

Quellenkodierung und Datenverdichtung

Zusammenfassung VL375-Kapitel 6, Christof Roduner (roduner@access.unizh.ch)

Quellenkodierung

Ziel der Quellenkodierung ist es, Daten (z.B. Bilder, Klänge etc.) so zu kodieren, dass sie möglichst wenig Platz beanspruchen und somit wirtschaftlich gespeichert bzw. übertragen werden können. Für eine möglichst optimale Kodierung werden normalerweise die Charakteristika der Datenquelle und -senke berücksichtigt. Die Quellenkodierung kann durch zwei verschiedene Teilprozesse erfolgen: Einerseits werden irrelevante Daten meist unter Berücksichtigung der menschlichen Wahrnehmung eliminiert. Dieser Vorgang ist verlustbehaftet und daher irreversibel. Andererseits wird die Redundanz in den Daten durch gängige verlustfreie Verfahren der Entropiekodierung entfernt.

Verlustfreie Verfahren

Shannon definierte den mittleren statistischen Informationsgehalt (H) einer Quelle, die Nachrichten bestehend aus einer Sequenz der Symbole x_1, x_2, \dots, x_N erzeugt, als $H = -\sum_{i=1}^N p(x_i) \log_2 p(x_i)$. Dabei steht $p(x_i)$ für die Auftretenswahrscheinlichkeit der Symbole x_i . Die mittlere Wortlänge ist definiert als $H_0 = \log_2 N$. Die Redundanz ergibt sich dann als $R = H_0 - H$. Diese Redundanz kann mittels geeigneter Kodierverfahren minimiert werden. Allerdings gilt die obenstehende Formel über den mittleren statistischen Informationsgehalt nur, wenn die einzelnen Symbole statistisch unabhängig voneinander auftreten. In realen Anwendungen (z.B. Bilder) tritt dieser Fall aber nie auf. Für den allgemeinen Fall mit statistischen Abhängigkeiten stehen daher komplexere Berechnungsverfahren zur Verfügung, die aber in der Praxis mit einem sehr hohen Rechenaufwand verbunden sind und daher nicht verwendet werden können. In realen Anwendungen stehen deshalb pragmatische Ansätze für die Datenverdichtung im Vordergrund. Beispiele dafür sind die Lauflängenkodierung, der Fano-Kode, der Huffman-Kode oder LZW.

Lauflängenkodierung (Run Length Encoding)

Dieses Kodierverfahren ersetzt eine Folge gleichartiger aufeinanderfolgender Symbole durch eine kürzere Repräsentation: Jedes Symbol wird nur noch ein Mal gespeichert. Dafür wird angegeben, wieviele Male es sich wiederholt. So wird z.B. aus der Folge ABBCCCCAAAAAADDDDD die kürzere Folge A1B2C4A6D5. Ein Spezialfall liegt bei einem Binärkode vor: Da es nur zwei Symbole (0, 1) gibt, die sich ständig abwechseln, muss nicht mehr vermerkt werden, um welches Symbol es sich handelt. Aus der Folge 000011101111 wird dann 4314. Die Lauflängenkodierung ist für Schwarz-/Weissbilder gut geeignet, da dort meist lange Folgen weisser Bildpunkte auftreten. Für Grau- und Frabbilder hingegen ist sie ungünstig.

Huffman-Kode

In den Symbolfolgen realer Anwendungen (z.B. Bild- oder Klangdaten) treten nicht alle Symbole gleich häufig auf. Der Huffman-Kode basiert nun auf der Idee, häufige Symbole durch kurze Bitfolgen zu kodieren, während seltene Symbole längere Bitfolgen erhalten. Vor der eigentlichen Kodierung muss ein Kodier- / Dekodierbaum aufgebaut werden, der die Charakteristika der Datenquelle berücksichtigt. Dies geschieht folgendermassen: Als erstes werden die Symbole nach absteigender Häufigkeit sortiert. Dann wird den beiden Symbole mit der geringsten Häufigkeit eine 0 bzw. 1 zur Kodierung zugeordnet. Dann werden die beiden Häufigkeiten zusammengefasst und neu in der sortierten Liste eingeordnet. Dieser Vorgang wird wiederholt, bis alle Symbole zusammengefasst sind. Mit diesem Verfahren wird ein Baum aufgebaut, der den Symbolen Bitfolgen variabler Länge zuordnet. Diese Bitmuster können später mit dem Dekodierbaum wieder eindeutig in die ursprünglichen Symbole übersetzt werden. Ein Problem bei diesem Verfahren sind die gravierenden Auswirkungen, die ein „gekipptes“ Bit in der kodierten Bitfolge hat.

Fano-Kode

Das Verfahren von Shannon-Fano ähnelt dem Huffman-Kode. Auch hier werden die Symbole gemäss ihrer Häufigkeit sortiert. Die Liste wird dann in zwei gleichhäufige Teillisten unterteilt, wobei der einen Hälfte eine 0, der anderen eine 1 zugeordnet wird. Dieser Vorgang wird nun rekursiv auf jede Teilliste mit mehr als zwei Einträgen angewendet. Im allgemeinen erzeugt dieser Algorithmus einen etwas redundanteren Kode als Huffman.

LZW

LZW (Lempel, Ziv, Welch) beschränkt sich nicht mehr nur auf die Kodierung einzelner Symbole, wie es in den obigen Verfahren der Fall war. Stattdessen wird ein Verzeichnis häufiger Symbolfolgen aufgebaut, denen neue Symbole zugeordnet werden. Als verlustfreies Verfahren erreicht LZW damit gute Kompressionsraten.

Faksimilekodierung (Gruppe 3, 1D)

Beim Fax-Verfahren wird die Vorlage zeilenweise abgetastet. Jede Zeile besteht aus 1728 schwarzen oder weissen Bildelementen. Bei der Kodierung werden in einem ersten Schritt die Lauflängen über schwarze oder weisse

Quellenkodierung und Datenverdichtung

Zusammenfassung VL375-Kapitel 6, Christof Roduner (roduner@access.unizh.ch)

Bereiche dieser Zeile bestimmt. Diesen Lauflängen werden dann mittels fixer Kodiertabelle Kodewörter variabler Länge zugeordnet. Die verwendete Tabelle entspricht im Prinzip einer Huffman-Kodierung mit festem Kodierbaum. Sie wurde unter Berücksichtigung einer grossen Zahl üblicher Dokumente erstellt und von CCITT/ITU standardisiert. Das Faksimileverfahren ist für Bilder oder Graufstufen ungeeignet (siehe RLE).

Verlustbehaftete Verfahren

Im Gegensatz zu verlustfreien Verfahren entfernen verlustbehaftete Verfahren nicht redundante, sondern irrelevante Information. Was relevant bzw. irrelevant ist, hängt von der speziellen Anwendung und der menschlichen Wahrnehmung ab.

JPEG

JPEG (Joint Photographic Expert Group) ist ein Standard zur verlustbehafteten Speicherung von Bilddaten. Um eine Bilddatenquelle in JPEG zu kodieren, sind folgende Schritte nötig:

1. Unterteilung des Bilds in 8x8 Punkte grosse Tiles
2. Diskrete Cosinus Transformation (DCT): Ähnlich wie die Fourier Transformation überführt sie jedes Tile vom Orts- in den Frequenzbereich. Die DCT ist rechenaufwendig und wird daher nicht auf das Gesamtbild, sondern auf die einzelnen Tiles angewendet. Im DCT-transformierten Tile finden sich in der oberen linken Ecke die niedrigen Frequenzen und in der unteren rechten Ecke die hohen Frequenzen.
3. Quantisierung: Da normale Bilder Tiefpasscharakter haben, kann die Genauigkeit der höheren Frequenzanteile und damit der Speicherbedarf reduziert werden. Dies wird erreicht, indem man jeden Wert des transformierten Tiles durch das entsprechende Element einer Quantisierungsmatrix dividiert. Dieser Vorgang ist natürlich verlustbehaftet. Für verschiedene Qualitätsstufen stehen daher standardisierte Matrizen zur Verfügung. Diese Matrizen weisen im oberen linken Bereich (niedere Frequenzen) jeweils geringere Werte auf als im unteren rechten (höhere Frequenzen). Das so entstandene quantisierte Tile besitzt daher im weniger relevanten hochfrequenten Bereich zahlreiche Nullwerte.
4. Zick-Zack-Anordnung: Die quantisierten Werte des Tiles werden nun – beginnend in der oberen linken Ecke – in einem Zick-Zack-Pfad ausgelesen. In der linearen Anordnung stehen so die niederfrequenten Werte am Anfang und die hochfrequenten am Schluss.
5. RLE-Kodierung: In der entstandenen Anordnung liegen speziell im hochfrequenten Bereich zahlreiche Nullen direkt nebeneinander. Diese werden mit einer verlustfreien RLE-Kodierung verdichtet.
6. Huffman-Kodierung: Die resultierenden Daten werden nun mit einer verlustfreien Huffman-Kodierung erneut verdichtet.

Die Dekodierung von JPEG-Bildern verläuft analog, aber in umgekehrter Reihenfolge. Mit diesem verlustbehafteten Verfahren erreicht man massiv höhere Kompressionsfaktoren (bis 100) als mit verlustfreien Methoden.

MPEG

MPEG ist ein Standard zur Komprimierung von Realbildsequenzen und Klängen. Die Bildkomprimierung basiert einerseits auf der Inter-Frame-Kodierung, andererseits auf der Intra-Frame-Kodierung. Die Intra-Frame-Kodierung verwendet das gleiche Prinzip wie JPEG. Die Inter-Frame-Kodierung jedoch bietet eine zusätzliche Verdichtung: In Videosequenzen bewegen sich meist nur einzelne Teile des Bildes, während z.B. der Hintergrund beim Übergang von einem Frame zum anderen gleich bleibt. So ist es möglich, dass sich ein Frame Teile des Gesamtbilds aus einem vorhergehenden oder nachfolgenden Frame „leiht“. Anstatt ein mehrfach auftretendes Teilbild erneut zu speichern, verweist ein Frame einfach auf den Bildausschnitt eines Nachbarframes. Allenfalls wird der so geliehene Ausschnitt mittels Verschiebevektor noch neu positioniert.

MP3

Das MP3-Format wird benutzt, um Klänge zu kodieren. Dabei macht sich dieses Verfahren psychoakustische Merkmale der menschlichen Wahrnehmung zu Nutze, um die Daten möglichst wirtschaftlich zu verdichten:

- Hörbereich: Das menschliche Ohr nimmt Töne erst ab einer bestimmten Lautstärke wahr. Diese Grenze liegt aber nicht für alle Frequenzen bei der gleichen Lautstärke. So erfordern höhere Frequenzen tendenziell eine grössere Lautstärke, bis sie wahrgenommen werden.
- Frequenzmaskierung: Frequenzen, die neben einem relativ lauten Ton liegen, werden von diesem überhört und können nicht mehr gleichzeitig wahrgenommen werden. Bei zunehmender Frequenz des relativ lauten Maskierungstons wird der so verdeckte Frequenzbereich breiter.
- Zeitmaskierung: Nach einem relativ lauten Ton benötigt das Ohr eine gewisse Zeitspanne, bis ein relativ leiser Ton ähnlicher Frequenz wahrgenommen werden kann.

MP3 komprimiert die Audiodaten nun auf eine verlustbehaftete Weise, indem es die maskierten (und für das Ohr damit irrelevanten) Bereiche entfernt. Eine weitere Verdichtung wird erreicht, indem beim Stereosignal die Summe des Links- und Rechtskanals (Mittelkodierung) und deren Differenz (Seitenkodierung) anstelle der Einzelsignale gespeichert werden. Da der Links- und Rechtskanal zu einem grossen Teil korrelieren, wird die Datenmenge so weiter verringert. Schliesslich wird auf den entstandenen Datenstrom eine verlustfreie Huffman-Kodierung angewendet.