

Skalierbare Videocodierung mittels Teilbandzerlegung

Von dem Fachbereich Elektrotechnik und Informationstechnik

der Universität Hannover

zur Erlangung des Grades eines

DOKTORS DER ELEKTROTECHNIK

Dr.-Ing.

genehmigte Dissertation

von

Dipl.-Ing. Ulrich-Lorenz Benzler

geboren am 21. Januar 1966, in Hannover

2002

Referent: Prof. Dr.-Ing. H. G. Musmann

Korreferent: Prof. Dr.-Ing. K. Jobmann

Tag der Promotion: 2. August 2002

Kurzfassung

In der vorliegenden Arbeit wird ein bewegungskompensierendes Teilband-DCT-Verfahren für die örtlich skalierbare Videocodierung entwickelt.

Die hierarchische, örtlich skalierbare Videocodierung eröffnet die Möglichkeit, lediglich Teile des codierten und übertragenen Datenstroms separat mit Hilfe eines so genannten Baselayer-Decoders zu Bildern niedriger örtlicher Auflösung zu decodieren. Im Zusammenhang mit ungleichmäßigem Fehlerschutz ermöglicht die Skalierbarkeit eine stetige Abnahme der Bildqualität im Fall von Übertragungsfehlern, während anderenfalls ein vollständiger Signalverlust auftreten kann. Dies ist besonders in der Mobilkommunikation von Bedeutung, da sich hier die Kanaleigenschaften über kurze Zeitabschnitte wesentlich verschlechtern können. Die Skalierbarkeit verbessert auch die Einsetzbarkeit von Multimedia-Servern, da sie den Zugriff auf codierte Videodaten mit unterschiedlichen Datenraten erlaubt, ohne dass die Videodaten in mehreren Auflösungen vorgehalten werden müssen. Dies ist zum Beispiel bei der Anbindung mehrerer Benutzer über heterogene Netzwerke erforderlich.

Um das Problem einer erhöhten Anzahl von codierten Abtastwerten zu vermeiden, das bei allen bisher standardisierten örtlich skalierbaren Videocodierungsverfahren auftritt, benutzt das in der vorliegenden Arbeit vorgestellte Verfahren eine Zerlegung des Videosignals in vier Teilbänder. Diese Teilband-Signale werden anschließend mittels einer bewegungskompensierenden Hybrid-8x8-DCT-Codierung ähnlich dem standardisierten MPEG-2-Verfahren codiert. Das tieffrequente Teilband stellt dabei die niedrige örtliche Auflösung dar, die separat im Baselayer-Bitstrom übertragen wird. Für die niedrige örtliche Auflösung wird eine eigene bewegungskompensierende Prädiktion durchgeführt, während die Prädiktion für die hohe örtliche Auflösung sowohl aus dem tieffrequenten als auch aus den hochfrequenten Teilbändern gewonnen wird. Um eine Aufteilung der Gesamtdatenrate in je 50% für Baselayer- und Enhancementlayer-Bitstrom zu ermöglichen, wird eine zusätzliche SNR-Skalierung des Baselayersignals eingeführt.

Die Effizienz des entwickelten skalierbaren Teilband-DCT-Codierungsverfahrens wird anhand experimenteller Ergebnisse überprüft. Es zeigt sich, dass eine Datenratenerhöhung von weniger als 10% im Vergleich zu einer nichtskalierbaren Codierung mit dem Referenz-MPEG-2-Coder benötigt wird, während das standardisierte örtlich skalierbare MPEG-4-SSP-Verfahren zu einer Datenratenerhöhung von mehr als 40% führt.

Stichworte: Videocodierung, örtlich skalierbare Codierung, Teilbandcodierung

Abstract

In this thesis, a motion compensating subband-DCT coding scheme for spatial scalable video coding is presented.

Spatial scalable video coding allows that parts of the encoded and transmitted data can be decoded separately by a so-called base layer decoder to reconstruct low resolution images. Combined with unequal error protection, this scalability provides a graceful degradation of the picture quality in the case of transmission errors, which otherwise can result in a complete signal loss. This feature is especially important in mobile communications, where the channel conditions can degrade significantly over certain periods in time. Scalability also enhances multimedia server applications, as it provides access to the video at different bit rates without the need to store the video at different resolutions. This is necessary for a multimedia server connected to different users over a heterogeneous network.

To overcome the problem of an increased number of encoded samples that is inherent to all standardized spatial scalable video coding schemes, the approach described in this thesis uses a decomposition of the video signal into four subbands. The subband signals are encoded using a motion compensating hybrid 8x8 DCT coding similar to the standardized MPEG-2 scheme. The low frequency subband represents the low resolution images, and is separately coded into the base layer bit stream. A separate motion compensating prediction is used for the low resolution images, while the prediction for the high resolution images is carried out using both the low and the high frequency subbands. To allow a distribution of the total bit rate into 50% for each the base and the enhancement layer bitstream, an additional SNR scalability for the base layer signal is introduced.

The efficiency of the developed spatial scalable subband-DCT coding scheme is verified by experimental results. It is shown that the proposed scheme is able to achieve less than 10% bit rate increase compared with a non-scalable MPEG-2 reference encoder, whereas the standardized MPEG-4 Spatial Scalable Profile (SSP) scheme leads to an increase in bit rate of more than 40%.

Keywords: Video Coding, Spatial Scalability, Subband Coding

Inhaltsverzeichnis

1. Einleitung	1
2. Grundlagen der skalierbaren Videocodierung	9
2.1. Charakterisierung von Videosequenzen	9
2.2. Hybrid-Codierung	10
2.3. Örtliche Skalierbarkeit	13
2.3.1. Simulcast-Codierung	14
2.3.2. Pyramiden-Codierung	15
2.3.3. Teilband-Codierung	17
2.4. SNR-Skalierbarkeit	18
3. Problem zeilenverschränkter Eingangssignale	23
3.1. Codierung eines zeilenverschränkten Signals	23
3.2. Codierung eines progressiven Hilfssignals	26
4. Konzept der örtlich skalierbaren Teilband-DCT-Codierung	29
5. Entwurf der Teilbandfilter für die örtliche Skalierung	33
5.1. Filterbänke mit exakter Rekonstruktion	33
5.1.1. Verallgemeinerte Quadraturspiegelfilter (GQMF)	36
5.1.2. Zweidimensionale Teilbandzerlegung	37
5.2. Optimierungskriterium und Randbedingungen für den Entwurf . .	38
5.2.1. Festlegung des Toleranzschemas	39

5.2.2. Optimierung der Codierungseffizienz	39
6. Bewegungskompensation und Codierung der Teilbandsignale	45
6.1. Bewegungsschätzung und -kompensation	45
6.2. Vorverstärkung der hochfrequenten Teilbänder	48
6.3. Entwurf der Wichtungsmatrizen für die Quantisierer	48
7. Entwurf der Quantisierung und Codierung für die SNR-Skalierung	57
7.1. SNR-Skalierung des Baselayersignals	57
7.2. Effizienz der Bitplane-Codierung	58
7.3. Drift im SNR-Baselayer	60
8. Experimentelle Ergebnisse	63
8.1. Vergleich mit dem MPEG-Codierungsverfahren	63
8.2. Bewertung der Ergebnisse	65
9. Zusammenfassung	67
A. Gleichungssystem für GQMF-Filterbänke	75

Abkürzungen und Formelzeichen

MPEG	ISO/IEC SC29 WG11 (Moving Pictures Experts Group)
ITU	International Telecommunication Union
MCP	bewegungskompensierende Prädiktion (Motion Compensating Prediction)
ME	Bewegungsschätzung (Motion Estimation)
EC	Entropie-Codierung
ED	Entropie-Decodierung
(I)DCT	(Inverse) Diskrete Cosinus-Transformation
SSP	örtliche Skalierbarkeit in MPEG (Spatial Scalable Profile)
PCM	Puls-Code-Modulation
PSNR	Spitzen-Signal-Rausch-Abstand (Peak-Signal-to-Noise-Ration)
SIF	Source Input Format (352x288 Bildpunkte)
MTF	Modulations-Übertragungsfunktion
AF	Analysefilterbank
SF	Synthesefilterbank
GQMF	Verallgemeinerte Quadraturspiegelfilter (Generalized Quadrature Mirror Filters)
GOP	MPEG-Bildergruppe (Group of Pictures)
I-Frame	intra-codiertes Bild
P-Frame	prädiktiv codiertes Bild
B-Frame	bidirektional prädiktiv codiertes Bild
$I_{x,y,t}$	horizontales, vertikales bzw. zeitliches Abtastintervall
$\vec{n} = (n_x, n_y)$	Ortsvektor
k	Zeitindex der Bilder einer Videosequenz
$N \times M$	Blockgröße der Bewegungskompensation
$\vec{d} = (d_x, d_y)$	Displacementvektor
s	Videosignal
s_0	hochaufgelöstes Videosignal
s_1^\downarrow	niedrig aufgelöstes Baselayersignal (tieffrequentes Teilbandsignal)
$s_2^\downarrow, s_3^\downarrow, s_4^\downarrow$	hochfrequente Teilbandsignale
\hat{s}	Prädiktionssignal
\tilde{s}	rekonstruiertes Signal

e	Prädiktionsfehlersignal
\tilde{e}	quantisiertes Prädiktionsfehlersignal
q	Quantisierungsfehler
$h_{1,2}$	Teilband-Analysefilter
$g_{1,2}$	Teilband-Synthesefilter
$e_{1..4}^\downarrow$	Teilband-Prädiktionsfehlersignale
Q_B	Quantisierer für SNR-Baselayersignal
Q_E	Quantisierer für SNR-Enhancementlayersignal
e_B	Prädiktionsfehler des SNR-Baselayersignals
e_E	Prädiktionsfehler des SNR-Enhancementlayersignals
$\tilde{s}_{Dec,Base}$	rekonstruiertes SNR-Baselayersignal
$\hat{s}_{Dec,Base}$	Prädiktionsignal im SNR-Baselayer-Decoder
QP	MPEG-Quantisierungsparameter zur Datenratenregelung
$w_{i,j}$	Elemente der Quantisierer-Wichtungsmatrix
$\tilde{s}_{1,Base}^\downarrow$	rekonstruiertes Baselayer-Basissignal

1. Einleitung

Die Übertragung und Speicherung von Videosignalen benötigt relativ große Datenraten und Speicherkapazitäten. Ein digitales Fernsehsignal entsprechend dem Studioformat ITU-R 601 [1] benötigt im Fall einer "Pulse Code Modulation" (PCM) eine Datenrate von 166 Mbit/s. Ein hochaufgelöstes "High Definition Television"-Signal (HDTV) [2] benötigt sogar eine PCM-Datenrate von 885 Mbit/s. Zum Vergleich beträgt die PCM-Datenrate eines Sprachsignals 64 kbit/s.

Zur Reduktion der Datenrate werden so genannte Quellencodierungsverfahren [3] eingesetzt. Im Allgemeinen sind die gesamten codierten Daten zur Rekonstruktion des Videosignals notwendig. Die codierten Daten skalierbarer Codierungsverfahren bestehen aus einem so genannten Baselayer- und einem oder mehreren Enhancementlayer-Bitströmen. Durch Decodierung lediglich des Baselayer-Bitstroms erhält man ein niedrig aufgelöstes Signal. In diesem Fall ist die örtliche, zeitliche oder die Amplituden-Auflösung des rekonstruierten Signals reduziert. Die Enhancementlayer-Bitströme beinhalten die zur Rekonstruktion des hochaufgelösten Signals benötigten komplementären Informationen.

Skalierbare Videocodierung ist für unterschiedliche Anwendungen von Interesse. In mobilen Kommunikationsanwendungen kann damit ein hierarchisches Übertragungssystem realisiert werden, bei dem das Baselayersignal durch eine leistungsfähige Kanalcodierung des Baselayer-Bitstroms besonders gut gegen Übertragungsfehler auf dem Kanal geschützt ist. Im Fall schlechter Übertragungsbedingungen kann so eine drastische Verschlechterung der Bildqualität vermieden werden, indem nur der stark geschützte Baselayer-Bitstrom decodiert und dargestellt wird. Dies bezeichnet man als graduelle Verschlechterung (graceful degradation). Für Videoserveranwendungen erlaubt die skalierbare Codierung unterschiedlichen Benutzern den Zugriff mit unterschiedlichen Datenraten, ohne dass eine mehrfache Speicherung des Videosignals in unterschiedlichen Auflösungen notwendig ist.

Diese Anwendungen benötigen einen Baselayer- und einen oder mehrere Enhancementlayer-Bitströme. In der Literatur werden hauptsächlich Verfahren mit zwei Stufen (layers) beschrieben, und es werden Vorschläge gemacht, wie diese zu mehrstufigen Verfahren erweitert werden können.

Im Allgemeinen muss die Funktionalität der Skalierbarkeit durch einen Rückgang der Codierungseffizienz im Vergleich zu einer nichtskalierbaren Codierung bezahlt werden. Die Effizienz von skalierbaren Codierungsverfahren kann auf zwei unterschiedliche Arten bewertet werden:

- Anhand der Erhöhung der Datenrate verglichen mit einer nichtskalierbaren Codierung bei gleicher Bildqualität (Angabe der Bildqualität z.B. mittels Spitzen-Signal-Rausch-Abstand (PSNR) für das hochaufgelöste Signal)
- Anhand der Verminderung der Bildqualität (PSNR) verglichen mit einer nichtskalierbaren Codierung bei gleicher Datenrate

Die untere Grenze der Codierungseffizienz stellt die so genannte "Simulcast-Codierung" dar. Dabei wird die gesamte Datenrate in zwei Teilbitströme aufgespalten. Im ersten Teilbitstrom wird das Signal der niedrigen Auflösung, im zweiten Teilbitstrom das gesamte Signal der hohen Auflösung übertragen.

Im Rahmen dieser Arbeit werden skalierbare Codierungsverfahren entwickelt und untersucht, die eine zweistufige Rekonstruktion bei den örtlichen Auflösungen gemäß ITU-R 601 und gemäß dem "Source Input Format" (SIF) [6] ermöglichen.

Das in den "Spatial Scalable Profiles" von MPEG-2 [7] und MPEG-4 [9] standardisierte skalierbare Videocodierungsverfahren (MPEG-SSP) benutzt eine so genannte Pyramidencodierung. Zur Erzeugung eines niedrig aufgelösten Signals aus dem ursprünglich hochaufgelösten Signal wird eine Tiefpassfilterung mit anschließender Abtastratenreduktion verwendet. Das entstehende niedrig aufgelöste Signal bildet das Eingangssignal des Baselayers und wird unabhängig mittels einer bewegungskompensierenden Hybrid-DCT-Codierung [32] codiert. Das Prädiktionsfehlersignal des Baselayers und die Displacementvektordinformation werden im Baselayer-Bitstrom übertragen.

Als Eingangssignal für den Enhancementlayer benutzt die Pyramidencodierung das gesamte hochaufgelöste Signal. Um die Codierungseffizienz im Vergleich zu einer Simulcast-Codierung zu erhöhen, wird das rekonstruierte Baselayersignal zur Unterstützung der Codierung des hochaufgelösten Signals verwendet. Zur Anpassung an das Abtastraster des hochaufgelösten Signals wird es in der Abtastrate erhöht und tiefpassgefiltert. Das resultierende interpolierte Signal kann benutzt werden, um die Enhancementlayer-Prädiktion zu verbessern. Es gibt 3 Modi für die Erzeugung des Enhancementlayer-Prädiktionssignals: Mittels des interpolierten Baselayersignals ("örtliche Prädiktion"), mittels des bewegungskompensierten hochaufgelösten Signals ("zeitliche Prädiktion") oder mittels des Mittelwertes beider Signale ("örtlich-zeitliche Prädiktion"). Für jeden so genannten Makroblock von 16x16 Bildpunkten wird derjenige Modus ausgewählt, der

zum jeweils geringsten Prädiktionsfehler führt. Das resultierende Prädiktionsfehlersignal wird ebenfalls DCT-codiert und zusammen mit den Enhancementlayer-Displacementvektoren und der Information über die benutzten Prädiktionsmodi im Enhancementlayer-Bitstrom übertragen.

Im Vergleich zur nichtskalierbaren Codierung erhöht sich bei der Pyramidencodierung die Displacementvektorinformation und die Anzahl der Abtastwerte, die in den beiden Stufen übertragen werden. Aus diesem Grund ist die Codierungseffizienz von MPEG-SSP nur geringfügig besser als die einer Simulcast-Codierung. Nach [13] beträgt das PSNR für eine nichtskalierbare MPEG-2-Codierung der ITU-R 601 Testsequenz "mobile & calendar" 30,06 dB bei einer Datenrate von 8 Mbit/s. Für eine MPEG-2-SSP-Codierung, bei der 3 Mbit/s für den Baselayer- und 5 Mbit/s für den Enhancementlayer-Bitstrom verwendet werden, beträgt das PSNR des rekonstruierten Signals 28,44 dB. Eine unabhängige Simulcast-Codierung, die 3 Mbit/s für das Signal der niedrigen sowie 5 Mbit/s für das Signal der hohen Auflösung verwendet, führt auf ein PSNR von 27,74 dB.

In [19] wird ein verbessertes Pyramidencodierungsverfahren vorgestellt. Es benutzt eine hierarchische bewegungskompensierende Prädiktion (Motion Compensating Prediction, MCP) mit skalierbaren Displacementvektoren für Base- und Enhancementlayer. Damit wird die bei MPEG-SSP zu beobachtende Erhöhung der Displacementvektorinformation vermieden, da für die beiden Stufen nicht mehr zwei unabhängige Sätze von Displacementvektoren übertragen werden müssen. Zusätzlich wird zur Codierung des Prädiktionsfehlersignals die DCT durch eine Vektorquantisierung ersetzt.

Das Verfahren erreicht für die SIF-Testsequenz "coastguard" bei einer Datenrate von 1,63 Mbit/s ein PSNR von 33,26 dB [20], während eine nichtskalierbare MPEG-2-Codierung ein PSNR von 33,45 dB bereits bei 1,2 Mbit/s erreicht. Demnach wird für dieses skalierbare Codierungsverfahren eine Datenratenerhöhung von mehr als 30% verglichen mit einer nichtskalierbaren MPEG-2-Codierung benötigt.

Ein anderer Ansatz, die so genannte "Frequenz-Skalierbarkeit", wird in [21] vorgestellt. Bei diesem Verfahren wird die MCP nur für das hochaufgelöste Signal durchgeführt, und das resultierende Prädiktionsfehlersignal wird DCT transformiert. Die 8x8 DCT-Koeffizienten werden aufgeteilt: Die 4x4 tieffrequenten DCT-Koeffizienten werden zusammen mit der Displacementvektorinformation im Baselayer-Bitstrom, die übrigen hochfrequenten DCT-Koeffizienten als Komplementärinformation im Enhancementlayer-Bitstrom übertragen.

Das Baselayersignal wird rekonstruiert, indem eine 4x4 inverse DCT (IDCT) auf die 4x4 tieffrequenten Koeffizienten der 8x8 DCT angewendet wird. Zur Rekonstruktion des hochaufgelösten Signals werden die im Baselayer-Bitstrom übertragenen 4x4 tieffrequenten Koeffizienten zusammen mit den übrigen, im

Enhancementlayer-Bitstrom übertragenen Koeffizienten in einer 8x8 IDCT im Enhancementlayer-Decoder rücktransformiert.

Das Signal, welches zur MCP im Baselayer-Decoder verwendet wird, kann jedoch ausschließlich aus dem rekonstruierten niedrig aufgelösten Baselayersignal erzeugt werden. Da der Encoder das rekonstruierte hochaufgelöste Signal zur Erzeugung des Prädiktionsfehlersignals verwendet, kommt es zu einer so genannten Drift im rekonstruierten Baselayersignal [25]. Diese Drift vermindert das PSNR des Baselayers erheblich. Weiterhin entstehen Artefakte durch fehlende DCT-Basisbilder, da das Baselayersignal mittels einer 4x4 IDCT der tieffrequenten Koeffizienten einer 8x8 DCT erzeugt wird. Dies führt zu einer nicht akzeptablen Bildqualität im Baselayersignal [25].

Die Verwendung einer 3D-örtlich-zeitlichen Teilbandzerlegung wird in [16] und [17] vorgeschlagen. Im Fall starker Bewegung innerhalb der Sequenz ergeben sich bei der 3D-Teilbandzerlegung hohe Energien in den zeitlich hochfrequenten Teilbändern, die zu einer geringen Codierungseffizienz führen. Um diesem Problem entgegenzuwirken, wird in [17] eine "Kamera-Pan-Kompensation" angewandt. Das PSNR der Standard-MPEG-Testsequenz "flower garden" in SIF-Auflösung beträgt jedoch lediglich 24,79 dB bei einer Datenrate von 1,5 Mbit/s, während eine nichtskalierbare MPEG-1-Codierung ein PSNR von 25,43 dB bereits bei 1 Mbit/s erreicht. Für dieses skalierbare Codierungsverfahren wird demnach eine Datenratenerhöhung von etwa 50% benötigt.

Ein kürzlich in [22] veröffentlichtes Verfahren zur 3D-Teilbandzerlegung, welches eine bewegungskompensierende Prädiktion vor der zeitlichen Teilbandzerlegung benutzt, erreicht bereits eine Codierungseffizienz, die mit einer nichtskalierbaren MPEG-4-Codierung vergleichbar ist. So beträgt das PSNR für die Standard-MPEG-Testsequenz "mobile & calendar" in SIF-Auflösung 30,64 dB bei einer Datenrate von 2,4 Mbit/s, während eine nichtskalierbare MPEG-4-Codierung ein PSNR von 30,61 dB bei 2,28 Mbit/s erreicht. Somit benötigt dieses Verfahren eine lediglich 5%-10% höhere Datenrate bei gleicher Bildqualität.

Die angewandte fünfstufige zeitliche Oktavbandzerlegung benötigt jedoch 16 Bildspeicher, was gleichzeitig zu einer Codier-Verzögerung von mehr als 16 Bildern führt. Dies ist erheblich aufwendiger als die 3 Bildspeicher, die für die nichtskalierbare MPEG-Codierung benötigt werden.

Neben den Verfahren zur Skalierbarkeit der örtlichen Auflösung sind aus der Literatur auch Verfahren zur Skalierung der Amplitudenauflösung, der so genannten SNR-Skalierung, bekannt. Das standardisierte MPEG-2 "SNR scalable profile" [7] verwendet eine zweistufige Quantisierung der DCT-Koeffizienten. In der ersten Stufe werden grob quantisierte Koeffizienten erzeugt, die im Baselayer-Bitstrom übertragen werden. Jeder der rekonstruierten grob quantisierten Koeffizienten wird anschließend vom jeweiligen Originalwert subtrahiert. Diese Dif-

ferenz, die den Quantisierungsfehler der ersten Stufe darstellt, wird in der zweiten Stufe nochmals feiner quantisiert und im Enhancementlayer-Bitstrom übertragen. Beide Stufen verwenden die gleichen Huffman-Tabellen zur Codierung der DCT-Koeffizienten. Diese Tabellen sind an die statistischen Eigenschaften des Baselayersignals angepasst. Da in der zweiten Stufe jedoch der Quantisierungsfehler der ersten Stufe codiert wird, unterscheiden sich die statistischen Eigenschaften dieses Signals erheblich von denen des Baselayersignals. Dies hat einen deutlichen Verlust an Codierungseffizienz zur Folge.

In der laufenden Standardisierung von MPEG-4 wird die zweistufige Quantisierung von MPEG-2 durch eine inhärent skalierbare "Bitplane"-Codierung ersetzt. Diese ist eine Realisierung der so genannten "embedded quantization" [15]. Es wurde gezeigt, dass diese Technik die Codierungseffizienz für zweistufige Systeme kaum vermindert. Ziel der laufenden Untersuchungen bei MPEG-4 ist es, die Codierungseffizienz auch für eine größere Stufenanzahl zu verbessern.

Zusammenfassend weisen die in der Literatur veröffentlichten Verfahren zur örtlichen Skalierung eine Codierungseffizienz auf, die deutlich unterhalb einer entsprechenden nichtskalierbaren Codierung liegt. Eine Ausnahme bildet die bewegungskompensierte 3D-Teilbandzerlegung, die jedoch zu einer erheblich aufwendigeren Realisierung führt. Für die SNR-Skalierung erweist sich das in MPEG-4 vorgeschlagene Verfahren der Bitplane-Codierung zumindest im Fall von zwei SNR-Stufen als effizient.

Ziel dieser Arbeit ist es, ein effizientes skalierbares Codierungsverfahren mit zwei Stufen örtlicher Auflösung zu entwickeln, das den Anstieg der Datenrate bei gleicher Bildqualität auf weniger als 10% verglichen mit einer nichtskalierbaren MPEG-2-Codierung begrenzt. Um verschiedene Anwendungen unterstützen zu können, soll die Aufteilung der Datenrate möglichst flexibel sein und eine Baselayer-Datenrate von bis zu 50% der Gesamtdatenrate erlauben [27].

Um die mit der Pyramidencodierung verbundene Erhöhung der Anzahl der Abtastwerte zu vermeiden, soll eine 4-Band-Teilbandzerlegung mit exakter Rekonstruktion des hochaufgelösten Eingangssignal verwendet werden. Dabei soll das tieffrequente Teilband im Baselayer- und die hochfrequenten Teilbänder im Enhancementlayer-Bitstrom übertragen werden. Sowohl im Baselayer als auch im Enhancementlayer soll eine bewegungskompensierende Hybrid-DCT-Codierung ähnlich dem MPEG-2-Standard verwendet werden, um eine hohe Codierungseffizienz bei vertretbarem Realisierungsaufwand sicherzustellen.

Zur Entwicklung und Realisierung dieses skalierbaren Codierungsverfahrens sind die folgenden drei Probleme zu lösen:

Das erste Problem betrifft den Entwurf der Teilbandfilter. Für den Filterentwurf ergeben sich zwei Forderungen: Zum Einen soll das Tiefpassfilter, welches das Signal für den Baselayer generiert, eine gute subjektive Bildqualität erzeugen. Zum Anderen soll die Teilbandfilterbank eine hohe Codierungseffizienz des gesamten Teilband-Codierungsverfahrens gewährleisten. Dem ersten Gesichtspunkt trägt das von MPEG vorgeschlagene SIF-Filter [8] Rechnung. Mit Festlegung des Tiefpassfilters sind jedoch durch die Bedingung der exakten Rekonstruktion bereits alle verbleibenden Filter vollständig bestimmt.

Die Tiefpassfilterkoeffizienten müssen so gewählt werden, dass sie beiden Forderungen genügen. Zur Bewahrung der subjektiven Bildqualität werden nur solche Tiefpassfilter zugelassen, deren Übertragungsfunktionen ein vorgegebenes Toleranzschema des MPEG-SIF-Filters einhalten, wobei die Impulsantwortlänge wie beim MPEG-SIF-Filter auf sieben begrenzt werden soll. Anschließend soll die Codierungseffizienz anhand der erzeugten Datenrate der codierten Teilbandsignale optimiert werden.

Das zweite Problem betrifft die für Base- und Enhancementlayer vorgesehene MPEG-2-Codierung. Sie muss für die Verarbeitung von Teilbandsignalen angepasst werden, um eine effiziente Codierung zu erreichen.

Normalerweise benötigen die verwendeten MPEG-2-Hybrid-DCT-Codierungen für Base- und Enhancementlayer jeweils eine eigene bewegungskompensierende Prädiktion mit zugehöriger eigener Displacementvektordinformation, um eine Drift zwischen Sender und Empfänger zu vermeiden [25]. Es soll untersucht werden, ob der damit verbundene Anstieg der Datenrate für die Displacementvektordinformation vermieden werden kann, indem die gleiche Information in Base- und Enhancementlayer verwendet wird.

Eine wichtige Anpassung der Codierung betrifft die Quantisiererwichtungsmatrizen. Um die so genannte Modulations-Übertragungsfunktion (MTF) [28] des menschlichen visuellen Systems zur Datenreduktion zu nutzen, verwendet das standardisierte MPEG-2-Verfahren Quantisiererwichtungsmatrizen für die DCT-Koeffizienten. Da die Teilbandsignale jedoch nur noch Teilspektren des Gesamtsignals enthalten, sind diese Matrizen nicht mehr gültig. Folglich müssen entsprechende Wichtungsmatrizen für die Teilband-DCT-Koeffizienten basierend auf den Referenz-MPEG-Matrizen bestimmt werden.

Das dritte Problem betrifft die Forderung einer flexiblen Datenratenaufteilung zwischen Base- und Enhancementlayer-Bitstrom. Die Aufteilung der Datenrate auf die tief- und hochfrequenten Teilbänder und damit auf Base- und Enhancementlayer-Bitstrom ist aber durch die oben beschriebene Optimierung der Filterbank sowie die statistischen Eigenschaften des Eingangssignals festgelegt. Um dennoch eine flexible Aufteilung der Datenrate zu erlauben, soll

eine zusätzliche SNR-Skalierung für das Baselayersignal eingeführt werden. Auf diese Weise können Teile der Information des tieffrequenten Teilbandes im Enhancementlayer-Bitstrom übertragen werden.

Die Codierungseffizienz des entwickelten Verfahrens soll letztlich für die Bewertung sowohl mit der oberen Grenze einer nichtskalierbaren Codierung als auch mit der unteren Grenze einer Simulcast-Codierung verglichen werden, wobei das MPEG-2 TM6-Verfahren [8] als Referenz verwendet wird.

In Abschnitt 2 der Arbeit werden die Grundzüge skalierbarer Codierungsverfahren vorgestellt. Der Schwerpunkt liegt dabei auf Verfahren zur örtlichen und zur SNR-Skalierung.

Das Problem zeilenverschränkter Eingangssignale in örtlich skalierbaren Codierungsverfahren wird in Abschnitt 3 dargestellt.

Abschnitt 4 beschreibt das entwickelte Konzept der örtlich skalierbaren Codierung, bestehend aus einer Kombination von Teilbandzerlegung und bewegungskompensierender Hybrid-DCT-Codierung. Der Entwurf der dafür benötigten Teilbandfilterbank ist Inhalt von Abschnitt 5.

Abschnitt 6 behandelt die Modifikationen des MPEG-2-Verfahrens zur Bewegungskompensation und Codierung der Teilbandsignale.

Die zur flexiblen Datenraten-Aufteilung benötigte SNR-Skalierung ist in Abschnitt 7 beschrieben.

In Abschnitt 8 werden die experimentellen Ergebnisse vorgestellt und bewertet.

2. Grundlagen der skalierbaren Videocodierung

2.1. Charakterisierung von Videosequenzen

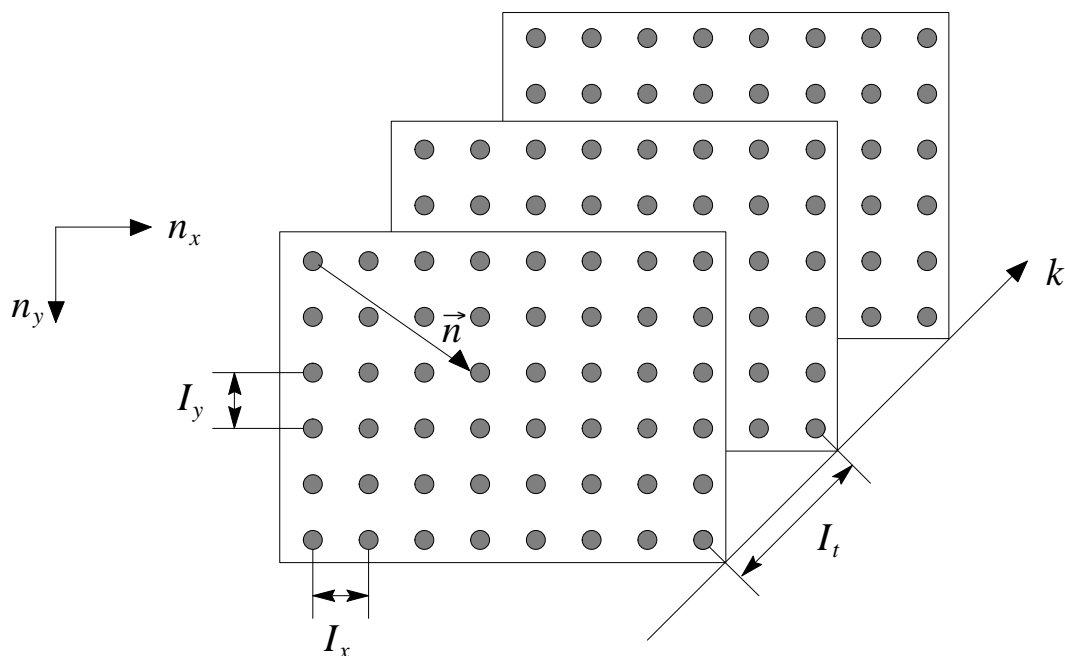


Abbildung 2.1.: Darstellung des örtlichen und zeitlichen Abtastrasters eines Videosignals

Nach der Abtastung kann ein kontinuierliches Videosignal $s = s(x, y, t)$ durch die diskrete Signaldarstellung $s = s(n_x I_x, n_y I_y, k I_t)$ beschrieben werden. Dabei bezeichnen I_x, I_y und I_t das horizontale, vertikale sowie das zeitliche Abtastintervall (siehe Abb. 2.1). Jeder Abtastwert beschreibt die Luminanz-Amplitude eines Bildpunktes an einer bestimmten Stelle des dreidimensionalen Abtastrasters.

Die ganzzahligen Koordinaten n_x, n_y, k sind dimensionslose Werte bezüglich der physikalischen Einheiten I_x, I_y und I_t . In Zusammenhang mit der Quellencodierung kann die Beschreibung des Signals s mit Hilfe der normalisierten Darstellung $s = s(\vec{n}, k)$ vereinfacht werden. Der Ortsvektor $\vec{n} = (n_x, n_y)$ und die Koordinate k beschreiben die horizontale, vertikale und zeitliche Position innerhalb des Abtastrasters. Für einen variablen Ortsvektor \vec{n} und einen festen Zeitpunkt k beschreibt $s = s(\vec{n}, k)$ die Abtastwerte eines einzelnen Bildes. Durch Variation von k erhält man eine Videosequenz.

Horizontal und vertikal benachbarte Abtastwerte besitzen einen Abstand von einem Bildpunkt, so dass $\delta n_x = \delta n_y = 1$ bezüglich der normalisierten Darstellung. Die Längenangabe "1 Bildpunkt" korrespondiert also zu der physikalischen Länge I_x in horizontaler und zu I_y in vertikaler Richtung. Entsprechend korrespondieren die Angaben $1/2$ Bildpunkt und $1/4$ Bildpunkt jeweils zu den physikalischen Längen $I_x/2, I_y/2$ und $I_x/4, I_y/4$. Die Angabe einer Blockgröße $N \times M$ beschreibt einen rechteckigen Block innerhalb eines Bildes mit insgesamt $N \cdot M$ horizontal und vertikal benachbarten Bildpunkten.

2.2. Hybrid-Codierung

Ein grundlegendes Verfahren zur Quellencodierung von Videosequenzen ist die bewegungskompensierende Hybrid-Codierung. Sie ist die Grundlage für zahlreiche standardisierte Codierungsverfahren der Bereiche Video-Kommunikation (ITU-T H.261 [4], ITU-T H.263 [5]), Multimedia (ISO-MPEG-1 [6]) und digitales Fernsehen (ISO-MPEG-2 [7], ISO-MPEG-4 [9]). Obwohl alle diese Standards Unterschiede im Detail aufweisen, bauen sie gleichwohl auf dem Prinzip der bewegungskompensierenden Hybrid-Codierung auf (siehe Abb. 2.2). Sie ist eine Kombination von Prädiktion und Intraframe-Codierung.

Die Prädiktion erzeugt ein Prädiktionssignal $\hat{s}(\vec{n}, k)$ für das aktuelle Signal $s(\vec{n}, k)$. Für Videosequenzen wird eine spezielle bewegungskompensierende Prädiktion (Motion Compensating Prediction, MCP) verwendet. Die MCP nutzt die statistischen Abhängigkeiten einer Videosequenz in zeitlicher Richtung zur Datenreduktion. In Abb. 2.3 ist sie unter der Annahme translatorisch bewegter Blockinhalte dargestellt.

Jedes Bild wird in rechteckige Blöcke der Größe $N \times M$ Bildpunkte unterteilt. Zur Prädiktion eines Blocks mit statischem Bildinhalt im Bild k wird ein Block der gleichen örtlichen Position im rekonstruierten Bild $\tilde{s}(\vec{n}, k - 1)$ verwendet. Ist Bewegung in der Sequenz enthalten, wird ein um den Displacementvektor $\vec{d} = (d_x, d_y)$ verschobener Block zur Prädiktion benutzt. Im Allgemeinen existiert für jeden Bildpunkt mit dem Ortsvektor \vec{n} ein Displacementvektor $\vec{d}(\vec{n})$. Es wird

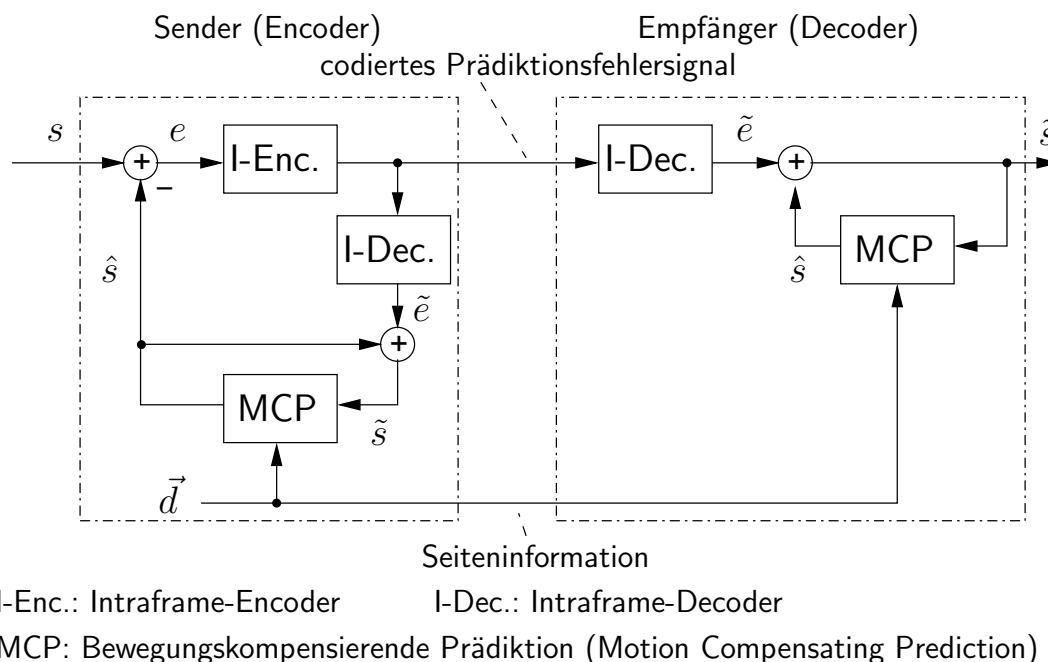


Abbildung 2.2.: Blockschaltbild der bewegungskompensierenden Hybrid-Codierung

angenommen, dass für alle Bildpunkte innerhalb eines Blocks derselbe Vektor gültig ist und nur ein Displacementvektor je Block übertragen wird. Der Displacementvektor beschreibt die Bild-zu-Bild-Verschiebung und ermöglicht die bewegungskompensierende Prädiktion, so dass

$$\hat{s}(\vec{n}, k) = \tilde{s}(\vec{n} - \vec{d}(\vec{n}), k - 1) \quad (2.1)$$

gilt. Die Gesamtheit aller Displacementvektoren $\vec{d}(\vec{n})$ wird als Displacementvektorfeld bezeichnet.

Zur Ausführung der bewegungskompensierenden Prädiktion benötigt die MCP in Abb. 2.2 einen Bildspeicher für das bereits übertragene und rekonstruierte Bild $\tilde{s}(\vec{n}, k - 1)$.

Im Anschluss an die Prädiktion wird der verbleibende Prädiktionsfehler $e = s - \hat{s}$ berechnet. In einer Hybrid-Codierung wird dieser Prädiktionsfehler von einer so genannten Intraframe-Codierung verarbeitet. Aus den codierten und übertragenen Daten rekonstruiert der Decoder zuerst das Prädiktionsfehlersignal \tilde{e} im Intraframe-Decoder. Dabei führt die Verwendung einer Quantisierung in der Intraframe-Codierung zu einem Quantisierungsfehler $q = e - \tilde{e}$.

Durch Addition des Prädiktionssignals \hat{s} zum rekonstruierten Prädiktionsfehlersignal \tilde{e} wird das rekonstruierte Signal $\tilde{s} = \hat{s} + \tilde{e}$ erzeugt (siehe Abb. 2.2). Dabei

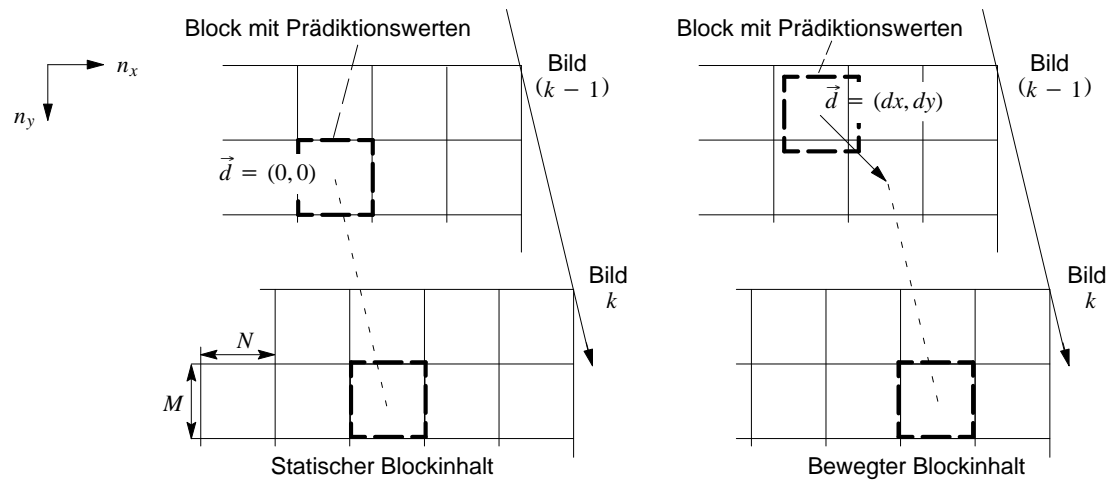


Abbildung 2.3.: Darstellung der blockweisen bewegungskompensierenden Prädiktion (MCP) für statischen und bewegten Blockinhalt

verwenden sowohl Encoder als auch Decoder dasselbe Prädiktionsignal \hat{s} . Folglich ergibt sich der Rekonstruktionsfehler des Videosignals ausschließlich durch den Quantisierungsfehler der Intraframe-Codierung, d. h. $s - \tilde{s} = e - \tilde{e} = q$. Grundlage für den Entwurf einer Intraframe-Codierung ist die Rate-Distortion-Theorie [31]. Ein Beispiel für eine Intraframe-Codierung ist in Abb. 2.4 dargestellt.

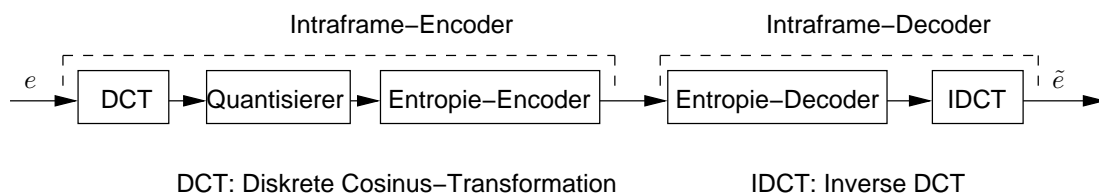


Abbildung 2.4.: DCT-Intraframe-Codierung

Zur Ausnutzung der statistischen Abhängigkeiten innerhalb eines Bildes wird eine zweidimensionale diskrete Cosinus-Transformation (DCT) mit Blöcken der Größe 8×8 Bildpunkte angewendet. Es entstehen 8×8 Transformationskoeffizienten, die weitgehend dekorreliert sind. Bei der anschließenden Quantisierung wird die begrenzte Auflösungsfähigkeit des menschlichen visuellen Systems ausgenutzt, die durch die so genannte Modulations-Transfer-Funktion (MTF) beschrieben ist und eine gröbere Quantisierung der hochfrequenten DCT-Koeffizienten erlaubt. Quantisierungsfehler, die nicht zu sichtbaren Störungen

im rekonstruierten Signal führen, werden als irrelevant zugelassen [32]. Eine redundanzreduzierende Entropie-Codierung der quantisierten DCT-Koeffizienten ermöglicht eine weitere Datenreduktion. Für die praktische Implementierung werden dabei in erster Linie die Huffman-Codierung [34] und die arithmetische Codierung [35] verwendet.

Die Hybrid-Codierung nach Abb. 2.2 verändert nicht die örtliche und zeitliche Auflösung des Videosignals. Das Original-Signal s und das rekonstruierte Signal \tilde{s} besitzen dasselbe Abtastgitter und unterscheiden sich, bedingt durch die Quantisierung, lediglich in ihren Amplitudenwerten.

2.3. Örtliche Skalierbarkeit

Die codierten Daten skalierbarer Codierungsverfahren bestehen aus einem Baselayer- und einem oder mehreren Enhancementlayer-Bitströmen. Durch Decodierung lediglich des Baselayer-Bitstroms erhält man ein niedrig aufgelöstes Signal, bei dem die örtliche, zeitliche oder die Amplituden-Auflösung des rekonstruierten Signals reduziert ist. Dies wird als örtliche, zeitliche bzw. SNR-Skalierung bezeichnet. Die Enhancementlayer-Bitströme beinhalten die kompletären Informationen, die in Kombination mit dem Baselayer-Bitstrom die Rekonstruktion des hochaufgelösten Signals ermöglichen.

In diesem und dem nachfolgenden Abschnitt werden die Grundlagen der örtlichen und der SNR-Skalierung behandelt.

Es wird ein zweistufiges örtlich skalierbares Codierungsverfahren betrachtet, das ein hochaufgelöstes Signal s_0 und ein Baselayersignal s_1^\downarrow mit halber örtlicher Auflösung in horizontaler und vertikaler Richtung zur Verfügung stellt, siehe Abb. 2.5. Dabei wird zunächst ein Eingangssignal in progressivem Format betrachtet (vgl. Abschnitt 3).

Durch die Verwendung eines linearen, zeitinvarianten Tiefpassfilters geht das Signal s_1 aus der Faltung des hochaufgelösten Signals s_0 mit der Impulsantwort h des Tiefpassfilters hervor,

$$s_1 = h * s_0. \quad (2.2)$$

Die Abtastwerte sind durch

$$s_1(\vec{n}, k) = \sum_{\vec{l}} h(\vec{l}) \cdot s_0(\vec{n} - \vec{l}, k) \quad (2.3)$$

beschrieben. Nach einer Abtaststratenreduktion von 2:1 in horizontaler und vertikaler Richtung ergibt sich das niedrig aufgelöste Baselayersignal zu

$$s_1^\downarrow = (h * s_0)^\downarrow. \quad (2.4)$$

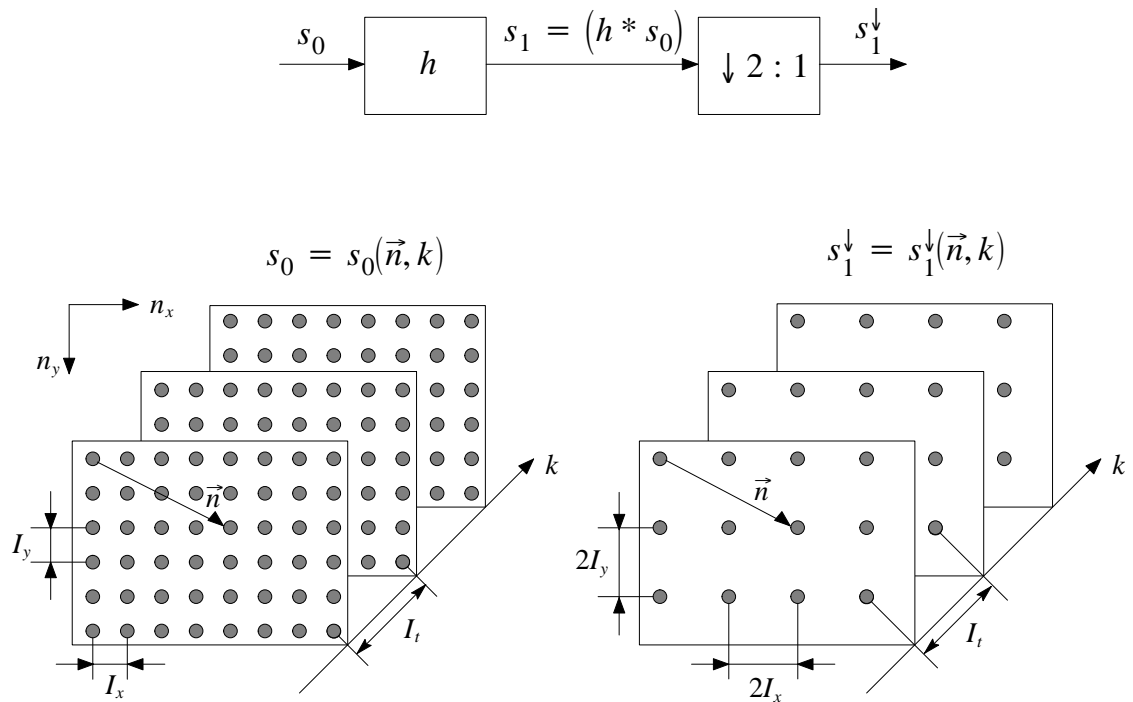
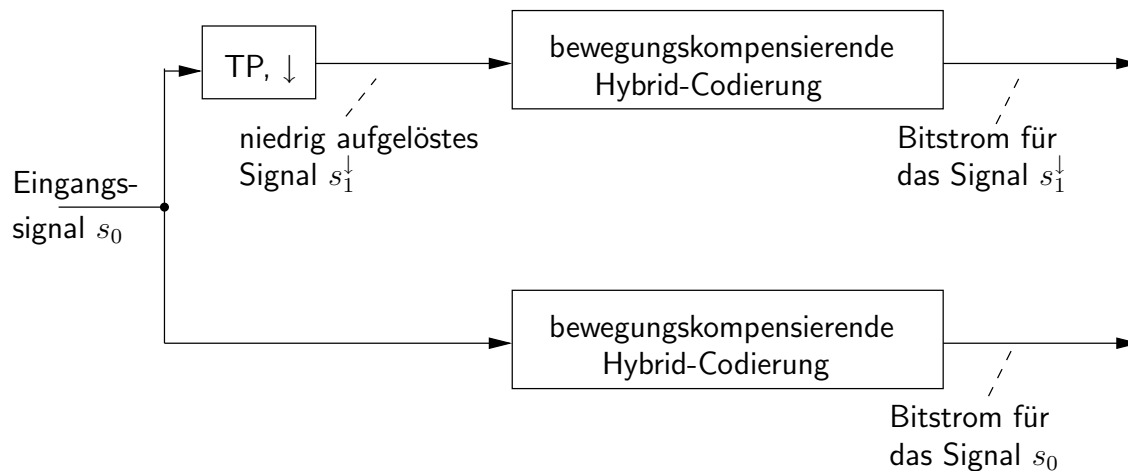


Abbildung 2.5.: Abtastraster des hochaufgelösten Signals s_0 und des niedrig aufgelösten Baselayersignals s_1^\downarrow im Fall progressiver Eingangssignale

Im Unterschied zur nichtskalierbaren Codierung erfordert die skalierbare Codierung des Signals s_0 immer eine separate Verarbeitung des niedrig aufgelösten Baselayersignals s_1^\downarrow . Diese separate Verarbeitung ist mit Einbußen in der Codierungseffizienz verbunden. Das Ziel dieser Arbeit ist es, ein örtlich skalierbares Codierungsverfahren zu entwickeln, das trotz dieser Forderung möglichst geringe Einbußen in der Codierungseffizienz bewirkt. Nachfolgend werden verschiedene aus der Literatur bekannte Lösungen und Ansätze zur örtlichen Skalierung vorgestellt.

2.3.1. Simulcast-Codierung

Es ist möglich, die beiden Signale s_0 und s_1^\downarrow getrennt zu codieren, indem zwei unabhängige Encoder entsprechend Abb. 2.2 verwendet werden (siehe Abb. 2.6). Dieses Verfahren wird als Simulcast-Codierung bezeichnet [36].



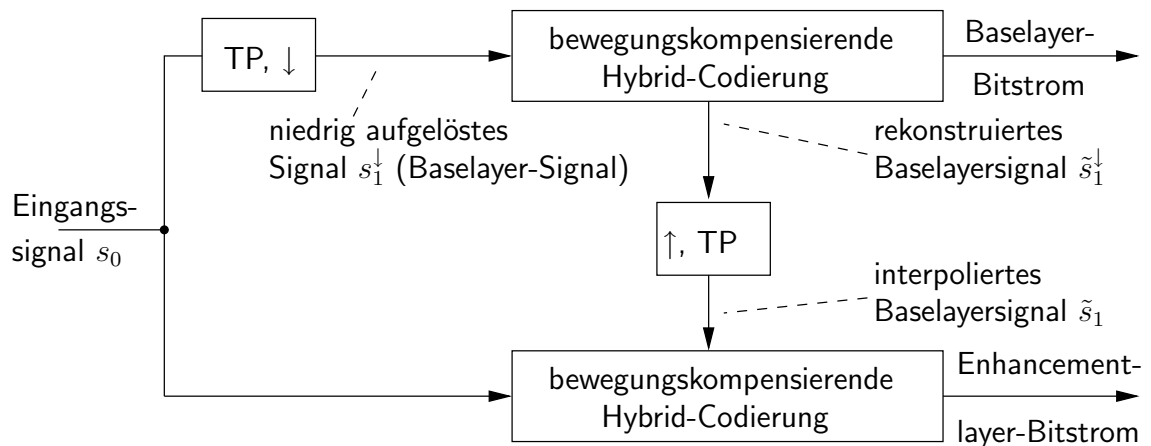
TP: Tiefpass-Filter ; ↓: Abtastratenreduktion

Abbildung 2.6.: Blockschaltbild des Encoders einer Simulcast-Codierung

Dabei wird zur Codierung des hochaufgelösten Signals s_0 die Datenrate $R_{SC}^{(0)}$ benötigt, die der Datenrate $R_{NS}^{(0)}$ einer nichtskalierbaren Codierung des Signals s_0 entspricht. Das niedrig aufgelöste Signal s_1^{\downarrow} benötigt eine zusätzliche Datenrate $R_{SC}^{(1)}$. Als Seiteninformation müssen die Displacementvektoren \vec{d}_1 und \vec{d}_0 übertragen werden. Verglichen mit einer nichtskalierbaren Codierung des Signals s_0 erhöht sich die Datenrate des Simulcast-Verfahrens also um den Anteil $R_{SC}^{(1)}$, der zur Codierung des Signals s_1^{\downarrow} zusätzlich übertragen werden muss, $R_{SC} = R_{SC}^{(0)} + R_{SC}^{(1)}$. Da das niedrig aufgelöste Signal s_1^{\downarrow} auch im hochaufgelösten Signal s_0 enthalten ist, wird es zweimal codiert und vermindert auf diese Weise die Codierungseffizienz.

2.3.2. Pyramiden-Codierung

Das in den "Spatial Scalable Profiles" von MPEG-2 [7] und MPEG-4 [9] standardisierte örtlich skalierbare Videocodierungsverfahren (MPEG-SSP) benutzt eine so genannte Pyramidencodierung, um die Codierungseffizienz im Vergleich zu einer Simulcast-Codierung zu erhöhen. Das Blockschaltbild ist in Abb. 2.7 dargestellt. Wie bei der Simulcast-Codierung wird zur Erzeugung des niedrig aufgelösten Signals s_1^{\downarrow} aus dem ursprünglich hochaufgelösten Signal s_0 eine Tiefpassfilterung mit anschließender Abtastratenreduktion entsprechend Abb. 2.5



TP: Tiefpass-Filter ; ↓: Abtastratenreduktion ; ↑: Abtastratenerhöhung

Abbildung 2.7.: Blockschaltbild des Encoders einer Pyramidencodierung entsprechend dem MPEG Spatial Scalable Profile (SSP)

verwendet. Das Signal s_1^\downarrow bildet das Eingangssignal des Baselayer-Encoders und wird unabhängig mittels einer bewegungskompensierenden Hybrid-DCT-Codierung [32] codiert.

Das Prädiktionsfehlersignal \tilde{e}_1^\downarrow des Baselayers und die Displacementvektorinformation \vec{d}_1 (vgl. Abb. 2.2) werden im Baselayer-Bitstrom übertragen.

Als Eingangssignal für den Enhancementlayer-Encoder benutzt die Pyramidencodierung ebenfalls das gesamte hochaufgelöste Signal s_0 . Um die Codierungseffizienz im Vergleich zur Simulcast-Codierung zu erhöhen, wird das rekonstruierte Baselayersignal \tilde{s}_1^\downarrow zur Unterstützung der Codierung des hochaufgelösten Signals s_0 verwendet. Zur Anpassung an das Abtastraster des hochaufgelösten Signals wird es in der Abtastrate erhöht und mit Hilfe eines Tiefpassfilters horizontal und vertikal interpoliert. Das resultierende Signal \tilde{s}_1 wird benutzt, um die Enhancementlayer-Prädiktion zu verbessern. In MPEG-SSP gibt es drei Modi für die Erzeugung des zu codierenden Enhancementlayer-Prädiktionsfehlersignals e_0 : Mittels des interpolierten Baselayersignals \tilde{s}_1 ("örtliche Prädiktion"), mittels des bewegungskompensierten hochaufgelösten Signals \hat{s}_0 ("zeitliche Prädiktion") oder mittels des Mittelwertes beider Signale ("zeitlich-örtliche Prädiktion"). Für jeden Makroblock von 16×16 Bildpunkten wird derjenige Modus ausgewählt, der zum jeweils geringsten Prädiktionsfehler führt. Das resultierende Prädiktionsfehlersignal e_0 wird wie das Baselayer-Prädiktionsfehlersignal \tilde{e}_1^\downarrow codiert

und zusammen mit den Enhancementlayer-Displacementvektoren \vec{d}_0 und der Information über die benutzten Prädiktionsmodi im Enhancementlayer-Bitstrom übertragen.

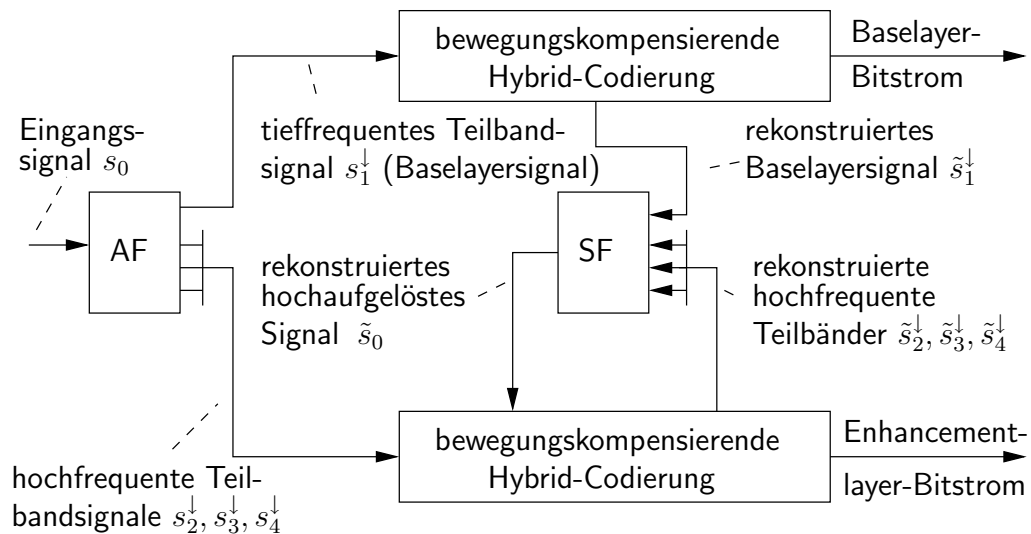
Im Vergleich zur nichtskalierbaren Codierung erhöht sich bei der Pyramidencodierung die Displacementvektorinformation und die Gesamtanzahl der Abtastwerte, die in den beiden Stufen übertragen werden. Aus diesem Grund ist die Codierungseffizienz von MPEG-SSP trotz der Unterstützung der Enhancementlayer-Codierung durch das interpolierte Baselayersignal nur geringfügig besser als die einer Simulcast-Codierung [13].

2.3.3. Teilband-Codierung

Die Erhöhung der Anzahl der zu codierenden Abtastwerte kann durch die Anwendung einer Teilbandzerlegung auf das Signal s_0 verhindert werden [37]. Abb. 2.8 zeigt ein zweistufiges Teilbandcodierungsverfahren, das eine Zerlegung des Signals in vier örtliche Teilbänder verwendet. Das Eingangssignal s_0 wird durch horizontale und vertikale Analysefilterung in die Teilbandsignale s_1^\downarrow , s_2^\downarrow , s_3^\downarrow und s_4^\downarrow zerlegt. Die Abtastrate der Teilbandsignale wird um den Faktor 2:1 sowohl in horizontaler als auch in vertikaler Richtung reduziert, so dass die Anzahl der Abtastwerte im Signal s_0 identisch zur Summe der Abtastwerte der vier Teilbandsignale ist. Das Baselayersignal s_1^\downarrow wird getrennt von den hochfrequenten Teilbändern s_2^\downarrow , s_3^\downarrow und s_4^\downarrow codiert. Durch Wahl des Tiefpass-Analysefilters zu $h_1 = h$ kann das Baselayersignal s_1^\downarrow identisch zum Baselayersignal der Pyramidencodierung (siehe Abb. 2.7) gewählt werden. In diesem Fall rekonstruieren die Baselayer-Decoder der Pyramiden- und der Teilbandcodierung identische Signale \tilde{s}_1^\downarrow .

Wie bei der Pyramidencodierung wird die komplementäre Information zur Rekonstruktion des hochaufgelösten Signals \tilde{s}_0 im Enhancementlayer-Bitstrom übertragen. Dieser enthält jedoch bei der Teilbandcodierung lediglich die drei Teilband-Prädiktionsfehlersignale \tilde{e}_2^\downarrow , \tilde{e}_3^\downarrow und \tilde{e}_4^\downarrow . Eine Synthese-Filterbank kombiniert das rekonstruierte Baselayersignal \tilde{s}_1^\downarrow mit den rekonstruierten hochfrequenten Teilbandsignalen \tilde{s}_2^\downarrow , \tilde{s}_3^\downarrow und \tilde{s}_4^\downarrow , um das rekonstruierte hochaufgelöste Signal \tilde{s}_0 zu gewinnen. Als Seiteninformation sind die Displacementvektoren \vec{d}_1 und \vec{d}_0 zu übertragen.

Im Vergleich zur Pyramidencodierung führt die Teilbandcodierung zu einer engeren Kopplung von Base- und Enhancementlayer. Da das rekonstruierte Baselayersignal \tilde{s}_1^\downarrow über die Synthesefilterbank direkt in die Rekonstruktion des hochaufgelösten Signals \tilde{s}_0 eingeht, müssen die Quantisierer der Teilbandsignale auf-



AF: Analysefilterbank ; SF: Synthesefilterbank

Abbildung 2.8.: Blockschaltbild des Encoders einer bewegungskompensierenden Teilbandcodierung

einander abgestimmt werden. Deshalb muss die Quantisierung des Baselayer-Prädiktionsfehlersignals e_1^\downarrow so eingestellt werden, dass sowohl das rekonstruierte Baselayersignal \tilde{s}_1^\downarrow als auch das rekonstruierte hochaufgelöste Signal \tilde{s}_0 die jeweiligen zulässigen Rekonstruktionsfehler nicht überschreiten.

Im Unterschied zur Teilbandcodierung erlaubt die Pyramidencodierung eine vollständig unabhängige Quantisierung der Prädiktionsfehlersignale e_0 und e_1^\downarrow .

2.4. SNR-Skalierbarkeit

Das Blockschaltbild eines SNR-skalierbaren bewegungskompensierenden Hybrid-Encoders und der zugehörigen Decoder ist in Abb. 2.9 dargestellt. Im Encoder wird das Prädiktionsfehlersignal e mittels eines ersten Quantisierers Q_B quantisiert, wodurch ein grob quantisiertes SNR-Baselayersignal \tilde{e}_B entsteht. Dieses Signal wird vom ursprünglichen Prädiktionsfehlersignal e subtrahiert; die Differenz wird mittels eines zweiten Quantisierers Q_E quantisiert und bildet das SNR-Enhancement-Signal \tilde{e}_E . Beide Signale werden einer Entropie-Codierung unterzogen und im SNR-Base- bzw. SNR-Enhancementlayer-Bitstrom zum

Empfänger übertragen.

Die Rekonstruktion des Signals \tilde{s} baut auf einem Prädiktionsfehlersignal $\tilde{e} = \tilde{e}_B + \tilde{e}_E$ auf, welches nur geringe Quantisierungsfehler enthält. Diese geringen Quantisierungsfehler sind das Ergebnis der zweistufigen Quantisierung Q_B gefolgt von Q_E . Verglichen mit einer nichtskalierbaren Codierung unterscheidet sich die Rekonstruktion von \tilde{s} lediglich dadurch, dass anstelle einer einstufigen eine zweistufige Quantisierung und Codierung durchgeführt wird. Folglich vermindert sich die Codierungseffizienz der SNR-skalierbaren Codierung nicht im Vergleich zu einer nichtskalierbaren Codierung, wenn die zweistufige Quantisierung und Entropiecodierung die gleiche Codierungseffizienz erzielt wie eine einstufige.

Allerdings baut im SNR-Baselayer-Decoder die Rekonstruktion des Signals $\tilde{s}_{Dec,Base}$ auf dem Signal \tilde{e}_B auf, welches große Quantisierungsfehler enthält, die durch Q_B erzeugt werden. Dadurch unterscheiden sich auch die Prädiktionssignale $\hat{s}_{Dec,Base}$ und \hat{s} . Dieser Unterschied wird als so genannte Drift der Prädiktionssignale bezeichnet. Die Drift kann sich über der Zeit, d.h. durch wiederholtes Durchlaufen der Prädiktionsschleife, akkumulieren. Sie vermindert die Bildqualität im rekonstruierten Baselayersignal $\tilde{s}_{Dec,Base}$, kann aber z.B. durch abschnittswise Neustart der Codierung beschränkt werden.

Im standardisierten MPEG-2 "SNR scalable profile" (MPEG-2 SNR) erzielt jedoch auch das rekonstruierte Signal \tilde{s} nicht die Codierungseffizienz einer nichtskalierbaren Codierung. Dies liegt an der bei MPEG-2 SNR verwendeten zweistufigen Quantisierung der DCT-Koeffizienten. In der ersten Stufe werden grob quantisierte Koeffizienten erzeugt, die im SNR-Baselayer-Bitstrom übertragen werden. Jeder der rekonstruierten grob quantisierten Koeffizienten wird anschließend vom jeweiligen Originalwert subtrahiert. Diese Differenz, die den Quantisierungsfehler der ersten Stufe darstellt, wird in der zweiten Stufe nochmals feiner quantisiert und im SNR-Enhancementlayer-Bitstrom übertragen. Beide Stufen verwenden die gleichen Huffman-Tabellen zur Codierung der DCT-Koeffizienten. Diese Tabellen sind an die statistischen Eigenschaften des SNR-Baselayersignals angepasst. Da in der zweiten Stufe jedoch der Quantisierungsfehler der ersten Stufe codiert wird, unterscheiden sich die statistischen Eigenschaften des SNR-Enhancement-Signals erheblich von denen des SNR-Baselayersignals. Dies hat einen Verlust an Codierungseffizienz zur Folge.

In der laufenden Standardisierung von MPEG-4 "Fine Granularity Scalability" (FGS) [11] wird die zweistufige Quantisierung von MPEG-2 durch eine inhärent skalierbare "Bitplane"-Codierung [11] ersetzt. Sie ist eine Realisierung der so genannten "embedded quantization" [15], die eine gemeinsame Quantisierung und Codierung von \tilde{e}_B und \tilde{e}_E erlaubt. Dabei wird das Signal vor der Codierung als Dualzahl dargestellt. Die Dualstellen gleicher Potenz mehrerer Koeffizienten werden zu einer so genannten Bitplane zusammengefasst, die mit Hilfe ei-

ner Lauflängencodierung codiert wird. Mit jeder übertragenen Bitplane wird der Quantisierungsfehler halbiert. Es wurde gezeigt, dass diese Technik die Codierungseffizienz für Systeme mit lediglich einem SNR-Enhancementlayer-Bitstrom kaum vermindert [11].

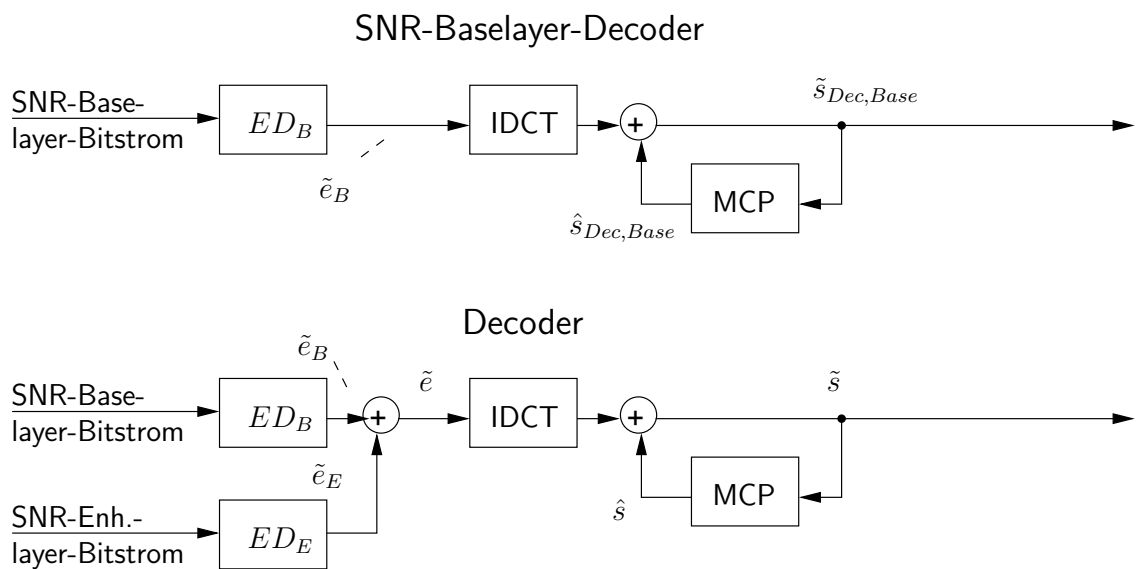
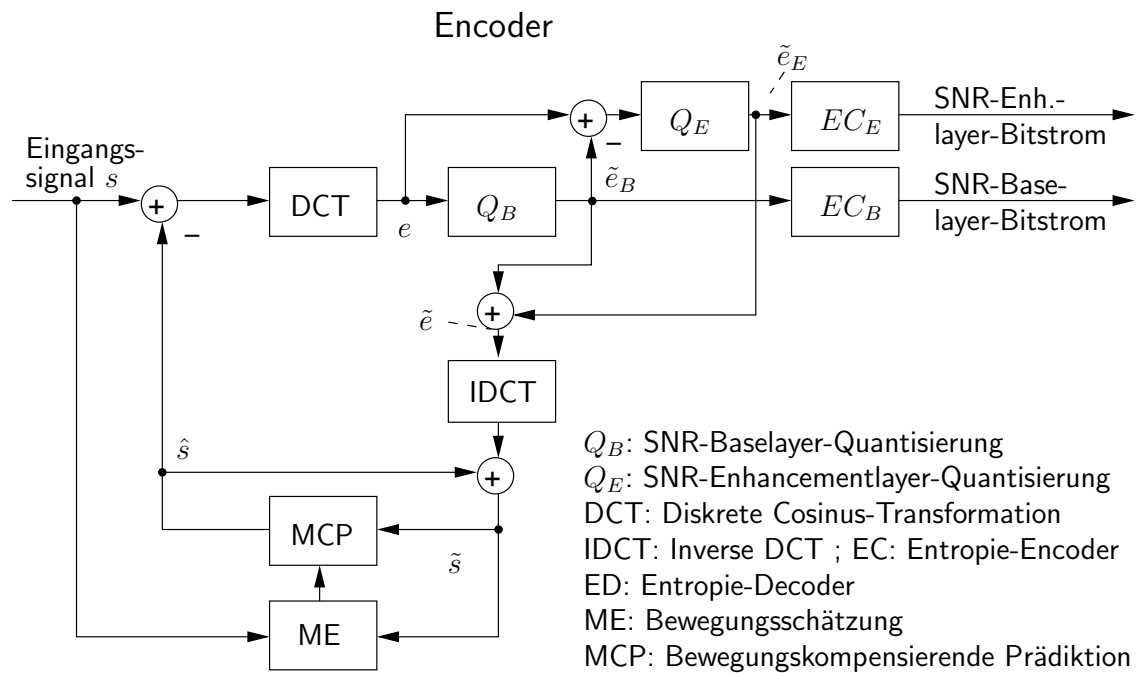


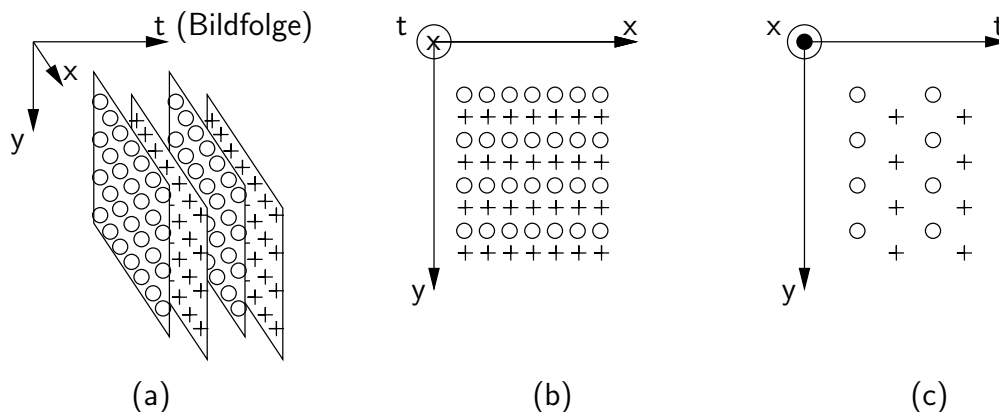
Abbildung 2.9.: Allgemeines Blockschaltbild eines SNR-skalierbaren bewegungskompensierenden Hybrid-Encoders und der zugehörigen Decoder

3. Problem zeilenverschränkter Eingangssignale

3.1. Codierung eines zeilenverschränkten Signals

Im Rahmen dieser Arbeit wird als Eingangssignal der hohen Auflösung ein zeilenverschränktes ("interlaced") Signal gemäß ITU-R 601 verwendet. Das zeitlich-örtliche Abtastraster dieses Signals ist in Abb. 3.1 dargestellt.

- : Abtastwerte der geraden (oberen) Halbbilder
- + : Abtastwerte der ungeraden (unteren) Halbbilder



(a) zeitlich-örtliche Position der Halbbilder im Zeilensprungverfahren

(b) Position der Zeilen und Bildpunkte zweier aufeinanderfolgender Halbbilder

(c) Position der Halbbilder und Zeilen in der $y - t$ -Ebene

Abbildung 3.1.: Darstellung des zeitlich-örtlichen Abtastrasters eines zeilenverschränkten Videosignals

In der örtlich skalierbaren Codierung ist zur Erzeugung des Baselayersignals eine Abtastratenreduktion nach der Tiefpassfilterung notwendig (vgl. Abb. 2.5). Bei der Verwendung eines zeilenverschränkten Eingangssignals entsteht da-

bei durch die vertikale Abtastratenreduktion im Baselayersignal ein zeilenverschränktes Abtastraster mit geometrisch ungleichmäßigem Zeilenabstand [29], siehe Abb. 3.2. Bei Wiedergabe des Baselayersignals auf einem Monitor der niedrigen örtlichen Auflösung macht sich dieser ungleichmäßige Zeilenabstand als starke Störung bemerkbar, da die Bildpunkte der ungeraden Halbbilder vertikal an einer geometrisch falschen Position dargestellt werden.

○ : Abtastwerte der geraden (oberen) Halbbilder
 + : Abtastwerte der ungeraden (unteren) Halbbilder

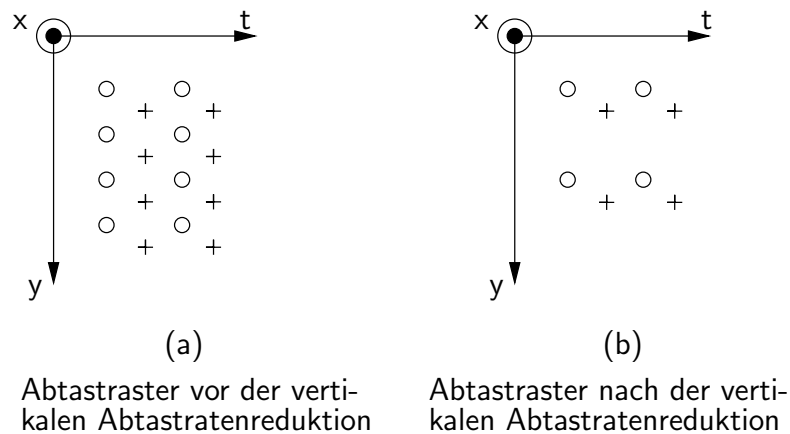


Abbildung 3.2.: Ungleichmäßiger Zeilenabstand der Abtastwerte in der $y - t$ -Ebene nach vertikaler Abtastratenreduktion der Halbbilder

Um dies zu verhindern, wird im standardisierten MPEG-2 "Spatial Scalable Profile" (SSP) vor der Filterung des Eingangssignals ein so genanntes De-Interlacing durchgeführt. Dabei handelt es sich um eine zeitlich-örtliche Filterung, bei der jedes zeilenverschränkte Halbbild mittels Zeileninterpolation zu einem nicht-zeilenverschränkten ("progressiven") Vollbild ergänzt wird, siehe Abb. 3.3 (a) und (b). Aus diesem progressiven Bildsignal wird durch Tiefpassfilterung und Abtastratenreduktion ein progressives Signal der niedrigen Auflösung erzeugt (Abb. 3.3 (c)). Durch einfaches Auslassen jeder zweiten Zeile (Re-Interlacing) lässt sich daraus ein zeilenverschränktes Signal niedriger Auflösung mit geometrisch gleichmäßigem Zeilenabstand gewinnen (Abb. 3.3 (d)), das codiert und im Baselayer-Bitstrom übertragen wird.

In MPEG-2 SSP wird das rekonstruierte Baselayersignal zur Prädiktion des Signals der hohen Auflösung benötigt. Dafür muss die Abtastrate erhöht werden, d.h. aus dem Abtastraster nach Abb. 3.3 (d) muss zunächst durch De-Interlacing das Abtastraster nach Abb. 3.3 (c), daraus durch örtliche Interpolation das Ab-

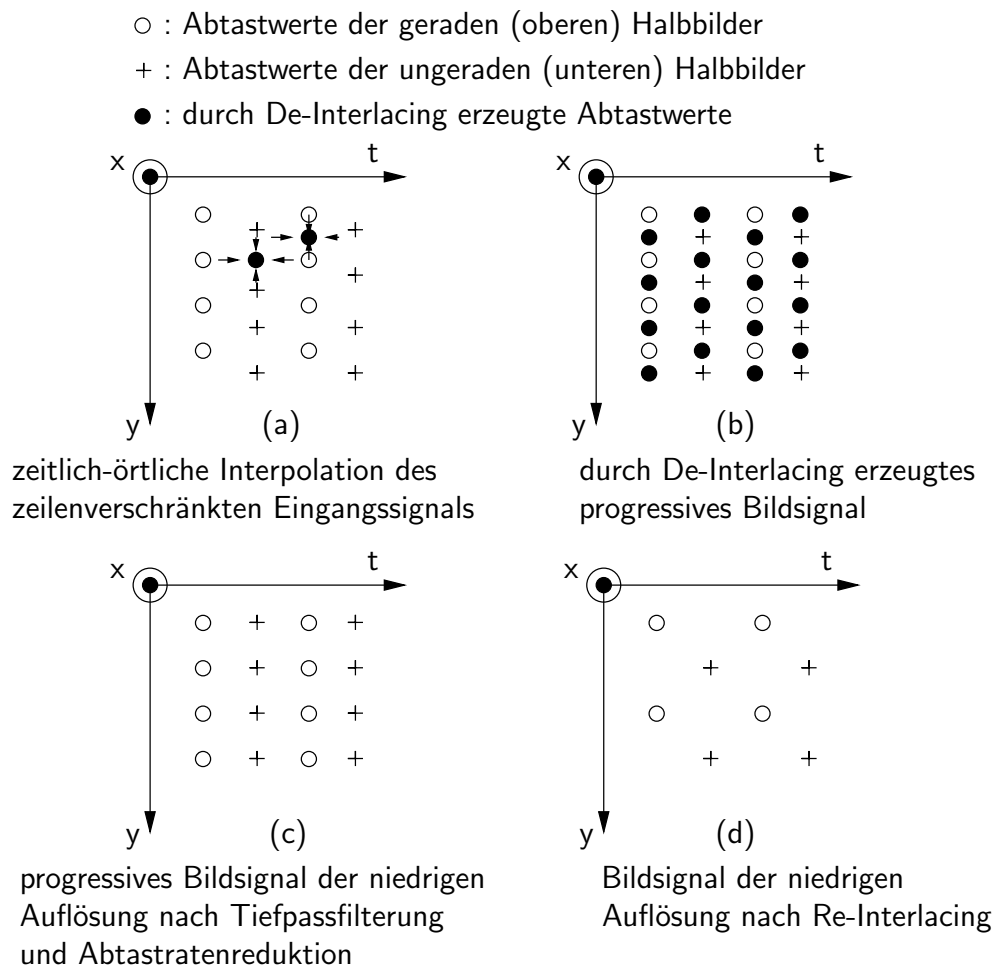
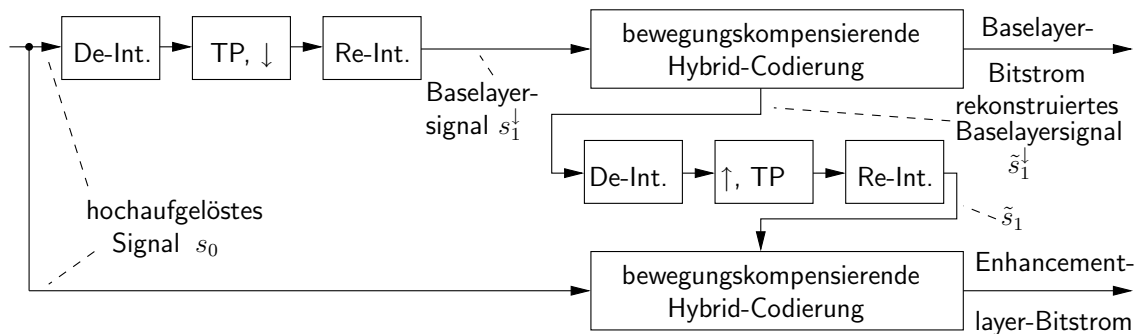


Abbildung 3.3.: Erzeugung der niedrigen Auflösung mittels De-Interlacing, Abtastratenreduktion und Re-Interlacing

tastraster nach Abb. 3.3 (b) und schließlich durch Re-Interlacing das benötigte Abtastraster nach Abb. 3.3 (a) erzeugt werden. Die dafür notwendigen zusätzlichen Komponenten, die auch im Decoder benötigt werden, sind in Abb. 3.4 dargestellt.

Da das De-Interlacing sehr aufwändig ist, wird in dieser Arbeit eine alternative Lösung untersucht, bei der kein De-Interlacing im Decoder erforderlich ist. Dazu wird kein zeilenverschränktes, sondern direkt das progressive Bildsignal codiert und übertragen. Da das progressive Bildsignal die doppelte Anzahl Abtastwerte verglichen mit dem zeilenverschränkten Eingangssignal besitzt, erscheint dieses Vorgehen zunächst nachteilig, weil eine Erhöhung der zu übertragenden Datenrate zu erwarten ist.



TP: Tiefpass-Filter ; ↓: Abtastratenreduktion ; ↑: Abtastratenerhöhung
 De-Int.: De-Interlacing ; Re-Int.: Re-Interlacing

Abbildung 3.4.: Blockschaltbild der MPEG-SSP-Pyramidencodierung bei zeilenverschränktem Eingangssignal

3.2. Codierung eines progressiven Hilfssignals

Um diesen Nachteil beurteilen zu können, werden in einer experimentellen Untersuchung sowohl das zeilenverschränkte Eingangssignal als auch das progressive Bildsignal der hohen Auflösung bei einem PSNR von 28 dB bis 34 dB mit dem MPEG-2-Verfahren codiert. Nach der Decodierung wird jede zweite Zeile des progressiven Bildsignals wieder entfernt (Re-Interlacing). Die Effizienz der Codierung des progressiven Signals und des zeilenverschränkten Signals wird anhand der Datenrate bei gleichem PSNR der decodierten zeilenverschränkten Signale verglichen (siehe Abb. 3.5).

Als De-Interlacing-Verfahren wird hierbei ein verbesserter Algorithmus nach [30] verwendet. Im Gegensatz zum De-Interlacing nach MPEG-2 SSP wird die zeitlich-örtliche Interpolationsfilterung dabei bewegungskompensiert durchgeführt. Zusätzlich beinhaltet der Algorithmus eine Klassifikation der zu interpolierenden Abtastwerte nach verläSSLicher und unsicherer Bewegungsinformation sowie nach frei werdendem und verdecktem Hintergrund. Abhängig von dieser Klassifikation werden zur Interpolation der einzelnen Abtastwerte verschiedene Filter verwendet.

Die Ergebnisse der Codierung für die beiden MPEG-Testsequenzen "Basketball" und "Bus" sind in Tabelle 3.1 dargestellt. Es zeigt sich am Beispiel der Testsequenz "Bus", dass im Fall der Codierung des progressiven Bildsignals bei einer hohen Bildqualität entsprechend einem PSNR von 34,3 dB die Datenrate von 6 Mbit/s lediglich um 9% größer ist als die Datenrate von 5,5 Mbit/s, die das zeilenverschränkte Signal bei gleichem PSNR benötigt. Bei geringerer Bildqualität

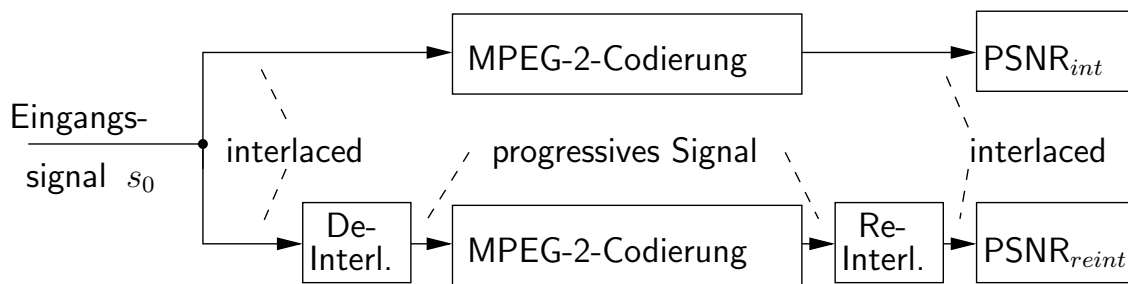


Abbildung 3.5.: Messaufbau zum Vergleich der Codierung von zeilenverschränktem und progressivem Signal

Datenrate	Bus		Basketball	
	PSNR _{int} [dB]	PSNR _{reint} [dB]	PSNR _{int} [dB]	PSNR _{reint} [dB]
3 Mbit/s	31,04	31,29	27,66	27,98
3,5 Mbit/s	31,92	32,02	28,61	28,73
4 Mbit/s	32,63	32,61	29,43	29,39
4,5 Mbit/s	33,25	33,12	30,17	29,97
5 Mbit/s	33,83	33,57	30,83	30,50
5,5 Mbit/s	34,31	33,98	31,43	30,96
6 Mbit/s	34,74	34,34	31,98	31,40

Tabelle 3.1.: Messergebnisse des Vergleichs, MPEG-Testsequenzen "Basketball" und "Bus" (ITU-R 601, 149 Bilder)

entsprechend einem PSNR von 32,6 dB (4 Mbit/s) ist kein Anstieg der Datenrate mehr zu beobachten. Bei einem noch geringeren PSNR benötigt die Codierung des progressiven Bildsignals sogar eine niedrigere Datenrate als die zeilenverschränkte Codierung. Eine Erklärung dafür ist, dass die durch das De-Interlacing hinzugefügten Zeilen lediglich mathematisch berechenbare, redundante Information beinhalten. Der Encoder ist in der Lage, diese Redundanz wieder zu entfernen und das progressive Bildsignal trotz doppelter Anzahl Abtastwerte bei nahezu gleicher Datenrate zu codieren wie das zeilenverschränkte Eingangssignal. Weiterhin fallen die bei zeilenverschränkter Codierung notwendigen speziellen Seiteninformationen zur Auswahl des Prädiktionsmodus bei einer rein progressiven Codierung weg. Darüberhinaus können die Intra-frames im Fall progressiver Darstellung effektiver codiert werden, indem das ungerade 2. Halbbild des Intra-Frames bereits bewegungskompensiert verarbeitet wird.

Diese Ergebnisse legen es nahe, in Zukunft nur noch progressiv codierte Bildsignale zu verwenden, mit dem Vorteil, dass auch progressive Computer- und

Flachbild-Monitore direkt zur Anzeige des Signals benutzt werden können. Für die Anzeige auf konventionellen zeilenverschränkten Monitoren ist lediglich ein Re-Interlacing, also ein Auslassen jeder zweiten Zeile, notwendig.

Die Erkenntnisse dieser Untersuchung werden nachfolgend in die Entwicklung des örtlich skalierbaren Teilband-Codierungssystems übernommen, indem dort ausschließlich progressive Signale verarbeitet werden.

4. Konzept der örtlich skalierbaren Teilband-DCT-Codierung

Die im Rahmen dieser Arbeit entwickelte bewegungskompensierende Teilbandcodierung geht von dem in Abb. 2.8 dargestellten Blockschaltbild einer örtlich skalierbaren Teilbandcodierung aus. Dabei soll sowohl im Baselayer als auch im Enhancementlayer eine Codierung ähnlich dem MPEG-2-Standard verwendet werden, um eine hohe Codierungseffizienz bei vertretbarem Realisierungsaufwand sicherzustellen.

Das Blockschaltbild des Encoders der bewegungskompensierenden Teilbandcodierung ist in Abb. 4.1 dargestellt. Nach der Analysefilterung des Eingangssignals s_0 wird das tieffrequente Teilband s_1^\downarrow als Baselayersignal im Baselayer-Encoder verarbeitet. Dieser stellt einen bewegungskompensierenden Hybrid-Coder mit eigener bewegungskompensierender Prädiktion (MCP) entsprechend Abschnitt 2.2 dar. Das durch die Prädiktion entstehende Baselayer-Prädiktionsfehlersignal e_1^\downarrow wird codiert und im Baselayer-Bitstrom übertragen. Prädiktion und Codierung des Baselayersignals werden dabei unabhängig von den hochfrequenten Teilbändern des Enhancementlayers durchgeführt, um eine unabhängige Decodierung des Baselayer-Bitstrom im Baselayer-Decoder zu ermöglichen. Würden auch Informationen der hochfrequenten Teilbänder bei der Erzeugung des Baselayer-Bitstroms berücksichtigt, wäre eine so genannte Drift der Prädiktions-signale zwischen Baselayer-Encoder und -Decoder die Folge, die zu einer Verminderung der Codierungseffizienz im Baselayer-Decoder führt [25].

Für die Codierung des Enhancementlayers wurde in [15] gezeigt, dass eine getrennte bewegungskompensierende Prädiktion der einzelnen Teilbandsignale s_2^\downarrow , s_3^\downarrow und s_4^\downarrow im Enhancementlayer zu einer Erhöhung des Prädiktionsfehlers führt, da diese so genannte "In-Band"-Bewegungskompensation durch Aliasing gestört wird. Aus diesem Grund werden in dieser Arbeit die Prädiktions-signale der drei Teilbänder des Enhancementlayers aus dem bewegungskompensierten Gesamtsignal \hat{s}_0 der hohen Auflösung gewonnen, siehe Abb. 4.1.

Nach der Synthesefilterung der rekonstruierten Signale \tilde{s}_1^\downarrow , \tilde{s}_2^\downarrow , \tilde{s}_3^\downarrow und \tilde{s}_4^\downarrow wird das rekonstruierte Signal \tilde{s}_0 der bewegungskompensierenden Prädiktion (MCP) zugeführt. Das entstehende Prädiktions-signal \hat{s}_0 wird in der Analysefilterbank in die Teilband-Prädiktions-signale $\hat{s}_1^{\downarrow*}$, \hat{s}_2^\downarrow , \hat{s}_3^\downarrow und \hat{s}_4^\downarrow zerlegt. Das Signal $\hat{s}_1^{\downarrow*}$ wird im

Enhancementlayer-Encoder nicht benötigt, da dieser nur die drei hochfrequenten Teilbänder überträgt. Die Teilband-Prädiktionssignale \hat{s}_2^\downarrow , \hat{s}_3^\downarrow und \hat{s}_4^\downarrow werden von den Teilbandsignalen s_2^\downarrow , s_3^\downarrow und s_4^\downarrow subtrahiert und die Teilband-Prädiktionsfehlersignale e_2^\downarrow , e_3^\downarrow und e_4^\downarrow codiert und im Enhancementlayer-Bitstrom übertragen.

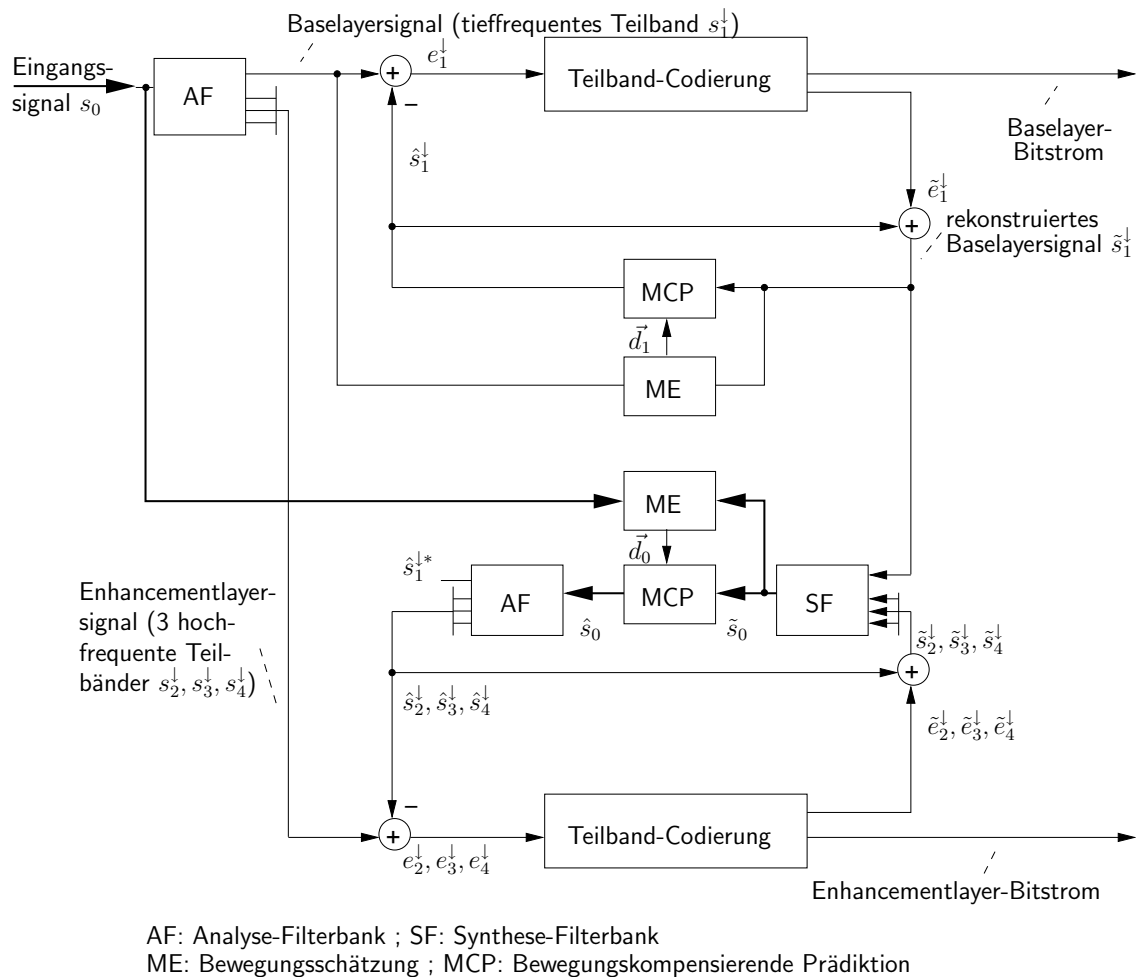


Abbildung 4.1.: Blockschaltbild des Encoders der bewegungskompensierenden Teilbandcodierung

Um eine hohe Codierungseffizienz zu erreichen, kann die Teilbandzerlegung mit einer 8×8 -DCT-Codierung der Teilbandsignale entsprechend dem MPEG-2-Verfahren kombiniert werden, siehe Abb. 4.2. Der Baselayer-Encoder stellt dann einen Standard-MPEG-Hybrid-DCT-Coder dar. Im Enhancementlayer werden für die drei Teilbandsignale drei DCT-Codierungen eingesetzt. Im Hinblick auf

den Realisierungsaufwand ist die Verwendung der DCT auch hier vorteilhaft, da somit das gleiche DCT-Codierungsmodul im Wechsel für Base- und Enhancementlayersignale verwendet werden kann. Durch die Abstratenreduktion der Teilbandsignale ist die Anzahl der zu codierenden Abtastwerte gleich der einer nichtskalierbaren Codierung des Signals s_0 . Der zusätzliche Realisierungsaufwand der gesamten örtlich skalierbaren Teilband-DCT-Codierung verglichen mit einer nichtskalierbaren MPEG-Codierung beschränkt sich damit auf die zusätzlichen Analyse- und Synthese-Filterbänke.

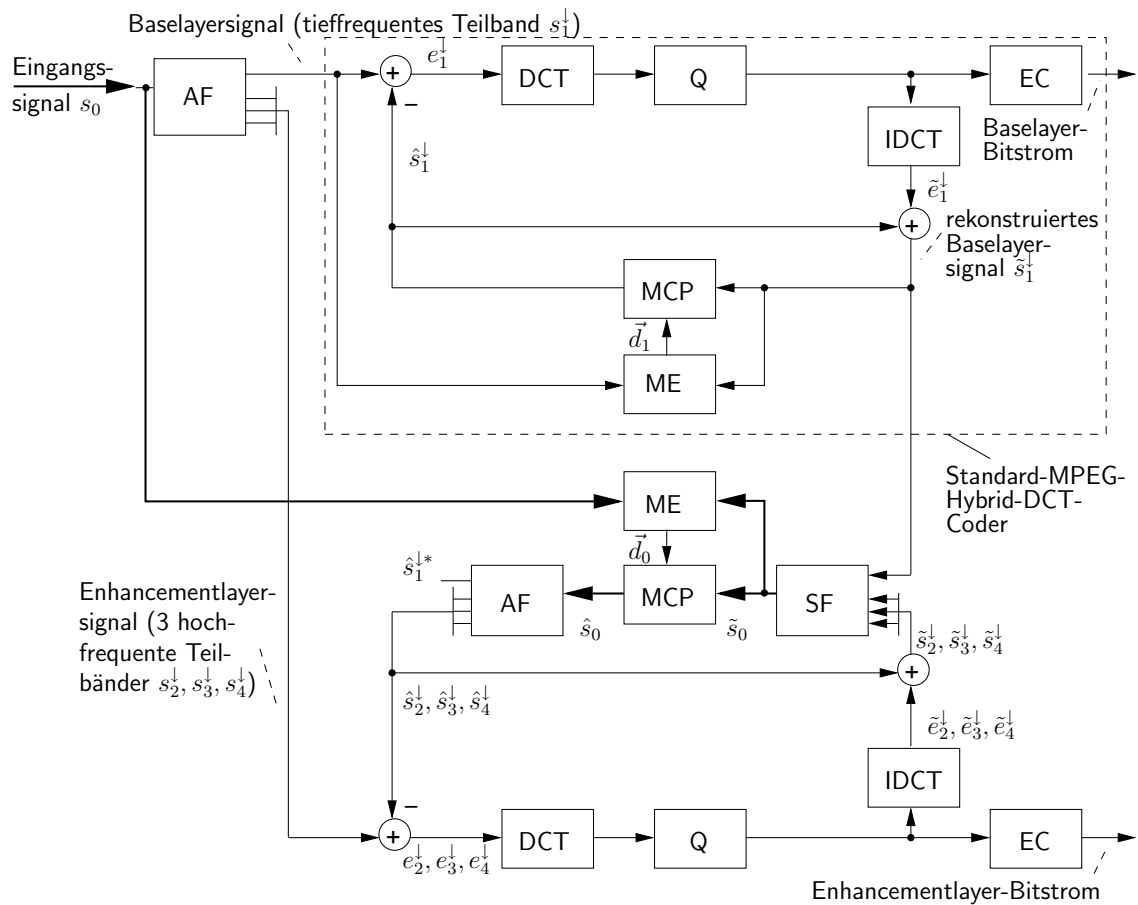


Abbildung 4.2.: Blockschaltbild des Encoders mit kombinierter Teilband-DCT-Codierung

5. Entwurf der Teilbandfilter für die örtliche Skalierung

5.1. Filterbänke mit exakter Rekonstruktion

Die in dieser Arbeit betrachtete Teilbandcodierung geht von einer zweidimensionalen örtlichen Analyse- und Synthesefilterbank aus. Die dabei verwendeten Teilbandfilter sind separierbar, so dass sich die zweidimensionale Filterbank als eine horizontale Filterbank, gefolgt von einer vertikalen Filterbank, darstellen lässt [41].

Das Prinzip der hier verwendeten Filterbänke mit der Eigenschaft der exakten Rekonstruktion wird in diesem Abschnitt anhand einer eindimensionalen Zweibandaufteilung erläutert. Anschließend wird darauf aufbauend die in dieser Arbeit verwendete zweidimensionale Teilbandzerlegung vorgestellt.

Die Analyse- und Synthesefilterbank einer eindimensionalen Zweibandaufteilung ist in Abb. 5.1 gezeigt.

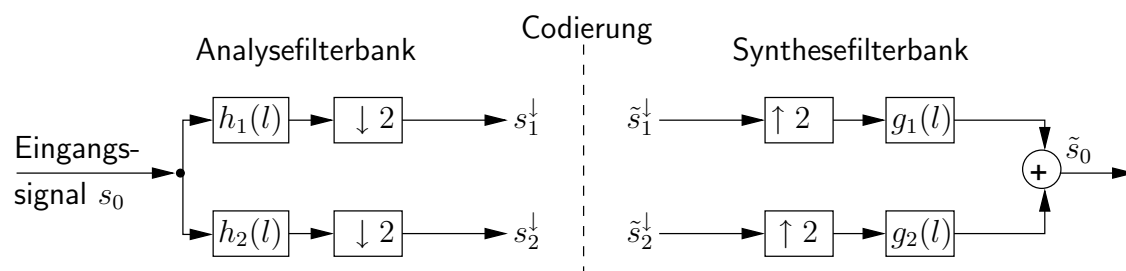


Abbildung 5.1.: Analyse- und Synthesefilterbank einer eindimensionalen Zweibandaufteilung

Das Eingangssignal s_0 wird in der Analysefilterbank in zwei Teilbandsignale s_1^\downarrow und s_2^\downarrow jeweils gleicher Frequenzbandbreite aufgespalten. Die Analysefilterbank setzt sich aus einem Tiefpassfilter mit der Impulsantwort $h_1(l)$ und einem Hochpassfilter mit der Impulsantwort $h_2(l)$ mit jeweils anschließender Abtastraten-

reduktion zusammen. Bei der Abtastratenreduktion wird jeder zweite Abtastwert ausgelassen, so dass die Abtastfrequenzen der Teilbandsignale s_1^\downarrow und s_2^\downarrow gegenüber der Abtastfrequenz des Eingangssignals halbiert sind. Es liegt also eine kritische Abtastratenreduktion vor. Der nach unten gerichtete Pfeil \downarrow symbolisiert die Abtastratenreduktion und zeigt an, dass gegenüber den Abtastwerten s_0 nur noch jeder zweite Abtastwert erhalten bleibt. Die Abtastwerte der Teilbandsignale s_1^\downarrow und s_2^\downarrow werden codiert und übertragen.

Die decodierten Teilbandsignale \tilde{s}_1^\downarrow und \tilde{s}_2^\downarrow werden der Synthesefilterbank, bestehend aus der Abtastratenerhöhung und den Teilbandfiltern $g_1(l)$ und $g_2(l)$, zugeführt. Bei der Abtastratenerhöhung wird die Abtastfrequenz der Teilbandsignale verdoppelt, indem zwischen den vorliegenden Abtastwerten der Teilbandsignale jeweils ein Abtastwert der Größe Null eingefügt wird. Am Ausgang der Synthesefilterbank besitzt das rekonstruierte Signal \tilde{s}_0 wieder die ursprüngliche Abtastfrequenz.

Im Folgenden soll für die Zweibandaufteilung gezeigt werden, welche Beziehungen zwischen den Teilbandfiltern $h_1(l)$, $h_2(l)$, $g_1(l)$ und $g_2(l)$ eingehalten werden müssen, damit die geforderte Eigenschaft der exakten Rekonstruktion erreicht wird.

Eine exakte Rekonstruktion kann mit dem Zweibandsystem überhaupt nur erreicht werden, wenn die Ausgangssignale der Analysefilterbank s_1^\downarrow und s_2^\downarrow ohne Informationsverlust direkt in die Synthesefilterbank gespeist werden, d.h. wenn

$$\tilde{s}_n^\downarrow = s_n^\downarrow \text{ mit } n = 1, 2. \quad (5.1)$$

Die weiteren Bedingungen zur Erzielung der exakten Rekonstruktion lassen sich einfach mit Hilfe der z-Transformation angeben [42]. Abtastratenreduktion und Abtastratenerhöhung eines Teilbandsignals werden dabei gemäß Abb. 5.2 zusammengefasst und als eine Multiplikation einer Folge von Abtastwerten mit der Folge $\{1, 0, 1, 0, \dots\}$ bzw. $(1 + (-1)^l)/2$ interpretiert. Für diese Folge gilt bei Anwendung der z-Transformation

$$\begin{aligned} Z \left\{ (1 + (-1)^l) \cdot s_0(l) \right\} &= S_0(z) + S_0(-z) \\ \text{mit } Z \{s_0(l)\} &= \sum_l s_0(l) \cdot z^{-l} = S_0(z) \\ Z \left\{ (-1)^l s_0(l) \right\} &= \sum_l s_0(l) \cdot z^{-l} \cdot (-1)^l = \\ &= \sum_l s_0(l) \cdot z^{-l} \cdot (-1)^{-l} = \sum_l s_0(l) \cdot (-z)^{-l} = S_0(-z) \end{aligned} \quad (5.2)$$

Die z-Transformierte des Ausgangssignals ist dann

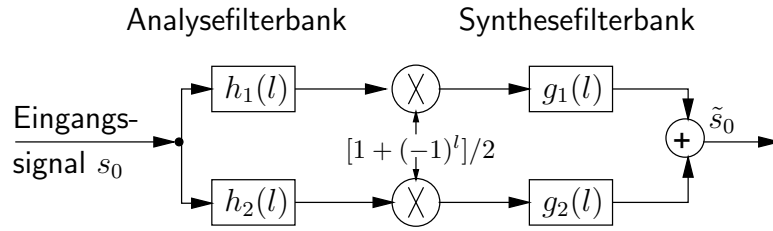


Abbildung 5.2.: Gesamtsystem zur eindimensionalen Zweibandaufteilung nach Zusammenfassung von Abtastratenreduktion und -erhöhung

$$\begin{aligned} \tilde{S}_0(z) = & \frac{1}{2} \left(G_1(z) \left(H_1(z) S_0(z) + H_1(-z) S_0(-z) \right) \right. \\ & \left. + G_2(z) \left(H_2(z) S_0(z) + H_2(-z) S_0(-z) \right) \right) \end{aligned} \quad (5.3)$$

$$\begin{aligned} \tilde{S}_0(z) = & S_0(z) \cdot \frac{1}{2} \left(G_1(z) H_1(z) + G_2(z) H_2(z) \right) + \\ & + S_0(-z) \cdot \frac{1}{2} \left(G_1(z) H_1(-z) + G_2(z) H_2(-z) \right). \end{aligned} \quad (5.4)$$

Die Terme mit $S_0(-z)$ in Gl. 5.4 sind so genannte Aliasing-Terme, die unerwünscht sind und sich zu Null kompensieren müssen.

Bei Teilbandfilterlängen L , d.h. $h(l) = [h(0), h(1), \dots, h(L-1)]$, ist für die exakte Rekonstruktion zu fordern, dass \tilde{s}_0 bis auf eine Verzögerung von $L-1$ Abtastwerten gleich s_0 ist:

$$\tilde{S}_0(z) = z^{-(L-1)} S_0(z). \quad (5.5)$$

Setzt man diese Forderungen in Gl. 5.4 ein, so lässt sich ein Gleichungssystem zur Berechnung der Synthesefilter aus den Analysefiltern aufstellen:

$$\begin{pmatrix} H_1(z) & H_2(z) \\ H_1(-z) & H_2(-z) \end{pmatrix} \begin{pmatrix} G_1(z) \\ G_2(z) \end{pmatrix} = \begin{pmatrix} 2z^{-(L-1)} \\ 0 \end{pmatrix}. \quad (5.6)$$

Die Lösung dieses Systems lautet:

$$\begin{aligned} G_1(z) &= \frac{2z^{-(L-1)} H_2(-z)}{D(z)} \\ G_2(z) &= \frac{2z^{-(L-1)} H_1(-z)}{D(z)} \end{aligned} \quad \text{mit } D(z) = H_1(z) H_2(-z) - H_1(-z) H_2(z). \quad (5.7)$$

Aufgrund der Forderung, dass die Impulsantworten g_1 und g_2 die endliche Länge L haben, muss der Nennerterm in Gl. 5.7 eine einfache Potenz von z multipliziert mit einer Konstanten c sein, d.h. $D(z)$ muss ein Verzögerungsglied mit der Verstärkung $c \neq 0$ sein. Da $D(z)$ ein Filter darstellt, das durch Faltung jeweils

zweier Impulsantworten der Länge L entsteht, muss diese Verzögerung $L - 1$ betragen. Die Bedingungen für die exakte Rekonstruktion lauten damit:

$$\begin{aligned} G_1(z) &= \frac{2}{c} H_2(-z) \\ G_2(z) &= -\frac{2}{c} H_1(-z) \\ H_1(z)H_2(-z) - H_1(-z)H_2(z) &= c z^{-(L-1)}. \end{aligned} \quad (5.8)$$

Durch inverse z-Transformation und Substitution von $F(z) = H_1(z)H_2(-z)$ bzw. $F(-z) = H_1(-z)H_2(z)$ erhält man nun die Bedingungen für die Impulsantworten der Teilbandfilter:

$$\begin{aligned} g_1(l) &= (-1)^l \frac{2}{c} h_2(l) \quad ; \quad g_2(l) = -(-1)^l \frac{2}{c} h_1(l) \quad (5.9) \\ f(l) (1 - (-1)^l) &= c \delta(l - L + 1) \quad \text{mit} \quad f(l) = h_1(l) * ((-1)^l h_2(l)). \quad (5.10) \end{aligned}$$

Dabei beschreibt Gl. 5.9 die Beziehungen zwischen Analyse- und Synthesefilterbank. Aus Gl. 5.10 lässt sich ein Gleichungssystem aufstellen, aus dem z.B. bei Vorgabe des Tiefpassfilters $h_1(l)$ das Hochpassfilter $h_2(l)$ bestimmt werden kann (siehe Anhang A). Mit Festlegung des Filters h_1 sind somit auch die anderen Filter h_2 , g_1 und g_2 bestimmt. Bei Aufstellung des Gleichungssystems kann man weiterhin sehen, dass die Impulsantwortlänge L gerade sein muss, da sonst die Gleichung für $l = L - 1$ nicht erfüllt werden kann.

Durch Vorgabe weiterer Bedingungen für die Teilbandfilter ergeben sich verschiedene Klassen von Filterbänken zur Zweiband-Aufteilung, von denen eine im Folgenden genauer beschrieben wird.

5.1.1. Verallgemeinerte Quadraturspiegelfilter (GQMF)

Für die Anwendung in der Codierung von Videosignalen wird von den Analyse- und Synthesefiltern gefordert, dass sie linearphasig sind, d.h. dass ihre Impulsantworten symmetrisch sind. Dies ist notwendig, um eine konstante Gruppenlaufzeit zu erhalten, wodurch alle Spektralkomponenten die gleiche, konstante Verschiebung erfahren. Diese Filterklasse wird als Verallgemeinerte Quadraturspiegelfilter (Generalised Quadrature Mirror Filter, GQMF) [23] bezeichnet.

Für die in dieser Arbeit betrachtete örtlich skalierbare Codierung ist weiterhin die Verwendung von Filtern mit ungerader Impulsantwortlänge L' notwendig, um die geometrische Position der Abtastwerte im Baselayersignal s_1^\downarrow nach Filterung und Abtastratenreduktion zu erhalten. Filter mit gerader Impulsantwortlänge L führen zu einer geometrischen Verschiebung der Abtastwerte um einen halben Bildpunkt bezogen auf das Abtastraster des hochaufgelösten Signals s_0 , welche

sich ungünstig auf die Bewegungskompensation innerhalb der örtlich skalierbaren Codierung auswirkt [15].

In [42] wird ein Verfahren vorgestellt, das den Entwurf von Filterbänken mit exakter Rekonstruktion auch für ungerade Impulsantwortlängen L' erlaubt. Zu diesem Zweck wird für die Aufstellung des Gleichungssystems nach Gl. 5.10 den Impulsantworten der Filter $h_1(l)$ und $h_2(l)$ jeweils eine Null hinzugefügt, wodurch sich für die Berechnung eine gerade Impulsantwortlänge $L = L' + 1$ ergibt. Wie in Anhang A gezeigt, ergibt sich durch die Forderung der Linearphasigkeit, dass die Verlängerung der Impulsantworten von h_1 und h_2 jeweils am entgegengesetzten Ende durchgeführt werden muss, was auch als abwechselnde Abtastung der Signale s_1^\downarrow und s_2^\downarrow interpretiert werden kann. Weiterhin ergibt sich als zusätzliche Bedingung, dass die Impulsantworten der Länge L im Fall von h_1 am Beginn mit zwei Nullen und am Ende mit einer Null, im Fall von h_2 am Ende mit einer Null besetzt sein müssen, d.h.

$$\begin{aligned} h_1(l) &= [h_1(0) = 0, h_1(1) = 0, h_1(2), h_1(3), \dots, h_1(L/2 - 1), h_1(L/2), h_1(L/2 - 1), \dots, 0] \\ h_2(l) &= [h_2(0), h_2(1), \dots, h_2(L/2 - 2), h_2(L/2 - 1), h_2(L/2 - 2), \dots, 0]. \end{aligned} \quad (5.11)$$

Die Impulsantwort von $h_1(l)$ hat $L/2 - 1$ freie Parameter, mit deren Festlegung auch alle anderen Filter bestimmt sind.

5.1.2. Zweidimensionale Teilbandzerlegung

Die in dieser Arbeit betrachteten separierbaren zweidimensionalen Filterbänke setzen sich stets aus zwei eindimensionalen Filterbänken zusammen, wobei die eine Filterung in horizontaler und die andere in vertikaler Richtung erfolgt. Abb. 5.3 zeigt das Blockschaltbild der zweidimensionalen Analyse- und Synthesefilterbank.

Alle Signale werden durch ortsdiskrete Abtastwerte in der zweidimensionalen (x,y) -Ebene dargestellt. In der Analysefilterbank wird das Eingangssignal zuerst einer horizontalen Filterung in x -Richtung (zeilenweise im Bild) unterzogen, der sich eine Abtastratenreduktion in horizontaler Richtung um den Faktor 2 anschließt. Es folgen die vertikale Filterung und Abtastratenreduktion um den Faktor 2 in y -Richtung (spaltenweise im Bild). Die resultierenden vier Teilbandsignale $s_1^\downarrow, s_2^\downarrow, s_3^\downarrow$ und s_4^\downarrow werden nach Codierung, Übertragung und Decodierung in der Synthesefilterbank zuerst der vertikalen Abtastratenerhöhung und Filterung zugeführt, bevor die horizontale Abtastratenerhöhung und Filterung stattfindet. Aufgrund der Separierbarkeit der zweidimensionalen Filterung ist die Reihen-

folge von horizontaler und vertikaler Verarbeitung in der Analyse- bzw. Synthesefilterbank prinzipiell vertauschbar.

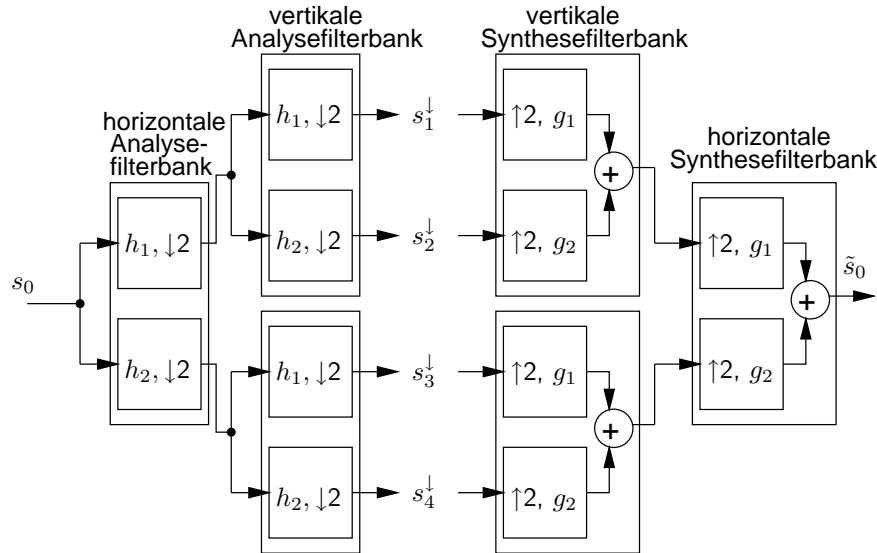


Abbildung 5.3.: Blockschaltbild der zweidimensionalen Analyse- und Synthesefilterbank

5.2. Optimierungskriterium und Randbedingungen für den Entwurf

Der Filterentwurf soll zwei Forderungen erfüllen: Zum Einen soll das Tiefpassfilter h_1 , das nach horizontaler und vertikaler Anwendung das niedrig aufgelöste Baselayersignal s_1^\downarrow erzeugt, eine gute subjektive Bildqualität liefern. Zum Anderen sollen die Teilbandfilterbänke, bestehend aus den Filtern h_1 , h_2 , g_1 und g_2 , eine hohe Codierungseffizienz der Teilbandzerlegung gewährleisten.

Der ersten Forderung trägt das von MPEG vorgeschlagene SIF-Filter [8] Rechnung. Es wurde von der ISO für die Konvertierung eines Videosignals vom ITU-R-601-Format in das SIF-Format entwickelt, unter dem Gesichtspunkt einer guten subjektiven Bildqualität und eines geringen Realisierungsaufwandes. Da auch in dieser Arbeit das Eingangssignal im ITU-R-601-Format und das Baselayersignal im SIF-Format vorliegen, ist das MPEG-SIF-Filter als Referenz geeignet. Das MPEG-SIF-Filter ist ein symmetrisches FIR-Filter mit einer Impulsantwortlänge von sieben.

Bei Verwendung des MPEG-SIF-Filters als Tiefpassfilter h_1 sind die verbleibenden Filter h_2 , g_1 und g_2 jedoch bereits vollständig bestimmt und damit auch die Codierungseffizienz festgelegt.

Um der zweiten Forderung Rechnung tragen zu können und die Codierungseffizienz zu berücksichtigen, wird für den Tiefpassfilterentwurf die Übertragungsfunktion des MPEG-SIF-Filters mit einem Toleranzbereich versehen und als Randbedingung für die Optimierung eingeführt. Zur Optimierung werden nur solche Filter zugelassen, die diese Randbedingung einhalten. Die Impulsantwortlänge von h_1 ergibt sich aus der Impulsantwortlänge des MPEG-SIF-Filters und den nach Gl. 5.11 hinzuzufügenden drei Nullen zu $L = 10$. Die Anzahl der freien Parameter beträgt damit $L/2 - 1 = 4$. Als Optimierungskriterium wird die zur Codierung der Teilbandsignale s_1^\downarrow , s_2^\downarrow , s_3^\downarrow und s_4^\downarrow benötigte Datenrate verwendet.

5.2.1. Festlegung des Toleranzschemas

Die bekannten Entwurfsverfahren [43] für digitale Tiefpassfilter beschreiben das Toleranzschema der Übertragungsfunktion mit den folgenden Parametern: δ_1 bzw. δ_2 beschreiben die maximale Amplitudenabweichung im Durchlass- bzw. im Sperrbereich, ω_D und ω_S beschreiben die Frequenzgrenzen des Durchlass- und Sperrbereiches (siehe Abb. 5.4).

Aus der Übertragungsfunktion des MPEG-SIF-Filters lassen sich die folgenden Parameter ablesen: $\delta_1 = 0.185$, $\delta_2 = 0.15$, $\omega_D = \frac{5}{16} \cdot \pi$ sowie $\omega_S = \frac{10}{16} \cdot \pi$. Zum Entwurf des Tiefpassfilters $h_1(l)$ werden die Werte δ_1 und δ_2 des MPEG-SIF-Filters übernommen. Die Grenzen des Durchlass- bzw. Sperrbereiches werden um jeweils $\frac{1}{16} \cdot \pi$ auf $\omega_D = \frac{4}{16} \cdot \pi$ sowie $\omega_S = \frac{11}{16} \cdot \pi$ erweitert, um Freiheitsgrade für die Optimierung der Codierungseffizienz zu gewinnen. Ohne diese Erweiterung gäbe es nur eine Lösung: das MPEG-SIF-Filter.

5.2.2. Optimierung der Codierungseffizienz

Aufgrund der gewählten Impulsantwortlänge $L = 10$ ergeben sich $L/2 - 1 = 4$ freie Parameter für den Entwurf des symmetrischen Tiefpassfilters $h_1(l)$. Zur Erhaltung der Bildhelligkeit, d.h. $H_1(\omega = 0) = 1$, muss zusätzlich die Bedingung

$$\sum_{l=0}^9 h_1(l) = 1 \quad (5.12)$$

erfüllt sein, wodurch sich die Anzahl der freien Parameter auf drei reduziert.

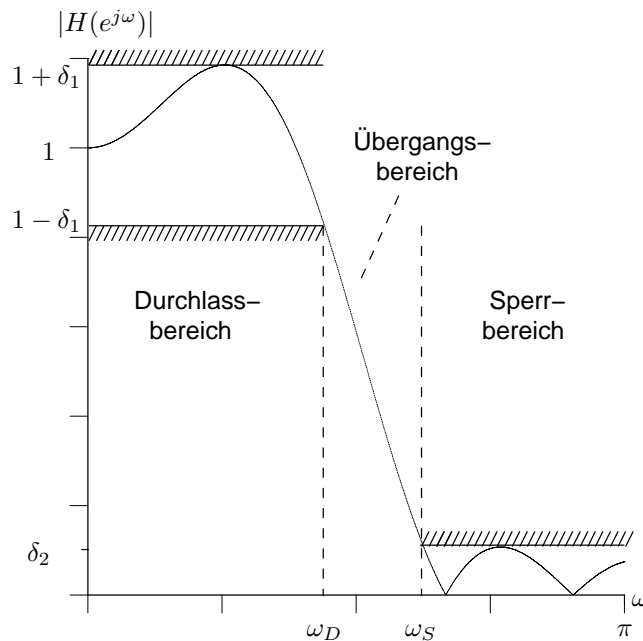


Abbildung 5.4.: Toleranzschema zum Entwurf eines Tiefpassfilters (nach [43])

Für eine einfache Implementierung der Filterbänke ist es vorteilhaft, die Filterkoeffizienten $h_i(l)$ als ganzzahlige Werte in der Einheit $[1/256]$ mit dem Wertebereich $\{-256 \dots 0 \dots 255\}$ anzugeben. Dies erlaubt eine Darstellung der Koeffizienten $h_i(l)$ als Binärzahlen mit jeweils 8 bit + 1 bit Vorzeichen.

Bei drei zu bestimmenden Filterkoeffizienten und einer Darstellung der Koeffizienten mit 9 bit ergeben sich $(2^9)^3 = 512^3 \approx 130.000.000$ mögliche Kombinationen. Die Berücksichtigung der Randbedingung schränkt diese Zahl auf 1715 Tiefpassfilter-Frequenzgänge ein, die das Toleranzschema einhalten.

Für jedes der mittels vollständiger Suche gefundenen Tiefpassfilter h_1 , die das Toleranzschema einhalten, werden die zugehörigen Teilbandfilter h_2 , g_1 und g_2 entsprechend Anhang A berechnet.

Um die Codierungseffizienz der Teilbandzerlegung abzuschätzen, werden die mit diesen Filtern erzeugten Teilbandsignale jeweils mit dem MPEG-2-Verfahren im Intra-Mode codiert. Diese Abschätzung ist möglich, da im nachfolgend erläuterten skalierbaren Codierungsverfahren die Teilbandsignale auf ähnliche Weise codiert werden.

Das zu entwickelnde skalierbare Codierungsverfahren beinhaltet jedoch auch eine bewegungskompensierende Prädiktion. Aus diesem Grund wird für die Messung der Datenrate als Eingangssignal der Teilbandfilterbänke ein Bildsignal s und ein Prädiktionsfehlerbildsignal e verwendet (vgl. Abb. 2.2). Um die Ein-

flüsse der Quantisierersteuerung zu erfassen, wird die Codierung jeweils bei den MPEG-Quantisierer-Stufenbreiten QP von 10, 20, 30 und 40 durchgeführt, so dass insgesamt 8 Testbilder verarbeitet werden.

Aus den 1715 Lösungsmöglichkeiten wird die Teilbandfilterbank, die zur geringsten mittleren Datenrate führt, ausgewählt. Abb. 5.5 zeigt die Verteilung der 1715 Filter in Abhängigkeit von der benötigten Datenrate. Aus Abb. 5.5 ist erkennbar,

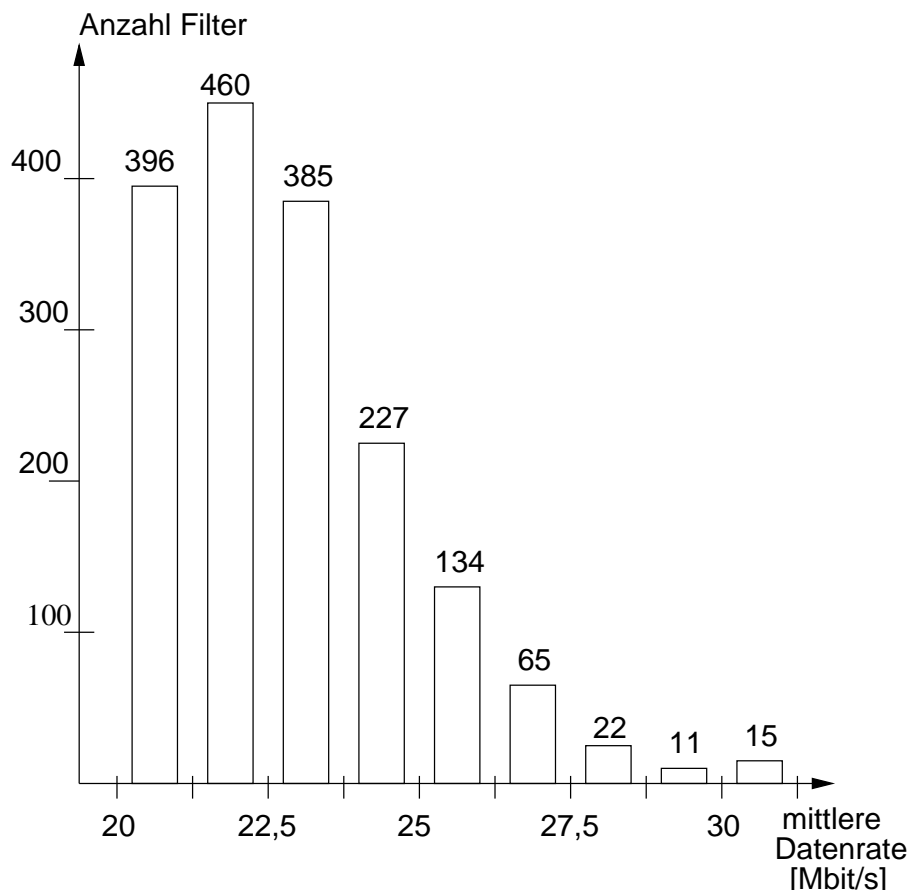


Abbildung 5.5.: Verteilung der Filterbänke, die das Toleranzschema einhalten, in Abhängigkeit von der Datenrate (MPEG-SIF-Filter benötigt 24,95 Mbit/s)

dass mehr als 85% der Filterbänke, die das Toleranzschema einhalten, eine geringere Datenrate benötigen als die Filterbank mit dem MPEG-SIF-Tiefpassfilter. Die Optimierung der Koeffizienten führt auf eine Filterbank mit einer mittleren Datenrate von 20,425 Mbit/s. Dies entspricht einer Einsparung der Datenrate von 18,1% im Vergleich zu der Filterbank mit dem MPEG-SIF-Tiefpassfilter, die eine mittlere Datenrate von 24,95 Mbit/s benötigt. Die subjektive Bildqualität

im Signal s_1^\dagger des optimierten Tiefpassfilters und des MPEG-SIF-Filters sind dabei gleich. Die Betragsfrequenzgänge der Analysefilterbank mit dem MPEG-SIF-Tiefpassfilter sowie die Betragsfrequenzgänge der optimierten Analysefilterbank sind in Abb. 5.6 dargestellt.

Zusätzlich konnte beobachtet werden, dass im Fall der optimierten Filterbank das PSNR im rekonstruierten Signal hoher Auflösung \tilde{s}_0 nach der Synthesefilterung um 0,8 dB größer ist als im Fall der Filterbank mit dem MPEG-SIF-Tiefpassfilter. Dies ist damit zu erklären, dass die optimierte Filterbank Teilbandsignale erzeugt, die in der nachfolgenden DCT-Codierung zu einer großen Anzahl von Koeffizienten führt, deren Amplituden unterhalb der Quantisiererstufenbreite liegen. Nach der Rate-Distortion-Theorie [31] führen solche Koeffizienten sowohl zu einer geringeren Datenrate als auch zu einer geringeren Quantisiererfehlerleistung und damit zu einem größeren PSNR des rekonstruierten Signals.

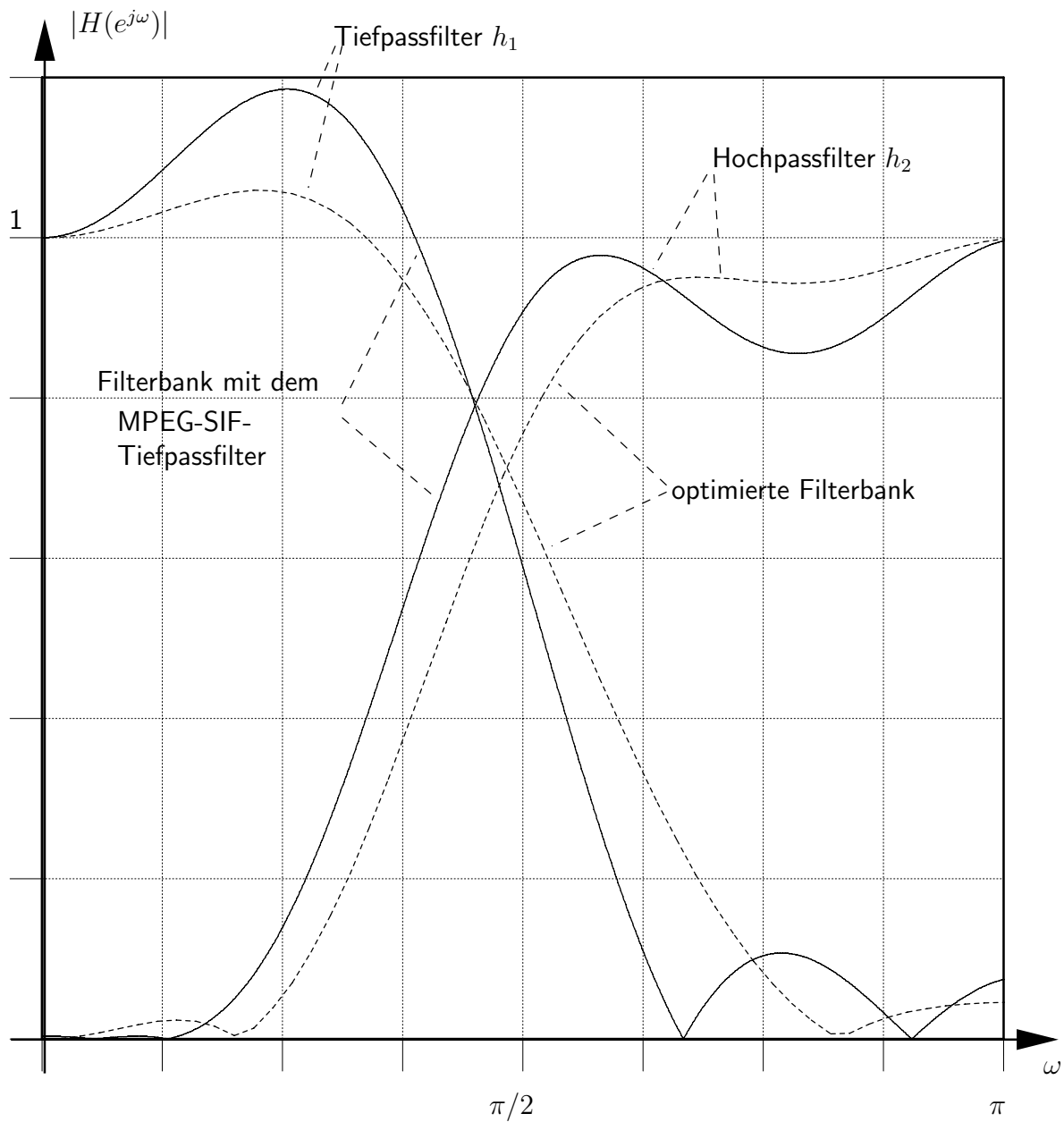


Abbildung 5.6.: Betragsfrequenzgänge der Analysefilterbank mit dem MPEG-SIF-Tiefpassfilter sowie der optimierten Analysefilterbank

(zugehörige Impulsantworten der Länge $L = 10$:

$$h_{1,MPEG} = \{0, 0, -29, 0, 88, 138, 88, 0, -29, 0\},$$

$$h_{2,MPEG} = \{13, 0, -27, -64, 156, -64, -27, 0, 13, 0\}$$

sowie

$$h_{1,opt} = \{0, 0, -12, -10, 79, 142, 79, -10, -12, 0\},$$

$$h_{2,opt} = \{9, 8, -18, -72, 146, -72, -18, 8, 9, 0\}$$

6. Bewegungskompensation und Codierung der Teilbandsignale

Im zu entwickelnden skalierbaren Teilband-DCT-Codierungsverfahren wird sowohl im Baselayer als auch im Enhancementlayer eine Codierung ähnlich dem MPEG-2-Standard verwendet, um eine hohe Codierungseffizienz bei vertretbarem Realisierungsaufwand sicherzustellen (vgl. Abschnitt 4). Zu diesem Zweck sind Anpassungen der MPEG-2-Codierung an die jeweiligen Teilbandsignale notwendig. Diese Anpassungen bestehen aus Modifikationen der bewegungskompensierenden Prädiktion sowie der Erzeugung geeigneter Quantisierungsmatrizen für die DCT-Koeffizienten der Teilbandsignale.

6.1. Bewegungsschätzung und -kompensation

Die entwickelte Teilband-DCT-Codierung wendet jeweils eine bewegungskompensierende Prädiktion für den Base- und den Enhancementlayer an, um eine drifffreie örtlich skalierbare Codierung zu ermöglichen (vgl. Abschnitt 4).

Dabei werden die folgenden Parameter verwendet:

Im Enhancementlayer wird die bewegungskompensierende Prädiktion auf Blöcken der Größe 16x16 Bildpunkte bezüglich des Abtastrasters der hohen Auflösung durchgeführt. Die Displacementvektoren besitzen eine Amplitudenauflösung von $1/2$ Bildpunkt. Diese Parameterwerte entsprechen denen des MPEG-2-Verfahrens.

Die in MPEG-2 verwendete blockbasierte Bewegungskompensation generiert ein Prädiktionssignal, das an den Blockgrenzen Signalsprünge aufweist. Diese führen in der Teilbandzerlegung zu unerwünschten, energiereichen hochfrequenten Koeffizienten. Um diese Signalsprünge zu vermeiden, wird im entwickelten System eine Prädiktion mit überlappenden Blöcken gemäß [40] verwendet. Im MPEG-2-Codierungsverfahren ist diese Maßnahme nicht erforderlich, da dort die Blockgrenzen der Bewegungskompensation und der DCT-Transformation zusammenfallen.

Zur Vermeidung eines Anstiegs der Datenrate für die Seiteninformation sollen in Base- und Enhancementlayer die gleichen Displacementvektordinformationen

verwendet werden. Die Codierung des Baselayersignals bleibt dabei weiterhin driftfrei. Aufgrund der reduzierten örtlichen Auflösung des Baselayersignals ist die Displacementvektorinformation im Baselayer für Blöcke der Größe 8×8 Bildpunkte gültig. Die Displacementvektor-Amplitudenauflösung entspricht in diesem Raster $1/4$ Bildpunkt. Da diese Parameter vom MPEG-2-Verfahren nicht verarbeitet werden können, muss die MPEG-2-Codierung im Baselayer erweitert werden. Die Erweiterung besteht aus einem Interpolationsfilter, das auch eine Bewegungskompensation mit $1/4$ -Bildpunkt Amplitudenauflösung der Displacementvektoren und einer Blockgröße von 8×8 Bildpunkten erlaubt. Diese Erweiterung wurde auch als Vorschlag in die ISO-Standardisierung eingebracht und ist in den aktuellen MPEG-4-Standard [9] aufgenommen worden, da sie auch die Codierungseffizienz verbessert [12].

Die in beiden Layern gemeinsam genutzten Displacementvektoren können entweder unter Verwendung des Baselayersignals oder des hochaufgelösten Signals geschätzt werden. Um die effizientere Variante zu bestimmen, wird der folgende Vergleich durchgeführt. Es werden zwei Displacement-Vektorfelder gleicher Amplitudenauflösung generiert: \vec{d}_0 , das unter Verwendung des hochaufgelösten Signals s_0 , sowie \vec{d}_1 , das unter Verwendung des Baselayersignals s_1^\dagger geschätzt wird. Die Codierung von Base- und Enhancementlayer wird für den Vergleich jeweils mit beiden Displacement-Vektorfeldern durchgeführt.

Der Vergleich der Codierung bei einer Datenrate von jeweils 3 Mbit/s für Base- und Enhancementlayer liefert das folgende Ergebnis: Die Codierung unter Verwendung der Vektoren \vec{d}_0 führt im Vergleich zur Codierung unter Verwendung der Vektoren \vec{d}_1 zu einer Verminderung des PSNR um 0,6 dB im rekonstruierten Baselayersignal. Im rekonstruierten hochaufgelösten Signal führen beide Vektorfelder in etwa zu gleichen Ergebnissen.

Im Allgemeinen liefert die Schätzung unter Verwendung des hochaufgelösten Signals präzisere Vektoren, da dieses Signal mehr hochfrequente Anteile enthält. Im Baselayersignal entstehen jedoch durch die Tiefpassfilterung und Abtastratenreduktion Aliasinganteile. Dieses Aliasing wird bei der Schätzung der Vektoren \vec{d}_1 berücksichtigt, da diese Vektoren direkt aus dem aliasingbehafteten Baselayersignal generiert werden. Bei der Schätzung der Vektoren \vec{d}_0 hingegen ist das durch die Tiefpassfilterung und Abtastratenreduktion erzeugte Aliasing nicht vorhanden, wodurch sich bei der Anwendung dieser Vektoren auf das aliasingbehaftete Baselayersignal ein deutlicher Anstieg der Prädiktionsfehlerleistung ergibt.

Aufgrund dieser Ergebnisse wird im entwickelten skalierbaren Teilband-DCT-Codierungsverfahren die Displacementvektorinformation \vec{d}_1 verwendet, die aus dem Baselayersignal gewonnen wird (siehe Abb. 6.1).

Zusätzlich zu den Displacementvektoren muss die Information über die verwendeten MPEG-2-Makroblocktypen signalisiert werden. In MPEG-2 wird für

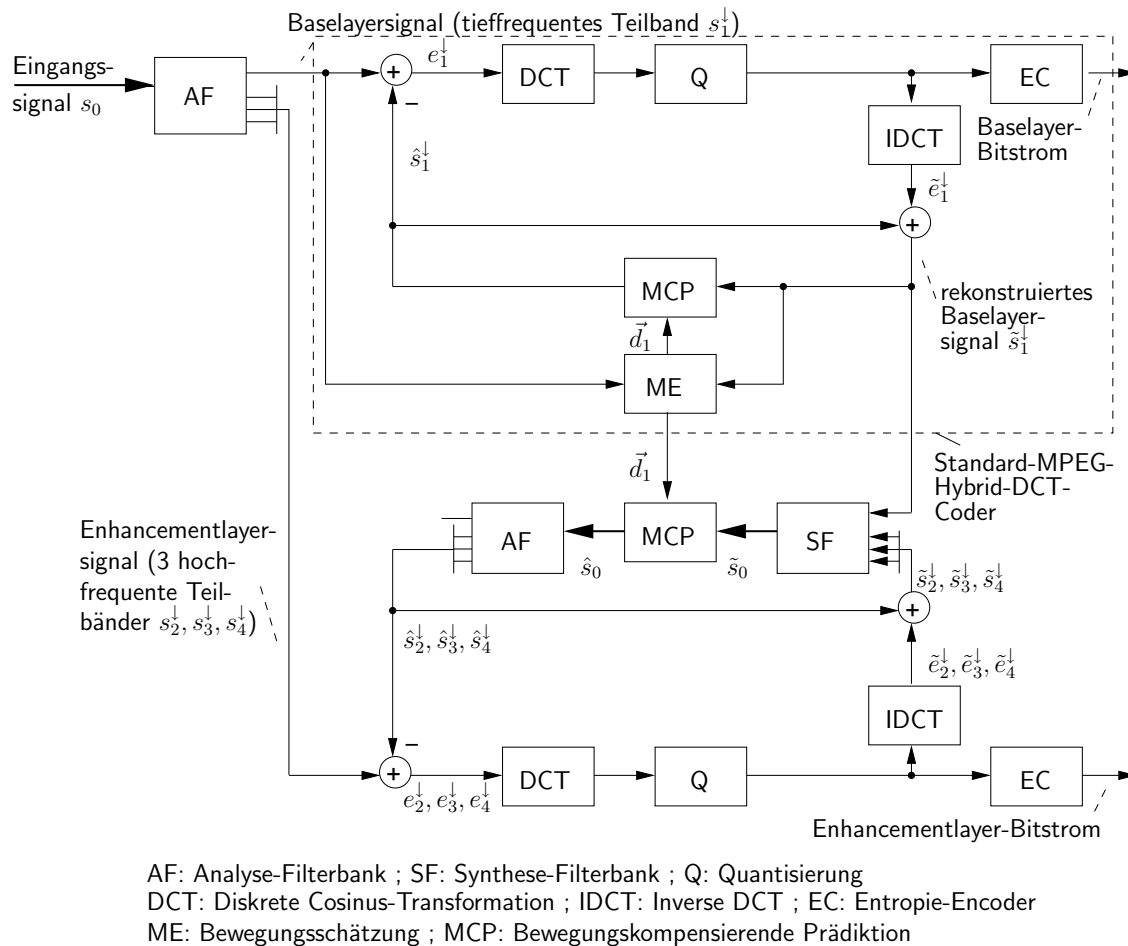


Abbildung 6.1.: Blockschaltbild des Encoders mit Schätzung der gemeinsamen Displacementvektoren unter Verwendung des Baselayersignals

jeden Block von 16x16 Bildpunkten ein Makroblocktyp übertragen. Dies wird auch im Baselayer des entwickelten Verfahrens beibehalten. Um die notwendige Seiteninformation gering zu halten, werden die Makroblocktyp-Informationen ausschließlich im Baselayer-Bistrom übertragen und auch für die korrespondierenden Blöcke von 32x32 Bildpunkten des Enhancementlayers verwendet.

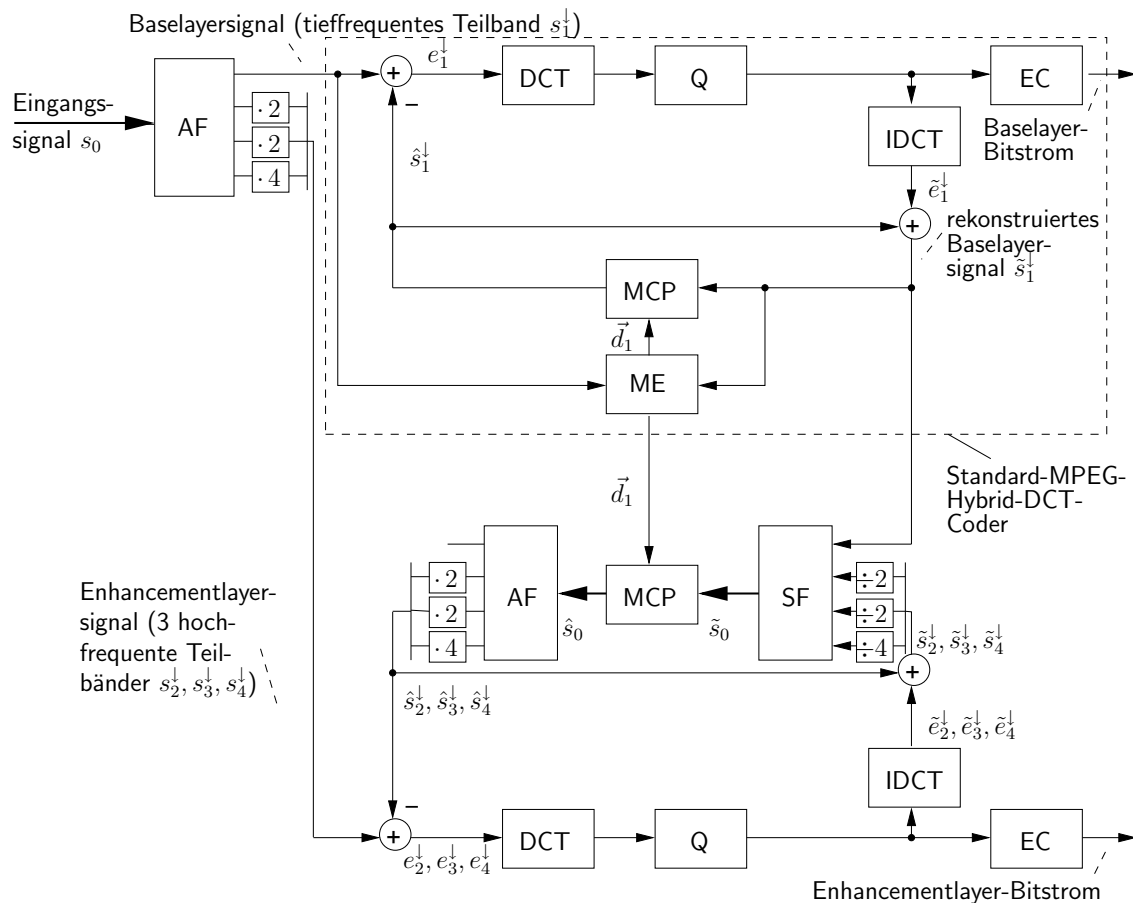
6.2. Vorverstärkung der hochfrequenten Teilbänder

Die Abtastwerte der hochfrequenten Teilbänder weisen eine geringere Varianz auf als die Abtastwerte des tieffrequenten Teilbandes. Da die DCT im entwickelten System in Anlehnung an die MPEG-2-Codierung ganzzahlige Eingangswerte in 8-bit-Darstellung verarbeitet, ergibt sich für die hochfrequenten Teilbänder ein schlechterer Signal-Rausch-Abstand als für das tieffrequente Teilband. Um diesen ungleichen Signal-Rausch-Abstand zu kompensieren, werden die hochfrequenten Teilbänder im entwickelten System nach der Analysefilterung verstärkt, siehe Abb. 6.2. Die Verstärkung um den Faktor zwei für die Teilbandsignale s_2^\downarrow und s_3^\downarrow sowie um den Faktor vier für das Teilbandsignal s_4^\downarrow wird vor der Synthesefilterung wieder rückgängig gemacht, so dass die Bedingung der exakten Rekonstruktion erfüllt bleibt.

6.3. Entwurf der Wichtungsmatrizen für die Quantisierer

Ein übertragenes Bildsignal muss nicht exakt mit dem Originalsignal übereinstimmen, um einen originalgetreuen Bildeindruck beim Betrachter hervorzurufen, da das menschliche visuelle System bestimmte Unterschiede nicht wahrnehmen kann. Die so genannte Irrelevanzreduktion nutzt diese Tatsache, um die zur Übertragung von Bildsignalen notwendige Datenmenge zu verringern. Dabei ist im Besonderen die Frequenzabhängigkeit der Kontrastempfindlichkeit zu nennen, die auch als Modulations-Übertragungsfunktion (Modulation Transfer Function, MTF) des menschlichen visuellen Systems bezeichnet wird [28]. Sie zeigt eine zu höheren Ortsfrequenzen hin stark abfallende Empfindlichkeit. Aus diesem Grund können hochfrequente Anteile im Bildsignal gröber quantisiert werden als tieffrequente, ohne dass sich die subjektive Bildqualität verschlechtert.

Um die MTF des menschlichen visuellen Systems zur Datenreduktion zu nutzen,



AF: Analyse-Filterbank ; SF: Synthese-Filterbank ; Q: Quantisierung
 DCT: Diskrete Cosinus-Transformation ; IDCT: Inverse DCT ; EC: Entropie-Encoder
 ME: Bewegungsschätzung ; MCP: Bewegungskompensierende Prädiktion

Abbildung 6.2.: Blockschaubild des Encoders nach Abbildung 6.1 mit Vorverstärkung der hochfrequenten Teilbänder

verwendet das MPEG-2-Verfahren zwei Quantisierer-Wichtungsmatrizen für die Koeffizienten der 8x8 DCT, jeweils eine für intra-codierte Makroblöcke (Intra-Matrix) und eine für prädiktiv codierte Makroblöcke (Inter-Matrix). Dabei wird ausgenutzt, dass es sich bei den DCT-Koeffizienten um eine Spektraldarstellung des Signals handelt, und man durch Gewichtung der Koeffizientenquantisierung eine spektrale Gewichtung der Quantisierungsfehler erreicht.

Jeder der 8x8 DCT-Koeffizienten wird mit einem gleichförmigen Quantisierer verarbeitet. Die Stufenbreite für die einzelnen Koeffizienten ist jedoch unterschiedlich. Sie wird mittels der Elemente der Quantisierer-Wichtungsmatrizen eingestellt und ergibt sich durch Multiplikation des jeweiligen Matrixelementes $w_{i,j}$ mit dem zur Regelung der Datenrate verwendeten globalen Quantisierungsparameter QP . Dabei kennzeichnet i die horizontale und j die vertikale Ortsfrequenz, die Ortsfrequenzen wachsen mit steigendem i bzw. j . Durch die unterschiedliche Stufenbreite $w_{i,j} \cdot QP$ erfahren die hochfrequenten Koeffizienten eine gröbere Quantisierung als die tieffrequenten (siehe Tabellen 6.1 und 6.2). Dies entspricht den oben genannten Eigenschaften des menschlichen visuellen Systems, welches im hochfrequenten Bereich weniger empfindlich ist als im tieffrequenten Bereich.

Da im zu entwickelnden skalierbaren Teilband-DCT-Codierungsverfahren die Teilbandsignale aufgrund der Filterung und Abtastratenreduktion nur noch Teilspektren des Eingangssignals enthalten, sind die MPEG-Matrizen für die Teilbandsignale nicht mehr gültig. An Stelle der beiden MPEG-Matrizen werden zur Quantisierung der vier Teilbandsignale jeweils vier unterschiedliche Intra- und Inter-Matrizen benötigt.

Aus den 8x8 Koeffizienten einer MPEG-Matrix, die für das Spektrum des Gesamtsignals gelten, müssen also vier 8x8-Matrizen für jeweils genau ein Viertel des Gesamtspektrums erzeugt werden. In einem ersten Ansatz wurden zu diesem Zweck die beiden 8x8 MPEG-Matrizen durch Einfügen von bilinear interpolierten Zwischenwerten zu 16x16 Hilfsmatrizen erweitert. Diese wurden in horizontaler und vertikaler Richtung in der Mitte geteilt, wodurch je vier 8x8 Teilmatrizen für jeweils ein Viertel des Spektrums entstehen.

Es zeigte sich jedoch, dass bei Anwendung des gleichen globalen Quantisierungsparameters QP in Verbindung mit den jeweils vier Teilmatrizen das PSNR der rekonstruierten Signale bei Teilbandzerlegung um 0,8 bis 1,3 dB geringer ist.

Die Ursache kann nur darin liegen, dass die durch bilineare Interpolation erzeugten Matrixelemente ungeeignet sind.

In einem zweiten Ansatz wurden deshalb zur Ermittlung geeigneter Matrizen zwei Messvorrichtungen aufgebaut, siehe Abb. 6.3. Als Eingangssignal für die Messvorrichtungen dienen das hochaufgelöste Eingangssignal s_0 (zur Bestimmung der Intra-Matrix) bzw. das nach der bewegungskompensierenden Prädiktion des hochaufgelösten Signals entstehende Prädiktionsfehlersignal e_0 (zur Be-

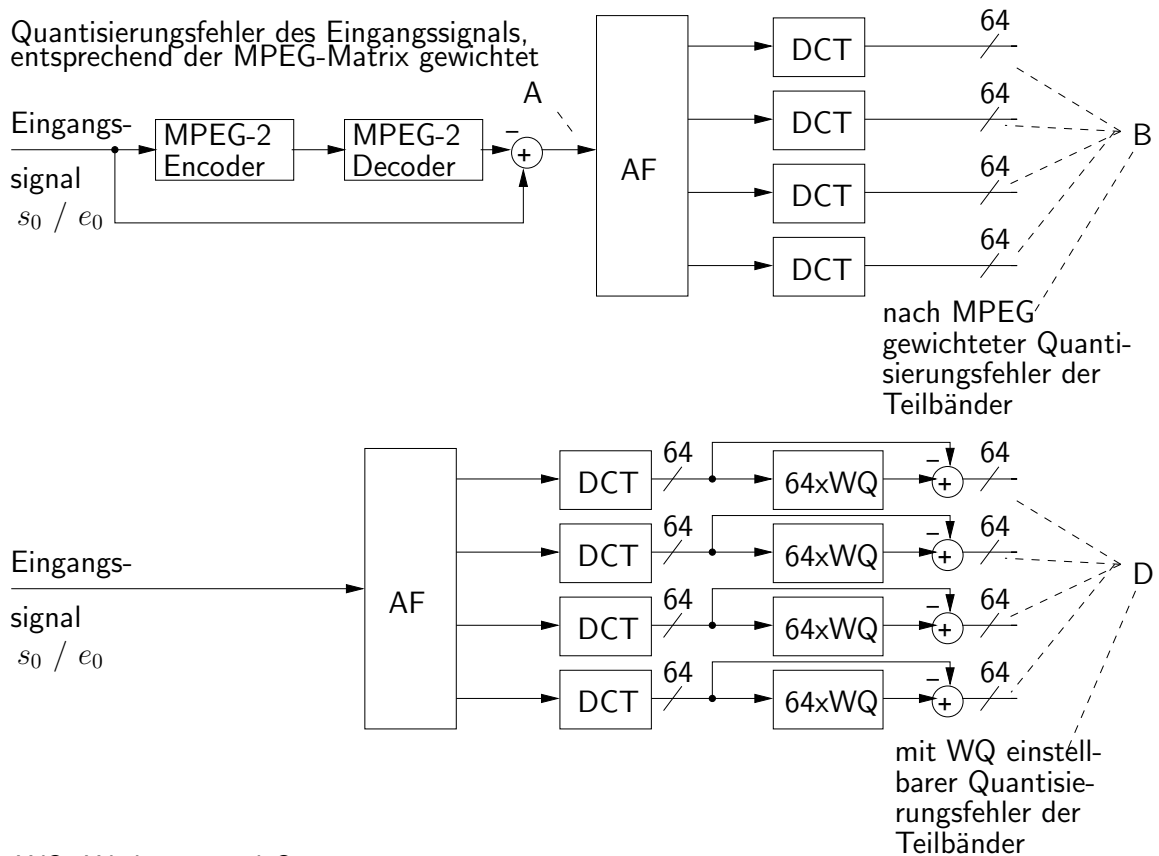
stimmung der Inter-Matrix).

Die erste Messvorrichtung besteht aus einem MPEG-2-Encoder und -Decoder, dessen rekonstruiertes Signal vom Eingangssignal subtrahiert wird. Damit liegt an Punkt "A" in Abb. 6.3 der entsprechend der Intra- bzw. Inter-MPEG-Matrix gewichtete Quantisierungsfehler an. Dieses Signal wird anschließend mit der in Abschnitt 5 entwickelten Analysefilterbank in Teilbandsignale zerlegt. Jedes der vier Teilbandsignale wird mit der 8x8 DCT transformiert. Damit kann an Punkt "B" gemessen werden, wie sich der gewichtete Quantisierungsfehler der MPEG-2-Codierung auf die DCT-Koeffizienten der Teilbandsignale verteilt.

Die zweite Messvorrichtung besteht aus der Teilbandzerlegung mit nachgeschalteter DCT der Teilbandsignale, sowie der Wichtung und Quantisierung der Teilband-DCT-Koeffizienten. Die gewichteten und quantisierten Teilband-DCT-Koeffizienten werden von den unquantisierten Koeffizienten subtrahiert. Die Elemente der zu bestimmenden Intra- bzw. Inter-Matrizen werden so eingestellt, dass die mittleren Quantisierungsfehlerleistungen der insgesamt jeweils $4 \times 64 = 256$ Teilband-DCT-Koeffizienten mit denen an Punkt "B" der ersten Messvorrichtung übereinstimmen.

Durch Verwendung eines repräsentativen Satzes von Testsignalen (ITU-R 601 Testsequenzen "Mobile&Calendar", "Flower Garden", "Bus" und "Basketball") sowie verschiedener Werte des globalen Quantisierungsparameters QP werden auf diese Weise die gesuchten Matrizen ermittelt. Die Ergebnisse der Messungen sind in den Tabellen 6.3 und 6.4 dargestellt.

Die Verstärkung der hochfrequenten Teilbänder gemäß Abschnitt 6.2 wird beim Entwurf der Quantisierer-Wichtungsmatrizen berücksichtigt, indem entsprechend der Verstärkung in der Analysefilterbank die Matrixelemente für die Teilbandsignale s_2^\downarrow und s_3^\downarrow mit dem Faktor zwei und für das Teilbandsignal s_4^\downarrow mit dem Faktor vier multipliziert werden. Diese Multiplikation ist in den Tabellen 6.3 und 6.4 bereits enthalten.



WQ: Wichtung und Quantisierung

AF: Analyse-Filterbank ; DCT: Diskrete Cosinus-Transformation

Abbildung 6.3.: Messvorrichtungen zur Ermittlung geeigneter Quantisierungsmatrizen für die Teilband-DCT-Koeffizienten

	→ i	(Index der horizontalen Ortsfrequenz)							
↓	8	16	19	22	26	27	29	34	
j	16	16	22	24	27	29	34	37	
(Index	19	22	26	27	29	34	34	38	
der	22	22	26	27	29	34	37	40	
vertikalen	22	26	27	29	32	35	40	48	
Orts-	26	27	29	32	35	40	48	58	
frequenz)	26	27	29	34	38	46	56	69	
	27	29	35	38	46	56	69	83	

Tabelle 6.1.: MPEG-2 Intra-Matrix mit den Elementen $w_{i,j}$

	→ i	(Index der horizontalen Ortsfrequenz)							
↓	16	17	18	19	20	21	22	23	
j	17	18	19	20	21	22	23	24	
(Index	18	19	20	21	22	23	24	25	
der	19	20	21	22	23	24	25	26	
vertikalen	20	21	22	23	25	26	27	28	
Orts-	21	22	23	24	26	27	28	30	
frequenz)	22	23	24	26	27	28	30	31	
	23	24	25	27	28	30	31	33	

Tabelle 6.2.: MPEG-2 Inter-Matrix mit den Elementen $w_{i,j}$

	$\longrightarrow i$	$(n = 1)$								$(n = 2)$							
↓	5	5	8	8	9	9	10	10	:	24	24	23	23	25	26	27	24
j	5	6	8	9	10	10	10	10	:	24	23	22	22	24	25	26	30
	8	8	8	9	11	10	11	11	:	26	25	24	24	26	28	28	33
	8	8	9	10	11	11	11	11	:	26	25	24	24	26	27	27	32
	9	9	11	11	12	12	11	12	:	28	26	26	25	26	28	27	34
	9	9	10	11	12	11	11	11	:	27	25	24	25	25	28	26	35
	9	10	10	10	11	11	11	11	:	26	24	23	24	25	28	26	48
	9	9	10	10	11	11	11	12	:	26	25	24	25	29	40	39	48
																
	22	22	25	25	26	25	25	26	:	58	55	53	55	58	68	56	96
	21	21	23	23	24	24	23	24	:	55	52	49	50	54	62	53	96
	20	20	22	22	22	22	22	24	:	53	49	50	53	56	72	57	96
	21	20	21	21	22	22	22	25	:	54	51	53	62	79	96	96	96
	20	19	19	20	21	22	23	27	:	55	54	59	75	96	96	96	96
	23	23	24	24	25	26	27	40	:	62	64	73	96	96	96	96	96
	20	19	20	20	21	22	25	35	:	59	63	70	96	96	96	96	96
	18	23	24	24	25	28	33	48	:	73	80	96	96	96	96	96	96
				$(n = 3)$				$(n = 4)$									

Tabelle 6.3.: Ermittelte vier Intra-Teilband-Matrizen mit den Elementen $w_{i,j}^n$

	\longrightarrow	i								\vdots	$(n = 1)$								\vdots	$(n = 2)$							
↓		8	8	8	8	9	9	8	8	\vdots	20	20	19	19	20	20	20	25									
j		8	8	8	8	9	9	8	8	\vdots	20	19	19	18	19	19	20	24									
		8	8	9	9	9	9	9	8	\vdots	21	20	20	20	21	21	22	28									
		9	8	9	9	9	9	9	9	\vdots	22	20	20	20	20	21	22	33									
		9	9	9	9	9	10	9	9	\vdots	22	21	21	20	21	22	23	47									
		9	9	9	9	9	9	9	9	\vdots	22	21	21	20	20	22	23	48									
		8	8	9	9	9	9	9	8	\vdots	22	20	20	20	21	22	23	48									
		8	8	9	9	9	9	9	8	\vdots	22	20	20	20	21	24	27	48									
.....																											
		20	20	22	22	22	22	22	22	\vdots	52	50	49	49	50	55	57	96									
		19	19	20	21	21	20	20	20	\vdots	50	47	45	44	47	50	56	96									
		18	18	20	20	20	20	19	19	\vdots	48	44	43	43	44	53	55	96									
		18	18	19	19	19	19	19	19	\vdots	47	44	44	45	45	60	85	96									
		18	18	18	19	20	20	19	20	\vdots	49	45	44	45	47	56	71	96									
		20	20	22	22	23	23	22	22	\vdots	53	49	49	49	52	61	89	96									
		19	19	20	21	21	21	21	21	\vdots	51	48	47	48	50	53	63	96									
		22	23	23	24	24	24	22	24	\vdots	58	53	53	54	56	58	70	96									
		$(n = 3)$								\vdots	$(n = 4)$																

Tabelle 6.4.: Ermittelte vier Inter-Teilband-Matrizen mit den Elementen $w_{i,j}^n$.

7. Entwurf der Quantisierung und Codierung für die SNR-Skalierung

7.1. SNR-Skalierung des Baselayersignals

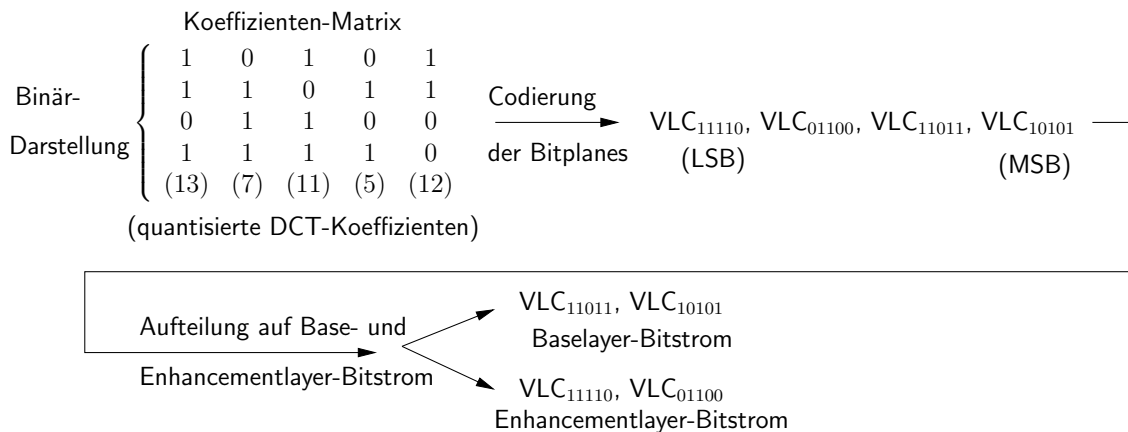
Das dritte Problem, das bei der Entwicklung des skalierbaren Teilband-DCT-Codierungsverfahrens gelöst werden muss, betrifft die flexible Datenratenaufteilung zwischen Base- und Enhancementlayer-Bitstrom. Erste Messungen am entwickelten skalierbaren Teilband-DCT-Codierungsverfahren zeigten, dass zwischen 70% und 80% der gesamten Datenrate zur Codierung des tieffrequenten Teilbandes im Baselayer benötigt werden. Die Ursache dafür ist, dass sich der Hauptteil der Energie des Videosignals auf das tieffrequente Teilband konzentriert. Zusätzlich werden sämtliche Seiteninformationen im Baselayer-Bitstrom übertragen. Im Hinblick auf zukünftige Anwendungen der skalierbaren Codierung ist jedoch eine flexible Datenratenaufteilung mit einem maximalem Baselayer-Anteil von 50% an der Gesamtdatenrate anzustreben[27].

Zu diesem Zweck wird eine zusätzliche SNR-Skalierung des Baselayers eingeführt. Dabei wird das Baselayersignal in zwei Teilsignale aufgeteilt: Ein Baselayer-Basisignal und ein Baselayer-Zusatzsignal. Das Baselayer-Zusatzsignal wird im Enhancementlayer-Bitstrom übertragen.

Zur Realisierung der SNR-Skalierung wird in dieser Arbeit die in MPEG-4 für die "Fine Granularity Scalability" (FGS) [10] vorgeschlagene "Bitplane"-Codierung [11] für die DCT-Koeffizienten eingesetzt. Diese Codierung verwendet zur Quantisierung der DCT-Koeffizienten so genannte "embedded quantizers" [15]. Zur Codierung wird zunächst eine Binärdarstellung der Koeffizienten verwendet: Die unter Verwendung des MPEG-Quantisierungsparameters QP quantisierten DCT-Koeffizienten werden entsprechend ihrer Zahlendarstellung im Binärsystem spaltenweise in eine Koeffizientenmatrix geschrieben (siehe Abb. 7.1). Für die Bitplanecodierung wird diese Koeffizientenmatrix beginnend mit dem "Most Significant Bit" (MSB) zeilenweise verarbeitet. Dabei werden die Nullen jeder Zeile ("Bitplane") mittels einer Lauflängendarstellung und Huffman-Codierung codiert.

Durch die Verwendung der Binärdarstellung ergibt sich mit jeder übertra-

genen "Bitplane" eine Halbierung des Quantisierungsintervalls. Der Encoder entscheidet zur SNR-Skalierung, welche der Bitplanes im Baselayer- bzw. im Enhancementlayer-Bitstrom übertragen werden. So ergibt sich bei der Decodierung des Baselayer-Bitstroms aus dem Beispiel in Abb. 7.1 eine um den Faktor vier (2^2) höhere Quantisiererstufenbreite als bei der Decodierung von Base- und Enhancementlayer-Bitstrom, da der Enhancementlayer-Bitstrom zwei zusätzliche Bitplanes enthält.



VLC_{nnnn}: Variable-Längen-Codewörter der Bitplanes

MSB: höchstwertige Bitplane (Most Significant Bitplane)

LSB: niederwertigste Bitplane (Least Significant Bitplane)

Abbildung 7.1.: Beispiel für die Bitplane-Codierung mit Aufteilung auf Baselayer- und Enhancementlayer-Bitstrom

Mit dem im Enhancementlayer-Bitstrom übertragenen Baselayer-Zusatzsignal wird der Quantisierungsfehler im rekonstruierten Baselayersignal \tilde{e}_1^\downarrow reduziert, wenn das Bildsignal der hohen Auflösung rekonstruiert wird (siehe Abb. 7.2).

7.2. Effizienz der Bitplane-Codierung

In einer ersten experimentellen Untersuchung zur Codierungseffizienz wird überprüft, ob die Verwendung der Bitplane-Codierung im Vergleich zur der bei MPEG-2 verwendeten so genannten zweidimensionalen Lauflängencodierung der DCT-Koeffizienten zu einem Anstieg der Datenrate führt. Zu diesem Zweck werden die aus den ITU-R 601 Testsequenzen "Bus" und "Basketball" erzeugten Baselayersignale bei verschiedenen Einstellungen des globalen Quantisierungsparameters QP sowohl mit der zweidimensionalen Lauflängencodierung nach

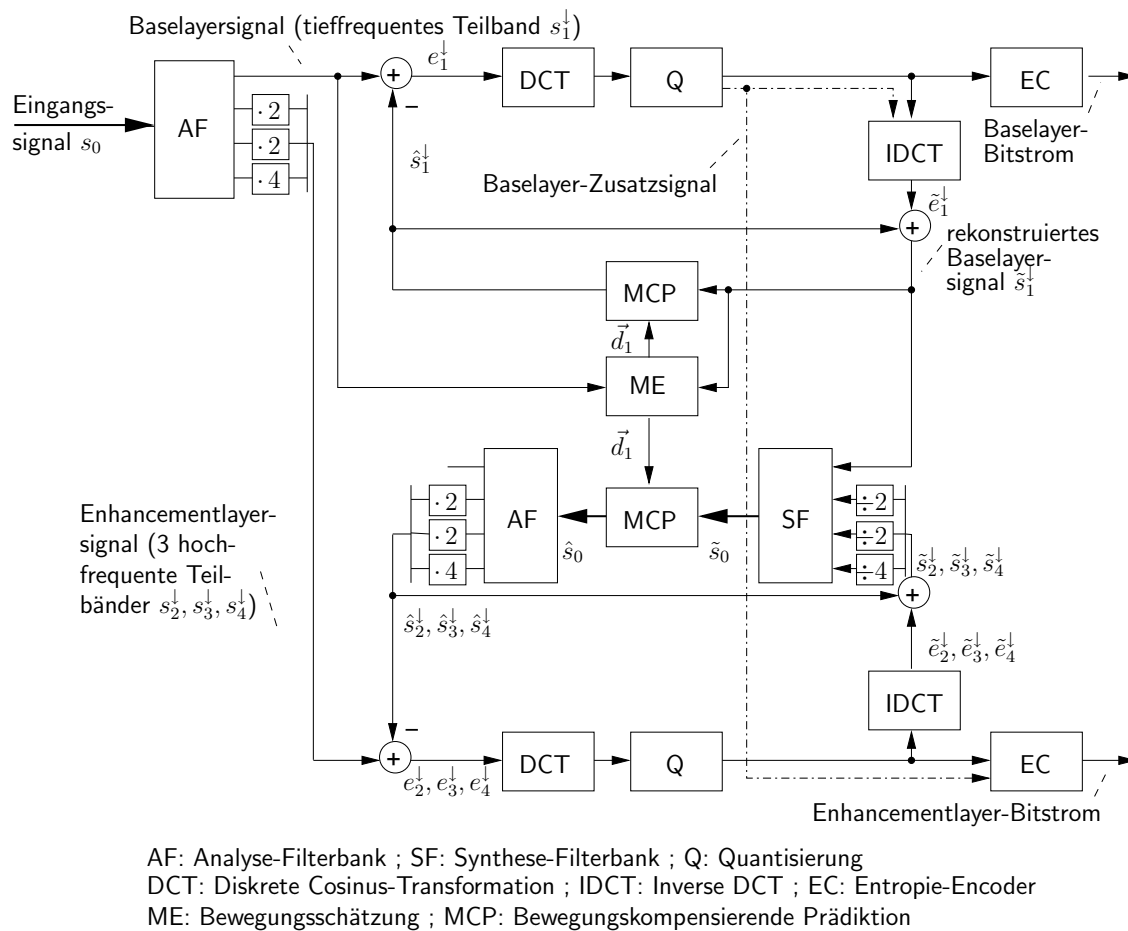


Abbildung 7.2.: Blockschaltbild des Encoders der kombinierten Teilband-DCT-Codierung mit zusätzlicher SNR-Skalierung des Baselayersignals

MPEG-2 als auch mit der Bitplane-Codierung nach MPEG-4 FGS codiert. Die Tabellen zur variablen Längencodierung nach MPEG-4 FGS sind in [11] angegeben. Da nur die Entropie-Codierung verändert wird, erzeugen beide Varianten das gleiche PSNR, aber unterschiedliche Datenraten. Die Ergebnisse sind in Tabelle 7.1 zusammengefasst.

Die Bitplane-Codierung nach MPEG-4 FGS führt für alle untersuchten Einstellungen von QP zu einer geringeren Datenrate als die zweidimensionale Lauflängencodierung nach MPEG-2. Die maximale Einsparung der Datenrate bei Verwendung der Bitplane-Codierung beträgt 7,95% gegenüber der Lauflängencodierung (Testsequenz Basketball, $QP = 12$). Aufgrund dieser Ergebnisse wird im entwickelten skalierbaren Teilband-DCT-Codierungsverfahren die Bitplane-Codierung nicht nur zur SNR-Skalierung im Baselayer, sondern auch zur Co-

QP	Bus		Basketball	
	Datenrate [Bit/Bild] MPEG-2 Codierung	Datenrate [Bit/Bild] Bitplane- Codierung	Datenrate [Bit/Bild] MPEG-2 Codierung	Datenrate [Bit/Bild] Bitplane- Codierung
12	139030	130748	186610	171777
16	100106	94701	136420	126730
20	82457	78275	112553	105361
24	70519	67074	95790	90283
28	61896	58888	83748	79258
32	54546	51990	73048	69447
36	49823	47507	66053	62988
40	46046	43964	60491	57855

Tabelle 7.1.: Datenraten der MPEG-2- und der Bitplane-Codierung, Testsequenzen "Bus" und "Basketball"

dierung der Teilband-DCT-Koeffizienten im Enhancementlayer eingesetzt. Damit besteht die Möglichkeit, auch eine SNR-Skalierung des Enhancementlayer-Signals durchzuführen.

7.3. Drift im SNR-Baselayer

Wie in Abschnitt 2.4 dargestellt, erzeugt eine Drift der Prädiktionssignale eine Verminderung der Codierungseffizienz im rekonstruierten Baselayer-Basissignal $\tilde{s}_{1,Base}^\downarrow$. Um dem entgegenzuwirken, wird ein Verfahren entwickelt, das den Einfluss von Drift im rekonstruierten Baselayer-Basissignal $\tilde{s}_{1,Base}^\downarrow$ begrenzt.

Zu diesem Zweck wird die Struktur von MPEG-2, die so genannte "Group of Pictures" (GOP), berücksichtigt. In dieser Arbeit wird die GOP-Struktur "N24M6" verwendet, siehe Abb. 7.3. In diesem Fall besteht eine GOP aus 24 Bildern. Der Parameter "M6" gibt an, dass der Abstand zwischen dem Intraframe-codierten Bild (I-frame) und dem ersten unidirektional prädizierten Bild (P-frame) bzw. zwischen zwei aufeinanderfolgenden P-frames sechs Bilder beträgt. Dazwischen werden jeweils fünf bidirektional prädizierte Bilder (B-frames) übertragen. Eine GOP besteht somit aus einem I-frame, drei P-frames und 20 B-frames. Jede GOP beginnt mit einem I-frame, um der Decodierung einen wahlfreien Zugriff im Abstand von 24 Bildern zu ermöglichen.

Drift kann sich nur ausgehend von den rekonstruierten I-frames und P-frames akkumulieren, da nur diese zur weiteren Prädiktion benutzt werden. Die B-

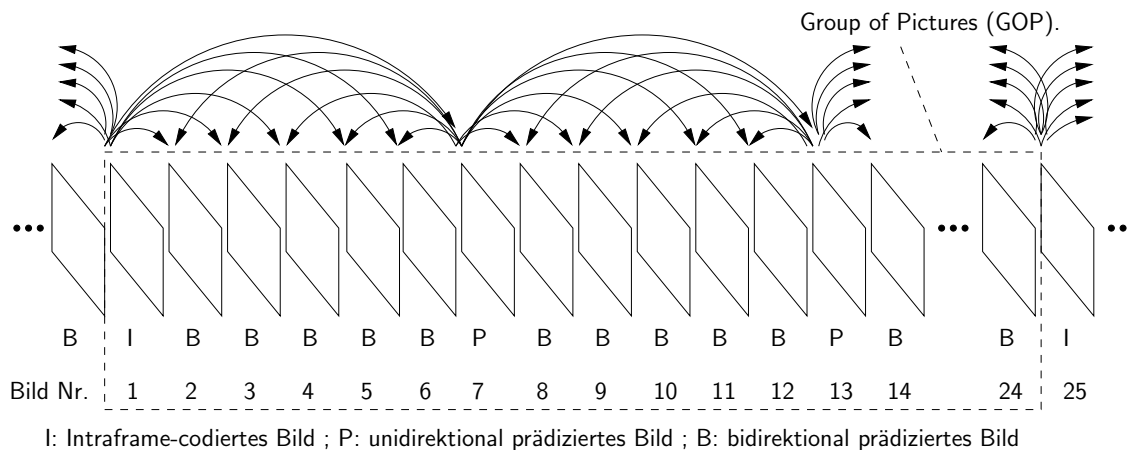


Abbildung 7.3.: Darstellung des Aufbaus und der Prädiktionsrichtungen innerhalb einer "Group of Pictures" (GOP) der Struktur "N24M6"

frames, die den größten Anteil der GOP bilden, werden nicht zur weiteren Prädiktion benutzt. Darüberhinaus wird die Drift der Prädiktionssignale mit jedem übertragenen I-frame zurückgesetzt, da die I-frames als Synchronisation wirken.

Durch eine unterschiedliche SNR-Skalierung für I-, P- und B-frames lässt sich die Akkumulation der Drift beeinflussen. Das rekonstruierte I-frame wird indirekt zur Prädiktion der 23 weiteren Bilder innerhalb der GOP verwendet, so dass sich Driftfehler ausgehend vom rekonstruierten I-frame am stärksten bemerkbar machen. Aus diesem Grund werden für die I-frames sämtliche codierte Daten des Baselayersignals im Baselayer-Bitstrom übertragen; es findet also keine SNR-Skalierung in I-frames statt.

Die drei P-frames innerhalb der GOP haben einen unterschiedlichen Einfluss auf die Akkumulation von Drift: Während das erste P-frame (Bild Nr. 7 in Abb. 7.3) zur Prädiktion der verbleibenden 22 Bilder innerhalb der GOP verwendet wird, reduziert sich diese Zahl auf 16 Bilder für das zweite P-frame (Bild Nr. 13) und auf 10 Bilder für das dritte P-frame (Bild Nr. 19). Daher wird eine abgestufte SNR-Skalierung für die P-frames eingeführt. Im ersten P-frame werden ca. 80% der Daten im Baselayer- und entsprechend ca. 20% im Enhancementlayer-Bitstrom übertragen. Für das zweite P-frame wird das Verhältnis auf ca. 70% zu 30% und für das dritte P-frame auf ca. 60% zu 40% eingestellt.

Die B-frames werden nicht zur Prädiktion weiterer Bilder verwendet, es ist somit keine Abstufung der SNR-Skalierung für die unterschiedlichen B-frames innerhalb der GOP notwendig. Die Einstellung der SNR-Skalierung für B-frames ergibt sich allein aus der Forderung, dass die Gesamtdatenrate des Teilband-DCT-Codierungsverfahrens im Verhältnis 50% zu 50% auf Baselayer-

und Enhancementlayer-Bitstrom aufgeteilt werden soll. Als Folge der oben genannten Einstellung der SNR-Skalierung in I- und P-frames ergibt sich daraus für die B-frames ein Datenratenverhältnis von ca. 50% für das Baselayer-Basissignal und 50% für das Baselayer-Zusatzsignal.

Im Vergleich zu einer einheitlichen SNR-Skalierung der I-, P- und B-frames ergibt sich für die entwickelte Einstellung der SNR-Skalierung nach Frametypen eine Verbesserung des PSNR um 0,4 dB bis 0,5 dB für das ausschließlich aus dem Baselayer-Bitstrom rekonstruierte Signal.

8. Experimentelle Ergebnisse

Zur Bewertung der Codierungseffizienz des entwickelten skalierbaren Teilband-DCT-Codierungsverfahrens werden Experimente unter Verwendung der ITU-R 601 Testsequenzen "Bus" und "Basketball" durchgeführt. Die Testsequenzen werden vor der Codierung mit dem in Abschnitt 3 beschriebenen De-Interlacing-Algorithmus [30] in ein progressives Format umgewandelt.

8.1. Vergleich mit dem MPEG-Codierungsverfahren

Für eine vergleichende Bewertung werden die folgenden Codierungsverfahren untersucht:

- Nichtskalierbare MPEG-2-Codierung als Referenz für die obere Grenze der Codierungseffizienz
- Simulcast-Codierung basierend auf MPEG-2 für die untere Grenze der Codierungseffizienz
- Pyramidencodierung mit MPEG-4 SSP als Referenz für eine standardisierte örtlich skalierbare Codierung
- Das entwickelte örtlich skalierbare Teilband-DCT-Codierungsverfahren einschließlich der SNR-Skalierung des Baselayersignals

Es werden zwei unterschiedliche Experimente durchgeführt:

- Codierung bei gleicher Gesamtdatenrate von 6 Mbit/s
- Codierung bei gleicher Bildqualität (gleicher Spitzen-Signal-Rausch-Abstand (PSNR) wie die nichtskalierbare MPEG-2-Codierung bei 6 Mbit/s)

In beiden Experimenten wird die Gesamtdatenrate für die Simulcast-, die MPEG-4-SSP- und die skalierbare Teilband-DCT-Codierung in jeweils 50% für Baselayer- und Enhancementlayer-Bitstrom aufgeteilt.

Die erzielten Ergebnisse für die rekonstruierten Signale \tilde{s}_0 der hohen Auflösung sind in Abb. 8.1 und Tabelle 8.1 dargestellt.

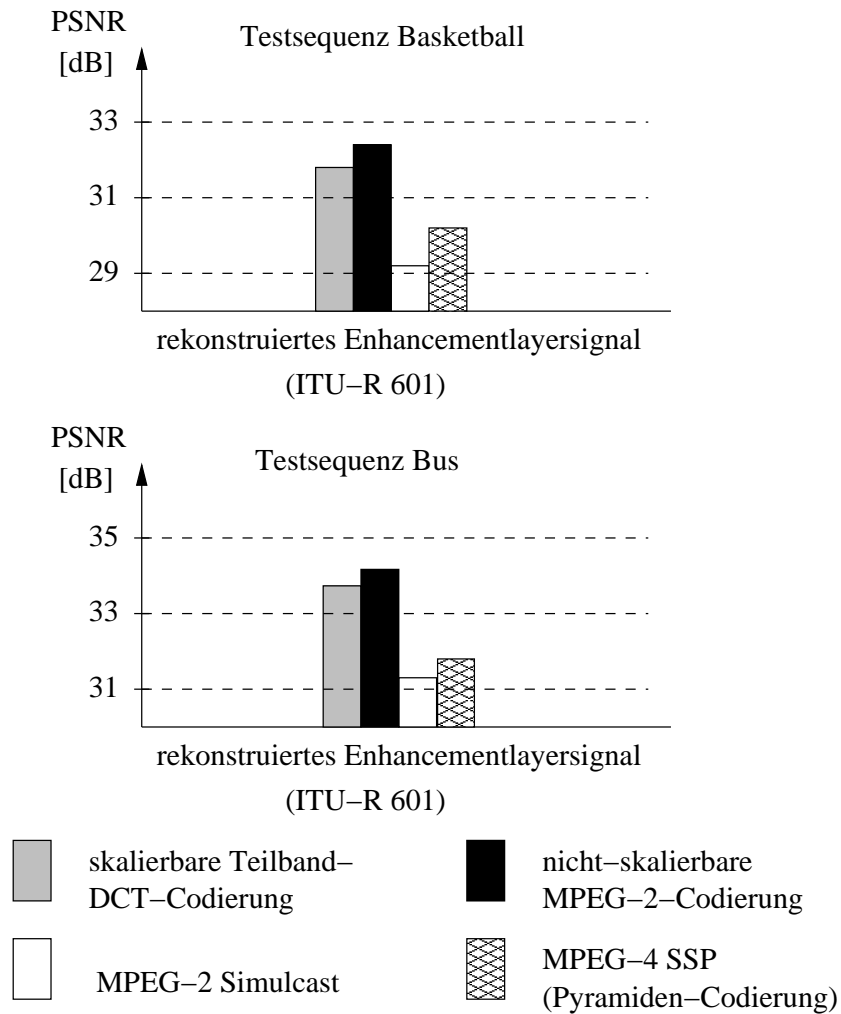


Abbildung 8.1.: Experimentelle Ergebnisse für das rekonstruierte Signal \tilde{s}_0 der hohen Auflösung bei Codierung mit gleicher Gesamtdatenrate von 6 Mbit/s

Testsequenz Basketball				
PSNR [dB]	Datenrate [Mbit/s] MPEG-2 nicht-skalierbar	Datenrate [Mbit/s] MPEG-2 Simulcast	Datenrate [Mbit/s] MPEG-4 Spatial Scalable	Datenrate [Mbit/s] skalierbare Teilband-DCT-Codierung
31,6	6,0	12,0	8,5	6,29

Testsequenz Bus				
PSNR [dB]	Datenrate [Mbit/s] MPEG-2 nicht-skalierbar	Datenrate [Mbit/s] MPEG-2 Simulcast	Datenrate [Mbit/s] MPEG-4 Spatial Scalable	Datenrate [Mbit/s] skalierbare Teilband-DCT-Codierung
34,1	6,0	12,0	8,93	6,54

Tabelle 8.1.: Experimentelle Ergebnisse für die Codierung bei gleichem PSNR des rekonstruierten Signals \tilde{s}_0 der hohen Auflösung (Bezugspunkt: nichtskalierbare MPEG-2-Codierung bei 6 Mbit/s)

8.2. Bewertung der Ergebnisse

Bezogen auf das hochaufgelöste Signal führt das standardisierte MPEG-4 SSP-Verfahren zu einer Verschlechterung im PSNR von 2,2 dB bzw. 2,4 dB bei gleicher Datenrate im Vergleich zur nichtskalierbaren MPEG-2-Codierung (vgl. Abb. 8.1). MPEG-4 SSP liegt dabei lediglich 0,5 dB bis 1,0 dB über der Simulcast-Codierung. Um den PSNR-Wert der nichtskalierbaren MPEG-2-Codierung zu erreichen, muss bei MPEG-4 SSP die Gesamtdatenrate von 6 Mbit/s auf mehr als 8,5 Mbit/s erhöht werden (vgl. Tabelle 8.1). Dies entspricht einer Datenratenerhöhung um mehr als 40%.

Im Gegensatz dazu erreicht das entwickelte skalierbare Teilband-DCT-Codierungsverfahren ein PSNR, das lediglich um 0,5 dB bis 0,6 dB unter dem PSNR der nichtskalierbaren MPEG-2-Codierung liegt (vgl. Abb. 8.1). Es liegt hingegen 2,4 dB bis 2,6 dB oberhalb der Simulcast-Codierung.

Zum Erreichen des PSNR der nichtskalierbaren MPEG-2-Codierung bei 6 Mbit/s benötigt das entwickelte Verfahren eine Datenratenerhöhung von lediglich 5%-9% auf 6,29 bzw. 6,54 Mbit/s (vgl. Tabelle 8.1). Damit ist das eingangs gesetzte Ziel einer maximalen Datenratenerhöhung von 10% im Vergleich zur nichtskalierbaren Codierung erreicht.

Die PSNR-Ergebnisse für das rekonstruierte Baselayersignal sind in Tabelle 8.2 dargestellt. Dabei wird die erhöhte Codierungseffizienz von MPEG-4 gegenüber

Testsequenz Basketball		
PSNR [dB] MPEG-2 Simulcast	PSNR [dB] MPEG-4 Spatial Scalable	PSNR [dB] skalierbare Teilband- DCT-Codierung
32,47	33,4	32,46

Testsequenz Bus		
PSNR [dB] MPEG-2 Simulcast	PSNR [dB] MPEG-4 Spatial Scalable	PSNR [dB] skalierbare Teilband- DCT-Codierung
31,27	32,01	30,97

Tabelle 8.2.: Experimentelle Ergebnisse für das rekonstruierte Baselayersignal \hat{s}_1^\downarrow bei Codierung mit einer Baselayerdatenrate von 3 Mbit/s

MPEG-2 durch einen Anstieg des PSNR um 0,8 dB bis 0,9 dB deutlich. Die PSNR-Ergebnisse des entwickelten skalierbaren Teilband-DCT-Codierungsverfahrens liegen aufgrund der driftbehafteten SNR-Skalierung des Baselayersignals unterhalb der MPEG-2-Codierung. Die Einbußen von 0.01-0.3 dB im rekonstruierten Baselayersignal aufgrund von Drift führen jedoch zu keiner sichtbaren Verschlechterung der subjektiven Bildqualität.

9. Zusammenfassung

In der vorliegenden Arbeit wird ein Verfahren zur effizienten skalierbaren Videocodierung mit zwei Stufen örtlicher Auflösung entwickelt und untersucht, das den aufgrund der Skalierung entstehenden Anstieg der Datenrate verglichen mit einer nichtskalierbaren MPEG-2-Codierung von mehr als 40% auf weniger als 10% verringert. Das Verfahren verwendet eine 4-Band-Teilbandzerlegung mit exakter Rekonstruktion des hochaufgelösten Eingangssignals. Dabei werden das tieffrequente Teilband im Baselayer- und die hochfrequenten Teilbänder im Enhancementlayer-Bitstrom übertragen. Sowohl im Baselayer als auch im Enhancementlayer wird eine bewegungskompensierende Hybrid-DCT-Codierung ähnlich dem MPEG-2-Standard verwendet. Um verschiedene Anwendungen unterstützen zu können, sollte die Aufteilung der Datenrate möglichst flexibel sein und eine Baselayer-Datenrate von bis zu 50% der Gesamtdatenrate erlauben.

Dazu waren im Wesentlichen drei Probleme zu lösen:

- Entwurf der Teilbandfilter mit den Forderungen der Bewahrung einer guten subjektiven Bildqualität des tieffrequenten Teilbandes durch Vorgabe eines Toleranzschemas für die Übertragungsfunktion des Tiefpassfilters sowie einer hohen Codierungseffizienz des gesamten Teilband-Codierungsverfahrens
- Anpassung der in Base- und Enhancementlayer verwendeten MPEG-2-Codierungen an die Teilbandsignale
- Einführung einer SNR-Skalierung für das Baselayersignal zur flexiblen Aufteilung der Datenrate

Durch die Vorgabe eines zeilenverschränkten ("interlaced") Eingangssignals gemäß ITU-R 601 ergibt sich für die örtlich skalierbare Codierung ein zusätzliches Problem: Infolge der Abtastratenreduktion nach der Tiefpassfilterung entsteht im Baselayersignal ein zeilenverschränktes Abtastraster mit geometrisch ungleichmäßigem Zeilenabstand [29]. Bei Wiedergabe des Baselayersignals auf einem Monitor niedriger örtlicher Auflösung macht sich dieser ungleichmäßige Zeilenabstand als starke Störung bemerkbar. Um dies zu verhindern, wird im standardisierten MPEG-2 "Spatial Scalable Profile" (SSP) vor der Filterung des Eingangssignals ein so genanntes De-Interlacing [30] durchgeführt, bei dem jedes Halbbild mittels Zeileninterpolation zu einem Vollbild ergänzt wird. Aus diesem

progressiven Bildsignal wird durch Tiefpassfilterung und Abtastratenreduktion ein progressives Signal der niedrigen Auflösung erzeugt. Durch einfaches Auslassen jeder zweiten Zeile (Re-Interlacing) lässt sich daraus ein zeilenverschränktes Signal niedriger Auflösung mit geometrisch gleichmäßigem Zeilenabstand gewinnen, das codiert und im Baselayer-Bitstrom übertragen wird.

Im MPEG-2 SSP wird das rekonstruierte Baselayersignal zur Prädiktion des Signals der hohen Auflösung verwendet. Aus diesem Grund ist auch im Decoder vor der Abtastratenerhöhung ein De-Interlacing erforderlich. Da das De-Interlacing sehr aufwändig ist, wird in dieser Arbeit eine alternative Lösung vorgeschlagen, bei der kein De-Interlacing im Decoder benötigt wird, indem durchgängig ein progressives Bildsignal verarbeitet wird.

Da das progressive Bildsignal die doppelte Anzahl Abtastwerte verglichen mit dem zeilenverschränkten Eingangssignal besitzt, erscheint dieses Vorgehen zunächst nachteilig, weil eine Erhöhung der zu übertragenden Datenrate zu erwarten ist. Um diesen Nachteil beurteilen zu können, wurde eine Voruntersuchung durchgeführt. Unter Verwendung des normalen MPEG-2-Codierungsverfahrens wurden sowohl das zeilenverschränkte Eingangssignal als auch ein daraus durch De-Interlacing gewonnenes progressives Bildsignal codiert und übertragen. Nach der Decodierung wurde jede zweite Zeile des progressiven Bildsignals wieder entfernt und die jeweils erforderliche Datenrate bei gleichem PSNR beider decodierter zeilenverschränkter Signale verglichen. Die Ergebnisse zeigen, dass im Fall der Codierung des progressiven Bildsignals bei einer hohen Bildqualität entsprechend einem PSNR von 34,3 dB die Datenrate lediglich um 9% größer ist. Bei geringerer Bildqualität entsprechend einem PSNR von 32,6 dB ist kein Anstieg der Datenrate mehr zu beobachten. Bei einem noch geringeren PSNR benötigt die Codierung des progressiven Bildsignals sogar eine niedrigere Datenrate als die zeilenverschränkte Codierung.

Der geringe Unterschied der Datenraten erklärt sich dadurch, dass die vom De-Interlacing hinzugefügten Zeilen lediglich mathematisch berechenbare, redundante Information beinhalten. Der Encoder ist in der Lage, diese Redundanz wieder zu entfernen und das progressive Bildsignal trotz doppelter Anzahl Abtastwerte bei nahezu gleicher Datenrate zu codieren wie das zeilenverschränkte Eingangssignal. Hinzu kommen eine Einsparung von Seiteninformation und eine bessere Datenreduktion der Intraframe-Codierung bei der Verwendung eines progressiven Bildsignals.

Diese Ergebnisse legen es nahe, in Zukunft nur noch progressive Bildsignale zu übertragen, mit dem Vorteil, dass sowohl konventionelle zeilenverschränkte Monitore als auch progressive Computer- und Flachbild-Monitore direkt zur Anzeige des Signals benutzt werden können.

Die Erkenntnisse dieser Voruntersuchung werden nachfolgend in die Entwicklung des örtlich skalierbaren Teilband-Codierungssystems übernommen, indem dort nur progressive Signale verarbeitet werden.

Zur Lösung des ersten Aufgabenpunktes dieser Arbeit geht der Teilbandfilter-Entwurf von einem Tiefpassfilter ähnlich dem MPEG-SIF-Filter mit vorgegebenem Toleranzschema aus, um die im Baselayer gewünschte Bildqualität zu gewährleisten. Das Tiefpassfilter wird entsprechend dem MPEG-SIF-Filter mit einer symmetrischen Impulsantwort der Länge sieben gewählt. Bei der Festlegung des Toleranzschemas für die Übertragungsfunktion werden als Grenzwerte für die Abweichung vom idealen Tiefpassfilter im Durchlass- und Sperrbereich die maximalen Abweichungen des MPEG-SIF-Filters übernommen. Die Breite des Übergangsbereiches zwischen Durchlass- und Sperrbereich wird im Vergleich zum MPEG-SIF-Filter von $3/32$ auf $5/32$ der Abtastfrequenz vergrößert, um Freiheitsgrade für die Optimierung der Codierungseffizienz zu gewinnen. Anderenfalls gäbe es als Lösung nur das MPEG-SIF-Filter.

Durch die gewählte Impulsantwortlänge von sieben ergeben sich drei freie Parameter für den Entwurf des symmetrischen Tiefpassfilters. Bei Beschränkung auf Filterkoeffizienten in einer Darstellung mit 9 Bit sowie der gewählten Erweiterung des Übergangsbereiches ergeben sich durch vollständige Suche 1715 mögliche Frequenzgänge, die das Toleranzschema einhalten. Für jedes dieser Tiefpassfilter werden die zugehörigen Filter der Analysefilterbank berechnet, und die mit diesen Filtern erzeugten Teilbandsignale anschließend mit dem MPEG-2-Verfahren codiert. Die Analysefilterbank, die zur geringsten Datenrate bei der Codierung führt, wird ausgewählt.

Diese Optimierung der Teilbandfilterbank führt im Vergleich zu einer Filterbank mit dem MPEG-SIF-Tiefpassfilter zu einer Einsparung der Datenrate von etwa 19% bei gleichzeitigem leichten Anstieg des PSNR im rekonstruierten hochauflösenden Signal. Auch die subjektive Bildqualität des optimierten Tiefpassfilters und die des MPEG-SIF-Filters sind dabei gleich.

Das zweite in dieser Arbeit zu lösende Problem betrifft die Anpassung der in Base- und Enhancementlayer verwendeten MPEG-2-Codierungen an die Teilbandsignale. Hier waren die notwendigen Modifikationen der bewegungskompensierenden Prädiktion und die Erzeugung geeigneter Quantisierungsmatrizen zu bearbeiten.

Die entwickelte Teilband-DCT-Codierung verwendet jeweils eine bewegungskompensierende Prädiktion für den Base- und den Enhancementlayer, um eine driftfreie örtlich skalierbare Codierung zu ermöglichen. Es ist bekannt, dass eine getrennte bewegungskompensierende Prädiktion der einzelnen Teilbandsignale des Enhancementlayers zu einer verminderten Codierungseffizienz führt, da die Bewegungskompensation durch Aliasing gestört wird. Aus diesem Grund werden in dieser Arbeit die bewegungskompensierten Prädiktionssignale der drei Teilbänder des Enhancementlayers aus dem rekonstruierten Gesamtsignal der hohen Auflösung gewonnen, indem dieses bewegungskompensiert wird und das entstehende Prädiktionssignal in einer zusätzlichen Analysefilterbank in Teilban-

signale zerlegt wird.

Die bewegungskompensierende Prädiktion der entwickelten skalierbaren Teilband-DCT-Codierung verwendet die folgenden Parameter: Im Enhancementlayer wird sie auf Blöcken der Größe 16x16 Bildpunkte bezüglich des Abtastrasters der hohen Auflösung durchgeführt, die Displacementvektoren besitzen eine Auflösung von $1/2$ Bildpunkt. Diese Parameterwerte entsprechen denen des MPEG-2-Verfahrens. Da die blockbasierte Bewegungskompensation ein Prädiktionssignal generiert, das in der Teilbandzerlegung zu unerwünschten energiereichen hochfrequenten Koeffizienten führt, wird eine Prädiktion mit überlappenden Blöcken verwendet. Im MPEG-2-Codierungsverfahren ist diese Maßnahme nicht erforderlich, da dort die Blockgrenzen der Bewegungskompensation und der DCT-Transformation zusammenfallen.

Zur Vermeidung eines Anstiegs der Datenrate für die Seiteninformation wird im Base- und Enhancementlayer die gleiche Displacementvektorinformation verwendet. Aufgrund der reduzierten örtlichen Auflösung des Baselayersignals ist diese Displacementvektorinformation im Baselayer für Blöcke der Größe 8x8 gültig. Die Displacementvektor-Auflösung entspricht in diesem Raster $1/4$ Bildpunkt. Da diese Parameter vom MPEG-2-Verfahren nicht unterstützt werden, wird die MPEG-2-Codierung im Baselayer um ein spezielles Interpolationsfilter für $1/4$ -Bildpunkt-Auflösung sowie eine 8x8-Bewegungskompensation erweitert. Diese Erweiterungen wurden als Vorschläge in die ISO-Standardisierung eingebracht und sind in den aktuellen MPEG-4-Standard aufgenommen worden.

Die zur Reduzierung der Seiteninformation in beiden Layern gemeinsam genutzten Displacementvektoren können entweder unter Verwendung des Baselayersignals oder des hochaufgelösten Signals geschätzt werden. In dieser Arbeit wird gezeigt, dass die Displacementvektorschätzung unter Verwendung des hochaufgelösten Signals zu einem deutlichen Anwachsen des Rekonstruktionsfehlers im Baselayer führt. Eine Displacementvektorschätzung unter Verwendung des Baselayersignals hingegen erhöht den Rekonstruktionsfehler im Enhancementlayer kaum messbar. Aus diesem Grund wird im entwickelten System die zweite Variante verwendet.

Um die so genannte Modulations-Übertragungsfunktion (MTF) des menschlichen visuellen Systems zur Datenreduktion zu nutzen, verwendet das MPEG-2-Verfahren eine Quantisierer-Wichtungsmatrix für die Koeffizienten der 8x8 DCT. Da die Teilbandsignale jedoch nur noch Teilspektren des Eingangssignals enthalten, ist diese Matrix nicht mehr gültig. An Stelle der MPEG-Matrix werden zur Quantisierung der vier Teilbandsignale vier unterschiedliche Matrizen für die jeweils 8x8 DCT-Koeffizienten benötigt. Da jedes Teilbandsignal ein Viertel des Gesamtspektrums enthält, ist zunächst eine bilineare Interpolation der 8x8 MPEG-Matrix zu einer 16x16 Hilfsmatrix durchgeführt worden, die anschließend in vier 8x8 Matrizen für die vier Teilspektren zerlegt wurde. Die so erzeugten Matrizen

erwiesen sich jedoch als nicht geeignet. Es zeigte sich, dass die entwickelte spezielle Teilbandzerlegung bei der Ermittlung der Matrizen berücksichtigt werden muss.

Zu diesem Zweck wurden zwei Messvorrichtungen aufgebaut. Die erste Messvorrichtung besteht aus einem MPEG-2-Encoder und -Decoder, dem eine Teilbandzerlegung mit jeweils einer DCT für die vier Teilbandsignale nachgeschaltet ist. Diese Messvorrichtung zeigt, wie sich der Quantisierungsfehler der MPEG-2-Codierung auf die DCT-Koeffizienten der Teilbandsignale verteilt.

Die zweite Messvorrichtung besteht aus einer Teilbandzerlegung mit nachgeschalteter DCT der Teilbandsignale, sowie der Wichtung und Quantisierung der DCT-Koeffizienten. Die Elemente der zu bestimmenden Matrizen wurden so eingestellt, dass die mittleren Quantisierungsfehlerleistungen der Koeffizienten in den vier Teilbändern mit denen der ersten Messvorrichtung übereinstimmen.

Durch Verwendung eines repräsentativen Satzes von Testsignalen wurden auf diese Weise die Elemente der gesuchten Matrizen ermittelt.

Das dritte Problem, das in dieser Arbeit gelöst werden musste, betrifft die flexible Datenratenaufteilung zwischen Base- und Enhancementlayer-Bitstrom. Erste Messungen am entwickelten skalierbaren Teilband-DCT-Codierungsverfahren zeigten, dass zwischen 70% und 80% der gesamten Datenrate zur Codierung des tieffrequenten Teilbandes im Baselayer benötigt werden. Die Ursache dafür ist, dass sich der Hauptteil der Energie des Videosignals auf das tieffrequente Teilband konzentriert. Zusätzlich werden sämtliche Seiteninformation im Baselayer-Bitstrom übertragen. Im Hinblick auf zukünftige Anwendungen der skalierbaren Codierung ist jedoch eine flexible Datenratenaufteilung mit einem maximalem Baselayer-Anteil von 50% an der Gesamtdatenrate anzustreben.

Zu diesem Zweck wird eine zusätzliche SNR-Skalierung des Baselayers eingeführt. Zur Realisierung der SNR-Skalierung werden in dieser Arbeit die in MPEG-4 für die "Fine Granularity Scalability" vorgeschlagenen "embedded quantizers" in Kombination mit einer "Bitplane"-Codierung eingesetzt. Infolge der SNR-Skalierung entstehen ein Baselayer-Basis- und ein Baselayer-Zusatzsignal. Mit dem im Enhancementlayer-Bitstrom übertragenen Baselayer-Zusatzsignal wird der Quantisierungsfehler im rekonstruierten Baselayersignal reduziert, wenn das Bildsignal der hohen Auflösung rekonstruiert wird.

Für die SNR-skalierbare Codierung des Baselayers wird ein Verfahren entwickelt, das den Einfluss von Drift im rekonstruierten Baselayer-Basisignal begrenzt. Drift kann sich nur ausgehend von den rekonstruierten Intra-codierten Bildern (I-frames) und den unidirektional prädizierten Bildern (P-frames) akkumulieren, da nur diese zur weiteren Prädiktion benutzt werden. Die bidirektional prädizierten Bilder (B-frames), die den größten Teil des Bitstroms bilden, werden nicht zur weiteren Prädiktion benutzt. Darüberhinaus wird die Drift mit jedem übertragenen I-frame zurückgesetzt.

Durch eine unterschiedliche SNR-Skalierung für I-, P- und B-frames lässt sich die Akkumulation der Drift beeinflussen. In dieser Arbeit wird gezeigt, dass die folgende Einstellung besonders günstig ist: In I-frames werden sämtliche codierten Daten für das Baselayersignal im Baselayer-Bitstrom übertragen, es findet also keine SNR-Skalierung statt. In P-frames werden abgestuft nach Position innerhalb der "Group of Pictures" ca. 60% bis 80% der Daten im Baselayer- und entsprechend ca. 20% bis 40% im Enhancementlayer-Bitstrom übertragen. Die Aufteilung für B-frames beträgt ca. 50% zu 50%. Durch diese Einstellung ergibt sich der angestrebte Anteil des Baselayer-Bitstroms von 50% an der Gesamtdatenrate. Im Vergleich zu einer SNR-Skalierung, die unabhängig vom Frametyp (I-, P- oder B-frame) zu einem Baselayer-Anteil von 50% an der Gesamtdatenrate führt, ergibt die vorgeschlagene Lösung eine Verbesserung des PSNR im rekonstruierten Baselayersignal um 0,4 dB bis 0,5 dB.

Zur Bewertung der Codierungseffizienz des entwickelten skalierbaren Teilband-DCT-Codierungsverfahrens werden Experimente unter Verwendung der ITU-R 601 Testsequenzen "Bus" und "Basketball" durchgeführt. Die Testsequenzen werden vor der Codierung mit einem De-Interlacing-Algorithmus in ein progressives Format umgewandelt. Die Gesamtdatenrate von 6 Mbit/s wird in jeweils 3 Mbit/s für den Base- und den Enhancementlayer-Bitstrom aufgeteilt. Für eine vergleichende Bewertung werden die folgenden Codierungsverfahren untersucht:

- Nichtskalierbare MPEG-2-Codierung als Referenz für die obere Grenze der Codierungseffizienz
- Simulcast-Codierung basierend auf MPEG-2 für die untere Grenze der Codierungseffizienz
- Pyramidencodierung mit MPEG-4 SSP als Referenz für eine standardisierte örtlich skalierbare Codierung
- Das entwickelte skalierbare Teilband-DCT-Codierungsverfahren

Bezogen auf das hochaufgelöste Signal führt das standardisierte MPEG-4 SSP-Verfahren zu einer Verschlechterung im PSNR von 2,2 dB bis 2,4 dB bei gleicher Datenrate im Vergleich zur nichtskalierbaren MPEG-2-Codierung. MPEG-4 SSP liegt lediglich 0,5 dB bis 1,0 dB über der Simulcast-Codierung. Um den PSNR-Wert der nichtskalierbaren MPEG-2-Codierung zu erreichen, müsste bei MPEG-4 SSP die Gesamtdatenrate von 6 Mbit/s auf über 8,5 Mbit/s erhöht werden. Dies entspricht einer Datenratenerhöhung um mehr als 40%.

Im Gegensatz dazu erreicht das entwickelte skalierbare Teilband-DCT-Codierungsverfahren ein PSNR, das lediglich um 0,5 dB bis 0,6 dB unter dem PSNR der nichtskalierbaren MPEG-2-Codierung liegt. Es liegt hingegen 2,4 dB bis 2,6 dB oberhalb der Simulcast-Codierung. Zum Erreichen des PSNR der nichtskalierbaren MPEG-2-Codierung bei 6 Mbit/s benötigt das entwickelte Verfahren

eine Datenratenerhöhung von lediglich 5%-9%. Damit ist das eingangs gesetzte Ziel erreicht.

Bei den PSNR-Ergebnissen für das rekonstruierte Baselayersignal wird die erhöhte Codierungseffizienz von MPEG-4 gegenüber MPEG-2 durch einen Anstieg des PSNR um 0,8 dB bis 0,9 dB deutlich. Die PSNR-Ergebnisse des entwickelten skalierbaren Teilband-DCT-Codierungsverfahrens liegen aufgrund der driftbehafteten SNR-Skalierung des Baselayersignals um 0,01 dB bis 0,3 dB unter denen der MPEG-2-Codierung. Einbußen in der subjektiven Bildqualität werden ab etwa 0,5 dB wahrnehmbar.

Das entwickelte skalierbare Teilband-DCT-Codierungsverfahren ist für eine Codierung in zwei örtlichen und zwei Amplitudenauflösungen ausgelegt. Die vorgestellte Lösung lässt eine Erweiterung auf eine größere Anzahl sowohl örtlicher als auch SNR-Auflösungsstufen zu. Darüberhinaus ist in Zukunft auch eine Kombination mit einer zeitlichen Skalierung möglich.

A. Gleichungssystem zum Entwurf linearphasiger GQMF-Filterbänke ungerader Impulsantwortlänge L'

Der Term $f(l) (1 - (-1)^l)$ auf der linken Seite von Gleichung 5.10 ist für gerade l unabhängig von h_1 und h_2 gleich Null. Weiterhin ist Gl. 5.10 nur für gerade Impulsantwortlängen L erfüllt, damit bei $l = L - 1$ auch die linke Seite der Gleichung ungleich Null ist.

Für gerade Impulsantwortlängen L lässt sich Gleichung 5.10 mit Hilfe der Faltungsformel $f(l) = h_1(l) * ((-1)^l h_2(l)) = \sum_{i=0}^{L-1} h_1(i) \cdot h_2(l-i) \cdot (-1)^{l-i}$ als Gleichungssystem entwickeln:

$$\begin{aligned}
 l = 2L - 3 : \quad & h_1(L-1) \cdot h_2(L-2) - h_1(L-2) \cdot h_2(L-1) & = 0 \\
 l = 2L - 5 : \quad & h_1(L-1) \cdot h_2(L-4) - h_1(L-2) \cdot h_2(L-3) + \\
 & + h_1(L-3) \cdot h_2(L-2) - h_1(L-4) \cdot h_2(L-1) & = 0 \\
 & \cdot \\
 & \cdot \\
 & \cdot \\
 l = L + 1 : \quad & -h_1(2) \cdot h_2(L-1) + h_1(3) \cdot h_2(L-2) - h_1(4) \cdot h_2(L-3) + \dots + \\
 & + h_1(L-3) \cdot h_2(4) - h_1(L-2) \cdot h_2(3) + h_1(L-1) \cdot h_2(2) & = 0 \\
 l = L - 1 : \quad & h_1(0) \cdot h_2(L-1) - h_1(1) \cdot h_2(L-2) + h_1(2) \cdot h_2(L-3) + \dots + \\
 & + h_1(L-3) \cdot h_2(2) - h_1(L-2) \cdot h_2(1) + h_1(L-1) \cdot h_2(0) & = c \\
 & \cdot \\
 & \cdot \\
 & \cdot \\
 l = 3 : \quad & h_1(3) \cdot h_2(0) - h_1(2) \cdot h_2(1) + h_1(1) \cdot h_2(2) - h_1(0) \cdot h_2(3) & = 0 \\
 l = 1 : \quad & h_1(1) \cdot h_2(0) - h_1(0) \cdot h_2(1) & = 0
 \end{aligned} \tag{A.1}$$

Um dieses Gleichungssystem zum Entwurf von Filterbänken mit ungerader Im-

pulsantwortlänge L' verwenden zu können, wird den Impulsantworten $h_1(l)$ und $h_2(l)$ zur Aufstellung des Gleichungssystems jeweils eine Null hinzugefügt, wodurch sich eine gerade Impulsantwortlänge $L = L' + 1$ ergibt.

Wird die Null bei beiden Impulsantworten am Anfang eingefügt, d.h. $h_1(0) = h_2(0) = 0$, entstehen durch die geforderte Linearphasigkeit die symmetrischen Impulsantworten

$$\begin{aligned} h_1(l) &= [0, h_1(1), h_1(2), \dots, h_1(L-2), h_1(L-1)] = \\ &= [0, h_1(1), h_1(2), \dots, h_1(2), h_1(1)] \\ h_2(l) &= [0, h_2(1), h_2(2), \dots, h_2(L-2), h_2(L-1)] = \\ &= [0, h_2(1), h_2(2), \dots, h_2(2), h_2(1)] \end{aligned} \quad (\text{A.2})$$

Damit ergibt sich nach Einsetzen von Gl. A.2 in Gl. A.1 für

$$\begin{aligned} l = L + 1 : & -h_1(2) \cdot h_2(1) + h_1(3) \cdot h_2(2) - h_1(4) \cdot h_2(3) + \dots + \\ & + h_1(3) \cdot h_2(4) - h_1(2) \cdot h_2(3) + h_1(1) \cdot h_2(2) = 0 \\ l = L - 1 : & -h_1(1) \cdot h_2(2) + h_1(2) \cdot h_2(3) - h_1(3) \cdot h_2(4) + \dots + \\ & + h_1(4) \cdot h_2(3) - h_1(3) \cdot h_2(2) + h_1(2) \cdot h_2(1) = c \end{aligned} \quad (\text{A.3})$$

Da die linken Seiten in Gl. A.3 bis auf das Vorzeichen identisch sind, ist dieses Gleichungssystem für $c \neq 0$ nicht lösbar.

Gleiches gilt auch für das Einfügen der Null am Ende beider Impulsantworten, d.h. $h_1(L-1) = h_2(L-1) = 0$.

Erst durch Einfügen der Null an den entgegengesetzten Enden von $h_1(l)$ und $h_2(l)$, d.h. $h_1(0) = h_2(L-1) = 0$, ergeben sich unterschiedliche linke Seiten für $l = L + 1$ und $l = L - 1$:

$$\begin{aligned} l = L + 1 : & h_1(3) \cdot h_2(0) - h_1(4) \cdot h_2(1) + \dots - h_1(2) \cdot h_2(3) + h_1(1) \cdot h_2(2) = 0 \\ l = L - 1 : & -h_1(1) \cdot h_2(0) + h_1(2) \cdot h_2(1) + \dots - h_1(2) \cdot h_2(1) + h_1(1) \cdot h_2(0) = c \end{aligned} \quad (\text{A.4})$$

Damit ist das Gleichungssystem auch für $c \neq 0$ lösbar.

Durch das Setzen von $h_1(0) = h_2(L-1) = 0$ ergibt sich jedoch für

$$\begin{aligned} l = 2L - 3 : & h_1(1) \cdot h_2(0) = 0 \\ l = 1 : & h_1(1) \cdot h_2(0) = 0, \end{aligned} \quad (\text{A.5})$$

d.h. entweder $h_1(1) = 0$ (und damit aufgrund der Symmetrie auch $h_1(L-1) = 0$) oder $h_2(0) = 0$ (und damit auch $h_2(L-2) = 0$). Dies bedeutet, dass entweder $h_1(l)$ oder $h_2(l)$ eine Impulsantwortlänge von $L'' = L - 3$ aufweisen muss.

Wird $h_1(1) = h_1(L-1) = 0$ gewählt, haben die Impulsantworten bezogen auf die für das Aufstellen des Gleichungssystems verwendete Länge L folgendes Aussehen:

$$\begin{aligned} h_1(l) &= [0, 0, h_1(2), h_1(3), \dots, h_1(L/2-1), h_1(L/2), h_1(L/2-1), \dots, h_1(3), h_1(2), 0] \\ h_2(l) &= [h_2(0), h_2(1), \dots, h_2(L/2-2), h_2(L/2-1), h_2(L/2-2), \dots, h_2(1), h_2(0), 0], \end{aligned} \quad (\text{A.6})$$

und für die von Null verschiedenen Koeffizienten nach Substitution von l durch l'' bzw. l' bezogen auf die ungeraden Impulsantwortlängen $L'' = L-3$ bzw. $L' = L-1$:

$$\begin{aligned} h_1(l'') &= [h_1(0), h_1(1), \dots, h_1(L''/2-1), h_1(L''/2), h_1(L''/2-1), \dots, h_1(1), h_1(0)] \\ h_2(l') &= [h_2(0), h_2(1), \dots, h_2(L'/2-1), h_2(L'/2), h_2(L'/2-1), \dots, h_2(1), h_2(0)]. \end{aligned} \quad (\text{A.7})$$

Bei Vorgabe von $h_1(l)$ sind somit die $L'+1/2 = L/2$ von Null verschiedenen Koeffizienten der Impulsantwort $h_2(l)$ zu bestimmen. Das Gleichungssystem A.1 liefert, bedingt durch seine Symmetrie, lediglich $L/2-1$ Bestimmungsgleichungen. Der verbleibende Freiheitsgrad von Eins wird bei der Codierung von Videosignalen i.A. dazu benutzt, die Unterdrückung des Gleichanteils im hochfrequenten Teilband, also $H_2(\omega=0) = 0$, sicherzustellen.

Literaturverzeichnis

- [1] ITU-R 601: International Telecommunication Union, "Recommendation ITU-R BT.601-5: Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios," Genf, 1995.
- [2] ITU-R 709: International Telecommunication Union, "Recommendation ITU-R BT.709-3: Parameter values for the HDTV standards for production and international programme exchange," Genf, 1998.
- [3] Robert M. Gray, "Entropy and Information Theory," Springer Verlag New York, 1990.
- [4] International Telecommunication Union: "Recommendation ITU-T H.261: Video codec for audiovisual services at p x 64 kbit/s," International Telecommunication Union, Genf, 1993.
- [5] International Telecommunication Union: "Recommendation ITU-T H.263: Video coding for low bit rate communication," International Telecommunication Union, Genf, 1998.
- [6] MPEG-1 : ISO/IEC JTC1/SC29/WG11, "International Standard ISO/IEC IS 11172 : Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," ISO/IEC, Genf, April 1992.
- [7] MPEG-2 : ISO/IEC JTC1/SC29/WG11, "Revised Text for ITU-T Recommendation H.262 — ISO/IEC 13818-2 : Information technology - Generic coding of moving pictures and associated audio information : video," ISO/IEC, Genf, März 1995.
- [8] MPEG-2, "Draft - Test Model 6," ISO/IEC JTC1/SC29/WG11, N0491, Juli 1993.
- [9] MPEG-4 : ISO/IEC JTC1/SC29/WG11, "ISO/IEC 14496:2000-2 : Information Technology - Coding of Audio-Visual Objects - Part 2: Visual," ISO/IEC, Genf, Dezember 2000.

-
- [10] MPEG-4 FGS, "Study on ISO/IEC 14496-2:1999/FDAM4 : Amendment 4: Streaming Video Profile," ISO/IEC JTC1/SC29/WG11, N3670, Oktober 2000.
- [11] W. Li, "Fine Granularity Scalability Using Bit-plane Coding of DCT coefficients," Doc. ISO/IEC JTC1/SC29/WG11 MPEG98/4204, Dezember 1998.
- [12] U. Benzler et.al., "Results of Core Experiment P8," Docs. ISO/IEC JTC1/SC29/WG11 M1814,M2212,M2404,... 1997-1999.
- [13] A. Vincent, D. Lauzon, P. Haighton, "Spatial Prediction in Scalable Video Coding," International Broadcast Convention, London, Conference Publication No. 413, pp. 244-249, September 1995.
- [14] U. Benzler, O. Werner, "Improving multiresolution motion compensating hybrid coding by drift reduction," Proc. Picture Coding Symposium 1996, Melbourne, Australien, März 1996.
- [15] F. Bosveld, "Hierarchical Video Compression using SBC," Delft: Universiteit Delft, Faculteit der Elektrotechniek, Dissertation, 1996.
- [16] J.-R. Ohm, "Three-Dimensional Subband Coding with Motion Compensation," IEEE Transactions on Image Processing, Vol. IP-3, No. 5, pp. 559-571, September 1994.
- [17] D. Taubman, A. Zakhor, "Multirate 3-D Subband Coding of Video," IEEE Transactions on Image Processing, Vol. IP-3, No. 5, pp. 572-588, September 1994.
- [18] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," IEEE Journal on Selected Areas in Communication, Vol. SAC-5, No. 7, pp 1140-1154, August 1987.
- [19] U. Horn, B. Girod, "Scalable Video Transmission for the Internet," Proceedings 8th Joint European Networking Conference, Edinburgh, Mai 1997.
- [20] U. Horn, B. Girod, bit streams for "Scalable Video Transmission for the Internet," <http://www.lnt.de/> Proceedings 8th Joint European Networking Conference, Edinburgh, Mai 1997.
- [21] M. R. Cinvalar, A. Puri, "Scalable video coding in frequency domain," Proc. SPIE Symp. Visual Commun., Image Processing (Boston, MA), Vol. 1818, pt. 3, pp. 1124-1134, November 1992.
- [22] S.-T. Hsiang, J. W. Woods, "Invertible Three-Dimensional Analysis/Synthesis System for Video Coding with Half-Pixel-Accurate Motion Compensation," Proc. SPIE Symp. Visual Commun., Image Processing (San Jose, CA), Vol. 3653, pt. 1, pp. 537-546, Januar 1999.

-
- [23] G. Wackersreuther, "Some New Aspects of Filters for Filter Banks," IEEE Transactions on Acoustic, Speech and Signal Processing, Vol. ASSP-34, No. 5, pp. 1182-1199, Oktober 1986.
- [24] U. Benzler, "Spatial Scalable Video Coding using a Combined Subband-DCT Approach," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No.7, Oktober 2000.
- [25] O. Werner, "Drift analysis and drift reduction for multiresolution hybrid video coding," EURASIP journal Signal Processing: Image Communication, Vol. 8, No. 5, pp. 387-409, Juli 1996.
- [26] John W. Woods (Ed.), "Subband Image Coding," Kluwer Academic Publishers, 1991.
- [27] A. Seeger "Hierarchical channel coding for Rayleigh and Rice fading," Proc. IEEE Globecom, Communication Theory Mini Conference, Phoenix, pp. 208-212, November 1997.
- [28] G. Seelmann, "Beitrag zur Optimierung des Bildaufbaus bei der progressiven Übertragung mittels Teilbandcodierung," Hannover: Universität Hannover, Fakultät für Maschinenwesen, Dissertation, 1996.
- [29] T. Chiang, D. Anastassiou, "Hierarchical Coding of Digital Television," IEEE Communications Magazine, Vol 32, No. 5, Mai 1994.
- [30] Vandendorpe, L., Cuvelier, L. et. al., "Motion-Compensated conversion from interlaced to progressive formats," EURASIP journal Signal Processing: Image Communication, Vol. 6, No. 3, pp. 193-211, Juni 1994.
- [31] T. Berger, "Rate Distortion Theory," Englewood-Cliffs, NJ: Prentice Hall, 1971.
- [32] H.G. Musmann, P. Pirsch, H.-J. Grallert, "Advances in Picture Coding," Proc. of the IEEE, Vol. 73, No. 4, pp. 523-548, April 1985.
- [33] H.G. Musmann, "Quellencodierung," Skript zur Vorlesung an der Universität Hannover, 1999.
- [34] D.A. Huffman, "A method for the construction of minimum redundancy codes," Proc. IRE, No. 40, pp. 1098-1101, 1962.
- [35] J. Rissanen, G.G. Langdon, "Universal modeling and coding," IEEE Transactions on Information Theory, Vol. IT-27, No. 1, pp. 12-23, Januar 1981.

- [36] N. Wells, P. Tudor, "Standardization of scalable coding schemes," Proc. IEEE ISCAS 94, Tutorials, Mai 1994.
- [37] J. W. Woods, S.D. O'Neil, "Subband Coding of Images," IEEE Transactions on Acoustic, Speech and Signal Processing, Vol. ASSP-34, No. 5, pp. 1278-1288, Oktober 1986.
- [38] M. Vetterli, C. Herley, "Wavelet and Filter Banks: Theory and Design," IEEE Transactions on Signal Processing, Vol. 40, No. 9, pp. 2207-2232, September 1992.
- [39] H. Schiller, "Modellierung und Codierung des Fehlersignals der bewegungskompensierten Interframe-Prädiktion von Bewegtbildfolgen," Hannover: Universität Hannover, Fakultät für Maschinenwesen, Dissertation, 1994.
- [40] Mattavelli, M. ; Nicoulin, A. ; Fernandez, G. "Overlapped motion compensation for subband coding of video sequences" EURASIP journal Signal Processing: Image Communication, Vol. 8, No. 2, pp. 149-160, März 1996.
- [41] U. Pestel-Schiller, "Optimierung der Filterbank für eine Teilbandcodierung bezüglich Redundanz- und Irrelevanzreduktion," Hannover: Universität Hannover, Fakultät für Maschinenwesen, Dissertation, 1996.
- [42] B. Edler, "Äquivalenz von Transformation und Teilbandzerlegung in der Quellencodierung," Hannover: Universität Hannover, Fakultät für Maschinenwesen, Dissertation, 1995.
- [43] A. V. Oppenheim, R. W. Schaffer, "Digital Signal Processing," Englewood-Cliffs, NJ: Prentice Hall, 1975.

Lebenslauf

Ulrich-Lorenz Benzler

geb. am 21.01.1966 in Hannover

Ausbildung

10.85 - 08.88 Norddeutscher Rundfunk, Landesfunkhaus Hannover
Ausbildung zum Funkelektroniker

Studium

10.88 - 08.93 Universität Hannover
Fachbereich: Elektrotechnik
Vertiefung: Nachrichtenverarbeitung
18.08.93 Diplom der Universität Hannover (Dipl.-Ing.)

Berufstätigkeit

09.93 - 03.01 Wissenschaftlicher Mitarbeiter
Abteilung: Codierung von Audio- und Videosignalen
Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
Universität Hannover

seit 04.01 Entwicklungsingenieur
Abteilung Forschung und Voraentwicklung Multimediasysteme
Robert Bosch GmbH
Hildesheim