

05. Bildcodierung und Kompression

Digitales Bild-Dateiformat

Wenn eine Zahl unendlich viele mögliche Werte annehmen kann, sprechen wir von kontinuierlichen oder analogen Daten. Computer können mit analogen Daten nicht rechnen, die Daten müssen daher in eine für den Computer bearbeitbare Version umgewandelt werden: die unendliche "Vielfalt" muss in diskrete (abgegrenzte, getrennte) Zahlenwerte umgewandelt werden. Diesen Umwandlungsprozess bezeichnet man als **Digitalisierung**, der beim Scannen (Abtasten) von Fotos oder eines dreidimensionalen Raumes beim Fotografieren mit digitalen Kameras erfolgt. Das Ergebnis ist in jedem Fall der Verlust der unendlichen Vielfalt, da digitale Daten begrenzt, also endlich sind. Ein digitales Bild ist somit eine numerische Repräsentation eines zweidimensionalen Bildes, welches vom Typ her aus Vektorbeschreibungen oder einem Raster (zweidimensionale Matrix) bestehen kann. Ein **Rasterbild** ist ein Raster von diskreten Werten (jeder Bildpunkt eines Bildrasters muss mit seinem Helligkeitswert bzw. Farbwert abgespeichert werden). Ein **Vektorbild** ist eine am Rechner entstandene Menge von Linien, Rechtecken und anderen Objekten. Rasterbilder beinhalten eine endliche Menge an digitalen Werten, die Bildelemente oder **Pixel** genannt werden, welche die kleinsten, individuellen Elemente in einem Bild darstellen.

Um digitale Bildinformationen zu erhalten, muss diese in einem Bilddatenformat abgespeichert werden. Bilddatenformate verwenden standardisierte Methoden zur Organisation und Abspeicherung digitaler Bilddaten. Einige dieser Formate können für das Anzeigen auf einem Computerbildschirm oder für Drucker gerastert werden, d.h. das Bild enthält ein Pixelraster, das für die Farbgebung jeder Rasterzelle eine Anzahl an Bits zur Verfügung stellt und der Farbtiefe des anzeigenden Gerätes entspricht. Ein Bild $L \times N$ mit L Zeilen und N Spalten, 2^B Graustufen und c Farbkomponenten kann in einem komprimierten, unkomprimierten oder Vektorformat abgespeichert werden und benötigt im unkomprimierten Fall einen Speicherplatz der Größe $L \times N \times B \times c$, z.B. bei einer Bildgröße von 1024×768 , 3 Farbkomponenten (RGB) und 256 (2^8) Graustufen ergibt das $1024 \times 768 \times 3 \times 8 = 18.87 \text{ Mbits} \rightarrow 2.36 \text{ MBytes}$. Im allgemeinen korreliert die Bilddatengröße positiv mit der Anzahl an Pixeln in einem Bild und der Farbtiefe oder Bits per Pixel eines Bildes. Es gibt hunderte von Bilddatenformaten, von denen die wichtigsten hier kurz erläutert werden.

Raster-Bildformate

Durch die Verwendung des Raw-Bildformates ermöglichen uns digitale Kameras den Zugang zu den tatsächlich von der Kamera aufgenommenen Bilddaten, das heißt dieses Format speichert für jeden Pixel den entsprechenden Farbwert ohne Nachbearbeitung (bei 1-Chip Kameras Rot-, Grün- und Blauwerte entsprechend des CFA-Musters). Diese RAW-Format stellt somit kein „richtiges“ Farbbild dar, es muss zur farbigen Anzeige noch umgewandelt werden, da ja 2/3 der Farbinformation aus den bestehenden Werten interpoliert werden muss. Konventionelle Raster-Bildformate sind zum Beispiel Bitmap (BMP) oder Portable Network Graphics (PNG). Ein modernes Raster-Bildformat ist in der Lage, zweidimensionale digitale Bilder beliebiger Breite, Höhe und Auflösung abzuspeichern, sei es in Farbe oder in Grauwerten unterschiedlicher Farbtiefe. Eine Rasterbilddatei besteht aus Strukturen fixer Größe (Header) und variabler Größe (bildabhängig), die in einer vordefinierten Sequenz aufscheinen, z.B. für BMP der Bitmap File Header, welcher die allgemeinen Informationen über die Bitmap Bilddatei (14 Bytes) speichert. Eine Fülle an Metainformationen wird in jedem individuellen Dateiformat gespeichert.

Vektor-Bildformate

Im Gegensatz zu den Raster-Bildformaten, bei denen die Daten die Charakteristik jedes einzelnen Pixels beschreiben, beinhaltet das Vektor-Bildformat eine geometrische Beschreibung, die für jede gewünschte Anzeigegröße problemlos gerendert werden kann. An einem bestimmten Punkt müssen alle Vektorgrafiken rasterisiert werden um auf einem digitalen Bildschirm angezeigt werden zu können. Plotter sind Drucker, die auch Vektordaten zum Zeichnen von Grafiken verwenden. Ein freier und offener internationaler Standard für die Speicherung von 2D-Vektorzeichnungen, Rasterbildern und Text ist das sogenannte Computer Graphics Metafile (CGM) Format. Der Standard wird in Bereichen wie technische Illustration, Kartografie, Visualisierung und elektronischen Publikationen verwendet. Alle grafischen Elemente werden in Quelltextdateien spezifiziert, die anschließend zu einer Binärdatei oder Textdarstellungen kompiliert werden. CGM stellt Instrumente für den Austausch von Grafikdaten bei der Darstellung zweidimensionaler grafischer Information zur Verfügung, unabhängig von einer bestimmten Anwendung, Plattform, Systems, oder Gerätes. Das CGM-Format besitzt als Metadatei (eine Datei die beschreibende Information über eine andere Datei enthält) und eine Vielzahl an möglichen Elementen, die ihrerseits Funktionen zur Verfügung stellen. CGM beinhaltet Instruktionen und Daten, um grafische Komponenten für das Rendern eines Bildes rekonstruieren zu können und verwendet hierbei objektorientierte Ansätze.

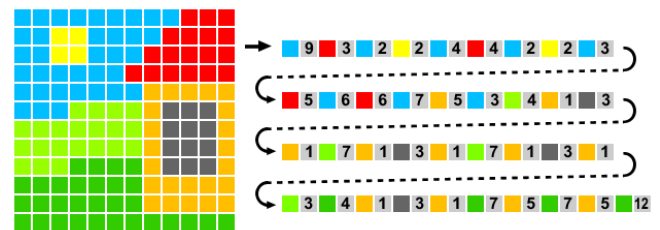
Ein weiteres Beispiel ist das Windows-Metafile (WMF), ein grafisches Dateiformat für Microsoft Windows Systeme, das im Jahr 1990 entwickelt wurde. Windows-Metafiles ermöglichen den Datenaustausch zwischen Anwendungen und beinhalten Vektorgrafiken und Bitmap Komponenten.

Bildkompression

Das Ziel der Bildkompression ist die Reduzierung irrelevanter und redundanter Bildinformation, um Daten in einer effizienten Form übermitteln oder speichern zu können. Bildkompression kann **verlustfrei** (engl. *lossless*) oder **verlustbehaftet** (engl. *lossy*) sein. Verlustfreie Kompression wird zu Archivierungszwecken und für medizinische Bilder, technische Zeichnungen, Cliparts oder Comics bevorzugt, da verlustbehaftete Kompression mit niedriger Bitrate Kompressionsartefakte aufweist. Verlustbehaftete Methoden eignen sich besonders für natürliche Bilder wie Fotografien in Anwendungen, wo ein kleiner (vom Menschen nicht erkennbarer) Genauigkeitsverlust akzeptabel ist, um eine Reduktion der Dateigröße zu erzielen. Die verlustbehaftete Kompression, die unbemerkbare Veränderungen produziert, wird auch als visuell verlustfrei (engl. *visually lossless*) bezeichnet.

Verlustfreie Kompression

Verlustfreie Datenkompression erfolgt durch Verwendung von Datenkompressionsalgorithmen, die eine exakte Rekonstruktion der Originaldaten von den komprimierten Daten aus ermöglichen. Sie wird in jenen Bereichen eingesetzt, bei denen die vollständige Erhaltung der Daten von Interesse ist. Kompressionsprogramme führen meistens zwei Schritte hintereinander aus:

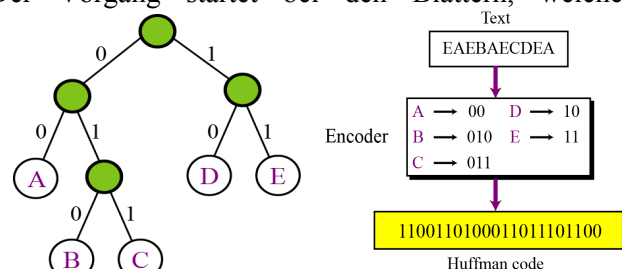


1. Generierung eines statistischen Modells der Eingabedaten.

2. Abbildung der Eingabedaten anhand des Modells auf eine Bitreihenfolge, wo „wahrscheinliche“ (oft vorkommende) Daten einen kürzeren Output erzeugen als „unwahrscheinliche“.

Einer der grundlegendsten Codierungsalgorithmen zur Produktion von Bitfolgen ist **Run Length Encoding (RLE)**. Dabei werden Sequenzen von Datenwerten, in der ein gleicher Datenwert in vielen aufeinanderfolgenden Datenelementen vorkommt (engl. runs), als einzelner Datenwert und dessen Anzahl an Wiederholungen abgespeichert. Beispiele für Daten, die solche Sequenzen enthalten, sind: einfache grafische Bilder wie Icons, Linienzeichnungen und Animationen. Diese Methodik eignet sich nicht bei Daten die wenige solcher Sequenzen beinhalten (z.B. Natürliche Bilder), da es dann sogar zu einer erheblichen Vergrößerung des Datensatzes kommen kann.

Eine andere Methode ist die **Huffman-Codierung**, welche eine spezielle Methode zur Auswahl der Repräsentation jedes Symbols verwendet. Sie erstellt kurze Zeichenketten für häufig vorkommenden Ausgangssymbole und längere Zeichenketten für selten vorkommende Ausgangssymbole (Entropiecodierung). Hierzu wird ein Binärbaum erstellt, dessen Knoten ein interner Knoten oder ein Blatt sein können. Am Beginn sind alle Knoten Blätter, die das Symbol selbst beinhalten, die Gewichtung (Häufigkeit des Vorkommens) eines Symbols und optional eine Verbindung zu einem Elternknoten, mit dem es möglich ist, auf einfachem Wege den Code (in umgekehrter Reihenfolge), beginnend von den Blättern aus, zu lesen. Interne Knoten beinhalten die Gewichtung, Verbindungen zu den Kinderknoten und optional eine Verbindung zu einem Elternknoten. Der Vorgang startet bei den Blättern, welche die Wahrscheinlichkeiten des Symbols, das sie repräsentieren, enthalten. Anschließend wird ein neuer Knoten erstellt, der als Kinder jene zwei Knoten/Blätter mit der niedrigsten Wahrscheinlichkeit hat, sodass die Summe der Wahrscheinlichkeiten der Kinder zu dessen Wahrscheinlichkeit wird. Von diesem Punkt an wird nur mehr der neu entstandene Knoten betrachtet. Dieser Vorgang wird solange wiederholt bis nur mehr ein Knoten übrig bleibt: der Huffman Tree.



Eine der bekanntesten Methoden ist das **Lempel Ziv (LZ)** Kompressionsverfahren. Der Algorithmus codiert 8-Bit-Datensequenzen als 12-Bit-Code mit fixer Länge. Die Codes von 0 bis 255 repräsentieren 1-Zeichen-Sequenzen, die aus dem 8 Bit Zeichen bestehen. Die Codes von 256 bis 4095 werden in einer Tabelle für Sequenzen erstellt und anhand deren Vorkommen in den Daten codiert. In jedem Schritt der Kompression werden Eingabebytes in einer Sequenz gesammelt bis das nächste Zeichen auftaucht, welches noch keinen Eintrag in der Tabelle hat. Der Code für diese Sequenz, ohne dieses Zeichens, wird ausgegeben und ein neuer Code (für die Sequenz und das Zeichen) wird zur Tabelle hinzugefügt. Dies funktioniert folgendermaßen:

1. Initialisiere die Tabelle mit allen Strings der Länge 1.
2. Finde den längsten String W in der Tabelle der zum aktuellen Input passt.

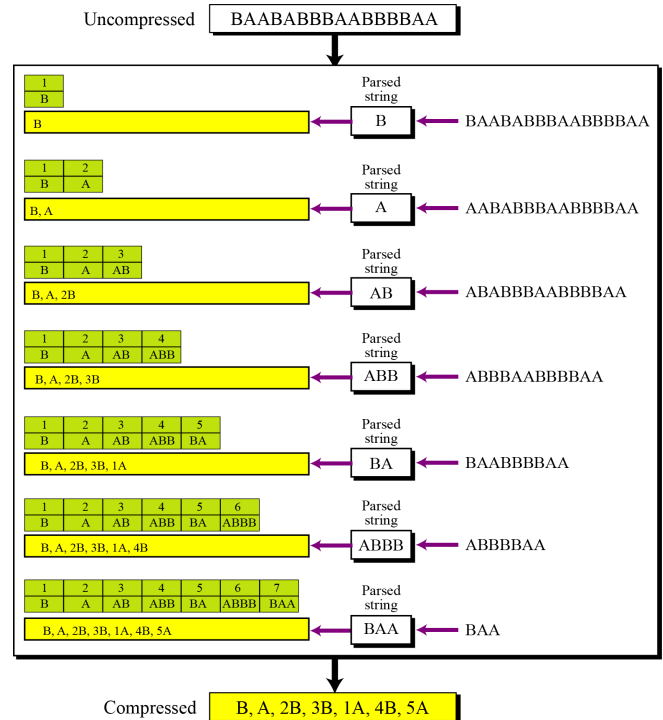
3. Gib den Tabellenindex für W aus, um W vom Input zu entfernen.
4. Füge W gefolgt von dem nächsten Symbol des Inputs der Tabelle hinzu.
5. Gehe zu Schritt 2.

Bildformate, die verlustfreie Kompression verwenden:

Das **Graphics Interchange Format (GIF)** ist ein Raster-Bildformat, das 1987 durch CompuServe eingeführt wurde und sich seit damals aufgrund der allgemein großen Unterstützung und Portierbarkeit im World Wide Web weit verbreitet hat. Das Format unterstützt bis zu 8 Bit pro Pixel und ermöglicht in einem einzigen Bild eine Palette von bis zu 256 unterscheidbaren Farben des 24-Bit RGB Farbraums. Es ist unbrauchbar für Fotografien, aber eignet sich für kleine Bilder wie Icons oder simple Animationen.

Portable Network Graphics (PNG) wurde entwickelt, um GIF zu verbessern und ersetzte daraufhin dieses Bilddatenformat, da es keine Patentlizenz benötigt. PNG unterstützt palettenbasierte Bilder (mit 24 Bit RGB-Paletten oder 32 Bit RGBA-Farben), Graustufenbilder (mit oder ohne Alphakanal) und vollfarbige nicht palettenbasierte RGB[A] Bilder (mit oder ohne Alphakanal). PNG wurde für die Übertragung von Bildern im Internet entwickelt und nicht für hochqualitative Druckgrafiken und unterstützt daher nur RGB und keine anderen Farbräume wie z.B. CMYK.

TIFF (Tagged Image File Format) ist ein populäres Dateiformat bei Grafikern, in der Verlagsindustrie sowie bei Amateur- und Profifotografen. Es wird von Adobe Systems verwaltet und wurde ursprünglich dahingehend entwickelt, dass Hersteller von Desktopscannern Mitte der 1980er einem allgemeinen Scan-Bilddatenformat zustimmen, um dadurch die Entwicklung von verschiedenen proprietären Formaten zu vermeiden. TIFF ist ein flexibles, anpassungsfähiges Datenformat, das Bilder und Daten in einem einzigen Datensatz mittels Hinzufügen von Headertags (Größe, Definition, Bilddatenanordnung, angewendete Bildkompression), die die Bildgeometrie definieren, verwaltet. Eine TIFF-Datei kann z.B. ein Container sein, der verlustbehaftet komprimierte JPEG- und verlustfrei komprimierte Bilder beinhaltet.



Verlustbehaftete Kompression

Verlustbehaftete Kompression ist eine Datencodierungsmethode, die Daten durch die Entfernung bestimmter Datenteile komprimiert. Verlustbehaftete Kompression orientiert sich an wahrnehmungspsychologischen Erkenntnissen und codiert Signale, die schwach oder gar nicht wahrgenommen werden, entsprechend "platzsparend". Die Ursprungsdaten werden dafür in eine Domäne transformiert, die den Informationsinhalt genauer wiedergibt.

JPEG

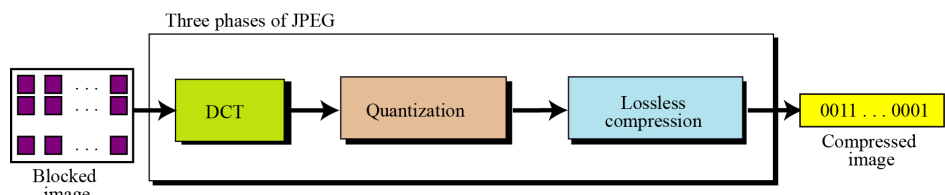
JPEG ist die am häufigsten verwendete Methode für verlustbehaftete Kompression von digitalen Bildern. Der Grad der Kompression kann angepasst werden und ermöglicht einen auswählbaren Kompromiss zwischen Speicherplatz und Bildqualität. Der Name „JPEG“ steht für "Joint Photographic Experts Group", der Name des Komitees, das den JPEG Standard entwickelt hat. Der Codierungsprozess besteht aus mehreren Schritten:

1. Das Ausgangsbild wird von RGB in den YCbCr-Farbraum umgerechnet. Das YCbCr-Modell stellt die Farbinformation als Helligkeitskomponente Y und zwei Farbkomponenten Cb (Blue-Yellow Chrominance) und Cr (Red-Green Chrominance) dar.

1. Die Farbauflösung (Cb und Cr) wird halbiert. Dieser Schritt ist durch die geringere Empfindlichkeit des Auges gegenüber feinen Farbdetails im Gegensatz zu feinen Helligkeitsunterschieden begründet.

2. Das Bild wird in 8x8 Pixelblöcke aufgeteilt und für jeden Block wird mit jedem Y-, Cb- und Cr-Datensatz eine Diskrete Cosinus Transformation (DCT) durchgeführt.

3. Die Amplituden einzelner Frequenzkomponenten werden quantisiert. Das menschliche Auge ist empfindlicher gegenüber kleinen Farb- oder Helligkeitsänderungen über größere Flächen als gegenüber hochfrequenten Helligkeits- oder Farbvariationen. Daher werden hochfrequente Komponenten mit geringerer Genauigkeit abgespeichert als niederfrequente Komponenten. Die Qualitätseinstellung des Codierers beeinflusst, in welchem



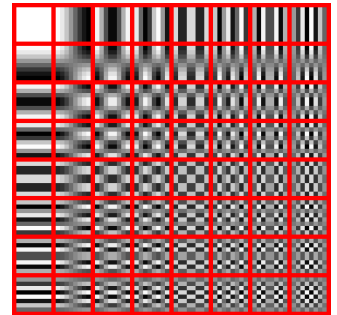
Ausmaß die Genauigkeit jeder Frequenzkomponente reduziert wird.

4. Die resultierenden Daten eines 8x8 Blocks werden zusätzlich mit einem verlustfreien Algorithmus komprimiert, eine Variante der Huffman Codierung.

Diskrete Cosinus Transformation

Die DCT ist eine Variation der Fouriertransformation, die beliebige Signale (jede Folge von Datenpunkten) durch Überlagerungen von Cosinuswellen mit verschiedener Frequenz und Amplitude darstellt. Die Pixelwerte werden als Frequenz und Amplitudenverteilung dargestellt. Dabei werden die unterschiedlichen Frequenzen des Signals sichtbar. Die zweidimensionale DCT für einen 8x8 Block ist gegeben durch:

$$F(u, v) = \sum_{x=0}^7 \sum_{y=0}^7 \alpha(u) \alpha(v) f(x, y) \cos\left[\frac{\pi}{8} \left(x + \frac{1}{2}\right) u\right] \cos\left[\frac{\pi}{8} \left(y + \frac{1}{2}\right) v\right]$$
$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{8}} & \text{falls } u = 0 \\ \sqrt{\frac{2}{8}} & \text{sonst} \end{cases}$$



wobei u die horizontale Ortsfrequenz für ganzzahlige Werte $0 \leq u < 8$, v die vertikale Ortsfrequenz für ganzzahlige Werte $0 \leq v < 8$, $\alpha(u)$ ein normalisierender Skalierungsfaktor für die Orthonormalität der Transformation, $f(x,y)$ der Pixelwert an den Koordinaten (x,y) und $F(u,v)$ der DCT-Koeffizient an den Koordinaten (u,v) ist. Die eigentliche verlustbehaftete Kompression erfolgt bei JPEG durch den Quantisierungsschritt, der die Gesamtgröße der DCT-Koeffizienten minimiert. Jeder DCT-Koeffizient wird durch einen vordefinierten Quantisierungswert dividiert und das Ergebnis auf den nächsten Integer-Wert gerundet. Die Quantisierungswerte sind für niedrige Frequenzen kleiner als die für hohe Frequenzen (aufgrund der vorher beschriebenen unterschiedlichen Empfindlichkeit des Auges für unterschiedliche Bildfrequenzen). Das Ergebnis ist, dass einige der hochfrequenten Komponenten zu 0 gerundet werden und die restlichen dieser Komponenten gering positiv oder negativ sind, was dazu führt, dass weniger Bit für die Repräsentation benötigt werden. Als Entropiecodierung wird meist eine Huffman-Kodierung verwendet. JPEG Kompression kann ein Bild zu einem Fünftel seiner Originalgröße reduzieren (ohne visuelle Verschlechterung). Je höher die Kompression (Quantisierung), desto mehr Artefakte treten auf (Blockformation). JPEG wurde für natürliche Bilder entwickelt und nicht für künstlich erstellte Bilder (Computergrafiken).

Video Kompression

Videokompression verwendet Techniken, um die Redundanz in Videodaten zu reduzieren. Kompressionsalgorithmen und Codecs kombinieren räumliche Bildkompression und zeitliche Bewegungskompensation. Die Mehrheit der Algorithmen verwendet verlustbehaftete Kompression, bei der große Mengen an Daten entfernt werden, der Unterschied aber visuell kaum erkennbar ist. Ein Kompromiss zwischen Videoqualität, Bearbeitungsaufwand der Kompression und Dekompression und den Systemvoraussetzungen ist nötig. Videokompression arbeitet mit quadratischen Bildausschnitten (Makroblöcke), deren Unterschiede von Frame zu Frame ermittelt werden. Beinhaltet das Video mehr Bewegung, muss die Kompression mehr Daten codieren, da sich eine größere Anzahl an Pixel ändert. Die **Interframe-Kompression** betrachtet ein bzw. zwei Frames vor oder nach dem aktuellen Frame in der Sequenz, um diesen zu komprimieren. Im Gegensatz dazu komprimiert die **Intraframe-Kompression** nur den aktuellen Frame (eigentlich eine Bildkompression, z.B. JPEG).

Die am häufigsten verwendete Methode zur Videokompression vergleicht jeden Frame im Video mit dem vorigen. Bereiche, in denen sich nichts verändert hat, werden durch einen einfachen Befehl in den nächsten Frame kopiert. Wenn sich ganze Makroblöcke eines Frames durch eine einfache Transformation verändern, befiehlt der Kompressor dem Dekompressor, dass dieser die Kopie Verschieben, Rotieren, Aufhellen oder Verdunkeln soll. Durch Interframekompression komprimierte Videos lassen sich problemlos abspielen, das Bearbeiten der Videos ist aber ggf. komplexer, wenn dadurch Teile des Videos neu codiert werden müssen. Eine der meist verwendeten Techniken zur Videokompression ist **MPEG (Moving Picture Experts Group)**. Sie wurde von einer Expertengruppe entwickelt, die sich 1988 durch ISO und IEC zusammengefunden hatte, um einen Standard für Audio- und Videokompression und deren Übermittlung zu entwickeln. MPEG arbeitet asymmetrisch, da die Codierung algorithmisch komplexer als die Decodierung ist. Im Broadcastingbereich ist dies von Vorteil, da die Zahl an benötigten billigen Decodierern groß ist und die der teuren Codierer klein. Der erste MPEG-1 Kompressionsstandard ist auf 1.5Mbit/s beschränkt, obwohl die Spezifikation höhere Bitraten verspricht. Er wurde entwickelt, um Film und Ton auf die Bitrate einer Compact Disc zu codieren. Er wird für Video CDs, SVCD und DVDs mit niedriger Videoqualität verwendet. Der MPEG-2 Standard hat einen größeren Umfang, unterstützt das Zeilensprungverfahren sowie High Definition Auflösung und ist bedeutend für digitales Fernsehen, digitales Kabelsignal, SVCD und DVD Video. MPEG-4 bietet eine noch effizientere Codierung von Videos und rückt näher in den Bereich von Computergrafik-Applikationen. Der MPEG-4 Decodierer kann als Renderingprozessor fungieren und der codierte Bitstream beschreibt 3D Formen und Oberflächentexturen. MPEG-4 wird neben MPEG-2 für Blu-ray Discs verwendet.