

Hierarchische DCT-Teilband-Codierung

Ulrich Benzler

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Schneiderberg 32, D-30167 Hannover
Tel.: +49 511 762-5047, Fax: +49 511 762-5052,
E-Mail: benzler@tnt.uni-hannover.de

Übersicht

Bei der digitalen Übertragung von Videodaten über Mobilfunkkanäle kann es zu sehr störenden Einbrüchen in der Bildqualität kommen. Mit dem Konzept eines hierarchischen Übertragungsverfahrens mit skalierbarer Videocodierung, deren örtliche Auflösung sich im Fall ungünstiger Empfangsbedingungen reduziert, soll der abrupte Abfall der Bildqualität vermieden werden.

Die zu diesem Zweck standardisierte Pyramiden-codierung nach den MPEG-2 und MPEG-4 Spatial Scalable Profiles (SSP) führt zu einer Datenratenerhöhung von mehr als 40% bei gleicher Bildqualität im Vergleich zu einer nicht-skalierbaren Codierung.

Um diese Verminderung der Codierungseffizienz zu vermeiden, wird ein neuer Ansatz basierend auf einer Teilbandzerlegung mit DCT-Codierung entwickelt. Allerdings benötigt die niedrige Auflösung dabei erheblich mehr als 50% der Gesamtdatenrate. Da die hierarchische Kanal-codierung und Modulation einen maximalen Datenratenanteil von 50% für die niedrige Auflösung erlaubt, wird zusätzlich eine sogenannte SNR-Skalierbarkeit eingeführt.

Die entwickelte DCT-Teilband-Codierung erfordert lediglich eine um 5-9% erhöhte Datenrate im Vergleich zur nicht-skalierbaren Codierung bei gleicher Bildqualität, so dass kaum eine Verminderung der Codierungseffizienz zu erkennen ist.

1 Einführung

Zur digitalen Übertragung von Videodaten auf Mobilfunkkanälen soll ein hierarchisches Übertragungsverfahren verwendet werden. Dazu muß der zu übertragende Datenstrom in einen sogenannten Baselayer- und einen Enhancementlayer-Datenstrom aufgeteilt werden. Der Baselayer

beinhaltet dabei nur die Daten für eine niedrige örtliche Auflösung des Bildsignals, der Enhancementlayer die zusätzlichen Daten für die hohe örtliche Auflösung. Bei günstigen lokalen Empfangsbedingungen werden Base- und Enhancementlayer decodiert und die hohe örtliche Auflösung auf einem großen Display dargestellt. Im Fall ungünstiger Empfangsbedingungen werden nur die Baselayer-Daten decodiert und als Rückfallmodus auf dem großen Display angezeigt. Unabhängig davon kann der Baselayer-Datenstrom von einem weniger komplexen Empfänger decodiert und auf einem kleinen Display, das lediglich die niedrige Auflösung anzeigen kann, dargestellt werden.

Für eine solche hierarchische Codierung digitaler Videosignale ist es notwendig, das Eingangssignal in Teilsignale mit unterschiedlichen örtlichen Auflösungen zu skalieren. Diese Skalierbarkeit erlaubt die Decodierung in unterschiedliche Auflösungen, ohne das gesamte Datenvolumen decodieren zu müssen. Im MPEG-2- und MPEG-4-Standard wird diese Funktionalität durch die "Spatial Scalable Profiles" (MPEG-SSP) [1] [2] zur Verfügung gestellt. Dabei wird eine Pyramiden-codierung benutzt, welche die niedrige Auflösung separat als Baselayer-Datenstrom codiert. Um Teile der Baselayer-Information für die hohe Auflösung nutzbar zu machen, werden die rekonstruierten Bilder auf das Format der hohen Auflösung interpoliert und als zusätzliches Prädiktionssignal zur Verfügung gestellt, siehe Abbildung 1.

Der resultierende Prädiktionsfehler wird im Enhancementlayer mit der Abtastrate des Eingangssignals codiert. Dadurch erhöht sich die Anzahl der insgesamt zu codierenden Abtastwerte im Vergleich zur nicht-skalierbaren Codierung um die Anzahl der Abtastwerte der niedrigen

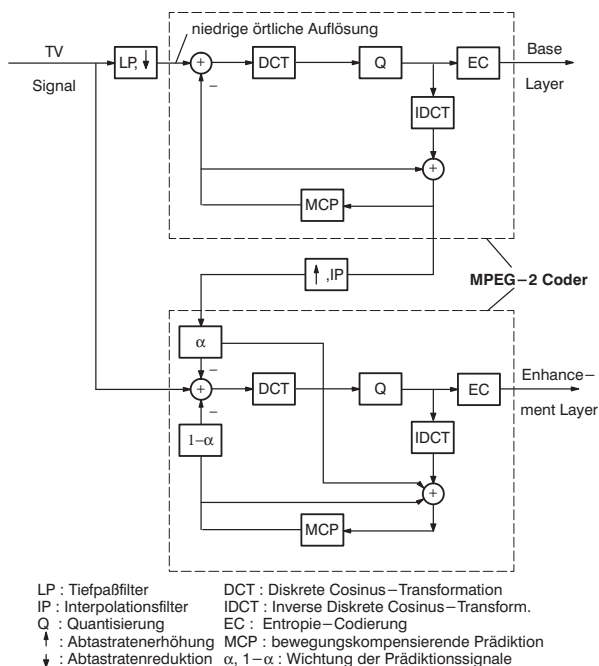


Abbildung 1: Pyramidencodierung nach MPEG Spatial Scalable Profile (SSP)

Auflösung. Dieser Nachteil läßt sich nicht durch die Bereitstellung des zusätzlichen Prädiktions-signals für die hohe Auflösung kompensieren, so daß die Gesamteffizienz des Verfahrens nur unwesentlich besser ist als eine Simulcast-Codierung, bei der beide Auflösungen unabhängig voneinander codiert werden.

2 Codierung mittels Teilbandzerlegung

Zur Vermeidung der erhöhten Anzahl Abtastwerte bei der Pyramidencodierung wurde ein Ansatz entwickelt, der auf einer zweidimensionalen Teilbandzerlegung des Eingangssignals mit kritischer Abtastung basiert. Dabei bildet das tieffrequente Teilband (LL-Band) die niedrige Auflösung, die ähnlich MPEG SSP mit einer separaten bewegungskompensierenden Prädiktion als Baselayer codiert und zum Empfänger übertragen wird. Dabei kommt eine speziell entwickelte bewegungs- und aliasingkompensierende Prädiktion zum Einsatz, die eine sehr effiziente Verarbeitung des Baselayer-Signals ermöglicht.

Dieses Prädiktionsverfahren ist auch für eine

nicht-skalierbare Codierung anwendbar, wobei es zu einer Datenrateneinsparung von etwa 30% bei gleicher Bildqualität führt. Aufgrund dieser Resultate wurde das Prädiktionsverfahren von der ISO/IEC im Rahmen von MPEG-4 standardisiert.

Im Enhancementlayer wird die bewegungskompensierende Prädiktion in der hohen Auflösung durchgeführt und die Information für die hochfrequenten Teilbänder codiert, so daß der Empfänger durch Decodierung von Base- und Enhancementlayer die hohe Auflösung erhält, siehe Abbildung 2. Die DCT-Codierung der Teilbänder wird dabei entsprechend dem MPEG-2-Standard durchgeführt, was eine einfache Implementierung des Systems als Erweiterung des MPEG-2 Referenz-Coders erlaubt.

2.1 Verteilung der Datenrate auf Base- und Enhancementlayer

Es wurden Untersuchungen zur Beurteilung der subjektiven Bildqualität in Abhängigkeit von der Datenratenaufteilung durchgeführt. Dabei ergab sich, daß bei einer vom Enhancementlayer unabhängigen Codierung des Baselayers für ein kleines Display, das lediglich die niedrige Auflösung anzeigen kann, ca. 50% der Gesamtdatenrate benötigt werden. Ein höherer Anteil der Baselayer-Daten am Gesamtdatenstrom ist auch aus Sicht der hierarchischen Kanalcodierung und Modulation nicht mehr sinnvoll [6].

Wird die so codierte niedrige Auflösung aber als LL-Band zusammen mit den hochfrequenten Teilbändern des Enhancementlayers zur Rekonstruktion der hohen Auflösung eingesetzt, wird damit nicht die für die hohe Auflösung angestrebte Bildqualität erreicht. Untersucht man die benötigte Datenrate für den dem LL-Band entsprechenden Frequenzbereich einer nicht-hierarchischen, einstufigen Codierung nach MPEG-2, so benötigt dieser je nach Testsequenz zwischen 65 und 80% der Gesamtdatenrate. Diesen Anteil benötigt auch das LL-Band des entwickelten Teilbandcoders. Diese Aufteilung läßt sich zum einen durch die Konzentration der Signalenergie auf die tiefen Frequenzen im Leistungsdichtespektrum begründen. Zum an-

deren fällt das Auflösungsvermögen des menschlichen visuellen Systems zu hohen Ortsfrequenzen hin ab. Deshalb können in den hochfrequenten Teilbändern größere Quantisierungsfehler zugelassen werden, und somit sind dort nur wenige Teilbandkoeffizienten ungleich Null, wodurch nur eine geringe Datenrate für deren Übertragung benötigt wird.

Zusätzlich zu den DCT-Koeffizienten des LL-Bandes muß im Baselayer-Datenstrom die gesamte Seiteninformation (Prädiktionsauswahl, Bewegungsinformation etc.) übertragen werden, was eine weitere Verschiebung des Datenratenverhältnisses zum Baselayer hin bewirkt.

2.2 SNR-Skalierbarkeit im Baselayer

Um der von den Verbund-Partnern [6] geforderten Datenrataufteilung von maximal 1:1 zwischen Base- und Enhancementlayer zu genügen, müssen Informationsanteile der DCT-Koeffizienten des LL-Bandes teilweise im Enhancementlayer übertragen werden. Zu diesem Zweck eignet sich eine sogenannte SNR-Skalierbarkeit. Dabei wird das LL-Band zunächst grob quantisiert und als Baselayer übertragen, so daß bei einer Darstellung auf dem kleinen Display keine sichtbaren Fehler erkennbar sind. Dabei stellt sich eine Baselayer-Datenrate von ca. 50% der Gesamtdatenrate ein (vergl. Abschnitt 2.1). Anschließend wird die für den Enhancementlayer benötigte feinere Quantisierung durchgeführt und die Differenz zum Baselayersignal gemeinsam mit den Informationen der hochfrequenten Teilbänder übertragen, siehe Abbildung 3. Hierfür werden "embedded quantizers" [5] in Verbindung mit einer Bitplane-basierten Codierung [3] der Koeffizienten eingesetzt.

3 Simulationsergebnisse

In experimentellen Untersuchungen wurden die entwickelte DCT-Teilband-Codierung und das standardisierte MPEG-4-SSP-Verfahren mit einer nicht-skalierbaren MPEG-2-Codierung sowie einer Simulcast-Codierung verglichen. Die Gesamtdatenrate von 6 Mbit/s wurde dabei zu je 3 Mbit/s auf Base- und Enhancementlayer-Datenstrom aufgeteilt.

Die MPEG-4 SSP Codierung führt zu einem Verlust von mehr als 2.2 dB PSNR gegenüber der nicht-skalierbaren Codierung und liegt damit nur gering über dem PSNR der Simulcast-Codierung (siehe Abbildung 4). Dieser Verlust an Bildqualität ist deutlich in den rekonstruierten Sequenzen zu erkennen. Eine Codierung bei gleicher Bildqualität führt zu einem Anstieg der Datenrate um mehr als 40%.

Im Gegensatz dazu liegen die PSNR-Werte der DCT-Teilband-Codierung lediglich 0.4 dB unter der nicht-skalierbaren MPEG-2-Codierung. Dies bedeutet, dass praktisch kein Verlust an Bildqualität erkennbar ist. Bei Codierung mit gleicher Bildqualität ist ein Anstieg der Datenrate von lediglich 5-9% zu beobachten, so dass sich nahezu kein Rückgang der Codierungseffizienz als Folge der zusätzlichen Skalierungs-Funktionalität einstellt.

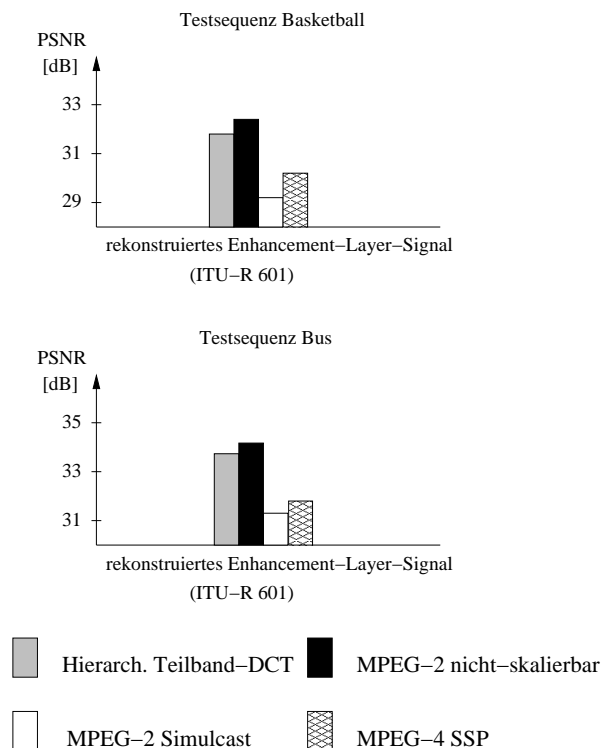


Abbildung 4: Simulationsergebnisse der Testsequenzen "Bus" und "Basketball"

Da die Codierung des Baselayers sowohl bei der Simulcast-Codierung als auch bei der MPEG-SSP-Codierung identisch zu einer nicht-skalierbaren Codierung der niedrigen Auflösung

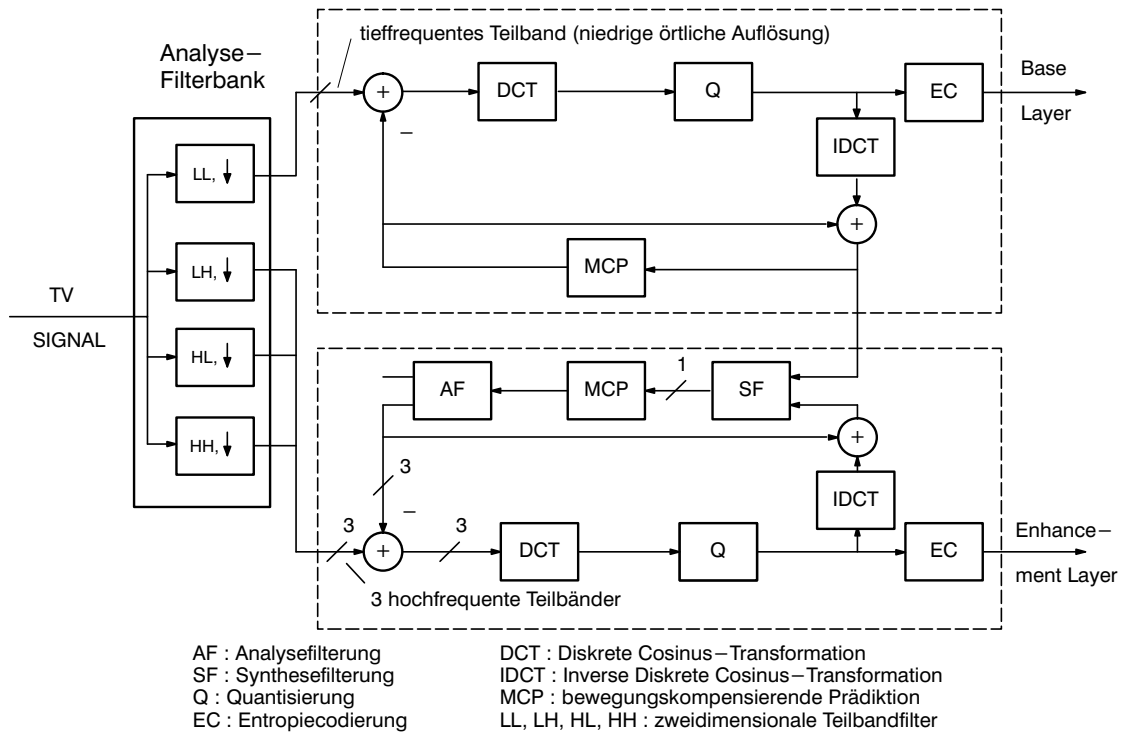


Abbildung 2: Hierarchische Codierung mittels Teilbandzerlegung

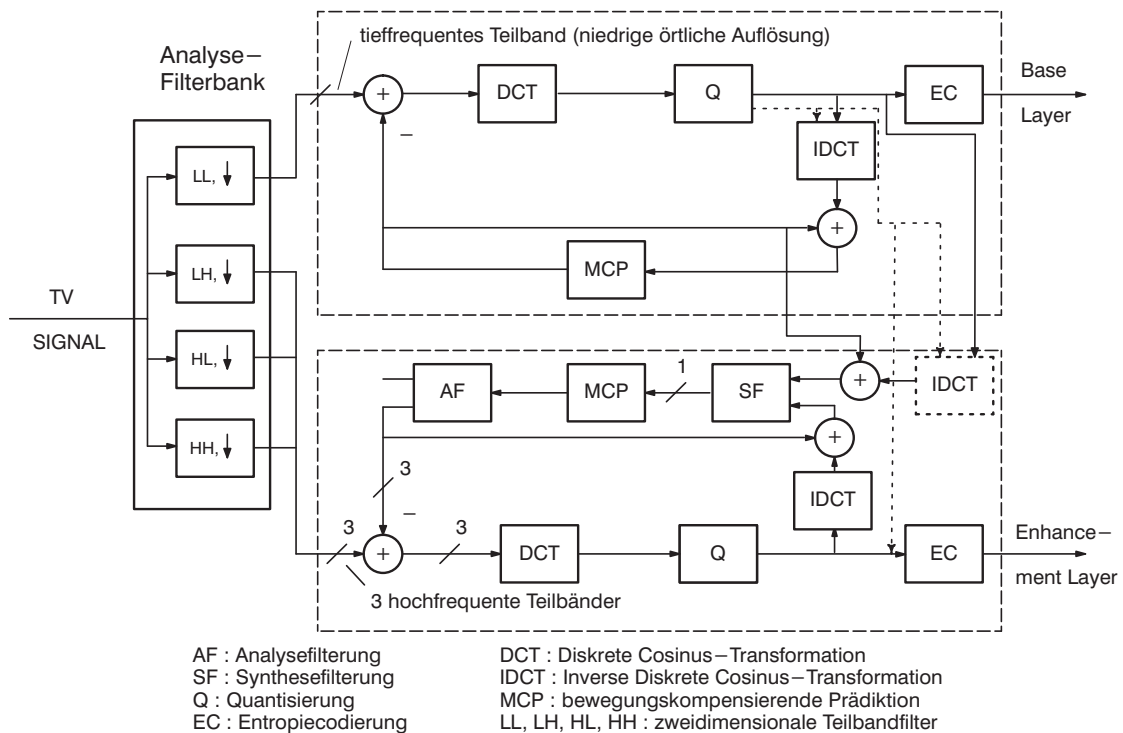


Abbildung 3: SNR-Skalierbarkeit im Baselayer

ist, erzeugen sie einen identischen Baselayer-Bitstrom von 3 Mbit/s.

Der Baselayer der DCT-Teilband-Codierung zeigt eine Qualität, die der nicht-skalierbaren Codierung der niedrigen Auflösung bei 3 Mbit/s vergleichbar ist. Die PSNR-Werte liegen 0.05 bis 0.5 dB unter denen der nicht-skalierbaren Codierung, was auf die SNR-Skalierbarkeit des Baselayers in der DCT-Teilband-Codierung zurückzuführen ist.

4 Zusammenfassung

Der vorgestellte Ansatz einer örtlich skalierbaren DCT-Teilband-Codierung für die hierarchische Videoübertragung vermeidet das von der standardisierten Pyramidencodierung nach MPEG-SSP bekannte Problem der Erhöhung der Anzahl von Abtastwerten durch die Verwendung einer Teilbandzerlegung.

Zur effizienten Verarbeitung des Baselayer-Signals wird eine speziell entwickelt bewegungs- und aliasingkompensierende Prädiktion verwendet, die aufgrund ihrer hohen Codierungseffizienz von der ISO/IEC im Rahmen von MPEG-4 standardisiert wurde.

Um den Randbedingungen der hierarchischen Kanalcodierung und Modulation hinsichtlich einer maximalen Baselayer-Datenrate von 50% zu genügen, wird eine SNR-Skalierbarkeit des Baselayers eingeführt. Die hierbei verwendeten "embedded quantizer" in Verbindung mit einer Bitplane-Codierung führen dazu, dass sich die Qualität des rekonstruierten Baselayer-Signal kaum von einer nicht-skalierbaren Codierung der niedrigen Auflösung unterscheidet.

Die Simulationsergebnisse zeigen, dass die standardisierte Pyramiden-Codierung mit MPEG-4-SSP zu einem Absinken des PSNR um 2.2 dB führt, während die DCT-Teilband-Codierung lediglich 0.4 dB unter dem PSNR einer nicht-skalierbaren Codierung liegt. Bei Codierung mit gleicher Bildqualität reduziert sich die zusätzlich erforderliche Datenrate von mehr als 40% für MPEG-4 SSP auf 5-9% für die DCT-Teilband-Codierung. Dies bedeutet, dass die zusätzliche Funktionalität der Skalierbarkeit mit der entwickelten DCT-Teilband-Codierung nahezu ohne Verlust an Codierungseffizienz ermöglicht wird.

Literatur

- [1] 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, Geneva 1995
- [2] MPEG-4: ISO/IEC JTC1/SC29/WG11: ISO/IEC 14496-2 : Information Technology - Coding of Audio-Visual Objects - Part 2: Visual, ISO/IEC, Geneva, December 1999.
- [3] W. Li et. al.: Bit-Plane Coding of DCT Coefficients *Doc. ISO/IEC JTC1/SC29/WG11 MPEG97/2691*, Oktober 1997
- [4] Werner, O.: Drift analysis and drift reduction for multiresolution hybrid video coding. In: *Signal Proc.: Image Commun.*, Vol. 8, 1996, S. 387–409
- [5] Bosveld, F.: Hierarchical Video Compression using SBC. Delft: Universiteit Delft, Faculteit der Elektrotechniek, Diss. 1996, S. 80–84
- [6] A. Seeger und U. Benzler: Hierarchische Quellen- und Kanalcodierung für digitale Videosignale, In: *ITG Fachbericht 146, ITG Fachtagung: Codierung für Quelle, Kanal und Übertragung*, März 1998, Aachen, S. 123–130