

Hauptseminar

Multimedia-Codierung und Kompression im Streaming Bereich

von

Christian Heider

Telematik

TU-Ilmenau

Betreuer: Thorsten Strufe

Inhaltsverzeichnis

Einführung

1. Methoden

- 1.1 Das menschliche Sehsystem
- 1.2 Farbtransformationen
- 1.3 Signaltransformationen
- 1.4 Quantisierung
- 1.5 Präcodierung
- 1.6 Entropiecodierung
- 1.7 Zukunftsorientierte Kodierungsverfahren

2. Standards

- 2.1 JPEG / JPEG2000
- 2.2 MPEG
- 2.3 H.26x

3. Streaming Media

- 3.1 RealNetworks
- 3.2 Microsoft
- 3.3 Apple

Einleitung

Das Internet entwickelte sich in den letzten Jahren zu einem Unterhaltungsmedium in dem Audio- und Videodaten in Echtzeit übertragen werden können. Doch um dies zu ermöglichen bedürfte es einiger Neuentwicklungen. Besonders im Bereich Kompression hat sich in den letzten einiges getan.

Aber wozu braucht man eigentlich die Datenkompression? Eine einfache Rechnung soll einmal verdeutlichen was es bedeutet ein Video ohne Kompression abzuspeichern. Wir nehmen folgende Daten an: 384x288 Bildpunkte, 24 bit Farbtiefe, 25 Bilder pro Sekunde und Ton in 44,1 kHz Sample-Rate bei einer Auflösung von 16 bit und in Stereo (entspricht VHS-Qualität), dann würde man pro Sekunde 8,5 Megabyte Platz benötigen. Und dies ist in Bezug auf das Internet und Multimedia nicht akzeptabel.

Im 1. Abschnitt möchte ich eine Reihe von Kodierungsmethoden vorstellen. Im 2. Abschnitt sollen die Kompressionsstandards im Bild und Videobereich dargestellt werden. Zum Abschluss noch ein kleiner Überblick über die großen Firmen die Streamingsoftware vertreiben.

1. Methoden

1.1 Das menschliche Sehsystem (HSV)

Als erstes sollte man sich im Klaren sein was das menschliche Auge erfassen kann. Denn nur mit diesen Erkenntnissen lassen sich effiziente Kompressionsalgorithmen gestalten.

Die Aufnahme der optischen Reize im Auge geschieht über die Stäbchen und Zapfen. Diese beiden Rezeptortypen besitzen unterschiedliche Eigenschaften und kommen in unterschiedlichen Situationen zum Einsatz. Während die Zapfen sowohl farb- als auch helligkeitsempfindlich sind, können die Stäbchen nur Helligkeitsunterschiede wahrnehmen. Dadurch dass die Stäbchen wesentlich helligkeitsempfindlicher kommen sie hauptsächlich beim Nachtsehen zum Einsatz. Die Zapfen hingegen sind nur für das Tages- und Farbsehen verantwortlich.

Folgende Eigenschaften sind des Weiteren zu beachten:

Lichtempfindlichkeit

Das menschliche Auge ist in der Lage durchschnittlich ca. 500 Helligkeitsstufen zu unterscheiden. Der Wert kann sich aber unter gewissen Umständen sogar verdoppeln. Die Lichtempfindlichkeit lässt sich jedoch nicht nur auf den Unterschied von Zapfen und Stäbchen zurückführen, sondern auch durch den Einsatz von rezeptiven Feldern wird diese Wirkung noch verstärkt.

Auflösungsvermögen

Das Auflösungsvermögen ist bei ruhenden Bildern am größten. Die örtliche Auflösung ist beim Farbsehen deutlich geringer als beim Helligkeitssehen. Der Grund dafür liegt vermutlich in der Informationsverarbeitung in den nachfolgenden Neuronenschichten.

Maskierungseffekte

Das menschliche Auge kann horizontale und vertikale Kontraste besser wahrnehmen als schräge Strukturen.

Mit Hilfe dieser Eigenschaften lassen sich bei der Kompression redundante Informationen aus den Bildern entfernen.

2.2 Farbsysteme und Farbtransformationen

Es existieren eine Menge von Farbräume, wobei jedes Vor- und Nachteile besitzt und meist für einen bestimmten Anwendungsbereich konzipiert ist. Ich will mich in dieser Arbeit nur auf die 2 wichtigsten zwei Farbräume beschränken.

Der RGB Farbraum wird durch eine additive Farbmischung der drei Grundfarben Rot (700 nm), Grün (546,1 nm) und Blau (435,8 nm) bestimmt. Die drei Vektoren sind linear unabhängig und spannen einen dreidimensionalen Raum auf (RGB-Einheitswürfel). Während RGB zur Bilddarstellung in der Monitortechnologie genutzt wird, basiert die digitale Video- und Bildtechnologie auf den Farbraum $YCbCr$. Die Luminanzkomponente Y beschreibt die Helligkeit und die Chrominanzkomponenten C_b und C_r enthalten den Farbverlauf. Die Transformation lautet:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0,299 & 0,587 & 0,144 \\ -0,169 & -0,331 & 0,500 \\ 0,500 & -0,419 & 0,081 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Die Wertebereiche die sich daraus ergeben sind:

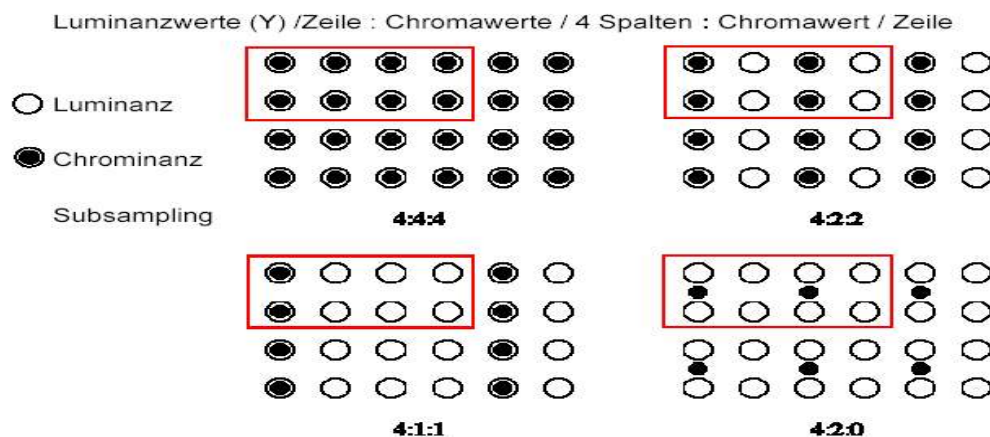
$$0 \leq Y \leq 255 \qquad -127,5 \leq Cb \leq 127,5 \qquad -127,5 \leq Cr \leq 127,5$$

Ausserdem wird festgelegt das die Y Komponente nur im Bereich [16...235] definiert ist und die Chrominanz C_b und C_r werden in den positiven Zahlenbereich [16...240] verschoben.

Farbunterabtastung

Da das menschliche Auge die farblichen Kontraste nicht so gut separieren kann, wird zur Vorbereitung der Kompression eine Unterabtastung der Chrominanzwerte vorgenommen. Dabei wurden verschiedene Formate definiert:

Format	Bits pro Bildpunkt	Beschreibung
4:4:4	8+8+8=24	Keine Unterabtastung
4:2:2	8+4+4=16	Horizontale Unterabtastung der Chrominanz $M = 2$
4:1:1	8+2+2=12	Horizontale Unterabtastung der Chrominanz $M = 4$
4:2:0	8+2+2=12	Horizontale und vertikale Unterabtastung $M = 2$



Mit dieser Unterabtastung lässt sich die Datenmenge schon beachtlich verringern ohne einen subjektiven Verlust.

1.3 Signaltransformationen

Diskrete Kosinus-Transformation: DCT

Die HVS zeigt, dass das menschliche Sehen beeinflusst ist von der Frequenz im Bild. Deshalb versucht man, die unterschiedlichen Frequenzen in einem Bild aufzuteilen, um die hohen Frequenzen, die nur einen geringen Einfluss auf die visuelle Wahrnehmung haben, von den niedrigen zu trennen.

Die DCT zerlegt das Bild in verschiedene Kosinuswellen unterschiedlicher Frequenz und Amplitude. Dabei entsteht eine Linearkombination von Kosinuswellen die das Bild

$$F(v, u) = \frac{C(v)C(u)}{2} \sum_{y=0}^7 \sum_{x=0}^7 f(y, x) \cos\left[\frac{(2x+1)v\pi}{16}\right] \cos\left[\frac{(2y+1)u\pi}{16}\right] \quad \forall v, u = 0, \dots, 7$$

interpretieren. Da Bilder einen zweidimensionalen Aufbau besitzen, ist eine zweidimensionale DCT für die Bildkompression relevant. In den meisten Fällen wird das Bild in Blöcke aufgeteilt (in der Regel 8x8), da es sonst zu Berechnungsschwierigkeiten kommen kann.

$$C(u) = C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{für } u, v = 0 \\ 1 & \text{sonst} \end{cases}$$

Dabei ist $f(y,x)$ das zweidimensionale Ausgangssignal mit den Pixelwerten in den Zeilen y und Spalten x . Der Wert mit der horizontalen und vertikalen Frequenz von null wird als der Gleichanteil (DC) bezeichnet. Alle anderen Werte in dem Block werden als Wechselstromkoeffizienten (AC) bezeichnet und zeigen den Farbwechsel und die Kanten im Bild an. Mit steigender Frequenz werden die Werte immer kleiner bis sie nach null konvergieren. Die DCT ist verlustfrei, doch auf Rechnern kommt es wegen des physikalischen Aufbaus zu fehlender Genauigkeit. Die DCT besitzt noch weitere Probleme. Dadurch dass die Basisfunktionen periodisch sind, wird eine bestimmte Frequenz in einem Abschnitt, die nirgendwo sonst auftritt, nicht gesondert berücksichtigt. D.h. dass die Frequenz über das komplette Ausgangssignal gelegt wird und zusätzliche Frequenzen hinzugezogen werden müssen, um sie in den Abschnitten in denen sie nicht vorkommt auszulöschen. Damit wird die Anzahl der Koeffizienten die ungleich Null sind größer. Ein weiteres Problem ist die Aufteilung in Blöcke. Denn an den Grenzen zwischen den Blöcken kann es durch die Quantisierung zu Unstetigkeiten kommen.

Wavelet Transformation

Die Theorie der Wavelet Transformation ist primär in den 80iger Jahren des letzten Jahrhunderts von J. Morlet, S. G. Mallat, Y. Meyer und I. Daubechies entwickelt wurden.

Die Wavelet Transformation ermöglicht eine Signalanalyse im Zeit- und Frequenzbereich. Während bei der DCT nur der Sinus und der Kosinus als analysierende Funktionen zulässig sind, kann bei der Wavelet Transformation die zu analysierende Funktion nahezu beliebig gewählt werden.

Ausgangspunkt ist das Basiswavelet Ψ . Um bei einer gegebenen Funktion die Frequenzbestandteile zu untersuchen, können die Wavelets gestreckt, gestaucht und verschoben werden. Während der Transformation kommen dann die verschiedenen Varianten des Basiswavelets zum Einsatz. Diese verschiedenen Varianten sind auch unter dem Namen Waveletfamilie bekannt. Das Basiswavelet muss aber verschiedene Bedingungen erfüllen:

1. Zulässigkeitsbedingung

$$0 < c_Y := 2\pi \int \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty$$

d.h. das Wavelet ist nur in einem bestimmten Bereich ungleich null

2. Integralbedingung

$$\int \Psi(x) dx = 0$$

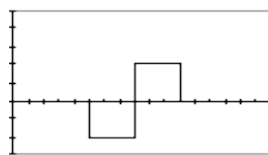
Die allgemeine Beschreibung einer Waveletfamilie kann durch folgende Gleichung dargestellt werden:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad a \in \mathbb{R}^+ \quad b \in \mathbb{R}$$

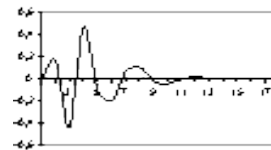
a - Skalierungsfaktor

b - Verschiebungsfaktor

Beispiele für Basiswavelets:



Haar-Wavelet



Daubechies-Wavelet

In der Bildverarbeitung wird hauptsächlich mit der Schnellen Wavelet Transformation (FWT) gearbeitet. Bei FWT wird eine sogenannte Multiskalenanalyse (Multiresolution Analysis MRA) verwendet. Die Umsetzung der MRA erfolgt folgendermaßen:

Mittels einer Skalierungsfunktion wird das Bild in verschiedene Auflösungen approximiert. Dies erfolgt durch einen Tiefpass-Filter, welcher im Bild eine Mittelwertbildung von benachbarten Pixeln bewirkt. Die „verlorenen“ Informationen werden durch ein geeignetes Wavelet beschrieben. Technisch wird dies durch einen Hochpass-Filter realisiert. Nach der Hoch- bzw. Tiefpassfilterung wird das Ergebnis um den Faktor 2 unter abgetastet. Anschließend kann eine weitere Transformationsstufe durchgeführt werden, wobei das Eingangssignal der Tiefpass-Anteil der vorhergehenden Stufe ist.

Dieser Prozess kann so lange durchgeführt werden bis die Größe des Eingangsbildes die Größe des Filters unterschreitet.

Dadurch, dass Bilder einen 2 dimensional Aufbau besitzen, wird der Prozess innerhalb einer Transformationsstufe noch einmal aufgeteilt. Als erstes wird das Bild nur zeilenweise gefiltert (Hoch- und Tiefpassfilter) und danach unter abgetastet. Im zweiten Schritt werden jetzt die jeweiligen Bilder aus dem Subsampling spaltenweise Hoch- und Tiefpass gefiltert und anschließend wieder unter abgetastet (Abb. 1).

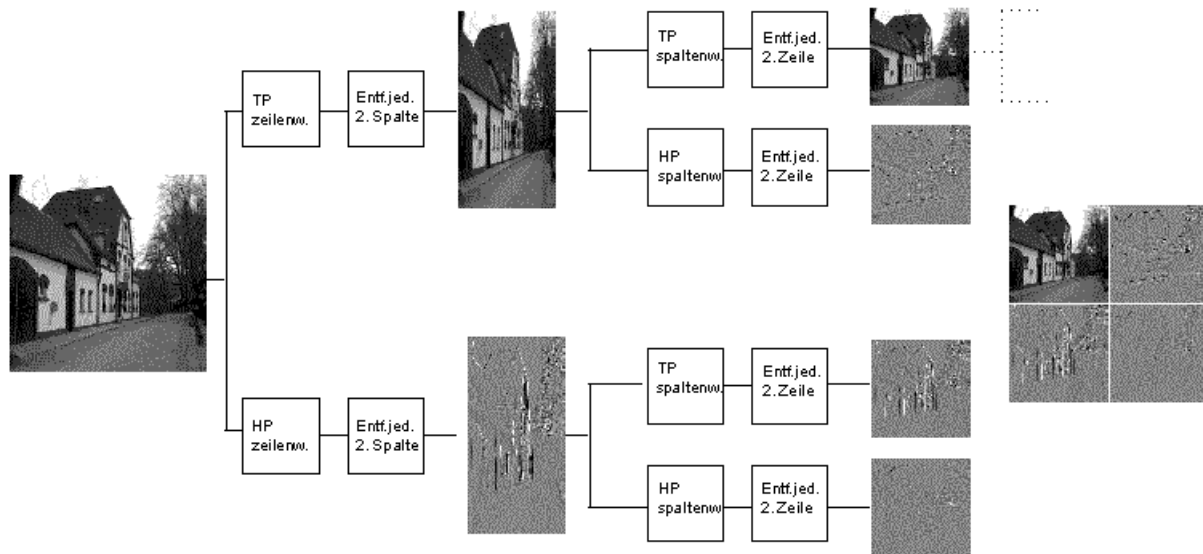


Abbildung 1 Prinzip der Wavelet Transformation

Die Wavelet Transformation besitzt mehrere Vorteile. Als erstes lassen sich die Bilddaten skalierbar laden, d.h. ein progressiver Bildaufbau ist möglich. Desweiteren hat man nur eine Datei aus der alle Bildqualitäten erzeugt werden können. Dies ermöglicht ein leichteres Handling mit den Bildern z.B. in Archiven.

Fraktale Kompression

Ziel der fraktalen Transformation ist es ein Gleichungssystem zu finden, das durch iterative Anwendung zum Bild führt. Da innerhalb eines Bildes viele Strukturen Ähnlichkeiten aufweisen, werden diese als Basisvektoren (Referenz) zur Transformation herangezogen.

Das Verfahren funktioniert folgendermaßen: Als erstes wird das Bild in nicht überlappende Regionen (Range-Blöcke) eingeteilt. Für jeden Range-Block wird nun ein ähnlicher Bildausschnitt (Domain-Block) gesucht, der sich durch eine geeignete Transformationsvorschrift (z.B. Drehung, Scherung, Streckung/Stauchung und Verschiebung) ergibt.

Die Dekodierung erfolgt solange durch die iterative Anwendung der Transformationsmatrizen, bis keine Änderung auftritt.

Der Vorteil des Verfahrens liegt in der Auflösungsunabhängigkeit und in der sehr guten Kompressionsrate. Jedoch benötigt das Verfahren bei der Encodierung einen hohen Rechenaufwand, der umso größer wird je mehr Blöcke oder Transformationsvorschriften hinzukommen. Durch die Erweiterung von Kontrast und Helligkeit kommt es zu einer besseren Ausnutzung der Ähnlichkeiten.

1.4 Quantisierung

Bei der Quantisierung wird einem analogen Wert x ein quantisierter Wert x_q zugewiesen. Es gibt mehrere Möglichkeiten der Quantisierung:

1. Gleichmäßige Quantisierung

Das entscheidende Merkmal hierbei ist das alle Quantisierungsintervalle die gleiche Breite besitzen. Der Quantisierungsfehler beträgt immer:

$$\frac{\Delta}{2} \leq e_q \leq \frac{\Delta}{2}$$

2. Sukzessive Approximation

Sinn dieses Verfahrens ist es zuerst eine grobe Signalinformation zu übertragen und diese dann schrittweise zu verfeinern. Nach der kompletten Übertragung sieht die Verteilung wie bei einer gleichmäßigen Quantisierung aus.

Als erstes wird der maximale Wert x_{\max} bestimmt. Daraufhin wird der Quantisierungsbereich in 2 gleichgroße Teile zerlegt und die Schwelle wird mit T bezeichnet. Im ersten Schritt wird nun eine binäre Entscheidung getroffen, ob der Wert über- oder unterhalb der Schwelle liegt. Im weiteren Verlauf des Algorithmus wird der betreffende Quantisierungsbereich wieder unterteilt, damit eine genauere Lokalisierung möglich ist. Die Intervalle können so bis zu einem Abbruchkriterium verfeinert werden.

3. Ungleichmäßige Quantisierung

Das Hauptmerkmal der ungleichmäßigen Quantisierung sind die unterschiedlichen Intervallbreiten. Die Quantisierungskennlinie ist durch die jeweilige Problemstellung gekennzeichnet. Es existieren eine Menge von verschiedenen Kennlinien: z.B. Wahrnehmungsdichteverteilungs-optimiert, Wahrnehmungsoptimiert, SNR-optimiert usw.

1.5 Präcodierung

In Bildern gibt es meist vielfältige Beziehungen und Abhängigkeiten zwischen den Zeichen (Symbolen). Die Präcodierung versucht die Intersymbolredundanz zu vermindern.

Lauf längencodierung

Die Lauf längencodierung ist ein sehr einfaches Verfahren, um mehrmals hintereinander vorkommende Zeichen oder Symbole zusammenzufassen. In diesem Abschnitt werden nur Signale aus einer binären Quelle betrachtet. Dabei gibt es mehrere Möglichkeiten:

1. Man zerlegt das Signal in 0- und 1-Blöcke und zählt dann die Anzahl der jeweiligen Werte. Beispiel: 1110000011111000 wird folgendermaßen kodiert (1,3), (0,5), (1,5), (0,3)

Dadurch, dass bei einer binären Quelle die Wertfolge immer alternierend ist, lässt sich diese auch zusammenfassen: 3 5 5 3 . Dabei muss aber der Anfangswert zuerst angegeben werden.

2. Eine weitere Möglichkeit ist es, die Position und die Länge eines Blockes anzugeben. Der zweite Block kann dann automatisch in die Lücken gefüllt werden. Beispiel für einen 0-Block: 1111001110000011 wird zu (4,2), (9,5)

Die Lauf längencodierung eignet sich besonders gut bei Schwarz-Weiß-Bildern. Da es bei Farbbildern selten vorkommt, dass mehrere Pixel hintereinander den gleichen Farbwert besitzen.

LZW-Algorithmus

Der LZW-Algorithmus ist Ende der 70iger Anfang der 80iger Jahre von A. Lempel, J. Ziv und T. Welch entwickelt wurden. Hierbei werden verschiedene Symbole zu Gruppen zusammengefasst und in ein Wörterbuch aufgenommen. Das Wörterbuch wird aus den unkomprimierten Daten aufgebaut. Innerhalb des Wörterbuches wird zu jeder neuen Symbolgruppe ein Codewort erzeugt. Kommt ein neues Signal, wird es mit dem Wörterbuch

verglichen und passende Muster werden durch den entsprechenden Code ersetzt. Andernfalls wird die neue Symbolgruppe ins Wörterbuch geschrieben und ein Codewort erzeugt.

Beispiel:

ABBABABAC

mit Wörterbuch:

1 = A, 2 = B, 3 = C

⇒ Kodiert: 1 2 2 4 7 3

Wörterbuch danach:

1 = A, 2 = B, 3 = C,

4 = AB, 5 = BB, 6 = BA,

7 = ABA, 8 = ABAC

1.6 Entropiecodierung

Schwerpunkt der Entropiecodierung ist es die Redundanz durch Ausnutzung der Symbolverteilung zu reduzieren. D.h., dass zu häufig auftretenden Symbolen kurze Codewörter und zu seltenen auftretenden Symbolen lange Codewörter zugewiesen werden.

Huffman-Codierung

Dieses Verfahren wurde 1952 von D. Huffmann entwickelt und garantiert in jedem Fall die optimale Codezuweisung. Durch folgendes Codebaummodell wird der Huffmann-Code erzeugt:

1. alle Symbole und ihre dazugehörigen Wahrscheinlichkeiten werden betrachtet
2. die Beiden mit der geringsten Wahrscheinlichkeit werden zu einem Knoten zusammengefasst
3. die Zweige werden mit 0 und 1 beschriftet
4. Weiterführung bis die Wurzel des Baumes oder eine Wahrscheinlichkeit von $p=1$ erreicht ist und alle Symbole codiert sind

Abbildung am Beispielwort „Betriebssystem“:

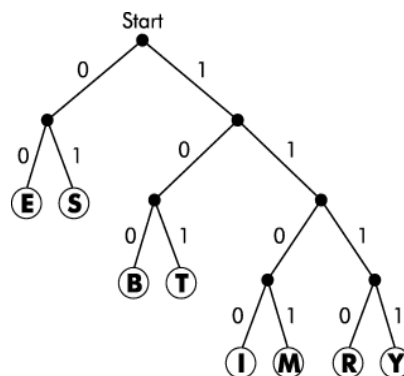


Abbildung 2 Beispiel für einen Codebaum

1.7 Zukunftsorientierte Kodierungsverfahren

Modellbasierte Kodierung

Der Nachteil an den blockbasierten Kodierungsverfahren ist, dass durch die Beschreibung mittels Blöcken und Bewegungsvektoren die Bildmanipulation und die Effizienz stark eingeschränkt sind. Dagegen wird bei dem modellbasierten Ansatz ein 3D Modell der Objekte erstellt. Dieses Modell kann parametrisch beschrieben werden und ermöglicht somit eine sehr effektive Datenreduktion.

Bei der Encodierung müssen als erstes die Parameter (Form, Textur, Mimik, Bewegung) der Objekte bestimmt werden. Zur Bestimmung der Parameter existieren mehrere Varianten:

1. mit Hilfe eines Laserscanners wird das Beispiel abgetastet und mittels einer Kamera wird noch die Textur aufgenommen
2. die Bildaufnahme geschieht mittels einer Stereokamera
3. „on the fly“ Objektgenerierung aus dem 2D Bilddaten, um es dann an ein vorhandenes Modell anzupassen (z.B. Gesichtsmodell)

Dieses Verfahren ist derzeit noch in der Experimentierphase und bedarf noch einiger Entwicklungsarbeit bis es zu einem stabilen und effektiven Verfahren für die Praxis wird.

Skalierbare Kodierung

Der Vorteil der skalierbaren Kodierung ist, dass man nur einen Datenstrom komprimieren muss um eine ganze Palette von Endgeräten und Übertragungsstrecken zu bedienen. Das Prinzip dabei ist folgendes: es wird ein Basislayer und mehrere „Enhancement Layer“ erzeugt. Der Basislayer besitzt eine geringere Datenrate und Qualität. Zur Verbesserung des Bildes werden solange Teile aus dem „Enhancement Layer“ hinzugenommen bis die höchste Qualitätsstufe erreicht ist.

Es gibt mehrere Arten der Skalierbarkeit:

- SNR Skalierbarkeit: mit zunehmender Anzahl von Layern verbessert sich die Qualität des Bildes
- Spatiale Skalierbarkeit: durch Unterabtastung wird die Bildauflösung beeinflusst
- Skalierbarkeit der Bildschärfe: Aufteilung in mehrere Frequenzbänder
- Temporale Skalierbarkeit
- Inhaltsbasierte Skalierbarkeit

Die Zukunftsorientierung entspricht nicht ganz der Wahrheit, da schon seit Jahren in den bestehenden Videostandard mehrere Skalierungsverfahren integriert sind. Jedoch besitzen diese Verfahren eine schlechtere Effektivität der Codierung, eine höher benötigte Bandbreite und einen komplexeren Encoder.

Ein weiteres Problem ist der sogenannte „Drift“ Effekt bei der Videoskalierung. Durch die bewegungskompensierte Prädiktion benötigen der Encoder und Decoder immer die selbe Prädiktion. Die Idee der skalierbaren Videokodierung skalierbaren Videokodierung ist es aber die Übertragung der Datenmenge zu verändern. Das heißt das es zu Fehlern kommt, da die Prädiktion zwischen Encoder und Decoder bei einer Übertragungsratenänderung unterschiedlich ist. Es existieren zwar mehrere Lösungsvorschläge, jedoch haben alle Ansätze nur einen bestimmten Arbeitsbereich und sind meist zu komplex.

Eine wesentlich bessere Effizienz bietet das Wavelet basierte Verfahren. Hierbei stehen mehrere Ansätze zur Verfügung, die aber zu diesem Zeitpunkt noch Gegenstand der Grundlagenforschung sind.

Zusammenfassung

Die hier vorgestellten Methoden sind der Schlüssel für die Multimedia Bildkodierung. In der Signaltransformation gehört die DCT schon zu den auslaufenden Modellen, da sie in ihrer Entwicklung schon voll ausgeschöpft ist. Die Zukunft gehört den Wavelet basierten Verfahren. Ihre bessere Effizienz und die leichtere Einbindung der skalierbaren Kodierung erweisen sich als vorteilhaft. Aber auch den modellbasierte Ansatz sollte man nicht unterschätzen, obwohl er meiner Ansicht nach nur in speziellen Anwendungen zum Einsatz kommen kann.

Bei der Präcodierung werden schon verbessert Algorithmen in der Praxis eingesetzt, z.B. ist bei MPEG-4 eine rückwärts dekodierbare variable Längenkodierung integriert. Diese ermöglicht es nicht den ganzen Block bei einem einzelnen Fehler zu verwerfen, sondern es wird zuerst bis zur Fehlerstelle dekodiert und dann vom Ende des Blockes zurück zur Fehlerstelle dekodiert. Somit bleibt nur noch die Fehlerstelle übrig.

Im allgemeinen kann man sagen das in allen Bereichen immer weiter geforscht wird und eine Stagnation derzeit nicht in Sicht ist. Das heißt für die Zukunft das es immer bessere Algorithmen geben wird. Das Problem der Komplexitätssteigerung lässt sich derzeit sehr gut mit der schnelleren Steigerung der Rechenleistung kompensieren.

Für die Zukunft wird aber auch auf mehr Robustheit und Sicherheit geachtet werden, insbesondere in Bezug auf Multimedia-Streaming im Internet und „wireless“ Verbindungen die eine führende Marktposition einnehmen werden.

2. Standards

2.1 JPEG/JPEG2000

JPEG

Die Joint Photographic Expert Group ist 1986 gegründet worden um einen neuen Standard zur Kompression von Bildern in Fotoqualität zu entwickeln.

Der Ablauf des Verfahrens im einzelnen:

Zuerst wird eine Farbtransformation vom RGB-Raum in den $YCbCr$ -Raum durchgeführt. Danach kommt es zum sogenannten Sampling, indem die Chrominanzkomponenten mit einer geringeren Auflösung bestimmt werden. Ein typischer Wert ist 4:1:1, indem der Luminanzanteil viermal höher ist als der Chrominanzanteil.

Als nächstes wird das Bild in 8×8 Pixelblöcke unterteilt. Ist das Bild nicht durch 8 teilbar, werden an den Rändern, durch spiegeln und periodisches fortsetzen, Werte angehängt. Jeder dieser Blöcke wird als ein diskretes Signal mit 64 Werten aufgefasst. Auf diese Blöcke wird die DCT angewendet, wobei jede Bildkomponente (Y, C_b, C_r) separat abgearbeitet wird.

Durch die DCT ist es möglich eine Aussage über die Variationen der Intensität, Farbe und Helligkeit zu treffen. Nachdem die Transformation aller Blöcke abgeschlossen ist, folgt die Quantisierung. Die Werte aus der Transformation werden durch die Quantisierungswerte aus der Quantisierungsmatrix geteilt und auf ganzzahlige Werte abgerundet. Dabei wird eine ungleiche Quantisierung benutzt, die die niedrigeren Frequenzen feiner abstuft als die höheren. Dadurch kommt es, dass die kleinen Werte bei den hohen Frequenzen zu null quantisiert werden.

63	56	40	36	27	23	12	14	8	16	19	22	26	27	29	34	8	4	2	2	1	1	0	0
64	66	51	39	37	34	16	18	16	16	22	24	27	29	34	37	4	4	2	2	1	1	0	0
61	65	52	36	34	36	16	17	19	22	26	27	29	34	34	38	3	3	2	1	1	1	0	0
49	53	52	13	41	36	20	19	22	22	26	27	29	34	37	40	2	2	2	0	1	1	1	0
43	41	32	14	34	17	11	8	22	26	27	29	32	35	40	48	2	2	1	0	1	0	0	0
37	30	22	15	14	12	12	6	26	27	29	32	35	40	48	58	1	1	1	0	0	0	0	0
28	21	14	15	15	16	13	4	26	27	29	34	38	46	56	69	1	1	0	0	0	0	0	0
20	14	16	13	13	10	9	0	27	29	35	38	46	56	69	83	1	0	0	0	0	0	0	0

DCT-kodierte Matrix
nach der

Quantisierungsmatrix

resultierende Matrix
Division

Bei der eigentlichen Kodierung werden die DC-Koeffizienten getrennt von den AC-Koeffizienten behandelt. Bei den DC-Koeffizienten wird nur der Differenzwert zum nächsten Block gespeichert und dann mit Hilfe des Huffman-Codes codiert. Dadurch, dass der Gleichanteil zwischen 2 benachbarten Pixeln meistens ähnlich ist, wird der Huffman-Code in 12 Kategorien eingeteilt, wobei kleine Unterschiede bei den DC-Koeffizienten mit einem kleinen Codewort versehen werden.

Die AC-Koeffizienten werden durch einen Zick-Zack-Algorithmus in einen Vektor überführt. Durch diese Anordnung der Ortsfrequenzen kommt es dazu, dass die hochfrequenten Werte mit null am Ende stehen.

Auf die Werte ungleich null wird wiederum der Huffman-Code angewendet.

JPEG2000

In den letzten Jahren ist klar geworden das der JPEG Standard nicht für die heutige Zeit ausreicht. Aus diesem Grund wurde 1996 die JPEG2000 Working Group gegründet. Ihr Ziel war es einen Kodierungsstandard zu entwickeln der verschiedene Bildtypen mit unterschiedlichen Charakteristiken in einem Verfahren vereinigt. Ausserdem sollte es möglich sein, unterschiedliche Bildmodelle (für Echtzeittransmission, limitierte Bandbreiten und Puffer usw.) unterzubringen. Der Standard der sich daraus entwickelte, wurde in 11 Teile aufgeteilt. Der erste Teil beschäftigt sich mit dem En- und Dekodierungsprozess.

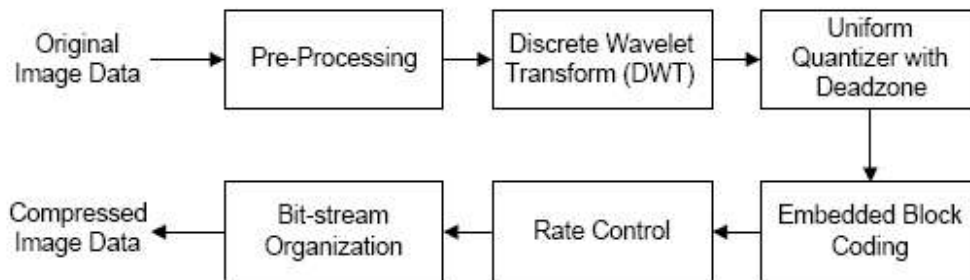


Abbildung 3 JPEG2000 Enkodierungsprozess

Im Pre-Processing Schritt wird das Bild erst einmal in rechteckige sich nicht überlappende Kacheln eingeteilt. Dies ist nötig um Speicherplatz bei der späteren Kompression zu sparen. Häufig werden Bilder im RGB Format präsentiert. Jedoch wird zur Kompression der $YCbCr$ Farbraum benutzt, da die einzelnen Schichten unabhängiger voneinander sind und die Komprimierung effektiver wird.

In JPEG2000 wird die diskrete Wavelet-Transformation benutzt um die einzelnen Kacheln in ihre Frequenzbänder aufzuteilen. Die einzelnen Komponenten werden dabei einzeln transformiert.

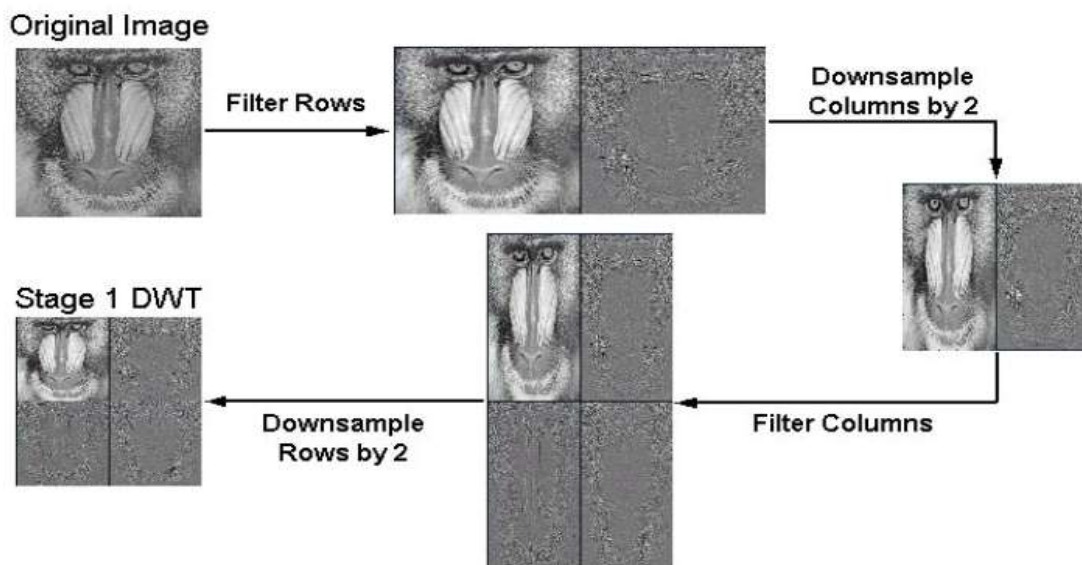


Abbildung 4 Transformation der Y-Komponente

In JPEG2000 ist es möglich 2 verschiedene Wavelet-Filter zu benutzen. Zum einen das verlustlose 5/3-Filter (Integer, reversible) und zum anderen das verlustbehaftete Daubechies 9/7-Filter (Fließkoeffizienten, irreversible).

Die Wavelet-Koeffizienten werden mit einem gleichmäßigen Quantisierer mit Todeszone quantisiert. Die Quantisierung ist fast immer verlustbehaftete, außer wenn beim 5/3-Filter mit einer Schrittweite $\Delta_b=1$ quantisiert wird.

$$q(u, v) = \text{sign}(a(u, v)) \frac{|a(u, v)|}{\Delta_b} \quad \begin{array}{l} a(u, v) - \text{ein Waveletkoeffizient} \\ q(u, v) - \text{quantisierter Waveletkoeffizient} \end{array}$$

$$\Delta_b = 2^{R_b - \epsilon_b} \left(1 + \frac{\mu_b}{2^{11}}\right)$$

Ansonsten lässt sich die Quantisierung mit Δ_b einstellen. Die Mantissen-Exponenten-Darstellung macht es möglich über einen großen Bereich gute Ergebnisse zu erzielen. Mit R_b und μ_b kann der Quantisierungsschritt Δ_b eingestellt werden, wobei R_b von der Bittiefe des Bildes abhängt. Ausserdem ist Δ_b auch für jedes Subband einzeln bestimmbar. Durch die Todeszone werden mehr Nullen generiert.

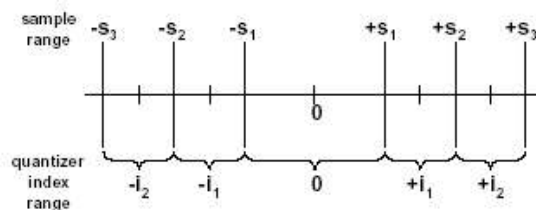


Abbildung 5 Struktur des Quantisierers mit Todeszone

Entropie-Kodierung

Bei JPEG2000 werden die Subbänder jeder Kachel in Code-Blöcke eingeteilt. Die Größe des Code-Blocks wird durch den Algorithmus bestimmt, wobei sie zwischen 4096 und 16 Koeffizienten liegt. Die Blöcke haben innerhalb eines Subbandes die selbe Größe.



Abbildung 6 Teilung der Subbänder in Code-Blöcke

Das embedded block coding gehört zu den adaptiven arithmetischen Kodierern. Jede Bit-Ebene eines Blockes wird durch eine Sequenz von 3 Kodierungsdurchgängen geschickt: significance propagation, magnitude refinement und clean up. Dies ermöglicht zum einen eine optimale Einbettung, so daß die Informationen die den größten Abfall in der Qualität hervorrufen als erstes codiert werden. Und zum anderen eine größere Anzahl an bit-stream Abbruchpunkten zu bestimmen, um eine feineren SNR Skalierbarkeit zu bekommen.

Bei significance propagation wird für jeden Bitwert der nicht signifikant ist, die Signifikanz seiner benachbarten Werte überprüft. Ein Wert ist dann signifikant, falls in einer früher betrachteten Bitebene für diesen Wert eine 1 als Binärwert festgestellt wurde. Ist einer seiner Nachbarn signifikant, so wird für diesen Bitwert ein Kontextsymbol generiert, was die Anzahl und die Stellung der signifikanten Nachbarn charakterisiert. Der Status der Signifikanz für

diesen Wert wird durch eines von 5 Kontextsymbolen codiert, welche eine Zusammenfassung der Nachbarn repräsentiert. Wird der Wert dabei signifikant, so wird sein Vorzeichen codiert. Im Verfeinerungsdurchlauf (magnitude refinement) werden für die Werte Kontextsymbole generiert, die in einer früheren Ebene als signifikant markiert wurden. Der letzte Durchgang (cleanup pass) behandelt alle Werte die bis jetzt noch nicht signifikant sind. Ist der betrachtete Wert und seine Nachbarn nichtsignifikant, so wird ein Symbol generiert, das diesen Zustand charakterisiert.

In 2.Ebene der Entropie-Kodierung wird der resultierende Bitstream für jeden Code-Block in Schichten (quality layers) einsortiert. Eine Schicht ist eine Sammlung von aufeinanderfolgenden Bitebenen von allen Code-Blöcken in allen Subbandern und allen Komponenten von jeder Kachel. Das bedeutet das bei der Übertragung jede zusätzliche Schicht die Bildqualität erhöht.

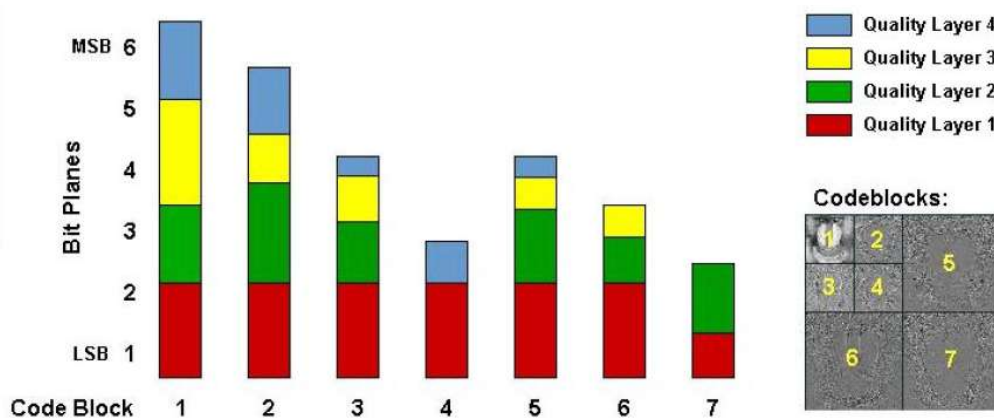


Abbildung 7 Schichtenverteilung für eine 2 stufige DWT für Subbänder die nur einen Code-Block besitzen

Während in der ersten Ebene der Entropie-Kodierung eine Anzahl von Bit-Streams (für jeden Code Block einen) generiert wurden ist, werden in der 2.Ebene die einzelnen Streams zu einem Code-Stream zusammengefasst und auf eine effiziente Art und Weise angeordnet. Dies ermöglicht ein SNR, Auflösungen, Region-of Interest und Skalierbarkeit.

Rate Control

Mit Hilfe des Rate Control kann der Code-Stream so verändert werden, damit eine bestimmte Bitrate erreicht wird. Dabei wird ein Vorprozess eingesetzt, der durch alle komprimierten Blöcke geht und die Größe bestimmt bei der die Bit-Streams abgeschnitten werden, um die optimale Bitrate zu erreichen.

Bitstream Organisation

Die Daten die aus der Entropie-Kodierung kommen werden als erstes in Pakete aufgeteilt. Ein Paket besteht aus einer Anzahl von Kacheln innerhalb eines Bezirkes. Dieser Bezirk kann unterschiedliche Größen haben, jedoch muss er die Dimension eines Vielfachen von 2 haben. Die Pakete bestehen aus dem Header und den komprimierten Daten. Durch das geordnete Zusammenfassen der Pakete entsteht ein Stream. JPEG2000 kennt 5 Möglichkeiten die Pakete anzuordnen (z.B. nach Qualität, Auflösung oder Position).

Motion JPEG2000

Motion JPEG2000 ist der 3. Teil des neuen internationalen JPEG2000 Standard. Das MJ2 Format ist für eine Bewegungssequenzen von JPEG2000 Bildern konstruiert wurden. Im Unterschied zu den derzeitigen Videostandards (z.B. MPEG und H.26x) die mit Intra- und Interbildern arbeiten, stützt sich Motion JPEG2000 nur auf Intra-bilder. D.h es wird dabei keine Bewegungskompensation vorgenommen.

Durch die Bereitstellung einer Reihe von Eigenschaften und die Nutzung von modernen Technologien ist Motion JPEG2000 ein flexibles Format. Es eignet sich besonders für qualitativ hochwertige Videos.

2.2 MPEG

Das Ziel der „Motion Pictures Experts Group“ war es einen Standard zu entwickeln der digitale Videodaten in Echtzeit wiedergibt. Bis heute sind 4 Standards von MPEG entwickelt wurden.

Als erstes soll das allgemeine Verfahren gezeigt werden, um dann später die Unterschiede der einzelnen Standards zu zeigen.

Als erstes wird das einzelne Bild in Makroblöcke der Größe 8×8 aufgeteilt. Danach kommt es zur Farbtransformation in den $Y C_r C_b$ Raum. Da das menschliche Auge bekannterweise Helligkeitsunterschiede besser als Farbunterschiede aufnimmt wird eine Unterabtastung der Chrominanzwerte vorgenommen (z.B. 4:2:2). Mittels der diskreten Cosinus-Transformation werden die Makroblöcke vom Ortsbereich in den Frequenzbereich transformiert. Durch die anschließenden Quantisierung werden die Koeffizienten durch einen festen Betrag geteilt. Diese Werte stehen in der Quantisierungstabelle und steigen von links oben nach rechts unten an. Nach der Rundung sieht man das gleiche Bild wie bei JPEG. Im nächsten Schritt wird mit Hilfe des Zick-Zack-Algorithmus die Matrix in einen Vektor geschrieben.

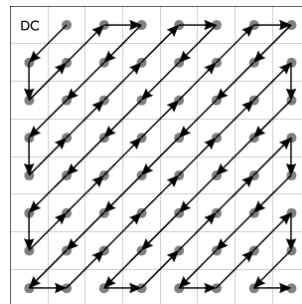


Abbildung 8 Zig-Zag-Algorithmus

Dieser Vektor durchläuft danach zuerst eine Lauflängencodierung und dann eine Huffman-Kodierung.

Im Endeffekt hat sich MPEG sehr stark an den JPEG Algorithmus angelehnt und nur kleine Änderungen vorgenommen.

Bewegungskompensation

Um nun bewegte Bilder zu sehen, wird eine schnelle Abfolge von mind 15 Einzelbildern in der Sekunde benötigt. Die Veränderung zwischen zwei benachbarten Bildern ist jedoch nur gering und meistens ist es so das sich die Objekte vor einem kaum veränderten Hintergrund bewegen.

Diese Bewegung wird herausgefiltert und durch geeignete Bewegungsvektoren dargestellt. Für die Suche nach solchen Bewegungen werden die Bilder wieder in Makroblöcke eingeteilt, da sonst der Aufwand zu groß wäre. Ist eine Ähnlichkeit zu einem Vorgängerbild gefunden, wird nur die Differenz zum vorhergehenden Bild komprimiert. Diese Differenz nennt man Prädiktionsfehler. Da aber die Ähnlichkeit zwischen den beiden Bildern nicht an der selben Stelle sein muss, ist es notwendig die räumliche Differenz zwischen ihnen noch zu ermitteln. Diese Differenz bezeichnet man als Bewegungsvektor.

MPEG-Bildtypen

Im MPEG-Standard existieren 3 verschiedene Kategorien von Bildtypen, die sich durch die Kodierung unterscheiden. Dadurch wird eine freier Zugriff und eine gewisse Flexibilität gewährleistet.

Intra-Pictures (I-Bild)

Die I-Bilder werden durch das JPEG-Format kodiert und komprimiert. Die Bewegungskompensation findet bei ihnen nicht statt, deshalb besitzen sie nur die niedrigste Komprimierungsrate (etwa 7:1). Sie werden als Zugriffspunkte beim Vor- und Rückspielen verwendet. Außerdem dienen sie als Referenz für P- und B-Bilder. Durchschnittlich wird nur jedes 5. Bild als I-Bild kodiert.

Predicted Pictures (P-Bilder)

Für P-Bilder wird ein I- oder P-Bild benötigt, das sie als Referenz benutzen. P-Bilder nutzen die räumlichen und inhaltliche Redundanz mittels der Bewegungskompensation aus. Damit erreichen sie eine weit bessere Komprimierung als Intra-Bilder (etwa 20:1). Jedoch kostet das mehr Zeit bei der En- und Dekodierung.

Bidirectional Pictures (B-Bilder)

B-Bilder sind von vorhergehenden und nachfolgenden Bildern abhängig, da sie ihre Informationen aus beiden Richtungen ziehen. Als Referenz können sowohl I- als auch P-Bilder benutzt werden. Jedoch können sie nicht mehr als Referenzbilder für andere genutzt werden. Dadurch bieten sie den größten Kompressionsfaktor der drei Bildtypen (1:50).

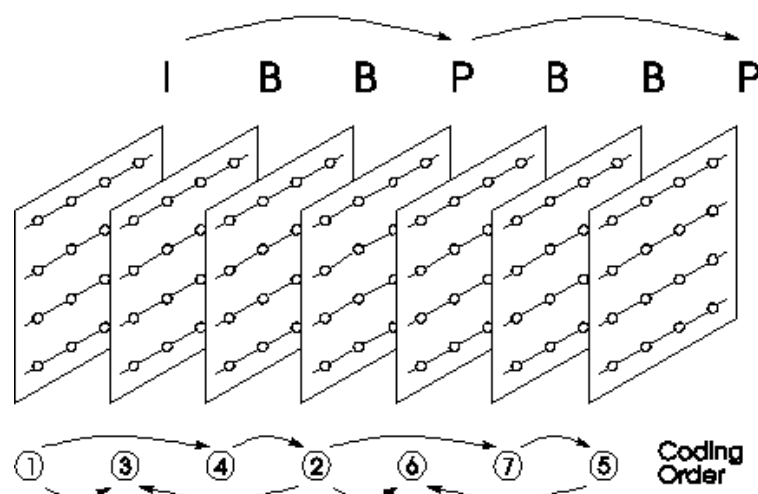


Abbildung 9 Beispiel für den Ablauf der verschiedenen Bilder

MPEG-1

1992 wurde der erste Teil der Standardisierung veröffentlicht. Damit ist die Kodierung von Bewegtbildern für digitale Speichermedien mit bis zu 1,5 MBit/s definiert worden.

Anwendungsbeispiele für MPEG-1 sind Video-CD, Interaktive Videospiele, CD-I usw.

MPEG-2

1994 wurde MPEG-2 verabschiedet und stellt eine Erweiterung zu MPEG-1 dar. Die Kodierung ist im wesentlichen dieselbe wie bei MPEG-1. Es wurde die Makroblockgröße von 8x8 auf 10x10 erhöht und die Genauigkeit der Quantisierung kann bis zu 10 bit eingestellt werden um eine bessere Qualität zu erreichen. Datenübertragungsraten von 1,5 bis 15 MBit/s können genutzt werden und unterstützt nun außer dem progressive scan auch den Interlaced Modus. Dadurch ist es möglich mit MPEG-2 Fernsehbilder zu übertragen. Durch die zusätzlichen Auflösungen ist es außerdem noch in der Lage verschiedene Anwendungen anzusprechen (z.B. normale Fernsehübertragungen, HDTV, DVD-ROM).

MPEG-4

MPEG-4 ist in erster Linie für Multimediale Anwendungen entwickelt worden. Das Ziel war es eine hohe Videoqualität bei Datenraten zwischen 10 Kbit/s und 1 Mbit/s. Die Neuerungen in MPEG-4 lassen sich kurz in 3 Gebiete aufteilen.

Inhaltsbezogene Interaktivität

Ein schnellerer Zugriff auf audiovisuelle Daten ist durch Indizierung, Hyperlinking, Laden, Herunterladen und Löschen von audiovisuellen Objekten gegeben. Durch die Bereitstellung von Syntax und Kodierungsschemata können Objekte, ohne dass die digitale Darstellung bekannt, innerhalb einer Szene verändert werden. Mittels der objektorientierten Kodierung ist ein wahlfreier Zugriff des Anwenders auf Objekte möglich (Ton, Bild und Daten).

Kompression

Es wurde eine höhere Qualität bei vergleichbarer Bitrate erreicht. Zusätzlich besteht die Möglichkeit, Objekte aus verschiedenen Blickrichtungen zu kodieren (Darstellung von 3D-Objekten).

Skalierbarkeit

Die räumliche und zeitliche Auflösung ist in MPEG-4 skalierbar. Außerdem können auch bestimmte Objekte hervorgehoben oder durch eine stärkere Auflösung dargestellt werden. Die räumliche Skalierung funktioniert allerdings nur bei kompletten Frames und kann nicht auf einzelne Videoobjekte angewendet werden.

MPEG-7

MPEG-7 stellt keinen neuen Kodierungsstandard dar, sondern stellt eine Beschreibung von Multimedia-Daten dar. D.h. es ist kein Kompressionsalgorithmus, sondern dient der Verwaltung und Charakterisierung von Multimediadaten.

2.3 H.26x

H.261

Der H.261 Standard ist 1990 von der CCITT (Comite Consultatif International Telegraphique et Telephonique) verabschiedet wurden. Der Komprimierungsalgorithmus ist für Videokonferenz und andere Videodienste im ISDN Bereich bei Bitraten von px64kBit/s entwickelt wurden. H.261 unterscheidet sich im Grundgerüst nicht sehr stark von MPEG, da die Erfahrungen und Entwicklung von H.261 in die Entwicklung von MPEG eingeflossen ist. Das Verfahren unterstützt 2 Bildformate: CIF (352 x 288) und QCIF (176 x 144). Die Bilder sind im $YCrCb$ kodiert und die Unterabtastung liegt bei 4:1:1. Vor der Transformation wird das Bild in 8x8 große Blöcke aufgeteilt. Nach der Transformation mit der DCT und der Quantisierung wird zum Abschluß noch mit Hilfe von statischen Huffman-Tabellen entropiecodiert.

Es sind 2 Bildtypen definiert: zum einen die Intra-Bilder (I-Frame) und zum anderen Inter-Bilder (P-Frames). Bei der Bewegungskompensation wird nur das vorhergehende Bild nach einem Referenzblock durchsucht.

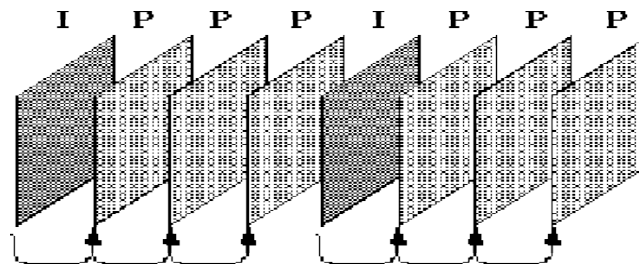


Abbildung 10 Ablauf bei H.261

H.263

H.263 ist eine Weiterentwicklung von H.261. Die größten Unterschiede zwischen den beiden:

1. eine größere Genauigkeit bei der Bewegungskompensation
2. durch eine hierarchische Datenstruktur kann eine bessere Fehlerkorrektur bei niedrigen Bitraten vorgenommen werden
3. anstatt der Huffman-Kodierung wird die Arithmetische Kodierung verwendet
4. Einsatz von PB-Bilder: besteht aus 2 Bildern die als Einheit kodiert werden. Es wird aus einem P-Bild, das durch Prädiktion aus dem vorhergegangenen P-Bild entstand und einem B-Bild aus dem vorhergegangenen und dem gerade kodierten P-Bild generiert.

Außerdem sind noch 3 weitere Bildformate aufgenommen wurden, um einen größeren Anwendungsbereich abzudecken.

H.264

Der H.264 Standard entstand in Zusammenarbeit der Video Coding Experts Group (VCEG) der ITU-T und ISO/IEC Moving Picture Experts Group (MPEG). Durch diese Zusammenarbeit besitzt er noch eine weitere Bezeichnung: MPEG-4 Part 10/Advanced Video Coding(AVC). Die Farbkodierung erfolgt im YUV-Format mit einer Unterabtastung von 4:2:0. Als erstes werden die Bilder wieder in Blöcke aufgeteilt und das erste Bild einer Sequenz ist ein Intra-Frame. Bei der Bewegungskompensation erlaubt es dieser Standard nun

auch eine Auswahl der Referenz aus mehreren Bildern. Dies ergibt eine bewegungskompensierte Langzeitvorhersage und führt zu einer höheren Effizienz der Bewegungskompensation. Dabei sind auch kleinste Bewegungen und Bewegungen ausserhalb des Raumes möglich. Die Transformation erfolgt in Blöcken der Größe 4x4. Statt der DCT wird nun die Integertransformation verwendet, um die Genauigkeit der Berechnungen zu erhöhen. Der Vorteil der Integertransformation liegt in der einfachen Transformationsmatrix und der kostengünstigen Implementierung. Bei der Quantisierung wird eine lineare Skala verwendet. Der H.264 Standard ist besonders für Videokonferenzen und Streaming geeignet.

Zusammenfassung

Alle hier Vorgestellten Standards zeigen besondere Eigenschaften und werden deshalb in verschiedenen Einsatzbereichen angewendet. Während JPEG2000 speziell für den Bereich der Einzelbilder entwickelt wurde, sind die anderen Standards auf den Videobereich genormt.

Im direkten Vergleich zwischen Motion JPEG2000, MPEG-2, MPEG-4 und H.264 muss sich der Standard der JPEG-Group aber nicht verstecken. Denn in Sachen Kompressionseffizienz ist er in fast allen Videoframegrößen vergleichbar (nur bei kleiner Videoframegröße und einer Kompression größer als 1:80 schwächt er ein klein wenig im Gegensatz zu MPEG-2 und MPEG-4). Der JPEG-Group ist zwar ein Vorstoß mit Motion JPEG2000 in den Videobereich gelungen, jedoch zeigen sich einige Nachteile. Die Berechnungskomplexität benötigt etwas mehr Kapazität und durch die fehlende Bewegungskompensation, kommt es außerdem zu einer erhöhten Dateigröße.

Der Vorteil von Motion JPEG2000 liegt in der error resilience evaluation, denn die ist um einiges besser ausgearbeitet. Außerdem hat es meiner Meinung nach mit der Wavelet-Transformation die zukunfrüchtigere Transformation im Gegensatz zu den DCT basierten Standards. Um nähere Information über den Vergleich zu erhalten will ich auf die Paper [4] und [5] verweisen.

Bei der Motion Expert Group werden derzeit große Bemühungen in eine Interframe-Waveletkodierung gesteckt. Aber vor 2006 ist nicht mit einem neuen internationalen Standard zu rechnen.

3. Streaming Formate

3.1 RealNetworks

Die Firma RealNetworks Inc. ist der Pionier in Sachen Streaming-Technologie. So ist es nicht verwunderlich das sie mit einer eigenen Streamingtechnologie für Audio- und Videodaten auf dem Markt vertreten sind. Ihr mittlerweile in der Version 9 vorliegende RealPlayer (wird jetzt RealOnePlayer bezeichnet) kann alle hauseigenen Audio und Videodaten abspielen. Die Spezifikationen über den internen Kompressionsalgorithmus ist ein wohl gehütetes Firmengeheimnis. Man kann aber davon ausgehen das es am Anfang auf der diskreten Kosinus-Transformation basierte. Dies ist nämlich die Basis für alle älteren Bewegtbildkompressionsverfahren. Jedoch durch den enormen Qualitätsgewinn nimmt man an, das zusätzlich die Wavelet- und Fraktalkompression integriert wurden sind.

RealNetworks stellt alle Werkzeuge die man zum Streaming benötigt bereit. In der aktuellen Version (als „Helix“ bezeichnet) werden der HelixProducer, der RealOnePlayer und der Helix Universal Server angeboten. Dabei sind nur die Basic-Versionen kostenlos und die Pro und Gold-Version kostenpflichtig. Seit neustem sind auch einige Teile der Helix Reihe freigegeben und können nach der Registrierung bei Real eingesehen werden.

Das Dateiformat ist eines der am fortgeschrittensten Formate in der Branche. Außer den Video- und Audiodaten ist es auch in der Lage Texte, Animationen und Präsentation wiederzugeben. Ein weiterer Vorteil ist, das es auf fast allen relevanten Betriebssystemen erhältlich ist.

Der Helix Universal Server unterstützt neuerdings auch die Dateiformate und Protokolle der anderen beiden großen Anbieter: Microsoft und Apple. Jedoch ist auf der Empfängerseite der jeweilige Player noch erforderlich.

Der HelixProducer encodiert die Rohdaten von Videokameras oder von anderen Dateiformaten (z.B. MPEG) in das hauseigene RealMedia-Format.

Mit dem sogenannten „SureStream“ ist es ihm möglich verschiedene Bitraten für verschiedene Zielgruppen zu erstellen. D.h. während der Übertragung zwischen den verschiedenen Bitraten zu wechseln, um individuell auf die Verbindungsqualität einzugehen.

Der RealOnePlayer kann neben seinem eigenem Format noch die meisten gängigen Formate (AVI, MPEG, MP3,...) abspielen.

Derzeit ist RealNetworks noch Marktführer in diesem Segment, jedoch Microsoft versucht ihnen diesen Rang abzunehmen (was ihnen in anderen Bereichen in der Vergangenheit schon gelungen ist).

3.2 Microsoft

Microsoft hat erst sehr spät mit der Entwicklung der Streamingtechnologie angefangen, da man die Tragweite nicht sofort erkannte. Jedoch durch den finanziellen Rückhalt und ein großes Engagement ist es ihnen gelungen den Entwicklungsvorsprung der anderen aufzuholen. Im Gegensatz zu den anderen Anbietern ist das komplette Softwarepaket nur für das eigene Betriebssystem verfügbar. Außerdem hält Microsoft leider immer noch nicht viel vom OpenSource Projekt.

Durch den späten Einstieg setzte man zu Anfang auf die enorme Leistungsfähigkeiten der MPEG-4 und MP3 Standards. Ab der Version 7 verzichtet man auf den offenen Standard und entwickelte einen neuen Codec mit verbesserter Bild und Tonqualität. Es ist anzunehmen das dieser eine Weiterentwicklung des MPEG-4 Standards ist. Das WMV-Format von Microsoft ist in der neusten Version (auch Version 9) so flexible wie das RealMedia-Format. D.h. es ist genau wie bei RealNetworks SureStream möglich, um verschieden Bitraten eines Streams in eine Datei zu schreiben.

Der Vorteil am Microsoft Produkt ist, dass es eine integrierte Rechteverwaltung bietet. Das macht es besonders für die Musikindustrie mit ihren Bezahlangeboten im Internet attraktiv.

3.3 Apple

Apple kann auf die längste Tradition im Streaming-Business zurückblicken. Mit QuickTime bieten sie seit Jahren einen offenen Standard für die Erstellung, Integration, Organisation und Vertrieb digitaler Medien. Der QuickTime Player ist sowohl für das eigene als auch für Windows verfügbar. Als Streaming Server bietet Apple ein OpenSource Projekt namens „Darwin“ für alle bekannten Betriebssysteme.

Apple hat, im Gegensatz zu seinen Konkurrenten nie eigene Codecs entwickelt. Sie benutzen sowohl Codecs anderer Hersteller z.B. B.Sorenson als auch offene Standards wie MPEG-4, H.26x, MP3 usw [9].

Die Flexibilität bei QuickTime wird durch den modularen und hierarchischen Aufbau möglich. D.h. eine Datei besitzt eine beliebige Anzahl von Spuren in der Text, Animation, Audio, MPEG usw enthalten sind. Dieses Konzept ist sehr flexible, da sich die Mediadaten beliebig kombinieren lassen und in Tracks organisiert werden.

Anders als bei RealNetworks und Microsoft weicht Quicktime von der SureStream Technologie ab. Bei ihrer Alternate Data Rate werden anstelle von mehreren Spuren in einem Movie verschiedene Movies generiert. Die Verknüpfung der einzelnen Movies geschieht durch ein Referenz Movie. Dieses überprüft vor der Übertragung die voreingestellte Übertragungsrate im Player um die bestmögliche Wiedergabequalität zu gewährleisten.

Durch das modulare Konzept, die langjährige Etablierung und das OpenSource Projekt bietet QuickTime in vielen Punkten ein hohes Innovationspotenzial. Und mit den großen Hollywoodstudios und der strategischen Zusammenarbeit mit RealNetworks hat Microsoft einen weiteren starken Konkurrenten im Bereich Streaming.

Zusammenfassung

Abschließend kann man sagen das alle Hersteller in Sachen Ton- und Videoqualität etwa auf dem selben Niveau liegen. Und mit der Zunahme von Streaming Angeboten im Internet ist anzunehmen das die Hersteller weitere Qualitätsverbesserungen vornehmen.

Bei der Wahl der Server-Plattform hängt es in erster Linie von den Ausstattungsmerkmalen des Produktes ab. Microsoft sticht dabei besonders mit seinem Rechtemanagement, Apple kann mit seinen guten Kontakten zur Filmindustrie und seiner großen Anzahl von Codecs konkurrenzieren und RealNetworks hat den Vorteil auf alle gängigen Betriebssystemen lauffähig sein.

Literaturverzeichnis

- [1] Tilo Strutz, *Bilddatenkompression: Grundlagen, Codierung, JPEG, MPEG, Wavelets*, Vieweg Studium Technik, Braunschweig, 2000
- [2] Offizielle JPEG Homepage www.jpeg.org
- [3] MPEG Homepage www.mpeg.org
- [4] Wei Yu, Ruibiao Qiu, Jason Fritts, *Evaluation of Motion-JPEG2000 for Video Processing*, Department of Computer Science, Washington University, St. Louis, 6.November 2001
- [5] Detlev Marpe, Valeri Georgev, Hans L. Cycon, und Kai U. Barthelb, *Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode*, Fraunhofer-Institute for Telecommunications, Heinrich-Hertz-Institute (HHI), Fachhochschule für Technik und Wirtschaft (FHTW), University of Applied Sciences
- [6] Offizielle RealNetworks Homepage, www.realnetworks.com/products/codecs/realvideo.html
- [7] Offizielle Apple Homepage, <http://developer.apple.com/referencelibrary/QuickTime/>
- [8] Offizielle Microsoft Homepage, <http://www.microsoft.com/germany/ms/windowsxp/warum/top10/imaging/mediaplayer9/>
- [9] Tobias Künkel, *Streaming Media*, Addison-Wesley, 2001
- [10] M. Rabbani und D.Santa Cruz, *The JPEG2000 Still-Image Compression Standard*, Oktober 2001
- [11] *FKT, Fachzeitschrift für Fernsehen, Film und Elektronische Medien*, Schiele&Schön 05/2004