuchungen bei der Bewegungsschätzung zeigen, ist hier das Speicherinterface oftmals der begrenzende Faktor. Als Lösungsansatz kann zum einen die algorithmische Komplexität für die Bewegungsschätzung verringert werden und zum anderen kann auch das Prozessordesign verbessert werden. Neben den klassischen Möglichkeiten, schnelleren Speicher bzw. einen schnelleren und größeren Cache einzusetzen, können auch die Multimedia-Einheiten selbst verbessert werden. AMD unterstützt z.B. in ihren neuen Prozessorfamilien Opteron und Athlon 64 einen SSE2-Befehlssatz, der die doppelte SIMD-Registeranzahl aufweist. Zudem könnten dann auch Aufgaben, für deren Berechnung Bitmasken notwendig sind, beschleunigt werden. Die Masken werden in den zusätzlichen Registern gespeichert. Dies ist z.B. bei der Quantisierung oder bei FLEX notwendig.

Des Weiteren könnten viele Berechnungen, die aktuell durch mehrere einzelne SIMD-Befehle "emuliert" werden, durch einen verbesserten Befehlssatz beschleunigt werden. So z.B. die Berechnung des Absolutbetrags, die Berechnung des *SAD*-Wertes mit 16 Bit breiten Werten oder die direkte Unterstützung von Bewegungsschätzung und -kompensation mit einer höheren Suchgenauigkeit.

Neben den Berechnungen auf Pixel- bzw. Koeffizienten-Ebene eignen sich noch weitere Aufgaben für eine SIMD-Bearbeitung. Allerdings ist hier meist die erzielbare Beschleunigung geringer. So z.B. bei der Codeckontrolle und auch der Lösung von Gleichungssystemen.

Die SIMD-Implementierung, der in diesem Abschnitt beschriebenen Funktionen, ist ausreichend, alle Testsequenzen wesentlich schneller zu codieren, als für Echtzeitbedingungen notwendig. Für eine weitere Beschleunigung kann, z.B. die DCT und IDCT mit SSE2 berechnet oder die Bewegungskompensation weiter parallelisiert werden. Die Datenratenkontrolle kann durch einen Einsatz der SIMD-Befehle im Steuerbereich nochmals um ca. den Faktor 2 beschleunigt werden.

## 6.4 Minimierung der Ende-zu-Ende-Verzögerung

Neben der Codiereffizienz spielt bei Kommunikationsanwendungen vor allem die erzielbare Ende-zu-Ende-Verzögerung eine wichtige Rolle. In Abbildung 6.5 sind die einzelnen Bearbeitungsmodule einer Video-Punkt-zu-Punkt-Verbindung dargestellt.

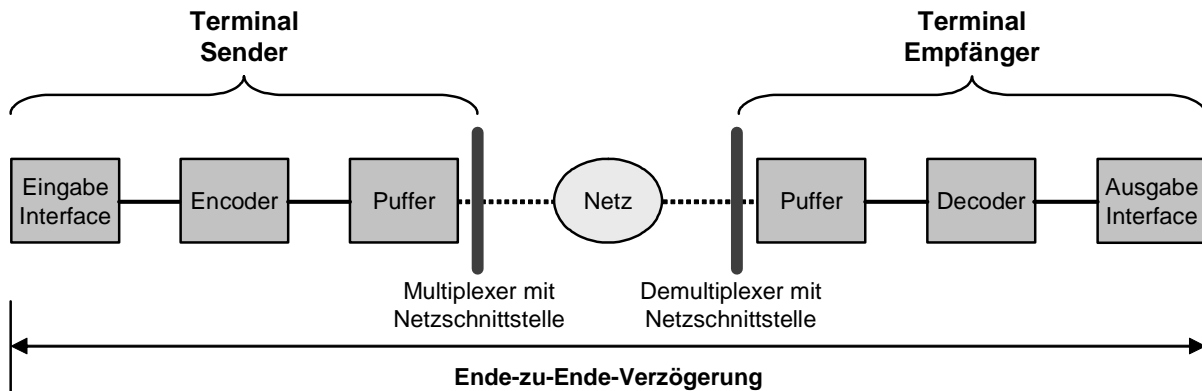


Abbildung 6.5: Komponenten einer Ende-zu-Ende-Kommunikation.

Bei der Standardisierung von H.263 ist beachtet worden, dass eine Videocodierung auf Makroblock-Ebene ohne Zwischenspeicherung und ohne Vorgriff auf zukünftige Makroblöcke realisierbar ist. Die Verfahren für die Datenratenkontrolle bei H.263 und auch bei MPEG-4 (siehe Abschnitt 2.3.3) benötigen allerdings für die Berechnung von Schätzparametern auf Bild-Ebene eine Zwischenspeicherung der Makroblöcke. Additiv zur Pufferung des Bitstroms im Encoder bzw. Decoder bei der Übertragung (siehe Kapitel 2) erhöht sich dadurch die tatsächliche Ende-zu-Ende-Verzögerung nochmals. Insbesondere bei verzögerungskritischen Anwendungen wirkt sich dies negativ aus. Als Lösungsansatz können zum einen die Übertragungspuffer verkleinert und zum anderen die Schätzparameter aus der Bildstatistik vorheriger Bilder abgeleitet werden. Beide Ansätze haben Einfluss auf die Codiereffizienz des Videocodecs.

Für eine bessere Abschätzung des Verhaltens ist es notwendig, die prinzipbedingte Ende-zu-Ende-Verzögerung einer Video-Punkt-zu-Punkt-Verbindung sowohl mit als auch ohne notwendige Zwischenspeicherung der Makroblöcke abzuschätzen. Zusätzlich soll anhand des neuen Verfahrens für die Datenratenkontrolle FLEX untersucht werden, welchen Einfluss die Zwischenpufferung auf die Codiereffizienz hat. In Verbindung mit den Ergebnissen aus Kapitel 5 hinsichtlich unterschiedlicher Übertragungspuffer werden die unterschiedlichen Lösungsansätze im Anschluss bewertet.

### 6.4.1 Ermittlung der Verarbeitungsverzögerung

Die folgende Abschätzung der Verzögerung bei einer IPPP...-Kommunikation<sup>1</sup> beruht auf der Annahme einer makroblockbasierten Pipelineverarbeitung und der daraus entstehenden systembedingten Verzögerung. Hierdurch kann auf die Verzögerung auf Funktionsblockebene geschlossen werden, wobei noch zusätzliche Bild- bzw. Makroblockpuffer berücksichtigt werden müssen.

Im Folgenden wird der eigentliche Videoencodier- bzw. Decodierprozess ohne Pufferung betrachtet, wobei von folgenden Annahmen ausgegangen wird:

- Kontinuierliche Codierung. Die Encodierung und Decodierung benötigen also gerade die Zeit, die zwischen zwei Bildern liegt. Alle Bestandteile der Verarbeitungskette - außer den Puffern im Encoder und Decoder - weisen damit exakt die Durchlaufzeit auf, die für eine Echtzeitbearbeitung der jeweiligen Bildgröße und Framerate notwendig ist.
- Konstante Framerate  $FR$ , die nicht von der Datenratenkontrolle beeinflusst wird. Die Framerate  $FR$  ergibt sich, wenn ausgehend von der Referenz-Framerate  $REF\_FR$  zwischen zwei codierten Bildern jeweils  $m$  Bilder übersprungen werden. Der Zusammenhang zwischen Referenz-Framerate  $REF\_FR$ , der Framerate  $FR$  und der Anzahl  $m$  bei der Codierung ausgelassenen Zwischenbildern ist durch  $FR = REF\_FR / (m + 1)$  festgelegt.
- Alle Bearbeitungsschritte des Codierprozesses (Einlesen, Codierung, Übertragen, Decodierung und Ausgeben) können parallel zueinander abgearbeitet werden und beginnen zum frühestmöglichen Zeitpunkt.

Nach den oben genannten Bedingungen muss der Encoder mit der Referenz-Framerate  $REF\_FR$  und dem gegebenen zeitlichen Abstand zwischen zwei Bildern  $t_{REF}$  bei  $m$  übersprungenen Bildern in der Zeit  $t_p^E$  die Codierung eines kompletten Bildes abgeschlossen haben:

$$t_p^E = (m + 1) \cdot t_{REF} \quad (6.2)$$

In Abbildung 6.6 ist der grundlegende funktionale Aufbau des SW-basierten H.263-Encoders, der dieser Arbeit zugrunde liegt, dargestellt. In Vorwärtsrichtung weist der Encoder  $n^E$  (hier:  $n^E = 5$ ) Funktionsblöcke auf. Dabei werden die Blöcke Inverse DCT (IDCT) und Inverse Quantisierung ( $Q^{-1}$ ) nicht mitgezählt, da deren Ergebnis erst für die Codierung des nächsten Bildes notwendig ist. Für eine Echtzeit-Kommunikation ist es also notwendig, dass alle  $N$  Makroblöcke des aktuellen Bildes die  $n^E$  Funktionsblöcke in der Zeit  $t_p^E$  durchlaufen.

---

1. Bei den PB-Bildern des H.263 werden P- und B-Makroblöcke jeweils zusammen codiert und übertragen. In diesem Fall muss auf den ersten Makroblock des folgenden P-Bildes gewartet werden, bevor mit dem Bearbeiten des aktuellen B-Bildes begonnen werden kann. Damit ergibt sich hier eine wesentlich höhere systembedingte Verzögerung, die in der folgenden Abschätzung nicht weiter berücksichtigt wird.

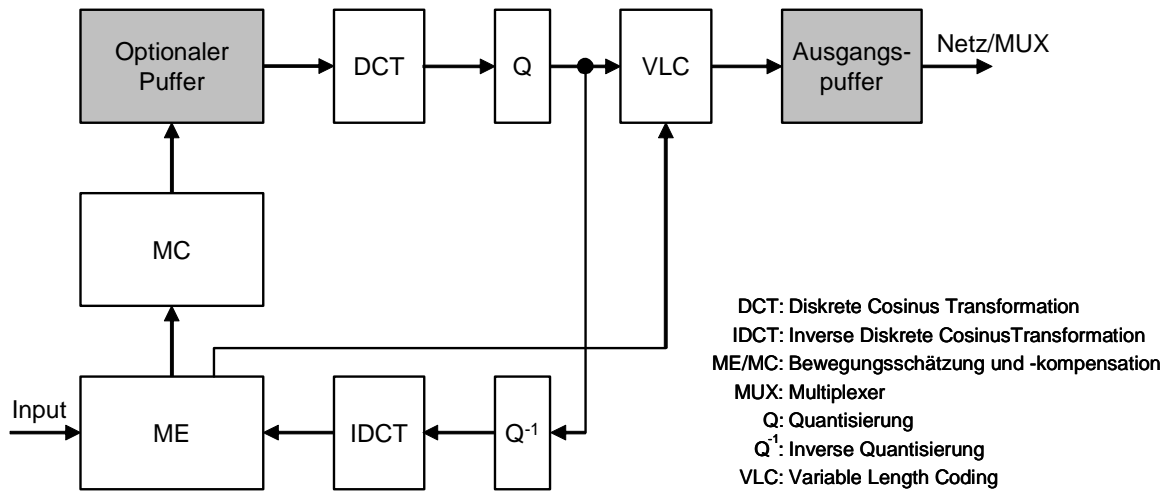


Abbildung 6.6: Verzögerungbestimmende Funktionsblöcke eines Hybrid-DCT Videoencoders.

Mit der Annahme einer makroblockbasierten Pipelineverarbeitung<sup>1</sup> und der konstanten Durchlaufzeit  $t_p^E$  für alle  $N$  Makroblöcke eines Bildes im Encoder ergibt sich bei einer Videocodierung ohne Zwischenpufferung eine Durchlaufzeit  $t_B^{E2}$  pro Funktionsblock von:

$$t_p^E = [(N - 1) + n^E] t_B^E = (m + 1) t_{REF} \text{ bzw.} \tag{6.3}$$

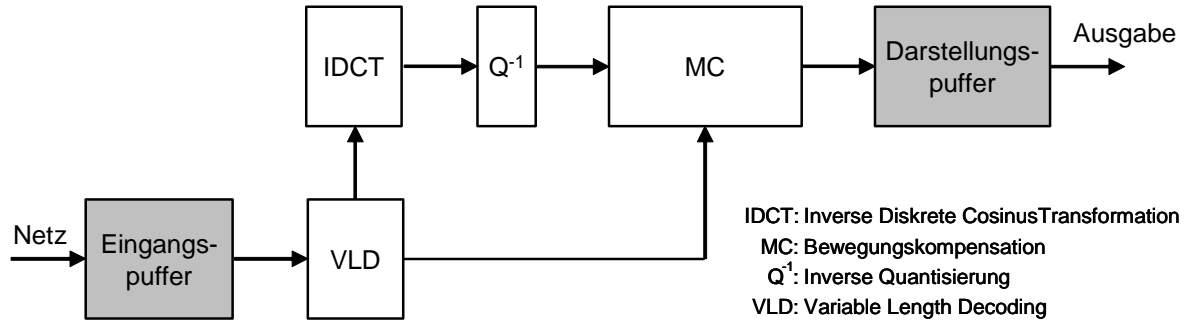
$$t_B^E = \frac{(m + 1) t_{REF}}{(N - 1) + n^E} \tag{6.4}$$

Ist es für die Encodierung notwendig, die einzelnen Makroblöcke zu speichern, wie in Abbildung 6.6 durch den optionalen Zwischenspeicher dargestellt, dann muss man zwischen der Anzahl der Blöcke vor ( $n_{VP}^E$ <sup>3</sup>) und nach ( $n_{NP}^E$ <sup>4</sup>) dem Zwischenpuffer unterscheiden. Nimmt man die gleiche Durchlaufzeit  $t_B^E$  pro Block wie in (6.4) an und benötigt die Auswertung der Zwischenpufferung keine Zeit<sup>5</sup>, so ergibt sich eine Durchlaufzeit  $t_{pD}$ <sup>6</sup> von:

$$t_{pD} = \{[(N - 1) + n_{VP}^E] + [(N - 1) + n_{NP}^E]\} t_B^E = [2(N - 1) + n^E] t_B^E \tag{6.5}$$

1. Jeder Makroblock wird sequenziell und unabhängig der Reihe nach in den verschiedenen Funktionsmodulen des Encoders bzw. Decoders bearbeitet. Sobald die Bearbeitung eines Makroblocks in einem Modul abgeschlossen ist, beginnt die Bearbeitung des im Bild folgenden Makroblocks.
2.  $B$ : Stellt den Bezug zu einem Funktionsblock her.
3.  $VP$ : Stellt den Bezug zu "Vor dem Puffer" her.
4.  $NP$ : Stellt den Bezug zu "Nach dem Puffer" her.
5. In dieser Abschätzung werden die einzelnen Makroblöcke solange gespeichert, bis alle Makroblöcke eines Bildes im Zwischenpuffer eingetroffen sind. Soll keine zusätzliche Bearbeitungsverzögerung notwendig werden, dann müssen alle Makroblöcke direkt beim Abspeichern im Zwischenpuffer bearbeitet/analysiert werden. Ansonsten tritt eine weitere Verzögerung auf.
6.  $D$ : Stellen den Bezug zu verzögert (Delay) her.

Ebenso muss der Decoder in der Zeit  $t_p^D = t_p^E$  alle Funktionsblöcke mit allen  $N$  Makroblöcken eines Bildes durchlaufen haben. In Abbildung 6.7 sind die  $n^D$  (hier:  $n^D = 4$ ) funktionalen Bearbeitungsmodulare eines Hybrid-DCT Videodecoders dargestellt. Es wird dabei von der gleichen Durchlaufzeit pro Block wie beim Encoder ausgegangen ( $t_B^D = t_B^E$ ).



**Abbildung 6.7:** Verzögerungbestimmende Funktionsblöcke eines Hybrid-DCT Videodecoders.

Damit ergibt sich eine Verarbeitungsverzögerung  $t_p^{ED}$  ohne Einbeziehung der Übertragungspuffer:

$$t_p^{ED} = [(N-1) + n^E]t_B^E + [(N-1) + n^D]t_B^D = (2N-2 + n^E + n^D) \frac{(m+1)t_{REF}}{(N-1) + n^E} \quad (6.6)$$

Für die Videocodierung mit Zwischenpufferung ergibt sich mit der gleichen Durchlaufzeit pro Block  $t_B^E$  die Verarbeitungsverzögerung  $t_{pD}^{ED}$  zu:

$$t_{pD}^{ED} = [2(N-1) + n^E]t_B^E + [(N-1) + n^D]t_B^D = [3(N-1) + n^E + n^D] \frac{(m+1)t_{REF}}{(N-1) + n^E} \quad (6.7)$$

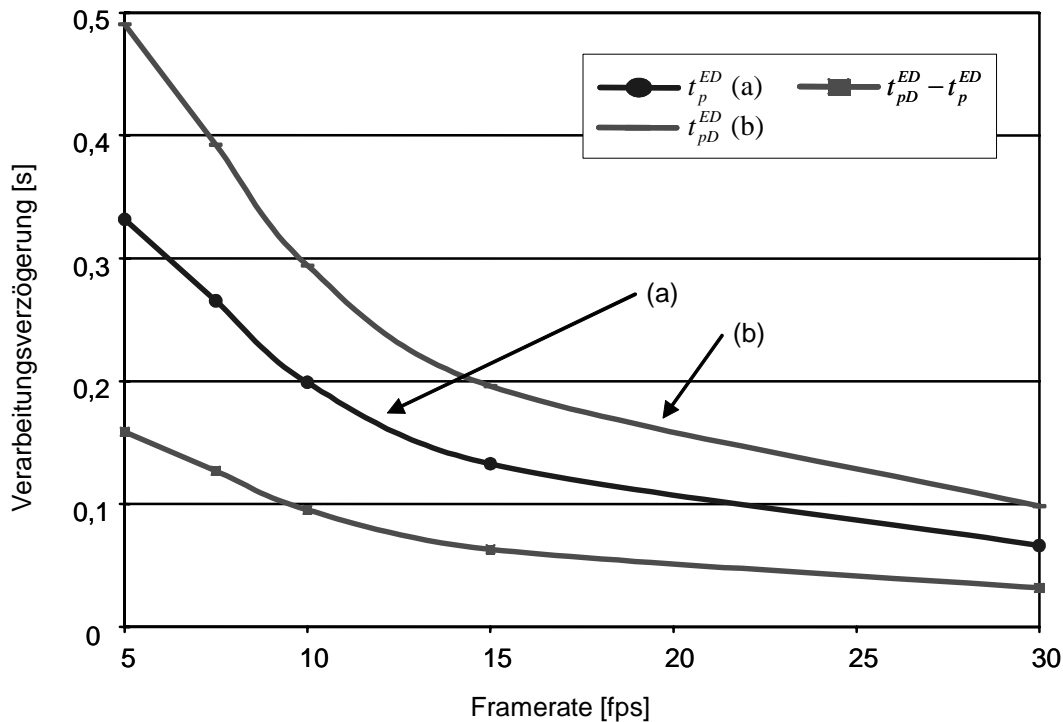
Die Ende-zu-Ende-Verzögerung  $\Delta T^S$  der Video-Punkt-zu-Punkt-Verbindung setzt sich aus der Verzögerung  $\Delta T$  zwischen dem Puffer im Encoder bzw. dem Puffer im Decoder (siehe Abschnitt 2.3.2) und der Verarbeitungsverzögerung  $t_p^{ED}$  bzw.  $t_{pD}^{ED}$  im Encoder und Decoder zusammen:

$$\Delta T^S = \begin{cases} \Delta T + t_p^{ED} & \text{ohne Zwischenpufferung} \\ \Delta T + t_{pD}^{ED} & \text{mit Zwischenpufferung} \end{cases} \quad (6.8)$$

Als Ende-zu-Ende-Verzögerung wird dabei die Differenz der Zeitpunkte definiert, an denen sich ein bestimmtes Bild am Ausgang des Empfängers bzw. am Eingang des Senders befindet.

1.  $S$ : Bezug auf System.

In Abbildung 6.8 sind die Verarbeitungsverzögerung  $t_p^{ED}$  ohne Zwischenpufferung von Codiererergebnissen und  $t_{pD}^{ED}$  mit Zwischenpufferung in Abhängigkeit von der Framerate dargestellt. Die Ergebnisse für CIF bzw. QCIF unterscheiden sich nur sehr gering ( $<10$  ms). In der Darstellung ist keine Pufferverzögerung enthalten ( $\Delta T = 0$  s). In Abhängigkeit von der effektiven Puffergröße und der damit gegebenen Verzögerung (z.B. 100 oder 250 ms) verschieben sich die Kennlinien entsprechend nach oben.



**Abbildung 6.8:** Verarbeitungsverzögerung mit (a) und ohne (b) Zwischenpufferung von Codiererergebnissen im Encoder und Decoder.

Die Verarbeitungsverzögerung hat auf die Ende-zu-Ende-Verzögerung der Videokommunikation die gleiche Auswirkung wie ein größerer Ausgangspuffer. Sie wirkt sich insbesondere bei der niederbitratigen Kommunikation stark störend aus. So ergeben sich z.B. für eine QCIF-Sequenz bei einer Codierung mit 7,5 fps und einer Ausgangspufferverzögerung von 0,2 s eine Ende-zu-Ende-Verzögerung von etwa 0,45 s ohne Zwischenpufferung und 0,55 s mit Zwischenpufferung (siehe Abbildung 6.8). Ebenso kann sowohl eine bildbasierte Vorverarbeitung der Daten, wie eine Erkennung von Bildobjekten oder auch eine entsprechende Nachbearbeitung, die tatsächliche Ende-zu-Ende-Verzögerung um bis zu mehrere Bilder erhöhen.

## 6.4.2 Lösungsansätze zur Minimierung der Verzögerung

Nach der Abschätzung aller verzögerungsbestimmender Faktoren im Videocodec können nun die verschiedenen Lösungsansätze für die Minimierung der Verzögerung hinsichtlich ihrer Realisierbarkeit und auch ihrer Leistungsfähigkeit untersucht und bewertet werden. Prinzipiell stehen für die Minimierung der Ende-zu-Ende-Verzögerung  $\Delta T^S$  nach (6.8) folgende drei Möglichkeiten zur Verfügung:

- Beschleunigung der Videoencodierung bzw. -decodierung. Durch den Einsatz schnellerer Hardware bzw. stärker optimierter Software - über die Echtzeitanforderungen hinaus - ist es möglich, die Durchlaufzeiten der Funktionsmodule im Encoder  $t_B^E$  und im Decoder  $t_B^D$  zu verringern. Dadurch können die Bearbeitungsverzögerungen  $t_p^{ED}$  und  $t_{pD}^{ED}$  direkt beeinflusst werden. So kann z.B. mit einer vierfachen Beschleunigung der Hardware oder Software (siehe Abschnitt 6.3) die Bearbeitungsverzögerung  $t_p^{ED}$  bei der Bildgröße CIF und einer Codierung mit 7,5 fps auf ca. 60 ms im Vergleich zur Abbildung 6.8 reduziert werden. Leistungsfähige Codecs, die nicht bis an ihre Grenzen genutzt werden, können damit durch die zur Verfügung stehende Rechenleistung die tatsächliche Ende-zu-Ende-Verzögerung deutlich verringern.
- Verkleinerung der effektiven Puffergrößen im Codec und damit der Verkleinerung der Zeit  $\Delta T$  (siehe Abschnitt 2.3.2). Größere Puffer sind für die Videoencodierung unkritischer, da die Schwankungen der Bitrate einfacher ausgeglichen werden können. Zudem kann das Überspringen von einzelnen Bildern effektiv verhindert werden. Eine Verkleinerung der Puffergrößen wirkt sich damit insbesondere in kritischen Bereichen störend aus. Zudem kann es dabei zu einer Verringerung der Codiereffizienz kommen. Die Untergrenze für die Zeit  $\Delta T$  in Abhängigkeit vom Bildmaterial und dem jeweiligen Übertragungsnetz liegt bei ca. 100 ms.
- Verzicht auf eine Zwischenspeicherung von Makroblöcken. Oftmals werden für die Datenratenkontrolle statistische Auswertungen auf Bild-Ebene vorgenommen, um den Regelungsprozess an die Anforderungen des Bildes anpassen zu können. So ist z.B. bei den im Abschnitt 2.3.3 beschriebenen Verfahren für die Datenratenkontrolle in MPEG-4 oder H.263, aber auch bei dem in dieser Arbeit entworfenen Verfahren FLEX eine Speicherung von Makroblöcken notwendig.

Im Folgenden werden Varianten von FLEX untersucht, die auf eine Zwischenpufferung von Makroblöcken verzichten.

In Abbildung 6.9 ist der Vergleich des neuen Verfahrens für die Datenratenkontrolle FLEX mit unterschiedlichen Modifikationen dargestellt. Auf Basis des in Kapitel 5 definierten Testsets werden folgende Verfahren gegenübergestellt:

1. FLEX ohne Modifikationen (siehe Kapitel 5) als Referenz: Hier tritt aufgrund der notwendigen Zwischenspeicherung eine zusätzliche Erhöhung der Verzögerung auf.
2. FLEX ohne Zwischenpuffer (Bezeichnung: FLEX\_oZP): Hier wird für die Abschätzung der Aktivität des ganzen Bildes  $k_i$  aus (4.20) auf die Werte des vorherigen Bildes zurückgegriffen. Der Wert  $k_i$  wird dabei pro Makroblock  $i$  bezogen auf die restlichen Makroblöcke  $i, \dots, N-1$  des aktuellen Bildes berechnet (siehe Abschnitt 4.3.1). Aufgrund der verschlechterten Schätzgenauigkeit kann die Analyse der DCT-Koeffizienten für die Erhöhung der Codiereffizienz aus Abschnitt 4.4.1 nicht verwendet werden. Zudem ist die Berechnung des Quantisierungsparameters auf Basis der Aktivität pro Makroblock  $i$  nicht sinnvoll, sodass  $t_i = m_{DCT_i}; \forall i \in [0, N-1]$  aus (4.20) für alle  $N$  Makroblöcke des aktuellen Bildes gesetzt wird. Ebenso muss aufgrund der fehlenden Werte auch auf die Kontrolle der Modellgenauigkeit mit entsprechender Korrektur der Zielbitmenge verzichtet werden (siehe Abschnitt 4.3.4.3).
3. FLEX ohne Zwischenpuffer mit Korrektur der Schätzparameter (Bezeichnung: FLEX\_oZP\_mK): Analog Fall 2 wird hier auf den Schätzwert  $k_i$  aus (4.20) des vorherigen Bildes zurückgegriffen. Allerdings wird der Schätzwert vor der Codierung des aktuellen Makroblocks  $i$  aktualisiert, indem die Werte des korrespondierenden Makroblocks  $i$  des vorherigen Bildes subtrahiert und die für den Makroblock  $i$  des aktuellen Bildes berechneten Werte addiert werden. Hierzu ist eine Speicherung der Werte aller  $N$  Makroblöcke des vorherigen Bildes notwendig, sodass ein zusätzlicher Speicherbedarf entsteht. Aufgrund der verbesserten Schätzgenauigkeit kann im Weiteren das neue Verfahren FLEX ohne weitere Einschränkungen verwendet werden. Lediglich auf die Berechnung des Quantisierungsparameters auf Basis der Aktivität pro Makroblock  $i$  wird - wie im Fall 2 - verzichtet.
4. FLEX ohne Zwischenpufferung mit Korrektur der Schätzparameter und additivem Ausgangspuffer (Bezeichnung: FLEX\_oZP\_mAP): Analog dem Fall 3, allerdings wird hier die Verringerung der Verarbeitungsverzögerung durch einen größeren Ausgangspuffer im Encoder kompensiert. Diese Variante weist damit die gleiche Ende-zu-Ende-Verzögerung wie FLEX aus Kapitel 5 auf. Es kann damit untersucht werden, ob dadurch die Codiereffizienz gesteigert werden kann.

Wie bereits in Kapitel 5 beschrieben, beziehen sich die oberen Verläufe auf die linke Achse und geben den absoluten  $PSNR$ -Wert der verschiedenen Einzeltests in dB an. Die rechte Achse bezieht sich auf die unteren Kurvenverläufe. Sie repräsentieren den Vergleich der Modifikationen von FLEX jeweils als  $PSNR$ -Differenz in dB zu FLEX, das in Kapitel 4 und 5 beschrieben ist. Bei positiven Ergebnissen weist FLEX jeweils einen höheren  $PSNR$ -Wert als das Vergleichsverfahren auf.



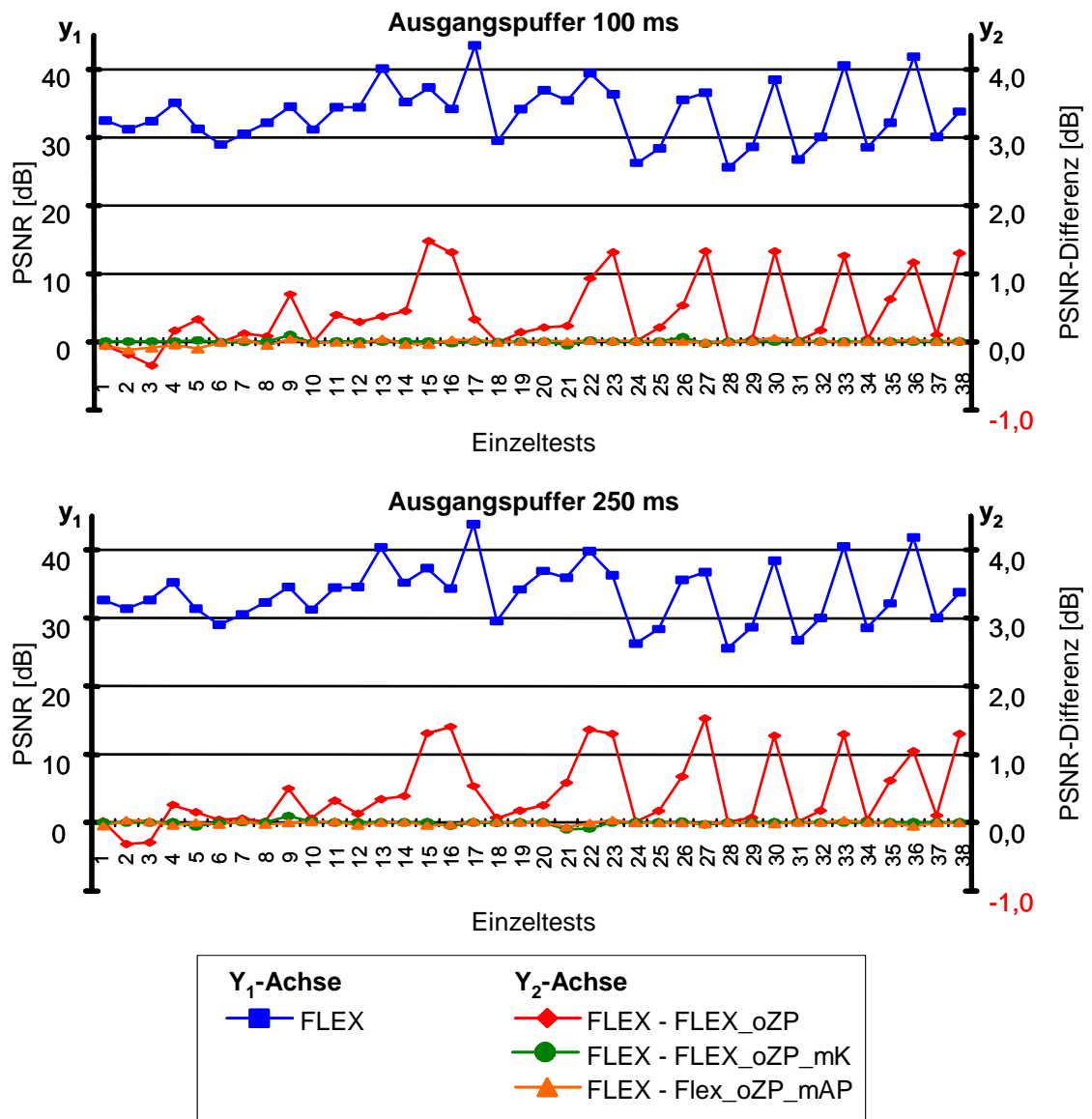


Abbildung 6.9: Qualitätsvergleich unterschiedlicher Varianten der Datenratenkontrolle FLEX.

Die Auswertung zeigt, dass eine Minimierung der Verzögerung von FLEX zu keiner Verschlechterung der Codiereffizienz führt, solange eine Korrektur der Schätzwerte pro Makroblock durchgeführt wird. Dies führt allerdings zu einem höheren Speicherbedarf, da Zwischenergebnisse für die Makroblöcke des letzten Bildes gespeichert werden müssen.

In Abbildung 6.9 ist zudem erkennbar, dass es keine Vorteile bietet, auf die Zwischenpufferung bei der Videoencodierung zu verzichten, um im Gegenzug den Ausgangspuffer im Encoder zu vergrößern. Aufgrund der verringerten Schätzgenauigkeit kann die Minimierung der Verzögerung in kritischen Situationen (z.B. Szenenschnitten) dazu führen, dass Bilder übersprungen werden. Die Adaptivität des Verfahrens für die Datenratenkontrolle verschlechtert sich unter Umständen dabei.

Insgesamt zeigt sich, dass bei verzögerungskritischen Anwendungen mit dem neuen Verfahren für die Datenratenkontrolle eine Ende-zu-Ende-Verzögerung von 300 ms bei 10 fps und einer Ausgangspuffergröße von 100 ms erreicht werden kann. Im Vergleich zu den Verfahren von H.263 oder MPEG-4 ist damit eine Reduzierung um 100 ms ohne Verlust an Codiereffizienz möglich.

Beim Einsatz schnellerer Codecs und einer schnellen Bewegungsschätzung ist eine weitere Reduzierung der Ende-zu-Ende-Verzögerung auf 140 ms bei 10 fps möglich. Hierzu muss der Codec allerdings um den Faktor 5 gegenüber Echtzeit schneller sein, wie es bei aktuellen Prozessoren durchaus möglich ist (siehe Abschnitt 6.3).

## 6.5 Zusammenfassung

Standardprozessoren, wie z.B. der Pentium 4 von Intel, sind aktuell in der Lage, Videosignale in hoher Qualität in Echtzeit zu codieren. Insbesondere der Einsatz der SIMD-Funktionseinheiten in den Prozessoren ermöglicht eine effektive Beschleunigung der Videocodierung. Die Leistungsfähigkeit der Prozessorerweiterung beruht zum einen auf einem speziell an die Anforderung der Signalverarbeitung optimierten Befehlssatz und zum anderen auf der parallelen Anwendung eines Befehls auf mehrere Daten.

Die Untersuchung der Einsatzmöglichkeiten der SIMD-Prozessorerweiterung bei der Videocodierung zeigt, dass mit Ausnahme der VLC alle Funktionsblöcke eines Videoencoders prinzipiell für eine Parallelisierung auf Basis des SIMD-Prinzips geeignet sind. Die Analyse des nicht optimierten H.263-Encoders zeigt, dass die Fullsearch Bewegungsschätzung mit über 95% den Hauptteil der notwendigen Bearbeitungszeit für die Encodierung benötigt. Für die Ermittlung der realen Leistungsfähigkeit der SIMD-Prozessorerweiterung bei der Videocodierung sind folgende Funktionsblöcke eines H.263-Software-Encoder in Standard-C auf einem Pentium 4 in Assembler parallelisiert worden: Fullsearch Bewegungssuche mit 1-Pixel Genauigkeit, Bewegungssuche mit 1/2-Pixel Genauigkeit, Bewegungskompensation, DCT, IDCT, Quantisierung, inverse Quantisierung und Teile der Datenratenkontrolle.

In Abhängigkeit der notwendigen Berechnungen und der Speichertransferleistung sind Beschleunigungsfaktoren zwischen Faktor 20 für die Bewegungsschätzung und keiner Beschleunigung für die IDCT erzielt worden. Insgesamt kann der H.263-Encoder durch die Parallelisierung der genannten Funktionen um den Faktor 13 beschleunigt werden. Eine weitere Reduzierung der notwendigen Rechenleistung kann zum einen durch eine Optimierung der Software-Architektur des Encoders und durch eine weitere Parallelisierung, z.B. von Schleifen oder weiteren Funktionen, erreicht werden. Zudem kann die algorithmische Komplexität, z.B. durch Verzicht auf Fullsearch bei der Bewegungsschätzung, verringert werden.

Ein wichtiger Aspekt für eine effektive Implementierung in Assembler ist dabei die Anpassung der Befehlsreihenfolge an die Prozessorarchitektur. Die jeweiligen Befehle verwenden für die Bearbeitung in Abhängigkeit von ihrer Aufgabe unterschiedliche Prozessoreinheiten, die weitgehend parallel zueinander arbeiten. Zusätzlich muss zwischen der Verzögerung eines Befehls und dem Durchsatz der jeweiligen Prozessoreinheit unterschieden werden. Der Durchsatz entspricht dabei der Anzahl der Taktzyklen, die gewartet werden muss, bis die jeweilige Verarbeitungseinheit erneut benutzt werden kann. Für viele Befehle ist der Durchsatz signifikant geringer als die Verzögerung. Durch die Optimierung der Befehlsreihenfolge wird die zur Verfügung stehende Rechenleistung des Prozessors deutlich gesteigert.

Die Analyse der Ergebnisse zeigt, dass vor allem die Berechnung statistischer Schätzparameter auf Basis einzelner Pixel oder Koeffizienten effektiv beschleunigt werden kann. Eine stärkere Beschleunigung des Encoders wird vor allem durch die begrenzte Anzahl von verfügbaren 64- oder 128-Bit breiten Registern verhindert. Dadurch ist es notwendig, Vergleichswerte oder auch Bitmasken häufig vom Speicher nachzuladen, sodass die zur Verfügung stehende Speichertransferleistung die effektive Rechenleistung des Prozessors reduziert. Zusätzlich werden für die Videocodierung wichtige Operationen, wie z.B. Bewegungsschätzung und -kompensation mit höherer Genauigkeit als 1-Pixel oder die Berechnung des *SAD*-Wertes für höhere Auflösung als 8 Bit, nicht direkt im Befehlssatz unterstützt. Durch die Erweiterung der SIMD-Funktionseinheiten um diese Aspekte kann die Leistungsfähigkeit zukünftiger Prozessoren bei der Videocodierung deutlich gesteigert werden.

Die Verarbeitungsgeschwindigkeit im Videocodec und der Einsatz von Puffern bestimmen im Wesentlichen die systembedingte Verzögerung und damit die Ende-zu-Ende-Verzögerung bei der Videocodierung. Eine Reduzierung der Verzögerung kann durch folgende Maßnahmen erreicht werden: Beschleunigung der Videoencodierung bzw. -decodierung, Verkleinerung der effektiven Puffergrößen im Codec und Verzicht auf eine Zwischenspeicherung von Makroblöcken während der Encodierung. Der zuletzt genannte Punkt erfordert, dass das in Kapitel 5 definierte Verfahren für die Datenratenkontrolle modifiziert wird. Die Quantisierungsparameter für die einzelnen Makroblöcke müssen ohne Kenntnis der folgenden Bildbereiche berechnet werden.

Die Untersuchung des neuen Verfahrens für die Datenratenkontrolle zeigt, dass es durch einfache Modifikationen möglich ist, auf die bisher notwendige Zwischenpufferung von Makroblöcken zu verzichten. Ohne einen nennenswerten Verlust der Codiereffizienz kann in Abhängigkeit von der Framerate die Ende-zu-Ende-Verzögerung zwischen 150 ms bei 5 fps und ca. 50 ms bei 30 fps reduziert werden.



## 7 Zusammenfassung, Bewertung und Ausblick

Die flächendeckende Einführung digitaler Netze und die Verfügbarkeit von multimediafähigen Endgeräten haben die Einsatzmöglichkeiten der Videokommunikation wesentlich erweitert. Seit über 10 Jahren basieren alle neuen Videostandards auf dem Prinzip eines Hybrid-DCT-Videocodierverfahrens mit Bewegungskompensation. Sowohl MPEG-4 als auch H.264/MPEG-4 AVC entsprechen dabei in der grundlegenden Funktion H.263. Durch Erweiterungen, insbesondere bei der Bewegungsschätzung, ist die Codiereffizienz gesteigert worden.

Zusätzlich ist die Funktionalität bei den aktuellen Standards erhöht worden. Während H.261, MPEG-1 und MPEG-2 für spezielle Anwendungsbereiche konzipiert worden sind, können aktuelle Standards ein breites Anwendungsspektrum unterstützen. Zudem stellen sie eine effiziente Übertragung über verschiedene Netze sicher. Die hierzu notwendige Adaptivität und der gezielte Einsatz unterschiedlicher Codierverfahren erhöhen jedoch deutlich den Kontroll- und Steueraufwand im Encoder. Daher werden an die qualitätsbestimmenden Module des Videoencoders, wie die Datenratenkontrolle, neue und höhere Anforderungen gestellt.

Die Aufgabe der Datenratenkontrolle ist die Anpassung der Ausgangsdatenrate des Videoencoders an die verfügbare Bandbreite des Netzes, um eine verzögerungsbegrenzte Übertragung sicherzustellen. Hierzu wird das Bildsignal nach der Dekorrelation quantisiert. Je nach Quantisierungsparameter wird vor der verlustfreien Entropiecodierung eine Verzerrung festgelegt. Die resultierende Bitrate nach der Entropiecodierung ist dabei abhängig von der Verteilungsdichtefunktion des Videosignals.

Eine Codierung mit konstanter Verzerrung resultiert damit in einer - vom Informationsgehalt des jeweiligen Bildes abhängigen - variablen Bitrate. Eine verzögerungsbegrenzte Übertragung erfordert die Anpassung der Bitrate an die gegebene Bandbreite des Netzes, indem die Verzerrung in Abhängigkeit vom Informationsgehalt des Bildes variiert wird. Im Encoder und Decoder wird hierzu ein Puffer benutzt, in dem der Bitstrom bei der Übertragung gespeichert wird. Droht ein Puffer über oder leer zu laufen, muss gegengesteuert werden. Die Verzerrung wird entsprechend erhöht bzw. verringert. Die Größe des Ausgangspuffers hat einen maßgeblichen Einfluss auf die Ende-zu-Ende-Verzögerung bei der Videokommunikation.

Die notwendige Stabilität der Datenratenkontrolle für die Bestimmung der Quantisierungsparameter kann durch eine Modellierung der Rate-Distortion-Charakteristik des Videosignals in Verbindung mit dem Encoder erreicht werden. Den funktionalen Zusammenhang zwischen Bitrate und Verzerrung beschreibt die Rate-Distortion-Theorie.

Die Modellierung der Rate-Distortion-Charakteristik kann anhand statistischer Modelle, wie z.B. Gauß- oder Laplacequellen, näherungsweise beschrieben werden. Für den Einsatz unter Echtzeitbedingungen werden die statistischen Modelle durch eine vereinfachte Modellierung approximiert und an das reale Verhalten des Encoders angepasst. Als Basis für die Vereinfachungen dienen aus dem jeweils interessanten Bereich ableitbare Einschränkungen der Signalstatistik.

Die Betrachtung bekannter, echtzeitfähiger Verfahren für die Datenratenkontrolle zeigt, dass sie bisher in Anlehnung an den jeweils zu unterstützenden Videostandard auf einen begrenzten Anwendungsbereich oder für spezielle Netze konzipiert worden sind. Dabei unterstützen neue Videocodecs ein breites Anwendungsfeld. Das bisherige Vorgehen, für jeden Bereich gezielt ein Verfahren für die Datenratenkontrolle zu entwickeln, ist damit oftmals nicht mehr durchführbar.

Ein möglicher Ansatz die notwendige Flexibilität zu erreichen, ist der parallele Einsatz mehrerer unterschiedlicher Verfahren jeweils für spezielle Einsatzbereiche. Dies führt zu einer erhöhten algorithmischen Komplexität, da ein stufenloser Übergang zwischen den verschiedenen Verfahren für die Datenratenkontrolle gewährleistet sein muss. Zudem steigt der Aufwand für die Implementierung. Es muss für jeden Teilbereich ein passendes Rate-Distortion-Modell mit den jeweiligen Schätzparametern implementiert werden.

Daher ist die Zielsetzung dieser Arbeit ein neues adaptives Verfahren für die Datenratenkontrolle mit dem Schwerpunkt Kommunikationsanwendungen zu konzipieren, das das breite Anwendungsspektrum aktueller Hybrid-DCT-Videocodiervorgängen unterstützt. Gefordert ist ein Verfahren, das eine effiziente Codierung unter unterschiedlichen Randbedingungen, wie z.B. Zieldatenrate, Ende-zu-Ende-Verzögerung oder Komplexität des Bildmaterials sicherstellt. Zudem soll es eine geringe Komplexität aufweisen.

Im ersten Teil dieser Arbeit werden die Anforderungen entsprechend den Vorgaben der Videostandardisierung und des Anwendungsbereichs Videokommunikation ermittelt. Als Anforderungen sind zu nennen: Ausgangsbitraten von einigen kBit/s bis zu 2 MBit/s, Bildmaterial unterschiedlicher Komplexität, eine im Voraus wählbare Ende-zu-Ende-Verzögerung und variable Ausgangsbitraten bei der Übertragung über zeitvariante Netze.

Die Untersuchung bekannter Ansätze für die Modellierung des Rate-Distortion-Verhaltens zeigt, dass im geforderten Anwendungsbereich hohe mittlere Fehler bei der Schätzung der Verzerrung bzw. Bitrate von über 25% auftreten. Eine Reduzierung des Schätzfehlers kann durch folgende Maßnahmen erreicht werden: eine exaktere Rate-Distortion-Modellierung für die prädiktive Beschreibung der Signalcharakteristik in Verbindung mit dem jeweiligen Videoencoder und der Einsatz von verbesserten Schätzmaßen.

Insbesondere für den Einsatz bei echtzeitfähigen Kommunikationsanwendungen ist es aus Kostengründen notwendig, die Systemkomplexität der Datenratenkontrolle zu berücksichtigen. Es sind analytisch lösbare Ansätze notwendig, die ohne Mehrfachcodierung und numerisch iterative Approximationen auskommen.

Für die Schätzung der Verzerrung bietet sich in Abhängigkeit von der Bitrate und der Signalcharakteristik eine Kombination aus zwei verschiedenen Modellen und eine adaptive Parametrisierung an. Dabei muss ein fließender Übergang zwischen den Modellen sichergestellt sein. Einheitliche Modellparameter und Schätzparameter sind für eine Reduzierung der notwendigen Rechenleistung und der Speicheranforderungen notwendig.

Diese Aspekte sind bei der funktionalen Spezifikation des neuen Verfahrens berücksichtigt worden. Ebenso ist ein modularer Aufbau mit einer Funktionsaufteilung in die zwei Bereiche Bitallocation und Makroblockebene, eine geringe algorithmische Komplexität und systembedingte Verzögerung gewährleistet worden. Darüber hinaus ist es insbesondere für die Leistungsfähigkeit unter realen Bedingungen notwendig, dass eine rasche Adaption auf variierende Randbedingungen und an das aktuelle Eingangssignal stattfindet.

Im zweiten Teil dieser Arbeit wird auf Basis der bisherigen Untersuchungsergebnisse ein neuer Ansatz für eine echtzeitfähige Rate-Distortion-Modellierung hergeleitet, der die Grundlage für das neue Verfahren für die Datenratenkontrolle bildet.

Die neue Rate-Distortion-Modellierung beruht auf einem verallgemeinerten Ansatz zur Beschreibung von nieder- bis hochbitratigen Videosignalen. Dabei ist eine parametrisierte Modellierung der Verzerrung zur Erhöhung der Genauigkeit und ein neuer Schätzparameter auf Basis der DCT-Koeffizienten eingesetzt worden. Gegenüber dem Schätzmaß Varianz kann dadurch im Bitratenbereich unter 0,3 Bit/Pixel eine Steigerung der Schätzgenauigkeit erreicht werden.

Auf Basis der hergeleiteten Rate-Distortion-Modellierung berechnet das neue Verfahren für die Datenratenkontrolle den Quantisierungsparameter auf Makroblock-Ebene durch eine Minimierung der Verzerrung unter Einhaltung der vorgegebenen Bitmenge. Die Leistungsfähigkeit der Rate-Distortion-Optimierung ist dabei abhängig von der Genauigkeit der prädiktiven Schätzung und damit auch von der geeigneten Parametrisierung.

Die Auswertung der Ergebnisse der Testsequenzen zeigt, dass die für die Anpassung der Rate-Distortion-Modellierung notwendigen Parameter mit der notwendigen Genauigkeit prädiktiv geschätzt werden können. Die Bestimmung der Parameter vereinfacht sich entgegen den theoretischen Betrachtungen in vielen Fällen durch nichtlineare Effekte im Videocodec, so z.B. der Eingrenzung des Wertebereichs der Quantisierungsparameter, dem erhöhten Signalisierungsaufwand für einen Wechsel des Quantisierungsparameters von Makroblock zu Makroblock bzw. den damit verbundenen Schätzfehlern.

Die Unterstützung optionaler Leistungsmerkmale von aktuellen Videostandards erfordert eine Erweiterung des Verfahrens. Folgende Ansätze sind zusätzlich konzipiert worden: Unterstützung von variierenden Ausgangsdatenraten, Modellierung von I- und B-Bildern bzw. Makroblöcken und die Aufteilung einer Szene in mehrere Objekte. Zudem ist im Design des neuen Verfahrens berücksichtigt worden, andere Verzerrungsmaße als die objektive Bildqualität für die Qualitätsbewertung heranzuziehen. Anhand einer Gewichtung pro Makroblock im Bild können subjektiv wichtige Bildbereiche gewichtet werden. Damit kann das Codierergebnis besser an die Wahrnehmung des Menschen angepasst werden. Für die Steuerung ist beispielhaft ein einfaches Bewertungsmodell implementiert worden.

Die Verifikation des neuen Verfahrens für die Datenratenkontrolle auf Basis der definierten Anforderungen zeigt, dass es die geforderte Flexibilität aufweist. Es erfüllt die Anforderungen aktueller Videocodecs, wie H.263 oder MPEG-4. Im Bitratenbereich von 10 kBit/s bis zu 2 MBit/s, verschiedener Ausgangspuffergrößen und bei Testsequenzen unterschiedlichster Charakteristik kann das neue Verfahren auf Teilbild-Ebene auf sich ändernde Bedingungen reagieren. Damit bildet es die Grundlage für eine stabile und effiziente Encodierung.

Im Vergleich zu den Referenzverfahren aus der Standardisierung von MPEG-4 und H.263 kann das neue Verfahren für die Datenratenkontrolle eine Steigerung der Codiereffizienz erreichen. Der *PSNR*-Gewinn bzw. die entsprechende Verringerung der Übertragungsdatenrate (siehe Abbildung 2.4) ist dabei abhängig von der Ausgangspuffergröße und den jeweiligen Randbedingungen, wie Sequenztyp oder Datenrate. Sie schwankt zwischen ca. -0,4 und 2,2 dB. Der mittlere Gewinn beträgt über 38 einzelne Tests mit unterschiedlichen Randbedingungen dabei ca. 0,7 dB. Nur in einzelnen Fällen sind die *PSNR*-Werte minimal geringer (bis zu 0,4 dB).

Bei gleich bleibender Bildqualität entspricht dies durch die gesteigerte Codiereffizienz jeweils in Abhängigkeit von der aktuellen Rate-Distortion-Charakteristik einer möglichen Reduzierung der Bitrate bis zu 30 % und mehr. Im Vergleich zum aktuellen Stand der Technik ist dies eine deutliche Steigerung der Codiereffizienz (siehe Abbildung 2.4).

Die Qualitätsschwankungen von Bild zu Bild sind ähnlich den Ergebnissen der Vergleichsverfahren. Allerdings sind die Qualitätsschwankungen innerhalb von einzelnen Bildern beim neuen Verfahren gegenüber den ebenfalls auf Makroblock-Ebene arbeitenden Vergleichsverfahren geringer.

Die Codiererergebnisse zeigen, dass es ausreicht, die Modellparameter zur Anpassung an die aktuelle Bildsignalstatistik pro Makroblock-Zeile zu berechnen. Die Bestimmung der Parameter kann auf Basis eines bereits codierten Bildes erfolgen. Bei der Datenratenkontrolle auf Bild-Ebene kann durch den Einsatz einer Rate-Distortion-Optimierung keine Steigerung der Codiereffizienz erreicht werden. Es zeigt sich, dass die Modellfehler aufgrund der prädiktiven Schätzparameter hierfür zu groß sind. Allerdings kann durch eine zusätzliche Analyse der DCT-Koeffizienten zur Modifikation der berechneten Quantisierungsparameter eine deutliche Steigerung der Codiereffizienz erreicht werden. Insbesondere bei Sequenzen mit festem Hintergrund ist eine Steigerung des *PSNR*-Werts von bis zu 1,6 dB möglich.

Im dritten Teil dieser Arbeit wird auf die Implementierungsaspekte eingegangen. Aktuelle Standardprozessoren sind in der Lage, Videosignale in Fernsehqualität in Echtzeit zu codieren. Insbesondere der Einsatz von SIMD-Prozessorerweiterungen ermöglicht eine effektive Beschleunigung der Videocodierung. Die Leistungssteigerung beruht auf einem speziell an die Anforderung der Signalverarbeitung optimierten Befehlssatz und auf der parallelen Anwendung eines Befehls auf mehrere voneinander unabhängige Daten.

Der verwendete H.263-Encoder kann durch den Einsatz der SIMD-Befehlsweiterung bei den Funktionen Bewegungsschätzung, DCT, Bewegungskompensation, Quantisierung/Dequantisierung, Datenratenkontrolle und IDCT um den Faktor 13 beschleunigt werden. In Abhängigkeit der notwendigen Berechnungen und der Speichertransferleistung der jeweiligen Funktionalität sind Beschleunigungsfaktoren zwischen Faktor 20 für die Bewegungsschätzung und keiner Beschleunigung gegenüber einer bereits in C-optimierten IDCT erzielt worden.



Die Analyse der implementierten Funktionen zeigt, dass vor allem die Berechnung statistischer Größen auf Basis einzelner Pixel oder Koeffizienten effektiv beschleunigt werden kann. Damit können auch die Schätzparameter für die Datenratenkontrolle effektiv berechnet werden. Eine stärkere Beschleunigung des Encoders wird vor allem durch die begrenzte Anzahl von Registern verhindert. Dadurch ist es notwendig, Vergleichswerte oder auch Bitmasken vom Speicher nachzuladen. Die zur Verfügung stehende Speichertransferleistung begrenzt die effektiv zur Verfügung stehende Rechenleistung.

Zusätzlich werden für die Videocodierung wichtige Operationen, wie z.B. die Berechnung des *SAD*-Wertes für höhere Auflösung als 8 Bit, nicht direkt im Befehlssatz unterstützt. Die Erweiterung der SIMD-Funktionseinheiten um diese Aspekte kann die Leistungsfähigkeit zukünftiger Prozessoren bei der Videocodierung deutlich steigern.

Die Verarbeitungsgeschwindigkeit im Videocodec und der Einsatz von Zwischenpuffern bestimmen wesentlich die systembedingte Verzögerung der Videocodierung und damit auch die tatsächliche Ende-zu-Ende-Verzögerung des Kommunikationssystems. Eine Reduzierung der Verzögerung kann durch folgende Maßnahmen erreicht werden: Beschleunigung der Videoencodierung bzw. -decodierung, Verkleinerung der effektiven Puffergrößen im Codec und Verzicht auf eine Zwischenspeicherung von Makroblöcken während der Encodierung. Der zuletzt genannte Punkt erfordert, dass die Quantisierungsparameter für die einzelnen Makroblöcke für die Datenratenkontrolle ohne Kenntnis der folgenden Bildbereiche berechnet werden müssen.

Die Untersuchung des neuen Verfahrens für die Datenratenkontrolle zeigt, dass es durch einfache Modifikationen möglich ist, auf die bisher notwendige Zwischenpufferung von Makroblöcken zu verzichten. Ohne nennenswerten Verlust der Codiereffizienz kann in Abhängigkeit von der Frame-rate die Ende-zu-Ende-Verzögerung zwischen 150 ms bei 5 fps und ca. 50 ms bei 30 fps reduziert werden.

Die Ergebnisse dieser Arbeit zeigen, dass es für einen breiten Anwendungsbereich möglich ist, in praktischen Systemen ein effizientes, echtzeitfähiges Verfahren für die Datenratenkontrolle mit einem im Vergleich zu anderen Verfahren geringen bzw. vergleichbaren Aufwand zu implementieren.

Die vorgeschlagene Rate-Distortion-Modellierung bzw. die Erweiterung des modularen Verfahrens für die Datenratenkontrolle können die Basis für zukünftige Untersuchungen und Entwicklungen darstellen.

Ein interessanter Aspekt ist die Weiterentwicklung der in MPEG-4 vorgeschlagenen Adaption an unterschiedliche Übertragungsnetze. Durch eine entsprechende Auswertung der Informationen des Netzes bzw. der Netzschnittstelle ist es möglich, spezielle Eigenschaften der Videocodierung zu berücksichtigen. Als Beispiele sind die Analyse der Paketfehler oder die Auswahl entsprechender Fehlerschutzstrategien für den Videobitstrom in Abhängigkeit von der Zeit zu nennen. Beide Faktoren beeinflussen die effektiv zur Verfügung stehende Bandbreite für die Übertragung des Videobitstroms. Die daraus ableitbaren Informationen können gezielt für die Anpassung der Ausgangsdatenrate des Encoders durch die Datenratenkontrolle genutzt werden. In diesen Bereichen kann eine zusätzliche Steigerung der Codiereffizienz erreicht werden.

Weitere Ergänzungen des Verfahrens für die Datenratenkontrolle sind im Bereich MPEG-4 und der darin gegebenen Unterstützung von verschiedenen Bildobjekten vorstellbar. Aktuell arbeitet das Verfahren mit einer konstanten Framerate, um die visuelle Wahrnehmung von Bewegungsabläufen nicht zu stören. Bei der Aufteilung der Bildszenen in mehrere Bildobjekte ist es allerdings oftmals sinnvoll, den zeitlichen Abstand für die Wiederholung der Objekte gezielt zu beeinflussen. Hierzu ist es notwendig, die vorgeschlagene Rate-Distortion-Modellierung zu erweitern.

Für den Einsatz des neuen Verfahrens für die Datenratenkontrolle bei H.264 sind ebenfalls Anpassungen notwendig. Der Basiscodex von H.264 ist ähnlich wie H.263 aufgebaut. Allerdings ist das Codierverfahren deutlich erweitert worden. So ist z.B. die Bewegungskompensation verbessert worden und anstelle der 2D-DCT wird auf 4 x 4 Blöcke eine separierbare Ganzzahlen-Transformation angewendet, die eine ähnliche Statistik wie die 2D-DCT liefert. Zudem ist die Quantisierung modifiziert worden. Diese Änderungen erfordern eine Anpassung der Rate-Distortion-Optimierung an die speziellen Eigenschaften dieses Codex. Darüber hinaus muss die Parametrisierung des Verfahrens für die Datenratenkontrolle an die neuen Funktionen des Codierverfahrens angepasst werden.

# Anhang



## Anhang A - Beschreibung der Testsequenzen

Bei der Standardisierung von MPEG-4, H.263 und H.264 sind für die Beurteilung der verschiedenen Verfahren die Bildsequenzen und Testbedingungen definiert worden (siehe Kapitel 5). Im Folgenden werden die verwendeten Testsequenzen mit jeweils 300 Bildern, entsprechend 10 s Laufzeit, charakterisiert und anhand von Einzelbildern visualisiert.

### A.1 Kurzbeschreibung der Sequenzen

Sequenzname	Kurzbeschreibung
<b>Klasse: A</b>	
<b>Charakteristik:</b> Wenig Details und Bewegung	
Container Ship	Ein Containerschiff bewegt sich langsam im Wasser. Codierfehler werden im ruhigen Wasser, das einen Großteil der Szene ausfüllt, deutlich sichtbar. Durch die Bewegung des Schiffes werden neue Bereiche aufgedeckt. Zudem fliegen (kleine) Möwen durch die Szene und ein kleines Motorboot bewegt sich im Hintergrund.
Mother & Daughter	Typische Szene einer Videokonferenz. Eine Mutter sitzt mit ihrer Tochter vor einem unbewegten Hintergrund mit wenig Details. Während der gesamten Szene werden nur die Arme und Hände bzw. der Kopf langsam bewegt.
<b>Klasse: B</b>	
<b>Charakteristik:</b> Mehr Details, wenig Bewegung oder wenig Details und mehr Bewegung	
Bowing	Zu Beginn der Szene ist nur ein detaillierter Hintergrund zu sehen. Im Anschluss kommt ein Mann vom Bildrand in das rechte Teilbild. Bewegt sich hin und her und tritt wieder rechts aus der Szene. Die Bewegungen und die verursachten Schatten verdecken dabei wechselnde Teile des Hintergrunds.
Coastguard	Am Anfang der Szene fahren zwei Schiffe aufeinander zu. 2/3 der Szene besteht aus Wasser und 1/3 aus einer steinigen Böschung bzw. Wiese. Die Kamera fängt im Anschluss das große Schiff ein, zentriert es im Bild und verfolgt es im weiteren Verlauf. Dabei entsteht der Eindruck, das Schiff steht und der Bildhintergrund bewegt sich.
Foreman	Zu Beginn der Szene erzählt ein Bauarbeiter vor einem Gebäude aus Betonteilen eine Geschichte und gestikuliert heftig. In der Mitte der Sequenz schwenkt die Kamera nach rechts. An einem Kran vorbei mit Blick auf den Himmel wird ein Rohbau sichtbar. In dieser Szene verweilt die Kamera, wackelt aber erkennbar.

**Tabelle A.1:** Kurzbeschreibung der Sequenzen.

Sequenzname	Kurzbeschreibung
Irene	Vor einem stehenden Hintergrund mit großen, homogenen Elementen (rund und eckig), erzählt eine Frau in der Gebärdensprache mit teilweise schnellen und heftigen Bewegungen eine Geschichte.
News	Zwei Sprecher lesen die Nachrichten vor. Die Hintergrundkulisse ist ein Studiomischpult mit kleinen Monitoren. Im Hintergrund befindet sich ein großer Monitor, in dem eine Ballettszene mit Szenenschnitten und starken Bewegungen eingespielt wird.
<b>Klasse: C</b> <b>Charakteristik:</b> Viele Details, mehr Bewegung oder mehr Details und viel Bewegung	
Mobile& Calendar	Vor einer Kinderzimmertapete mit vielen Details und klaren Strukturen fährt ein Spielzug, der einen Ball vor sich herschiebt. Zudem bewegt sich ein Kalender mit vielen Details und scharfer Schrift auf und ab. Die Kamera bewegt sich während der gesamten Szene.
Stephan	Das Spiel eines Tennisspielers wird von der Kamera eingefangen. Die Kamera bleibt während des Ballwechsels beim gleichen Spieler, sodass schnelle Kamerabewegungen die Szene dominieren. Ständig werden neue Bereiche des Tennisplatzes gezeigt. Als Hintergrund dienen beschriftete Werbetafeln und die Zuschauerränge.
Table Tennis	In einem Raum wird ein Tischtennispiel mit der Kamera verfolgt. Die Kamera ist dabei jeweils auf den Spieler mit Ballbesitz gerichtet. Es treten Kameranäherungen mit schnellen Bewegungen auf. Als Hintergrund dient eine fein strukturierte Tapete und ein Plakat mit Text und Strichzeichnungen.
<b>Klasse: E</b> <b>Charakteristik:</b> Gemischt natürliches und synthetisches Bildmaterial	
Children	Zwei Kinder werfen sich einen Ball zu, der zu Boden fällt. Als synthetischer Hintergrund dient eine abstrahierte Fantasielandschaft mit Bäumen, Wasser und Wiesenelementen. Durch die Szene fliegt ein nahezu bildfüllender teiltransparenter Textblock, der sich verkleinert und in der linken oberen Ecke sichtbar bleibt.
Weather	Eine Frau liest eine Wettervorhersage vor. Als Hintergrund dient eine Landkarte mit Wettersymbolen. Während der Szene bewegt sich die Frau vor der Karte und zeigt mit den Händen auf verschiedene Bereiche der Wetterkarte.

Tabelle A.1: Kurzbeschreibung der Sequenzen.

## A.2 Darstellung der Sequenzen



Bild 0

Bild 99

Bild 299

Abbildung A.1: Sequenz "Container Ship" der Klasse A.



Bild 0

Bild 99

Bild 299

Abbildung A.2: Sequenz "Mother&Daughter" der Klasse A.



Bild 0

Bild 99

Bild 299

Abbildung A.3: Sequenz "Bowing" der Klasse B.



Bild 0



Bild 99



Bild 299

**Abbildung A.4:** Sequenz "Coastguard" der Klasse B.



Bild 0



Bild 199



Bild 299

**Abbildung A.5:** Sequenz "Foreman" der Klasse B.



Bild 0



Bild 99



Bild 299

**Abbildung A.6:** Sequenz "Irene" der Klasse B.





Bild 0



Bild 99



Bild 299

Abbildung A.7: Sequenz "News" der Klasse B.



Bild 0



Bild 99



Bild 299

Abbildung A.8: Sequenz "Mobile&amp;Calendar" der Klasse C.



Bild 0



Bild 99



Bild 299

Abbildung A.9: Sequenz "Stephan" der Klasse C.



Bild 0

Bild 99

Bild 299

Abbildung A.10: Sequenz "Table Tennis" der Klasse C.



Bild 0

Bild 99

Bild 299

Abbildung A.11: Sequenz "Children" der Klasse E.



Bild 0

Bild 199

Bild 299

Abbildung A.12: Sequenz "Weather" der Klasse E.

## Anhang B - Ergänzende Auswertungen zu Kapitel 5

Als Ergänzung zu den Auswertungen im Abschnitt 5.3 sind in diesem Kapitel zusätzlich folgende Testergebnisse aufgeführt:

- Auflistung der Testergebnisse für die Verifikation der Leistungsfähigkeit des neuen Verfahrens für die Datenratenkontrolle FLEX im Vergleich zu den Referenzverfahren.
- Visualisierung des Verlaufs der gemittelten *PSNR*-Werte pro Bild für die Sequenzen Mobile&Calendar und Foreman.

### B.1 Verifikation der Leistungsfähigkeit von FLEX

Für den Leistungsvergleich des neuen Verfahrens für die Datenratenkontrolle FLEX wird dessen Codiereffizienz im Vergleich zu den Verfahren MPEG4, MPEG4MB und TMN8 ermittelt. Hierzu wird jeweils der gemittelte *PSNR*-Wert einer codierten Sequenz bestimmt. Die Datenrate, die Framerate und die Größe des Ausgangspuffers werden vorgegeben. Im ersten Schritt werden für unterschiedliche Sequenzen die gemittelten *PSNR*-Werte der Verfahren für die Datenratenkontrolle über definierte Bitratenbereiche in Abhängigkeit von der Bildgröße ermittelt. Dabei werden unterschiedliche Ausgangspufferverzögerungen untersucht.

Die Ermittlung des *PSNR*-Verlaufs basiert, wie bereits im Kapitel 5 beschrieben, auf folgende Testbedingungen:

- QCIF-Sequenzen: Bitratenbereich von 10 - 128 kBit/s mit 8 Einzelmessungen, Framerate bis einschließlich 15 kBit/s gleich 7,5 fps bzw. für höhere Bitraten gleich 10 fps.
- CIF-Sequenzen: Bitratenbereich von 128 - 2048 kBit/s mit jeweils 6 Einzelmessungen, Framerate bis 256 kBit/s gleich 10 fps, für 512 kBit/s gleich 15 fps bzw. für höhere Bitraten gleich 30 fps.

Jede Sequenz wird mit folgenden Ausgangspufferverzögerungen untersucht: 100, 250, 500 und 1000 ms. In Abhängigkeit von den Ergebnissen wird der Bitratenbereich für einzelne Sequenzen so weit eingegrenzt, dass alle Verfahren für die Datenratenkontrolle interpretierbare Ergebnisse liefern. Der Anteil der übersprungenen Bilder darf bei maximal 10 % der Gesamtsequenzlänge liegen. Eine detaillierte Aufstellung der jeweiligen Parameter befindet sich in Tabelle B.1 für die QCIF-Sequenzen bzw. in Tabelle B.2 für die CIF-Sequenzen.



QCIF Sequenz	Bitrate [kBit/s] (Framerate [fps])							
	10 (7,5)	15 (7,5)	24 (10)	48 (10)	64 (10)	112 (10)	128 (10)	256 (10)
Bowing			x	x	x	x	x	x
Children				x	x	x	x	x
Coastguard				x	x	x	x	x
Container Ship			x	x	x	x	x	x
Foreman				x	x	x	x	x
Irene	x	x	x	x	x	x	x	x
Mother&Daughter	x	x	x	x	x	x	x	x
News		x	x	x	x	x	x	x
Stephan						x	x	x
Table Tennis				x	x	x	x	x
Weather			x	x	x	x	x	x

**Tabelle B.1:** Auflistung der einzelnen Tests der QCIF-Sequenzen.

CIF Sequenz	Bitrate [kBit/s] (Framerate [fps])					
	128 (10)	256 (10)	512 (15)	1024 (30)	1536 (30)	2048 (30)
Bowing	x	x	x	x	x	x
Children		x	x	x	x	x
Coastguard		x	x	x	x	x
Container Ship	x	x	x	x	x	x
Foreman		x	x	x	x	x
Irene	x	x	x	x	x	x
Mobile&Calendar		x	x	x	x	x
Mother&Daughter	x	x	x	x	x	x
News	x	x	x	x	x	x
Stephan			x	x	x	x
Table Tennis		x	x	x	x	x
Weather	x	x	x	x	x	x

**Tabelle B.2:** Auflistung der einzelnen Tests der CIF-Sequenzen.

Für die Auswertungen in den Tabellen B.3 und B.4 sind die einzelnen Ergebnisse pro Sequenz, Bitrate, Framerate und Bildformat mit jeweils einer Ausgangspufferverzögerung ermittelt worden. Im Anschluss ist dann jeweils die Differenz zwischen den gemittelten *PSNR*-Werten in dB von FLEX und dem jeweiligen Vergleichsverfahren berechnet worden. Bei positiven Werten weist FLEX eine höhere Codiereffizienz aus. Die Ergebnisse pro Vergleich sind dann für jede Sequenzklasse zusammengefasst worden, indem der Mittelwert, der maximale und der minimale Wert der Einzelergebnisse bestimmt worden sind. In Tabelle 5.3 im Kapitel 5 sind die Ergebnisse der Tabellen B.3 und B.4 zusammengefasst.

Klasse	Sequenzen	Bildformat: QCIF Qualitätsvergleich bei verschiedenen Bitraten u. Sequenzen: Mid. / Max. / Min. relativ zu FLEX in PSNR [dB]		
		TMN8	MPEG4	MPEG4MB
<b>Ausgangspufferverzögerung: 100 ms</b>				
A	Mother & Daughter, Container Ship	0,08 / 0,37 / -0,16	0,34 / 0,98 / 0,06	0,72 / 1,39 / 0,17
B	Foreman, News, Coastguard, Irene, Bowing	0,13 / 0,57 / -0,10	0,94 / 5,25 / 0,17	0,84 / 1,66 / 0,19
C	Table Tennis, Stephan	0,15 / 0,39 / -0,11	0,55 / 0,96 / 0,35	0,82 / 0,98 / 0,63
E	Weather, Children	1,03 / 1,75 / 0,41	0,90 / 1,70 / 0,44	0,23 / 1,14 / -0,32
<b>Ausgangspufferverzögerung: 250 ms</b>				
A	Mother & Daughter, Container Ship	0,19 / 0,38 / -0,03	0,13 / 0,46 / -0,12	0,29 / 0,64 / -0,05
B	Foreman, News, Coastguard, Irene, Bowing	0,18 / 0,51 / -0,07	0,60 / 3,49 / -0,08	0,25 / 0,63 / 0,01
C	Table Tennis, Stephan	0,22 / 0,41 / -0,05	0,32 / 0,59 / 0,12	0,30 / 0,64 / 0,15
E	Weather, Children	0,39 / 1,40 / -0,36	0,64 / 1,28 / 0,20	0,28 / 1,16 / -0,24
<b>Ausgangspufferverzögerung: 500 ms</b>				
A	Mother & Daughter, Container Ship	0,28 / 0,53 / 0,05	0,16 / 0,67 / 0,02	0,34 / 0,60 / 0,01
B	Foreman, News, Coastguard, Irene, Bowing	0,29 / 0,68 / 0,05	0,43 / 2,50 / -0,06	0,31 / 0,70 / 0,03
C	Table Tennis, Stephan	0,30 / 0,53 / 0,07	0,25 / 0,60 / -0,01	0,34 / 0,58 / 0,22
E	Weather, Children	0,53 / 1,46 / -0,19	0,54 / 1,20 / 0,17	0,31 / 1,08 / -0,13
<b>Ausgangspufferverzögerung: 1000 ms</b>				
A	Mother & Daughter, Container Ship	0,36 / 0,69 / 0,03	0,08 / 0,79 / -0,19	0,30 / 0,58 / -0,11
B	Foreman, News, Coastguard, Irene, Bowing	0,41 / 0,86 / 0,08	0,31 / 2,42 / -0,09	0,30 / 0,69 / 0,05
C	Table Tennis, Stephan	0,36 / 0,70 / 0,06	0,09 / 0,29 / -0,05	0,31 / 0,51 / 0,17
E	Weather, Children	0,65 / 1,63 / 0,04	0,45 / 1,07 / 0,20	0,28 / 1,10 / -0,10
<b>Gesamt:</b>		<b>0,30 / 1,63 / -0,36</b>	<b>0,43 / 5,25 / -0,19</b>	<b>0,43 / 1,70 / -0,24</b>

**Tabelle B.3:** Zusammenfassung des Qualitätsvergleichs der QCIF-Sequenzen.

Klasse	Sequenzen	Bildformat: CIF Qualitätsvergleich bei verschiedenen Bitraten u. Sequenzen: Mid. / Max. / Min. relativ zu FLEX in PSNR [dB]		
		TMN8	MPEG4	MPEG4MB
<b>Ausgangspufferverzögerung: 100 ms</b>				
A	Mother & Daughter, Container Ship	0,17 / 0,54 / -0,02	0,51 / 1,41 / 0,03	0,5 / 1,12 / -0,02
B	Foreman, News, Coastguard, Irene, Bowing	0,09 / 0,51 / -0,47	1,12 / 3,88 / -0,03	0,44 / 0,95 / -0,03
C	Table Tennis, Stephan, Mobile&Calendar	0,12 / 0,33 / -0,02	0,60 / 2,57 / -0,01	0,28 / 0,65 / 0,10
E	Weather, Children	0,55 / 1,44 / 0,02	1,14 / 2,03 / -0,01	0,49 / 2,19 / -0,07
<b>Ausgangspufferverzögerung: 250 ms</b>				
A	Mother & Daughter, Container Ship	0,20 / 0,43 / -0,01	0,43 / 1,11 / -0,10	0,38 / 0,82 / -0,01
B	Foreman, News, Coastguard, Irene, Bowing	0,14 / 0,48 / -0,01	0,42 / 4,96 / -0,11	0,33 / 0,85 / -0,02
C	Table Tennis, Stephan, Mobile&Calendar	0,18 / 0,46 / 0,03	0,17 / 0,77 / -0,06	0,26 / 0,61 / 0,11
E	Weather, Children	0,64 / 1,68 / 0,01	0,49 / 1,25 / -0,11	0,31 / 1,78 / -0,13
<b>Ausgangspufferverzögerung: 500 ms</b>				
A	Mother & Daughter, Container Ship	0,28 / 0,49 / 0,02	0,24 / 1,09 / -0,12	0,42 / 0,90 / 0,01
B	Foreman, News, Coastguard, Irene, Bowing	0,22 / 0,59 / 0,01	0,23 / 2,24 / -0,10	0,38 / 0,91 / 0,01
C	Table Tennis, Stephan, Mobile&Calendar	0,25 / 0,51 / 0,07	0,15 / 0,80 / -0,09	0,30 / 0,62 / 0,17
E	Weather, Children	0,83 / 1,82 / 0,13	0,54 / 1,52 / 0,10	0,40 / 1,68 / 0,03
<b>Ausgangspufferverzögerung: 1000 ms</b>				
A	Mother & Daughter, Container Ship	0,33 / 0,57 / 0,03	0,14 / 0,58 / -0,08	0,37 / 0,85 / 0,00
B	Foreman, News, Coastguard, Irene, Bowing	0,34 / 0,73 / 0,03	0,27 / 2,53 / -0,08	0,39 / 0,91 / 0,00
C	Table Tennis, Stephan, Mobile&Calendar	0,34 / 0,58 / 0,15	0,08 / 0,56 / -0,10	0,29 / 0,60 / 0,16
E	Weather, Children	1,04 / 2,08 / 0,29	0,56 / 1,08 / 0,09	0,40 / 1,60 / 0,03
<b>Gesamt:</b>		<b>0,36 / 2,08 / -0,47</b>	<b>0,44 / 4,96 / -0,12</b>	<b>0,37 / 2,19 / -0,13</b>

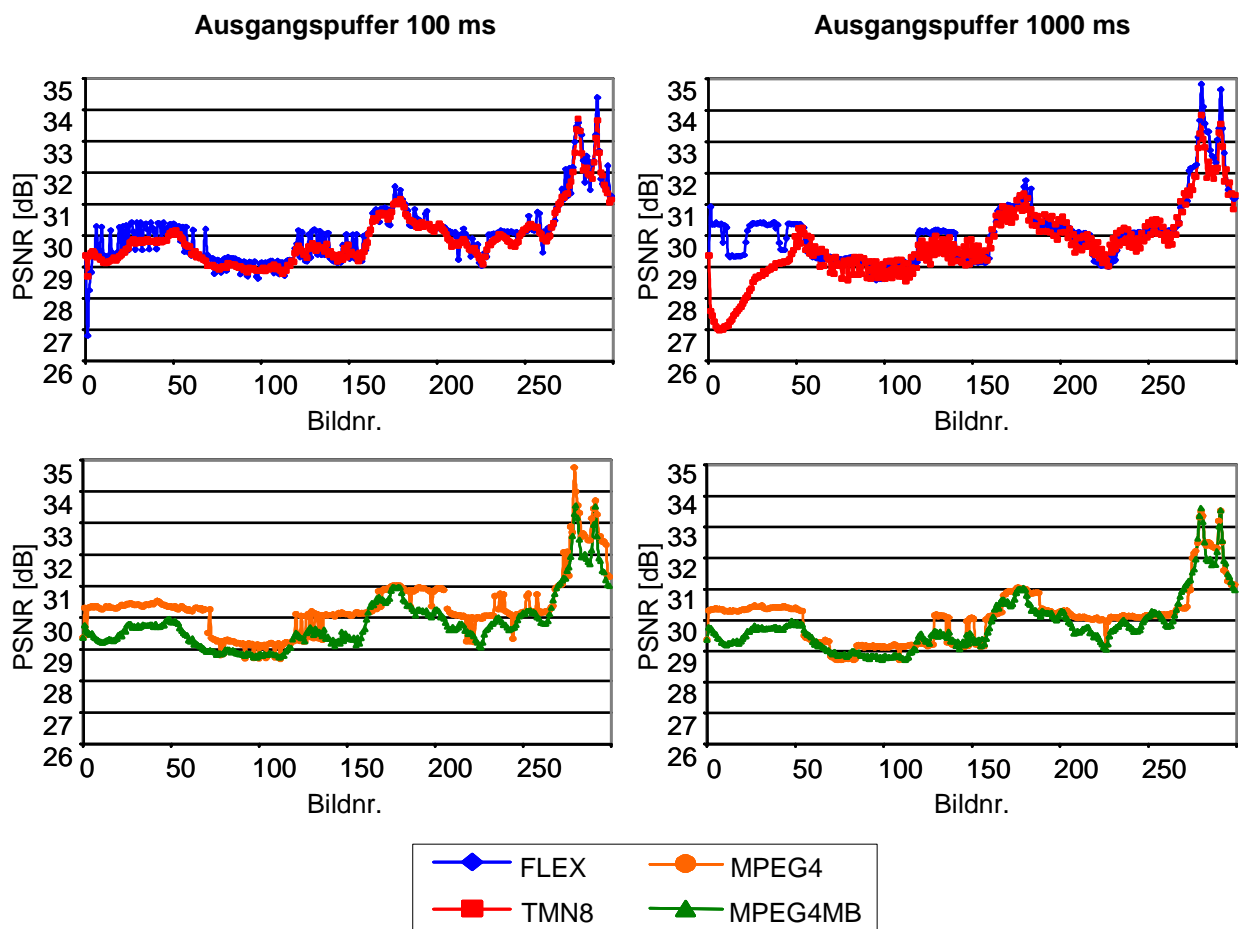
Tabelle B.4: Zusammenfassung des Qualitätsvergleichs der CIF-Sequenzen..

## B.2 Qualitätsanalyse der Verfahren pro Bild

Als Ergänzung zu der Auswertung in Kapitel 5 sind folgende Testergebnisse visualisiert worden:

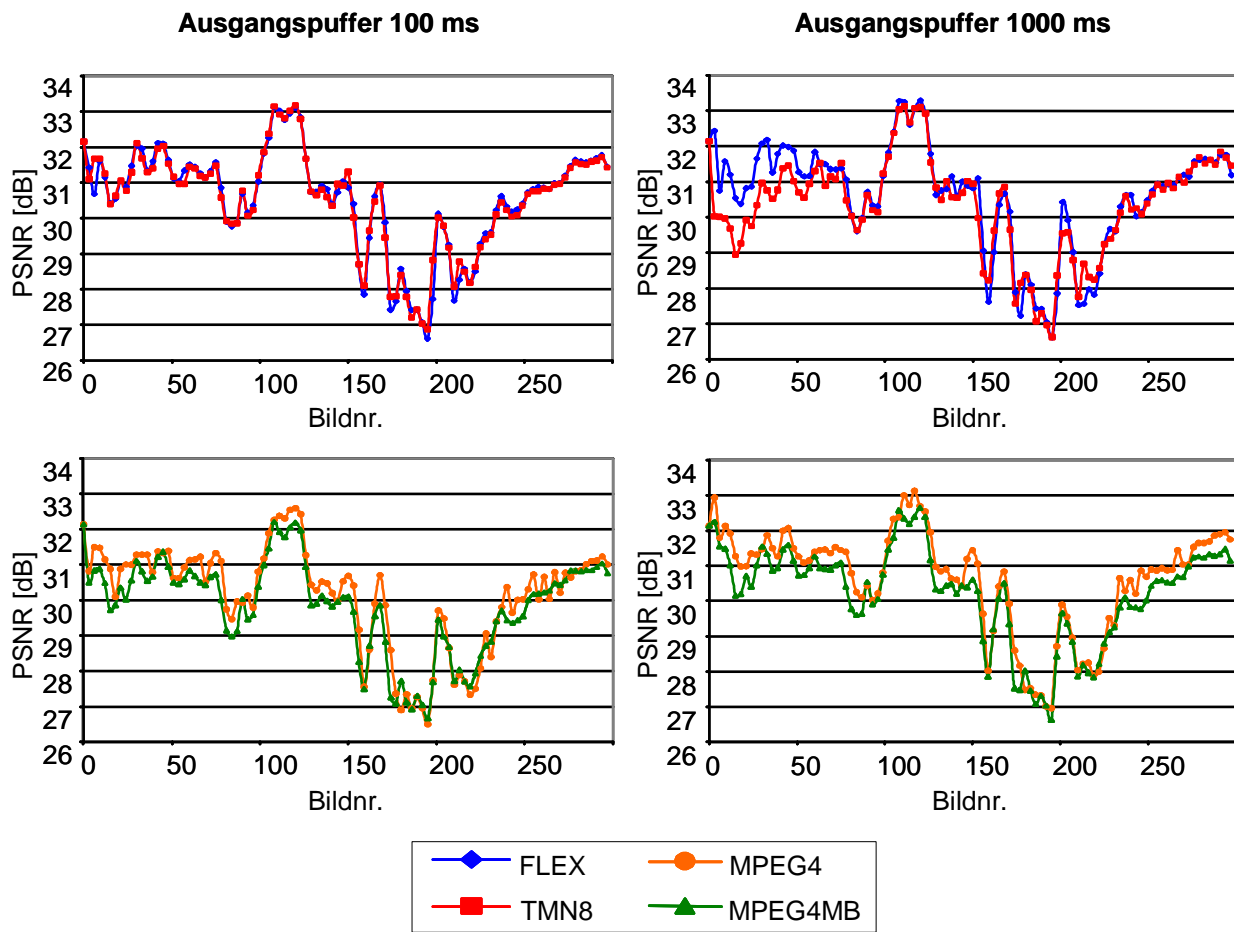
- In Abbildung B.1 ist der Verlauf der *PSNR*-Werte der Einzelbilder der CIF-Sequenz Mobile&Calendar für die Verfahren TMN8, FLEX, MPEG4 und MPEG4MB bei einer IPPP...-Codierung mit einer Bitrate von 2048 kBit/s bei einer Framerate von 30 fps dargestellt.
- In Abbildung B.2 ist der Verlauf der *PSNR*-Werte der Einzelbilder der QCIF-Sequenz Foreman für die Verfahren TMN8, FLEX, MPEG4 und MPEG4MB bei einer IPPP...-Codierung mit einer Bitrate von 48 kBit/s bei einer Framerate von 10 fps dargestellt.

Dabei sind die Verfahren TMN8 bzw. FLEX jeweils in den oberen Diagrammen dargestellt und MPEG4 bzw. MPEG4MB in den unteren.



**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, 2048 kBit/s, 30 fps

**Abbildung B.1:** Qualitätsvergleich der Einzelbilder der CIF-Sequenz Mobile&Calendar für FLEX und den Referenzverfahren.



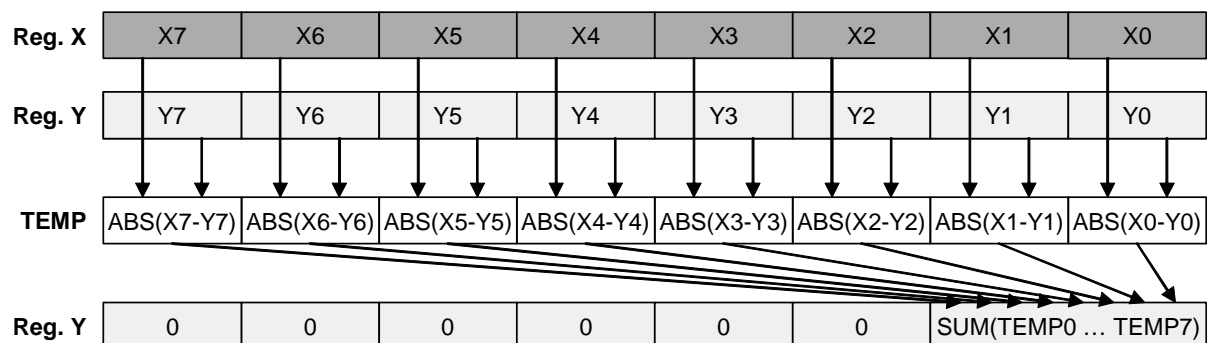
**Bedingungen:** H.263 ohne Optionen, IPPP...-Codierung, 48 kBit/s, 10 fps

**Abbildung B.2:** Qualitätsvergleich der Einzelbilder der QCIF-Sequenz Foreman für FLEX und den Referenzverfahren.



## Anhang C - Ergänzungen zu SIMD-Prozessorerweiterungen in Kapitel 6

Die ersten SIMD-Erweiterungen führte Intel bereits mit dem Pentium unter dem Namen MMX (MultiMedia eXtension) ein. Die Befehlsweiterung nutzte die acht 64 Bit breiten Fließkomma-register für SIMD-Integeroperationen mit jeweils 8, 16 oder 32 Bit Datenbreite<sup>1</sup>. In Abbildung C.1 ist das Funktionsprinzip für die Berechnung der  $SAD(x_R, y_R)$  zwischen zwei unterschiedlichen Bildbereichen  $x_R$  und  $y_R$  anhand des SIMD-Befehl PSADBW (Compute Sum of Absolute Differences) dargestellt. Beide Bildbereiche bestehen in diesem Beispiel aus je 8 aufeinander folgenden Pixeln, die jeweils eine Auflösung von 8 Bit aufweisen. Das Register X ist mit den 64 Bit des Bildbereiches  $x_R$  und das Register Y mit den 64 Bit des Bildbereiches  $y_R$  initialisiert worden. Als temporäre Zwischenergebnisse werden die absoluten Differenzen der korrespondierenden Pixel in den beiden Registern berechnet. Die jeweils 8 Bit breiten Einzelergebnisse werden aufsummiert und das 16 Bit breite Ergebnis wird im Register Y abgespeichert.



**Abbildung C.1:** SIMD-Funktionsbeispiel für die Berechnung des  $SAD$ -Wertes zweier 64-Bit Register mit 8 Bit breiten Einzeldaten.

Mit dem Pentium III sind unter dem Namen SSE (Streaming SIMD Extensions) zusätzlich Erweiterungen definiert worden: acht neue 128 Bit breite Register, zusätzliche SIMD-Ganzzahlenbefehle und Fließkomma-Befehle mit einfacher Genauigkeit. Die aktuellen SSE2-Prozessorerweiterungen sind abwärtskompatibel zu den vorherigen Implementierungen und umfassen folgende Funktionalitäten [41], [43], [45]-[47]:

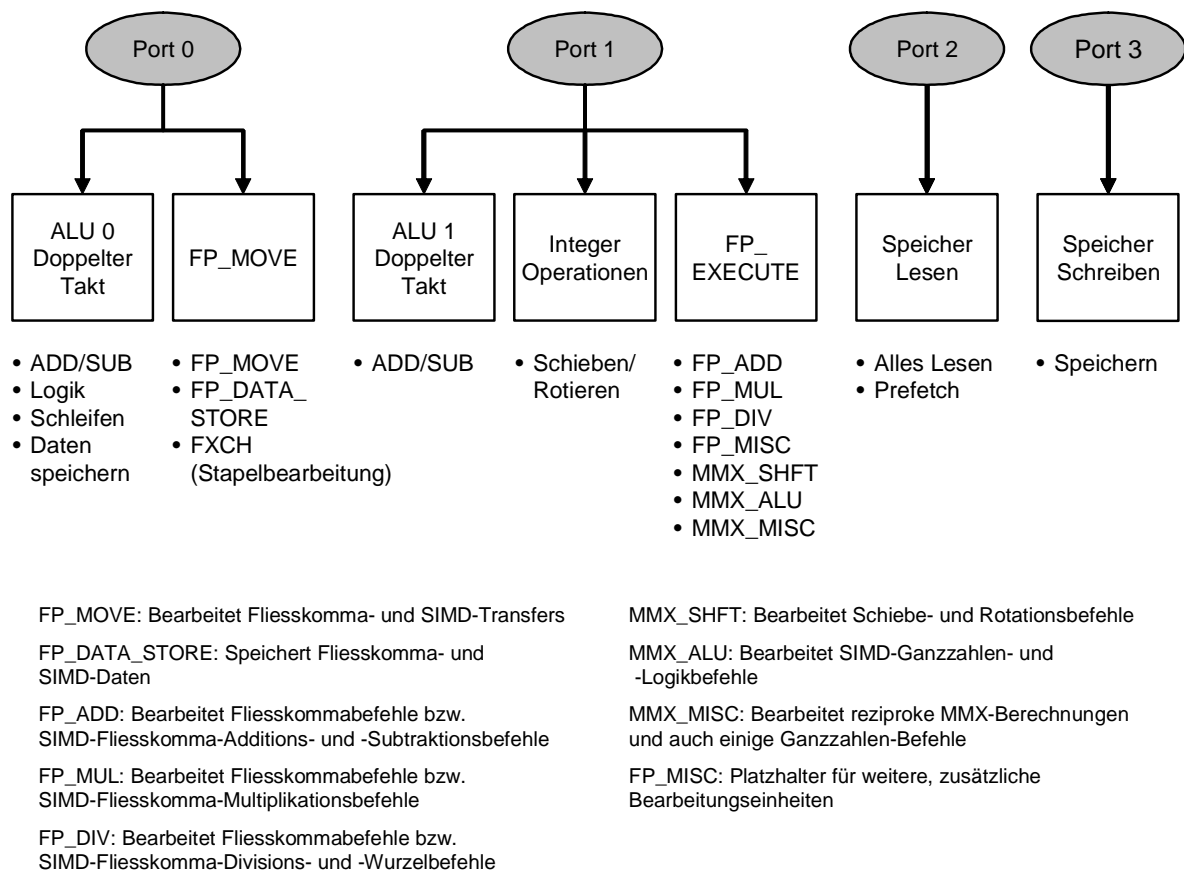
1. Bezeichnungen für Daten mit unterschiedlicher Bitbreite: Byte (8 Bit), Word (16 Bit), Double Word (32 Bit), Quad Word (64 Bit) und Double Quad Word (128 Bit).

- Acht 64 Bit breite Register (MMX-Register), die auch vom arithmetischen Koprozessor (x87) der IA-32 Architektur benutzt werden.
- Acht zusätzliche 128 Bit breite Register (XMM-Register).
- Transferbefehle vom und zum Speicher bzw. zwischen einzelnen Registern.
- Befehle für die Formatkonvertierung.
- Befehle für einfach genaue Fließkommaberechnungen (Floating Point, FP).
- Befehle für doppelt genaue Fließkommaberechnungen.
- Befehle für Ganzzahlen-Berechnungen (Integer).
- Befehle zur Kontrolle des Caches und für die Beeinflussung der Befehlsreihenfolge.

Eine Vorgabe für den Entwurf des Pentium 4 ist es gewesen, eine möglichst hohe Taktfrequenz zu ermöglichen. Daher werden die Befehle in Abhängigkeit von ihrer Komplexität in einen oder mehrere so genannte  $\mu$ -Operationen ( $\mu$ ops) zerlegt und in tiefen Pipelines abgearbeitet. Pro Taktzyklus kann der Prozessor parallel mehrere  $\mu$ -Operationen in Abhängigkeit ihres Typs auf die einzelnen Ports, die Eingänge der Pipelines, verteilen.

Speicheroperationen werden aufgeteilt in das Speichern der Daten und der Adressen. Zudem können die Befehle auch unter bestimmten Voraussetzungen begrenzt umsortiert werden, um eine bessere Auslastung zu ermöglichen. In Abbildung C.2 sind die einzelnen Ports dargestellt. Jede Pipeline besitzt dabei mehrere Verarbeitungseinheiten. Die einzelnen Ports haben folgende Eigenschaften:

- Port 0: In der ersten Hälfte des Taktzyklus unterstützt der Port entweder den Transfer einer Fließkommazahl, eine arithmetische Berechnung (Arithmetik, Logik) oder die Speicherung von Daten. Zusätzlich kann in der zweiten Hälfte des Taktzyklus erneut eine arithmetische Berechnung oder die Speicherung von Daten bearbeitet werden.
- Port 1: In der ersten Hälfte des Taktzyklus unterstützt der Port die Ausführung einer Fließkomma-Operation, einer SIMD-Operation (Transfers werden in Port 0 ausgeführt), einer Ganzzahlenrechenoperation (Multiplizieren, Schieben oder Rotieren) mit normaler Geschwindigkeit oder einer arithmetischen Berechnung (Arithmetik, Logik, Schleifen). Zusätzlich kann in der zweiten Hälfte des Taktzyklus erneut eine arithmetische Berechnung bearbeitet werden.
- Port 2: Unterstützt eine Ladeoperation pro Takt.
- Port 3: Unterstützt die Speicherung einer Adresse pro Takt.



**Abbildung C.2:** Ports und Verarbeitungseinheiten der Pentium 4 Architektur [41].

Aufgrund der Aufteilung der einzelnen Befehle in  $\mu$ -Operationen und die Verarbeitung in Pipelines ist es notwendig, zwischen der Verzögerung eines Befehls und dem Durchsatz zu unterscheiden:

- **Verzögerung (Latency):** Anzahl der Taktzyklen, die der Prozessor benötigt, bis alle  $\mu$ -Operationen des Befehls komplett abgearbeitet sind.
- **Durchsatz (Throughput):** Anzahl der Taktzyklen, die gewartet werden muss, bevor die gleiche Verarbeitungseinheit erneut benutzt werden kann. Für viele Befehle ist der Durchsatz signifikant geringer als die Verzögerung.

In Tabelle C.1 sind für einige Befehle, die in der Videocodierung häufig eingesetzt werden, die Verzögerung, der Durchsatz und die ausführende Verarbeitungseinheit aufgeführt. Aufgrund der parallelen Verarbeitungseinheiten und der verschiedenen Ports kann durch eine geschickte Befehlsreihenfolge der Durchsatz deutlich erhöht werden.

Befehle	Verzögerung <sup>a</sup> [Taktzyklus]	Durchsatz [Taktzyklus]	Verarbeitungseinheiten	Kurzbeschreibung <sup>b</sup>
MOV	0,5	0,5	ALU	Move (IA-32 Befehl): Kopiert max. 32 Bit breite Daten zwischen verschiedenen Registern bzw. dem angegebenen Speicher.
ADD/ SUB	0,5	0,5	ALU	Add/Subtract (IA-32 Befehl): Addiert bzw. subtrahiert max. 32 Bit breite Daten in Registern bzw. dem angegebenen Speicher.
MOVDQA/ MOVDQU	6	1	FP_MOVE	Move Aligned/Unaligned Double Quadword (SIMD-Befehl): Kopiert 128 Bit breite Daten in ein XXM-Register bzw. vom und zum Speicher. Bei "aligned" muss die Speicheradresse ein ganzzahliges Vielfaches von 16 Byte sein.
PSADBW	4	2	MMX_ALU	Compute Sum of Absolute Differences (SIMD-Befehl): Berechnet die Summe der absoluten Differenz zwischen den einzelnen vorzeichenlosen 8 Bit Daten zweier SIMD-Register bzw. dem angegebenen Speicher.
PSRAW/ PSRAD	2	2	MMX_SHFT	Shift Packed Data Right Arithmetic (SIMD-Befehl): Bitweises schieben der 16 bzw. 32 Bit breiten Daten in einem SIMD-Register um die angegebene Anzahl nach rechts mit Berücksichtigung des Vorzeichens.
PADDB/ PADDW/ PADDD	2	2	MMX_ALU	Add Packed Byte/Word/Doubleword Integers (SIMD-Befehl): Addiert die einzelnen 8, 16 oder 32 Bit breiten Daten von SIMD-Registern bzw. dem angegebenen Speicher.
MOVD	6	2	MMX_MISC, MMX_SHFT	Move Doubleword (SIMD-Befehl): Verschiebt 32-Bit von einem SIMD-Register in ein IA-32 Register.
MOVD	10	1	FP_MOVE, FP_MISC	Move Doubleword (SIMD-Befehl): Verschiebt 32-Bit von einem IA32-Register in ein SIMD-Register.
PMULHUW	8	2	FP_MUL	Multiply Packed Unsigned Integers and Store High Result (SIMD-Befehl): Multipliziert die einzelnen vorzeichenlosen 16 Bit breiten Daten von SIMD-Registern bzw. dem angegebenen Speicher zu 32 Bit breiten Einzelergebnissen und speichert die oberen 16 Bit der Ergebnisse ab.
<p>a. Die Angaben beziehen sich jeweils auf Befehle, die jeweils nur Register nutzen. Sobald ein Speicherzugriff notwendig ist, verlängert sich die Verzögerung um mind. 6 Taktzyklen. Der Durchsatz bleibt allerdings gleich.</p> <p>b. SIMD-Register ist ein Sammelbegriff für MMX-Register mit 64 Bit Breite bzw. XXM-Register mit 128 Bit Breite.</p>				

**Tabelle C.1:** Beispiele für Bearbeitungszeiten verschiedener Pentium 4 Befehle [41], [46].

# Verzeichnis der Abkürzungen, Definitionen und Nomenklatur

Im Rahmen dieser Arbeit wird eine weitgehend einheitliche und eindeutige Nomenklatur verwendet. Sofern vereinzelt Doppelbezeichnungen verwendet werden, geht aus dem Zusammenhang eindeutig der Bezug bzw. die Bezeichnung hervor.

Zur Vereinfachung der Lesbarkeit dieser Arbeit sind folgende drei Verzeichnisse jeweils in alphabetischer Reihenfolge aufgelistet:

- Abkürzungsverzeichnis
- Ausgewählte Definitionen und Fachbegriffe
- Modellsignale, Codierungsparameter, Variablen und mathematische Operationen

## Abkürzungsverzeichnis

In dieser Arbeit werden Abkürzungen verwendet, so weit sie in der Fachliteratur üblich sind. Sofern ihre Bedeutung nicht offensichtlich ist, werden sie direkt im inhaltlichen Zusammenhang erläutert. Bei wichtigen oder mehrfach genannten Abkürzungen ist zusätzlich noch eine kurze Erklärung in der Aufstellung vorhanden.

- AAN: Arai-Agui-Nakajima: Bezeichnet einen schnellen Algorithmus für die Berechnung der DCT.
- ALU: Arithmetic Logic Unit: Bezeichnung für eine Verarbeitungseinheit im Prozessor.
- AMD: Advanced Micro Devices, Inc.
  - AR: AutoRegressiv.
- ARM: ARM Holdings plc.
- ARQ: Automatic Retransmission reQuest.
- ATM: Asynchronous Transfer Mode.
- AVC: Advanced Video Coding.
  - B: Bidirectional predicted.
- CBR: Constant BitRate.
- CCIR: Comité Consultatif International du Radiodiffusion.
- CCITT: Comité Consultatif International Télégraphique.
  - CD: Compact Disk.
  - CIF: Common Intermediate Format.

- CU: Control Unit.
- DAB: Digital Audio Broadcast.
- DCT: Diskrete Cosinus Transformation: Bezeichnet auch das gleichnamige Funktionsmodul im Encoder.
- DPCM: Difference Pulse Code Modulation.
- DSL: Digital Subscriber Line.
- DVB: Digital Video Broadcast.
- DVD: Digital Versatile Disk.
- FKT: Fachzeitschrift "Fernseh- und KinoTechnik".
- FP: Floating Point.
- fps: Frames per second (Bilder/s).
- GOB: Group Of Blocks.
- GOP: Group Of Pictures.
- GSM: Global System for Mobile Communication.
- HDTV: High Definition TeleVision.
- HRD: Hypothetical Reference Decoder.
- HW: HardWare.
- I: Intra.
- IA: Intel Architecture.
- IDCT: Inverse Diskrete Cosinus Transformation: Bezeichnet auch das gleichnamige Funktionsmodul im Encoder oder Decoder.
- IEC: International Electrotechnical Commission.
- IEEE: Institute of Electrical and Electronics Engineers, Inc.
- Intel: Intel Corporation.
- IO: Input Output.
- IP: Internet Protokoll.
- ISDN: Integrated Services Digital Network.
- ISO: International Organization for Standardization.
- ITU: International Telecommunication Union.
- LAN: Local Area Network.
- LLM: Loeffler-Ligtenberg-Moschytz: Bezeichnet einen schnellen Algorithmus für die Berechnung der DCT.
- MAD: Mean Absolut Differences.
- MB: MakroBlock: Ein 16 x 16 Pixel großer Bereich des Luminanz-Bildes und der zwei korrespondierenden Chrominanzkomponenten.
- MC: Motion Compensation (Bewegungskompensation): Die Kompensation der Bewegungen zwischen einem oder mehreren Bildern. Bezeichnet auch das gleichnamige Funktionsmodul im Encoder oder Decoder.
- ME: Motion Estimation (Bewegungsschätzung): Die Schätzung der Bewegungen zwischen einem oder mehreren Bildern. Bezeichnet auch das gleichnamige Funktionsmodul im Encoder.
- MMX: MultiMedia Extensions: Dient auch als Bezeichnung für die zugehörigen Register.
- MPEG: Moving Picture Experts Group: Eine Gruppe von Industrie- und Forschungsteilnehmern, die sich mit der Definition von Standards für die Bewegtbildcodierung beschäftigt.
- MSE: Mean Squared Error.

- MUX: MULTipleXer.
- MV: Motion Vector (Bewegungsvektor): Richtung- und Entfernungsinformation, die für die Beschreibung der Bewegung der Makroblöcke von einem Bild zum nächsten verwendet wird.
- NB: NebenBedingung.
- OSI: Open Systems Interconnection.
- P: Predicted.
- PAN: Personal Area Network.
- PB: H.263 unterstützt, P- und B-Bilder als eine Einheit (PB-Bild) zu codieren.
- PC: Personal Computer.
- PCM: Pulse Code Modulation.
- PSNR: Peak Signal to Noise Ratio.
- PU: Processing Unit.
- Q, Quant: Quantisierung, Quantisierung mit der Stufenhöhe Q bzw. Quantisierungsstufe Q: Bezeichnet auch das gleichnamige Funktionsmodul im Encoder.
- $Q^{-1}$ , IQuant: Inverse Quantisierung bzw. inverse Quantisierung mit der Stufenhöhe Q: Bezeichnet auch das gleichnamige Funktionsmodul im Encoder oder Decoder.
- QoS: Quality of Service.
- QP: QuantisierungsParameter: Eine Variable, die im Encoder und Decoder als Skalierung für die Quantisierung bzw. inverse Quantisierung der DCT-Koeffizienten benutzt wird. Oftmals auch gleichbedeutend für Q benutzt, wobei gilt:  $2QP = Q$ .
- RC: RateControl (Datenratenkontrolle): Bezeichnet auch das gleichnamige Funktionsmodul im Encoder.
- RD: Rate-Distortion.
- SAD: Sum of Absolut Differences.
- SIMD: Single Instruction Multiple Data.
- SPIE: The International Society for Optical Engineering.
- SSD: Sum of Squared Differences.
- SSE: Streaming SIMD Extensions: Ein neuer Befehlssatz für den Pentium III.
- SSE2: Streaming SIMD Extensions 2: Ein neuer Befehlssatz für den Pentium 4.
- SW: SoftWare.
- TV: TeleVision.
- U: Siehe YUV-Farbraum.
- UMTS: Universal Mobile Telecommunication System.
- V: Siehe YUV-Farbraum.
- VBR: Variable BitRate.
- VLC: Variable Length Coding: Bezeichnet auch das gleichnamige Funktionsmodul im Encoder.
- VLD: Variable Length Decoding: Bezeichnet auch das gleichnamige Funktionsmodul im Decoder.
- VM: Verification Model.
- VQEG: The Video Quality Experts Group.
- WLAN: Wireless Local Area Network.
- Y: Siehe YUV-Farbraum.
- YUV: YUV-Farbraum: Ein Farbbild wird im YUV-Farbraum durch die Luminanz (Y) und die zwei Farbdifferenzkomponenten U und V (Chrominanz) beschrieben.

## Ausgewählte Definitionen und Fachbegriffe

Die folgende Liste enthält eine Zusammenstellung von ausgewählten Definitionen und Fachbegriffe. Soweit möglich werden sie im inhaltlichen Zusammenhang erläutert. Die Definitionen sind aus der Fachliteratur, wie z.B. Videostandards der ITU-T [73], [74], [76], übernommen und bei Bedarf erweitert worden.

- 16CIF: 16-fach Common Intermediate Format (CIF):  
Mit einer Auflösung von Luminanz: 1408 x 1152 und Chrominanz: 704 x 576.
- 4CIF: 4-fach Common Intermediate Format (CIF):  
Mit einer Auflösung von Luminanz: 704 x 576 und Chrominanz: 352 x 288.
- B-Bild (Bidirectional Predicted Frame): Ähnlich dem P-Bild, allerdings wird der Inhalt durch Interpolation zwischen vergangene-  
nem und zukünftigen Bild vorhergesagt. Die Abweichungen werden wieder mittels  
DCT codiert.
- Bild (Frame, Picture): Ein Bild besteht aus einer Pixelmatrix für die Luminanz und zwei entsprechenden  
Matrizen für die Chrominanz.
- Bild- oder Szenenobjekte: Ein Bild bzw. eine Bildszene wird z.B. aus einem Hintergrundbild (Objekt 1) und Vor-  
dergrundobjekten (Objekte 2 ... n) zusammengesetzt, die getrennt behandelt werden.
- Bitallocation: Reservierung einer Bitmenge (Bitallocation) für die Codierung eines oder teilweise  
auch mehrerer Bilder. Dabei wird meist der aktuelle Pufferfüllstand, die augenblickliche  
Datenrate und die Komplexität bzw. die Codiermethode der Bilder abgeschätzt und  
berücksichtigt.
- Bitrate, Videobitrate, Datenrate: Allgemein definiert als die Anzahl der Bits dividiert durch eine Zeiteinheit. Im Bezug  
auf die Videocodierung wird dabei oftmals auch die Anzahl der Bits des Bitstroms divi-  
diert durch die entsprechende Länge der Testsequenz (einschließlich des ersten Bildes)  
verstanden. Teilweise ist auch eine abschnittsweise Berechnung der Bitrate nach dem  
gleichen Prinzip notwendig. Oftmals wird auch der Begriff Datenrate anstelle von  
Bitrate verwendet.
- Bitstrom: Eine Sequenz von Bits, die eine Repräsentation der codierten Bilder einer oder mehre-  
rer Bildsequenzen darstellen. Zudem können noch Steuerinformationen für die Deco-  
dierung enthalten sein.
- Block: Eine N x M (N-Spalten und M-Reihen) große Pixelmatrix oder eine N x M große  
Matrix von Transformationskoeffizienten.
- Byte: Eine Sequenz von 8 Bit. Das höchstwertige Bit steht links.
- Charakteristik/ Komplexität des Bildsignals: Typische Eigenschaften des Bildmaterials, die Einfluss auf die Art und Weise der  
Codierung haben, wie z.B. Anzahl und Geschwindigkeit von bewegten Bildinhalten,  
Detailreichtum oder Häufigkeit von Szenenwechseln. Man spricht auch von der Kom-  
plexität des Bildsignals.
- Chroma bzw. Chrominanz: Eine Bezeichnung für eine Pixelmatrix oder ein einziges Pixel, das für die Farbkompo-  
nenten-Darstellung eine der beiden zwei Farbdifferenzkomponenten U oder V reprä-  
sentiert.
- CIF: Common Intermediate Format:  
Mit einer Auflösung von Luminanz: 352 x 288 und Chrominanz: 176 x 144.
- Codec: Sammelbezeichnung für Encoder und Decoder.
- Codiereffizienz: Zusammenhang zwischen der Bitrate, die notwendig ist, um eine vorgegebene Verzer-  
rung zu erreichen bzw. der Zusammenhang zwischen Höhe der Verzerrung bei vorgege-  
bener Bitrate.



- Codiermethode: Auswahl eines Prädiktionsverfahrens oder anderer Methoden für die Codierung eines Makroblocks oder eines Bildes (z.B. Intracodierung oder Intercodierung).
- Codiertes Bild: Eine codierte Repräsentation eines Bildes. In Abhängigkeit des Bildtyps kann das Bild einen Bezug zum vorher codierten Bild enthalten (P-Bild) oder auch zum vorherigen und zum nachfolgenden Bild (B-Bild). Ein codiertes I-Bild enthält alle notwendigen Informationen, um es ohne Bezug zu anderen Bildern komplett zu rekonstruieren.
- Codiertes Videosignal (Bildsignal, Video-, Bild-, Sequenz): Siehe Bitstrom.
- Codierte Videobitrate, -datenrate: Die Bitrate oder Datenrate des codierten Videobitstroms, die nach dem Ausgangspuffer des Encoders beim Übergang zum Multiplexer gemessen wird.
- Codierte Videosequenz bzw. Bildfolge: Bitstrom bestehend aus der codierten Information der einzelnen Bilder bzw. deren Teilbereiche und den notwendigen Steuerdaten für den Decoder bzw. für die Synchronisation bei Fehlern. Die einzelnen Elemente des Bitstroms sind meist hierarchisch aufgebaut und in folgende Bereiche unterteilt: Sequenz, Einzelbild, GOB, Makroblock, Block.
- Decoder: Gerät, Bauteil oder Software für das Ausführen des Decodierprozesses.
- Decodierprozess bzw. Decodierung: Ein in einer Vorschrift, wie z.B. eine Recommendation oder ein Standard, spezifizierter Prozess, der den Bitstrom liest und daraus die decodierten Bilder erzeugt.
- Decodiertes Bild: Rekonstruktion eines codierten Bildes aus dem Bitstrom, dabei kann auch ein Bezug zu vorherigen und auch zukünftigen Bildern notwendig sein.
- Double Quad Word: Eine Sequenz von 128 Bit. Das höchstwertige Bit steht links.
- Double Word: Eine Sequenz von 32 Bit. Das höchstwertige Bit steht links.
- Ebene (Layer): Eine zusammengehörige Einheit in einer hierarchischen Anordnung von Elementen. Im Videobitstrom enthalten höhere Layer die niedrigeren, wobei die einzelnen Einheiten z.B. bei H.263 wie folgt definiert sind: Codierte Videosequenz, Bild, GOB (optional), Makroblock, Block. Der Begriff "Ebene" wird ebenfalls für die Strukturierung von Netzwerken verwendet.
- Echtzeitcodierung: Die Zeitdauer für die Codierung eines Bildes ist geringer oder gleich der Dauer, bis das nächste Bild der Videosequenz zur Codierung ansteht. Bei der Verarbeitung kann allerdings eine Verzögerung (z.B. durch Pufferung) entstehen.
- Encoder: Gerät, Bauteil oder Software für das Ausführen des Encoderprozesses.
- Encoderprozess bzw. Encodierung: Ein nicht spezifizierter Prozess, der aus dem Bildsignal einen Bitstrom erzeugt. Der Bitstrom entspricht allerdings einer Vorschrift, wie z.B. einer Recommendation oder einem Standard.
- Ende-zu-Ende-Kommunikation bzw. Ende-zu-Ende-Kommunikationssystem: (Interaktive) Verbindung zwischen zwei Endgeräten (Terminals) über ein oder mehrere Übertragungsnetze. Ein Kommunikationssystem setzt sich im Wesentlichen aus den Einzelkomponenten: Anwendung, Videocodierung und Übertragungsnetz zusammen.
- FLEX: Bezeichnung des neuen Verfahrens für die Datenratenkontrolle, das auf Bild- und Makroblock-Ebene arbeitet und im Rahmen dieser Arbeit konzipiert, untersucht und bewertet wird.
- Flexible Videostandards: Flexible Videostandards, wie z.B. H.263, MPEG-4 oder H.264, können für unterschiedlichste Multimediaanwendungen eingesetzt werden. Sie stellen zudem die Übertragung der codierten Inhalte über verschiedene Netze sicher. Dagegen sind MPEG-1 oder H.261 speziell auf den jeweiligen Einsatzzweck ausgerichtet.
- Flexible Datenratenkontrolle: Bezeichnung für ein Verfahren für die Datenratenkontrolle, das im Gegensatz zu den bisherigen Verfahren nicht auf einen speziellen Einsatzzweck ausgerichtet ist. Vielmehr werden die neuen Möglichkeiten flexibler Videostandards unterstützt bzw. Teilbereiche davon, wie z.B. das breite Spektrum der Kommunikationsanwendungen.

- Framerate: Das Inverse der Zeitdauer von einem bereits codierten Bild bis zum aktuellen Bild, das gerade codiert wird. Oftmals wird auch ein Mittelwert über mehrere Bilder berechnet (siehe mittlere Framerate).
- H.120: Recommendation CCITT H.120 "Codecs for Videoconferencing using Primary Digital Group Transmission".
- H.261: Recommendation ITU-T H.261 "Video Codec for Audiovisual Services at p x 64 kBit/s".
- H.263: Recommendation ITU-T H.263 "Video Coding for Low Bitrate Communication". Erweiterungen der ersten Version von H.263 werden auch mit H.263+ und H.263++ bezeichnet.
- H.264: Recommendation ITU-T H.264 "Advanced video coding for generic audiovisual services".
- I-Bild (Intra Frame): Sämtliche Bildinformation wird, z.B. mittels DCT-Koeffizienten, gespeichert. Dies ist z.B. für das erste Bild einer Sequenz notwendig (Intracodierung).
- Interkodierung (Inter Coding): Codieren eines Blockes, Makroblocks oder eines Bildes mittels einer Inter-Prädiktion.
- Inter-Prädiktion (Inter prediction): Information wird durch Verweise auf ein oder mehrere (decodierte) Bilder beschrieben. Zusätzlich wird die Fehlerkompensation der Prädiktion mittels DCT-Koeffizienten gespeichert.
- Intracodierung (Intra coding): Codieren eines Blockes, Makroblocks oder eines Bildes ohne Verweise auf zukünftige oder vorherige Bilder. Sämtliche Bildinformation wird mittels DCT-Koeffizienten gespeichert, wie es z.B. für das erste Bild einer Sequenz notwendig ist.
- Inverse Transformation: Teil des Decodierprozesses bei dem eine Ansammlung von Transformationskoeffizienten in den Ortsbereich transformiert wird.
- IPPP...-Codierung: Bezeichnung für die Intracodierung des erste Bildes (I-Bild) und der restlichen als P-Bilder (IPPP...-Bildfolge).
- ITU-T: ITU Telecommunication Standardization Sector: einer der drei Sektoren der ITU.
- Level: Eine definierte Zusammenfassung von Einschränkungen des Wertebereichs von Syntax-Elementen oder Variablen in einer Vorschrift, wie z.B. eine Recommendation oder ein Standard. Die gleiche Zusammenfassung ist für alle Profiles definiert, wobei der Level immer gleich definiert ist. Einzelne Implementierungen können in bestimmten Grenzen auch unterschiedliche Level für das jeweilige Profile unterstützen.
- Luminanz bzw. Luma: Ein Farbbild wird im YUV-Farbraum durch die Luminanz (Y), also die Graustufen, und die zwei Farbdifferenzkomponenten U und V (Chrominanz) beschrieben.
- Makroblock: Ein Makroblock (MB) besteht aus 4 Luminanz-Blöcken (Y-Blöcke) und aufgrund der Unterabtastung der Farbinformation je einem Block für die U- und V-Farbinformation. Jeder Block hat dabei eine Auflösung von 8 x 8 Pixel.
- Makroblock-Skipping: Siehe skipped Makroblock.
- Mittlere Framerate: Die Gesamtzahl der encodierten Bilder dividiert durch die Länge der Sequenz (in Sekunden). Die Verzögerung des Codecs wird dabei nicht berücksichtigt.
- MPEG-1: ISO/IEC 11172 "Coded Representation of Moving Pictures and Associated Audio for Digital Storage Media up to about 1.5 MBit/s".
- MPEG-2: ISO/IEC 13818 "Generic Coding of Moving Pictures and Associated Audio Information".
- MPEG-4: ISO/IEC 14496 "Information Technology - Coding of audio-visual Objects -"
- MPEG4: Bezeichnung eines Verfahrens für die Datenratenkontrolle, das auf Bild-Ebene arbeitet und im Multimedia-Standard der Moving Picture Experts Group mit der offiziellen Bezeichnung ISO/IEC 14496 "Information Technology - Coding of audio-visual Objects -" [57] beschrieben wird.

- MPEG4MB:** Bezeichnung eines Verfahrens für die Datenratenkontrolle, das auf Makroblock-Ebene arbeitet und im Multimedia-Standard der Moving Picture Experts Group mit der offiziellen Bezeichnung ISO/IEC 14496 "Information Technology - Coding of audio-visual Objects -" [57] beschrieben wird.
- P-Bild (Predicted Frame):** Die Bildinformation wird durch Verweise auf das vorherige Bild und die Fehlerkompensation der Prädiktion, z.B. mittels DCT-Koeffizienten, gespeichert (Intercodierung).
- Pentium:** Bezeichnung für eine Prozessorfamilie der Intel Corporation.
- Prädiktion, Prädiktionsbild, Prädiktionsfehler:** Ermittelt eine Schätzung auf Basis von Prädiktoren von Teilbildern oder Datenelementen (Ausführung des Prädiktionsprozesses). Als Beispiel besteht das Prädiktionsbild im Encoder aus den Blöcken bereits rekonstruierter Bilder, auf die die Bewegungsvektoren verweisen. Der Prädiktionsfehler ist die Abweichung des Prädiktionsbildes zum jeweiligen Originalbild.
- Profile:** Eine bestimmte Untermenge des Syntaxes einer Recommendation oder eines Standards.
- QCIF:** Quarter Common Intermediate Format:  
Mit einer Auflösung von Luminanz: 176 x 144 und Chrominanz: 88 x 72.
- Quad Word:** Eine Sequenz von 64 Bit. Das höchstwertige Bit steht links.
- Quelle (Source):** Ausdruck für das Videomaterial bzw. -signal.
- Referenzbild (Reference frame):** Ein Referenzbild kann für die Inter-Prädiktion verwendet werden, wenn ein P- oder B-Bild codiert oder decodiert wird.
- Sequenzlänge:** Die Länge der Videosequenz für den Test.
- Single-Pass-Verfahren:** Ohne Mehrfachcodierung. Jeder Makroblock und jedes Bild werden nur einmal codiert.
- Skipped Makroblock:** Ein Makroblock für den keine Daten im Bitstrom übertragen werden, außer der Information, dass er übersprungen wird.
- Sub-QCIF:** Sub-Quarter Common Intermediate Format:  
Mit einer Auflösung von Luminanz: 128 x 96 und Chrominanz: 64 x 48.
- TMN11:** Video codec test model TMN11.
- TMN8:** Video codec test model TMN8 bzw. Bezeichnung für ein Verfahren für die Datenratenkontrolle, das auf Makroblock-Ebene arbeitet und in TMN8 beschrieben ist.
- Verzögerung des Ende-zu-Ende-Videoübertragungssystems bzw. -kommunikationssystems:** Die Verzögerung des Videocodecs vom Einlesen des Bildes im Encoder bis zur Ausgabe durch den Decoder. Dabei wird die Übertragungsbitrate, die Signalvorverarbeitung, die Bearbeitungszeit und die Übertragungszeit berücksichtigt. Oftmals wird hier erst die Verzögerung nach dem Einschwingvorgang betrachtet, also nach der Übertragung des ersten bzw. der ersten Bilder.  
Bei der "anfängliche Verzögerung des Ende-zu-Ende-Videoübertragungssystems bzw. -kommunikationssystems" wird speziell die Verzögerung des ersten Bildes betrachtet.
- XXM:** Bezeichnung für 128 Bit breite Register der SSE- bzw. SSE2-Befehlserweiterung.
- Word:** Eine Sequenz von 32 Bit. Das höchstwertige Bit befindet sich links.

## Modellsignale, Codierungsparameter, Variablen und mathematische Operationen

Die folgende Liste enthält die Bezeichnungen von Modellsignalen und mathematischen Operationen, sowie Bezeichnungen von Codierungsparametern, Symbolen und verwendeten Variablen. Dabei werden folgende Gruppen unterschieden:

- Hochgestellte Indizes und deren Bezug bzw. Bedeutung.
- Tiefgestellte Indizes und deren Bezug bzw. Bedeutung.
- Symbole, Operatoren und Sonstiges.

Bei den Indizes sind auch (mehrfache) Kombinationen möglich.

### Hochgestellte Indizes

*B*: *B*-Bild oder *B*-Makroblock.

*D*: Decoder.

*E*: Encoder.

*F*: Bild (Frame).

*GOP*: Group of Pictures.

*I*: *I*-Bild oder *I*-Makroblock.

*MB*: Makroblock (*MB*).

*MO*: Mehrere Objekte.

*N*: TransportNetz.

*O*: Einzelnes Objekt.

*P*: *P*-Bild oder *P*-Makroblock bzw. Puffer.

*S*: System.

## Tiefgestellte Indizes

- 1, 2, ... : Laufende Nummerierung.
- B* : Block bzw. Funktionsblock.
- c* : Konstante.
- C** : Codebuch.
- CH* : Wechsel (CHange), so kann z.B. in einem Frame der Quantisierungsparameter von Makroblock zu Makroblock wechseln.
- D* : Verzögert (Delay) bzw. Verzerrung (Distortion).
- DCT* : Stellt den Bezug zu den DCT-Koeffizienten her.  $B_{DCT_i}$  ist z.B. die Bitmenge für die Codierung der DCT-Koeffizienten eines Makroblocks oder eines Bildes.
- EFF* : Effektiv-Wert.
- FR* : Framerate.
- H* : Hohe Bitraten (High).
- i* : Laufindex (z.B. zur Kennzeichnung eines Blocks, Makroblocks, Objekts oder Bildes).
- j* : Laufindex (z.B. zur Kennzeichnung eines Blocks, Makroblocks, Objekts oder Bildes).
- l* : Laufindex (z.B. zur Kennzeichnung eines Blocks, Makroblocks, Objekts oder Bildes).
- L* : Niedrige Bitraten (Low).
- LF* : Vorheriges Bild (Last Frame).
- Li* : Linearer Bereich für den Verlauf der Verzerrungskennlinie.
- M* : Mittlere Bitraten (Mid).
- max* : Maximalwert.
- n* : Laufindex (z.B. zur Kennzeichnung eines Blocks, Makroblocks, Objekts oder Bildes).
- NCH* : Kein Wechsel, z.B. konstanter Quantisierungsparameter für jeden Makroblock in einem Bild (No CHange).
- norm* : Normiert, sodass alle Werte  $\leq 1$  sind.
- NP* : Nach dem Puffer.
- opt* : Rate-Distortion-Optimierung.
- over* : Die Bitmenge für die Codierung des Overheads, wie die Übertragung der Bewegungsvektoren oder der Steuerinformationen.
- p* : Bearbeitung (processing), so ist z.B.  $t_p$  die Zeit für die Bearbeitung eines Bildes.
- Qu* : Quadratischer Bereich, so z.B. beim Verlauf der Verzerrungskennlinie.
- R* : Rechteckiger Ausschnitt eines Bildes mit z.B.  $M \times N$  Pixel.
- re* : Bitmenge, die für die Codierung des restlichen Bildes noch zur Verfügung steht.
- s* : Schätzung.
- sf* : Summenkennzeichnung, z.B. ist  $d_{sf_i}$  die Bitmenge für die teilweise abgeschlossene Codierung des aktuellen Bildes bis zum Makroblock *i*.

*REF*: Referenz-Framerate.

*SO*: Signalisierungsoverhead, z.B. ist  $B_{SO}^F$  die Bitmenge pro Bild zur Signalisierung wechselnder Quantisierungsparameter zwischen aufeinander folgenden Makroblöcken.

*u*: Laufindex für die Spektralkoeffizienten der 2D-DCT.

*v*: Laufindex für die Spektralkoeffizienten der 2D-DCT.

*VP*: Vor dem Puffer.

*x*: Bezug zum Signal  $x$ .

## Symbole, Operatoren und Sonstiges

- $\hat{\phantom{x}}$ : Schätzung.
- $\&$ : Und-Operator.
- $!(\dots)$ : Nicht-Operator.
- 0x: Kennzeichnet Hexadezimaldarstellung von Zahlen.
- $a$ : Modellparameter für die Adaption an die aktuelle Signalstatistik.
- $A$ : Amplitude eines digitalen Signals, z.B.  $A = 255$  bei 8 Bit Auflösung bzw. Anzahl der Luminanz-Pixel in einem Makroblock.
- $a_{max_i}$ : Maximalwert der absoluten DCT-Koeffizienten aller Y-Blöcke des Makroblocks  $i$ .
- ADD: Add (IA-32 Befehl).
- $B_i$ : Geplante oder geschätzte Bitmenge, z.B. für das  $i$ -te Bild oder den  $i$ -ten Makroblock.
- $BPPmaxKb$ : Maximale Bitmenge pro Bild, die zwischen Encoder und Decoder auf Basis von definierten Standardwerten signalisiert wird.
- $br$ : Puffergröße in Abhängigkeit von der Bitrate  $R$ .
- $bs_n$ : Pufferfüllstand zum Zeitpunkt des Bildes  $n$ .
- $BS, BS_n$ : Puffergröße. Die Abhängigkeit von der Zeit kennzeichnet ein Laufindex, wie z.B.  $n$ .
- $b_t$ : Zielfüllhöhe des Puffers.
- $b_{tmax}$ : Vorgabe für die maximale Befüllungshöhe des Puffers.
- $c$ : Modellparameter für die Adaption an die aktuelle Signalstatistik.
- $C$ : Codebuch, z.B. für die Abbildung des Vektors  $\mathbf{X}$  auf den Vektor  $\mathbf{Y}$ .
- $C_D$ : Stellt die Menge aller Codes dar, die eine Codierung des Ausgangsvektors mit der gleichen Verzerrung  $D$  ermöglichen.
- $C_n, C$ : Bitmenge, die im Intervall zwischen den Bildern  $n$  und  $n + 1$  über den Übertragungskanal transportiert werden kann bzw. falls konstant (unabhängig von der Zeit) ohne Index.
- $c_{uv}, c_{juv}$ : Spektralkoeffizienten  $c_{uv}$  der orthonormalen 2D-DCT-Transformation eines rechteckigen  $U \times V$  großen Bildbereichs mit  $0 \leq u < U$  und  $0 \leq v < V$ . Dabei werden meist die Indizes  $u$  und  $v$  verwendet. Der Index  $j$  kennzeichnet den jeweiligen Block.
- $cdqp_{uv}$ : Mit dem Quantisierungsparameter  $QP$  quantisierter und wieder dequantisierter Spektralkoeffizienten  $c_{uv}$  der orthonormalen 2D-DCT-Transformation eines rechteckigen  $U \times V$  großen Bildbereichs mit  $0 \leq u < U$  und  $0 \leq v < V$ . Dabei werden meist die Indizes  $u$  und  $v$  verwendet.
- $cqp_{uv}$ : Mit dem Quantisierungsparameter  $QP$  quantisierte Spektralkoeffizienten  $c_{uv}$  der orthonormalen 2D-DCT-Transformation eines rechteckigen  $U \times V$  großen Bildbereichs mit  $0 \leq u < U$  und  $0 \leq v < V$ . Dabei werden meist die Indizes  $u$  und  $v$  verwendet.
- $D$ : Distortion/Verzerrung, z.B. für ein Bild oder Objekt.
- $d_n$ : Nach der Codierung ermittelte tatsächliche Bitmenge für z.B. das  $n$ -te Bild.
- $D(Q)$ : Distortion eines Signals, das mit einem Quantisierungsparameter  $Q$  quantisiert wird.

- $D(R)$ : Inverse Funktion zu  $R(D)$ .
- $E\{\dots\}$ : Erwartungswert des Arguments.
- FP\_ADD: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet Fließkomma- bzw. SIMD-Fließkomma-Addition- und -Subtraktionsbefehle.
- FP\_DATA\_STORE: Verarbeitungseinheit im Pentium 4 Prozessor: Speichert Fließkomma- und SIMD-Daten.
- FP\_DIV: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet Fließkomma- bzw. SIMD-Fließkomma-Divisions- und -Wurzelbefehle.
- FP\_MISC: Verarbeitungseinheit im Pentium 4 Prozessor: Platzhalter für weitere, zusätzliche Verarbeitungseinheiten.
- FP\_MOVE: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet Fließkomma- und SIMD-Transfers.
- FP\_MUL: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet Fließkomma- und SIMD-Fließkomma-Multiplikationsbefehle.
- $H$ : Entropie (Mittelwert des Informationsgehalts).
- $hb$ : Steuervariable für die Datenratenkontrolle FLEX.
- $H(j_2|j_1)$ : Bedingte Entropie eines Ereignisses  $j_2$ , wenn die Information über das Ereignis  $j_1$  bereits vorliegt.
- $H(j_2;j_1)$ : Mittlerer Transinformationsgehalt.
- $H(\mathbf{X};\mathbf{Y})|_{\mathbf{C}}$ : Mittlerer Transinformationsgehalt zwischen den Vektoren  $\mathbf{X}$  und  $\mathbf{Y}$  bei Verwendung des Codebuches  $\mathbf{C}$ .
- $i(j)$ : Informationsgehalt des Ereignisses  $j$ .
- $i(j_2|j_1)$ : Bedingter Informationsgehalt eines Ereignisses  $j_2$ , wenn die Information über das Ereignis  $j_1$  bereits vorliegt.
- $i(j_2;j_1)$ : Transinformationsgehalt ist die Reduktion an Informationsgehalt, wenn das Wissen über die statistischen Zusammenhänge zwischen  $j_1$  und  $j_2$  beim Auftreten von  $j_2$  ausgenutzt wird.
- $j$ : Zufälliges Ereignis von  $J$  möglichen Ereignissen.
- $J$ : Menge von möglichen Ereignissen.
- $K$ : Anzahl von Bildpunkten, Bildern oder Blocklänge eines Codes.
- $k_i$ : Hilfsvariable für Berechnungen auf Basis des jeweiligen Makroblocks  $i$ .
- $lf$ : Steuervariable für die Datenratenkontrolle FLEX.
- $m$ : Laufindex für Bildpixel.
- $m_{DCT}^2, m_{DCT_i}^2$ : Parameter für die Schätzung der Bildkomplexität, anstelle der Varianz  $\sigma^2$ . Bei Angabe des Index  $i$  wird der Bezug auf einen Makroblock  $i$  hergestellt. Der Doppelindex z.B. von  $m_{DCT_{nj}}$  adressiert dabei den jeweiligen Makroblock  $j$  im Bild  $n$  mit jeweils  $N$  Makroblöcken.
- $M$ : Anzahl von z.B. Bildern in einer Sequenz oder Makroblöcken in einem Bild.
- $MAD(x_R, y_R)$ : Mean Absolut Differences eines rechteckigen Bereichs, definiert durch  $x_R$  und  $y_R$ .
- $\max(a_1, \dots, a_n), \max$ : Ergebnis ist das größere Element der Liste  $a_1, \dots, a_n$  bzw. Maximal/Maximalwert.
- $\min(a_1, \dots, a_n), \min$ : Ergebnis ist das kleinste Element der Liste  $a_1, \dots, a_n$  bzw. Minimal/Minimalwert.



- $mk_i$ : Hilfsvariable für Berechnungen auf Basis des jeweiligen Makroblocks  $i$ .
- MMX\_ALU: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet SIMD-Ganzzahlen- und -Logikbefehle.
- MMX\_MISC: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet reziproke MMX-Berechnungen und auch einige Ganzzahlen-Befehle.
- MMX\_SHFT: Verarbeitungseinheit im Pentium 4 Prozessor: Bearbeitet Schiebe- und Rotationsbefehle für SIMD-Ganzzahlen.
- MOV: Move (IA-32 Befehl).
- MOVD: Move Doubleword (SIMD-Befehl).
- MOVDQA: Move Aligned Double Quadword (SIMD-Befehl).
- MOVDQU: Move Unaligned Double Quadword (SIMD-Befehl).
- $MSE(x_R, y_R)$ : Mean Squared Error eines rechteckigen Bereichs, definiert durch  $x_R$  und  $y_R$ .
- $n$ : Laufindex, z.B. für Bildpixel oder Bilder.
- $N$ : Anzahl von z.B. Bildern in einer Sequenz oder Makroblöcken in einem Bild.
- $nff$ : Anzahl der zukünftigen Bilder.
- $nsf$ : Anzahl der Bilder, die bezogen auf die aktuelle Framerate  $FR$  übersprungen werden.
- $P_x$ : Leistung des Bildsignals  $x$ .
- PADDB: Add Packed Byte Integers (SIMD-Befehl).
- PADDD: Add Packed Doubleword Integers (SIMD-Befehl).
- PADDW: Add Packed Word Integers (SIMD-Befehl).
- $p(j)$ : Wahrscheinlichkeit des Ereignisses  $j$ .
- $p(j_1, j_2)$ : Verbundwahrscheinlichkeit der Ereignisse  $j_1$  und  $j_2$ .
- $p(j_1|j_2)$ : Bedingte Wahrscheinlichkeit für die Wahrscheinlichkeiten  $p(j_1)$ ,  $p(j_2)$  der Ereignisse  $j_1$  und  $j_2$ .
- PSADBW: Compute Sum of Absolute Differences (SIMD-Befehl).
- $PSNR(x_R, y_R)$ : Peak Signal to Noise Ratio eines rechteckigen Bereichs, definiert durch  $x_R$  und  $y_R$ .
- PSRAD: Shift Packed Data Right Arithmetic Doubleword (SIMD-Befehl).
- PSRAW: Shift Packed Data Right Arithmetic Word (SIMD-Befehl).
- $Q_i$ : Quantisierung des Bildes  $i$  oder des Makroblocks  $i$  mit der Stufenhöhe  $Q_i$ . Abkürzend wird auch vom Quantisierungsparameter  $Q$ , Quantisierungsparameter der Höhe  $Q$  bzw. der Quantisierung  $Q$  gesprochen.
- $QP$ : Quantisierungsparameter der Stufenhöhe  $2QP = Q$  (siehe auch  $Q_i$ ).
- $qp\_mul\_mask$ : Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- $R$ : Rechteckiger Ausschnitt eines Bildes mit z.B.  $M \times N$  Pixel.
- $R, R'$ : Bitrate: eine konstante Bitrate wird dabei mit  $R$  und eine variable Bitrate mit  $R'$  gekennzeichnet.
- $R(D)$ : Kleinste mögliche Bitrate bei der vorgegebenen Verzerrung  $D$ .
- $R(Q)$ : Bitrate für ein Signal, das mit einem Quantisierungsparameter der Höhe  $Q$  quantisiert wird.

- $R_{uv}$ : Bitrate für die Codierung des DCT-Koeffizient  $c_{uv}$ .
- reciprodiv\_cor\_mask*: Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- reciprodiv\_mask*: Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- REF\_FR*: Referenz-Framerate.
- $S_i$ : Mittlere absolute Differenz *MAD*, z.B. für das  $i$ -te Bild oder den  $i$ -ten Makroblock.
- $SAD(x_R, y_R)$ : Sum of Absolut Differences eines rechteckigen Bereichs, definiert durch  $x_R$  und  $y_R$ .
- s\_add\_mask*: Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- sign\_mask*: Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- $SSD(x_R, y_R)$ : Sum of Squared Differences eines rechteckigen Bereichs, definiert durch  $x_R$  und  $y_R$ .
- s\_sub\_mask*: Bezeichnung einer Bitmaske für die SIMD-Implementierung.
- SUB**: Subtract (IA-32 Befehl).
- $\Delta t$ : Pufferverzögerung.
- $\Delta T$ : Verzögerung zwischen Puffer im Encoder bzw. dem Puffer im Decoder.
- $t_i$ : Modellparameter pro Makroblock  $i$ .
- $t_n$ : Zeitpunkt des  $n$ -ten Bildes bezogen auf die Referenzframerate bzw. Zeitpunkt an dem das  $n$ -te Bild vom Puffer entfernt wird.
- $t_{REF}, t_{FR}$ : Zeitlicher Abstand zwischen zwei Bildern bei der Referenz-Framerate *REF\_FR* (Index *REF*) bzw. bei einer bestimmten Framerate (Index *FR*).
- $\overline{tm}$ : Mittelwertberechnung über  $M$  Bilder mit jeweils  $N$  Makroblöcken des Modellparameters  $t_{lj}$  und des Parameters für die Schätzung der Bildkomplexität  $m_{DCT_{lj}}$  mit dem Schleifenindex  $j$  für den jeweiligen Makroblock und dem Index  $l$  für das jeweilige Bild:
- $$\overline{tm} = \frac{1}{M} \sum_{l=0}^{M-1} \sum_{j=0}^{N-1} t_{lj} \cdot m_{DCT_{lj}}.$$
- $x, x(m, n)$ : Bezeichnungen für Bilder, Bildausschnitte, Bildpunkte oder ein Bildsignal.
- X**: Bezeichner für zu einem Vektor zusammengefasster Bildpunkte.
- $X, x$ : Bezeichner für einen Modellparameter bzw. ein Register.
- $y, y(m, n)$ : Bezeichnungen für Bilder, Bildausschnitte, Bildpunkte oder ein Bildsignal.
- Y**: Bezeichner für zu einem Vektor zusammengefasster Bildpunkte.
- $Y, y$ : Bezeichner für ein Register.
- $\alpha_i$ : Gewichtungsfaktor für den Makroblock  $i$ .
- $\epsilon$ : Hilfsvariable.
- $\sigma^2, \sigma$ : Varianz eines Signals ( $\sigma^2$ ) bzw. dessen Standardabweichung ( $\sigma$ ).
- $\mu_x$ : Mittelwert des Bildsignals  $x$ .

## Literaturverzeichnis

- [1] Advanced Micro Devices, Inc., [Online]: <http://www.amd.com>.
- [2] S. M. Akramullah, I. Ahmad, M. L. Liou, "Optimization of H.263 video encoding using a single processor computer: performance tradeoffs and benchmarking", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 8, pp. 901 - 915, Aug. 2001.
- [3] S. Aramvith, I.-M. Pao, M.-T. Sun, "A rate-control scheme for video transport over wireless channels", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 5, pp. 569 - 580, Mai 2001.
- [4] ARM Holdings plc., [Online]: <http://www.arm.com>.
- [5] Bluetooth Special Interest Group, [Online]: <http://www.bluetooth.org>.
- [6] C. van den Branden Lambrecht, O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system", *Proceedings of the SPIE*, Vol. 2668, *Digital Video Compression: Algorithms and Technologies 1996*, pp. 450 - 461, März 1996.
- [7] D. Brash, "The ARM architecture version 6 (ARMv6)", *ARM White Paper*, Jan. 2002, [Online]: <http://www.arm.com>.
- [8] I. N. Bronstein, K. A. Semendjajew, "Taschenbuch der Mathematik", ISBN 3-322-00259, Teubner Verlag, Leipzig, 1989.
- [9] D. C. Brown, "A bitrate control algorithm for the berkeley MPEG-1 video encoder", [Online]: <http://bmrc.berkeley.edu/research/publications/1997/144/dbrown.html>.
- [10] J.-J. Chen, H.-M. Hang, "Source model for transform video coder and its application - part II: variable frame rate coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 2, pp. 299 - 311, Apr. 1997.
- [11] T. Chiang, Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 1, pp. 246 - 250, Feb. 1997.
- [12] J. Choi, D. Park, "A stable feedback control of the buffer state using the controlled Lagrangian multiplier method", *IEEE Transactions on Image Processing*, Vol. 3, No. 5, pp. 546 - 557, Sept. 1994.
- [13] J. J. Chen, D. W. Lin, "Optimal bit allocation for coding video signals over ATM networks", *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 6, pp. 1002 - 1015, Aug. 1996.
- [14] G. Cote, B. Erol, M. Gallant, F. Kossentini, "H.263+: video coding at low bit rates", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 7, pp. 849 - 866, Nov. 1998.
- [15] W. Ding, "Joint encoder and channel rate control of VBR video over ATM networks", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 2, pp. 266 - 278, Feb. 1997.
- [16] W. Ding, B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 1, pp. 12 - 19, Feb. 1996.
- [17] W. Ding, B. Liu, "Rate-quantization modeling for rate control of MPEG video coding and recording", *Proceedings of SPIE Volume: 2419, Digital Video Compression: Algorithms and Technologies 1995*, pp. 139 - 150, April 1995.

- [18] DivXNetworks, Inc., [Online]: <http://www.divx.com>.
- [19] E. D. Frimout, J. Biemond, R. L. Lagendik, "Forward rate control for MPEG recording", Proceedings of SPIE Volume: 2094, Visual Communications and Image Processing 1993, pp. 184 - 194, Okt. 93.
- [20] B. Furht, "Processor architectures for multimedia: a survey", Proceedings of Multimedia Modeling Conference, World Scientific, pp. 89 - 109, Singapore, Nov. 1997.
- [21] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding", IEEE Transactions on Image Processing, Vol. 9, No. 2, pp. 173 - 183, Feb. 2000.
- [22] H. Gish, J. N. Pierce, "Asymptotically efficient quantizing", IEEE Transactions on Information Theory, Vol. 14, No. 5, pp. 676 - 683, Sep. 1968.
- [23] D. D. Giusto, M. Pinna, C. Perra, "Bit allocation strategies for quality control in H.263 video transmission", COST 254 Workshop, Bordeaux, Frankreich, 23. - 24. März 2000.
- [24] R. M. Gray, "Entropy and information theory", Stanford University/CA, Information Systems Laboratory, Department of Electrical Engineering, 2000, [Online]: <http://ee.stanford.edu/~gray/it.html>.
- [25] R. M. Gray, L. D. Davisson, "Introduction to statistical signal processing", Stanford University/CA, Information Systems Laboratory, Department of Electrical Engineering, 2002, [Online]: <http://ee.stanford.edu/~gray/sp.html>.
- [26] R. H. M. Hafez, G. R. Rajugopal, "Adaptive rate controlled robust video communications over packet wireless networks", ACM/Baltzer Mobile Networks and Applications Journal, Vol. 3, No. 1, pp. 33 - 47, 1998.
- [27] H. M. Hang, J. J. Chen, "Source model for transform video coder and its application - part I: fundamental theory", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 2, pp. 287 - 298, Apr. 1997.
- [28] G. Hauske, "Systemtheorie der visuellen Wahrnehmung", ISBN 3-519-06156-2, B.G. Teubner Verlag, Stuttgart, 1994.
- [29] Z. He, Y. K. Kim, S. K. Mitra, "Low-delay rate control for DCT video coding via  $\rho$ -domain source modeling", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 8, pp. 928 - 940, Aug. 2001.
- [30] J. L. Hennessy, D. A. Patterson, "Computer Architecture A Quantitative Approach", ISBN 1558600698, Morgan Kaufmann Publishers, Inc., 1990.
- [31] Home of the XviD codec, [Online]: <http://www.xvid.org>.
- [32] Homepage der Moving Picture Experts Group (MPEG), [Online]: <http://www.chiariglione.org/mpeg/index.htm>.
- [33] C. Hsu, A. Ortega, "A Lagrangian optimization approach to rate control for delay-constrained video transmission over burst-error channels", ICASSP'98, Seattle, USA, 12 - 15 Mai 1998.
- [34] C. Hsu, A. Ortega, M. Khansari, "Rate control for robust video transmission over burst-error wireless channels", IEEE Journal on selected Areas in Communications, Vol. 17, No. 5, pp. 756 - 773, Mai 1999.
- [35] C. Hsu, A. Ortega, M. Khansari, "Rate control for robust video transmission over wireless channels", Proceedings of Visual Communication and Image Processing, VCIP 1997, San Jose, USA, Feb. 1997.
- [36] C. Y. Hsu, A. Ortega, A. R. Reibman, "Joint selection of source and channel rate for VBR transmission under ATM policing constraints", IEEE Journal on selected Areas in Communications, Vol. 15, No. 6, pp. 1016 - 1028, Aug. 1997.
- [37] H. Hutzelmann, J. U. Klöcking, M. Zeller, H. Steckenbiller, R. Knorr, "Communication network, method for transmitting a signal, network connecting unit and method for adjusting the bit rate of a scaled data flow", Patentanmeldung WO 99/40753, 12. Aug. 1999.

- [38] H. Hutzemann, M. Zeller, J. U. Klöcking, H. Steckenbiller, "Intelligentes Schnittverfahren mit einer einfachen Bewegungserkennung für die 4:3/16:9 Bildformatkonvertierung", Fernseh- und Kino-Technik, Nr. 4/1998, S. 206 - 208, April 1998.
- [39] Intel, "A fast precise implementation of 8x8 discrete cosine transform using the streaming SIMD extensions and MMX instructions version 1.0", Intel Bestellnummer: 742474-001, 1999.
- [40] Intel, "Block-matching in motion estimation algorithms using streaming SIMD extensions 2 (SSE2) version 2.0", Intel Bestellnummer: 248605-001, Juli 2000.
- [41] Intel, "IA-32 Intel architecture optimization reference manual", Intel Bestellnummer: 248966-009, 2003.
- [42] Intel, "Intel architecture MMX technology in business applications", Intel Bestellnummer: 243367-002, Juni 1997.
- [43] Intel, "Intel Pentium 4 processor extended instruction set techniques - taking advantage of streaming SIMD extension 2 (SSE2) technology", [Online]: <http://www.intel.com>.
- [44] Intel, "Software development strategies for streaming SIMD extensions version 2.1", Intel Bestellnummer: 243648-002, 1999.
- [45] Intel, "The IA-32 Intel architecture software developer's manual, volume 1: basic architecture", Intel Bestellnummer: 245470-012, 2003.
- [46] Intel, "The IA-32 Intel architecture software developer's manual, volume 2: instruction set reference", Intel Bestellnummer: 245471-112, 2003.
- [47] Intel, "The IA-32 Intel architecture software developer's manual, volume 3: system programming guide", Intel Bestellnummer: 245472-212, 2003.
- [48] Intel, "Using streaming SIMD extensions 2 (SSE2) in motion compensation for video decoding and encoding version 2.0", Intel Bestellnummer: 248607-001, Juli 2000.
- [49] Intel, "Using streaming SIMD extensions 2 (SSE2) to implement an inverse discrete cosine transform version 2.0", Intel Bestellnummer: 248670-001, Juli 2000.
- [50] Intel, "Using streaming SIMD extensions in a fast DCT algorithm for MPEG encoding version 1.2", Intel Bestellnummer: 243651-002, Jan. 1999.
- [51] Intel, "Using streaming SIMD extensions in a motion estimation algorithm for MPEG encoding version 1.2", Intel Bestellnummer: 243652-002, Jan. 1999.
- [52] Intel Corporation, [Online]: <http://www.intel.com>.
- [53] Intel Software College, [Online]: <http://www.intel.com/software/products/college/ia32/sse2/sse2lic.htm>.
- [54] International Telecommunication Union, [Online]: <http://ftp3.itu.int/av-arch/video-site/sequences>.
- [55] ISO/IEC 11172 "Information technology – coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s", Teil 1 - 5, 1993 - 1999.
- [56] ISO/IEC 13818, "Information technology – generic coding of moving pictures and audio", Teil 1 - 10, 1996 - 2000.
- [57] ISO/IEC 14496, "Information technology – coding of audio-visual objects", Teil 1 - 17, 1999 - 2003.
- [58] ISO/IEC JTC1/SC29/WG11, "Contribution to rate control Q2 experiment: a quantizer control tool for achieving target bit rates accurately", Coding of Moving Pictures and Associated Audio, M1812, Feb. 1997.
- [59] ISO/IEC JTC1/SC29/WG11, "Descriptions of rate control and texture coding core experiments on coding efficiency in MPEG-4 video", Coding of Moving Pictures and Associated Audio, M2258, Juli 1997.

- [60] ISO/IEC JTC1/SC29/WG11, "Detailed algorithm technical description for ITU-T VCEG draft H.26L algorithm in response to video and DCinema CfPs", Coding of Moving Pictures and Associated Audio, M7512, Juli 2001.
- [61] ISO/IEC JTC1/SC29/WG11, "Evaluation methods and procedures for MPEG-4 tests (rev. 3)", Coding of Moving Pictures And Associated Audio, N1557, Feb.1997.
- [62] ISO/IEC JTC1/SC29/WG11, "Levels for MPEG-4 visual profiles", Coding of Moving Pictures and Associated Audio, N4507, Dez. 2001.
- [63] ISO/IEC JTC1/SC29/WG11, "Main results of the AVC complexity analysis", Coding of Moving Pictures and Associated Audio, N4964, Juli 2002.
- [64] ISO/IEC JTC1/SC29/WG11, "MPEG-4 profiles under consideration", Coding of Moving Pictures and Associated Audio, N5237, Okt. 2002.
- [65] ISO/IEC JTC1/SC29/WG11, "MPEG-4 testing and evaluation procedures document", Coding of Moving Pictures and Associated Audio, N999, Juli 1995.
- [66] ISO/IEC JTC1/SC29/WG11, "MPEG-4 video verification model version 18.0", Coding of Moving Pictures and Associated Audio, N3908, Jan. 2001.
- [67] ISO/IEC JTC1/SC29/WG11, "Optimization model version 2.0", Coding of Moving Pictures and Associated Audio Information, N3675, Okt. 2000.
- [68] ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-4 standard", Coding of Moving Pictures and Associated Audio, N4668, März 2002.
- [69] ISO/IEC JTC1/SC29/WG11, "Production of video test sequences", Coding of Moving Pictures and Associated Audio, N5109, Juli 2002.
- [70] ISO/IEC JTC1/SC29/WG11, "Requirements for AVC codec", Coding of Moving Pictures and Associated Audio, N4672, März 2002.
- [71] ISO/IEC JTC1/SC29/WG11, ITU-T SG16, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)", Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, März 2003.
- [72] ITU-T Recommendation H.261, "Video codec for audiovisual services at p x 64 kbit/s", März 1993.
- [73] ITU-T Recommendation H.263, "Video coding for low bit-rate communication", version 1, Nov. 1995, version 2, Jan. 1998.
- [74] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", Mai 2003.
- [75] ITU-T, Study Group 15, "Video codec test model, TMN8", Juni 1997.
- [76] ITU-T, Study Group 16, "Draft for H.263++, annexes U, V und W to recommendation H.263", Nov. 2000.
- [77] ITU-T, Study Group 16, "Recommended simulation conditions for H.26L", Jan. 2001.
- [78] ITU-T, Study Group 16, "Video codec test model, TMN11", Monterey, Finland, Feb. 1999.
- [79] V. Jung, H.-J. Warnecke, "Handbuch für die Telekommunikation", ISBN 3540427953, 2. Auflage, Springer Verlag, Berlin, 2002.
- [80] A. Kaup, "MPEG-Standards: Techniken und Entwicklungstrends", Fernseh- und Kino-Technik, Nr. 6/2001, S. 352 - 362, Juni 2001.
- [81] J.U. Klöcking, H. Hutzelmann, M. Zeller, H. Steckenbiller, R. Knorr, "Echtzeit-Multimediakommunikation in Computernetzen", ITG-Fachtagung: Internet - frischer Wind in der Telekommunikation, ITG-Fachbericht 153, VDE-Verlag, Berlin, Offenbach, 1998.
- [82] J. Konrad, "Visual communications of tomorrow: natural, efficient and flexible", IEEE Communications Magazin, Vol. 39, No. 1, pp. 126 - 133, Jan. 2001.

- [83] I. Kontoyiannis, "Model selection via rate-distortion theory", Conference on Information Sciences and Systems, Princeton Universität, USA, 15. - 17. März 2000.
- [84] P. M. Kuhn, G. Diebel, S. Herrmann, A. Keil, H. Mooshofer, A. Kaup, R. Mayer, W. Stechele, "Complexity and PSNR-comparison of several fast motion estimation algorithms for MPEG-4", Proceedings of SPIE Volume: 3460, Applications of Digital Image Processing XXI, pp. 486 - 499, Okt. 1998.
- [85] J. Lee, B. W. Dickinson, "Rate-distortion optimized frame type selection for MPEG encoding", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 3, pp. 501 - 510, Juni 1997.
- [86] B. K. Lee, L. K. John, "Implications of programmable general purpose processors for compression/encryption applications", 2002, [Online]: <http://citeseer.nj.nec.com/509431.html>.
- [87] J.-W. Lee, A. Wang, Y.-S. Ho, "Object-based rate allocation with spatio-temporal trade-offs", Proceedings of SPIE Volume: 4671, Visual Communications and Image Processing 2002, pp. 374 - 384, Jan. 2002.
- [88] A. Lewis, "MPEG-4 over wireless networks", ARM White Paper, 2000, [Online]: <http://www.arm.com>.
- [89] L. J. Lin, A. Ortega, "Bitrate control using piecewise approximated rate-distortion characteristics", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, No. 4, pp. 446 - 459, Aug. 1998.
- [90] L. J. Lin, A. Ortega, C. - C. J. Kuo, "Rate Control using spline-interpolated R-D characteristics", Proceedings VCIP, pp. 111 - 122, Orlando, USA, März 1996.
- [91] H. Liu, M. El Zarki, "Adaptive source rate control for real-time wireless video transmission", Mobile Networks and Applications 3, pp. 49 - 60, 1998.
- [92] K. Meyberg, P. Vachenaer, "Höhere Mathematik 1", ISBN 3540517987, Springer Verlag, Berlin, Heidelberg, 1990.
- [93] F. Moscheni, F. Dufaux, H. Nicolas, "Entropy criterion for optimal bit allocation between motion and prediction error information", Proceedings VCIP, pp. 604 - 612, Nov. 1993.
- [94] Motorola, Inc., [Online]: <http://www.motorola.com>.
- [95] A. N. Netravali, B. G. Haskell, "Digital pictures: representation, compression, and standards (applications of communications theory)", ISBN: 030644917X, Plenum Pub Corp., 2. Auflage, 1995.
- [96] J.- R. Ohm, "Digitale Bildcodierung", ISBN 3540585796, Springer Verlag, Berlin, Heidelberg, 1995.
- [97] A. Ortega, "Optimal bit allocation under multiple rate constraints", Proceedings of the Data Compression Conference, DCC 1996, Snowbird, USA, April 1996.
- [98] A. Ortega, M. Khansari, "Rate control for video coding over variable bit rate channels with applications to wireless transmission", Proceedings of the 2nd. IEEE International Conference on Image Processing ICIP 1995, Washington, USA, Okt. 1995.
- [99] A. Ortega, K. Ramchandran, M. Vetterli, "Optimal trellis-based buffered compression and fast approximations", IEEE Transaction on Image Processing, Vol. 3, No. 1, pp. 26 - 40, Jan. 1994.
- [100] C. Perra, M. Pinna, D. D. Giusto, "H.263+ rate control at fixed objectiv quality", 10th International Packet Video Workshop, Forte Village Resort, Sardinien, Italien, Mai 2000.
- [101] Peter Pirsch, "VLSI implementations for image communications", ISBN 0444887903, Elsevier Verlag, 1993.
- [102] J. G. Proakis, "Digital communications 4th edition", ISBN 0072321113, McGraw-Hill, 2000.
- [103] R. Rauschert, A. Kaup, "A simple multiple video objects rate control algorithm for MPEG-4 real time applications", ITG Conference on Source and Channel Coding, ITG-Fachbericht 159, pp. 237 - 244, München, 17. - 19. Jan. 2000.

- [104] E. C. Reed, F. Dufaux, "Constrained bit-rate control for very low bit-rate streaming-video applications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 7, pp. 882 - 889, Juli 2001.
- [105] J. Ribas-Corbera, S.- M. Lei, "A frame-layer bit allocation for H.263+", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 7, pp. 1154 - 1158, Okt. 2000.
- [106] J. Ribas-Corbera, S.- M. Lei, "Rate control in DCT video coding for low-delay communications", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 1, pp. 172 - 185, Feb. 1999.
- [107] J. Ribas-Corbera, D. L. Neuhoff, "On the optimal motion vector accuracy for block-based motion-compensated video coders", *Proceedings of SPIE Volume: 2668, Digital Video Compression: Algorithms and Technologies 1996*, pp. 302 - 314, März 1996.
- [108] J. Ribas-Corbera, D. L. Neuhoff, "Optimizing motion vector accuracy in block-based video coding", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 4, pp. 497 - 511, April 2001.
- [109] J. I. Ronda, M. Eckert, F. Jaureguizar, N. Garcia, "Rate control and bit allocation for MPEG-4", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 8, pp. 1243 - 1258, Dez. 1999.
- [110] E. Schrüfer, "Signalverarbeitung", ISBN 3446165630, Carl Hanser Verlag, München, 1991.
- [111] C. E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, Vol. 27, pp. 379 - 423 und 623 - 656, Okt. 1948.
- [112] H. Song, J. Kim, C. - C. J. Kuo, "Improved H.263+ rate control via variable frame rate adjustment and hybrid I-frame coding", [Online]: <http://citeseer.nj.nec.com/147236.html>.
- [113] H. Song, J. Kim, C.-C. J. Kuo, "Real-time encoding frame rate control for H.263+ video over the internet", *Signal Processing: Image Communication (Elsevier Science)*, Vol. 15, No. 1 - 2, pp. 127 - 148, Sept. 1999.
- [114] H. Song, C.-C. J. Kuo, "Rate control for low-bit-rate video via variable-encoding frame rates", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 4, pp. 512 - 521, April 2001.
- [115] W. Stechele, "Algorithmic complexity, motion estimation and a VLSI architecture for MPEG-4 core profile video codecs", *Proceedings of VLSI-TSA International Symposium on VLSI Technology, Systems and Applications, Taiwan*, 18. - 20. April 2001.
- [116] G. J. Sullivan, T. Wiegand, "Rate-distortion optimization for video compression", *IEEE Signal Processing Magazine*, Vol. 15, No. 6, pp. 74 - 90, Nov. 1998.
- [117] G. J. Sullivan, T. Wiegand, T. Stockhammer, "Using the draft H.26L video coding standard for mobile applications", [Online]: [http://www.lnt.e-technik.tu-muenchen.de/mitarbeiter/stockhammer/puplications/icip2001\\_h26l.doc](http://www.lnt.e-technik.tu-muenchen.de/mitarbeiter/stockhammer/puplications/icip2001_h26l.doc).
- [118] H. Sun, W. Kwok, M. Chien, C. H. J. Ju, "MPEG coding performance improvement by jointly optimizing coding mode decisions and rate control", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 3, pp. 449 - 458, Juni 1997.
- [119] B. Tao, H. A. Peterson, B. W. Dickinson, "A rate-quantization model for MPEG encoders", *Proceedings ICIP*, Vol. I, pp. 338 - 341, Santa Barbara, USA, Okt. 1997.
- [120] A. Vetro, H. Sun, "Joint rate control for coding multiple video objects", *Proceedings IEEE Workshop Multimedia Signal Processing*, pp. 181 - 186, Princeton, USA, Juni 1997.
- [121] A. Vetro, H. Sun, Y. Wang, "MPEG-4 rate control for multiple video objects", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 1, pp. 186 - 199, Feb. 1999.
- [122] The Video Quality Experts Group (VQEG), [Online]: <http://www.its.bldrdoc.gov/vqeg>.



- [123] H. Wang, S. Chang, "A highly efficient system for automatic face region detection in MPEG video", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 7, No. 4, pp. 615 - 628, Aug. 1997.
- [124] Q. Wang, Z. Xiong, F. Wu, S. Lim, "Optimal rate allocation for progressive fine granularity scalable video coding", *IEEE Signal Processing Letters*, Vol. 9, No. 2, pp. 33 - 39, Feb. 2002.
- [125] P. H. Westerink, R. Rajagopalan, C. A. Gonzales, "Two-pass MPEG-2 variable-bitrate encoding", *Journal of Research and Development, Digital multimedia technology*, Vol. 43, No. 4, IBM Bestellnummer: G322-0218, 1999.
- [126] T. Wiegand, "Digital Image Communication", Kurs der Technischen Universität Berlin, 2003, [Online]: <http://bs.hhi.de/~wiegand/DIC.html>.
- [127] T. Wiegand, M. Lightstone, D. Mukherjee, T. G. Campbell, S. K. Mitra, "Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, No. 2, pp. 182-190, April 1996.
- [128] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 560- 576, Juli 2003.
- [129] W. Xu, G. Hauske, "Subjektive Qualitätsschätzung datenkomprimierter und gestörter Bilder: Modellierung und Optimierung", *Proceedings of ITG Conference on Source and Channel Coding*, München, BRD, 26. - 28. Okt. 1994.
- [130] W. Xu, G. Hauske, P. Filip, M. Ruf, "Subjective rating of picture coding algorithms", *SPIE's Symposium on Visual Communications and Image Processing '95*, Taipei, Taiwan, 23. - 26. Mai 1995.
- [131] A. Y. K. Yan, M. L. Liou, "Adaptive predictive rate control algorithm for MPEG videos by rate quantization method", *Proceedings Picture Coding Symposium*, pp. 619 - 624, Berlin, BRD, Sept. 1997.
- [132] Y. Yang, S. S. Hemami, "Rate control for VBR video over ATM: simplification and implementation", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 9, pp. 1045 - 1058, Sep. 2001.
- [133] M. Zeller, W. Aicher, P. Kuhn, W. Stechele, "Hard- und Softwareaspekte der Videocodierung - Teil 2: Videoprozessoren - Prinzipien und Architekturen", *F&M 11/97*, Nov. 1997.
- [134] M. Zeller, M. Heidrich, "Adaptive Verfahren der Datenratenkontrolle für anwendungs- und netzübergreifende Videocodierung", 3. ITG-Fachtagung Netze und Anwendungen, Neue Kommunikations-Anwendungen in modernen Netzen, Duisburg, BRD, 2002.
- [135] M. Zeller, M. Heidrich, "Anwendungs- und netzübergreifende Videocodierung: Adaptive Verfahren für die Datenratenkontrolle", *ITG-Workshop Breitbandige Zugangsnetze und integrierte Inhausnetze*, Berlin, BRD, 7. Okt. 2002.
- [136] M. Zeller, M. Heidrich, R. Steffen, "Haus-Kommunikationssystem, Transportplattform für ein Haus-Kommunikationssystem und intelligentes Netzabschlußgerät für ein Haus-Kommunikationssystem", *Internationale Patentanmeldung PCT/EP00/07799*, Anmeldedatum 13.08.1999.
- [137] M. Zeller, H. Hutzelmann, J. U. Klöcking, H. Steckenbiller, "Transcodierung mit Schwerpunkt Video am Beispiel einer standardübergreifenden MCU für Breit- und Schmalbandkommunikation", *ITG-Fachbericht 144, Multimedia-Anwendungen, Technologie, Systeme*, 7. Dortmunder Fernsehseminar, Dortmund, BRD, 29. Sept. - 1.Okt. 1997.
- [138] M. Zeller, P. Kuhn, A. Hutter, W. Stechele, "Hard- und Softwareaspekte der Videocodierung - Teil 1: Prinzipien und Standards", *F&M 5/97*, Mai 1997.
- [139] M. Zhou, "Optimization of MPEG-2 video encoding", ISBN 3826528174, Shaker Verlag, Aachen, 1997.

