

Multimedia 2: Technologien Sommersemester 2009

Christian Breiteneder

Zusammenfassung

Foliensatz 1

Anforderungen von Mediaservern sind komplett anders als existierende Netzwerkserver:

- Echtzeitnatur von Multimediadaten erfordert Quality of Service Garantien (Garantien für periodischen Abruf und Übertragung von Daten)
- Solche Garantien können durch Zugangskontrolle und Ressourcenreservierung geboten werden
- Datenbereitstellungs- und Planalgorithmen müssen so designt sein dass Ressourcenreservierung über eine ganze Sitzung Bestand hat.

Eigenschaften der Abfrageeintreffen (Properties of Request Arrival)

Räumliche Lage (spatial locality)

- Besagt dass auf manche Dateien wahrscheinlich öfter zugegriffen wird als auf andere.
- Eintreffprozess kann mit Hilfe von Zipf's Gesetz modelliert werden
- Wird benutzt für Datenplatzierung in Speicherhierarchien verschiedener Speichertypen

Zeitliche Lage (temporal locality)

- Abfrageankunft kann zeitlich geclustered sein
- Wichtig für das Entscheiden des Service- und Netzwerkmodells des Servers

On-Demand Service Typen

Pay-Per-View (PPV): Server entscheidet über Zugriffszeiten; Server multicastet zu bestimmten Zeiten

Near Video-On-Demand (NVD) oder Shared Viewing with Constraints (SVC)

- Beobachtung: Abfrageeintreffen zeigen wahrscheinlich signifikante zeitliche oder räumliche Lage
- Der Server bearbeitet und akzeptiert Abfragen in Gruppen
- Neuer Client kann eine variable Zugriffslatenz haben
- Client wird Mitglied einer Multicast-Gruppe

True-Multimedia-On-Demand (TMOD) oder Dedicated Viewing

- Server behandelt jeden Zugriff unabhängig
- Paradigma für personalisierte, interaktive Multimedia Auslieferung
- Schwerwiegende Konsequenzen auf den Durchsatz

Netzwerkmodell

Speicherkosten sind der Hauptfaktor der Gesamtkosten eines MODS

3 Wege um Netzwerk- und Speicherkosten zu skalieren

- Hierarchische Lösungen
- Caching
- Teilen (Sharing)

Mediaserver Designprobleme

- Applikationsanforderungen
 - Zeitabhängigkeit
 - Offene Schnittstellen
 - Client, Netzwerk und Server Fähigkeiten
- Geschäftsentwicklungsanforderungen (business deployment requirements)
 - Skalierbarkeit
 - Zuverlässigkeit
 - Dynamische Anpassung an Auslastung (Arbeitsbelastung)
- Architektonische Anforderungen
 - Topologie
 - Kostenperformanz

Performanzmetrik – um Speicherserverarchitekturen zu vergleichen und zu klassifizieren

- Gleichzeitigkeit: Anzahl der Clients die unabhängig gleichzeitig auf eine Datei zugreifen
- Zugriffs- und Bearbeitungslatenz: sollte unabhängig von Serverauslastung < 1 sec sein
- Speicherkapazität und Speicher, Netzwerkdurchsatz pro Dollar
- Skalierbarkeit
- Erweiterbarkeit

Einfacher Media Server

- Applikationsserver (einer pro Applikation) empfängt Applikationsanweisungen vom Client und wandelt sie in Media Server Anweisungen um

- Kontrollserver
 - Zugriffskontrolle
 - Optimierung um Servereffizienz zu steigern
 - Verstecken der Konfigurationskomplexität vor dem Applikationsserver
- Datenserver: verantwortlich für tatsächliche Abfrage und Lieferung

Verteilter Media Server:

- Subkomponenten (Einfache Media Server) können unterschiedlich ausgerichtet werden
- Gefahr von ungleicher Lastaufteilung zwischen den einzelnen Servern
- Eine Auslastungs-Ausgleich-Policy braucht neue Hardware und erhöht die Komplexität der Software
- Großangelegte Server die aus kleineren Komponenten bestehen haben Vorteile dadurch: Stapelverarbeitung von mehreren Client Anforderungen für die gleichen Dateien oder Interval Caching.

Media Server Hardware

Geräteschnittstellen: fungieren als Datenpuffer und Geschwindigkeitsanpasser, unterstützen den Prozessor

RAID – Redundant Array of Inexpensive Disks (gute grafische Übersicht auf Folie 48f)

- 2 gegensätzliche Konzepte
 - Data striping für verbesserte Performance
 - Redundanz für verbesserte Zuverlässigkeit
- Raid Level 0
 - Keine Redundanz, nur Striping
 - Vorteile: Günstige Kosten, optimale Schreibperformance
 - Nachteile: geringe Zuverlässigkeit, nichtoptimale Leseperformance
- Raid Level 1
 - Spiegelung verwendet doppelt so viele Disks wie ein nichtredundantes Array.
 - Immer 2 Kopien der Information
 - Wird oft in Datenbankenapplikationen eingesetzt wo Verfügbarkeit und Übertragungsrate wichtiger als Speichereffizienz sind.
 - Vorteile: optimale Leseperformance, hohe Zuverlässigkeit
 - Nachteile: hohe Kosten, mäßige Schreibperformance
- Raid Level 0+1
 - Kombination von Spiegelung und Striping
 - Vorteile: optimale Schreib-/Lesegeschwindigkeit, Zuverlässigkeit
 - Nachteile: hohe Kosten
- Raid Level 2
 - Speicherähnliches ECC – Speichersysteme verwenden „billigere“ Hamming Codes anstatt Spiegelung für Wiederherstellung defekter Komponenten
 - Hamming Codes enthalten Paritätsinformationen für unterschiedliche sich überschneidende Teilsätze der Komponenten. Die Anzahl der redundanten Disks ist proportional zum Logarithmus der Anzahl der Disks.
 - Wenn eine einzige Komponente ausfällt werden mehrere redundante Disks gebraucht um die defekte Disk zu identifizieren, aber nur eine wird für das Wiederherstellen der verlorenen Information benötigt.
 - Korrektur bei Ausfall einer Platte, Hammingcode nur auf bestimmten Platten
 - Vorteile: Hamming Code ECC, optimale Lesegeschwindigkeit, einfacher Controller
 - Nachteile: geringe Schreibgeschwindigkeit, hoher Overhead bei kleinen Daten
- Raid Level 3
 - Bit-verschränkte Parität (Bit-Interleaved Parity)
 - Einzige Paritätsplatte – Controller kann einfach feststellen welche Disk defekt ist
 - Daten werden bitweise über die Datendisk verschränkt
 - Jeder Schreibzugriff greift auf alle Datendisks zu, jeder Schreibzugriff zusätzlich noch auf die Paritätsdisk
 - Nur eine Anfrage zu einer Zeit behandelbar
 - Wird oft in Anwendungen verwendet die eine hohe Bandbreite aber nicht so hohe I/O Raten brauchen
 - Einfacher zu implementieren als Levels 4, 5 und 6
 - 1 Paritätslaufwerk, Laufwerkköpfe synchronisiert
 - Vorteile: hohe Transferrate, Zuverlässigkeit, geringer Datenoverhead, einfache Implementierung (Vgl. 4, 5, 6)
 - Nachteile: geringe I/O-Rate
- Raid Level 4
 - 1 Paritätslaufwerk (Blockparität), Laufwerkköpfe nicht synchronisiert, blockweises Striping
 - Vorteile: hohe Leseperformance, hohe Transaktionsrate bei Leseaufgaben, geringer Datenoverhead
 - Nachteile: komplexer Controller, schlechte Performance bei Schreibaufgaben, Zuverlässigkeit (Rekonstruktion bei Plattenausfall ineffizient)
- Raid Level 5
 - ECC-Blockwerte verteilt über alle Platten
 - Vorteile: hohe Schreib-/Leseperformance, hohe Transaktions- und Transferrate, gleichmäßige mechanische Belastung

- Nachteile: Controller Komplexität, schwierige Datenrekonstruktion
- Raid Level 6
 - Datenrekonstruktion bei Ausfall 2er Platten, 2 Paritäten pro Platte (sonst wie Level 5)
 - Vorteile: Zuverlässigkeit
 - Nachteile: schlechte Schreibperformance, schwierige Datenrekonstruktion, Komplexer Controller, Overhead

Media Server Architekturen

- Großangelegte (large-scale) Server – Problem?
 - Speicher und Netzwerk I/O Flaschenhals
- Architekturen großangelegter Server
 - Verteilte speicherbasierte geclusterte MM Server
 - Shared Memory MIMD Multiprozessor Maschinen
 - Massive parallele SIMD Maschinen

QoS Spezifikationen

- QoS Anforderungen können explizit von der Applikation spezifiziert werden (z.B. für Videokonferenzen mit den Geräteinformationen)
- Oder implizit spezifiziert in einem assoziierten Metafile, ein einem spezifischen Teil des Files oder als assoziierte Attribute (z.B. VOD mit dem Medienfile)
- Die flexibelste Implementierung ist die implizite die von Applikationen überschrieben werden darf.

Explizite vs. Implizite QoS Spezifikationen

- Explizit
 - Einfachere Implementierung
 - Erlaubt Applikationen das Überschreiben, nützlich für dynamische Anpassung der QoS Anforderungen
- Implizit
 - Einfacherer Prozess der Applikationsentwicklung
 - Bietet einen Migrationsspfad für nicht-MM Applikationen weil der Server QoS Support anbieten kann
 - Speicherung von assoziierten Attributen kann notwendig sein.

QoS Parameter

- Bandbreitanforderungen – Bereitstellung spezifischer Datenraten (muss nicht konstant sein, durchschnittliche Rate)
- Spitzenbandbreitanforderungen – maximale Dauer der Spitzenbandbreite
- Burst size oder workahead – maximale Datenmenge die ein Stream bei konstanter Rate übersteigen (exceed?) kann
- Delay – maximale Verzögerung
- Verlustwahrscheinlichkeit – maximale tolerierbare Verlustwahrscheinlichkeit (Kontrollinformation empfindlicher)

Kapazitätsschätzung

- Statistische Zugangskontrolle – Kontrolle des Zugangs der Benutzeranfrage um dem kombinierten QoS die Anforderungen von allen gleichzeitigen Anfragen zu garantieren
- Naive Methode: Ressourcenmanager enthält eine Tabelle mit der bewerteten Bandbreite aller Komponenten
- Kalibrierung – Schätzung der Bandbreitenkapazität der Serverkomponenten durch Messung

Serverkalibrierung

- Der Kalibrierungsprozess nimmt als Eingang die Konfiguration des Videoservers und läuft einen Benchmark den der Server unterstützt ohne die QoS Anforderungen zu verletzen
- Der Kalibrierungsprozess kann möglicherweise nicht zwischen den Kapazitäten unterschiedlicher Komponenten unterscheiden (z.B. solche in einer Pipeline Konfiguration)
- Solche Komponenten können in logische Komponenten vereinigt werden

Serverkalibrierungsprozess

1. Komponenten des Systems als Komponentengraph repräsentieren
2. Vereinige Komponenten deren Kapazitäten nicht durch Messung unterschieden werden können; das sind Komponenten für welche das Verhältnis des Datenflusses dazwischen fix ist; 2 Hauptfälle
 - a. Daten fließt entweder durch beide Komponenten oder durch keine von beiden (z.B. Diskadapter und Disk)
 - b. Semantik der Konfiguration (z.B. Disk Striping)
3. Betrachte ein Knotenset des Graphen als Anfang (z.B. reduzierte Knoten die Speichergeräte beinhalten) und ein anderes als Ziel (z.B. Netzwerkadapter); für jedes Paar zähle alle Pfade die ein ausgewähltes Anfangs- und ein Zielpaar verbinden.
4. Berechne das Gewicht jedes Pfades als den Bruchteil des gesamten Datenflusses zwischen den Paaren
5. Um die Kapazität der logischen Komponenten zu messen spiele durch Wählen passender Paare und Kontrollieren deren Datenrate systematisch Datenströme ab.
6. Die Bandbreitenkapazität jedes logischen Knotens wird geschätzt durch die Sättigung
 - a. Sortiere die Pfade die durch den Knoten gehen anhand ihrer Gewichte
 - b. Wähle den Pfad mit dem höchsten Gewicht und verfolge ihn mit maximalem Fluss (Sättigung möglich durch Flaschenhals in einem anderen Knoten)
 - c. Zusätzlicher Fluss im gewählten Knoten kann erzeugt werden indem man die restlichen Pfade anhand ihrer Gewichte auswählt und jeden Pfad bis zu einem maximalen Fluss folgt bevor man den nächsten Pfad auswählt.
7. Die resultierende Bandbreitenkapazität des gewählten Knotens ist die Summe aller gleichzeitigen Flüsse in allen Pfaden.

Einrichten eines logischen Kanals

- Der Ressourcenmanager sieht den Mediaserver als ein verbundenes Set von logischen Komponenten
- Um einen logischen Kanal für eine Anfrage zu erzeugen müssen die logischen Komponenten ausgewählt und für Ressourcen reserviert werden.
- Das Wählen von Komponenten für das Bedienen einer Anfrage ist nicht trivial: verschiedene Komponentensets können die Anfrage behandeln.

Komponentenauswahl

- Kriterien für die Komponentenauswahl
 - Optimierung
 - Komponentencharakteristika
 - Servertopologie – Komponentenauswahl kann auf der Komponentenverbindung basieren
- Auch nach Anwendung mehrerer Kriterien können verschiedene Pfade mit verschiedenen Komponentensets existieren
- Finale Auswahl: Lastausgleich; verschiedene Strategien, z.B. pfadbasierende Komponentenauswahlpolicy

Clientanforderungsplanung

- Planungsprobleme
 - VCR Kontrolloperationen – ein neuer logischer Kanal ist erforderlich für die Wiederaufnahme; andere Operationen (Vorspulen) haben auch unvorhersehbaren Bedarf an Ressourcenanforderungen
 - Contingency channel policy (Unvorhergesehener Fall Kanal Policy) – behandelt solche Anfragen; legt ein paar statistisch geteilte contingency channels für unvorhersehbaren Bedarf beiseite
 - Algorithmen für Reoptimierung oder Erleichterung des höheren Bedarfs für eine Ressource durch Benutzung eines anderen verfügbaren Typs
 - Gemeinsame Datenströme – teilen gesamten oder Teil eines logischen Kanals auf mehreren Clients (Batching policies, multicast)
 - Datencenteransatz
 - Serverkapazität proportional zu der Anzahl an Mediafiles
 - Intervallcaching
 - Zeitvariierende Arbeitslast

Clientverhalten

- Minimale Nichteinhaltungszeit / Maximale Wartezeitgarantie (Bild Folie 71)

Kanalplanung

- Minimierung langzeitiger Nichteinhaltungswahrscheinlichkeit
- Minimierung kurzzeitiger Spitzennichteinhaltungswahrscheinlichkeit
- Minimierung der durchschnittlichen Wartezeit
- Fairness eine Planungspolitik ist fair wenn die Nichteinhaltungswahrscheinlichkeit für alle Videoanfragen gleich ist
- Wiederaufnahmeverzögerung

VCR Kontrolloperationen

- Pause und Wiederaufnahme Implementierung ist einfach; Hauptkomplexität: genug Bandbreite für Wiederaufnahme
- Andere Operationen (Vor- und zurückspulen)
 - Spezielle Dateien: Bandbreitenanforderungen verschieden; diskretes Geschwindigkeitsset, Speicheroverhead
- Schnelles Abspielen: erhöhte Clientimplementierungskomplexität; von momentanen Decodingsstandards nicht unterstützt
- Scan – gewählte frames werden vom Server zum Client transferiert

Contingency Channel Policy Performance

- Fallstudie: nimmt die Arbeitslast von Folie 71 und ein System mit 100 Videos. Die Ankunftsrate liegt bei 50 Requests/min. Der Server hat einen Durchschnitt von 6000 aktiven Kunden. Es wird angenommen dass die Kunden einmal per Video pausieren und die durchschnittliche Pause wird mit 15 min angesetzt.
- Nichteinhaltungswahrscheinlichkeit kleiner als 5%
- 95% der VCR Kontrollanfragen sollen unter 30 Sekunden erfüllt werden.

Batching Policies

- Einfaches NVOD
 - Maximale Zeit die ein Kunde zu warten hat ist $1/3$ der Spielzeit ($N_{\text{objects}} * T_{\text{play}} / N_{\text{logical.server.chan}}$) (bei 3 Channels wobei der Film in 3 Stücke zerteilt ist und jeder Channel immer gerade ein anderes Stück abspielt)
- Pyramid Broadcasting
 - Die Serverbandbreite wird in $N_{\text{pyr.ch}}$ logische Kanäle geteilt (z.B. 2 logische Kanäle, jeder mit der 3fach benötigten Bandbreite)
 - Jedes Video wird wiederum in $N_{\text{pyr.ch}}$ Segmente unterteilt deren Größen um $\alpha_{\text{pyr.ch}}$ erhöht werden
 - Wenn $N_{\text{pyr.ch}}$ erhöht wird verringert sich die Größe des ersten Segments
 - Das i-te Segment von allen Videos wird wiederholt auf Kanal i gebroadcastet
 - Um Video i abzuspielen startet der Client den Download und das Abspielen von O_{i1} sobald verfügbar
 - Das Schema erfordert genug lokalen Speicherplatz in der Ordnung der Größe des größten Segments
 - Der Wert von $\alpha_{\text{pyr.ch}}$ für ununterbrochene Wiedergabe muss es möglich sein den Download des nächsten Segments zu starten bevor das aktuelle zu Ende ist. Maximaler Wert von $\alpha_{\text{pyr.ch}}: N_{\text{logical.server.chan}} / (N_{\text{objects}} * N_{\text{pyr.chan}})$

- Dynamic Batching Policies
 - NVOB policies brauchen Wissen über die Zugriffsmuster der Videoobjekte
 - Um mit Unsicherheiten klarzukommen gibt es batching policies die dynamisch populäre Videos entdecken und mehrere Benutzer in einem einzelnen Stream bedienen.
 - 3 verschiedene Arten
 - FCFS (First Come First Serve) policy: eine Anfragewarteschlange; time-of-service Garantie, fair
 - MQL (Maximum Queuing Length) policy: greedy policy, jedes Video hat eine eigene Warteschlange, hohe Nichteinhaltungsrate für „kalte Videos“
 - GGSC-FCFS (Group Guaranteed Server Capacity – First Come First Serve) policy: versucht die Durchschnittswartezeit für Anfragen zu verringern
 - Auswirkung von dynamic batching policies
 - Die 3 policies unterscheiden sich bezüglich ihrer Fähigkeit quantitative Maße wie Kunden Nichteinhaltung und Unfairness zu beeinflussen
 - FCFS und GGSC-FCFS können time-of-service Garantien bieten, MQL nicht
 - In GGSC-FCFS kann die Wartezeit berechnet werden (Wiedergabezeit, Anzahl der Kanäle)
 - FCFS vs. MQL: FCFS hat eine geringere Nichteinhaltungswahrscheinlichkeit als MQL
 - Beispiele (Folien 84f): Annahme eines minimalen Nichteinhaltungszeitmodells, wo Zuschauer mindestens $T_{ren.min}$ warten, gefolgt von einer exponentiell verteilten Zeitspanne mit 5 min Durchschnitt; Benutzeranfragen werden als Poissonankunftsprozess mit Ankunftsrate λ modelliert; Filmlänge ist 2h; λ variiert zwischen 2 und 25 pro Minute (300-3000 Benutzer)
 - Für kleine Serverkapazitäten hat MQL eine niedrigere Kurzzeitannteinhaltungswahrscheinlichkeit.

Scheduling in Systemkomponenten

- CPU-Scheduling für kontinuierliche Medien unterscheidet sich vom herkömmlichen Scheduling, weil kontinuierliche Auslieferung periodische Anfragen mit „harten deadlines“ erfordert
- Klassische Papers über Scheduling von periodischen Tasks studierten Tasks mit den folgenden Charakteristika:
 - Der einzige Task mit harter Echtzeit-Deadline ist ein periodischer Task
 - Die Deadline für einen Task ist das nächste Mal wenn eine Anfrage für den Task erzeugt wird
 - Die Tasks sind unabhängig voneinander
 - Die Ausführungszeit eines Tasks ist konstant und hängt nicht ab von der Zeit oder einer Ausführungssequenz
 - Nichtperiodische Tasks haben keine Deadlines
 - Tasks können in Beschlag genommen werden bevor sie fertig sind.
- Diese Charakteristika passen besser zu CPU-Scheduling als anderen Komponenten
- Solche Tasks können mit Hilfe einer static priority policy oder einer dynamic priority policy geplant werden.
 - Rate monotonic scheduling – static: assoziiert 1 eines fixen Sets von Prioritätsleveln zu jedem Task; Tasks mit größerer Frequenz bekommen eine höhere Priorität; Tasks werden in Prioritätsreihenfolge ausgeführt.
 - Deadline scheduling – dynamic priority-based policy: Deadline für den Task ist die Zeit wenn die nächste Anfrage für diesen Task erzeugt wird. Die Policy setzt die Priorität jedes Tasks auf seine Deadline. Der Task mit der niedrigsten Priorität (Deadline) wird zuerst ausgeführt.
- Beide Policies nehmen an dass nichtperiodische Anforderungen im Background ausgeführt werden
 - Kann in unnötig langen Antwortzeiten resultieren
 - Keine Garantie für diese Anfragen möglich

Hierarchisches Scheduling

- Der Prozessor wird unter Taskgruppen partitioniert; jede Gruppe unterteilt den Prozessor in Subgruppen oder führt Tasks in der Gruppe aus.
- Dies erlaubt jeder Applikation eine passende Scheduling policy zu implementieren
- Prinzipiell sind lower-level policies die zwischen Tasks scheitern applikationsabhängig.
- Eine wichtige Anforderung ist die Isolation verschiedener Untergruppen voneinander (Partitionierung unabhängig der Arbeitsauslastung)

Foliensatz 2

Nachtrag Zipf's Gesetz

- George Kinsley Zipf (amerikanischer Sprachwissenschaftler)
- Untersucht statistische Verteilung von Wörtern in verschiedenen Sprachen
- In jeder Sprache gibt es Wörter die oft und andere die weniger oft genutzt werden.
- Ranking der Wörter einer Sprache nach Häufigkeit
- Beobachtung: Verteilung in verschiedenen Sprachen fast identisch
- N Anzahl der Elemente
- k Rang
- s Verteilungsparameter
- $f(k; (s, N)) = \frac{1/k^s}{\sum_i 1/i^s}$

Fortsetzung Hierarchical Scheduling

- Ideale Policy: perfect fair-share policy die unterschiedliche Gruppen nach dem Round-Robin Scheduling bedient und nur ein unendlich kleines Quantum an CPU-Zeit nimmt (praktisch unrealisierbar)
- Abweichung von diesen Maßen können für Vergleiche verwendet werden
- Erreichbar: Approximationen der fair-share policy
 - Weighted fair queuing (WFQ); Nachteile: komplexe Berechnung und Information über die Ausführungszeiten der Tasks
 - Start-time fair queuing

Das Speichersubsystem

- Block Placement
 - Suchoverhead ist einer der Hauptoverheads während der Diskblock-Bereitstellung
 - Traditionelle block placement policies – Benutzung der Information aufgrund von Dateizugriffsmustern
 - Contiguous allocation (zusammenhängend Zuteilung)
 - Organ-pipe placement
 - Constrained-placement policies
 - Alle Diskblöcke für eine Datei werden zusammenhängend auf der Disk platziert; eliminiert dateiinternes (intrafile seek) Suchen; kein interfile seeks und auch keine Suche aufgrund von zufälligem Zugriff
 - Sehr gut für read-only Medien geeignet
- Contiguous Allocation
 - Unter Lese/Schreib-Arbeitslast hat die Policy mit der Fragmentation des Speichers zu tun
 - Verdichten der Mediendaten
 - Zeitraubend; Datei kann unerreichbar sein
 - Großer Overhead
 - Belegungsraum als Vielfaches der Blockgröße für I/O Operationen
- Everest Contiguous Allocation
 - Adaption der Contiguous Policy um den Verdichtungs-overhead zu minimieren
 - Blöcke werden Daten nur in zusammenhängenden Segmenten von ω^i zugewiesen
 - Freie Blöcke werden in freien Segmenten organisiert die in „freien Segmenten Listen“ verwaltet werden (eine Liste für jede mögliche Segmentgröße ω^i)
 - Die Policy verwaltet ein Maximum von $\omega-1$ Segmenten in einer Liste
 - Wenn das Löschen einer Datei dazu führt dass mehr als $\omega-1$ freie Segmente der Größe ω^i entstehen, werden belegte Segmente migriert bis mindestens ω zusammenhängende freie Segmente der Größe ω^i übrig sind.
 - Diese Segmente werden dann in ein Segment der Größe ω^{i+1} zusammengeführt.
- Organ-Pipe Placement
 - Versucht den interfile-Suchoverhead zu reduzieren
 - Studien zeigen dass die optimale Platzierung von Blöcken einem Orgelpfeifen-Prinzip folgt wobei die am öftesten verwendeten Blöcke in der Mitte der Disk sind.
 - Seltener verwendete Blöcke werden abwechselnd um die heißesten Blöcke platziert
- Constrained-Placement Policies
 - Diskutierte Policies sind Adaptierungen von konventionellen Blockplatzierungs-policies
 - Policies ziehen Vorteil aus dem sequentiellen Zugriff auf Medienströme um Suchoverhead zu beschränken
 - REBECA (Region based block allocation method): versucht den Suchoverhead zugunsten der Startverzögerung für neue Streams“ herunterzuhandeln“
 - Strand-based allocation
 - REBECA
 - Speichergerät ist in eine fixe Anzahl von zusammenhängenden Regionen unterteilt
 - Zugriffe auf viele Blöcke innerhalb einer Region werden von einem Scanschritt bedient
 - Mehr Regionen verringern den Suchoverhead aber erhöhen die Anfangslatenz
 - Aufeinander folgende Blöcke eines Medienobjekts s werden entsprechend der Reihenfolge der Bewegungen auf der Platte gespeichert.

Retrieval

- Von Applikationen benötigte Daten müssen pünktlich ohne Verzögerungen erlangt werden
- Disk Scheduling Policies erreichen das durch Planen der Anfragen für Blöcke so dass Streams die Blöcke bekommen bevor sie gebraucht werden
- Zwei Kategorien
 - Round-based policies – I/O Anfragen werden nur geplant durch periodische Planungsrunden, Annahme dass die Wiedergaberate konstant ist
 - Fixed block size policies
- Round-robin
- Scan policies: zuerst outward und dann inward
- Group Sweeping Scheduling (GSS)

- Kombiniert Scan und Round-Robin, resultiert in einer Policy mit geringem Suchoverhead (scan) und mit Bufferanforderungen einer Runde
- Policy
 1. Teile die V aktiven Ströme in M_{gr} Gruppen
 2. Bediene die M_{gr} Gruppen in round-robin Reihenfolge
 3. Bediene die Ströme in jeder Gruppe mit der Scan Policy

File Placement

- Operationsprobleme
 - Effiziente Nutzung von Speichergeräten
 - Platzierung basiert auf erwarteter Arbeitslast
 - Erwartete Arbeitslast kann sich ändern (periodisches Update, Online-Rekonfiguration)
 - Die gesamte Platzierung der Medienobjekte kann in Koordination mit der Placement policy gemacht werden.
 - Placement policy kann in die Policies für die Verwaltung der Speicherhierarchie integriert werden.
- Policies
 - Popularity-based assignment – wird verwendet um Medienobjekte auf einem Set identischer Disks zu speichern
 - Policy versucht die durchschnittliche Anzahl der Kunden die wegen Überlastung der Speichergeräte rausgeworfen werden zu verringern.
 - Policy verwendet Zugriffswahrscheinlichkeiten für jede Datei i an einem Tag t (Popularität)
 - Annahme dass Zugriffswahrscheinlichkeiten über 24h gleich sind
 - Abschätzung von Wahrscheinlichkeiten durch lineare Interpolation und tägliche Rekonfiguration
- BSR (bandwidth-to-space) policy
 - Versucht heiße und kalte und große und kleine Medienobjekte in heterogene Speichergeräte zu mischen (Striping Groups)
 - Policy charakterisiert jedes Gerät anhand seines Bandbreiten-zu-Raum Verhältnisses
 - Jedes Medienobjekt wird durch die erwartete Bandbreitenanforderung an die Kopie und den erforderlichen Speicherplatz charakterisiert
 - Policy erstellt und löscht dynamisch Kopien
- Strand-Based Allocation
 - Strand (Litze) – eine Sequenz von kontinuierlich aufgenommenen Audio samples oder Video frames
 - Mediablock – Sequenz von aufeinanderfolgenden Sektoren die Medieneinheiten speichern; Größe eines Mediablocks vom Strand s_i wird M_i gekennzeichnet
 - Media gap – Die Unterteilung zwischen aufeinanderfolgenden Blöcken eines Strands; Größe eines Media gaps von s_i wird mit G_i gekennzeichnet.
 - Jeder Medienblock M_i und gap G_i umfasst eine Anzahl an Sektoren; (M_i, G_i) bezieht sich auf ein Speichermuster von strand s_i
- Storage Patterns (genaue Instruktionen Folien 122ff)
 - Kontinuität in der Bereitstellung kann garantiert werden wenn jeder Medienblock am Abspielgerät zur oder vor der geplanten Abspielzeit vorhanden ist
 - Zufällige vs. Zusammenhängende Position; kontinuierliche Wiedergabe ist nicht möglich vs. Kopieroverhead; deswegen
 - Blockgröße von unten beschränkt
 - Gap-Größe von oben beschränkt
 - Merging two Strands
 - Um den Speicher effizient zu nutzen müssen Blöcke von s_2 in den Gaps des Musters von s_1 gespeichert werden, die Kontinuität darf nicht verletzt werden.
 - Merge Condition
 - Ein Strand s_2 kann man mit s_1 zusammenlegen wenn der Anteil von belegtem Medienplatz von s_2 nicht den Anteil von leerem Platz in s_1 überschreitet
 - Layout of Media Blocks
 - Wenn das Muster von s_2 spärlich im Vergleich zu den Gaps in s_1 ist, werden Medienblöcke von s_2 früher gelesen als sie angezeigt werden was zu Spitzen in den Pufferanforderungen führt

Multimedia Information Caching

- Die Entwicklung und der Einsatz von großangelegten Medienservern werden behindert durch das Fehlen von Server- und Netzwerkbandbreite
- Caching von Dokumenten im lokalen Speicher kann die Kosten der erforderlichen Retrieval Bandbreite senken
- Z.B. speichern von oft benutzten Dokumenten im Client oder dazwischenliegenden Distributionsserverknoten
- Verschiedene Anforderungen von verschiedenen Applikationen (MOD vs mddb)
- Zusammengesetzte Dokumente aus vielen kleinen Mediensegmenten resultieren in Weite bzw. Fluktuationen
- Bandbreitenanforderungsprofil kann geglättet werden indem man die angeforderten Datenblöcke im Voraus geholt werden

- Das Zugriffsmuster innerhalb eines Multimediadokuments ist möglicherweise nicht gleich und Wiedergabesequenzen müssen nicht immer linear sein -> Caching und Prefetching der betroffenen Datenblöcke
- Datenblöcke die kontinuierlich empfangen werden heißen „continuous object presentation unit (COPU)“

Document Access Patterns (Folie 135)

- COPU wird übersprungen
- COPU wird vorgespult
- COPU wird rückwärts abgespielt

Prefetching, Buffering, Caching

- Buffering und Caching betreffen den Einsatz von Primärspeicher um Verzögerungen und Overhead beim Zugriff auf den Sekundärspeicher zu vermeiden
- Unterscheidung zwischen Buffering und Caching liegt in den Performancezielen, Applikationsanforderungen und Primärspeicherverwendung
 - Buffering wird verwendet um Zugriffsverzögerungen zu vermeiden
 - Caching wird verwendet um Zugriffsoverhead und –verzögerungen zu vermeiden
- Caching Ziele
 - Erhöhen der Serverkapazität – Speichern aller oder Teile häufig verwendeter Medienobjekte im Serverspeicher
 - Reduzierung der Zugriffslatenz – Zugriffszeit ändert sich mit dem Ort wo sich Daten in der Speicherhierarchie befinden, z.B. können Daten die sich im Speicher befinden direkt an die Clients geliefert werden
 - VOD, VCR control operations
 - Reduzierung der Netzwerkbandbreitenanforderungen – ein großangelegter Server besteht aus einem lokalen Server (liefert Daten an Clients) und Remote Speicherservern (speichern Originalkopien der Daten); Lokale Server cachen Daten auf lokalen Disks und im Speicher um Kommunikationsoverhead zu vermeiden
 - Verteilen der Auslastung auf Speichergeräte – abhängig von der ursprünglichen Platzierung der Daten und der Popularität kann der Flaschenhals variieren
 - Durch einen shared Cache auf dem gemeinsamen Übertragungsweg der selektiv Daten von den Flaschenhals-Servern cacht kann die Last besser über alle Systemressourcen verteilt werden
 - Unterstützung von Datenmigration in Speicherorganisationen – caching kann verwendet werden um populäre Daten zu replizieren und die Last zu verteilen; Verwenden von gecachten Daten als permanente Kopien und Entfernen alter Kopien die selten verwendet werden
- Cache Management Policies
 - Charakterisierung
 - Stromabhängiges cachen vs. Block-level caching
 - Traditionelles caching (block-level) kann keine kontinuierliche Auslieferung garantieren
 - Große Medienobjekte müssen als Ganzes gecacht werden.
 - Traditionelle Konzepte basieren auf dem „Heißen Daten-Konzept“ (kleinere Größe)
 - Neue heiße Objekte in den Cache zu bekommen kann länger dauern
 - Aber: Beziehungen zwischen Wiedergabeströmen können hergestellt werden um einen Folgestrom mit dem Cache zu unterstützen
 - Memory vs. Storage Caching
 - Ziele von memory-caching sind die Verbesserung der Cache-Platzausnutzung
 - Storage caching muss Platz und Bandbreite besser ausnützen
 - Caching vs. Prefetching
 - Verwandter Aspekt von Caching: Eher Prefetching von Blöcken die von einem Strom verwendet werden als gemeinsames Verwenden von Blöcken zwischen Strömen; Prefetching kann Abweichungen in der Antwortzeit verbergen und daher Jitter (Schwankungen) vermeiden.
 - Gemeinsame Probleme
 - Adaptive Arbeitslast
 - Policies sollten mit dynamischen Änderungen in der Arbeitslast klarkommen
 - Zurechtkommen von großangelegten verteilten Multimediasystemen die aus vielen verschiedenen Servern bestehen
 - Sicherstellen guter Cache Hits in jedem Server
 - Ausbalancieren der Last über alle Server
 - Unterstützung für VCR control
 - Problem tritt auch in anderen Aspekten auf (batching, load balancing)
 - Contingency policies (GIC)
 - Integration mit anderen Ressourcenoptimierenden Policies
 - Blöcke die von einem Client empfangen werden können von anderen die nahe folgen wiederverwendet werden
 - Batching
 - Stream chasing

Memory Caching Policies

- Stromabhängiges caching
 - Viewer enrollment window
 - Single buffer
 - Wenn mehrere Streams auf das gleiche Medienobjekt zugreifen und dem ersten Stream ein großer Puffer zugewiesen wird können nachfolgende Streams auch den Puffer verwenden
 - Partitioned buffer
 - Verbesserung durch Unterteilung des Puffers und Zuteilung zu Viewgruppen
 - Mehrere Enrollment Windows
 - Weitere Verbesserungen durch adaptives Unterteilen des Puffers
 - BASIC
 - Heißt so wegen der Einfachheit des Ansatzes
 - Wenn ein Puffer erzeugt werden soll wählt BASIC den Puffer der den Block enthält auf den von den existierenden Clients für die längste Periode nicht zugegriffen wird (wenn Datenrate gleich bleibt)
 - Hoher Implementierungsoverhead weil policy die Zugriffsmuster und Beziehungen zwischen den Streams bei jedem Zyklus und jeder Blockveränderung neu schätzt.
 - DISTANCE
 - Verwendet Distanzbeziehungen zwischen Strömen
 - Die Distanz (d_i) eines Clients (c_i) von einem sofort darauffolgenden Client (c_{i+1}) wird durch das offset der Blockanzahl berechnet
 - Für einen Client mit keinen direkt folgenden Clients wird die Distanz als sehr groß angenommen
 - Policy ordnet alle Clients in aufsteigender Reihenfolge nach ihren Distanzen
 - In jedem Zyklus wird zuerst verbrauchter Puffer aus dem letzten Zyklus freigemacht (c_{bi}) in der aufsteigenden Reihenfolge von d_i
 - Kleinerer Implementierungsoverhead als BASIC weil die Distanzkalkulation sich nur ändert wenn ein Client stoppt, pausiert oder ein neuer hinzukommt
 - Interval caching
 - Datenblöcke zwischen einem Paar aufeinanderfolgender Streams S_{ij} und $S_{i(j+1)}$ heißen Intervall
 - Zwei assoziierte Streams heißen preceding und following stream
 - Durch cachen eines laufenden Intervalls kann der nachfolgende Stream durch die Blöcke des vorhergehenden Streams bedient werden
 - Die Zeit eines Intervalls wird geschätzt als die Zeitdifferenz zwischen den beiden Streams beim Lesen des gleichen Blocks
 - Anzahl der Blöcke die gebraucht werden um ein Intervall zu speichern – Cache-Anforderung eines Intervalls
 - Um das Cache Hit Ratio zu verbessern und I/O zu verringern sortiert die policy die Intervalle bezüglich ihrer Größe und cacht nur die kürzesten.
 - Geringer Implementierungsoverhead, Änderungen treten nur bei Neuankünften oder Ende auf
 - Generalized interval caching (GIC)
 - IC nicht optimal für kleine Medienobjekte (keine laufenden Intervalle können gemacht werden)
 - Integriert stream-aware cache management genauso wie temporal locality
 - Für kleine Objekte: Zeitintervall zwischen zwei aufeinanderfolgenden Zugriffen auf das gleiche Objekt
 - Cache Anforderung ist die Größe des Objekts; in diesem Fall wird das ganze Objekt gecacht
 - Größe des letzten Intervalls wird als voraussichtliche Intervallgröße angenommen
 - Split and merge (SAM)
 - Least/most relevant for presentation (L/MRP)
 - Hypergraph access hints

Foliensatz 3

Definition Information Retrieval

- Information Retrieval (IR) (Informationswiedergewinnung, gelegentlich Informationsbeschaffung)
- Betrachtet Informationssysteme in Bezug auf ihre Rolle im Prozess des Wissenstransfers vom menschlichen Wissensproduzenten zum Informations-Nachfragenden
- Beschäftigt sich mit computergestützter, inhaltsorientierter und unscharfer Suche in unstrukturierten Datenmengen

Definition Multimodal

- Multi: viel-, vielfach, Viel-
- Modal: die Art und Weise bezeichnend (mit der etwas geschieht oder gedacht wird)
- Multimodal Information Retrieval oft synonym mit (Inhaltsbasiertes) Multimedia Information Retrieval

Grundprinzip

- Aus einem Medien- oder Multimediaobjekt werden inhaltsbasierte Merkmale extrahiert (feature extraction)
- Diese Merkmale werden in einem Merkmalsvektor zusammengefasst

- Merkmalsvektoren werden gespeichert
- Die Merkmalsextraktion erfolgt für alle Objekte einer Kollektion
- Anfrage: welche zu einem Anfrage-Medienobjekt ähnlichen Objekte enthält die Kollektion?
- Vergleich von Merkmalsvektoren (Anfragevektor vs. Kollektionsvektoren), Ähnlichkeitssuche
- Reihung (Ranking) der Ergebnisse
- Iteration durch Modifikation der Anfrage oder Relevanz-Feedback

Information Retrieval Modelle

Überblick

- IR-Systeme lassen sich nach den zugrunde liegenden IR-Modellen klassifizieren
- IR-Modell legt die Art der Realisierung der folgenden Komponenten eines IR-Systems fest:
 - Interne Dokumentendarstellung
 - Anfrageformulierung und interne Anfragedarstellung
 - Vergleichsfunktion zwischen jeweils zwei Dokumenten beziehungsweise zwischen Anfrage und einem Dokument
- Modelle
 - Boolesches Modell: Dokumente werden als Mengen von Indextermen repräsentiert. Suche über Terme in Termmengen; boolesche Junktoren
 - Konzepte der Mengentheorie und der booleschen Algebra
 - Klare Semantik, sehr einfaches Modell
 - Gewicht eines Terms bezogen auf ein Text-Dokument binär
 - Jedes Dokument wird intern durch die Menge von Indextermen mit Gewicht „1“ repräsentiert
 - In Anfrage werden Terme angegeben, die durch boolesche Junktoren, also durch „and“, „or“ und „not“, kombiniert werden
 - Innerhalb der Vergleichsfunktion werden die durch die Anfrage spezifizierten Anfrageterme in den jeweiligen Dokumenten auf Enthaltensein getestet.
 - Ergebnis eines Termtests ist ein boolescher Wert
 - Boolesche Junktoren zur Kombination einsetzbar
 - Normalformen
 - Für die Berechnung des Anfrageergebnisses wird die Anfrage normalisiert
 - Disjunktive (DNF) oder konjunktive (KNF) Normalform
 - DNF: Disjunktion von Konjunktionstermen (Veroderung)
 - KNF: Konjunktion von Disjunktionstermen (Verundung)
 - Nachteile
 - Exaktes Modell: aufgrund binärer Gewichte keine Ähnlichkeitssuche
 - Größe des Ergebnisses: oft zu viele Dokumente oder keine
 - Boolesche Junktoren: Schwierigkeiten vieler Anwender
 - Milderung
 - Exaktes Modell: Umwandlung von Konjunktionen in Disjunktionen; Stufen der Relevanz
 - Größe des Ergebnisses: einige Systeme haben zweistufiges Suchverfahren (faceted query)
 - Anfrage formuliert und verfeinert, ohne jedoch das Ergebnis anzuzeigen
 - Vollständiges Ergebnis
 - Boolesche Junktoren: all und any statt and und or
 - Fuzzy-Modell: Erweiterung von booleschem Modell
 - Anfragen mit Hilfe von booleschen Junktoren
 - Fuzzy-Theorie zur Milderung der zu scharfen Enthaltenseinsbedingung von Termen in Dokumenten
 - Graduelle Zugehörigkeit von Dokumenten zu Termen
 - Fuzzy-Menge
 - Eine Fuzzy-Menge $A = \{ \langle u; m_A(u) \rangle \}$ über einem Universum U ist durch eine Zugehörigkeitsfunktion $m_A: U \rightarrow [0, 1]$ charakterisiert, welche jedem Element u des Universums U einen Wert $m_A(u)$ aus dem Intervall $[0, 1]$ zuordnet.
 - Menge aller gespeicherten Dokumente: Universum; Term: Fuzzy-Menge
 - Ein Fuzzy-Wert $\mu_t(d1)$ des Dokuments $d1$ bezüglich des Terms t drückt aus, wie stark der Term das Dokument charakterisiert
 - Anfrage analog zum booleschen Modell
 - Jeder Suchterm erzeugt Fuzzy-Menge
 - Anfrage in disjunktive Normalform übergeführt
 - Fuzzy-Mengenoperationen auf Fuzzy-Mengen ausgeführt
 - Ergebnis entsprechend den Zugehörigkeitswerten sortiert
 - Zugehörigkeitswerte
 - Veile Ansätze für die Berechnung
 - Z.B. Term-zu-Term Korrelationsmatrix

- Dokumentenzugehörigkeitswert
 - Zugehörigkeitswert eines Dokumentes d_j zu einem Term t_j :
 - Aggregation aller Korrelationswerte zu allen im Dokument auftretenden Termen durch eine Funktion
- Vektorraummodell
 - Sehr weit verbreitetes Retrieval-Modell
 - Dokumente werden als Vektoren eines Vektorraums aufgefasst
 - Überführung des Retrieval-Problems in das Gebiet der Linearen Algebra
 - Kann überall dort eingesetzt werden, wo Medienobjekte durch eine feste Anzahl numerischer Merkmalswerte dargestellt werden können und sich Ähnlichkeiten auf dieser Basis berechnen lässt
 - Unterstützt im Gegensatz zum booleschen Modell das Konzept der Ähnlichkeit
 - Die Ähnlichkeit wird zwischen zwei Vektoren berechnet
 - Anfrage wird durch einen Vektor repräsentiert (Folien 192ff)
 - Viele Möglichkeiten zur Berechnung der Ähnlichkeit (z.B. Cosinusmaß)
 - Auch Ähnlichkeit über Distanzfunktionen

Relevanzfeedback

Motivation

- IR-Prozess enthält häufig Iterationen
- Anfrageverfeinerung durch Nutzerinteraktion
- Erste Ergebnisliste für den Suchenden oft nicht zufriedenstellend

Anfragemodifikation – Gründe

- Vage Vorstellung über Suchergebnis
- Schlechte Anfrageformulierung
- Unbekannte Datenkollektion
- Keine relevanten Dokumente verfügbar

Arten der Nutzerreaktion

- Browsing – sequentielle Suche, meist wenig sinnvoll
- Manuelle Anfragemodifikation
- Relevanzfeedback
 - Bewertung der Dokumente der Ergebnisliste entsprechend ihrer Relevanz zur Anfrage durch Nutzer
 - Das IR-System modifiziert die Anfrage

Bewertung von Dokumenten – Berücksichtigung folgender Aspekte

- Anzahl der zu bewertenden Dokumente: < 10
- Reduzierte Darstellung der Ergebnisdokumente
- Art der Bewertung
 - Relevant und keine Bewertung
 - Relevant, irrelevant und keine Bewertung
 - Gestufte Relevanzwerte
- Bewertungsgranulat: unterschiedliche Bewertungen eines Dokumentes bezüglich verschiedener Dokumenteneigenschaften; erhöhter Bewertungsaufwand
 - Mehrere Anfrageobjekte
 - Ähnlichkeit aufgrund verschiedener Eigenschaftswerte
- Pseudorelevanz: automatische Bewertung

Bewertungsauswertung – Dokumentenbewertung kann auslösen

- Anfragemodifikation
- Modifikation von Nutzerprofilen
- Modifikation der Dokumentbeschreibungen
- Modifikation des Suchalgorithmus
- Modifikation von Anfragetermgewichten

Verfahren von Rocchio (Folien 204ff)

- Eigentlich für Textdokumente entwickelt, aber auf andere Medientypen übertragbar
- Modifikation von Termgewichten des Anfragevektors im Vektorraummodell
- Termgewichte relevanter Dokumente werden verstärkt und die Termgewichte irrelevanter Termgewichte abgeschwächt
- Verschiebung des Anfragepunktes innerhalb des Vektorraums in Richtung der relevanten Dokumente

Bewertung von Retrieval Systemen

Definitionen

- False alarms – Dokumente die vom Retrieval-System irrtümlicherweise als relevant zurückgeliefert werden
- False dismissals – Dokumente die fälschlicherweise vom Retrieval-System als irrelevant eingestuft wurden und damit im Ergebnis nicht erscheinen
- Correct alarms

- Correct dismissals
- Variablen gelten jeweils gegenüber einer Anfrage q
- Gesuchte Ergebnismenge: $fd + ca$
- Berechnete Ergebnismenge: $ca + fa$
- Dokumentkollektion: $fd + ca + fa + cd$
- Precision: $P_q = \frac{ca}{ca+fa}$
- Recall: $R_q = \frac{ca}{ca+fd}$
- Fallout: $F_q = \frac{fa}{fa+cd}$

Precision-Recall-Paare

- Abhängigkeit des Precision- und Recall-Wertes von der Größe der Ergebnismenge kann in Form einer Linie in einem Precision-Recall-Diagramm dargestellt werden
- Inkrementelles Vergrößern der der Ergebnismenge

Content-based Image Retrieval

Taxonomie von Image Retrieval Typen

- Retrieval by browsing (RBR) – Browsing kann an richtigen Bildern oder thumbnails durchgeführt werden
- Retrieval by objective attributes (ROA) – eine Anfrage wird mit Hilfe von Meta-Attributen, logischen Attributen oder einer Kombination (ähnlich zu DB-Anfragen – perfekter Treffer) formuliert
- Retrieval by spatial (räumlichen) constraints (RSC) – Anfragen basieren auf relativen räumlichen Beziehungen zwischen den Bildobjekten, z.B. Direkte Beziehungen, Nachbarschaft, Überschneidung und Eingrenzung
 - Relaxed RSC queries - erhalte Bilder die so viele gewünschte räumliche Beziehungen wie möglich zufriedenstellen (Ranking erforderlich)
 - Strict RSC queries – erhalte Bilder die präzise alle räumlichen Beziehungen des Anfragebilds zufriedenstellen
- Retrieval by spatial constraints (con't.)
- Wenig Objekte sind in eine Anfrage eingebunden – explizite Spezifikation der gewünschten räumlichen Beziehungen möglich
- Viele Objekte sind in eine Anfrage eingebunden – eine Anfrage ist spezifiziert durch Platzieren von Icons (entsprechen Domänenobjekten) in einem speziellen Fenster (sketch pad window, spw)
- Retrieval by semantic attributes (RSA) – eine Anfrage ist in Begriffen des Domänenkonzepts spezifiziert
- Retrieval by feature similarity (RFS) – eine Anfrage ist durch Auswählen von Features von Beispielbildern spezifiziert; es werden Bilder erhalten die Objekte mit ähnlichen Features enthalten; oft als „content-based“ retrieval bezeichnet

Motivation

- Konventionelle Formen
 - Formale Erfassung + inhaltliche Erschließung
 - Inhaltliche Erschließung enorm aufwändig
 - Manchmal nicht durchführbar (Bilder im Internet)
- Content Based Image Retrieval
 - Bilder aufgrund des dargestellten Inhalts finden
 - Merkmale aus Bildern extrahieren
 - Ähnlichkeitsmessung (keine exakte Übereinstimmung der Bilder)

Anwendungsgebiete

- Medienagenturen
 - Große Bilddatenbanken
 - Manuelle Vergabe von Schlüsselwörtern ist zeit- und kostenintensiv
- Markenzeichen
 - Prüfung ob ähnliche Markenzeichen existieren
- Produktkataloge
 - Kunde erinnert sich eher an Aussehen eines Produkts als an dessen Name oder Artikelnummer
- Kriminalistik
 - Verfahren zur Gesichtserkennung
 - Identifikation von Fingerabdrücken
- Medizin
 - Suche ähnlicher gelagerter Krankheitsfälle
 - Vergleich von Krankheitsbildern

CIBR-Systemarchitektur (Folie 222)

Anfragearten

- Browsing
 - Benutzer interagiert mit dem System
 - Stöbern in vorgegebenen Kategorien
- Textsuche
 - Suche mit Schlagwörtern

- Visuelle Beschreibung
 - Query-by-example – Vorgabe eines Beispielbildes
 - Query-by-sketch - Benutzer skizziert gesuchtes Bild (abhängig von Talent, Technik ,Genauigkeit, ...)
 - Query-b-template – Benutzer definiert Suchbild durch Auswahl aus Farb- und Texturkomponenten

Merkmale

Merkmale Überblick

- Primitive Merkmale
 - Auch Wahrnehmungsmerkmale und Low-Level-Features genannt
 - Einfache Attribute wie Farbe, Textur, Form, räumliche Beziehung von Objekten
 - Werden automatisch aus Bildern extrahiert und miteinander verglichen
- Semantische Merkmale
 - Erkennen von Objekten
 - System benötigt Wissen über Aussehen der Objekte
 - Rollen (Tennisspieler) und Szenen (Sonnenuntergang)
 - Problematik liegt im Aufstellen des Modells (Größe, Form, Bewegung, ...)

Farbmerkmale

- Globale Eigenschaften von Farbverteilung
 - Farbhistogramme: z.B. 64 Farben von hellrot bis dunkelblau
- Lokale Eigenschaften von Objektfarben
 - Objekterkennung durch farbbasierte Aufteilung – Gruppieren benachbarter Pixel mit ähnlichen Farbeigenschaften in einfache Gruppen; die Performance des Ansatzes ist ganz gut wenn er auf Bilder mit halbwegs homogenen Regionen angewandt wird, aber er kann leicht aussetzen wenn starke Hervorhebungen oder Schatten existieren

Farbmodelle

- RGB
 - Gleichmäßige Quantisierung führt zu redundanten und fehlenden Farbwerten
 - Distanzmaße nicht zufriedenstellend (z.B. euklidische Distanz)
 - Hue-Saturation-Brightness
 - Sowohl uniforme (z.B. L^*u^*v) als auch nicht uniforme Farbräume (z.B. HSV)
 - Bei uniformen Farbräumen Ähnlichkeitsmessung durch euklidische Distanz möglich

Histogramm – Eigenschaften

- Einfache und effiziente Repräsentation des Farbinhalts
- Einfache Berechnung
- Unempfindlich bezüglich Änderung der Bildauflösung, Kamerarotation und Zoomen
- Empfindlich gegenüber Änderung der Lichtverhältnisse
 - Verbesserung bei RGB durch Normalisieren der Farbwerte
 - Verbesserung bei HSB durch nicht Berücksichtigen der Helligkeitswerte
- Keine Information über räumliche Verteilung der Farbe

Histogramm – Distanzmaße

- Distanz zwischen Histogrammen – Maß für Ähnlichkeit
- L1- und L2-Maß (Euklidische Distanz)
- Histogramm-Schnitte
 - Unabhängig von der Anzahl der Bins im Histogramm
- Weighted Euclidian Distance
 - Berücksichtigt Zusammenhang zwischen einzelnen Bins

MPEG-7 Farbdeskriptoren – Überblick

- Dominant Color – vorherrschende Farben repräsentieren Bilder bzw. Regionen
 - N dominante Farben plus Prozentsatz der Pixel, die diesen Farben entsprechen
 - Eigenschaften bezüglich räumlicher Verteilung
 - Kompakte und effiziente Beschreibung
- Scaleable Color – ein mittels Haar-Transformationen kodierte Farbhistogramm
 - Farbhistogramm im HSB-Farbraum
 - Quantisierung uniform, 256 Bins fix
 - Kodiert mittels Haar-Transformation
 - Skalierbar durch Weglassen von Koeffizienten
 - Ähnlichkeitsmessung durch Distanzberechnung zwischen Histogrammen (z.B. L1-Maß)
- Color Structure – Farbhistogramm kombiniert mit räumlicher Information
 - Histogramm wird mit räumlicher Information erweitert
 - Quadrat von 8x8 Pixel über das Bild bewegt
 - Enthaltene Farben tragen zum Histogramm bei
 - Farben im HMMD-Farbraum, nicht-uniforme Quantisierung
 - 32-256 Bins

- Color Layout – räumliche Anordnung der Farben im Bild
 - Beschreib räumliche Verteilung der Farben
 - 8x8 großes Feld gefüllt mit repräsentativen Farben der entsprechenden Bildregionen (Farbmittelwerte)
 - Farben im YCrCb-Raum, DCT
 - Auswahl und Kodierung von üblicherweise 12 Koeffizienten
 - Ähnlichkeitsmessung durch Vergleich der Koeffizienten
 - Einfache und schnelle Berechnung
- Group of Pictures – Erweiterung von Scaleable Color auf eine Bildfolge eines Videos
 - Erweiterung des Scaleable Color Deskriptor
 - Errechnung eines Histogramm aus den Farbhistogrammen der Einzelbilder
 - Mittelwert, Median oder Durchschnitt
- Zur zusätzlichen Beschreibung obiger Deskriptoren
 - Color space
 - Unterstützt werden RGB; YCbBr, HSV und HMMD
 - Color Quantization
 - Uniforme Quantisierung eines Farbraums (Definition der Bins)
 - Wird üblicherweise mit Dominant Color Deskriptor verwendet

Texturmerkmale

Überblick

- Texturen nur schwer beschreibbar (retrieval-by-example)
- Textur auch bei Graustufenbildern aussagekräftig
- Eigenschaften von Texturen
 - Körnigkeit – Feinheitsgrad der Textur
 - Periodizität – Regelmäßigkeit der Textur
 - Ausrichtung der räumlichen Anordnung der Textur
 - Eigenschaften sind abhängig von Entfernung zu Kamera und vom Betrachtungswinkel

Verfahren

- Strukturell – Lage und Ausrichtung der Strukturelement
 - Besonders geeignet für regelmäßige Texturen
- Statistisch – Verteilung der Helligkeitswerte im Bild
 - Autokorrelationsfunktion – Maß für Körnigkeit
 - Co-Occurrence-Matrix – Kontrast
 - Simultaneous Auto Regressive Model (SAR) – Pixel werden durch die Helligkeitswerte ihrer Nachbapixel ausgedrückt (RISAR, MRSAR)
- Markov Random Fields
 - Stochastisches Modell – Pixel werden in Abhängigkeit ihrer Nachbapixel modelliert
- Fraktale Modelle
 - Zur Beschreibung von sehr unregelmäßigen Texturen
 - Beispiel – Erosionen, Küstenlinien

Foliensatz 4

Texturen

Tamura Modell – 3 Features

- Kontrast – statistische Verteilung von Pixelintensitäten
- Grobkörnigkeit – Maß der Granularität; basiert auf bewegenden Mittelwerten über Fenster verschiedener Größe pro Pixel
- Richtungsabhängigkeit – Gradientenvektor auf jedem Pixelwert (Graulevel Unterschied) und Winkel

Andere Modelle

- RISAR – Rotation Invariant SAR
- MRSAR Modell – MultiResolution SAR; Beschreib Granularitäten durch Repräsentieren der Texturen durch eine Vielzahl an Auflösungen -> SAR oder RISAR wird auf jedes Level angewandt
- Wold (Heideland) Features – Textur wird beschrieben durch
 - Periodizität – hoch, für starke harmonische Komponenten
 - Richtungsabhängigkeit – hoch für starke dahinschwindende Komponenten
 - Zufälligkeit – hoch für schwach strukturierte Texturen

Brodatz Datenbank

- De facto Standard für das Evaluieren von Texturalgorithmen
- Fotoalbum das 112 Graustufenbilder (512x512 8-bit) enthält, jedes steht für eine andere Texturklasse

Texturevaluierungsalgorithmus von Brodatz

- 9 Subbilder (128x128) werden vom Zentrum jedes Brodatz-Bildes extrahiert. Features werden berechnet indem das Modell auf alle von diesen entstandenen 1008 Subbilder angewandt wird. Die Distanz von jedem Testsubbild zu allen anderen wird dann berechnet
- 2 Ergebnisse
 - Retrieval Rate – Prozentsatz der n nächsten Subbilder die zur gleichen Klasse wie die Testsubbilder gehören
 - Average Retrieval Rate – jedes Subbild wird als Testsubbild berechnet

Brodatz – Vergleich

- Diskussion
 - Tamura
 - Nachteile – gleiche Genauigkeit aber langsamer als SAR
 - Vorteile – Rotationsinvariant; Tamura Features sind näher zur menschlichen visuellen Wahrnehmung
 - MRSAR
 - Vorteile – höchste Genauigkeit
 - Nachteile – Berechnen kleinerer Auflösungen des Bilds wenn der Bildbereich zu klein zum Starten ist

Texturanalyse im Frequenzbereich

- Frequenzbereich
 - Beschreibt räumliche Verteilung der Helligkeitswerte
 - Häufige Helligkeitswechsel entsprechen hohen Frequenzen
 - Errechnet durch Fourier-Transformation
- Frequenzbild (Folie 257f)
 - Position gibt die Frequenz an
 - Tiefe Frequenzen im Mittelpunkt
 - Je weiter entfernt vom Mittelpunkt desto höher die Frequenz

MPEG-7 Texturdeskriptoren

- Texturbrowsing Deskriptor
 - Beschreibt Ausrichtung, Regelmäßigkeit und Körnigkeit der Textur
 - Ausrichtung (direction)
 - Beschreibt vorherrschende Richtung der Textur
 - 6 verschiedene Richtungen (0° bis 150° in 30°-Schritten)
 - 2. Richtung optional
 - Körnigkeit (coarseness)
 - Feinheitsgrad für jede Richtung der Textur – 0 (fein) bis 3 (grob)
 - Regelmäßigkeit (regularity)
 - 4 Abstufungen für die Regelmäßigkeit – 0 (gänzlich unregelmäßig) bis 3 (gleichmäßiges Muster)
- Homogenitätstexturdeskriptor (Folie 262)
 - Beschreibung regelmäßiger Texturen
 - Wichtig für Luft- und Satellitenaufnahmen
 - Berechnung basiert auf Frequenzbild
- Kantenhistogrammdeskriptor
 - Räumliche Verteilung/Anordnung von Kanten im Bild
 - Auch für sehr unregelmäßige Muster geeignet
 - Häufig kombiniert mit Farbdeskriptor (z.B.: ColorLayoutDescriptor)
 - 5 Arten von Kanten
 - 4 richtungsbezogene (0°, 45°, 90° und 135°)
 - 1 nichtrichtungsspezifische (isotrop)

Formmerkmale

- 2 Arten von Formen
 - Umrissbasiert (contour shape)
 - Form eines Objekts aufgrund seiner Umrisslinie repräsentiert (Kantenextraktion)
 - Bereichsbasiert (region shape)
 - Zusammenhängende oder unterbrochene Regionen

MPEG-7 Shape Descriptors (Vier Methoden zur Beschreibung von Formeigenschaften)

- Region Shape – Pixel innerhalb eines bestimmten Bereichs (Objektmomente)
 - Beschreib sowohl einfache als auch komplexe Formen (Folie 267)
 - Transformationsinvariante Objektmomente
 - Der Region-Shape-Deskriptor beschreibt
 - Einfache Formen
 - Komplexe Formen
 - Verfahren zur Berechnung von Region Shape Deskriptoren
 - GMD – Geometric Moment Deskriptor
 - GD – Grid Deskriptor
 - ZMD – Zernike Moment Deskriptor

- Contour Shape – Umriss eines Objekts (Curvature Scale Space Darstellung)
 - Beschreibung eines Objekts / Objektausschnittes
 - Curvature Scale Space (CSS)
 - Vorgehensweise 1
 - Berechnung der Randkoordinaten
 - Normieren, sodass Objekte mit unterschiedlicher Anzahl von Raumkoordinaten verglichen werden können
 - Umriss des Objekts wird geglättet bis keine Wendestellen mehr existieren
 - Vorgehensweise 2
 - Glättung mittels Gauß-Filter
 - Bestimmung von Wendestellen relativ zu einem fixen Startpunkt zu jedem Glättungsschritt
 - Erstellen des CSS-Diagramms
 - Zu jeder Konvexität oder Konkavität der Kontur existiert ein Peak im CSS-Bild
 - Ermittlung der Lage und Höhe der Maxima der einzelnen Peaks
 - Vorgehensweise 3
 - Ähnlichkeitsmessung durch Histogrammvergleich
 - Invariant gegenüber Skalierung und Rotation
- Shape 3D Spektrum – Anzahl von lokalen Konvexitäten einer 3D Oberfläche (Histogrammform)
 - Beschreibung lokaler geometrischer Eigenschaften
 - Shape Index (SI)
 - Beschreibt Konvexität in einem Punkt
 - Funktion der beiden Hauptkrümmungen in diesem Punkt
 - Skalierungs-, translations-, und rotationsabhängig
 - Elementare Formen
 - SI für Punkte der gesamten 3D Objektoberfläche
 - SI für Scheitelpunkte bei Gitternetzen
 - Histogramm (Shape Spektrum)
- Shape 2D/3D – Beschreibung einer 3D Form durch mehrere 2D Deskriptoren (aus verschiedenen Winkeln)
 - 3D Objekt wird aus verschiedenen Winkeln als 2D Bild beschrieben
 - Ähnlichkeitsvergleich mit Bildern aus gleichen Betrachtungswinkeln
 - Anwendung aller visuellen Deskriptoren
 - Contour shape
 - Region shape
 - Color deskriptoren
 - Texturdeskriptoren

Weitere Merkmale

- Räumliche Anordnung von Objekten in einem Bild
- Kantenbilder (query-by-sketch) (Folien 280ff)
 - Query-by-sketch
 - Bildsuche mittels handgefertigter Skizze
 - Erfolgsquote abhängig von Fähigkeit des Zeichners und Genauigkeit der Skizze
 - Merkmale abhängig von verwendeter Technik
 - Umriss, Schattierung, Farben
 - Kantenbilder
 - Bilder müssen als Kantenbilder repräsentiert werden
 - Position der Objekte muss nicht exakt stimmen
- Beziehung von Objekten zueinander
 - Richtungsbezogen – Position zueinander (links, rechts, etc), Entfernung, Winkel zwischen Objekten
 - Topologie – „gleich wie“, „innerhalb von“, „berührt“, überlappt“, Beziehungen sind invariant gegenüber Translation, Rotation, Skalierung
- Beispiel – 2D-String (Folie 278ff)
 - 2D-String Repräsentationen
 - Projektion der Objekte auf die x-Achse
 - Projektion der Objekte auf die y-Achse
 - Erweiterungen zu 2D-String
 - Berücksichtigung von Entfernung und Ausrichtung der Objekte
 - Zusätzliche Beziehungen (A überlappt B, A enthält B, ...)
 - Umgang mit Segmenten von Objekten
- 2D-String Repräsentation ändert sich bei Translation nicht

Gesichtserkennung

- 2 Aufgabenstellungen
 - Gesichtsdetektion – das Erkennen von Gesichtern in Bildern

- Gesichtsidentifikation – das Wiedererkennen eines Gesichts
- Anwendungsgebiete
 - Zugangskontrollen für sicherheitskritische Bereiche
 - Identifizierung von Straftätern
- Charakteristische Punkte von Gesichtern: Folie 284
- Geometrische Merkmale
 - Abstand der Augen, Nasenlänge, Breite des Mundes, ...
 - Horizontale und vertikale Kanten des Gesichts ermitteln
 - Charakteristische Punkte des Gesichts
- Ähnlichkeitsmessung
 - Distanzen einzelner Punkte zueinander – abhängig von Lage, Ausrichtung und Größe des Gesichts
 - Normierung der Distanzen
 - Relationen von Distanzen – Abstand rechtes Auge zum rechten Rand zu Abstand linkes Auge zum linken Rand

Verfahren zur Gesichtserkennung

- Template Matching
 - Erstellung von Schablonen für Gesichter bzw. Gesichtsteile (Mund, Nase, ...)
 - Maß der Ähnlichkeit – Übereinstimmungen der Schablonen zueinander
- Ähnlichkeitsmessung im Frequenzbereich
 - Diskrete Fouriertransformation
 - Maß der Ähnlichkeit wird durch Koeffizientenvergleich bestimmt
- Eigenfaces
 - Bilder werden als Vektoren repräsentiert (Komponenten = Pixel des Bildes)
 - Bilder, die Gesichter darstellen, bilden Unterraum des gesamten Vektorraums
 - Eigenvektoren des Unterraums mittels PCA ermitteln
 - Gesichter werden durch Linearkombinationen der Eigenvektoren approximiert (Eigenfaces)
 - Maß der Ähnlichkeit – Distanz zweier Eigenfaces

Probleme bei der Gesichtserkennung (Faktoren, die die Gesichtserkennung erschweren)

- Bildqualität (Schärfe, Kontrast, Helligkeit)
- Unterschiedliche Beleuchtung, Probleme mit harten Schatten
- Unterschiedliche Kamerawinkel
- Kopfposition und -haltung (Frontalansicht, Seitenansicht)
- Gesicht unterschiedlich durch Bart, Make-Up, Tragen einer Brille, ...
- Gesichtsausdruck verändert das Gesicht (traurig, wütend, fröhlich, ...)

User Interfaces for Formulation Queries

- Query durch visuelle Vorlagen
- Zeichnen einer Query
- Query durch visuelle Beispiele (Blobworld; Folie 292f)

Indexingmethoden

- Schlüsselattribute für Bilder sind Featurevektoren die mit einem Punkt im multidimensionalen Featureraum korrespondieren
- Ähnlichkeits-Queries korrespondieren zum nächsten Nachbarn oder zu Bereichssuchen
- Schnelle Query Prozesse erfordern multidimensionale Indexierungstechniken
 - R-Trees
 - Linear quadtrees
 - Grid files
- Die meisten Indexingmethoden explodieren exponentiell wenn die Nummer der Dimensionen steigt
- Für die Effizienz sind Hohe Dimensionstechniken nicht besser als sequenzielles Scannen
- Suchzeit mit linearen Quadrees ist proportional zur Hyperfläche der Query Region und wächst exponentiell
- Mit Grid-Files wächst das Directory exponentiell mit den Dimensionen
- Limitierungen von Indexingmethoden sind zurückzuführen auf 2 Annahmen:
 - 1) Distanz im Suchraum ist euklidisch
 - 2) Die Dimensionalität des Suchraums ist halbwegs gering
 - Suchraum sollte in einem Vorschrift gefiltert werden welcher false hits, aber keine fals dismissals erlaubt
 - Feature Vektoren werden in geeigneteren Raum gemappt in welchem alle Distanzen die Distanzen im aktuellen Suchraum unterbewerten werden.

R-Trees (Folien 297f)

- R-Tree basiertes Indexing kann für höhere Dimensionen effizient sein wenn der Fan-out > 2 und die Dimensionalität unter 20 bleibt
- Ein geometrisches Objekt wird durch seine minimum bounding rectangle (MBR) repräsentiert

- Nicht-Blätter-Knoten enthalten Einträge der Form (R, ptr) wobei ptr ein Zeiger auf einen Kindknoten ist und R der MBR ist der alle Rechtecke in den Kindknoten abdeckt
- Blattknoten enthalten Einträge der Form (obj-id, R) wobei obj-id ein Zeiger auf die Objektbeschreibung und R der MBR des Objekts ist
- Die Hauptinnovation ist das Elternknoten überlappen dürfen; deshalb kann ein R-Tree mindestens 50% Platzausnutzung garantieren

Foliensatz 5

Videoretrieval

Videsegmentation (Schnitterkennung)

- Hierarchische Gliederung
 - Video -> Episode -> videoshot -> frame
- Terminologie
 - Shot oder Take bezieht sich auf eine zusammenhängende Aufnahme von Videoframes die eine kontinuierliche Aktion in Zeit und Raum wiedergeben
 - Video Segmentation ist der Prozess des Entscheidens der Grenzen zwischen aufeinanderfolgenden Kamerashots
- Shot-Level Organisation von Video wird als bestpassend betrachtet für Videobrowsing und content-based retrieval
- Gruppierung in
 - Frameunterscheidungstechniken
 - Pixelvergleiche
 - Framehistogrammvergleiche
 - Kameraoperationen
 - Techniken auf komprimiertem Video

Frameunterscheidungstechniken

- Paarweises Pixelvergleichen – ein Pixel wird als anders beurteilt wenn die Differenz zwischen den Intensitätswerten von 2 aufeinanderfolgenden Frames einen gegebenen Grenzwert übersteigt
- Segment-Boundary (Segmentgrenzen) – wenn mehr als eine angegebene Prozentzahl der Totalanzahl der Pixel sich geändert hat
 - Potenzielles Problem: Metrik ist sensibel bezüglich Kamerabewegung, Lichtänderungen, Objektbewegung und Rauschen
- Likelihood ratio: basierend auf der zweiten Ordnung von statistischen Charakteristika von Intensitätswerten von korrespondierenden Regionen (Blöcken) aufeinanderfolgender Frames
 - Zwei aufeinanderfolgende Frames werden in ein Set von Beispielflächen unterteilt
 - Eine Kamerapause kann bestimmt werden wenn die totale Anzahl von Beispielflächen deren Likelihood ratio den Grenzwert übersteigt groß genug
 - Vorteil gegenüber paarweisem Pixelvergleich – Likelihood ratio erhöht das Toleranzlevel zu langsamer und kleiner Objektbewegung von frame zu frame. Das macht es unwahrscheinlicher dass diese Effekte irrtümlich als Kamerapausen interpretiert werden
 - Nachteil – wenn die zwei verglichenen Beispielflächen den gleichen Mittelwert und Varianz haben aber eine komplett verschiedene Wahrscheinlichkeitsdichtefunktion wird keine Änderung erkannt. Zum Glück ist solch eine Situation sehr unwahrscheinlich

Histogrammtechniken (Folien 310ff)

- Das Histogramm eines Frames bietet eine globale Beschreibung; $H_i(j)$ ist der Histogrammwert des i-ten Frames, wobei j eine von G möglichen Graustufen ist; die Differenz zwischen dem i-ten Frame und seinem Nachfolger wird gegeben durch: $SD_i = \sum_{j=1}^G |H_i(j) - H_{i+1}(j)|$ Falls die Gesamtdifferenz SD_i größer als ein bestimmter Grenzwert ist wird ein Segmentboundary deklariert
- Die Gleichung kann auch auf Farbkanalhistogramme angewandt werden
- Histogrammtechniken sind weniger sensibel auf Objektbewegung als paarweise Pixelvergleichsalgorithmen
- Frame Histogramm Differencing (Folie 314)

Motion Continuity

- Bewegung ist eine wichtige Eigenschaft jeglicher Videoquelle
- Optical Flow – Vektorfeld wo jeder Vektor eine Geschwindigkeit des entsprechenden Blocks repräsentiert
- Bewegungsvektoren werden durch Blockmatching berechnet, indem zu jedem Block in einem Frame i, b_i , ein Vektor D_{bi} zugewiesen wird welcher den den ganzen Block displaced (translates)
- Folien 316ff Gleichungen

Gradual Transitions

- Kamerapause – hohe Pulse
- Gradual Transitions – Pulse mit höheren Werten als deren Nachbarn aber wesentlich niedriger als der Abschaltgrenzwert

Twin-Comparison Approach (Folie 320ff)

- Erfordert 2 Abschaltgrenzwerte: T_b wird für die Kamerapauseerkennung und T_s für die Spezialeffekterkennung verwendet
- Wenn eine Differenz zwischen T_s und T_b fällt wird der entsprechende Frame markiert und mit den folgenden Frames verglichen
- Der Endframe einer Transition wird entdeckt wenn die Differenz zwischen aufeinanderfolgenden Frames sich auf weniger als T_s verringert, während der Vergleich mit dem markierten Frame sich auf einen Wert größer als T_b erhöht hat.
- $SD_{p,q}$ – die Unterschiede zwischen aufeinanderfolgenden Frames definiert durch die Differenzmetrik
- $SD'_{p,q}$ – berechnet wenn $T_b > SD_{p,q} > T_s$; die akkumulierte Differenz zwischen dem aktuellen Frame und dem potenziellen Starting F_s Frame einer Transition
- T_s – der Grenzwert der verwendet wird um den startenden Frame F_s entdecken
- T_b – der Grenzwert der verwendet wird um den endenden Frame F_e und Kamerapausen zu entdecken

Multi-Pass Approach

- Ansatz um die Bearbeitungszeit zu verkürzen
- Erster Durchlauf
 - Ein kleineres T_b wird gesetzt
 - Die Auflösung wird kurzzeitig geopfert -> hoher Skip-Faktor: erhöht $SD_{p,q}$ während gradual transitions
 - Kamerapausen und gradual transitions werden entdeckt (auch Falschmeldungen)
 - Weitere Durchläufe – kurzzeitig erhöhte Auflösung wird verwendet um die potenziellen Segmentboundaries besser zu erkennen

Motion Analysis

Kameraoperationserkennung

- Änderungen mit Spezialeffekttransitionen müssen von Änderungen durch Kamerabewegungen unterschieden werden
- Änderungen wegen Kamerabewegung tendieren dazu darauffolgende Differenzwerte des gleichen Grades wie die der Gradual transition zu erzeugen. Deswegen ist es notwendig Muster von Bildbewegung zu erkennen die durch Kameraoperationen erzeugt werden.
- 2 Ansätze
 - Motion vector analysis
 - Frame wird in Blöcke unterteilt
 - Blockmatching – jedem Block wird ein Vektor zugeteilt
 - Optical Flow – Vektorfeld wo jeder Vektor eine Geschwindigkeit des entsprechenden Blocks auf einer Bildebene repräsentiert
 - Die Verteilung von Bewegungsvektoren in einem Frame kann folgendes erkennen:
 - Panning (Schwenken) and tilting (Kippen) (gleiche Richtung der Vektoren)
 - Zooming (alle Vektoren raus) in and zooming out (alle Vektoren rein)
 - Bestimmte Bewegungsmuster
 - Objektbewegung
 - Kann auch bestimmte Aktionen erkennen
 - Bewegung von Mensch oder Tier – laufen, hüpfen, gehen
 - Objektbewegung – Bewegen von Autos, Trucks, Flugzeugen, Helikoptern
 - Eine Beschreibung kann verwendet werden um eine generelle raue Beschreibung der Szene zu erzeugen
 - Video X-ray (Würfel)
 - Raumzeitliches Bild
 - Annahme: Hintergrund enthält unverwechselbare Muster
 - Panning: schräge Linien auf der oberen Ansicht
 - Tilting: schräge Linien auf der seitlichen Ansicht
 - Zooming – Linien divergieren oder konvergieren

Bewegungsdeskriptoren

Überblick

- Bewegung ist ein wichtiges visuelles Merkmal zur Charakterisierung von Videosequenzen
- Oft mit anderen Merkmalen kombiniert
- Einsatz im Video-Retrieval, Video-Browsing und in der Überwachungsindustrie
- Beispiel – Unterscheidung von Nachrichtensprechereinstellungen und eingespielten Beiträgen

MPEG-7 Motiondeskriptoren

- Camera Motion – Kamerabewegung
- Motion Trajectory – Bewegungstrajektorie (Ablauf, Flugbahn)
- Motion Activity – Bewegungsaktivität

Kamerabewegung

- Fixed – keine Kamerabewegung

- Track left/right – horizontale Kamerafahrt
- Boom up/down – vertikale Kamerafahrt
- Dolly forward/backward – Kamerafahrt entlang der optischen Achse
- Pan left/right – horizontaler Schwenk
- Tilt up/down – vertikaler Schwenk
- Roll – Drehung um die optische Achse

Bewegungstrajektorie

- Beschreibung der Bewegungen einzelner Objekte
- Vorgangsweise
 - Verfolgung eines für das Objekt repräsentativen Punktes (z.B. Schwerpunkt)
 - Erstellen einer Liste von Schlüsselpunkten (x,y,z,t)
 - Interpolation zwischen den Schlüsselpunkten
 - Nicht lineare Interpolation durch Angabe von Interpolationsparametern möglich

Parametrische Bewegungsmodelle

- Beschreibt die Bewegung von Objekten in Videos als zweidimensionale parametrische Bewegungsmodelle
- Geometrische Modelle – Translation, Rotation und Skalierung und perspektivische Modelle
- Repräsentation komplexer Bewegungsabläufe – Kombination der Modelle
- Bewegungsmodelle sind anwendbar auf beliebige Objekte (Bildregion = Gruppe von Pixel)
- Objektretrieval durch Angabe eines Beispiels oder Anfragen wie „gesucht sind rotierende Objekte“

Bewegungsaktivität

- Videosequenzen unterscheiden sich aufgrund der Bewegungsaktivität
 - Beispiel – Sportbeiträge weisen höhere Aktivität auf als Nachrichtenbeiträge
- Bewegungsintensität – Beschreibung durch Integer-Wert im Bereich von [1,5]
 - 1 – keine bzw. sehr geringe Bewegungsaktivität
 - 5 – sehr hohe Bewegungsaktivität
 - Wert errechnet sich aus Standardabweichung der Beträge der Bewegungsvektoren

Videorepräsentation

Repräsentierung und Klassifizierung

- Wie kann man eine Stunde Video in nur ein paar Minuten schauen und trotzdem den Inhalt korrekt wahrnehmen?
 - Video Content Abstraction – Prozess des Extrahierens einer Präsentation von visuellen Informationen inkl. Hervorstechender Features, typischer Stil und aller Hauptsubjekte
 - Video Icon Konstruktion – Keyframe ist der Frame der den hervorstechenden Inhalt des Shots präsentiert

Video Icons (mehrere Ansätze)

- Iconbasierend auf einem extrahieren Frame des Shots, mit Pseudotiefe für Repräsentation der Dauer, Pfeilen und Zeichen für die Repräsentation von Objekt- und Kamerabewegungen
- Darstellung eines Bilds dass den globalen visuellen Inhalt des Shots repräsentiert
 - Hervorstechende Standbilder
 - Videospace icon
 - Videomap

Interactive Tools (Folie 249ff)

- Movie icon (micon) – Video wird durch ein 3d Volumen dargestellt
- Interactive micons – Examinierung und Manipulationsumgebung
- Paper-vieo – „chard-based“ video browser
- Video panorama – Anzeige des aufgenommen Videoplatzes
- Videoscope – Videoinhaltsanalysierer
- Sound Browsr – entdeckt das Vorhandensein von Musik

Key-Frame Extraction

- Vorteil – einfache Berechnung: wenn signifikante Inhaltsänderungen auftreten zwischen dem aktuellen Frame des Shots und dem letzten Keyframe wird der aktuelle Frame als Keyframe ausgewählt – color, texture, motion
- Nachteil – nicht besonders repräsentativ

Video Browsing Tools

- Timeline und strata (Schicht) browsers
- Hierarchische Browser

Video Content Indication (Präsentieren Teil des Inhalts für besseres Verständnis)

- Schnittstellen für MM-Inhaltsanzeige sollten folgendes bieten
 - Weiche Abschnittsübergängen zwischen den Beobachtungsmodi
 - Sinn des Überblicks – Interface sollte ein breites Spektrum an ähnlichem (related) Inhalt bieten
 - Sinn des Unterteilens
 - Effektive Präsentation – sehr intuitive Repräsentationen werden benötigt, nicht einfach für zeitliches Material wie Video
 - Attraktivität – Interface sollte den Originalanzeigestil so gut wie möglich behalten.

Segmentation of Compressed Video

- Hardware Dekompression um vorher verwendete Methoden zu verwenden
- Software Dekompression – weitaus weniger effizient
- Wenn Hardware Dekompression nicht verfügbar ist – verwende Metriken die die direkt auf die komprimierten Repräsentationen angewandt werden
 - DCT Koeffizienten – (JPEG, MPEG)
 - MPEG motion vectors

Algorithmen die auf DCT Koeffizienten basieren

- MPEG-Video – besteht aus I,P,B Frames (nur I-Frames mit DCD Koeffizienten kodiert)
- GOP Struktur bestimmt Häufigkeit der I-Frames
- Vorteile – da nur ein kleiner Anteil aller Videoframes I-Frames ist wird die Berechnungszeit reduziert; wegen des hohen Skip-Faktors werden gradual transitions und breaks während eines Durchlaufs entdeckt
- Nachteile – Verlust der zeitlichen Auflösung kann false positives nach sich ziehen

DCT Koeffizientenkorrelation

- DCT-Koeffizienten nacheinanderfolgender Frames JPEG-komprimierter Videos werden verglichen

DCT Blockvergleich

- Paarweiser DCT-Blockvergleich

Bewegungsvektorbasierende Algorithmen

- Bewegungsvektoren im MPEG Datenstrom
 - P-Frames – single set an motion vectors
 - B-Frames – two sets of motion vectors, for- and backward
- Feld von Bewegungsvektoren in einem Video
 - Innerhalb eines Kamershots – ziemlich kontinuierliche Änderungen
 - Zwischen verschiedenen Shots – Kontinuität wird unterbrochen
- Definition M-
 - P-Frame – Anzahl valider Bewegungsvektoren
 - B-Frame – Kleiner Anzahl an validen Bv (forward and backward)

Hybrider Ansatz zur Unterteilung

- Erster Durchlauf – DCT Vergleich (I-Frames) mit hohem Skip-Faktor um Regionen potenzieller Breaks, gradual transitions, camera operations und object motions zu entdecken
- Zweiter Durchlauf – DCT Vergleich mit kleinerem Skip-Faktor der nur auf die Nachbarschaft potenzieller ... angewandt wird; löscht falsche Positive
- Weitere Durchläufe – Bewegungsbasierter Vergleich entweder auf dem ganzen Video oder nur in den vorher entdeckten Sequenzen bevor die Durchläufe die DCT Resultate verifizieren und die Genauigkeit verbessern

Evaluierung der Algorithmen (Folien 369ff)

- Diskussion
 - Breaks – Bewegungsvektoren und Hybridansatz entdeckten alle Breaks, DCT weniger effektiv; falsche Entdeckung war hauptsächlich weil nur I-Frames verwendet werden können
 - Gradual Transitions – Bewegungsvektoren versagten komplett
 - Fazit – Hybridansatz bietet höchste Genauigkeit im Entdecken von breaks und gradual transitions
- Kameraoperationsalgorithmus entdeckte erfolgreich alle 4 Kameraschwenks in den Testdaten

Fallstudien

Fernsehnachrichten

- Automatische Extraktion semantischer Information eines Videos nur möglich wenn seine Struktur auf Domänenwissen basiert
- Beispiel – Fernsehnachrichten
 - Räumliche Struktur – anchorperson shots
 - Zeitliche Struktur – in der Reihenfolge der Shots

Nachrichtenvideoparsingalgorithmus (Folien 377ff)

- Zeitliche Segmentierung (siehe Video Segmentierung)
- Klassifizierung der Shots
 - Definieren eines Modells für einen A-Shot (Anchorperson shot)
 - Entwickeln von Ähnlichkeitsmaßen die verwendet werden können um diesem Modell zu entsprechen
 - Zeitliches Strukturmodell eines kompletten Nachrichtenprogramms – Sequenz und Episodenidentifizierung
- Visuelle Abstraktion – k-Frames für jeden Shot um seinen visuellen Inhalt zu repräsentieren

Matching other Shot Types

- Start- und Endsequenzen – werden als Signatur einer Produktion verwendet, sehr konsistent verwendet, hat eine fixe und vordefinierte zeitliche und räumliche Struktur, kann einfach identifiziert werden
- Werbungen – meist vordefinierte Startsequenzen wenn das Programm von einer Werbepause zurückkehrt
- Wetterberichte – bedeutende Variation, Allgemein tendiert der Frame der gezeigt wird dazu den ganzen Shot über identisch zu sein, oft eine vordefinierte Startsequenz

- Nachrichtensendungen – zu viel Varietät um ein Strukturmodell unterstützen zu können; kann identifiziert werden als nicht konform zu einem Modell

Basketball

Überblick

- Erkennen von Schlüsselszenen in Sportvideos
 - Beispiele
 - Tore und Fouls im Fußball
 - Korbwürfe im Basketball
- Vorgehensweise
 - Empirische Beschreibung von Schlüsselereignissen
 - Aufstellen von Zeitmodellen, die das Auftreten der Schlüsselereignisse in eine typische Reihenfolge bringen
 - Erstellen von Low-Level Merkmalen, die zum Erkennen der Schlüsselereignisse herangezogen werden können.

Schlüsselereignisse

- Jubelnde Menge
 - Erhöhter Geräuschpegel im Audiosignal ist ein Indiz für einen geglückten Korbwurf
- Einblendung des Spielstandes
 - Nach einem Treffer wird die Spielstandsanzeige aktualisiert
- Richtungsänderung der Spieler
 - Nach einem Treffer bewegen sich die Spieler wieder vom Korb weg (Richtungswechsel der Spieler und der Kameraführung)

5 verschiedene Zeitmodelle (Folie 388)

Merkmale Audioanalyse

- MPEG1-Audio
- Amplituden werden normal kodiert
 - Aktueller Normierungsfaktor leitet sich aus größtem Amplitudenwert des jeweiligen Frequenzbandes und Frames ab
- Merkmal zum Erkennen der jubelnden Menge ergibt sich aus der Summe der Normierungsfaktoren

Merkmale Videoanalyse

- Texterkennung
 - Spielstandsanzeige als Text
 - Charakteristisch sind scharfe Kanten
 - Kanten sind im Frequenzspektrum als hohe Frequenzen erkennbar
- Bewegungsrichtung
 - Richtungsänderung der Spieler erkennbar an Kameraschwenks

Indexing

Indexing and Retrieval

- Zugriff auf Videodaten durch gespeicherte Metadaten: Shots, Object Motion, camera movement, textual descriptions of objects and events, etc
- Metadaten decken eine Sequenz an Frames (Frameintervalle)
- Segment Index Trees (SR-Trees) – Adaptierung von R-Bäumen für Segmentintervalle; ein Knoten speichert ein Intervall (statt MBR); Elternknoten enthält die Intervalle der Kinder; effizienter Mechanismus um Intervall- und Punktdaten in einem Index zu indexieren; ein neues einzufügendes Intervall kann splitten (Folie 394)

Audio Retrieval

Einleitung/Motivation

- Weltweit beherbergen Archive eine gigantische Anzahl an Video- und Tondokumenten
- Exponentielles Wachstum von Musikangeboten
- Nur automatische Indizierung kann solche Archive auf Dauer nutzbar machen

Audio vs. Visuelles Retrieval

- Heute wird mehr Augenmerk auf visuellen Aspekt gelegt
- Akustischer Teil oft (noch) außer Acht gelassen
- In vielen Fällen Ton aussagekräftiger als Bild
 - Beispiel Videoszene mit Dialog -> Gesprochenes aufschlussreicher als Bild

Audio-Indexierung und -retrieval

- Einfachste Methode: über Titel und Dateiname
 - Sehr verbreitet
 - Namen allerdings unvollständig und subjektiv – schwierig zu finden
 - Außerdem keine Möglichkeit, Audioaufnahmen zu finden, die so klingen wie etwas was gerade zu hören ist
- Inhalt verwenden
 - Vergleiche Messwert für Messwert
 - Wenig erfolgversprechend, da Unterschiede in Abtastrate und Auflösung nicht berücksichtigt
 - Daher Merkmale (Features) extrahieren und nutzen
 - Mittlere Amplitude

▪ Frequenz-Verteilung

Allgemeiner Ansatz

- Klassifikation – in verbreitete Typen wie Sprache, Musik, Geräusch
- Differenzierte Behandlung jeder Klasse – z.B. Sprache: Spracherkennung und Indexierung des Textes
- Anfragen – ebenso klassifiziert, verarbeitet und indexiert
- Retrieval – beruht auf der Ähnlichkeit der Anfrage-Merkmale mit den Merkmalen der gespeicherten Tondokumente

Klassifikation

- Verschiedene Typen verlangen unterschiedliche Verarbeitung und unterschiedliche Indexierungstechniken
- Verschiedene Typen haben unterschiedliche Bedeutung für eine Anwendung
- Sprache ist der wichtigste Typ, und es gibt heute recht erfolgreiche Spracherkennungstechniken und –systeme
- Die Typinformation selbst ist in einigen Anwendungen sehr nützlich
- Der Suchraum reduziert sich auf eine Klasse

Grundlagen

Retrieval Prozess

- Beispiele für Anfragen
 - Query durch Summen
 - Finde ein bestimmtes Wort in einem Nachrichtenarchiv
 - Finde einen bestimmten Tierlaut in einem Tierlaute-Archiv

Repräsentationsraum

- Abfrage-Audiosignal und Audiosignale der Datenbank werden in Repräsentationsraum (e) transformiert
- Vergleich erfolgt im Repräsentationsraum
- In Abhängigkeit von der Aufgabe: viele Repräsentationsräume und entsprechend viele Ähnlichkeitsmaße
- 2 Gründe für e
 - Abfrage nicht notwendigerweise in Waveform
 - e muss für die konkrete Aufgabe diskriminieren: e ist extrem abhängig vom konkreten Problem
- e ist ein Raum von Audio Features

Segmentierung des Signals

- Frame (Folie 408)
 - Jeder Frame ist an einer bestimmten Zeit t positioniert
 - Charakterisiert das Signal zur Zeit t
 - Multiplikation des Drucksignals mit einem Fenster(-signal) (positioniert zur Zeit t) ergibt einen Frame
 - t korrespondiert mit dem Zentrum oder einem der Endpunkte
 - Windows können verschiedene Formen haben (z.B. Hamming, Bartlett, Dreieck, Rechteck)

Spektrogramm (Folie 410)

- Einfache Darstellungen haben Grenzen
 - Zeit-Domäne zeigt Frequenzanteile eines Signals nicht
 - Frequenz-Domäne zeigt nicht wann Frequenzen auftreten
- Kombinierte Darstellung – Spektrogramm
 - X-Achse: Zeit, y-Achse: Frequenzanteile
 - Schwärzung (Farbe) eines Punkts: Energie der Frequenz zu dieser Zeit
- Analysen – z.B. Regelmäßigkeit des Auftretens von Frequenzen, Musik vs. Geräusch

Vorgangsweise

- Eingangssignal wird blockweise verarbeitet
- Überlappende Segmente des Signals werden verwendet
- Sinusoide Fensterfunktionen in Abb. Deuten Signalausschnitte an, auf die sich die Analyse in einem Schritt „konzentriert“
- Punktweise Multiplikation von Signalblock mit Fensterfunktion
- Resultierendes Signal wird fouriertransformiert
- Aneinanderreihung der Spektralvektoren liefert Zeit-Frequenzdarstellung des Signals

Short Time FT (STFT) – Folie 412

Spektrogramm – Folie 413

Audio Klassifikation

- Sprache
 - Männliche oder weibliche Sprache
- Musik
 - Arten von Musik
- Umgebungsgeräusche (z.B. Tierlaute)

Sprache

- Bandbreite vergleichsweise gering, 100 bis 7000Hz
- Zentroid deshalb niedriger als bei Musik
- Häufige Pausen (zwischen Worten und Sätzen)

- Höherer Anteil der Stille
- Charakteristische Struktur: Folgen von Silben, die aus kurzen Perioden von Friktionen (Konsonanten) bestehen, auf die längere Perioden von Vokalen folgen
- Während der Friktionen hohe Nulldurchlaufrate, ZC variiert stärker

Musik

- Hohe Bandbreite, 16 bis 20.000Hz
- Zentroid deshalb höher
- Niedriger Anteil der Stille
- Ausnahmen: Soloinstrument, A-Capella-Gesang
- Nulldurchlauf variiert nicht so stark
- Regular beat

Vorgang Klassifikation

- Schritt für Schritt
 - Ein Merkmal nach dem anderen
 - Z.B. erst Zentroid; wenn hoch: Musik
 - Dann Anteil der Stille; wenn niedrig: Musik
 - Dann ZC-Variabilität; wenn niedrig: Solo-Musik; sonst: Sprache
 - Reihenfolge wichtig; algorithmische Komplexität und Differenzierungsvermögen einfach zu berechnen und hohe Differenzierung zuerst
 - Ein Merkmal allein auch schon nutzbar: nur ZC: bis zu 90% korrekt klassifiziert; nur Anteil der Stille: bis zu 82%
- Feature Vektoren
 - Werte einer Menge von Merkmalen berechnet und zu Vektor zusammengefasst
 - Training: Durchschnittsvektor (Referenzvektor) einer jeden Klasse bestimmen
 - Für neues Audio Feature – Vektor berechnen und Distanz zu den Referenzvektoren ermitteln (meist euklidisch)

Spracherkennung (Folie 420f)

- Nach der Klassifikation
- Techniken
 - Time Warping (Sprechgeschwindigkeit)
 - Hidden Markov Models
 - Neuronale Netze

Musikindexierung

- Noch in den Anfängen
- 3 Arten
 - Strukturierte Musik
 - Synthetische Musik (MIDI)
 - Aufgezeichnete Musik

Indexierung strukturierter Musik

- Keine Extraktion von Merkmalen erforderlich
- Sogar exakte Übereinstimmung als Suchmethode denkbar
- Allerdings könnten Instrumente nicht übereinstimmen
- Ähnlichkeit schwierig zu definieren
- Eine Möglichkeit: nur den Pitch-Wechsel berücksichtigen (Up, Down, Repeat – U, D, R)
- Retrieval durch Zeichenkettenvergleich

Beispiele: Folien 424ff

Audiomerkmale (im Spektrogramm Bereich)

Merkmale Einteilung

- Unterschiedliche Einteilungen in der Literatur
- Sprache, Musik, Geräusch
- Frames oder Clips
- Subjektivität / Interpretation (z.B. Timbre)
- Nach Repräsentationsräumen (Zeit, Frequenz, oder spezieller)

Volume

- Auch Short Time Energy, Lautstärke

Bandenergie

- Energie innerhalb eines Frequenzbereichs (Band) $f_0 - f_1$
- Band Energy Ratio
- Histogramm Approximation

Medianfrequenz

- Median des Spektrums eines Frames, auch Centroid Frequenz oder Brightness genannt
- Lagemaß

Bandbreite

- Streuungsmaß

Anteil der Stille

- Silence Ratio
- Anteil der Messwerte an der Gesamtzahl, die einer Periode (!) der Stille angehören
- Zwei Schwellenwerte
 - Amplitudenwert, unterhalb dessen Stille angenommen wird
 - Anzahl unmittelbar aufeinanderfolgende Messwerte, die mindestens still sein müssen, um eine Stilleperiode zu bilden

Harmonie

- Spektrale Komponenten oft Vielfache der niedrigsten und lautesten Frequenz („fundamental frequency“)
- Musik in der Regel harmonischer als andere Geräusche
- Prüfung, ob eine Tonaufnahme harmonisch ist: dominante Komponenten Vielfache der fundamentalen Frequenz?
- Beispiel: Flöte spielt Note G4; Spitzen bei den Frequenzen 400, 800, 1200, 1600Hz usw.
- F, 2f, 3f, 4f usw. Harmonische der Note

Tonhöhe

- Nur periodische Klänge (Instrumente, Stimme)
- Perkussion dagegen nicht
- Subjektiv; verwandt, aber nicht gleichbedeutend mit der fundamentalen Frequenz, die oft als Näherung verwendet wird
- Viele unterschiedliche Berechnungsverfahren

Nullkreuzungsrate

- Zero Crossing Rate
- Anzahl der Vorzeichenwechsel in einem Frame
- Diskreter Fall

Probleme

- Außer Vorzeichenwechselrate alle diskutierten Features verwandt mit Spektrogramm
- Wahl von Fenster und Fensterlänge hat entscheidenden Einfluss auf Merkmale
- Redundante Features

Foliensatz 6

Distanz und Ähnlichkeit

Distanzfunktionen

Einleitung

- Distanzfunktionen vergleichen die Merkmale zweier Medienobjekte
- Invarianz – drückt aus welche Merkmale zum Vergleich nicht herangezogen werden sollten

Definition

- Binäre Funktion mit folgenden Eigenschaften
 - Selbstidentität
 - Positivität
 - Symmetrie
 - Dreiecksungleichung

Distanzfunktionen

- Einfache Distanzfunktion
- Euklidische Distanzfunktion
- Minkowski Distanzfunktion
- m-Einheitskreise
- gewichtete Minkowski-Distanz
- Einheitskreise
- Quadratische Distanz
- Mahalanobis Distanzfunktion
- Quadratische Pseudodistanz
- Einheitskreis
- Bottleneck Distanz

Ähnlichkeitsmaße

- Objekte werden als ähnlich wahrgenommen, wenn sie bei Menschen zu ähnlichen Reizen (Stimuli) führen
- Keine allgemein akzeptierte Definition von Ähnlichkeit
- Ähnlichkeitsmodelle in Mathematik, Statistik, Bildverarbeitung und Mustererkennung
- Ähnlichkeitsmaß: Funktion, die einem Paar von Objekten eine reelle Zahl aus $[0,1]$ zuordnet
- Wert 1 korrespondiert mit maximaler Ähnlichkeit

Distanz vs. Ähnlichkeit

- Viele Ansätze verwenden Distanzfunktionen auf Featurewerten
- Distanzwerte werden auf $[0,1]$ abgebildet
- Distanzeigenschaften für Ähnlichkeitsempfinden zu restriktiv (Untersuchungen in der Psychologie)
- Bedeutet nicht automatisch, dass Distanzfunktionen für Ähnlichkeitsmaße ungeeignet sind.
- Nur nicht grundsätzlich für alle Anwendungen geeignet

Probleme

- Selbstidentität: gilt nicht grundsätzlich
- Positivität: von Tversky als allgemeine Bedingung für menschliches Ähnlichkeitsempfinden widerlegt
- Symmetrie: Rollentausch macht Unterschied
- Dreiecksungleichung: Unterschiede zwischen 2 Objekten werden zu hoch bewertet wenn kein drittes für den Vergleich vorliegt

Symmetrieprobleme

- Abfrage mit A als Suchbild vs. B als Suchbild
- Salient features

Dreiecksungleichung

- Unähnlichkeit zwischen A und B wird stärker eingeschätzt als Summe der Unähnlichkeiten zu C

Ähnlichkeitsabstand

- Unähnlichkeitsmaß
- Mindesteigenschaften:
 - Dominanz
 - Konsistenz
 - Transitivität
- Eigenschaften sind allgemeiner als Distanzeigenschaften (z.B. Symmetrie nicht gefordert)
- Bei Anwendung einer monoton wachsenden Funktion auf Werte eines Abstandsmaßes bleiben Eigenschaften erhalten

Grenzen von Ähnlichkeitsmaßen

- Weltwissen spielt Rolle bei menschlicher Wahrnehmung
- Ebenen der Inhaltsverarbeitung
 - Syntaktisch (ohne Bedeutung der Objekte)
 - Semantisch (Ähnlichkeitsvergleich)
 - Pragmatisch (Interpretation, thematische Kategorien)

Pre-Attentive vs. Attentive Wahrnehmung

- Pre-attentive Wahrnehmung in den ersten 250ms (ohne Interpretation = Weltwissen)
- Features der preattentiven Phase
 - Linienorientierung
 - Länge, Breite, Größe von Objekten
 - Krümmung
 - Anzahl von Objekten
 - Farbe und Intensität von Objekten

Ähnlichkeitsmaße

- Viele verschiedene Funktionen und Maße vorgeschlagen
- Viele Kombinationen
- Leider keine allgemein anerkannte Kombination
- Viele Alternativen zur Auswahl

Feature-Kontrast-Modell

- Matching
- Monotonie
- Unabhängigkeit

Kosinusmaß

Abstandsfunktion

Aggregation von Werten

- Ähnlichkeitswerte
- Monotonie
- Strikte Monotonie
- Stetigkeit
- Idempotenz
- Unabhängigkeit von der Reihenfolge

Generalisiertes Mittel

Umwandlungsfunktionen

Dynamische Sensibilität

Modifikation

Parametrisierbare Funktion

Foliensatz 7

Content Description

MPEG-7

- Motivation
 - Immer mehr audiovisuelle Information ist digital erhältlich
 - Neue Wege zu produzieren, anzubieten, zu filtern, zu suchen und verwalten der digitalisierten MM-Information
 - Breitband wird angeboten und Audio-/Videoqualität wird immer besser und schneller zu erreichen
 - Informationswert hängt oft davon ab wie leicht Information gefunden, erhalten, verwaltet, gefiltert werden kann.
 - Benutzer werden mit so vielen Inhalten konfrontiert dass der effiziente Zugriff fast unmöglich ist
 - Inhalt effizient zu identifizieren und zu verwalten wird immer schwieriger wegen des schieren Volumens
 - Das Problem ist nicht nur auf DB-Retrieval Applications begrenzt (auch bei MM-Editing, Broadcast-Channel-Selection, Multimedia directory services)
- Lösung
 - ISO/IEC Standard; früher: Multimedia Content Description Interface
 - Sowohl Menschen als auch Maschinen die audiovisuelle Information bearbeiten sind im Rahmen von MPEG-7
 - Anbieten von Tools
 - Aktive Leute: Broadcasters, Elektronikerzeuger, Inhaltserzeuger und –manager, Publishers...
- Ziele
 - Beschreibung von MM-Content
 - Flexibilität in Datenmanagement
 - Globalisierung und Interopabilität von Datenressourcen
 - Versucht zu standardisieren:
 - Set von Description Schemes und Descriptors um Daten zu beschreiben
 - Sprache um Description Schemes zu spezifizieren, z.B. Description Definition Language (DDL)
 - Schema um die Description zu kodieren

Terminologie

- Feature: Unterscheidendes Merkmal von Daten das jemandem etwas bedeutet
 - Features brauchen eine bedeutungsvolle Repräsentation (Descriptor) und eine Instanzierung (Descriptor Value) für ein gegebenes Datenset
 - Beispiele: Farbe eines Bilds, fundamentale Frequenz eines Sprachsegments, Rhythmus eines Audiosegments, Kamerabewegung in einem Video, Genre eines Musikstücks, Titel eines Films, Schauspieler in einem Film
- Descriptor: repräsentiert ein Feature. Ein Descriptor definiert die Syntax und die Semantik einer Featurerepräsentation
 - Muss präzise die Semantik des Features, den assoziierten Datentyp, erlaubte Werte und eine Interpretation der Descriptor Values definieren
 - Beispiel: Color: string. Datentyp kann zusammengesetzt sein, z.B.: RGB-Color: [integer, integer, integer]
 - Mehrere Descriptors können ein Feature beschreiben (verschiedene relevante Anforderungen)
 - Beispiele: enumerierte Listen, Color moments, Histogramme die Farbe repräsentieren
 - Syntaxbeispiel: <name, typekind, spec> value </name>
- Descriptor-Value: Instanzierung eines Descriptors für ein gegebenes Datenset. Sie werden mittels Description Schemes kombiniert um eine Description zu formen.
- Description Scheme: spezifiziert die Struktur und Semantik der Beziehungen zwischen seinen Komponenten, welche Descriptors oder Description Schemes sein können.

Description Scheme

- Wird mittels DDL generiert
- Unterschied zwischen Description Scheme und Descriptor ist dass ein Descriptor sich um die Repräsentierung eines Features kümmert während sich das Description Scheme um die Struktur der Description kümmert.

Terminologie

- Description: besteht aus Descriptor Scheme (Struktur) und einem Set von Descriptor Values (Instanzierungen) die die Daten beschreiben. Ein Descriptor enthält bzw. bezieht sich auf ein volles oder teilweise instanziiertes Description Scheme.
- Coded Description: Description die kodiert wurde um relevante Anforderungen wie Kompression, Effizienz, Direktzugriff, etc. zu erfüllen
- Description Definition Language (DDL): erlaubt die Erstellung und Modifizierung von Description Schemes und Descriptors.

MPEG-7 in der Praxis: Folie 513

DDL-Anforderungen

- Compositional capabilities

- Transformational capabilities
- Unique identification
- Data types
- Relationships innerhalb und zwischen DS
- Relationships zwischen Description und Daten

MPEG-7 Bestandteile

1. MPEG-7 Systems: tools
2. DDL
3. Audio
4. Visual
5. MM Description Schemes
6. Reference Software
7. Conformance

Foliensatz 8

Digital and Interactive Television

Überblick über digitales Fernsehen

- neue Technologie für Übertragen/Empfangen von Fernsehen
- zusätzliche 6MHz angeboten; 19,39 Mb/s Datenstrom
- ersetzt analoges Signal
- neue Fernsehgeräte oder spezielle Set-top Boxen notwendig

Hauptthemen

- erste neue große Innovation in fast 60 Jahren des Fernsehens
 - Digital ist analog überlegen
 - Flexibilität: macht Computer/TV aus TV
 - Vielseitigkeit: HD oder mehrere Signale, Daten
 - Effizienz: Weniger Energie zum Übertragen notwendig
 - Qualität: kein Rauschen/Ghosting, perfektes Signal
 - Widescreen
 - Audio: 5 Kanaltone

Kosten

- Übertragung: 2 Signale für die nächsten 6 Jahre
- Neue DTV Transmitter, Antennen
- Lokale Produktion in Digital oder HDTV
- Neues Seitenverhältnis 16:9 statt 4:3 -> Aufnahmen für beide?

Standardkörper und Vereinbarungen

- European Telecommunications Standards Institute (ETSI)
- Digital Video Broadcasting (DVB)
- Advanced Television Systems Committee (ATSC)
- Digital Audio Visual Council (DAVIC)
- European Cable Communication Association (ECCA)
- Federal Communications Commission (FCC)

Building Blocks of Digital TV

- Kompression und Kodierung
- Modulation
 - Quadrature Amplitude Modulation (QAM): bis zu 40Mbit/s; Kabelgesellschaften
 - Quadrature Phase Shift Keying (QPSK): Satellitenumgebung, 10Mbit/s
 - Coded Orthogonal Frequency Division Multiplexing (COFDM): verschiedene Signalträger, funktioniert gut in zugebauten Gegenden
- Conditional Access System
- Netzwerkübertragungstechnologien
 - Digital über Hybrid Fiber-coax (HFC)
 - **Digital über Wireless Cable??? (WTF???)**
 - Multichannel multipoint distribution system (MMDS)
 - Local multipoint distribution system (LMDS)
 - Digital über terrestrisch (DTT)
 - Digital über Direct Broadcast Satellite (DBS)
 - Netzwerkmanagement

DVB

- DVB Konsortium hat ein Set von Standards für MPEG-2 basierendes Digitalfernsehen veröffentlicht

- Wird in Europa, Asien und Australien verwendet
- Erlaubt SDVT und HDTV
- Frameraten von 25 oder 30Hz
- Für PAL oder SECAM; MPEG-2 Hauptprofil auf Hauptlevel (MP@ML)
- Auf Konsumentenseite verbindet das Consumer Multimedia Home-Network (CMHN) MM Equipment daheim
- Konsument verwendet den Service mit einer Multimedia Home Plattform (MHP) kompatiblen settop-box, ein Standardteil des CMHN.
- MHP ist eine Erweiterung des DVB Digital TV Standards für Konsumenten Settop-Boxen.
- Es definiert eine Java API für value-added services
- Für value-added Applications die nichtlokale Interaktivität implementieren, DVB interactive services, ist ein Feedback Kanal für zweiweg-Informationsaustausch verfügbar (wird vom ISP angeboten; Verwendung IP-basierter Protokolle)

System Reference Architecture

- Vertraut auf die folgenden Schlüsselstandards
 - MPEG-2 Transport Streams (TS) transmission technology standards
 - DVB-SI Service Information Standards
 - DVB data broadcasting standards basierend auf MPEG-2 und dem digital storage media command and control (DSM-CC) Protokoll.
 - DVB interactive service standards inkludieren DVB Feedback Kanalstandards

DVB Übertragungsstandards

- Übertragung von MPEG-2 TS in 3 physikalischen Broadcast Medien:
 - Satellit (DVB-S, QPSK)
 - Kabel (DVBC, QAM)
 - Terrestrisch (DVB-T, COFDM)
- Standars teilen gewöhnliche technische Elemente die einen hohen Integrationslevel für TV Broadcasting und Receiver Equipment ermöglichen.

DVB Architektur Komponenten

- DVB Broadcast Service Provider (BSP)
- DVB Interactive Service Provider (ISP)
- Service Providers (SPs)
- Service Editors (SEs)
- Broadcast Channels
- Feedback Channel
- Consumer Multimedia Home Network (CMHN)

Hauptfeatures von Set-tops

- Dekodieren des eingehenden digitalen Signals
- Verifizierung von Zugriffsrechten und Sicherheitslevels
- Anzeige von Kinoqualitätsbildern am Fernsehgerät
- Ausgabe von digitalem Surroundsound
- Bearbeiten und Rendern von Internet und Interactive Services

Set-top Architektur

- Architektur ist sehr ähnlich wie ein Standard-Multimedia-Computer und in folgende Kategorien unterteilt
- Systemboard
- Tuner – hauptsächlich digitale und analoge Eingangssignale. Unterteilt in:
 - Broadcast in band – isoliert einen physikalischen Kanal von einem Multiplex von Kanälen und konvertiert ihn in ein Baseband
 - Out-of-Band Tuner – bietet Abonnements ein Gemisch von interaktiven Services
 - Return Path Tuner: Schickt Daten zurück zum interaktiven Service Provider
 - Demodulator – konvertiert modulierte Signal in einen digitalen Bitstrom
 - Modulator – konvertiert digitalen Bitstrom in ein modulierte Signal
 - Demultiplexer – Splittet MPEG-2 Streams in separate Streams: Video, Audio, Daten von Interaktiven Serv.
 - Decrypter – verhindert nicht autorisierte Benutzer am Programmsehen
 - Decoder – Datendekompression
 - CPU und Speicher
 - Modems
 - High-Speed Interfaces – Echtzeitkommunikation mit DVD-Playern, Camcorders, CD Playern, USB, Firewire
 - Smart Card Reader – autorisiert Abonnements den Zugang zu digitalen TV Services
 - Wireless Keyboard – ermöglicht Abonnements Internetzugriff und interaktive TV Services

Set-Top Betriebssysteme

- Anforderungen und Restriktionen
 - Sehr robust und zuverlässig
 - Limitierte Hardwareressourcen

- Fähigkeit gleichzeitig eine Anzahl an Tasks zu bearbeiten
- Muss open computer und Internetstandards wenn möglich annehmen
- Layers
 - Kernel
 - Loader
 - Driver
 - APIs
- Kein Standard set-top OS

Set-top Middleware

- Middleware entspricht den Application, Presentation und Session Layers des OSI Modells
- Wird verwendet um set-top Applikationen von den darunterliegenden Hardware und Netzwerkkomponenten zu isolieren
 - Virtuelle Maschinen
 - Multimedia Home Platform

MHP

- Ermöglicht die Aufnahme und Präsentation von Applikationen in einem offenen Framework
- Applikationen von Service Providern werden interoperabel mit verschiedenen MHP Implementierungen sein
- Applikationen, Netzwerke und MHP Terminals können von unabhängigen Providern verfügbar sein
- Definiert Java-basierte API für value-added services in DVB
- Beruht auf DVB Digital TV System Reference mit standardisiertem Broadcast und Feedback Channels
- Offener Standard basierend auf Java
- Standardisiert Schnittstelle zwischen Software der Box (Middleware) und TV-Applikation
- Ist kein Subjekt einer Lizenz
- Funktioniert auf Low- und Highend-Produkten
- Besteht aus User Terminal (set-top Box, PC, Peripheriegeräte), Standardmiddleware, APIs
- Basiert auf Internet- und Webstandards und bietet so Kompatibilität und Konvergenz zwischen TV und Internet

MHP für die Zuseher

- Enhanced Broadcasting – z.B. Erweiterter Teletext auf HTML basierend, mehr Information in besserem Layout
- Interactive Broadcasting – z.B. Televoting (Backchannel erforderlich)
- Internet Service – volle Verbindung zwischen Home-TV Equipment und Internet

Kategorien von Digitalen Set-tops

- Broadcast TV – kann nur Daten des MPEG2-Stroms verarbeiten
- Enhanced TV – Rückkanal: VOD, E-Commerce, Internetbrowsing
- Advanced Services – Ähnlichkeit zu einem MM-Desktop; schneller Rückkanal

OCAP

- Technical Level
 - Verschiedene DVB-Technologien und Spezifikationen die nicht in den USA Kabelumgebungen verwendet werden, werden entfernt und durch ihre funktionale Äquivalente laut GEM ersetzt
- Business Level
 - OCAP erweitert MHP mit Unterstützung für eine Monitoranwendung welche den Kabel MSO (Netzwerkoperator) zuständig für die policy des Receivers macht.

Relationship MHP, OCAP

- Ursprünglich basierte OCAP auf Version 1.0.0 der MHP Spezifikation
- Neuere Versionen verwenden GEM statt MHP als Grundlage
- Zusätzlich verwendet OCAP Elemente von MHP die nicht in GEM enthalten sind
- Das heißt es verwendet manche MHP-Elemente über GEM und manche direkt

GEM

- Globally Executable MHP
- Erlaubt Organisationen Spezifikationen zu definieren die auf MHP basieren zusammen mit DVB
- Ziele
 - Erhöhung der Interopabilität zwischen GEM-basierten Spezifikationen verschiedener Organisationen
 - Erhöhung der Präsenz von MHP Komponenten
 - Berücksichtigung von lokalen geschäftlichen und technischen Constraints
- Keine Standalone Spezifikation, sondern ein Framework um Spezifikationen auf MHP zu definieren
- GEM Spezifikationen erlaubt Inhaltsautoren Applikationen zu schreiben die direkt interoperabel zwischen verschiedenen GEM-basierten Receivern sind.
- GEM Spezifikation listet die Teile von MHP die DVB Technologie oder marktspezifisch sind.