

# Motion Pictures

Die Motion Pictures Expert Group ([www.mpeg.org](http://www.mpeg.org)) entwickelte die MPEG-Standards für digitales Video, nach denen nicht nur Videobilder, sondern auch die dazugehörigen Audiosignale komprimiert werden. Während MPEG-2 jetzt schon für DVD und Digitalfernsehen eingesetzt wird, ist MPEG-4 als universelles Format für jegliche digitale Übertragung gedacht. Die Kompressionsraten steigen nochmals deutlich...

Werner Krause

Als erstes Ergebnis wurde 1991 der Videokompressionsstandard MPEG-1 vorgestellt, der ursprünglich für die Datenübertragungsrate eines CD-ROM bei Single-Speed mit 1400 Bit/s vorgesehen war. Video nach MPEG-1 mit einer Auflösung in PAL 352x288 Pixel - zeigt also das Videobild nur mit einem Viertel der Originalgröße - unterstützt ausschließlich eine konstante Geschwindigkeit beim Auslesen der Daten.

Schon MPEG-2 als Nachfolger (seit 1994) erlaubt variable oder adaptive Datenraten - die Datenmenge wird beim Auslesen dem tatsächlichen Bedarf der Bildinhalte angepasst. Digitale Fernsehstudioqualität im Bereich bis zu 8000 Bit/s wird derzeit auch erfolgreich für hochauflösendes Fernsehen (HDTV) eingesetzt. Ursprünglich war für hochwertige HDTV-Signale ein MPEG-3 Format angedacht. Doch umfangreiche Tests zeigten, dass der MPEG-2 Standard ausreichend gute Qualität lieferte. MPEG-3 wurde deshalb nie eingeführt. Alle weltweit laufenden Projekte zur Einführung neuer digitaler Normen setzen seitdem auf MPEG-2.

1999 wurde MPEG-4 verabschiedet. Vorgesehen für Videoübertragungen bei geringer Bandbreite (Echtzeit-Anwendungen, Netmeeting, Internet-Streaming) lag der Schwerpunkt mehr auf extremer Datenreduktion als in der Optimierung der Bildqualität.

## Physiologie

Ein wirklich hoher Komprimierungsfaktor wird in erster Linie durch Weglassen wenig relevanter Reizinformationen (gemessen an der physiologischen Wahrnehmungsfähigkeit) erzielt. Im optischen und akustischen Spektrum sind die menschlichen Sinne nicht linear empfindlich, sondern reagieren in Folge von bereits aufgenommenen Reizen dynamisch. Solche Wahrnehmungstoleranzen werden zur Datenreduktion genutzt, die sich durch komplexe Filterfunktionen bewerkstelligen lässt. Dazu müssen die Bildinformationen transformiert werden.

Zwischen aufeinanderfolgenden Bildern eines Films besteht eine sehr hohe Ähnlichkeit, sie korrelieren in weiten Bereichen miteinander. Die Kompression setzt zunächst zwei Teilschritte voraus: DCT (*Discrete Cosinus Transformation*) und Quantisierung.

## Diskrete Cosinus Transformation

Die Frames (Einzelbilder eines Videos) werden in Blöcke zu je 8 x 8 Pixel unterteilt. Auf jeden Block kommen 64 Bildpunkte, die einer mathematischen Umformung unterzogen werden. Das Ergebnis der DCT (Diskrete Cosinus Transformation) sind je Block 64 Koeffizienten.

Die Informationen über Helligkeitswerte werden in eine Frequenz-Darstellung umgewandelt. Diese Transformation selbst bleibt bis auf Rundungsfehler verlustlos und bringt noch keine Reduktion der Datenmenge. Die 64 Bildpunkte pro Block werden in eine Form übertragen,

## Bildkompression nach einem Modell aus der Wahrnehmungsphysiologie:



Von jedem Pixel wird die Helligkeit definiert,



die Farbinformationen aus jeweils 2 x 2 Bildpunkten interpoliert.



Zusammengesetzt ergibt sich eine Halbierung der Bildinformation ohne wesentlichen Informationsverlust.

die es einem Kompressionsalgorithmus möglich macht, Redundanzen zu erkennen und zu reduzieren.

## Quantisierung

Die 64 Koeffizienten werden quantisiert: Das Verfahren entfernt vorrangig jene Bildinformationen, die unter der Wahrnehmungsschwelle liegen. Bestimmte Anteile innerhalb der Pixelblöcke können vernachlässigt werden. Die Quantisierung erfolgt mit Hilfe spezieller Matrizen, die in Versuchsreihen ermittelt wurden. Sie stützen sich auf psychovisuelle Modelle, die die Motion Picture Expert Group von der menschlichen Wahrnehmung entwickelt hat. Im Prinzip ist MPEG eine Weiterentwicklung des JPEG-Standards.

Da das Auge für Unterschiede in der Helligkeit empfindlicher ist als für Farbunterschiede, wird auf vier Helligkeitsinformationen nur eine Farbinformation gespeichert, bei der Rekonstruktion der Bilder werden die dazwischenliegenden Pixel interpoliert. Stellt man die Werte in Byteform dar, so ergibt sich hierdurch schon eine Halbierung der Bildinformation ohne wesentlichen Informationsverlust. Speziell bei der Farb- und Helligkeitswahrnehmung werden die Pixelinformationen, die im Grafikbereich typischerweise im RGB-Modell (rot/grün/blau) vorliegen, in eine YCbCr-Darstellung transformiert. YCbCr ist ein Akronym für die Luminanz- und Chrominanzkomponenten. Dabei wird der Tatsache Rechnung getragen, dass das menschliche Auge eine besonders hohe Grün-Empfindlichkeit aufweist, während die Empfindlichkeit gegenüber Rot geringer und gegenüber Blau am geringsten ist. Bei dieser Transformation wird das RGB-Farbsignal in seinen Helligkeitsanteil (Y) und in zwei Farbanteile (Cb und Cr) zerlegt.

Es sind einerseits Erkenntnisse zur subjektiven Farb- und Kontrastsensibilität, aber auch rein technische Grenzen, die die Vorgaben zu verschiedenen Filterfunktionen beeinflussen: Details, die in ihrer Frequenz das Darstellungsraster übersteigen, können nicht mehr richtig wiedergegeben werden, sie verursachen oft störende Effekte (z.B. Moiree) und tragen nicht mehr sinnvoll zur korrekten Bildinformation bei. Nach dem Herausfiltern dieser Bereiche bleiben beim Speichern aufeinanderfolgender Einzelbilder nur mehr jene Daten erhalten, die innerhalb einer Sequenz qualitativ für das Wahrnehmungsvermögen von Bedeu-



tung sind. Nach solchen Kodierungskriterien lassen sich je nach Art der Bildinformationen mittlere Kompressionsraten von 1:100 bis fast 1:800 erreichen.

Die Quantisierung ist immer verlustbehaftet und dafür verantwortlich, dass sich an Kontrastkanten beim Dekomprimieren oft störende Artefakte (Pixelblöcke) bemerkbar machen. Da sich Bildinformationen sprunghaft ändern können, divergieren auch die DCT-Koeffizienten entsprechend stark. Dabei stößt die Dekompression (*Inverse DCT*) bald an mathematische Grenzen.

Neben einer rein statischen Differenzkodierung müssen bei MPEG auch Objekt- und Kamerabewegungen berücksichtigt und entsprechend verarbeitet werden.

### Motion Compensation

Ein speziell entwickeltes Verfahren (Motion Compensation) reduziert sich wiederholende - und deshalb unnötige - Informationen aus identischen Bildteilen aufs Notwendigste. Der Hauptanteil aller Einzelbilder im Videostrom wird in Differenzblöcken mit Bewegungsvektoren definiert.

Ein modifizierter JPEG-Algorithmus kodiert drei verschiedene Arten von Einzelbildern auf dem Weg zum MPEG-Video:

- Die **I-Frames** (*intra*) sind völlig eigenständige Bilder, die unabhängig von anderen Bildern wie einzelne JPEG Bilder komprimiert werden. I-Frames tauchen regelmäßig in MPEG-Datenströmen auf, auf die der Decoder aufsetzen kann, ohne die vorangegangenen Bilder analysieren zu müssen. I-Frames werden benötigt, um Szenenwechsel einwandfrei darzustellen und aufsummierte Rundungsfehler zu vermeiden.



**I-Frame: Intraframes sind Vollbilder, entsprechen einer JPEG - Kompression.**



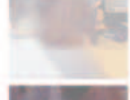
**B-Frames (Bidirectional predicted Interframes): enthalten am wenigsten Daten, werden aus I-Frames und P-Frames abgeleitet.**



**P-Frame (Predicted Interframe): speichert nur Veränderungen zu I-Frames.**



**B-Frames machen in ihrer Gesamtmenge den größten Anteil des Datenstroms aus. Der Encoder übernimmt Bildinformationen aus dem P-Frame und dem I-Frame aus der nächsten GOP.**



**I-Frame der nachfolgenden GOP dient als Referenz vorangegangener B-Frames.**

1



2



3



4



1 **GOP: Eine Anzahl von Videoframes wird zu einer Group Of Pictures zusammengefasst.**

2 **Picture in einer bestimmten Position innerhalb der GOP, in Makroblöcke unterteilt.**

3 **Ein Makroblock umfasst 16 x 16 Pixel und besteht aus 4 Blöcken.**

4 **Ein Block 8 x 8 Pixel enthält die Werte für Helligkeit und Farbe.**

- **P-Frames** (*predicted*) sind Einzelbilder, die lediglich die Differenz zum vorangegangenen I-Frame enthalten.
- **B-Frames** (*bidirectional*) werden ausgehend von einem vorangegangenen und dem folgenden Bild berechnet.

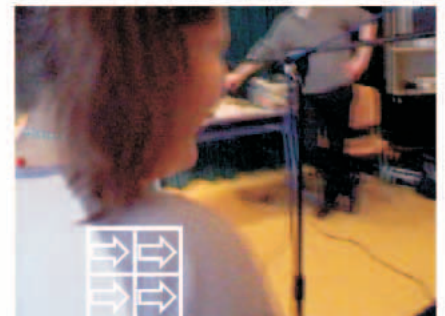
Man unterscheidet zwischen referenzierten Bildern (Interframes) und solchen, welche nur auf sich selbst verweisen, sogenannte Intraframes (Originalbilder). Sie sind im normalen Ablauf notwendig, um einerseits die Abbildungsfehler nicht zu groß werden zu lassen, andererseits, um Bildsequenzen direkt innerhalb des Videostroms anzuspringen, oder sie schneller vorwärts oder rückwärts abzuspielen. Es wird zwischen Predicted Frames (P-Frames) und Bidirektional Predicted Frames (B-Frames) unterschieden. Die P-Frames orientieren sich an Intraframes oder voranlaufende P-Frames, die B-Frames an das je nächste vor- und nachlaufende Intra- oder P-Frame.

Je nachdem, ob ein P-Frame auf ein nachfolgendes oder voranlaufendes Bild referenziert wird, spricht man von einem Forward- oder Backward-Predicted-Frame. Da ein MPEG-Decoder B-Frames nur darstellen kann, wenn er den nachfolgenden referenzierten P-Frame kennt, müssen die Bilder in einer geeigneten Reihenfolge abgespeichert werden - die nicht immer der Reihenfolge der Darstellung entspricht. In der Praxis hat die doppelte Referenz einen immensen Vorteil: Der Fehler zwischen zwei aufeinanderfolgenden I-Frames nimmt nicht stetig zu, sondern hat sein Maximum in der Mitte, außerdem wird das Rauschen kleiner. Nachteil ist der erhöhte Rechenaufwand, der bei asymmetrischen Anwendungen für den Endanwender jedoch kaum ins Gewicht fällt. Asymmetrische Anwendungen zeichnen sich durch aufwendige Codierung, aber schnelle Dekodierung aus.

Die Voraussage von P- und B-Frames erfolgt bewegungskompensiert. Dazu wird jedes Bild in Makroblöcken bestehend aus

16 x 16 Bildpunkten eingeteilt. Bei jedem Makroblock in P- und B-Frames ist getrennt vermerkt, ob er völlig neu kodiert ist oder nur die Differenz zum Vor- und/oder Nachfolger-P/I-Frame enthält. Sowohl für das Vorgänger- als auch für das Nachfolgebild kann angegeben werden, um wie viel dort der Makroblock verschoben werden muss, bevor interpoliert wird.

Die Verwendung der jeweiligen Bildtypen bestimmt wesentlich sowohl Qualität als auch Kompressionsgrad des MPEG-Videos. So verbessert man durch I-Frames die Qualität, während durch B-Frames der Kompressionsgrad erhöht wird. Insbesondere ist der Abstand zweier



**Der Encoder fasst Makroblöcke zusammen, die bei Bewegungen über mehrere Frames hinweg verschoben werden. Verschiebungsvektoren reduzieren die Datenmenge erheblich.**



I-Frames im Datenstrom ein Qualitätsmaßstab.

Ein MPEG-Datenstrom besteht aus Sequenzen von *Groups-Of-Pictures* (GOP). Jede GOP enthält genau einen I-Frame als erstes Bild.

### Makroblock

Eine Vielzahl von Bewegungen lassen sich in Form einfacher Verschiebungen in der Ebene beschreiben. So besteht ein langsamer Kameraschwenk zum größten Teil aus verschobenen Bildbereichen, nur der Rand enthält wirklich neue Daten. Alle übrigen Bildbereiche lassen sich im Prinzip für jeden Makroblock (16x16 Pixel, umfasst 4 Blöcke) in einem einzigen Verschiebungsvektor ohne Informationsverlust ausdrücken. Gewisse Toleranzen innerhalb von Makroblöcken sind vorgesehen: Bei geringen Abweichungen wird auf eine Angleichung des Bildinhaltes vollständig verzichtet und nur der Bewegungsvektor übertragen. Bewegt sich der Fehler innerhalb bestimmter Grenzen, so wird nur die Differenz des betreffenden Makroblocks mitkodiert und übertragen. Bei stärkeren Änderungen wird auf den Verschiebungsvektor verzichtet und der Block als Ganzes neu kodiert.

### Run Length Encoding

Die RLE-Kodierung (übersetzt: Lauflängenkodierung) ist eines der einfachsten und am häufigsten eingesetzten verlustfreien Kompressionsverfahren. Es beruht darauf, dass sich in Bildern Pixel ein und derselben Farbe häufig wiederholen. Da ein Bild bei seiner Digitalisierung in einen einzigen langen Datenstrom übergeht, wird nicht jeder identische Punkt einzeln kodiert, sondern nur einmal zusammen mit der Häufigkeit seiner Wiederholungen und der Koordinaten. Bei größeren Datenmengen führt die RLE-Kompression zu guten Ergebnissen.

### Huffmann Kompression

Eine verlustfreie und variable Längenkodierung, die die Häufigkeitsverteilung benutzt, um Binärcodes von Bildpunkten, die sich oft wiederholen, zu verkürzen. In einer Hierarchie (Huffmann-Baum) werden gleichwertige Pixel, die die maximale Fläche im Bild abdecken, in kürzester Bitlänge beschrieben. Übliche Kompressionsraten liegen zwischen 15 und 75 Prozent.

### Encoder

Bei der Auswahl der Quantisierungstabelle, der Erstellung von Bewegungskompensationsvektoren und beim Interpolieren in Makroblöcken muss der Encoder sehr viel Rechenleistung investieren. Zudem wird der entstandene Stream zwei weiteren - diesmal verlustfreien - Standardkompressionsverfahren unterzogen: der RLE-Kodierung - sie nutzt die durch die Quantisierung vervielfachten Wiederholungen von gleichen Pixeln aus - und einer modifizierte Huffmann-Kompression.

Die Bildqualität hängt stark vom Encoder getroffenen Aufwand ab. Der Standard

definiert nur die Funktionalität eines Decoders, sodass große Freiräume für die Hersteller bestehen. Daraus erklären sich auch die deutlichen Qualitätsunterschiede.

Ein Encoder kann in einem Modus mit variabler Bitrate (VBR) arbeiten, bei dem die Quantisierungstabellen konstant sind und die anfallende Datenrate von der Komplexität des Bildes abhängt. Beim MPEG-2 Standard kommen unter anderem noch Möglichkeiten der getrennten Behandlung von geraden und ungeraden Zeilen hinzu, um mit Quelldaten im Zeilensprungverfahren (Interlacing) zurechtzukommen. Statt 25 kompletten Bildern speichern Videoaufzeichnungsgeräte nämlich für jede Sekunde 50 Halbbilder. In einem Halbbild sind nur die geraden, im andern nur die ungeraden Zeilen enthalten.

Ferner besteht die Möglichkeit, auf einem Kanal ein grobes Bild zu übertragen und auf einem zweiten Kanal das Differenzsignal zum qualitativ hochwertigen Bild. MPEG ist daher qualitativ skalierbar. Dadurch wird es beispielsweise möglich, beim digitalen Fernsehen zusätzlich zu den normalen Fernsehsignalen noch ein hochauflösendes 16:9 Format (HDTV) auszustrahlen, ohne übermäßig mehr Bandbreite zu belegen. Ähnlich arbeiten auch DVDs, auf denen ein Breitwandbild mit allen Zeilen abgelegt wird.

### Decoder

Die Decodierung ist weitaus einfacher, weshalb MPEG als asymmetrisches Verfahren bezeichnet wird. Dies ist auch sinnvoll, da die Komprimierung in der Regel nur ein einziges Mal stattfindet, die Decodierung jedoch bei jeder Wiedergabe.

### MPEG-2 Profile

MPEG-2 baut grundsätzlich auf den Videokompressionsroutinen von MPEG-1 auf und bietet eine ganze Reihe sogenannter "*Coding Tools*" an. Diese werden in Profilen gruppiert, um verschiedene Funktionalitäten zu realisieren. Aus Gründen der Standardisierung muss der MPEG-2-Decoder den Qualitätsbereich von ISDN-Videokonferenz über DVD-Wiedergabe bis zur Satellitenübertragung abdecken.

MPEG-2 beinhaltet auch eine Audiokompression, die bei 20% der Originalgröße immer noch sehr gute Audioqualität erlaubt. Die Datenrate bei einer typischen MPEG-2-Wiedergabe für Bild und Ton liegt bei etwa 1,2 MBit/s.

### MPEG-4

Für Anwendungen niedriger und niedrigster Bandbreite wurde an MPEG-4 wurde schon seit 1996 für den Mobilfunkbetrieb intensiv geforscht. In Kombination der MPEG-1 und -2 Techniken mit zusätzlichen Multimedia-Elementen kam eine spezielle Version der DCT (*Discrete Cosinus Transformation*) zum Einsatz, durch die die Algorithmen der *Motion Compensation* und Audiokompression weiterentwickelt werden konnten.

Mit MPEG-4 kodierte Videos erfahren grundsätzlich eine dramatische Reduktion des Platzbedarfs. Während für einen Kinofilm mit 2 Stunden Spieldauer im MPEG-2-Verfahren eine DVD mit rund 8GB Kapazität nötig ist, kann der gleiche Film in einer VHS deutlich übersteigenden Qualität und in Stereo-CD-Ton auf einer gewöhnlichen CD-R mit 650MB gespeichert werden.

Erstmals wird bei einem Bild neben der Bewegung (Motion) und der Textur auch die sogenannte Form (Shape) übergeben. Damit ist es möglich, mit Hilfe von "Sprites" beliebig konturierte Standbilder über die sich bewegenden Bilder zu schieben. Sprites können sehr hoch komprimiert werden.

MPEG-4 vereint zahlreiche unterschiedliche Ansätze, multimediale Inhalte komprimiert darzustellen, mit extremer Fehlertoleranz (durch rückwärts lesbare Codes), die für den mobilen Funkbetrieb unerlässlich ist. Außerdem gestattet MPEG-4, bei der Übertragung von Standbildern im Codec von DCT zu Wavelet zu wechseln. Dadurch lässt sich die Kompressionsrate gegenüber der DCT um mindestens 30-50% verbessern. Die von JPEG her bekannten blockartigen Artefakte treten nicht mehr auf. Dazu kommt die Unterstützung skalierbarer Information, bei der Bilddaten in verschiedenen Auflösungen in einem Bitstrom gleichzeitig gesendet werden.

Die Tatsache, dass MPEG-4 durch reine Software-Decoder auf einem üblich ausgerüsteten Büro- oder Heim-PC in Echtzeit bildschirmfüllend abgespielt werden kann, verhilft diesem Dateiformat derzeit zu einer weltweit rasanten Verbreitung. Die aufkommenden schnellen Internet-Verbindungen machen ab rund 512 kBit/s Datenrate auch Video-On-Demand-Anwendungen in Echtzeit (virtuelle Videothek, Video-Datenbanken, Internet-Multimedia-Präsentationen, etc.) möglich.

Als Ausgangsmaterial zur Encodierung in MPEG-4 können grundsätzlich alle analogen und digitalen Datenquellen genutzt werden. Analoge Quellen müssen dabei vordigitalisiert werden. Wie komplex die Funktionalität von MPEG-4 ist, zeigt sich per Definition in über 40 Profilen, die spezielle Anwendungsmöglichkeiten darstellen. Darunter befinden sich Profile für Audio, Video, Grafik und 3D-Darstellung.

### Quellen

- Jörg Behrend, Multimediasprachen und Standards im Internet, MPEG- Datenformate für Videosignale, Universität Tübingen, 2001 (PDF-Dokument)
- A. Dombi Engineering, Publizieren in MPEG 4, Stäfa, Schweiz ([www.adenet.ch/MP4video.htm](http://www.adenet.ch/MP4video.htm))
- Björn Eisert, Was ist MPEG? Ein Erklärungsversuch, 1995 ([www.cybersite.de/german/service/tutorial/mpeg](http://www.cybersite.de/german/service/tutorial/mpeg))
- Daniel Warner, Videokomprimierung & DVD-Ripping von Anfang an, 2002 Planet Intermedia, ISBN 3-8155-6044-6