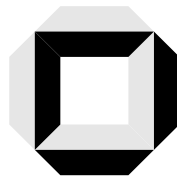


MULTIMEDIA DATENSPEICHERUNG UND DATENBANKEN (Tutorial)

A. Klingert, C. Knörzer, A. Schmitt

Interner Bericht
Nr. 32/94



Universität Karlsruhe

Fakultät für Informatik

Institut für Betriebs- und Dialogsysteme
Lehrstuhl Prof. Dr. A. Schmitt

1994

Zusammenfassung

Multimedia hat heute in Technik, Forschung und Wirtschaft einen Stellenwert erreicht, der es notwendig macht, sich mit dieser Problematik vertraut zu machen.

Wir wollen versuchen aufzuzeigen, welche die wichtigsten Anwendungsmöglichkeiten für diese neue Technik sind und welchen Anforderungen Multimediasysteme idealerweise genügen sollten. Darüberhinaus erklären wir, warum in mancher Hinsicht Diskrepanzen zwischen Wunsch und Wirklichkeit zu verzeichnen sind.

Diese Defizite beziehen sich vor allem auf mangelndes Problemverständnis, aber auch auf Aspekte wie Rechenleistung sowie fehlende Infrastruktur und flächendeckende Standards. Wir zeigen das anhand des Problems, eine Bewegbilddatenbank zu implementieren (insbesondere Wiederfinden und Kompression) sowie anhand des verteilten Bewegbilddatenspeichers in einem herkömmlichen Workstation-LAN.

Inhaltsverzeichnis

1	Multimedia-Anwendungen	1
1.1	Zum Begriff Multimedia	1
1.2	Die wichtigsten Anwendungen	1
2	Multimedia-Daten	3
2.1	Monomediale Daten	3
2.2	Multimediale-Daten	4
3	Operationen auf Multimedia-Daten	5
4	Stand der Technik	6
4.1	Analyse der technischen Grenzen	6
4.2	Zusammenfassung der technischen Grenzen	8
5	Datenkompression	9
5.1	JPEG	10
5.2	H.261 (px64)	14
5.3	MPEG	14
5.4	Kompression und Qualität	18
6	Transport von Bewegtbildern im LAN	19
6.1	Technische Randbedingungen	19
6.2	Messungen	20
6.3	Fazit	22
7	Trends und Ausblick	23
A	Anhang	26

1 Multimedia-Anwendungen

1.1 Zum Begriff Multimedia

Der Begriff Multimedia heißt, wörtlich genommen, nichts weiter, als daß mehrere (unterschiedliche) Medien an einer Quelle, während eines Transports oder an einer Senke vereinigt vorliegen. Mit Medien ist dabei nicht das physikalische Medium gemeint, auch nicht ein Medium bezogen auf die Wahrnehmung, sondern ein weit abstrakterer Begriff: Wir unterscheiden Medien nach der Darstellung und nach dem Informationstypus (Semantik), der transportiert wird [MW91]. Die Klassifikation von Medien kann jedoch nach zahlreichen weiteren Kriterien durchgeführt werden.

Beispiele für Klassifikationsmöglichkeiten sind [Ste93] :

- nach der Art der Speicherung oder Kodierung,
- nach der Art der Wahrnehmung,
- nach der Art der Darstellung
- sowie nach den angesprochenen Sinnesorganen und
- der zeitlichen Änderbarkeit eines Mediums.

Es bietet sich an, obige wörtliche Interpretation des Multimedia-Begriffes als „Multimedia im weiteren Sinne“ zu bezeichnen. Ein herkömmliches Fenstersystem beispielsweise ist ein multimediales System in diesem Kontext, denn Text und Graphik sind hier erfolgreich intergriert.

Diese Art von multimedialer Anwendung schwebt einem jedoch meistens nicht vor, wenn von Multimedia die Rede ist. Man meint vielmehr die Kombination von Medien, die sowohl zeitveränderlich sind (zeitkontinuierlich oder -diskret) und solche, die weitgehend zeitinvariant sind. Zu ersteren gehören Ton (Sprache, Musik, Geräusche) und Bewegtbild (Video, Stummfilm); zu letzteren gehören Texte (formatiert, unformatiert), Graphiken und Rasterbilder. Wir sprechen bei diesen Kombinationen von „Multimedia im engeren Sinne“. Von dieser Art Multimedia soll im folgenden hauptsächlich die Rede sein.

Dabei erheben sich die Fragen welche Medienkombinationen für welche Anwendungen relevant sind und welche Anwendungen welches wirtschaftliche Gewicht haben.

1.2 Die wichtigsten Anwendungen

Die wichtigsten Anwendungen für Multimedia [GER94, Wir94] sind:

Spiele Visuelle Medien (Graphiken, Rasterbild, Bewegtbild) werden mit Audiodaten kombiniert. Dabei spielen Realzeitaspekte eine große Rolle. Multimedialen Spielen wird ein bedeutendes wirtschaftliches Gewicht zugetraut.

Auch wenn man sich nicht mit der Zielsetzung dieser Anwendungen identifiziert, muß man doch gestehen, daß sie in der Lage sind, die Entwicklung nachhaltig zu beeinflussen; insbesondere deshalb, weil in den nächsten Jahren vermutlich ein exponentielles Wachstum zu erwarten ist.

Computer Based Training (CBT) Interaktives Lernen ist grundsätzlich nichts Neues. Eine neue Dimension wird durch die Kombination von Ausprobieren (anhand von Graphiken), Nachlesen (Text) und Demonstrieren (anhand von Video, Ton) möglich. Ein Programmiersystem läßt sich vermutlich leichter durch ein interaktives und durch Querverweise angereichertes System erlernen als durch sequentielles Lesen der Handbücher; insbesondere, wenn man sofort probieren kann, was man gelernt hat – wie es bei Lernprogrammen über Programmieren oder bei Simulationsprogrammen der Fall ist.

Diese Technik wird experimentell bereits angewandt. Kommerzielle Systeme gibt es im Bereich des Fremdsprachenlernens. Beobachtet und prognostiziert wird ein lineares Wachstum dieser Sparte.

Videokonferenzen Neben der Kombination von Bewegtbild und Ton werden auch eingescannte Graphiken und Texte übermittelt. Es besteht Bedarf an Datentransport über unterschiedliche Distanzen und für alle Typen von Medien. Synchronisation der einzelnen Medien an der jeweiligen Senke ist dabei ein wichtiger Aspekt. Das wirtschaftliche Potential von Anwendungen in diesem Bereich ist hoch, wenn auch schwer abzuschätzen.

	Sprache	Musik	Geräusch	Bewegtbild	Text	Bild
Spiele	-	wichtig	wichtig!	wichtig!	-	wichtig!
CBT	bedingt	-	bedingt	wichtige Ergänzung	wichtig	wichtig
Video-konferenz	wichtig	-	bedingt	wichtig !	zweitrangig	wichtig
Kooper. Arbeiten	wichtig	-	bedingt	als Ergänzung	wichtig	wichtig!
Interakt. TV	bedingt	wichtig	wichtig	wichtig	-	-
Kiosk	bedingt	wichtig	-	wichtige Ergänzung	wichtig	wichtig
Informations-systeme	bedingt	bedingt	-	wichtige Ergänzung	wichtig	wichtig

Tabelle 1: Bedeutung der einzelnen Medien für die wichtigsten Anwendungsgebiete. Das Ausrufezeichen bei manchen Einträgen soll ihre besonders wichtige Stellung unterstreichen. Unter der Bezeichnung 'Bild' sind Graphiken und Rasterbilder zusammengefaßt, die sich bezüglich der Relevanz für die Anwendungsgebiete nicht unterscheiden.

Kooperatives Arbeiten Kooperatives Arbeiten deckt einen weiten Bereich an Anwendungen ab. Es reicht vom gemeinsamen Editieren von Texten und Graphiken bis hin zur gemeinsamen Montage eines Triebwerks in Canada, die der italienische Ingenieur an seiner Workstation verfolgt und mit seinen technischen Zeichnungen vergleicht.

Im Vordergrund steht der Transport der Medien; die Synchronisation verschiedener Medien spielt eine entscheidende Rolle. Zusätzliche Aspekte sind Sicherheit und Konsistenz der bearbeiteten Dokumente. Eine stärkere Verbreitung des kooperativen Arbeitens hängt maßgeblich von den Lösungen, die mit der Einbettung in die Benutzungsschnittstelle (simultanes Zeigen, Malen, Tippen, Sprechen) verbunden sind.

Die wirtschaftliche Zukunft dieser Anwendung wird sehr unterschiedlich bewertet.

Interaktives Fernsehen Als eine Anwendung, die mit Sicherheit einen Markt finden wird, steht das 'Videospiel per Fernsehkanal' bereits fest. Dieses wird vor allem durch den Einstieg von Videospiele-Konzernen in den amerikanischen TV-Markt deutlich. Spiele werden zunächst ohne Kosten probierhalber vorgestellt und befristet zur Verfügung gestellt. Der Benutzer kann über einen geeigneten Rückkanal das Spiel ausprobieren. Anschließend wird dasselbe Spiel auf einem anderen Medium (CD-ROM) verkauft. Interaktives Fernsehen dient somit als Erweiterung der Marketing-Möglichkeiten. Für weitere Applikationen mangelt es bislang an tragfähigen Ideen und Visionen.

Auf Dauer ist interaktives Fernsehen ohne zusätzliche Hardware beim Konsumenten nicht denkbar. Das wiederum hat eine eher langfristige Entwicklung zur Folge. In diesem Zusammenhang fällt häufig das Schlagwort „Powerstation“ als eine nicht näher definierte Verbindung von PC und TV.

Multimedia-Kiosk Ein Multimedia-Kiosk erlaubt das Nachfragen, Bestellen und Buchen mithilfe eines multimedialen Terminals rund um die Uhr. Man unterscheidet lokale Kioske (z.B. das Hotelverzeichnis des Verkehrsvereins Graubünden) und verteilte, die von einer Zentrale automatisch und periodisch aktualisiert werden können. Verteilte Kiosksysteme sind recht aufwendig. Lokale finanzieren sich oft bereits über eingebaute Anzeigen.

Worin liegt der potentielle Nutzen dieser Einrichtungen?

Kioske können automatisch ihre eigene Akzeptanz protokollieren und ein Benutzerprofil erzeugen (derzeit zu ca. 70% von Männern konsultiert). Sie sind häufig vielsprachig, vermitteln Wissen unbewußt anhand von Bildern und Videos und können mehrere Branchen mit der gleichen Benutzungsschnittstelle abdecken. Im Zusammenhang mit der Diskussion um die Ladenöffnungszeiten kann Kiosken eine besonders hohe wirtschaftliche Bedeutung zukommen.

Beispiele für zentrale Systeme zum Ausschuchen und Buchen werden derzeit von großen Reiseveranstaltern (Global Explorer, MultiMeier) auf den Markt gebracht. Für 1995 wird mit einem gegenüber 1994 verdoppelten Umsatzvolumen bei Kiosken gerechnet.

Informationssysteme Es gibt grundsätzlich zwei Typen von multimedialen Informationssystemen. Wir unterscheiden „online“-Systeme, bei denen Transport von Daten über ein digitales Netz eine wesentliche Rolle spielt und „offline“-Systeme, bei denen üblicherweise CDs als physikalische Medien benutzt werden. Üblich ist eine von Graphik und Text dominierte Darstellung, die von einzelnen Video-Sequenzen und Ton untermalt werden kann. Beispiele für solche Informationssysteme gibt es in großer Zahl: Nachschlagewerke, Kataloge, Wartungshandbücher, Gesetzestexte und Gerichtsurteile.

Zeitschriften werden schon heute in dieser Form (meist online) angeboten, und Enthusiasten sprechen bereits von *Mediamorphose*. So ist beispielsweise die Wirtschaftswoche auf diese Art (partiell) bereits Mittwochs verfügbar, wohingegen sie Abonnenten erst ab Freitag ausgeliefert wird. 17 verschiedene Zeitungen und Zeitschriften werden derzeit als Informationssysteme in Deutschland angeboten. Man erwartet, daß im Bereich der Zeitungen und Zeitschriften Multimediasysteme bis in 6 Jahren ca. 10-15% Anteil am Umsatz (europaweit) haben werden. Insgesamt wird von einem eher langsamen Wachstum des Informationssysteme-Marktes ausgegangen.

Tabelle 1 faßt noch einmal zusammen, welche Anwendungsgebiete auf welche Medien und Medienkombinationen angewiesen sind. Es zeigt sich deutlich, wie wichtig die Integration von Bewegtbild und Bild ist, auf die wir in dieser Abhandlung noch genauer eingehen werden. Als nächstes wollen wir die einzelnen Medien etwas genauer betrachten.

2 Multimedia-Daten

Erst das Arbeiten an multimedialen Problemen hat dazu geführt, daß man sich der monomedialen Eigenschaften der bisherigen Systeme bewußt wurde. Der Übergang zu multimedialen Anwendungen geschieht dadurch, daß man nach Gemeinsamkeiten der einzelnen Medien fahndet, um ihre Behandlung zu vereinheitlichen und damit die Implementierung und die Interaktionen für den Benutzer zu vereinfachen.

2.1 Monomediale Daten

Mit multimedialen Daten verbindet man häufig die Vorstellung vollständig digitaler Systeme. Daß das keine notwendige Voraussetzung für Multimedia ist, darauf wurde von verschiedenen Autoren bereits hingewiesen [Ste93].

Man spricht auch dann von einem Multimediasystem, wenn nur analoge Geräte (Videorekorder, analoges Audio) beteiligt sind, die Steuerung der Komponenten jedoch digital erfolgt. Der Trend besteht demgegenüber in einer zunehmenden Digitalisierung und Integration dieser Komponenten. Bedingt durch technische Unzulänglichkeiten sind zahlreiche Übergangsvarianten notwendig und wichtig. Da wir uns auf zukünftige Multimedia-Anforderungen konzentrieren, gehen wir von digitalen Daten aus.

Text Er kann in Form von formatierten oder unformatierten Daten vorliegen; er kann linear angeordnet sein (Papierform) oder mit Querverweisen, Struktur und Indexierung (Hypertext) versehen. Heutige Datenbank-Systeme verarbeiten fast ausschließlich formatierte alphanumerische Daten, die wir im folgenden mit Text gleichsetzen.

Graphik Unter Graphikdaten sollen solche Beschreibungen eines Bildes verstanden werden, die in der Lage sind, mathematisch exakt ein Bild auf verschiedenen Ausgabemedien (Bildschirm, Drucker etc.) zu reproduzieren.

In digitaler Form sind hier binäre Dateiformate und textuelle Dateiformate (PostScript) gebräuchlich. Man bildet also häufig Graphiken auf ein textuelles Medium ab, um sie zu speichern. Dies ist ein Beispiel einer sogenannten *Medientransformation*.

	Audio	Bewegtbild	Text	Graphik, Rasterbild
Einheiten der Verwaltung	Satz .. Symphonie Zeitdauer	Sequenz Frame Zeitdauer	Satz .. Seite, Kapitel Kompendium	Einzelbild Seite
wichtigste Parameter	Auflösung Anzahl Bit	Format zeitl. Aufl. Anzahl Bit	formatiert? linear?	Format Farbmodell Anzahl Bit

Tabelle 2: Dateneinheiten, mit denen Medien kombiniert werden

Rasterbild Es gibt kaum ein Medium, für das es mehr Formate und Kodierungen gibt als für Rasterbilder. Grundsätzlich unterscheidet man Farbtabellebilder und Vollfarbenerbilder, komprimierte Speicherung und unkomprimierte Speicherung sowie die Archivierung mittels Bilddatenbank oder – weniger strukturiert – als Dateien. Ein Vollfarbenerbild (24Bit) der Größe 800×600 Bildpunkte hat in unkomprimierter Form eine Größe von 1.4 MB.

Digitales Audio Die Kodierung von Audio erfolgt per zeitlicher Abtastung und anschließender Quantisierung. Dabei entscheidet die Abtastrate über die Qualität der Aufzeichnung. Die Qualität des Telefons entspricht einer Abtastung mit 8kHz und einem 4 Bit Wert. Die Qualität der CD mit 44kHz und 16 Bit erzeugt für einen Audiostrom von einer Stunde 600MB (stereo, unkomprimiert).

Bewegtbild Wir isolieren hier die Bewegung vom Ton (Audio), auch wenn diese Trennung künstlich erscheinen mag. Da die Audiodatenmenge im Verhältnis zur Datenmenge des digitalen Bewegtbildes fast vernachlässigbar ist, entscheidet die Handhabung letzterer über einsetzbare Algorithmen. Auch hier spielt die Abtastrate eine Rolle. Allerdings hat sich die Video-Abtastung 25 bzw. 30(USA) Frames/sec (bis auf wissenschaftliche Zwecke) durchgesetzt. Für PAL-kodierte Daten (767×575 , YUV im Verhältnis 4:2:2) braucht man für eine Stunde Video ca. 74GB (unkomprimiert) zur Speicherung.

Monomediale Daten setzen sich zusammen aus Rohdaten, Registrierungsdaten sowie aus identifizierenden Daten. Bei Rasterbildern bilden die Pixelwerte die Rohdaten, Größe und Farbtiefen-Information die Registrierungsdaten, und die identifizierenden Daten werden z.B. durch Ort und Zeit der Entstehung beschrieben. Einzelne dieser Komponenten können auch entfallen oder werden nur zur Speicherung gebraucht und vorübergehend hinzugefügt [MW91].

2.2 Multimediale-Daten

Wie bereits aus Tabelle 1 ersichtlich, sind im professionellen Bereich die Integration von Text, Graphik und Rasterbild von besonderer Bedeutung und heute auch bezüglich mancher der unten aufgezählten Operationen vollzogen.

Allerdings darf eine Integration an der Benutzungsschnittstelle nicht darüber hinwegtäuschen, daß Archivierung und Wiederfinden der Daten weitgehend auf Dateibasis und nach Medien getrennt funktioniert. Beispiele hierfür sind Hypertext-Systeme wie Mosaic und diverse Online-Manuals.

Eine weitere wichtige Medienkombination besteht aus Bewegtbild, Audio und evtl. Rasterbildern sowie Graphiken. In diesem Bereich bestehen noch einige Hindernisse, die in der Hardware bzw. der Systemsoftware begründet liegen. Existierende Hypermedia-Systeme zielen vor allem auf das Auge des Benutzers; ihre interne Realisierung ist häufig noch nicht aus einem Guß.

Um verschiedene Medien überhaupt erst kombinieren zu können, muß man die Einheiten der Kombination festlegen. Tabelle 2 soll dabei helfen, die im folgenden verwendeten Einheiten und ihre wichtigsten Parameter zu identifizieren.

3 Operationen auf Multimedia-Daten

Die wichtigsten Operationen auf Daten, die im Zusammenhang mit multimedialer Verarbeitung eine Rolle spielen, sind folgende [MW91]:

Archivierung Zunächst gehört zum Archivieren das Schleusen der Daten von Eingabegeräten (oder einem Netz) zu einem geeigneten physikalischen Speichermedium.

Dieses Speichermedium muß den Anforderungen der evtl. notwendigen Geschwindigkeit sowie der benötigten Datenmengen angepaßt sein. Daher kann es nötig werden, Daten gewissermaßen vorzuarchivieren, d.h. auf einem schnellen, kleinen Speichermedium zwischenspeichern, wo sie editiert werden können, um sie dann auf einen voluminösen aber langsameren Speicher auszulagern.

Häufig werden quasi-simultan zum Schreiben der Rohdaten in einer Datenbank zusätzliche identifizierende Daten und Registrierungsdaten erfaßt. Registrierungsdaten können benutzt werden, um den Zustand eines Datensatzes zu beschreiben (Archivierung, Editierstadium etc...).

In bezug auf echte Multimedia-Daten ist es wichtig, daß der Vorgang für einzelne Medien überlappend bzw. nacheinander gesteuert werden muß.

Wiederfinden Das Wiederfinden multimedialer Daten kann reduziert werden auf das mehrfache Wiederfinden einzelner Medien. Dies wiederum funktioniert anhand von:

1. identifizierenden Daten (textuell und formatiert: herkömmliche Datenbank)
2. Registrierungsdaten (textuell und formatiert: herkömmliche Datenbank)
3. Rohdaten (bzw. Teilen davon)

Von Interesse, weil in den meisten Datenbanken noch nicht realisiert, ist das Suchen und Wiederfinden auf Rohdaten. Handelt es sich um das Medium Text (eindimensional), dann kann man mit Volltextsuche, basierend auf Metriken des „Pattern Matching“, zum Ziel kommen.

Bei Bildern helfen Metriken auf Farben oder Inhalten weiter, die ebenfalls in der Bildverarbeitung erarbeitet wurden. Eingesetzt worden sind solche Verfahren jedoch allenfalls in Versuchssystemen. Zum einen gibt es dabei das Problem der Anfrage (meist Spezifikation eines Bildes durch Text), zum anderen technische Grenzen, was die Geschwindigkeit angeht. Dazu mehr im folgenden Kapitel.

Anzeigen Zur visuellen Kontrolle, aber auch zum Editieren müssen verschiedene Medien kombiniert angezeigt werden. Das führt die gesamte Problematik der Benutzungsschnittstellen ein. Dadurch berührte Problemfelder sind: Einsammeln und Auslesen der Daten aus dem Speicher, Synchronisation einzelner Medien und Übersichtlichkeit der Navigationsmöglichkeiten.

Löschen Um multimediale Daten zu löschen, muß man natürlich alle Schritte des Archivierungsvorgangs (oder Vorarchivierungsvorgangs) rückgängig machen. Zusätzlich müssen im Falle, daß nur einzelne Medien aus einem multimedialen Satz gelöscht werden, Querverweise entfernt werden.

Transportieren Man unterscheidet – ganz grob – Transport von Medien mit Dringlichkeit und solchen ohne Dringlichkeit. Von hoher Dringlichkeit spricht man, wenn an der Senke innerhalb einer bestimmten Zeit oder synchron zu einem anderen Vorgang Medien vorliegen müssen.

Zum Transportieren gehören:

- Verpacken (incl. Quellkodierung und Kompression),
- Kontrolle der Zeitbedingungen,
- Auspacken von Einheiten.

Fusionieren Unter Multimedia-Datenfusion versteht man das räumliche und zeitliche „In bezug setzen“ von verschiedenen Medien, die sich auf ein und dasselbe Objekt beziehen. Häufig ist dies ein sichtbares Problem an der Anzeige-Operation, es kann aber auch ganz unabhängig davon gesehen

werden. Ein Beispiel:

- + Bewegtbild vom Galopp eines Pferdes,
- + Geräusch des Hufschlags,
- + animierte Graphik zum Bewegungsablauf des Pferdeskeletts,
- + Textinformation zur Pferderasse oder über beobachtete Anomalien,
- + Standbild des Pferdes.

Das Zuschalten des Tones zum Bild, das Einblenden des Textes sowie das Umschalten auf die Animation sind Fälle, in denen Datenfusion gebraucht wird.

Inwieweit die Medienintegration bezüglich der beschriebenen Operationen stattgefunden hat, soll anschließend geklärt werden.

4 Stand der Technik

Hier soll der Frage nachgegangen werden, welche Operationen auf Multimedia-Daten heute ausgeführt werden können, welche Hilfsmittel es dafür gibt und wodurch technische Grenzen gesetzt sind.

4.1 Analyse der technischen Grenzen

Archivierung Vorab kann festgehalten werden: Die technischen Grenzen der Archivierung werden bzgl. sämtlicher Teilvorgänge durch die Bewegbilddaten erreicht.

Das Einschleusen von Mediendaten ist nur durch die Datenmenge pro Zeit begrenzt, die ein System in der Lage ist, ein- oder auszuschleusen. Diese Parameter werden durch System- und Peripherie-Bus beschränkt. Bewegtbilder in vollem Videoformat (767×575) können auch auf einer Workstation nur mit ca. 3-5 Frames/sec ohne Spezialhardware digital archiviert werden.

Das Gesamtdatenvolumen und die dadurch notwendigen Speicherkapazitäten liegen im Bereich von Gigabyte für wenige Minuten Spielzeit. Zugriffszeit und Speicherkapazität des physikalischen Mediums stellen eine Einschränkung dar:

Vorarchivierung erfolgt häufig durch Festplatten mit mittlerem bis hohem Datenvolumen, Zugriffszeiten im Bereich von 10-20 msec und maximalen Datentransferraten von ca. 20 MB/sec, die Archivierung dann u.U. auf Bändern bzw. Magneto-Optischen-Laufwerken (max. 1.4 GB) mit deutlich längeren Zeiten für einen Zugriff (Magneto-Optische-Laufwerke: Zugriff mind. 20 msec, Transfer max. 2 MB/sec)

Datenbanken, die sämtliche Multimedia-Daten verwalten, bezeichnet man als Multimedia-Datenbanken (MMDB). Bislang gibt es nur fortgeschrittene Experimentalsysteme [MW91]. Generell werden zwei Wege eingeschlagen. Einerseits werden bestehende relationale Systeme um neue Objekte und Operationen erweitert. Andererseits werden vollständig neue Systeme nach dem abstrakteren, objektorientierten Ansatz entworfen.

Lösungsansätze für die Leistungsprobleme bestehen entweder in der Reduktion auf kleinere Formate (CIF, QCIF s.u.) oder in massiver und schneller Kompression (JPEG, MPEG, Wavelets, s.u.) bzw. in einer grundsätzlichen Parallelisierung der Vorgänge. Dabei stellt die Formatreduktion sicherlich nur eine temporäre Lösung dar, während alle weiteren Varianten ganz erhebliche Rechenleistungen bis hin zu Spezialhardware erfordern.

Wiederfinden Das Suchen und Vergleichen auf Basis von Registrierungs- und identifizierenden Daten stellt keine technischen Probleme dar (eindimensionales Pattern Matching). Technische Grenzen werden durch das Suchen auf Rohdaten erreicht; hier vor allem bei Audio-, Bild- und Videodaten.

Audiodaten sind zwar primär zeitlich linear angeordnet, sie haben aber die zusätzliche Dimension des Frequenzspektrums, so daß rechenaufwendige Matching-Verfahren notwendig werden. Ein weiteres Problem ist die Spezifikation der Anfrage. Zwar sind beispielsweise sprecherspezifischer

Dialekt und Tonhöhe trainierbar, dennoch ist eine eindeutige Metrik nicht trivial und erfordert häufig eine Transformation, die wiederum erheblichen Rechenaufwand verursacht. Generell sind solche Systeme auf Workstation-Basis möglich.

Es gibt verschiedene Ansätze, dieses Problem anzugehen. Wir wollen uns hier auf zwei Varianten beschränken:

1. In dem Projekt REDI (später IMAID genannt) wurde versucht, Bilddaten zunächst nach Großthemen zu trennen (Satellitenaufnahmen, medizinische Aufnahmen etc.). Anschließend wurden Bildanalyse-Methoden benutzt, um die extrahierten Linien sowie einfache Objekte mit Hilfe des Kontextes zu themenbezogenen Textobjekten zu machen. Diese wurden dann als Daten in eine relationale Datenbank übernommen [CF80]. Als Anfrage konnte ein Beispielfeld oder eine Skizze verwendet werden.
2. Mit Iconic Indexing bezeichnet man den Vorgang, benachbarte Bildpunkte eines Bildes zu einem Wert zu verschmelzen. Das reduziert die Größe des Bildes und damit den Suchraum, und man kann Schritt für Schritt (je nach Bedarf) die Bilder vereinfachen.

Eine Zusammenfassung hierzu findet man in [MW91].

Bewegtbilder müssen entsprechend entweder nach einzelnen Bildern oder nach Sequenzen (Bewegungen) durchsucht werden. Letzteres erfordert einen 3D-Suchraum, der zusammen mit der immensen Datenfülle und den mehr oder minder heuristischen Metriken auch auf leistungsfähigen Arbeitsplatzrechnern nicht in vertretbaren Rechen- bzw. Antwortzeiten realisierbar ist.

Notwendige Konsequenz ist die Kompression der Bild- und Bewegtbilddaten, denn eine Verkleinerung der Formate ändert an der Problemstellung nichts.

Anzeigen Die Leistungsprobleme im Zusammenhang mit der Anzeige von Medien erfordern Zusatz-Hardware, die allerdings schon in den Bereich der Consumer Systeme vordringt.

Wesentlich zäher gestalten sich die Lösungen, die die Handhabung von Multimedia-Systemen leichter machen sollen. Das Phänomen, den Kontext einer Recherche zu verlieren, wurde unter der Bezeichnung „Lost in Hyperspace“ bekannt. Am MIT entwickelt, bildet *Cybermap* eine Lösungsmöglichkeit an. Es erlaubt die Navigation in großen Datenbeständen anhand von graphischen Karten, die hierarchisch angeordnet sind und die Filterfunktionen und Browsing-Funktionen integriert haben.

Es gibt Demonstrationsimplementierungen auf dem Mac und eine parallelisierte Version mittels C^* auf der Connection Machine-2.

Löschen Es sind keine spezifischen oder nachhaltigen Beschränkungen bekannt.

Transport Dringlicher Transport von Audiodaten aber auch von Videodaten sind im LAN generell möglich. Dabei ist die Qualität von Audio im Bereich der CD, die Qualität von komprimiertem Video (wie unten erläutert) auf ca. 8-10 Frames pro Sekunde beschränkt.

Im MAN und WAN Bereich kommt es maßgeblich auf die eingesetzte Netzhardware und die Kompression an, ob vergleichbare Werte erreicht werden können. So sind Audio-Konferenzen im Internet mit mäßiger Qualität an der Tagesordnung (μ - law-Kodierung: 8Bit, 8kHz).

Fusionieren Ein Problem der Datenfusion ist es, ein Beschreibungsformat zu finden, mit dem die Fusion definiert werden kann. SGML, HTML sind Beispiele für solche Formate. Das Hauptproblem der Datenfusion ist es jedoch, eine Synchronisation zwischen einzelnen Dateneinheiten verschiedener Medien herzustellen. Ist diese Synchronisation nicht durch die Eingabedaten vorgegeben (Live-Synchronisation), so muß sie künstlich, z.B. manuell, hergestellt werden (synthetische Synchronisation).

Man kann die Synchronisation anhand globaler, relativ vager Operatoren beschreiben, z.B.: die Medien verlaufen zeitlich *parallel*, *nacheinander* oder *unabhängig voneinander*. Exakter jedoch

sind Operatoren, die die genauen Zusammenhänge zwischen den einzelnen Medieneinheiten beschreiben. Das ist jedoch relativ aufwendig. Benutzt man eine gemeinsame Zeitachse (indirekte Adressierung), so muß man jede Medieneinheit in Bezug zu dieser Achse setzen. Definiert man den Bezug von einzelnen Medieneinheiten hingegen untereinander (direkte Adressierung), so gibt es 13 verschiedene Relationen zweier Medieneinheiten und eine große Zahl an Relationen, die festgelegt werden müssen [Ste93].

Wiederfinden und Anzeigen sind von allen erwähnten Operationen, die rechenintensivsten Vorgängen; Archivierung dann, wenn Kompressionsmethoden eingesetzt werden.

4.2 Zusammenfassung der technischen Grenzen

Drei technische Grenzen stellen sich als besonders einschränkend dar: Da ist erstens die Datenfülle, vor allem im Bereich der Videodaten, also mit anderen Worten die Archivierung von Bewegtbilddaten in einem anwendungsangemessenen digitalen Format. Zur Lösung dieses Problems werden im folgenden Abschnitt Kompressionsverfahren vorgestellt.

Und zum anderen gibt es die Leistungsaspekte der inhaltlichen Suche in Bildern, also den Aspekt 'Wiederfinden von Bewegtbildsequenzen und Bildern'. Auch hier können komprimierte Bilder eingesetzt werden. Dennoch sind die Rechenzeiten erheblich: durch die eingesetzte Kompression sowie durch das Errechnen verschiedener Metriken.

Der dritte Grenzbereich ist der Transport (mit hoher Dringlichkeit) in Netzen. Die momentan verbreiteten LANs wie Ethernet und Token-Ring mit ca. 10 MBit/sec erlauben keine Echtzeitübertragung mit voller Video-Auflösung. Was mit diesen Netzen erreicht werden kann, wird im letzten Abschnitt aufgezeigt. Wird volle Video-Auflösung und volle zeitliche Auflösung verlangt, so müssen neuere Netztechnologien eingesetzt werden. FDDI ist hier das aktuelle Schlagwort. Mit nominal 10facher Geschwindigkeit (gegenüber Ethernet) können bei dieser Technik Daten transportiert werden. Diese Tatsache zusammen mit der in FDDI-2 garantierten isochronen Übertragung ermöglichen einen zuverlässigen Transport in Echtzeit auch bei vollem Videosignal und 24 Frames pro Sekunde.

Im Weitverkehrsbereich sind Netze mit weniger als 100 kBit/sec Leitungen (ISDN, WIN) üblich. Nur eine kleine Zahl an Institutionen sind in Deutschland an das sogenannte Mbone (mind. 1MBit/sec) angeschlossen, das eine den LANs vergleichbare Geschwindigkeit erlaubt. Nur hier ist ATM überhaupt relevant. Die damit verbundenen Kosten sind jedoch nicht unerheblich.

Kompression im Ortsraum		
Subsampling	mäßige Kompression	schlechte Qualität
Quantisierung	mäßige Kompression	schlechte Qualität
Entropiekodierung	mäßige Kompression	verlustfrei

(a)

Kompression im Frequenzraum		
Tiefpass	mäßige Kompression	leichter Qualitätsverlust
Quantisierung	mäßige Kompression	nicht sichtbarer bis leichter Qualitätsverlust
Entropiekodierung	gute Kompression	verlustfrei
Quantisierung + Entropiekodierung	sehr gute Kompression	nicht sichtbarer bis leichter Qualitätsverlust

(b)

Tabelle 3: Kompression im Orts- und Frequenzraum

5 Datenkompression

Wie bereits mehrmals erwähnt, benötigen Rasterbilder und vor allem Bewegtbilder sehr viel Speicherplatz. Eine Stunde Video würde unkomprimiert in voller Auflösung mit 74 GByte jeden üblichen Massenspeicher sprengen. Aus diesem Grund wurden eine Vielzahl von Kompressionsverfahren für Bilddaten entwickelt. Die ersten Verfahren arbeiteten direkt mit den Rohdaten (s. Tabelle 3a).

Unter *Subsampling* versteht man dabei das Herabsetzen der Auflösung. Zum Beispiel sind beim PAL- und NTSC-Format die Farbkomponenten (CbCr bzw. UV) gegenüber der Helligkeit (Y) gröber aufgelöst (im Verhältnis 4:2:2). Bei der *Quantisierung* wird die Farbtiefe reduziert. Bis vor wenigen Jahren hatte beispielsweise eine Workstation in der Regel maximal 8 Bit Farbtiefe. Um häßliche Farbränder zu vermeiden, wurden die Bilder gedithert. Diese beiden Verfahren sind für Bilddaten nicht sonderlich effektiv, da die geringen Kompressionsfaktoren nur durch starke Qualitätseinbußen erkaufte werden. Geeigneter ist hierfür die *Entropiekodierung*, bei der man sich zunutze macht, daß benachbarte Pixel sehr ähnlich sind oder daß bestimmte Werte häufiger auftreten als andere. Zu nennen wäre da die *Runlength* Kodierung (z.B. im FAX- und PCX-Format) und die *Ziv-Lempel(-Welch)*-Kodierung (z.B. im GIF-Format). Die Entropiekodierung an sich ist verlustfrei.

Bessere Ergebnisse erzielt man, wenn die Bilddaten in den Frequenzraum transformiert werden. Mögliche Transformationen sind in Tabelle 4 aufgelistet. Die Fouriertransformation (FT) wird in der Praxis nicht für die Bilddatenkompression angewendet, da an den Bildrändern und bei scharfen Konturen wellenförmige Artefakte entstehen. Bei der Diskreten Kosinus-Transformation (DCT) sind diese Artefakte wesentlich geringer und sie hat daher mittlerweile eine weite Akzeptanz und Verbreitung (z.B. im JPEG- und MPEG-Format, s.u.) gefunden. Die Wavelet Transformation (WT) ist momentan noch ein weites

Diskrete Fouriertransformation	$F(\omega) = \sum f(x)e^{-j\frac{\omega x}{N}}$
Diskrete Kosinus-Transformation (DCT)	$F(u) = \frac{C(u)}{2} \sum f(x) \cos \frac{(2x+1)u\pi}{2N}$
Diskrete Wavelet-Transformation	$F(a, b) = a^{-\frac{1}{2}} \sum f(x)\psi\left(\frac{x-b}{a}\right)$

Tabelle 4: Transformationen in den Frequenzraum

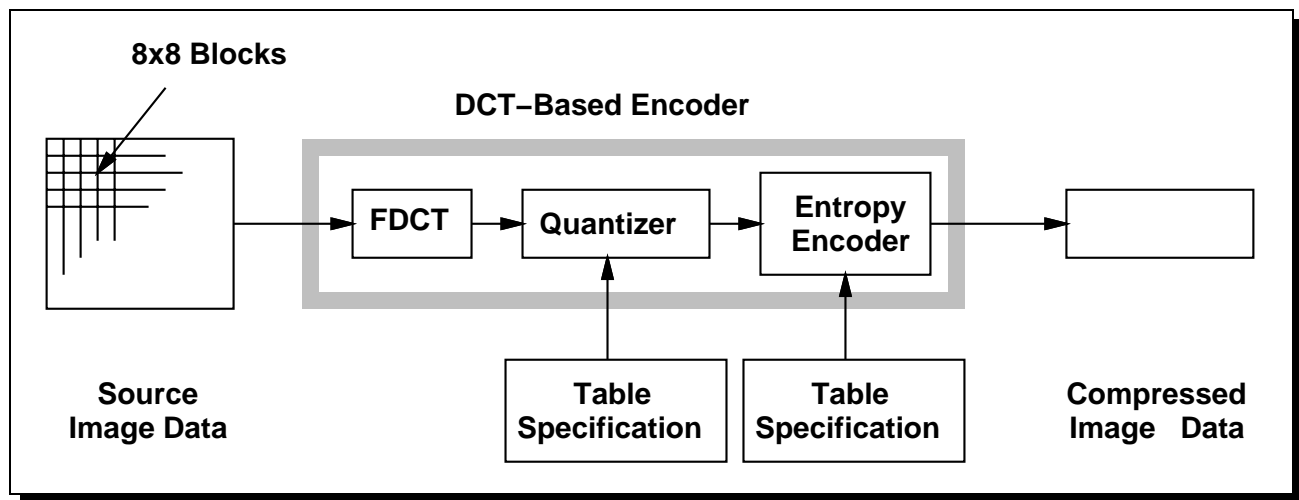


Abbildung 1: Prinzip der JPEG-Codierung [JPE92]

Forschungsgebiet, jedoch lassen sich jetzt schon bessere Kompressionsraten erzielen als bei der DCT. Daher hat die WT gute Chancen, ebenfalls in den JPEG-Standard aufgenommen zu werden.

Wendet man nun im Frequenzraum Kompressionsverfahren wie Tiefpassfilterung (analog zum Sub-sampling: hohe Frequenzen werden einfach weggeworfen) oder Quantisierung an, so erhält man (im Vergleich zum Ortsraum) bei gleichem Kompressionserfolg geringere Qualitätsverluste (siehe Tabelle 3b). Die Entropiekodierung erzielt sogar wesentlich höhere Kompressionsraten, da die Werte im Frequenzraum eine charakteristischere Verteilung besitzen als im Ortsraum. Die Kombination von Quantisierung und Entropiekodierung liefert schließlich die besten Ergebnisse.

5.1 JPEG

Die JPEG-Kodierung (JPEG = Joint Photographic Expert Group) entstand aus gemeinsamen Aktivitäten der ISO und CCITT und ist seit 1992 ein internationaler Standard [JPE92].

JPEG wurde entwickelt, um Farb- oder Grauwert-Bilder zu komprimieren. Dieses Verfahren ist für Einzelbilder gedacht, wird jedoch oft auch auf Bildsequenzen angewendet (Motion JPEG). Für Bewegtbildsequenzen ist aber das MPEG-Verfahren (siehe 5.3) besser geeignet.

Die JPEG-Kodierung ist ein sehr flexibles Format. Es unterstützt beliebige Bildformate und ist unabhängig von der Pixelform (Verhältnis Höhe:Breite). Das Bild kann bis zu vier Farbkanäle mit unterschiedlichen Auflösungen enthalten. JPEG ist jedoch „farbenblind“, d.h. die Interpretation der Kanäle (z.B. RGB oder YUV) ist Aufgabe der Anwendung bzw. des umschließenden Datenformats wie z.B. JFIF¹ oder TIFF². Die Qualität und die damit verbundene Kompressionsrate sind einstellbar: von verlustfreier Kodierung (Kompressionsrate ca. 2:1) bis zu sichtbaren Qualitätsverlusten (Kompressionsrate 20:1 und mehr). Bei der verlustfreien Kodierung wird ein Prädiktionsverfahren mit anschließender Entropiekodierung verwendet. Die verlustbehaftete Kodierung benutzt anstatt des Prädiktionsverfahrens eine Diskrete Kosinus-Transformation (DCT) mit anschließender Quantisierung.

JPEG-Modi

JPEG ist eigentlich nicht *ein* einziges Kompressionsverfahren, sondern es umfaßt eine ganze Reihe von sogenannten Kompressionsmodi:

- verlustbehaftet, sequentiell DCT-Komprimiert (Baseline Mode):

Dies ist die einfachste DCT-Kodierung und sie reicht für die meisten Anwendungen aus. Dieser

¹JFIF = JPEG File Interchange Format

²TIFF ist ein Tag-Basiertes Fileformat von Aldus Corporation für Speicherung und Austausch von Rasterbildern

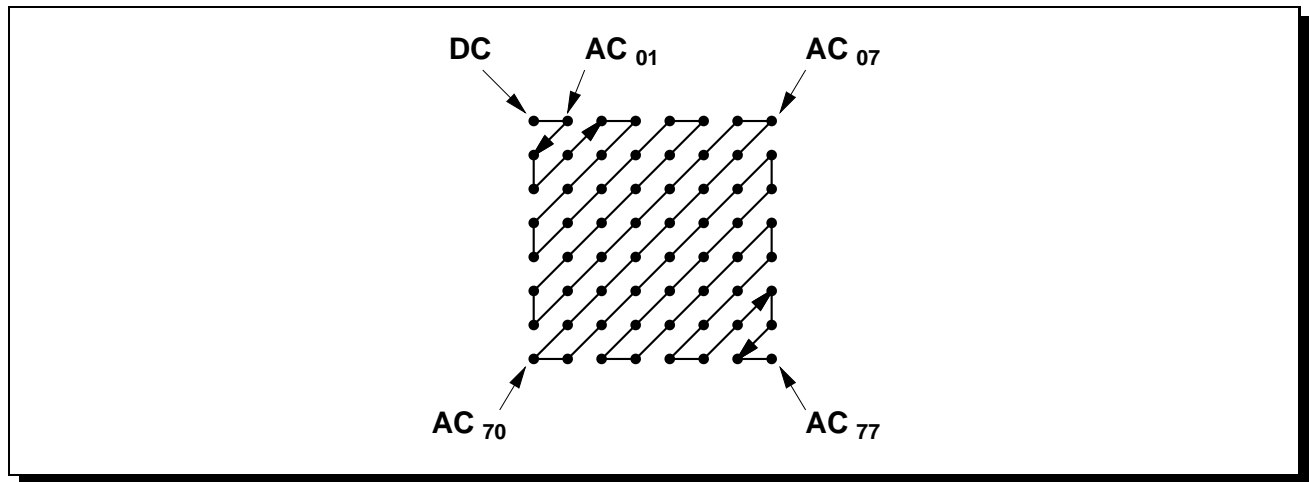


Abbildung 2: Zickzack-Kurs durch einen 8x8 Block [JPE92]

Modus muß von jedem JPEG Dekoder unterstützt werden. Die folgenden Abschnitte beziehen sich in erster Linie auf diesen Modus.

- verlustbehaftet, DCT-Komprimiert (erweiterter Modus):
Dieser Modus stellt weitere Alternativen zur Verfügung. So werden zum Beispiel eine höhere Farbtiefe des Quellbildes und vier anstatt zwei Quantisierungstabellen (s.u.) unterstützt. Weiterhin kann das Bild anstatt sequentiell auch progressiv abgelegt werden. Dabei wird zuerst ein grobes Bild kodiert, das dann sukzessive verfeinert wird.
- verlustfreie Kodierung:
Bei diesem Modus ist eine fehlerfreie Rekonstruktion möglich, was jedoch durch einen geringeren Kompressionsfaktor erkauft wird.
- hierarchische Kodierung:
Das Bild wird in verschiedenen Auflösungen abgelegt. Zuerst wird das Bild in einer sehr groben Auflösung kodiert, die Stufe für Stufe erhöht wird. Bei jedem Schritt werden nur die Differenzen zur nächst niedrigeren Auflösung weiterverarbeitet.

Der Baseline Modus

Bei der JPEG-Kompression wird, wie in Abbildung 1 gezeigt, verfahren.

i. Blockbildung Die einzelnen Farbkkanäle (z.B. Y, Cb und Cr) des Bildes werden in 8x8 Blöcke aufgeteilt, um den Aufwand der nachfolgenden Kosinus-Transformation zu beschränken. Nachteil dieser Vorgehensweise ist, daß bei starker (nachfolgender) Quantisierung häßliche Farbsprünge an den Blockgrenzen auftreten können.

ii. Transformation (DCT) Auf jeden 8x8-Block wird eine Kosinus-Transformation angewendet. Die eindimensionale DCT $F(u) = \frac{C(u)}{2} \left[\sum_{x=0}^7 f(x) \cos \frac{(2x+1)u\pi}{16} \right]$ wird auf jede Zeile und anschließend auf jede Spalte angewandt. Als geschlossene Formel erhält man:

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right]$$

wobei

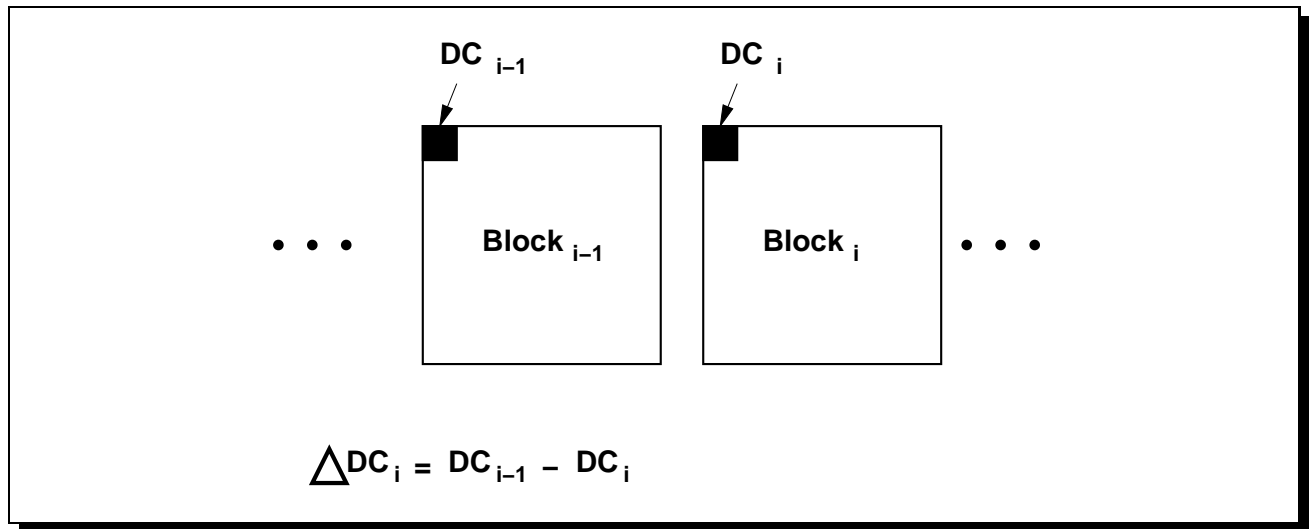


Abbildung 3: Differenz zwischen zwei aufeinanderfolgenden DC-Werten [JPE92]

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{wenn } u, v = 0 \\ 1 & \text{sonst} \end{cases}$$

$F(0,0)$ enthält den Blockmittelwert des 8×8 Blockes (DC-Komponente, Gleichstromanteil), während die anderen Koeffizienten die Amplituden der Frequenzen (AC-Komponenten, Wechselstromanteile) enthalten.

iii. Quantisierung Bei der Quantisierung hat jeder Koeffizient der Transformaten einen eigenen Quantisierungsfaktor. Diese werden in einer 8×8 Matrix, der sogenannten Quantisierungstabelle, abgelegt.

$$F^Q(u, v) = \text{Integer} \left(\frac{F(u, v)}{Q(u, v)} \right)$$

Im Basis-Modus können zwei, im erweiterten Modus vier derartige Quantisierungstabellen angegeben werden, zwischen denen hin- und hergeschaltet werden kann. Daher ist es möglich den einzelnen Farbkanälen verschiedene Quantisierungstabellen zuzuordnen.

iv. Entropiekodierung Die Koeffizienten werden, wie in Abbildung 2 dargestellt, im Zickzack-Pfad durchlaufen und damit sequentialisiert. Bei der Kodierung wird zwischen den DC- und den AC-Komponenten unterschieden:

DC-Komponenten Es werden nicht die Werte selbst, sondern nur die jeweilige Differenz zum vorhergehenden DC-Wert in einer Huffman-Kodierung abgelegt (Abbildung 3).

AC-Komponenten Für die AC-Komponenten wird eine Kombination aus Huffman- und Lauflängenkodierung verwendet. Die Folge von AC-Komponenten enthält sehr viele Nullen, da bei der Quantisierung niedrige Werte auf Null abgebildet werden. Daher ist es sinnvoll, nur die Werte ungleich Null zu kodieren, wobei angegeben werden muß, wieviele Nullen sich zwischen den einzelnen Werten befinden:

`<Anzahl Nullen1 / Wert1 / Anzahl Nullen2 / Wert2 / ...>`

Die Werte selbst können wiederum als Tupel (n, code_n) kodiert werden, wobei n die Anzahl der benötigten Bits und code_n die Kodierung des Wertes mit n Bit darstellt. Wir haben nun die Folge:

`<Anzahl Nullen1 / n1 / Code1,n / Anzahl Nullen2 / n2 / Code2,n / ...>`

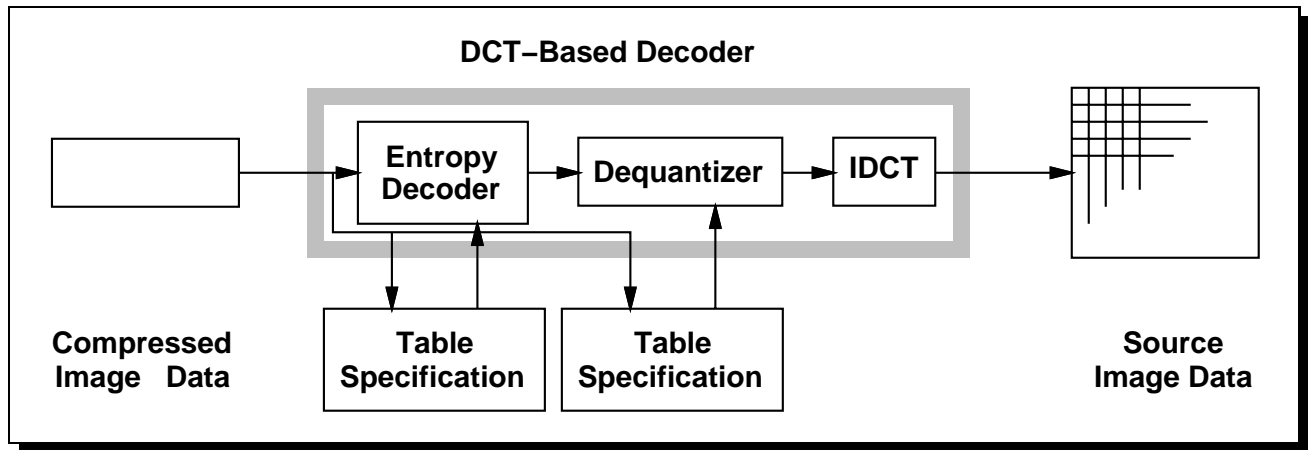


Abbildung 4: Prinzip der JPEG-Dekodierung [JPE92]

Sowohl die Anzahl der Nullen als auch die Anzahl der benötigten Bits (n) unterliegen einer statistischen Verteilung und werden daher als Tupel huffmankodiert³:

$\langle \text{HuffCode}(\text{Anz. Nullen}_1, n_1) / \text{Code}_{1,n} / \text{HuffCode}(\text{Anz. Nullen}_2, n_2) / \text{Code}_{2,n} / \dots \rangle$

Dekompression

JPEG ist symmetrisch. Das bedeutet, daß die Kodierung und die Dekodierung ungefähr den gleichen Aufwand haben. Zuerst muß der Kode mit variabler Länge entschlüsselt werden. Das geschieht mit Hilfe der Tabellen die bei der Entropiekodierung benutzt wurden. Anschließend werden die Daten mit der Quantisierungsmatrix multipliziert und letztendlich mittels der Inversen Kosinus-Transformation (IDCT) zurücktransformiert. Diese ist vollkommen analog zur DCT:

$$f(x, y) = \frac{1}{4} \left[\sum_{u=0}^7 \sum_{v=0}^7 C(u)C(v)F(u, v) \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \right]$$

Aufwand

Das Format 767 x 575 kann in 6912 8x8 Blöcke⁴ aufgeteilt werden. Bei einem Farbbild mit dem Abtastverhältnis von 4:2:2 ergibt das 6912 Luminanz- und 6912 Chrominanzblöcke (3456 Cb- und 3456 Cr-Blöcke).

Bei einem trivialen Ansatz, bei dem jeder Bruch und jeder Kosinuswert „just in time“ berechnet werden, benötigt man pro 8x8 Block 32896 Multiplikationen, 12224 Additionen, 8256 Shiftoperationen und 8192 Cosinusberechnungen. Wenn vorab nach $C(a, b) = \cos \frac{(2a+1)b\pi}{16}$ eine Tabelle berechnet wird, reduziert sich der Aufwand auf 8320 Multiplikationen, 4032 Additionen und 64 Shiftoperationen pro 8x8 Block.

Da die DCT der Fouriertransformation sehr ähnlich ist, kann eine Technik angewandt werden, die der Fast Fourier Transformation (FFT) entspricht. Bei einem Algorithmus nach M. Vetterli (wie in [Wal91] beschrieben) ist die Anzahl der nötigen Operationen:

462 Additionen, 54 Multiplikationen und 6 Shifts.

Pro Bild wären das dann 6.386.688 Additionen, 746.496 Multiplikationen und 82.944 Shift-Operationen. Bei 25 Bildern pro Sekunde ergibt das insgesamt 180 Millionen Ganzzahloperationen pro Sekunde für die DCT.

³Bei der Kodierung nach Huffman werden Kodeworte unterschiedlicher Länge verwendet. Die Länge hängt von der zu erwartenden Häufigkeit des Wortes ab.

⁴Gegebenenfalls (wie in diesem Fall) werden die nicht vollständig gefüllten Blöcke beliebig aufgefüllt.

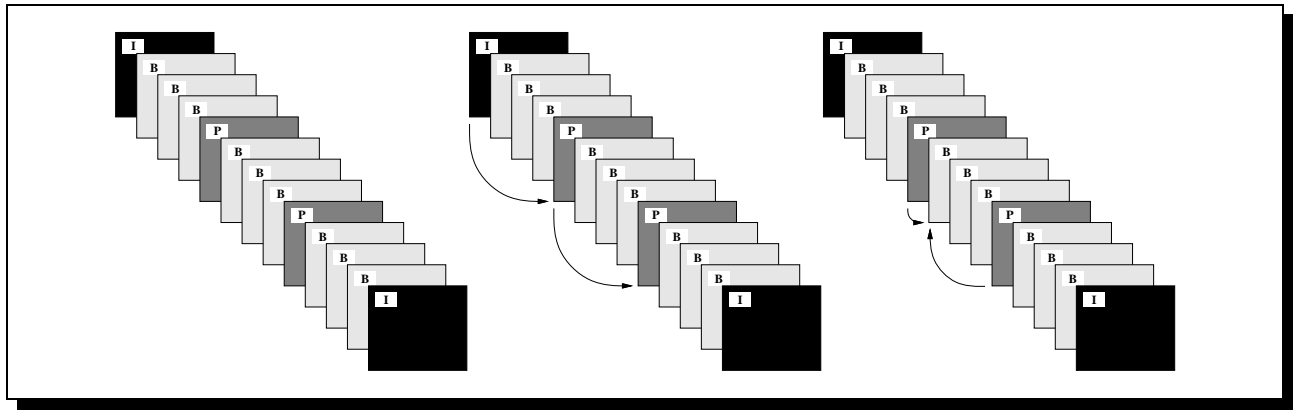


Abbildung 5: I-, P- und B-Frames bei MPEG

5.2 H.261 (px64)

H.261 ist eine CCITT Empfehlung zur Komprimierung von Bewegtbildern in Echtzeit für Schmalband-Übertragungsnetze. Entsprechend der vorhandenen Übertragungskapazität und der geforderten Bildqualität werden in H.261 mehrere Datenraten ($p \times 64$ KBit/s $p=1,2,\dots,30$) definiert. Es verwendet standardmäßig das Format QCIF⁵ mit 176x144 Pixel für die Luminanz- und 88x72 Pixel für die Chrominanzkomponente. Optional kann auch das Format CIF⁶ mit doppelter Auflösung verwendet werden (352x288 für Luminanz, 176x144 für Chromnanz).

Unter H.261 kann ein Bild auf zwei verschiedene Weisen kodiert werden — entweder als sogenannte Intra- oder als Interframes:

Intraframe Bei Intraframes wird wie bei JPEG jeder 8x8 Block mittels DCT transformiert, quantisiert und anschließend entropiekodiert.

InterFrame Hier wird für jedem Makroblock⁷ ein Bereich im vorhergegangenen Bild gesucht, das ihm am ähnlichsten ist. Die Position dieses Bereiches wird relativ zur Position des Makroblockes mittels eines Bewegungsvektors⁸ bestimmt. Schließlich wird die Differenz zwischen dem Makroblock und dem ermittelten Bereich transformiert, quantisiert und entropiekodiert, wenn ein gewisser Schwellwert überschritten ist, ansonsten wird nur der Bewegungsvektor weitergegeben.

Die Quantisierung ist rückgekoppelt: die Schrittweite der Quantisierung hängt von dem Füllungsgrad eines Übertragungsspeichers ab. Hierdurch wird eine konstante Datenrate gewährleistet.

5.3 MPEG

MPEG wird in der ISO/ICE JTC1/SC29/WG11 entwickelt. Es definiert Algorithmen zur Kodierung von Bildsequenzen mit Audio. Ähnlich wie in H.261 ist der Bildaufbau genau festgelegt: So enthält jedes Bild drei Kanäle: die Luminanzkomponente (Y) und zwei Chrominanzkomponenten (Cb und Cr). Die Luminanzkomponente hat horizontal und vertikal die doppelte Auflösung wie die Chrominanzkomponenten (4:2:0). Das Format 768x576 darf (bei MPEG-1) nicht überschritten werden. Weiterhin werden acht verschiedene Bildwechselfrequenzen (23,976 Hz, 24 Hz, 25 Hz, ... , 60 Hz) und 14 verschiedene Seitenverhältnisse eines Pixels festgelegt. Die maximale Datenrate ist auf 1,5 MBit/s begrenzt.

Bei MPEG können die Einzelbilder auf vier verschiedene Arten kodiert werden:

⁵QCIF = Quarter Common Intermediate Format

⁶CIF = Common Intermediate Format

⁷Ein Makroblock besteht aus vier 8x8 Luminanz- und zwei 8x8 Chrominanzblöcken

⁸Die Berechnung von Bewegungsvektoren ist optional. Die Bewegungsvektoren können daher alle gleich dem Nullvektor sein

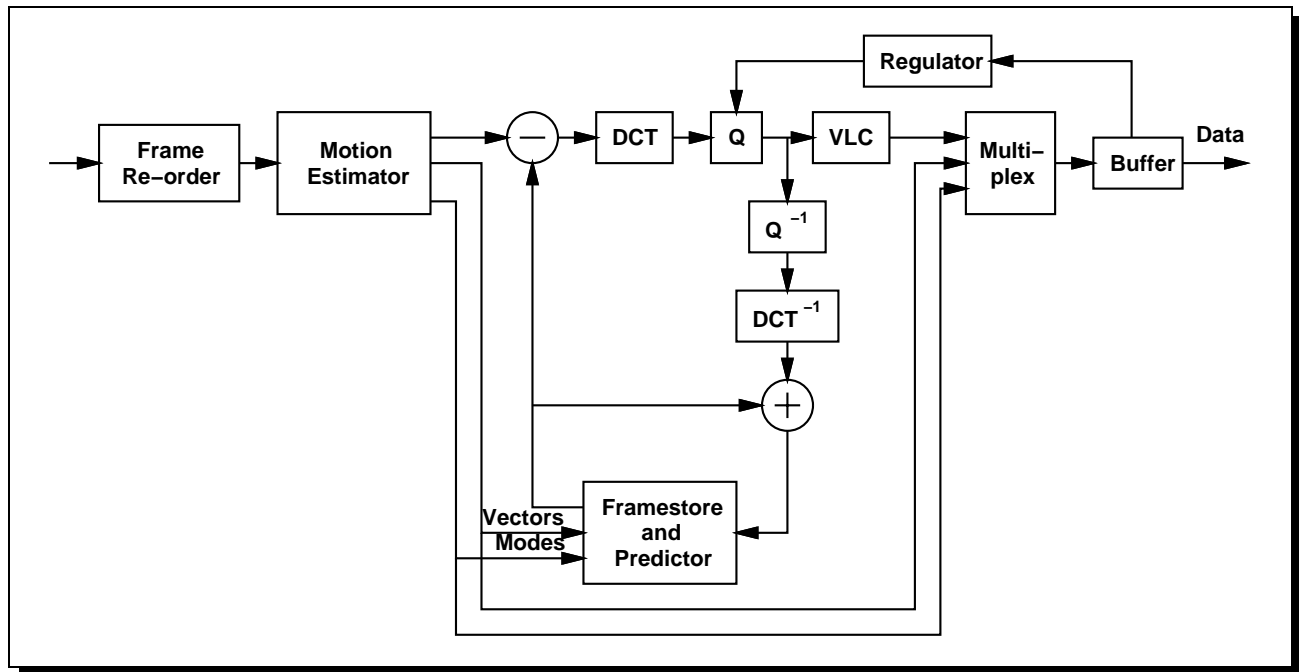


Abbildung 6: Prinzip der MPEG-Kodierung [MPE91]

I-Frame Die I-Frames entsprechen den Intraframes von H.261. Auch hier werden die 8x8 Blöcke wie in JPEG DCT-transformiert, quantisiert und entropiekodiert. Sie haben die geringste Kompressionsrate, jedoch sollte mindestens jedes 15-te Bild als I-Frame kodiert sein.

P-Frame Die P-Frames sind den Interframes von H.261 sehr ähnlich, jedoch beziehen sie sich nicht auf das unmittelbar vorhergehende Bild, sondern auf das letzte I- oder P-Frame (Abbildung 5b). In der Regel befinden sich dazwischen einige B-Frames (s.u.).

B-Frame Bei den B-Frames werden Prädiktionen in zwei Richtungen vorgenommen, d.h. die Bewegungsvektoren beziehen sich sowohl auf das letzte wie auch auf das nächste I- bzw P-Frame (Abbildung 5c).

D-Frame Diese Kodierung wird für das schnelle Durchlaufen von Bildsequenzen benötigt. Hierbei werden nur die DC-Komponenten der DCT kodiert. Diese Art der Kodierung wird jedoch selten verwendet.

Auch bei MPEG wird die Quantisierung durch den Füllungsgrad eines Übertragungspuffers reguliert, um eine gleichbleibende Datenrate zu erhalten (Abbildung 6).

Aufwand

Bei der Dekodierung von MPEG-Sequenzen (Abbildung 7) ist der Aufwand vergleichbar mit der Dekodierung von JPEG-Bildfolgen, und benötigt daher ungefähr 180 Millionen Ganzzahloperationen pro Sekunde für die DCT bei einer Framegröße von 767 x 575 Pixel und 25 Bilder/sec. Hinzu kommt der Aufwand für den Entropiedekoder und die Dequantisierung.

Dagegen liegt bei der MPEG-Kodierung (Abbildung 6) der größte Aufwand bei der Suche der Motion-Vektoren. Die Effektivität der Kompression hängt stark davon ab, ob gute Motion-Vektoren gefunden werden oder nicht. Daher werden die besten Kompressionsfaktoren nur bei der sehr aufwendigen vollständigen Suche zu erreichen sein.

Für eine vollständigen Suche benötigt man jedoch pro P-Frame 1.100 Milliarden Additionen, bei B-Frames das Doppelte. Kodiert man beispielsweise eine Sekunde mit der Kodierreihenfolge „IBBPBB-

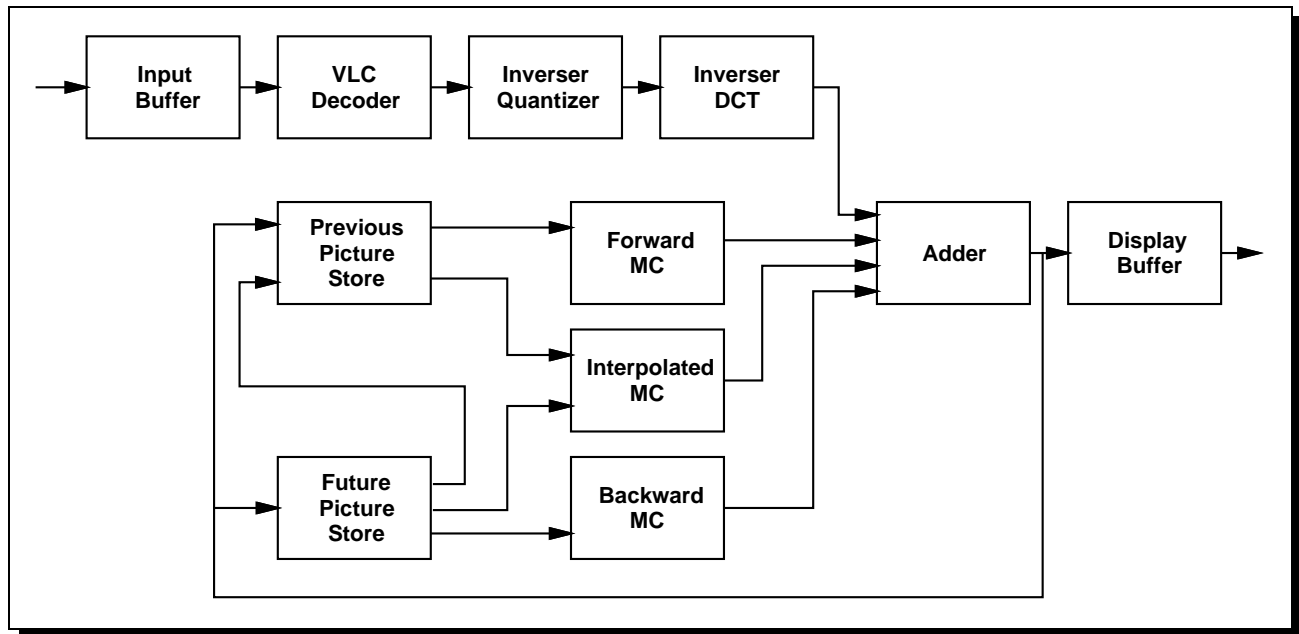


Abbildung 7: Prinzip der MPEG-Dekodierung [MPE91]

„PBBIBBPBPPBBIBBPBPF“ (3 I-Frames, 6 P-Frames und 16 B-Frames) ergäbe das allein für die Suche der Motion-Vektoren 42.400 Milliarden Ganzzahloperationen.

Daher benutzen viele MPEG-Encoder Heuristiken und Einschränkungen bei dieser Suche. So kann zum Beispiel der Suchbereich eingeschränkt werden. Bei dem zum MPEG verwandten Verfahren H.261 wird der Suchbereich auf +/-15 Pixel eingeschränkt. Der Suchaufwand sinkt hierbei auf 2,7 Milliarden Ganzzahloperationen (nur noch 0,15%) pro P-Frame. Andere Encoder benutzen Suchheuristiken wie die Gradientensuche.

Audiokodierung

Das Audiosignal wird in MPEG mittels der schnellen Fouriertransformation (FFT) in 32 sich nicht überlappende Frequenzbänder zerlegt und mit Hilfe eines psychoakustischen Modells quantisiert. Im Anschluß kann eine Entropiekodierung nach Huffman durchgeführt werden (Abbildung 8).

Videostrom

Aufgrund der bidirektionalen Prädiktionen bei B-Frames können die Frames nicht in ihrer natürlichen Reihenfolge bearbeitet und abgelegt werden. Sie können erst dann bearbeitet werden, wenn das nachfolgende P- bzw. I-Frame abgearbeitet wurde. So ergibt die Beispielfolge aus [MPE91]

„I₀ B₁ B₂ P₃ B₄ B₅ P₆ B₇ B₈ P₉ B₁₀ B₁₁ I₁₂ B₁₃ B₁₄ P₁₅“ folgende Bearbeitungsreihenfolge:
 „I₀ P₃ B₁ B₂ P₆ B₄ B₅ P₉ B₇ B₈ I₁₂ B₁₀ B₁₁ P₁₅ B₁₃ B₁₄“

MPEG-2

Die maximale Datenrate wurde auf 4 - 10 MBit/s angehoben. Es werden die volle PAL-, NTSC- und HDTV-Auflösungen (bis 1920 x 1080 x 30) und das Zeilensprungverfahren unterstützt.

MPEG-4

MPEG-4 definiert die Kompression von Bildsequenzen in geringer Qualität und sehr niedrigen Bitraten. Die Bildgröße wurde auf 176 x 144 reduziert bei 10 Bildern/sec. Die Bitrate liegt bei 4800 bis 64.000 bit/sec und ist daher besonders für Video-Telephone (z.B. über ISDN) geeignet.

Bild	Kompr.faktor	QFaktor	subjektive visuelle Qualität
Schnecke	$\frac{1}{10}$	6	vom Original nicht unterscheidbar
	$\frac{1}{20}$	18	vom Original nicht unterscheidbar
	$\frac{1}{50}$	66	kaum sichtbare Artefakte
	$\frac{1}{100}$	416	starke Veränderungen
Kugeln	$\frac{1}{10}$	24	kaum sichtbare Veränderungen
	$\frac{1}{20}$	88	sichtbare Artefakte an Rändern von gleichmäßigen Flächen
	$\frac{1}{50}$	555	Artefakte im ganzen Bild sichtbar
	$\frac{1}{100}$	2500	sehr starke Veränderungen
Baum	$\frac{1}{10}$	26	vom Original nicht unterscheidbar
	$\frac{1}{20}$	84	leichte Artefakte in kontrastreichen Regionen
	$\frac{1}{50}$	357	sichtbare Veränderungen bei gleichmäßigen Flächen (Himmel)
	$\frac{1}{100}$	1000	sehr starke Veränderungen
Kirche	$\frac{1}{10}$	10	vom Original nicht unterscheidbar
	$\frac{1}{20}$	26	kaum sichtbare Artefakte
	$\frac{1}{50}$	132	sichtbare Veränderungen bei gleichmäßigen Flächen (Himmel) und in kontrastreichen Regionen
	$\frac{1}{100}$	714	sehr starke Veränderungen

Tabelle 5: Bildqualität nach verlustbehafteter JPEG-Kompression

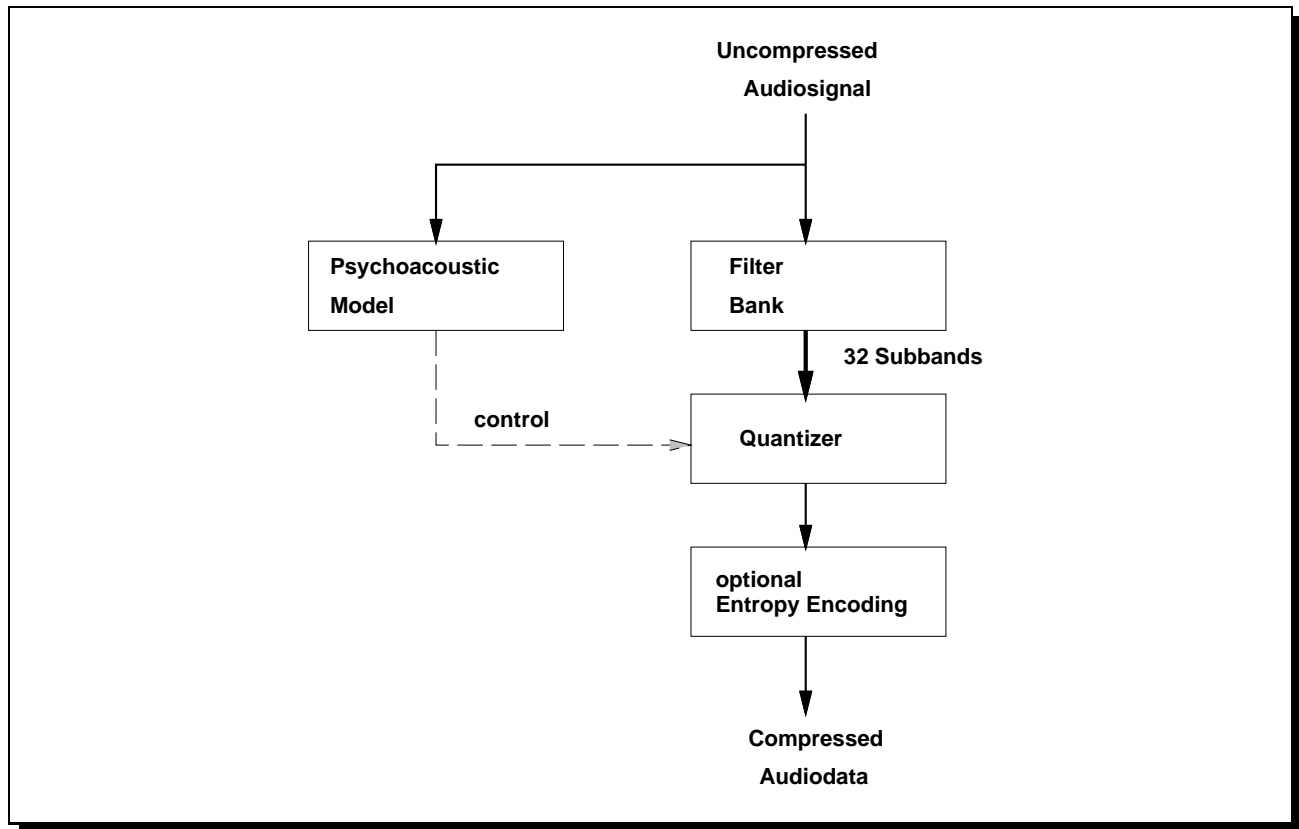


Abbildung 8: Audiokodierung bei MPEG [MPE91]

JPEG	$\frac{1}{20}$
Kompression mit Wavelet Transformation	$\frac{1}{50}$
MPEG (Bildsequenz)	$\frac{1}{90}$

Tabelle 6: Erreichbare Kompressionsfaktoren pro Bild bei „guter“ Bildqualität

5.4 Kompression und Qualität

Die Qualität des Bildmaterials nach verlustbehafteter Kompression ist sowohl vom Bildinhalt als auch von der Stärke der Kompression abhängig. In Tabelle 5 befindet sich eine Gegenüberstellung von Kompressionsfaktor und subjektiver Bildqualität. Die Bilder „Schnecke“ und „Kugeln“ wurden mittels Raytracing von einem Computer generiert. Die beiden anderen Bilder sind Fotografien eines Baumes und einer Kirche. Das Bild des Baumes befindet sich, zusammen mit einigen Ausschnittvergrößerungen, im Anhang.

In Tabelle 6 werden für einige Kompressionsverfahren die erreichbaren Kompressionsfaktoren bei „guter“ Bildqualität aufgelistet. „Gute“ meint hier, daß Veränderungen und Artefakte nur durch genaues Hinsehen aufgefunden werden.

6 Transport von Bewegtbildern im LAN

Der Transport von Bewegtbildern hat sich als eine der Operationen herausgestellt, die besonders stark durch technische Einschränkungen betroffen sind. Es stellt sich die Frage, inwieweit die Kompression der Daten an dieser Problematik etwas ändert. Wir sind ihr im LAN-Bereich nachgegangen und haben versucht, durch Messungen die Grenzen zu ergründen.

Motivation und Anforderungen

Die Motivation für Bewegtbild-Transport im LAN ist dadurch gegeben, daß beispielsweise in Fernsehstationen oder Redaktionen folgende Situation vorliegt: Es gibt ein zentrales oder verteiltes Bewegtbildarchiv und dezentrale Arbeitsplätze zum Anzeigen, Editieren und Fusionieren von Bewegtbilddaten (evtl. mit anderen Medien). Geht man von einem herkömmlichen Ethernet-LAN aus, so stellt sich die Frage, inwiefern ein ausreichend schneller Transport der Daten gewährleistet werden kann.

Ausreichend wäre ein Transport der Video-Format-Daten in Echtzeit anliefern kann. Das bedeutet 25 Frames pro Sekunde. Durch Puffern der Daten an der Senke kann man für kurze Sequenzen mit weniger Leistung auskommen. Man muß davon ausgehen, daß für längere Filme die Echtzeit-Bedingung einzuhalten ist, wenn man nicht Frames ausläßt oder kleinere Formate wählt.

Software-Architektur

Aus verschiedenen Gründen erscheint eine Realisierung des Gesamtsystems als Client/Server-Konfiguration als sinnvoll. Zum einen kontrolliert dann ein Server den Zugriff auf langfristig archivierte Videodaten, und zum anderen ist dies der Ort, wo erheblicher Massenspeicher (Terabyte-Bereich) konzentriert ist, der nicht einfach dupliziert werden kann.

Arten des Transports

Versucht man, die Daten simultan zum Transport darzustellen, so braucht man einen sogenannten *isochronen Transport*. Das bedeutet, daß jeder Frame eine exakt definierte – und für alle Frames gleiche – Verweilzeit im Netz hat. Dies ist nicht immer erforderlich. Einige Frames lassen sich im allgemeinen lokal puffern, so daß *synchroner Transport* ausreicht. Synchroner Transport garantiert die maximale Verweilzeit eines Frames. Handelt es sich um geringe Datenmengen oder liegt keine Dringlichkeit vor, so kann man auch *asynchronen Transport* benutzen. Hierbei werden keine Garantien über die Verweilzeit abgegeben.

6.1 Technische Randbedingungen

Ein Ethernet-LAN ist charakterisiert durch folgende Eigenschaften, die sich zunächst auf Schicht 1 und 2 des ISO/OSI-Referenzmodells beziehen:

- eine Bus-Topologie und die daraus resultierende flexible Anzahl der angeschalteten Client-Rechnern,
- eine beschränkte Übertragungsrate von maximal 10 MBit/sec,
- ein Zugriffsprotokoll per CSMA/CD, das keinen Zugriff innerhalb einer bestimmten Zeit garantiert,
- und einen Kollisions-Overhead im Falle von Hochlastbetrieb.

Auf den höheren OSI-Schichten sind verschiedene Realisierungen anzutreffen. Im wesentlichen unterscheidet man TCP/IP und UDP. Die maßgeblichen Unterscheidungsmerkmale sind in Tabelle 7 zusammengetragen. Die anschließenden Messungen wurden mithilfe einer verbreiteten TCP/IP Implementierung durchgeführt.

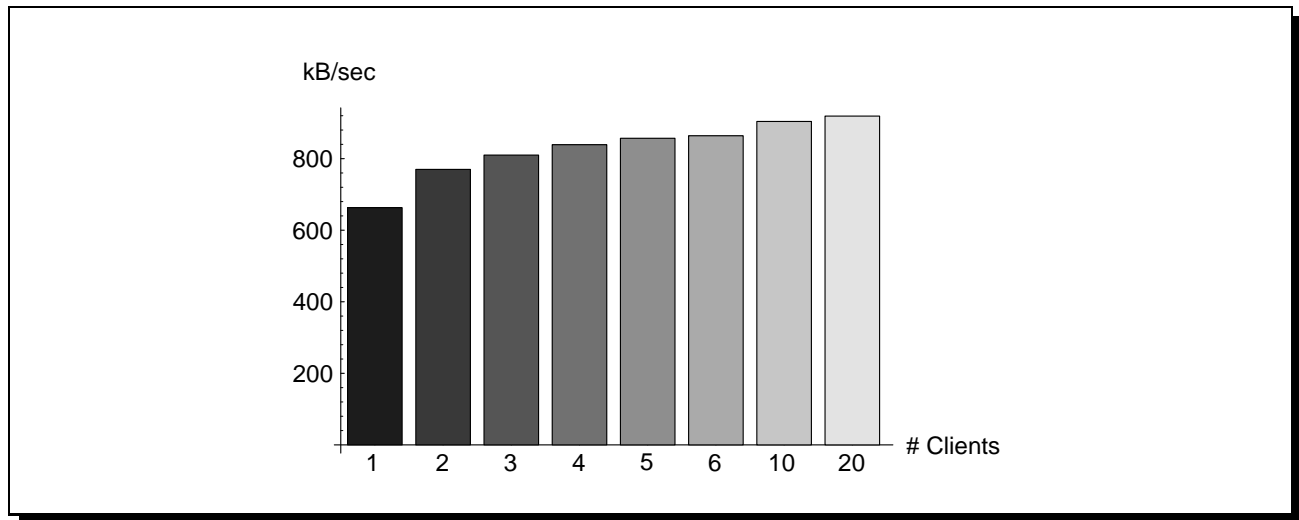


Abbildung 9: Datendurchsatz am Server bei n Clients in kB

Protokolleigenschaften	UDP	TCP
verbindungsorientiert	nein	ja
maximale Größe von Einheiten	ja	nein
timeout/retransmission	nein	ja
duplicate detection	nein	ja
Sequenzüberwachung	nein	ja
Flußkontrolle	nein	ja

Tabelle 7: Wesentliche Unterscheidungsmerkmale der Schicht-4-Dienste UDP und TCP

6.2 Messungen

Versuchsanordnung

Bei den folgenden Messungen wurde von einer variablen Anzahl von sogenannten Video-Clients und genau einem Video-Server ausgegangen. Diese Situation ist in Abbildung 10 skizziert. Der Server besorgt Video-Sequenzen auf Anforderung, indem er seine Speichermedien durchsucht, die Daten auf das Netz bringt und sie damit den Clients zur Verfügung stellt. Es wurde folgende Werte gemessen:

Wieviele Clients bekommen wieviele Frames pro Sekunde? Wieviele Frames kann der Server maximal in einer Sekunde austeilen? Welchen Einfluß hat die Qualität, also die Kompressionsrate der Bilder, auf diese Zahlen? Und schließlich, was bringt der Übergang von RGB-Frames zu JPEG-Frames bzw. zu MPEG in dieser Beziehung?

Zum Messen wurden drei verschiedene Videosequenzen benutzt, deren charakteristische Merkmale in Tabelle 8 aufgeführt sind. Der Quantisierungsfaktor ist dabei ein Qualitätsmerkmal der JPEG-Transformation. Je größer der Wert, desto schlechter die Bildqualität und desto besser die Kompression. Mit dem Standard-Wert 100 versehene Bilder sind vom Original nur sehr selten unterscheidbar. Bei einem QF von 400 sind zwar bereits deutliche Unterschiede gegenüber dem Original zu verzeichnen; das Bild ist jedoch ohne Probleme zu erkennen.

Zu bemerken ist, daß der Verarbeitungsaufwand im einzelnen Client nicht zu Buche schlägt, da hierzu einerseits Spezialhardware zur Verfügung stand und der restliche Aufwand andererseits vernachlässigbar ist. Eine weitere Randbedingung war, daß Netzknoten, die keine Video-Clients waren, nur minimale zusätzliche Last erzeugen.

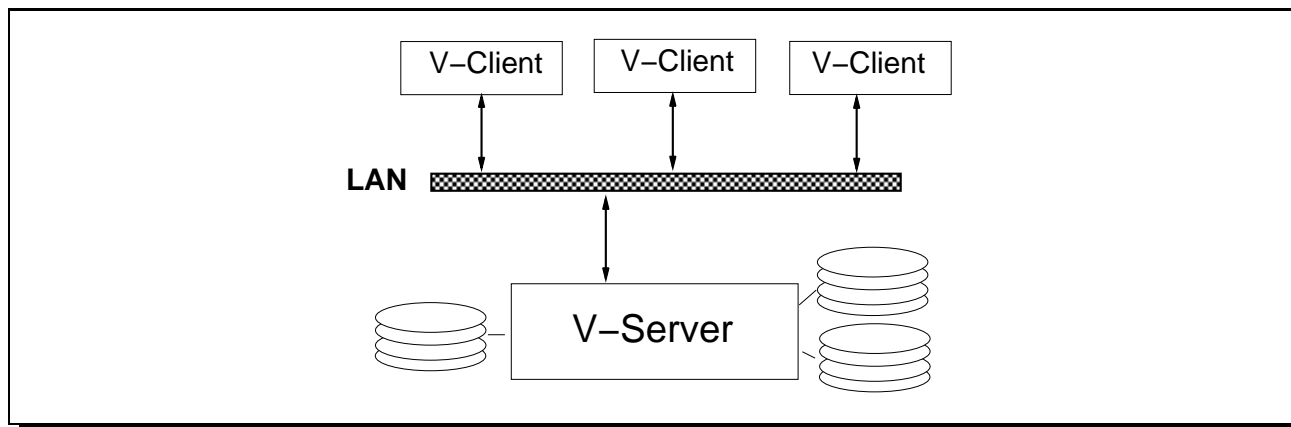


Abbildung 10: Meßanordnung

Video-Sequenz	QF	Frames	Datenmenge (KB)	KB/Frame
1. Kamera	100	596	39 050	65
2. Occursus	100	598	24 379	41
3. Kamera	400	600	13 279	22

Tabelle 8: Video-Sequenzen für die Leistungstests (zum Quantisierungsfaktor QF siehe Text)

Leistungsanalyse aus der Sicht des Servers

Anhand von Abbildung 9 ist folgendes zu beobachten: Mit zunehmender Zahl der Video-Clients steigt der Datendurchsatz im gesamten Netz nur marginal. Bei 20 Video-Clients erreichen wir mit über 800 kB/sec (< 7 Mbit/sec) fast den Grenzbereich des Netzes. Die Messungen in kB sollen zunächst einmal die Kapazität unabhängig von Video-Formaten und Kodierung demonstrieren.

In Abbildung 11 ist die Abhängigkeit der transportierten Datenmenge von der Video-Sequenz und dementsprechend auch von der Kodierung und Kompression abgetragen; daher auch die Angabe in Frames pro Sekunde. Deutlich ist der Unterschied zwischen der stärker (3) und der weniger stark komprimierten Version (1) einer Sequenz zu erkennen. Man erreicht den dreifachen Durchsatz (fast 30 Frames pro Sekunde bei einem Client) bei einer Kompression, die fast um einen Faktor vier besser ist und bei einem eher geringen Qualitätsverlust.

Leistungsanalyse aus der Sicht der Clients

Aus der Sicht eines einzelnen Clients ist interessant, inwieweit sich die Auslieferung eines Frames maximal verzögert, oder, anders ausgedrückt, wie nahe an eine isochrone Übertragung man kommen kann.

Aus den bisher gemachten Messungen heraus gehen wir von einer Framerate von 10 Frames pro Sekunde (komprimiert) aus. Wir wollen sehen, ob mit dieser Rate Echtzeit zu erreichen ist. Abbildung 12 zeigt eine Folge von 100 Frames und ihre zeitliche Abweichung (in Sekunden) von der optimalen Auslieferungszeit. Hierzu ist nach oben die Zeit zwischen zwei empfangenen Frames aufgetragen. Der Nullpunkt der Zeitachse stellt die isochrone Auslieferungszeit dar. Alle Frame-Pakete haben eine kleine zeitliche Ausdehnung (durch einen senkrechten Strich angedeutet), solche, die ihr Zentrum unterhalb der x -Achse haben, können noch als synchron akzeptiert werden.

Es ist deutlich zu sehen, daß die meisten Frames die Echtzeitbedingung einhalten und gepuffert werden müssen; dies gilt vor allem nach einer kurzen Anlaufphase des Gesamtsystems. Allerdings werden einige auch deutlich zu spät ausgeliefert. Aus der Anzahl der Frame-Pakete, die zu spät oder zu

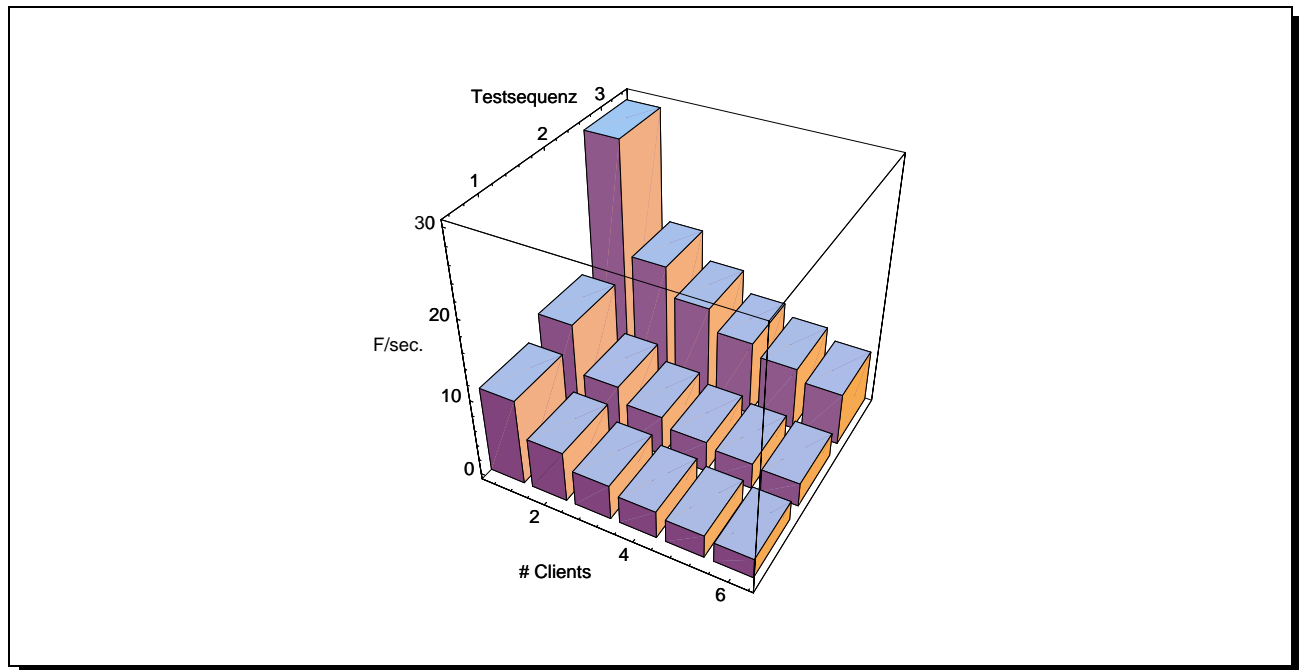


Abbildung 11: Leistungsanalyse in Frames pro Sekunde (F/sec) bei verschiedenen Video-Sequenzen

früh ankommen läßt sich die Größe eines Ringpuffers errechnen, der eingesetzt werden muß, um an der Benutzungsschnittstelle Echtzeit zu simulieren.

Mit der Anzahl der Verbindungen (Anzahl der Clients), die bedient werden müssen, sinkt naturgemäß die Anzahl der Frames, die pro Client ausgeliefert werden können. In welchem Maße das der Fall ist, dazu gibt Abbildung 13 Auskunft. Die Gesamtnetzlast erhöht sich zwar mit jedem Client noch ein wenig, übersteigt aber nicht die bereits gemessene Grenze von 7 MBit/sec.

Die Rolle der Formate

Ob eine Bewegtbildsequenz als einzelne RGB-Frames, als JPEG-baseline-komprimierte Frames oder als MPEG-Movie vorliegt, hat gravierende Auswirkungen auf das Datenvolumen, das diese Sequenz einnimmt.

In Abbildung 14 ist zu sehen, wie sich in PAL-Auflösung die drei Kodierungsformen zueinander verhalten. Der Schritt von RGB nach JPEG bringt einen Faktor 20 an Platz- und letztendlich an Zeitbedarf. Reduziert man die Anzahl der Frames auf 10 F/sec so ist man bei JPEG immer noch um einen Faktor 3 schlechter als eine MPEG-1-Kodierung (ganz unten – nicht exakt dieselbe Frame-Größe) der gleichen Sequenz. In der Mitte ist gezeigt, daß JPEG mit 10 F/sec und CIF-Format vom Datenvolumen her ziemlich genau der MPEG-1 (volles Format, 25 F/sec) Kodierung entspricht.

Aus dem Gesagten kann gefolgert werden, daß bereits in heutigen LANs unter Einsatz von MPEG ein Transport von Bewegtbilddaten (in PAL-Größe, 25 F/sec) annähernd in Echtzeit möglich ist.

6.3 Fazit

Folgende Faustregeln bringen die Messungen auf den Punkt:

- Ohne Qualitätsverlust ist kein dringlicher Transport über längere Zeiträume in heutigen LANs möglich.
- Mit JPEG-baseline-kodierten Bildfolgen erreicht man Echtzeit nur bei einer zeitlichen Reduktion auf 10 Frames pro Sekunde (PAL) oder bei einer örtlichen Begrenzung auf ein Viertel (CIF) der Bildgröße. Dies gilt, wenn man eine gute bis ordentliche Qualität voraussetzt.
- Mit MPEG-1 kann man recht nahe an Echtzeit im vollen Format kommen. Dabei ist allerdings die Verarbeitungszeit bei der (De)-Kompression nicht mehr zu vernachlässigen und die kurzzeitige Pufferung

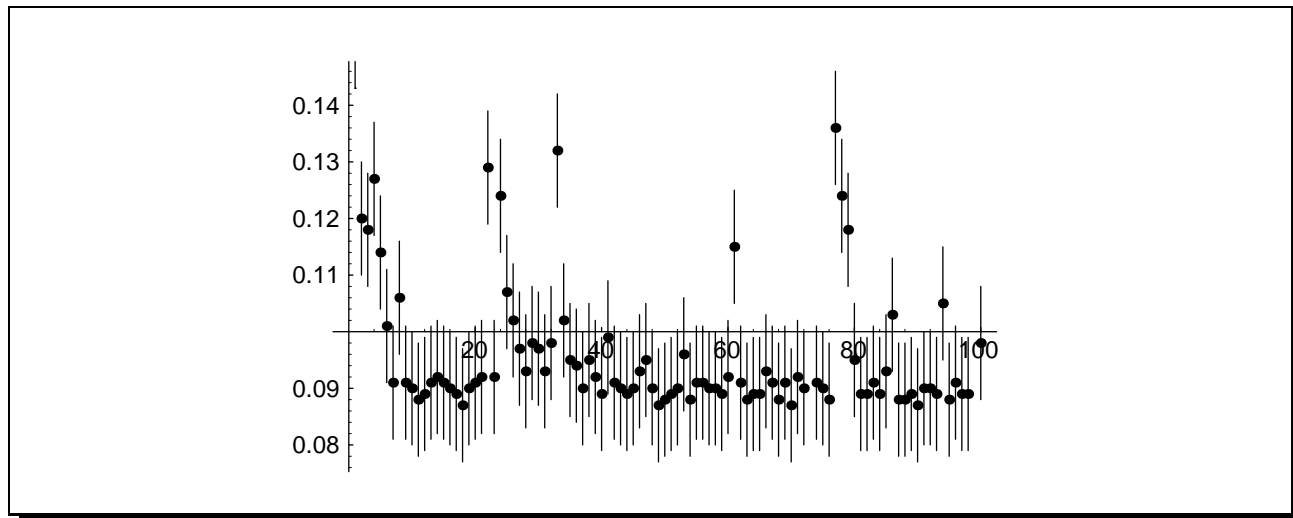


Abbildung 12: Anlieferungszeit aus der Sicht eines Clients für einzelne Frames unter Netzbelastung

der Daten notwendig.

– Ein wesentlich besseres Verhalten im lokalen Netz ist mit aufwendigen FDDI-Installationen zu erreichen.

Generell kennt die FDDI-Spezifikation Bandbreiten-Reservierung und sogar einen synchronen Transport-Modus. Allerdings arbeiten die meisten heute ausgelieferten Systeme nur asynchron und geben damit bei höheren Anforderungen auch keine Zeitgarantie.

Erst mit dem in der Normung begriffenen FDDI-2 soll Ernst gemacht werden: Hier soll es feste Zeitslots für dringliche Massendaten geben und isochrone Übertragung.

7 Trends und Ausblick

Nur für eine kleine Zahl von Formaten und Algorithmen gibt es ISO-Standards. Für einige Medien existieren sogenannte de facto Standards, die von Firmen entwickelt wurden und je nach Marktgewicht an Bedeutung gewinnen. Was fehlt sind flächendeckende und medienübergreifende Normen. Das Problem ist dabei der lange Normungsweg und die dem gegenüberstehenden kurzen Innovationszyklen. Bis eine Spezifikation den Weg vom *New Work Item* bis zum internationalen Standard *IS*⁹ geschafft hat, ist sie im schlimmsten Fall bereits veraltet.

Ein Lehrbeispiel hierfür ist die CD. Auf physikalischer Ebene deckt die Normung (Red Book) sämtliche Varianten ab. Bereits auf der Ebene der Datenorganisation (Operating System, File Structure) erfassen High Sierra- und ECMA-Standards alle CD-Formate außer der reinen Audio-CD. Auf höherer Ebene (Bild, Video) ist noch weniger Gemeinsamkeit zu finden. Bezüglich API und Datenbank-Aspekten gibt es schließlich disjunkte Standards: z.B. CD-I Green Book, CD-ROM X/Open XCDR.

⁹Der gesamte Normungsweg enthält NWI (*New Work Item*), dann WD (*Working Draft*), dann DP (*Draft Proposal*) und schließlich DIS (*Draft International Standard*) sowie IS. Dabei sind in einigen Schritten gemeinsame und nationale Abstimmungen, notwendig und bestimmte Schritte können auch mehrfach wiederholt werden.

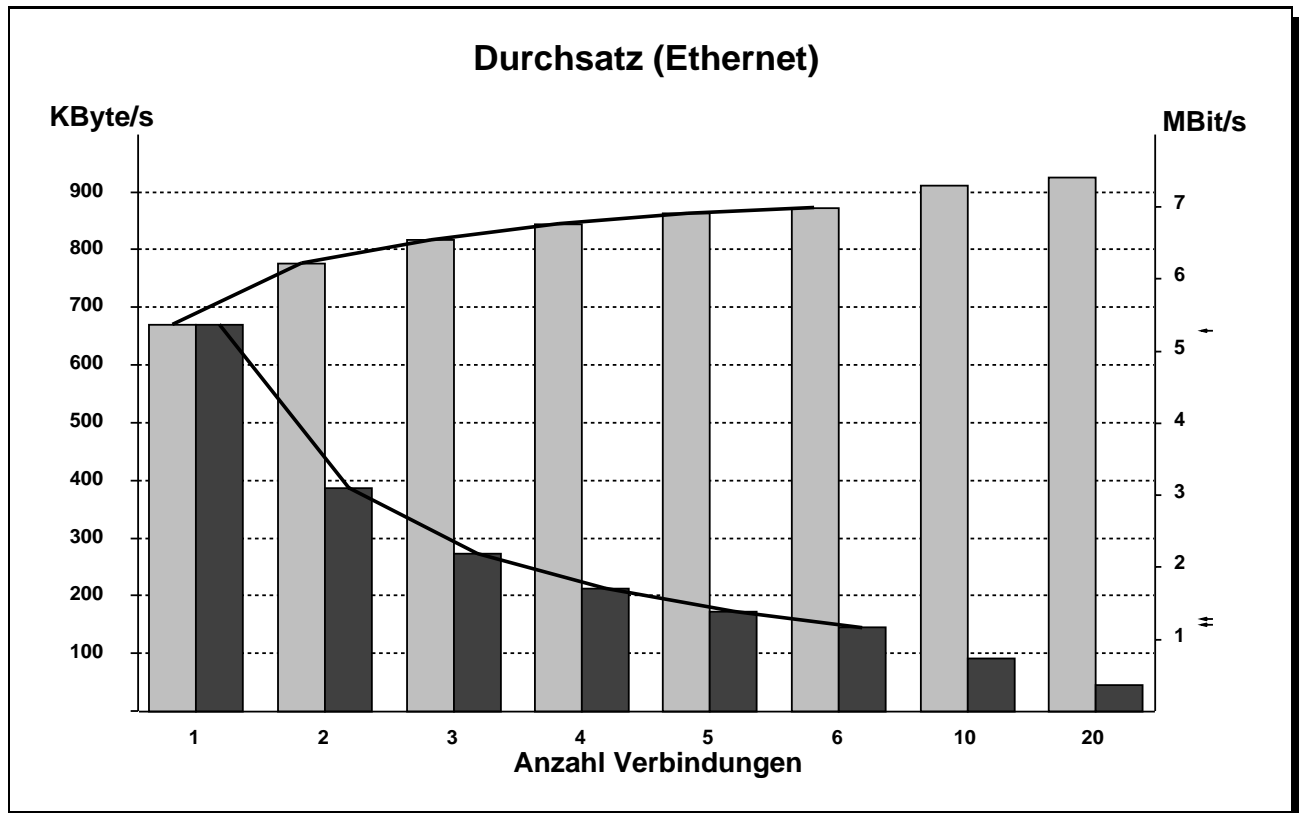


Abbildung 13: Leistungsbetrachtungen bei n Client-Verbindungen (dunkle Säulen) im Vergleich zum Gesamtdurchsatz des Netzes (helle Säulen)

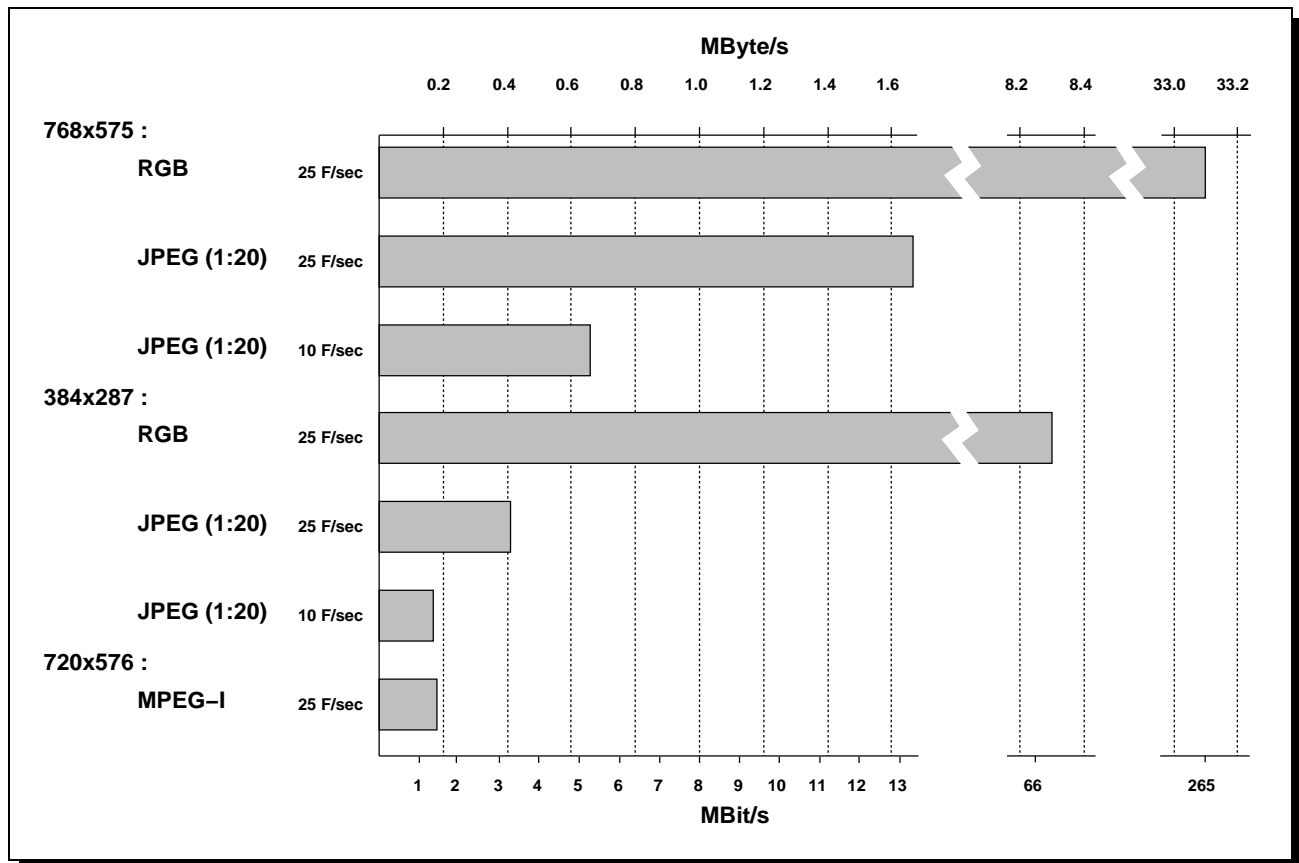


Abbildung 14: Datenmengen für RGB-Frames, JPEG-Frames und MPEG-Sequenzen bei verschiedenen Formaten



Abbildung 15: Die Fotografie eines Baumes als Testbild



Abbildung 16: Ausschnitte aus Bild 15. Oben ein Teil des Originals — darunter Ausschnitte des Bildes nach Kompression auf 1/10 (mitte links), 1/20 (mitte rechts), 1/50 (unten links) und 1/100 (unten rechts) der Originalgröße

Literatur

- [CF80] S.K. Chang und K.S. Fu(Hrsg.). **Pictorial Information Systems**. Springer-Verlag, Berlin, Heidelberg, 1980.
- [GER94] U. Glowalla, E. Engelmann und G. Rossbach(Hrsg.). **Multimedia '94 – Grundlagen und Praxis**. Springer-Verlag, Berlin, Heidelberg, 1994.
- [JPE91] Digital Compression and Coding of Continuous-Tone Still Images, Part II, Compliance Testing. ISO/IEC JTC1 Committee Draft 10918-2, Dec 1991.
- [JPE92] Digital Compression and Coding of Continuous-Tone Still Images, Part I, Requirements and Guidelines. ISO/IEC JTC1 Draft International Standard 10918-1, Jan 1992.
- [MPE91] Coding of Moving Pictures and Associated Audio for Digital Storage Media, Part I+II. ISO/IEC JTC1 CD 10172-1,-2, Nov 1991.
- [MW91] K. Meyer-Wegener. **Multimedia-Datenbanken**. Verlag Teubner, Stuttgart, 1991.
- [PM93] W.B. Pennebaker und J.L. Mitchell. **JPEG Still Image Data Compression Standard**. van Nostrand Reinhold, New York, 1993.
- [Ste93] R. Steinmetz. **Multimedia-Technologie – Einführung und Grundlagen**. Springer-Verlag, Berlin, Heidelberg, 1993.
- [Wal91] Gregory K. Wallace. The JPEG Stillpicture Compression Standard. *Communications of the ACM*, 34(4):30–44, Apr 1991.
- [Wir94] Analysen zur Multimedia Entwicklung. Wirtschaftswoche, div. Ausgaben 1993/94.