



Vier Jahrzehnte Videocodierung

Four Decades Video Coding

Hans Georg Musmann, Universität Hannover,
Reiner Lippmann, Deutsches Zentrum für Luft- und Raumfahrt e. V., Braunschweig

Zusammenfassung Es wird die Entwicklung der Videocodierung in den vergangenen 40 Jahren anhand der wesentlichen Meilensteine dargestellt. Das Ergebnis dieser Entwicklung führte zu den ersten weltweit einheitlichen Codierungsstandards für digitale Videosignale, die Voraussetzung zur Einführung des digitalen Fernsehens, der DVD oder auch der digitalen Multimedia-Kommunikation über das Internet waren. Sie ermöglichen es beispielsweise, die PCM-Datenrate eines Videosignals im ITU 601-Format von 166 Mbit/s auf etwa 4 Mbit/s zu reduzieren. Abschließend werden die zur Standardisierung vorgeschlagenen neuen Codierungsverfahren erläutert, mit de-

nen die Übertragungsrate sogar bis auf 2 Mbit/s und darunter verringert werden kann. ▶▶▶ **Summary** The development of video coding over 40 years is reviewed along its main mile stones. It resulted in the first worldwide standards for digital video coding. The developed standards enabled the introduction of digital television, of the DVD, and of multimedia communications via the internet. By digital video coding the data rate of a PCM video signal in ITU 601-format can be reduced from 166 Mbit/s to about 4 Mbit/s. Finally new coding tools going to be standardized are explained which allow a further reduction of the data rate down to 2 Mbit/s and below.

KEYWORDS I.4.2 [Image Processing] *Compression, Video Coding*

1 Einleitung

Die ersten Veröffentlichungen über digitale Videocodierung sind in den 50er Jahren erschienen. Die erste Konferenz, die sich speziell diesem Thema widmete, wurde im April 1969 von Professor Thomas Huang am MIT in Cambridge, USA, veranstaltet. Aus dieser Konferenz ging eine bis in die Gegenwart bestehende Folge von inzwischen 23 so genannten Picture Coding Symposia hervor, die zeigt, wie aktuell dieses Thema noch heute ist. Aus den Beiträgen dieser Symposia ist die geschichtliche Entwicklung gut zu verfolgen.

Eine Datenreduktion von digitalen Videosignalen kann sowohl mit irrelevanzreduzierenden als auch mit redundanzreduzierenden Verfahren erreicht werden. Die Irrelevanzreduktion nutzt Wahrnehmungseigenschaften des Auges aus, indem sie zur Datenreduktion beispielsweise die Quantisie-

rungsstufen der Abtastwerte des Luminanz- oder Chrominanzsignals gerade so weit vergrößert, dass die Quantisierungsfehler noch unter der Sichtbarkeitsschwelle bleiben. Die Redundanzreduktion nutzt unterschiedliche Wahrscheinlichkeiten und statistische Abhängigkeiten der Abtastwerte zur Datenreduktion aus, indem beispielsweise häufig auftretende Abtastwerte mit kurzen und selten auftretende Abtastwerte mit langen Codierwörtern codiert werden.

2 Meilensteine der Vergangenheit

2.1 Intraframe-Codierung

Erste Berichte über Untersuchungen zur digitalen Codierung von Videosignalen wurden aus den USA bekannt, speziell von den Bell Laboratories, Murray Hill und dem MIT, Cambridge. Diese frühen Codierungsverfahren nutzen die Redundanz und Irrelevanz eines einzelnen

Bildes zur Datenreduktion aus. Sie werden daher als Intraframe-Codierung bezeichnet.

Kretzmer [1] wie auch Schreiber [2] teilen das Bildsignal mit Hilfe eines Tiefpassfilters in tieffrequente und hochfrequente Signalanteile und quantisieren die hochfrequenten Signalanteile in Anpassung an die Kontrastempfindlichkeit des Auges gröber als die tieffrequenten Anteile, um die Datenrate zu reduzieren, vgl. Bild 1. Je geringer die Kontrastempfindlichkeit ist, umso gröber kann die Quantisierung sein.

Schon 1952 führte Harrison [3] Experimente durch, um mit Prädiktion eine effizientere Übertragung von Fernsehsignalen zu erreichen. Der von ihm verwendete analoge lineare Prädiktor ist aus Verzögerungsleitungen und Verstärkerröhren aufgebaut.

Graham [4] griff 1958 den auf Cutler [5] zurückgehenden Vor-

Bild 1 Kontrastempfindlichkeit des Auges als Funktion der örtlichen Frequenz nach Robson [19]. Parameter ist die zeitliche Frequenz.

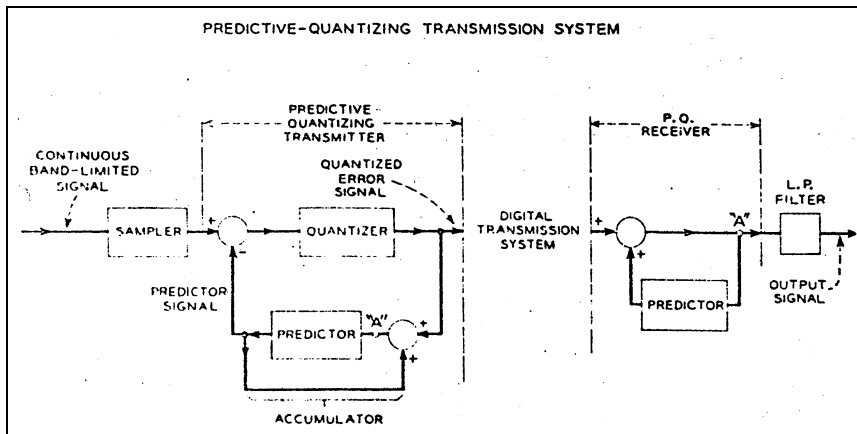
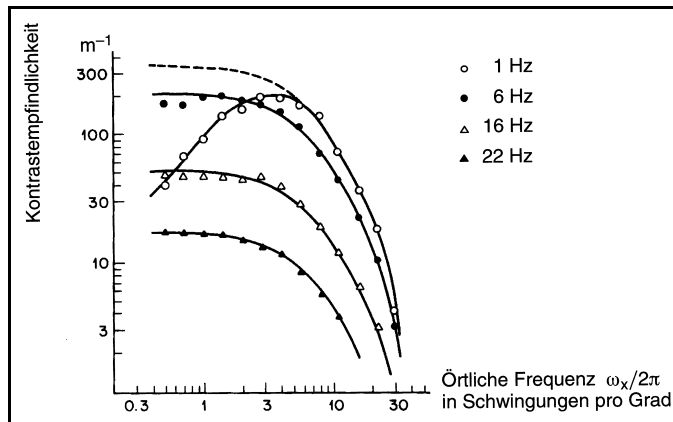


Bild 2 Blockschaltbild eines DPCM-Systems nach Graham [4].

schlag des „differential quantization“ auf und wendete diesen erstmals zur digitalen Prädiktiven Codierung von Videosignalen an. Er bezeichnete seinen Vorschlag als „Predictive-quantizing transmission system tailored to the observer's perception“ (Bild 2). Es handelt sich um eine differentielle Pulsmodulation (DPCM) mit einem Quantierer für den Prädiktionsfehler, der an die Wahrnehmungseigenschaften

des Auges und dabei speziell an die durch Helligkeitssprünge verursachte Maskierung angepasst ist (siehe Bild 3). An Helligkeitssprüngen auftretende große Prädiktionsfehler werden größer als kleine Prädiktionsfehler quantisiert.

Unter der Leitung von C.C. Cutler wurde um 1964 an den Bell Laboratories, Holmdel, eine Arbeitsgruppe eingerichtet, die später von J.O. Limb [6] geleitet wurde und

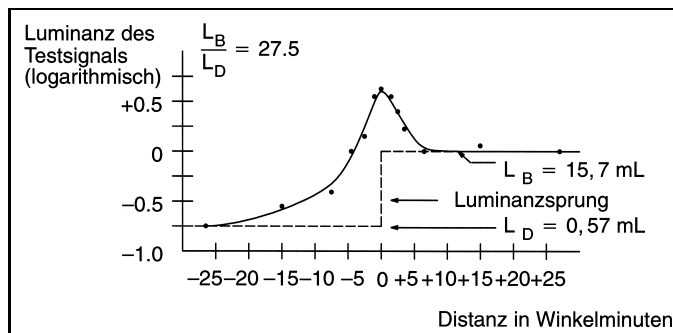


Bild 3 Sichtbarkeitsschwelle eines Testsignals als Funktion der Distanz zu einem Luminanzsprung nach Fiorentini [20].

sich zu jener Zeit zur führenden Forschungsinstitution auf dem Gebiet der digitalen Videocodierung entwickelte. Der Schwerpunkt der Untersuchungen lag auf der prädiktiven Codierung, speziell auf der Optimierung der DPCM.

O'Neal [7] veröffentlichte 1966 eine mathematische Beschreibung der DPCM und berechnete den Codierungsgewinn einer DPCM mit eindimensionalem Prädiktor im Vergleich zur PCM. Dieser Gewinn beträgt 12 dB, gemessen mit dem Signalrauschverhältnis, das durch den Quantisierungsfehler bestimmt wird. Es folgten Veröffentlichungen auch von andernorts mit Vorschlägen zur Verbesserung der Prädiktion und Quantisierung der DPCM durch Anwendung adaptiver Techniken, z.B. Connor [8], Bostelmann [9] oder durch Anpassung der Quantisierungskennlinien an die Maskierungseigenschaften des Auges, z.B. Thoma [10] sowie Sharma und Netravali [11].

Um etwa 1970 begann die prädiktive Codierung von Farbfernsehsignalen. Es gab zwei Ansätze. Im Fall der so genannten Composite-Codierung wird nach Vorschlägen von Devereux [12] oder Thomson [13] das Luminanzsignal mit den modulierten Chrominanzsignalen als ein Signal codiert. Die so genannte Component-Codierung codiert das Luminanzsignal Y und die Chrominanzsignale R-Y, B-Y als drei getrennte Signale. Da die Component-Codierung von der unterschiedlichen Modulation der Chrominanzsignale bei NTSC, PAL und SECAM unabhängig ist, ermöglichte sie die Entwicklung eines weltweit einheitlichen digitalen Codierungsstandards und setzte sich damit auch durch.

Neben der prädiktiven Codierung wurde bereits damals auch die Anwendung einer zweidimensionalen Transformations-Codierung vorgeschlagen.

Erste Veröffentlichungen gehen auf Enomoto und Shibata [14], Andrews und Pratt [15], Habibi und Wintz [16] sowie Woods und

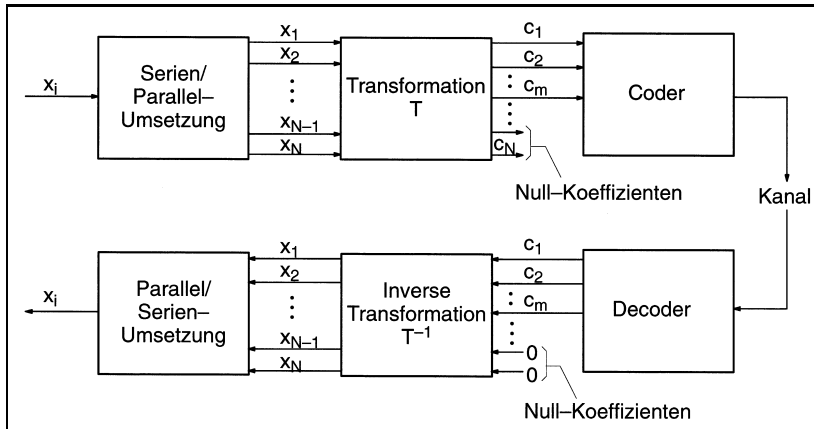


Bild 4 Blockschaltbild einer eindimensionalen Transformations-Codierung.

Huang [17] zurück. In diesen Arbeiten werden die Walsh-Hadamard-Transformation, die theoretisch optimale Karhunen-Loeve-Transformation (KLT) sowie die Diskrete Fourier-Transformation (DFT) untersucht. Bild 4 zeigt das Blockschaltbild einer eindimensionalen Transformations-Codierung. Zur Videocodierung werden zweidimensionale Transformationen auf Blöcke von $N \times N$ Bildpunkten angewendet. Nach der blockweisen Transformation werden die $N \times N$ Koeffizienten quantisiert und codiert.

Erst 1974 wird von Ahmed, Natarajan und Rao [18] die Diskrete Cosinus-Transformation (DCT) vorgeschlagen. Da die Koef-

fizienten einer Transformation ein periodisches Signal darstellen, können Signalsprünge an den Blockgrenzen auftreten, die im Fall der DCT vermieden werden, wie ein Vergleich von Bild 5b und Bild 5c zeigt. Demzufolge enthalten die hochfrequenten Koeffizienten der DCT geringere Energie als die der DFT und der übrigen Transformationen, sodass die DCT-Codierung eine höhere Codierungseffizienz aufweist. Die DCT wird normalerweise mit Blockgrößen von 8×8 Bildpunkten, neuerdings auch 4×4 Bildpunkten durchgeführt.

Die prädiktive Codierung wie auch die Transformations-Codierung haben das Ziel, aus korrelierten Abtastwerten unkorrelierte

Prädiktionsfehler bzw. unkorrelierte Transformations-Koeffizienten zu erzeugen, die anschließend quantisiert und codiert werden. Demzufolge ist auch die Datenreduktion durch Ausnutzung der Redundanz nahezu gleich. Unterschiede ergeben sich bezüglich der durch Quantisierung erreichten zusätzlichen Irrelevanzreduktion.

Die Quantisierung der prädiktiven Codierung nutzt die Maskierung durch Helligkeitssprünge aus (Bild 3), die Quantisierung der Transformations-Codierung nutzt die so genannten Kontrastempfindlichkeit (Bild 1) des Auges aus. Ein Überschreiten der Sichtbarkeitschwelle führt im Fall der prädiktiven Codierung zu sichtbarem Rauschen an Helligkeitssprünge, im Fall der Transformations-Codierung zu so genannten Blockartefakten.

Unter Verwendung einer Intraframe-Codierung kann die Datenrate einer PCM-Codierung von 8 Bit je Abtastwert auf etwa 2 bis 3 Bit je Abtastwert ohne sichtbare Codierungsfehler reduziert werden.

2.2 Interframe-Codierung

Um auch die Bild-zu-Bild-Redundanz auszunutzen, muss das vorangegangene Bild einer Bildsequenz gespeichert und eine so genannten Interframe-Codierung angewendet werden. Der 1969 von Mounts [21] veröffentlichte legendäre „Conditional Replenishment Coder“ der Bell Laboratorien war die Basis vieler weltweit nachfolgender Forschungsarbeiten. Als Bildspeicher wurde dabei in den Bell Laboratorien unter anderem ein Magnetrommelspeicher eingesetzt. Bild 6 zeigt ein Blockschaltbild des Coders. Es werden nur jene Bildpunkte übertragen, die sich von Bild zu Bild verändert haben, die übrigen werden aus dem gespeicherten vorangegangenen Bild übernommen.

Im Vergleich zur Intraframe-Codierung kann mit dem Conditional Replenishment Coder eine erheblich größere Datenreduktion erreicht werden. Für monochrome Kopf-Schulter-Bildsequenzen, wie

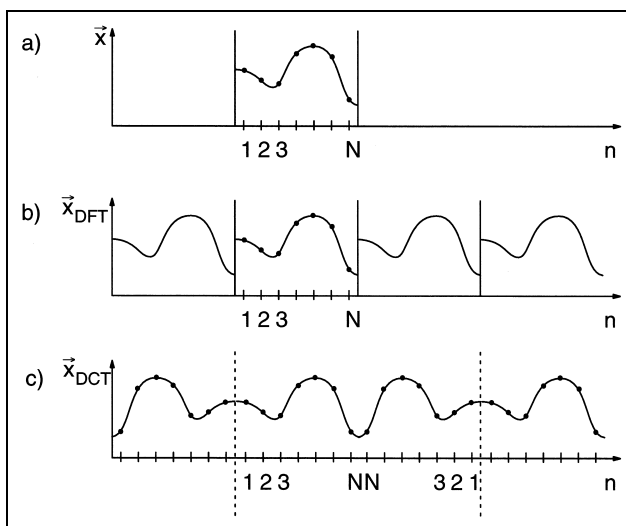


Bild 5a–c Periodisches Signal eines zu codierenden eindimensionalen Blocks nach Anwendung einer DFT- und DCT-Codierung und -Decodierung. (a) zeigt den zu codierenden Block, (b) das Signal nach DFT-Codierung und -Decodierung und (c) das Signal nach DCT-Codierung und -Decodierung.

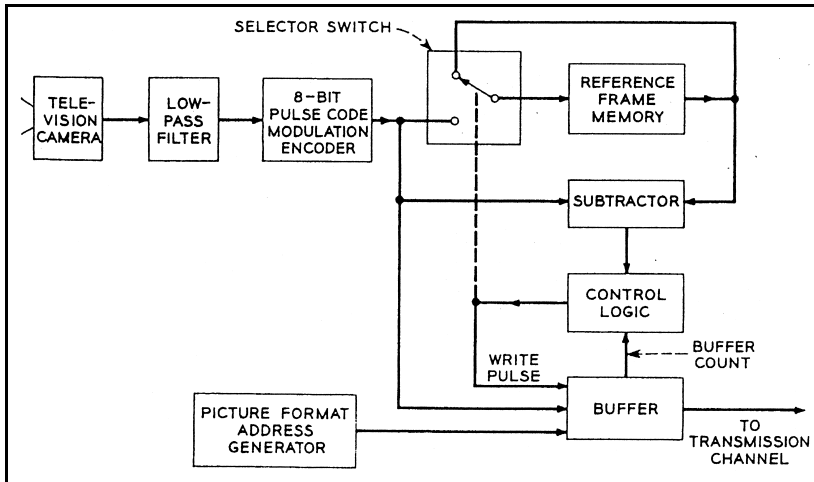


Bild 6 Conditional Replenishment Coder nach Mounts [21].

sie für Bildfernseh-Anwendungen typisch sind, wird im Mittel eine Datenrate von nur noch etwa 1 Bit je Abtastwert benötigt.

In den USA und Japan wurden unter Verwendung dieses Prinzips Codierungssysteme für Videokonferenz-Anwendungen mit 1,544 Mbit/s [22; 23], in Europa mit 2,044 Mbit/s Übertragungsrate entwickelt [24]. Diese Datenraten entsprechen der ersten Stufe der inzwischen standardisierten PCM-Hierarchie der Übertragungsnetze.

Musmann und Klie [25] demonstrierten 1979 erstmals die mit einem erweiterten Conditional Replenishment Coder und einer Datenrate von 64 kbit/s erreichte Bildqualität. Die Erweiterung bestand in einer zusätzlichen Bild-zu-Bild-Filterung der geänderten Bildpunkte. Das Experiment galt dem Ziel, Bildfernsehsignale zu gleichen Übertragungskosten wie die eines Fernsehsignals zu übertragen.

Die erste Monographie zum Thema Bildcodierung wurde 1979 von W.K. Pratt [36] herausgegeben. Sie beschreibt die Anfänge der ersten 20 Jahre einschließlich Faksimile-Codierung.

Ein bedeutender nächster Meilenstein war der auf Rocca [26] zurückgehende Vorschlag einer so genannten bewegungskompensierenden Bild-zu-Bild-Prädiktion, bei der die Bewegung von Objekten gemessen und zur Erzeugung eines

Prädiktionsbildes verwendet wird. Als dieser Vorschlag 1972 aufkam, glaubte niemand daran, dass eine solche Idee einmal realisiert werden könnte. Zunächst musste das Problem der Bewegungsschätzung gelöst werden. Erste Veröffentlichungen dazu erschienen 1975 [27]. Man schätzte die translatorische Bewegung mit dem so genannten Displacement-Vektor. Von der anfänglichen Idee bis zu den ersten Ergebnissen eines Codierungssystems mit Bewegungskompensation dauerte es sieben Jahre. Netravali und Robbins beschreiben in [28] ein Codierungssystem mit Rückwärtsprädiktion und bildpunktweiser rekursiver Bewegungsschätzung. Sie zeigen in Rechnersimulationen, dass die Übertragungsrate um etwa 30%

im Vergleich zum Conditional Replenishment Coder reduziert werden kann. Im Gegensatz zu einem Codierungssystem mit Rückwärtsprädiktion muss bei einer Vorwärtsprädiktion der Displacement-Vektor einer translatorischen Bewegung zusätzlich als Seiteninformation übertragen werden. Um die Seiteninformation gering zu halten, schlagen J.R. Jain und A.K. Jain in [29] eine blockbasierte bewegungskompensierte Prädiktion vor, die im Bild 7 veranschaulicht ist. Für einen zu codierenden Block $z = (x, y)$ in Bild k wird zunächst ein Displacement-Vektor $D(d_x, d_y)$ gemessen und übertragen. Zur Prädiktion des Blocks $z = (x, y)$ in Bild k wird anschließend die um den Displacement-Vektor versetzte Bildinformation aus Bild $k-1$ verwendet.

Noch war man aber von der Zweckmäßigkeit der Anwendung einer bewegungskompensierten Prädiktion nicht überzeugt. Daher untersuchten Ishiguro und Iinuma [30] die Codierungseffizienz der bewegungskompensierten Prädiktion in Abhängigkeit von der Bewegungsgeschwindigkeit und verglichen diese mit der einer einfachen Interframe- und Intraframe-Codierung.

In einer mathematischen Analyse zeigt Girod [31] 1987, dass die bewegungskompensierte Prädiktion in Verbindung mit einer Intraframe-Codierung des Prädiktions-

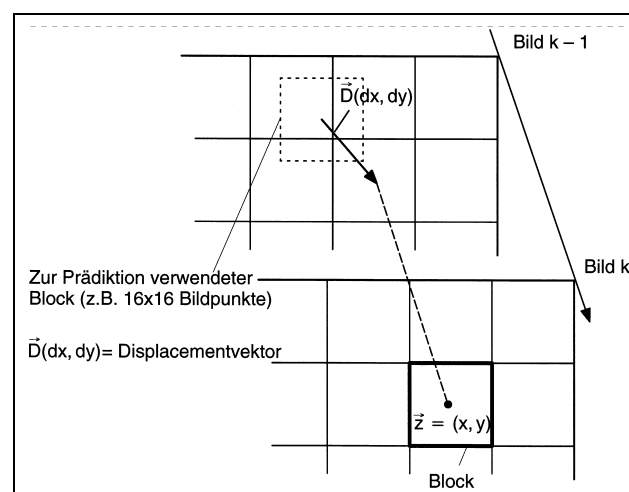


Bild 7 Darstellung einer blockbasierten bewegungskompensierten Prädiktion.

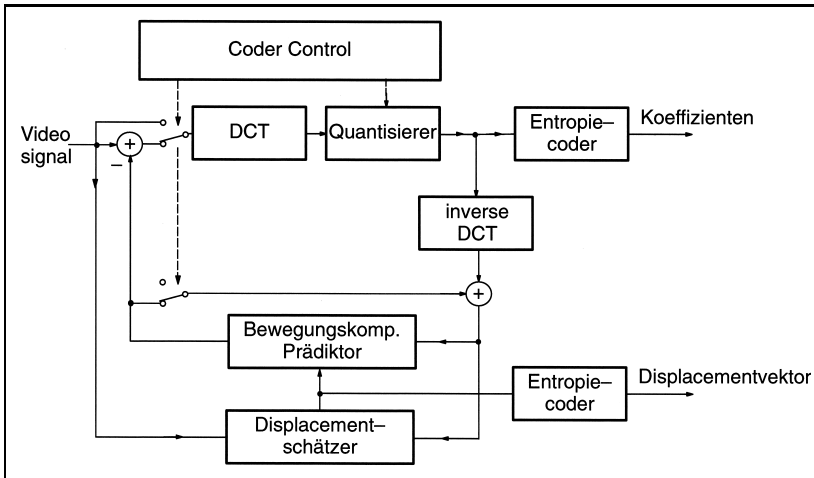


Bild 8 Blockschaltbild einer bewegungskompensierten Hybridcodierung nach [35].

fehlers nur dann einen Codierungsgewinn bringt, wenn die Varianz des Displacement-Schätzfehlers kleiner als $0,1 \text{ pel}^{-2}$ ist (pel – Kurzform für Bildpunkt). In Konsequenz dieser Erkenntnis bemühte man sich, die Schätzgenauigkeit der Displacement-Schätzer zu verbessern, wie z. B. Koga et al. [32] und Bierling [33]. In einem Übersichtsaufsatz von Musmann et al. [34] werden die verschiedenen Ansätze zur Displacement-Schätzung beschrieben und verglichen.

Netravali und Haskell [37] stellten in mühevoller Arbeit sowohl die Grundlagen wie auch die inzwischen entwickelten Codierungstechniken in einem umfangreichen, 1988 veröffentlichten Werk zusammen, das wohl als das bedeutendste Handbuch der Bildcodierung angesehen werden kann und inzwischen auch in zweiter Auflage erschienen ist.

Etwa zur gleichen Zeit begann die Study Group XV des CCITT (inzwischen umbenannt in ITU-T), den ersten digitalen Video-Codierungsstandard H.261 für Bildfernsprechen und Videokonferenz zu entwickeln. Der Chairman war Sakae Okubo. Das Ergebnis der Standardisierung war die bewegungskompensierte Hybridcodierung gemäß Bild 8.

Die bewegungskompensierte Prädiktion wird entsprechend Bild 7 mit so genannten Makroblöcken

von 16×16 Bildpunkten vorgenommen. Die verbleibenden Prädiktionsfehler werden anschließend einer Diskreten Cosinus-Transformation (DCT) mit Blöcken von 8×8 Elementen zugeführt. Die DCT erzeugt Blöcke mit 8×8 Koeffizienten, die zunächst quantisiert und nach einem Zig-Zag-Scan der Koeffizienten gemäß Bild 9 codiert werden. Der Zig-Zag-Scan reiht die Koeffizienten entsprechend abnehmender Varianz. Nach dem letzten Koeffizienten mit einer Amplitude größer Null wird zur Einsparung von Datenrate nur noch ein so genanntes EOB-Zeichen (end of block) gesendet. Neben den auf diese Weise codierten P-frames gibt es im Codierungsstandard H.261 auch I-frames mit Intraframe-Codierung. B-frames mit bidirektionaler Prädiktion wurden

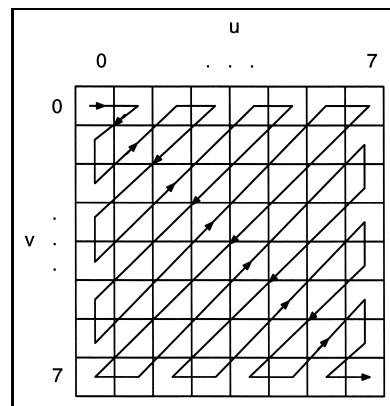


Bild 9 Zig-Zag-Scan für Blöcke mit 8×8 DCT-Koeffizienten.

erst in späteren Codierungsstandards vorgesehen.

Die bewegungskompensierte Hybridcodierung wurde die Basis einer Reihe nachfolgender Codierungsstandards, die von der International Telecommunications Union (ITU) und der International Standardization Organisation (ISO-MPEG) für unterschiedliche Anwendungen entwickelt wurden.

3 ITU- und MPEG-Standards

Anknüpfend an die in Abschnitt 2.2 dargestellte Entwicklung des ITU-T-Codierungsstandards H.261 entwickelte die ITU-T Study Group XV in den nachfolgenden Jahren aufgrund neuer Anwendungsanforderungen den in [38] beschriebenen, weiteren Codierungsstandard H.263. Ziel war es, neben den auf ISDN-Kanäle ausgerichteten Übertragungsbitraten von $n \times 64 \text{ kbit/s}$ ($n = 1, 2, \dots, 30$) des H.261-Standards auch Bitraten von weniger als 64 kbit/s zu ermöglichen, wie sie u. a. für Übertragungen über Analogkanäle des Fernsprechnetzes und für Mobilfunkverbindungen benötigt werden.

Ein wichtiger Schritt zu entsprechend erhöhter Leistungsfähigkeit der Codierung ging von den in Abschnitt 2.2 angeführten Überlegungen zum Einfluss des Displacement-Schätzfehlers aus und bestand in der Verwendung einer auf $\frac{1}{2} \text{ pel}$ -Genauigkeit verbesserten Displacement-Schätzung. Sie erfordert eine Interpolation zur Erzeugung zusätzlicher Bildpunkte in $\frac{1}{2}$ -pel-Abstand. In Vergleichsuntersuchungen konnten Girod et al. [39] am Beispiel der bekannten Videotestsequenz „Foreman“ zeigen, dass die mit einer H.261-Codierung bei 64 kbit/s erzielte subjektive Bildqualität mit einer H.263-Codierung bereits bei weniger als der Hälfte dieser Bitrate erreicht wird. Zur Erhöhung der Leistungsfähigkeit der Codierung tragen im Wesentlichen bei:

- die verbesserte Genauigkeit der Displacement-Vektoren von $\frac{1}{2} \text{ pel}$, die den Hauptbeitrag liefert,

Standard ¹⁾	Video-format	Displacement-Schätzung		Bitrate	Anwendung
		Blockgröße	Genauigkeit		
H.261 1993 [35]	CIF QCIF	16 × 16 pel	1 pel (ganzzahlig)	n × 64 kbit/s (n=1,2,...,30)	Bildfern- sprechen, Video- konferenz
H.263 1998 [38]	CIF QCIF sub QCIF	16 × 16 pel 8 × 8 pel	1/2 pel	n × 8 kbit/s	Bildfern- sprechen, Video- konferenz, Streaming für Internet

¹⁾ inkl. Erscheinungsjahr der ITU-T Recommendation lt. Literaturzitat

Bild 10 Merkmale der ITU-T-Codierungsstandards H.261 und H.263.

sowie die Betriebsart-Optionen:

- Vier Displacement-Vektoren je Makroblock
- Bewegungskompensation mit überlappenden Blöcken
- Codierung von PB-frames als Einheit
- Arithmetische Codierung anstelle von Huffman-Codes.

Charakteristische Merkmale der beiden Codierungsstandards stellt die Tabelle nach Bild 10 zusammenfassend gegenüber.

Ende der 1980er Jahre begann die auf Initiative von Hiroshi Yasuda gegründete Moving Picture Experts Group (MPEG) der ISO mit der Entwicklung von Codierungsstandards für Video- und zugehörige Audioinformation für Anwendungen im Bereich digitaler Speicher- und Übertragungsmedien, wie CD-ROM, DVD und Digitales Fernsehen. Aus der ersten Phase dieser Tätigkeit ging im Jahre 1993 der MPEG-1-Standard [40] hervor. Ziel war insbesondere die Nutzung des Speichermediums CD-ROM mit kurzer Zugriffszeit für die Videodarbietung in einer mit VHS-Aufzeichnungen vergleichbaren Bildqualität, basierend auf dem CIF-Videoformat und einer durch die CD-ROM-Eigenschaften vorgegebenen Video-Bitrate von 1,15 Mbit/s am Coderausgang. Grundlage des Standards ist eine bewegungskompensierte Hybridcodierung mit $\frac{1}{2}$ pel-Genauigkeit der Displacement-Vektoren. Einer der weiteren Unterschiede

zum ITU-T-Standard H.261 besteht in der Verarbeitung der Bildsequenz in der Form von Bildgruppen (*Group of Pictures*), wobei das erste Bild einer Bildgruppe immer als I-frame codiert wird und für die übrigen Bilder einer Gruppe verschiedene Verarbeitungsmodi (Intraframe, bewegungskompensierende oder bidirektionale bewegungskompensierende Prädiktion) vorgesehen sind. Dadurch wird z.B. auch bei Szenen mit Bildschnitten und Trickfunktionen eine schnelle Anpassung der Codierung sowie eine Decodierung beginnend bei einer beliebigen Bildgruppe ermöglicht.

Die Nachfolgephase dieser Standardisierungsaktivität der ISO galt der Entwicklung eines auf MPEG-1 aufbauenden Standards für Videosignale im Format ITU-R BT.601, d.h. einer Bildauflösung entspre-

chend der des Rundfunkfernsehens. Anwendungshintergrund waren insbesondere Digitales Fernsehen und DVD. Der mit dieser Zielsetzung im Jahr 1995 entstandene Standard MPEG-2 [41] erweitert MPEG-1 durch zusätzliche Betriebsarten der bewegungskompensierten Prädiktion, welche die Zeilensprungstruktur des vorgegebenen Abtastrasters berücksichtigen. MPEG-2 benötigt für die Bildqualität des Rundfunkfernsehens nur noch eine Datenrate von 4 Mbit/s. HDTV-Signale können entsprechend mit Bitraten um 16 Mbit/s codiert werden. MPEG-2 wurde darüber hinaus zusammen mit dem ITU-T-Standard H.263 zur Grundlage für den im Jahre 2000 spezifizierten MPEG-4 Standard [42], der auch objektbasierte Werkzeuge in die Codierung einbezieht, die neben Textur- und Bewegungsmerkmalen auch Formmerkmale der Videoszene berücksichtigen. Vorschläge derartiger Codierungen, auch modellbasierte Codierungen genannt, gehen auf Forchheimer [52], Geuen [53] und Aizawa et al. [54] zurück. Abhängig vom verwendeten Objektmodell unterstützen sie die Codierung spezieller Bildinhalte. Der MPEG-4 Standard ist in einem speziellen Beitrag dieses Themenheftes ausführlicher beschrieben. In Bild 11 sind ausgewählte Merkmale der MPEG-Standards als Übersicht zusammengestellt.

Standard ²⁾	Video-format	Displacement-Schätzung		Bitrate	Anwendung
		Blockgröße	Genauigkeit		
MPEG-1 1993 [40]	CIF	16 × 16 pel	1/2 pel	1,15 Mbit/s	CD-ROM
MPEG-2 1995 [41]	TV HDTV	16 × 16 pel	1/2 pel	4...9 Mbit/s 16...36 Mbit/s	Digitales Fernsehen, DVD
MPEG-4 2000 [42]	QCIF CIF TV Video objekte	16 × 16 pel 8 × 8 pel globale Displacement- Schätzung	1/4 pel	32 kbit/s ...4 Mbit/s	Bildkommunikation, Streaming für Internet

²⁾ inkl. Erscheinungsjahr des Standards lt. Literaturzitat

Bild 11 Merkmale der ISO-Codierungsstandards MPEG-1, MPEG-2 und MPEG-4.

4 Gegenwärtige Entwicklung

Die ITU-T Video Experts Group und die ISO/IEC Moving Picture Experts Group haben Ende 2001 ein so genanntes Joint Video Team (JVT) gegründet mit dem Ziel, die laufende Standardisierung von H.26L und MPEG-4 zusammenzuführen. Der zukünftige gemeinsame Codierungsstandard wird voraussichtlich die Bezeichnung H.264 bzw. MPEG-4 Part 10 AVC erhalten [43]. Im Vergleich zum H.263 und MPEG-4 Simple Profile wird der neue Codierungsstandard u.a. folgende zusätzliche Komponenten enthalten, die wesentlich zur Verbesserung der Codierungseffizienz beitragen: Intraframe-Prädiktion für I-Bilder, Filterung des Prädiktionssignals zur Verringerung von Blockartefakten, variable Blockgrößen für die Bewegungsschätzung, Displacement-Vektorschätzung mit erhöhter Genauigkeit, bewegungskompensierte Prädiktion mit mehreren Referenzbildern und Context-Based Adaptive Entropy Coding.

Die *Intraframe-Prädiktion für I-Bilder* wird noch vor der DCT-Codierung der 4×4 Blöcke angewendet. Zur Prädiktion der Bildpunkte eines 4×4 Blockes werden Referenzbildpunkte aus bereits übertragenen benachbarten Blöcken verwendet und damit die statistische Abhängigkeit von Block-zu-Block ausgenutzt. Zur Auswahl der Referenzbildpunkte stehen mehrere Modi zur Verfügung, die als Seiteninformation übertragen werden müssen.

Die *Filterung des Prädiktionssignals* mit einem so genannten „de-blocking filter“ soll Signalsprünge an den Blockgrenzen verringern, die bei geringer Datenrate sichtbar werden können.

Die Einführung *variabler Blockgrößen für die Bewegungsschätzung* und bewegungskompensierte Prädiktion erfordert die Übertragung mehrerer Displacement-Vektoren je Makroblock. Mit der Unterteilung eines Makroblocks kann die bewegungskompensierte Prädiktion ins-

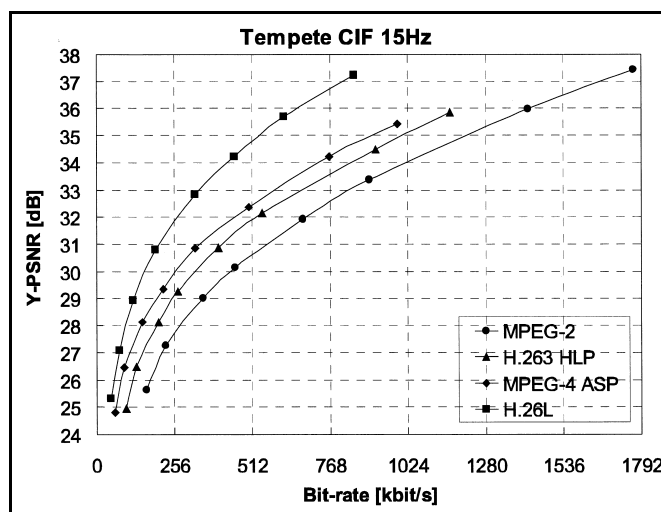


Bild 12 Signalrauschabstand als Funktion der Datenrate für verschiedene Codierungen, gemessen mit der Testsequenz Tempete nach [48].

besondere an den Rändern bewegter Objekte verbessert werden, wo Bildteile eines Makroblockes unterschiedliche Bewegung aufweisen. Es werden die Blockgrößen 16×8 , 8×8 , 8×4 und 4×4 pel zur Verfügung gestellt. Verglichen mit einer festen Blockgröße von 16×16 pel kann trotz der zusätzlich zu übertragenden Displacement-Vektoren die Datenrate auf diese Weise um etwa 10 bis 15% reduziert werden [44].

Die *Displacement-Vektoren* werden nunmehr mit $\frac{1}{4}$ pel-Genauigkeit anstelle von $\frac{1}{2}$ pel-Genauigkeit geschätzt. Die größere Genauigkeit erfordert ein dazu angepasstes Interpolationsfilter wie von Benzler [45] gezeigt wurde. Nur dann bringt die damit erzielte verbesserte bewegungskompensierte Prädiktion auch eine zusätzliche Datenreduktion von etwa 20%.

Eine weitere Verbesserung der bewegungskompensierten Prädiktion kann erreicht werden, indem nicht nur ein, sondern *mehrere vorausgehende Referenzbilder für die Prädiktion* bereitgehalten werden. Das zur Prädiktion jeweils verwendete Referenzbild muss dem Decoder über Seiteninformation mitgeteilt werden. Diese, auch als Multiframe-Prädiktion bezeichnete Technik geht auf Wiegand, Zhang und Girod [46] zurück. Sie reduziert die Datenrate um 5 bis 10% [44].

Für die Entropie-Codierung sowohl der DCT-Koeffizienten als auch der Displacement-Vektoren

stehen zwei neue Verfahren zur Auswahl, denen beiden gemeinsam ist, dass sie die Codierung eines Ereignisses abhängig vom Kontext verändern. Als Kontext wird beispielsweise der Bildinhalt eines benachbarten, bereits übertragenen Blocks verwendet. Eine solche Codierung bringt Codierungsgewinne, wenn die zu codierenden Ereignisse vom Kontext statistisch abhängig sind. Im Vergleich zum *Context-Adaptive Variable Length Coding* (CAVLC) verwendet die *Context-based Adaptive Binary Arithmetic Coding* (CABAC) eine arithmetische Codierung zur Entropie-Codierung. Sie ist aufwändiger, ermöglicht aber auch eine größere Datenreduktion von etwa 10% [44].

Die angeführten Erweiterungen der bewegungskompensierten Hybrid-Codierung reduzieren die Datenrate insgesamt um etwa 50%, wie Messungen der subjektiven Bildqualität ergeben haben [47]. Bild 12 zeigt dazu Ergebnisse objektiver Bewertungen anhand der operationellen Rate-Distortion-Funktionen des Luminanzsignals Y. Insbesondere ist zu entnehmen, dass bei einem PSNR von 37 dB die Datenrate der H.26L-Codierung nicht einmal halb so groß wie die der MPEG-2-Codierung ist.

5 Ausblick

Inzwischen sind weitere Vorschläge zur Erhöhung der Codierungseffizienz bekannt geworden, die in

zukünftige Standardisierungen einbezogen werden können. Dazu zählen: Displacement-Schätzung mit nochmals erhöhter Genauigkeit von $\frac{1}{8}$ pel, adaptive Filter für die Interpolation und Prädiktion, DCT mit adaptiver Blockgröße, objektbasierte Codierung und skalierbare Wavelet-Codierung.

Wedi [49] zeigt, dass die Datenrate bei Anwendung von *Displacement-Vektoren mit $\frac{1}{8}$ pel-Genauigkeit* anstelle von $\frac{1}{4}$ pel-Genauigkeit bei hoher Datenrate um 2 Mbit/s um nochmals 10% vermindert werden kann. Bei niedrigen Datenraten unter 150 kbit/s bringt dieser Vorschlag bisher keinen Gewinn [48].

Darüber hinaus schlägt Wedi [50] auch die Anwendung *adaptiver Interpolationsfilter* für die bewegungskompensierte Prädiktion vor, die für jedes Bild berechnet und deren Koeffizienten als zusätzliche Seiteninformation übertragen werden müssen. Damit ist eine weitere Datenreduktion um etwa 5 bis 10% möglich.

Nachdem bereits variable Blockgrößen für die Bewegungsschätzung und die bewegungskompensierte Prädiktion eingeführt sind, liegt es nahe, auch eine daran angepasste *variable Blockgröße für die DCT* anzuwenden, wie es Wien und Mayer [51] vorschlagen. Bei niedrigen Datenraten kann mit diesem Ansatz die Datenrate um etwa 5%, bei hohen Datenraten um etwa 10% reduziert werden.

Mit den erläuterten neuen Vorschlägen wird in Zukunft insbesondere bei hohen Datenraten, wie sie für Videosignale in ITU-601 Format, d. h. für Fernsehsignale, benötigt wird, die Datenrate um insgesamt etwa 25%, d. h. auf 1,5 Mbit/s zu reduzieren sein. Demzufolge bietet sich in Zukunft die Möglichkeit, Videosignale in Fernsehqualität aus dem Internet über eine Fernsprechanschlussleitung mit DSL-Modem zum Teilnehmer zu übertragen. Dem Fernsehen wird sich damit eine neue Dimension eröffnen.

Untersuchungen von Eisert et al. [55] sowie von Kampmann [56]

zur *objektbasierten Codierung* deuten darauf hin, dass im Fall von Kopf-Schulter-Bildern, wie beispielsweise bei Videokonferenzen oder Nachrichtensprechern, mit diesem Ansatz für Bildsequenzen in QCIF oder CIF sogar nur 6 kbit/s Datenrate benötigt werden. Damit wären in Zukunft sogar Bewegtbildübertragungen über schmalbandige Mobilfunkkanäle wie GSM möglich.

Um auch Teile codierter Daten separat decodieren zu können, muss eine Codierung skalierbar sein. Eine skalierbare Codierung kann beispielsweise dazu dienen, einen Bildverlust bei Übertragungsfehlern zu vermeiden. Hierzu bieten sich besonders *Wavelet-Codierungen* an, die zurzeit intensiv u. a. von Ohm [57] sowie Hsiang und Woods [58] untersucht werden.

Literatur

- [1] E.R. Kretzmer: „Reduced-alphabet representation of television signals“, IRE Nat. Conv. Rec., vol. 4, pt. 4, pp. 140–147, 1956.
- [2] W.F. Schreiber, C.F. Knapp: „TV bandwidth reduction by digital coding“, IRE Convention Record, Part IV, pp. 88–98, 1958.
- [3] C.W. Harrison: „Experiments with linear prediction in television“, Bell Syst. Tech. J. vol. 31, pp. 764–783, July 1952.
- [4] R.E. Graham: „Predictive quantizing of television signals“, IRE Wescon Convention Record, vol. 2, pt 4, 1958 pp. 147–157.
- [5] C.C. Cutler: „Differential quantization of communication signals“, U.S. Patent 2 605 361, July 1952.
- [6] J.O. Limb: „Source receiver encoding of television signals“, Proc. IEEE, vol. 55, pp. 364–380, Mar. 1967.
- [7] J.B. O’Neal: „Predictive quantizing systems (differential pulse code modulation) for the transmission of television signals“, Bell Syst. Tech. J. vol. 45, pp. 689–722, May 1966.
- [8] D.J. Connor, R.F. Pease, W.G. Scholes: „Television coding using two-dimensional spatial prediction“, Bell Syst. Tech. J. vol. 50, pp. 1049–1061, Mar. 1971.
- [9] G. Bostelmann: „A simple high quality DPCM system for videotelephony using 8 Mbit per second“, Nachrichtentechnische Zeitschrift, vol. 27, pp. 115–117, Feb. 1974.
- [10] W. Thoma: „Optimizing the DPCM for signals using a model of the human visual system“, Proc. of the IEEE Int. Zürich Seminar on Digital Communications pp. C3 (1) – C3 (7), Mar. 1974.
- [11] D.K. Sharma, A.N. Netravali: „Design of quantizers for DPCM coding of picture signals“, IEEE Trans. Comm., com-25, pp. 1267–1274, Nov. 1977.
- [12] M.A. Devereux: „Digital video: Differential coding of a PAL signal based on differences between samples one subcarrier period apart“, BBC Res. Dep. Rep. 1973/6.
- [13] J.E. Thomson: „Differential coding of composite color television signals using chrominance-corrected prediction“, IEEE Trans. Comm. com-22, pp. 1106–1113, Aug. 1974.
- [14] H. Enomoto, K. Shibata: „Features of Hadamard transformed television signals“, 1965 Nat. Conf. IECE, Japan, paper No. 881.
- [15] H.C. Andrews, W.K. Pratt: „Fourier transform coding of images“, Proc. of the Hawaii Int. Conf. System Science, pp. 677–679, Jan. 1968.
- [16] A. Habibi, P.A. Wintz: „Image coding by linear transformation and block quantization“, IEEE Trans. Commun. Tech. com-19, pp. 50–63, Feb. 1971.
- [17] J.W. Woods, T.S. Huang: „Picture bandwidth compression by linear transformation and block quantization“, Symposium Picture Bandwidth compression, M.I.T. Cambridge, MA, 1969.
- [18] N. Ahmed, T. Natarajan, R. Rao: „Discrete cosine transform“ IEEE Trans. Comput. vol C-23, pp. 90–93, Jan. 1974.
- [19] J.G. Robson: „Spatial and temporal contrast sensitivity functions of the visual system“, J. Opt. Soc. Am. vol 56, pp. 1141–1142, 1966.
- [20] A. Fiorentini, M.T. Zoli: „Detection of a target superimposed to a step pattern of illumination“, Atti. Fond. G. Ronchi, vol 22, pp. 207–217, 1967.
- [21] F. Mounts: „A video encoding system with conditional picture-element replenishment“, Bell Syst. Tech. J. 48, pp. 2545–2554, Sep. 1969.

- [22] H. Yasuda, F. Kanaya, H. Kawanishi: „1.544 Mbit/s transmission of tv-signals by interframe coding system“, IEEE Trans. on Commun., COM-24, No 10, pp. 1175–1180, Oct. 1976.
- [23] B.G. Haskell, P.T. Gordon, R.L. Schmidt, J.V. Scattaglia: „Interframe Coding of 525-line, monochrome television at 1.5 Mbits“, IEEE Trans. on Commun., vol COM-25, pp. 1339–1348, Nov. 1977.
- [24] J.E. Thomson: „European collaboration on picture coding research for 2 Mbit/s transmission“, IEEE Trans. on Commun., vol-COM-29, pp. 2003–2004, Dec. 1981.
- [25] H.G. Musmann, J. Klie: „TV transmission using 64 kbit/s transmission rate“, Proc. of the Internat. Conf. on Communications, Boston, pp. 23.3.1–23.3.5, June 1979.
- [26] F. Rocca, S. Zanoletti: „Bandwidth reduction via movement compensation on a model of the random video process“, IEEE Trans. Commun. vol COM-20, pp. 960–965, Oct. 1972.
- [27] J.O. Limb, J.A. Murphy: „Measuring the speed of moving images from television signals“, IEEE Trans. on Commun. vol COM-23, No 4, pp. 474–478, Apr. 1975.
- [28] A.N. Netravali, J.D. Robbins: „Motion compensated television coding – Part 1“, Bell Syst. Tech. J. vol 58, pp. 631–670, Mar. 1979.
- [29] J.R. Jain, A.K. Jain: „Displacement measurement and its application in interframe image coding“, IEEE Trans. on Commun. vol COM-29, pp. 1799–1806, Dec. 1981.
- [30] T. Ishiguro, K. Iinuma: „Television bandwidth compression transmission by motion compensated interframe coding“, IEEE Commun. Mag. pp. 24–30, Nov. 1982.
- [31] B. Girod: „The efficiency of motion-compensating prediction for hybrid coding of video sequences“, IEEE J. on Selected Areas in Commun., vol SAC-5 No 7, pp. 1140–1154, Aug. 1987.
- [32] T. Koga, K. Jinuma, A. Hirano: *Y. Jijima and T. Ishiguro*, „Motion-compensated interframe coding for video conferencing“, Proc. of the NTC 81, pp. G 5.3.1–G 5.3.5, Dec. 1981.
- [33] M. Bierling: „Displacement estimation by hierarchical block matching“, Proc. of the 3rd SPIE Symp. on Visual Commun. and Image Processing, USA, pp. 942–951, Nov. 1988.
- [34] H.G. Musmann, P. Pirsch, H.J. Grallert: „Advances in picture coding“, Proc. of the IEEE, vol 73, No 4, pp. 523–548, Apr. 1985.
- [35] ITU-T Recommendation H.261: Video codec for audiovisual services at $p \times 64$ kbit/s, 1993.
- [36] W.K. Pratt (Ed.): Image Transmission Techniques, New York: Academic Press, 1979.
- [37] A.N. Netravali, B. Haskell: Digital Pictures, New York: Plenum Press, 1988.
- [38] ITU-T Recommendation H.263: Video coding for low bit rate communication, Geneva, 1998.
- [39] B. Girod, E. Steinbach, N. Färber: „Comparison of the H.263 and H.261 Video Compression Standards“, Standards and Common Interfaces for Video, SPIE, vol. CR 60, pp. 233–251, Oct. 1995.
- [40] 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, Geneva, 1993. (MPEG-1).
- [41] ISO/IEC.IS 13818-1: Information technology – generic coding of moving pictures and associated audio information systems, Geneva, 1995 (MPEG-2 Systems).
- [42] ISO/IEC.IS 14496-2000-2 Information technology – coding of audio-visual objects – part 2: visual, Geneva, 2000 (MPEG-4 video).
- [43] ITU-T Rec. H.264/ISO/IEC 11496-10. „Advanced video coding“, Final Committee Draft, Document JVT-F100, Dec. 2002.
- [44] Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, „Data transfer and storage complexity analysis of the AVC/JVT codec on a tool-by-tool basis“, JVT-D138 doc, July 2002.
- [45] U. Benzler: „Results of core experiment P8 (Motion and aliasing compensating prediction)“, DOC. ISO/IEC ITC1/SC29/WG11, MPEG 97/2625, Oct. 1997.
- [46] T. Wiegand, X. Zhang, B. Girod: „Long-term memory motion compensated prediction“, IEEE Trans. on Circuits and Syst. for Video Technology, vol 9, No 1, pp. 70–84, Feb. 1999.
- [47] ISO/IEC JTC1/SC29/WG11 N4240, „Preliminary results of subjective assessment of responses to video call for new tools to further improve coding efficiency“, July 2001.
- [48] A. Joch, F. Kossentini, H. Schwarz, T. Wiegand, G. Sullivan: „Performance comparison of video coding standards using Lagrangian coder control“, Proc. of the IEEE ICIP 2002, part II, pp. 501–504, Sep. 2002.
- [49] T. Wedi: „Hybrid video coding based on high-resolution displacement vectors“, Proc. SPIE, vol 4310, Visual Commun. and Image Processing (VCIP 2001), pp. 186–193, Jan. 2001.
- [50] T. Wedi: „Adaptive interpolation filter for motion compensated prediction“, Proc. of the IEEE ICIP 2002, part II, pp. 509–512, Sep. 2002.
- [51] M. Wien, C. Mayer: „Adaptive block transform for hybrid video coding“, Proc. SPIE, vol 4310 Visual Commun. and Image Processing, (VCIP 2001), pp. 153–162, Jan. 2001.
- [52] R. Forchheimer, O. Fahlander: „Low bit-rate coding through animation“, in Proc. Int. Picture Coding Symp., Davis, CA, pp. 113–114, 1983.
- [53] W. Geuen, F. Kappeli: „Modelling of moving 3-D objects in image sequences“, in Proc. Int. Picture Coding Symp., Tokyo, Japan, paper 3. 21, 1986.
- [54] A. Aizawa, H. Harishima, T. Saito: „Model-based image coding system-construction of a 3-D model of a person's face“, in Proc. Int. Picture Coding Symp., Stockholm, Sweden, paper 3.11, 1987.
- [55] P. Eisert, T. Wiegand, B. Girod: „Model-aided coding: A new approach to incorporate facial animation into motion-compensated video coding“, IEEE Trans. on Circuits and Syst. for Video Technology, vol 10, No.3, pp. 344–358, Apr. 2000.
- [56] M. Kampmann: „Analysis-Synthesis Coding based on an automatic model of a human person“, Dissertation at the University of Hannover, Germany, Feb. 2002.
- [57] J.-R. Ohm: „Three-Dimensional Subband Coding with Motion Compensation“, IEEE Trans. on Image

Processing, vol. IP-3, no. 5, Sep. 1994, pp. 559–571.

- [58] S.-T. Hsiang, J.W. Woods: „Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank“, Signal Processing: Image Commun., vol. 16, pp. 705–724, May 2001.



1



2

1 Prof. Dr.-Ing. Hans Georg Musmann leitet seit 1972 das Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung der Universität Hannover. Der Schwerpunkt seiner Forschungsarbeiten liegt bei der Bild- und Toncodierung für zukünftige neue digitale Kommunikationssysteme. Er war Chairman der ISO MPEG/Audio, die den ISO Audio Coding Standard MP3 für die Multimedia-Technik entwickelt hat. Adresse: Universität Hannover, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Appelstr. 9A, D-30167 Hannover, Tel.: +49 (511) 762-5317, Fax: +49 (511) 762-5333, E-Mail: musec@tnt.uni-hannover.de

2 Dr.-Ing. Reiner Lippmann war nach einer Industrietätigkeit im Entwicklungsbereich der Siemens AG, von 1966 bis 1999 Senior Scientist am Deutschen Zentrum für Luft- und Raumfahrt e.V., Braunschweig sowie zwischenzeitlich Visiting Fellow an der University of New South Wales, Canberra, Australien. Sein Forschungsgebiet ist die Bord-Boden-Übertragung quellencodierter Video-Luftbildszenen. Adresse: Geysostr. 2, D-38106 Braunschweig, Tel.L +49 (531) 345941, E-Mail: rei.lippmann@t-online.de

Neu!

J. Handke

Multimedia im Internet

Konzeption und Implementierung



Oldenbourg

Jürgen Handke
Multimedia im Internet
Konzeption und Implementierung
2003. 394 Seiten
€ 44,80
ISBN 3-486-27217-9

„Multimedia im Internet“ richtet sich an Internet-Entwickler: programmierende Entwickler und Projektmanager, die über die verschiedenen multimedialen Elemente verfügen und sie nun zu komplexen Systemen zusammenstellen.

Grundsätzliches zur Konzeption und Realisierung multimedialer Anwendungen wird ebenso behandelt wie die Aspekte der Implementierung und der Effizienzsteigerung.

Die komplexe Lernumgebung "The Virtual Linguistics Campus", eine eLearning-Plattform, bei der Multimedia in bisher nicht gekanntem Maße verwirklicht wird, dient dabei als Hintergrund.

In kleinen Schritten wird gezeigt, wie die verschiedenen multimedialen Elemente, die bei CD-ROM-Entwicklungen zum Standard gehören, auch für das Internet entwickelt werden können. Dabei wird die Verknüpfung der verschiedenen Elemente zu interaktiven multimedialen Systemen beschrieben.

Oldenbourg Wissenschaftsverlag
Rosenheimer Straße 145
D-81671 München
Telefon 0 89 / 4 50 51-0
Fax 0 89 / 4 50 51-204

Weitere Informationen zum Buch:
www.oldenbourg-verlag.de

Oldenbourg

