Multimedia (188.931)

Ausarbeitung – Prüfungsordner – PO

Zusammengefügt aus Martin Tintels MM 1 und MM 2 Ausarbeitungen.

Jänner 2013, Version 0.91

Datentypen

Erklären Sie kurz die Begriffe Dynamic Range und Dithering. [2] bzw. Erklären Sie kurz Cluster Dithering. Weshalb wird Clustering Dithering eingesetzt? [2.5]

Dynamic Range

Bei achromatischem Licht (nur Grauwert) stellt die Dynamic Range das Verhältnis zwischen der minimalen und maximalen Intensität (Helligkeit) dar. Wie viele Intensitäten in einem Bild zur Darstellung benötigt werden, hängt vom verwendeten Medium ab. 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 (Punktmustern). Dies funktioniert, da das menschliche Auge so feine Punktmuster vermischt (schwarz/weiß → grau) zu einem Grauton und wird vom menschlichen Auge als Schattierungen wahrgenommen.

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). So werden z.B. bei Druckern auf die weiße Fläche des Papiers mehr oder weniger große Punkte aufgetragen. Man versucht dabei die Punkte so zu setzten, dass sie keine regelmäßigen Muster erzeugen. Ist dies nicht möglich, so nimmt man diagonale Linien.

Cluster Dithering

Ditherintensität zentriert auf die Matrizenmitte. Line patterns 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.

Was versteht man unter Halftone Approximation, wofür wird sie eingesetzt und worauf ist dabei zu achten? [2]

Das Dithering ist eine Technik in der Computergrafik, um bei Bildern mit geringer Farbtiefe die Illusion einer größeren Farbtiefe zu erzeugen. Approximierung von Intensitäten unter der Verwendung von Dot-Patterns.

Es werden die fehlenden Farben durch eine bestimmte Anordnung aus verfügbaren Farben/ Pixeln nachgebildet und dadurch harte Übergänge zwischen den Farben/ Pixeln 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)

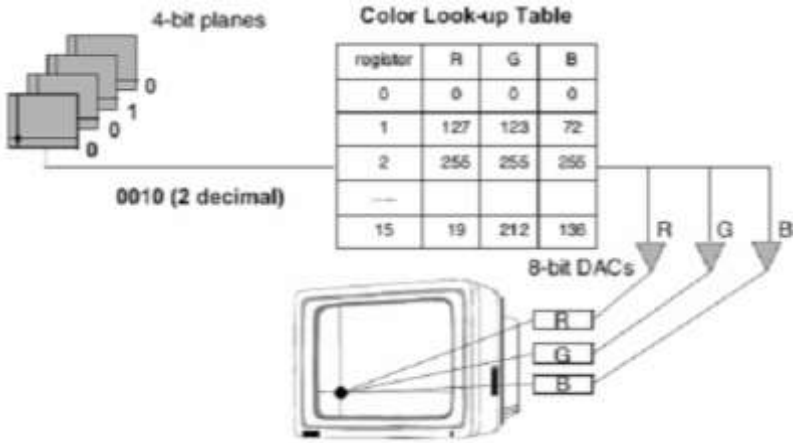
Es gibt viele verschiedene Algorithmen, diese arbeiten oft auch mit einer Fehlerstreuung (Übertragung von Überläufen oder Rundungsfehlern auf benachbarte Bildpunkte für eine feinere Darstellung).

Was sind indizierte Farben (indexed colors) und weshalb werden sie verwendet? [2]

Es wird der Farbwert eines Pixels (max. 256 Farben = 2^8 Bit = 1 Byte) durch einen Index in Color Maps oder aus einer CLUT (Color Lookup Table) ausgelesen.

Diese Maps oder Tables sind entweder vordefiniert oder Teil des Bildes. Jeder Pixel verweist auf einen Eintrag in diesem Table der dem jeweiligen Farbwert entspricht.

Es dient der platzsparenden Speicherung von Bildern, was jedoch nur bis zu einer bestimmten Bildgröße der Fall ist. Ab einer gewissen Grenze, werden Farbinformation besser direkt gespeichert. (Index der CLUT wird zu groß)



Erklären Sie kurz das Prinzip des Zeilensamplings (Line Sampling) für digitales Video. Was bedeuten die n:m:l Angaben im CCIR 601 Videoformat? [3] bzw. Was bedeuten die Komponenten n:m:l in CCIR 601 Video? [1.5]

Diese Angabe bedeutet, dass die Komponente Y (Helligkeit) n-Mal und die Farbkomponenten U und V m- und l-Mal pro Taktrate gesampled werden. Die Möglichen Taktmultiplikatoren sind: 1,2,3 oder 4.

Sampling

Das Konvertieren von einem analogen zu einem digitalen Video nennt man sampling. Dabei wird das analoge Signal zu bestimmten Zeitpunkten abgetastet und für die jeweiligen Werte eine digitale Repräsentation gesucht. Die einzelnen Frames werden dann durch ein Pixelarray dargestellt. Hier

sollte man nach dem Abtasttheorem von Nequist/ Shannon anwenden (Abtastrate = 2* höchste Frequenz im Signal), um einen Informationsverlust zu vermeiden.

CCIR 601 Standard (line sampling)

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

- m steht hierbei für die Basisrate, mit der Y abgetastet wird.
- n und l bestimmen, mit welcher Rate die beiden Farbkanäle (CC) abgetastet werden.
- Die Basisrate für das Sampling beträgt standardmäßig 3.375MHz.
- m,n und l enthalten nur noch den Multiplikator („Taktmultiplikatoren“) dieser Rate. Dieser kann 1,2,3 oder 4 sein.

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).

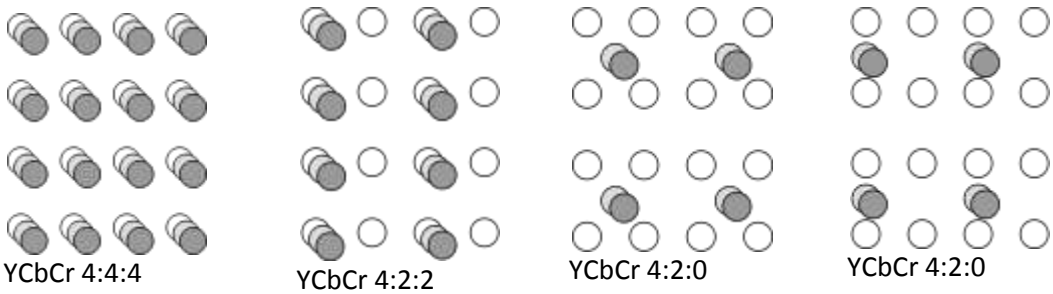
- YCbCr 4:4:4 Chrominanz-Auflösung identisch zur Luma-Auflösung
- YCbCr 4:1:1 Chrominanz-Auflösung horizontal geviertelt und vertikal unverändert

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.



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

Beispiel

4:4:4 …. Pro Pixel werden alle Werte abgetastet. Die ergibt die höchstmögliche Qualität (Studio).
4:2:2 …. Nur von jedem 2. Pixel pro Zeile werden Farbwerte abgetastet (Broadcast)
4:1:1 …. Nur von jedem 4. Pixel pro Zeile werden Farbwerte abgetastet (VHS)

Erklären Sie kurz die Begriffe skew, jitter und delay. Nennen Sie 2 Hardwarekomponenten für die Behebung dieser Probleme. [3]

Skew

Mit Skew wird die Zeitdifferenz eines Signals zu seinem korrekten Abspielzeitpunkt (Videosignal) angegeben. Er entsteht beispielsweise aufgrund von Übertragungsdivergenzen (Wenn z.B. das Lesen der Scanline für Video länger dauert, als das für Audio, diese aber synchron abgespielt werden müssen. (Zeitversetztes Lesen, gleichzeitiges Abspielen)). Des Weiteren können solche Divergenzen auch bei der Übertragung über verschiedene Medien entstehen. (Asynchronität zweier Streams).

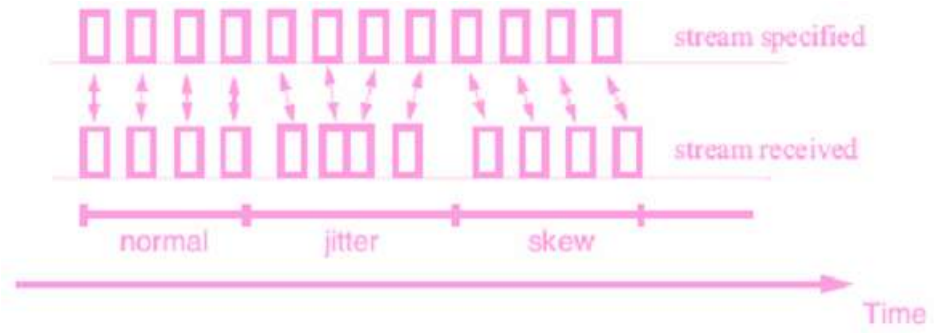
Jitter

Als Jitter bezeichnet man ein Taktzittern bei der Übertragung von Digitalsignalen, eine leichte Genauigkeitsschwankung im Übertragungstakt (Clock).

Entsteht bei einer unterschiedlichen Verzögerung der Signale eines Datenstroms. Abrupte Störung des Signals in Signalkomponenten z.B. hervorgerufen durch Fehlerwahrscheinlichkeit beweglicher Teile. Lösung: Ein Sync Generator kann helfen, da er einen stabilen Zeittakt vorgibt.

Allgemeiner ist Jitter in der Übertragungstechnik ein abrupter und unerwünschter Wechsel der Signalcharakteristik. Dies kann sowohl Amplitude als auch Frequenz und Phasenlage betreffen.

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.



Delay

Verzögerung des gesamten Streams (im Gegensatz zu Skew, wo die Verzögerung eines Streams gegenüber einem anderen gemessen wird). Audio und Video-Signal sind übertragungsbedingt verzögert. Müssen dadurch wieder synchronisiert werden.

Lösung: Frame Delay-Buffer hilft mehre externe Sources zu synchronisieren bzw. mittels Sync-Generator Problem zu beheben.

Erklären Sie Abtastfrequenz (sampling frequency), sample size und Quantisierung. Wie lautet das Abtasttheorem? [5]

Sample Frequency / Abtastfrequenz

Gibt an, wie oft (mit welcher Frequenz pro Zeitintervall) eine Abtastung eines analogen Signals vorgenommen wird. Je höher der Wert, desto besser ist die Qualität des resultierenden digitalen Signals, da ein geringerer Informationsverlust besteht. Minimale Abtastrate zur verlustlosen Rekonstruktion des Originalsignals: Nyquistfrequenz → min. doppelte im Signal höchste vorkommende Frequenz (2x fmax)

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.

Sample Size

Die Samplesize bestimmt die Aufteilung der Amplitudenskala bei der Digitalwandlung eines analogen Signals und wird in Bits angegeben. Man kann es also als Speicherplatz, der pro abgetasteten Wert verwendet wird verstehen. Daher gibt an, mit welcher Genauigkeit der Wert gespeichert wird (8bit, 16bit,…).

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.

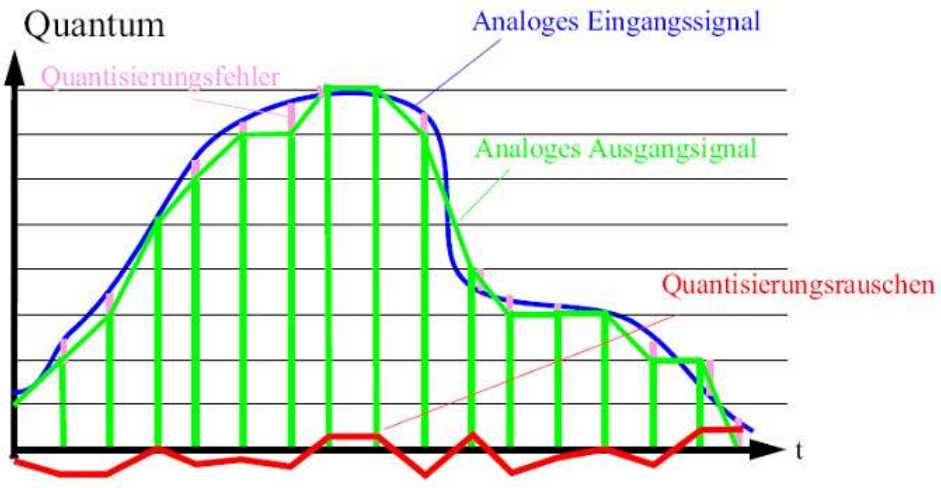
Quantisierung

Näherung des analogen Signals an diskrete Funktionen und Werte. Hierbei werden den abgetasteten Spannungswerten diskrete Zahlenwerte zugeordnet. Der Gesamtspannungsbereich wird in so genannte Quantisierungsintervalle Q unterteilt und der abgetastete Wert dem nächstgelegenen diskreten Zahlenwert zugeordnet. Hierbei entstehen Rundungsfehler, die sich im so genannten Quantisierungsrauschen bemerkbar machen. Sind die Quantisierungsintervalle von konstanter Größe, spricht man von linearer Quantisierung (übliche Methode in der Audiotechnik).

Um eine möglichst gute Qualität zu erreichen soll Q möglichst klein sein, es ergibt sich aus der Länge des Datenwortes. 8bit (256 Intervalle): Q=Vpp/(28) = 4*10-3xVpp

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)

Bei der nichtlinearen Quantisierung findet eine Unterteilung in unterschiedlich große Intervalle statt. Hierbei gilt, je kleiner die Werte, desto kleiner die Intervalle und umgekehrt. Diese Quantisierung ist somit aufwändiger zu berechnen, erreicht aber bei geringeren Abtastfrequenzen die bessere Qualität und wird daher in Systemen mit geringer Übertragungsbandbreite eingesetzt.



Abtasttheorem nach Nyquist-Shannon

Besagt, dass Signale über der Nyquist-Frequenz (= Hälfte der verwendeten Sampling-Frequenz) nicht mehr fehlerfrei (ohne Informationsverlust) dargestellt werden können. Daher soll man die Sampling-Frequenz auf das doppelte der max. Frequenz des Eingangssignals festlegen, um Artefakte im gesampleten Signal zu vermeiden (Abtastrate = 2* höchste Frequenz im Signal).

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? [3]

Beide dienen der Video Repräsentation. Werden für die Übertragung des Videosignals genutzt.

Component Video

Digital Component Video (D1/D5, SMPTE RP125): Trennung der Signale in Helligkeit (Y) und Farbe (Cr/Cb), 27MB/s Datenrate, subsampled Farbsignale 4:2:2.

Composite Video:

Digital Composite Video (D2/D3, SMPTE 244M): 14.3MB/s Datenrate, subsampled Farbsignale 4:2:2.

Component Video
+ mehrere Signale
- benötigt mehr Bandbreite (27 MB/s)
- Synchronisation aufwändiger
- Senden und Verkabelung komplizierter
+ Jede Primärfarbe und Helligkeit eigenes Signal
+ Bessere Qualität
+ Wird im Profi Bereich z.B. Editing, Encoding,.. verwendet
+ Jedes Signal enthält RGB + luma/ chroma Info

Composite Video
- in einem Signal gemischt
+ einfache Handhabung
+ einfacher zu übertragen (weil BB 14,3 MB/s)
+ einfachere Verkabelung
- Interferenzen können auftreten
- schlechtere Qualität
~ wird in Haushalten verwendet (VHS)
+ in einziger Trägerwelle übertragen

Erklären Sie kurz den Begriff Interlacing bzw. Interleaving und führen Sie aus, was er im Kontext der Datentypen Audio, Images und Video bedeutet [3] bzw. Erklären Sie kurz den Begriff Interleaving im Kontext des Datentyps Audio. Welche Vor- bzw. Nachteile haben interleaved Audioformate im Vergleich zu non-interleaved Formaten? [2]

Interlacing ist ein Zeilensprungverfahren und der „Content" setzt sich in Durchgängen zusammen.

Interlacing Video

Anstatt ein Frame progressiv aufzubauen werden zwei Halbbilder erzeugt, die jeweils um eine Zeile verschoben sind und nacheinander angezeigt werden. Das geht
deshalb, weil der Phosphor im TV-Gerät oder Monitor nachleuchtet und dieser Trick dem menschliche Auge daher nicht auffällt. Zugleich spart man damit gewaltig Bandbreite, hat aber den Nachteil, dass das De-Interlacing sehr aufwändig ist.

Interlacing Image

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.

Das Bild wird nicht Zeile für Zeile aufgebaut, sondern in mehreren Durchläufen. Beim ersten Durchlauf wird 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.

Interleaving

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

Interlacing in dem Sinn hat an sich nichts mit der RGB-Codierung zu tun. Das wäre dann wieder Interleaving. Man könnte zwar die drei Farbkanäle in einem File separat speichern, das hätte jedoch Nachteile, da man das File immer komplett geladen haben muss, bevor man es richtig anzeigen kann. Vom Prinzip her würde das dann (vereinfacht) so aussehen:
RRRRRRR...GGGGGG...BBBBBBB...

Beim Interleaving werden linear durchgezählte Objekte (z.B. Video und Audio oder RGB) in eine verschachtelte Form gebracht, um Burstfehler zu vermeiden, indem diese durch das Interleaving zu Einzelbitfehler, die leicht zu korrigieren sind, ausgemerzt werden.

Im Interleaved-Verfahren werden für jeden einzelnen Bildpunkt oder zumindest Bildbereich die einzelnen Kanäle gleich hintereinander gespeichert, also RGBRGBRGB - damit ist es möglich die Bildpunkte sofort auszugeben.

Ohne Interleaving
Fehlerfreie Übertragung: aaaabbbbccccddddeeeeffffgggg
Übertragung mit einem Burstfehler: aaaabbbbccc____deeeeffffgggg

Nun die gleichen Daten mit Interleaving:
Fehlerfreie Übertragung: abcdefgabcdefgabcdefgabcdefg
Übertragung mit einem Burstfehler: abcdefgabcd____bcdefgabcdefg

De- Interleaved Übertragung mit einem Burstfehler
aa_abbbbccccdddde_eef_ffg_gg

Jetzt fehlen zwar von a, e, f und g je ein Bit, aber das kann korrigiert werden, weil jeweils nur ein Bit und nicht die ganzen Sequenzen cccc und dddd verloren sind.

Vorteile: Gegen Burstfehler abgesichert**.** Brauchen für Sicherung keine zusätzlichen Daten.
Nachteile: Latenzzeit erhöht sich.

Interleaving Video

Anstatt die drei Farbkanäle in einem File separat zu speichern (RRRRRRR….GGGGGGGGG…BBBBBBBBBB…), 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.

Interleaving 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 der folgenden Frames enthält. Der Abstand zwischen den Audioteilen 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)

Interleaving Images

Pixel Interleaving, Raster Interleaving und Frame Interleaving. Die unterschiedlichen Interleaving Schemen mit 4 Reihen und 5 Spalten würde so aussehen:

Pixel interleaved looks like:
RGBRGBRGBRGBRGB row 1
RGBRGBRGBRGBRGB row 2
RGBRGBRGBRGBRGB row 3
RGBRGBRGBRGBRGB row 4

Raster interleaved looks like:
RRRRRGGGGGBBBBB row 1
RRRRRGGGGGBBBBB row 2
RRRRRGGGGGBBBBB row 3
RRRRRGGGGGBBBBB row 4

Frame interleaved looks like:
RRRRRRRRRRRRRRRRRRR all the red
GGGGGGGGGGGGGGGGGGG all the green
BBBBBBBBBBBBBBBBBBB all the blue

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

- Analoges Video ist kostengünstiger in Bezug auf Speicherplatz und Übertragung (mehrere GB große Videos).
- Die Digitalisierung von Videos benötigt immer noch viel Zeit.
- Analoges Video ist einfacher zu senden.

Erklären Sie kurz die Begriffe chroma keying und navigational video. [2] bzw. Erklären Sie kurz die Begriffe Chroma Keying und Luminance keying. [2]

Keying beschreibt das Freistellen von Bildelementen vom Hintergrund

Luminanz Key

Der Luminanz Key hat seinen Namen aus dem Videosignalbereich in welchem Farbe und Helligkeit getrennt übertragen werden. Man separiert 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/ Bild sichtbar. Das Verfahren ist eher für Effekte interessant. Kann weniger kontrolliert werden, wird aber oft bei Graphikelementen verwendet.

Der Luma Keyer errechnet aus einem farbigen RGB-Bild ein monochromes Graustufenbild. An dieser Stelle kann das Mischverhältnis der drei Farbkanäle justiert werden um Vordergrund und Hintergrund kontraststark zu trennen. Die Stanzmaske wird folglich aufgrund starker Helligkeitsdifferenzen innerhalb des Graustufenbildes erzielt. Um ein gutes Ergebnis zu erreichen solle der zu keyende Gegenstand sehr hell bis weiß gegenüber einem sehr dunklen oder sogar schwarzen Hintergrund sein; Oder umgekehrt.

Chroma Keying:

Auch der Chroma Key hat seinen Namen vom getrennten Videosignal. Bei Chroma Keying nimmt man als Hintergrund eine bestimmte Farbe (man verwendet gerne blau – weil es eine Komplementärfarbe zu den Hautfarben darstellt - könnte im Prinzip aber jede Farbe sein). Damit kann man dann das oder die Objekte leicht vom Hintergrund separieren und über jeden anderen Hintergrund/ Video darüberlegen. Es wird somit eine Maske (Alpha Kanal) gemacht die besagt, welche Bereiche transparent sind.

Ein Mixer ersetzt eine Farbe in einem Signal, durch Daten eines anderen Signals (z.B.: Blue-Box-Effekt). Im Videobild werden Teile einer ausgewählten Farbe von einer anderen Videosequenz ersetzt. Auch bekannt als Bluebox-Effekt bzw. Bluescreen-Technik für den Austausch von Bild-Hintergründen.

Das gleiche kann man auch genauso im Luminance-Bereich anwenden, da separiert man dann aufgrund von Helligkeitsverwerten. Durch die Beschränkung auf “eine” Farbe ist diese Methode selektiver und flexibler als der Luma Key. Intern wechselt der Keyer vom RGB Farbraum in HSV (Hue, Saturation, Value), weil er auf Farbnuancen, Sättigungsgrade und die Helligkeit angewiesen ist, was nicht zu den Attributen von RGB zählt. Daher gekeyed wird in HSV, wobei die Justierung ebenso in RGB, YUV, RGBCMYL oder den separaten Kanälen (rot, grün, blau) stattfinden kann.

Navigational Video (=^ navigable Video)

Erstellt eine „virtuelle“ Kamerafahrt aus Einzelbildern. Hierbei wird zuerst ein Objekt aus möglichst vielen Richtungen in geringem Abstand zueinander aufgenommen (mit Winkelabständen von beispielsweise 5°). Nun kann man diese Einzelbilder beliebig aneinander reihen und so den Eindruck einer bewegten Kamera erzeugen und einen realistischen Rundumblick bieten (wenn man sie nicht linear abspielt!). Man erhält den Eindruck einer Kamerafahrt und 360°-Ansicht (Panorama).

Erklären Sie kurz den Begriff Multimediasystem im engeren und weiteren Sinne. [2.5]

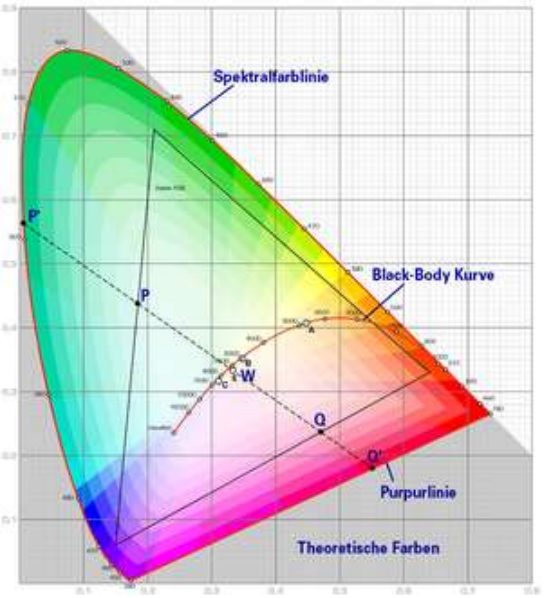
Ein Multimediasystem ist charakterisiert durch computergesteuerte, integrierte Produktion, Manipulation, Präsentation, Speicherung und Kommunikation unabhängiger Information, welche zumindest durch ein kontinuierliches (zeitabhängiges) und ein diskretes (zeitunabhängiges) Medium kodiert ist.

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

Erklären Sie kurz den CIE Farbraum. [2]

Der CIE Farbraum wird verwendet um die anderen Farbräume (CMYK, RGB, usw.) zu kalibrieren, da alle in diesem Modell enthalten sind.

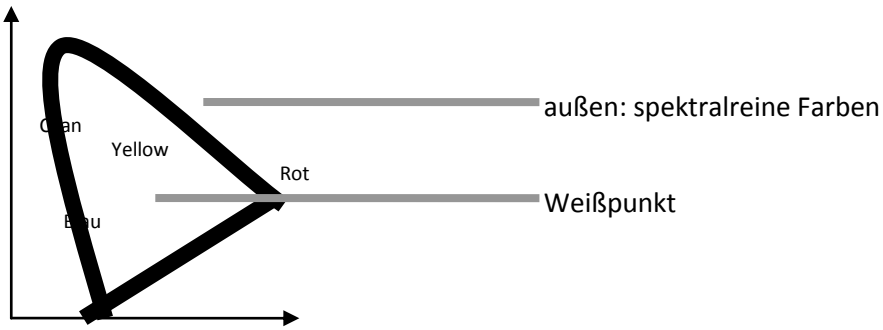
Darstellung von Farben beruhend auf dem menschlichen Wahrnehmungsapparat. Additive Farbmischung. Grundlage ist ein sogenannter Normalbeobachter mit einem Sichtfeld von 2° (CIE1931) bzw. 10°(CIE1964). Normfarbwerte X (virtuelles rot), Y (virtuelles grün), Z (virtuelles blau)



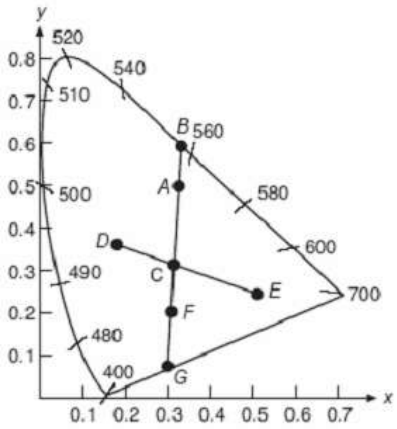
Darstellung: Querschnitt durch den Raum; alle Farben befinden sich innerhalb der Fläche.
Wellenlänge: gebogene, äußere Linie.
Purpurgerade: gerade Verbindungslinie zwischen Rot und Blau.

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. Deshalb kann CIE benutzt werden um Farben zu identifizieren und analysieren, oder um die Leistung eines CRT zu messen (wie viel des CIE Farbraums kann er darstellen)



Auf der äußeren Kurve liegen die reinen Spektralfarben, ca. in der Mitte liegt Weiß(C). Wenn eine Farbe A im Raum zwischen Weiß und der Spektralkurve liegt so bekommt man die dominante Wellenlänge, indem man die Gerade, die durch C und A geht bis zur Kurve verlängert. Der Schnittpunkt zeigt dann die dominante Farbe an. Würde der Schnittpunkt auf der Geraden liegen die den Raum abschließt, so wird die dominante Wellenlänge als das Komplement von A genommen.



Dominate Wellenlänge: zwischen Farbe und Weißpunkt Gerade ziehen und bis Spektralkurve verlängern.
Komplementärfarbe: liegt gespiegelt zum Weißpunkt auf selben Gerade.

Verwendung:

- zur Kalibrierung von anderen Farbmodellen (RGB, CMYK,…).
- zur Identifikation und Analyse von Farben.
- zur Messung für CRT Leisten („Performancemessung“), sprich wie viel vom CIE Farbraum. wird durch CRT Abgedeckt.

Diskutieren Sie kurz 3 Kriterien ihrer Wahl zum Vergleich von Videokomprimierungsverfahren. [3]

Verlustfreie oder verlustreiche Komprimierung

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.

Echtzeit Komprimierung

Zeitaufwand für En- und Decodierung. 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)

Interframe- oder Intraframe-Kompression

Interframe-Kompression erstellt alle Bilder (ausgenommen Keyframes) aufbauend auf das Vorgängerbild. Intraframe komprimiert nur Einzelbilder.

Erklären Sie kurz Schalldruckpegel und Lautstärkepegel. Wozu dienen diese Werte? [3]

Schalldruckpegel

L = 20*log(p/p0) [dB] mit Bezugsschalldruck p0 (1kHz, gerade noch hörbar20µPa). Die Einheit des Schalldrucks ist Pa und die des Schalldruckpegels dB. Er dient zur wahrnehmungsgemäßen Beschreibung von Lautstärken und wird in Dezibel gemessen.

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.

Somit sind 10 Geigen, die gleichzeitig spielen, doppelt so laut wie nur eine Geige. dB Angaben werden benötigt um einen gigantischen Hörbereich in einen sinnvollen kleineren Wertebereich einzuschränken.

Lautstärkepegel

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 Lautstärkeeindruck zu erzielen. Welcher Schalldruckpegel für einen Einzelton bei welcher Frequenz erforderlich ist, um jeweils den gleichen Lautstärkeeindruck zu erzielen, ist in den "Kurven gleicher Lautstärkepegel" (Isophone) beschrieben. Z. B.: Ton mit 100Hz und Schalldruckpegel L=60dB → 50 Phon.

Die wahrgenommene Lautstärke ist also nicht nur abhängig vom Schalldruckpegel, sondern auch von der Frequenz. Um so alle Töne, die für uns gleich laut klingen, mit einem gleichen Pegelwert beschreiben zu können, wurde der Lautstärkepegel LN mit der Einheit Phon definiert.

Beschreiben Sie kurz das Prinzip von MIDI und erklären Sie die 2 MIDI-Konzepte Clock und Palette [2.5] bzw. Beschreiben Sie kurz das Prinzip von MIDI und erklären Sie die 2 MIDI-Konzepte Channel und Voice [2.5]

Das Musical Instrument Digital Interface 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, die am MIDI-Gerät eingestellt werden kann.
- Maß: Teile pro Viertelnote
- versieht die Messages mit Timestamps
- Tempo: bpm

Palette

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

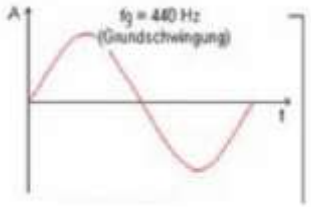
Voice

- ist der Teil des Synthesizers, welcher den Sound produziert.

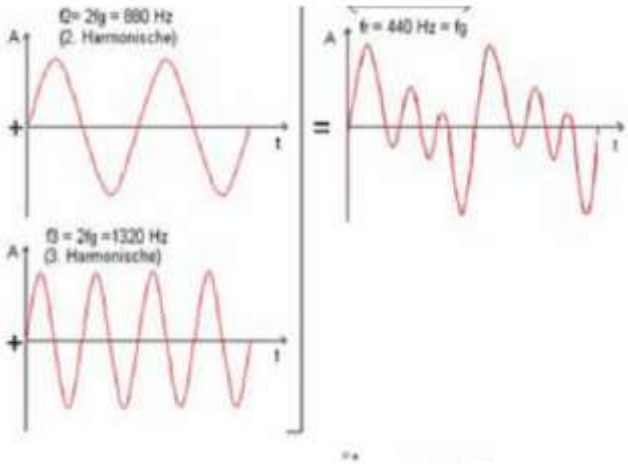
- Synthesizer können mehrere (16, 20, 24, 32, 64,…) voices haben, wobei jede unabhängig und simultan arbeitet um Sound mit verschiedenen Tonlagen und Klangfarben zu produzieren.

Erklären Sie kurz die Begriffe Ton, Klang und Geräusch [1.5]

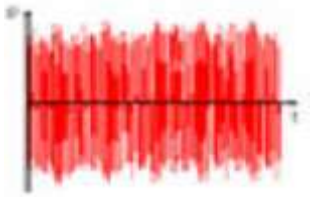
Als **Ton** wird eine einzelne Sinusschwingung bezeichnet. P(t) ist eine Sinusschwingung (mit einer bestimmten Frequenz – 1/t)



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



Bei einem **Geräusch** ist kein Grundton mehr erkennbar und nur noch eine Ansammlung von verschiedenen Frequenzen vorliegt. Aperiodische Schwingung (unregelmäßige Schwingung).

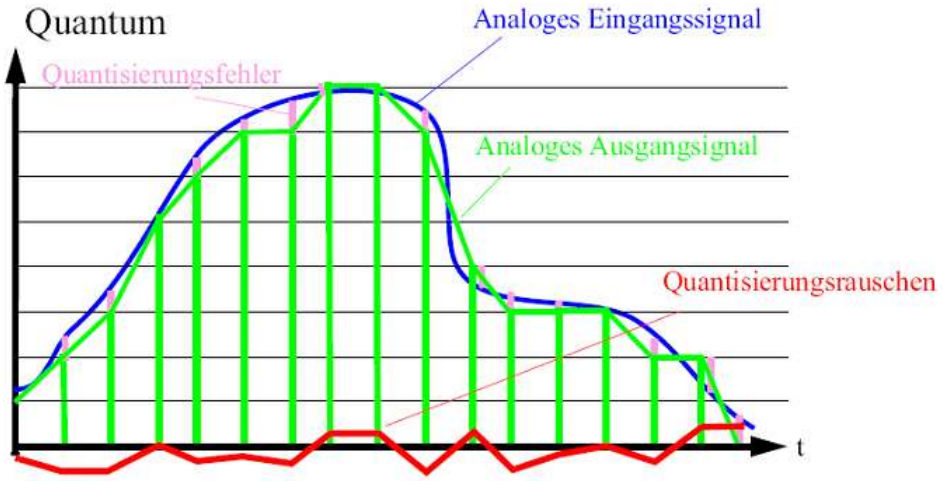


Was ist und wie entsteht Quantisierungsrauschen? Auf welche Weise kann es reduziert werden? [2]

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 (höhere Sampleauflösung =Anzahl der Quantisierungspunkte) reduziert werden und andererseits bei der linearen Quantisierung, sofern der Signal-Rausch-Abstand groß genug ist, ausmaskiert werden.



Welchen Zweck haben Distribution Amplifiers und Timebase Correctors? [1.5]

Distribution Amplifier

Ein Distribution Amplifier ist ein NF-Verstärker mit mehreren Ausgängen. Die Ausgänge haben alle das gleiche Ausgangssignal und können an unterschiedliches Equipment angeschlossen werden. Funktional handelt es sich um eine Signalverteilung mit Verstärker (gibt ein Eingangssignal verstärkt an mehreren Ausgängen aus). (splits 1 input signal to 2 or more outputs)

Timebase Correctors

Baut Signal wieder auf („Rekonstruiert"), um die Timing-Störungen zu entfernen, die durch VTR/VCR (video cassette recorder) entstehen, zu eliminieren.

Erklären Sie kurz 2 Arten von Video Time Codes Ihrer Wahl. [1.5]

Mit auf dem Band geschriebene Time Codes:

- LTC: Longitudinal Time Code: Zeitcode meist im Audiotrack aufgezeichnet; entlang des Bandes
- VITC: Vertical interval TC; in den vertikalen Austastlücken (vertical blanking) aufgezeichnet; im Nachhinein nicht mehr korrigierbar; kann nur gleichzeitig mit dem Video aufgezeichnet werden

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

- Wird extra zu jedem Bild gespeichert und ist eindeutig.
- non-drop frame time code – FF in [0, 29]
- drop frame time code FF in [2, 29], außer bei jeden 10. Minute [10, 29]

Was ist ein Frequenzspektrum? Wie unterscheiden sich die Frequenzspektren von Klängen und Geräuschen? [3]

Frequenzspektrum

Jede periodische Schwingung beliebiger Wellenform (= Klang) kann als eine Überlagerung von Grundschwingung 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. Diskret/kontinuierlich.

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 ist eine Color Look-up Table? Weshalb wird sie eingesetzt? [2]

Die LookUp Table (LUT) ist eine hardwaremäßig geschaltete (zwischen Datenspeicher und Darstellungsgerät) Liste von Wertepaaren, die jedem Grauwert r einen neuen Grauwert s = T(r) zuordnet. Sie wird daher bei Punktoperatoren (Grauwertabbildungen) genutzt.

Größe: Anzahl der Farbwerte

Kleiner Bereich genützt: ineffizienter als direkte Indizierung (schlechte Speicherplatzausnutzung) (Speicherung des Farbwertes direkt ohne Tabelle)

Quantisierung – fein –grob

Fein: besser, kleinerer Quantisierungsfehler, genauer, besonders wichtig, wenn sich der Funktionswert nur langsam ändert. In diesen Fällen werden räumlich große Teile auf denselben digitalen (Grau)wert quantisiert und die plötzlichen Änderungen im digitalen Wert (= Sprung von einer Quantisierungsstufe zur nächsten), die an zufälligen Stellen auftreten, neigen dazu, falsche Konturen erscheinen zu lassen und somit Scheinobjekte zu definieren.
Nachteil: höherer Speicherbedarf, höherer Rechenaufwand

Rastern ->fein – grob

Abtasten/Quantisieren

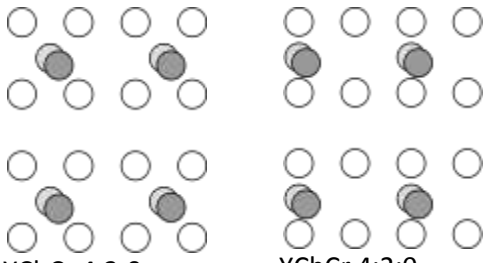
Abtastfrequenz

Frequenz, wie oft ein Signal abgetastet wird. Alle 1/f Zeitpunkte wird der Signalwert abgetastet.

Was ist die Besonderheit des Samplings für digitale 4:2:0 Videoformate? [2]

Das Luminanzsignal wird mit 13,5 MHz (=^ 3.375 MHz * 4) abgetastet. Bei der Abtastung der Farbdifferenzsignale macht man sich zu nutze, dass das menschliche Auge in vertikaler Richtung eine erheblich geringeres Auflösungsvermögen besitzt als in der Horizontalen. Abwechselnd wird eine Bildzeile im Verhältnis 4:2:2 und die darauffolgende Zeile im Verhältnis 4:0:0 digitalisiert. Zeilen, für die keine Farbinformation aufgezeichnet wurde, übernehmen das Farbsignal der darüber liegenden Zeile.

Oder einfacher ausgedrückt: Hierbei wird das Farbsignal in jeder zweiten Zeile überhaupt nicht abgetastet, und in jeder ersten nur halb so oft wie das Luminanzsignal, was die Datenrate im Vergleich zur 4:2:2-Abtastung noch ein Mal deutlich reduziert.


YCbCr 4:2:0 YCbCr 4:2:0

Die 4:2:0 Abtastung wird im PAL-DV Standard verwendet. Daher benötigt Digital Video trotz einer 5:1 Kompression weniger Speicherplatz (ca.3,5 MB/sec) als das ähnliche MJPEG-Verfahren. Dieses

benötigt bei gleicher Kompression mit 4:2:2 Abtastung ca. 4,4 MB/sec bei PAL-Vollbild (786 x 576 Pixel).

Plattformen

In welchem Kontext spielen Real-Time-Video und Production-Level-Video eine Rolle und was unterscheidet die beiden Formate? [3]

Real-Time-Video und Production-Level-Video sind 2 Varianten der Video-Kompression im Rahmen von DVI (Digital Video Interactive).

Production Level (PLV) bzw. Presentation Level Video

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. Braucht lange zum Codieren, decodieren dann in Echtzeit.
- 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!

Vorteile

- Hohe Qualität
- Decodieren ist in Echtzeit möglich.
- Kompression viel besser (wird daher eingesetzt, wenn ein Video nur ein Mal komprimiert und oft abgespielt wird (z.B. wenn es als Film verkauft wird)).

Nachteile

- Lange Übertragungszeiten
- Hoher Rechenaufwand, daher nicht in Echtzeit komprimiert

Real-Time-Video (RTV)

Kompressionsformat, mit dem Videosignale in Echtzeit digitalisiert, komprimiert, codiert, decodiert 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 (z.B. bei Videokonferenzen) in Echtzeit komprimiert und mit "Video for Windows" dargestellt werden.

- 128x120, mit Zeilen/Spalten-Verdopplung und Interpolation 256x240
- RTV- Intraframe-Kodierung: Differenz zwischen aktuellem und darüberliegendem Pixel => viele Nullen.
- RTV-Interframe-Kodierung: Differenz von Pixeln gleicher Position in aufeinanderfolgenden Bildern => viele Nullen.
- anschließend RLE+ Huffman
- einfaches, aber schnelles Verfahren
- symmetrisch
- Kompression in Echtzeit

Vorteile

- Multimediaapplikationen zunächst in RTV (= ELV = Edit Level Video),nachfolgend PLV

- codieren und decodieren in Echtzeit. Wichtig bei Videokonferenzen.

Nachteile

- Schlechtere Qualität
- Erzielt nicht so eine gute Kompression (da in Echtzeit codiert und decodiert werden muss)

Was sind die technologischen Unterschiede zwischen Audio-CD und CD-ROM? [2] bzw. Erklären Sie kurz die Unterschiede zwischen CD-Audio und CD-ROM. [1.5]

CD-ROM

- zusätzlicher Layer („Schicht“) für Fehlerdetektion und Fehlerkorrektur
- Unterteilung des Speicherplatzes in aufeinander folgende Datenblöcke (98 frames, 2352 bytes)
- 2 Blockformate
    - Mode1 – 2048 bytes Anwendungsdaten
    - Mode2 – 2333 bytes Anwendungsdaten

CD-Audio

- verwendet nur CLV (constant linear velocity)
- frame mit 588 channel bits, 200 bits nach dem Dekodieren (je AudioCD 20MB nicht nutzbar)
- Enthält nur ausschließlich Audio Daten und keine Multimedialen oder textuellen Zusatzinformationen.
- Audio Daten werden unkomprimiert gespeichert.

Erklären Sie kurz die QuickTime Medienorganisation auf dem konzeptionellen Level. [4] bzw. Erklären Sie kurz die konzeptionelle Medienorganisation von QuickTime. [3]

Die Speicherung erfolgt sowohl in sequenziellen als auch hierarchisch organisierten Dateneinheiten, sogenannten Atomen. Diese Atome bestehen aus: size, „four char code“ (bspw. Moov, mdia, trak) sowie content section.

- **Data entity:** Repräsentiert den tatsächlichen Speicher der für die a/v Daten benutzt wird.
- **Media entity:** ist eine zeitliche Sequenz, in media time gemessen. Jede Entity hat einen Medientyp (entweder a (audio) oder v (video)). Die Elemente referenzieren Speicherregionen.
- **Track entity:** beschreibt die Anordnung von Media Entitäten. Jede Entitäten ist in einer Movie Entity enthalten. Zeit wird in Movie TCS gemessen.
- **Movie entity:** Ist eine Gruppe von Track Entities. Spezifiziert die Zeitskalierung und Dauer. Es beschreibt also die Anordnung, Dauer, Verzögerung („Offset“) von Tracks innerhalb des Movies.

Erklären Sie kurz das Zeitkonzept in QuickTime. [2]

- Zeit ist explizit
- Werte werden als Integer ohne Vorzeichen (Unsinged Integers) repräsentiert
- Die Zeitskalierung wird in units per second angegeben
- Die Dauer wird mit dem Maximalen Zeitwert angegeben.
- Die Timebase („Zeitbasis“) gibt die Playbackrate an.
- Verwendung eines Time Coordinate Systems (TCS) für Skalierung und Dauer

Worin unterscheiden sich die 2 Formate von Videodiscs? Listen Sie kurz Vor- und Nachteile dieser Formate. Welches Format war Vorbild für die CD-Audio? [2.5]

Die Videodisc steht am Scheitelpunkt der digitalen und analogen Medien, ist aber selber ein digitales Medium. Alle Videodiscs sind flache Scheiben und vom Durchmesser mit einer Langspielplatte zu vergleichen. In den siebziger Jahren konkurrierten verschiedene Videodisc-Systeme miteinander. Das Problem aller Systeme bestand darin, gewaltige Informationsmengen kodiert auf relativ begrenzten Flächen unterzubringen.

Das System, das am meisten verbreitet war, ist LV (Laser Vision). Dabei wurde ein Laser verwendet, um kleine Pits (Vertiefungen) in die Oberfläche der einer Master-Disc zu brennen. Die Pits waren in einer langen Spirale, die in der Mitte startete, angeordnet. Von der Master-Disc wurde ein Abdruck gemacht, und aus diesem wurden die Discs für den Verkauf gegossen. Gelesen wurden die Videodiscs von einem Laser, der die Pits abtastete und dadurch in binäre Signale umwandelte. Logischerweise waren diese Discs Read-Only.

LV – Discs werden in 2 Formaten angeboten:

CAV (constant angular velocity)

Angewendet wird dieses Format bei DVD- RAM und ist auch Vorbild für CD-DA. Schallplatten, Laserdisc (LD),…

- konstante Winkelgeschwindigkeit (Platte dreht sich mit konstanter Geschwindigkeit)
- Je weiter der Lesekopf nach außen bewegt wird, desto höher ist die relative Geschwindigkeit zwischen Kopf und Plattenoberfläche
- Die Datenrate pro Umdrehung bleibt gleich
- Dadurch ist die Dichte der Daten nach außen hin immer geringer
- CAV Discs haben ein flexibles Playback: Pause (Still, Bild einfrieren), Play, Forward, Reverse in verschiedenen Geschwindigkeiten.

Vorteil: flexible Abspielmöglichkeiten (freeze frame, vorwärts oder rückwärts abspielen mit verschiedenen Raten) durch Datenanordnung
Nachteil: schlechte Nutzung der Kapazität

CLV (constant linear velocity)

Standard für CD-ROM, Audio Compact Disc (Audio- CD)

- konstante lineare Geschwindigkeit (Absenkung der Drehzahl nach außen hin)
- Je weiter der Lesekopf nach außen bewegt wird, desto mehr wird die Umdrehungsgeschwindigkeit gesenkt.
- Die relative Geschwindigkeit zwischen Lesekopf und Oberfläche bleibt annähernd gleich.
- Datendichte und Datenrate ebenfalls.

Vorteil: doppelte Kapazität (Speicherplatz) im Vergleich zu CAV
Nachteil:

- längere Suche und schlechtere Positionsmöglichkeit (nicht frameadressierbar)
- nicht so gute Playback-Funktionen
- schlechtere Zugriffszeit

Vorbild für die CD- Audio

CD Audio wurde von Philips und Sony entwickelt, in Anlehnung an die LV Videodisc.

Unterschiede:

- Echtes digitales Medium (LV benutzt nur analoge Informationen).

- Die CD verwendet nur eine konstante lineare Geschwindigkeit (CLV).

Beschreiben Sie kurz CD-i. Beschreiben Sie kurz die wichtigsten Unterschiede zwischen CD-ROM und CD-i. [3]

CD-i ist die erste MM-Plattform Anfang der 90er Jahre gewesen. Es ist ein komplettes System im Gegensatz zu CD- Rom. Sowohl logische als auch physikalische Organisation von Multimediainformationen 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 archictecture, die eine offene Form der CD-i ist und beide Formate lesen kann)

Die CD-i gibt dem Abspielgerät zusätzliche Informationen über den auf der CD vorhanden Medientyp (Video, Bilder, Audio, Daten), die alle in verschiedenen vorgegebenen Standards mit unterschiedlichen Qualitätsstufen gespeichert werden können.

Auf CD-i-Playern lassen sich Audio-, Photo- und VideoCDs als auch interaktive CDs mit Spielen oder Lernsoftware abspielen. Mit ihr waren schließlich erstmals interaktive Videos möglich.

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 Abbildung von Media Time zur Speicheradresse in QuickTime. Welche Datenstrukturen werden dafür benötigt? [3.5]

Zu einer bestimmten Media Time werden entsprechende Samples abgefragt welche Chunks zugeordnet sind, diese beinhalten die Information über die Speicheradresse.

Media Value: Die Media Value weißt die Samples den einzelnen Chunks zu.

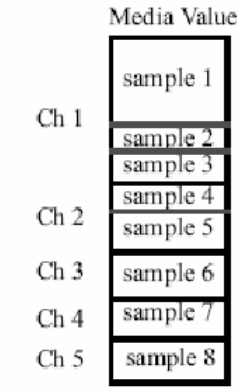
Time to Sample

Bei Time to Sample werden die Samples der Zeit zugeordnet
Sample number: Sample Nummer
Sample Span: Anzahl der einander gereihten Samples der gleichen Größe
Sample Start Time: Anfangszeit des ersten Samples
Sample Duration: Dauer eines einzelnen Samples

Sample to Chunk

Effiziente Speicherung der Zuordnung von Samples zu den Chunks
Chunk Number: Chunk Nummer
Chunk Span: Anzahl der einander gereihten Chunks der gleichen Größe
Start Sample in Chunk: Nummer der ersten Samples im jeweiligen Chunk
Samples/ Chunk: Anzahl der Samples im jeweiligen Chunk
Encoding: Welches Encoding wird hier verwendet

Hierzu benötigt man für die Zeit die Datenstruktur media time, für die Samples media und für die Chunks BLOB (Binary Large Objects) (data entities)



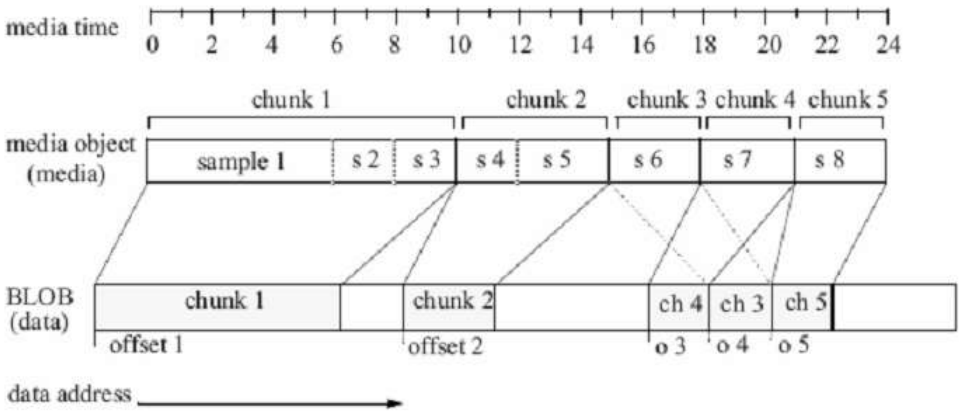
Time-to-sample

sample number	sample span	sample start time	sample duration
1	1	0	6
2	3	6	2
5	4	12	3

Sample-to-chunk

chunk number	chunk span	start sample in chunk	samples /chunk	encoding
1	1	1	3	E1
2	1	4	2	E1
3	2	6	1	E1
5	1	8	1	E2

Durch diese Tabellen wird Speicherplatz gespart. 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. Welche Besonderheit haben die Tabelleneinträge, z. B. bei aufeinanderfolgenden Samples gleicher Länge? [5]

Algorithmus

1) Rechne Movie-Time mit Hilfe des TCS (Time Coordinate System) 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 des vorherigen.
- Die Sample-To-Chunk-Tabelle enthält Chunks nicht, wenn sie die gleiche Sample-Anzahl wie der vorherige Chunk haben und gleich codiert sind.

Datenkompression

Worin unterscheidet sich die I-Frame-Kodierung des MPEG-Verfahrens von der JPEG-Kodierung [2] bzw. Worin unterscheiden sich die Quantisierungsschritte im JPEG- und MPEG-Verfahren? [2]

JPEG Verfahren

Die Quantisierung erfolgt, je nach Qualitätsstufe (bspw. 95%), mit der selben Quantisierungsmatrix. Hierbei wird jeder der Werte aus der Frequenzmatrix durch den zugehörigen Wert in einer vordefinierten Quantisierungsmatrix dividiert und auf Integer gerundet => es entsteht eine Gewichtung.

MPEG Verfahren

Die Quantisierung wird an die Datenrate angepasst. Quantisierung hat konstanten Wert für alle DCT Koeffizienten. Wenn nun sich die Datenrate erhöht und einen bestimmten Grenzwert übersteigt, erhöht die Quantisierung die Schrittweite. Wenn aber nun sich die Datenrate verringert, wird die Quantisierung mit feinerer Granularität (Schritten) durchgeführt.
Bei I-Frames ist der Quantisierungswert für alle Makroblöcke der gleiche. Bei P-Frames wird er 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

Beschreiben Sie kurz die MPEG-4 Videokodierung. [5]

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.

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. [6]

```
build ordered list of symbols (increasing probability)
do while list contains at least 2 elements
        construct tree using the first two elements in list
        add parent node for the union of these elements and
        compute probability
        mark edges by '0' and '1' (the '1'-edge points to the element with greater probability)
        delete the first 2 elements in list; insert parent into list (on right position)
```

end

Was können Sie über die AC und DC Koeffizienten der DCT aussagen, wenn der zu transformierende 8x8 Block eines Bildes aus einem Schachbrettmuster besteht? [2]

Das erste Pixel im Block (links oben) ist der DC-Koeffizient. Er hat die geringste Frequenz und gibt die Grundfarbe des gesamten Blocks vor. Alle übrigen Pixel (die AC-Koeffizienten) haben höhere Frequenzen und können vom menschlichen Auge nicht mehr so gut unterschieden werden, wie niedrigere.

Der DC Koeffizient liegt an (0,0) und gibt die mittlere Graustufe des Blocks an. Die AC Koeffizienten sind entsprechend hoch, da ein Schachbrettmuster extrem hohe Ortsfrequenzen aufweist. 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? [2]

Scharfe Schnittkante → 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 (da die Überschneidung der 2 Flächen eine hohe Differenz angeben).

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? [2.5]

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 handeln, in dem sich zum Beispiel zwei verschiedenfärbige Flächen horizontal mit einer scharfen Kante („horizontale Linie“) schneiden.

Erklären Sie kurz den Begriff Quantisierung im Kontext der Sampling-Theorie und im Kontext von Komprimierungsverfahren. [2]

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

Bei Komprimierungsverfahren werden benachbarte Werte auf einen neuen Wert abgebildet. Jeder Wert in einer Frequenzmatrix wird durch seinen zugehörigen Wert in der Quantisierungstabelle dividiert und auf Integer gerundet um möglichst viele AC-Koeffizienten abzuschwächen (also auf 0 zu setzen) 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. Hohe Frequenzen werden also stärker komprimiert (eher auf 0 gesetzt), da das menschliche Auge diese schlecht wahrnimmt.

Erklären Sie kurz das Konzept Komposition im MPEG-4 Systems Standard. [3]

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 kreiert, welche audiovisuelle Szenen formen) beruhend auf den Konzepten von VRML. Autoren können die Beschreibung im Textformat generieren, möglich auch mit Hilfe eines Authoringtools.

Komposition

- Beschreibt ein Modell zur Komposition von Multimediaobjekten
- Behandelt nur 2D Inhalt
- Gekoppelt mit Streaming Media wie z.B. Video, Audio, Streaming Text.
- Ist auch im Stande Synchronisation durchzuführen
- Die Szenenbeschreibung wird in einer binären Repräsentation kodiert, nämlich in der BIFS (Binary Format for Scene Description) der Effizienz wegen.
- Eine Multimedia Szene beinhaltet hierarchische Strukturen und wird als Graph repräsentiert. Jedes Blatt im Graph repräsentiert ein Mediaobjekt.
- Alle Knoten und Blätter der Graphen der Szene werden an den zugehörigen Stream empfangen oder gesendet. Alle Kanten und Grafenblätter, die Streamingsupport benötigen um Medieninhalte zu erhalten, sind logisch mit den DekodingPipelines verbunden.

Beschreiben Sie kurz die Facial Animation und Facial Definition Parameters im Face Animation-Teil von MPEG-4. [2] bzw. Erklären Sie kurz die Facial Animation und Facial Definition Parameter im Face Animation-Teil von MPEG-4. [2]

Facial Animation beschäftigt sich mit den Parametern für Gesichtsanimation und Gesichtsdefinition.

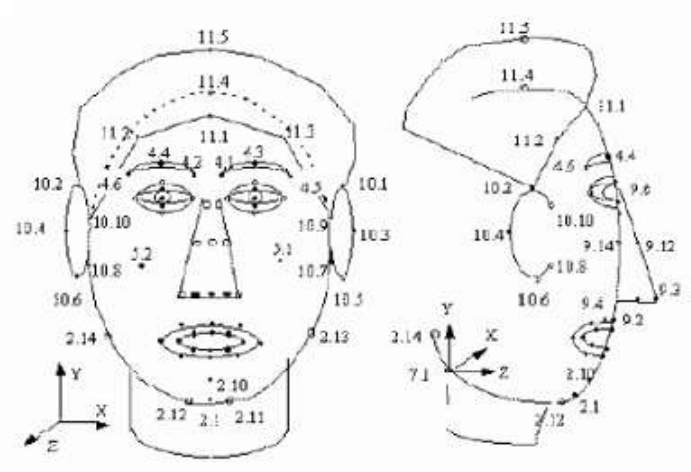
Facial Animation Parameters (FAP)

Erlauben das Anwenden von einem einzelnen Set von Parametern, unabhängig von dem verwendeten Gesichtsmodell das benutz wird. Die meisten FAPs beschreiben atomare Bewegungen von Gesichtsfeatures (Gesichtszügen), andere (expressions und visemes) definieren viel komplexere Deformationen/ Deformierungen.

Viseme definieren die Position des Mundes zusammenhängend mit den Phonemen. Sie sind auf die Position und Bewegungsabläufe einzelner Gesichtsteile (Mund, Augen…) spezialisiert. FAPs können entweder arithmetisch oder mit DCT encodet werden

Facial Definition Parameters (FDP)

Werden verwendet um das Receiver Terminal Default Face Model zu kalibrieren (adaptieren oder verändern der Form), oder um eine komplett neue Gesichtsmodelgeometrie und Textur zu erstellen und/oder zu übertragen.



Erklären Sie knapp den Unterschied zwischen Inter- und Intraframe-Synchronisation. Geben Sie 2 Geräte an, die der Synchronisation dienen. [3] bzw. Was versteht man unter Intra-Flow Synchronization und welche Störungen soll sie vermeiden? [2.5]

Inter-Frame-Synchronisation

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.

Intra-Frame-Synchronisation

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.

Zusammengefasst

- Interframe/ Interflow: Zusammenführen und Synchronisieren von 2 Streams (z.B. Audio und Video) um auseinanderlaufen von Ton / Bild zu verhindern.
- Intraframe/ Intraflow: Synchronisation eines Stream (z.B. von Videodaten) untereinander um ruckeln zu vermeiden.

Erklären Sie kurz die verschiedenen Frametypen der MPEG-Komprimierung. [1.5] bzw. Erklären Sie kurz die P- und B-Frame Kodierung der MPEG-Komprimierung. [2]

I-Frames (intracoded Frames)

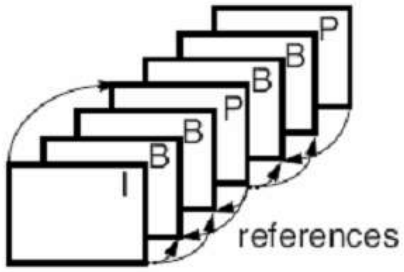
Ein I-Frame ist ein kodiertes Vollbild. Benötigen keine Referenz auf andere Bilder und werden wie stehende Bilder behandelt. MPEG verwendet JPEG für I-Frames.
Komprimierung muss in Echtzeit passieren, deswegen ist die Komprimierungsrate die geringste innerhalb von MPEG.
I-Frames sind Punkte für den Zufallszugriff 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. Nur die Differenz zum vorhergehenden Bild wird weiter behandelt. Es ist möglich die Bewegung der Makroblöcken zu berücksichtigen – Motion Compensated Prediction

Zeitliche Redundanz: Bestimme den letzen 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 von einander 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. Sind dem P-Frame ähnlich, verwenden zur Kodierung allerdings den vorhergehenden und den folgenden Frame. Wenn Motion Compensation verwendet wird, werden Bewegungsvektoren in die „Vergangenheit“ und in die „Zukunft“ gespeichert.

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 = Block BlockNext, B = 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. Nur das Matrixelement a(00) der DCT-Koeffizienten wird verwendet. 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.

Beschreiben Sie kurz die Aufgabe von D-Frames im Rahmen von MPEG. Wie werden D-Frames kodiert? [2]

D-Frames sind intraframe-kodiert und werden bei MPEG für schnelles vorwärts bzw. rückwärts spulen verwendet (fast forward, fast rewind Modi). Nur die DC Komponente wird bei D-Frames kodiert, die AC Koeffizienten werden verworfen. Daher bestehen D-Frames nur aus den niedrigsten Frequenzen eines Bildes. Sie verwenden nur eine Art von Makroblocks und werden nur in MPEG-1 benutzt.

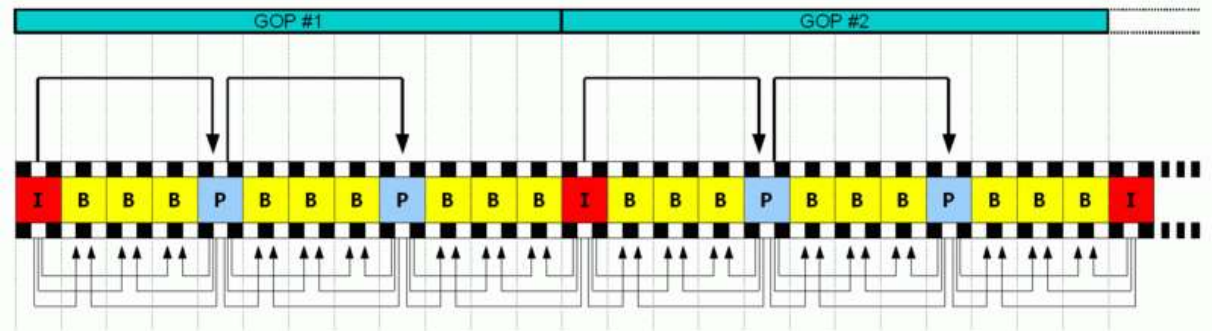
Was versteht man unter Group of Pictures? Wie hängen Group of Pictures Muster mit der Datenrate zusammen? [2.5]

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 ist jeweils eine zusammenhängende Folge von I-, P- und B-Frames. Z.B. IPPBPPBPP*IPPBPPBPP* (den Beginn der neuen GOP hab ich kursiv markiert).

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 (und die sind ja die aufwendigsten!) 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.

Erklären Sie kurz die Begriffe Video Object und Video Object Plane. In welchem Kontext werden sie verwendet? [3]

Werden bei der MPEG-4 Codierung verwendet.

Video Object Plane

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

Video Objects

Aufeinanderfolgende VOPs, die zum selben physikalischen Objekt in einer Szene (z.B. ein Mensch) 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.

Für welche Einsatzbereiche würden Sie das MotionJPEG-Verfahren dem MPEG-Verfahren vorziehen? [1.5]

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

Formulieren Sie in Pseudocode einen Algorithmus, der in einer quantisierten DCT-Matrix die Runlength-Kodierung durchführt. [3.5]

```
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
                 }
        }
```

Anders formuliert

Für alle Matrixeinträge (also 64)
    Vergleiche Matrixeintrag (i) mit Matrixeintrag (i+1)
    Wenn beide gleich, dann Zähler um eines erhöhen

Wenn die Anzahl der gleichen Buchstaben größer 3 und nächster Matrixeintrag (i+1) nicht gleich sind, dann gib „!“ + Matrixeintrag (i) + Anzahl der Buchstaben aus
Sonst wenn Matrixeintrag (i+1) nicht gleich gibt Matrixeintrag (i) aus
Wenn gleich geh weiter
Erhöhe i um 1

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

Profil

Seit MPEG2 dienen Profiles der Definition eines Liste von Codierungs-Möglichkeiten, zusammen mit Levels definieren sie die Mindestanforderungen an eine Decoder-Klasse. Ein Profile legt fest welche Tools und Funktionen dazu verwendet werden können einen bitstream zu erzeugen und wie dies zu geschehen hat.

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 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.

Level

Die Level legen innerhalb eines Profiles wichtige Bildparameter fest. Hierzu gehören u.a. die unterschiedlichen Bildauflösungen (z.B. Mainlevel ML => 720x576), 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.

Beispiele

Profile: Simple, Main, 4:2:2, SNR, spatial, high, multiview
Levels: low (SIF), main (CCIR 601), high-1440, high (HDTV)

Nennen und erläutern Sie kurz 2 Arten von Artefakten bei der Audiokodierung. [2.5]bzw. Beschreiben Sie kurz 2 der häufigsten Artefakte bei der Audiokompression. [2.5]

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. Z.B. vor einem lauten plötzlichen Geräusch (zum Beispiel Schlagzeug) sind klirrend-metallische Artefakte hörbar. 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. Stärkenänderung beginnen, insbesondere wenn ein Bereich mit sehr niedrigen Amplituden vorausgeht.

roughness, double-speak

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

Bandbreitenverlust

Nennen und erläutern Sie kurz 2 Arten von Scalable Bit Streams in MPEG-2. [3] bzw. Wozu dienen Spatial Scalability und SNR Scalibility in MPEG-2? Was haben sie gemein und was sind die Unterschiede? [3] bzw. Wozu dienen Temporal Scalability and SNR Scalibility und was sind die Unterschiede? [2.5]

Es gibt 4 skalierbare Modi – MPEG2 Video wird in verschiedene Layer aufgeteilt, meist mit der Absicht Videodaten zu priorisieren:

- **Spatial scalability:** Kodiert einen Basislayer mit niedrigeren Samplingdimensionen (Auflösungen) als die höheren (darüberliegenden) Layer. Dann hat man die normale Auflösung und die reduzierte
- **Data partitioning:** Teilt den Block von 64 (8x8) quantisierten Koeffizienten in 2 Bitstreams. Der erste, höher priorisierte Bitstream beinhaltet die kritischeren niederfrequenten Koeffizienten und Randinformationen (DC Werte, Bewegungsvektoren)
- **SNR scalability:** Macht auch mehrere Channels mit gleicher Samplingrate verschiedene Qualität (Bildqualität) wegen unterschiedlicher 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

Unterschiede - Gemeinsamkeiten von Spatial Scalability und SNR Scalibility

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

Unterschiede - Gemeinsamkeiten von Temporal Scalability und SNR Scalibility

Temporal Scalability hat zwei Streams mit unterschiedlicher Priorität und verwendet den ersten Stream als Prognose. SNR ist frei skalierbar (beliebige Anzahl an Streams).

Erklären Sie kurz die 4 Arten(Modi) der Makroblock Codierung von B-Frames in MPEG-4 [3]

Vorwärts

Referenzframe ist vorheriger I- oder P-Frame, kodiert wird ein Vektor und Fehlerbild (=^ Prädiktionsfehler)

Rückwärts

Referenzframe ist nachfolgender I- oder P-Frame, kodiert wird ein Vektor und Fehlerbild (=^ Prädiktionsfehler)

Interpolationsmodus

Beide Vektoren (aus Vorwärts- und Rückwärtsprädiktion) warden übertragen. Übertragenes Fehlerbild resultiert aus Interpolation beider Referenzwerte.

Direkter Modus

Zwei Vektoren wie beim Interpolationsmods, aber von einem Vektor abgeleitet:

- Vektor linear skaliert nach Zeit
- Nur Delta Vektor kodiert
- Bewegungsvektor im nachfolgenden Frame, gleiche Position wie Makroblock im akt. B-Frame

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 [2.5]

Verschiedene Möglichkeiten der Darstellung, je nach Beschriftung der Kanten!
A = 01 B = 1 C = 001 D = 0000 E = 0001

Ermitteln Sie eine Huffman-Codierung der Symbole V, W, X, Y und Z; Wahrscheinlichkeiten ihres Auftretens: p(V)=0.2, p(W)=0.4, p(X)=0.18, p(Y)=0.15, p(Z)=0.07 [2.5]

V = 01 W = 1 X = 001 Y = 0000 Z = 0001

Was sind die wesentlichsten Unterschiede zwischen MPEG-1 und MPEG-2 Video? [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
- Progressivesund interlaced Framekodierung
- zusätzliche Prediction Modi sowie zugehörige Makroblöcke
- 4 Skaliermodi (verschiedene Qualitätsstufen in einem Videostream)
- Verbeserte Bildqualität (Quantisierung, Zig-zag-scan)
- Erhöhung der Genauigkeit der Bewegungsvektoren auf halbe Bildpunkte
- erweiterte Fehlerredundanz durch spezielle Vektoren bei I-Frames
- Präzision der DCT (diskreten Cosinus-Transformation) wählbar

Was versteht man unter Pre-Echo? Wo tritt dieses Phänomen auf? [2]

Pre-Echo

Pre-Echo ist ein Artefakt, das bei der Audio-Kodierung entsteht. Vor einem lauten plötzlichen Geräusch (zum Beispiel Schlagzeug) sind deutliche Artefakte zu hören. 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.

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. Z.B. vor einem lauten plötzlichen Geräusch (zum Beispiel Schlagzeug) sind klirrend-metallische Artefakte hörbar. 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. Stärkenänderung beginnen, insbesondere wenn ein Bereich mit sehr niedrigen Amplituden vorausgeht.

Post-Echo

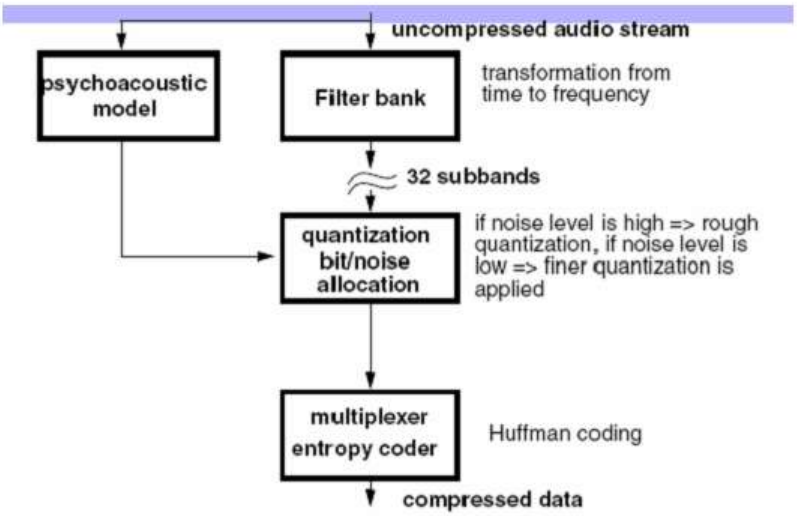
Nach einem plötzlichen Geräusch sind deutliche Artefakte zu hören

Beschreiben Sie kurz die wesentlichen Schritte der MP3-Kodierung. Erklären Sie insbesondere Quantisierung und Kodierung. [5] bzw. Erklären Sie knapp die Schritte der MP3-Kodierung. [5]

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 Eingang- Signal in Subbänder geteilt. Dieser Prozess wird auch als Zeit-Frequenz-Zuordnung bezeichnet.
- Der nächste Schritt der Kodierung ist die Implementierung des psychoakustischen Modells. Es wird eine Schätzung des tatsächlichen Maskierungsgrenzwert errechnet.
- Im dritten Schritt erfolgt die Quantisierung der Subbands mit der Information aus dem psychoakustischen Modell
- Der letzte Schritt ist die endgültige Erzeugung der komprimierten Datei, wobei Huffman codiert wird und die Datei durch aneinander reihen der Frames erzeugt wird.



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

Welchem Zweck dienen BIFS in MPEG-4? Welche Informationen enthalten sie? [2.5]

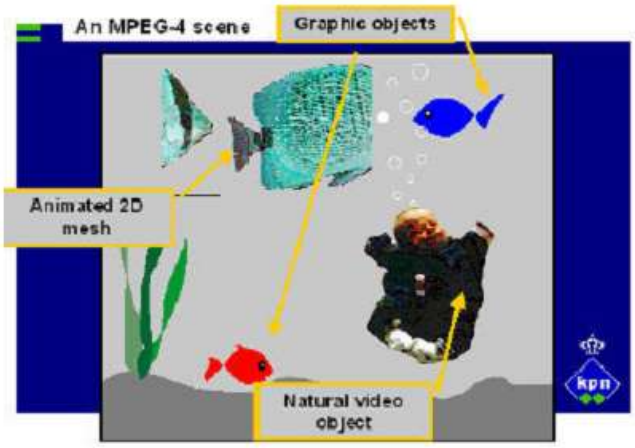
BIFS ist die Abkürzung für Binary Format for Scenes, eine auf VRML97 basierende und in MPEG-4 part 11 (ISO/IEC 14496-11, „Szenenbeschreibung und Anwendungs-Engine“) standardisierte Beschreibungssprache für zwei- und dreidimensionale multimediale audiovisuelle interaktive Inhalte. Sie wird binär codiert und wird bei mpeg4 verwendet um den Szenegraph einer Komposition zu speichern.

Örtlich-zeitliche Zuordnung einzelner Objekte eines Datenstromes können unabhängig voneinander beschrieben werden. 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 Studiotechnik als Blue-Keying bekannt ist.

Waren MPEG-1 und MPEG-2 noch Standards, um ausschließlich Audio- und Videodaten zu codieren, war MPEG-4 von Anfang an als "Toolbox" zur Codierung und Übertragung von (nicht notwendigerweise) interaktiven audiovisuellen Inhalten geplant. Wenn im folgenden MPEG-4 erwähnt wird, ist immer der Szenenbeschreibungs-Aspekt gemeint, d.h. die audiovisuellen Inhalte werden als sog. Szene beschrieben, die aus natürlichen (z.B. Videosequenzen oder aufgenommene Audiospuren) und synthetischen Objekten (z.B. 2D- und 3D-Grafiken) bestehen kann. Dazu kommt in MPEG-4 erstmalig das Konzept der objektbasierten Codierung zum Einsatz, bei dem die einzelnen Objekte in einer Szene (z.B. Hintergrund, Darsteller, 3D- und 2D-Objekte, Sprache, Hintergrundmusik etc.) getrennt codiert und übertragen werden. Bei der Wiedergabe wird die Szene dann wieder zusammengesetzt.

Die Vorteile der objektbasierten Codierung und Übertragung

- Sie ermöglicht eine äußerst effiziente Übertragung bzw. Speicherung von multimedialen Inhalten. So reicht es bspw. aus, den Hintergrund der Szene einmalig als Bild zu übertragen (solange die Kamera nicht bewegt wird).
- Für jedes Objekt kann ein passender Codec aus einem der MPEG-Standards verwendet werden (z.B. Sprachcodec, Still-Image-Codec, Videocodec etc.).
- Einzelne Objekte können bei der Produktion von anderen Szenen einfach wiederverwendet werden.
- Die Szene kann im Wiedergabegerät an dessen Fähigkeiten angepasst werden.
- Es sind interaktive Inhalte möglich.



Der Betrachter kann sich in dreidimensionalen Szenen eingeschränkt oder frei bewegen. Weiterhin sind Aktionen (z.B. Start eines Video- und/oder Audio-Clips, Weiterleitung auf einen Webshop, in dem man den betreffenden Artikel erhalten kann etc.) beim Anklicken eines Objektes in der Szene denkbar.

Erklären Sie kurz die Huffman-Kodierung von DC-Koeffizienten. [3.5]

Bei der Huffman-Kodierung werden Zeichenketten innerhalb der zu kodierenden Daten durch kürzere Zeichenketten ersetzt. Häufig vorkommende Zeichenketten werden durch möglichst kurze Symbole ersetzen, selten auftretende Zeichenketten durch längere Symbole.

Die innerhalb von JPEG verwendeten Tabellen beruhen auf der Anzahl der DC-Koeffizienten mit dem Wert Null, die einem Koeffizienten mit einem Wert ungleich Null vorausgehen. Üblicherweise findet die in definierte Huffmantabelle Verwendung.

Diff Values	SSSS (number of bits needed to encode Diff
0	0 bit
-1, 1	1 bit
-3, ..., 3	2 bits
-7, ..., 7	3 bits
.	
.	
.	
-2047, ..., 2047	11 bits

Hier werden den Diff Values die benötigten Bits zur Encodierung zugeordnet (SSSS). Nun behandelt man die SSSS (Anzahl der für die Diff-Kodierung benötigten bits) wie Huffmancode Symbole, wobei deren Wahrscheinlichkeit p für alle 12 Varianten (0 -11 bits) ausgerechnet und dadurch ein Huffman-Baum erstellt wird.

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

Beispiel

Wenn SSSS=2 den Huffman Code 001 und die Differenz -3 hat → sende 00100
001: HuffmanCode für 2. 00 entspricht -3 in Zweierkomplementdarstellung

Erklären Sie das Prinzip der Lauflängenkodierung (Run Length Encoding). [1.5]

Die Run-Length Codierung ist ein Kompressionsverfahren zur verlustfreien(!) Kompression von

Daten. Es nutzt sich wiederholende Zahlenwerte in Folge, die es mit der Anzahl der Zahlenwerte zusammen als kurzen Code hinterlegt. Es ist ein Entropy-coding Algorithmus, inhaltsabhängige Codierung.

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

Methode

1. Wenn die selben Bytes mindestens 4 mal hintereinander vorkommen, zähle die Anzahle, wie oft sie hintereinander vorkommen
2. Schreibe die komprimierten Daten in der Form: gezähltes Byte + ‚!' + Vorkommensanzahl

Beispiel

unkomprimiert: ABCCCCCCCCCDEFFFFGGG
kompriniert: ABC!9DEF!4GGG

Was sind Quantisierungstabellen, wie sind sie aufgebaut und weshalb setzt man sie ein? [2]

Quantisierungstabelle

Eine Quantisierungstabelle ist eine 8x8 Pixel große Matrix (hat also 64 Elemente), mit einem DC-Koeffizienten (erstes Pixel links oben) und 63 AC-Koeffizienten. Bei der JPEG- Quantisierung wird für jeden Wert dieser Tabelle 8 Bit verwendet.

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.

Weshalb setzt man sie ein?

Zur Komprimierung der Koeffizienten bei JPEG. Die 64 Koeffizienten (=Quantisierung/ Frequenzmatrix die durch DCT gewonnenen wurde) werden durch die Werte in der Quantisierungstabelle jeweils dividiert und ergeben die komprimierten Werte der einzelnen Koeffizienten (und werden anschließend auf Integer gerundet).

Was sind die Aufgaben des Synchronization Layers in MPEG-4? [1.5]

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

Die Informationen, die in den Synchronization Layers -Headers enthalten werden, behalten die korrekte Time Base für die grundlegenden Decoder und für den Empfängerterminal, plus die korrekte Synchronisierung in der Darstellung der grundlegenden Medienobjekte in der Szene bei.

Die Clock Referenzen Mechanismus unterstützt Timing des Systems, und das Mechanismus der Time Stamps. Unterstützt Synchrounisierung der unterschiedlichen Medien.

Was versteht man unter Perceptual Measurement Techniques und wofür werden sie eingesetzt? [1.5]

Zur Messung der Qualität von Codecs.

- State-of-art ist nicht ausreichend, um große Skala und die gut vorbereitete Hörtests zu überholen.
- Jedoch sind Perceptual Measurement Techniques 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 (Int. Telecommunications Union,Radiocommunications sector) entwickelte eine Empfehlung für das System PEAQ. (Perceptual Evaluation of AudioQuality)

Welchen Zweck hat die Diskrete Cosinus-Transformation in Kompressionsverfahren (z.B. JPEG)? [2] bzw. Welchen Zweck haben Transformationen in den Frequenzraum im Kontext von Kompressionsverfahren (z.B. DCT bei JPEG)? [2]

Menschen haben nur begrenzte Fähigkeit hohe Frequenzen wahrzunehmen. Bilder die ihre hohe Frequenzen verringern könnten dabei einen hohen Grad an Qualität gewinnen während sie ziemlich beträchtlich komprimiert werden.

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 („Outlines“) von Bildobjekten, währenddessen hohe Frequenzkomponenten feine Strukturen darstellen. Dies ist möglich, weil das menschliche Auge nur begrenzt fähig ist hohe Ortsfrequenzen wahrzunehmen.

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),…,S(7,7)) sind die sogenannten AC-Koeffizienten, wobei die Frequenz von zumindest einer Richtung ungleich 0 ist.

Um die Hochfrequenzen zu verringern müssen die Daten nach Frequenzen sortiert werden->DC

Was sind - abgesehen von Leistungsunterschieden - die wichtigsten Unterschiede zwischen MP3- und AAC-Kodierung? [2]

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). Kein Filter Bank Design, sondern alleine MDCT als Basistransformationen.

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

Was versteht man unter Differential Encoding? Nennen Sie Beispiele, wo das Verfahren eingesetzt wird. [2]

- Differential Encoding ist Teil des source codings.
- Betrachtet man eine Sequenz von Symbolen S1, S2, S3… mit von Null verschiedenen Werten, die sich aber nicht sehr stark unterscheiden, so berechne die Differenz zum vorhergehenden Wert.
- Nullen können mit Lauflängenkodierung kodiert werden.

Beispiele:

- Statische Hintergründe (Videokonferenzen, Nachrichten,…).
- Still Images

- MPEG verwendet die Kodierung (→ **motion compensation**, motion vectors): 8x8 Blöcke werden verglichen; areas similar, only shifted, e.g., to the right (motion vector)

Was versteht man unter Structured Audio in MPEG-4? [2.5]

- 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? [2.5]

- Der composition Stream (BIFS) hat seine 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 (**DTS**, decoding time stamp) bereit sein muss und wann die Kompositionseinheit am Kompositorinput (**CTS**, composition time stamp) 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.
    - For example, the start of a video clip represents the relative offset between the composition time stamp of the scene and the start of the video display.

Erklären Sie kurz die wesentlichen Charakteristika des MPEG-4 Composition Streams. [3]

- The composition stream is treated differently from others because it provides the information required by the terminal to set up the scene structure and map all other elementary streams to the respective media objects.
- The composition stream (BIFS) has its own time base.
- Time stamps associated to the composition stream specify at what time the access units for composition must be ready at the input of the composition information decoder. (CTS)

Erklären Sie kurz den Skalierfaktor und den Zusammenhang mit Kompressionsverfahren. [2]

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. [2]

Kritische Abtastung digitaler Filterbanken

- Wortlängen bleiben in Subbands unverändert
- Abtastrate in Subbands reziprok zur Anzahl der Subands
- 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. [2]

Zwei Töne erklingen gleichzeitig mit unterschiedlicher Lautstärke und ähnlicher Frequenz.

Maskierungseffekt – Ohr hört nur den lauten Ton

- leiser Ton wird "maskiert"
- der laute Ton mindert Empfindlichkeit des Gehörs im umliegenden Frequenzbereich
- leiser Ton liegt unter neuer Hörschwellenkurve

Maskierungsschwellwert

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. [2]

- 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)

Daraus ergibt sich z.B. bei der MP3 Kodierung, dass man die Tatsache nutzt, dass der Mensch nur einen bestimmten Frequenzbereich wahrnehmen kann. Der nicht hörbare Bereich wird weggeschnitten. Dadurch wird die Datengröße reduzieren, ohne eine echte Kompression durchgeführt zu haben.

Ähnlich verhält es sich mit der kritischen Bandbreite. Der Frequenzbereich, der vom Menschen nicht getrennt wahrgenommen werden kann, wird bei der Kompression berücksichtigt und weggeschnitten, damit man auch alle Töne getrennt wahrnehmen kann.

Retrieval und Indexing

Nennen Sie ein von geometrischen Transformationen unabhängiges Image-Feature. Diskutieren Sie kurz Vor- und Nachteile. [2]

Farbe- bzw. die Verteilung von Farbe innerhalb von Bildern. Wird in Form von Histogrammen dargestellt. Ein Histogramm gibt die Anzahl der Pixel je vorkommender Farbe an. Die Anzahl der Farben kann dabei frei festgelegt werden.

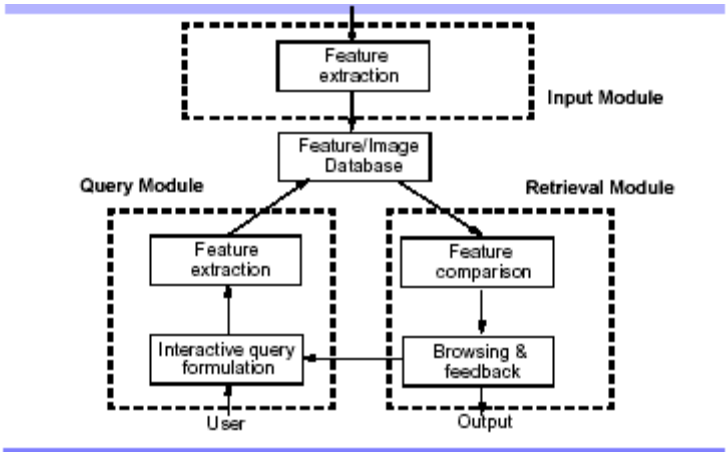
Vorteil der Histogramme ist die Unabhängigkeit gegenüber Translationen und Rotationen entlang der View-Achse. Verändert sich nur wenig, wenn sich der Ansichtswinkel oder der Maßstab verändert.

2 Histogramme können folgendermaßen verglichen werden:

- Intersection: Anzahl der Pixel mit gleicher Farbe, eventuell normalisiert auf eines der beiden Bilder.
- Distance: Alle Farben werden im Histogramm unabhängig voneinander betrachtet. D.h. keine Ähnlichkeiten zwischen Zellen. Lösung über Distanz $D(I,Q) = (I-Q)_t A(I-Q)$.

Beschreibung Sie kurz die Architektur eines Content-Based Image Retrieval-Systems [3]

Ziel ist die Suche in einer Bilddatenbank nach entsprechenden Kriterien.



1. Input Mode

Beim Import der Bilder in die Datenbank sind die entsprechenden Features zu extrahieren und zusammen mit den Bildern in der Datenbank abzulegen.

2. Query Mode

Wird nach Bildern gesucht, hat der Benutzer die entsprechenden Anfragen zu formulieren (RBR, ROA, RSC…). Dies sollte im besten Fall interaktiv geschehen können. Aus diesen Anfragen werden die für die Suche benötigten Features in der Datenbank ermittelt und als Basis für die Suche verwendet.

3. Retrieval Mode

Die von der Benutzeranfrage ermittelten Features werden mit denjenigen der Bilder in der Datenbank verglichen. Dieser Vergleich liefert mitunter eine Menge von Ergebnisbildern, die der Anfrage entsprechen. Daher kann das Ergebnis im nächsten Schritt manuell durchsucht und entsprechend weiter eingeschränkt werden (➔ Neue Query). Vorgang solange, bis das gewünschte Ergebnis erreicht wurde.

Was ist die Brodatz Datenbank und weshalb ist sie von Bedeutung? [2] bzw. Beschreiben Sie kurz die Evaluierung von Textur-Features mit Hilfe der Brodatz Datenbank [3.5]

Diese DB ist der de facto Standard für die Evaluierung von Textur- Algorithmen und gibt an wie gut unterschiedliche Algorithmen Texturen in Bildern erkennen können. Sie besteht aus einem Fotoalbum mit 112 Graustufen-Bildern (512x512x8 Bit), welche unterschiedlichen Texturklassen repräsentieren.

Texturauswahl- Algorithmus

9 Subimages werden vom Zentrum jedes Brodatz-Bildes extrahiert. Die Features werden berechnet, indem das Modell aufgefordert wird, jedes der 1008 Subimages abzuschätzen. Die Distanz jedes einzelnen Test- Subimages zu allen anderen wird dann berechnet.

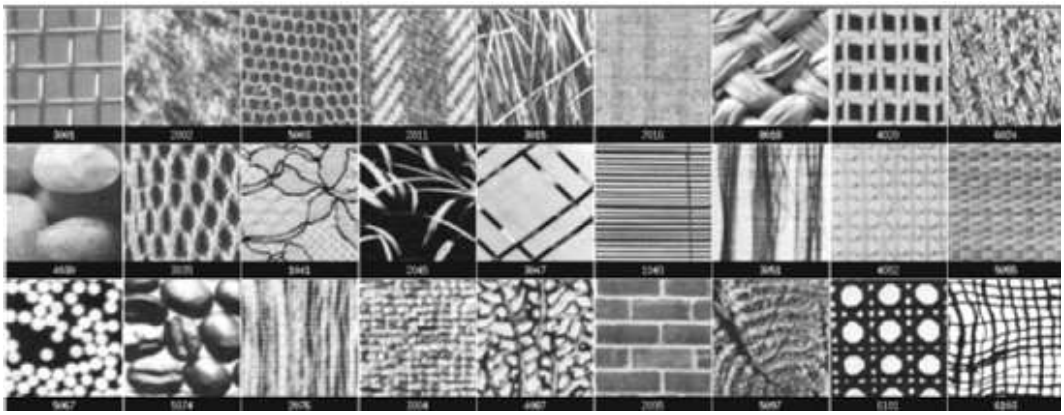
Anders gesagt: Zur Überprüfung des jeweiligen Algorithmus wird nun jedes der 112 Bilder (Klassen) in 9 kleinere Subbilder unterteilt. Man erhält nun in Summe 1008 Einzelbilder wobei jeweils 9 zusammengehören. Dabei wird auf 2 Arten die Qualität des jeweiligen Texturalgorithmus überprüft:

Retrieval Rate

Aus den 1008 Bildern wird nun „eines“ ausgewählt und als Referenz verwendet. Der jeweilige Algorithmus liefert nun eine gewisse Anzahl an Bildern zurück, die dem Referenzbild entsprechen sollen. Unter Retrieval Rate versteht man nun das Verhältnis wie viele der gefundenen Bilder sich einer Klasse (bestehend aus jeweils 9 Subbildern) zuordnen lassen. Beispiel: Werden 10 Bilder gefunden aber nur 7 dieser Bilder entsprechen einer bestimmten Klasse so spricht man von einer Retrieval Rate von 70%.

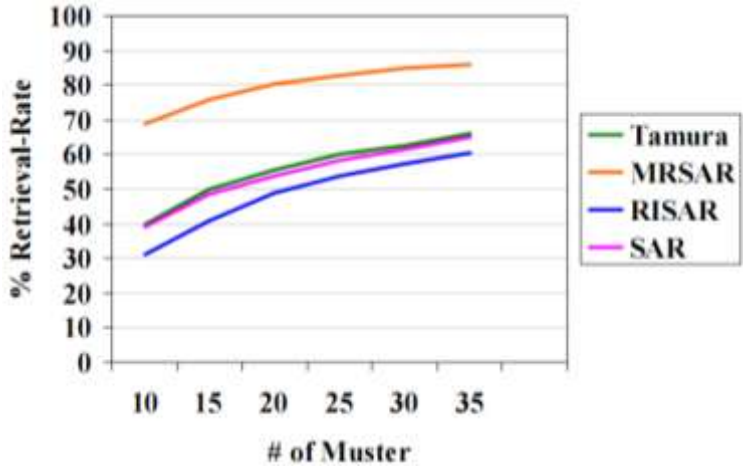
Average Retrieval Rate

Jedes Subimage wird als Test-Subimage berechnet. Also jedes der 1008 Bilder wird als Testbild (Test-Subbild) verwendet und das Ergebnis für verschiedene n-Werte dargestellt. Somit werden „alle“ 1008 Subbilder miteinander verglichen.



Die anderen Modelle im Vergleich

- Tamura:
  - **Vorteil:** Rotations invariant; Die Tamura-Features sind der menschlichen Wahrnehmung ähnlicher.
  - **Nachteil:** Hat die gleiche Genauigkeit, ist aber langsamer als SAR
- MRSAR:
  - **Vorteil:** hat die höchste Genauigkeit
  - **Nachteil:** berechnet niedrigere Auflösungen des Bildes, wenn die Imageregion zu klein ist.



Beschreiben Sie kurz die allgemeinen Kategorien, mit denen zeitabhängige Ströme unterschieden werden können. Ordnen Sie diese Kategorien MPEG-1 Video und CD-Audio zu. [3.5]

???

MPEG-7, MPEG-21 und Content Description

Erklären Sie kurz die Ziele von MPEG-7 [2]

- Weniger als Kompressionsformat gedacht
- Beschreibung von Multimedia-Inhalten durch Metadaten, um das Suchverfahren, Anzeigestrategien, aber auch Konsistenzprüfungen zu verbessern (z.B. Vision wäre es Videos nach Schlagwörtern zu suchen. Wenn man z.B. als Suchwort „gewaltfreie Szenen“ eingibt,

liefert eine Suchmaschine alle Videos, die dem Kriterium entsprechen, deswegen werden auch Metadaten benötigt)
- Flexibilität im Datenmanagement
- Globalisierung und Interoperabilität von Datenressourcen

MPEG-7 bestrebt folgende Punkte zu standardisieren

- Ein Set von Description Schemes und Descriptors um Daten zu beschreiben
- Eine Sprache um Desciption Schemes zu spezifizieren, wie die Description Definition Language (DDL)
- Ein Schema zur Kodierung der Description.

Erklären Sie kurz die Konzepte Deskriptor, Description Scheme und DDL im Kontext von MPEG-7. [3]

Eine Description besteht aus einem Description Scheme (Struktur) und einer Reihe von Description Werten (Umschreibungen), die die Daten beschreiben. Eine Description beinhaltet oder zeigt auf Eine vollständige oder teilweise umschriebene Description Scheme.

Deskriptor(D)

Ein Deskriptor ist eine Repräsentation eines Features (Feature ~ bestimmte Eigenschaft der Daten, die etwas für irgendjemanden bedeutet) und wird dazu verwendet die unterschiedlichen Eigenschaften (Features) von multimedialen Inhalten zu beschreiben. 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/ beschreiben, ebenso wie multiple Deskriptoren möglich sind (z.B. Histogramm). Ein Deskriptor kann wie folgt aussehen: <name, typekind, spec> value </name>

Beispiele:

- Color: string.
- RGB-Color: [integer, integer, integer]

Description Scheme (DS)

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

Description Definition Language (DDL)

Beschreibt die Syntax der Description Scheme. Description Definition Language ermöglicht Erstellung neuer Deskription Schemas und Deskriptoren. Außerdem können damit bestehende Deskription Schemas erweitert und modifiziert werden. DDL basiert auf XML, ist aber keine Modellierungssprache! Sie stellt den Kern von MPEG-7 dar.

Folgende Anforderungen werden abgedeckt

- **Compositional capabilities:** Möglichkeiten der Entwicklung: Es müssen neue Description-Schemata und Deskriptoren entwickelt werden können, aus vielen alten müssen neue generiert werden, die auch dann alle MPEG-7 kompatibel sein sollten.
- **Transformational capabilities:** Wiederverwendung, Erweiterung und Vererbung von existierenden DS und Deskriptoren.

- **Unique identification:** DDL bietet Mechanismen zur eindeutigen Identifizierung von D-Schemata und Deskriptoren für eindeutige Verweise.
- **Data types:** Datentypen müssen definiert sein: Primitives bereitstellen - Mechanismus um Deskriptoren mit verschiedenen Medientypen zu kombinieren: z.B. Text, integer, time/time index. Muss einen Mechanismus anbieten um Deskriptoren Daten mehrerer Medientypen mit eigener Struktur (audio, video, av-presentations,..) zuzuordnen oder auch histogram, graph, RGB, audio, audio-visual presentation,triggers,....
- **Relationships** innerhalb von DS und zwischen DS
- **Realtionship** zwischen Beschreibung und Daten.

Erklären Sie sehr kurz die Terminal Architektur von MPEG-7. [4]

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 geparts und die Content description wird rekonstruiert.

Video Retrieval - Visual Retrieval – Visuelles Information Retrieval

Beschreiben Sie kurz 3 Werkzeuge für die Inhaltsanzeige (Content Indication) von Video.[2]

Content Indication ist der Prozess von explizitem Präsentieren eines Teils des Inhalts für ein besseres Verständnis.

Video Icons

Icon, basierend auf einem Frame, der von einem Shot extrahiert wird, mit Pseudotiefe für die Repräsentation von Dauer, Richtungspfeilen und Zeichen fürdie Repräsentation von Objekt und Kamerabewegung.

Video map

Shot-Referenz, Intensitäten-Histogramm, Intensitäts-Durchschnitt, Histogramm-Differenzen, Video X-Ray (Ansicht von oben), Video X-Ray (Ansicht von der Seite), Farb-Histogramm

Keyframe-Extrahierung

- ***Vorteil*:** einfache Berechnung: wenn signifikante Inhaltsänderungen zwischen dem aktuellen Frame des Shots und dem letzten Keyframe auftreten, wird der aktuelle Frame als Keyframe ausgewählt – signifikante Merkmale sind Farbe, Textur und Bewegung
- ***Nachteil*:** nicht sehr repräsentativ

Video Panorama

Frames des Videos werden gestitched (also zu einem Panoramabild verschmolzen). In einem Bild ist z.B. Sprungschanze als Panoramabild vorhanden, in dem der Springer öfter auftaucht (je nach aktueller Position des Springers im verwendeten Frame).

Weiters gibt es

- Entstehung eines Bildes, das den globalen visuellen Inhalt des Shots repräsentiert
- Herausragende Standbilder
- Videospace icon (3D Video Icons)
- Videomap
- movie icon (micon): Video wird als 3D Gebilde dargestellt.
- Interactive micons: Auswertung und Bedienungsumgebung
- Paper video: auf Diagramme basierender Video Browser
- Videoscope: Analyseprogramm für Video- Content
- Sound Browser: zeigt das Vorhandensein des jeweiligen Sounds (Stimme, Musik,…)

Erklären Sie kurz in welchem Kontext die Erkennung von Kamerabewegungen eine Rolle spielt. Erklären Sie kurz die prinzipielle Arbeitsweise der Erkennung. [3.5] bzw. Erklären Sie kurz ein *Video X-Ray Interface*. Wie sind dabei Kamerabewegungen zu erkennen. [2.5]

Veränderungen, die mit Specialeffect- Übergängen assoziiert werden, müssen von Veränderungen durch Kamerabewegungen, unterschieden werden.

Veränderungen wegen Kamerabewegungen tendieren dazu, aufeinanderfolgende Differenzwerte von derselben Reihenfolge, wie die von schrittweisen Übergängen, hervorzurufen. Daher ist es notwendig die Muster von Bildbewegungen, die durch Kamerabewegung hervorgerufen werden, zu entdecken.

Es gibt zwei Zugänge

- Bewegungsvektor-Analyse (Motion Vector Analysis)
- (Video X-Ray), Visualisierung des 1. Verfahrens

Motion Vector Analysis (Bewegungsvektor-Analyse)

Die einzelnen Frames werden in Blöcke unterteilt, wobei jedem Block ein Bewegungsvektor zugewiesen wird. Daraus entsteht ein Vektorfeld, mit Hilfe dessen erkannt werden kann, ob es sich um

- Panning (Schwenk) oder tilting (Kippen): Alle Vektoren zeigen in Richtung des Modalvektors (mit geringen Abweichungen)
- zooming (in oder out): Summe der Vektoren ist 0.
    - Zoom in: Alle Vektoren zeigen vom Zentrum nach außen.
    - Zoom out: umgekehrt.
    - Spezialfall, wenn Zoom-Mittelpunkt außerhalb des Bild-Zentrums liegt ➔ mit Hilfe der Annahme, dass der Zoom-MP innerhalb des Bildes liegen muss gelöst
- um vordefinierte Bewegungsmuster handelt: z.B. Bewegungen von Menschen, Tieren, Objekten. Dazu kann eine entsprechende Beschreibung verwendet werden.

Ablauf

1. Frame i wird in Blöcke aufgeteilt
2. Block matching: Jedem Block wird ein Vektor zugewiesen, der den Block versetzt.
3. Optional flow: die Vektoren erzeugen ein Vektorfeld in dem jeder Vektor die Geschwindigkeit des entsprechenden Blocks auf die Bildfläche darstellst
4. Die Verteilung der Bewegungsvektoren in einem Frame kann
    a. Panning und tilting (Schwenken und Kippen der Kamera)

b. Zooming (in und out)
c. Und vordefinierte Bewegungsmuster

erkennen

Video X-Ray Interface

Annahme, dass der Hintergrund entsprechende Unterscheidungsmerkmale bietet.

- Panning (Schwenk) wird durch schräge Linien in der oberen Ansicht erkannt (unter Vorstellung eines des Würfels).
- Tilting (Kippen): wird durch schräge Linien in der seitlichen Ansicht erkannt.
- Zooming wird durch auseinander- oder zusammenlaufende Linien erkannt.

Erklären Sie kurz 2 Ansätze zur Erkennung der Kameraoption "Zoom" und nennen Sie Vor- und Nachteile der Ansätze. [3] bzw. Nennen Sie 2 Verfahren zur Erkennung von Kamerazooms. Erklären Sie kurz die Arbeitsweise eines dieser Verfahren. [2]

- Der Kamerafokus ist im Zentrum des Frames → alle Bewegungsvektoren zeigen entweder zu diesem Zentrum (zoom out) oder gehen von diesem Zentrum nach außen (zoom in) → Vektoren heben einander auf → der Zoom ist erkannt.
  - Vorteil: Zoom leicht zu erkennen → Summe der Vektoren = 0
  - Nachteil: Fokus muss im Frame-Zentrum liegen
- Der Fokus liegt wo anders (aber innerhalb des Frames) → der Bewegungsvektor in der obersten Reihe in der linkesten Spalte & der in der untersten Reihe und der rechtesten Spalte sind invers zueinander → Zoom-Operation.
  - Vorteil: flexibler bezüglich Fokus als 1. Ansatz
  - Nachteil: Aufwendiger

Wozu dienen Motionfeatures? Beschreiben Sie kurz 2 dieser Features. [2.5]

Mit Hilfe von Motionfeatures kann man Kamerabewegungen bzw. Objektbewegungen erkennen.

Bewegungsanalyse – Zooming

Annahme: der Fokus ist im Zentrum eines Frames und keine Objektbewegung findet statt.
Algorithmus: ist die Summe der Bewegungsvektoren gleich null → Zoom ist entdeckt

- Zoom In: alle Pfeile zeigen nach außen
- Zoom Out: alle Pfeile zeigen nach innen

Es ist gar nicht so einfach, das Zentrum eines Zooms zu finden. Es werden nur vertikale und horizontale Komponenten verglichen, wenn deren Betrag größer als ein bestimmter Schwellwert ist, dann wurde ein Zoom entdeckt.

Sicht von außen

Dieser Ansatz setzt nur voraus, dass das Zentrum des Fokus sich innerhalb der Grenzen eines Frames befindet.

Bewegungsanalyse – Objektbewegung

Bewegungsanalyse wird gebraucht um gewisse Aktionen, die innerhalb einer Szene stattfinden, zu definieren

- Menschliche oder tierische Bewegungen, wie laufen, springen, gehen, …
- Objektbewegung: sich bewegende Autos, LKWs, Flugzeuge, Hubschrauber, …

Eine Beschreibung kann dazu verwendet werden eine rohe Beschreibung der Szene zu liefern.

Beispiel

„Ein springendes Tier aus der Sicht einer sich bewegenden Kamera“ ist eine rohe Beschreibung eines Videoclips von einem Känguru, aus einem danebenherfahrenden Auto aufgenommen.

Was versteht man unter Block Matching? [2]

Es wird berechnet, wie sich die einzelnen Blöcke von Frame zu Frame verschoben haben.

- Verschieben sich die Blöcke einer einzelnen Region → Objekt bewegt sich;
- Verschiebt sich alles (einheitlich) → Kamera bewegt sich

Verwendest man z.B. um die Bewegungsvektoren zu berechnen.

Was versteht man unter Motion Continuity und wie wird sie berechnet? [3] bzw. Erklären Sie kurz Motion Smoothness. Wie kann Motion Smothness für die Schnitterkennung eingesetzt werden? [3]

Bewegung spielt in Video Material eine große Rolle. Motion Continuity beschreibt, wie „glatt“ eine Bewegung abläuft in dem es das Verhältnis von Bewegung zu Geschwindigkeit berechnet und ist somit ein Maß für die Motion Smoothness. Die Bilder werden in kleine Blöcke unterteilt. Die Bewegungsvektoren werden mittels block matching berechnet.

Es wird also berechnet, wie sich die einzelnen Blöcke von Frame zu Frame verschoben haben.

- Verschieben sich die Blöcke einer einzelnen Region → Objekt bewegt sich
- Verschiebt sich alles (einheitlich) → Kamera bewegt sich

Mathematische Berechnung

- Für jeden Block $\boldsymbol{b_i}$, $\boldsymbol{W_{1i}(b)}$ ist definiert als $\boldsymbol{W_{1i}(b)} = \begin{cases} \boldsymbol{1\ if\ |D_{bi}| > t_s} \\ \boldsymbol{0\ otherwise} \end{cases}$
- Ähnlich, $\boldsymbol{W_{2i}(b)}$ ist definiert als $\boldsymbol{W_{2i}(b)} = \begin{cases} \boldsymbol{1\ if\ |D_{bi+1} - D_{bi}| > t_m} \\ \boldsymbol{0\ ansonsten} \end{cases}$
- $\boldsymbol{t_s}$ und $\boldsymbol{t_m}$: kleine Schwellenwerte
- $\sum \boldsymbol{bW_{1i}(b)}$ - Anzahl signifikanter Bewegungsvektoren in Frame i
- $\sum \boldsymbol{bW_{2i}(b)}$ - Anzahl Bewegungsvektoren in Frame i welche signifikant von ihrem entsprechendem Vektor in Frame i+1 abweichen
- Die Smoothness $\boldsymbol{W_i}$ von Frame i: $\boldsymbol{W_i} = \frac{\sum \boldsymbol{bW_{1i}(b)}}{\sum \boldsymbol{bW_{2i}(b)}}$ wenn $\boldsymbol{W_i}$ < T =>Kameraschnitt

Erklärung

Berechnung Motion Vektor eines Blocks in einem Frame:
1 wenn ein Bewegungsvektor einen gewissen Treshold übersteigt.
0 wenn eben nicht.

Berechnung Differenz der einzelnen Motionvektoren zwischen zwei aufeinander folgenden Frames:
1 wenn Betrag der Differenz größer ist als ein gewisser Treshold.
0 wenn das nicht zutrifft.

Jetzt schaut man sich das Verhältnis der beiden Summen dieser Indikatorfunktionen an. Dprich man summiert alle Vektoren in einem Bild und teilt diese Summe durch die Summe der Differenzindikatorfunktion (also das erste dividiert durch das zweite).

Deutung des Ergebnisses

Je größer dieser Wert ist, desto smoother ist die Bewegung. Da der Teiler unten nur kleine Differenzen der Bewegungen aufweist (die beiden Frames unterscheiden sich nur geringfügig) und eine kleine Differenz natürlich auch eine smoothere Bewegung mit sich bringt.

Je kleiner der Wert, also je größer der Teiler (viele Differenzindikatoren), desto sprunghafter die Bewegung.

Das finden eines Kameraschnittes mithilfe der Motion Smothness

Man kann für diesen zuvor berechneten Quotienten auch einen Treshold festlegen und sagen: ist mein Wert kleiner als mein Treshold, also eine vermeidlich ziemlich ruckartige Bewegung im gesamten Frame, dann interpretiere diesen Frame als Kameraschnitt.

Beschreiben Sie kurz den *Multi-Pass Ansatz* zur Videoschnitterkennung. Weshalb verwendet man diesen Ansatz? [3]

Man will damit die Bearbeitungszeit reduzieren indem man ein Art Filter, welche Frames in Ordnung sind und welche weiter untersucht werden müssen, verwendet.

1. Durchgang

Tb wird niedriger angesetzt → Auflösung wird vorübergehend heruntergesetzt (weniger Frames pro Sekunde) → großer Sprungfaktor: steigert SD während schrittweisen Übergängen → Sowohl Schnitte als auch Übergänge werden erkannt (manchmal auch zu viele)

Weitere Durchgänge

Gesteigerte temporäre Auflösung wird verwendet um potentielle Segmentgrenzen zu erkennen

Worin besteht Ihrer Meinung nach das Hauptproblem bei der Videosegmentierung? [1.5]

Segmentierung findet ohne Berücksichtigung des Inhalts statt. Algorithmen bieten nicht die Möglichkeiten einer manuellen Segmentierung durch einen Menschen ➔ Qualität der Ergebnisse.

Erklären Sie kurz die Klassifikation von Shots für das Parsing von News Videos. [4]

1. Ein Modell eines A-shots (Anchorperson shot (Bei einer Konferenz (oder auch Debatte) gibt es immer einen Reporter, der die zentrale Rolle übernimmt. Dieser wird als Anchorperson bezeichnet)) wird definiert.
2. Entwicklung von Ähnlichkeitsmaßen, die im Vergleich dieser Modelle verwendet werden (die man mit diesen Vorlagen vergleichen kann)
3. Zeitliches Strukturmodell von einem gesamten Nachrichtenprogramm – Identifikation von Sequenz und Episoden.

Welche Schnitterkennungsverfahren für komprimierte Videos kennen sie? Beschreiben Sie ein Verfahren kurz. [3] bzw. Nennen Sie 3 Verfahren zur Schnitterkennung in Videos. Erklären Sie kurz die Arbeitsweise eines dieser Verfahren. [2.5] bzw. Welche Verfahren zur Schnitterkennung in MPEG-komprimierten Videos kennen Sie? Erklären Sie kurz die Arbeitsweise eines dieser Verfahren. [4]

Frame Differenz Techniken

- Pixelvergleich: Ein Pixel wird als unterschiedlich bewertet, wenn der Unterschied zwischen den Intensitätswerten in zwei aufeinanderfolgenden Frames einen bestimmten Schwellwert t übersteigt.
    - Schnitt wenn mehr als ein gegebener Prozentsatz der Gesamtzahl der Pixel sich verändert hat.

- Likelihood Ratio: Basiert auf statistischen Charakteristika 2. Ordnung der Intensitätswerte von einander entsprechenden Regionen (Blöcke) zweier aufeinanderfolgender Frames. Ein Kamera-Break kann festgestellt werden, wenn die Gesamtzahl der Sample-Gebiete, deren Likelihood-Ratio den Schwellwert übersteigt, ausreichend groß ist (hängt davon ab, wie der Frame aufgeteilt wird)
- Histogramm Techniken: Die Histogramm-Techniken reagieren weniger sensibel auf Objektbewegung als der Paarweise- Pixelvergleichs-Algorithmus.

Algorithmen, die auf DCT-Coeffizienten basieren

- Paarweise DCT Block Vergleiche
  - Vorteil: da nur ein kleiner Teil der Video Frames I-Frames sind, wird die Berechnungszeit reduziert.
  - Nachteil: der Verlust von zeitlicher Auflösung kann zu falschen Positiven führen
- DCT Coefficients Correlation

Algorithmen, die auf Bewegungsvektoren basieren

- Motion Vectoren im MPEG Datenfluss:
  - P- Frames: einzelne Gruppe von Motion Vectoren
  - B- Frames: zwei Gruppen von Motion Vectoren (vorwärts und rückwärts)
- Feld von Motion Vectoren in einem Video
  - Innerhalb einer Aufnahme- relativ kontinuierliche Änderungen
  - Zwischen verschiedenen Aufnahmen – Kontinuität wird gestört
- Definition M:
  - P- Frame: Zahl zulässiger Motion Vectoren
  - B- Frames: kleinere Zahl zulässiger Motion Vectoren (vorwärts und rückwärts)
- M< T => Kameraschnitt

Gemischter Ansatz zur Partitionierung

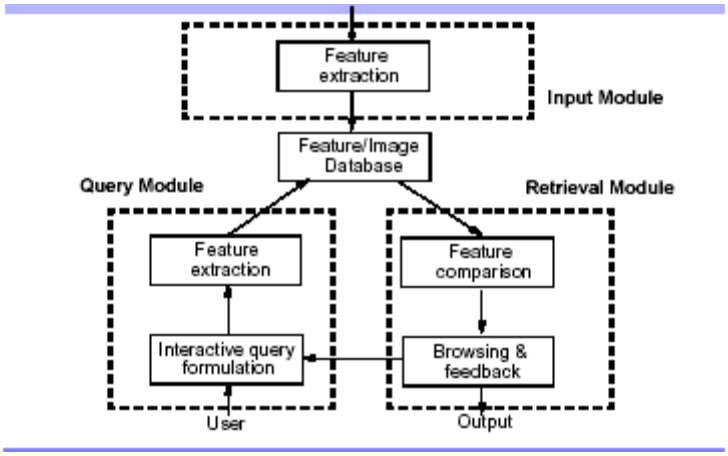
- 1. Schritt: DCT-Vergleich der I-frames mit einem großen Skip-Faktor um Regionen von potentiellen Schnitten, Übergängen, Kameraoperationen oder Objektbewegungen zu erkennen.
- 2. Schritt: DCT-Vergleich mit geringerem Skip-Faktor in der Umgebung von potentiellen Schnitten, Übergängen,... → eliminiert zuerst falsch erkannte.
- weitere Schritte: Bewegungsbasierte Vergleiche entweder auf dem gesamten Video oder auf den zuvor erkannten Sequenzen verifizieren die DCT-Resultate und verbessern die Genauigkeit.

Evaluierung

- Motion Vektor und hybrides Verfahren besser um Schnitte zu erkennen (DCT hat auch falsche Schnitte erkannt – da nur I-Frames verwendet wurden);
- bei Übergängen hat die MotionVektor-Methode allerdings versagt.

Beschreibung Sie kurz die Architektur eines Content-Based Image Retrieval-Systems [2] bzw. Beschreiben Sie kurz die Architektur eines Image Retrieval Systems. [2]

Ziel ist die Suche in einer Bilddatenbank nach entsprechenden Kriterien.



1. Input Mode

Beim Import der Bilder in die Datenbank sind die entsprechenden Features zu extrahieren und zusammen mit den Bildern in der Datenbank abzulegen.

2. Query Mode

Wird nach Bildern gesucht, hat der Benutzer die entsprechenden Anfragen zu formulieren (RBR, ROA, RSC…). Dies sollte im besten Fall interaktiv geschehen können. Aus diesen Anfragen werden die für die Suche benötigten Features in der Datenbank ermittelt und als Basis für die Suche verwendet.

3. Retrieval Mode

Die von der Benutzeranfrage ermittelten Features werden mit denjenigen der Bilder in der Datenbank verglichen. Dieser Vergleich liefert mitunter eine Menge von Ergebnisbildern, die der Anfrage entsprechen. Daher kann das Ergebnis im nächsten Schritt manuell durchsucht und entsprechend weiter eingeschränkt werden (➔ Neue Query). Vorgang solange, bis das gewünschte Ergebnis erreicht wurde.

Wozu dient die *Brodatz*-Datenbank? Beschreiben Sie kurz, wie diese Datenbank eingesetzt wird. [3.5] bzw. Erklären Sie kurz, wie verschiedene auf *Texturfeatures* beruhende Retrievalverfahren mit Hilfe der *Brodatzdatenbank* verglichen werden können. [3.5] bzw. Beschreiben Sie kurz die Evaluierung von Texture- Features mit Hilfe der Brodatz Datenbank.

Diese DB ist der de facto Standard für die Evaluierung von Textur- Algorithmen und gibt an wie gut unterschiedliche Algorithmen Texturen in Bildern erkennen können. Sie besteht aus einem Fotoalbum mit 112 Graustufen-Bildern (512x512x8 Bit), welche unterschiedlichen Texturklassen repräsentieren.

Texturauswahl- Algorithmus

9 Subimages werden vom Zentrum jedes Brodatz-Bildes extrahiert. Die Features werden berechnet, indem das Modell aufgefordert wird, jedes der 1008 Subimages abzuschätzen. Die Distanz jedes einzelnen Test- Subimages zu allen anderen wird dann berechnet.

Anders gesagt: Zur Überprüfung des jeweiligen Algorithmus wird nun jedes der 112 Bilder (Klassen) in 9 kleinere Subbilder unterteilt. Man erhält nun in Summe 1008 Einzelbilder wobei jeweils 9 zusammengehören. Dabei wird auf 2 Arten die Qualität des jeweiligen Texturalgorithmus überprüft:

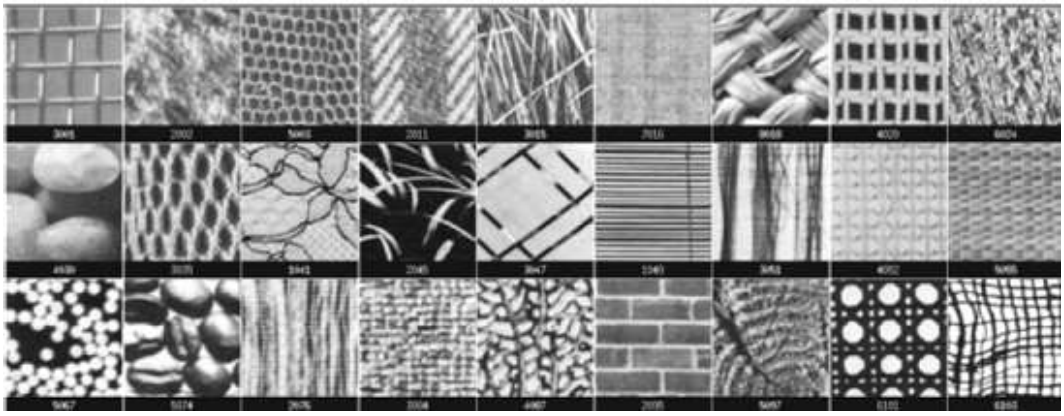
Retrieval Rate

Aus den 1008 Bildern wird nun „eines“ ausgewählt und als Referenz verwendet. Der jeweilige Algorithmus liefert nun eine gewisse Anzahl an Bildern zurück, die dem Referenzbild entsprechen

sollen. Unter Retrieval Rate versteht man nun das Verhältnis wie viele der gefundenen Bilder sich einer Klasse (bestehend aus jeweils 9 Subbildern) zuordnen lassen. Beispiel: Werden 10 Bilder gefunden aber nur 7 dieser Bilder entsprechen einer bestimmten Klasse so spricht man von einer Retrieval Rate von 70%.

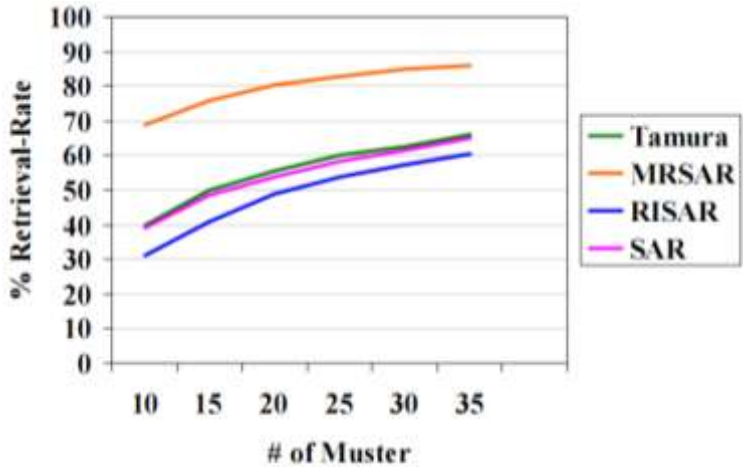
Average Retrieval Rate

Jedes Subimage wird als Test-Subimage berechnet. Also jedes der 1008 Bilder wird als Testbild (Test-Subbild) verwendet und das Ergebnis für verschiedene n-Werte dargestellt. Somit werden „alle“ 1008 Subbilder miteinander verglichen.



Die anderen Modelle im Vergleich

- Tamura:
  - **Vorteil:** Rotations invariant; Die Tamura-Features sind der menschlichen Wahrnehmung ähnlicher.
  - **Nachteil:** Hat die gleiche Genauigkeit, ist aber langsamer als SAR
- MRSAR:
  - **Vorteil:** hat die höchste Genauigkeit
  - **Nachteil:** berechnet niedrigere Auflösungen des Bildes, wenn die Imageregion zu klein ist.



Erklären Sie kurz die 2 wichtigsten *Performance* Evaluierungsmaße und ihren Zusammenhang. [2]

Performance Evaluation: Die normalüblichen Maße, um die Performance eines retrieval Systems abzuschätzen, sind:

Recall: Verhältnis der Anzahl der „gefundenen und relevanten Elemente“ zu der Anzahl der „relevanten Elemente“ in der Sammlung.

Precision: Verhältnis der Anzahl der „gefundenen und relevanten Elemente“ zu der Anzahl der „gefundenen Elemente“ in der Sammlung.

Verbesserung des Recalls verursacht geringere Precision und umgekehrt. User muss entscheiden, welches Maß vorzuziehen ist.

Multimedia Standards und Content Description

Erklären Sie kurz die Ziele von MPEG-7. [2]

Als schließlich Audio-Visuelle (AV) Daten in digitaler Form verfügbar wurden, war es kaum noch möglich, die immensen Datenmengen, deren Struktur und Inhalt, als sinnvoll verwertbare, wertvolle Information zu kapseln.

Ziel von MPEG-7 ist es nun, Multimediainhalte so zu beschreiben, dass für Mensch und Maschine der Informationsgehalt maximiert wird, sprich es werden intuitive beschreibungswerkzeuge zur Verfügung gestellt, die es Menschen und automatischen Systemen (Computern) erlauben, eine Datenbasis zu implementieren, die für einen qualitativen Zugang zu den Inhalten unumgänglich ist.

Durch einen gemeinsamen Standard für die inhalts- und Datenbeschreibung ist es nun möglich, audio-visuelle Daten weltweit auszutauschen und organisationsübergreifend zu arbeiten. Weiters können die Datenbeschreibungen flexibel eingesetzt werden, was einen generellen Mehrwert für die Verwaltung und das Management der Daten mit sich bringt.

Der gemeinsame Standard (mpeg-7) stützt sich hierbei auf

- Eine Menge von Deskription Scheme und Deskriptors, um die Struktur und den Inhalt der Daten zu beschreiben.
- Eine Sprache, die Description Definition Language, die es erlaubt neue Deskription Scheme zu erstellen oder sie zu modifizieren.
- Ein Schema für die Kodierung der Beschreibungen.

Ziele

- "Multimedia Content Description Interface"
- Abdeckung der Anforderungen von Menschen und automatischen Systemen
- Zurverfügungstellung eines Sets von audiovisuellen Beschreibungs-Tools als Basis für Applikationen und deren qualitativen Zugang zu Inhalten
- Standardisierte Beschreibung von Multimedia Content
- Flexibilität im Datenmanagement
- Globalisierung und Interoperabilität von Datenressourcen

Standardisierung

- eines Sets von Description Schemens und Deskriptoren zur Beschreibung der Daten
- einer Sprache, um die DS (Description Scheme) zu spezifizieren (DDL) - textuell und maschinenlesbar
- eines Schemas für die Codierung der Beschreibung (Description)

Was versteht man unter einer *MPEG-7 Description*? [1] bzw. Erklären sie kurz die Konzepte Deskriptor, Description Scheme und DDL im Kontext von MPEG-7.

Eine Description besteht aus einem Description Scheme (Struktur) und einer Reihe von Description Werten (Umschreibungen), die die Daten beschreiben. Eine Description beinhaltet oder zeigt auf Eine vollständige oder teilweise umschriebene Description Scheme.

Deskriptor(D)

Ein Deskriptor ist eine Repräsentation eines Features (Feature ~ bestimmte Eigenschaft der Daten, die etwas für irgendjemanden bedeutet) und wird dazu verwendet die unterschiedlichen Eigenschaften (Features) von multimedialen Inhalten zu beschreiben. 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/ beschreiben, ebenso wie multiple Deskriptoren möglich sind (z.B. Histogramm). Ein Deskriptor kann wie folgt aussehen: <name, typekind, spec> value </name>

Beispiele:

- Color: string.
- RGB-Color: [integer, integer, integer]

Description Scheme (DS)

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

Description Definition Language (DDL)

Beschreibt die Syntax der Description Scheme. Description Definition Language ermöglicht Erstellung neuer Deskription Schemas und Deskriptoren. Außerdem können damit bestehende Deskription Schemas erweitert und modifiziert werden. DDL basiert auf XML, ist aber keine Modellierungssprache! Sie stellt den Kern von MPEG-7 dar.

Nennen und erklären Sie kurz 3 Anforderungen an die *MPEG-7 DDL*. [2.5]

Description Definition Language erlaubt die Erstellung von neuen DS (Description Scheme) und Deskriptoren sowie die Erweiterung und Modifikation von bestehenden DS. Die DDL stellt den Kern von MPEG-7 dar.

Folgende Anforderungen werden abgedeckt

- **Compositional capabilities:** Möglichkeiten der Entwicklung: Es müssen neue Description-Schemata und Deskriptoren entwickelt werden können, aus vielen alten müssen neue generiert werden, die auch dann alle MPEG-7 kompatibel sein sollten.
- **Transformational capabilities:** Wiederverwendung, Erweiterung und Vererbung von existierenden DS und Deskriptoren.
- **Unique identification:** DDL bietet Mechanismen zur eindeutigen Identifizierung von D-Schemata und Deskriptoren für eindeutige Verweise.
- **Data types:** Datentypen müssen definiert sein: Primitives bereitstellen - Mechanismus um Deskriptoren mit verschiedenen Medientypen zu kombinieren: z.B. Text, integer, time/time index. Muss einen Mechanismus anbieten um Deskriptoren Daten mehrerer Medientypen mit eigener Struktur (audio, video, av-presentations,..) zuzuordnen oder auch histogram, graph, RGB, audio, audio-visual presentation,triggers,....

- **Relationships** innerhalb von DS und zwischen DS
- **Realtionship** zwischen Beschreibung und Daten.

Welche Konzepte stellt MPEG-7 für die Navigation zur Verfügung? Erklären Sie kurz 2 dieser Konzepte. [2.5]

MPEG-7 stellt DSs (Description Scheme) zu Verfügung, die Navigation und Zugang von audiovisuellem Inhalt unterstützen. Folgendes wird spezifiziert:

- Summaries – aktiviert effizientes Browsen, navigieren, entdecken, visualisieren…
- Views and Partitions – erlaubt multi-resolution und progressive access
- Variations – spezifiziert die Relationen zwischen verschiedenen AV-Inhalten, die eine Auswahl verschiedener Variationen von Inhalt unter verschiedenen Lieferbedingungen erlaubt.

Erklären Sie kurz das Konzept *Segment Description Scheme* und seine Anwendung. [3]

Das Segment DS ist das Kernelement (core element) der Content Description. Mit dem Segment DS werden zeitliche und räumliche Aspekte beschrieben. Es adressiert die Beschreibung von physikalischen und logischen Aspekten von audiovisuellem Inhalt. Das Segment DS wird verwendet um segment trees zu erstellen. Das Segment DS ist eine abstrakte Klasse und hat 5 Unterklassen.

Die 5 Unterklassen

- AudioVisual Segment DS
- Audio Segment DS
- Still Region DS
- Moving Region DS
- Video Segment DS.
- (Anmerkung: 5 Unterklassen laut Folien; laut anderen Quellen sind es 9!)

Beispiel

Ein zeitliches Segment kann eine Menge von Samples sein, die von dem Audio Segment DS beschrieben wird. Ein räumliches Segment kann ein Teil eines Bildes oder Frames in einer Videosequenz sein, das durch das Still Region DS beschrieben wird. Pro Segment können CreationInformation, Verwendung, Media Location und Text Anmerkungen angegeben werden.

Was versteht man unter *Variation Description Schemes* und wofür werden sie verwendet? [2.5]

Ermöglicht Servern eine Alternative basierend auf Möglichkeiten des Terminals, Netzwerkmodalitäten und Präferenzen der Benutzer auszuwählen. Das Variation DS spezifiziert die verschiedenen Variationen von audiovisuellem Inhalt (AV-Inhalt) wie Zusammenfassungen, komprimierte oder niedrig aufgelöste Version, verschiedene Sprachen und Modalitäten (Audio, Video, Bild, Text).

Eine der Hauptfunktionen vom Variation DS ist dem Server oder Proxy eine passende Variation des audiovisuellen Inhalts zu den Fähigkeiten der Endgeräte, Netzwerkbedingungen und Benutzervorlieben zu liefern. Genau diese Variationen werden vom Variation DS beschrieben. Die Variationen beziehen sich auf erst kürzlich gelieferten AV-Inhalt oder auf AV-Inhalt anderer Quellen.

Die variation fidelity value gibt die Qualität im Vergleich zum Original an. Das variation type attribute gibt die Art (Zusammenfassung, Auszug, Sprachübersetzung, Farbreduktion, Kompression…) der Veränderung an.

Infos

Nach Änderungen des Studienplans für Medieninformatik und Visual Computing wurde Multimedia 1 und Multimedia 2 zu einer neuen LVA namens Multimedia (VO) zusammengelegt.

Ich habe Martin Tintels Ausarbeitungen für Multimedia 1 und 2 nach bestem Wissen und Gewissen zusammengefügt. Weiters habe ich einige Sachen korrigiert und geändert. Es ist gut möglich und sehr wahrscheinlich, dass diese Ausarbeitung nicht vollständig und auch nicht fehlerfrei ist.

Neue Versionsnummer: 0.91.

Über die Ausarbeitung

Die Fragen stammen aus dem Informatik Forum, MTB, LVA Homepage, von Kollegen die die Prüfung machten bzw. aus einer Sammlung die ich mir über die Jahre anlegte.

Ich habe die Ausarbeitung so gut es geht gemacht, aber trotzdem können sich Fehler einschleichen! Falls man welche findet, bitte per E- Mail oder PM an mich weiter leiten damit ich sie ausbessere!

Die Antworten stammen teilweise aus Ausarbeitungen von anderen Studenten (die richtigen Namen standen nicht immer dabei. Also bitte einfach bei mir melden, damit ich die Namen dann unten dazu geben kann zu den anderen!) und Ausarbeitung von mir. Ich habe aber oft die Antworten aus diversen Ausarbeitungen, meine Antworten,… zusammengefasst, sodass man das ganze Wissen konzentriert in der Ausarbeitung hat. Den Rest der Fragen die unbeantwortet sind, versuche ich von Version zu Version weiter auszuarbeiten und zu beantworten.

Da ich es gut finde alle POs, Lösungen, Ausarbeitungen,… in einer Datei gesammelt zu haben, damit man nur die lernen braucht und auch keine Redundanzen hat, machte ich diese Ausarbeitung.

Bei einigen Sachen war ich mir nicht ganz sicher ob sie so stimmen. Deswegen habe ich diese Sätze rot geschrieben (zum leichten erkennen das man da „aufpassen“ sollte, da es womöglich nicht stimmt). Grün ist wenn es mehrere Antworten gibt und die noch zusammengefasst werden müssen, oder der Text auf Englisch ist.

Inkludierte Prüfungsangaben

5. Februar 2003, 20. März 2003, 6. Juni 2003, 2. Oktober 2003, 29. Januar 2004, 29. April 2004 29. Oktober 2004, 10. Dezember 2004, 25. Januar 2005, 7. April 2005, 23. Juni 2005, 20. Oktober 2005, 27. Januar 2006, 24. März 2006, 30. Mai 2006, 9. Oktober 2006, 26. Jänner 2007, 2. April 2008, 7. Mai 2008.

Falls jemand Angaben / Fragen hat die hier nicht zu finden sind, wäre es auch sehr nett sie mir zukommen zu lassen!

Zusätzliche Informationen

Version:	0.9
LVA Webseite:	http://www.ims.tuwien.ac.at/teaching_detail.php?ims_id=188135
Neuste Version:	http://stud4.tuwien.ac.at/~e0402913/uni.html
Ausarbeitung:	Martin Tintel (mtintel)

Ausarbeitungen auf die aufgebaut wird:	Marcel Nürnberg (multimedia%20i%20fragenkatalog.pdf) Elisabeth Wetzinger (mm1_po ausarbeitung.pdf) Martin Tintel (Pruefungordner 2004- 2005 leicht überarbeitet.doc) Antworten von Christoph R. aus dem Forum
Quellen:	http://de.wikipedia.org/wiki/YCbCr-Farbmodell http://www.slashcam.de/multi/Glossar/-Buchstabe--4.html#2 http://de.wikipedia.org/wiki/BIFS http://de.wikipedia.org/wiki/MPEG-2 http://de.wikipedia.org/wiki/JPEG