T. WIEGAND

Videocodierstandards der ITU-T

Die ITU-T hat seit Anfang der 80er Jahre eine Reihe von sehr erfolgreichen Standards für die Videocodierung geschaffen. Die Vielzahl dieser Standards wird in diesem Beitrag beschrieben, wobei die Anforderungen und technischen Lösungen von H.26L, dem derzeitigen Projekt der ITU-T, detailliert erklärt werden. Die präsentierten Simulationsergebnisse ermöglichen einen Vergleich der Verbesserung der Codiereffizienz der einzelnen technischen Komponenten. Dabei wird gezeigt, dass sich schon mit dem gegenwärtigen Stand des H.26L-Projekts eine Bitratenreduktion von 40%, bei gleicher Qualität des decodierten Videosignals gegenüber H.263+, dem leistungsfähigsten der gebräuchlichen Videocodierstandards der ITU-T, erreichen lässt.

Dr.-Ing. Thomas Wiegand, Leiter der Arbeitsgruppe „Codierung und Übertragung visueller Information“ in der Abteilung Bildsignalverarbeitung am Heinrich-Hertz-Institut, in Berlin, ist der „Associated Rapporteur“ der „Video Coding Experts Group“ (VCEG) der ITU-T (SG 16/WP 3/Q.6) und damit einer der beiden Leiter des H.26L-Projekts.

1. Einleitung

Algorithmen zur Videokompression sind wesentliche Bestandteile für die effiziente Übertragung von Videosequenzen. Die Notwendigkeit für Videokompression ergibt sich aus der Diskrepanz zwischen der Bitrate des unkomprimierten Videosignals und der Bitrate der zur Verfügung stehenden Übertragungskanäle. Das unkomprimierte Videosignal besteht aus einer zeitlich geordneten Sequenz von Bildern, die häufig mit 25 oder 30 Bildern pro Sekunde progressiv abgetastet werden. Weiterhin sind für Farbbilder drei Farbkomponenten notwendig, um einen ausreichenden Farbraum darzustellen. Nimmt man an, dass jedes Bild einer Videosequenz die relativ niedrige QCIF-Auflösung hat und dass jeder der 176 × 144 Abtastwerte pro Bild mit 8 Bits dargestellt und nur mit 10 Bildern/s codiert wird, so ist, um diese niedrig aufgelöste Sequenz von Bildern zu übertragen, immer noch eine Bitrate von mehr als 6 Mbit/s erforderlich.

Andererseits arbeiten die heutigen Übertragungskanäle bei viel niedrigeren Bitraten, insofern es sich nicht um Broadcastkanäle handelt. Beispielsweise übertragen V.34-Modems maximal 33,4 kbit/s über analoge Telefonleitungen. Obwohl sich die DSL- und Glasfasertechniken auf dem Vormarsch befinden, sind Bitraten unter 100 kbit/s typisch für die meisten Internetverbindungen. Für die drahtlose Übertragung sind geeignete Bitraten für Videosequenzen nur eingeschränkt vorhanden. Die drahtlosen Netzwerke der zweiten Generation, wie GSM, übertragen 10 bis 15 kbit/s, was für Videosequenzen zu wenig ist. Höhere Bitraten werden bei den drahtlosen Netzwerken der dritten Generation erreicht. Jedoch bleibt die Bitrate eine wertvolle Ressource, sodass die effiziente Übertragung von Videosequenzen auch in Zukunft wichtig sein wird.

Die Effizienzsteigerung des Videoübertragungssystems kann dadurch erreicht werden, dass einerseits die Effizienz der Videokompression erhöht wird, andererseits die Robustheit des Systems gegen Übertragungsfehler durch die Anpassung der komprimierten Videodaten an die Übertragungsstrecke verbessert wird. Beide Ansätze werden bei der Entwicklung des H.26L-Standards berücksichtigt, der in diesem Beitrag vorgestellt wird. Dabei wird zunächst auf die Grundlagen der Codierung von Videosequenzen eingegangen und im Anschluss daran ein Überblick der ITU-T-Standards zur Videocodierung präsentiert. Schließlich wird der derzeitige Status des H.26L-Projekts erklärt und ein Effizienzvergleich zum vorausgehenden ITU-T-Standard H.263 durchgeführt.

Fachbegriffe und Abkürzungen

QCIF = Quarter Common Intermediate Format (176×144 Pixel @ 30 Bilder/s = 9,115 Mbit/s)
Entropie = untere Schranke für die mittlere Länge eines Codes bei gegebener Verteilung der Codesymbole
CR-Codierung = Conditional Replenishment
DPCM = Differential Pulse Code Modulation
CABAC = context-basierte adaptive binäre arithmetische Codierung
VCEG = Video Coding Experts Group

2. Codierung von Videosequenzen

Eine einfache Methode zur verlustbehafteten Codierung von Bewegbildsequenzen stellt die getrennte Verarbeitung jedes einzelnen Bildes dar, die zum Beispiel mit einem Bildcodierstandard wie JPEG [1,2] vorgenommen werden kann. Dabei wird im „Baseline“-Verfahren von JPEG eine Aufteilung des Bildes in uniforme Blöcke von 8×8 Bildpunkten vorgenommen. Diese Blöcke werden einer 2-D-Transformation, wie der in JPEG spezifizierten diskreten Cosinus-Transformation (DCT), unterzogen. Die resultierenden Koeffizienten werden quantisiert und mit Codes variabler Länge übertragen. Bei dieser Entropiecodierung der DCT-Koeffizienten wird das Verbundereignis — An-

zahl der im „Zig-Zag-Scan“ über den 8×8-Block zuvor zu Null quantisierten DCT-Koeffizienten und Wert des ungleich Null quantisierten DCT-Koeffizienten — gemeinsam codiert. Dieser Ansatz nennt sich Intra-Codierung. Er findet unter anderem Verwendung, wenn bei einem Videocodiersystem die Kompressionseffizienz weniger wichtig ist als die Recheneffizienz, oder um bei der Editierung von Sequenzen einen schnellen Zugriff auf alle Bilder zu erlauben.

Vor etwa 20 Jahren wurde zur Steigerung der Codiereffizienz auf dem Gebiet der Videokompression der Übergang von Intra- zu Inter-Bild-Algorithmen vollzogen. Dabei konnten wesentlich niedrigere Bitraten, unter Ausnutzung der statistischen Abhängigkeiten zwischen den Bildern, erzielt werden, was durch eine drastische Erhöhung des Speicher- und Rechenaufwands erkauft wurde. Eine einfache Methode zur Ausnutzung von Bild-zu-Bild-Abhängigkeiten wird „Conditional Replenishment“ (CR) genannt. CR-Codierung ist die einzige Methode der Inter-Bild-Codierung in der ersten Version der ITU-T-Empfehlung H.120 [3]. Bei der CR-Codierung teilt ein Steuersignal dem Decoder mit, ob ein entsprechender Bildblock vom vorher decodierten Bild einfach kopiert wird, weil sich nichts oder wenig in diesem Bereich geändert hat, oder ob dieser Block im Intra-Modus codiert wird. Es sei hier noch angemerkt, dass die Intra-Codierung bei H.120 auf einem DPCM-Verfahren [4] basiert und nicht auf Transformationscodierung, wie es in heutigen Standards benutzt wird. Der Vorgang des Kopierens von Bildinhalt aus dem vorher decodierten Bild in die gleiche Position im zu codierenden Bild ist auch in heutigen Standards implementiert und wird Skip-Codierung genannt. Die Codieroption, ganze Bildblöcke mit geringer Bitrate (das heißt, das Steuersignal mit einer Codewortlänge von 1 bit oder weniger) zu codieren, liefert bei Videosequenzen mit wenig Bewegung in der Bildebene schon erhebliche Gewinne, wie später gezeigt werden wird.

Die CR-Codierung weist den entscheidenden Nachteil auf, dass das Prädiktionsfehlersignal (Differenz zwischen Original und Prädiktion) nicht weiter reduziert werden kann. Daher hat man Prädiktionsfehlercodierung als zusätzliche Option in der zweiten Version von H.120 [3] hinzugefügt. Dabei wird das Prädiktionsfehlersignal wiederum als Bildsignal aufgefasst und bekannte Methoden der Bildcodierung zur Übertragung angewandt. Jedoch ist auch diese Methode wenig effizient, wenn die Sequenz starke Bewegungen in der Bildebene aufweist. Daher lässt sich ein weiterer Effizienzgewinn durch die Kompensation der „Bild-zu-Bild“-Bewegung erreichen, wobei das zuvor codierte Bild gegen das zu codierende Bild ausgerichtet wird. Diese verbesserte Näherungsmethode zur Ausnutzung statistischer „Bild-zu-Bild“-Abhängigkeiten wird bewegungskompensierte Prädiktion genannt und wurde in der ITU-T-Empfehlung H.261 [5] standardisiert. Bei der bewegungskompensierten Prädiktion kann das Prädiktionsfehlersignal ebenfalls weiter reduziert werden. Das Zusammenspiel von Bewegungskompensation und Prädiktionsfehlercodierung wird Hybridcodierung genannt. Dabei werden die zur Bewegungskompensation benötigten Parameter in allen standardbasierten Videocodecs als Seiteninformation übertragen und erfordern zusätzliche Bitrate. Der beschriebene Algorithmus wird im folgenden Inter-Codierung genannt und das dazugehörende Blockdiagramm eines Hybridcodierers ist im **Bild 1** veranschaulicht.

Alle der ITU-T-Empfehlung H.261 bis heute nachfolgenden internationalen Videocodierstandards MPEG-1 [6], MPEG-2/H.262 [7] (ein gemeinsames Projekt von ISO/IEC JTC 1/SC 29/WG 11 und ITU-T/SG 16), H.263 [8], MPEG-4[1)] [9] und auch das sich derzeit in Bearbeitung befindende H.26L-Projekt der ITU-T „Video Coding Experts Group“ (VCEG) [10], haben dieselbe Struktur der Hybridcodierung. Dabei wird das Bild in Makroblöcke, von 16×16 Bildpunkten des Luminanzsignals und den dazugehörigen Blöcken von 8×8 Bildpunkten der beiden Chrominanzsignale, aufgeteilt. Zur Vereinfachung der Beschreibung wird in den folgenden Erklärungen mit dem Bildsignal nur die Luminanzkomponente gemeint.

Jeder Makroblock kann mit einer der Codieroptionen

1. Intra-Codierung,
2. Skip-Codierung oder
3. Inter-Codierung

übertragen werden.

Bei der Inter-Codierung gibt es bei den neueren Standards H.263 und MPEG-4 sowie dem H.26L-Projekt eine weitere Aufteilung der Inter-Codieroption, die zur Bewegungskompensation mit Blöcken variabler Größe korrespondiert. Im Kapitel 3 wird auf die Videocodierstandards der ITU-T näher eingegangen.

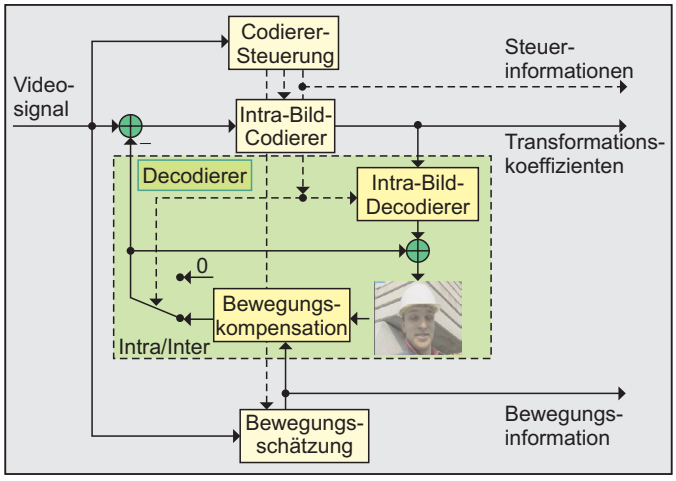

Bild 1. Blockschaltbild eines hybriden Videocodecs

3. Überblick der ITU-Videocodierstandards

H.120 [3] ist der erste internationale Standard, der für Videocodierung entworfen wurde. H.120 wurde von der ITU-T (damals noch CCITT) entwickelt und im Jahr 1984 eingeführt. Dabei werden die „Bild-zu-Bild“-Abhängigkeiten durch das Verfahren des „Conditional Replenishments“ (CR), im Zusammenspiel mit DPCM und skalierbarer Codierung mit Codes variabler Länge, ausgenutzt. Eine zweite Version von H.120 wurde 1988 verabschiedet. Bei dieser zweiten Version ist Bewegungskompensation und Hintergrundprädiktion hinzugefügt worden. (Bei Hintergrundprädiktionsverfahren wird eine Schätzung des Szenenhintergrunds durchgeführt, die bei der bewegungskompensierten Prädiktion benutzt wird und zu verbesserten Ergebnissen führen kann, wenn der Hintergrund bei Vordergrundbewegung aufgedeckt wird. Seit H.120 hat kein weiterer Standard ein solches Verfahren für die interaktive Bildkommunikation spezifiziert.) H.120 wurde bei Bitraten

[1)] MPEG-4 stellt dabei eine Abweichung dar, weil die Syntax von H.263 nur der Startpunkt von MPEG-4 war, die dann um die Funktionalität der objektbasierten Videoverarbeitung erweitert wurde. Die objektbasierten Technologien stellen somit den eigentlichen Fokus des MPEG-4-Projektes dar.

zwischen 1544 und 2048 kbit/s betrieben und ist heute nicht mehr gebräuchlich.

H.261 [5] ist der erste wirklich erfolgreiche Videocodierstandard gewesen. H.261 macht Bildkommunikation bei gebräuchlichen Bitraten zwischen 80 und 320 kbit/s möglich. In der H.261-Spezifikation findet sich zum ersten Mal die bis heute übliche Bildaufteilung in Makroblöcke von 16×16 Bildpunkten der Luminanzkomponente und 8×8 Bildpunkten der beiden Chrominanzkomponenten. Jeder Makroblock kann bei H.261 jeweils einer Intra-, Skip-, oder Inter-Codierung unterzogen werden. Die Bewegungskompensation wird bei der Inter-Codierung unter Benutzung eines Verschiebungsvektors durchgeführt, welcher eine in ganzen Bildpunkten angegebene horizontale und vertikale Verschiebung spezifiziert. Diese Beschränkung der Bewegungskompensation auf Ganz-Pixel-Genauigkeit wurde aus Komplexitätsgründen festgelegt. Die in H.261 standardisierte Prädiktionsfehlercodierung wird mit der DCT über einen Block von 8×8 Bildpunkten und einer Entropiecodierung der DCT-Koeffizienten auch in späteren Standards einschließlich MPEG-4 und H.263++ als grundlegende Struktur benutzt. H.261 wurde im Jahr 1991 von der ITU-T verabschiedet. Eine zweite Version von H.261 wurde 1993 verabschiedet und beinhaltet eine rückwärts kompatible Option zur Übertragung von hochauflösenden Graphiken.

MPEG-2/H.262 [7] ist ein gemeinsames Projekt von ISO/IEC JTC 1/SC 29/WG 11 und ITU-T/SG 16 gewesen. MPEG-2/H.262 ist im Jahr 1994 verabschiedet worden. Wichtige technische Neuerungen von MPEG-2/H.262 gegenüber H.261 sind bi-direktionale Prädiktion (B-Bilder) und Halb-Pixel-Bewegungskompensation, die auch schon im MPEG-1-Standard [6] spezifiziert worden sind. Diese Codieroptionen bewirken besonders bei hohen Bitraten eine verbesserte Codiereffizienz. Weiterhin ist in MPEG-2/H.262 die effiziente Verarbeitung von zeilenverschränkt abgetasteten Bildsignalen, sowie eine Codieroption zur Erzeugung einer skalierbaren Repräsentation des codierten Videosignals eingeführt worden. Die Bitraten, für die MPEG-2/H.262 entworfen wurde, liegen zwischen 4 und 30 Mbit/s.

H.263 [8] ist anfangs für die Bildkommunikation bei niedrigen Bitraten (zwischen 10 und 30 kbit/s) entworfen worden, um die Schwächen von H.261 in diesem Bereich auszugleichen. Jedoch hat es sich gezeigt, dass sich mit H.263 auch bei höheren Bitraten bessere Resultate als mit H.261 erzielen lassen. H.263 ist in seiner ersten Version im Jahr 1996 von der ITU-T verabschiedet worden. Die wesentlichen Neuerungen dieses Standards bestehen in der Codieroption zur Bewegungskompensation mit Blöcken variabler Größe (zusätzlich zu den Luminanzblöcken von 16×16 Bildpunkten können auch Luminanzblöcke von 8×8 Bildpunkten benutzt werden), überlappenden Fenstern bei der Bewegungskompensation, Bewegungsvektoren, die auch in extrapolierte Bereiche außerhalb des Referenzbildes zeigen und eine 3D-Entropiecodierung der DCT-Koeffizienten. Die 3D-Entropiecodierung benutzt das Verbundereignis „End-Of-Block“, Anzahl der im „Zig-Zag-Scan“ zuvor zu Null quantiserten DCT-Koeffizienten und Wert des „ungleich Null“ quantisierten DCT-Koeffizienten. Bei sehr niedrigen Bitraten (<80 kbit/s) liefert H.263 eine Bitratenersparnis gegenüber H.261 von etwa 50%, bei gleicher Qualität des rekonstruierten Signals. Bei höheren Bitraten beträgt diese Ersparnis etwa 30%.

Die zweite Version von H.263 (H.263+) ist von der ITU-T im Januar 1998 verabschiedet worden. In H.263+ sind zwölf weitere Anhänge dem H.263-Standard hinzugefügt worden, die die Codiereffizienz verbessern und die Benutzung verschiedenster Funktionalitäten, wie zum Beispiel die Verarbeitung flexibler Bildformate, ermöglichen. Bemerkenswert ist, dass bei H.263+ zum ersten Mal in einem internationalen Standard, die verbesserte Übertragung von komprimierten Videosignalen über verlustbehaftete paketvermittelte Verbindungen unterstützt worden ist.

Die dritte Version von H.263 (H.263++) ist im November 2000 von der ITU-T verabschiedet worden. In H.263++ wurden weitere drei Anhänge hinzugefügt, die die Bewegungskompensation von mehreren zuvor decodierten Referenzbildern [11], die Aufteilung der Syntaxelemente in so genannte „Partitions“ und weitere Funktionalitäten, einschließlich der exakten Spezifikation der inversen DCT, enthalten.

H.26L [10] ist das derzeitige Projekt der ITU-T VCEG. Das Projekt ist im Jahr 1997 begonnen worden, als klar wurde, dass es möglich sein würde, einen gegenüber H.263 signifikant verbesserten Videocodierstandard zu entwerfen. Dabei hat sich die ITU-T VCEG folgende Ziele gesetzt:

— Verbesserte Codiereffizienz:
Der H.26L-Standard soll eine Syntax spezifizieren, die es möglich macht, bei ähnlichen Optimierungsmethoden eine Halbierung der Bitrate gegenüber H.263+ zu erreichen.

— Verbesserte Netzwerkanpassung:
Die bei H.263 und MPEG-4 erst verspätet untersuchten Übertragungseigenschaften der Videocodiersyntax sollen bei H.26L von Anfang an bei der Entwicklung des Standards berücksichtigt werden. Hierzu gehören hauptsächlich paketvermittelte Übertragungen im Internet und Mobilfunkanwendungen für drahtlose Netzwerke der dritten Generation. Dazu werden auch Techniken der Videocodierung untersucht, die eine Verringerung der Qualitätseinbußen des decodierten Videosignals bei Paketverlusten ermöglichen.

— Einfache Videocodiersyntax:
Der Entwurf von H.26L soll zu einer einfachen und sauberen Videocodiersyntax führen, unter Vermeidung der Vielzahl der Syntaxoptionen der H.263- und MPEG-4-Standards.

Die Erreichbarkeit dieser Ziele wird sich am Ende des Projekts zeigen, das für Sommer/Herbst 2002 vorgesehen ist. Die technische Arbeit für den H.26L-Standard ist durch einen von der ITU-T VCEG ausgeschriebenen „Call for Proposals“ eingeleitet worden. Drei der Vorschläge sind damals in die engere Wahl gekommen. Dabei hat jede der vorschlagenden Firmen oder Institute (Nokia [12], Heinrich-Hertz-Institut [13], Telenor [14]) nachgewiesen, dass ihre Entwürfe zu verbesserten Codierergebnissen gegenüber H.263+ führen. Die Evaluierung der Vorschläge brachte die ITU-T VCEG zu der Entscheidung für den Telenor-Vorschlag, der dann im August 1999 zum TML-1 geworden ist. Die Abkürzung TML-1 steht für „Test Model Long-Term“, Version 1, wobei der Terminus „Long-Term“ noch aus einer Planung der ITU-T SG16 aus dem Jahr 1993 stammt. Seit August 1999 ist die Effizienz des TML stetig verbessert worden.

Die jetzige Version, TML-7 [10], enthält folgende technische Merkmale:

- Die Struktur des TML ist die eines blockbasierten Hybridcodierers und beruht auf dem Konzept, das schon auf H.261 zurückgeht. Dabei gibt es wieder die Einteilung des zu codierenden Bildes in Makroblöcke, die zu Blöcken der Luminanzkomponente von 16×16 Bildpunkten und zu Blöcken der beiden Chrominanzkomponenten von 8×8 Bildpunkten korrespondieren.
- Die Bewegungskompensation wird mit Blöcken variabler Größe durchgeführt, wobei jede Blockgröße zu einer speziellen Aufteilung des Makroblocks korrespondiert. Diese Makroblockaufteilung wird durch ein Steuersignal mitgeteilt und ermöglicht die

Bewegungskompensation mit Blöcken der Größe 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 und 4×4. Das Bewegungsmodell ist translatorisch, und es können mehrere Referenzbilder verwendet werden. Die Bewegungsvektoren können mit einer Viertel-Pixel-Genauigkeit und optional auch mit einer Achtel-Pixel-Genauigkeit übertragen werden [15]. Dazu hat es weiterführende Beiträge gegeben, die die Benutzung eines affinen Bewegungsmodells zum Gegenstand haben [12].

- Die Prädiktionsfehlercodierung wird nicht mehr, wie bei H.261 oder H.263, mit einer DCT, sondern mit einer Integer-Transformation durchgeführt. Weiterhin wird die Transformation nicht mehr für Blöcke der Größe von 8×8 Bildpunkten berechnet, wie es bei H.261 oder H.263 der Fall war, sondern mit 4×4 Bildpunkten. Zur Entropiecodierung der Transformationskoeffizienten wird wieder ein 2-D-Verbundereignis verwendet. Auch zur Größe der Blöcke für die Prädiktionsfehlercodierung hat es Vorschläge gegeben, die wiederum eine Vergrößerung der Blöcke auf Seitenlängen von 8 und 16 Bildpunkten zum Inhalt haben [16].

- Die Entropiecodierung wird mit einer einzigen Tabelle von Codes variabler Länge durchgeführt. Dieser Code ist eine reversibler „Exp-Golomb-Code“, dessen einfache Konstruktion und Decodierbarkeit hervorzuheben ist. Weiterhin kann die Entropiecodierung optional auch mit „context-basierter adaptiver binärer arithmetischer Codierung“ (CABAC) durchgeführt werden [17]. Bei dem CABAC-Verfahren wird eine Anpassung des Videocodecs an veränderliche Statistiken des Videosignals bei unterschiedlichen Abtast- und Bitraten vorgenommen. Weiterhin können Abhängigkeiten zwischen Symbolen ausgenutzt werden und auch nichtganzzahlige Bits für Symbole übertragen werden.

- Die in H.263+ und MPEG-4 nachträglich eingeführten „slices” (Sequenzen von aufeinander folgenden Makroblöcken) sind bei H.26L von Anfang an spezifiziert. Damit wird die Netzwerkanpassung der codierten Videodaten erheblich verbessert. Weiterhin ist ein „Network Adaptation Layer“ in der Spezifikation vorgesehen, welcher die Anpassung der Videodaten an relevante Übertragungskanäle vorsieht.

Bild 2 zeigt einige der beschriebenen Komponenten für das Blockdiagramm des hybriden Videocodecs.

4. Effizienz der ITU-T-Videocodierstandards

Die Codiereffizienz der zuvor beschriebenen ITU-T-Videocodierstandards soll hier vergleichend untersucht werden. Dabei wird als Verzerrungsmaß das PSNR in dB verwendet. Das PSNR für einen Bildbereich A ist definiert als

$$\mathrm{PSNR}_A(s,s') = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}_A(s,s')}\right)$$

mit

$$\mathrm{MSE}_A(s,s') = \frac{1}{|A|}\sum_{(x,y)\in A}\left(s[x,y]-s'[x,y]\right)^2,$$

wobei s das Originalsignal und s' das decodierte Signal sind.

Das PSNR wird schon seit einiger Zeit in der Bildcodierung als Qualitätsmaß genutzt. Dabei zeigen Erfahrungswerte aus der Videocodierung, dass bei einem PSNR von 35 dB eine für die meisten Betrachter annehmbare Qualität des Videosignals widergegeben wird. Als zweite Variable in den Experimenten wird die zur Rekonstruktion des Videosignals am Decoder notwendige Bitrate genutzt.

Bild 3a zeigt die Resultate für die Sequenz „Foreman“, die eine sehr starke Bewegung in der Bildebene aufweist, verursacht sowohl durch die der Person in der Szene als auch durch die Kamera. **Bild 3b** zeigt die Resultate für die Sequenz „Container Ship“, die nur wenig Bewegung bei einer statischen Kamera aufweist. Beide Sequenzen wurden in QCIF-Auflösung codiert und mit 10 Bildern pro Sekunde abgetastet. Für beide Sequenzen sind die Codierergebnisse für sieben Fälle dargestellt:

1. Intra-Codierung:
   Jedes Bild wird mit dem JPEG [1] ähnlichen Intra-Codiermodus von H.263 codiert.
2. CR-Codierung:
   Zusätzlich zum Intra-Codiermodus können Bildbereiche im Skip-Modus codiert werden. CR-Codierung ist zuerst in der 1. Version von H.120 im Jahr 1984 standardisiert worden [3]. Hier wird für den H.263-Coder neben der Intra-Codierung auch die Skip-Codierung für Makroblöcke zugelassen.
3. Bild-Differenzcodierung:
   Der CR-Codierung wird die weitere Option der Bild-Differenzcodierung hinzugefügt. Die Bild-Differenzcodierung ist zuerst in der 2. Version von H.120 im Jahr 1988 standardisiert worden [3]. Hier ist das Resultat für H.263 mit zu Null gesetzten Bewegungsvektoren gezeigt.
4. Bewegungskompensation mit Ganz-Pixel-Genauigkeit:
   Bei der Bild-Differenzcodierung werden auch Bewegungsvektoren, die auf das Bildpunktraster des Referenzbildes fallen, zugelassen. Die Bewegungskompensation mit Ganz-Pixel-Genauigkeit ist zuerst in dieser Form in H.261 im Jahr 1990 standar-

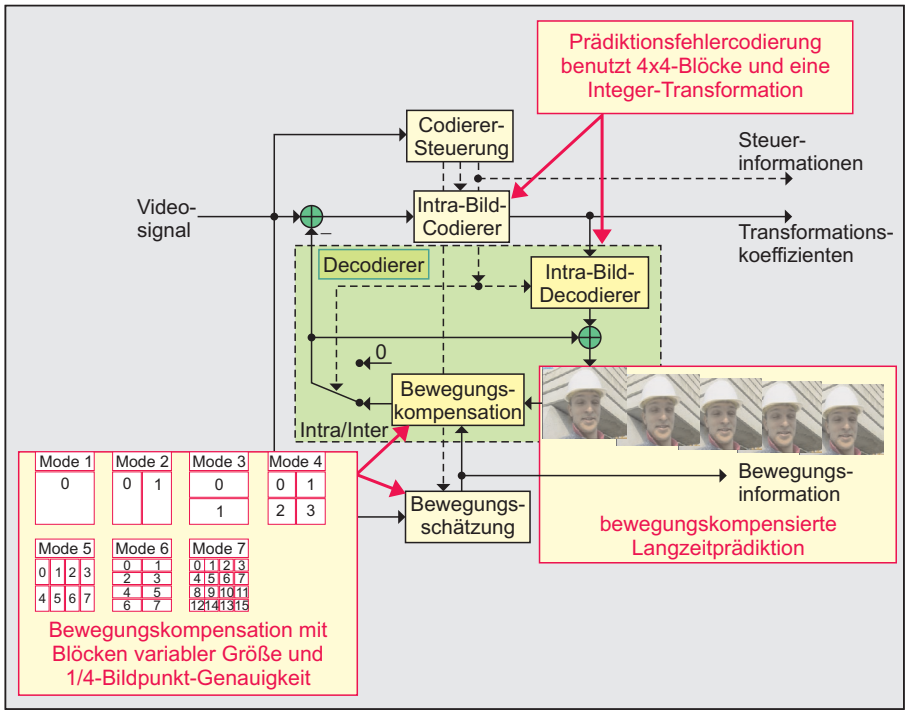

Bild 2. Erweiterungen des Hybridcodierers bei H.26L.

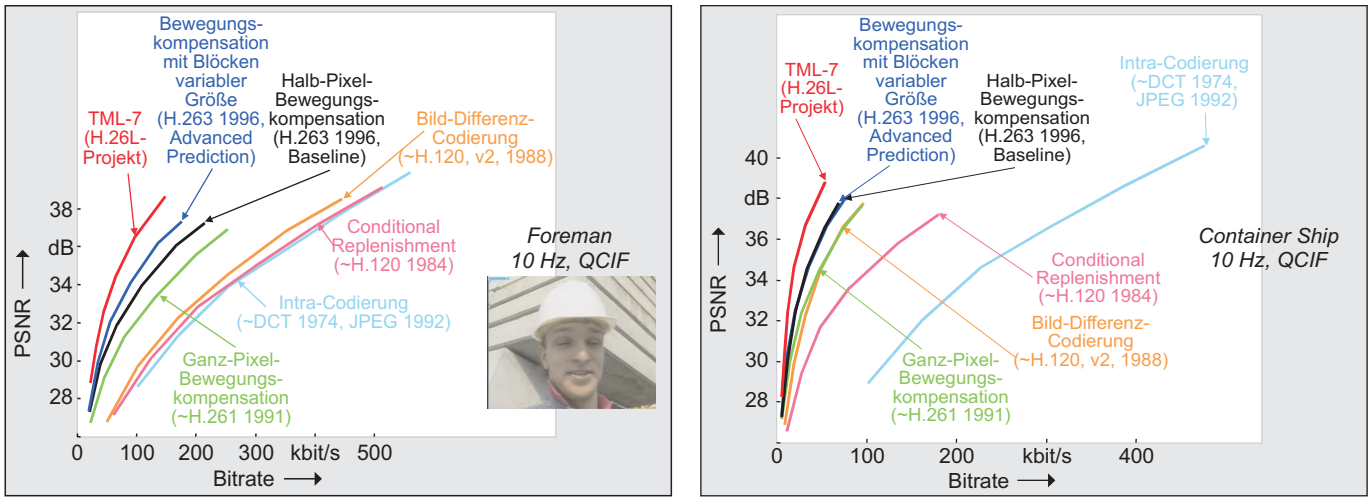

Bild 3. Effizienz verschiedener Codieroptionen und ITU-T-Videocodierstandards. a) links: Ergebnisse für die Videosequenz *Foreman,* b) rechts: Ergebnisse für *Container Ship*. Mit Hilfe des H.263-Standards werden die Fälle simuliert, die die Entwicklung in der Videocodierung seit H.120 veranschaulichen sollen. Schließlich wird das beste Ergebnis, das mit H.263 erreicht wird, mit dem gegenwärtigen Stand des H.26L-Projektes verglichen, das durch die TML-7 repräsentiert wird

disiert worden. Dabei ist anzumerken, dass in H.261 noch zusätzlich ein Schleifenfilter angewendet wird, der die Effizienz verbessert [5]. Hier ist das Resultat für H.263 unter Auslassung der Halb-Pixel-Positionen gezeigt.

5. Bewegungskompensation mit Halb-Pixel-Genauigkeit:
   Es werden Bewegungsvektoren, die auch auf interpolierte Zwischenpositionen des Referenzbilds fallen können, zugelassen. Bewegungskompensation mit Halb-Pixel-Genauigkeit ist zuerst in MPEG-1 im Jahr 1993 standardisiert worden. Hier ist das Resultat für den H.263-Coder im Baseline-Modus gezeigt [8].
6. Bewegungskompensation mit Blöcken variabler Größe:
   Es wird Bewegungskompensation mit Halb-Pixel-Genauigkeit nicht nur für Luminanzblöcke der Größe von 16×16 Bildpunkten durchgeführt, sondern eine Aufteilung der 16×16-Blöcke in Blöcke von 8×8 Bildpunkten ist ebenso möglich. Das gezeigte Resultat ist durch den von der ITU-T VCEG empfohlenen Video-Codierer TMN-10 erzielt worden.
7. TML-7-Codec:
   Die öffentlich zugängliche Software [18] für den TML-7-Codec [10] wird unter Verwendung aller Codieroptionen zur bewegungskompensierten Prädiktion mit maximal zehn möglichen Referenzbildern verwendet.

Die Experimente für die Fälle 1 bis 6 sind mit einem H.263-Coder durchgeführt worden, wobei verschiedene Möglichkeiten zur bewegungskompensierten Prädiktion zugelassen sind. Dabei ist in allen sechs Fällen die gleiche Syntax für die Intra-Codierung und die Prädiktionsfehlercodierung benutzt worden. Beim siebten Experiment ist hingegen der TML-7-Coder verwendet worden, der den gegenwärtigen Stand des H.26L-Projektes repräsentiert.

Für beide Videosequenzen liefert die Intra-Codierung (Fall 1) die schlechtesten Resultate. Um eine Rekonstruktionsqualität der Videosequenzen von 35 dB PSNR zu erreichen, ist bei „Foreman“ eine Bitrate von 300 kbit/s und bei „Container Ship“ eine Bitrate von 230 kbit/s notwendig. Das Hinzufügen des Skip-Modus bei der CR-Codierung (Fall 2) hat eine vollkommen unterschiedliche Wirkung auf die Resultate für die beiden Sequenzen. Während bei „Foreman“ fast keine Bitratenreduktion erreicht wird, ergibt sich bei „Container Ship“ eine Reduktion um 55%, und bei gleicher Rekonstruktionsqualität von 35 dB sind nur noch 100 kbit/s notwendig. Auch der nächste Schritt, das Hinzufügen der Prädiktionsfehlercodierung (Fall 3), bringt für „Foreman“ nur geringfügige Verbesserungen der Codiereffizienz, während bei „Container Ship“ eine weitere Halbierung der Bitrate möglich wird (das heißt, die 35-dB-Rekonstruktionsqualität können schon mit 50 kbit/s erreicht werden). Jedoch führt jedes weitere Hinzufügen von H.263-Optionen (Fälle 4 bis 6) kaum zu verbesserter Codiereffizienz bei der Videosequenz „Container Ship“. Hingegen führt das Hinzufügen der weiteren H.263-Optionen (Fälle 4 bis 6) bei der Videosequenz „Foreman“ zu erheblichen Verbesserungen. Dabei ergibt der Fall 4, die Bewegungskompensation mit Ganz-Pixel-Genauigkeit, eine Bitratenreduktion von 270 kbit/s (Fall 3) auf 170 kbit/s bei der betrachteten Rekonstruktionsqualität von 35 dB. Die weiteren Codieroptionen, die Bewegungskompensation mit Halb-Pixel-Genauigkeit (Fall 5) und die Bewegungskompensation mit Blöcken variabler Größe (Fall 6), führen für „Foreman“ zu weiteren signifikanten Bitratenreduktionen (für Fall 5 eine Bitrate von 140 kbit/s; für Fall 6 eine Bitrate von 110 kbit/s), um die Videosequenz „Foreman“ bei 35 dB Rekonstruktionsqualität decodieren zu können. Für beide Videosequenzen korrespondiert der Fall 6 zu dem mit H.263 besten erreichten Resultat und muss daher direkt mit dem für H.26L und mit dem TML-7-Coder (Fall 7) erzielten Resultat verglichen werden. Deutlich ist zu erkennen, dass für beide Videosequenzen eine Verbesserung der Codiereffizienz erreicht wird. Bei der Videosequenz „Foreman“ kann die Bitrate auf 70 kbit/s gesenkt werden, was einer Bitratenreduktion von 36% gegenüber H.263+ (Fall 6) entspricht. Bei der Videosequenz „Container Ship“ kann durch H.26L die Bitrate von 38 kbit/s (Fall 6: H.263+) auf 22 kbit/s (Fall 7: H.26L) gesenkt werden, was einer Bitratenreduktion von 42% entspricht. Es sei noch angemerkt, dass gegenüber der für Intra-Codierung notwendigen Bitrate ein Faktor 5 für „Foreman“ und etwa Faktor 10 für „Container Ship“ erreicht wird.

5. Ausblick

Die ITU-T hat seit Anfang der 80er Jahre eine Reihe von sehr erfolgreichen Standards für die Videocodierung spezifiziert. Dabei ist seit H.120, dem ersten internationalen Standard, der überhaupt für die Videocodierung gemacht wurde, eine stetige Verbesserung der Technik erreicht

worden. H.261 ist der zweite Standard der ITU-T für Videocodierung gewesen und hat dabei die Grundlagen für die heutigen Standards gelegt. Für MPEG/H.262 hat die ITU-T mit ISO/IEC ihre Kompetenzen gebündelt und so den bisher erfolgreichsten Standard für die Videocodierung gemacht. Sowohl das digitale Fernsehen als auch die DVD benutzen MPEG-2/H.262, um nur zwei der Anwendungen zu nennen. Der Nachfolger von MPEG-2/H.262 in der ITU-T, der H.263-Standard, hat die Grundlage für die erfolgreiche kommerzielle Nutzung der Videotelefone geschaffen.

H.263 ist für lange Zeit der effizienteste Standard zur Codierung von Videosequenzen bei niedrigen Bitraten gewesen. Jedoch kann auch die Effizienz von H.263 noch weiter verbessert werden, was eindrucksvoll durch den Stand des H.26L-Projekts nachgewiesen wird. Das im Sommer 1999 begonnene Standardisierungsprojekt befindet sich gerade auf halbem Wege zur Vervollständigung (Sommer/Herbst 2002). Dabei wird beim Entwurf eine Vielzahl hochgesteckter Ziele verfolgt. Unter anderem soll mit einer einfachen und klaren Syntax sowohl eine bessere Anpassung an paketvermittelte Übertragungsmedien als auch ein Faktor zwei an Bitratenreduktion gegenüber H.263 erreicht werden. Bei letzterem Ziel ist man schon ein gutes Stück vorangekommen. Simulationen, die hier exemplarisch für zwei Testsequenzen aufgeführt worden sind, zeigen, dass gegenwärtig eine Bitratenreduktion von etwa 40% erreicht wird. Die verbleibende Zeit wird von der ITU-T VCEG genutzt werden, um die selbstgesteckten Ziele schließlich zu erreichen.

Schrifttum

[1] ISO/IEC JTC1 und ITU-T (früher CCITT): Digital Compression and Coding of Continuous-Tone Still Images. ISO/IEC 10918-1 – ITU-T-Empfehlung T.81 (JPEG), 9/1992

[2] Pennebaker, W. B.; Mitchell, J. L.: JPEG: Still Image Data Compression Standard. Van Nostrand Reinhold, New York, USA, 1993

[3] ITU-T: Codec for Videoconferencing Using Primary Digital Group Transmission. ITU-T-Empfehlung H.120, 1. Version: 1984, 2. Version: 1988

[4] Jayant, N. S.; Noll, P.: Digital Coding of Waveforms. Prentice Hall (1984), Englewood Cliffs, HJ, USA

[5] ITU-T: Video Codec for Audiovisual Services at px64 kbit/s. ITU-T-Empfehlung H.261, 1. Version: 11/1990, 2. Version: 3/1993

[6] ISO/IEC JTC1: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s – Part 2: Video. ISO/IEC 11172-2 (MPEG-1), 3/1993

[7] ISO/IEC JTC1: Generic Coding of Moving Pictures and Associated Audio Information – Part 2: Video. ITU-T-Empfehlung H.262 – ISO/IEC 13818-2 (MPEG-2), 11/1994

[8] ITU-T: Video Coding for Low Bit-Rate Communication. ITU-T-Empfehlung H.263, 1. Version: 11/1995, 2. Version: 1/1998, 3. Version: 11/2000

[9] ISO/IEC JTC 1/SC 29 14496-2 (MPEG-4 Visual, version 2)

[10] H.26L Test Model Long Term Number 7 (TML-7), Doc. VCEG-M81, 4/2001

[11] Wiegand, Th.; Zhang, X.; Girod, B.: Long-Term Memory Motion-Compensated Prediction. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9 (1999), No. 1 (Feb.), pp. 70 to 84

[12] Nokia Research Center: MVC Video Codec - Proposal for H.26L. Doc. Q15-F-24, Seoul, Korea, 11/1998

[13] Heising, G.; Marpe, D.; Cycon, H. L.; Petukhov, A. P.: HHI-proposal for ITU-T H.26L. ITU-T SG 16, Doc. Q15-G-36, Berlin, 2/1999

[14] Bjontegaard, G.: Enhancement of the Telenor proposal for H.26L. Doc. Q15-G-25, Berlin, 8/1999

[15] Wedi, Th.: 1/8-pel Displacement Vector Resolution for TML-6. Doc. VCEG-M45, Austin, Texas/USA, 4/2001

[16] Wien, M.; Dahlhoff, A.: Adaptive Block Transforms. Doc. VCEG-M62, Austin, Texas/USA, 4/2001

[17] Marpe, D.; Blättermann, G.; Heising, G.; Wiegand, T.: Further Results for CABAC entropy coding scheme. Doc. VCEG-M59, Austin, Texas/USA, 4/2001

[18] Die Software und die ITU-T VCEG-Dokumente können durch anonymous-ftp zu standard.pictel.com/video-site heruntergeladen werden