

Fortschritte in der Multimediakommunikation

von

Dr.-Ing. Bernd Edler

Dipl.-Ing. Sven Klomp

Dipl.-Ing. Marco Munderloh

Dipl.-Ing. Matthias Narroschke

Dipl.-Ing. Dieu Thanh Nguyen

Dipl.-Math. Nikolce Stefanoski

Dipl.-Ing. Yuri Vatis

Prof. Dr.-Ing. Jörn Ostermann

Institut für Informationsverarbeitung (TNT) /

Laboratorium für Informationstechnologie (LFI),

Leibniz Universität Hannover

Inhaltsverzeichnis

- 1 Einleitung
- 2 Fortschritte in der Videocodierung
 - 2.1 Grundlagen der Videocodierung
 - 2.1.1 Prädiktion
 - 2.1.2 Transformation und Quantisierung
 - 2.1.3 Redundanz mindernde Codierung
 - 2.2. Techniken zur weiteren Reduktion der Datenrate
 - 2.2.1 Nicht separierbares adaptives Interpolationsfilter
 - 2.2.2 Displacementvektorauflösung von 1/8-pel
 - 2.2.3 Adaptive Codierung des Prädiktionsfehlers im Orts- und im Frequenzbereich
 - 2.3 Vergleich der Codierungseffizienz
- 3 Distributed Video Coding
 - 3.1 Unterschiedliche Modelle in der Videocodierung
 - 3.2 Aufbau eines DVC-Coders
 - 3.3 Vorteile des DVC
- 4 Prädiktive Codierung von 3D-Netzsequenzen
 - 4.1 Prädiktive nicht-skalierbare Codierung
 - 4.2 Prädiktive örtlich-skalierbare Codierung
 - 4.3 Ergebnisse der Netzcodierung
- 5 Videostreaming
 - 5.1 Multimedia Streaming System: Stand der Technik und zu lösende Probleme
 - 5.2 Skalierbares Videostreaming - Aktuelle Forschung am TNT
 - 5.2.1 Skalierbare Erweiterung von H.264/AVC
 - 5.2.2 Congestion Control basierend auf einer Bandbreitenabschätzung
 - 5.2.3 Fehlerverschleierung für skalierbare Videocodierung
- 6 Audiocodierung und -signalverarbeitung
 - 6.1 Parametrische Audiocodierung: MPEG-4 HILN
 - 6.2 Kombination verschiedener Codierverfahren
 - 6.3 Filterbänke für parametrische Erweiterungen
 - 6.4 Signalverarbeitung für Cochlea-Implantate
- 7 Zusammenfassung
- 8 Referenzen
- 9 Verfasser-Portraits

1 Einleitung

Die Bedeutung der digitalen Multimediasignalverarbeitung hat mit Einführung der so genannten MP3-Player, der DVD-Abspielgeräte und des digitalen Fernsehens im täglichen Leben stark zugenommen. Die schnelle Verbreitung dieser Geräte wurde zum einen durch die Leistungssteigerung der Mikroelektronik, zum anderen durch die Entwicklung und Einführung internationaler Standards wie MPEG-1 [5], MPEG-2 [6] und MPEG-4 [7] auf dem Gebiet der Audio- und Videocodierung ermöglicht. Im Vergleich zu analogen Verfahren ermöglicht die digitale Codierung eine effizientere Nutzung von Speicherplatz und Übertragungsbandbreite. Schon die ersten Standards DVB-S [4] für digitales Satellitenfernsehen und DVB-C [3] für digitales Kabelfernsehen ermöglichten es, durch Einsatz der MPEG-Standards bis zu zehn Programme dort zu übertragen, wo bisher mit analoger Technik nur ein einziges Programm Platz gefunden hat. Bei besserer Bild- und Tonqualität gab es nun die Möglichkeit, zusätzlich neben dem eigentlichen Fernsehsignal weitere so genannte Mehrwertdienste anzubieten. Am bekanntesten sind sicherlich die Ausstrahlung von mehreren Tonspuren für Surroundsound und die elektronische Fernsehzeitung (EPG).

Durch die Standardisierung der Codierung audiovisueller Daten wurde es der Industrie ermöglicht, einfache und kostengünstige Massenprodukte zu bauen, die für die Wiedergabe praktisch aller bedeutenden Audio- und Videoinhalte geeignet sind. Das Institut für Informationsverarbeitung (TNT) der Leibniz Universität Hannover hat an dieser Standardisierung im Rahmen der ISO/IEC SC29 WG11 Gruppe, die auch unter dem Kürzel MPEG für Moving Pictures Experts Group bekannt ist, aktiv mitgearbeitet. Am Institut entwickelte Technik wird im Bereich Audiocodierung für MP3-Player und Advanced Audio Coding (AAC) sowie im Bereich Videocodierung für DVDs, digitales Fernsehen und zukünftiges HDTV eingesetzt. In Zukunft wird neben der Codierung von Audio und Videosignalen auch die Codierung dreidimensionaler Objekte, deren Oberfläche durch so genannte dreidimensionale Netze beschrieben wird, an Bedeutung gewinnen. Der Einsatz solcher Objekte in Verbindung mit Videos ermöglicht zum Beispiel gezielte auf Benutzergruppen zugeschnittene Werbung in Spielfilmen.

In diesem Artikel werden die neuesten Entwicklungen im Bereich der Audio- und Videocodierung dargestellt. Verwandte Arbeitsgebiete des Instituts werden ebenfalls beleuchtet. So finden die im Zusammenhang mit den Arbeiten an der MP3-Audiocodierung entwickelten Algorithmen zur Modellierung der menschlichen Wahrnehmung heute Einsatz in der Steuerung von Cochlea-Implantaten, die es ertaubten Menschen ermöglichen, wieder zu hören. Neben weiteren Fortschritten bei der Effizienz der Videocodierung werden Verfahren entwickelt, die die im Allgemeinen hohe Rechenleistung eines Videoencoders auf den Decoder verlagern. Hiermit werden Systeme mit vielen Encodern und wenigen Decodern, also Sensornetze oder Überwachungskameras, wirtschaftlich betrieben werden können. Prinzipien der Videocodierung ermöglichen eine effiziente Codierung bewegter dreidimensionaler Netze.

Im Vergleich zu traditionellen Telekommunikationsnetzen stellen moderne Kommunikationsnetze wie WLAN, GPRS, UMTS und DSL für eine Anwendung keine konstante Bandbreite zur Verfügung, da in diesen paketbasierten Netzen aus Kostengründen kein Quality-of-Service-Merkmal implementiert wurde. Deshalb entwickelte sich die Echtzeitübertragung von Audio- und Videodaten über diese Netze in ein aktives Forschungsgebiet. Die Echtzeitübertragung ist auch unter dem Namen Streaming bekannt. Am Institut wurde der erste Streamingserver entwickelt, der die Übertragung eines MPEG-4 AVC/H.264 [10], [24], [25] Videostroms in Echtzeit über Netze mit stark variabler Datenrate zeigte. Solche Techniken sind für ein leistungsfähiges internetbasiertes Fernsehen (IPTV) aber auch für Videotelefonie, die heute im Rahmen der Instant Messenger- und Internettelefonieanwendungen schnell an Bedeutung gewinnt, von großer Bedeutung.

Die genannten Arbeiten werden in den folgenden vier Kapiteln vorgestellt. In Kapitel 2 werden Videocodierverfahren beschrieben, die im Vergleich zum aktuellen Videocodierstandard MPEG-4 AVC/H264 eine weitere Steigerung der Codiereffizienz ermöglichen. Aufbauend auf Erkenntnissen der Informationstheorie wird in Kapitel 3 das Distributed Video Coding vorgestellt, dass eine Verlagerung der Rechenkapazität vom Videocoder auf den Decoder ermöglichen wird. Die Codierung dreidimensionaler Netze wird in Kapitel 4 erläutert. Streaming für die Übertragung von Video über moderne internetbasierte Kommunikationsnetze wird in Kapitel 5 beschrieben. In Kapitel 6 werden Audiocodierung und die Anwendung von in der Audiocodierung entwickelten Gehörmodellen in Cochlea-Implantaten vorgestellt.

2 Fortschritte in der Videocodierung

In den letzten Jahren hat die Speicherung und die Übertragung von digitalen Videosignalen eine zunehmende Bedeutung erfahren. Digitale Videosignale in Fernsehauflösung werden heutzutage auf der DVD (Digital Versatile Disc) gespeichert oder mittels DVB (Digital Video Broadcasting). Des Weiteren speichern heutzutage digitale Videorekorder Videosequenzen auf einer Festplatte oder direkt auf einer DVD ab. Derzeit befinden sich die HD-DVD und die Blu-ray Disc in der Einführung, auf der hoch aufgelöste HDTV-Signale [12] (High Definition Television) gespeichert werden. Infolge der immer schneller werdenden Internetzugänge im Heimbereich öffnet sich derzeit auch das Internet mehr und mehr als Übertragungsmedium für digitale Videosignale.

Da die Übertragungskanäle in der Regel Kanalkapazitäten besitzen, die zu gering sind für eine Übertragung von Videosequenzen in unkomprimierter Form, ist eine effiziente Quellencodierung notwendig. Ein entsprechend dem Studioformat ITU-R 601 [11] PCM-codiertes (Pulse Code Modulation)-Fernsehsignal benötigt beispielsweise eine Datenrate von 166 Mbit/s. Durch das von der ISO/IEC MPEG (Motion Picture Experts Group) standardisierte Quellencodierungsverfahren für Videosequenzen MPEG-2 [6] kann die Datenrate von 166 Mbit/s auf etwa 4-6 Mbit/s reduziert werden, ohne dass für einen menschlichen Betrachter Unterschiede zum PCM-codierten Signal sichtbar werden. MPEG-2 wird für die Speicherung digitaler Videosignale auf der DVD und die Übertragung mittels DVB eingesetzt. Die Übertragungskanäle unter Verwendung von DSL oder UMTS stellen jedoch in der Regel eine geringere Datenrate als 4-6 Mbit/s zur Verfügung. Aus diesem Grund ist eine Übertragung eines Videosignals in voller TV-Auflösung über solche Kanäle mit einer MPEG-2-Codierung nur unter Einbußen in der Bildqualität möglich. Eine Weiterentwicklung von MPEG-2 stellt der MPEG-4 Standard dar. Dort kann die Datenrate bei gleicher Bildqualität auf ca. 3-4 Mbit/s reduziert werden. Für eine Übertragung eines Videosignals in voller TV-Auflösung über DSL oder UMTS ist diese Datenrate aber in der Regel noch immer zu hoch. Gemeinsam veröffentlichten ISO/IEC MPEG und ITU-T VCEG Ende 2003 den Standard H.264 bzw. MPEG-4 part 10 Advanced Video Coding (AVC) [10], [24], [25]. Abkürzend wird dieser Standard auch als H.264/AVC bezeichnet. Mithilfe dieses Standards konnte die Datenrate für die Codierung eines Fernsehsignals auf etwa 1.8-3 Mbit/s reduziert werden. Ende 2004 wurden die so genannten Fidelity Range Extensions verabschiedet, die H.264/AVC um Codierungstechniken für Videosequenzen mit hoher Auflösung (HDTV) erweitern. Bild 1 stellt die für ein Fernsehsignal erforderliche Datenrate in Abhängigkeit der Zeit dar.

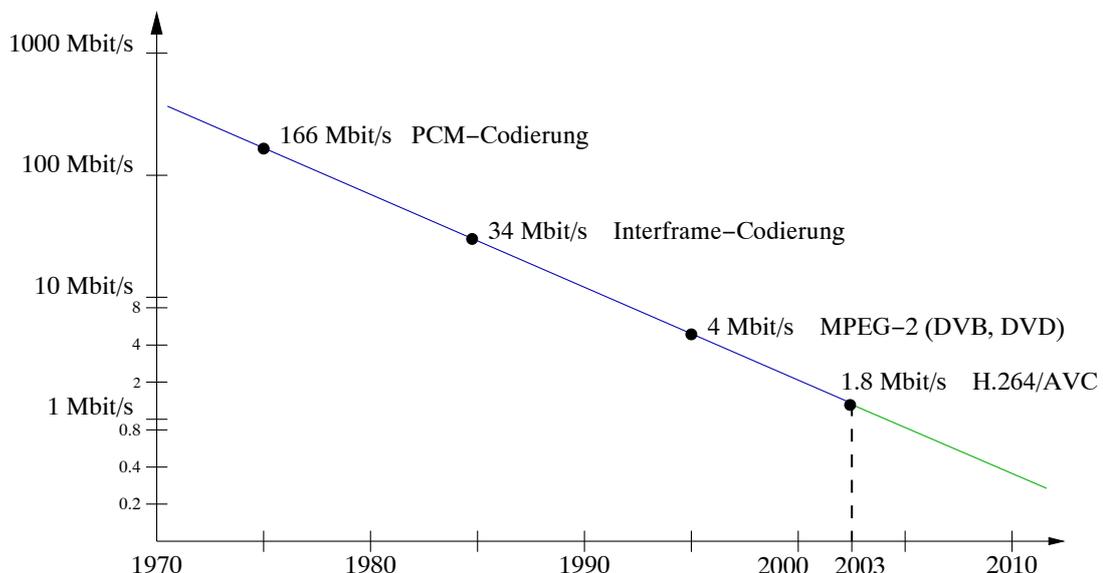


Bild 1 Erforderliche Datenrate für die Codierung eines Fernsehsignals in Abhängigkeit von der Zeit.

In diesem Kapitel werden die grundlegenden Codierungstechniken des H.264/AVC-Standards dargestellt. Darauf aufbauend wird gezeigt, durch welche Codierungstechniken die Datenrate im Vergleich zu H.264/AVC bei gleicher Bildqualität weiter reduziert werden kann.

2.1 Grundlagen der Videocodierung

Alle bisher standardisierten Verfahren zur Codierung von Videosignalen [5], [6], [7], [8], [9], [10] basieren auf dem Prinzip der Hybridcodierung. Dabei wird die Redundanz des Videosignals in zeitlicher Richtung durch eine Prädiktion reduziert. Das zu codierende Bild wird hierfür in Blöcke aufgeteilt. Für jeden dieser Blöcke wird entweder eine Inter- oder eine Intra-Prädiktion durchgeführt. Bei der Inter-Prädiktion handelt es sich um eine bewegungskompensierende Prädiktion, bei der die zeitliche Korrelation zwischen aufeinanderfolgenden Bildern ausgenutzt wird. In bereits übertragenen Bildern, den sogenannten Referenzbildern, wird ein korrespondierender Block mit dem gleichen Bildinhalt gesucht und zur Prädiktion verwendet. Die Position dieses Blockes wird mit einem Displacementvektor beschrieben. Bei der Intra-Prädiktion wird lediglich die örtliche Korrelation innerhalb eines Bildes ausgenutzt. Dieser Mode wird z. B. bei Szenenwechseln verwendet, da dort eine Prädiktion aus vorangegangenen Bildern nicht sinnvoll ist. Das verbleibende Prädiktionsfehlersignal wird mit einer Transformationscodierung codiert, wobei die örtliche Redundanz reduziert wird. Bei der Transformationscodierung wird der Prädiktionsfehler blockweise transformiert und die sich ergebenden Koeffizienten quantisiert und Redundanz mindernd codiert. Die quantisierten Koeffizienten werden zurücktransformiert und nach Addition des Prädiktionssignals in einem Bildspeicher abgelegt. Die abgelegten Bildsignale können dann zur weiteren Prädiktion verwendet werden. Bild 2 zeigt das verallgemeinerte Blockschaltbild eines hybriden Videoencoders gemäß H.264/AVC.

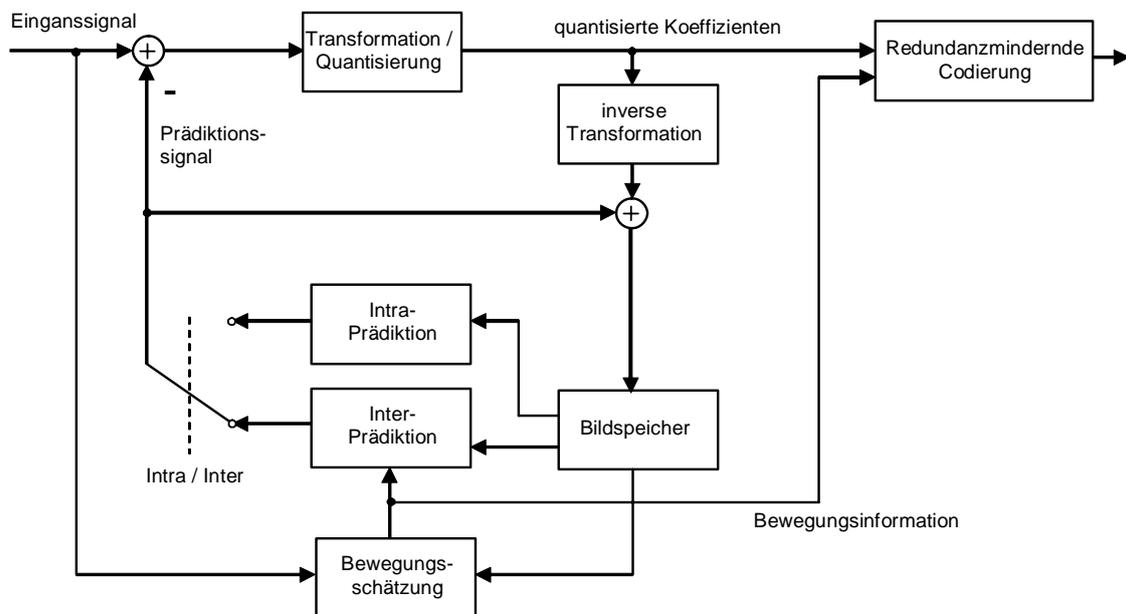


Bild 2 Darstellung eines verallgemeinerten Blockschaltbildes eines hybriden Encoders gemäß H.264/AVC.

2.1.1 Prädiktion

Das aktuell zu codierende Bild wird in Blöcke aufgeteilt. Für jeden dieser Blöcke wird entweder eine Intra- oder eine Inter-Prädiktion durchgeführt. Bei der Intra-Prädiktion werden alle Abtastwerte eines Blockes ausschließlich aus bereits rekonstruierten Abtastwerten desselben Bildes prädiziert. Diese Prädiktionsart wird im Wesentlichen für das erste Bild einer Szene verwendet. Im Gegensatz dazu erfolgt die Inter-Prädiktion aus bereits decodierten Bildern. Die Inter-Prädiktion ist besonders effizient, wenn der Bildinhalt des aktuellen Bildes große Ähnlichkeit mit dem Bildinhalt bereits decodierter Bilder aufweist. Ein Beispiel hierfür ist die Aufnahme einer Szene mit einem Kameraschwenk. Es wird für jeden Block ein korrespondierender Referenzblock ähnlichen Bildinhaltes in einem bereits decodierten Bild gesucht. Die Position des Referenzblockes wird mithilfe eines Displacementvektors beschrieben. In H.264/AVC haben diese Displacementvektoren eine Amplitudenauflösung von 1/4 Bildpunkt. Aus diesem Grund können sie im Referenzbild auch auf eine Position zeigen, die zwischen dem Abtastraster liegt. Diese Positionen werden als Subpel-Positionen bezeichnet. Bild 3 stellt diesen Fall dar. Um das Bildsignal an Positionen zwischen dem Abtastraster zu rekonstruieren, werden unterschiedliche, fest eingestellte Interpolationsfilter verwendet. Verweist der Vektor auf eine Halb-pel

Position, wird das Prädiktionssignal mittels eines separierbaren 6-tap FIR (Finite Impulse Response) Interpolationsfilters berechnet. Prädiktionswerte an Viertel-pel Positionen werden durch Mittelung der Werte der benachbarten Integer-pel bzw. Halb-pel-Positionen berechnet (2-tap FIR Interpolationsfilter).

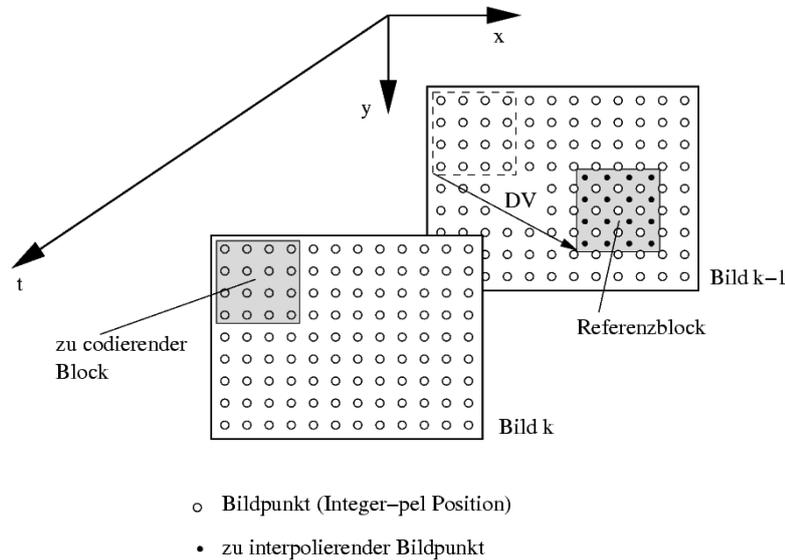


Bild 3 Darstellung der bewegungskompensierenden Prädiktion für den Fall eines Displacementvektors (DV) mit einer Amplitudenauflösung von weniger als einem Bildpunkt.

2.1.2 Transformation und Quantisierung

Für die Codierung des entstehenden Prädiktionssignals wird dieses blockweise einer zweidimensionalen DCT-approximierenden Integertransformation unterworfen. Ziel der Transformation ist es, aus korrelierten Abtastwerten unkorrelierte Koeffizienten zu erzeugen. Die Größe der Transformation ist abhängig von der Größe der Blöcke in der bewegungskompensierenden Prädiktion. Bei Blockgrößen in der bewegungskompensierenden Prädiktion von größer oder gleich 8×8 Bildpunkten kann die Größe der Transformation zu 8×8 oder 4×4 Bildpunkten gewählt werden, wobei die Auswahl als Seiteninformation übertragen wird. Bei kleineren Blockgrößen in der bewegungskompensierenden Prädiktion ist lediglich eine 4×4 -Transformation erlaubt.

Die Quantisierung der entstehenden Koeffizienten erfolgt mittels eines skalaren Quantisierers. H.264/AVC unterstützt hierbei die Berücksichtigung der menschlichen Wahrnehmungseigenschaften und quantisiert höherfrequente Signalanteile entsprechend gröber als niederfrequente. Da die verwendete Transformation eine Spektraldarstellung des Signals erzeugt, wird die frequenzabhängige Quantisierung durch eine gewichtete Quantisierung der Koeffizienten erzielt. Bild 4 zeigt die unterschiedlichen Quantisierungsstufenbreiten für die Koeffizienten eines 8×8 -Blockes.

		f_x							
		9	13	15	17	19	21	22	24
	f_y	13	13	17	19	21	22	24	25
		15	17	19	21	22	24	25	27
		17	19	21	22	24	25	27	28
		19	21	22	24	25	27	28	30
		21	22	24	25	27	28	30	32
		22	24	25	27	28	30	32	33
		24	25	27	28	30	32	33	35

Bild 4 Darstellung der Quantisierungsstufenbreiten für Blöcke mit 8×8 -Koeffizienten.

2.1.3 Redundanz mindernde Codierung

Die quantisierten Koeffizienten werden mittels des so genannten Zigzag-Scans nach fallender Energie angeordnet, s. Bild 5. Zur Redundanz mindernden Codierung der quantisierten Koeffizienten unterstützt H.264/AVC die zwei Codierungen CAVLC (Context-Based Adaptive Variable Length Code) und CABAC (Context Adaptive Binary Arithmetic Coding) [13]. Beide Codierungen nutzen die Anordnung nach fallender Energie aus. Bei der CAVLC handelt es sich um eine Codierung mittels fest eingestellter Codeworttabellen nach Huffman [14], die zustandsabhängig umgeschaltet werden. CABAC ist eine binäre arithmetische Codierung, die zustandsabhängig umgeschaltet wird und eine fortlaufende Anpassung an eine sich ändernde Statistik ermöglicht. Im Gegensatz zu CAVLC benötigt CABAC eine um etwa 10 % geringere Datenrate für die gleiche Bildqualität.

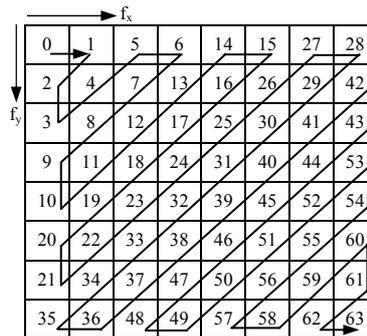


Bild 5 Darstellung des Zigzag-Scans für Blöcke mit 8x8 Koeffizienten.

2.2 Techniken zur weiteren Reduktion der Datenrate

2.2.1 Nicht separierbares adaptives Interpolationsfilter

Wie bereits beschrieben, haben die Displacementvektoren so genannte Subpel-Auflösung. Um solche Displacementvektoren zu schätzen und zu kompensieren, soll das Bildsignal vom Referenzbild an Subpel-Positionen mithilfe der Interpolation erzeugt werden. Aufgrund eines nichtidealen Tiefpassfilters, das bei dem Bildaufnahmeprozess angewendet wird, wird Aliasing erzeugt, der die Interpolation und somit die bewegungskompensierende Prädiktion verfälscht. Neben Aliasing gibt es weitere Störeinflüsse, die die Effizienz der bewegungskompensierenden Prädiktion und damit der Codiereffizienz begrenzen. Ein solcher Störeinfluss wird von fehlerhaften Displacementvektoren erzeugt. Der Displacementschätzfehler ergibt sich zum Einen aus der begrenzten Amplitudenauflösung und zum Anderen aus dem einfachen, blockbasierten und translatorischen Abbildungsmodell, welches bei der Bewegungsschätzung verwendet wird. Ein weiterer Störeinfluss geht von den Quantisierungsfehlern im Referenzbild aus, die sich aus der Quantisierung des Prädiktionsfehlers ergeben [1].

Im TNT wird an Lösungen gearbeitet, die solche Störeffekte beheben und damit die Codiereffizienz bestehender Codierverfahren steigern, ohne die Komplexität des Encoders und des Decoders wesentlich zu erhöhen. Hierfür wird ein zeitvariantes nicht separierbares Interpolationsfilter verwendet.

Im Folgenden wird das adaptive Filter detailliert beschrieben [16]. Für jede Subpel-Position SP ($a \dots o$), s. Bild 6, wird ein individueller Satz von Filterkoeffizienten analytisch berechnet. Wenn die zu interpolierende Subpel-Position a, b, c, d, h oder l ist (s. Bild 6), wird ein eindimensionales 6-tap Filter berechnet, wobei für die Subpel-Positionen a, b, c die Abtastwerte $C1 - C6$ und für die Subpel-Positionen d, h, l die Abtastwerte $A3 - F3$ verwendet werden. Für die restlichen Subpel-Positionen e, f, g, i, j, k, m, n und o , wird ein zweidimensionales nicht separierbares 6x6-tap-Filter berechnet. Für alle Subpel-Positionen werden die Filterkoeffizienten dermaßen berechnet, dass die resultierende Prädiktionsfehlerenergie minimiert wird. Diese Filterkoeffizienten werden zu dem Empfänger als Seiteninformation übertragen.

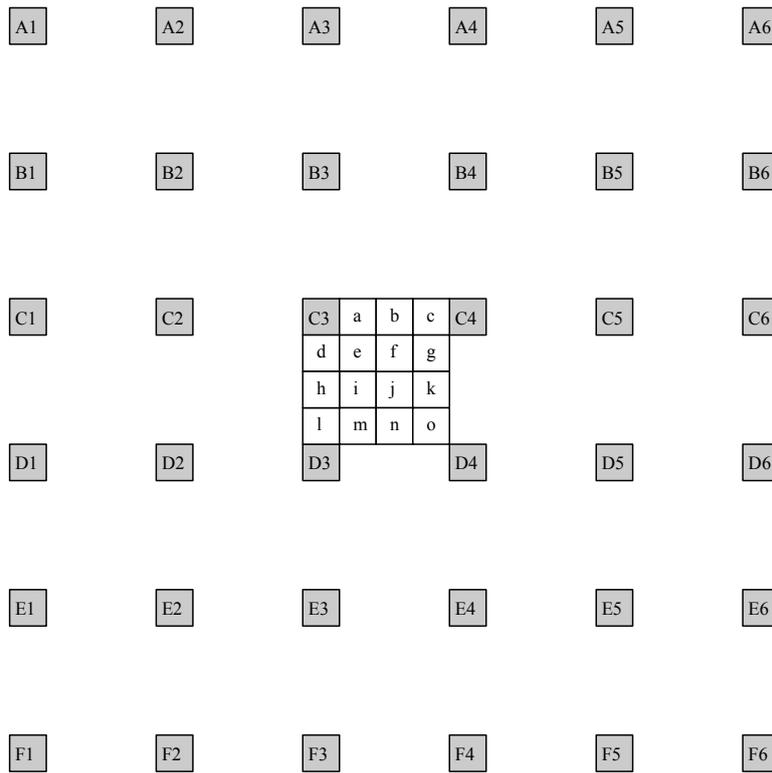


Bild 6 Fullpel- (grau) und Subpel-Positionen (weiß) für Viertelpel-Interpolation.

Man nehme an, dass $h_{00}^{SP}, h_{01}^{SP}, \dots, h_{54}^{SP}, h_{55}^{SP}$ die 36 Filterkoeffizienten eines 6x6-tap 2D-Filters sind, die für eine bestimmte Subpel-Position SP verwendet werden. Dann ist der zu interpolierende Wert p^{SP} ($a \dots o$) ist durch eine zweidimensionale Faltung wie folgt berechnet:

$$p^{SP} = \sum_{i=1}^6 \sum_{j=1}^6 P_{i,j} h_{i-1,j-1}^{SP} \quad (1)$$

wobei $P_{i,j}$ eine Fullpel-Position ($A1 \dots F6$) ist.

Die Berechnung der Filterkoeffizienten und die Bewegungskompensation sind in den folgenden Schritten durchgeführt:

Displacementvektoren $d_t = (mvx, mvy)$ werden für jedes zu codierende Bild mithilfe des Standard-Wiener-Filters geschätzt.

2D-Filterkoeffizienten $h_{i,j}$ werden für jede Subpel-Position SP unabhängig berechnet, indem die resultierende Prädiktionsfehlerenergie minimiert wird:

$$\left(e^{SP} \right)^2 = \sum_x \sum_y \left(S_{x,y} - \sum_i \sum_j h_{i,j}^{SP} P_{\tilde{x}+i, \tilde{y}+j} \right)^2 \quad (2)$$

mit $\tilde{x} = x + \lfloor mvx \rfloor$, $\tilde{y} = y + \lfloor mvy \rfloor$, wobei $S_{x,y}$ das Originalbild und $P_{x,y}$ das Referenzbild sind, i, j - die Filterindizes, mvx und mvy die geschätzten Displacementvektoren und der $\lfloor \dots \rfloor$ -Operator die floor-Funktion, die die geschätzten Displacementvektoren mvx und mvy auf die nächstliegenden und vom Betrag her kleineren Fullpel-Positionen abbildet. Dies ist ein notwendiger Schritt, da das vorher decodierte Referenzbild die Information nur an solchen Fullpel-Positionen enthält. Somit wird für jede Subpel-Position $a \dots o$ ein individueller Filtersatz ermittelt, indem die erste Ableitung von

$(e^{SP})^2$ im Bezug auf den jeweiligen Filterkoeffizienten $h_{i,j}^{SP}$ berechnet und auf null gesetzt wird. Die Anzahl der sich daraus ergebenden Gleichungen ist gleich der Anzahl der Filterkoeffizienten, die für die aktuelle Subpel-Position SP verwendet werden.

$$\begin{aligned} 0 &= \frac{\partial (e^{SP})^2}{\partial h_{k,l}^{SP}} = \frac{\partial}{\partial h_{k,l}^{SP}} \left(\sum_x \sum_y \left(S_{x,y} - \sum_i \sum_j h_{i,j}^{SP} P_{\tilde{x}+i, \tilde{y}+j} \right)^2 \right) \\ &= \sum_x \sum_y \left(S_{x,y} - \sum_i \sum_j h_{i,j}^{SP} P_{\tilde{x}+i, \tilde{y}+j} \right) P_{\tilde{x}+k, \tilde{y}+l}, \forall k, l \in \{0;5\} \end{aligned} \quad (3)$$

Unter der Annahme, dass die Bildeigenschaften symmetrisch sind, wird das gleiche Filter für Subpel-Positionen e, g, m und o mit 21 Filterkoeffizienten verwendet. Für die Subpel-Positionen f, i, k und n ergeben sich 18 Filterkoeffizienten, für j 6, für b und h 3 und schließlich für a, c, d und l 6 Filterkoeffizienten.

In diesem Schritt werden neue Displacementvektoren geschätzt, wobei die Referenzbilder mit dem bereits berechneten Filter interpoliert werden. Dieser Schritt ermöglicht die Reduktion von Fehlern, erzeugt durch die Bewegungsschätzung, Aliasing, Kamerarauschen, etc. auf der einen Seite und optimiert die Rate-Distortion-Performance auf der anderen Seite.

Die Filterkoeffizienten werden quantisiert, Redundanz mindernd codiert und als Seiteninformation übertragen [17].

2.2.2 Displacementvektorauflösung von 1/8-pel

Aufgrund einer geringen Amplitudenauflösung der Displacementvektoren kann es zu Displacement-schätzfehlern kommen, die dazu führen können, dass relativ große Prädiktionsfehler verbleiben. Die Codierung großer Prädiktionsfehler erfordert eine hohe Datenrate. Um die Displacementschätzfehler und damit die erforderliche Datenrate zu reduzieren, wurde am TNT ein Verfahren entwickelt, welches die Amplitudenauflösung von 1/4-pel auf 1/8-pel erhöht [1], [15]. Dieses Verfahren kann mit den zuvor beschriebenen adaptiven Interpolationsfiltern kombiniert werden.

2.2.3 Adaptive Codierung des Prädiktionsfehlers im Orts- und im Frequenzbereich

Bisher wurde der Prädiktionsfehler blockweise einer Transformation unterworfen. Das Ziel dieser Transformation ist, unkorrelierte Koeffizienten zu erhalten, die unabhängig voneinander codiert werden können. Die Effizienz dieser Transformationscodierung ist hoch, wenn die Abtastwerte des Prädiktionsfehlersignals untereinander korreliert sind. In Situationen, in denen die Abtastwerte nur schwach oder nicht korreliert sind, bringt die Transformation keinen Vorteil. Verbesserungen in der bewegungskompensierenden Prädiktion, wie z. B. die separierbaren adaptiven Prädiktionsfilter und die Displacementvektorauflösung von 1/8-pel, führen dazu, dass die verbleibende Korrelation im Prädiktionsfehlersignal geringer wird, sodass diese Situationen zunehmend eintreten. Solche Blöcke können direkt im Ortsbereich ohne Transformation codiert werden. Aus diesem Grund wurde am TNT eine Prädiktionsfehlercodierung entwickelt, die die Blöcke adaptiv entweder mittels Transformation im Frequenzbereich oder ohne Transformation direkt im Ortsbereich codiert [19], [21], [22], [23]. Die Auswahl der Codierung erfolgt für jeden Block individuell anhand der benötigten Datenrate und des entstehenden Quantisierungsfehlers. Das Blockschaltbild dieser Codierung ist in Bild 7 dargestellt.

2.3 Vergleich der Codierungseffizienz

Beim Vergleich der Codierungseffizienz verschiedener Standards [18] ist festzustellen, dass bei Rundfunkqualität von 35 dB PSNR der 1998 entwickelte Standard H.263 (High Latency Profile) im Vergleich zu MPEG-2 aus dem Jahr 1994, der noch heute aktiv eingesetzt wird (DVD, DVB-T), die erforderliche Datenrate um etwa 25 % reduziert. Eine weitere Reduktion der Datenrate in Höhe von ca. 20 % lässt sich mit MPEG-4, entwickelt im Jahr 2000, erzielen. Betrachtet man den jüngsten Standard H.264/AVC, der im Jahr 2003 verabschiedet wurde, so reduziert er die Datenrate um weitere 40 % (siehe Bild 7).

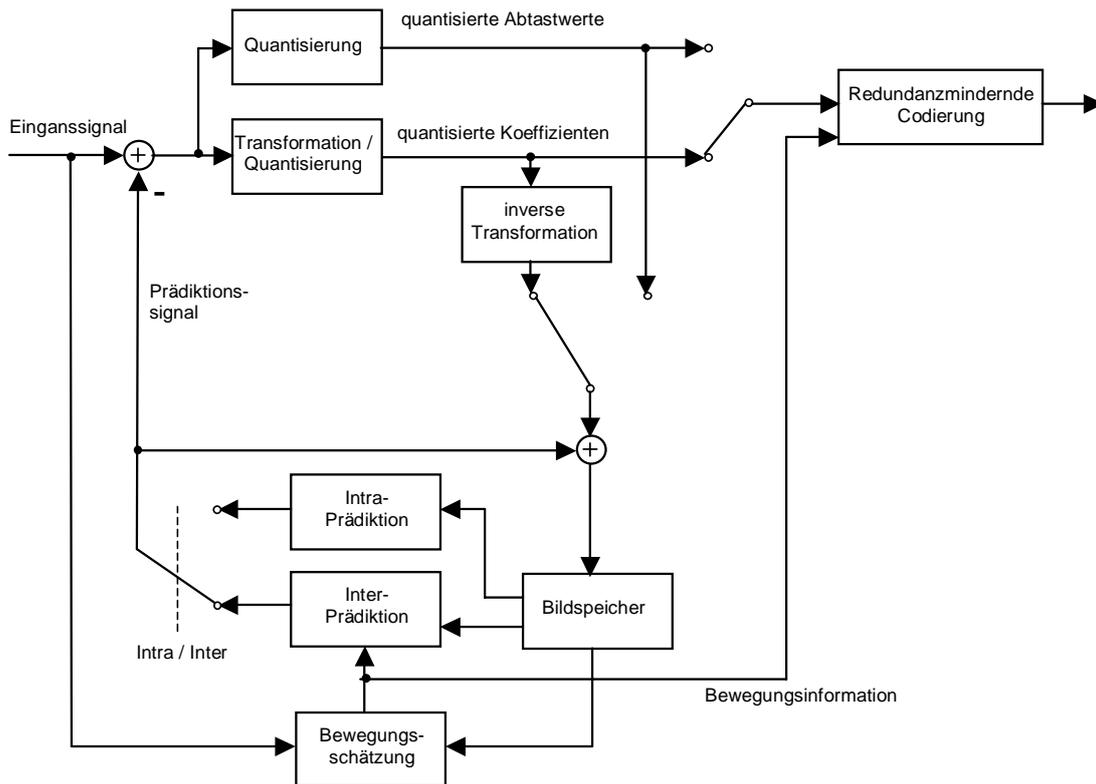


Bild 7 Darstellung eines verallgemeinerten Blockschaltbildes eines hybriden Encoders mit adaptiver Prädiktionsfehlercodierung.

Obwohl die Datenrate in den letzten 10 Jahren weit über 50 % reduziert werden konnte, ist eine Sättigung noch nicht erreicht. Die am TNT entwickelten Techniken ermöglichen durchschnittlich weitere 12 % Reduktion, verglichen mit H.264/AVC. Aus diesem Grund wurden diese Techniken in die sogenannte KTA (Key Technical Area) Software [20] aufgenommen. Bei dieser Software handelt es sich um die Exploration-Software der VCEG. Tabelle 1 stellt den Vergleich der Codierungseffizienz verschiedener Standards anhand der repräsentativen Testsequenz Mobile&Calendar dar. Weitere Vergleiche können unter [18], [19] nachgelesen werden.

Tabelle 1 Vergleich der Codierungseffizienz von verschiedenen Standards

	Mittlere Datenratensparnisse im Vergleich zu:		
Coder	MPEG-4 ASP	H.263 HLP	MPEG-2 MP
H.264/AVC MP	37.44 %	47.58 %	63.57 %
MPEG-4 ASP	-	16.65 %	42.95 %
H.263 HLP	-	-	30.61 %

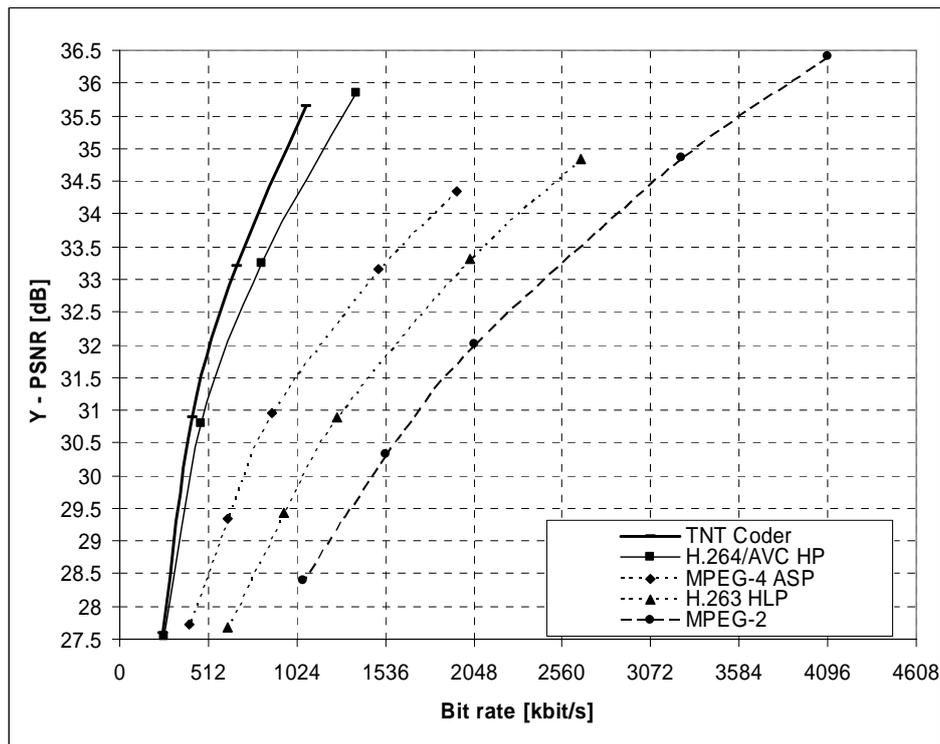


Bild 8 Darstellung des gemessenen Signalrauschabstandes in Abhängigkeit von der Datenrate für die Codierungen MPEG-2, H.263 (High Level Profile), MPEG-4 (Advanced Simple Profile), H.264/AVC (High Profile) sowie für den TNT-Coder. Sequenz: Mobile&Calendar in CIF, 30Hz.

3 Distributed Video Coding

3.1 Unterschiedliche Modelle in der Videocodierung

Wie in Kapitel 2 beschrieben, werden in den heute standardisierten Videocodern die Statistiken der Quellen am Encoder bestimmt und ausgenutzt, wodurch es zu einer ungleichen Verteilung der Komplexität zwischen Encoder und Decoder kommt. So ist zum Beispiel der Encoder eines H.264 Systems fünf bis zehn mal komplexer als der Decoder. Dieses Modell wurde hauptsächlich durch Anwendungen wie Broadcasting, Video-on-Demand und der DVD vorangetrieben, bei denen das Video einmal codiert und mehrmals decodiert wird.

Distributed Video Coding (DVC) setzt ein anderes Modell ein, um effiziente Kompressionen zu erreichen. Bei DVC ist es die Aufgabe des Decoders die Quellenstatistiken - teilweise oder ganz - zu bestimmen. Diese Verschiebung der Aufgaben auf den Decoder erlaubt die Entwicklung von einfachen Encodern und komplexeren Decodern. Dieser neue, von keinem Standard berücksichtigte Ansatz ist für mobile Videokameras, kabellose Überwachungssysteme, Einwegkameras, Sensorenetzwerke, verschiedene medizinische Anwendungen, etc. besonders geeignet, da aufgrund von unzureichendem Speicher und Energie einfachere Encoder nötig sind.

DVC basiert hauptsächlich auf den Theoremen von Slepian-Wolf [26] und Wyner-Ziv [27]. Diese Theoreme besagen, dass es möglich ist, zwei statistisch abhängige Quellen in einem dezentralen System (getrennte Codierung, gemeinsame Decodierung) mit der gleichen Rate zu übermitteln, wie bei einem System mit gemeinsamer Codierung und Decodierung. Obwohl diese Theoreme schon in den siebziger Jahren vorgestellt wurden, fanden sie erst um die Jahrtausendwende Anwendung in der Videocodierung.

3.2 Aufbau eines DVC-Coders

Der am TNT verwendete Coder wurde in dem Gemeinschaftsprojekt DISCOVER (Distributed Coding for Video Services, [28]) entwickelt. In Bild ist die zugrunde liegende Architektur des DISCOVER-codec dargestellt.

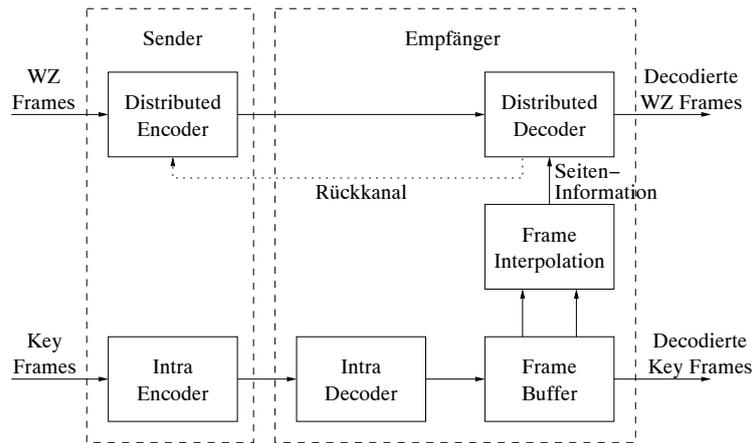


Bild 9 Aufbau eines DVC-Coders.

Am Encoder wird die Sequenz in Wyner-Ziv und Key Frames aufgeteilt, die die zwei statistisch abhängigen Quellen darstellen. Die Key Frames werden unabhängig von den WZ-Frames konventionell als Einzelbilder mit Hilfe des im Kapitel 2 beschriebenen H.264 Videocoders übertragen. Auf der Empfangsseite können zwei zeitlich benachbarte Key Frames verwendet werden, um eine Schätzung des WZ-Frames zu berechnen. Dieses interpolierte Bild wird auch als Seiteninformation bezeichnet und stellt ein gestörtes Abbild des WZ-Frames dar. Mit Hilfe von Parity-Bits, die vom Encoder gesendet werden, können die Störungen in der Seiteninformation eliminiert werden. Ist die Seiteninformation zu gestört um das Bild rekonstruieren zu können, fordert der Decoder weitere Parity-Bits über den Rückkanal an. Zur Berechnung und Decodierung der Parity-Bits wird ein Turbo Coder [29] genutzt, der bei großen Blocklängen sehr effizient arbeitet. Da der WZ-Bitstrom nur aus Parity-Bits besteht, die zur Korrektur der Seiteninformation nötig sind, hängt die Effizienz des Systems stark von der verwendeten Interpolationstechnik ab. Aus diesem Grund wird die Seiteninformation mithilfe einer bewegungskompensierenden Interpolation [30] berechnet. Obwohl die Bewegungskompensation sehr ähnlich zu der in Kapitel 2.1.1 vorgestellten Inter-Prädiktion ist, können Bewegungsschätzer aus der konventionellen Videocodierung hier nur bedingt genutzt werden, da bei DVC das Originalbild nicht zur Verfügung steht. Doch trotz dieser Einschränkung konnte am TNT gezeigt werden, dass eine Subpel genaue Schätzung der Bewegungsvektoren [31] zu einer weiteren Verbesserung der Seiteninformation führt.

3.3 Vorteile des DVC

Durch Distributed Video Coding ist es möglich, die Komplexität zwischen Encoder und Decoder flexibler zu verteilen. Der Gewinn entsteht am Decoder, der die Korrelationen auswertet, die bei der präzifizierenden Codierung am Encoder ausgenutzt werden. So konnte die Effizienz gegenüber Intracodierung bei gleichbleibender Komplexität des Encoders wie in Bild 10 erhöht werden. Für die Coastguard Sequenz konnte der Signalrauschabstand um bis zu 1,2 dB erhöht werden. Bei der Flowergarden Sequenz liegen die Gewinne deutlich über 2 dB, da hier die Bewegungskompensation eine gute Interpolation der Seiteninformation liefert. Jedoch liegt die Performanz noch weit unter konventioneller Inter-codierung. Doch das steigende Forschungsinteresse an DVC zeigt, dass ganz neue Anwendungen entstehen können, die aufgrund begrenzter Ressourcen am Encoder bis heute noch nicht möglich waren.

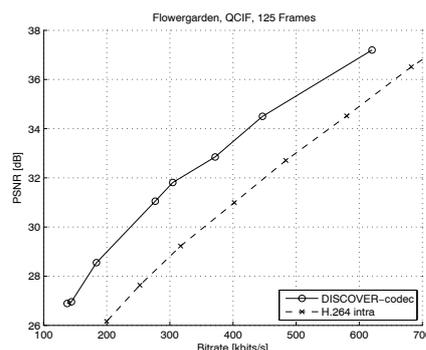
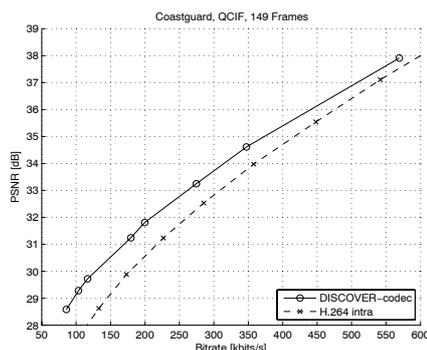


Bild 10 Rate-Distortion Kurven der QCIF-Sequenz Coastguard (links) und Flowergarden (rechts).

4 Prädiktive Codierung von 3D-Netzsequenzen

Grafikkarten für PCs werden stets leistungsfähiger, und das zu immer erschwinglicheren Preisen. Neuerdings werden auch mobile Endgeräte, wie Smartphones, mit Grafikchips ausgerüstet. Derartige Entwicklungen fördern nicht nur eine ständige Verbesserung bestehender Anwendungen, sondern sie ermöglichen auch die Entwicklung neuartiger Anwendungen, wie z. B. 3D-Fernsehen, Fernchirurgie in der Medizin oder neuartige interaktive Spiele für die Unterhaltungsbranche. Eine effiziente Kompression von zeitlich-veränderlichen 3D-Inhalten bekommt in diesem Zusammenhang eine große Bedeutung.

Zeitlich-veränderlicher 3D-Inhalt wird in seiner einfachsten Form durch eine Folge von 3D-Netzen bzw. Frames dargestellt. Frames bestehen aus zwei Typen von Daten: Konnektivität bzw. Knoten und Kanten, und 3D-Positionen, die als Attribute an den Knoten gespeichert sind. In Bild 11 sind die geometrischen Realisierungen von 3 Frames zu sehen. Oft haben alle Frames solch einer Sequenz dieselbe Konnektivität, während sich die 3D-Positionen von Frame zu Frame ändern können. Im Folgenden betrachten wir 3D-Netzsequenzen bestehend aus F Frames und V Knoten pro Frame. Jedem Knoten v eines Frames f wird eine 3D-Position p_v^f zugeordnet. Da wir annehmen, dass die Konnektivität unverändert bleibt über die Zeit, wird diese nur einmal codiert [35], [36]. Der größte Teil einer codierten 3D-Netzsequenz besteht demnach aus 3D-Positionen.

Am TNT wurden Verfahren zur prädiktiven Codierung von 3D-Positionen entwickelt. Im Folgenden werden zwei Verfahren vorgestellt: ein Verfahren zur *nicht-skalierbaren* Codierung und ein Verfahren zur *örtlich-skalierbaren* Codierung von 3D-Positionen.

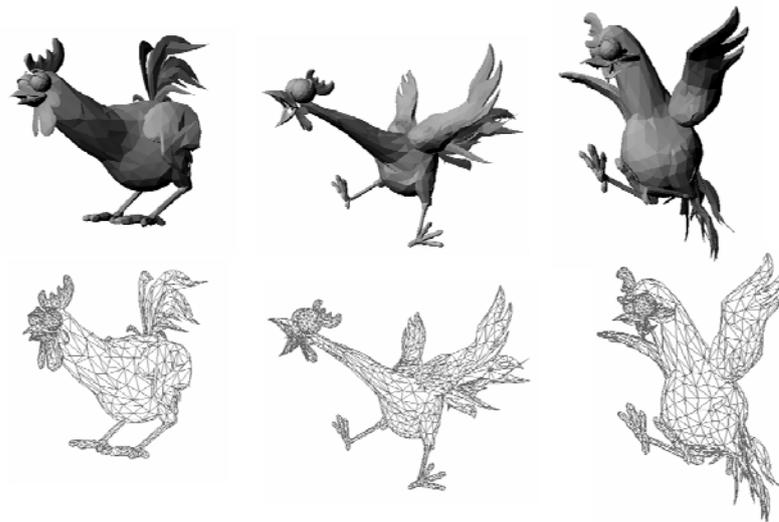


Bild 11 3 Frames einer 3D-Netzsequenz.

4.1 Prädiktive nicht-skalierbare Codierung

In der prädiktiven nicht-skalierbaren Codierung [32] werden alle 3D-Positionen in einer vordefinierten Reihenfolge codiert

$$p_1^1, \dots, p_V^1, \dots, p_1^f, \dots, p_V^f, \dots, p_1^F, \dots, p_V^F,$$

indem zunächst alle Knoten innerhalb eines Frames f codiert werden, bevor die Knoten des darauffolgenden Frames $f+1$ codiert werden. Die Reihenfolge, in der die Knoten innerhalb eines Frames f codiert werden, wird mithilfe eines Ausbreitungsalgorithmus, der sich entlang von Kanten der Konnektivität ausbreitet, definiert. Dabei werden Regionen bereits besuchter und noch zu besuchender Knoten gebildet (s. Bild 12). Nachdem alle Knoten besucht wurden, steht die Reihenfolge der Codierung innerhalb eines Frames fest.

Der Codierungsalgorithmus verwendet das Paradigma der prädiktiven Codierung. Dabei wird jedes Mal, wenn ein neuer Knoten besucht wird, seine 3D-Position p_v^f durch die 3D-Positionen bereits codierter Nachbarknoten prädiziert. Der dabei entstehende Prädiktionsfehler wird dann codiert. Der

Ausbreitungsalgorithmus sorgt immer dafür, dass bereits codierte 3D-Positionen in der Nachbarschaft des neuen zu codierenden Knotens existieren.

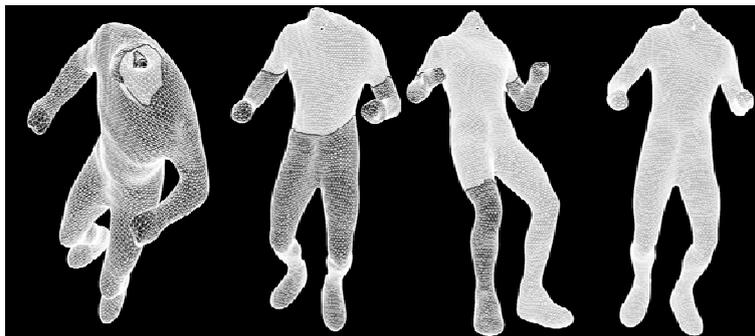


Bild 12 Illustration der Reihenfolge des Durchlaufs durch die Knoten eines Netzes. Die graue Region stellt den Bereich der bereits besuchten Knoten dar, während die schwarze Region noch durchlaufen werden muss.

In Bild 13 ist der Extended-Lorenzo-Prädiktor (ELP) illustriert. Er verwendet für die Prädiktion die 3D-Positionen benachbarter Knoten. Dazu werden ein benachbartes Dreieck aus Frame f und das korrespondierende Dreieck aus Frame $f-1$ verwendet. Durch Parallelogrammbildung wird ein Bewegungsvektor für Knoten v aus Frame $f-1$ prädiziert (gestrichelt dargestellt). Da die 3D-Position von diesem Knoten p_v^{f-1} bereits codiert ist, erhält man durch Addieren mit dem zuvor prädizierten Bewegungsvektor einen Prädiktionswert $\text{pred}_{\text{ELP}}(v, f)$ für p_v^f . Es entsteht der in Bild 13 durch einen kurzen Pfeil dargestellte Prädiktionsfehler.

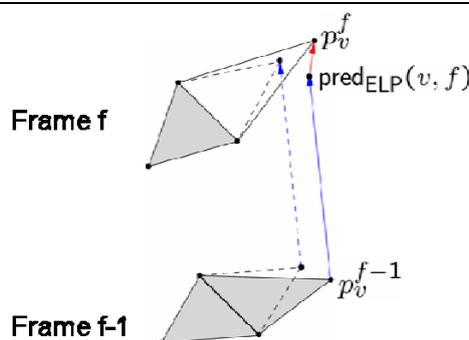


Bild 13 Der Extended-Lorenzo-Prädiktor. Graue Dreiecke gehören zur bereits codierten Region.

Entstehende Prädiktionsfehler werden anschließend quantisiert und entropiecodiert. Eine gute Prädiktion gewährleistet, dass die meisten Prädiktionsfehler in der Nähe von 0 liegen. Um sich den verändernden statistischen Eigenschaften von Frame zu Frame während der Entropiecodierung anzupassen, wird in diesem Bereich eine adaptive arithmetische Codierung der Ordnung 0 [37] angewandt. Größere Prädiktionsfehler, die selten vorkommen, werden mit Golomb-Codes [38] codiert.

4.2 Prädiktive örtlich-skalierbare Codierung

Die immer größer werdende Anzahl an unterschiedlichen Zugriffsmodi für 3D-Inhalte, wie Zugriff über das Internet oder lokale Netzwerke, Zugriff mittels mobiler Endgeräte etc., macht es nötig, die Bitraten komprimierter 3D-Inhalte den Übertragungsraten des Netzwerkes und die 3D-Inhalte selbst den Endgeräten anzupassen. Um dieser Forderung entgegenzukommen, wird die skalierbare Codierung eingesetzt.

In der prädiktiven örtlich-skalierbaren Codierung werden alle 3D-Positionen von Frame zu Frame in der Ordnung $1, \dots, f, \dots, F$ codiert. Innerhalb eines Frames f werden allerdings die 3D-Positionen layerweise in der Reihenfolge

$$v \in V_1, V_2, \dots, V_l, \dots, V_{L-1}, V_L$$

codiert, beginnend mit allen Knoten aus dem Base-Layer V_1 und endend mit allen Knoten aus dem höchsten Layer V_L . Welche Knoten zu einem Layer V_l gehören, wird durch Netzvereinfachung bestimmt. Dazu wird das Netz in so genannte Patches zerlegt (s. Bild 14). Ein Patch ist eine Menge von Dreiecken, die einen gemeinsamen Knoten besitzen. Durch Entfernen der mittleren Knoten aller Patches wird eine Menge von Knoten erhalten, die einen Layer V_l bilden. Eine nachträgliche Retriangulierung der zurückbleibenden Löcher liefert ein Frame mit geringerer örtlicher Auflösung. Durch eine Iteration dieser Schritte wird eine disjunkte Zerlegung der Menge aller Knoten $V = \bigcup_{l=1}^L V_l$ erhalten.

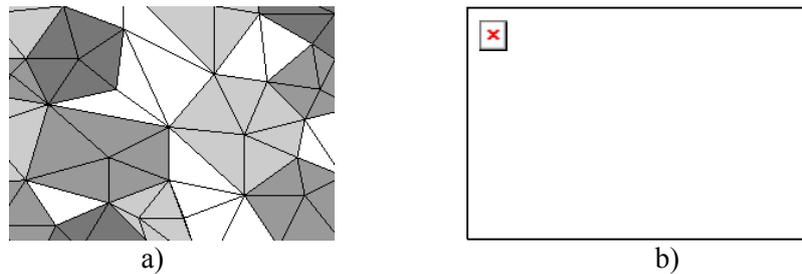


Bild 14 a) Zerlegung eines Netzes in Patches. Grau eingefärbte Dreiecke sind zu Patches zugeordnet. Weiße Dreiecke konnten zu keinen Patches zugeordnet werden.

b) Das Netz nach Entfernung der mittleren Knoten eines jeden Patches und Retriangulierung der zurückbleibenden Löcher.

Hierdurch ist es möglich, eine 3D-Netzsequenz layer-weise zu codieren (s. Bild 15). Dazu werden die 3D-Positionen eines Layers V_l aus Frame f aus den bereits codierten 3D-Positionen des darunterliegenden Layers V_{l-1} und des Frames $f-1$ prädiziert. Der dabei entstehende Prädiktionsfehler wird anschließend quantisiert und entropiecodiert.

Mit einer örtlich-skalierbaren Codierung ist es nötig nur einmal layer-weise zu codieren, während die Decodierung an allen möglichen Arten von Endgeräten mit der besten zulässigen Qualität erfolgen kann. Hierbei kann die Bitrate bzw. die Qualität über die Anzahl der gesendeten Layer gesteuert werden.

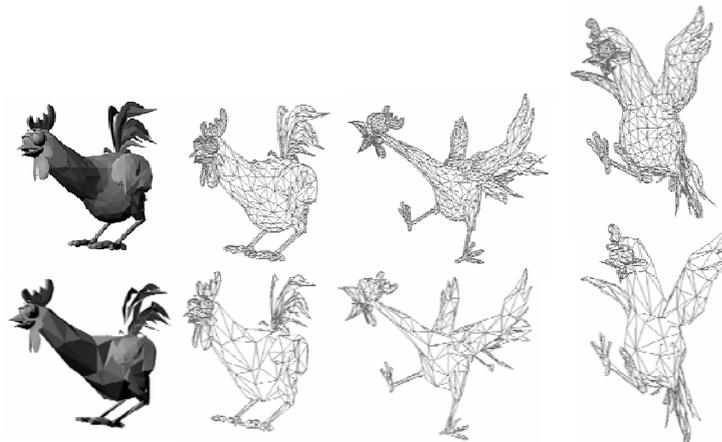


Bild 15 4 Frames einer 3D-Netzsequenz. Die obige Sequenz wird aus der unteren durch Hinzunehmen der Daten weiterer Layer erzeugt.

4.3 Ergebnisse der Netzcodierung

Die beiden vorgestellten Verfahren wurden experimentell ausgewertet anhand der 3D-Netzsequenz *Chicken Crossing*, bestehend aus 400 Frames und 3030 Knoten pro Frame. Als Fehlermaß wurde eine normalisierte knotenbasierte L_2 -Norm verwendet, die in Bild 16 mit d_a bezeichnet wird. Die Bitrate wurde in der Einheit Bits pro Knoten und Frame (bpvf) gemessen. Die Qualität der codierten Netze wurde variiert durch sukzessive Änderung der Quantisierung.

In Bild 16 sind die Ergebnisse der Auswertung dargestellt. Eine Codierung mit einem Layer entspricht einer prädiktiven nicht-skalierbaren Codierung, während Codierungen mit mehreren Layern durch eine prädiktive örtlich-skalierbare Codierung erhalten wurden. Die Ergebnisse zeigen, dass in dem Bereich unterhalb von 7 bpvf eine skalierbare Codierung einer nicht-skalierbaren vorzuziehen ist. Je mehr Layer verwendet wurden für die Codierung, desto besser wurde sie. Der Grund dafür ist, dass die Prädiktion während einer skalierbaren Codierung auf Interpolation basiert, die robuster gegen Quantisierungsrauschen ist, während die Prädiktion bei einer nicht-skalierbaren Codierung auf Extrapolation basiert, und somit anfälliger gegenüber Quantisierungsrauschen ist. Sollte die Eigenschaft der Skalierbarkeit nicht von Nöten sein, so ist für eine hochqualitative Codierung (oberhalb von 7 bpvf) allerdings die nicht-skalierbare Codierung vorzuziehen.

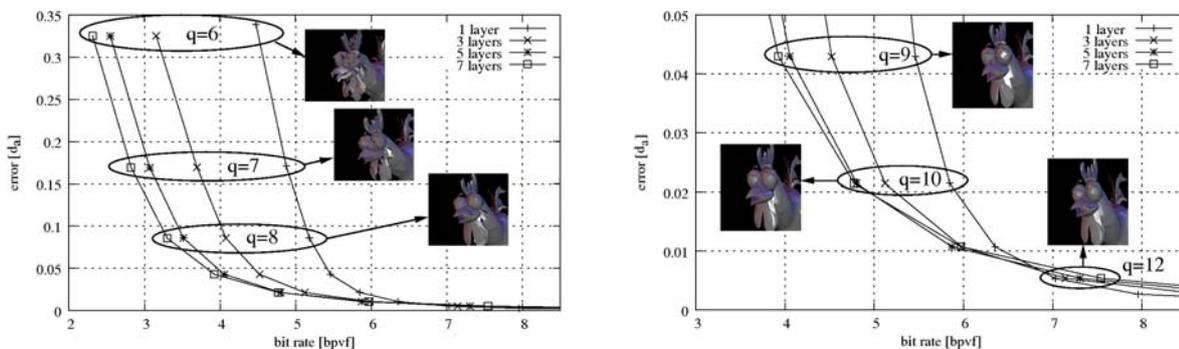


Bild 16 Experimentelle Ergebnisse.

5 Videostreaming

5.1 Multimedia Streaming System: Stand der Technik und zu lösende Probleme

Zusammen mit der Zunahme von Hochgeschwindigkeits-Internetzugängen hat das Streaming von Multimediadaten über das Internet in der letzten Zeit an Interesse gewonnen. Anders als beim download-then-play Service werden die Multimediadaten beim Streaming beim Empfänger nicht gespeichert, sondern nur kurz gepuffert, bevor sie abgespielt werden. Seit Mitte 2004 kam Apples AirPort Express mit großem Erfolg auf den amerikanischen Markt. Das Gerät ist eine 802.11g-WLAN-Basisstation in der Größe einer Steckdose und ausgerüstet mit Audiostreaming-Technologie, um Musikstücke vom Computer an die konventionelle Musikanlage über Funkweg zu streamen. Im Mitte September 2006 hat Apple iTV vorgestellt. Die iTV Box hat etwa die Größe eines MacMini und kann sogar Filme aus dem iTMS (iTunes Music Store) auf den Fernseher gestreamt bringen. In Deutschland bieten Telekom und Alice ab Herbst 2006 den Kunden IPTV (Internet Protocol Television) in HDTV-Qualität an. IPTV ist ein Dienst zur Übertragung von Fernsehprogrammen und Filmen über ein digitales Datennetz, das das dem Internet zugrunde liegende Internet Protokoll (IP) verwendet und eine konstante garantierte Bitrate (CBR) anbietet. Die IPTV-Angebote werden nur Kunden mit VDSL 50 Mbit/s (Very high bitrate Digital Subscriber Line) oder mindestens DSL-16000 Anschlüssen anboten. Die Videodaten sind im veralteten, doch noch beliebten MPEG-2 Format codiert.

Dass Streaming-Produkte nur begrenzt angeboten werden, liegt in erster Linie an den sehr hohen Anforderungen, die Streaming-Applikationen über das Internet an Reaktionszeit und Fehlerverträglichkeit stellen. Dabei spielt das Kommunikationsnetzwerk eine große Rolle. Beispielhaft sind der Protokollstapel und die Paketverschachtelung für Multimedia-Streaming über Ethernet (LAN) in Bild 17 dargestellt. Neben den gewöhnlichen Protokollen wie IP, TCP und UDP werden zusätzlich RTSP- und RTP/RTCP-Protokoll für eine Streaming-Sitzung verwendet. Das RTSP-Protokoll ist für den Auf- und

Abbau einer Sitzung sowie die Kontrolle innerhalb der Sitzung zuständig. Für die Übertragung der Multimediadaten und die Kontrolle dieser Übertragung sind das RTP- bzw. RTCP-Protokoll verantwortlich. Ein RTP-Paket enthält die Multimediadaten als Nutzdaten und einen Header, der die notwendigen Informationen für die Abspielreihenfolge und die Synchronisation der Multimediadaten enthält. Jede weitere untere Schicht im Protokollstapel fügt einen entsprechenden Header hinzu. Am Ende entsteht ein Ethernetrahmen, der über das physikalische Medium übertragen wird. Die RTSP-Pakete werden über eine zuverlässige TCP-Verbindung und die RTP/RTCP-Pakete über die unzuverlässige UDP-Verbindung gesendet. Der Grund dafür ist, dass die Multimediadaten verzögerungsempfindlich und in gewisser Hinsicht tolerant gegenüber Paketverlust sind. Im Gegensatz dazu gibt TCP am Empfänger die ankommenden Pakete nicht an die höhere Schicht weiter, solange ein vorhergegangenes Paket in der Reihe noch fehlt. Während TCP komplex mit umfangreichen Header-Informationen, Bestätigungs- und Flussregelungsmechanismen entworfen wurde, verfügt UDP nur über einen Header mit Paketlänge und Checksum. Von daher muss die Applikationsschicht über UDP die Flussregelung (in Englisch: Congestion Control) vornehmen, um die gesendete Datenrate an Änderungen der Netzwerkbedingungen zu adaptieren.

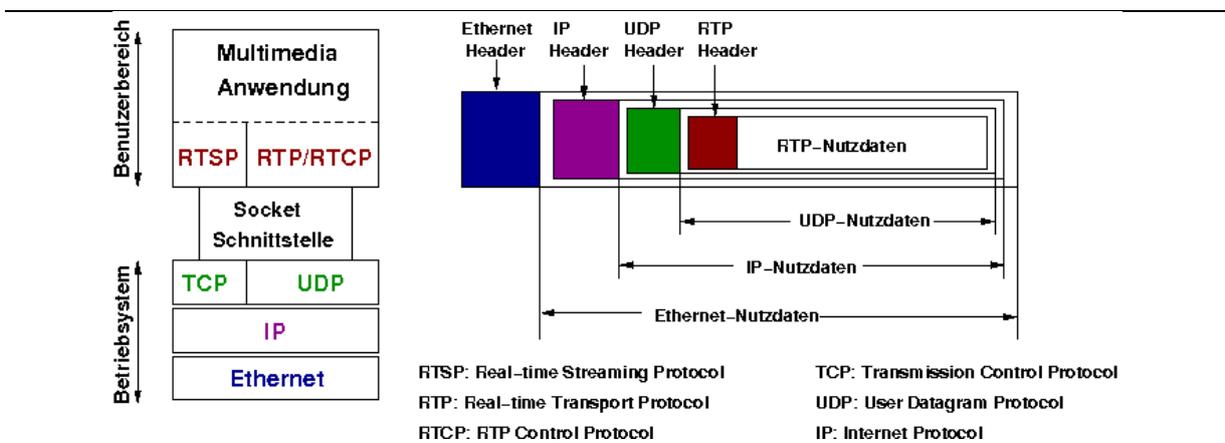


Bild 17 Protokollstapel und Paketverschachtelung für Multimedia-Streaming.

Die Entstehung von Änderungen im IP-Netzwerk hinsichtlich Bandbreiten, Verzögerungszeit oder Paketverluste veranschaulicht das Szenario in Bild 18. Zuerst empfängt Host A1 ein Video per Streaming von Host B1. Da die Videodatenrate sehr groß ist, belegt die Streaming-Applikation fast die ganze Bandbreite der Strecke vom Haus bis zum ersten Router. Kurz danach fängt Host A2 an, von Host B2 eine Datei herunterzuladen. Da das Internet nach dem Besteffortprinzip funktioniert, versucht es die beiden Applikationen unterzubringen, aber diese konkurrieren miteinander. Damit sinkt die verfügbare Bandbreite für die Streaming-Applikation. Das Internet ist im Backbone zwischen den Routern überdimensioniert. Nur am Rand, wo mehrere Nutzer an einem Router anschließen sind, ist die Bandbreite aus wirtschaftlichen Gründen niedrig. Wenn der Puffer eines Routers überlastet wird, verwirft er die ankommenden Pakete, was zum Paketverlust führt. Weiterhin erhöht sich die Bearbeitungszeit für Pakete in der Warteschlange und erhöht damit die Latenzzeit und ihre Varianz. Wenn Host B1 die gesendete Datenrate nicht rechtzeitig adaptiert, kann dies zur Zwangsbeendigung der Sitzung führen, da der Puffer auf Host A1 leer läuft und die Videowiedergabe in Echtzeit erfolgen muss.

Die Fähigkeit zur Datenratenadaptierung ist erforderlich, um Streaming zu einer erfolgreichen kommerziellen Anwendung zu machen. Dies beinhaltet sowohl einen intelligenten Congestion Control des Streamingsystems als auch die Datenratenadaptierungsfähigkeit des Videoformats selbst. Im kommerziellen Bereich wurde das Thema Congestion Control bisher umgangen, indem das Streamingsystem nur im Home-Netz wie bei Apple oder im Internet nur für sehr breitbandige Anschlüsse wie im Fall der IPTV-Angebote von Telekom und Alice realisiert wird. Hierbei gibt es keine Konkurrenz oder die Einflüsse der konkurrierenden Datenströme können vernachlässigt werden. In der Forschung wurde bisher TCP-friendly Congestion Control wie z.B. die binomialen Congestion Control Methoden [39] entwickelt. Im Hintergrund steht, dass fast die gesamte Kommunikation im Internet auf TCP basiert. Die binomialen Congestion Control Methoden haben zwei Nachteile. Erstens fängt eine Streamingsitzung mit einer niedrigen Datenrate (slow-start) an. Zweitens ist die Datenratenadaptierung von der

vorherigen gesendeten Datenrate abhängig und eine gute Adaptierung erfolgt nur bei einer ausreichenden Anzahl von Bestätigungspaketen.

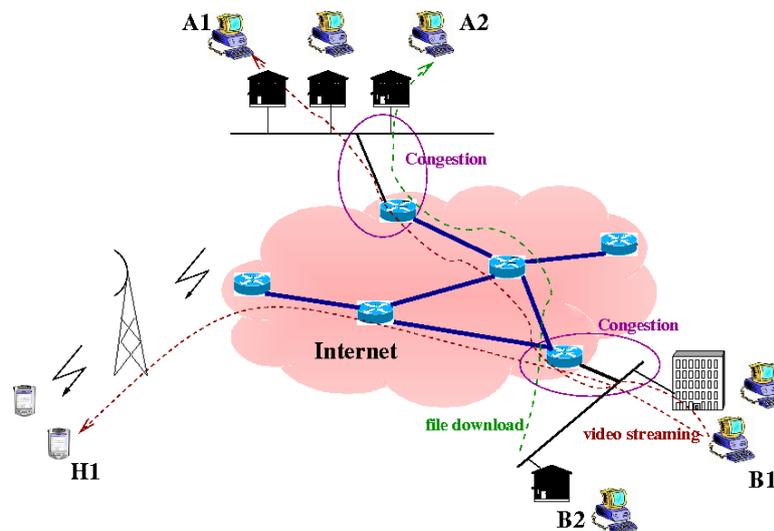


Bild 18 Ursache der Überlastung und des Paketverlusts beim Streaming über ein IP-Netzwerk.

Die Datenratenadaptierungsfähigkeit des Videoformats ist bis heute ein aktuelles Forschungsthema. Die bisherigen Videocodierungsstandards ermöglichen die Adaptierung der Datenrate durch Weglassen der inter-prädizierten Frames, der sogenannten P- und B-Frames, oder die Änderung der Quantisierungsparameter, die aber einen Zugriff auf den Encoder oder eine Transcodierung erfordert. Unter Verwendung standardisierter Videocodierverfahren wird die Datenratenadaptierung heutzutage durch die Simulcast-Lösung realisiert. Darunter versteht man eine Verwendung von mehreren Bitströmen mit gleichem Videoinhalt aber unterschiedlichen Datenraten. Der Streamingserver wechselt zum entsprechenden Bitstrom, falls sich die verfügbare Bandbreite zum Empfänger verändert. Diese Lösung kann auch zur Bedienung von Empfängern mit Zugangskanälen unterschiedlicher Bandbreiten verwendet werden.

5.2 Skalierbares Videostreaming – Aktuelle Forschung am TNT

Der neueste Videostandard H.264/AVC zeigt eine neue Steigerung der Codierungseffizienz. Nachdem der Standard 2003 verabschiedet wurde, geht die Forschung in Richtung eines neuen skalierbaren Videostandards. Der Unterschied zwischen der skalierbaren Videocodierung und der nicht-skalierbaren ist, dass der skalierbare Bitstrom aus einem Base-Layer und mehreren Enhancement-Layern besteht. Je mehr Layer der Decoder aus einem skalierbaren Bitstrom verarbeitet, desto höher wird die Qualität des empfangenen Videosignals hinsichtlich Auflösung, Framerate und Quantisierungsartefakten. Im Januar 2005 entstand der erste Entwurf der skalierbaren Erweiterung von H.264/AVC [40]. Da die Skalierbarkeit die Datenratenadaptierung beim Streaming erheblich erleichtert, konzentriert das Institut für Informationsverarbeitung (TNT) die Forschungsaktivitäten im Bereich der Multimediakommunikation über das Internet auf skalierbares Videostreaming unter Anwendung der skalierbaren Erweiterung von H.264/AVC. Im ersten Schritt wird ein Congestion Control Algorithmus für das skalierbare Streamingsystem entwickelt. Darüber hinaus soll eine für Streaming geeignete Fehlerverschleierungsmethode für die skalierbare Erweiterung von H.264/AVC entwickelt werden, die einerseits eine möglichst geringe Interaktion mit der Videocodierung und somit einen geringen Rechenaufwand erfordert und andererseits die Gesamtdatenrate nur möglichst wenig erhöht. Dabei ist zu beachten, dass Paketverluste nicht nur auf Grund von Netzwerküberlastung, sondern auch durch Kanalstörung, wie bei der Funkverbindung entstehen können.

5.2.1 Skalierbare Erweiterung von H.264/AVC

Die skalierbare Erweiterung von H.264/AVC ermöglicht in einem GOP (Group of Pictures) die örtliche, zeitliche sowie Qualitäts-Skalierbarkeit. Die örtliche Skalierbarkeit wird durch eine Unterabtastung der Originalframes um einen Faktor 2 erreicht und erzeugt damit die so genannten Spatial-Layer. In jedem Spatial-Layer wird die hierarchische B-Frame oder MCTF (Motion Compensated Temporal

Filtering)-Methode für die zeitliche Skalierbarkeit verwendet, um mehr so genannte Temporal-Level zu gewinnen. Ein Temporal-Level enthält die Daten der Videosequenz mit einer geringeren zeitlichen Auflösung. Für jedes Frame eines Temporal-Levels kann eine Qualitätsskalierbarkeit zum Einsatz kommen, indem mehrere, immer feinere Quantisierungsstufen hintereinander auf die Transformationskoeffizienten der Prädiktionsfehler angewendet werden. Dabei entstehen die FGS-(Fine Grain Scalability) Layer. Der Vorteil dieser FGS-Methode ist, dass die FGS-Layer beliebig abgeschnitten werden können, ohne den Decodierungsprozess zu stören. Da zwischen den Spatial-Layern, Temporal-Levels sowie FGS-Layern eine Prädiktion verwendet wird, um die Codierungseffizienz zu erhöhen, ergeben sich Abhängigkeiten zwischen den Layern. In Bild 19 ist die Datenratenverteilung eines skalierbaren Bitstroms mit zwei Spatial-Layern 0 und 1, drei bzw. vier Temporal-Levels für jeden Spatial-Layer und drei FGS-Layern für jeden Temporal-Level. Der FGS-Baselayer des kleinsten Spatial-Layers und des kleinsten Temporal-Levels wird als Base-Layer bezeichnet, alle weiteren Layer als Enhancement-Layer. Gezeigt ist jeweils die Gesamtdatenrate, die sämtliche niedrigen Layer bzw. Level einschließen. Dieser Bitstrom bietet einen Datenratenbereich von 75kbit/s bis zu 970kbit/s.

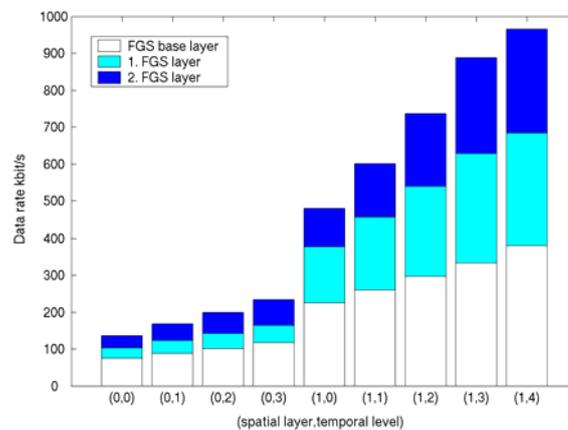


Bild 19 Datenratenverteilung eines skalierbaren Bitstroms mit den zwei Spatial-Layern 0 und 1, drei bzw. vier Temporal-Levels für jeden Spatial-Layer und drei FGS-Layern für jeden Temporal-Level. Gezeigt wird jeweils die Gesamtdatenrate, die sämtliche niedrigen Layer bzw. Level einschließen.

5.2.2 Congestion Control basierend auf einer Bandbreitenabschätzung

Das am TNT entwickelte Streamingsystem, unter Verwendung der skalierbaren Erweiterung von H.264/AVC mit der Congestion Control, wird in Bild 20 dargestellt. Anders als gewöhnliche Streamingsysteme ist unser System mit einem Bitstrom-Extractor und einem auf Bandbreitenabschätzung basierenden Congestion Control Algorithmus ausgerüstet. Der Bitstrom-Extractor gewinnt aus einem skalierbaren Bitstrom die Layer, die der von dem Congestion Control Algorithmus gewünschten Datenrate entsprechen. Die Bandbreite zwischen Server und Client wird mithilfe der RBPP (Receiver-based Packet Pair) Methode [41] abgeschätzt. Dabei werden die gesendeten RTP-Pakete für die Schätzung ausgenutzt. Die Schätzung erfolgt auf dem Client, und die geschätzte Bandbreite wird zusammen mit der Information des Pufferzustandes am Client über RTSP zum Server gesendet. Weiterhin werden die RTP-Pakete des Base-Layers bestätigt. Mit Informationen wie der geschätzten Bandbreite, der Bestätigung für die RTP-Pakete des Base-Layers und dem Pufferzustand des Clients berechnet der Congestion Control Algorithmus auf dem Server eine geeignete zu sendende Datenrate [42].

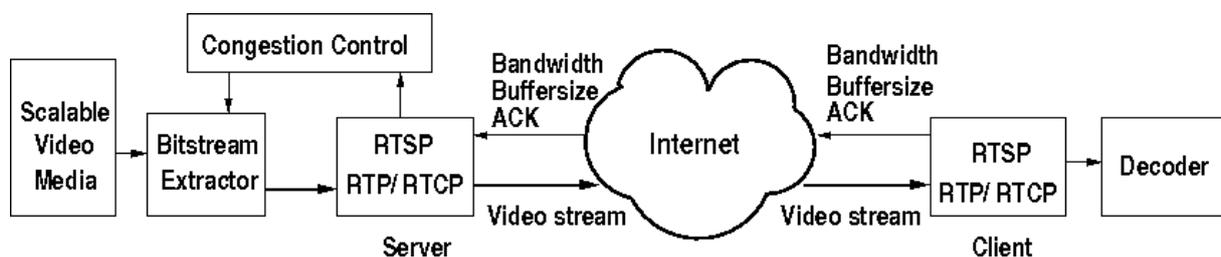


Bild 20 Skalierbares Videostreamingsystem mit dem auf Bandbreitenabschätzung basierendem Congestion Control Algorithmus.

Um die Performance des entwickelten Systems zu testen, wurde eine Simulationsumgebung wie in Bild 21 aufgebaut. Der Streamingserver befindet sich auf Host A und der Client auf Host C. Die Pakete zwischen A und C werden über Host B geroutet, auf dem die Simulationssoftware NS-2 [43] installiert ist. Mit Hilfe von NS-2 werden zwei Router R1 und R2 zwischen Host A und C simuliert. Die Kanalkapazität zwischen R1 und R2 ist 1500 kbit/s. In einem zweiminütigen Simulationsdurchlauf ist die UDP-Applikation von 2s bis 22s und die Telnet-Applikation von 22s bis 120s aktiv. Die Streamingsitzung benutzt die Software JSVM 3.0 [44] für die skalierbare Videocodierung.

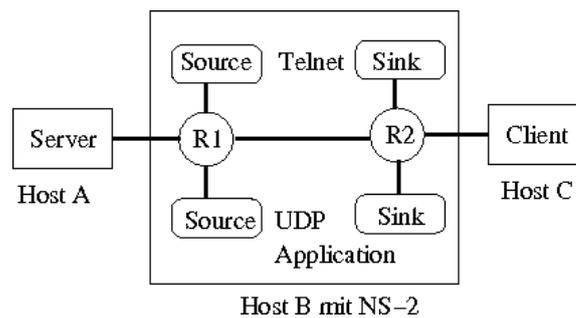


Bild 21 Simulationsumgebung für skalierbares Videostreaming .

Bild 22 zeigt den Vergleich zwischen dem Streamingsystem mit dem Binomialen Congestion Control Algorithmus SQRT [39] und dem auf Bandbreitenabschätzung basierendem Congestion Control Algorithmus in Hinsicht auf die gesendete Datenrate und die empfangenen Layer. Die Congestion Control mit der Bandbreitenabschätzung löst das slow-start Problem der binomialen Congestion Control und bewältigt die starke Bandbreitenschwankung durch die Konkurrenz der UDP- und Telnet-Applikation [42], [43]. Dabei wurde der größte Teil der starken Bandbreitenschwankungen durch FGS ausgeglichen, sodass die zeitliche oder örtliche Auflösung nur sehr selten gewechselt werden muss, wobei die höhere zeitliche Auflösung aufgrund der Bewegungsempfindlichkeit des menschlichen Auges bevorzugt wurde.

5.2.3 Fehlerverschleierung für skalierbare Videocodierung

Die skalierbare Erweiterung von H.264/AVC besitzt weiterhin die Struktur des H.264/AVC-Standards mit VCL (Video Coding Layer) und NAL (Network Abstraction Layer). Hierbei erzeugt VCL die codierten Daten und NAL bereitet sie für die Paketisierung der unteren Schicht im Protokollstapel vor, indem die codierten Daten aus dem VCL in die NAL-Units mit NAL-Header aufgeteilt werden. In der skalierbaren Erweiterung von H.264/AVC wird jeder FGS-Layer eines Frames in eine NAL-Unit gepackt. Die NAL-Units eines GOPs werden vom niedrigsten bis zum höchsten Temporal-Level angeordnet. Für jeden Temporal-Level werden die NAL-Units von dem niedrigsten bis zum höchsten Spatial-Layer gegliedert.

Aufgrund der Abhängigkeiten zwischen den Layern muss der Base-Layer immer vorhanden sein, damit die minimale Qualität wenigstens gezeigt werden kann und die Decodierung nicht unterbrochen wird. In einem GOP wird die NAL-Unit des Base-Layers mit einer FEC (Forward Error Correction) Methode geschützt. Die Bitstrom-Parameter der Enhancement-Layer werden analysiert, um Verluste von NAL-Units zu detektieren. Für die korrekt empfangenen NAL-Units wird auf den Ergebnissen dieser Analyse eine Neusortierung durchgeführt, um die bestmögliche örtliche und zeitliche Auflösung zu erzielen. Ergeben sich mehrere Kombinationsmöglichkeiten mit unterschiedlichen örtlichen und zeitlichen Auflösungen, so wird bei Videosignalen mit schnellen Bewegungen die höhere zeitliche Auflösung bevorzugt. Andernfalls wird die höhere örtliche Auflösung gewählt [45].

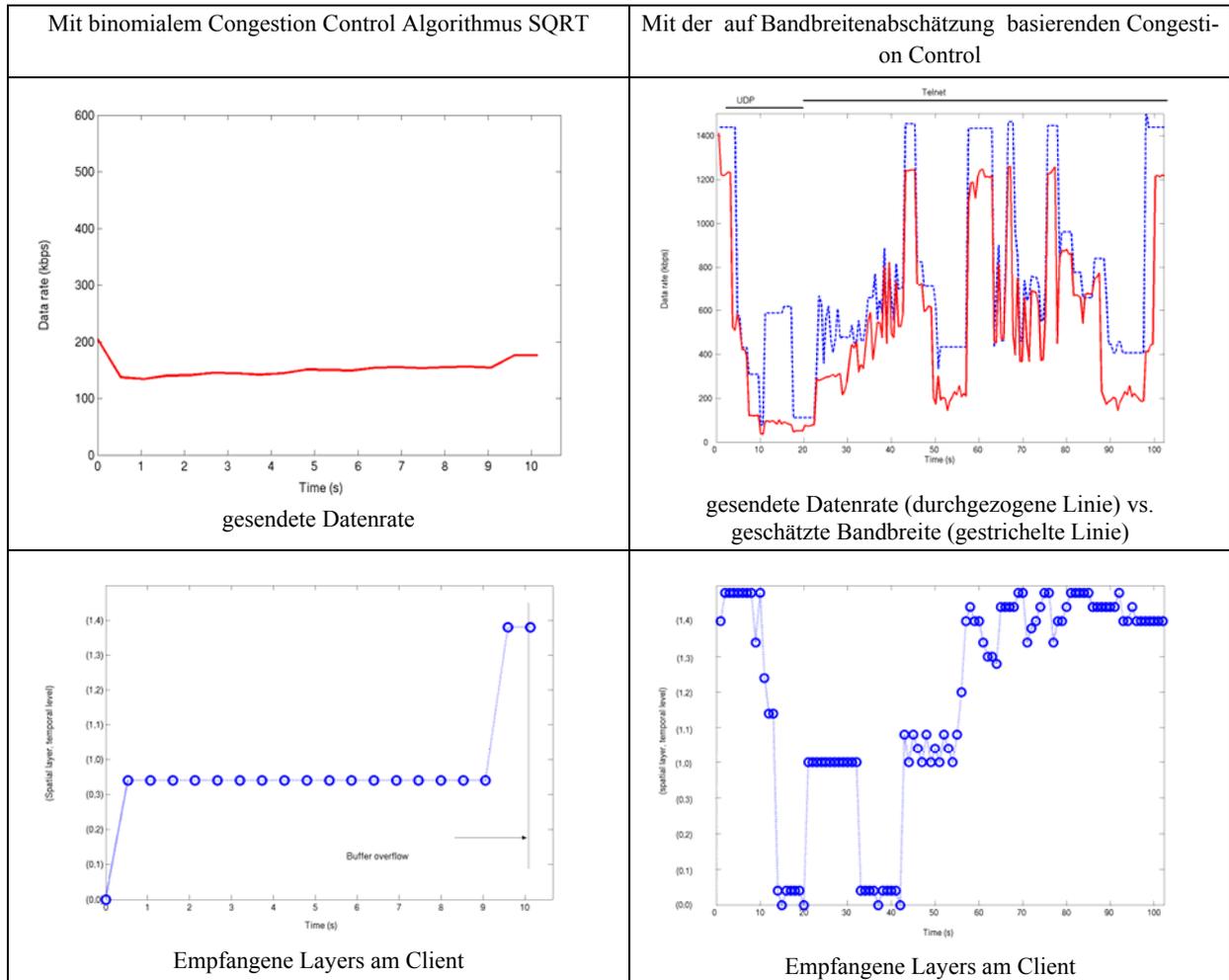


Bild 22 Vergleich zwischen binomialer Congestion Control SQRT und auf Bandbreitenabschätzung basierender Congestion Control in Hinsicht auf gesendete Datenrate und empfangene Layer.

6 Audiocodierung und -signalverarbeitung

Die meisten der heute weit verbreiteten Audiocodierverfahren basieren auf dem Prinzip der Transformations- bzw. Teilbandcodierung. Hierbei wird das Eingangssignal zunächst einer Spektralzerlegung mittels einer Transformation bzw. einer Analyse-Filterbank unterzogen. Anschließend werden die resultierenden Spektralkomponenten quantisiert und codiert. Bei der Decodierung werden die Spektralkomponenten durch eine Rücktransformation bzw. eine Synthesefilterbank wieder zu einem breitbandigen Audiosignal zusammengeführt. Gegenüber einer direkten Quantisierung und Codierung des Eingangssignals hat dieser Ansatz den Vorteil, dass durch eine geeignete Steuerung der Quantisiererstufenhöhen der zeitliche und spektrale Verlauf des resultierenden Quantisierungsfehlers gut an die Wahrnehmungseigenschaften des Gehörs angepasst werden kann. Einerseits sollen nämlich für eine maximale Datenratenreduktion die Quantisiererstufenhöhen so groß wie möglich gewählt werden, andererseits soll die Klangqualität des Audiosignals nicht oder nur so gering wie möglich verfälscht werden. Der maximale Gewinn bei unveränderter Klangqualität lässt sich also erzielen, wenn die Quantisierungsfehler noch gerade unterhalb ihrer Wahrnehmungsschwelle liegen. Diese Wahrnehmungsschwelle ist wegen der sogenannten Verdeckungswirkung (Maskierung) im Gehör abhängig vom Audiosignal [46]. So kann ein lauterer Ton einen größeren Quantisierungsfehler verdecken als ein leiserer Ton. Außerdem ist die Verdeckungswirkung eines einzelnen Tones auf Störsignale in seiner spektralen Nachbarschaft stärker als auf Störsignale in entfernter liegenden Frequenzbereichen. Für eine optimale Steuerung der Quantisiererstufenhöhen, die direkt die Größe der Quantisierungsfehler beeinflussen, ist also eine laufende Analyse des Eingangssignals hinsichtlich seiner momentanen Verdeckungswirkung erforderlich. Diese Aufgabe übernimmt in der Regel ein so genanntes Psychoakustisches Modell.

Aus dem Institut für Informationsverarbeitung und dem mit ihm eng kooperierenden Laboratorium für Informationstechnologie (LFI) kamen in diesem Bereich wesentliche organisatorische und wissenschaftliche Beiträge zur Entwicklung der MPEG Audiocodierungsstandards [47], [48], [49], [50], [51]. Von diesen werden heute unter anderem MPEG Audio Layer2 im digitalen Rundfunk (DAB, DVB) und Layer3 (mp3) für die digitale Speicherung eingesetzt [52]. Auch der Nachfolgestandard MPEG AAC [53], der zunehmend für den Musikvertrieb über das Internet eingesetzt wird, enthält Beiträge aus dem TNT/LFI [54].

Wegen der Wichtigkeit der Wahrnehmungseigenschaften des Gehörs wurde in der Folge ein in Software realisiertes Gehörmodell entwickelt, welches in hohem Maße neuere Erkenntnisse der Physiologie einbezieht [55].

6.1 Parametrische Audiocodierung: MPEG-4 HILN

Neuere MPEG-Standardisierungsaktivitäten hatten zum einen den mobilen Einsatz bei geringeren Datenraten als auch die Unterstützung von interaktiven Anwendungen zum Ziel. Der MPEG-4-Standard beschreibt daher eine objektbasierte Repräsentation eines Audiosignals. Ein Audioobjekt kann dabei das Signal eines einzelnen Instruments, ein Sprachsignal, ein komplettes Musikstück oder sogar Beschreibungen für synthetische Sprache bzw. Musik enthalten. Daher werden verschiedene Codierverfahren unterstützt, sodass das optimale Verfahren bei der Codierung ausgewählt werden kann. Mit dem HILN-Verfahren (Harmonic and Individual Lines plus Noise) wurde ein am TNT/LFI entwickeltes parametrisches Codierverfahren in den Standard aufgenommen, welches sich besonders für den Datenratenbereich von 6 bis 16 kbit/s eignet [56]. Hierbei wird das Signal im Encoder automatisch in harmonische Töne, einzelne Sinustöne und Rauschkomponenten zerlegt. Für diese Komponenten werden nun nicht die Signalformen direkt codiert, sondern lediglich Parameter, welche die wesentlichen Eigenschaften beschreiben. Dies sind für die Sinustöne beispielsweise Frequenzen und Amplituden.. Im Decoder werden mithilfe dieser parametrischen Beschreibung die einzelnen Komponentensignale synthetisiert und zu einem Ausgangssignal zusammengesetzt. Erstmals wurde im HILN-Encoder ein psychoakustisches Modell zur Steuerung eines parametrischen Codierverfahrens eingesetzt. Dieses dient insbesondere zur Angabe der Wichtigkeit einzelner Signalkomponenten für den Klangeindruck, sodass gegebenenfalls die unwichtigsten weggelassen werden können (Bild 23).

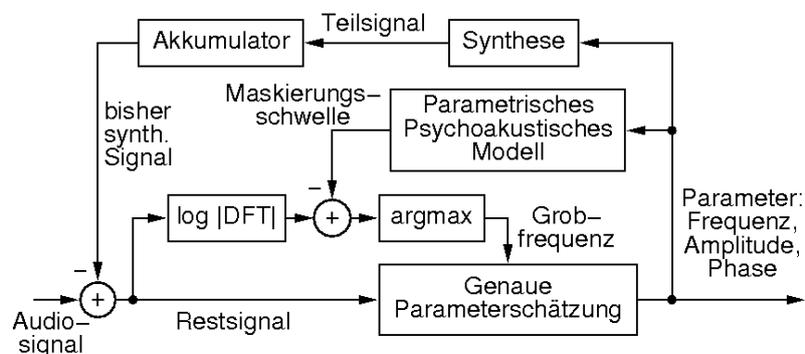


Bild 23 Analyse-Synthese mit psychoakustischem Modell in einem HILN-Encoder.

6.2 Kombination verschiedener Codierverfahren

Im Rahmen der MPEG-4 Standardisierung stellte sich heraus, dass noch kein Verfahren existierte, welches im unteren Datenratenbereich sowohl für Sprach- als auch für Musiksignale eine befriedigende Qualität lieferte. Daher wurde im Rahmen des EU-Projektes ARDOR (Adaptive Rate Distortion Optimized sound codeR) der Ansatz verfolgt, mehrere verschiedene Teilcoder derart zu kombinieren, dass sowohl das Eingangssignal als auch die zur Verfügung stehende Datenrate auf optimale Weise aufgeteilt wird [57]. Ein wesentlicher Beitrag des Instituts bestand hierbei neben der Implementierung und Optimierung einer transformationsbasierten Codierung in einer neuartigen Codierungstechnik für die Steuerdaten, die von einem psychoakustischen Modell generiert wurden. Wegen der unterschiedlichen Anforderungen der verschiedenen Teilcoder wurden Steuerdaten mit hoher Frequenz- und Zeitauflösung benötigt. Dabei kommt nun mit SPECK (Set Partitioning in Embedded bloCKs) ein Verfahren zum Einsatz, welches ursprünglich für die Bildcodierung entwickelt wurde. Im ARDOR-Codec

wird es auf, gemäß ihrer zeitlichen und spektralen Reihenfolge, zweidimensional angeordnete Erregungsmuster angewendet (Bild 24) [58]. Die Erregungsmuster wiederum können nun für die Steuerung verschiedener Teilcoder eingesetzt werden. Beispielsweise lassen sich aus ihnen Maskierungsschwellen für eine Transformationscodierung gewinnen oder es kann die Stärke eines synthetischen Rauschsignals zum Auffüllen von Signallücken berechnet werden [59].

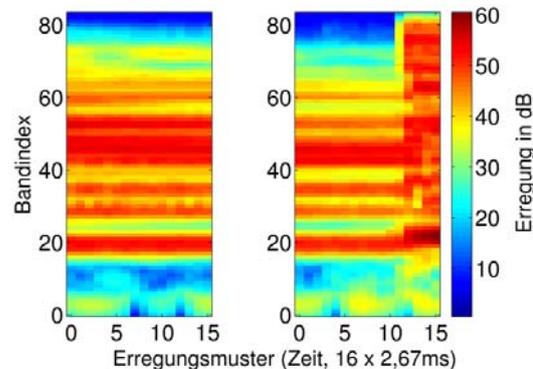


Bild 24 Zeit-Frequenzdarstellung der Erregungsmuster eines quasi-stationären Signalabschnitts (links) und eines Signalabschnitts mit einem Anschlag (rechts).

Weitere aktuelle Studien befassen sich mit der Optimierung einer Vektorquantisierung für Spektralkomponenten und der verlustlosen Codierung für Spektralkomponenten einer Transformationscodierung [60].

6.3 Filterbänke für parametrische Erweiterungen

In den neuesten Erweiterungen des MPEG-4-Audiocodierungsstandards wie z. B. MPEG-4 SBR [61] kommen Techniken zum Einsatz, bei denen Teile des Signals mit einer Transformationscodierung repräsentiert werden, die dann mit Hilfe zusätzlich übertragener Parameter erweitert werden. Hierzu zählt zum einen das SBR-Verfahren (Spectral Band Replication) und zum anderen die parametrische Stereo- bzw. Mehrkanalcodierung. Bei SBR wird nur ein relativ stark bandbegrenztetes Signal auf herkömmliche Weise codiert, aus dem anschließend durch Kopieren, Verschieben und Gewichten von Spektralkomponenten höherfrequente Signalanteile generiert werden (Bild 25) [61]. Bei der parametrischen Stereocodierung nach dem BCC-Prinzip (Binaural Cue Coding) wird nur ein Kanal, z. B. eine entsprechend gewichtete Summe der beiden Eingangskanäle übertragen, das dann mittels Parametern wie Pegeldifferenzen, Phasendifferenzen und Korrelationskoeffizienten wieder zu einem Stereosignal aufgeweitet wird. Entsprechend wird bei der parametrischen Mehrkanalcodierung verfahren, nur, dass hier auch mehr als ein Kanal mittels Transformationscodierung übertragen werden kann [63]. Diese Techniken beruhen im Wesentlichen auf einer komplexwertigen Teilband-Darstellung für das Audiosignal, welche eine um den Faktor 2 erhöhte effektive Abtastfrequenz für die Teilbandsignale aufweist. Gegenüber der Transformation bzw. Filterbank für die eigentliche Codierung ist hier nämlich keine kritische Abtastung erforderlich, da die komplexwertigen Spektralkomponenten nicht explizit übertragen werden. Eine kritische Abtastung bedeutet, dass im zeitlichen Mittel die Gesamtanzahl aller Spektralkomponenten gleich der Anzahl der Abtastwerte des Eingangssignals ist. Die komplexwertige Darstellung erhält man durch Hintereinanderschaltung einer reellwertigen Synthesefilterbank und einer komplexwertigen Analysefilterbank. Dieser relativ hohe Berechnungsaufwand lässt sich durch ein Verfahren reduzieren, welches Approximationswerte für die Imaginärteile der gewünschten komplexwertigen Darstellung berechnet. Diese können durch eine Linearkombination zeitlich und spektral benachbarter Realteile gewonnen werden [64]. Bild 26 zeigt die notwendigen Datenraten zur Codierung von Mehrkanal-Audiosignalen mit der häufig verwendeten 5.1-Kanal-Konfiguration.. Die parametrische Mehrkanalcodierung (Spatial Audio Coding) benötigt nur 64 kbit/s für gute Qualität.

Auch für Anwendungen, die eine Anhebung oder Absenkung unterschiedlicher Spektralkomponenten erfordern, z.B. für Echounterdrückung oder Equalizer-Funktionalitäten, hat die komplexwertige Spektraldarstellung deutliche Vorteile gegenüber der kritisch abgetasteten, reellen Darstellung. Letztere würde beispielsweise bei einer starken Dämpfung eines einzelnen Teilbandsignals aufgrund der endli-

chen Flankensteilheiten der Teilbandfilter unkompensierte Aliasingkomponenten aus Signalanteilen benachbarter Teilbänder erzeugen. Das oben beschriebene Approximationsverfahren hat bei diesen Anwendungen jedoch zwei wesentliche Nachteile. Zum einen entstehen systematische Fehler in dem aus der komplexwertigen Darstellung rekonstruierten Ausgangssignal, selbst wenn keine Modifikation an den Spektralkomponenten durchgeführt wird. Zum anderen bewirken die Filter, die zur Berücksichtigung der zeitlichen Nachbarschaften erforderlich sind, eine hohe Systemverzögerung. Aus diesem Grunde wurde ein Verfahren entwickelt, welches nur dazu dient, Aliasingkomponenten zu kompensieren, die durch Unterschiede zwischen Verstärkungsfaktoren benachbarter Spektralkomponenten entstehen. Im Falle gleicher Verstärkungsfaktoren bleiben die reellwertigen Spektralkomponenten unverändert erhalten, sodass die Rekonstruktionseigenschaften der ursprünglichen Filterbank erhalten bleiben. Aus diesem Grunde können nun auch Filter kürzerer Länge eingesetzt werden, da sich dadurch nur Auswirkungen auf die Aliasing-Reduktion ergeben. Bei einer hinreichenden Aliasing-Reduktion lässt sich mit diesem Verfahren die Systemverzögerung für die Signalmodifikationseinheit in etwa halbieren [65].

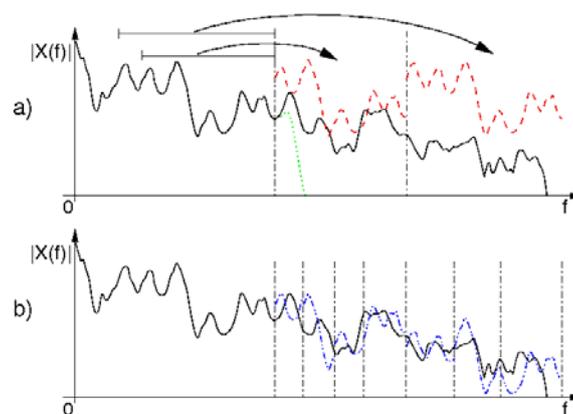


Bild 25 Kurzeitspektren bei Spectral Band Replication: Eingangssignal vor (—) und nach Bandbegrenzung (---), Ergebnis von Kopieren und Verschieben (- -) und Gewichten (- · · -).

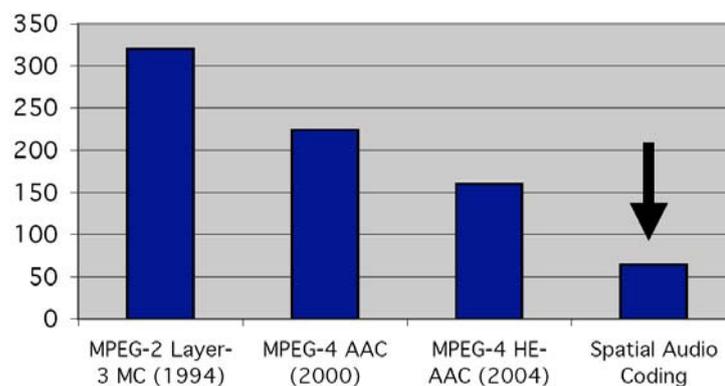


Bild 26 Datenraten für die Codierung von 5.1-Kanal-Audiosignalen bei „guter Qualität“ [66].

6.4 Signalverarbeitung für Cochlea-Implantate

Aus den Arbeiten in den Bereichen Psychoakustik und Physiologie des Gehörs entstanden intensive Kontakte zum Hörzentrum Hannover, einer Einrichtung der Medizinischen Hochschule Hannover. Gemeinsam wird nun an der Entwicklung verbesserter Signalverarbeitungsalgorithmen für Cochlea-Implantate gearbeitet. Diese dienen im Falle einer vollständigen Ertaubung zur Wiederherstellung eines gewissen Hörvermögens. Dazu werden betroffenen Patienten Elektroden in das Innenohr implantiert, wo bei einem gesunden Gehör die Umwandlung der Schallschwingungen in Nervensignale und zugleich eine spektrale Zerlegung des Schallsignals erfolgt. Eines der modernsten Implantate be-

sitzt beispielsweise 22 Elektroden, von denen jede einem bestimmten Frequenzbereich zugeordnet werden kann. Ein elektrischer Impuls auf einer dieser Elektroden kann nun einen Reiz hervorrufen, der einem kurzen Ton im betreffenden Frequenzbereich nahe kommt. Allerdings ergibt sich gegenüber dem gesunden Gehör nur eine sehr grobe Frequenzauflösung. Außerdem können die Elektroden nur sequentiell angesteuert werden, da sich ansonsten unerwünschte Interaktionen ergäben. Die Aufgabe eines üblicherweise nicht mit-implantierten Sprachprozessors ist es nun, das über ein Mikrofon aufgenommene Schallsignal zu analysieren und Anregungsmuster für die Elektroden zu erzeugen. Dabei wird das Eingangssignal zunächst in 22 Frequenzbänder aufgeteilt, die jeweils den Elektroden zugeordnet sind. Aus den über kurze Zeitabschnitte gemittelten Signalleistungen werden anschließend Impulsstärken gewonnen. Würde man nun jeweils alle 22 Elektroden sequentiell stimulieren, ergäbe sich eine zu geringe zeitliche Auflösung, da die Datenübertragungsrate zu den Elektroden begrenzt ist. Daher wurde bisher das so genannte ACE-Verfahren (Advanced Combinational Encoding) eingesetzt, welches aus den M (hier: $M=22$) Bändern eine Untermenge von N Bändern, jeweils mit der maximalen momentanen Signalleistung, auswählt. Übliche Werte für N liegen im Bereich von 4 bis 8. Dieses Verfahren hat allerdings den Nachteil, dass häufig Bänder ausgewählt werden, die zusammen nur einen schmalen Bereich des Hörspektrums abdecken. Ein neues Verfahren wurde entwickelt, bei dem die Auswahl der zu stimulierenden Elektroden von einem Psychoakustischen Modells derart gesteuert wird, dass die N für den Klangeindruck bzw. für die Sprachverständlichkeit wichtigsten Bänder gewählt werden [67]. Bei der Entwicklung dieses PACE (Perceptual Advanced Combinational Encoding) genannten Verfahrens waren die Erfahrungen mit dem oben beschriebenen HILN-Codierverfahren sehr hilfreich. Nachdem klinische Sprachverständlichkeitstests mit Patienten eine deutliche Erhöhung des Anteils richtig verstandener Sätze andeuteten (Bild 27), wurden mithilfe einer PACE-DSP-Implementierung so genannte chronische Studien durchgeführt. Bei diesen wird den Patienten eine ausreichende Zeit gegeben, um sich im Alltag jeweils an eines der zu bewertenden Verfahren zu gewöhnen. Abschließend konnten sie sogar kurzfristig selbst das ihnen angenehmste Verfahren auswählen. Nachdem auch diese Studie Vorteile für PACE aufzeigte, wird es mittlerweile schon in der Serienfertigung als Standardverfahren für Sprachprozessoren eingesetzt. Aktuelle Studien zielen auf eine Verbesserung der Sprechererkennung und der Tonhöhenwahrnehmung ab.

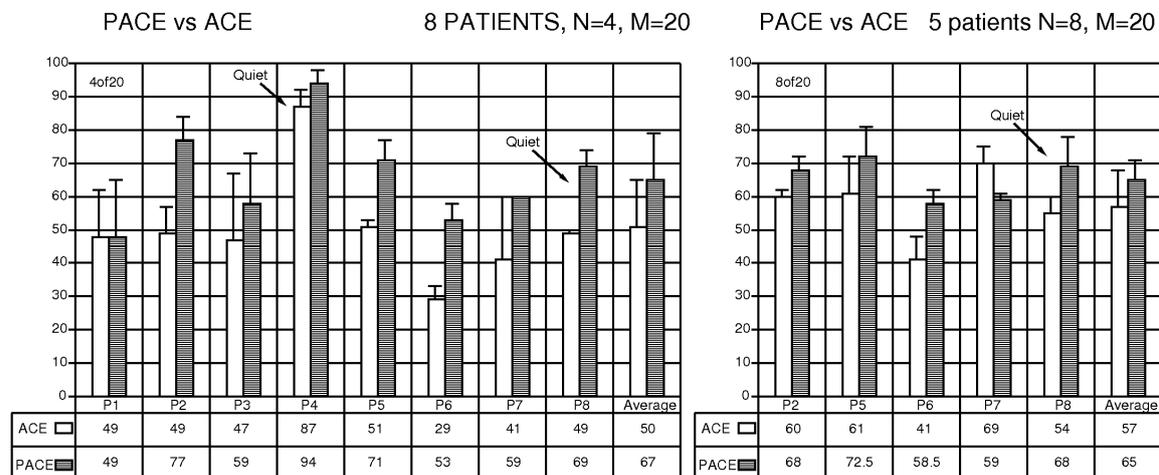


Bild 27 Ergebnisse von Sprachverständlichkeitstests von PACE im Vergleich zu ACE.

7 Zusammenfassung

Aufgrund der Fortschritte in der Multimediakommunikation können Übertragungskanäle immer effizienter genutzt werden. Daher stehen heute und vermehrt in der Zukunft zusätzliche Kanäle für die Übertragung von Fernsehprogrammen, zum Teil auch in HDTV, zur Verfügung. Gleichzeitig ermöglicht die effiziente Codierung der digitalen Daten und die Verbreitung des Internets mit Zugängen über DSL, GPRS, und UMTS auch eine kostengünstige Nutzung dieser Übertragungskanäle für die private und individuelle Multimediakommunikation.

Im Vergleich zum heute auf der DVD und im digitalen Fernsehen DVB-T eingesetzten Videocodierstandard MPEG-2 kann die Codiereffizienz um mehr als 50 % gesteigert werden. Fortschritte in der Bewegungskompensation und Prädiktionsfehlercodierung zeigen, dass bei dem neuesten Videocodierstandard MPEG-4 AVC/H.264 noch mindestens 12 % der Datenrate eingespart werden kann. Während die heute kommerziell eingesetzten Videoencoder eine deutlich höhere Rechenleistung als die zugehörigen Decoder erfordern, erlaubt das Konzept des Distributed Video Coding eine Verlagerung eines Großteils der Rechenleistung auf den Decoder. Dieses Konzept kann zum Beispiel in der Überwachungstechnik zu kostengünstigeren Systemen führen.

Mit der Einführung des Internets als das dominierende Übertragungsnetz auch für die Multimediakommunikation entstehen im Vergleich zu konventionellen Netzen mit konstanter Datenrate neue Aufgaben. Am Beispiel der skalierbaren Videocodierung, die die Extraktion verschiedener zeitlicher und örtlicher Auflösungen des Videosignals aus einem Datenstrom ermöglicht, wurde gezeigt, dass ein Videostreamingserver aufgebaut werden kann, der die robuste Übertragung des Videos auch bei sehr stark schwankenden zur Verfügung stehenden Datenraten erlaubt.

Audiocodierverfahren haben in den letzten Jahren zu weiteren Einsparungen bei der Datenrate geführt. Dabei ergeben sich auch Synergieeffekte mit der Bildcodierung, welche zum Einsatz des SPECK-Verfahrens bei der Codierung von Audioseiteninformationen führte. Aktuelle Forschungsarbeiten beschäftigen sich mit der Codierung von Mehrkanalaudiosignalen. Ein 5.1-Kanal-Audiosignal kann heute bei guter Qualität mit einer Rate von 64 kbit/s codiert werden. Zur Steuerung der Audiocoder werden psychoakustische Modelle eingesetzt. Der Einsatz dieser Modelle zur Verbesserung der Signalverarbeitungsalgorithmen für Cochlea-Implantate führt zu einer Erhöhung der Sprachverständlichkeit um durchschnittlich 17 %.

Neben den klassischen Multimediadaten Audio und Video werden in Zukunft auch bewegte dreidimensionale Modelle übertragen werden müssen. Angeregt durch die Videocodierung werden Verfahren der örtlichen und zeitlichen Prädiktion zur Vorhersage der Objektbewegungen genutzt. Dadurch kann die erforderliche Datenrate für die Codierung eines Stützpunktes des Netzes von 24 bit auf 5 bit reduziert werden.

Trotz der Effizienzsteigerungen bei der Codierung multimedialer Daten, der Erhöhung der Übertragungsbandbreiten und der Ausweitung der Speicherkapazitäten in den letzten Jahren sind weiterhin Fortschritte bei der Codiereffizienz erforderlich und zu erwarten. Diese werden dringend benötigt, um die Übertragung immer hochwertigerer Daten wie Mehrkanalaudio oder HDTV auch für interaktive Anwendungen zu ermöglichen. Die Thematik Codierung bleibt somit spannend!

Danksagung

Die Sequenz *Chicken Crossing* wurde von Andrew Glassner, Tom McClure, Scott Benza und Mark Van Langeveld erstellt. Diese kurze Sequenz aus Konnektivität und 3D-Positionen darf nur verwendet werden, um Kompressionsverfahren für Geometrie zu vergleichen. Das Frame dargestellt in Kapitel 4 stammt aus den Arbeiten [33], [34].

8 Literatur

- [1] T. Wedi, H. G. Musmann, „*Motion and Aliasing Compensated Prediction for Hybrid Video Coding*“, IEEE Transactions on Circuits and Systems for Video Technology, Volume 13, Number 7, Juli 2003.
- [2] Ulrich Reimers, „*Digital Video Broadcasting – The International Standard for Digital Television*“, Springer-Verlag, 2005.
- [3] ETSI EN 300 429: „*Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for cable systems*“, V1.2.1, 1998.
- [4] ETSI EN 300 421: „*Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for 11/12 GHz satellite services*“, V1.1.2, 1197.
- [5] 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.
- [6] MPEG-2: ISO/IEC JTC1/SC29/WG11 and ITU-T, „*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 and ITU-T, Genf, März 1995.
- [7] MPEG-4: ISO/IEC JTC1/SC29/WG11, „*ISO/IEC 14496:2000-2: Information on Technology - Coding of Audio-Visual Objects - Part 2: Visual*“, ISO/IEC, Genf, Dezember 2000.
- [8] H.261: International Telecommunication Union, „*Recommendation ITU-T H.261: Video codec for audiovisual services at p x 64 kbit/s*“, International Telecommunication Union, Genf, 1993.
- [9] H.263: International Telecommunication Union, „*Recommendation ITU-T H.263: ideo coding for low bit rate communication*“, International Telecommunication Union, Genf, 1998.

- [10] Joint Video Team of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), „*Advanced Video Coding (AVC) - 3rd Edition*”, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10), Juli 2004.
- [11] 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*”, International Telecommunication Union, Genf, 1995.
- [12] ITU-R 709: International Telecommunication Union, „*Recommendation ITU-R BT.709-3: Parameter values for the hdtv standards for production and international programme exchange*”, International Telecommunication Union, Genf, 1998.
- [13] D. Marpe, H. Schwarz und T. Wiegand, „*Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard*”, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 13, No. 7, Seiten. 620–636, Juli 2003.
- [14] D. A. Huffman, „*A Method for Construction of Minimum-Redundancy Codes*”, in Proc. IRE, 1952, No 40, Seiten 1098–1101.
- [15] J. Ostermann, M. Narroschke, „*Motion compensated prediction with 1/8-pel displacement vector resolution*”, ITU-T Q.6/SG16, doc. VCEG-AD09, Hangzhou, China, Oktober, 2006
- [16] Y. Vatis, B. Edler, D. T. Nguyen, J. Ostermann, „*Motion-and Aliasing-compensated Prediction using a two-dimensional non-separable Adaptive Wiener Interpolation Filter*”, Proc. ICIP 2005, IEEE International Conference on Image Processing, Genova, Italien, September 2005
- [17] Y. Vatis, B. Edler, I. Wassermann, D. T. Nguyen, Jörn Ostermann, „*Coding of Coefficients of two-dimensional non-separable Adaptive Wiener Interpolation Filter*”, Proc. VCIP 2005, SPIE Visual Communication & Image Processing, Beijing, China, Juli 2005
- [18] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, G.J. Sullivan, „*Rate-Constrained Coder Control and Comparison of Video Coding Standards*”, IEEE Transactions on Circuits and Systems for Video Technology, Volume 13, Number 7, Juli 2003
- [19] M. Narroschke, Y. Vatis, „*Increasing the coding efficiency by adaptive interpolation filters, adaptive prediction error coding, and 1/8-pel displacement vector resolution*”, ISO/IEC JTC1/SC29/WG11 Doc. M13137, Montreux, CH, April 2006
- [20] Key Technical Area Software of the ITU-T, Version jm10.2aif1.0, SG16, Q6: available at <http://www.tnt.uni-hannover.de/~vatis/hta>, 2005
- [21] M. Narroschke, „*Extending H.264/AVC by an adaptive coding of the prediction error*”, Proceedings of the Picture Coding Symposium 2006, Beijing, China, April 2006
- [22] M. Narroschke, „*Adaptive prediction error coding in spatial and frequency domain for H.264/AVC*”, ITU-T Q.6/SG16, doc. VCEG-AB06, Bangkok, Thailand, Januar, 2006
- [23] M. Narroschke, „*Adaptive prediction error coding in spatial and frequency domain with a fixed scan in the spatial domain*”, ITU-T Q.6/SG16, doc. VCEG-AD07, Hangzhou, China, Oktober, 2006
- [24] T. Wedi, M. Narroschke, „*Der H.264/AVC Videocodierstandard*”, Telekommunikation Aktuell, Berichte aus Forschung und Entwicklung in Informationstechnik und Telekommunikation, 58. Jahrgang, Heft 07-08, Juli-August 2004
- [25] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, „*Video coding with H.264/AVC: Tools, Performance, and Complexity*”, IEEE Circuits and Systems Magazine, Volume 4, Number 1, First Quarter 2004
- [26] J. Slepian, J. Wolf, „*Noiseless coding of correlated information sources*”, IEEE Trans. on Inform. Theory, Vol. 19(4), pp. 471-480, 1973
- [27] A. Wyner, J. Ziv, „*The rate-distortion function for source coding with side information at the decoder*”, IEEE Trans. on Inform. Theory, Vol. 22(1), pp. 1-10, 1976
- [28] DISCOVER (Distributed Coding for Video Services), an European Commission Future and Emerging Technologies (FET) project, www.discoverdvc.org
- [29] C. Berrou, A. Glavieux, P. Thitimajshima, „*Near Shannon limit error-correcting coding and decoding: Turbo Codes*”, IEEE International Conference on Communications, Vol. 2, pp. 1064-1070, Geneva, Switzerland, May 1993
- [30] J. Ascenso, C. Brites, F. Pereira, „*Content Adaptive Wyner-Ziv Video Coding Driven by Motion Activity*”, IEEE International Conference on Image Processing, Atlanta, USA, Oktober 8-11, 2006
- [31] S. Klomp, Y. Vatis, J. Ostermann, „*Side Information Interpolation with Sub-pel Motion Compensation for Wyner-Ziv Decoder*”, International Conference on Signal Processing and Multimedia Applications (SIGMAP), Setúbal, Portugal, August 7-10, 2006, pp 178-182
- [32] N. Stefanoski, J. Ostermann, „*Connectivity-Guided Predictive Compression of Dynamic 3D Meshes*”, Proc. ICIP 2006, IEEE International Conference on Image Processing, Atlanta, USA, October 2006
- [33] N. Anuar, I. Guskov, „*Extracting Animated Meshes with Adaptive Motion Estimation*”, Proceedings of the 9th International Fall Workshop on Vision, Modeling, and Visualization, November 2004
- [34] P. Sand, L. McMillan, J. Popovic, „*Continuous Capture of Skin Deformation*”, ACM Transactions on Graphics. 22(3), pp. 578-586, 2003
- [35] S. Gumhold and W. Strasser, „*Real time compression of triangle mesh connectivity*”, SIGGRAPH '98, ACM Press, pp. 133-140, 1998

- [36] P. Alliez and M. Desbrun, „*Valence-driven connectivity encoding for 3d meshes*”, Comput. Graph. Forum, vol. 20, no. 3, 2001
- [37] M., R. M. Neal, I.H. Witten, „*Arithmetic coding revisited*”, ACM Trans. Inf. Syst., vol. 16, no. 3, pp. 256-294, 1998
- [38] K. Sayood, „*Introduction to Data Compression*”, Morgan Kaufmann, 1996
- [39] D. Bansal and H. Balakrishnan, „*Binomial Congestion Control*”, IEEE INFOCOM, April 2001.
- [40] J. Reichel, H. Schwarz and M. Wien, „*Scalable Video Coding - Working Draft I*”, Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Doc. JVT-N020, January 2005.
- [41] V. Paxson, „*Measurements and Analysis of End-to-End Internet Dynamics*”, PhD Dissertation, Computer Science Department, University of California at Berkeley, 1997.
- [42] D.T. Nguyen and J. Ostermann, „*Streaming and Congestion Control using Scalable Video Coding based on H.264/AVC*”, 15th International Packet Video Workshop, Hangzhou, China, April 2006.
- [43] Network Simulator ns-2. <http://www.isi.edu/nsnam/ns/index.html>
- [44] J. Reichel, H. Schwarz and M. Wien, „*Joint Scalable Video Model JSVM-3*”, Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Doc. JVT-P202, July 2005.
- [45] D.T. Nguyen, M. Shaltev and J. Ostermann, „*Error Concealment in the Network Abstraction Layer for the Scalability Extension of H.264/AVC*” 1st International Conference on Communications and Electronics, Hanoi, Vietnam, October 2006.
- [46] E. Zwicker, „*Psychoakustik*”, Springer-Verlag, Berlin-Heidelberg-New York, 1982
- [47] H.G. Musmann, „*The ISO audio coding standard*”, IEEE GLOBECOM, San Diego, pp. 511-517, Dezember 1990
- [48] B. Edler, „*Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen*”, Frequenz, Band 43, Nr. 9, pp. 252-256, September 1989
- [49] B. Edler, „*Aliasing Reduction in Subbands of Cascaded Filter Banks with Decimation*”, Electronics Letters, Vol. 28, No. 12, pp. 1104-1105, Juni 1992
- [50] H. Fuchs, „*Die Hörtests im Rahmen der ISO/MPEG-Audio Standardisierung*”, 17. Tonmeistertagung, Karlsruhe, November 1992
- [51] H.G. Musmann, „*Genesis of the MP3 audio coding standard*”, IEEE Transactions on Consumer Electronics, Vol. 52, No. 3, August 2006
- [52] ISO/IEC JTC1/SC29/WG11, „*IS11172-3: Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - Part 3: Audio*”, Genf, 1993
- [53] ISO/IEC JTC1/SC29/WG11, „*IS13818-7: Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC)*”, Genf, 1997
- [54] H. Fuchs, „*Improving MPEG Audio Coding by Backward Adaptive Linear Stereo Prediction*”, Preprint 4086, 99th AES Convention, New York, September 1995
- [55] F. Baumgarte, „*Ein psychophysiologisches Gehoermodell zur Nachbildung von Wahrnehmungsschwellen fuer die Audiocodierung*”, Dissertation, Universität Hannover, Februar 2000
- [56] H. Purnhagen, B. Edler, „*Objektbasierter Analyse/Synthese Audio Coder für sehr niedrige Datenraten*”, ITG-Fachtagung Codierung für Quelle, Kanal und Übertragung, Aachen, März 1998
- [57] ARDOR „*Adaptive Rate-Distortion Optimised sound coder*”, EU Project IST-2001-34095, <http://www.hitech-projects.com/euprojects/ardor/index.htm>
- [58] O. Niemeyer, B. Edler, „*Efficient Coding of Excitation Patterns Combined with a Transform Audio Coder*”, Preprint 6466, 118th AES Convention, Barcelona, Mai 2005
- [59] S. v.d. Par, V. Kot, N. v. Schijndel, „*Scalable Noise Coder for Parametric Sound Coding*”, Preprint 6465, 118th AES Convention, Barcelona, Mai 2005
- [60] N. Meine, B. Edler, „*Improved Quantization and Lossless Coding for Subband Audio Coding*”, Preprint 6468, 118th AES Convention, Barcelona, Mai 2005
- [61] ISO/IEC JTC1/SC29/WG11, „*IS14496-3:2005: Information technology - Generic coding of moving pictures and associated audio information - Part 3: Audio (3rd edition)*”, Genf, 2005
- [62] M. Dietz, L. Liljeryd, K. Kjørling, O. Kunz, „*Spectral Band Replication, a Novel Approach in Audio Coding*”, Preprint 5553, 112th AES Convention, München, April 2002
- [63] C. Faller, F. Baumgarte, „*Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression*”, Preprint 5574, 112th AES Convention, München, April 2002
- [64] B. Edler, S. Geyersberger, „*Arrangement and method for the generation of a complex spectral representation of a time discrete signal*”, Int. Patentanmeldung WO 2004/013839, Juli 2002
- [65] B. Edler, „*Aliasing Reduction for Gain Control with Critically Sampled Filter Banks*”, First International Conference on Communications and Electronics (ICCE'06), Hanoi, Vietnam, Oktober 2006
- [66] S. Geyersberger J. Herre P. Kroon C. Faller, „*Spatial Audio Coding – An Enabling Technology for Bitrate-Efficient and Compatible Multi-channel Audio Broadcasting*”, 7th Workshop on Digital Broadcasting, Erlangen, Deutschland, September 2006
- [67] W. Nogueira, A. Buechner, Th. Lenharz, B. Edler, „*A Psychoacoustic NofM type speech coding strategy for cochlear implants*”, Eurasip Journal on Applied Signal Processing - Special issue on DSP in Hearing Aids and Cochlear Implants, Vol. 18, pp. 3044-3059, November 2005

9 Verfasserportraits



Bernd Edler studierte von 1979 bis 1985 Elektrotechnik an der Universität Erlangen-Nürnberg. Von 1985 bis 1986 war er in der christlichen Studentenarbeit „Campus für Christus“ tätig. Anschließend war er bis 1993 wissenschaftlicher Mitarbeiter am Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung der Leibniz Universität Hannover. Seit 1993 ist er für den Bereich Systemtechnik des Laboratoriums für Informationstechnologie der Leibniz Universität Hannover verantwortlich. 1994 promovierte er mit dem Thema „Äquivalenz von Transformation und Teilbandzerlegung in der Quellencodierung“. Von 1998 bis 1999 war er Gastwissenschaftler bei den Bell Laboratories von Lucent Technologies in Murray Hill, USA.

Sven Klomp studierte Elektrotechnik mit dem Schwerpunkt Nachrichtenverarbeitung an der Leibniz Universität Hannover. Ende 2004 schloss er das Studium zum Dipl.-Ing. mit Auszeichnung ab. Seit dem arbeitet er am Institut für Informationsverarbeitung an der Leibniz Universität Hannover als wissenschaftlicher Mitarbeiter an seiner Promotion. Seine Interessen liegen in der Bildverarbeitung und der Videocodierung mit Schwerpunkt auf Distributed Video Coding.



Dipl.-Ing. Marco Munderloh, Jahrgang 1977, studierte von 1999 bis 2004 an der Technischen Universität Ilmenau den Studiengang Ingenieurinformatik. Nach seiner Arbeit am Fraunhofer Institut für Digitale Medientechnologie (IDMT) im Bereich Wellenfeldsynthese wechselte er 2005 als wissenschaftlicher Mitarbeiter an das Institut für Informationsverarbeitung der Leibniz Universität Hannover und beschäftigt sich mit Multimedia Daten-Links auf Basis von OFDM.

Matthias Narroschke, Jahrgang 1974, erlangte 2001 das Diplom in Elektrotechnik an der Universität Hannover mit Auszeichnung. Derzeit arbeitet er als wissenschaftlicher Mitarbeiter am Institut für Informationsverarbeitung an der Leibniz Universität Hannover mit dem Ziel der Promotion. 2003 wurde er „Ober-Ingenieur“. Seine Hauptarbeitsgebiete sind die Videocodierung, die Bildverarbeitung und das Internet-Streaming. Für das beste Diplom im Bereich der Elektrotechnik bekam er 2001 den Robert-Bosch-Preis verliehen. Herr Narroschke ist aktiver Beitragender der Motion Picture Experts Group (MPEG) der ISO/IEC SC29.



Dieu Thanh Nguyen studierte Elektrotechnik mit der Studienrichtung Nachrichtentechnik an der Leibniz Universität Hannover. Ende 2003 schloss sie das Studium zum Dipl.-Ing. ab. Seit Januar 2004 arbeitet sie als wissenschaftliche Mitarbeiterin an dem Institut für Informationsverarbeitung. Ihre Forschungsinteressen sind Videocodierung und Übertragung der codierten Videodaten über IP-Netzwerke. Sie war Stipendiat der Friedrich-Ebert-Stiftung und erhielt 2001 die Auszeichnung für hervorragende Studienleistungen und soziales Engagement von der Leibniz Universität Hannover.

Jörn Ostermann schloss das Studium der Elektrotechnik 1988 ab und promovierte 1994 an der Universität Hannover. Von 1993 bis 1994 arbeitete er bei AT&T Bell Labs und anschließend bis 2003 bei AT&T Labs-Research. Seit 2003 leitet er das Institut für Informationsverarbeitung an der Leibniz Universität Hannover. Er ist ein Fellow des IEEE, ein ehemaliger Distinguished Lecturer der IEEE CAS Society, Empfänger eines ISO Awards und des AT&T Standards Recognition Awards. Er ist Coautor eines Lehrbuchs zum Thema Video Communications. Er veröffentlichte mehr als 100 Artikel und 22 Patente. Seine Interessen und Arbeitsschwerpunkte liegen auf dem Gebiet der Audio- und Videosignalverarbeitung sowie der Kommunikation über IP-Netzwerke.



Nikolce Stefanoski studierte Mathematik mit Studienrichtung Informatik und Anwendungsfach Nachrichtentechnik an der Leibniz Universität Hannover. Er schloss das Studium zum Dipl.-Math. im Sommer 2003 ab. Der Fachbereich für Informatik zeichnete ihn für seine Diplomarbeit aus dem Bereich der stochastischen Algorithmenanalyse aus. Seit dem promoviert er am Institut für Informationsverarbeitung an der Leibniz Universität Hannover. Seine Interessen und Arbeitsschwerpunkte liegen in der Verarbeitung und Codierung von Geometriedaten, sowie in der Stochastik.

Yuri Vatis erlangte 2003 das Diplom in Technischer Informatik an der Technischen Universität Berlin. Seitdem arbeitet er als wissenschaftlicher Mitarbeiter am Institut für Informationsverarbeitung an der Leibniz Universität Hannover mit dem Ziel der Promotion. Seine Hauptarbeitsgebiete sind konventionelle Videocodierung sowie Distributed Video Coding. Yuri Vatis arbeitet aktiv in den Standardisierungsgruppen JVT, MPEG und VCEG der ISO/IEC und ITU-T mit. Seine Arbeiten bezüglich der adaptiven Interpolationsfilterung sind in die derzeitige Exploration Software von VCEG aufgenommen.

