

Multimedia 1: Daten und Formate

PO-Ausarbeitung

Elisabeth Wetzinger

09.05.2007

INHALTSVERZEICHNIS

Datentypen	6
Erklären Sie kurz den Begriff Multimediasystem im engeren und im weiteren Sinne.	6
Was versteht man unter Halftone Approximation, wofür wird sie eingesetzt und worauf ist dabei zu achten?	6
Erklären Sie kurz Schalldruckpegel und Lautstärkepegel. Wozu dienen diese Werte?	6
Was ist und wie entsteht Quantisierungsrauschen? Auf welche Weise kann es reduziert werden?	7
Beschreiben Sie kurz das Prinzip von MIDI und erklären Sie die MIDI-Konzepte Clock und Palette und Voice	7
Welchen Zweck haben Distributon Amplifiers und Timebase Correctors?	7
Erklären Sie kurz 2 Arten von Video Time Codes Ihrer Wahl.	8
Erklären Sie kurz die Begriffe chroma keying und navigationl video.	8
Erklären Sie kurz die Begriffe <i>dynamic range</i> und <i>dithering</i> .	8
Erklären Sie kurz das Prinzip des Zeilensamplings (Line Sampling) für digitales Video. Was bedeutet die n:m:l Angabe im CCIR 601 Videoformat?	9
Erklären Sie kurz die Begriffe <i>skew</i> , <i>jitter</i> und <i>delay</i> . Nennen Sie 2 Hardwarekomponenten für die Behebung dieser Probleme.	10
Erklären sie Abtastfrequenz (sampling frequency), sample size und Quantisierung. Wie lautet das Abtasttheorem?	10
Erklären Sie die Zusammenhänge anhand der Bandbreite von ISDN.	10
Erklären Sie kurz die Begriffe <i>component video</i> und <i>composite video</i> . Was sind Vor- und Nachteile? Welche Rolle spielen <i>component video</i> und <i>composite video</i> für digitale Videoformate?	11
Erklären Sie kurz den Begriff <i>Interlacing</i> bzw. <i>interleaving</i> und führen Sie aus, was er im Kontext der Datentypen Audio, Images und Video bedeutet.	11
Was sind indizierte Farben (indexed colors) und weshalb werden sie verwendet?	13
Erklären Sie kurz den CIE-Farbraum.	13
Nennen Sie 2 Gründe, weshalb Video – als einziger Medientyp – in MM-Applikationen auch in analoger Form eingesetzt wird.	13
Erklären Sie kurz die Begriffe <i>Ton</i> , <i>Klang</i> und <i>Geräusch</i> .	14
Erklären Sie kurz die Begriffe <i>chroma keying</i> und <i>luminance keying</i> .	14
Plattformen	15
In welchem Kontext spielen <i>Real-Time</i> Video und <i>Production-Level</i> Video eine Rolle und was unterscheidet die beiden Formate?	15
Was sind die technologischen Unterschiede zwischen Audio-CD und CD-ROM?	15
Beschreiben Sie kurz die wichtigsten Unterschiede zwischen CD-ROM und CD-i.	15
Erklären Sie kurz die <i>Quicktime</i> Medienorganisation auf dem konzeptionellen Level.	16
Erklären Sie kurz das Zeitkonzept in <i>Quicktime</i> .	16
Worin unterscheiden sich die zwei Formate von Videodiscs? Listen Sie kurz Vor- und Nachteile dieser Formate.	16
Welches Format war Vorbild für die CD-Audio?	16
Erklären Sie kurz die Abbildung von Media Time zur Speicheradresse in Quickttime. Welche Datenstrukturen werden dafür benötigt?	16
Geben Sie eine verbale Beschreibung des Algorithmus, der für einen gegebenen QT-Track und einen gegebenen Zeitpunkt, die Adresse des zu diesem Zeitpunkt abzuspielenden Samples ermittelt.	17
	2

Was ist ein Frequenzspektrum? Wie unterscheiden sich die Frequenzspektren von Klängen und Geräuschen?	17
Was versteht man unter <i>Differential Encoding</i> ? Nennen Sie Beispiele, wo das Verfahren eingesetzt wird.	17
Kompression	18
Erklären Sie das Prinzip der <i>Laufängenkodierung</i> (Run Length Encoding).	18
Formulieren Sie einen Pseudocodealgorithmus, der in einer quantisierten DCT-Matrix die Runlength-Kodierung durchführt.	18
Diskutieren Sie kurz 3 Kriterien Ihrer Wahl zum Vergleich von Videokomprimierungsverfahren.	18
Erklären Sie kurz die Begriffe <i>Profiles</i> und <i>Levels</i> im Kontext von MPEG-2.	19
Erklären Sie kurz die Huffman Kodierung von DC-Koeffizienten.	19
Formulieren Sie in Pseudocode einen Algorithmus, der für die im Array <i>Symbole</i> gespeicherten Zeichen eine Huffman-Kodierung durchführt. Das Feld <i>Symbole</i> enthält in der 1. Komponente das zu kodierende Zeichen, in der 2. Komponente seine Wahrscheinlichkeit und in der 3. Komponente die zu errechnende Kodierung.	20
Welchen Zweck hat die <i>Diskrete Cosinus-Transformation</i> in Kompressionsverfahren (zb JPEG).	20
Erklären Sie kurz den Begriff <i>Quantisierung</i> im Kontext der Sampling-Theorie und im Kontext von Komprimierungsverfahren.	20
Worin unterscheiden sich die Quantisierungsschritte im JPEG- und MPEG-Verfahren?	21
Was sind <i>Quantisierungstabellen</i> und weshalb setzt man sie ein?	21
Erklären Sie kurz die P- und B-Frame Kodierung der MPEG-Komprimierung	21
Beschreiben Sie kurz die Aufgabe von D-Frames im Rahmen von MPEG. Wie werden D-Frames kodiert?	22
Was versteht man unter <i>Group of Pictures</i> ? Wie hängen Gop Muster mit der Datenrate zusammen?	22
Nennen Sie kurz 2 Arten von Artefakten bei der <i>Audiokodierung</i> .	23
Nennen und erläutern Sie kurz 2 Arten von <i>Scalable Bit Streams</i> in MPEG-2.	23
Was können Sie über die AC und DC Koeffizienten der DCT aussagen, wenn der zu transformierende 8x8 Block eines Bildes aus einem Schachbrettmuster besteht?	23
Was können Sie über die AC und DC Koeffizienten der DCT aussagen, wenn im zu transformierenden 8x8 Block eines Bildes 2 Flächen in einer horizontalen Linie schneiden?	24
Was können Sie über einen 8x8 Pixelblock aussagen, der nach der DCT-Transformation nur in den ersten beiden Spalten Werte aufweist, die von 0 verschieden sind?	24
Ermitteln Sie eine Huffmann-Codierung der Symbole A, B, C, D und E:	24
Wahrscheinlichkeiten ihres Auftretens: $p(A) = 0.25$, $p(B) = 0.41$, $p(C) = 0.18$, $p(D) = 0.14$, $p(E) = 0.02$	24
Wozu dienen Spatial Scalability und SNR Scalability? Was haben sie gemein und was sind die Unterschiede?	24
Was sind die wesentlichsten Unterschiede zwischen MPEG-1 und MPEG-2?	24
Für welche Einsatzbereiche würden Sie das Motion JPEG-Verfahren dem MPEG-Verfahren vorziehen?	25
Was versteht man unter <i>Pre-Echo</i> ? Wo tritt dieses Phänomen auf?	25
Beschreiben Sie kurz die wesentlichen Schritte der <i>MP3-Kodierung</i> . Erklären Sie insbesondere Quantisierung und Kodierung.	25
Was sind – abgesehen von Leistungsunterschieden – die wichtigsten Unterschiede zwischen MP3- und AAC-Kodierung?	26
Worin unterscheiden sich die I-Frame-Kodierung des MPEG-Verfahrens von der JPEG-Kodierung?	26
Erklären Sie knapp den Unterschied zwischen <i>Inter-</i> und <i>Intraframesynchronisation</i> . Geben Sie 2 Geräte an, die der Synchronisation dienen.	26
Erklären Sie kurz die verschiedenen Frametypen der MPEG-Komprimierung.	27

Was versteht man unter <i>Perceptual Measurement Techniques</i> und wofür werden sie eingesetzt?	28
Beschreiben Sie kurz die MPEG-4 Videocodierung	28
Erklären Sie kurz die Begriffe <i>Video Object</i> , <i>Video Object Plane</i> und <i>Video Object Layer</i> . In welchem Kontext werden sie verwendet?	28
Erklären Sie die Begriffe <i>Video Object</i> und <i>Video Object Plane</i> . In welchem Kontext werden sie verwendet?	28
Welchem Zweck dienen <i>BIFS</i> in MPEG-4? Welche Informationen enthalten sie?	29
Erklären Sie kurz das Konzept <i>Komposition</i> im MPEG-4 Systems Standard.	29
Was sind die Aufgaben des <i>Synchronization Layers</i> in MPEG-4.	29
Beschreiben Sie kurz die <i>Facial Animation</i> und <i>Facial Definition Parameters</i> im Face Animation-Teil von MPEG-4.	29
Erklären Sie kurz den <i>Skalierfaktor</i> und den Zusammenhang mit Kompressionsverfahren.	30
Erklären Sie kurz die <i>kritische</i> Abtastung	30
Erklären Sie kurz den <i>Maskierungseffekt</i> und den Zusammenhang mit Kompressionsverfahren.	30
Erklären Sie kurz <i>Kritische Bandbreite</i> und den Zusammenhang mit Kompressionsverfahren.	31
Was versteht man unter Structured Audio in MPEG-4?	31
Wie erfolgt die zeitliche Komposition (<i>Temporal Composition</i>) in MPEG-4 Datenströmen?	31
MM Programmierung	33
Nennen und erklären sie die wichtigsten Eigenschaften, nach denen zeitabhängige Ströme (Timed Streams) klassifiziert werden können.	33
Nennen und diskutieren Sie kurz zwei Gründe für das Konzept <i>Ableitung</i> (<i>Derivation</i>).	33
Nennen Sie 3 Gründe, weshalb die Multimedia – Programmierung fast ausschließlich dem objektorientierten Ansatz folgt.	33
Was versteht man unter dem Begriff <i>Configuration</i> und wie wird das Konzept im vorgestellten Framework realisiert?	33
Worin unterscheiden sich Ihrer Meinung nach die im vorgestellten Framework definierten <i>Components</i> vom allgemeinen Begriff <i>Komponenten</i> wie sonst im SE verwendet? Was haben die beiden Konzepte gemeinsam?	34
Wozu dienen die <i>Ports</i> von Komponenten? Unter welchen Umständen dürfen Ports zu verschiedenen Komponenten verbunden werden?	34
Was versteht man unter <i>padding</i> und weshalb wird <i>padding</i> eingesetzt?	34
Was ist die Aufgabe der Klasse <i>Transform</i> im vorgestellten MM-Framework und was ist an den Objekten dieser Klasse bemerkenswert?	34
Was versteht man unter der „dualen Natur“ zeitabhängiger Medien?	34
Was versteht man unter <i>Composition</i> und wie wird das Konzept im vorgestellten Framework realisiert.	35
Was sind function objects? In welchen Klassen des Frameworks werden sie implementiert?	35
Was versteht man unter abgeleiteten Objekten (derived objects), wozu dienen sie und wie werden sie im vorgestellten Framework realisiert?	35
Erklären Sie kurz den Begriff Quality of Service.	35
Was ist ein Framework und welche Anforderungen werden an ein MM-Framework gestellt?	35
Was leistet die Programmierungsabstraktion Interpretation? Nennen Sie zwei Sachverhalte, die die Interpretation erschweren.	36
MPEG-7 und MPEG-21	37

Erklären Sie kurz die Ziele von MPEG-7	37
Erklären Sie kurz die Konzepte <i>Deskriptor</i> , <i>Description Scheme</i> und <i>DDL</i> im Kontext von MPEG-7.	37
Erklären Sie sehr kurz die Terminal Architektur von MPEG-7.	37
Erklären Sie kurz die Konzepte Rights Data Dictionary (RDD) und Rights Expression Language (REL). Inwelchem Kontext treten sie auf und wie spielen sie zusammen?	38
Erklären Sie kurz die Ziele von MPEG-21.	38

DATENTYPEN

Erklären Sie kurz den Begriff Multimediasystem im engeren und im weiteren Sinne.

Ein Multimediasystem ist charakterisiert durch computergesteuerte, integrierte Produktion, Manipulation, Präsentation, Speicherung und Kommunikation unabhängiger Information, welche zumindest durch ein kontinuierliches (zeitabhängiges) und ein diskretes (zeitunabhängiges) Medium kodiert ist.

Im weiteren Sinne wird es verwendet um die Verarbeitung individueller Bilder und Texte zu beschreiben, obwohl kein kontinuierliches Medium vorhanden ist.

Was versteht man unter Halftone Approximation, wofür wird sie eingesetzt und worauf ist dabei zu achten?

Approximierung von Intensitäten unter der Verwendung von Dot-Patterns; dadurch wird bei Bildern mit niedriger Farbtiefe die Illusion einer höheren Farbtiefe erzeugt. Es werden die fehlenden Farben durch eine bestimmte Anordnung aus verfügbaren Farben nachgebildet und dadurch harte Übergänge zwischen den Farben vermieden. Das menschliche Auge nimmt diese als Mischung der einzelnen Farben wahr.

Verwendung: am häufigsten bei Farbverläufen

zu beachten: Linepatterns und andere Artefakte vermeiden, Patterns sollen wachsen (Minimierung des Kontureffekts), Patterns sollen sich vom Zentrum ausbreiten (Effekt wachsender Punktgrößen)

Es gibt viele verschiedene Algorithmen, diese arbeiten oft auch mit einer Fehlerstreuung (Übertragung von Überläufen oder Rundungsfehlern auf benachbarte Bildpunkte für eine feinere Darstellung).

Erklären Sie kurz Schalldruckpegel und Lautstärkepegel. Wozu dienen diese Werte?

Schalldruckpegel: $L = 20 \cdot \log(p/p_0)$ [dB] mit Bezugsschalldruck p_0 (1kHz, gerade noch hörbar $20\mu\text{Pa}$)

Der Schalldruckpegel dient zur sinnvollen Darstellung und Beschreibung des Hörbereichs, denn durch die logarithmische Darstellung kann die Spanne von der Hörschwelle bis zur Schmerzgrenze (1: 2 Millionen) in einem kleineren Wertebereich dargestellt werden (0 – 120dB). Außerdem kommt die logarithmische Darstellung dem Menschen insofern zu Gute, weil sich durch Experimente herausstellte, dass das Empfinden von Lautstärkedifferenzen einem logarithmischen Gesetz folgt und somit diese Darstellung dem Empfinden mehr gerecht wird als die Angabe eines absoluten Druckamplitudenwertes.

Lautstärkepegel: [phon] Der Lautstärkepegel ist ein Vergleichsmaß. Er beschreibt, welchen Schalldruckpegel ein Sinuston mit einer Frequenz von 1000 Hz haben müsste, damit dieser genauso laut empfunden wird wie der betrachtete Schall. Bei einer Schall-Frequenz von 1000 Hz stimmen Schalldruckpegel, gemessen in dB SPL (sound pressure level), und Lautstärkepegel, gemessen in Phon, überein. Für Sinustöne anderer Frequenzen sowie für komplexe Schallereignisse sind dagegen andere Schalldruckpegel erforderlich, um den gleichen Lautstärkeindruck zu erzielen. Welcher Schalldruckpegel für einen Einzelton bei welcher Frequenz erforderlich ist, um jeweils den gleichen Lautstärkeindruck zu erzielen, ist in den "Kurven gleicher Lautstärkepegel" (Isophone) beschrieben. Z. B.: Ton mit 100Hz und Schalldruckpegel $L=60\text{dB}$ → 50 Phon

Was ist und wie entsteht Quantisierungsrauschen? Auf welche Weise kann es reduziert werden?

Das Quantisierungsrauschen ist die Folge des Quantisierungsfehlers. Dieser entsteht bei der Quantisierung analoger Signale: Das analoge Signal wird mit einer bestimmten Frequenz abgetastet und der gemessene Amplitudenwert wird auf den nächsten diskreten Wert gerundet (AD-Wandlung). Die Differenz zwischen tatsächlichem analogen Wert und quantisiertem, diskretem Wert bezeichnet man als Quantisierungsfehler, welcher kleiner gleich $0,5 Q$ (Quantisierungsintervall) ist. Wird das diskrete Signal wieder in ein analoges gewandelt, entsteht bei dessen Wiedergabe das Quantisierungsrauschen, welches die akustische Auswirkung des Quantisierungsfehlers ist und den Schalldruckpegel L_{noise} erzeugt.

Reduktion: Das Rauschen kann zum Einen durch eine erhöhte Auflösung beim Quantisieren reduziert werden und andererseits bei der linearen Quantisierung, sofern der Signal-Rausch-Abstand groß genug ist, ausmaskiert werden.

Beschreiben Sie kurz das Prinzip von MIDI und erklären Sie die MIDI-Konzepte Clock und Palette und Voice

Skizze (S. 127)

MIDI ist ein Protokoll, das die Kommunikation von Computern, Synthesizern, Keyboards und anderen Musikinstrumenten ermöglicht, es stellt sozusagen die Schnittstelle zwischen den Devices dar, über die die Kommunikation untereinander möglich ist.

Clock:

gibt die Taktrate vor, versieht die Messages mit Timestamps; Maß: Teile pro Viertelnote, Tempo: bpm

palette:

Aufgrund gespeicherter Werte aus einer Palette werden die entsprechenden gespeicherten Sounds wiedergegeben.

voice:

ist der Teil des Synthesizers, welcher den Sound produziert. Synthesizer können mehrere (16, 20, 24, 32, 64,...) voices haben, wobei jede unabhängig und simultan arbeitet um Sound mit verschiedenen Tonlagen und Klangfarben zu produzieren.

Welchen Zweck haben Distribution Amplifiers und Timebase Correctors?

Distribution Amplifier: Er gibt ein Eingangssignal verstärkt an mehreren Ausgängen aus. Die Ausgänge haben alle dasselbe Eingangssignal. Es handelt sich um eine Signalverteilung mit Verstärkung.

Timebase Corrector: TBCs; Rekonstruiert das Signal um Timingfehler, die durch VTR/VCR (video cassette recorder) entstehen, zu eliminieren.

Erklären Sie kurz 2 Arten von Video Time Codes Ihrer Wahl.

SMPTE time code (HH:MM:SS:FF)

non-drop frame time code – FF in [0, 29]

drop frame time code FF in [2, 29], außer bei jeden 10. Minute [10, 29]

TimeCode recorded on tape:

LTC: Longitudinal Time Code: meist im Audiotrack aufgezeichnet; entlang des Bandes

VITC: Vertical interval TC; in den vertikalen Austastlücken aufgezeichnet; im Nachhinein nicht mehr korrigierbar; kann nur gleichzeitig mit dem Video aufgezeichnet werden

Erklären Sie kurz die Begriffe chroma keying und navigationl video.

chroma keying:

Der Begriff Farbstanze oder englisch Chroma Key bezeichnet beim Farbfernsehen ein Verfahren, bei dem ein bestimmter Chroma-Wert durch ein anderes Bild ersetzt wird. Die Bildteile mit entsprechender Farbe werden gleichsam durchsichtig. Chromakeying wird hauptsächlich für die Bluescreen-Technik eingesetzt.

Navigational video / navigable video: Erstellung einer „virtuellen“ Kamerafahrt aus Einzelbildern. Dabei wird ein Objekt von allen Seiten mit Winkelabständen von beispielsweise 5° aufgenommen. Diese Bilder werden dann aneinandergereiht und man erhält einen Eindruck einer Kamerafahrt und 360°-Ansicht (Panos).

Erklären Sie kurz die Begriffe dynamic range und dithering.

Dynamic range: Bei achromatischem Licht stellt die *Dynamic range* das Verhältnis zwischen minimalen und maximalen Intensitäten dar. Wie viele Intensitäten in einem Bild zur Darstellung nötig sind, hängt davon ab, mit welchem Medium es verwendet wird. Die Intensitäten reichen aus, wenn die Übergänge in der Darstellung kontinuierlich sind. (hilfe: Dithering)

Dithering: (siehe auch Halftone Approximation)

Approximierung von Intensitäten unter der Verwendung von Dot-Patterns. Dies funktioniert, da das menschliche Auge so feine Punktmuster vermischt (schwarz/weiß → grau); dadurch wird bei Bildern mit niedriger Farbtiefe die Illusion einer höheren Farbtiefe erzeugt. Es werden die fehlenden Farben durch eine bestimmte Anordnung aus verfügbaren Farben nachgebildet und dadurch harte Übergänge zwischen den Farben vermieden. Das menschliche Auge nimmt diese als Mischung der einzelnen Farben wahr.

Verwendung: am häufigsten bei Farbverläufen, z.B. auch bei Druckern (4 Farben)

Clusterdithering:

Ditherintensität auf zentriert auf die Matrizenmitte.

Linepatterns und andere Artefakte vermeiden, Patterns sollen wachsen (Minimierung des Kontureffekts),

Patterns sollen sich vom Zentrum ausbreiten (Effekt wachsender Punktgrößen)

Es gibt viele verschiedene Algorithmen, diese arbeiten oft auch mit einer Fehlerstreuung (Übertragung von Überläufen oder Rundungsfehlern auf benachbarte Bildpunkte für eine feinere Darstellung). Sind Linien nicht zu vermeiden, verwende diagonale Linien.

Erklären Sie kurz das Prinzip des Zeilensamplings (Line Sampling) für digitales Video. Was bedeutet die n:m:l Angabe im CCIR 601 Videoformat?

Stellt eine Norm zur Digitalisierung analoger Videosignale dar. Basiert auf dem YCC-Farbmodell (Y = Luminanz, CC = 2 Farbkanäle). Es gibt mehrere CCIR „Familien“, die sich durch ihre unterschiedlichen Werte für m:n:l unterscheiden.

m: Basisrate, mit der Y abgetastet wird

n, l: bestimmen, mit welcher Rate die Farbkanäle abgetastet werden

Standard: Basisrate = 3,375MHz; m, n, l erhalten einen Multiplikator dieser Rate

Einer der großen Vorteile von YCbCr ist, dass die Abtastrate der Chrominanz-Kanäle niedriger als die des Y Kanals sein kann, ohne dass es zu einer spürbaren Verringerung der zu gewährleistenden Qualität kommt (Chroma Subsampling).

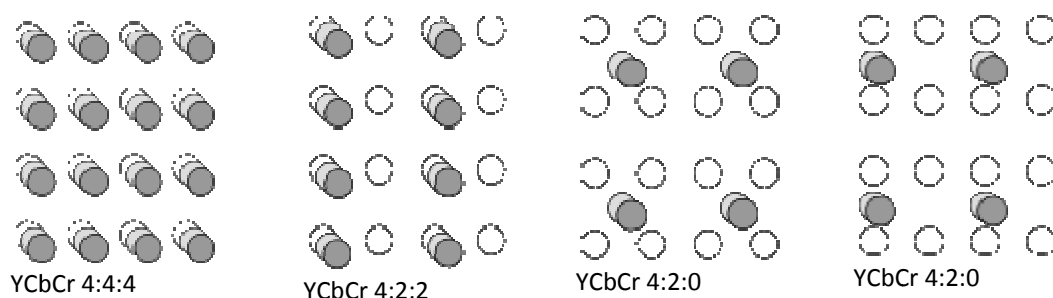
- YCbCr 4:4:4 Chrominanz-Auflösung identisch zur Luma-Auflösung
- YCbCr 4:1:1 Chrominanz-Auflösung horizontal geviertelt und vertikal unverändert

Bei YCbCr 4:2:2 handelt es sich um horizontale Unterabtastung der Farbinformation (Chrominanz) um den Faktor 2, somit wurde die Chrominanz-Auflösung horizontal halbiert. Die 4:2:2-Kodierung entspricht der "Studioqualität" für digitales Video (ITU-R 601-4) und wird von professioneller Aufnahme- und Schnitthardware verwendet.

Bei YCbCr 4:2:0 wird die Chrominanz-Auflösung sowohl horizontal als auch vertikal halbiert. Für Übertragung und für den Heimgebrauch wird typischerweise 4:2:0 verwendet.

Es gibt weiterhin noch Unterschiede über das Zentrum der Chrominanzwerte. Diese können zentriert oder nicht zentriert sein:

Zur Beschreibung wird eine so genannte n:m:C-Notation verwendet. Sie gibt wieder, wie oft Cb und Cr im Vergleich zu Y abgetastet werden.



Weißer Kreise symbolisieren hier die Abtastpositionen für die Luminanz, graue Kreise die Positionen für die Chrominanz.

Erklären Sie kurz die Begriffe *skew*, *jitter* und *delay*. Nennen Sie 2 Hardwarekomponenten für die Behebung dieser Probleme.

Skew:

Mit *Skew* wird die Zeitdifferenz eines Signals zu bezüglich seinem korrekten Abspielzeitpunkt (Videosignal) angegeben. Er entsteht beispielsweise aufgrund von Übertragungsdivergenzen (Audiosignal wird schneller übertragen als das Videosignal, beide müssen aber synchron abgespielt werden). Des Weiteren können solche Divergenzen auch bei der Übertragung über verschiedene Medien entstehen. (Asynchronität zweier Streams)

Jitter:

Als Jitter bezeichnet man ein Taktzittern bei der Übertragung von Digitalsignalen, eine leichte Genauigkeitsschwankung im Übertragungstakt (Clock). Allgemeiner ist Jitter in der Übertragungstechnik ein abrupter und unerwünschter Wechsel der Signalcharakteristik. Dies kann sowohl Amplitude als auch Frequenz und Phasenlage betreffen.

In der Netzwerktechnik wird mit Jitter außerdem die Varianz der Laufzeit von Datenpaketen bezeichnet. Dieser Effekt ist insbesondere bei Multimedia-Anwendungen im Internet (zum Beispiel Video-Streaming und IP-Telefonie) störend, da dadurch Pakete zu spät eintreffen können, um noch zeitgerecht mit ausgegeben werden zu können. Dies wirkt sich wie eine erhöhte Paketverlustrate aus. Der Effekt wird durch einen sog. Jitterbuffer reduziert, allerdings zum Preis von zusätzlichem Delay, was bei streaming-Anwendungen nichts ausmacht, bei Dialoganwendungen aber ebenfalls stört.

Jitter kann auch bei der Aufzeichnung von Fernsehsignalen auf Videorekordern entstehen. Hier wird der Jitter durch mechanische Toleranzen verursacht. Häufig erscheinen diese Fehler als ein Wackeln des Bildes.

Die Einheit, in der der Jitter durch Messinstrumente gemessen wird, ist die Zeit: die Schwankung der Periodendauer wird typischerweise in μs gemessen.

Jitter entsteht zum Beispiel durch Bedämpfung des Hochfrequenzanteils eines Frequenzspektrums auf einer langen Kabelstrecke. Das nennt man dann deterministischen Jitter. Andere Ursachen sind Rauschen (zufälliger Jitter, random Jitter) oder Symbolübersprechen (datenabhängiger Jitter). Um die Qualität einer Übertragungsstrecke beurteilen zu können, versucht man oft, die Ursachen für den vorhandenen Jitter zu finden. Dazu versucht man, den Gesamtjitter in die oben erwähnten Anteile aufzuspalten. Dieses Vorgehen nennt man Jitterseparation.

Delay

Verzögerung des gesamten Streams (im Gegensatz zu *Skew*, wo die Verzögerung eines Streams gegenüber einem anderen gemessen wird).

Behebung mittels Sync-generator oder frame (delay) buffer.

Erklären sie Abtastfrequenz (sampling frequency), sample size und Quantisierung. Wie lautet das Abtasttheorem? Erklären Sie die Zusammenhänge anhand der Bandbreite von ISDN.

Mit *Abtastrate*, auch Samplingrate, bezeichnet man in der Signalverarbeitung die Häufigkeit, mit der ein Signal pro Zeitintervall abgetastet wird. Der Abstand zwischen den Abtastzeitpunkten ist das Abtastintervall. Ist er konstant, heißt die Abtastrate auch Abtastfrequenz oder Samplingfrequenz. Die Wahl einer konstanten Abtastfrequenz vereinfacht die Weiterverarbeitung des Signals.

Je höher die Abtastrate, desto besser die Qualität des zeitdiskreten Signals. Minimale Abtastrate zur verlustlosen Rekonstruktion des Originalsignals: Nyquistfrequenz \rightarrow min. doppelte im Signal höchste vorkommende Frequenz ($2 \times f_{\text{max}}$)

sample size:

Die Samplesize bestimmt die Aufteilung der Amplitudenskala bei der Digitalwandlung eines analogen Signals und wird in Bits angegeben. Ist beispielsweise die Samplesize 8bit, so kann der gesamte im analogen Signal vorkommende dynamische Bereich in 256 diskrete Teile unterteilt werden. Bei der Sampling wird die Amplitude des analogen Signals durch den nächstgelegenen diskreten Wert (aus den 256 im Falle von 8bit) angenähert. Somit erhält man ein amplitudendiskretes Signal.

Quantisierung

abgetasteten Spannungswerten werden diskrete Zahlungswerte zugeordnet. Gesamtspannungsbereich in Quantisierungsintervalle Q unterteilt → kontinuierlicher Wert nächst gelegenen Zahlenwert zugeordnet. Um eine möglichst gute Qualität zu erreichen soll Q möglichst klein sein, es ergibt sich aus der Länge des Datenwortes. 8bit (256 Intervalle): $Q = V_{pp}/(2^8) = 4 \cdot 10^{-3} \times V_{pp}$

Man unterscheidet zwischen *Linearer Quantisierung* (Q konstant, üblich in Audiotechnik) und *nichtlinearer Quantisierung* (Q von unterschiedlicher Größe → kleine Werte: kleiner Quantisierungsfehler, große Werte: großer Quantisierungsfehler)

ISDN:

digitale Übertragung bis zum Endgerät

8bit Quantisierung (Raster mit 256 Einteilungen zur Digitalisierung der Amplitude).

8kHz Abtastrate (Abtastung 8000x pro Sekunde).

2-B/1-D Kanäle: B-Kanal – Nutzsignal á 64kbit/s; D-Kanal – 16kbit/s Steuerkanal (Verbindungsaufbau, -abbau, Rückruf-bei-Besetzt,...)

Erklären Sie kurz die Begriffe *component video* und *composite video*. Was sind Vor- und Nachteile? Welche Rolle spielen *component video* und *composite video* für digitale Videoformate?

component video:

Jede Primärfarbe sowie das Luminanzsignal werden als separate Signale gesendet. – RGB oder eine Luma-Chromatransformation (YUV, YIQ, $YCrCb$) Deswegen werden mehr Bandbreite sowie eine gute Synchronisation der Signale benötigt. Verwendung für Editing und Encoding im professionellen Bereich.

digital component video (D1/D5), SMPTE RP125): Trennung der Signale in Helligkeit und Farbe, 27MB/s Datenrate, subsampled Farbsignale 4:2:2

Composite video:

Die Signale werden gemixt und auf einer einzigen Trägerwelle übertragen. Vorteile sind die einfachere Übertragung, weniger Verkabelung, jedoch ist die Qualität nicht so gut (Interferenzen), verwendet im Haushalt (VHS).

digital composite video (D2/D3, SMPTE 244M): 14.3MB/s Datenrate, subsampled Farbsignale 4:2:2

Erklären Sie kurz den Begriff *Interlacing* bzw. *interleaving* und führen Sie aus, was er im Kontext der Datentypen Audio, Images und Video bedeutet.

Interlacing (Zeilensprungverfahren):

Video:

Anstatt Zeile für Zeile eines Bildes darzustellen, wird es in zwei Halbbilder aufgeteilt, wobei eines aus den geraden und das andere aus den ungeraden Bildzeilen besteht. Diese Halbbilder werden abwechselnd angezeigt. Dies ist möglich, weil der Bildschirm nachleuchtet und somit das Auge zu träge

ist, die Halbbilder als solche zu erkennen und damit nicht auffällt. Vorteil: Bandbreite sparen; Nachteil: aufwändiges Deinterlacing

Image:

Das Bild wird nicht Zeile für Zeile aufgebaut, sondern in mehreren Durchläufen. Beim ersten Durchlauf werden eine bestimmte Anzahl von Zeilen übersprungen, in den nächsten Durchläufen werden die übersprungenen Zeilen dann fortlaufend dargestellt. Sichtbar ist dieser Vorgang insofern, als dass das Bild anfangs unscharf ist und dann mit der Zeit immer „schärfer“ wird.

Als Interlace bzw. **Interlacing** bezeichnet man ein Speicherverfahren für Grafiken, bei dem das Bild in mehreren Schichten abgespeichert wird. Dies ermöglicht den schnellen Aufbau eines Übersichtsbildes, wenn die Grafik von einem langsamen Medium geladen wird. Die Darstellung wird dann während des Ladevorgangs immer feiner.

Interleaving

Der Begriff Verschränkung oder englisch Interleaving (engl. to interleave „verschachteln, überlappen“) bezeichnet den Prozess, mehrere linear durchzählbare Objekte in einer speziellen Reihenfolge anzuordnen.

Audio

wird verwendet um die Audiospur mit der Bilderfolge zu einem Stream zu verbinden. Nach einer festgelegten Anzahl Frames wird ein Stück des Audiostreams gespeichert, der die Tonspur die folgenden Frames enthält. Der Abstand zwischen den Audiotteilen darf nicht zu groß sein, da sonst das Abspielen ohne Aussetzer nicht gewährleistet ist. Je enger der Abstand ist, desto mehr Systemressourcen werden aber in Anspruch genommen, da der Computer die für das Abspielen des Videos zuständigen Geräte, sehr oft mit eher kleinen Datenmengen beliefern muss. D.h. also, dass Audio und Video in abwechselnden Blöcken gespeichert werden.

Vorteile: Einfache Synchronisation von Audio/Video, weniger Streams.

Nachteile: siehe oben (Abstand)

Video:

Anstatt die drei Farbkanäle in einem File separat zu speichern (RRRRRR...GGGGGGGG...BBBBBBBB...), werden diese interleaved für jeden einzelnen Bildpunkt oder zumindest Bildbereich in die einzelnen Kanäle gleich hintereinander gespeichert (RGBRGBRGBRGB...). Somit muss das File nicht immer komplett neu geladen werden, damit es richtig angezeigt werden kann, sondern die Bildpunkte können sofort ausgegeben werden.

Image:

Using the different interleaving schemes, an image with 4 rows and 5 columns would look like the following depending on the interleaving scheme.

Pixel interleaved looks like:

RGBRGBRGBRGB	row 1
RGBRGBRGBRGB	row 2
RGBRGBRGBRGB	row 3
RGBRGBRGBRGB	row 4

Raster interleaved looks like:

RRRRRGGGGGBBBBB	row 1
RRRRRGGGGGBBBBB	row 2
RRRRRGGGGGBBBBB	row 3
RRRRRGGGGGBBBBB	row 4

Frame interleaved looks like:

RRRRRRRRRRRRRRRRRR	all the red
GGGGGGGGGGGGGGGGGG	all the green
BBBBBBBBBBBBBBBBBBBB	all the blue

Was sind indizierte Farben (indexed colors) und weshalb werden sie verwendet?

Farben werden durch einen Index in Color maps oder Color Lookup Tabellen (CLUT) repräsentiert, welche entweder vordefiniert sind oder Teil des Bildes. Jeder Pixel verweist auf einen Eintrag in dieser Tabelle, der dem jeweiligen Farbwert entspricht. Vorteil: Platzsparende Bildspeicherung. Verwendet beim Color-Mapping.

Erklären Sie kurz den CIE-Farbraum.

Darstellung von Farben beruhend auf dem menschlichen Wahrnehmungsapparat

Additive Farbmischung

Grundlage ist ein sogenannter *Normalbeobachter* mit einem Sichtfeld von 2° (CIE1931) bzw. 10° (CIE1964).

Normfarbwerte X (virtuelles rot), Y (virtuelles grün), Z (virtuelles blau)

Darstellung: Querschnitt durch den Raum; alle Farben befinden sich innerhalb der Fläche

Wellenlänge: gebogene, äußere Linie

Purpurgerade: gerade Verbindungslinie zwischen Rot und Blau

Spektralfarben: höchster Sättigungsgrad; am äußeren Kurvenrand, reines Weiß in Zentrumsnähe (C)

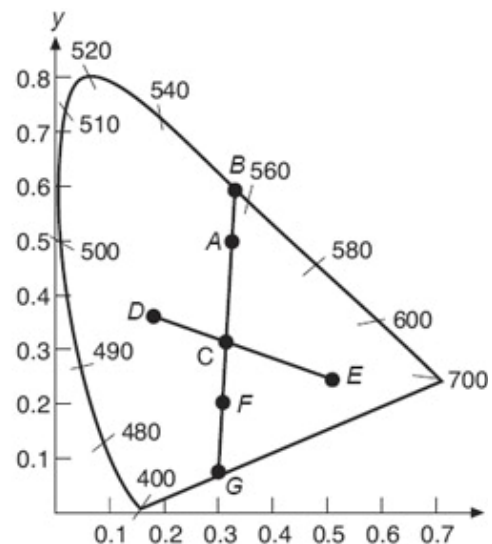
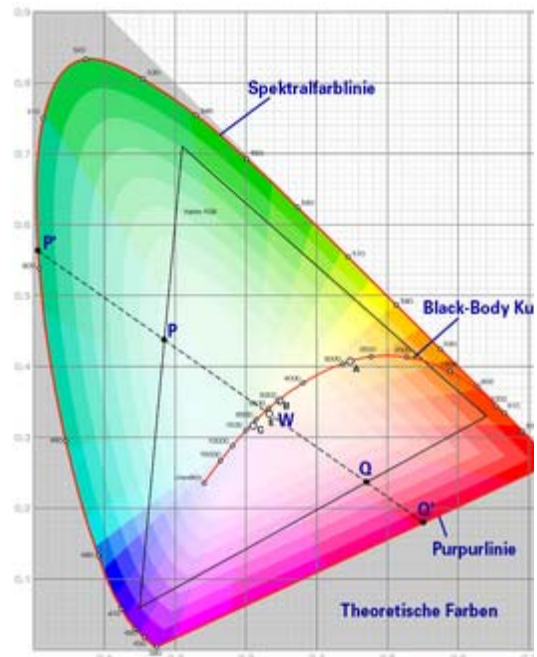
Mittelpunktvalenz: $x = 0,33$, $y = 0,33$ (C)

Wenn A die betrachtete Farbe ist, so wird sie durch die Wellenlänge B dominiert. Wird B mit C gemischt, so erhält man A.

Die in F dominierende Wellenlänge G ist definiert als das Komplement der dominierenden Wellenlänge B in A.

Werden Komplementärfarben gemischt (E mit D, F mit A), so erhält man weiß.

Verwendung: Kalibrierung anderer Farbmodelle, Identifikation und Analyse von Farben, Performancemessung von CRT's



Nennen Sie 2 Gründe, weshalb Video – als einziger Medientyp – in MM-Applikationen auch in analoger Form eingesetzt wird.

Billige Speicherkosten
Digitalisierung sehr zeitaufwändig
analoges Video ist einfacher zu senden.

Erklären Sie kurz die Begriffe *Ton*, *Klang* und *Geräusch*.

Ton:

Als Ton wird eine einzelne Sinusschwingung bezeichnet. $P(t)$ ist eine Sinusschwingung (mit einer bestimmten Frequenz – $1/t$)

Klang:

Klang entsteht durch Überlagerung eines Grundtons (hörbar als Tonhöhe) und Obertönen ($f = n \cdot f_g$) (hörbar als Klangfarbe) und stellt eine allgemeine periodische Funktion $p(t)$ mit einer Frequenz f_g dar.

Geräusch:

Bei einem Geräusch ist kein Grundton mehr erkennbar, die Funktion $p(t)$ ist aperiodisch.

Erklären Sie kurz die Begriffe *chroma keying* und *luminance keying*.

chroma keying:

Der Begriff Farbstanze oder englisch Chroma Key bezeichnet beim Farbfernsehen ein Verfahren, bei dem ein bestimmter Chroma-Wert durch ein anderes Bild ersetzt wird. Die Bildteile mit entsprechender Farbe werden gleichsam durchsichtig. Chromakeying wird hauptsächlich für die Bluescreen-Technik eingesetzt.

Luminance keying:

Chroma bedeutet, dass man als Hintergrund eine spezielle Farbe verwendet - das gleiche kann man auch genauso im Luminance-Bereich anwenden, da separiert man dann aufgrund von Helligkeitswerten.

(Bei diesem Verfahren werden die Bildteile eines Videos transparent, die einer bestimmten Helligkeit entsprechen. An diesen Stellen wird ein darunter gelegtes zweites Video sichtbar. Das Verfahren ist eher für Effekte interessant. Kann weniger kontrolliert werden, wird aber oft bei Graphikelementen verwendet.)

Erklären Sie kurz den Begriff Interleaving im Kontext des Datentyps Audio. Welche Vor- bzw. Nachteile haben interleaved Audioformate im Vergleich zu non-interleaved Formaten?

In welchem Kontext spielen *Real-Time Video* und *Production-Level Video* eine Rolle und was unterscheidet die beiden Formate?

Kontext: DVI (digital video interactive)

RTV:

Kompressionsformat, mit der Videosignale in Echtzeit digitalisiert, komprimiert und auf der Festplatte gespeichert werden können; Real Time Video (RTV) repräsentiert die einfache Qualitätsstufe unter Digital Video Interactive (DVI). Bei RTV können die Videosequenzen in Echtzeit komprimiert und mit "Video for Windows" dargestellt werden. Eine Variante von RTV ist Indeo.

128x120, mit Zeilen/Spalten-Verdopplung und Interpolation 256x240, RTV-Intraframe-Codierung: anschließend RLE+Huffman, einfaches, aber schnelles Verfahren, symmetrisch; Vorteil: MM-Applikationen – zunächst RTV, dann PLV codieren und decodieren in Echtzeit (Videokonferenzen)

PLV:

Production Level Video (PLV) repräsentiert eine hohe Qualitätsstufe unter Digital Video Interactive (DVI). Bei PLV werden die Videosequenzen offline von einem Superrechner komprimiert und können dann mit Personal Computer (PC) und DVI-Karte wiedergegeben werden. Mit diesem Verfahren werden Kompressionen von 150:1 erreicht.

nur I, P-Bilder (keine B-Bilder); Luminanz 256x240, 24bit (true color), mit Zeilen/Spalten-Verdopplung und Interpolation: 512x480;

Kompressionsalgorithmus nicht freigegeben, Kompressionszeit/Dekompressionszeit 100:1

Was sind die technologischen Unterschiede zwischen Audio-CD und CD-ROM?

CD-ROM	zusätzlicher Layer für Fehlerdetektion und Fehlerkorrektur eingeteilt in aufeinanderfolgende Datenblöcke (98 frames, 2352 bytes) 2 Blockformate (Mode1 – 2048 bytes application data, Mode2 – 2333 bytes application data)
CD-audio	verwendet nur CLV (constant linear velocity) frame mit 588 channel bits, 200 bits nach dem Dekodieren (je AudioCD 20MB nicht nutzbar)

Beschreiben Sie kurz die wichtigsten Unterschiede zwischen CD-ROM und CD-i.

CD-i ist die erste MM-Plattform Anfang der 90er Jahre gewesen

Es ist ein komplettes System. Sowohl logische als auch physikalische Organisation von Multimediam Informationen auf der CD. Hardware Interfaces für CD-i Player, Softwareinterfaces für CD-i Programme. Im Gegensatz zur CD-ROM sind auf der CD-i für jeden Medientyp Formate definiert, der Prozessor ist in den Player integriert → eine CD-i ist auf einem PC nicht mehr abspielbar, jedoch aber CD-ROM XA (extended architecture, die eine offene Form der CD-i ist und beide Formate lesen kann)

Auf der Cd-i sind folgende Medientypen festgelegt:

Audio (CD-DA, Level A, Level B, Level C)

Image (RGB 5:5:5, DYUV, CLUT (256, 128, 16 Einträge), RL (runlength coding, 128 oder 8 Einträge))
Video (MPEG-1 enkodiert)
Text und Graphik (applikationsabhängig)

Erklären Sie kurz die *Quicktime* Medienorganisation auf dem konzeptionellen Level.

- Data entity: repräsentiert den tatsächlichen für a/v – Daten verwendeten Speicher
- Media entity: zeitliche Sequenz [media time]; jede Entität hat einen Medientyp (audio/video), Elemente referenzieren Speicherregionen.
- Track entity: Beschreibt die Anordnung von Media Entitäten. Jede Entität ist in einer Movie Entity enthalten, die Zeit wird in movie TCS gemessen.
- Movie entity: ist eine Gruppe von Track Entities und spezifiziert Zeitskalierung und Dauer. Beschreibt die Anordnung, Dauer und Offset der Tracks innerhalb des Movies.

Erklären Sie kurz das Zeitkonzept in *Quicktime*.

Zeit ist explizit
Werte werden als Integer ohne Vorzeichen repräsentiert
Zeitskala: units per second
Dauer: maximaler Zeitwert
Zeitbasis: Playbackrate
time coordinate system TCS, scale, duration)

Worin unterscheiden sich die zwei Formate von Videodiscs? Listen Sie kurz Vor- und Nachteile dieser Formate. Welches Format war Vorbild für die CD-Audio?

- CAV – konstante Winkelgeschwindigkeit (Platte dreht sich mit konstanter Geschwindigkeit)
Vorteil: flexible Abspielmöglichkeiten (freeze frame, vorwärts oder rückwärts abspielen mit verschiedenen Raten)
Nachteil: schlechte Nutzung der Kapazität
DVD-RAM, Vorbild für die CD-DA
- CLV – konstante lineare Geschwindigkeit (Absenkung der Drehzahl nach außen hin)
Vorteil: doppelte Kapazität im Vergleich zu CAV
Nachteil: längere Suche und schlechtere Positionsmöglichkeit (nicht frameadressierbar), schlechtere Zugriffszeit
Standard für CD-ROM

Erklären Sie kurz die Abbildung von Media Time zur Speicheradresse in *Quicktime*. Welche Datenstrukturen werden dafür benötigt?

Zu einer bestimmten Media Time werden entsprechende Samples abgefragt welche Chunks zugeordnet sind, diese beinhalten die Information über die Speicheradresse.
Die verwendeten Datenstrukturen sind die Media und Data Entity.

Geben Sie eine verbale Beschreibung des Algorithmus, der für einen gegebenen QT-Track und einen gegebenen Zeitpunkt, die Adresse des zu diesem Zeitpunkt abzuspielenden Samples ermittelt.

- 1) Rechne Movie-Time mit Hilfe des TCS des Media-Entities in Media-Time um.
- 2) Bestimme zum diskreten Zeitpunkt das entsprechende Sample mit Hilfe der Time-To-Sample-Tabelle.
- 3) Bestimme zum Sample den Chunk mit Hilfe der Sample-To-Chunk-Tabelle. In diesem Chunk sind die Daten enthalten, die zum gefragten Zeitpunkt abgespielt werden müssen. Die Adresse des Chunks ist bekannt.

Besonderheiten in den Tabellen:

Die Time-To-Sample-Tabelle enthält Samples nicht, wenn sie die gleiche Länge haben das vorherige.

Die Sample-To-Chunk-Tabelle enthält Chunks nicht, wenn sie die gleiche Sample-Anzahl wie der vorherige Chunk haben und gleich codiert sind.

Was ist ein Frequenzspektrum? Wie unterscheiden sich die Frequenzspektren von Klängen und Geräuschen?

Frequenzspektrum:

Jede periodische Schwingung beliebiger Wellenform (= Klang) kann als eine Überlagerung von Grundschiwingung und Oberschwingungen dargestellt werden. Trägt man in einem Diagramm die Amplituden aller beteiligten Schwingungen in Abhängigkeit der Frequenz auf, so erhält man das sogenannte Frequenzspektrum

Diskretes Spektrum:

Klänge enthalten nur Schwingungen, deren Frequenzen ein ganzzahliges Verhältnis zur Grundfrequenz aufweisen. Kontinuierliches Spektrum: Allgemeine Schallereignisse (Geräusche) enthalten eine unendliche Anzahl von Einzelschwingungen, deren Frequenzwerte sich kontinuierlich über die x-Achse erstrecken. Es entsteht eine kontinuierliche mathematische Funktion: $p=p(f)$

Was versteht man unter *Differential Encoding*? Nennen Sie Beispiele, wo das Verfahren eingesetzt wird.

Differential Encoding ist Teil des Sourcecodings.

Betrachtet man eine Sequenz von Symbolen $S_1, S_2, S_3...$ mit von null verschiedenen Werte, die sich aber nicht sehr stark unterscheiden, so berechne die Differenz zum vorhergehenden Wert.

Nullen können mit Lauflängenkodierung kodiert werden.

MPEG verwendet die Kodierung (\rightarrow motion compensation, motion vectors

KOMPRESSION

Erklären Sie das Prinzip der *Laufängenkodierung* (Run Length Encoding).

Entropy-coding Algorithmus, inhaltsabhängige Codierung

Ersetzung einer Sequenz gleicher aufeinanderfolgender Bytes durch deren Anzahl, gekennzeichnet durch ein spezielles Flag!

Methode: Wenn die selben Bytes mindestens 4 mal hintereinander vorkommen, zähle die Anzahl, wie oft sie hintereinander vorkommen

Schreibe die komprimierten Daten in der Form: gezähltes Byte + '!' + Vorkommensanzahl

Beispiel: unkomprimiert: ABCCCCCCCCDEFFFFGGG

komprimiert: ABC!9DEF!4GGG

Formulieren Sie einen Pseudocodealgorithmus, der in einer quantisierten DCT-Matrix die Runlength-Kodierung durchführt.

```
i=0;
Solange (i < 63)
{
  Wenn (Matrix(i) != Matrix(i+1)) (im Zig-Zag-Scan){
    Gib Matrix(i) aus
    i=i+1
  }
  Sonst{
    count=i+1
    Solange ((Matrix(i)=Matrix(count) AND (count<63)){
      count=count+1
    }
    count=count-i
    z = 0;
    Wenn (count >= 4){
      Gib Matrix(i) aus
      Gib ! aus
      Gib count aus
    }
    Sonst{
      solange (z < count) gib Matrix(i) aus
    }
    i=count+1
  }
}
```

Diskutieren Sie kurz 3 Kriterien Ihrer Wahl zum Vergleich von Videokomprimierungsverfahren.

Bildqualität: Ist die Bildqualität ausschlaggebend und eine hohe Qualität benötigt, ist es notwendig, auf verlustfreie Kompressionsverfahren zurückzugreifen. Dem stehen allerdings eine niedrigere Kompression und damit eine größere Größe gegenüber. Optimal wäre eine hohe Qualität bei höchstmöglicher Kompression.

Zeitaufwand für En- und Decodierung, Komplexität: In vielen MM-Anwendungen ist es notwendig, dass die

Verarbeitung von Videos in Echtzeit möglich ist. (Videoschnitt, aber auch Videokonferenzen,...) Je komplexer ein Verfahren ist, desto schwieriger ist dies zu ermöglichen. Daher muss ein Kompromiss zwischen Qualität und Zeitaufwand gefunden werden.

Editierbarkeit, Navigation: Wie einfach und gut ist bzw. bleibt das komprimierte Video editierbar? Aber auch – ist ein fast forward, fast backward möglich?

Datenrate: Ist die Datenrate hoch, werden hohe Bandbreiten zur Übertragung benötigt, allerdings ist die Qualität auch viel besser. MPEG-1 ist für niedrige Datenraten ausgelegt, MPEG-2 hingegen benötigt hohe Datenrate, hat aber eine hohe Qualität (bis 15Mbit/sec)

Erklären Sie kurz die Begriffe *Profiles* und *Levels* im Kontext von MPEG-2.

Das Angebot unterstützter Codierung ist in Profile und Level unterteilt. Für jedes Profil/Level bietet MPEG-2 die Syntax für den kodierten Bitstream sowie die Dekodierungsanforderungen an.

Ein Profil ist eine definierte Teilmenge der gesamten Bitstream Syntax. Innerhalb eines Profils ist ein Level als eine Reihe von Bedingungen für die Parameter des Bitstreams (zum Beispiel für die Auflösung, die maximale Bitrate)

Profile: Simple, Main, 4:2:2, SNR, spatial, high, multiview

Levels: low (SIF), main (CCIR 601), high-1440, high (HDTV)

Profile und Levels stehen in hierarchischer Beziehung. Die Syntax, die von einem höheren Profil/Level unterstützt wird, inkludiert alle syntaktischen Elemente der niedrigeren Profile/Levels. Dekoder eines bestimmten Profils/Levels sollten demnach auch in der Lage sein Bitstreams niedrigerer Profile/Levels zu dekodieren – Ausnahme: Dekoder des simple Profils am main-Level müssen auch das main Profil für Low-Level Bitstreams dekodieren können (MPEG-1).

Ein Profile definiert aus der Menge der von der Norm zur Verfügung gestellten Verfahren eine bestimmte Anzahl, die in diesem Profile genutzt werden darf. Jeder Decoder, der dieses Profile beherrscht, kann den Datenstrom verarbeiten. Die Level legen innerhalb eines Profils wichtige Bildparameter fest. Hierzu gehören u.a. die unterschiedlichen Bildauflösungen, die für die Decodierung benötigten Speichergrößen und die maximalen Größen der Bewegungsvektoren. Die Profiles und Levels sind abwärtskompatibel. Das bedeutet, daß ein höheres Profile oder ein höherer Level alle niederwertigen Profiles und Levels mit beinhaltet.

Erklären Sie kurz die Huffman Kodierung von DC-Koeffizienten.

JPEG:

Kategorisieren der DC Werte in DC Codetabellen

Verschiedene Amplitudenkategorien für DC-Koeffizienten (12 Kategorien)

Behandle SSSS (Anzahl der für die Diff-Kodierung benötigten bits) als Huffman-Symbol, hole $p(0)$, $p(1)$,... $p(11)$ und erstelle einen Huffman-Baum → Huffman Code für SSSS

Für jede Kategorie wird ein zusätzliches Bitfeld angefügt um die Differenz, die in der Kategorie tatsächlich auftrat, eindeutig zu identifizieren.

wenn SSSS=2 den Huffman Code 001 und die Differenz -3 hat → sende 00100

001: HuffmanCode für 2, 00 entspricht -3 in Zweierkomplementdarstellung

Formulieren Sie in Pseudocode einen Algorithmus, der für die im Array *Symbole* gespeicherten Zeichen eine Huffman-Kodierung durchführt. Das Feld *Symbole* enthält in der 1. Komponente das zu kodierende Zeichen, in der 2. Komponente seine Wahrscheinlichkeit und in der 3. Komponente die zu errechnende Kodierung.

Erstelle einen Array *e*, der die sortieren *Symbole[z][p][c]* aufsteigend nach *p* enthält;

```
Für alle Elemente in e {
    Für jedes Zeichen aus e[i]
        c[e[i].charAt(j)][c] = „1“ + c[e[i].charAt(j)][c];
    Für jedes Zeichen aus e[i+1]
        c[e[i+1].charAt(j)][c] = „0“ +
        c[e[i+1].charAt(j)][c];
    e[e[i]+e[i+1]][p] = Symbole[e[i]][p] + Symbole[e[i+1]][p];
    lösche e[i] und e[i+1];
    sortiere e neu;
}

Symbole.c = c;
```

Welchen Zweck hat die *Diskrete Cosinus-Transformation* in Kompressionsverfahren (zb JPEG).

Mittels Diskreter Cosinus-Transformation werden die 8x8 Pixelblöcke eines unkomprimierten Bildes von der Zeitebene in die Frequenzebene transformiert. Dadurch erhält man Koeffizienten $S(u,v)$, wobei $S(0,0)$ die kleinste Frequenz(0) in beiden Richtungen des Blockes aufweist und damit DC-Koeffizient genannt wird; er stellt die Grundfarbe des Blocks dar. Die anderen ($S(0,1), \dots, S(7,7)$) sind die sogenannten AC-Koeffizienten, wobei die Frequenz von zumindest einer Richtung ungleich 0 ist.

Anhand der Darstellung im Frequenzraum können zur Kompression der Daten hohe Frequenzen reduziert werden bei Beibehaltung einer sehr guten Qualität. Niedrige Frequenzen entsprechen den groben Konturen von Bildobjekten, währenddessen hohe Frequenzkomponenten feine Strukturen darstellen. Dies ist möglich, weil das menschliche Auge nur begrenzt fähig ist hohe Ortsfrequenzen wahrzunehmen.

Erklären Sie kurz den Begriff *Quantisierung* im Kontext der Sampling-Theorie und im Kontext von Komprimierungsverfahren.

Sampling:

die abgetasteten analogen Werte werden diskreten Zahlenwerten zugeordnet (quasi eine Rundung). Der gesamte Spannungsbereich wird dazu in sogenannte Quantisierungsintervalle aufgeteilt und die abgetasteten Werte werden dem nächstgelegenen diskreten Wert zugeordnet.

Komprimierung

Jeder Wert in einer Frequenzmatrix wird durch seinen zugehörigen Wert in der Quantisierungstabelle dividiert und auf Integer gerundet um möglichst viele AC-Koeffizienten abzuschwächen oder gar zu eliminieren. Die

Quantisierungsmatrix ist deshalb so aufgebaut, dass deren Werte von links oben nach rechts unten zunehmen, da bei der Frequenzmatrix korrespondierend die Frequenzen von links oben nach rechts unten zunehmen und somit mehr komprimiert werden.

Worin unterscheiden sich die Quantisierungsschritte im JPEG- und MPEG-Verfahren?

JPEG:

Ziel: hinauswerfen von Bits (Kürzung)

Uniforme Quantisierung: Dividieren die Koeffizientenwerte $S(u, v)$ durch N und runde das Ergebnis

Anhand der Quantisierungstabelle wird entschieden, wie viele Bits gekürzt werden sollen. Sie besteht aus 64 Elementen, wobei jedes 8-bit verwendet: Q_{uv}

Die neuen komprimierten Werte ergeben sich durch die Verwendung der Tabellen: $S_{q_{uv}} = S_{uv} / Q_{uv}$

zwei verschiedene Standardtabellen: Helligkeit und Farbe

MPEG:

AC Koeffizienten der B- und P-Frames haben üblicherweise große Werte, I-Frames haben kleinere Werte

I-Frames: alle den selben Quantisierungswert; B-, P-Frames: für jeden Block separat ermittelt

Die MPEG-Quantisierung ist angepasst:

Wenn die Datenrate über einen Schwellwert steigt, vergrößert die Quantisierung die Schrittweite
wenn die Datenrate absinkt wird die Quantisierung mit einer feineren Körnung ausgeführt

Conclusion:

JPEG-Quantisierung ist uniform, MPEG-Quantisierung wird der aktuellen Gegebenheit angepasst

Was sind Quantisierungstabellen und weshalb setzt man sie ein?

Warum: zur Komprimierung der Koeffizienten bei JPEG. Die Koeffizienten werden durch die Werte in der Tabelle jeweils dividiert und ergeben die komprimierten Werte der einzelnen Koeffizienten.

Die Tabellen bestehen aus 64 Elementen zu je 8 bit. Die Werte oben links sind kleiner, da sie den niedrigeren Frequenzen zugeordnet werden und damit werden die niedrigeren Frequenzen im Vergleich weniger stark komprimiert. Dies hat den Grund, dass niedrigere Frequenzen vom menschlichen Auge sehr gut wahrgenommen werden, höhere hingegen zunehmend schlechter. Daher werden hohe Frequenzen zunehmend mehr komprimiert.

Erklären Sie kurz die P- und B-Frame Kodierung der MPEG-Komprimierung

P-Frame Codierung – predictive-coded Frames

Sie benötigen Informationen des vorhergehenden I-Frames und/oder vorhergehender P-Frames für Kodierung sowie Dekodierung. Bei der P-Frame Codierung macht man sich aufeinanderfolgende Bilder zu Nutze, in welchen Teile sich nicht verändern bzw. nur „verschoben“ werden.

Zeitliche Redundanz: Bestimmen den letzten P- oder I-Frame, der dem aktuellen Block am ähnlichsten ist.

Verwende die *Motion estimation* Methode am Encoder.

Motion Estimation Method: Für jeden Makroblock wird ein Vektor errechnet, der angibt, wo dieser in dem darauffolgenden Referenzbild seine ähnlichste Entsprechung findet. Die beiden Makroblocks werden anschließend voneinander subtrahiert und DCT-kodiert. Die Dekodierung/Kodierung erfolgt also mit Infos bzgl. vorangegangener I- oder P-Bilder und die Kompressionsrate liegt daher auch höher als bei I-Bildern.

B-Frame Codierung (bidirektional predictiv-coded)

benötigen Information des vorhergehenden und nachfolgenden I und/oder P-Frames

Motionvektor des vorhergehenden und nachkommenden Referenzframes

Durchschnitt der Blöcke des vorhergehenden und späteren Frames

Für die B-Frame-Kodierung ist es nötig, auch das Vorgängerbild einzubeziehen. $A = \text{Block BlockNext}$, $B = \text{Blockc-BlockLast}$; Das Ergebnis ergibt sich aus einem Multiplexing von A , B und $(A+B)/2$. Es wird anschließend DCT-Kodiert. Die Dekodierung/Kodierung erfolgt also mit Infos bzgl. vorangegangener und später auftretender I- oder P-Bilder und die Kompressionsrate ist daher am höchsten.

Beschreiben Sie kurz die Aufgabe von D-Frames im Rahmen von MPEG. Wie werden D-Frames kodiert?

DC-kodierte Frames

Verwendung für fast forward, fast rewind Modi

DC-Parameter sind DCT-codiert, AC Koeffizienten werden vernachlässigt

D-Frames beinhalten somit die niedrigsten Frequenzen des Bildes

Es wird nur ein Typ von Makroblöcken verwendet und nur in MPEG-1.

Was versteht man unter Group of Pictures? Wie hängen Gop Muster mit der Datenrate zusammen?

Die **Group of Pictures (GOP)** ist eine Gruppe von aufeinanderfolgenden [MPEG](#)-Bildern, engl. [Frames](#), innerhalb eines [MPEG](#)-codierten Films bzw. [Videostreams](#). Jeder MPEG-codierte Film bzw. Videostream besteht aus aufeinanderfolgenden GOPs. Aus den darin enthaltenen MPEG-Bildern werden die sichtbaren Einzelbilder generiert.

Eine GOP kann folgende Bildtypen enthalten:

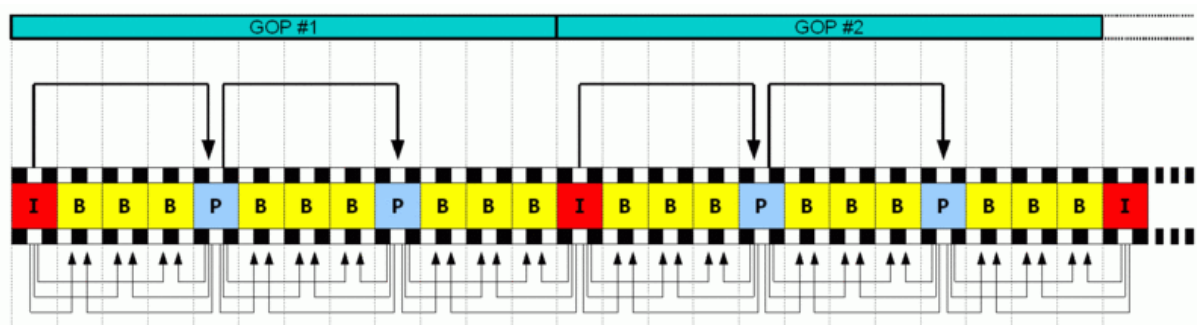
I-Bild bzw. **I-Frame** (engl. intra-coded picture) Referenzbild, entspricht einem Standbild und ist von anderen Bildtypen unabhängig. Jede GOP beginnt mit diesem Bildtyp.

P-Bild bzw. **P-Frame** (engl. predictive-coded picture) enthält Differenz-Informationen aus dem vorhergehenden I- oder P-Frame.

B-Bild bzw. **B-Frame** (engl. bidirectionally predictive-coded pictures) enthält Differenz-Informationen aus dem vorhergehenden und/oder nachfolgenden I- oder P-Frame.

D-Bild bzw. **D-Frame** (engl. DC direct coded picture) dient dem schnellen Vorlauf.

Eine GOP beginnt stets mit einem I-Frame. Danach folgen ein oder mehrere P-Frames, jeweils mit einigen Frames Abstand. In den verbleibenden Lücken befinden sich B-Frames. Mit dem nächsten I-Frame beginnt eine neue GOP. Die folgende Grafik zeigt einen typischen Aufbau von GOPs innerhalb einer MPEG-Datei. Die Pfeile kennzeichnen die Referenzen.



Zusammenhang mit der Datenrate:

Bei größeren GOP's werden weniger I-Frames benötigt. Da I-Frames aber die geringste Komprimierung erreichen (Intraframe-Kompression) und B-, und P-Frames viel besser komprimiert werden können (und kleiner sind), sinkt mit zunehmender GOP-Größe und damit abnehmender I-Frame-Anzahl die Datenrate.

Nennen Sie kurz 2 Arten von Artefakten bei der *Audiokodierung*.

Bandbreitenverlust

Pre-Echos

Die Verschlechterung der Zeitauflösung hat zur Folge, dass ein Pre-Echo-Effekt auftreten kann. Als Pre-Echo bezeichnet man solche Geräusche, die vor dem eigentlichen Signal entstehen. Beobachten kann man es bei solchen Signalen, die mit einer großen, plötzlichen Lauts Um dem Pre-Echo-Effekt vorzubeugen, werden bei der MDCT unterschiedliche Transformationsblocklängen gewählt. Lange Blöcke werden bei stationären, sich also zeitlich wenig ändernden Signalen eingesetzt, um die Kompression zu erhöhen. Kurze Blöcke werden bei sich schnell ändernden Signalen verwendet, um den Bereich, in dem Pre-Echo auftreten kann zu verkleinern. tärkenänderung beginnen, insbesondere wenn ein Bereich mit sehr niedrigen Amplituden vorausgeht.

roughness, double-speak

Bei niedrigen Bitraten und niedrigeren Samplingfrequenzen tritt eine Fehlanpassung bzw. ein Ungleichgewicht zwischen der zeitlichen Auflösung des Coders und den Anforderungen, die bestimmte Signale aufgrund ihrer Zeitstruktur stellen, auf.

Nennen und erläutern Sie kurz 2 Arten von *Scalable Bit Streams* in MPEG-2.

Spatial scalability Kodiert einen Basislayer mit niedrigeren Samplingdimensionen (Auflösungen) als die höheren Layer

Data partitioning Teilt den Block von 64 quantisierten Koeffizienten in 2 Bitstreams. Der erste, höher priorisierte Bitstream beinhaltet die kritischeren niederfrequenten Koeffizienten und Randinformationen (DC Werte, Bewegungsvektoren)

SNR scalability: Kanäle sind mit derselben Samplerate kodiert, aber mit verschiedener Bildqualität (durch Quantisierung)

Temporal scalability Bitstream mit höherer Priorität kodiert Video mit niedrigerer Framerate, mittlere Frames können in einem zweiten Bitstream kodiert werden unter der Verwendung der Rekonstruktion des ersten Bitstreams als Prognose

Was können Sie über die AC und DC Koeffizienten der DCT aussagen, wenn der zu transformierende 8x8 Block eines Bildes aus einem Schachbrettmuster besteht?

DC → mittel

AC → ungleich null, konzentriert zu den höheren Frequenzen (entspricht einem idealen Schachbrettmuster)

Was können Sie über die AC und DC Koeffizienten der DCT aussagen, wenn im zu transformierenden 8x8 Block eines Bildes 2 Flächen in einer horizontalen Linie schneiden?

Scharfe Schnitkante → höchste Frequenzen

horizontale Linie → vertikal in der DCT – Darstellung

DC-Koeffizient: mittel (wenn sich beide Flächen genau in der Mitte schneiden und eine schwarz und eine weiß ist; spiegelt die Grundfarbe des Blocks wider)

AC-Koeffizienten: erste (und zweite) Spalte von links ungleich 0, die anderen sollten 0 sein.

Was können Sie über einen 8x8 Pixelblock aussagen, der nach der DCT-Transformation nur in den ersten beiden Spalten Werte aufweist, die von 0 verschieden sind?

Es treten in vertikaler Richtung (im Bild, nicht in der transformierten Darstellung!) nur sehr niedrige Frequenzen $s(1,0)$ auf, in horizontaler Bildrichtung jedoch alle → es dürfte sich um ein Bild haben, in dem sich zum Beispiel zwei verschiedenfarbige Flächen horizontal mit einer scharfen Kante schneiden.

Ermitteln Sie eine Huffman-Codierung der Symbole A, B, C, D und E:

Wahrscheinlichkeiten ihres Auftretens: $p(A) = 0.25$, $p(B) = 0.41$, $p(C) = 0.18$, $p(D) = 0.14$, $p(E) = 0.02$

Verschiedene Möglichkeiten, je nach Beschriftung der Kanten.

Wozu dienen Spatial Scalability und SNR Scalability? Was haben sie gemein und was sind die Unterschiede?

SNR scalability: Kanäle sind mit derselben Samplerate kodiert, aber mit verschiedener Bildqualität (durch Quantisierung)

Spatial scalability Kodiert einen Basislayer mit niedrigeren Samplingdimensionen (Auflösungen) als die höheren Layer

Unterschiede, Gemeinsamkeiten:

beides skalierbare Codiermodi von MPEG-2, aber spatial scalability verwendet 2 fixe Layer, SNR ist wirklich frei skalierbar.

Was sind die wesentlichsten Unterschiede zwischen MPEG-1 und MPEG-2?

MPEG-2 hat zusätzlich:

zusätzliche Syntax für effiziente Kodierung von interlaced Videos

10-bit DCT DC Koeffizienten (MPEG-1 hat 8 bits)

Skalierbare Erweiterungen, die eine Aufteilung eines kontinuierlichen Videosignals in 2 oder mehrere kodierte Bitstreams erlaubt, die das Video mit verschiedenen Auflösungen, Bildqualitäten oder Bildraten darstellen.

mehrere Bildformate
Progressives und interlaced Framekodierung
zusätzliche Prediction Modi
4 Skaliermodi
Verbesserte Bildqualität (Quantisierung, Zig-zag-scan)

Erhöhung der Genauigkeit der Bewegungsvektoren auf halbe Bildpunkte
erweiterte Fehlerredundanz durch spezielle Vektoren bei I-Frames
Präzision der DCT wählbar

Für welche Einsatzbereiche würden Sie das Motion JPEG-Verfahren dem MPEG-Verfahren vorziehen?

- MotionJPEG ist vorzuziehen, wenn man gezielt auf Einzelbilder zugreifen möchte (z.B. nonlinearer Videoschnitt, Navigable Video)
- Bei Harddisc-Recording. Verfahren, bei denen viele unterschiedliche Datenraten zum Einsatz kommen.

Was versteht man unter *Pre-Echo*? Wo tritt dieses Phänomen auf?

Pre-Echo ist ein Artefakt, das bei der Audio-Kodierung entsteht.

Die Verschlechterung der Zeitauflösung hat zur Folge, dass ein Pre-Echo-Effekt auftreten kann. Als Pre-Echo bezeichnet man solche Geräusche, die vor dem eigentlichen Signal entstehen. Beobachten kann man es bei solchen Signalen, die mit einer großen, plötzlichen Lauts um dem Pre-Echo-Effekt vorzubeugen, werden bei der MDCT unterschiedliche Transformationsblocklängen gewählt. Lange Blöcke werden bei stationären, sich also zeitlich wenig ändernden Signalen eingesetzt, um die Kompression zu erhöhen. Kurze Blöcke werden bei sich schnell ändernden Signalen verwendet, um den Bereich, in dem Pre-Echo auftreten kann zu verkleinern. tärkenänderung beginnen, insbesondere wenn ein Bereich mit sehr niedrigen Amplituden vorausgeht.

Beschreiben Sie kurz die wesentlichen Schritte der MP3-Kodierung. Erklären Sie insbesondere Quantisierung und Kodierung.

MP3 bezeichnet nicht ein unabhängiges Verfahren zur Audiokomprimierung, sondern entspringt der Familie der MPEG-I Codecs. MP3 ist die Abkürzung für MPEG-1-Layer-III, und bietet gegenüber den Layer-I und Layer-II Codecs zusätzliche Funktionalität.

Folgende Algorithmen finden in allen drei Layern Verwendung:

- Im ersten Schritt des Kompressionsprozesses wird das digitale Eingangssignal in Subbänder geteilt. Dieser Prozess wird auch als Zeit-Frequenz-Zuordnung bezeichnet.
- Der nächste Schritt der Kodierung ist die Implementierung des psychoakustischen Modells
- Im dritten Schritt erfolgt die Quantisierung und Encodierung jedes einzelnen aus Stufe 1 und 2 hervorgegangenen Samples in den Subbändern.
- Der letzte Schritt ist die endgültige Erzeugung der komprimierten Datei, wobei die digitalisierten Audiodaten in einzelne Frames unterteilt werden.

Zusätzlich zu diesen 4 Schritten sieht MPEG1-Layer III noch zwei weitere Schritte vor, die zwar die Zuverlässigkeit des psychoakustischen Modells erhöhen und den Ausgangsdatenstrom so klein wie möglich halten, aber allerdings zu einer längeren Encodierungszeit bei der Erstellung der Ausgangsdatei führen:

- Neben der Implementierung des psychoakustischen Modells aus Schritt 2 kommen noch mehrere Filter- (z.B.: Hoch- Tiefpassfilter) und "Diskrete Cosinus-Transformations"-Algorithmen zum Einsatz.

- In der letzten Phase, in der die Frames vorbereitet werden, ist ein komplexer Algorithmus hinzugefügt, der die Framegröße nach bestimmten Parametern variieren lässt. Es wird auch das sogenannte reserve Bit gesetzt, dass eventuell auftretende kritische Verzerrungen (Artefakte) bei der Encodierung des Samples löst.

Quantisierung und Kodierung:

Die Spektralkomponenten werden quantisiert und kodiert. Störungen soll unter einem Schwellwert gehalten werden.

zwei verschachtelte Iterationsschleifen (rate-loop und Noise-control-loop)

quantisiert wird mittels eines potenzförmigen Quantisierers (größere Werte werden automatisch mit weniger Genauigkeit codiert)

Quantisierte Werte werden nach Huffman codiert

Was sind – abgesehen von Leistungsunterschieden – die wichtigsten Unterschiede zwischen MP3- und AAC-Kodierung?

MP3: MPEG-1 Layer 3

AAC: MPEG-2 Advanced Audio Coding

AAC: bessere Komprimierung, rückwärtskompatible Mehrkanalkodierung fügt die Möglichkeit von vorwärts- und rückwärtskompatibler Codierung von Mehrkanalsignalen (inklusive 5.1 Kanalkonfigurationen)

Zusätzlich erweitert AAC den MPEG-1 Audiostandard um das Kodieren mit niedrigeren Samplingfrequenzen (16kHz, 22.05kHz und 24kHz)

Worin unterscheiden sich die I-Frame-Kodierung des MPEG-Verfahrens von der JPEG-Kodierung?

"i-frames use 8x8 blocks defined within a macroblock, on these blocks DCT is performed. Quantization is by constant value for all DCT coefficients."

Im gelben Skriptum habe ich noch einen weiteren Unterscheid gefunden, der aber ziemlich ins Detail geht. Es wird ja bei den DC-Koeffizienten die Differenz zum vorherigen berechnet. Diese Differenz wird nicht direkt abgespeichert, sondern man nimmt die Anzahl der benötigten Bits für diese Differenz, bestimmt dafür einen Huffman-Code, und speichert schließlich das Tupel (<Huffman code für die Bit-Anzahl>, <diff>). Das ganze gilt bei JPEG. Bei MPEG gibt es einen kleinen Unterschied: hier wird nur für die häufigsten Längen (Bit-Anzahlen) ein Huffman-Code generiert. Die seltenen werden mit einer Escape-Sequenz direkt gespeichert.

I-Frames werden mit einer konstanten Quantisierung komprimiert und bei JPEG werden Quantisierungsmatrizen definiert.

Erklären Sie knapp den Unterschied zwischen *Inter-* und *Intraframesynchronisation*. Geben Sie 2 Geräte an, die der Synchronisation dienen.

Inter-Frame-Synchronisation meint die Zusammenführung von 2 Strömen von unterschiedlichen Quellen, so dass die gleichen Frames und die gleichen Zeilen genau zeitgleich ankommen, so dass man sie dann vernünftig verbinden kann (z.B. mit einem video production switcher). Man verwendet dazu einen timebase corrector, der

von einem sync generator einen Takt bekommt.

Dagegen meint die Intra-stream-Synchronisation die zeitliche Korrektur eines einzelnen Streams, um die sonst auftretenden Fehler (jitter und skew) zu vermeiden. Dazu verwendet man einen Frame Delay Buffer, der das eingehende Signal zwischenspeichert und genau zu bestimmten Zeitpunkten weitergibt (getriggert von einem sync generator), egal wann es eingetroffen ist.

Interframe - Sync von Audio und Video um auseinanderlaufen von Ton / Bild zu verhindern

Intraframe - Sync von Videodaten untereinander um ruckeln zu vermeiden

Erklären Sie kurz die verschiedenen Frametypen der MPEG-Komprimierung.

I-Frames (intracoded Frames)

Benötigen keine Referenz auf andere Bilder und werden wie stehende Bilder behandelt. MPEG verwendet JPEG für I-Frames.

Komprimierung muss in Echtzeit passieren, deswegen ist die Komprimierungsrate die geringste innerhalb von MPEG.

I-Frames sind Punkte für den Zufallszugriff in einem MPEG-Stream

I-Frames verwenden 8x8 Blöcke innerhalb eines Makroblocks, auf diesen wird DCT ausgeführt und für alle DCT Koeffizienten mit konstantem Wert quantisiert.

P-Frame Codierung – predictive-coded Frames

Sie benötigen Informationen des vorhergehenden I-Frames und/oder vorhergehender P-Frames für Kodierung sowie Dekodierung. Bei der P-Frame Codierung macht man sich aufeinanderfolgende Bilder zu Nutze, in welchen Teile sich nicht verändern bzw. nur „verschoben“ werden.

Zeitliche Redundanz: Bestimme den letzten P- oder I-Frame, der dem aktuellen Block am ähnlichsten ist.

Verwende die *Motion estimation* Methode am Encoder.

Motion Estimation Method: Für jeden Makroblock wird ein Vektor errechnet, der angibt, wo dieser in dem darauffolgenden Referenzbild seine ähnlichste Entsprechung findet. Die beiden Makroblocks werden anschließend voneinander subtrahiert und DCT-kodiert. Die Dekodierung/Kodierung erfolgt also mit Infos bzgl. vorangegangener I- oder P-Bilder und die Kompressionsrate liegt daher auch höher als bei I-Bildern.

B-Frame Codierung (bidirektional predictiv-coded)

benötigen Information des vorhergehenden und nachfolgenden I und/oder P-Frames

Motionvektor des vorhergehenden und nachkommenden Referenzframes

Durchschnitt der Blöcke des vorhergehenden und späteren Frames

Für die B-Frame-Kodierung ist es nötig, auch das Vorgängerbild einzubeziehen. $A = \text{Block BlockNext}$, $B = \text{Block-BlockLast}$; Das Ergebnis ergibt sich aus einem Multiplexing von A, B und $(A+B)/2$. Es wird anschließend DCT-kodiert. Die Dekodierung/Kodierung erfolgt also mit Infos bzgl. vorangegangener und später auftretender I- oder P-Bilder und die Kompressionsrate ist daher am höchsten.

D-Frames (DC-kodierte Frames)

Verwendung für fast forward, fast rewind Modi

D-Frames beinhalten somit die niedrigsten Frequenzen des Bildes

DC-Parameter sind DCT-codiert, AC Koeffizienten werden vernachlässigt

Es wird nur ein Typ von Makroblöcken verwendet und nur in MPEG-1.

Was versteht man unter *Perceptual Measurement Techniques* und wofür werden sie eingesetzt?

Zur Messung der Qualität von Codecs.

State-of-art ist nicht ausreichend um große Skala und die gut vorbereitete Hörtests zu überholen.

PMT sind aber bis zu dem Punkt gekommen, in dem sie eine sehr nützliche Ergänzung zu den Hörtests sind und sie in einigen Fällen ersetzen können.

ITU-R Task Group 10/4 entwickelte eine Empfehlung für das System PEAQ.

Beschreiben Sie kurz die MPEG-4 Videocodierung

Der MPEG-4-Standard besteht generell aus einem Systemteil, einem Audio-Teil und einem Visual-Teil. Jeder einzelne Videoframe wird in eine Anzahl von willkürlich geformten Regionen unterteilt, die video object planes (VOP) genannt werden. Aufeinander folgende VOPs die zum selben physikalischen Objekt in einer Szene (z.B. ein Mensch) gehören werden als video objects bezeichnet. Die Form, Bewegung und Textur der VOPs, die zum selben VO gehören, werden jeweils in einem separaten Video object layer (VOL) enkodiert. Relevante Informationen die benötigt werden, um jedes der VOLs zu identifizieren, sowie wie die verschiedenen VOLs zusammengesetzt werden, werden auch kodiert. Dadurch wird dann auch selektives Dekodieren möglich und außerdem erlaubt es uns einzelne object levels getrennt zu skalieren. So kann man z.B. ein Objekt, das als separates VO kodiert wurde vergrößern und in eine neue Umgebung platzieren.

Erklären Sie kurz die Begriffe *Video Object*, *Video Object Plane* und *Video Object Layer*. In welchem Kontext werden sie verwendet?

Kontext: MPEG-4 Codierung

Video Object Plane:

Jeder Videoframe ist in eine Anzahl willkürlich geformter Regionen, die sogenannten Video Object Planes, segmentiert.

Video Objects:

Aufeinanderfolgende VOPs, die zum selben physikalischen Objekt in einer Szene gehören, werden als Video Objects bezeichnet.

Video Object Layer:

in einem Video Object Layer sind die Form-, Bewegungs- und Texturinformationen von VOPs, welche zum selben VO gehören, kodiert. Außerdem sind relevante Informationen, welche zur Identifikation jedes VOLs und wie verschieden VOLs zusammengesetzt sind, kodiert. Dies erlaubt eine selektive Dekodierung und bietet Object-Level Skalierbarkeit.

Erklären Sie die Begriffe *Video Object* und *Video Object Plane*. In welchem Kontext werden sie verwendet?

Kontext: MPEG-4 Codierung

Video Object Plane:

Jeder Videoframe ist in eine Anzahl willkürlich geformter Regionen, die sogenannten Video Object Planes, segmentiert.

Aufeinanderfolgende VOPs, die zum selben physikalischen Objekt in einer Szene gehören, werden als Video Objects bezeichnet.

Welchem Zweck dienen *BIFS* in MPEG-4? Welche Informationen enthalten sie?

BIFS: Binary Format for Scene Description. MPEG-4 definiert für mehr Effizienz einen Weg um Szenenbeschreibung binär zu enkodieren.

BIFS ist die Abkürzung für Binary Format for Scenes, eine auf VRML97 basierende und in MPEG-4 part 11 (ISO/IEC 14496-11, „Szenenbeschreibung und Anwendungs-Engine“) standardisierte Beschreibungssprache für zwei- und dreidimensionale multimediale audiovisuelle interaktive Inhalte. Sie wird binär codiert.

BIFS ist eine Szenenbeschreibungssprache, die in MPEG-4 eingesetzt wird und mit der die örtlich-zeitliche Zuordnung einzelner Objekte eines Datenstromes unabhängig voneinander beschrieben werden kann. Mit BIFS können Objekte beschrieben, hierarchisch geordnet und in einem zwei- oder dreidimensionalen Raum unter Berücksichtigung der Zeitachse gruppiert werden.

BIFS setzt auf die virtuelle Modellierungssprache VRML und kann verschiedene Szenen miteinander kombinieren. So könnten beispielsweise die Teilnehmer an einer Videokonferenz aus den verschiedenen Studio-Umgebungen ausgestanzt und in einem virtuellen Studio vereint werden. Bei dieser Technik kommt die Bluescreen-Technik zum Einsatz, die in der Studioteknik als Blue-Keying bekannt ist.

Erklären Sie kurz das Konzept *Komposition* im MPEG-4 Systems Standard.

Komposition ist ein Modell zur Beschreibung der Zusammenstellung einer komplexen MM-Szene (bestehend aus audiovisuellen Objekten (audio, video, natürlich, synthetisch, 2D, 3D, streamd oder downloaded), welche zusammengehörige MediaObjects kreiere, welche audiovisuelle Szenen formen) beruhend auf den Konzepten von VRML.

Autoren können die Beschreibung im Textformat generieren, möglich auch mit Hilfe eines Authoringtools. Der Effizienz wegen definiert der Standard das Binary Format for Scene Description.

MM-Szenen werden als hierarchische Strukturen, repräsentiert als Graphen, verstanden. Jedes Blatt des Graphen repräsentiert ein Mediaobjekt. Alle Kanten und Grafenblätter, die Streamingsupport benötigen um Medieninhalte zu erhalten, sind logisch mit den DekodingPipelines verbunden.

Was sind die Aufgaben des *Synchronization Layers* in MPEG-4.

Hinzufügen von MPEG-4 spezifischen Informationen für Timing und Synchronisation der kodierten Medien. Elementare Streams werden in paketierte, Header mit Timinginformation (Taktreferenzen) und Synchronisationsdaten hinzugefügt.

Beschreiben Sie kurz die *Facial Animation* und *Facial Definition Parameters* im Face Animation-Teil von MPEG-4.

Facial Animation

Beschäftigt sich mit den Parametern für Gesichtsanimation und Gesichtsdefinition.

FAP:

erlauben das Anwenden eines einzelnen Sets von Parametern unabhängig von dem verwendeten Gesichtsmodell. Die meisten FAPs beschreiben atomare Bewegungen von Gesichtszügen, andere (expressions und visemes) definieren komplexere Deformierungen.

Viseme definieren die Position des Mundes zusammenhängend mit den Phonemen.

Sie sind auf die Position und Bewegungsabläufe einzelner Gesichtsteile spezialisiert. FAPs können entweder arithmetisch kodiert werden oder mit DCT.

FDP:

werden zur Kalibrierung des Empfängerterminal Standardgesicht oder zur Übertragung komplett neuer Gesichtsmodellgeometrie und –textur verwendet.

Erklären Sie kurz den *Skalierfaktor* und den Zusammenhang mit Kompressionsverfahren.

- Skalierung ist eine Abwandlung der Bandrauschunterdrückungssysteme für Tonbandgeräte (DolbyA, B, C)
 - kleine Signalanteile vor Bandaufzeichnung verstärkt
 - bei Wiedergabe um selben Faktor gedämpft
 - Bandrauschen mit selben Faktor gedämpft
- Skalierung bei MPEG Kodierung
 - kleine Werte vor Kodierung mit Skalierfaktor multipliziert
 - bei Dekodierung mit selben Faktor dividiert
 - Quantisierungsrauschen um Skalierfaktor gedämpft
- Skalierung verbessert SNR für leise Signale

Erklären Sie kurz die *kritische Abtastung*

- Kritische Abtastung digitaler Filterbanken
 - Wortlängen bleiben in Subbands unverändert
 - Abtastrate in Subbands reziprok zur Anzahl der Subbands
 - Reduzierte Abtastrate in Subbands wird als kritische Abtastrate bezeichnet
- Kritische Abtastung der Subbands gewährleistet, dass Filterbank keinen Einfluss auf Datenmenge hat
 - Es werden zwar mehr Kanäle erzeugt, aber deren Abtastfrequenzen sind dementsprechend geringer.
 - Kritische Abtastrate ergibt sich aus Nyquist-Shannon-Theorem

Erklären Sie kurz den *Maskierungseffekt* und den Zusammenhang mit Kompressionsverfahren.

zwei Töne erklingen gleichzeitig mit unterschiedlicher Lautstärke und ähnlicher Frequenz

Maskierungseffekt – Ohr hört nur den lauten Ton, der leise Ton wird „maskiert“, denn der Ton mindert Empfindlichkeit des Gehörs im umliegenden Frequenzbereich. Der leise Ton liegt unter neuer

Hörschwellenkurve: Maskierungsschwellwert ist minimale Lautstärke, die der leise Ton aufweisen müsste, um nicht maskiert zu werden

- allgemeines Audiosignal besteht aus vielen Frequenzanteilen
 - jeder Frequenzanteil beeinflusst die Hörschwellenkurve
 - Audiosignal variiert mit der Zeit

- Hörschwelle zu jedem Zeitpunkt verschieden
- Psychoakustisches Modell
 - Berechnet jeweils aktuelle Hörschwellenkurve
- Variable Quantisierung
 - Frequenzkomponenten so quantisiert, dass Quantisierungsrauschen gerade unter aktueller Hörschwellenkurve
 - Wo maskiert wird, grobe Quantisierung, kleine Wortlängen

Erklären Sie kurz *Kritische Bandbreite* und den Zusammenhang mit Kompressionsverfahren.

- Psychoakustischer Effekt des Gehörs
- Frequenzen gleichzeitig erklingender Töne müssen bestimmten Mindestabstand haben, um als Töne verschiedener Tonhöhe wahrgenommen zu werden
- Kritische Bandbreite – umfasst jeweils einen Frequenzbereich, innerhalb welchem unser Gehör Töne nicht differenzieren kann
 - bezieht sich nur auf gleichzeitig gespielte Töne (bei sequenziell gespielten Tönen ist Frequenzauflösung des Gehörs wesentlich höher)
- elementar für viele psychoakustische Effekte und MPEG Kompression
- Töne unter 500 Hz – kritische Bandbreite konstant 100 Hz
Töne über 500 Hz – kritische Bandbreite ca. 20% der Frequenz selbst (entspricht ca. kleiner Terz)

Was versteht man unter *Structured Audio* in MPEG-4?

Erlaubt die Erstellung von synthetischen Sounds aus kodierten Inputdaten.

Spezielle Synthesesprache (Structured Audio Orchestra Language – SAOL) erlaubt die Definition eines synthetischen Orchesters, dessen Instrumente Klänge wie reale Instrumente erzeugen oder vorgespeicherte Sounds verarbeiten.

MPEG-4 standardisiert die Methode der Synthesebeschreibung, nicht die Methoden um Sound zu generieren.. Download von Noten/Partituren in den Bitstream kontrolliert die Synthese

Noten/Partituren ähneln Skripten in einer speziellen Sprache (Structured Audio Score Language – SASL), sie bestehen aus seinem Set von Befehlen für die verschiedenen Instrumente, wobei Befehle auf verschiedene Instrumente und zu verschiedenen Zeiten angewendet werden können.

Wie erfolgt die zeitliche Komposition (*Temporal Composition*) in MPEG-4 Datenströmen?

Der composition Stream (BIFS) hat sein eigene Zeitbasis. Auch wenn die Zeitbasen für Komposition und elementaren Datenstreams verschieden sind, müssen sie konsistent sein, abgesehen von Translation und Skalierung der Zeitachse.

Timestamps, welche bei den elementaren Medienstreams hinzugefügt werden, legen fest, zu welcher Zeit die Zugriffseinheit für ein Medienobjekt am Dekoderinput bereit sein muss und wann die Kompositionseinheit am Kompositorinput bereit sein muss.

Timestamps des Kompositionstreams bestimmen, wann die Zugriffseinheiten für Komposition am Input des Kompositionsinformationdekoders bereitgestellt sein müssen.

Felder innerhalb einer Szenenbeschreibung beinhalten auch einen Zeitwert. Er stellt entweder eine Dauer oder einen Moment als Zeit dar.

- stellen eine relative Zeit zum Timestamp des BIFS Elementarstreams dar.

Nennen und erklären sie die wichtigsten Eigenschaften, nach denen zeitabhängige Ströme (Timed Streams) klassifiziert werden können.

Ein Timed stream ist eine endliche Folge von Tupeln der Form $\langle e_i, s_i, d_i \rangle$, $i=1, \dots, n$. Jede Mediasequenz basiert auf einem Medientyp M und einem diskreten Zeitkoordinatensystem D .

Eigenschaften:

Homogenität (Gleichwertigkeit): Elementdeskriptoren sind konstant

Kontinuität (Beständigkeit): $s_{i+1} = s_i + d_i$ für $i=1, \dots, n-1$

Frequenz: d_i konstant für $i = 1, \dots, n-1$

Datenrate: Elementgröße/Dauer ist konstant

Uniformität: Elementgröße und Dauer sind konstant

Nennen und diskutieren Sie kurz zwei Gründe für das Konzept *Ableitung* (Derivation).

Niedrigere Speicher- und Netzwerkkosten: Durch die diversen Schritte der Ableitung, also z.B. Verkettung (Konkatenation) oder Konvertierung (Konversion) wird über allem Speicherplatz gespart. Modifikationen gehen effizienter vonstatten und Ableitung sorgt für Datenunabhängigkeit.

Nennen Sie 3 Gründe, weshalb die Multimedia – Programmierung fast ausschließlich dem objektorientierten Ansatz folgt.

Kapselung (encapsulation)

keine Software-Altlasten oder –überbleibsel (neues Gebiet)

Bedarf von Erweiterungsmöglichkeit (Spezialisierung kann neue Hardware, Medienformate, Standards, Kompressionstechniken usw. aufnehmen)

Bedarf an Cross-Plattform Entwicklung (abstrakte Interfaces reduzieren die Plattformabhängigkeit)

Was versteht man unter dem Begriff *Configuration* und wie wird das Konzept im vorgestellten Framework realisiert?

Komposition von Komponenten (MPEs), Streams und Medien zu einer Applikation

Configuration soll dynamisch ändernde Konfigurationen mit einer willkürlichen Anzahl von Komponenten unterstützen.

Weiters soll sie Verteilung, Authoring und Programmierung unterstützen.

Worin unterscheiden sich Ihrer Meinung nach die im vorgestellten Framework definierten *Components* vom allgemeinen Begriff *Komponenten* wie sonst im SE verwendet? Was haben die beiden Konzepte gemeinsam?

Components unterscheiden sich von Komponenten durch die Ports. Diese können als Schnittstelle gesehen werden, über denen die Streams gesendet bzw. empfangen werden können.

Wozu dienen die *Ports* von Komponenten? Unter welchen Umständen dürfen Ports zu verschiedenen Komponenten verbunden werden?

Components (MPE)

Streams werden über Ports gesendet und empfangen.

Verbundene Ports müssen plug-kompatibel sein

Ports dürfen zu verschiedenen Komponenten verbunden sein, wenn

- einer ein Output- und der andere ein Inputport ist.
- beide Ports plug-kompatibel sind.
- das Hinzufügen der Verbindung das fan-limit an keinem der Ports überschreitet.

Was versteht man unter *padding* und weshalb wird *padding* eingesetzt?

Padding ist Auffüllen von ungenutztem Platz mit leeren Daten zwecks Erfüllung von Formatvorgaben. z.B. Ethernet Pakete müssen ≥ 64 byte sein. wenn du weniger senden willst musst du sinnlose Bytes anhängen.

Was ist die Aufgabe der Klasse *Transform* im vorgestellten MM-Framework und was ist an den Objekten dieser Klasse bemerkenswert?

Implementiert function objects (Objekte, die sowohl eine Funktion als auch ihre Parameter kapselt).

Instanzvariablen von Transformklassen speichern Parameter, und eine „apply“ Methode löst die Funktion aus und gibt das Ergebnis zurück.

Alle Medienoperationen können von function objects repräsentiert werden.

Man kann neue Operationen hinzufügen, ohne existente Klassen und Applikationen zu modifizieren

Effiziente Lagerung und Bearbeitung der abgeleiteten Objekte.

Was versteht man unter der „dualen Natur“ zeitabhängiger Medien?

Zeitbasierte Medien können gesehen werden als:

Pools (für medienunabhängige Operationen, z.B. Kopieren)

Streams (für medienabhängige Operationen, z. B. Editieren)

Was versteht man unter *Composition* und wie wird das Konzept im vorgestellten Framework realisiert.

Komposition ist die Spezifikation von zeitlichen und/oder örtlichen Zusammenhängen zwischen einer Gruppe von Medienobjekten.

Das Resultat von Komposition wird Multimedia Objekt genannt, die ursprünglichen Elemente sind die „components“.

Was sind function objects? In welchen Klassen des Frameworks werden sie implementiert?

Die Transformklasse implementiert function objects – Objekte, die sowohl Funktionen als auch Parameter umschließen. Function Objects sollten mit Vorsicht verwendet werden, denn Leistungsverlust und möglicher Verlust des Type-Checkings können die Folge sein. Jede Art von Media-Operationen (Konvertierungen, zeitliche Transformationen etc.) können durch function objects repräsentiert werden.

Was versteht man unter abgeleiteten Objekten (derived objects), wozu dienen sie und wie werden sie im vorgestellten Framework realisiert?

Die Ableitung (D) eines Medienobjekts o_1 von einem Medienobjektset O ist die eine Abbildung der Form $D(O, P_D) \rightarrow o_1$, mit P_D als Set von für D spezifischen Parametern. o_1 wird dann abgeleitetes Objekt genannt. Abgeleitete Objekte verweisen auf die Medienobjekte und verwendeten Parameterwerte.

Erklären Sie kurz den Begriff Quality of Service.

Repräsentiert die Applikationsanforderungen für eine gegebene Ressource.

Unscharfe oder statistische Beschreibung der Parameter der Service-übergabe spiegelt Toleranz in menschlicher Wahrnehmung wider

ermöglicht ein Ressourcenmanagement und Trade-offs zwischen sich widersprechenden Zielen

in allen Levels soll es kompatibel, übertragbar und zugänglich für eine systemweite Optimierung sein

Typische QoS Parameter: min. und max. Auflösung, erlaubte Fehlerrate, akzeptierter Jitter und Delay-Schranken

Manche Applikationen benötigen strenge oder deterministische Garantien für das Service, andere brauchen nur statistische Garantien für eine bestimmte angebotene Performance.

Was ist ein Framework und welche Anforderungen werden an ein MM-Framework gestellt?

Frameworks

bieten eine allgemeine Architektur für die definierte Kooperation und Kommunikation von Objekten

bieten Templates für eine Gruppe von Objekten, welche gemeinsam die Verantwortung für einen bestimmten Problembereich managen

legen vordesignte Lösungen für eine Klasse von Problemen fest

Trennen die generischen und spezifischen Teile einer Lösung und strukturieren generische Parts zu kollaborierten Objekten

Wörtlich übersetzt bedeutet Framework (Programm-)Gerüst, Rahmen oder Skelett. Darin wird ausgedrückt, dass ein Framework in der Regel eine Anwendungsarchitektur vorgibt. Dabei findet eine Umkehrung der Kontrolle statt: Der Programmierer registriert konkrete Implementierungen, die dann durch das Framework gesteuert und benutzt werden, statt – wie bei einer Klassenbibliothek – lediglich Klassen und Funktionen zu benutzen. Wird das Registrieren der konkreten Klassen nicht fest im Programmcode verankert, sondern wird „von außen“ konfiguriert, so spricht man auch von Dependency Injection.

Ein Framework definiert insbesondere den Kontrollfluss der Anwendung und die Schnittstellen für die konkreten Klassen, die vom Programmierer erstellt und registriert werden müssen. Frameworks werden also im Allgemeinen mit dem Ziel einer Wiederverwendung „architektonischer Muster“ entwickelt und genutzt. Da solche Muster nicht ohne die Berücksichtigung einer konkreten Anwendungsdomäne entworfen werden können, sind Frameworks meist domänenspezifisch oder doch auf einen bestimmten Anwendungstyp beschränkt. Beispiele sind Frameworks für grafische Editoren, Buchhaltungssysteme oder elektronische Warenhäuser im World Wide Web.

Anforderungen an ein MM-Framework:

Wirtschaftlichkeit der Konzepte

offen – es sollte möglich sein neue Medientypen, neue Datenrepräsentationen und neue Hardwaremöglichkeiten zu implementieren.

queryable – es soll Interfaces für Abfrageumgebungen spezifizieren unter Berücksichtigung derer Möglichkeiten

Distribution – es soll helfen Applikationen zu partitionieren um Verteilung zu ermöglichen

Skalierbarkeit – es soll skalierbare Medienrepräsentation unterstützen

High-level Interfaces – für Synchronisation, Medienkomposition, Devicekontrolle, Datenbankintegration und simultane Medienprocessingaktivitäten

Was leistet die Programmierungsabstraktion Interpretation? Nennen Sie zwei Sachverhalte, die die Interpretation erschweren.

Aufbereitung der Daten, sodass man damit arbeiten kann.

Eine Interpretation eines Streams (BLOB) S ist eine Abbildung von S auf ein Set von Mediaelementen.

Erschwerende Faktoren:

Uneinheitlichkeit: verschiedengroße Elemente, verschiedene Kodierungsmethoden und Parameter
out-of-order Elemente

Interleaving und Padding

nicht destruktives Editieren und Skalierbarkeit

Erklären Sie kurz die Ziele von MPEG-7

Beschreibung von MM-Inhalt
flexibles Datenmanagement
Globalisierung und Interoperabilität von Datenressourcen.

MPEG-7 bestrebt folgende Punkte zu standardisieren:

ein Set von Description schemes und Deskriptoren um Daten zu beschreiben

eine Sprache um Description schemes zu spezifizieren (beispielsweise Description Definition Language – DDL)

a scheme for coding the description

Erklären Sie kurz die Konzepte *Deskriptor*, *Description Scheme* und *DDL* im Kontext von MPEG-7.

Ein Deskriptor ist eine Repräsentation eines Features. Er definiert die Syntax und Semantik der Featurerepräsentation. Ein Deskriptor muss die Semantic des Features, den assoziierten Datentyp, erlaubte Werte und eine Interpretation der Deskriptorwerte präzise definieren. Mehrere Deskriptoren können ein einzelnes Feature repräsentieren, ebenso wie multiple Deskriptoren möglich sind. Ein Deskriptor kann wie folgt aussehen: <name, typekind, spec> value </name>

Ein Description Scheme spezifiziert die Struktur und Semantik von Beziehungen zwischen seinen Komponenten, welche sowohl Deskriptoren als auch Description Schemes sein können. Der Unterschied zwischen Deskriptor und Description Scheme ist, dass ein Deskriptor sich mit der Repräsentation eines Features beschäftigt, das Description Schema jedoch mit der Struktur einer Featurebeschreibung.

Description Definition Language ermöglicht Erstellung neuer Description Schemas und Deskriptoren.

Außerdem können damit bestehende Description Schemas erweitert und modifiziert werden.

Anforderungen:

DDL unterstützt die Fähigkeit neue D-Schemata und Deskriptoren zu erstellen.

DDL erlaubt die Wiederverwendung, Erweiterung und Vererbung bestehender Deskriptoren und D-Schemas.

DDL bietet Mechanismen zur eindeutigen Identifizierung von D-Schemata und Deskriptoren für eindeutige Verweise.

DDL bietet ein Set primitiver Datentypen (Text, Integer, Real, Date,...) und muss einen Mechanismus anbieten um Deskriptoren Daten mehrerer Medientypen mit eigener Struktur (audio, video, av-presentations,...) zuzuordnen

Erklären Sie sehr kurz die Terminal Architektur von MPEG-7.

Das Terminal ist die Einheit, welche kodierte Repräsentationen von MM-Content nutzt. Dabei kann es sich um eine stand-alone-Applikation oder um einen Teil eines Applikationssystems handeln:

application

Transmission/Storage medium: verweist auf die niedrigeren Layer der Übermittlungsinfrastruktur (Netzwerk Layer und darunter genauso wie Speicherung); diese Layer übermitteln gemultiplexte Streams zum Delivery Layer.

Delivery Layer: umfasst Mechanismen für Synchronisation, Framing und Multiplexen von MPEG-7 Content.
Compression Layer: der Fluss von Access Units (entweder textuell oder binär kodiert) wird gepart und die Content description wird rekonstruiert.

**Erklären Sie kurz die Konzepte Rights Data Dictionary (RDD) und Rights Expression Language (REL).
Inwelchem Kontext treten sie auf und wie spielen sie zusammen?**

Kontext: MPEG-21, digital rights management

RDD: Verzeichnis mit Schlüsselbedingungen

The MPEG Rights Data Dictionary standard (ISO/IEC 21000-6) has been developed to meet the needs of rights owners and consumers by providing a consistent vocabulary for digital rights management.

ein Verzeichnis mit Schlüsselbestimmungen, die benötigt werden um Rechte von denen, die Dis steuern zu beschreiben, inklusive intellektueller Property Rechte.

Rechte sind eindeutig ausgedrückt durch die Verwendung einer standardisierten syntaktischen Konvention und kann auf alle Bereiche angewendet werden.

RDD spezifiziert wie weit Bestimmungen definiert werden unter Steuerung/Herrschaft einer Registrierungsautorität (registration authority).

RDD System unterstützt außerdem die Abbildung und Transformation von Metadaten einer Terminologie eines Namespaces in das eines anderen.

REL: maschinenlesbare Sprache, die Rechte und Permissions unter Verwendung von Bedingungen, wie sie im RDD definiert sind, deklariert

Beitrag flexible und interoperable Mechanismen zur Unterstützung von transparenten und erweiterten Verwendung von digitalen Ressourcen

schützt digitalen Content und anerkennt die Rechte, Konditionen und Kosten, die für digitalen Content spezifiziert sind

beabsichtigt weiters das Unterstützen von Zugangsspezifikation und Verwendungskontrollen für digitalen Content in Fällen, wo finanzieller Austausch/Wechsel/Markt nicht Teil der Nutzungsbestimmungen ist und das Unterstützen des Austauschs sensiblen und privaten digitalen Contents.

Erklären Sie kurz die Ziele von MPEG-21.

Ermöglichung von transparenter und erweiterter Verwendung von MM-Ressourcen über eine große Auswahl von Netzwerken und Devices

Kontrolle/Betrachtung von Elementen, die entweder existieren oder in Entwicklung sind um eine Infrastruktur für die Übermittlung und den Konsum von MM-Content aufzubauen

Ziele:

Verstehen, wie Elemente zusammenpassen

Bestimmen neuer Standards, die benötigt werden, wenn Lücken in der Infrastruktur bestehen.

Ausführen der Integration von verschiedenen Standards