

Karsten Sühling, Dr. Heiko Schwarz, Dr. Thomas Wiegand

Effizienter kodieren

Details zum kommenden Videostandard H.264/AVC

In jüngster Zeit tauchen im Zusammenhang mit Videokodierung immer häufiger die Begriffe 'H.264', 'Advanced Video Coding (AVC)' oder 'MPEG-4 Part 10' auf. Sie alle bezeichnen einen noch zu verabschiedenden Videokodierstandard, der nochmals effizientere Kompression verspricht als sämtliche Vorgänger, sodass er sich für die verschiedensten Einsatzszenarien eignen sollte – vom UMTS-Streaming bis hin zur HD-DVD. Welche Tricks und Algorithmen ihm dazu verhelfen sollen, beleuchtet der folgende Artikel.



Das Jahr 2002 stand ganz im Zeichen von MPEG-4, auch wenn die Einführung zunächst durch Lizenzquerelen verzögert wurde. Obwohl MPEG-4 sehr effiziente Videokodierung ermöglicht, ist damit noch immer nicht das Ende der Fahnenstange erreicht. Bereits im Jahr 1997 stieß die Video Coding Experts Group (VCEG) der ITU-T unter dem Namen H.26L ein ehrgeiziges Projekt an, das deutlich effizientere Videokodierung verspricht als alle aktuellen Verfahren. Im Dezember 2001 vereinten VCEG und die ISO/IEC Moving Picture Experts Group (MPEG) ihre Kräfte,

um das Design des nun als H.264/AVC bezeichneten Videokodierverfahrens fertig zu stellen. Inzwischen ist dies geschehen und H.264/AVC wird vermutlich noch im Frühjahr mit identischem Wortlaut von der ITU-T als 'Recommendation H.264' und von der ISO/IEC unter dem Namen 'MPEG-4 Part 10/Advanced Video Coding (AVC)' als internationaler Standard verabschiedet [1].

Wegen der gegenüber allen anderen Videokodierstandards deutlich erhöhten Kodiereffizienz setzt insbesondere die Telekommunikationsbranche bei Videoübertragung (Videokonfe-

renzen, Streaming) über UMTS große Hoffnungen in H.264/AVC, aber auch bei der TV-Branche ist das Interesse groß. Reichte etwa für 'normales' Digital-TV (DVB) die Kodiereffizienz von MPEG-2 aus, das die Bitrate des unkomprimierten Videosignals von 166 MBit/s (25 Bilder/Ss, bis zu 720 × 576 Bildpunkte, 16 Bit Farbtiefe) auf im Mittel zwei bis fünf MBit/s reduziert, stellt das auch für Europa angekündigte High Definition TV (siehe auch Artikel auf S. 216) mit unkomprimierten Datenströmen von bis zu 885 MBit/s neue Herausforderungen. HDTV blockiert selbst bei MPEG-2-Kompression vier bis fünf herkömmliche digitale TV-Kanäle. Ähnliche Probleme bereitet die Archivierung: Kommende HD-

DVDs müssten entweder den fünffachen Speicherplatz bieten (Stichwort: Blu-Ray Disc, Advanced Optical Disc) oder eben ein deutlich effizienteres Kodierverfahren einsetzen.

Im Folgenden geben wir einen Einblick in die Technik, die diese Effizienzsteigerung erreichen soll, ziehen einen Vergleich mit der Kodiereffizienz anderer Verfahren und stellen einige Betrachtungen zur Komplexität an.

Schichtensystem

Wie auch alle vorherigen Standards von VCEG und MPEG definiert H.264/AVC nur die Syntax und Semantik des Bitstroms sowie den Dekodierprozess. Die Umsetzung des Kodierers samt seiner Kodierstrategie und Opti-

mierungen bleibt den Herstellern und Programmierern der Applikation überlassen – wichtig ist nur, dass der erzeugte Bitstrom standardkonform ist.

H.264/AVC unterscheidet zwischen dem Video Coding Layer (VCL), der die eigentlichen kodierten Videodaten verarbeitet, und dem Network Abstraction Layer (NAL), der entsprechend dem eingesetzten Übertragungsverfahren oder Speichermedium die Daten des VCL formatiert und mit Zusatzinformationen versieht. Das Verfahren kodiert Video im YCbCr-Format (mit 4:2:0 Chroma-Unterabtastung, s. Kasten 'Begrifflichkeiten') und verarbeitet sowohl progressive als auch zeilenverschränkte Bilder ('Interlaced'). Halbbilder kodiert H.264/AVC entweder als zwei unabhängige Bilder oder aber zusammen als ein Vollbild.

Wie bei den Vorläufern MPEG-1/-2/-4, H.261 oder H.263 setzt der VCL von H.264/AVC auf blockbasierte, hybride Videokodierung. Der Algorithmus nutzt örtliche und zeitliche Abhängigkeiten des Videosignals zur Vorhersage ('Prädiktion') sowie eine Transformationskodierung für die Übertragung des Restfehlers aus.

Während der VCL für eine möglichst effiziente Kodierung der Videodaten sorgt, ist der NAL für die geeignete Formatierung zur Übermittlung oder Speicherung der Daten zuständig. Dazu verpackt der NAL die

VCL-Daten in 'NAL Units', deren Format für paketorientierte (etwa übers Internet) und byteorientierte (etwa über Satellit, Kabel oder ISDN) Übertragung identisch ist. In letzterem Fall werden die NAL Units mit einem Startcode-Präfix versehen, um eine Abtrennung der einzelnen NAL Units im Bytestrom zu ermöglichen.

Wie die meisten anderen Videokompressionsverfahren unterteilt H.264/AVC das zu kodierende Bild zunächst in Blöcke. Das erste Bild der Videosequenz oder Bilder, die als Random Access Point (etwa zum 'Spulen') in der Videosequenz dienen, wird 'intra'-kodiert, das heißt, man verwendet nur im zu kodierenden Bild selbst enthaltene Informationen zur Kodierung. Für jeden Bildpunkt eines Blocks im Intra-Bild berechnet man in Abhängigkeit von bereits übertragenen, örtlich benachbarten Blöcken eine Prädiktion. Welche Bildpunkte und welches Verfahren dazu benutzt werden, entscheidet der Kodierer. Er überträgt seine Entscheidungen zum Dekodierer mit Hilfe von Seiteninformationen, sodass diese Entscheidung dort eindeutig nachvollzogen werden kann.

Für alle anderen Bilder benutzt man typischerweise eine 'Inter'-Kodierung. Sie verwendet zur Prädiktion der Bildpunkte die Daten der zuvor übertragenen Bilder – auch als 'Bewegungs-

kompensation' bezeichnet. Für die Inter-Prädiktion führt der Kodierer eine Bewegungsschätzung durch, bei der ein Referenzbild und eine Bild-zu-Bild-Verschiebung ausgewählt werden, die er ebenfalls als Seiteninformation überträgt. Letztere legt fest, wie die Bildpunkte eines Blocks in das Prädiktionsbild kopiert – das heißt bewegungskompensiert – werden.

Sowohl Intra- als auch Inter-Kodierung reduzieren den Restfehler, der sich aus der Differenz zwischen Original und Prädiktion ergibt. Der Restfehler wird zur weiteren Reduktion einer Transformationskodierung unterzogen. Dabei werden die Bildpunkte in Transformationskoeffizienten umgewandelt, welche anschließend skaliert und quantisiert werden. Am Schluss folgt eine Entropiekodierung und die Übertragung zusammen mit den Seiteninformationen (für Intra- oder Inter-Prädiktion).

Die Quantisierung stellt den eigentlichen verlustbehafteten Schritt bei der Kodierung dar: Sie reduziert den Informationsgehalt des Bildes. Typischerweise bedeutet eine Reduktion des Informationsgehalts eine Verringerung der sichtbaren Details. Da man dabei jedoch die Eigenschaften des menschlichen visuellen Systems berücksichtigt, nimmt man auch eine stärkere Quantisierung kaum als störend wahr. Dieses Prinzip ist der

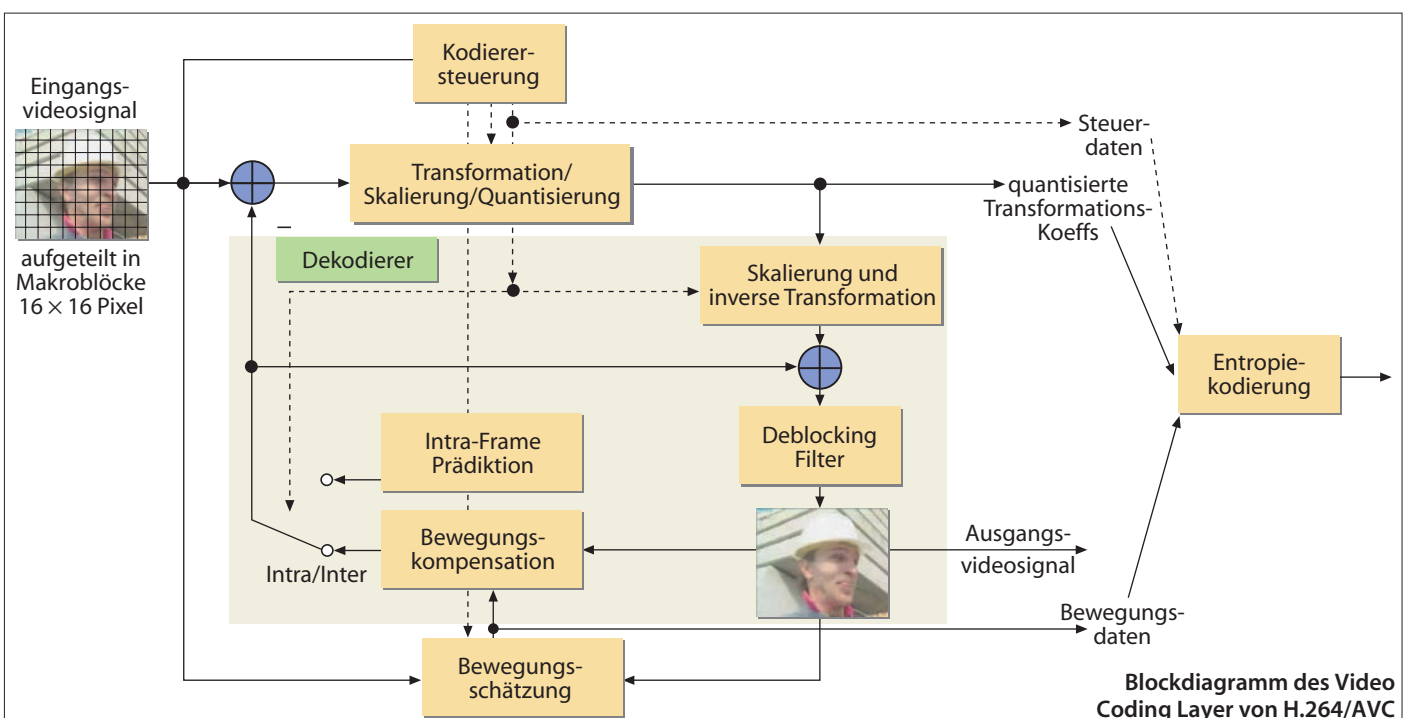
Grund für die im Vergleich zur Originalbitrate sehr geringen Übertragungsbitraten.

Um die Bewegungsschätzung korrekt durchführen zu können, müssen am Kodierer dieselben Referenzbilder vorhanden sein wie am Dekodierer. Würde man zur Erzeugung des Prädiktionsbildes die Originalbilder benutzen, entstünde ein falsches Restfehlerbild und die Signale von Kodierer und Dekodierer würden auf Dauer auseinander driften. Daher enthält der Kodierer auch immer einen kompletten Dekodierer. Um den kodierten Restfehler zu rekonstruieren, führt der Kodierer daher die inverse Skalierung und Transformation durch wie der Dekodierer. Danach wird das Restfehler-signal zum Prädiktionsbild addiert und ein Deblocking-Filter darauf angewendet. Daraus erhält der Kodierer das dekodierte Videosignal.

Die Beschreibung des VCL gilt (bis auf einige wenige Ausnahmen) für alle existierenden Videokompressionsverfahren. Die eigentlichen Neuerungen von H.264/AVC beschreiben wir im Folgenden.

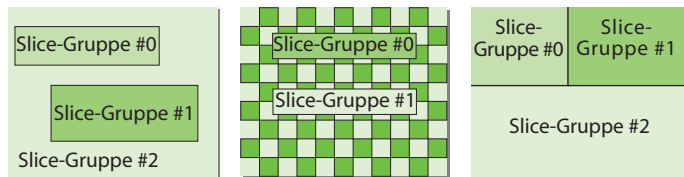
Block- und scheibenweise

Jedes Einzelbild wird in gleich große quadratische 'Makroblöcke' unterteilt. Diese sind



für die Luminanzkomponente 16×16 Bildpunkte und für die beiden Chrominanzkomponenten jeweils 8×8 Bildpunkte groß. Für alle Bildpunkte führt man entweder eine örtliche oder zeitliche Vorhersage durch und überträgt das aus Originalsignal und Prädiktionssignal entstehende Restfehlersignal mittels Transformationskodierung. Dazu wird jede Farbkomponente weiter in Blöcke unterteilt. Diese Blöcke unterzieht man einer Block-Transformation, quantisiert wie oben beschrieben die Transformationskoeffizienten und überträgt sie mittels einer Entropiekodierung.

Makroblöcke organisiert H.264/AVC in so genannten 'Slices': Darunter versteht man einen Bildbereich (Sequenz von Makroblöcken), der sich unabhängig von anderen Bildbereichen dekodieren lässt. Die Daten eines Slice werden in einer NAL Unit (und damit typischerweise in einem IP-Paket) übertragen.



Muster für Slice-Gruppen: Die Makroblöcke mit den gleichen Farben gehören zu einer Slice-Gruppe und werden getrennt übertragen.

H.264/AVC definiert fünf Slice-Typen: Der einfachste ist das I-Slice (wobei 'I' für Intra-Prädiktion steht), der alle Makroblöcke kodiert, ohne Informationen aus anderen Bildern der Sequenz dafür zu benutzen. Soll aus anderen Bildern prädiziert werden, kommen P- und B-Slices zum Einsatz ('P' für predictive und 'B' für bi-predictive). Die beiden übrigen Slice-Typen SP ('Switching P') und SI ('Switching I') ermöglichen ein effizientes Umschalten zwischen Bitströmen, die bei verschiedenen Bitraten kodiert wurden.

Zur effizienten Verschleierung von Übertragungsfehlern am Dekodierer setzt H.264/AVC 'Flexible Macroblock Ordering' (FMO) ein. Damit definiert man ein Muster, das Makroblöcke einer oder mehreren Slice-Gruppen zuordnet. Jede dieser Slice-Gruppen wird in getrennten NAL

Units übertragen. Beim Verlust einer Slice-Gruppe kann H.264/AVC die Informationen aus angrenzenden, korrekt empfangenen Slice-Gruppen benutzen, um den Übertragungsfehler (effizienter) zu verschleiern. Die erlaubten Muster reichen dabei von regelmäßigen Anordnungen (etwa einem Schachbrettmuster) über rechteckige Bildbereiche bis hin zu völlig frei definierbaren Formen.

Jeder Makroblock wird in einem Kodiertyp übertragen, der jeweils vom zugehörigen Slice-Typ abhängt. Für alle Slice-Typen stehen verschiedene Intra-Kodiertypen zur Verfügung, die jedem Block einen von mehreren Intra-Prädiktoren zuordnen. Neben einer Gleichsignal-Prädiktion, bei der alle Bildpunkte eines Blocks mit demselben Wert vorhergesagt werden, definiert H.264/AVC verschiedene Richtungsprädiktoren, die Ähnlichkeiten zu Nachbarbildpunkten ausnut-

zen. Sind etwa diagonale Streifen im Bild, werden diese durch die Prädiktoren für den zu kodierenden Bildinhalt fortgesetzt.

Zusätzlich zu den Intra-Kodiertypen setzt H.264/AVC bei Bedarf verschiedene Inter-Kodiertypen in P- oder B-Slices ein. Beim Inter-Typ können die Makroblöcke weiter in Blöcke von $M \times N$ Bildpunkten unterteilt werden. Für die Inter-Prädiktion jedes $M \times N$ -Blocks wird mit Hilfe von Seiteninformationen das benutzte Referenzbild und eine Bild-zu-Bild-Verschiebung des Blocks übertragen.

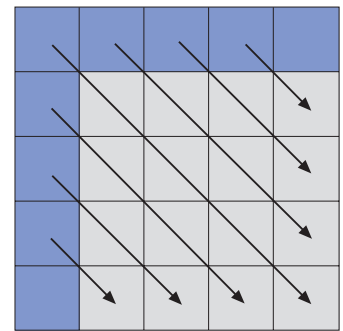
Neu bei H.264/AVC ist die so genannte bewegungskompensierte Langzeitprädiktion. Dabei wählt es die Referenz für die Bewegungskompensation aus mehreren Bildern aus – bei älteren Kompressionsverfahren konnte nur ein Bild verwendet werden.

Dazu müssen sowohl Kodierer als auch Dekodierer die Referenzbilder, die für die Inter-Prädiktion benötigt werden, in einem Referenzbildpuffer speichern. Dieses Verfahren führt zu einer höheren Effizienz der Bewegungskompensation – allerdings einhergehend mit erhöhtem Speicherplatzbedarf, was dank rapide fallender Speicherpreise für die Kodiererimplementierung nicht mehr so stark ins Gewicht fällt.

Nicht selten kommt es vor, dass sich Objekte in aufeinander folgenden Bildern nur geringfügig bewegen oder die Bewegungen kein Vielfaches des Abstandes der Bildpunkte sind. Um auch solche Bewegungen zur Prädiktion nutzen zu können, interpoliert H.264/AVC die Positionen zwischen den Bildpunkten und führt die Bewegungssuche anhand dieser Ergebnisse durch. Die Bild-zu-Bild-Verschiebungen werden dann – ebenso wie beim 'Advanced Simple Profile' von MPEG-4 – mit einer Genauigkeit von einem Viertel des Bildpunktabstandes übertragen (auch 'Quarter Sample' genannt). H.264/AVC erlaubt übrigens auch Bild-zu-Bild-Verschiebungen, die auf Bereiche außerhalb des Bildes verweisen. In diesem Fall erweitert man das Bild durch Wiederholung der Bildpunkte am Rand.

Außer den genannten Inter-Kodiertypen gibt es in P-Slices auch den so genannten Skip-Kodiertyp, bei dem einfach angenommen wird, dass der Bildinhalt des damit kodierten Makroblocks unverändert bleibt und somit direkt aus dem zuvor dekodierten Bild kopiert werden kann.

H.264/AVC erweitert das Konzept der B-Slices gegenüber MPEG-2 und MPEG-4, indem es auch Bilder mit B-Slices als Referenz für die Bewegungskompensation nutzt. Der Unterschied zwischen P- und B-Slices schmilzt damit auf die Art der Inter-Prädiktion zusammen. So kann das Prädiktionssignal in B-Slices aus dem gewichteten Mit-



Beispiel einer Intra-Prädiktion für diagonale Streifen: Die grauen Bildpunkte berechnet man aus den zuvor dekodierten blauen Bildpunkten.

tel zweier unterschiedlicher Prädiktionswerte gebildet werden, weshalb sie zwei verschiedene Referenzbildpuffer benutzen. Der Kodierer steuert dabei, welches Bild in welchen Puffer gelangt.

B-Slices verfügen über dieselben Makroblock-Unterteilungen wie P-Slices. Die Kodierung der Bewegungsvektoren entspricht der von P-Slices mit entsprechenden Modifikationen, da benachbarte Blöcke unterschiedliche Prädiktoren besitzen können.

Ade, DCT!

Ähnlich wie andere Kompressionsverfahren nutzt H.264/AVC eine Transformationskodierung zur Übertragung des Restfehlers. Der Unterschied ist jedoch, dass die Transformation auf 4×4 -Blöcken stattfindet und H.264/AVC anstelle einer Diskreten Cosinus-Transformation (DCT) eine separierbare Integertransformation einsetzt, die jedoch ähnliche Eigenschaften wie eine 4×4 -DCT besitzt. Die inverse Transformation ist exakt durch Operationen mit Integern definiert und vermeidet so Abweichungen durch Berechnungsungenauigkeiten.

Für die Quantisierung der Koeffizienten setzt H.264/AVC eine lineare Skala ein. Für jeden Ma-

	16x16	16x8	8x16	8x8
M Typen	0	0 1	0 1	0 1 2 3
8x8 Typen	8x8	8x4	4x8	4x4
	0	0 1	0 1	0 1 2 3

Blocktypen für die Bewegungskompensation bei H.264/AVC: Makroblöcke (oben) und 8×8 -Blöcke (unten)

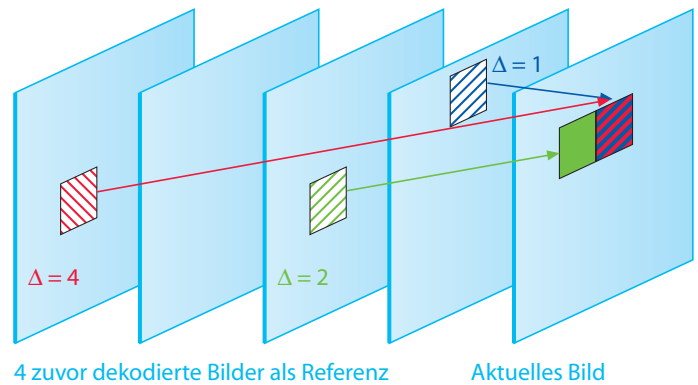
kroblock wird einer von 52 möglichen Quantisierern durch den Quantisierungsparameter (QP) ausgewählt. Die Quantisierer sind so festgelegt, dass eine Erhöhung des QP um eins einer um circa zwölf Prozent niedrigeren Bitrate entspricht. Die Transformationskoeffizienten werden in einem Zickzackmuster gescannt und mittels Entropiekodierung übertragen. Es ist übrigens möglich, alle Transformationen in H.264/AVC nur mittels Additionen und Shift-Operationen auf 16-Bit-Integern zu implementieren, was dem Einsatz in günstigen DSPs entgegenkommt.

H.264/AVC unterstützt zwei Methoden der Entropiekodierung. Beim Variable Length Coding (VLC) wird eine einzige unendliche Codewort-Tabelle für alle Syntaxelemente außer den Transformationskoeffizienten benutzt. Somit wird, anstatt eine eigene VLC-Tabelle für jedes einzelne Syntaxelement zu entwerfen, nur die Abbildung der Werte auf die Codewort-Tabelle anhand von statis-

tischen Eigenschaften angepasst. Als Codewort-Tabelle verwendet H.264/AVC einen Exp-Golomb-Code, der sehr einfache und regelmäßige Dekodiereigenschaften besitzt – mehr dazu in [2].

Zur Übertragung der quantisierten Transformationskoeffizienten setzt H.264/AVC eine aufwendigere Methode namens 'Context-Adaptive Variable Length Coding' (CAVLC) ein. CAVLC schaltet in Abhängigkeit von bereits übertragenen Daten zwischen verschiedenen VLC-Tabellen für das zu kodierende Element um. Da die VLC-Tabellen anhand von Statistiken sorgfältig konstruiert wurden, erzielt man so eine deutliche Leistungssteigerung gegenüber dem Exp-Golomb-Code.

Die Effizienz der Entropiekodierung erhöht sich nochmals, wenn man Context-Adaptive Binary Arithmetic Coding (CABAC) benutzt. Einerseits ermöglicht der Einsatz arithmetischer Kodierung die Zuordnung gebrochener Bit-Anzahlen für



Bei der bewegungskompensierten Langzeitprädiktion überträgt H.264/AVC für jeden Teilblock zusätzlich zur Bild-zu-Bild-Verschiebung einen Bildreferenzparameter.

jedes Symbol eines Alphabets, was besonders für Wahrscheinlichkeiten deutlich größer als 0,5 von Vorteil ist. Andererseits bietet die Anwendung adaptiver Codes die Anpassung an veränderliche Symbolstatistiken. Eine weitere wichtige Eigenschaft von CABAC ist die Kontextmodellierung: Dabei zieht man die Statistiken bereits kodierter Daten zur

Schätzung bedingter Wahrscheinlichkeiten heran. Letztere benutzt man zum Umschalten zwischen mehreren Wahrscheinlichkeitsmodellen. Der Kern der arithmetischen Kodiermaschine und die dazugehörige Wahrscheinlichkeitsschätzung sind in H.264/AVC als Methoden mit niedriger Komplexität implementiert, die nur Tabellen- und

Begrifflichkeiten

Chroma-Unterabtastung der Farbkomponenten eines Bildes. Bilder, die im RGB-Farbraum repräsentiert sind, können durch eine Transformation in den YCbCr-Farbraum und anschließende Unterabtastung der Chrominanzsignale Cb und Cr für das menschliche Auge im Allgemeinen verlustfrei wiedergegeben werden. Übliche Unterabtastungen sind etwa 4:2:2, bei der das Chrominanzsignal in horizontaler Richtung in der halben Auflösung der Luminanzkomponente vorliegt und 4:2:0, wobei die Auflösung sowohl in horizontaler als auch in vertikaler Richtung halbiert wird.

Chrominanz: Farbkomponenten (Cb und Cr) des YCbCr Signals

Common Intermediate Format (CIF): Bildformat mit 352×288 Bildpunkten

Entropiekodierung: Methode der Redundanzreduktion, bei der häufig auftretende Symbole mit wenigen Bits und seltener auftretende Symbole mit vielen Bits übertragen werden und sich somit im Mittel eine verringerte Anzahl von Übertragungsbits ergibt

Interlaced (zeilenverschränkt): Bilddarstellungsweise bei TV und Video. Das Bild wird dabei nacheinander als zwei Halbbilder dargestellt. Das erste Halbbild enthält dabei die Zeilen mit geraden Zeilennummern, das zweite Halbbild die Zeilen mit ungeraden Zeilennummern. Die Halbbilder sind dabei zeitlich um die halbe Bildwiederholrate versetzt.

Joint Video Team (JVT): Zusammenschluss von VCEG und MPEG mit dem Ziel, H.264/AVC fertig zu stellen

Luminanz: Helligkeitskomponente (Y) des YCbCr-Signals

Moving Picture Experts Group (MPEG – ISO/IEC JTC 1/SC 29/WG11): Expertengruppe der ISO/IEC JTC 1, die für die Entwicklung der Standards MPEG-1, MPEG-2 (H.262), MPEG-4, MPEG-7 und MPEG-21 verantwortlich ist

Network Abstraction Layer (NAL): Schicht in H.264/AVC, die die Daten des VCL entsprechend dem eingesetzten Übertragungsverfahren oder Speichermedium formatiert und mit Zusatzinformationen versieht

Prädiktion: Vorhersage von Werten zu kodierender Bildpunkte anhand bereits dekodierter Bildpunkte

Progressiv: Bildformat, bei dem zu einem Zeitpunkt das gesamte Bild vorliegt, während im Unterschied dazu beim Interlaced-For-

mat die beiden Halbbilder zu unterschiedlichen Zeitpunkten vorliegen

Peak Signal to Noise Ratio (PSNR): Fehlermaß für die Abweichung zweier Bilder in Dezibel (dB). Ein höheres PSNR entspricht im Allgemeinen einer höheren Bildqualität. Da es sich bei diesem Maß unter anderem um eine Durchschnittsbildung handelt, muss das Ergebnis nicht immer mit dem visuellen Eindruck übereinstimmen, welcher noch durch viele andere Faktoren wie die Art des Videos, des Displays et cetera beeinflusst wird.

Quarter CIF (QCIF): Bildformat mit 176×144 Bildpunkten

Quantisierung: Methode zur Reduzierung des Informationsgehalts von Zufallssignalen. Dabei geschieht eine Zuordnung von Ersatzwerten zu den Signalamplituden. Eine Reduzierung des Informationsgehalts wird erreicht, wenn die Ersatzwerte weniger Bits zur Übertragung als die Signalamplituden selbst erfordern. Die Abweichung zwischen Amplitude und Ersatzwert wird als Quantisierungsfehler bezeichnet.

Random Access Point: Zeitpunkt im Video-Bitstrom, an dem mit der Dekodierung begonnen werden kann. Random Access Points sind insbesondere bei der Fernsehübertragung wichtig, um das Schalten zwischen verschiedenen Fernsehkanälen zu ermöglichen.

Transformation: Umwandlung von N Bildpunkten in N Transformationskoeffizienten. Bei H.264/AVC ist $N = 16$, bei allen anderen Standards ist $N = 64$. Die inverse Transformation wird dazu benutzt, um die quantisierten Transformationskoeffizienten in Bildpunkte zurückzuwandeln.

Video Coding Experts Group (VCEG – ITU-T SG16 Q.6): Video-Expertengruppe der ITU-T, die die Videokodierstandards H.261, H.262 (MPEG-2), H.263 und H.264 entwickelt hat

Video Coding Layer (VCL): Schicht von H.264/AVC, in der die Verarbeitung der kodierten Videodaten durchgeführt wird

YCbCr: Farbraum, der aus einer Helligkeitskomponente (Y – Luminanz) und zwei Farbkomponenten (Cb und Cr – Chrominanz) besteht. Im Unterschied zu anderen Farbräumen ermöglicht YCbCr die Trennung von Farb- und Helligkeitsinformation und deren getrennter Kodierung (siehe auch Chroma-Unterabtastung).

Shift-Operationen verwenden. Im Vergleich zu CAVLC bietet CABAC im Durchschnitt zehn bis 15 Prozent Bitratenreduktion bei der Kodierung von TV-Bildern bei gleicher Qualität.

Walk around the blocks

Steht für die Kodierung eines Makroblocks aufgrund der vorgegebenen Datenrate ein zu geringes Bitreservoir zur Verfügung, treten bei blockbasierten Kodierverfahren mehr oder minder starke 'Blockartefakte' zu Tage, weil die Blöcke stärker quantisiert und daher die Blockränder ungenauer rekonstruiert werden.

Weil diese Artefakte gemeinhin als gravierendste Störung angesehen werden, begegnet

H.264/AVC diesem Problem mit einigem Aufwand. Es definiert ein 'Schleifen-Filter' innerhalb der Kodierschleife, dessen Filterstärke adaptiv gesteuert wird. Dadurch lassen sich Blockartefakte relativ gut kaschieren, ohne die Schärfe des Bildinhalts stark zu beeinflussen – die subjektiv empfundene Videoqualität steigt deutlich. Ein weiterer positiver Nebeneffekt des Filters: Es reduziert die Bitrate um etwa

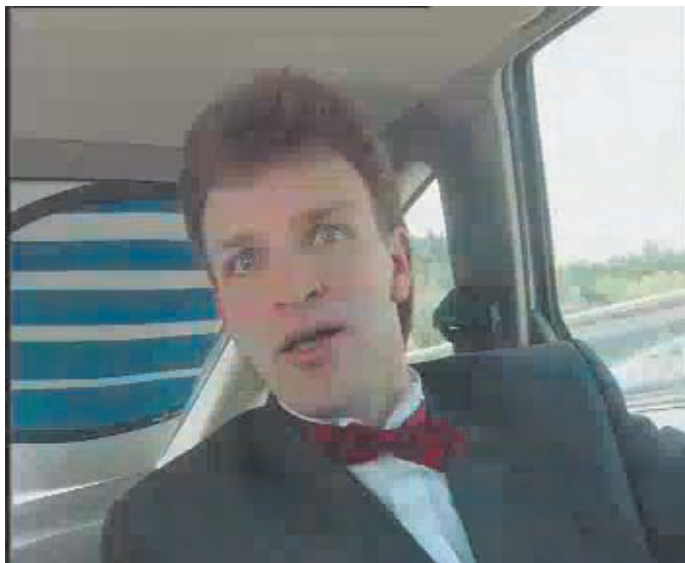
fünf bis zehn Prozent bei derselben objektiven Qualität wie ungefiltertes Video.

Wie bereits erwähnt, deckt H.264/AVC ein breites Spektrum unterschiedlicher Anwendungsgebiete ab – von Streaming bei niedrigsten Bitraten bis hin zu Video in HDTV-Qualität. Zur Klassifikation bestimmter Parameter und Anwendungsbereiche definiert man Profile und Level, die die Kommunikation zwischen

verschiedenen Anwendungen des H.264/AVC-Standards mit ähnlichen funktionalen Anforderungen sicherstellen sollen. 'Profile' definieren etwa eine Auswahl an Kodiermethoden, die zur Erzeugung eines standardkonformen Bitstroms eingesetzt werden dürfen, 'Level' legen Grenzen für verschiedene Parameter des Bitstroms und damit für die Hardware des Dekodierers fest.

Alle Dekodierer, die zu einem entsprechenden Profil konform sind, müssen alle darin enthaltenen Funktionen unterstützen. Kodierer müssen hingegen nicht alle Komponenten des Profils einsetzen, aber natürlich einen standardkonformen Bitstrom erzeugen. Für H.264/AVC sind drei Profile definiert:

	Durchschnittlicher Bitratengewinn gegenüber		
Kodierer	MPEG-4 ASP	H.263 HLP	MPEG-2
H.264/AVC	38,62 %	48,80 %	64,46 %
MPEG-4 ASP	–	16,65 %	42,95 %
H.263 HLP	–	–	30,61 %



Effizienz des Schleifen-Filters ('Deblocking') für stark komprimierte Bilder: links ohne Filter, rechts mit Filter

Baseline unterstützt alle Kodieroptionen von H.264/AVC, außer den folgenden zwei Gruppen von Kodiermethoden:

- Gruppe 1: B-Slices, gewichtete Prädiktion, CABAC, Kodiermethoden für zeilenverschränkte Bildsignale
- Gruppe 2: SP und SI Slices

Main unterstützt zusätzlich Gruppe 1, jedoch nicht Flexible Macroblock Ordering, welches in Baseline enthalten ist.

Extended unterstützt zusätzlich zu Baseline die Gruppen 1 und 2 außer CABAC.

Die Level-Definitionen sind jeweils identisch, wobei eine spezielle Implementierung für ver-

schiedene Profile unterschiedliche Level unterstützen kann. Es gibt elf Level, die die Anforderungen an den Dekodierer, angefangen von QCIF-Bildgrößen bis hin zu Bildgrößen, für digitales Kino definieren.

Schlagabtausch

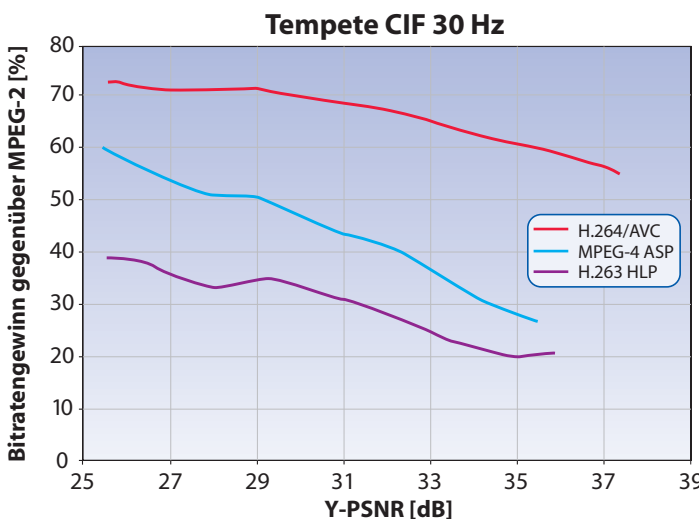
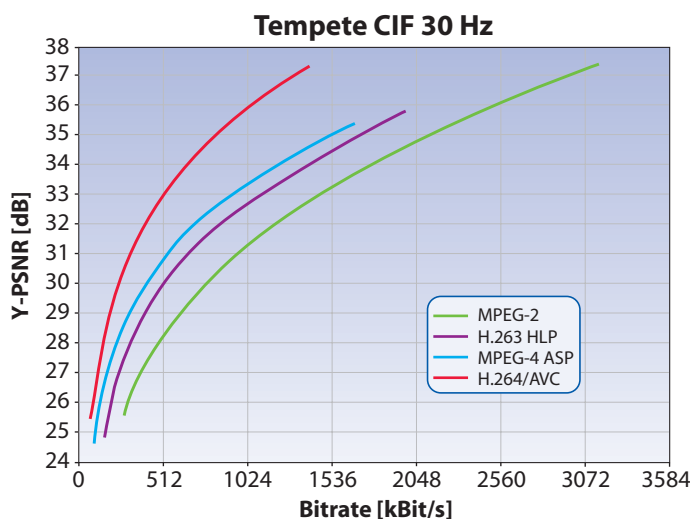
Nun stellt sich natürlich die Frage, ob sich der ganze Aufwand überhaupt lohnt. Um die Kodiereffizienz von H.264/AVC einordnen zu können, haben wir (die Autoren) es mit seinen Vorgängern MPEG-2 Visual [3], H.263++ [4] und MPEG-4 Visual [5] für verschiedene QCIF- und CIF-Sequenzen verglichen (siehe Glossar). Dazu wurden die

Abweichungen der mit dem jeweiligen Codec verarbeiteten Videos zum Quellmaterial Bild für Bild verglichen (Peak Signal to Noise Ratio, siehe Glossar) – ein höherer Wert entspricht gemeinhin einer besseren Bildqualität.

Für den Vergleich erzeugten wir MPEG-2-Videoströme im bekannten MP@ML (Main Profile at Main Level, wie es bei DVB oder Video-DVD zum Einsatz kommt), für H.263++ setzten wir das High Latency Profile (HLP) ein, bei MPEG-4 Visual das Advanced Simple Profile (ASP) – das Flaggschiff der MPEG-4-Videokodierung. Zusätzlich wurden die empfohlenen Nachverarbeitungsschritte genutzt.

Für den Herausforderer H.264/AVC kam ein JM-2.0-Kodierer (Joint Model, siehe [6]) mit den Optionen des Main-Profiles zum Einsatz. Bei allen getesteten Formaten haben wir nur das erste Bild intra-kodiert und zwischen den P-Bildern jeweils zwei B-Bilder eingefügt. Es wurde eine Bewegungsschätzung mit voller Suche im Bereich von ± 32 Bildpunkten bei allen Kodierern zusammen mit einer Lagrange-Coder-Steuerung durchgeführt [7]. Bitraten haben wir mittels fester Quantisierungsparameter eingestellt.

Bei den Tests erzielte H.264/AVC deutlich bessere Ergebnisse als die anderen Verfahren. Verantwortlich dafür sind im We-



Vergleich des PSNR in Abhängigkeit von der verwendeten Bitrate (links). Auf der rechten Seite sind die Bitratengewinne gegenüber MPEG-2 entlang des PSNR der Luma-Komponente aufgetragen.



Die Tempete-Sequenz in CIF mit 15 Bildern/s jeweils mit 128 kBit/s; es ist dabei zu erkennen, dass H.264/AVC bei der gewählten Bitrate die Details deutlich klarer darstellt als MPEG-4.

sentlichen das sehr effiziente Bewegungsmodell und die Entropiekodierung mit CABAC. Allerdings kann man nur schwer einzelnen Elementen die Steigerung der Kodiereffizienz zuordnen. Vielmehr summieren sich mehrere Detailverbesserungen auf.

Komplexität

Obwohl H.264/AVC ebenso wie seine Vorläufer nur den Decoder und das Bitstrom-Format spezifiziert, verlangt das Standardisierungsgremium für jedes vorgeschlagene Kodierverfahren eine Beispiel-Implementierung, um sowohl die Durchführbarkeit als

auch die erwartete Effizienzerhöhung nachzuweisen. Nimmt das Standardisierungsgremium ein Kodierverfahren an, implementiert es derjenige in die Referenzsoftware, der es vorgeschlagen hat. Während der Entwicklung von H.264/AVC sind so über 100 Vorschläge von mehr als 20 verschiedenen Unternehmen in die Referenzsoftware eingeflossen, was dazu führte, dass sie langsam, unübersichtlich und für praktische Implementierungen quasi unbenutzbar wurde. Aus diesem Grund überschätzen beispielsweise Analysen, die wie auf Basis dieser Implementierung erstellt wurden, die tatsächliche Komplexität des

H.264/AVC-Kodierers um Größenordnungen und die des Decoders um einen Faktor zwei bis drei.

Auf der letztjährigen IBC zeigte die Firma Videolocus jedoch bereits einen H.264/AVC-Video-Strom mit 1 MBit/s in DVD-Qualität, den ein Pentium-4-Computer in Echtzeit kodiert, wenn auch damals mit Hilfe von Hardwarebeschleunigung auf einer FPGA-Karte, die die Bewegungsschätzung, Schätzung der Intra-Prädiktion, Entscheidungsstatistiken und Video-Vorverarbeitung unterstützte.

Im Oktober 2002 zeigte UB Video (www.ubvideo.com) im H.264/AVC-Baseline-Profil kodier-

tes Video in CIF-Auflösung auf einem 800-MHz-Laptop. Er kodierte 49 Bilder pro Sekunde, beim Dekodieren brachte er es auf 105 Bilder pro Sekunde; beides parallel schaffte der Rechner mit immerhin 33 Bildern pro Sekunde. Die für Echtzeitvideokonferenzen optimierte Kodierimplementierung erzielte bei für ein solches Szenario typischem Material eine nur um zehn Prozent höhere Bitrate als die deutlich langsamere Referenzsoftware.

Wie viele andere Unternehmen – unter anderem Broadcom, Nokia und Motorola – entwickelt auch das Heinrich-Hertz-Institut (HHI) in Berlin H.264/AVC-Echt-



Tempete CIF mit 30 Bildern/s: H.264/AVC erreicht bereits bei 512 kBit/s etwa die gleiche visuelle Qualität wie MPEG-4 mit etwa 1 MBit/s.

zeitlösungen. Eine Software-Implementierung für Highend-PC (jenseits von 2,5 GHz) spielt in TV-Auflösung (4CIF) vorliegendes H.264-Material problemlos in Echtzeit ab und kodiert 20 Bilder pro Sekunde in CIF bei weniger als zehn bis 15 Prozent Bitraten-erhöhung gegenüber der langsamen Referenzsoftware im Main-Profile. Die Dekodierer-Implementierung des HHI wurde auch auf einen ARM922-Prozessor portiert und liefert dort sechs Bilder pro Sekunde in CIF-Auflösung und 25 Frames pro Sekunde in QCIF.

Fazit

Die Kodiereffizienz von H.264/AVC übertrifft alle existierenden Standards etwa um den Faktor zwei. Insbesondere im Vergleich zu MPEG-2, das die Basis für digitales Fernsehen weltweit ist, wird eine Steigerung um den Faktor 2,25 bis 2,5 erreicht. Diese Verbesserungen ermöglichen viele neue Einsatzgebiete, etwa im Bereich der Fernseh-

übertragung und HD-DVD sowie der 3G-Funknetze und xDSL. Obwohl H.264/AVC beim Dekodierer etwa zwei- bis dreimal und der Kodierer sogar bis zu fünfmal so komplex ist wie MPEG-2, ist es aufgrund des Anstiegs der allgemein verfügbaren Rechenleistung weniger ressourcenfressend als MPEG-2 zum Erscheinungszeitpunkt.

Ein weiterer wichtiger Punkt, der für H.264/AVC (seine Vorläufer und MPEG) spricht: Es handelt sich um einen offenen Standard. Prinzipiell kann jeder Hersteller einen H.264/AVC-Kodierer und/oder -Dekodierer bauen. Über das entsprechende Lizenzmodell ist noch nicht entschieden. Zudem gibt es keine Abhängigkeit von proprietären Formaten wie sie heute im Internet üblich sind, was vor allem für die Broadcast-, die DVD- und die Telekommunikations-Industrie außerordentlich wichtig ist. (vza)

Karsten Sühning ist Forscher am Fraunhofer Institut für Nachricht-

tenttechnik – Heinrich-Hertz-Institut (HHI) – und ist für die Koordination der Referenz-Software von H.264/AVC verantwortlich.

Dr. Heiko Schwarz ist Forscher am HHI und hat verschiedene Vorschläge in den Standardisierungsprozess für H.264/AVC eingebracht sowie aktiv an der gesamten Entwicklung der Spezifikation und der Software von H.264/AVC mitgearbeitet.

Dr. Thomas Wiegand ist Gruppenleiter für Bildkommunikation am HHI. Weiterhin ist er sowohl einer der Leiter der ITU-T VCEG und des ITU-T|ISO/IEC JVT, das H.264/AVC standardisiert, als auch der Editor der H.264/AVC-Spezifikation.

Literatur

- [1] Thomas Wiegand (Editor), 2002: Joint Final Committee Draft. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), Doc. JVT-E010, März

2003: <http://bs.hhi.de/~wiegand/JVT.html>

- [2] Exp-Golomb-Codebaum: www.vcodex.fsnet.co.uk/h264_vlc.pdf
- [3] ITU-T and ISO/IEC JTC1, 1994: Generic coding of moving pictures and associated audio information – Part 2: Video, ITU-T Recommendation H.262 – ISO/IEC 13818-2 (MPEG-2), November 1994
- [4] ITU-T, 1998: Video coding for low bitrate communication, ITU-T Recommendation H.263, Version 1: November 1995, Version 2 (H.263+): January 1998, Version 3 (H.263++): November 2000
- [5] ISO/IEC JTC1, 2000: Coding of audio-visual objects – Part 2: Visual. ISO/IEC 14496-2 (MPEG-4 Visual version 1), April 1999, Amendment 1 (Version 2), February 2000
- [6] H.264/AVC-Referenzimplementierung: <http://bs.hhi.de/~suehring/tml/>
- [7] Lagrange-Coder: <http://bs.hhi.de/~wiegand/icip01c.pdf>

