

1. JPEG

1.1 JPEG Kompression

JPEG ist die Abkürzung für Joint Photographic Experts Group und wurde in den späten 80zigern gegründet. Es ist ein weitverbreiteter Bildkompressionsstandard, der in erster Linie Einzelbilder oder Mehrfachbilder (Motion-JPEG) in der Datengröße reduzieren sollte, ohne einen allzu großen Qualitätsverlust davonzutragen.

Bevor jedoch der Algorithmus angewandt werden kann, müssen die analoge RGB Werte auf digitale Werte (Y Luminanzwert = Helligkeitswert, U und V Chrominanz) umgerechnet werden. Dies erfolgt mit folgenden Formeln:

$$Y = 0,299 * R + 0,587 * G + 0,114 * B$$

$$U = 0,169 * R - 0,3316 * G + 0,5 * B$$

$$V = 0,5 * R - 0,4186 * G - 0,0813 * B$$

Der Y- Anteil ist von Wichtigkeit, da sich mit diesen Signal bereits SW-Bilder erkennen lassen. Der G- Anteil (grün) ist sehr dominant ausgeprägt, da das Auge im Grünbereich am empfindlichsten ist. U und V sind Farbwerte. Helligkeitsveränderungen werden stärker wahrgenommen als Farbänderungen. Bei JPEG werden die Chrominanzwerte mit halber Auflösung abgetastet, da sie nicht von allzu großer Relevanz sind.

Zwei wichtige Arten sind definiert:

1.1.1 Baseline Verfahren

Die Y, U und V Werte werden in 8x8 Pixel Blöcken aufgeteilt und einer zweidimensionalen DCT (diskreter Cosinus Transformation) unterworfen. Dabei werden die Werte des 8x8 Pixelblocks von einem zeitkontinuierlichen Signal in ein Frequenzspektrum übergeführt.

Formel für 2D DCT:

<p>Formel 1: 2D FDCT</p> $C(u, v) = \frac{1}{4} C_u C_v \sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2y+1)u\pi}{16} \cos \frac{(2x+1)v\pi}{16}$
<p>Formel 2: 2D IDCT</p> $f(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C_u C_v C(u, v) \cos \frac{(2y+1)u\pi}{16} \cos \frac{(2x+1)v\pi}{16}$
$C_u, C_v = \frac{1}{\sqrt{2}} \text{ für } u, v = 0 \quad C_u, C_v = 1 \text{ für } u, v = 1 \dots 7$

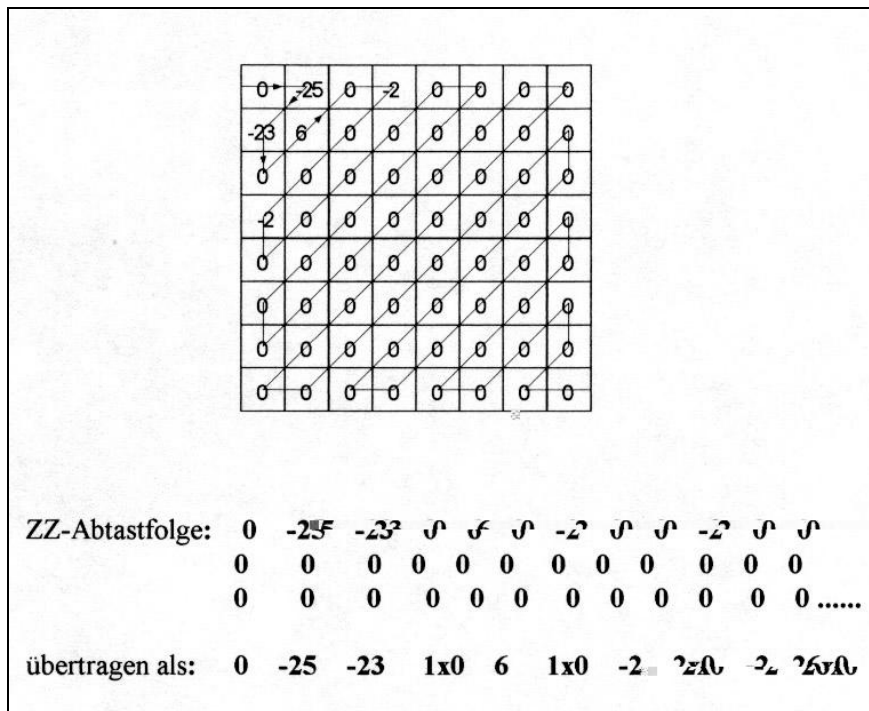
Diese 3 8x8 Matrizen werden nun in 8x8 DCT- Koeffizientenmatrizen umgewandelt. Nun entsteht eine Matrix die folgend interpretiert werden kann: Ganz links oben sind Werte für grobe Strukturen untergebracht (NF), und unten ganz rechts sind Details und feine Strukturen vorhanden. Die Werte links oben sind nun wichtiger als rechts unten.

Die DCT ist eine umkehrbare Funktion und bewirkt keinen Informationsverlust sowie auch keine Datenverdichtung.

Der nächste Schritt ist die Quantisierung (Stufung der Kompressionsrate): Hochfrequente Teile (viele Details) können mit einer höheren Quantisierungsstufe „vernichtet“ werden, wobei das Auge sowas nicht wahrnimmt (im besonderen für die Chrominanzwerte).

Strukturlose Teile stellen eine DC dar. Viele Werte der rechten unteren Ecke werden zu Null quantisiert. Eine Datenverdichtung hat noch nicht stattgefunden. Es entsteht somit eine Quantisierungsmatrix. Im nächsten Schritt wird die Matrix im ZigZag-Scan (siehe nächste Seite) abgetastet. Die Koeffizienten werden aneinandergereiht. Begonnen wird mit den DC-Werten, gefolgt von niederfrequenten AC- und hochfrequenten Werten.

Zig Zag-Scan:



Nun können Kodierungs-Verfahren zum Einsatz kommen, da immer mehr Nullfolgen auftreten. So werden z.B. 20 Nullen durch die Zahl 20 ersetzt. Die Huffmann – (ähnlich dem Morsealphabeth - häufig vorkommende Werte durch Codes ersetzen) und Variable Length Kodierung (VLC) wird hierbei verwendet. Am Ende der letzten Null wird ein EOB (End of Block) geschrieben, dass dem Dekodieren mitteilt, dass der Rest nur Nullen sind.

1.1.2 Lossless Verfahren

Dieses Verfahren basiert auf einem komplett anderen Algorithmus bei der Kodierung. Hierbei wird nämlich eine Vorhersage getroffen über die 3 nächsten kausalen Pixel-Nachbarn, wobei 7 verschiedene Prediktoren definiert sind. Der Vorhersage Fehler ist entropisch kodiert mit Huffmann (Entropie= die Messung von Zufälligkeiten). (siehe auch MPEG-2 Verfahren)

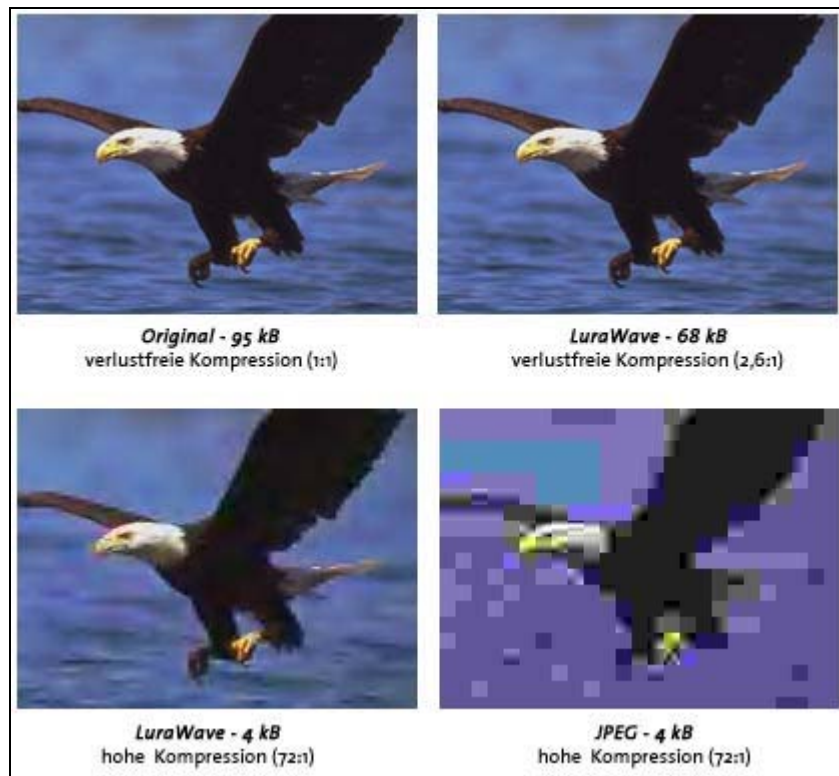
1.2 JPEG 2000 Kompression

Dieses Verfahren befindet sich im Moment noch in der Entwicklung. Der Standard sollte in Herbst 2001 eingeführt werden. Die Vorteile von JPEG2000 sind:

- Höherer frei wählbare Kompressionsrate
- Geringe Ladezeiten
- Keine blockartigen Artefakte wie bei JPEG (bei hoher Kompression)

- Passwortschutz für verschiedene Auflösungen
- Qualitativ hervorgehobene Bereiche eines Bildes (Portrait: Gesichtsbereich mehr Details)
- Nachladen von Bildqualität auf Benutzerwunsch

Vergleich: JPEG – JPEG 2000 (LWL-Format der Firma Luratech) LWL (Luratech Wavelet Dateiformat)



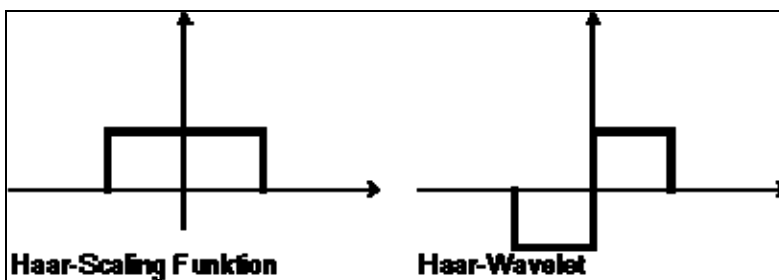
Das JPEG2000 Verfahren basiert auf dem Wavelet Verfahren. Hierbei wird eine sogenannte Wavelet- Transformation durchgeführt, die ähnlich der DCT ist.

Was ist ein Wavelet?

Ist eine Art Ersetzungsmethode der Fourier Transformation, wobei der Fehler (Fensterbildung), welcher entsteht, nahe Null ist. Familien solcher Funktionen heißen Wavelets.

Die Basisfunktion sind Scaling –Funktionen und Wavelets. Die Analyse eines Bildes beginnt mit einem „mother wavelet“, welches sich über das gesamte Bild erstreckt. Dies erlaubt eine Analyse von Bilddaten ohne Fenstereffekte. Die einfachste Wavelet- Funktion ist das Haar Wavelet.

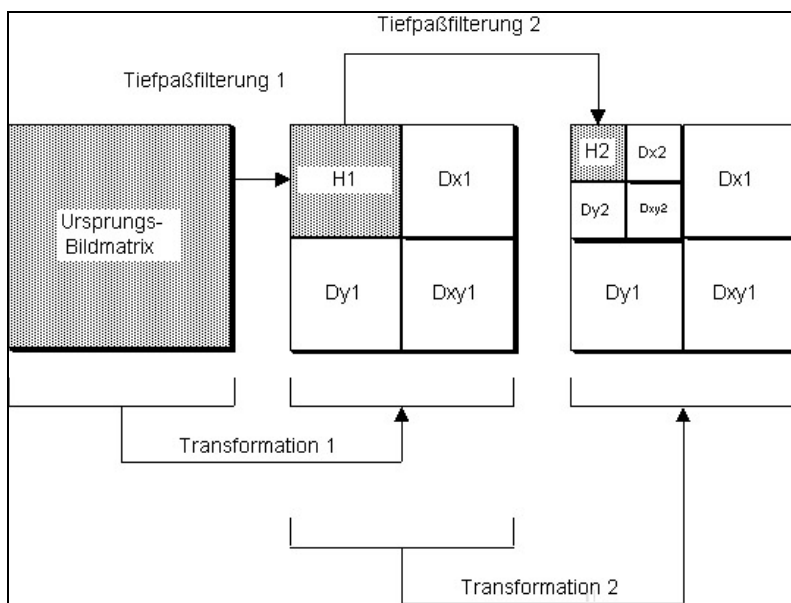
Faltung:



In diesem einfachen Fall bedeutet eine Analyse (eine Faltung) mit den Haar-Funktionen die Berechnung des Mittelwertes und der Differenz zwischen zwei benachbarten Pixelwerten. Die Ergebnisse werden unterabgetastet und als Tiefpaß und Hochpaßanteile gespeichert. Der Tiefpaßanteil wird mit den Haar-Funktionen weiter analysiert.

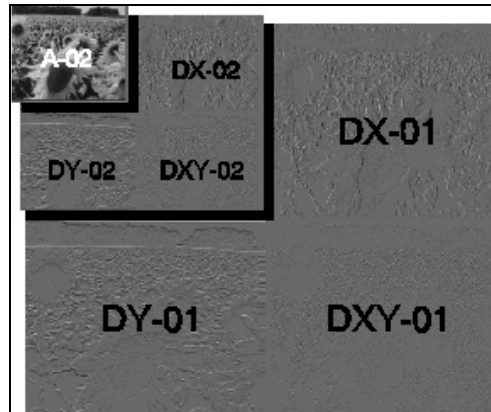
Zu guter Letzt ist das wavelet-transformierte Bild durch eine gewisse Anzahl immer kleiner werdender Hochpassanteile und einen Tiefpassanteil gekennzeichnet.

Prinzip des Haar-Wavelets:



Die Ursprungsmatrix wird in 3 Hochpassteile (Dx_1 , Dy_1 , Dxy_1) und in einen Tiefpassanteil (H_1) zerlegt. Die Zerlegung erfolgt nun rekursiv (wieder jeweils 3 HP und 1 TP-Anteil). Die Hochpassanteile der ersten Stufe bleiben jedoch erhalten. Das Ursprungsbild wird hierarchisch in immer kleinere Teile zerlegt. In den Hochpassanteilen der ersten Stufe werden feine Strukturen erkannt, in den darauf folgenden gröbere.

Beispiel:



Das Ursprungsbild (links) ist in drei Transformationsstufen zerlegt worden. Das kleine Bild im rechten transformierten Abbild stellt den Tiefpaßanteil der dritten Transformationsstufe dar. Alle übrigen Abbildbereiche repräsentieren Hochpaßanteile.

Aber erst andere spezielle Funktionen für die Transformation (z.B.: Daubechies Wavelets oder biorthogonale Wavelets) führen zu weitaus besseren Resultaten als JPEG.

Um eine Aussage über die Qualität eines Bilddatenkompressionsverfahrens treffen zu können wird der **PSNR** (Peak-Signal-to-Noise-Ratio) verwendet. Er ist ein quantitatives Maß für den Fehler, der durch die Bildveränderung hervorgerufen wurde.

Formel:

$$PSNR = 10 \cdot \log \frac{peak^2}{noise}$$

Peak:

$$peak = \max[f(x,y)]$$

Noise:

$$noise = \frac{1}{N \cdot M} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} [f(x_i, y_j) - \tilde{f}(x_i, y_j)]^2$$

$f(x,y)$ ist das Originalbild

$\tilde{f}(x,y)$ das Bild aus den komprimierten Daten

Weiters ist auch der optische Eindruck von Wichtigkeit, der wiederum sehr gut sein kann, obwohl der PSNR schlecht ist.

Die Wavelet Transformation ist abhängig von einigen Parametern. Dazu gehören, die Anzahl der Quantisierungsstufen für die Koeffizienten der Transformation sowie die Auswahl der Filterfunktion.

2. MPEG

2.1 MPEG Audio Layer 3

MPEG (Moving Pictures Experts Group) ist ein Kompressionsstandard für Video und Audiodateien.

Das MPEG Audio Layer3 Verfahren kurz MP3 wurde 1987 an der Universität von Erlangen entwickelt, um Musikstücke zu komprimieren. Hierbei werden oftmals Audio CDs, deren Datenaufkommen nahezu enorm ist (1min = 10,4MB!) bis auf ein Zwölftel komprimiert, „ohne einen Qualitätsverlust“ zu haben. Um diese hohe Kompressionsrate zu erreichen bedarf es einen eigenen Verfahren namens „*Perceptual Audio Coding*“. Dieses Verfahren nutzt die Gegebenheiten des menschlichen Hörapparates, z.B. den Verdeckungseffekt.

Das Signal wird in 32 Teilbänder zerlegt, wobei die Normierung der verschiedenen Pegel in den Teilbändern mittels Skalenfaktoren (jedes Teilband von 2 bis 15 Bit) quantisiert wird. Die Bitzuweisung der einzelnen Bänder übernimmt das psychoakustische Modell des Kodierers. Um eine präzise Auslösung des Frequenzspektrums zu erhalten, wird die schnelle Fourier Transformation (FFT) eingesetzt. Zusammen mit den Skalenfaktoren kann nun die Bitzuweisung erfolgen. Steuer- und Synchronisationsfunktionen werden in den Datenstrom „gemultiplext“. Um nun eine Kompressionsrate von 8-12 zu erhalten, werden die 32 Teilbänder auf 576 *nichtgleichbreite* erhöht. Die Quantisierung erfolgt anschliessend nicht linear sondern gemäß Kennlinien, an Anlehnung an das menschliche Gehör. Zum Schluss kommt noch die Huffman Kodierung zum Einsatz.

Vergleich von MPEG Layer 3 zu anderen MPEG Audio Kompressionsverfahren:

Verfahren	Kompression	Datenmenge	Anwendungsbereiche
Originaldaten	keine, 1:1	1411,200 KBit/s = 172 KByte/s	Audio-CD, etc.
MPEG Layer-1	1:4	384 KBit/s = 48KByte/s	Digital Compact Cassette (DCC)
MPEG Layer-2	1:6 bis 1:8	256-192 KBit/s = 32-24 KByte/s	Digital Audio Broadcast (DAB), CD-I, DVD
MPEG Layer-3	1:10 bis 1:12	128-112 KBit/s = 16-14 KByte/s	ISDN, Satellitenradio, Internet-Audio

2.2 MPEG-4

Die Entwicklung von MPEG-4 begann in 1997 und wurde in den ersten Monaten von 1999 bereits ein internationaler Standard. Heute wird es vor allem in digitalen Video Kameras oder für Echtzeit Video Übertragungen eingesetzt.

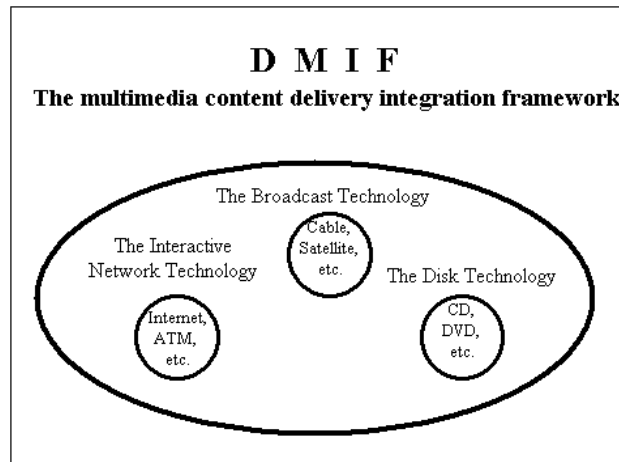
Drei Haupteinsatzgebiete von MPEG-4:

- Digitales Fernsehen
- Interaktive Graphikprogramme
- Interaktives Multimedia (WWW, ASF (advanced streaming format))

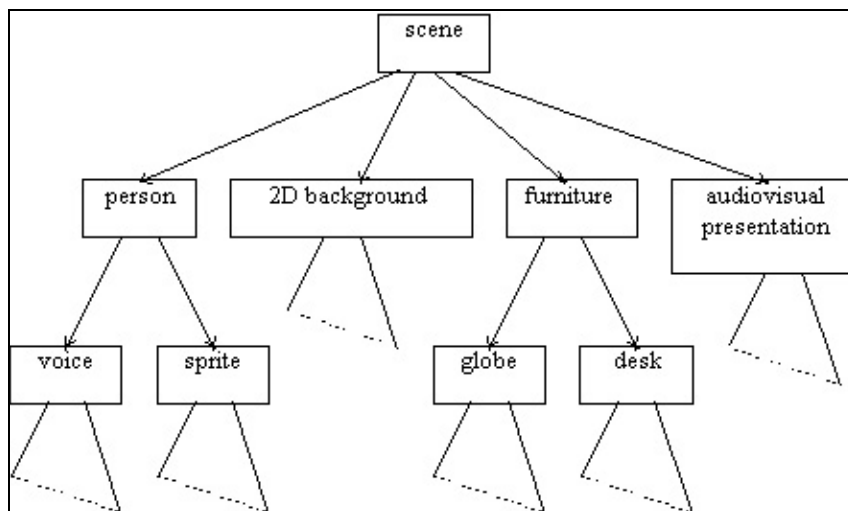
MPEG beinhaltet ein nützliches Feature zur Einbindung von Benutzerrechten und sog. „Authorinformation“. Weiters ist es möglich, dem Inhalt eine Interaktivität zu verleihen. Es wird jedoch unterschieden in drei Gruppen zwischen den einzelnen Anwendungsgebieten:

Benutzerapplikation
Multimedia Broadcast
Mobile Kommunikation } DMIF (Delivery Multimedia Integration Network)

DMIF:



Aufbau und Eigenschaften einer Szene:



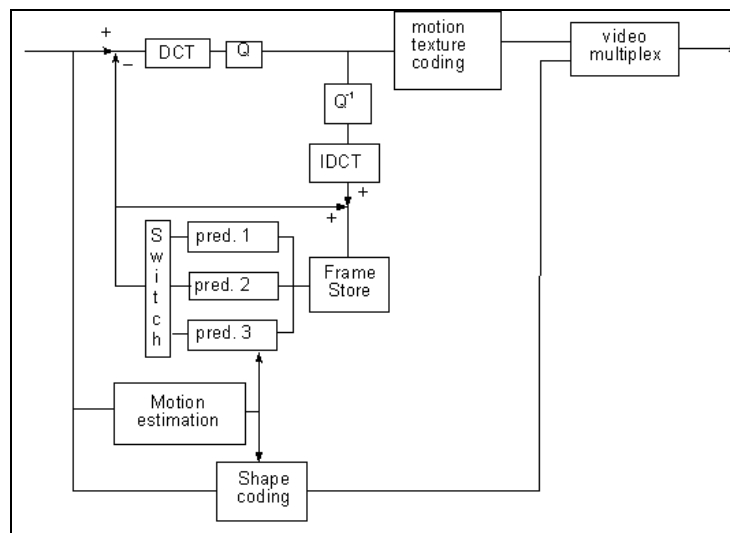
Die Szene wird logisch und in einer Hierarchie aufgebaut. Primitive Objekte können zu Gruppen zusammengefasst werden. Auch Transformationen von geometrischen und akustischen Elementen sind möglich (VRML= Virtual Reality Modeling Language). Die eigentliche Plazierung einer Szene wird jedoch durch ein festgelegtes Koordinatensystem beschrieben. Interaktivität wird einer Szene durch Veränderung des Blickwinkels sowie durch Änderung des akustischen Eindrucks verliehen.

Hauptfunktionalität von MPEG-4:

Wie bereits beschrieben ist der DMIF entscheidend für die Art der Übertragung. Damit wird die Art des Netzwerks und die Übertragungsart festgelegt. Zum System ist zu sagen, dass Datenstreams (ES= Elementary Streams) kodiert werden und dann separat gespeichert und übertragen werden. Auch eigene Objekte für die Interaktion können erschaffen werden. Die Beziehung von Objekten zu den Datenstreams kann mittels „Object Descriptoren“ hergestellt werden. Eine wichtige Eigenschaft ist auch das MPEG-4 Transport Layer unabhängig ist und somit bei verschiedenen Systemen eingesetzt werden kann (Narrowband-ISDN, ATM, usw.). Bei der Audio Funktionalität können bestimmte Sprachkodierungen zum Einsatz kommen (2-24kbit/s). Weiters kann für die Sprach Synthese Lippensynchronisation aber auch TTS (Text to Speech) verwendet werden. Die Bitraten von MPEG-4 Audio variieren von niedrigen Bitraten bis hin zu Multichannelübertragungen.

Zu den visuellen Features (MPEG-4 Video) zählen Hybrid coding (Natur Bilder + Computeranimation) mit den Bitraten von 5kbit/s- 10Mbit/s. Die Auflösung geht hin bis zu HDTV. Die Kompression erfolgt von akzeptabel bis hin zu fast verlustfrei.

Kodierungsschema:

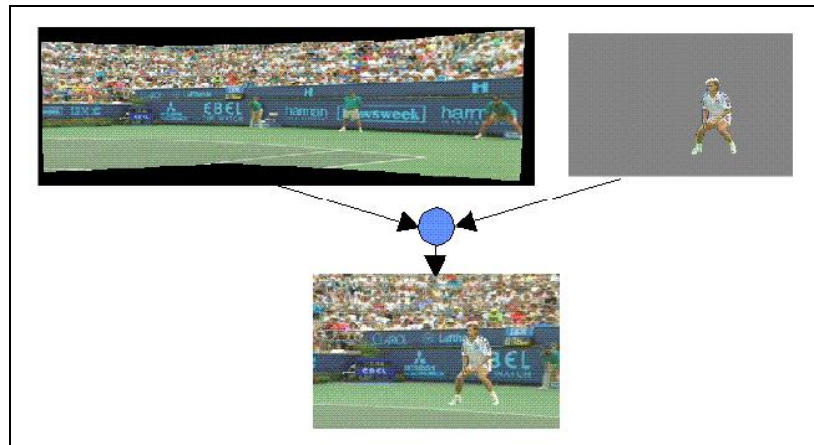


Die Kodierung ist recht flexibel und kann auf rechteckige als auch auf unregelmäßige Formen angewandt werden. Als Basis der Kodierung kommt „Shape coding“, „Motion compensation“ und DCT basiertes Kodieren (8x8) zum Einsatz. Da auch der Inhalt eines Videos bei MPEG-4 berücksichtigt wird, können manche Sequenzen besser komprimiert werden als die restlichen. Dafür wird die Vorhersage Technik genutzt. (Pred.) Diese wird in zwei Teile unterschieden:

- Normal 8x8 oder 16x16 Motion Compensation.
- GMC (Global Motion Compensation).

Dabei wird eine Szene in den Hintergrund (statisches Bild) und den sich bewegendem Teil aufgesplittet (Hintergrund wird auch als Sprite bezeichnet).

Beispiel:



In dem Beispiel wird die Szene in den Hintergrund und in den Tennisspieler unterteilt. Der Hintergrund wird als erstes zum Empfänger übertragen und bleibt in einen sog. Sprite buffer gespeichert. In den darauffolgenden Frames werden nur die für den Empfänger relevanten Hintergrundbilder übertragen. Dadurch kann sich der Empfänger den Hintergrund rekonstruieren. Der sich bewegende Teil wird aufgeteilt und unregelmäßige Formen werden separat übertragen. Nun werden jeweils Vordergrund und Hintergrund zusammengemischt für jeden einzelnen Frame.

Zusätzlich werden auch eine einzelne Texturen mit der „Shape-adaptive DCT“ kodiert und die Kodierungseffizienz zu erhöhen. Die Kompressionsrate ist somit vom Inhalt abhängig.