Seminarvortrag „Kompressionsalgorithmen“

Florian Pflug

Betreuer: Martin Oehler

18. Juni 2007

Inhaltsverzeichnis

Motivation

Die Datenraten herkömmlicher und kommender Systeme überschreiten teilweise die Grenzen des Handhabbaren. So liegt die Datenrate bei der klassischen Audio-CD bereits bei 2*16 bit*44,1 kHz = 1,44 Mbit/s, was einen üblichen 1 Mbit/s-DSL-Anschluss bereits auslasten würde. Bei aktuellen Systemen (Blueray-Disc, HD-DVD) werden Rohdatenraten von 40 Mbit/s im Audio- und etwa 800 Mbit/s im Videobereich erreicht, was selbstverständlich ohne Datenratenreduktion mit handelsüblichen Systemen überhaupt nicht übertragbar oder speicherbar wäre. Datenreduktion tut also not.

Grundlagen Kodierung

Bevor Daten reduziert werden können, müssen einige Grundbegriffe der Kodierungs- und Informationstheorie eingeführt werden [1].

Der Informationsgehalt eines Symbols x_i mit der Auftrittswahrscheinlichkeit $p(x_i)$ ist definiert als

$$I(x_i)=ld(\frac{1}{p(x_i)})$$

Anschaulich bedeutet dies, dass mit steigender Auftrittswahrscheinlichkeit der Informationsgehalt des Symbols sinkt.

Da nun der Informationsgehalt der Symbole bekannt ist, lässt sich der mittlere Informationsgehalt der Nachricht bestimmen:

$$H(x_i)=\sum_i p(x_i)\cdot I(x_i)$$

H(x) wird auch als Entropie bezeichnet und gibt die mittlere Kodewortlänge an.

Die maximale Entropie oder der größte mittlere Informationsgehalt ist dann erreicht, wenn alle Symbole gleichverteilt sind, ihre Auftrittswahrscheinlichkeit also identisch ist. Dann gilt:

$$H(x_i)=H_0=ld(n)$$

Eine Nachricht besteht aus mehreren Teilen: Einem redundanten/nicht redundanten, und einem irrelevanten/relevanten.

Unter Redundanz versteht man den Teil der Nachricht, der keine Information enthält, da diese bereits implizit oder explizit in der Nachricht enthalten war. Ein Beispiel hierfür sind natürliche Sprachen, die so redundant sind, dass die Information selbst bei Streichung aller Konsonanten erhalten bleibt („Sprchn snd shr rdndnt“).

Irrelevanz bezeichnet den Nachrichtenteil, der für den Empfänger nicht wahrnehmbar oder nicht von Bedeutung ist. Ein Beispiel hierfür sind Frequenzen über 15-20 kHz (je nach individuellem Hörvermögen) bei Tonaufzeichnungen für den Menschen.

Nun muss bei der Datenreduktion unterschieden werden zwischen

- verlustloser Kompression, bei der ausschließlich der redundante Anteil reduziert und im Idealfall entfernt wird (PAQ, ZIP),
- und verlustbehafteter Kompression, bei der zusätzlich der irrelevante Anteil reduziert/entfernt wird (MPEG-4 Part 3, JPEG).

Während beim ersten Verfahren also keine Daten verloren gehen, da die entfernte Redundanz jederzeit wieder hinzugefügt werden kann, ist dieses beim zweiten Verfahren nicht der Fall. Im Beispiel der tiefpassgefilterten Tonaufzeichnung bedeutet dies, dass die Informationen der Frequenzanteile überhalb der Grenzfrequenz unwiederbringlich verloren gegangen sind.

Kodierverfahren

Kodierverfahren zur Redundanzreduktion [2] sind z.B.:

1. Die Prädiktionskodierung im Orts- bzw. Zeitbereich (Abbildung 1). Hierbei wird versucht, das folgende Symbol aus den bisherigen vorherzusagen. Bei geeigneten Daten entspricht dies dann einer Redundanzreduktion. Durch einen nachgeschalteten Quantisierer lässt sich eine Irrelevanzreduktion realisieren.

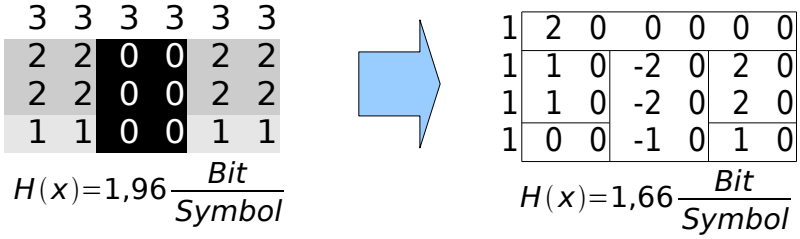

Abbildung 1: Redundanzminderung durch Prädiktion

2. Die Transformationskodierung im (Orts-)Frequenzbereich. Das Signal wird in einen geeigneten mathematischen Raum transformiert. Dadurch wird eine Konzentration der Signalenergie erreicht. Ein Beispiel hierfür ist die diskrete Kosinustransformation (s.u.).

3. Die Hybridkodierung, bei der zuerst eine Prädiktion, danach eine Transformation durchgeführt wird. Praktisch alle aktuellen Videokodierverfahren arbeiten nach diesem Prinzip.

Diskrete Kosinustransformation [2]

Die diskrete Kosinustransformation (DCT, „Discrete Cosine Transformation“) ist ein Verfahren zur Transformationskodierung. Dazu wird ein Ausschnitt des Signals *s(x,y)* (bei Bilddaten im Allgemeinen in Blöcken von 8x8 Bildpunkten) mittels der Transformationsgleichung

$$c(f_x,f_y)=\frac{1}{4}\cdot K(f_x)\cdot K(f_y)\cdot\sum_{x=0}^{7}\sum_{y=0}^{7}s(x,y)\cdot$$
$$\cdot\cos((2x+1)\cdot f_x\cdot\frac{\pi}{16})\cdot$$
$$\cdot\cos((2y+1)\cdot f_y\cdot\frac{\pi}{16})$$
$$\text{wobei } K(f)=\begin{cases}\frac{1}{\sqrt{2}}, \text{für } f=0\\ 1, \text{sonst}\end{cases}$$

in den Frequenzbereich mit den Transformationskoeffizienten *c(fx,fy)* gewandelt. Anschaulich entspricht dieses einer Nachbildung des Signals mit einer Überlagerung von Kosinustermen verschiedener Perioden (Abbildung 2).

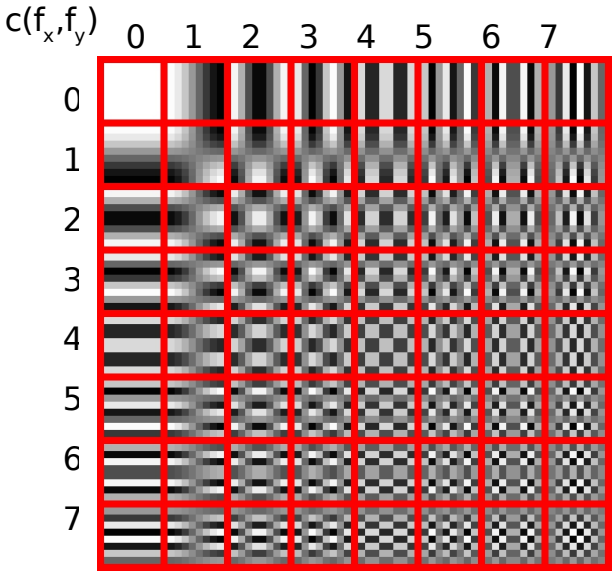

Abbildung 2: Basisfunktionen der DCT [3]

Als Ergebnis dieser Transformation ergibt sich eine Konzentration der Signalenergie bei den Koeffizienten niedriger Ordnung.

Lauflängenkodierung

Die Lauflängenkodierung ist ein einfaches Verfahren zur Redundanzreduktion, arbeitet also verlustlos. Die Idee ist, dass bei sich wiederholenden Symbolen nur ein Multiplikator und das Symbol übertragen wird [4]:

AAAAABBBBBBCCCCCCC => 5A6B7C

Bei geeignetem Quellmaterial lassen sich so bereits große Datenmengen einsparen (siehe auch „Transitional timing analysis“). Probleme können bei diesem Verfahren auftreten, wenn die Multiplikatoren und Daten den selben Zeichensatz verwenden. Dann muss gewährleistet werden, dass zwischen beiden unterschieden werden kann, es muss also Zeichenstopfen („char stuffing“) durchgeführt werden (Verfahren, bei dem Signal- von Nutzdaten durch ungültige, nicht im Zeichensatz enthaltenen Zeichen abgetrennt wird).

Beispiele für Kompressionsalgorithmen

JPEG (ITU T.81)

Das JPEG-Verfahren laut ITU T.81 beschreibt eine Methode zur verlustbehafteten Kompression von Bilddaten. Die Vorgehensweise ist dabei wie folgt [2]:

1. Transformation der Signalwerte *s(x,y)* mittels der DCT in Blöcken à 8x8 Bildpunkten in den Frequenzbereich. Dabei konzentriert sich die Signalenergie wie beschrieben bei den Koeffizienten niedriger Ordnung. Da sich dabei aber der Wertebereich stark vergrößert, hat noch keine Datenreduktion stattgefunden; Die Transformation ist

außerdem, abgesehen von Rundungsfehlern,.verlustlos umkehrbar (IDCT, inverse Kosinustransformation).

2. Division jedes Koeffizienten der Transformationsmatrix mit einem entsprechenden Koeffizienten einer Gewichtungsmatrix. Die Gewichtungsmatrix wurde empirisch ermittelt, sodass (bezogen auf der menschliche Auge) tatsächlich nur Irrelevanz entfernt wird (solange nicht zu stark quantisiert wird). Als Ergebnis ergibt sich eine Matrix, die bei Koeffizienten hoher Ordnung viele Nullen und bei Koeffizienten geringer Ordnung relativ kleine Werte (bezogen auf die ursprüngliche Transformationsmatrix) enthält.

3. Nun folgt eine Zick-Zack-Abtastung von oben links nach unten rechts. Es entsteht also eine Zahlenfolge von 64 Werten pro DCT-Block, an erster Stelle steht der Gleichstrom-Koeffizient (0,0) (die mittlere Bildhelligkeit). Von diesem Koeffizienten wird der Gleichstrom-Koeffizient des vorherigen DCT-Blocks abgezogen (Prädiktion). Damit verringert sich erneut der Wertebereich.

4. Nach der Zick-Zack-Abtastung ergeben sich etwa ab der Mitte der Zahlenfolge lange Reihen von Nullen. Die Folge wird also mittels einer Lauflängenkodierung zusammengefasst und abschließend mittels einer Huffman-Kodierung oder einer arithmetischen Kodierung entropiekodiert. Die JPEG-Kompression ist hiermit abgeschlossen.

MPEG-4 Part 3 (AAC)

MPEG-4 Part 3 (auch: „Advanced Audio Coding“, „AAC“) ist eine verlustbehaftete Audiokompressionstechnik des MPEG-Konsortiums und stellt einen Nachfolger des erfolgreichen MPEG-1 Layer 3-Standards („MP3“) dar. Neben generellen Verbesserungen wie höheren Abtastraten (bis zu 96 kHz) und einer größeren Anzahl von Audiokanälen (bis zu 48) bietet dieser Codec vor allem eine deutlich bessere Qualität bei geringeren Datenraten ([5]).

AAC arbeitet mit der modifizierten diskreten Kosinustransformation („MDCT“) und variablen Transformationslängen. Der wesentliche Unterschied zur DCT besteht darin, dass sich hier die Blöcke überlappen und somit Kompressionsartefakte an den Blockgrenzen verringert werden.

Die variablen Transformationslängen (entweder ein Block mit 1024 Abtastwerten oder acht Blöcke mit 128 Abtastwerten) bieten eine einstellbare Frequenz- oder Zeitbereichsauflösung ([6]). Bei transienten lauten Signalen wird die kleine Transformationslänge für eine hohe Zeitauflösung gewählt, um so genannte Pre-Echo-Effekte zu vermeiden, die zu wahrnehmbaren Kompressionsartefakten kurz vor diesem Signal führen würden. Die große Transformationslänge und die damit verbundene hohe Frequenzauflösung bietet hingegen eine bessere Grundlage für das psychoakustische Modell, welches zeitliche Vor- und Nachverdeckungs- und Maskierungseffekte ausnutzt, und damit eine bessere Kodiereffizienz.

Der AAC-Standard ist modular aufgebaut und stellt verschiedene Werkzeuge zur Quellkodierung von Audiosignalen zur Verfügung. Für die verbesserte Qualität bei sehr

geringen Datenraten (<64 Kibit/s) sind im Wesentlichen die folgenden zwei Werkzeuge verantwortlich:

- „Parametric Audio Coding“: Hierbei wird das Signal in eine Überlagerung von drei verschiedenen Objekten zerlegt, nur noch deren Parameter werden danach übertragen. Mit dieser Technik lassen sich große Datenmengen einsparen und gleichzeitig lässt sich das Signal beim Abspielen leicht verändern (andere Tonhöhe, Abspielgeschwindigkeit), da nur die Parameter geändert werden müssen. Die möglichen Objekte sind:

    - Rauschen, es wird die Rauschbandbreite und -amplitude übertragen. Im Empfänger wird ein quasizufälliges Signal erzeugt, welches mit den übertragenen Parametern angepasst wird.

    - Sinuston, übermittelt wird die Frequenz und Amplitude. Damit ist das Signal bereits vollständig beschrieben.

    - Harmonischer Ton, dieser wird beschrieben durch die Frequenz und Amplitude des Grundtones und einer Hüllkurve über die Obertöne. Die Frequenzen der Obertöne sind immer ganzzahlige Vielfache der Grundfrequenz. Die menschliche Stimme besteht etwa aus solchen Tönen.

- „Spectral Band Replication“: Das Problem bei der Kodierung von Audiosignalen bei geringen Datenraten ist, dass das Quantisierungsrauschen so stark ansteigt, dass es über den Mithörschwellen des Nutzsignals liegt und damit hörbar wird.

  Um dieses zu vermeiden, wird die Bandbreite des Signals mittels eines Tiefpasses begrenzt, je geringer die Datenraten, desto mehr wird herausgefiltert. Die Idee der SBR ist nun, die Korrelation zwischen den hohen und tiefen Frequenzen auszunutzen. Dazu wird im Enkoder vor der Tiefpassfilterung die Hüllkurve über die hohen Frequenzen und weitere Hilfsdaten über die Beziehung zwischen hohen und tiefen Frequenzen für den Dekoder bestimmt (SBR-Daten). Diese Daten werden komprimiert und dem Bitstrom hinzugefügt.

  Im Dekoder wird mit diesen Daten dann der hohe Frequenzbereich annähernd wiederhergestellt. Nicht-SBR-fähige Dekoder ignorieren diese Daten einfach, das Verfahren ist also rückwärtskompatibel.

Kanalkodierung [1]

Da nun die Redundanz der Nachricht weitestgehend entfernt wurde, ist sie sehr anfällig gegenüber Störungen geworden. Vor der Übertragung über einen realen und damit gestörten Kanal muss jetzt also möglichst gezielt Redundanz hinzugefügt werden. D.h. die Redundanz sollte angepasst auf die Kanaleigenschaften (Bitfehler-, Büschelfehlerwahrscheinlichkeit) sein.

Bei geringer Bitfehlerwahrscheinlichkeit und sehr geringer Büschelfehlerwahrscheinlichkeit wäre ein Hammingkode geeignet. Dieser Kode garantiert bei einstellbaren Blocklängen die Korrektur genau eines Bitfehlers. Bei höheren Fehlerwahrscheinlichkeiten und großen Blocklängen werden also unkorrigierbare Fehler immer wahrscheinlicher. Dafür ist der Hammingkode mit einer einfachen Schieberegisterschaltung realisierbar.

Bei hoher Einzelbitfehlerwahrscheinlichkeit bietet sich der Faltungskode an. Dieser Kode arbeitet mit einem Bitstrom und verteilt eingehende Bits auf mehrere Ausgangsbits, er „verschmiert“ die Information also zeitlich und örtlich. Dieser Kode kann mit längerer Gedächtnistiefe (also größeren „Verschmierungslänge“) sehr mächtig werden, sodass er mit komplizierteren Blockkodes mithalten kann.

Der Reed-Solomon-Kode ist effizient bei hohen Büschelfehlerwahrscheinlichkeiten. Er arbeitet mit Datensymbolen, also Gruppen von mehreren Bits. Die Menge der hinzuzufügenden Redundanz ist einstellbar und legt die maximale Anzahl erkennbarer und korrigierbarer Symbolfehler fest. Mit dem Einsatz eines nachgeschalteten Interleavers und einem weiteren (inneren) Reed-Solomon-Kodes kann der Kode eine große Anzahl von Fehlern korrigieren (sowohl Büschel-, als auch Einzelbitfehler, im Einsatz zum Beispiel bei CD und DVD).

Anwendung bei Messdaten

„Fast verlustlose“ Kompression

Bei vielen Messdaten, wie z.B. Medizindaten, sind verlustbehaftete Kompressionsalgorithmen nicht erwünscht, da die Gefahr besteht, dass für die Diagnostik oder Auswertung relevante Informationen verloren gehen. Andererseits sind die Kompressionsraten verlustloser Verfahren oftmals zu gering, um große Rohdatenraten ausreichend zu verringern. Einen Ausweg aus diesem Dilemma können so genannte „fast verlustlose“ Kompressionsalgorithmen („near lossless“) bieten ([7]). Die Idee ist, dass man vor der Irrelevanzreduktion einen Bereich definiert, in dem sich der Kompressionsfehler bewegen darf (z.B. „95% der Daten werden bitgenau wiederhergestellt“ oder „die Graustufe eines Bildpunktes ändert sich maximal um 1“). Während das erste Beispiel keine wirkliche Besserung gegenüber herkömmlichen verlustbehafteten Verfahren ist (der Haarriss eines Knochenbruchs auf einem Röntgenbild könnte ja genau in den verlustbehafteten 5% des Bildes liegen und damit unsichtbar werden), stellt das zweite Verfahren eine gute Alternative dar.

Während die Kodiervorschrift sehr einfach ist ($x_{original} - x_{rekonstruiert} < d$), stellt sich die Frage nach der praktischen Realisierung. Diese funktioniert ähnlich wie die Dekodierung eines Faltungskodes, mittels eines Trellis-Diagramms und nach dem Viterbi-Algorithmus, bei dem alle möglichen und erlaubten Symbolreihen durchprobiert und diejenige mit der geringsten Entropie ausgewählt wird.

Eine Verfeinerung dieser Technik ist es, die Messdaten vorher zu klassifizieren und je nach Relevanz entweder verlustlos oder -behaftet zu komprimieren. Im Beispiel des EP1536567 (Verfahren zur Datenkompression der Messdaten eines Prüfmolchs für Rohrleitungen, [8]) bedeutet dies, dass Messdaten an Korrosions- und Schweißstellen verlustlos und ansonsten verlustbehaftet komprimiert werden. Die Unterscheidung zwischen beiden Klassen findet über die Differenz des gleitenden Mittelwerts zum aktuellen Messwert statt, sobald diese einen bestimmten Wert überschreitet, wird die „Verlustlos-Klasse“ gewählt.

Transitional Timing Analysis [9]

Diese Methode zur Redundanzreduktion findet bei Logikanalysatoren Verwendung. Bei sich zeitlich nur langsam veränderlichen Signalen und hohen Abtastraten wird normalerweise viele Male der selbe Wert aufgenommen und gespeichert. Eine Lauflängenkodierung im Nachhinein würde die Datenmenge zwar erheblich reduzieren, wäre aber aufwändig, da die Daten ja zuerst zwischengespeichert werden müssen. Die Idee ist also, Messwerte zusammen mit einem Zeitstempel nur dann zu speichern, wenn sie sich auch geändert haben. So werden längere Messungen mit einem geringen Speicherplatzbedarf möglich gemacht.

Ausblick

Kryptographie

Neben der Redundanzerhöhung bei der Kanalkodierung ist auch die Kryptographie sehr wichtig geworden. Die Ziele hierbei sind:

- Authentizität: Die Echtheit der empfangenen Daten soll sichergestellt sein (Schutz vor Manipulationen oder unentdeckten Fehlern)
- Verschlüsselung: Unlesbarkeit der zu übertragenden Daten für Unbefugte (bei vertraulichen Messwerten, geheimen Daten)

Algorithmen arbeiten dabei nach dem symmetrischen oder asymmetrischen Prinzip. Bei der symmetrischen Verschlüsselung verwenden Sender und Empfänger den selben Schlüssel (das selbe Geheimnis) zur Ver- und Entschlüsselung, was das Verfahren sehr einfach macht, mit der Kompromittierung des Schlüssels ist die gesamte Kommunikation aber vollständig transparent.

Asymmetrische Verfahren verwenden für die Verschlüsselung einen anderen Schlüssel („öffentlicher Schlüssel“) als für die Entschlüsselung („privater Schlüssel“). Mit dem öffentlichen Schlüssel können Daten nur verschlüsselt werden (und digitale Signaturen des Inhabers überprüft werden, s.u.) und er wird jedem zur Verfügung gestellt, der verschlüsselte Daten an den Eigentümer des Schlüsselpaares übermitteln möchte. Der private Schlüssel wird vom Empfänger geheim gehalten und dient der Entschlüsselung empfangener Nachrichten und der Erstellung digitaler Signaturen, mit der die Authentizität gesendeter Nachrichten sichergestellt werden kann.

Zukünftige Kompressionsalgorithmen

Kommende Kompressionsalgorithmen werden eine deutlich erhöhte Komplexität aufweisen. So ist es bei den aktuellsten Algorithmen heutzutage schon üblich, mit mehreren Statistikmodellen gleichzeitig zu arbeiten und stets jenes auszuwählen., welches für den aktuell betrachteten Datenblock das beste Ergebnis liefert (z.B. PAQ8f: Vier verschiedene Prädiktionsmodelle, [10]). Gleichzeitig wird auch die Gedächtnistiefe, in welcher nach redundanten Mustern gesucht wird, immer größer (etwa bei rzip: 900 MB, [11]) und praktisch nur noch von der Arbeitsspeichergröße begrenzt. Weitere

Verbesserungen in der Kompressionsrate bei verlustlosen Verfahren lassen sich durch kontextabhängige Kompression erreichen, wenn bei verschiedenen Quellmedien (etwa Texte, Bilder oder Töne) verschiedene spezialisierte Algorithmen angewandt werden. PAQ8 verwendet beispielsweise bei Texten eine Wortprädizierung unterschiedlicher Tiefe, während bei JPEG-Bildern die Huffmankodierung und die DCT rückgängig gemacht wird, um danach mehrere Blöcke zusammenzufassen und so die Länge der Huffman-Kodewörter zu verringern.

Der aktuellste Stand der Kompressionstechniken lässt sich bei diversen Wettbewerben verfolgen, bei denen es um die Erreichung möglichst großer Kompressionsraten bei gegebenen Quelldaten geht. So wird z.B. beim „Hutter-Prize“ ([12]) eine 100 MiB große Textdatei auf aktuell 16,5 MiB reduziert (zum Vergleich: Zip erreicht mit der aufwändigsten Kompression gerade mal 33 MiB, BZIP2 28 MiB).

Literaturverzeichnis

[1] Prof. Kürner, „Skript Codierungstheorie", 2006
[2] Prof. Reimers, „Skript Bildkommunikation II", 2007
[3] DCT-Basisfunktionen: http://commons.wikimedia.org/wiki/Image:Dctjpeg.png
[4] Joachim Schwarz, Guido Sörmann, „Kompressionsalgorithmen", 1995
[5] Stefan Meltzer, Gerald Moser, „MPEG-4 HE-AAC v2 -audio coding for today's digital media world", Ausg. , Nr. , 2006
[6] Peter Doliwa, „MPEG-4 Advanced Audio Coding", 2004
[7] Nasir Memon, Xuan Kong, „Context-Based Lossless and Near-Lossless Compression of EEG Signals", Ausg. 3, Nr. 3, 1999
[8] Method for data compression: http://www.freepatentsonline.com/EP1536567.html
[9] Logic Analyzers from NCI: http://www.nci-usa.com/features_software_step1.htm
[10] The PAQ Data Compression Programs: http://www.cs.fit.edu/~mmahoney/compression/
[11] rzip-Algorithmus: http://rzip.samba.org/
[12] The Hutter Prize: http://prize.hutter1.net/