Multimedia

Oliver Vornberger, Fachbereich Mathematik/Informatik, Universität Osnabrück

**‘‘Multimedia’’, inzwischen in aller Munde, wurde zum vielstrapazierten Begriff und avancierte gar zum Wort des Jahres 1995. Offenbar verbergen sich dahinter ‘‘viele Medien’’, über die Politiker und Bürger, Verlage und Leser, Anbieter und Kunden die unterschiedlichsten Vorstellungen entwickeln.
Grenzen und Machbarkeit von Multimediasystemen werden durch die algorithmischen Grundlagen diktiert, auf denen ein EDV-gestütztes Management multimedialer Objekte beruht. In diesem Artikel sollen daher die wichtigsten Verfahren zur Digitalisierung und Komprimierung der Medien Text, Bild, Grafik, Audio und Video vorgestellt werden.**

String Präfix-String-Index	String Erweiterungs-character	Code
	a	1
	b	2
	c	3
1	b	4
2	a	5
4	c	6
3	b	7
5	b	8
8	a	9
1	a	10
10	a	11
11	a	12

Abbildung 1: Stringtabelle nach Einlesen von `ababcbababaaaaaa`*, z.B. bezeichnet Code 9 den String* `baba`*.*

Abbildung 2: 200 × 248 Grauwertbild vor und nach der JPG-Kompression mit Faktor 20

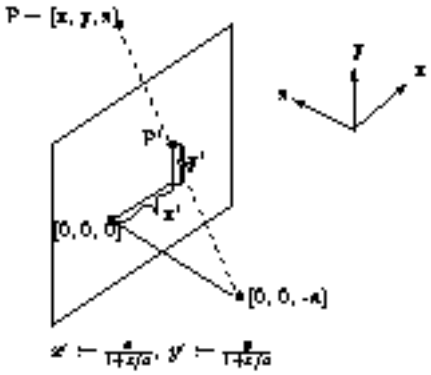


Abbildung 3: Bestimmung des Projektionspunkts P' aus den 3D-Koordinaten des Punkts P und des Augenpunkts mit Hilfe der Strahlensätze

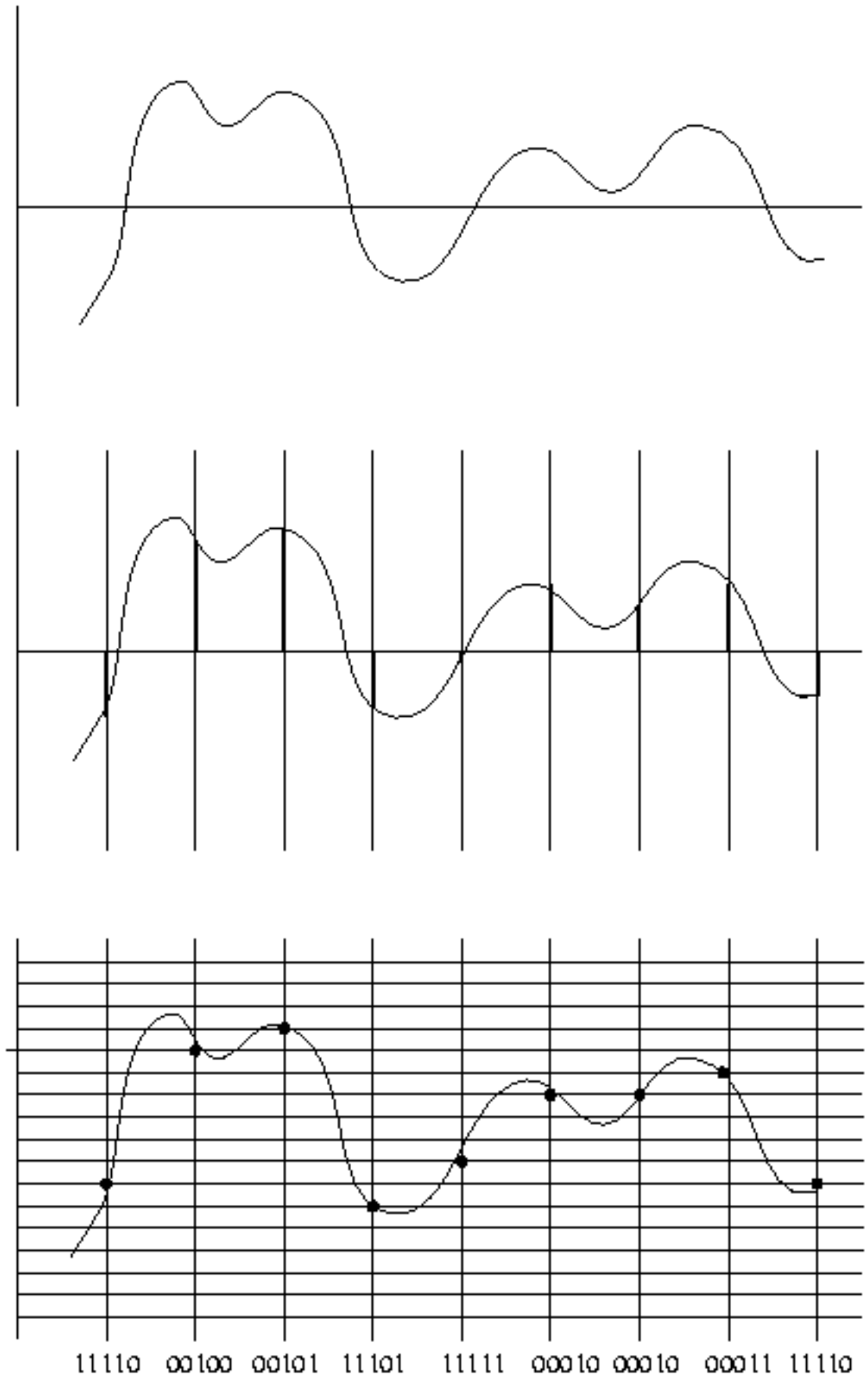
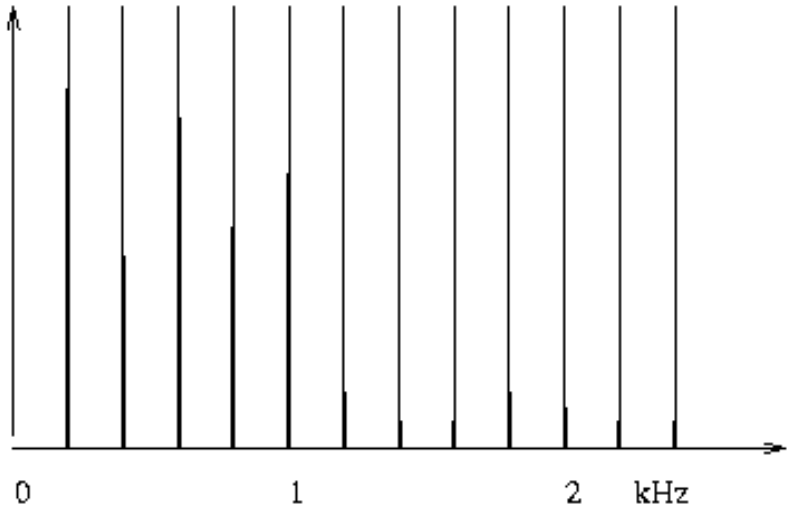


Abbildung 4: Diskretisierung einer analogen Schallwelle nach Zeit und Intensität (Ergebnis im 2er-Komplement)



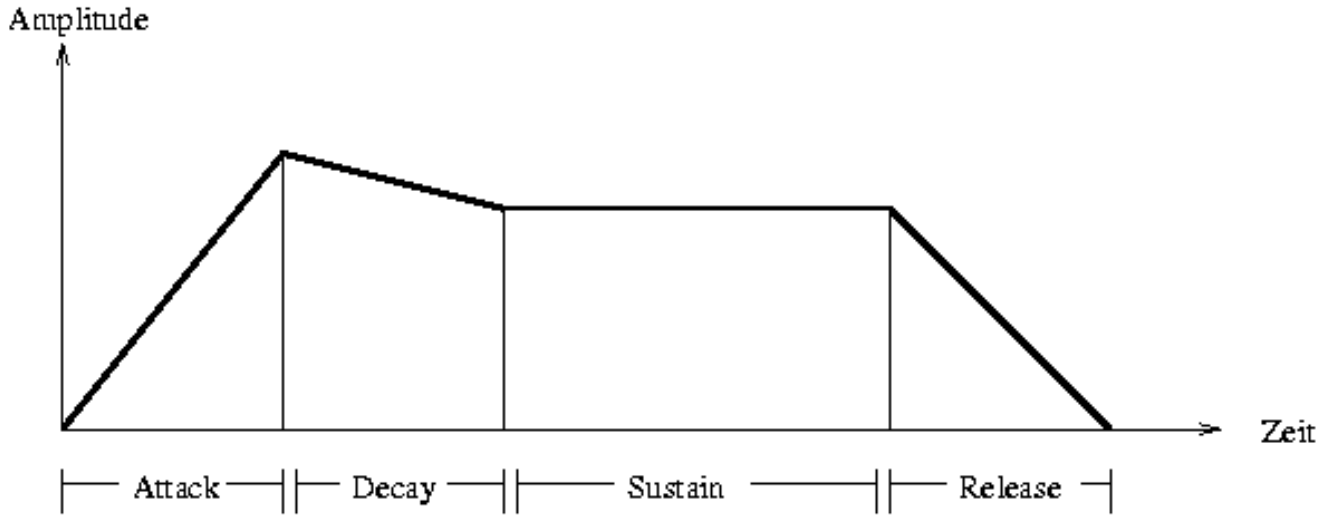


Abbildung 5: Charakterisierung des Klangs eines Musikinstruments durch Obertonspektrum und Hüllkurve

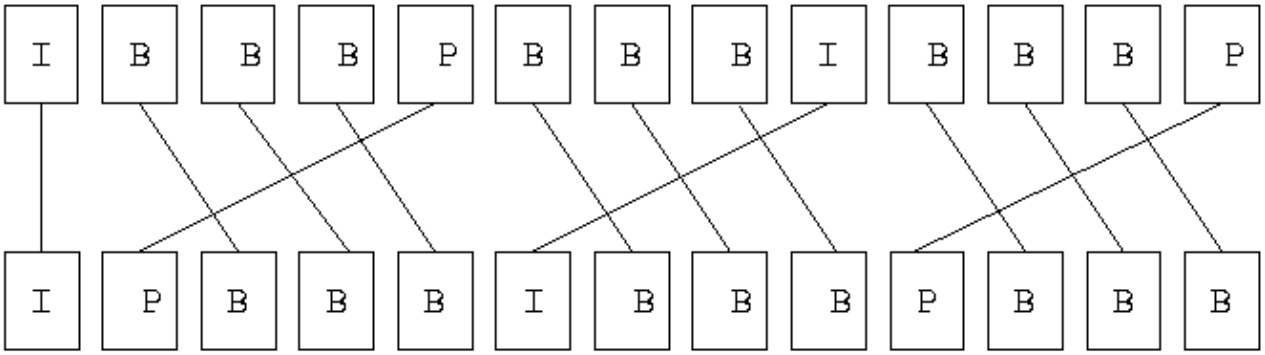


Abbildung 6: Umstellung der Bildfolge beim MPEG-Verfahren

Definition

Meyers Enzyklopädisches Lexikon aus dem Jahre 1975 hilft uns bei der Begriffsbildung. Wir lesen:

> *Multi- [lat.: viel], als Präfix*
>
> *Medium [lat. das in der Mitte Befindliche], allgemein Mittel, vermittelndes Element, insbesondere (in der Mehrzahl) Mittel zur Weitergabe und Verbreitung von Information durch Sprache, Gestik, Schrift und Bild*.

Übertragen in das Zeitalter von Bits und Bytes bedeutet dies, daß audiovisuelle Daten in einem einzigen Dokument integriert sind; d.h., Text, Bild und Ton sind in diversen digitalen Formaten repräsentiert und können interaktiv abgerufen und manipuliert werden. Hierbei stecken Mensch und Maschine gleichermaßen den Rahmen: Zum einen wird bei der Wahrnehmung von Licht- und Tonreizen die subjektive Empfindung durch charakteristische Eigenarten von Auge und Ohr bestimmt. Zum anderen diktiert die gegenwärtige Rechnertechnologie gewisse Vorgaben in puncto Prozessorleistung, Plattenkapazität und Transferrate. Beide Randbedingungen resultieren in eng gesteckten Grenzen für die Digitalisierung und Komprimierung der verschiedenen Medien. In den nächsten Kapiteln sollen daher in systematischer Weise die algorithmischen Grundlagen bei der digitalen Manipulation von Text, Bild und Ton behandelt werden.

Text

Text ist sicherlich das älteste Medium, welches der Mensch zum Speichern von Informationen verwendet. Auf der Grundlage eines Zeichensatzes stellt sich die Digitalisierung recht einfach dar: Für jedes verwendete Symbol wird ein Byte abgelegt. Bereits 1952 entwickelte Huffman eine Komprimierung, die auf der relativen Häufigkeit der beteiligten Buchstaben beruht. Hierzu wird aufgrund einer Analyse des vorliegenden Textes ein variabel langer Code konstruiert: Häufig vorkommende Buchstaben werden mit wenigen Bits, selten vorkommende Buchstaben werden mit vielen Bits kodiert. Normale Schriftstücke schrumpfen dadurch auf etwa 60 %. Um den im Laufe eines Textes sich ändernden Häufigkeiten Rechnung zu tragen und um das Übermitteln der im Preprocessing entstandenen Huffman-Tabelle zu vermeiden, wurde eine adaptive Variante entwickelt, in der Komprimierer und Dekomprimierer dynamisch mit Fortschreiten des Textes die Buchstabenhäufigkeit und daraus resultierende Codes aktualisieren.

Codes fester Länge, nämlich 14-Bit-Adressen, verwenden tabellengesteuerte Verfahren. Sukzessive wird eine Stringtabelle gefüllt, welche die im Text enthaltenen Zeichenketten in kompakter Form speichert. Beim Scannen des Dokuments wird solange wie möglich ein Match mit einem Tabelleneintrag gesucht; das erste abweichende Zeichen wird dann zusammen mit einem Pointer auf den bereits vorhandenen Anfangsstring in die Tabelle eingetragen (siehe Abbildung 1). Ein wichtiger Vertreter dieses Prinzips ist das Programm *PKZIP*, welches Kompressionsraten von 40-50% erreicht.

Bild

Ein Teil des elektromagnetischen Spektrums wird vom Auge wahrgenommen, nämlich Schwingungen mit Wellenlängen zwischen 780 mm (rot) und 380 mm (violett). Durch Mischen verschiedener Spektralfarben entstehen neue Farbeindrücke, und so hat sich das RGB-Modell bewährt, welches zur Spezifikation einer Farbe die Intensitäten der drei beteiligten Grundfarben Rot, Grün, Blau festlegt. Der Mensch kann etwa 350.000 verschiedene Farbtöne unterscheiden, und für eine ausreichende Differenzierung von Grauwerten wären etwa 80 Helligkeitsabstufungen ausreichend. Aus Effizienzgründen benutzt man trotzdem ein Byte zum Quantisieren, woraus sich $256^3 \approx 16$ Millionen Farbkodierungen ergeben.

Bei der Digitalisierung eines Schwarz-Weiß- oder Farbfotos entsteht durch das zeilenweise Abtasten in einem Flachbettscanner eine Matrix von Bildpunkten. Zur Darstellung auf einem Computermonitor mit VGA-Auflösung werden 480 Zeilen zu je 640 Pixeln verwendet. Für jede einzelne Pixelausprägung wird ein Bit, ein Byte oder ein Byte-Tripel benötigt, je nachdem, ob es sich um ein Schwarzweißbild, ein Grauwert- oder ein sogenanntes *True-Color*-Bild mit Rot-/Grün-/Blau-Anteilen handelt. Somit liegt der Platzbedarf bei 37 KByte, 300 KByte bzw. knapp 1 MByte.

Das einfachste Verfahren zur Komprimierung stellt das *Run Length Encoding* dar: Identische, aufeinanderfolgende Zeichen werden durch ein Tupel `<Anzahl, Wert>` zusammengefaßt. Diese Technik bringt nur Einsparungen bei künstlich erzeugten Farbflächen, in denen viele identische Farbtöne ohne Nuancen vorkommen.

Beim Einscannen von Schriftstücken durch ein Faxgerät entstehen Schwarz-Weiß-Bilder, die eine charakteristische Verteilung von Pixelsequenzen aufweisen. Hierfür wurde durch ein internationales Gremium eine Huffman-Tabelle festgelegt, die die unterschiedlichen Häufigkeiten berücksichtigt und durch variabel lange Codes die zu übertragende Datenmenge zwischen zwei Fax-Geräten reduziert.

Durch Farbfotographien entstandene Bilder können auf ein Drittel ihrer Größe schrumpfen, indem für jeden Bildpunkt statt eines RGB-Tripels ein Index in Byte-Größe gespeichert wird, welcher in eine Farbtabelle mit 256

Farbwerten verweist. Hierzu wird mit Hilfe des *Median-Cut-Algorithmus* der RGB-Würfel gemäß der beobachteten Farbwerte so lange in Subwürfel gleichmächtiger Pixelinstanzen zerteilt, bis sich 256 Repräsentanten gefunden haben. Um die durch das Quantisieren entstandenen Farbsprünge zu mildern, bedient man sich des *Farbdithering*. Hierbei wird der originale Farbton einer Fläche durch ein geeignetes Muster von Repräsentantenfarben angenähert, die in ihrer Gesamtheit den visuellen Originaleindruck simulieren.

Alle 16 Millionen Farben eines *True-Color*-Bildes lassen sich wieder rekonstruieren aus einer Datei im JPEG-Format, benannt nach der *Joint Photographers Expert Group*. Grundlage bildet die Diskrete Cosinus-Transformation, welche für jede Grundfarbe Teilmatrizen der Größe 8×8 in jeweils 64 Frequenzkoeffizienten transformiert. Hierdurch wird der Bildinhalt (ohne Informationsverlust) als Überlagerung von 2-dimensionalen Schwingungen unterschiedlicher Frequenzen dargestellt. Dabei verlangen Subblöcke mit uniformen Farbflächen wenige langwellige Schwingungen, filigrane Details jedoch das gesamte Frequenzspektrum. Interessanterweise läßt sich das Auge täuschen, da es ein Unterdrücken der hochfrequenten Anteile nicht bemerkt. Also werden die durch die Transformation ermittelten Koeffizienten mit unterschiedlichen Faktoren quantisiert (hierdurch läßt sich der Kompressionsgrad beeinflussen) und ihre Sequenz in der Reihenfolge ihrer subjektiven Wichtigkeit einer Huffman-Kodierung unterzogen. Kompressionsraten von bis zu 1:15 lassen sich erreichen, ohne daß wahrnehmbare Störungen auftreten. Abbildung 2 läßt bereits einige Artefakte erkennen.

Zahlreiche Bildverarbeitungsprogramme sind auf dem Markt, die Bilddateien in unterschiedlichsten Formaten einlesen, erzeugen und manipulieren können. Durch geometrische Operationen lassen sich Bildteile verschieben, skalieren und rotieren; durch Filter-Operationen lassen sich Kanten entdecken, Kontraste verstärken und Objekte weichzeichnen. Da ein 100 ASA Kleinbilddia eine Auflösung von 2000 dpi (Dots per Inch) in sich birgt, erzeugt das Einscannen einer 24 mm $\times$ 36 mm Filmfläche etwa 2000 x 3000 = 6 Millionen Pixel, die in unkomprimierter Form etwa 18 MByte verursachen. Auch die auf Kodak-Photo-CDs abgelegten Bilder benötigen diesen Speicherplatz.

Für eine zügige Bearbeitung solcher Pixelmengen sind also Hauptspeicher mit mindestens 32 MByte zwingend erforderlich.

Grafik

Unter Grafik verstehen wir die Generierung einer (ggf. fotorealistischen) Darstellung anhand der Beschreibung einer 3-dimensionalen Szene inkl. Beleuchtung und Kamerastandpunkt. Typische Einsatzgebiete sind

- CAD (Architektur & Maschinenbau),
- Visualisierung (von Meßergebnissen),
- Simulation (von physikalischen/ chemischen/biologischen Vorgängen),
- Unterhaltung (*Virtual Reality*).

Im Gegensatz zum fertigen Bild liegt also hier ein Objekt (z.B. ein brauner Holzwürfel) in Form einer geometrischen Beschreibung vor, die deutlich weniger Platz als das dadurch spezifizierte Aussehen benötigt. Ein Objekt wird dabei approximiert durch einen Polyeder, der sich aus ebenen Flächen zusammensetzt, die aus mehreren Kanten mit je zwei 3D-Endpunkten bestehen und gewisse Materialeigenschaften besitzen.

Zur Anzeige eines Objekts durchlaufen seine Bestandteile einen sehr rechenintensiven Prozeß, die sogenannte *Viewing Pipeline*, die aus der Repräsentation unter Berücksichtigung des Betrachterstandpunkts eine perspektivische Projektion berechnet und die Oberflächen gemäß des verwendeten Materials und der beteiligten Lichtquellen einfärbt. Die mathematischen Grundlagen für die Projektion sind einfache geometrische Strahlensätze, die mit wenigen Operationen 3D-Punkte auf Bildschirmkoordinaten abbilden können (siehe Abbildung 3). Das Einfärben stützt sich auf Lichtbrechungs- und Spiegelungsgesetze, welche unter Berücksichtigung der Materialeigenschaften zu jeder Farbkomponente Rot, Grün und Blau ihre Intensität aus der Sicht des Betrachters ermittelt.

Unhandlich wird der Vorgang durch die hohe Zahl der Bildpunkte, für die nach der Projektion jeweils einzeln die Einfärbung berechnet werden muß. Die Bandbreite reicht von der Einheitsfarbe pro Fläche (*Flat-Shading*) bis zur pixelweisen Berücksichtigung der Oberflächenkrümmung, definiert durch eine Approximation der Flächennormalen in den Endpunkten (*Phong Shading*). Zur weiteren Verbesserung des visuellen Eindrucks kann man eine einstellbare Unebenheit der Oberfläche durch ein ‘‘Zittern’’ der Oberflächennormale erreichen (*Bump Mapping*) und vorgefertigte Materialmuster, wie z.B. Marmor, auf die einzufärbende Fläche abbilden (*Texture Mapping*).

Kommerzielle Computergrafik-Software enthält typischerweise einen 3D-Editor zum interaktiven Modellieren der Szene in Realzeit. Ermöglicht durch die Repräsentation der Körper mittels weniger Kontrollpunkte lassen sich zahlreiche geometrische Verformungen auf ausgewählte Objekte anwenden. Durch Flächenextrusion oder Drehvorschriften können aus 2D-Polygonen 3D-Körper ‘‘wachsen’’. Nach dem Szenenaufbau erzeugt der Renderer je nach gewünschter Darstellungsqualität in Sekunden oder Stunden das fertige Bild, welches manchmal von einem realen Photo nicht zu unterscheiden ist. Eine schrittweise Modifizierung von Objekt- oder Betracht-

erstandpunkt bringt Bewegung ins Spiel. Zum Beispiel kann eine Kamerafahrt durch ein Gebäude durch Definition eines Bewegungspfads für den Betrachterstandpunkt simuliert werden. Ergebnis einer solchen Animation, für die Stunden oder Tage veranschlagt werden müssen, ist eine Folge von Einzelbildern. Je nach Einsatzgebiet werden diese mit geringer Qualität direkt von der Festplatte als Film abgespielt oder in hochauflösender Qualität auf einen Videorecorder mit Einzelbildaufnahme überspielt zur späteren Wiedergabe auf einem Fernsehmonitor.

Audio

Der Begriff *Audio* stammt von dem lateinischen Wort *audire* (hören) und dient als Sammelbegriff für akustisch wahrnehmbare Signale. Eine periodische Luftschwingung wird vom menschlichen Ohr als Ton empfunden; die Höhe des Tons wird durch die Frequenz bestimmt, die Lautstärke durch die Amplitude, die Klangfarbe durch die Schwingungsform. Der hörbare Bereich liegt zwischen 20 Hz und 20 KHz. Der Abstand eines Tons bis zu dem mit doppelter bzw. halber Frequenz erzeugten Ton nennt man *Oktave*. Sie wird in 12 Halbtöne unterteilt (Frequenzfaktor zum Vorgänger $\sqrt[12]{2}$). Das normale Ohr kann Tonhöhendifferenzen von etwa einem zwanzigstel Halbtonschritt wahrnehmen.

Die Schallintensität wird definiert als Leistung pro Fläche (Watt pro qm), als *Schallpegel* bezeichnet man den 10fachen dekadischen Logarithmus (dezibel) vom Verhältnis zweier Schallintensitäten. Eine Zunahme von 10 dB wird erreicht durch eine Verzehnfachung der Leistung. Ein trainiertes Ohr kann eine Zunahme von 1 dB wahrnehmen; die Schmerzgrenze liegt bei etwa 130 dB.

Bei der Digitalisierung werden kontinuierlich-analoge Signale in gleichbleibenden Intervallen abgetastet und die ermittelten Werte quantisiert. Es entsteht somit eine Folge diskreter Werte, deren Qualität von der Abtastfrequenz (z.B. 10 kHz) und der Auflösung (z.B. 8 Bit für 256 Quantisierungswerte) abhängt (siehe Abbildung 4).

Der Dynamikbereich des menschlichen Ohrs beträgt etwa 100 dB. Etwa 6 dB entsprechen einem Verdoppeln der Amplitude. Bei binärer Kodierung werden also 16 Bit benötigt, um $16 \times 6 = 96$ dB abzudecken. Eine Auflösung von 8 Bit führt vor allem bei leisen Tönen zu einem deutlich hörbaren Quantisierungsrauschen, da leichtes Signalschwanken zu großen diskreten Sprüngen führen kann.

Zur Vermeidung von *Aliasing*-Effekten muß die Abtastfrequenz mindestens doppelt so groß sein wie die höchste vorkommende Frequenz (Abtasttheorem von Nyqist). Da sich der hörbare Bereich bis 20 KHz erstreckt, ist eine Abtastfrequenz von mindestens 40 KHz erforderlich. Der im *Red Book* definierte Standard für die Audio-CD sieht einen Audiobitstrom von 1411200 Bit/sec vor. Bei einer Auflösung von 16 Bit für zwei Stereokanäle erlaubt dies eine Frequenz von 44100 Hz. Ein 3-Minuten Musikstück in HiFi-Stereo-Qualität benötigt daher etwa 30 MByte. Ein Telefongespräch im Frequenzbereich von 200-3200 Hz verlangt nur eine Abtastfrequenz von 8 KHz mit 8 Bit Auflösung Mono und begnügt sich daher mit 500 KByte pro Minute, also etwa 1/20 der CD-Datenrate.

MPEG-1-Audio Layer I ist ein Komprimierungsstandard der *Motion Picture Expert Group*, welcher die von einer Audio-CD lieferbare Tonqualität beibehält bei einer Datenrate von 2×192 KBit/sec. Dies entspricht einer Reduktion um den Faktor 3.5. Somit kann ein Single-Speed-CD-ROM-Laufwerk mit einer Transferleistung von etwa 1.34 MBit/sec zur Wiedergabe eines Spielfilms etwa 1/4 seiner Bandbreite für den komprimierten Ton, 3/4 seiner Bandbreite für das komprimierte Video verwenden.

Ähnlich wie bei der JPEG-Bildkomprimierung werden die Audio-Abtastwerte durch eine Fourier-Transformation aus dem Zeitbereich in den Frequenzbereich umgesetzt. Durch eine Spektralanalyse wird ermittelt, welche Frequenzen in welchem Maße am Ausgangssignal beteiligt sind. Auf Grundlage eines psychoakustischen Modells wird nun der Maskierungseffekt zwischen den einzelnen Frequenzen ermittelt. Z.B. verdeckt eine Frequenz von 1000 Hz einen um mindestens 18 dB leiseren Ton von 1100 Hz oder einen um 45 dB leiseren Ton von 2000 Hz. In der Umgebung einer starken Frequenz ist daher ein gewisser, nicht hörbarer Grundpegel einer anderen Frequenz akzeptabel, die dadurch weniger Bits zur Kodierung der restlichen Amplitude benötigt. Layer II und III, insbesondere geeignet für niedrigere Abtastraten bis runter zu 64 KBit (ISDN), erreichen durch komplexere Implementierungen eine weitergehende Kompression. MPEG-2-Audio bietet als Weiterführung zusätzlich Mehrkanal- und Surround-Sound sowie Abtastraten von 16 KHz und 24 KHz an.

Midi

Die Kodierung von Musik, welche nicht als analoges Gesamtsignal vorliegt, sondern elektronisch erzeugt wird, begnügt sich mit deutlich weniger Platz. Wie im Kapitel *Audio* bereits erwähnt, läßt sich ein Ton charakterisieren durch

- die Höhe (Schwingungsfrequenz),
- die Lautstärke (Schwingungsamplitude) und

- die Klangfarbe (Schwingungszusammensetzung).

Unter Schwingungszusammensetzung versteht man die Intensitäten der beteiligten Obertöne (ganzzahlige Vielfache des Grundtons), welche den charakteristischen Unterschied zwischen dem Kammerton A, erzeugt von einer Geige, und dem Kammerton A, erzeugt von einer Oboe, ausmachen. Zur Nachahmung des Klangs eines Musikinstruments muß ein Tongenerator daher die richtige Mischung von Obertönen verwenden sowie den für das Musikinstrument typischen dynamischen Lautstärkeverlauf, definiert durch die Hüllkurve, nachahmen (siehe Abbildung 5). Bei festgelegter Klangfarbe besteht nun die Beschreibung eines Musikstücks aus der Folge der Frequenz-, Amplituden- und Tondauerwerte.

Die Sprache MIDI (*Musical Instrument Digital Interface*) verfolgt genau dieses Prinzip. Basierend auf einer von Musikgeräteherstellern definierten Auswahl von einigen Hundert Klangfarben werden die angeschlossenen Synthesizer auf jedem der bis zu 16 beteiligten Kanäle mit Steuerinformationen über die Wahl der Klangfarbe und über den Melodieverlauf versorgt. Dies geschieht durch eine Sequenz von meistens drei Byte langen Kommandos, die über eine serielle Schnittstelle mit 31250 Baud an den Klangerzeuger übertragen werden. Z.B. folgt dem NOTE-ON-Befehl ein Byte mit der gewünschten Note (diskretisiert in 128 Halbtonschritten) und ein Byte mit der gewünschten Anschlagsstärke (diskretisiert über 256 Werte).

Bei einem Musikstück mit Schlagzahl 120 im Vier-Vierteltakt ergeben sich 30 Takte pro Minute. Pro Takt mögen etwa 10 Ereignisse (z.B. Note an, Note aus), kodiert jeweils in 3 Bytes, stattfinden. Erweitert auf 16 Kanäle ergibt dies einen Datenfluß von $30 \times 10 \times 3 \times 16 = 14$ KByte pro Minute. Verglichen mit einem HiFi-Audio-Strom von 10 MByte pro Minute resultiert daraus eine Platzersparnis von 14:10000 = 1:700.

Ursprünglich als reines Echtzeit-Protokoll entwickelt, enthält der MIDI-Standard inzwischen auch Vokabeln zum Diskretisieren der Zeitachse, um den dynamischen Ablauf in einer Datei zu konservieren. Dieser Vorgang wird von sogenannten *MIDI-Sequenzern* übernommen, welche die am PC eintreffenden MIDI-Daten in Echtzeit verarbeiten und abspeichern können. Ungenauigkeiten im Timing können durch Quantisierung korrigiert werden. Ein integrierter Noteneditor erlaubt die Darstellung des Musikstücks in konventioneller Notenschrift in der gewünschten Tonart sowie die interaktive Manipulation von musikalischen Parametern wie Tonhöhe, -stärke und -länge.

Video

Der Begriff Video stammt vom lateinischen Wort *videre* (sehen) und bezeichnet Sequenzen von Einzelbildern, die zur Bildschirmausgabe geeignet sind.

Als Aufnahmegerät verwendet man seit den 70er Jahren typischerweise CCD-Kameras (*Charge Coupled Device*), die hinter einem Linsensystem Tausende von Speicherzellen aufweisen, die sich bei Lichteinfall aufladen. Consumer Camcorder verwenden 1/4-Zoll Chips mit der PAL-Auflösung von $768 \times 576 = 440.000$ Bildpunkten. Durch die unvollkommene Speicherwirkung des Auges ist für eine flimmerfreie Wiedergabe eine Bildwiederholfrequenz von 50 Hz erforderlich. Diese wird dadurch erreicht, daß zwei ineinander verschränkte Halbbilder (*interlaced modus*) mit halber Zeilenzahl jeweils mit 25 Hz aufgebaut werden. Bei 24 Bit Farbtiefe pro Pixel entsteht somit eine Transferrate von $768 \times 576 \times 3 \times 25 = 32$ MByte/sec. Auf eine CD mit 660 MByte Speicherkapazität würde demnach ein 20-Sekunden-Film passen. Aber zum Abspielen durch ein Single-Speed-CD-ROM-Laufwerk wäre die für Video verfügbare Transferrate (75 % von 1.34 MBit/sec) um den Faktor 250 zu niedrig.

Drei Komponenten tragen zur Datenreduktion bei:

1. Vereinfachung des Videosignals (*Subsampling*, Faktor 8),
2. Ausnutzung räumlicher Redundanz (JPEG, Faktor 10-15),
3. Ausnutzung zeitlicher Redundanz (MPEG, Faktor 2-3).

- Zu 1.

Üblicherweise besteht ein PAL-Fernsehbild aus zwei räumlich verschränkten und zeitlich versetzten Halbbildern mit halbierter Zeilenzahl. Da bei sich bewegenden Motiven zwei aufeinanderfolgende Halbbilder im Abstand von 1/50 sec = 20 msec kein konsistentes Vollbild ergeben, stützt sich die Kompression auf ein Folge von Halbbildern mit einer Frequenz von 25 Hz. Entsprechend der halbierten vertikalen Auflösung wird auch die horizontale reduziert, und es bleiben 384×288 Pixel übrig. Die Farbinformation wird ohne Informationsverlust in das YUV-Modell überführt, welches den drei Farbwerten Rot (R), Grün (G), Blau (B) einen Luminanzwert (Y) und zwei Farbdifferenzen (U,V) zuordnet:

```
Y = 0.30*R + 0.59*G + 0.11*B
U = (B-Y)*0.493
V = (R-Y)*0.877
```

Da das Auge für Helligkeitssprünge sensitiver ist als für Farbdifferenzen, kann man die Y-Matrix in der vollen Auflösung belassen und in den U- und V-Matrizen jeweils 4 benachbarte Pixelwerte mitteln. Dieses Subsampling liefert daher pro Halbbild eine 384×288 Matrix und zwei 192×144 Matrizen. Verglichen mit der ursprünglichen Auflösung bedeutet dies eine Reduktion um den Faktor 8.

- Zu 2. und 3.

Eine Videosequenz wird in Gruppen unterteilt. Alle Gruppen haben die gleiche Anzahl von Bildern. Die Gruppen werden zusammenhängend komprimiert und erlauben den unmittelbaren Zugriff nur auf das Ausgangsbild der Gruppe. Da drei Zugriffsmöglichkeiten pro Sekunde möglich sein sollen, enthält eine Gruppe 25/3 = 8 Bilder. Es gibt

– *I-Picture* (*Intra-Coded-Picture*)

Das Anfangsbild einer Gruppe wird auf der Basis von 8×8 Macro-Blöcken einer JPEG-Kompression unterworfen, d.h. Diskrete Cosinus-Transformation, Quantisierung, Lauflängenkodierung der Koeffizienten mit Huffman-Tabellen.

– *P-Picture* (*Predictive Coded Picture*)

Ein *P-Picture* wird mit Bezug auf ein Referenzbild kodiert, welches durch ein vorangegangenes *I-Picture* gegeben ist. Dabei wird ausgenutzt, daß sich in zeitlich aufeinanderfolgenden Einzelbildern gewisse Bildbereiche komplett verschieben, z.B. durch einen Kameraschwenk oder durch ein sich bewegendes Objekt. Es wird daher zu jedem Macro-Block des *P-Pictures* ein möglichst ähnlicher Macro-Block im Referenzbild gesucht. Gespeichert wird der Verschiebevektor und eine Differenzmatrix mit den beobachteten Pixelabweichungen.

– *B-Picture* (*Bidirectionally predictive Coded Picture*)

Ein *B-Picture* bezieht sich auf ein vorangegangenes *P-* oder *I-Picture* und auf ein nachfolgendes *P-* oder *I-Picture.*

Es wird daher zu jedem Macro-Block des *P-Pictures* ein möglichst ähnlicher Macro-Block in der Interpolation zweier Referenzbilder gesucht.

Zur Kompression und Dekompression muß die ursprüngliche Bildfolge umgestellt werden (siehe Abbildung 6). Während der Anzeige des ersten *I-Bildes* wird das nachfolgende *P-Bild* dekodiert. Die nächsten *B-Bilder* werden unmittelbar nach ihrer Dekodierung angezeigt. Nun kann das bereits dekodierte *P-Bild* angezeigt werden, während das nächste *I-Bild* dekodiert wird.

Da Videorecorder ihre Bildsequenzen mit 25 Hz abliefern, muß die Kodierung in Echtzeit erfolgen. Zur Erzeugung einer MPEG-Datei wird oft zunächst mit Hilfe einer Hardwarekompression eine Folge von *I-Pictures* erzeugt, genannt *Motion-JPEG*. Daraus entsteht dann offline per Software das MPEG-Format.

Im Gegensatz zum linearen Videoschnitt, bei dem ausgewählte Clips auf dem Originalband nur durch sequentielles Spulen angefahren und dann kopiert werden können, bietet der digitale Videoschnitt den wahlfreien Zugriff auf jede einzelne Szene, allein begrenzt durch die Kapazität der Festplatte, auf der das komprimierte Material vorliegt. Weiterhin können Überblendungen, Filter und Verzerrungen nahezu unbegrenzter Vielfalt eingebaut werden. Schließlich vereinfacht sich auch die nachträgliche Vertonung, da Audiotracks längs der Zeitachse im Schneidefenster millisekundengenau plaziert werden könnnen. Jedoch liefern die heutigen Consumer-Videokarten mit Kompressionsraten von etwa 1:10 solche Artefakte, daß sich das entkomprimierte Video deutlich von einer analogen Kopie unterscheidet.

Fazit

Prozessorleistung, Plattenkapazität und Algorithmik haben inzwischen einen Stand erreicht, mit dem unter günstigen Voraussetzungen eine Illusion der Wirklichkeit vorgegaukelt werden kann. Mit zunehmender Verbreitung des Internets und dem Wunsch seiner Benutzer nach multimedialen Daten wachsen die Anforderungen an die Komprimierungsverfahren und Übertragungsbandbreiten. Virtuelle Welten können daher auch in Zukunft nur durch die Symbiose von ausgeklügelter Software und ausgereizter Hardware ermöglicht werden. □