Cross-Media-Retrieval

Thomas Zerbach

Proseminar Multimediadatenbanken und Retrieval
Institute for Web Science and Technologies
FB4 – Computer Science
Universität Koblenz-Landau, 56070 Koblenz, Germany
tzerbach@uni-koblenz.de

Abstract. Nach einer Studie der IDC Central Europe GmbH von 2008[1] wächst die weltweite Datenmenge um jährlich rund 60 Prozent. Dadurch rücken effiziente Verfahren zur Suche von Daten immer mehr in den Fokus der Forschung. Damit ein Benutzer möglichst schnell und effizient an für ihn relevante Informationen gelangt, müssen Daten so auf Datenbanken abgelegt werden, dass der Benutzer mit möglichst geringen Abweichungen genau die Informationen erhält, nach denen er sucht. Bei der Suche nach einem bestimmten Medientyp muss sich dabei nicht nur auf Textanfragen beschränkt werden, sondern es kann dank neuartiger Anwendungen jeglicher Medientyp als Suchparameter verwendet werden. Die nachfolgende Ausarbeitung beschäftigt sich mit den theoretischen Aspekten des Cross-Media-Retrieval und veranschaulicht das Thema anhand von einigen Praxisbeispielen.

Keywords: Cross-Media-Retrieval, Suche, Medien, Datenbanken

1 Einführung

Das Mediennutzungsverhalten hat sich in den letzten Jahrzehnten stark zugunsten der elektronischen Medien entwickelt. Dabei werden selbst die klassischen Medien immer mehr durch Computer und das Internet verdrängt. Durch günstigere elektronische Geräte haben mehr Nutzer die Möglichkeit, sich beliebige multimediale Inhalte zu beschaffen. Dabei ist besonders wichtig, dass dem Nutzer eine geeignete Infrastruktur und Verfahren zur Verfügung gestellt werden, mit deren Hilfe er mit möglichst geringem Aufwand an relevante Informationen gelangt. Auf der Seite des Anbieters bedeutet dies, dass Multimediadatenbanken Informationen bereitstellen müssen, die über angepasste Suchverfahren mit möglichst hoher Treffergenauigkeit vom Benutzer

1 http://www.macwelt.de/artikel/_News/353708/idc_studie_weltweite_datenmenge_waechst_jaehrlich_um_60_prozent

abgerufen werden können. Gleichzeitig werden von Benutzerseite Anforderungen an das Suchverfahren gestellt, im Bezug auf u.a. Intuitivität, Fehlertoleranz, Genauigkeit und Geschwindigkeit.

Das Cross-Media-Retrieval (CMR) beschreibt dabei Ansätze für intelligente Suchverfahren, bei denen es Nutzern ermöglicht wird, medienübergreifende Suchanfragen zu stellen („Cross-Media“). Das bedeutet, dass Text, Ton, Bild oder Bewegtbild sowohl Eingabe-, als auch Ausgabeparameter sein können. Um dies umzusetzen müssen Suchmethoden vorhanden sein, die mit den speziell dafür bearbeiteten Daten einer Multimediadatenbank interagieren können.

Im zweiten Abschnitt wird das CMR zunächst in den allgemeineren Kontext des Information-Retrieval eingeordnet und daraufhin werden in Kapitel 3 die Besonderheiten des CRM herausgearbeitet. In Kapitel 4 folgen anschauliche Beispiele zur Thematik und abschließend werden die Ergebnisse dieser Ausarbeitung in der Zusammenfassung gebündelt und ein grober Ausblick gegeben.

2 Information-Retrieval in Multimediadatenbanken

Multimediadatenbanken unterscheiden sich erheblich von relationalen Datenbanksystemen im Bezug auf Speicherung, Indexierung und Suche von Daten. In relationalen Datenbanken liegen explizite Informationen in Form von Attributen vor und Suchanfragen werden anhand fest vorgegebener Syntax in *SQL*[2] gestellt. Dabei müssen Grundkenntnisse über die Struktur der in der Datenbank vorliegenden Daten vorhanden sein, da im Suchterm Attribute übergeben werden müssen, die mit der internen Repräsentation übereinstimmen. Die Ergebnismenge generiert sich dabei aus den zum Suchterm gefundenen Treffern, alle Daten die sich nicht in dieser geschlossenen Menge befinden werden vernachlässigt. Für Multimediadatenbanken kommen diese festgelegten Strukturen nicht in Frage, da es möglich sein soll unscharfe Suchanfragen zu stellen, die nicht ausschließlich exakte Treffer liefern. Wie in Abschnitt 3 näher erläutert wird, soll das IR-System unterschiedliche Medientypen als Suchparameter verwenden können, um einen eindeutigen Abgleich mit gleichen oder ähnlichen Datenbankobjekten anhand ihres Inhalts durchführen zu können. So kann z.B. der Ausschnitt eines Musikstücks dazu verwendet werden, Zusatzinformationen zum Interpreten abzurufen oder Musikstücke mit ähnlicher Audiospur auszugeben. In den folgenden Unterkapiteln werden die einzelnen Schritte näher erläutert, die nötig sind um die Daten mit Metainformationen anzureichern, um im nächsten Schritt über Suchfunktionen möglichst effizient Ergebnisse zu liefern.

2.1 Indexierung

2 SQL: Datenbanksprache zur Definition, Abfrage und Manipulation von Daten

Ziel der Indexierung innerhalb des Information-Retrievals (IR) ist es, die interne Datenrepräsentation so zu erweitern, dass auch implizit vorliegende Informationen erfasst und über die Suchfunktion innerhalb der Multimediadatenbank abrufbar gemacht werden. Dazu muss eine semantische Interpretation der Daten stattfinden und die dadurch gewonnenen Metadaten in die interne Datenrepräsentation überführt werden.

Zum jetzigen Zeitpunkt wird die Anreicherung der Daten mit Metadaten meist manuell durchgeführt und ist daher aus mehreren Gesichtspunkten eher unzureichend für die korrekte und vor allem objektive Erfassung von inhaltsbezogenen Informationen. Bei diesem Verfahren handelt es sich meist um beschreibende Texte, Überschriften oder Schlagwörter, die von Personen an das Objekt angehängt werden. Vor allem das Sammeln verschiedenartiger Informationen, die zum Vergleich von unterschiedlichen Medienobjekten herangezogen werden können, stellt die Forschung vor große Herausforderungen. Eine möglichst passende Annäherung an die menschliche Wahrnehmung, die ausschlaggebend für das Ähnlichkeitsempfinden ist, gestaltet sich dabei sehr schwierig.

Es existieren einige Ansätze für Algorithmen, die auf einer niedrigen („low-level“) Ebene automatisiert Informationen extrahieren können. Explizit vorliegende Informationen werden dazu verwendet, Rückschlüsse auf die tatsächliche Semantik des Medienobjektes zu ziehen. Implizite Informationen, die sich nicht aus der digitalen Information ableiten lassen, entsprechen dabei der höchsten („high-level“) Ebene und können bislang nicht erfasst werden. So können einem Foto zwar Informationen zur Farbzusammensetzung entnommen werden, an welchem Ort das Foto entstanden ist, lässt sich jedoch ohne manuelle Annotationen in der Regel nicht direkt nachvollziehen.

Bei der Erzeugung der Merkmale auf low-level Ebene werden sogenannte Feature-Werte dazu herangezogen einen Feature-Index für jedes Medienobjekt zu erstellen, der bei Suchanfragen dazu genutzt wird eine Ähnlichkeitsbestimmung zwischen Medienobjekt und Suchanfrage durchzuführen. Bei einem Feature-Wert handelt es sich um ein wie bereits zuvor beschriebenes low-level Merkmal des Objektes, das in ein spezielles Format, den Feature-Index überführt wird, um die Effizienz des Suchvorgangs zu erhöhen. Allgemein ist festzuhalten, dass es im Bereich der Indexierung von Multimediaobjekten zahlreiche Ansätze gibt, die unterschiedliche Erfolgsquoten erzielen. Eine Standardisierung hat in diesem Bereich noch nicht stattgefunden, daher stellen die folgenden Ausführung keinen Anspruch an Vollständigkeit.

2.1.1 Zerlegung und Normalisierung

Zu Beginn der Vorverarbeitung eines Multimediaobjektes wird es zunächst in seine nicht komplexen Bestandteile zerlegt. Bei einem komplexen Multimediaobjekt handelt

es sich um ein Datenformat, das verschiedene Medientypen in sich vereinigt wie bspw. eine Präsentation, die Texte, Bilder und Videos enthalten kann und daher in Einzelbestandteile zerlegt werden muss.

Bei der Normalisierung wird das Medienobjekt in eine Normalform überführt. Dabei werden Einflüsse von Störfaktoren unterdrückt, die für die restliche Verarbeitung von keiner Relevanz sind wie bspw. das Rauschen einer Audiodatei.

2.1.2 Segmentierung

Bevor die Eigenschaften eines Multimediaobjektes näher analysiert werden können, muss es zunächst anhand syntaktischer Merkmale in verschiedene Segmente unterteilt werden, die später herangezogen werden um Informationen zu extrahieren. Ein Problem bei allen Segmentierungsverfahren ist, dass durch die Automatisierung oft semantisch inkonsistente Segmente erstellt werden und diese daher für die weitere Verarbeitung keinen Wert mehr haben. Die Segmentierung gestaltet sich nach Medientyp unterschiedlich.

Bei **Audiodateien** können, wie in Abbildung 1 zu erkennen, die verschiedenen Tonspuren dazu genutzt werden, um Segmente zu bilden. So können je nach Audioprofil Tonspuren klassifiziert werden, wie Sprache, Musik, Störgeräusche oder Stille. Dabei werden in späteren Schritten alle Audioklassen für die Indexierung verwendet, bis auf diejenigen, die keine Informationen enthalten (Stille). Nach der Aufteilung in Klassen folgt die *Clusterbildung*[3]. Je nach Klasse können bspw. kurze Tonpausen dazu genutzt werden, einzelne Abschnitte zu bündeln. Besonders bei der Sprache eröffnet dies Möglichkeiten Sinnabschnitte zu bilden, aus denen mit Spracherkennungstools Textinformation extrahiert werden können.

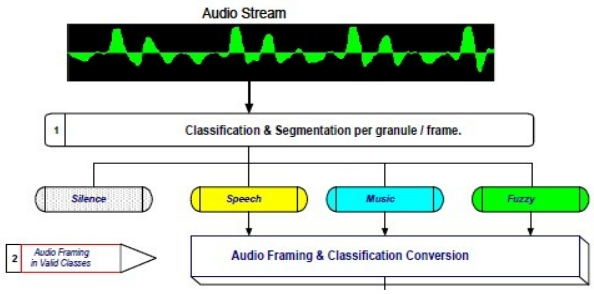

Abbildung 1: Schema zur Audiosegmentierung (Quelle: [1])

In **Bildern** werden Segmente abhängig von Farbe, Textur, Position und Form der darin befindlichen Pixeln gebildet. Um dies zu erreichen, stehen verschiedene Verfahren zur Verfügung. Häufig verwendet wird die Ecken- und die Kantenerfassung. Beide Methoden beruhen auf unterschiedlichen Algorithmen, jedoch ist die Funktionsweise ähnlich. Je nachdem welches Farbmodell zu Grunde gelegt

3 Clusterbildung: Bündelung

wird (bspw. RGB oder CMYK), wird ein Farbhistogramm zum Bild erstellt um Vorder- und Hintergrund voneinander unterscheiden zu können. Die Unterscheidung wird durch unterschiedliche Helligkeitsebenen erreicht, da der Vordergrund meist heller als der Hintergrund ist. In Abbildung 2 sind die einzelnen Schritte der

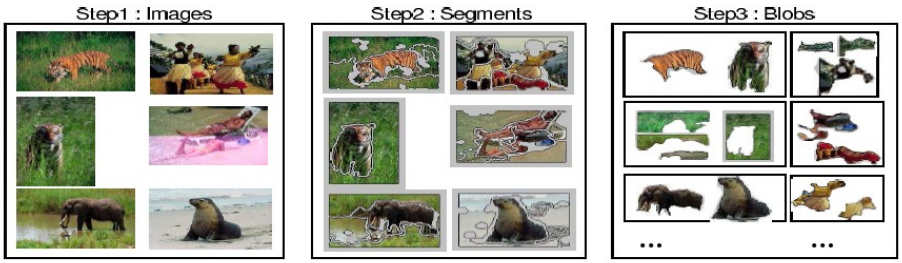

Abbildung 2: Segmentierungsschritte von Bilddaten (Quelle: [2])

Bildsegmentierung dargestellt. Im zweiten Schritt sind die Ausgangsbilder mit den hervorgehobenen, erkannten Segmenten zu sehen. Bereiche, die größer als ein vordefinierter Schwellenwert sind, werden mit in die Menge relevanter Segmente aufgenommen, deren Feature-Werte in weiteren Schritten extrahiert werden sollen. Aus Abbildung 2 lässt sich zudem ein Beispiel für die semantische Inkonsistenz der automatisiert erstellten Segmente entnehmen. Im oberen linken Bereich aus Schritt 3 wird der Körper eines Tigers mit in die relevante Menge an Segmenten aufgenommen, sein Kopf wurde jedoch vom Algorithmus einem anderen Segment zugeteilt.

In **Videos** bestehen auf Grund der Reichhaltigkeit des Mediums unterschiedliche Möglichkeiten Segmentierungen vorzunehmen. Anwenden lassen sich die gleichen Verfahren für sowohl Audio-, als auch Bildsegmentierung. Um sich Zusammenhänge einzelner Frames zu Nutze zu machen, gibt es zusätzlich zwei weitere Verfahren, die zeitliche Segmentierung und die Bewegungssegmentierung.

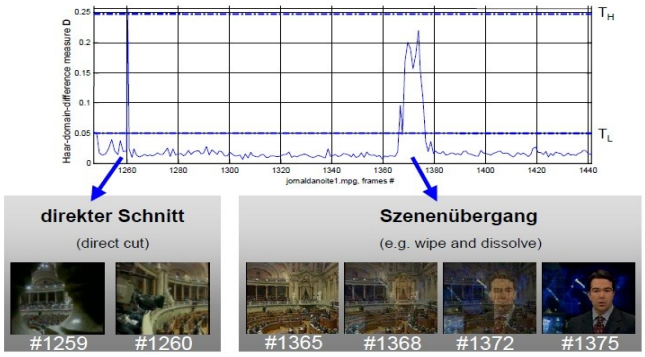

Abbildung 3: Zeitliche Segmentierung (Quelle: [7])

Bei der zeitlichen Segmentierung wird die Twin-Comparison-Methode angewandt, um an Hand von Histogrammdifferenzen verschiedene Videoabschnitte im Bezug auf den Szenenwechsel zu kategorisieren. Wie in Abbildung 3 zu erkennen, wird ein

oberer und unterer Schwellenwert der Differenzen ermittelt, um bestimmen zu können, wann ein Schnitt oder eine Überblendung vorliegt. Bei einem Schnitt ist die Histogrammdifferenz sehr hoch und es werden beide Schwellenwerte überschritten, da zwei komplett unterschiedliche Bilder aufeinanderfolgen. Bei einer Überblendung wird der Bildwechsel in die Länge gezogen, daher sind die Unterschiede aufeinanderfolgender Bilder geringer und es wird nur der untere Schwellenwert überschritten.

Bei der Bewegungssegmentierung werden Vektoren dazu genutzt, Bildverläufe zu erfassen und die Videodatei zu unterteilen. Dabei wird eine Raster, bestehend aus sogenannten Makroblöcken über ein Bild gelegt und registriert, welche Bereiche sich von Frame zu Frame voneinander unterscheiden. Erfasst werden die Richtung und die Geschwindigkeit der Bewegung einzelner Blöcke in einer bestimmten Zeitspanne. Die Möglichkeit mit Hilfe von Makroblöcken Bewegungen zu erkennen und auszuwerten, ist im MPEG-Standard integriert.

Von allen vorgestellten Methoden zur Videosegmentierung wird, wie in [1] und [3] beschrieben, häufig die Audiospur zur exakten Segmentierung genutzt. Der Grund dafür ist die Einzigartigkeit und Stabilität der enthaltenen akkustischen Informationen während der gesamten Abspieldauer.

2.1.3 Feature-Extraktion

Nach der Segmentierung folgt der ausschlaggebende Schritt, der den stärksten Einfluss auf die Effektivität des IR-Systems hat. Aus den segmentierten Medienobjekten werden inhaltstragende Eigenschaften extrahiert und in Form von Feature-Werten gespeichert, welche bei Suchanfragen für die Ähnlichkeitsbestimmung herangezogen werden. Dies geschieht in der Regel automatisiert mit auf den jeweiligen Medientyp abgestimmten Algorithmen. Jedoch handelt es sich bei den gewonnenen Werten in der Regel nur um Low-Level-Feature-Werte, da diese den Inhalt meist nur auf einer sehr oberflächlichen Ebene beschreiben und nicht zwangsläufig objektive Informationen enthalten. Die semantische Ausdruckskraft der Feature-Werte kann hier teils stark von der vom Menschen wahrgenommenen Ausdruckskraft abweichen. Die dabei entstehende, sogenannte semantische Lücke stellt eines der größten Probleme der automatischen Feature-Extraktion dar. Der Vorgang der Feature-Extraktion gliedert sich in zwei Schritte, an die unterschiedliche Anforderungen gestellt werden.

Zunächst erfolgt die **Feature-Erkennung**, bei der die bereits zur Segmentierung herangezogenen Eigenschaften des Medienobjektes erfasst werden. Für diesen Schritt ist die Adäquatheit von großer Bedeutung, dies bedeutet, dass sich der Inhalt des Medienobjekts möglichst angemessen in den erkannten Werten widerspiegelt. Zudem ist die Effizienz des Erkennungsprozesses und das Herausfiltern von unbedeutenden

Informationen wichtig. Unbedeutend sind bspw. Invarianzen, die lediglich eine Störung darstellen und das Ergebnis verfälschen.

Beim nächsten Schritt, der **Feature-Aufbereitung**, ist es wichtig das vor allem nur wenige, besonders aussagekräftige Feature-Werte zur weiteren Verarbeitung verwendet werden. Zudem sollten Abhängigkeiten zwischen einzelnen Feature-Werten zusammen mit vernachlässigbaren Werten entfernt werden.

Je nach Medientyp gestaltet sich der Vorgang der Feature-Extraktion äußerst schwierig. Die gewonnenen Low-Level-Feature-Werte führen in der Ergebnismenge einer Suchanfrage zu einem erhöhten Anteil an unrelevanten Informationen. Eine High-Level-Semantik kann bislang nur durch manuelles Hinzufügen von Feature-Werten erreicht werden. Dieser Prozess ist jedoch sehr zeit- und kostenintensiv. Allerdings gibt es um dem entgegenzuwirken einige Ansätze, die die automatisierte Feature-Extraktion um Methoden erweitern, die helfen den Inhalt eines Medienobjektes genauer zu erfassen. Wie in [2] und [3] beschrieben, können mit Textannotationen höhere Treffergenauigkeiten erzielt werden. Dazu wird eine Testmenge an Medienobjekten manuell bearbeitet, damit daraus Verhältnisse von Feature-Werten zu assoziierten Begriffen abgeleitet werden können. Daraus lassen sich Wahrscheinlichkeitsfunktionen herleiten die beschreiben, wie wahrscheinlich es ist, dass ein bestimmtes Wort mit einem Medienobjekt assoziiert wird, sofern zuvor ein bestimmter Feature-Wert extrahiert wurde. Wurde bspw. aus Abbildung 2 ein Feature-Wert extrahiert, der eine größere orangene Fläche mit schwarzen Streifen beschreibt, ist die Wahrscheinlichkeit größer, dass das Bild zutreffend mit dem Begriff „Tiger“ annotiert werden kann.

2.2 Information-Retrieval

Um die gewonnenen Werte nutzen zu können, müssen clientseitig Suchfunktionaliäten vorhanden sein, die eine effiziente Interaktion mit der Datenbank ermöglichen. Die Umsetzung der Suchfunktion und die Strukturierung der Kommunikation zwischen Nutzer und Datenbank, hängt dabei direkt davon ab, über welches Endgerät der Nutzer auf die Datenbank zugreift. Die Kommunikation über ein mobiles Endgerät findet in der Regel über eine Client-Server Architektur statt.

Abbildung 4 veranschaulicht den gesamten IR-Prozess, sowohl auf Seite des Clients als auch auf Server Seite. Dabei sind, wie in der Abbildung veranschaulicht, sowohl vor Absenden der Suchanfrage als auch nach Empfang des Suchergebnisses je nach Endgerät Konvertierungen der Mediendaten in ein geeignetes Format notwendig. Zu Beginn einer Suchanfrage hat der Nutzer die Möglichkeit, unscharf formulierte Anfragen zu stellen. Dies ist möglich, da wie in Abbildung 4 dargestellt, auch auf der Seite des Anfragers zunächst die Suchanfrage auf Feature-Werte untersucht werden muss. Dieser Vorgang ist nötig, um die Anfrage mit den gesuchten Daten vergleichbar zu machen. Der Vergleich von Suchanfrage mit den in der Datenbank

vorhandenen Medienobjekten erfolgt durch eine Gegenüberstellung ihrer Feature-Werte. Diese liegen in der Regel in Listen vor, die u.a. Punkte, Binärdaten, Intervalle oder Sequenzen enthalten können. Um einen effizienten Vergleich durchzuführen, werden daher die Feature-Werte als Feature-Vektoren aufgefasst. Die Ähnlichkeit wird daraufhin über die Distanz einzelner Vektoren zueinander berechnet. Je geringer die Distanz, desto relevanter ist das zu der Anfrage gefundene Medienobjekt.

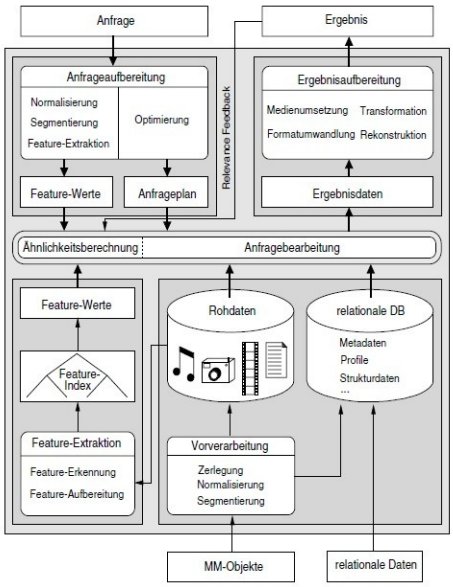

Abbildung 4: Schema eines IR-Prozesses (Quelle: [5])

In Abbildung 5 ist der Vergleich zweier Feature-Werte in Form von Bitvektoren dargestellt. Dabei wird der in der Anfrage übergebene Feature-Wert mit dem korrespondierenden Wert eines Dokumentes der Datenbank anhand bestimmter Positionen ihrer Bitfolge miteinander verglichen. Im angegebenen Beispiel wurde ein Dokument gefunden, dessen Feature-Wert mit dem der Anfrage übereinstimmt. Die eingerahmten Bereiche werden zusätzlich dazu verwendet einen Rankingwert zu erstellen. Im Beispiel stimmen in beiden Bereichen die Bitfolgen bis auf eine Position überein. Das bedeutet, dass es sich um keinen exakten Treffer zur Suchanfrage handelt, das gefundene Dokument jedoch trotzdem von hoher Relevanz ist. Nachdem die Relevanz aller in Frage kommender Objekte bewertet ist, wird je nachdem welche zusätzlichen Angaben der Nutzer bei der Suche angegeben hat, das Ergebnis aufbereitet und nach Relevanz sortiert ausgegeben. Sollte die Ergebnisliste nicht den Wünschen des Nutzers entsprechen, kann die Anfrage beliebig verfeinert und ein neuer Suchvorgang gestartet werden.

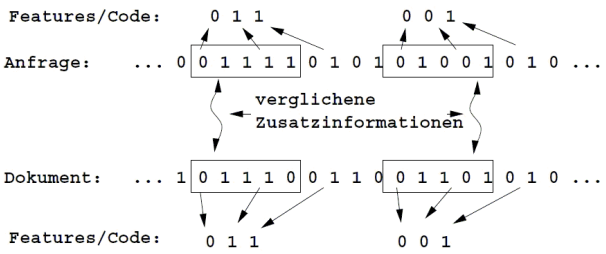

Abbildung 5: Vergleich zweier Feature-Werte (Quelle: [6])

3 Einordnung des Cross-Media-Retrieval

Das CMR stellt eine Erweiterung des IR dar. Im Fokus steht dabei der Ausbau der Suchfunktionalitäten auf der Seite des Clients. Während sich das klassische IR vorwiegend auf die Bearbeitung von textuellen Suchanfragen konzentriert, wird es im CMR ermöglicht jeglichen Medientyp als Suchparameter zu verwenden. Aus diesem Bereich gibt es zwar rechnergestützte Modelle, jedoch beschränken sich die meisten von ihnen auf Texteingaben. Computer verfügen zwar über die nötigen Eingabegeräte für Audio, Video und Bildformate, jedoch führt der Trend immer mehr hin zur mobilen Informationsbeschaffung. Die Forschung fokussiert sich dabei zunehmend auf die Entwicklung von Applikationen für Smartphones. Diese verfügen über die Möglichkeiten alle Medientypen problemlos über die vorhandenen Eingabegeräte wie Kamera, Tastatur oder Mikrofon zu erfassen und für Suchanfragen zu nutzen. Erfolgreiche Applikationen zeichnen sich in diesem Bereich durch die Möglichkeit aus, Anfragen über verschiedenartige Medientypen zu ermöglichen und diese angemessen schnell auszuwerten. In [1] wird dabei konkret auf die Umsetzung von Retrieval Techniken über die Eingabe von Audio- bzw. Videoinhalten zum IR auf Multimediadatenbanken eingegangen. In Abbildung 4 wird die dabei angestellte Gegenüberstellung von einer Video- zu einer Audioanfrage verdeutlicht. Ein Film

Abbildung 6: Gegenüberstellung der Resultate von Video- (A) und Audioanfrage (B) zu einem Helikoptervideo (Quelle: [2])

über Helikopter (*Thumbnail*[4] oben links) wird dazu verwendet, um nach Videos mit ähnlichen Inhalten zu suchen. Dabei wurde bei Suchanfrage A anhand der Videoinformation gesucht und in B die Audiospur zur Suche genutzt. Wie in der Abbildung zu erkennen und in 2.1.2 bereits erwähnt, liefert der Retrieval Prozess anhand der Audioinformation Werte, die der tatsächlichen Semantik eher entsprechen als die Werte der Videoinformation.

Die Möglichkeit zu schaffen, auch komplexe Medienobjekte als Suchparameter zu verwenden, stellt CMR Systeme vor einige Herausforderungen bei der Umsetzung der Client-Server-Architektur. Die zum Teil sehr laufzeitaufwändige Extraktion der

4 Thumbnail: Bildvorschau

Feature-Werte muss echtzeit auf dem Client durchgeführt werden. Dadurch werden erhöhte Anforderungen an die Endgeräte gestellt, die nicht nur in der Lage sein müssen Multimediadaten über Eingabegeräte wie Kamera oder Mikrofon zu generieren, sondern sie müssen auch über die nötige Rechenleistung verfügen, um Suchanfragen in einer angemessenen Zeit abarbeiten zu können. Besonders im Bereich der mobilen Endgeräte, die verstärkt in den Vordergrund bei der Entwicklung von CMR Systemen rücken, stellen die erhöhten Anforderungen ein Problem dar. Hinzu kommt, das mobile Endgeräte nur eine sehr limitierte Anzahl von Medienformaten unterstützen. Aus diesem Grund sind bei der Kommunikation mit dem Server Konvertierungen notwendig, die die Dauer der Anfragebearbeitung erhöhen. Jedoch gibt es in diesem Bereich trotzdem einige Beispiele, die passende Lösungen für die aufgeführten Probleme bereitstellen, welche im folgenden Abschnitt kurz beschrieben werden.

4 Praxisbeispiele von Anwendungen

Die Verbreitung von CMR Systemen hat dank der schnellen technischen Entwicklung von Computern und mobilen Endgeräten stetig zugenommen. Ein Großteil davon wird aktuell von Forschungsgruppen bereitgestellt, jedoch gibt es bereits einige kommerzielle Applikationen. Durch die Möglichkeit jeden beliebigen Medientyp für eine Suchanfrage zu verwenden, wird es dem Nutzer erspart persönlich wahrgenommene Eigenschaften eines Medienobjektes in Form von Text verbalisieren zu müssen und ermöglicht dadurch die direkte und komfortable Suche nach gewünschten Informationen. Im Folgenden werden für unterschiedliche Bereiche repräsentative Beispiele genannt, in denen CMR Funktionalitäten bereits Anwendung finden. Bei der aufgeführten Auswahl werden nur Beispiele für Systeme genannt, bei denen auch eine semantische Interpretation der Multimediaobjekte stattfindet.

4.1 Rechnergestütztes CMR

Im rein rechnergestützten Bereich gibt es wegen den in Kapitel 3 Abschnitt 1 bereits genannten Gründen nur sehr wenige Anwendungen. Ein Beispiel was hier zu nennen wäre ist **SoundFisher**[5]. Dabei handelt es sich um ein Audiotool, das anhand einer vorgegebenen Audiodatei eine Ähnlichkeitssuche durchführt und ähnlich klingende Songs zur Auswahl stellt.

5 http://www.soundfisher.com/

4.2 Webbasiertes CMR

Ein sehr interessantes Beispiel für den Bereich webbasiertes CMR ist ein Forschungsprojekt der Fachhochschule für Technik und Wirtschaft Berlin namens **pixolution**[6]. Die webbasierte Anwendung ermöglicht die semantische Suche nach Bildern und nutzt dabei große Bilddatenbanken von u.a. Yahoo und Flickr. Die semantische Bildsuche basiert bei der Anwendung auf der sogenannten Relevance Feedback Methode. Dabei kann zunächst eine Liste an Bildern ausgegeben werden, die einem bestimmten Suchbegriff entsprechen. Nachdem man ein Bild auswählt, was am ehesten dem gewünschten Suchergebnis entspricht, kann eine erneute Ergebnisfilterung durchgeführt werden, bei der die Feature-Werte des markierten Bildes berücksichtigt werden, um weitere möglichst relevante Bilder anzeigen zu lassen.

4.3 Mobiles CMR

Google Goggles[7] ist eine neuartige Smartphone Applikation und vereint unterschiedlichste CMR Funktionalitäten. Eine davon ermöglicht, über die Kamera des Smartphones visuelle Suchanfragen zu stellen. Dabei wird ein Bild dazu verwendet, das anvisierte Objekt zu identifizieren und für den Nutzer, bei erfolgreicher Identifikation, automatisch in Google eine Suchanfrage zu stellen. Es wird aber noch zusätzlich ermöglicht, durch die Übertragung von GPS Koordinaten Videoaufnahmen echtzeit mit Informationen anzureichern. Durch die Kommunikation mit einer der größten Datenbanken der Welt, steht jederzeit eine große Menge an potentiell relevanten Informationen zur Verfügung.

Shazam[8] ermöglicht es, über eine QBE (query-by-example) Operation einen Auszug eines Songs dazu zu nutzen, sich zu ihm Zusatzinformationen ausgeben zu lassen. Dazu gehören Information zum Interpreten und falls vorhanden kann zusätzlich das passende Video zum Song aufgerufen werden. Es besteht zudem direkt die Möglichkeit den Song über die Anzeige der Ergebnisse online zu erwerben. In diesem Fall sind sogar mehrere Medientypen Teil der Ausgabe.

5 Zusammenfassung

Diese Ausarbeitung hatte zum Ziel, einen Überblick über das CMR zu verschaffen. Dazu wurde zunächst in die allgemeinere Thematik des IR eingeführt und grundlegende Konzepte zur semantischen Suche innerhalb einer Multimediadatenbank

6 http://www.pixolution.de/
7 http://www.google.com/mobile/goggles/
8 http://www.shazam.com/

erläutert. Daraufhin wurde das CMR in den Kontext eingeordnet und abschließend einige Beispiele aufgeführt, die bereits CMR Funktionalitäten umsetzen.

Festzustellen war, dass es sich um ein sehr intuitives Suchverfahren handelt, das durch die Interpretation des Inhalts eines Medienobjekts Nutzern einen erhöhten Anteil an relevanten Informationen liefern kann. Auf der anderen Seite wurden Probleme und mögliche Schwachstellen des Verfahrens aufgeführt, die durch die Erweiterung der Fähigkeiten des Suchverfahrens auf Seite des Clients entstehen, im besonderen Hinblick auf die Entwicklung von mobilen Applikationen.

Die Entwicklung von CMR Anwendungen für die breite Masse befindet sich noch im Anfangsstadium. Immer mehr Hardware-Endgeräte werden intuitiver gestaltet, um den Benutzer die Interaktion mit dem System zu erleichtern. Aus diesem Grund werden sich auch in Zukunft immer mehr CMR Anwendungen etablieren und Nutzern die Suche nach gewünschten multimedialen Informationen erleichtern. Bereits jetzt haben sich, wie in Kapitel 4 aufgeführt einige Systeme durchgesetzt, die audio- und bildbasierte Suchanfragen ermöglichen. Vorstellbar ist, dass vor allem im Bereich von videogestützten Suchanfragen in Zukunft neue Anwendungen entworfen werden. Dabei sind Forschungsgebiete wie bspw. die *Augmented Reality*[9] von Interesse, bei denen während der Videoaufnahme der Bildschirm mit Informationen angereichert wird. Dies erfolgt bei Applikationen wie bspw. Google Goggles momentan über GPS Koordinaten und einen Kompass. Es ist aber vorstellbar, dass aus dem Bereich CMR Funktionen wie die Feature-Extraktion aus Videosequenzen dynamisch umgesetzt werden, um die Informationsbeschaffung zu optimieren.

6 Literaturverzeichnis

[1] I. Ahmad, F. A. Cheikh, S. Kiranyaz, and M. Gabbouj.
„*Audio-based queries for video retrieval over java enabled mobile devices*“.
Multimedia on Mobile Devices II, USA, 2006.

[2] J. Jeon, V. Lavrenko, and R. Manmatha,
„*Automatic image annotation and retrieval using cross-media relevance models”.*
26th Intl. ACM SIGIR Conf., Toronto, Canada, 2003.

[3] Deschacht, K. Moens, M.-F.,
“Finding the Best Picture: Cross-Media Retrieval of Content“.
NUMB 4956, Germany, 2008

9 Augmented Reality: Erweiterte Realität

[4] A. Fujii, K. Itou, T. Akiba, T. Ishikawa,
"A cross-media retrieval system for lecture videos".
Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003), Geneva, Switzerland, 2003

[5] I.Schmitt,
"Retrieval in Multimedia-Datenbanksystemen".
Datenbank-Spektrum, pp. 28-35, Germany, 4/ 2002

[6] F. Kurth,
"Beiträge zum effizienten Multimediaretrieval".
Friedrich-Wilhelms-Universität Bonn, Germany, 2004

[7] M. Höynck
"Videosegmentierung auf Basis von MPEG-7 Deskriptoren"
RWTH Aachen, Germany, 2007