

X . media . press



Roland Schmitz · Roland Kiefer
Johannes Maucher · Jan Schulze
Thomas Suchy

Kompendium Medieninformatik

Mediennetze

Mit 106 Abbildungen und 11 Tabellen

Roland Schmitz
Roland Kiefer
Johannes Maucher
Jan Schulze
Thomas Suchy
Hochschule der Medien
Nobelstr. 10
70569 Stuttgart
schmitz@hdm-stuttgart.de
www.medieninformatik.hdm-stuttgart.de

Bibliografische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.ddb.de> abrufbar.

ISSN 1439-3107
ISBN-10 3-540-30224-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-30224-7 Springer Berlin Heidelberg New York

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media
springer.de

© Springer-Verlag Berlin Heidelberg 2006

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften. Text und Abbildungen wurden mit größter Sorgfalt erarbeitet. Verlag und Autor können jedoch für eventuell verbliebene fehlerhafte Angaben und deren Folgen weder eine juristische Verantwortung noch irgendeine Haftung übernehmen.

Satz und Herstellung: LE- \TeX , Jelonek, Schmidt & Vöckler GbR, Leipzig
Umschlaggestaltung: KünkelOpka Werbeagentur, Heidelberg
Gedruckt auf säurefreiem Papier 33/3100 YL - 5 4 3 2 1 0

Einleitung

„The medium is the message“ – treffender und kürzer, als es Herbert Marshall McLuhan in den 60er Jahren des letzten Jahrhunderts auf den Punkt gebracht hat, kann man die wechselseitige Beziehung zwischen dem Inhalt einer Kommunikation und ihrer Präsentationsform auch heute nicht beschreiben. Versucht man davon ausgehend, den zu Grunde liegenden Sachverhalt etwas präziser zu fassen, so stellt man fest: Die Techniken zur Erzeugung, zum Transport, zur Speicherung und zur Darstellung einer Botschaft sind ebenso entscheidend für ihre Wahrnehmung, wie ihr eigentlicher Inhalt. Damit wird klar, was der viel benutzte (und strapazierte) Begriff der „Medien“ eigentlich bedeutet:

Medien sind dem Wortsinn nach „Vermittler“. Sie dienen der Speicherung, Darstellung und Übermittlung von Informationen. Im heutigen Sprachgebrauch meint man mit Medien aber häufig auch die Informationen selbst und unterscheidet nicht deutlich zwischen der Information und ihrem Träger.

Medien

Die übertragenen Informationen können in unterschiedlich strukturierter Form und Codierung vorliegen. Eine Strukturierung in Medien kann auf unterschiedliche Arten herbeigeführt werden, etwa mittels ergänzender Informationen, so genannter Metadaten, oder durch Eingliederung in eine Ordnungsstruktur, etwa eine hierarchische Gruppierung.

Die *Medieninformatik* beschäftigt sich speziell mit *digitalen Medien*, das sind zum einen digital (also in Form von Bits) codierte Arten von Informationen, zum anderen die Träger dieser Informationen. Die Art der Codierung, das heißt die Vorschrift, wie die ursprünglichen Informationen in Bitform darzustellen sind (und wie umgekehrt die Bits als Information zu interpretieren sind) bestimmt über den Medientyp, wie zum Beispiel Text, Dokument, Bild, Audio, Video. Die Digitalisierung stellt dabei einen entscheidenden Schritt dar: Durch sie wird es möglich, die Informationen von ihrem physikalischen Träger zu trennen. Diese Trennung trägt maßgeblich zur immer weiter wachsenden Mobilität unserer Kommunikation und unseres Arbeitslebens

Medieninformatik

insgesamt bei. Die Informatik stellt hierfür die theoretischen Grundlagen und Methoden der Informationsverarbeitung auf Rechnersystemen zur Verfügung. Die Medieninformatik zeigt, wie diese Methoden speziell auf digitale Medien anzuwenden sind.

Damit prägt die Medieninformatik unseren Alltag, weil sie letztlich darüber mitentscheidet, was und wie viel aus der täglichen Informationsflut in unseren Köpfen ankommt. Die Beherrschung ihrer Konzepte und Techniken ist für unser aller Zukunft maßgebend. Aus diesem Grund versteht sich die Medieninformatik auch nicht nur als rein technische Disziplin, sondern umfasst auch gestalterische, psychologische und ökonomische Aspekte. Diese Vielfalt spiegelt sich auch in den unterschiedlichen Inhalten der vielen verschiedenen, an deutschen Hochschulen angebotenen Studiengängen mit dem Namen „Medieninformatik“ wider.

Das Buch

Das „Kompendium der Medieninformatik“ befasst sich vor allem mit den technischen Aspekten der Medieninformatik und bildet damit den Schwerpunkt der Lehrinhalte im Studienbereich Medieninformatik an der Hochschule der Medien in Stuttgart ab. Trotz dieses Schwerpunkts auf den Informatik-Aspekten der Medieninformatik unterscheidet sich das Kompendium der Medieninformatik inhaltlich deutlich von klassischen Lehrbüchern der Informatik. Die Grundlagen der Informatik werden nur insoweit behandelt, als sich durch die Anwendung auf Mediendaten neue und spezifische Gesichtspunkte ergeben. In allen anderen Fällen gehen wir davon aus, dass der Leser mit den Grundlagen der Informatik bereits vertraut ist. Zielgruppen dieses Buches sind somit Studenten der Informatik und Medieninformatik im Hauptstudium sowie Praktiker, die bereits Erfahrung mit den Anwendungen der Informatik in der Industrie gesammelt haben und sich nun speziell über Anwendungen in den Medien informieren wollen.

Der vorliegende erste Teil des Kompendiums der Medieninformatik beschäftigt sich mit „Mediennetzen“. Damit sind die Techniken gemeint, die beim Transport von Informationen durch unterschiedliche Netze zum Einsatz kommen. Es geht also um die „Verpackung“ und den Transport digitalisierter Informationen. Kompressionsverfahren, Mediensicherheit, Multimedianeetze und Mobile Netze sind hier die Stichworte.

Darauf aufsetzend behandeln wir in einem zweiten Teil des Kompendiums Aspekte der praktischen Informatik und ihre Anwendung in der Medientechnik, wie die Entwicklung von Multimedia-Anwendungen, die Grundlagen der Computergrafik, Theorie und Praxis von Mediendatenbanken. Hinzu kommen einige ausgewählte Anwendungen der Medieninformatik wie die Content-Related-Technologien, mit denen mediale Informationen in möglichst effizienter Weise organisiert, strukturiert und an die richtigen Empfänger verteilt werden können.

Wir begleiten in diesen beiden Teilen also die digitalen Mediendaten auf ihrem Weg von der Erzeugung und Kodierung über ihren Transport durch drahtgebundene oder drahtlose Netze bis hin zum Endnutzer und decken damit eine ganze Wertschöpfungskette ab.

Die Autoren des Kompendiums sind allesamt ausgewiesene Experten ihres Fachs und stehen (oder standen) in aktiver Lehrtätigkeit an der Hochschule der Medien Stuttgart. Die Inhalte des Kompendiums basieren auf Lehrveranstaltungen, die von den Autoren im Rahmen der Studiengänge Medieninformatik (Bachelor/Master) und Information Systems (Bachelor/Master) gehalten werden. Wir danken unseren Studenten für zahlreiche wertvolle Kommentare und Verbesserungsvorschläge zu diesen Lehrveranstaltungen. Ohne diese Anregungen wäre dieses Buch nicht realisierbar gewesen.

Inhaltsverzeichnis

| | | |
|----------|--|----------|
| 1 | Kompressionsverfahren für Video und Audio | 1 |
| 1.1 | Einführung in die Informationstheorie | 2 |
| 1.2 | Grundlagen der Datenkompression | 4 |
| 1.3 | Elementare Verarbeitungsschritte..... | 5 |
| 1.3.1 | Unterabtastung und Quantisierung..... | 6 |
| 1.3.2 | Entropiecodierung..... | 8 |
| 1.3.3 | Prädiktion..... | 11 |
| 1.3.4 | Signaltransformationen..... | 12 |
| 1.4 | Medienspezifische Verarbeitungsschritte | 17 |
| 1.4.1 | Interframe-Kompression | 17 |
| 1.4.2 | Psychoakustische Kompression..... | 20 |
| 1.5 | Codecs im Überblick..... | 23 |
| 1.5.1 | Video-Codecs..... | 24 |
| 1.5.2 | Audio-Codecs | 27 |
| 1.6 | Die Standards JPEG und JPEG 2000 | 29 |
| 1.6.1 | JPEG | 29 |
| 1.6.2 | JPEG 2000 und Motion JPEG 2000 | 35 |
| 1.7 | Videokompression mit MPEG | 38 |
| 1.7.1 | MPEG-1..... | 39 |
| 1.7.2 | MPEG-2..... | 44 |
| 1.7.3 | MPEG-4..... | 50 |
| 1.8 | Audiokompression mit MPEG..... | 60 |
| 1.8.1 | MPEG-1 Audio..... | 61 |
| 1.8.2 | MPEG-4 Audio Lossless Coding..... | 67 |
| 1.9 | Verfahren der Sprachkompression | 70 |
| 1.9.1 | Signalformcodierer | 71 |
| 1.9.2 | Parametrische Codecs..... | 73 |
| 1.9.3 | Hybride Codecs | 74 |
| | Literatur | 78 |



| | | |
|----------|--|------------|
| 2 | Mediensicherheit | 83 |
| 2.1 | Einführung | 84 |
| 2.1.1 | Sicherheitsanforderungen für Digitale Medien | 84 |
| 2.1.2 | Sicherheitsmechanismen | 85 |
| 2.2 | Digitale Wasserzeichen..... | 92 |
| 2.2.1 | Anwendungsgebiete für Digitale Wasserzeichen ... | 93 |
| 2.2.2 | Einbettungstechniken | 97 |
| 2.2.3 | Angriffe | 106 |
| 2.3 | DRM-Systeme | 108 |
| 2.3.1 | Komponenten von DRM-Systemen | 109 |
| 2.3.2 | Mobile DRM-Systeme | 112 |
| 2.3.3 | Angriffe auf DRM-Systeme | 116 |
| 2.3.4 | Kommerzielle DRM-Systeme..... | 117 |
| 2.3.5 | Ökonomische und Rechtliche Aspekte..... | 121 |
| 2.4 | Ausblick | 123 |
| | Literatur..... | 124 |
| | | |
| 3 | Multimedia-Netze..... | 127 |
| 3.1 | Einführung | 127 |
| 3.2 | Multimedia-Transportnetze..... | 129 |
| 3.2.1 | Einleitung..... | 129 |
| 3.2.2 | High-Speed in Metronetzen | 130 |
| 3.2.3 | High-Speed im Weitverkehrsnetz | 133 |
| 3.3 | Quality of Service (QoS) | 143 |
| 3.3.1 | Einführung..... | 143 |
| 3.3.2 | Problemstellungen und Lösungsansätze | 145 |
| 3.3.3 | ATM-QoS | 148 |
| 3.3.4 | IP-QoS in Weitverkehrsnetzen | 151 |
| 3.4 | Next Generation Networks NGN..... | 156 |
| 3.4.1 | Einleitung..... | 156 |
| 3.4.2 | Voice-over-IP | 159 |
| 3.4.3 | IP Next Generation – IPv6..... | 176 |
| 3.4.4 | Multicast..... | 180 |
| 3.4.5 | Fazit | 181 |
| | Literatur..... | 181 |
| | | |
| 4 | Mobile Multimedia | 183 |
| 4.1 | Begriffsbestimmungen..... | 185 |
| 4.2 | Dienste und Architekturen zukünftiger mobiler Multimedia Netze | 186 |
| 4.2.1 | Mobilität und Multimedia – Anforderungen..... | 186 |
| 4.2.2 | Mobile Netze der Gegenwart..... | 190 |
| 4.2.3 | Architektur zukünftiger mobiler All-IP Netze am Beispiel des UMTS Release 5 | 202 |

| | | |
|-------|---|------------|
| 4.3 | Entwicklung Mobiler Multimedia-Anwendungen..... | 208 |
| 4.3.1 | Mobile Multimedia Endgeräte..... | 209 |
| 4.3.2 | Mobile Multimedia Software-Plattformen..... | 213 |
| 4.3.3 | Mobile Multimedia Datenformate | 272 |
| | Literatur | 280 |
| | Autorenverzeichnis | 285 |
| | Index..... | 287 |

Abkürzungen

| | |
|----------|--|
| 3GPP | 3rd Generation Partnership Project |
| 3GPP2 | 3rd Generation Partnership Project 2 |
| AAC | Advanced Audio Coding |
| AAL | ATM Adaption Layer |
| AAZ | Advanced Audio Zip |
| ABR | Available Bit Rate |
| AC-3 | Audio Coding 3 |
| ACELP | Algebraic Code Excited Linear Prediction |
| ADM | Add Drop Multiplexer |
| ADPCM | Adaptive Differential Pulse Code Modulation |
| AIS | Alarm Indication Signal |
| ALS | Audio Lossless Coding |
| AMR | Adaptive Multirate |
| AMR-WB | Adaptive Multirate Wideband |
| APS | Automatic Protection Switching |
| ASF | Advanced Streaming/Systems Format |
| ATM | Asynchronous Transfer Mode |
| AU | Administrative Unit |
| AUG | Administrative Unit Group |
| AVC | Advanced Video Coding |
| AVI | Audio Video Interleave |
| BAB | Binary Alpha Block |
| BBC | British Broadcasting Company |
| B-Frame | Bidirectionally predicted Frame |
| CBR | Constant Bit Rate |
| CDV | Cell Delay Variation |
| CDVT | Cell Delay Variation Tolerance |
| CELP | Code Excited Linear Prediction |
| CER | Cell Error Ratio |
| CIDR | Classless Inter Domain Routing |
| CLP | Cell Loss Priority |
| CLR | Cell Loss Ratio |
| CRC | Cyclic Redundancy Check |
| CS | Call Server |
| CS-ACELP | Conjugate Structure Algebraic Code Excited Linear Prediction |
| CVR | Cell Delay Variation |
| DAB | Digital Audio Broadcast |
| DBR | Deterministic Bit Rate |
| DCC | Digital Compact Cassette |
| DCI | Digital Cinema Initiatives |



| | |
|----------|--|
| DCT | Diskrete Kosinustransformation |
| DECT | Digital Enhanced Cordless Telecommunications |
| DiffServ | Differentiated Services |
| DRM | Digital Radio Mondiale |
| DS | Differentiated Services |
| DSCP | Differentiated Services Code Point |
| DSCP | DiffServ Code Point |
| DVB-S | Digital Video Broadcast Satellite (MPEG-2) |
| DVB-T | Digital Video Broadcast Terrestrial |
| DVD | Digital Versatile Disc |
| DVMRP | Distance Vector Multicast Routing Protocol |
| E-ADPCM | Enhanced Adaptive Differential Pulse Code Modulation |
| ECC | Embedded Communication Channel |
| EDGE | Enhanced Data Rates for GSM Evolution |
| EFR | Enhanced Full Rate |
| ENUM | TElephone NUmber Mapping |
| ES | Elementary Stream |
| FFT | Fast Fourier Transform |
| FIR | Finite Impulse Response |
| FR | Full Rate |
| FWT | Fast Wavelet Transform |
| GFC | Generic Flow Control |
| GFP | Generic Frame Procedure |
| GOP | Group Of Pictures |
| GPRS | General Packet Radio Services |
| GS | Guaranteed Service |
| GSM | Global System for Mobile Communications |
| HD | High Definition |
| HILN | Harmonic and Individual Lines and Noise |
| HR | Half Rate |
| HVXC | Harmonic Vector Excitation Coding |
| IDCT | Inverse DCT |
| IEC | International Electrotechnical Commission |
| IETF | Internet Engineering Task Force |
| I-Frame | Intra-Frame |
| IIR | Infinite Impulse Response |
| IIS | Fraunhofer Institut für integrierte Schaltungen |
| IMS | IP Multimedia Subsystem |
| IMS | IP Multimedia System |
| IntServ | Integrated Services |
| IP | Internet Protocol |
| ISDN | Integrated Services Digital Network |
| ISO | International Organization for Standardization |
| ITU | International Telecommunications Union |
| ITU | International Telecommunication Union |
| JBIG | Joint Bi-Level Image Group |
| JPEG | Joint Photographic Experts Group |
| JVT | Joint Video Team |
| LAN | Local Area Network |
| LCAS | Link Capacity Adjustment Scheme |
| LD-CELP | Low-Delay Code Excited Linear Prediction |
| LDP | Label Distribution Protocol |
| LER | Label Edge Router |

| | |
|------------|---|
| LPAC | Lossless Predictive Audio Coding |
| LPC | Linear Prediction Coefficients |
| LSR | Label Switch Router |
| MA | Monkey's Audio |
| MAC | Media Access Control |
| MAN | Metropolitan Area Network |
| MCU | Multipoint Control Unit |
| MDCT | Modified DCT |
| MELPe | Enhanced Mixed Excitation Linear Prediction |
| MGC | Media Gateway Controller |
| MJPEG 2000 | Motion JPEG 2000 |
| MJPEG | Motion JPEG |
| MMD | Multimedia Domain |
| MMS | Multimedia Messaging Service |
| MOS | Mean Opinion Score |
| MOS | Mean Opinion Score |
| MP3 | MPEG Audio Layer III |
| MPEG | Moving Picture Experts Group |
| MPEG-LA | MPEG Licensing Authority |
| MPLS | Multiprotocol Label Switching |
| MP-MLQ | Multipulse Maximum Likelihood Quantization |
| MSOH | Multiplexer Section Overhead |
| MTU | Maximum Transmission Unit |
| NAL | Network Abstraction Layer |
| NAPT | Naming Authority Pointer Records |
| NAPTR | Naming Authority Pointer Records |
| NGN | Next Generation Network |
| OAM | Operation, Administration and Management |
| OSPF | Open Shortest Path First |
| PCM | Pulse Code Modulation |
| PCR | Peak Cell Rate |
| PCR | Peak Cell Rate |
| PDH | Plesiochronous Digital Hierarchy |
| P-Frame | Predicted Frame |
| PHB | Per-hop-behaviour |
| PHB | Per Hop Behaviour |
| PIM | Protocol Independent Multicast |
| PNS | Perceptual Noise Substitution |
| POH | Path Overhead |
| PVC | Permanent Virtual Channel |
| QoS | Quality of Service |
| RFC | Request for Comment |
| RPE-LTP | Regular Pulse Excitation with Long-Term Prediction |
| RPR | Resilient Packet Ring |
| RSOH | Regenerator Section Overhead |
| RSVP | Resource Reservation Protocol |
| RSVP | Resource Reservation Protocol |
| RTCP | Real Time Transmission Control Protocol |
| RTP | Real-Time Transport Protocol |
| RTP | Real-Time-Protocol |
| RTSP | <i>Real Time Streaming Protocol</i> |
| SAR | Segmentation and Reassembly |
| SB-ADPCM | Subband Adaptive Differential Pulse Code Modulation |



| | |
|--------|---------------------------------------|
| SDDS | Sony Dynamic Digital Sound |
| SDH | Synchronous Digital Hierarchy |
| SDP | Session Description Protocol |
| SIP | Session Initiation Protocol |
| SIP | Session Initiation Protocol |
| SMPTE | Society of Motion Picture Engineers |
| SOH | Section Overhead |
| SONET | Synchronous Optical Network |
| STANAG | Standardization Agreement |
| STM | Synchronous Transfer Module |
| STS | Synchronous Transport Signal |
| SVC | Switched Virtual Channel |
| TCP | T ransmission Control Protocol |
| TMN | Telecommunications Management Network |
| TOS | Type of Service |
| TU | Tributary Unit Group |
| TUG | Tributary Unit Group |
| UAC | User Agent Client |
| UAS | User Agent Server |
| UBR | Unspecified Bit Rate |
| UBR | Unspecified Bit Rate |
| UDP | User Datagram Protocol |
| VBR | Variable Bit Rate |
| VBR | Variable Bit Rate |
| VC | Virtual Container |
| VC-1 | Video Coding 1 |
| VCAT | Virtual Concatenation |
| VCL | Video Coding Layer |
| VLAN | Virtual LAN |
| VMR | Variable Multirate |
| VMR-WB | Variable Multirate Wideband |
| VO | Video Object |
| VOL | Video Object Layer |
| VOP | Video Object Plane |
| VPN | Virtual Private Network |
| VQEG | Video Quality Experts Group |
| VRML | Virtual Reality Modeling Language |
| VS | Visual Sequence |
| VSELP | Vector Sum Excited Linear Prediction |
| WAN | Wide Area Network |
| WM | Windows Media |
| WMA | schnelle Wavelet-Transformation Audio |
| WMV | Windows Media Video |

1 Kompressionsverfahren für Video und Audio

Jan Schulz

Kontinuierliche Medien wie Audio- und Videosequenzen stellen in Bezug auf erforderliche Datenrate und benötigten Speicherplatz hohe Anforderungen an die verarbeitenden Systeme. Beispielsweise besitzt eine unkomprimierte Bildsequenz mit der Auflösung von 640×480 Pixeln bei einer Bildrate von 25 Bildern pro Sekunde und einer Farbtiefe von 8 Bit pro RGB-Kanal eine Datenrate von:

$$640 \cdot 480 \cdot 25 \text{ Hz} \cdot 8 \text{ Bit/Kanal} \cdot 3 \text{ Kanäle} = 184.320.000 \text{ Bit/s} \approx 184 \text{ MBit/s.}$$

10 Minuten einer solchen Bildsequenz benötigen Speicherplatz im zweistelligen Gigabyte-Bereich. Um also Audio- und Videosequenzen übertragen, verarbeiten und speichern zu können, ist der Einsatz effizienter Kompressionsverfahren fast immer unverzichtbar.

Dies gilt für den klassischen Fernsbereich ebenso wie z. B. für die Wachstumsbranche der Mobilien Multimediadienste (vgl. Kap. 4). Unter der Sammelbezeichnung High Definition Television (HDTV) starten derzeit immer mehr Fernsehprogramme in höherer Auflösung, während Auflösung in der Mobilfunkbranche aufgrund der Größe der Endgeräte nur ein untergeordnetes Kriterium darstellt. Hier steht – mehr noch als bei HDTV – die Datenrate im Blickpunkt, die bei der ersten UMTS-Version z. B. bis zu 384 kBit/s beträgt. HDTV-Programme benötigen dagegen mit etwa 15 MBit/s (DVB-S) weitaus höhere Datenraten. Diese ausgeprägten Unterschiede beginnen mit der Annäherung von Kommunikations- und Unterhaltungsbranche auf dem Konsumentenmarkt zu schwinden, was sich auch bei neuen Verfahren der Datenkompression widerspiegelt. Während frühere Kompressionsstandards wie z. B. MPEG-1 stark auf einen bestimmten Einsatzzweck zugeschnitten wurden, sind neuere Standards wie bspw. H.264/AVC viel flexibler einsetzbar und werden in beiden oben genannten Branchen verwendet.

Trotz aller Annäherung bei der Auslieferung von Medieninhalten hat die Film- und Fernsehbranche jedoch nach wie vor Bedarf an

Motivation für Datenkompression

Annäherung von Fernseh- und Kommunikationsbranche auf dem Konsumentenmarkt

Bleibender Bedarf an hoher Qualität

Verfahren, mit denen Produktionen möglichst ohne Informationsverluste komprimiert werden können. Dies ist z. B. an den Fidelity Range Extensions ersichtlich, welche H.264/AVC u. A. um eine Möglichkeit ergänzen, die volle Farbinformation des Videomaterials zu erhalten.

Inhalte des Kapitels

Im Folgenden werden zunächst einige Konzepte der Informationstheorie und Begriffe aus dem Bereich der Datenkompression erläutert, bevor in Kap. 1.3 grundlegende Verarbeitungsschritte dargestellt werden, die sich in vielen Kompressionsverfahren wieder finden. Mit Interframe- und psychoakustischer Kompression werden in Kap. 1.4 zwei Verfahren für die Video- bzw. Audiokompression beleuchtet, welche medienspezifische Besonderheiten ausnutzen und in Kombination mit anderen Verarbeitungsschritten die Kompressionsrate stark erhöhen können.

Kapitel 1.5 gibt einen Überblick über zahlreiche Kompressionsverfahren für Audio- und Videodaten, wobei insbesondere Standards der ITU und der MPEG-Gruppe betrachtet werden.

Nachdem in Kap. 1.6 die JPEG-Kompression als Ausgangspunkt zahlreicher Verfahren zur Videokompression erläutert wurde, behandelt Kap. 1.7 ausführlich die Videokompression nach MPEG. Anhand von MPEG-1 wird hierbei die grundlegende Funktionsweise der MPEG-Standards dargestellt, so dass in den weiteren Unterkapiteln zu MPEG-2 und MPEG-4 darauf aufgebaut werden kann.

Kapitel 1.8 widmet sich der Audiokompression und stellt mit MPEG-1 und MPEG-4 ALS Verfahren vor, denen zwei unterschiedliche Funktionsweisen zu Grunde liegen. Während MPEG-1 Audio ein psychoakustisches Modell verwendet, basiert MPEG-4 ALS im Wesentlichen auf Prädiktion. Als Spezialfall der Audiokompression wird in Kap. 1.9 abschließend die Sprachkompression erläutert. Weitere Informationen hierzu finden sich auch in Kap. 3.

1.1 Einführung in die Informationstheorie

Informationstheorie

Die Informationstheorie wurde durch Shannon begründet, der die Übertragung von Zeichen zwischen Nachrichtenquelle und Nachrichtensenke betrachtete und daraus mathematische Methoden entwickelte, um Information quantitativ zu bestimmen [Shn48]. Erkenntnisse der Informationstheorie werden u. A. bei Entropiecodierungen (Kap. 1.3.2) eingesetzt, die Nachrichten möglichst kompakt darstellen und so Datenkompression erzielen.

gedächtnislose Nachrichtenquelle

Im Folgenden werden diskrete gedächtnislose Nachrichtenquellen mit endlichem Alphabet betrachtet, die in der Informationstheorie einen Spezialfall darstellen. Eine solche Quelle versendet zu diskreten

Zeitpunkten $t = 0, 1, \dots$ die Symbole eines endlichen Alphabets A . Die Wahrscheinlichkeit für das Auftreten eines Zeichens $z \in A$ zum Zeitpunkt t hängt dabei nicht vom Zeitpunkt t und nicht von den Zeichen ab, die zu den vergangenen Zeitpunkten $0, \dots, t-1$ gesendet wurden. Gedächtnislose Quellen sind in der Realität selten zu finden, da hier häufig Abhängigkeiten zwischen einzelnen Quellsymbolen bestehen. Sie genügen jedoch, um grundlegende Konzepte und Begriffe der Informationstheorie zu erläutern, wie sie im Weiteren benötigt werden. Für eine Verallgemeinerung der Konzepte sei z. B. auf [Gra98] verwiesen.

Der Informationsgehalt $I(z)$ eines Zeichens $z \in A$, das von einer diskreten gedächtnislosen Nachrichtenquelle mit endlichem Alphabet erzeugt wird, berechnet sich aus der Wahrscheinlichkeit, mit der die Nachrichtenquelle das Zeichen z produziert:

Informationsgehalt

$$I(z) = \log_2 \frac{1}{W(z)} [\text{bit}]$$

Mit abnehmender Auftrittswahrscheinlichkeit eines Zeichens steigt folglich sein Informationsgehalt. Tritt mit $W(z) = 1$ stets dasselbe Zeichen auf, so ist sein Informationsgehalt gleich 0.

Betrachtet man eine Symbolfolge z_0, z_1, \dots, z_n , so lässt sich der mittlere Informationsgehalt dieser Folge berechnen, der auch als Entropie erster Ordnung bezeichnet wird. Unter der Voraussetzung, dass die Elemente der Zeichenfolge statistisch nicht voneinander abhängen, ergibt sich diese Entropie H aus der Summe der gewichteten Informationsgehalte der Einzelelemente:

Entropie

$$H = \sum_{i=0}^n W(z_i) \cdot I(z_i) [\text{bit/Symbol}]$$

Den höchsten Wert hat die Entropie demnach bei einer Gleichverteilung der Symbole. Je stärker sich die Auftrittswahrscheinlichkeiten um einen bestimmten Wert konzentrieren, desto geringer ist die Entropie.

Ist die gesamte Symbolfolge z_0, z_1, \dots, z_n schon im Voraus bekannt, so entspricht die Auftrittswahrscheinlichkeit eines Zeichens der relativen Häufigkeit des Zeichens in der Symbolfolge. Bei der Audio- und Videokompression ist dies jedoch selten der Fall, und die Wahrscheinlichkeiten müssen geschätzt werden.

Wenn man die Elemente einer Zeichenfolge auf binäre Codewörter abbildet, wird die mittlere Codelänge interessant. Ist l_i die Länge des Codeworts, mit dem das Zeichen z_i codiert wird, errechnet sich die mittlere Codelänge \bar{l}_i aus der Summe gewichteter Codelängeln:

mittlere Codelänge

$$\bar{l}_i = \sum_{i=0}^n W(z_i) \cdot l_i [\text{Bit/Symbol}]$$

Codierungsredundanz

Die Differenz zwischen Entropie und mittlerer Codelänge wird Codierungsredundanz genannt. Shannon zeigte, dass immer ein Code existiert, dessen mittlere Codelänge kleiner als $H+1$ ist. Hohe Redundanz ist bei der Datenkompression kontraproduktiv. Kompressionsverfahren zielen daher unter anderem auf die Minimierung der Codierungsredundanz ab.

Bei Nachrichten bzw. Symbolfolgen, die im Rahmen der Audiokompression betrachtet werden, handelt es sich um diskrete Samples eines digitalen Tonsignals. Bei Bilddaten sind es die Helligkeits- oder Farbwerte von Pixeln, die bei sequenzieller Betrachtung der Zeilen ebenfalls diskrete Symbolfolgen darstellen. Daher werden die Begriffe Zeichenfolge, Symbolfolge, Signal und Nachricht im Folgenden synonym verwendet.

1.2 Grundlagen der Datenkompression

Quellencodierung

Datenkompression ist eine Form der Codierung, welche die Symbolfolgen einer Nachrichtenquelle möglichst platzsparend abspeichert. Codierung zum Zwecke der Datenkompression wird häufig auch als Quellencodierung bezeichnet, um sie von Kanalcodierungen bei der Datenübertragung und kryptographischen Codierungen im Bereich der Datensicherheit abzugrenzen. Im Folgenden wird diese Unterscheidung nicht mehr getroffen, und der Begriff Codierung ist immer als Quellencodierung zu verstehen.

*verlustfreie und verlustbehaftete
Kompression*

Mit verlustfreien und verlustbehafteten Verfahren lassen sich zwei grundlegende Arten der Datenkompression bzw. Quellencodierung unterscheiden. Die verlustfreie Kompression verwirft ausschließlich redundante Informationen und ist daher voll reversibel. Durch Redundanzreduktion verlustfrei komprimierte Informationen können fehlerfrei wiederhergestellt werden. Die verlustbehaftete Kompression nimmt dagegen den Verlust von irrelevanten Informationen in Kauf. Während sich Redundanz eindeutig definieren lässt, ist Irrelevanz stark subjektiv geprägt. Irrelevant sind jene Informationen, die den Empfänger nicht interessieren oder nicht von ihm wahrgenommen werden können. Die Irrelevanzreduktion ist irreversibel. Einmal verworfene Informationen lassen sich nicht wieder zurückgewinnen.

*symmetrische und asymmetrische
Kompression*

Ein weiteres Kriterium für die Klassifizierung von Kompressionsverfahren erhält man, indem der zeitliche bzw. rechnerische Aufwand für die Kompression A_K mit dem Aufwand für die Dekompression A_D in Beziehung gesetzt wird. Ist der Aufwand identisch, spricht man von symmetrischer Kompression. Dagegen handelt es sich um asymmetrische Kompression, wenn $A_K > A_D$ ist. Bei der Kompression von Audio-

und Videodaten ist häufig ein geringer Aufwand bei der Decodierung erwünscht, da diese für eine flüssige Wiedergabe des Mediums in Echtzeit erfolgen muss. Verfahren für die Audio- und Videokompression sind daher zumeist asymmetrischer Natur.

Ein Algorithmus oder Schaltkreis, der eine Codierung oder Datenkompression vornimmt, wird als Coder oder Encoder bezeichnet. Umgekehrt macht ein Decoder die Kompression wieder rückgängig. Spricht man von einem Kompressionsverfahren als Ganzes, so verwendet man den zusammengesetzten Begriff Codec.

Coder/Encoder, Decoder, Codec

Als Maß für die quantitative Bewertung eines Kompressionsverfahrens kann die Kompressionsrate herangezogen werden, bei der die ursprüngliche Datenmenge zur komprimierten Datenmenge ins Verhältnis gesetzt wird. Durch Irrelevanzreduktion lassen sich weitaus höhere Kompressionsraten erzielen als mit Verfahren, die ausschließlich auf Redundanzreduktion beruhen.

Kompressionsrate

Die Kompressionsrate als rein objektives Kriterium ist bei der verlustbehafteten Irrelevanzreduktion nur bedingt aussagekräftig. Daher unterscheidet man bei der Bewertung von verlustbehafteten Kompressionsverfahren häufig zwischen objektiver und subjektiver Qualität. Als einfaches Kriterium für die Beurteilung der subjektiven Qualität hat sich der Mean Opinion Score (MOS) etabliert, der aus dem Bereich der telefonischen Sprachübertragung stammt. Für dieses Einsatzgebiet existieren Empfehlungen der International Telecommunication Union (ITU), einem internationalen Zusammenschluss zur Koordination von Entwicklungen im Telekommunikations- und Fernsehbereich. Der MOS wird durch Versuchspersonen ermittelt, die anhand einer fünfstufigen Skala qualitative Bewertungen zu unterschiedlich komprimierten Signalen abgeben. Im einfachsten Fall wird anschließend der arithmetische Mittelwert dieser Bewertungen gebildet [ITU96d] [ITU03b].

*objektive und subjektive Qualität,
Mean Opinion Score (MOS)*

Die Ermittlung der subjektiv empfundenen Qualität durch Versuchspersonen ist ein aufwändiges Verfahren. Daher wurde im Videobereich die Video Quality Experts Group (VQEG) ins Leben gerufen, die sich mit der Standardisierung objektiver Messmethoden zur Bewertung subjektiver Qualität bei Videokompressionsverfahren befasst.

Video Quality Experts Group (VQEG)

1.3 Elementare Verarbeitungsschritte

Bei der Audio- und Videokompression verwendet man Verfahren, die sequentiell verschiedene grundlegende Verarbeitungsschritte durchführen, um besonders hohe Kompressionsraten zu erzielen. Meist entstehen dadurch hybride Methoden, die Irrelevanz- und Redundanzreduktion kombinieren.

hybride Kompressionsverfahren

1.3.1 Unterabtastung und Quantisierung

Unterabtastung Durch Unterabtastung lässt sich die Datenrate eines Signals auf einfache Weise reduzieren. Bei digitalen Signalen wird dabei eine bestimmte Anzahl der diskreten Signalwerte (engl.: Samples) verworfen, und nur jeder n -te Wert mit $n = 2, 3, \dots$ wird beibehalten. Dies resultiert in einer Reduktion der Datenrate um den Faktor $1/n$. Die verworfenen Samples lassen sich nicht fehlerfrei wiederherstellen, weshalb es sich hierbei um einen verlustbehafteten Vorgang handelt.

Abtasttheorem, Aliasfrequenz Durch die Unterabtastung können bei der Rekonstruktion des Signals Aliasfrequenzen entstehen, die sich in Videosignalen beispielsweise durch störende Muster äußern. Diese Aliasfrequenzen treten nach dem fundamentalen nachrichtentheoretischen Abtasttheorem [Kot33][Shn48] immer dann auf, wenn die Frequenz der diskreten Signalwerte f_a durch die Unterabtastung kleiner wird als die doppelte Bandbreite des Signals. Um die Entstehung von Aliasfrequenzen zu verhindern, kann das Signal vor der Unterabtastung einer Bandpassfilterung mit der Grenzfrequenz $f_g = f_a/2n$ [Hz] unterzogen werden.

Chrominanz und Luminanz, YCrCb-Farbraum Bei digitalem Video findet man häufig eine Unterabtastung der Farbinformation (Chrominanz), da diese vom menschlichen Auge schlechter aufgelöst werden kann als Helligkeitsinformation (Luminanz). Um Luminanz und Chrominanz zu trennen, werden die Videobilder in den YCrCb-Farbraum konvertiert. Dieser ist in der ITU-Empfehlung BT.601 [ITU95a] für digitale Fernsehstudioteknik definiert und wird aus den RGB-Kanälen durch folgende Produktmatrix gewonnen:

$$\begin{pmatrix} Y \\ Cr \\ Cb \end{pmatrix} = \begin{pmatrix} 0,299 & 0,587 & 0,114 \\ 0,500 & -0,419 & -0,081 \\ -0,169 & -0,331 & 0,500 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

4:2:2-Subsampling In dieser weitgehend historisch bedingten Formel wird die Luminanz von der Y-Komponente, und die Chrominanz von den Komponenten Cr und Cb repräsentiert. Bei einer Unterabtastung der Chrominanz nach BT.601 werden für die Komponenten Cr und Cb jeweils nur halb so viele Pixelwerte aufgewendet wie für die Luminanz, was als 4:2:2-Subsampling bekannt ist. Noch geringer ist die Abtastung der CrCb-Komponenten bei der Kompression nach dem MPEG-1-Standard (Kap. 1.7.1), wo 4:2:0-Subsampling zum Einsatz kommt (Abb. 1.1).

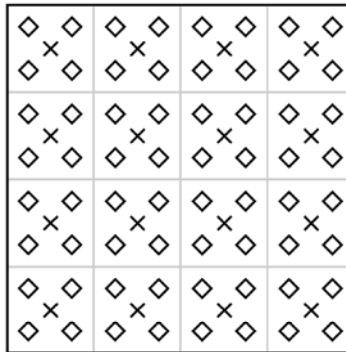
4:2:2-Subsampling nach ITU-R BT.601



◇ Luminanz-Wert

⊗ Luminanz-Wert und Chrominanz-Wert

4:2:0-Subsampling nach MPEG-1



× Chrominanz-Wert

Abb. 1.1:

4:2:2- und 4:2:0-Subsampling

Quantisierung ist eine weitere, sehr einfache Methode der Datenreduktion. Anstatt einzelne Samples zu verwerfen, wird hierbei die Genauigkeit reduziert, mit der sie abgespeichert werden. Man unterscheidet in skalare Quantisierung und Vektorquantisierung.

Quantisierung

Die skalare Quantisierung ordnet jedem Signalwert einen quantisierten Wert aus einer endlichen Wertemenge zu. Die Zuordnung erfolgt dabei im einfachsten Fall linear auf Basis eines Rasters mit Intervallen fester Länge. Alle Samples innerhalb eines bestimmten Intervalls werden dabei auf denselben quantisierten Wert abgebildet, wodurch verlustbehaftete Datenkompression entsteht. Anstelle eines festen Rasters können auch unterschiedliche Intervallbreiten gewählt werden, um bestimmte Werte stärker zu quantisieren als andere. Dies kann beispielsweise Sinn machen, wenn sich dadurch Einschränkungen der menschlichen Wahrnehmung ausnutzen lassen und wird als nichtlineare Quantisierung bezeichnet. Die Intervallbreiten bei nichtlinearer Quantisierung werden in Form von sog. Quantisierungskennlinien festgelegt.

skalare Quantisierung, nichtlineare Quantisierung

Die Vektorquantisierung berücksichtigt n Signalwerte gleichzeitig, die als Vektor des n -dimensionalen Raums aufgefasst werden. Wie die skalare Quantisierung stellt auch die Vektorquantisierung einen verlustbehafteten Vorgang dar. Ein Vektorquantisierer der Dimension n und Größe s bildet Eingabevektoren auf eine endliche Menge C ab, die aus s Ausgabevektoren besteht. Für die Wahl der Ausgabevektoren aus C können verschiedene Kriterien herangezogen werden. Im einfachsten Fall kommt das euklidische Abstandsmaß der Vektoren zum Einsatz. Die Menge C der Ausgabevektoren wird als Codebuch bezeichnet. Die größte Herausforderung bei der Vektorquantisierung ist die Wahl eines geeigneten Codebuchs. Dieses muss in einer Trainingsphase mit Hilfe

Vektorquantisierung, Codebuch

charakteristischer Signalvektoren optimiert und so an typische Signalstatistiken angepasst werden. Ein verbreiteter Algorithmus zur Codebuch-Erstellung ist der LBG-Algorithmus [LBG80].

Vektorquantisierung in der Videokompression teilt Einzelbilder in quadratische Blöcke auf, wobei gleichartige Blöcke durch einen generischen Block aus einem Codebuch ersetzt werden. Der Einsatz von Vektorquantisierung in Codecs wie Sorenson oder Cinepak trug in den neunziger Jahren zur Verbreitung von digitalem Video auf Heimcomputern bei [WNW94]. Ein Audio-Codec auf Basis von Vektorquantisierung ist z. B. SoundVQ, der unter der Bezeichnung TwinVQ entwickelt [IMM95] und inzwischen als Teil von MPEG-4 Audio standardisiert wurde (Kap. 1.8).

1.3.2 Entropiecodierung

Entropiecodierungen nutzen die statistische Verteilung von Symbolen innerhalb einer Zeichenkette aus. So lässt sich die mittlere Codelänge der Entropie annähern und damit die Codierungsredundanz mindern. Da Entropiecodierungen ausschließlich redundante Informationen verwerfen, arbeiten sie verlustfrei.

1.3.2.1 Huffman-Codierung

Die Huffman-Codierung [Huf52] geht davon aus, dass die einzelnen Symbole einer Nachricht statistisch unabhängig voneinander sind, und codiert diese separat. Symbole, die eine hohe Auftrittswahrscheinlichkeit besitzen, werden mit kürzeren Codes abgespeichert als Werte mit geringer Auftrittswahrscheinlichkeit.

grafische Ermittlung

Um eine Huffman-Codierung auf grafische Weise zu ermitteln, werden die Quellsymbole als Codebäume mit einem Knoten betrachtet. In jedem Knoten wird die Auftrittswahrscheinlichkeit des jeweiligen Symbols eingetragen. Die verschiedenen Knoten werden dann wie folgt zu einem einzelnen Baum zusammengefügt:

1. Die zwei Knoten mit der geringsten Auftrittswahrscheinlichkeit werden zu einem neuen Knoten zusammengefasst.
2. Dem neu entstandenen Knoten wird die Summe der beiden Auftrittswahrscheinlichkeiten zugewiesen.
3. Die neu entstandenen Zweige erhalten als Beschriftung 0 bzw. 1.
4. Man beginnt von vorne, wenn die Wurzel des Baumes mit der Auftrittswahrscheinlichkeit 1 noch nicht erreicht ist.

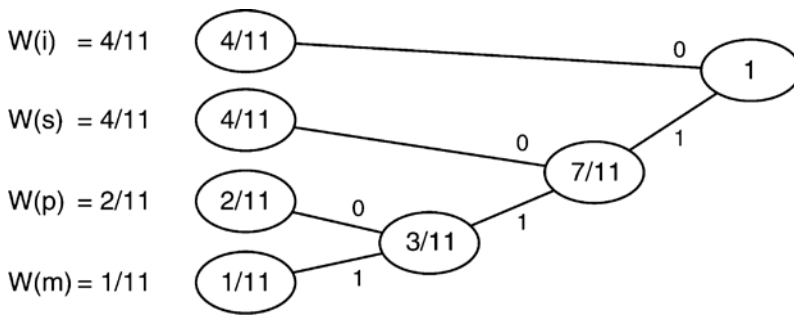


Abb. 1.2:
Grafische Entwicklung einer
Huffman-Codierung

Ausgehend von den Blattknoten kann am Ende das Codewort für jedes Symbol abgelesen werden. Die grafische Entwicklung einer Huffman-Codierung für die Zeichenkette *mississippi* ist in Abb. 1.2 dargestellt. Die Huffman-Codierung besitzt eine Entropie von $1,82 \text{ bit/Symbol}$ und eine mittlere Codelänge von $1,91 \text{ Bit/Symbol}$. Die Codierungsredundanz ist demnach nahe null. Dabei wurde jedoch nicht berücksichtigt, dass der Codebaum für die Decodierung bekannt sein muss und daher zusätzlich abgespeichert wird.

Da die Huffman-Codierung einzelne Symbole codiert, und diese immer durch eine ganzzahlige Folge von Bits dargestellt werden müssen, ist die Codierung nicht in allen Fällen optimal. Eine Codierungsredundanz von 0 Bits wird nur dann erreicht, wenn der Informationsgehalt eines Symbols und die zugeordnete Codelänge übereinstimmen. Dies ist der Fall, wenn die Auftretswahrscheinlichkeiten der Bedingung $W(z_i) = 2^{-n}$ mit $n = 1, 2, 3, \dots$ genügen. Ein Vorteil der Huffman-Codierung ist, dass sie einfach implementiert werden kann, und es sich nur um ein geringfügig asymmetrisches Verfahren handelt.

Eigenschaften

1.3.2.2

Arithmetische Codierung

Die arithmetische Codierung rückt von der separaten Symbolcodierung ab und erzeugt einen einzigen Code für die gesamte Nachricht. So wird Redundanz auch dort effizient vermindert, wo sich eine Huffman-Codierung als ungünstig erweisen würde. Trotzdem lässt sich die arithmetische Codierung sequentiell durch Abarbeiten einzelner Symbole ermitteln. Wie die Huffman-Codierung geht die arithmetische Codierung davon aus, dass die Nachrichtensymbole statistisch unabhängig voneinander sind.

Das Funktionsprinzip der arithmetischen Codierung ist es, die gesamte Zeichenfolge z_0, z_1, \dots, z_n durch eine rationale Zahl aus dem Einheitsintervall $[0,1)$ zu codieren. Dies geschieht wie folgt:

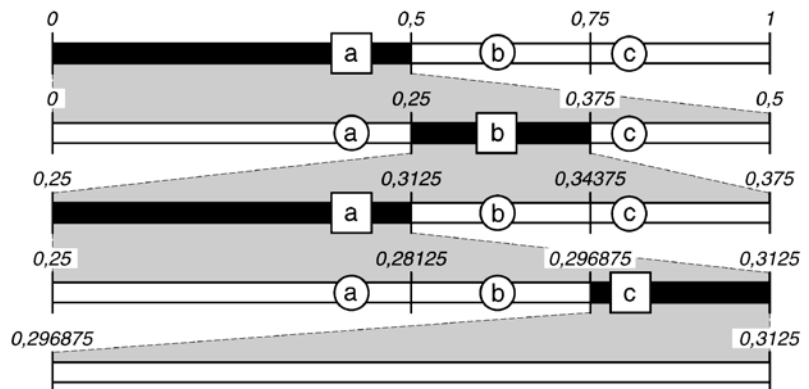
Funktionsprinzip

1. Das Einheitsintervall $[0,1)$ wird ausgewählt.
2. Das aktuell gewählte Intervall wird so in Teilintervalle eingeteilt, dass die Größen der Teilintervalle mit den Auftretswahrscheinlichkeiten $W(z_i)$ der Symbole z_0, z_1, \dots, z_n korrespondieren.
3. Das Teilintervall, das dem aktuellen zu codierenden Symbol z_i entspricht, wird ausgewählt.
4. Ist z_i das letzte Symbol der Zeichenfolge, so kann diese durch eine beliebige Zahl aus dem Teilintervall codiert werden. Ansonsten setzt man bei Schritt 2 fort.

Beispiel

Das beschriebene Verfahren wird in Abb. 1.3 am Beispiel der Zeichenfolge *abac* mit den Auftretswahrscheinlichkeiten $W(a) = 0,5$ sowie $W(b) = W(c) = 0,25$ dargestellt. Die Zeichenkette kann in diesem Fall durch eine Zahl aus $[0,296875, 0,31275)$ codiert werden. Die Decodierung wendet das eben beschriebene Verfahren rückwärts an und muss daher nicht explizit erläutert werden.

Abb. 1.3:
Arithmetische Codierung der
Zeichenkette *abac*



Entstehung

Die theoretische Grundlage der arithmetischen Codierung wurde durch [Shn48] gelegt. Trotz des relativ einfachen Prinzips scheiterte eine algorithmische Umsetzung lange Zeit an der endlichen Genauigkeit von Fließkommazahlen auf Computern. Mit zunehmender Nachrichtenlänge werden die Teilintervalle sehr klein und können mit endlicher Fließkomma-Genauigkeit u. U. nicht mehr eindeutig unterschieden werden. Schließlich gelang der Nachweis, dass bestimmte endlich lange Zahlendarstellungen noch die benötigte Unterscheidbarkeit bieten. Einer der ersten Algorithmen für die technische Umsetzung der arithmetischen Codierung wurde von IBM veröffentlicht [RiL79].

1.3.3 Prädiktion

Die Prädiktion zieht vorhergegangene und damit bekannte Symbole einer Nachricht heran, um Voraussagen über unbekannte Folgesymbole treffen zu können. Dies ist sinnvoll, da zwischen benachbarten Symbolen einer Nachricht häufig Abhängigkeiten bestehen. So ist in der deutschen Sprache beispielsweise eine hohe Korrelation zwischen den Buchstaben *Q* und *U* zu finden. Auch die Pixel eines Bildes können korrelieren, wenn sie demselben abgebildeten Objekt zugeordnet sind. Die Prädiktion stammt ursprünglich aus dem Bereich der Sprachverarbeitung, wo sie zur Kompression und Spracherkennung verwendet wird [AtS67].

Die Abweichung zwischen vorausgesagtem Prädiktionswert und tatsächlich auftretendem Folgesymbol wird als Prädiktionsfehler bzw. Residuum bezeichnet. Seien $\hat{s}(n)$ ein Prädiktionswert und $s(n)$ das tatsächlich auftretende Folgesymbol, so ist der Prädiktionsfehler $e(n) = s(n) - \hat{s}(n)$. Mit Hilfe von Prädiktionsfehlern lassen sich Nachrichten eindeutig decodieren, indem der Decoder ebenfalls Prädiktionswerte errechnet und diese durch Addition der Prädiktionsfehler korrigiert.

Prädiktionsfehler bzw. Residuum

Durch die Symbolkorrelation bedingt, entsteht bei der Prädiktion eine starke Ungleichverteilung der Prädiktionsfehler, was wiederum eine geringere Entropie bedeutet. Daher ist die Prädiktion ein Verarbeitungsschritt, welcher die Bedingungen für eine nachfolgende Entropiecodierung verbessert.

Die Prädiktion von Folgesymbolen erfolgt im einfachsten Fall als lineare Prädiktion durch das unmittelbar vorhergehende Symbol. Dieses kann entweder direkt verwendet oder vorher einer Gewichtung unterzogen werden. Der Gewichtungsfaktor wird als Prädiktorkoeffizient a_k bezeichnet und muss meist fortlaufend an das Signal angepasst werden. Wenn mehr als ein Vorgängersymbol zur Prädiktion herangezogen wird, so handelt es sich um eine Prädiktion höherer Ordnung. Dies kann z. B. bei der Prädiktion von Bildinhalten sinnvoll sein, da sich ein Bildpixel genauer vorhersagen lässt, wenn mehrere umliegende Pixel, das sog. Template, berücksichtigt werden.

lineare Prädiktion

Die lineare Prädiktion p -ter Ordnung erfolgt mit p Prädiktorkoeffizienten a_k , die auch als Linear Prediction Coefficients (LPC) bezeichnet werden. Der Prädiktionswert $\hat{s}(n)$ errechnet sich dann aus p vorangegangenen Symbolen als FIR-Filter nach:

Linear Prediction Coefficients (LPC)

$$\hat{s}(n) = \sum_{k=1}^p a_k \cdot s(n-k)$$

Es muss folglich ein FIR-Filter mit den Filterkoeffizienten a_k entwickelt werden, das die zukünftigen Symbole möglichst gut auf Basis vergangener Symbole vorhersagen kann. Dieses Filter wird Analyse- oder Prädiktionsfilter genannt, während das korrespondierende Gegenstück des Decoders als Synthese- oder LPC-Filter bezeichnet wird. Gibt man das Prädiktionsfehler-Signal, das vom Analysefilter errechnet wurde, unverändert in das Synthesefilter hinein, so kann das ursprüngliche Signal ohne Fehler rekonstruiert werden.

Eine etwas andere, wenn auch verwandte Aufgabe stellt sich mit der AR-Modellierung auf dem Gebiet der parametrischen Sprachkompression (Kap. 1.9.2). Dort wird ein Synthese-Filter gesucht, das ein stationäres Anregungssignal, wie z. B. ein weißes Rauschen, möglichst gut an das zu codierende Sprachsignal annähert. Zwischen linearer Prädiktion und AR-Modellierung besteht ein sehr enger Zusammenhang, der in [KaK02] ausführlich erläutert wird.

Prinzipiell arbeitet die Prädiktion verlustfrei, sofern die errechneten Prädiktionsfehler nicht quantisiert werden. Quantisierung bietet sich an, wenn die Prädiktionsfehler rational sind, was eine weitere Verarbeitung erschweren würde.

In vielen Verfahren zur Bildkompression wird die Prädiktion nicht direkt auf Bildpunkte angewendet. Statt dessen wird das Bild zunächst in eine Repräsentation überführt, die sich besser für den Einsatz einer prädiktiven Codierung eignet. Dies geschieht bspw. durch Signaltransformationen, die im Folgenden erläutert werden.

1.3.4 Signaltransformationen

Wie die Prädiktion zielen auch Signaltransformationen darauf ab, die Korrelation zwischen Nachrichtensymbolen zu verringern, und so die Effizienz nachfolgender Verarbeitungsschritte zu steigern. Um eine anschaulichere Darstellung zu ermöglichen, liegen den folgenden Erläuterungen Bilddaten zugrunde. Signaltransformationen sind jedoch nicht auf solche Daten beschränkt.

1.3.4.1 Diskrete Kosinustransformation

Die diskrete Kosinustransformation (DCT) als Sonderfall der Fouriertransformation existiert in verschiedenen Ausprägungen, die von [Wan84] erstmals in vier Transformationstypen eingeteilt wurden. Bei der Bilddatenkompression wird die DCT-II nach [ANR74] verwendet. Die DCT-II und die zugehörige inverse Transformation werden im Folgenden nur noch als DCT bzw. IDCT bezeichnet. Die DCT zur

Transformation zweidimensionaler Bilder wird auf quadratische Teilbereiche des Bildes angewendet und arbeitet auf Basis einzelner Bildpixel. Bei einem quadratischen Bildausschnitt von $N \times N$ Pixeln wird jedes Pixel $f(x,y)$ wie folgt transformiert:

$$F(u,v) = \frac{2}{N} \cdot C_u C_v \cdot \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cdot \cos\left[\frac{(2x+1) \cdot u \cdot \pi}{2N}\right] \cdot \cos\left[\frac{(2y+1) \cdot v \cdot \pi}{2N}\right]$$

Die IDCT ist:

$$f(x,y) = \frac{2}{N} \cdot \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C_u C_v \cdot F(u,v) \cdot \cos\left[\frac{(2x+1) \cdot u \cdot \pi}{2N}\right] \cdot \cos\left[\frac{(2y+1) \cdot v \cdot \pi}{2N}\right]$$

Dabei gilt für die beiden Gleichungen:

$$C_u = \begin{cases} \frac{1}{\sqrt{2}} & \text{für } u = 0 \\ 1 & \text{für } u \neq 0 \end{cases} \quad \text{und} \quad C_v = \begin{cases} \frac{1}{\sqrt{2}} & \text{für } v = 0 \\ 1 & \text{für } v \neq 0 \end{cases}$$

Die DCT transformiert das Bildsignal vom Orts- in den Frequenzbereich und zerlegt es dazu in Kosinus-Basisfunktionen unterschiedlicher Frequenz. Jeder Bildausschnitt mit $N \times N$ Pixeln resultiert in $N \times N$ Koeffizienten, die als zweidimensionale Frequenzangaben interpretiert werden können. Hohe Frequenzen repräsentieren im Ortsbereich feine Bildstrukturen, niedrige Frequenzen dagegen homogene Flächen.

Orts- und Frequenzbereich

In Abb. 1.4 ist der Ausschnitt eines Graustufenbildes zu sehen. Es handelt sich dabei um einen Bildbereich aus 8×8 Pixeln, wie er z. B. auch vom JPEG-Standard verwendet wird (Kap. 1.6.1). Das Graustufenbild ist mit 8 Bit pro Pixel quantisiert, weshalb die einzelnen Pixel Graustufenwerte von 0 bis 255 besitzen können.

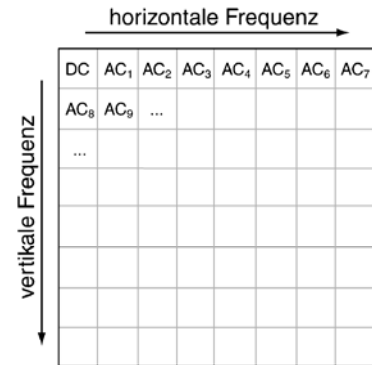
Beispiel



*Abb. 1.4:
Quadratischer Pixelblock
eines Graustufenbildes*

Abb. 1.5:
DCT-Koeffizienten eines Pixelblocks

| | | | | | | | |
|--------|--------|-------|-------|-------|------|------|-------|
| 806,8 | -136,5 | 56,6 | -59,3 | -94,5 | -2,7 | 7,9 | -8,8 |
| -137,6 | -71,0 | 35,7 | 20,4 | 12,3 | 24,4 | 12,2 | 3,3 |
| 47,6 | 9,8 | -67,8 | 46,9 | 88,6 | 12,0 | -2,1 | -4,3 |
| 1,9 | 15,3 | -3,5 | -21,2 | -8,2 | 19,0 | 18,8 | -11,5 |
| -33,5 | 43,8 | 34,6 | -66,5 | -31,3 | 37,5 | 22,2 | 1,7 |
| -17,6 | 0,4 | -37,8 | -52,1 | 7,1 | 13,8 | 1,0 | 0 |
| 1,1 | 8,4 | 0,4 | 15,6 | 14,7 | 3,9 | 5,3 | -9,6 |
| -5,5 | 10,1 | 15,6 | 1,3 | -11,0 | -2,0 | 2,0 | -8,6 |



DC-Koeffizient und AC-Koeffizienten

Abbildung 1.5 zeigt die gerundeten DCT-Koeffizienten des transformierten Pixelblocks. In Anlehnung an die englischen Bezeichnungen für Gleich- und Wechselstrom lassen sich die Koeffizienten in einen DC-Koeffizient und 63 AC-Koeffizienten unterscheiden. Der DC-Koeffizient, der häufig auch als Gleichanteil bezeichnet wird, entspricht dem Anteil der Frequenz 0 Hz in beiden Bildachsen und repräsentiert den mittleren Farbton bzw. Grauwert des 8×8 Pixelblocks. Die AC-Koeffizienten entsprechen Frequenzen größer 0 Hz in horizontaler und vertikaler Richtung. So entspricht der Koeffizient AC₇ beispielsweise der höchsten Frequenz, die im Pixelblock innerhalb der horizontalen Bildachse auftritt. Im Ortsbereich ist dies das dichtest mögliche Muster senkrechter Streifen innerhalb des Pixelblocks.

Weiterverarbeitung durch Entropie- und Lauflängencodierung

Da viele Bilder hauptsächlich aus flächigen Bereichen und nur zu einem geringen Teil aus scharfen Kanten bestehen, werden DCT-Koeffizienten, die höhere Frequenzen repräsentieren, sehr häufig niedrige Werte annehmen, wie auch in Abb. 1.5 beispielhaft zu erkennen ist. Die DCT-Koeffizienten sind also oft ungleich verteilt, und weisen somit eine geringe Entropie auf. Dies machen sich Verfahren zur Bild- und Videokompression zu Nutze und unterziehen die Koeffizienten einer nachfolgenden Entropiecodierung. Indem die Koeffizienten zuvor quantisiert werden, lässt sich die Kompression weiter optimieren. Durch die Quantisierung nehmen viele der kleineren Koeffizienten den Wert 0 an und können durch eine Lauflängencodierung zusammengefasst werden.

DCT-Koeffizienten können durch die IDCT theoretisch ohne Informationsverluste in die ursprünglichen Graustufenwerte zurück transformiert werden. In der Praxis lassen sich DCT und IDCT jedoch nur annähernd genau berechnen und die diskrete Kosinustransformation ist verlustbehaftet.

1.3.4.2

Wavelet-Transformation

Obwohl ein erstes Wavelet schon sehr früh im Rahmen der Haar-Transformation [Haa10] beschrieben wurde, entstand eine zusammenhängende Wavelet-Theorie erst in den 80er Jahren, nachdem Grossmann u. Morlet [GrM84] den Wavelets ihren Namen gegeben und eine genauere Beschäftigung mit dem Thema ausgelöst hatten.

Wie die DCT wandelt die Wavelet-Transformation das Bildsignal in den Frequenzbereich. Anstelle einer Kosinusfunktion werden dabei jedoch die Wavelets einer Wavelet-Familie als Transformationskern verwendet. Diese Wavelets erhält man durch einfache Modifikationen eines Mutter-Wavelets, von denen einige in Abb. 1.6 zu sehen sind. Das Haar-Wavelet ist sehr einfach aufgebaut und besitzt nur eingeschränkte Funktionalität, da es u. A. nicht differenzierbar ist. Aufgrund ihrer mathematischen Eigenschaften werden in der Signalverarbeitung oft Daubechies-Wavelets [Dau88] eingesetzt. Diese Wavelets lassen sich nicht funktionsanalytisch darstellen, sondern werden iterativ erzeugt. Durch die Wahlmöglichkeit eines Mutter-Wavelets und durch die Verwendung verschiedener Wavelets einer Wavelet-Familie ist die Wavelet-Transformation flexibler als die DCT mit ihrer festgelegten Basisfunktion.

Mutter-Wavelets können durch Verschiebungen auf der Abszisse und durch Skalierungen modifiziert werden. Durch Dehnung des Wavelets lassen sich stationäre Signalabschnitte erfassen, für die man eine gute Frequenzauflösung erhält, während sich gestauchte Wavelets mit hoher Frequenz für wechselhafte Signalabschnitte eignen, in denen eine gute zeitliche bzw. örtliche Auflösung gewünscht ist. Wavelets ermöglichen demnach eine Kombination aus Frequenzanalyse und zeitlicher bzw. örtlicher Analyse. Funktionswerte von Wavelets sind außerhalb eines engen Intervalls 0, während Kosinusfunktionen unendlich periodisch verlaufen. Durch diesen kompakten Träger wirken Wavelets immer nur auf einen Teilbereich des Bild- oder Tonsignals, und eine Aufteilung in Blöcke, wie z. B. bei der DCT, ist nicht mehr erforderlich.

Funktionsprinzip, Wavelet-Arten

Modifikation von Mutter-Wavelets

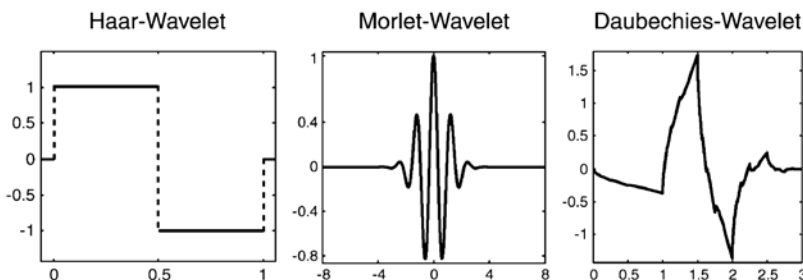


Abb. 1.6:
Wavelets

Multiskalen-Analyse

Die Multiskalen-Analyse [Mal89] führt eine Skalierungsfunktion als Basis der Wavelet-Familie ein und ermöglicht so eine effiziente algorithmische Umsetzung der Transformation. Diese wird als schnelle Wavelet-Transformation (FWT) bezeichnet und lässt sich als Kombination spezieller Hoch- und Tiefpass-Filter auffassen.

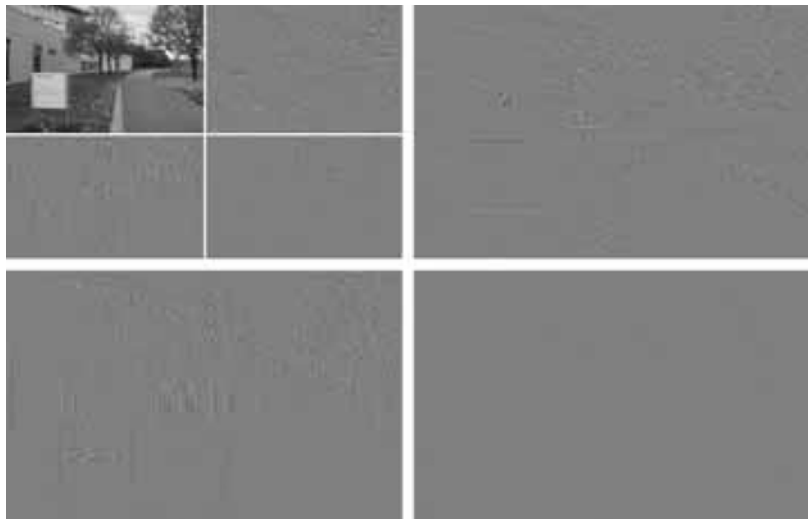
schnelle Wavelet-Transformation (FWT)

In der Bilddatenkompression wird die FWT meist nacheinander auf die Zeilen und Spalten des Bildes angewendet. Zunächst werden die Zeilen des Ausgangsbildes transformiert. Durch den Tiefpass-Filter erhält man eine geglättete, unscharfe Version des Bildes, der Hochpass resultiert in feinen Detailinformationen. Zusätzlich wird die Bildauflösung durch Subsampling reduziert, indem jedes zweite Pixel verworfen wird. Derselbe Vorgang wird nun auf die Spalten angewendet, wodurch man insgesamt vier verkleinerte Versionen des Ausgangsbildes erhält. Die Bildversion, deren Spalten und Zeilen Tiefpass-gefiltert wurden, kann nun mehrere Male auf dieselbe Weise weiterverarbeitet werden, wobei in jedem Verarbeitungsschritt vier weitere verkleinerte Bildversionen entstehen. Die drei anderen Bildversionen enthalten Detailinformationen, die mit senkrechten, waagerechten bzw. diagonalen Kanten innerhalb des Bildinhaltes korrespondieren. Durch die schrittweise Verarbeitung durch Filter erhält man eine unscharfe Repräsentation des Originalbildes in sehr geringer Auflösung, sowie mehrere Bilder mit Detailinformationen in unterschiedlichen Auflösungen.

Beispiel

In Abb. 1.7 ist dies anhand der ersten zwei Verarbeitungsschritte mit einem Haar-Wavelet zu sehen. Bei der Dekompression wird die gering aufgelöste Bildrepräsentation sukzessive durch die Detailinforma-

*Abb. 1.7:
Zwei Verarbeitungsschritte der
schnellen Wavelet-Transformation*



tionen angereichert, was als hierarchische Kompression bezeichnet wird. Hierarchische Kompression ist auch schon im JPEG-Standard (Kap. 1.6.1) als spezieller Modus verfügbar, allerdings auf Basis einer DCT. Die Dekompression des Bildes muss nicht komplett erfolgen, sondern kann gestoppt werden, wenn eine ausreichend hohe Auflösung erreicht wurde.

Die Wavelet-Transformation kann in der Bildverarbeitung anstelle der DCT eingesetzt werden, wodurch sich die eingangs geschilderten Vorteile der Wavelets nutzen lassen. Wie bei der DCT werden die resultierenden Koeffizienten durch eine Quantisierung und Entropiecodierung weiterverarbeitet. Die FWT ist bei Wahl eines geeigneten Wavelets und zugehöriger Skalierungsfunktion ein verlustfreier Vorgang. Trotz des Subsamplings lässt sich das Bild ohne Informationsverlust rekonstruieren. Erst durch eine entsprechende Quantisierung wird die Wavelet-Kompression zum verlustbehafteten Vorgang.

Eigenschaften

1.4 Medienspezifische Verarbeitungsschritte

Neben elementaren Verarbeitungsschritten, die sich für die Datenkompression sehr vielseitig einsetzen lassen, existieren auch medien-spezifische Verarbeitungsschritte, bei denen die Charakteristiken bestimmter Medientypen ausgenutzt werden.

1.4.1 Interframe-Kompression

Bei der Kompression von Videodaten lassen sich besonders hohe Kompressionsraten erzielen, wenn zusätzlich zu örtlichen Redundanzen innerhalb eines Einzelbildes auch zeitliche Redundanzen zwischen aufeinander folgenden Bildern berücksichtigt werden. In vielen Fällen sind Folgebilder einer Videosequenz sehr ähnlich, da sich von Bild zu Bild nur begrenzte Bereiche unterscheiden.

Werden auch zeitliche Redundanzen für die Kompression ausgenutzt, so spricht man von Interframe-Kompression. Im Gegensatz dazu verarbeitet die Intraframe-Kompression einzelne Videobilder separat und berücksichtigt somit nur örtliche Redundanz.

Interframe- und Intraframe-Kompression

Ein möglicher Ansatz für die Interframe-Kompression ließe sich realisieren, indem die Prädiktion nach Kap. 1.3.3 um eine Dimension für die Zeit erweitert wird. Um jedoch exaktere zeitliche Vorhersagen zu erhalten, werden Bewegungen innerhalb des Bildinhalts abgeschätzt. Dabei beschränkt man sich in der Regel auf translatorische