UNIVERSITÄT MANNHEIM

Multimodale Suche II
Verfahren und Systeme

Seminararbeit

eingereicht im: Juni 2006

von: Wolfgang Steitz
geboren am 09. Nov 1981
in Bad-Kreuznach

Matrikelnummer: 933793

Universität Mannheim
Lehrstuhl für Praktische Informatik 4
D – 68131 Mannheim
Internet: http://www.informatik.uni-mannheim.de/pi4

Inhaltsverzeichnis

Abbildungsverzeichnis **iii**

Abkürzungsverzeichnis **iv**

1 Motivation und [Please insert into preamble]berblick **1**
1.1 Motivation 1
1.2 Überblick 1

2 Multimodale Suche **2**
2.1 Multimodalität 2
2.2 Multimodale Suche 2

3 Verfahren **4**
3.1 Multimodal Fusion 4
3.2 Relevance Feedback 4

4 Systeme **8**
4.1 Photo-to-Search 8
4.1.1 Identische Bilder Suche 8
4.1.2 Hybride Bildersuche 10
4.1.3 Extraktion von Schlüsselbegriffen 10
4.1.4 Ergebnispräsentation 11
4.2 Mutual Relevance Feedback 11
4.3 iView 13
4.4 QBIC 14

5 Fazit **15**

Literaturverzeichnis **16**

Abbildungsverzeichnis

2.1 Multimodale Anfrage 3

3.1 Multimodal Fusion 5
3.2 Relevance Feedback Kreislauf 6

4.1 Aufbau des Photo-to-Search Systems 9
4.2 Erstellen der Hash-Signatur für die Identische Bilder Suche . . . 9
4.3 Screenshot Mutual Relevance Feedback 12
4.4 Screenshot Suchergebnisse 13

Abkürzungsverzeichnis

CBIR	Content-based image retrieval
DVL	Digital Video Library
OCR	Opticalt Character Recognition
QBIC	Query By Image Content
RF	Relevance Feedback
SVM	Support Vector Machine
WWW	World Wide Web

1 Motivation und Überblick

1.1 Motivation

Bei den heute verwendeten Suchmaschinen werden fast ausschließlich textbasierte Suchanfragen verwendet, dabei werden multimediale Dokumente häufig zu ungenau beschrieben. Durch multimodale Anfragen, also durch die Kombination von Text und Bilder und / oder Videos innerhalb einer Suchanfrage, können häufig bessere Ergebnisse erziehlt werden. 'Ein Bild sagt mehr als tausend Worte.' Der Nutzer kann mit Hilfe eines Bildes seine Intention bei der Suche besser ausdrücken, dadurch kann das Suchverfahren bessere Ergebnisse liefern. In dieser Arbeit sollen Verfahren und Systeme vorgestellt werden, die solche multimodalen Anfragen unterstützen.

1.2 Überblick

Die Arbeit ist folgendermaßen aufgebaut: Nach der Einführung in Kapitel 1 werden zunächst in Kapitel 2 die Begriffe multimodal und multimodale Suche kurz erklärt. Danach werden in Kapitel 3 einige Verfahren vorgestellt, die häufig bei Systemen mit multimodaler Suche zum Einsatz kommen. In Kapitel 4 werden dann einige ausgewählte Systeme vorgestellt. Schliesslich folgt in Kapitel 5 das Fazit.

2 Multimodale Suche

2.1 Multimodalität

Aus Systemsicht bezeichnet der Begriff Multimodalität die Fähigkeit eines Systems zur Kommunikation mit dem Nutzer über verschiedene Kommunikationskanäle. (Janko Calic, 2005) Betrachtet man multimediale Dokumente, bezeichnet Multimodalität die verschiedenen Kommunikationskanäle durch die der Nutzer den Inhalt wahrnimmt (Amir u. a., 2005) Denkt man an Webseiten sind die verschiedenen Modalitäten unter anderem der Text, die Struktur, Metadaten, Links, die Zahl der Webseiten die auf sie verlinken usw. Bei Bildern hat man die visuellen Merkmale auf niedriger Ebene (z.B. Farbhistogramme), Metadaten wie Größe, Keywords, Format, und bei WWW Bildern auch umgebende Texte und Hyperlinks. Verschiedene Modalitäten von Videos sind zB. die einzelnen Szenen, die visuellen Merkmale der einzelnen Bilder des Streams, die Untertitel, der Inhalt. Natürlich kann man sich noch viele weitere Modalitäten überlegen.(Kennedy u. a., 2005)

2.2 Multimodale Suche

Bei der Multimodalen Suche wird sich auf die Kommunikation zwischem dem User und dem System fokusiert.

Von Multimodaler Suche spricht man zum einen wenn mehrere Modalitäten der Dokumente durchsucht werden, zum anderen wenn die Suchanfrage selbst multimodal ist. In dieser Arbeit liegt der Schwerpunkt auf Verfahren und Systemen mit multimodalen Anfragen. Die Suchanfrage besteht hier nicht nur aus Text, wie bei den heute üblichen Suchverfahren, sondern es besteht die Möglichkeit Bilder, Zeichnungen oder sogar Videos in die Suchanfrage mit einzubeziehen. Eine Beispiel für solch eine multimodale Anfrage ist in Abbildung 2.2 zu sehen.

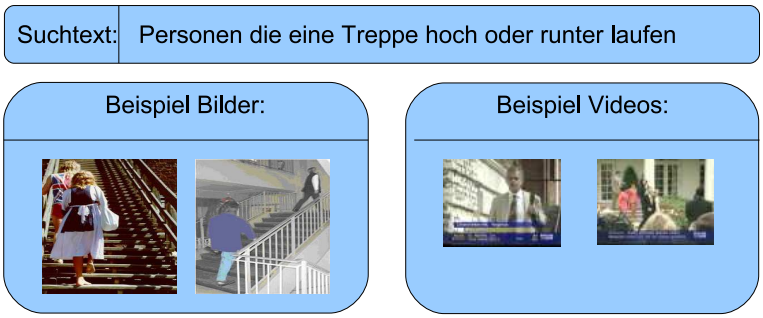

Abbildung 2.1: Beispiel einer multimodalen Anfrage an eine Video Datenbank, mit Text, Bildern und Video Shots.

3 Verfahren

In diesem Kapitel werden zwei Verfahren vorgestellt die in Systemen mit multimodaler Suche häufig zur Anwendung kommen.

3.1 Multimodal Fusion

Bei diesem Ansatz wird im ersten Schritt pro Modalität der Suchanfrage eine unimodale Suche gestartet. Angenommen die Suchanfrage besteht aus Text und einem Bild, dann müssen zwei unimodale Suchen ausgeführt werden. Die eine über eine normale textbasierte Suche, die andre z.B. über CBIR, also einer Suche nach Bildern mit ähnlichem Inhalt. Im zweiten Schritt müssen nun die Resultate der unimodalen Suchen kombiniert werden um die relevanten Suchergebnisse zu erhalten. Diese Fusion der Resultate lässt sich abhängig und unabhängig von der Suche ausführen. Bei dem unabhängigen Modell, wird immer die gleiche Fusionsstrategie verwendet. Dies führt zwar zur einer hohen Performance, allerdings nicht unbedingt zu den besten Ergebnissen, da die verschiedenen Modalitäten der Suche nicht immer gleich wichtig sind. Anders bei der abhängigen Suche, hier wird eine Fusionsstrategie abhängig von den Suchresultaten angewendet. Allerdings ist es nicht möglich für jede Art der Suchanfrage eine best mögliche Fusionsstrategie festzulegen. Dieses Problem wird durch 'Such-Klassen' gelöst. Hier wird in jeder Such-Klasse die gleiche Fusionsstrategie verwendet. Als Klassen werden unterschiedliche Unterteilungen vorgeschlagen. Zum Beispiel: Personen, Sport, Finanzen etc. (Kennedy u. a., 2005). Ein Beispielsystem ist in Abbildung 3.1 zu sehen. Hier wird eine Video Datenbank mit einer multimodalen Anfrage durchsucht. Für jede Modalität wird der entsprechende Index durchsucht. Aus diesen einzelnen Ergebnislisten wird dann durch gewichtete Fusion eine Gesamtergebnisliste erstellt.

3.2 Relevance Feedback

Ein weiteres Verfahren was im Bereich der multimodalen Suche häufig genutzt wird, ist das Relevance Feedback. Hier wird versucht die Suchergebnisse durch Feedback vom Nutzer schrittweise zu verbessern. Im Bereich der textbasierten Suche wird daran schon seit mehr als 40 Jahren geforscht und spielt eine viel wichtigere Rolle als im Multimedia Bereich. Bei einem Relevance Feedback Verfahren zur textbasierten Suche wird häufig versucht die Gewichte der Suchbegriffe anhand des Feedbacks anzupassen. Bei positivem Feedback erhöhen sich die Gewichte bzw werden erniedrigt bei negativem Feedback. Anhand der Häufigkeit mit der Worte in positiv bzw negativ bewerteten Dokumenten auftreten, werden die Gewichte neu berechnet. (Amir u. a., 2005)

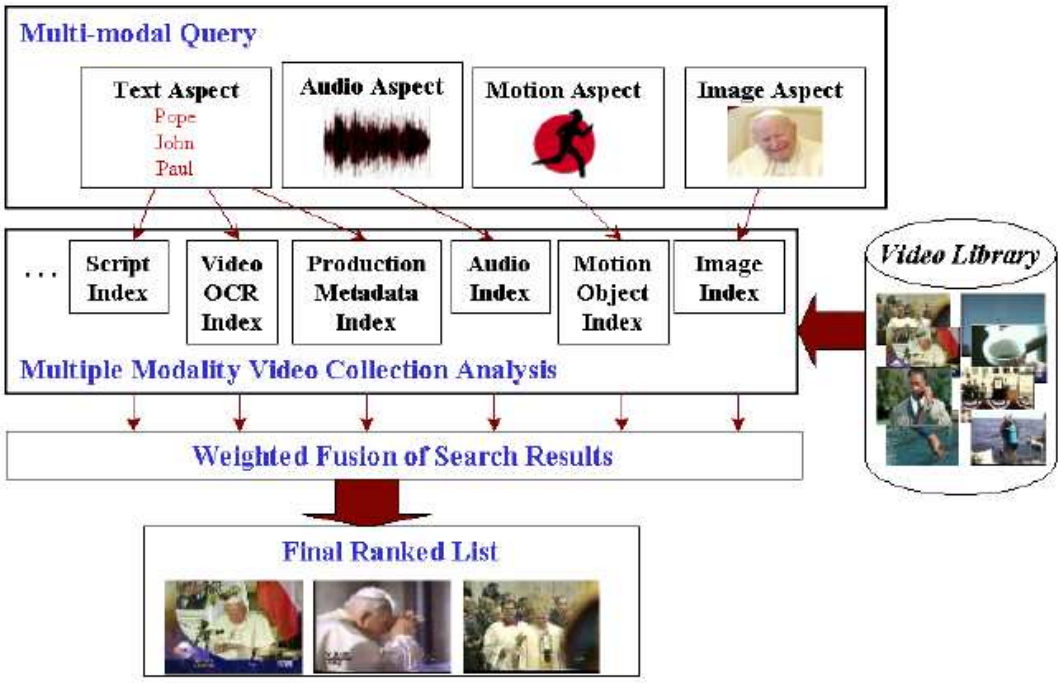

Abbildung 3.1: Multimodal Fusion

Mit Hilfe von Relevance Feedback kann man zum einen die Rangliste der Suchergebnisse optimieren, zum anderen aber auch die Suche selbst verändern um bessere Ergebnisse zu erzielen. Ein Relevance Feedback Verfahren zur Bildersuche läuft folgendermassen ab: Der Nutzer startet die Suche mit einer textbasierten Anfrage oder auch mit Hilfe eines Beispielbildes. Die Suchmaschine präsentiert dem Nutzer daraufhin das Ergebnis der textbasierten Suche bzw dem Beispielbild ähnelnden Bilder. Jetzt bekommt der Nutzer die Möglichkeit Feedback bezüglich der Suchergebnisse abzugeben, er bewertet einige der gefundenen Bilder als 'relevant' oder 'irrelevant' bzw positiv oder negativ. Von diesem Feedback versucht die Suchmaschine die visuellen Merkmale der Bilder zu lernen und liefert dem Nutzer verbesserte Suchergebnisse. Dieser Prozess wird mehrere Runden wiederholt, bis der Nutzer schliesslich mit dem Ergebnis der Suche zufrieden ist. Der Suchmechanismus sollte dabei versuchen die Interaktion zwischen Nutzer und System zu minimieren ((M. Ferecatu u. Boujemaa, 2004)). Dieser Kreislauf des Relevance Feedbacks ist in Abbildung 3.2 zu sehen.

Ein Relevance Feedback Verfahren besteht aus 2 Komponenten: der Learner und der Selector. Die Hauptaufgabe des **Learners** ist es das Suchziel des Nutzers zu erahnen und somit für jedes Bild in der Datenbank entscheiden zu können ob es relevant für den Nutzer sein könnte. Dazu nutzt er die Trainingsdaten, also die markierten Bilder aus den verschiedenen Feedback Runden, und manchmal auch Informationen aus vorhergegangenen Suchen. Diese Aufgabe des Learners ist aus verschiedenen Gründen nicht sehr einfach. Zum einen ist die Menge der

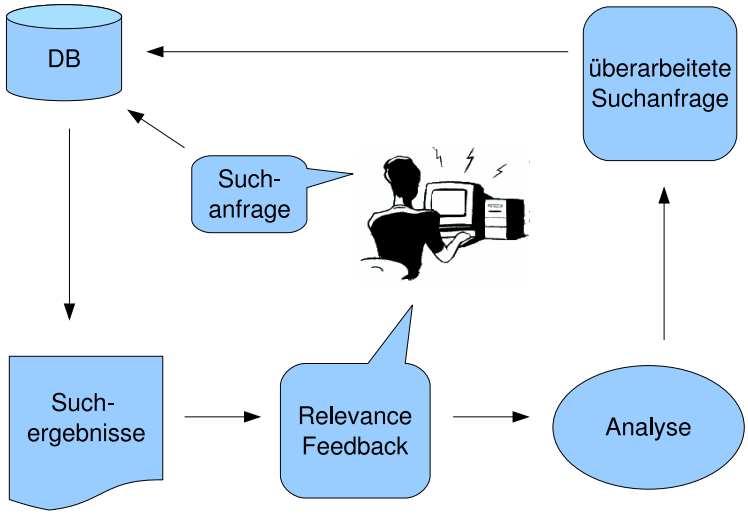

Abbildung 3.2: Relevance Feedback Kreislauf

Trainingsdaten meistens sehr niedrig, dazu kommt dass es üblicherweise mehr positive Beispiele (markiert als relevant) als negative Beispiele (markiert als irrelevant) gibt. Aus diesen wenigen Informationen ist es sehr schwer das häufig sehr komplexe Suchziel des Nutzers zu erahnen. Außerdem muss das Lernen möglichst schnell ablaufen, damit nur wenige Runden benötigt werden und der Nutzer nicht zu lange warten muss. (M. Ferecatu u. Boujemaa, 2004)

Die zweite Komponente, der **Selector**, ist dafür zuständig die Bilder auszuwählen die dem Nutzer präsentiert werden sollen. Dabei gibt es zwei gegensätzliche Ziele.

1. Dem Nutzer soll der aktuelle Wissensstand des Learners gezeigt werden, also werden dem Nutzer so viel als möglich relevante Bilder gezeigt
2. Der Wissentransfer zwischen Nutzer und System soll maximiert werden. Dem Nutzer sollen also möglichst viele Informationen entlockt werden, um relevante und irrelevante Bilder besser unterscheiden zu können.

Die meisten derzeitigen RF Ansätze konzentrieren sich auf das erste Ziel und versuchen dem Nutzer immer möglichst viele relevante Bilder zu präsentieren. Das hat den Vorteile, dass der Nutzer schon sehr schnell an gute Ergebnisse kommt, außerdem ist die Umsetzung einfacher. Allerdings dauert der Lernprozess dadurch länger, es sind mehr Runden nötig um optimale Ergebnisse zu erzielen. Das zweite Ziel des Learners ist wesentlich schwieriger zu erreichen, damit befasst sich das Active Learning auf das hier nicht näher eingegangen werden soll. Um die die beiden gegensätzlichen Ziele zu vereinen werden Hybride Strategien vorgeschlagen. Hier wird in den ersten Runden versucht den Wissentransfer zu maximieren, aber dann in späteren Runden mehr auf das Präsentieren von relevanten Ergebnissen fokussiert. (M. Ferecatu u. Boujemaa, 2004)

Pseudo-Relevance Feedback

Pseudo-Relevance Feedback automatisiert das Verfahren. Der Nutzer muss kein Feedback mehr geben, sondern das System tut das automatisch. Eine Möglichkeit ist es die top Ergebnisse der ersten Suche als die relevantesten anzusehen. Diese Dokumente werden untersucht und für die Suche verwendet. Das Feedback kommt hier also nicht vom Nutzer, sondern von der Suchmaschine selbst. So könnte man z.B. bei einer textbasierten Bildersuche die ersten Ergebnisse verwenden um damit eine multimodale Suche mit Text und Bildern starten und damit möglicherweise das Ergebnis der Suche verbessern.

4 Systeme

In diesem Kapitel werden vier Systeme vorgestellt die multimodale Suche bzw. multimodale Anfragen unterstützen.

4.1 Photo-to-Search

Mobile Endgeräte werden immer häufiger genutzt um das Web zu durchsuchen. Allerdings sind die derzeitigen mobile Suchmaschinen wie Google Mobile, Google SMS oder Yahoo! Mobile rein textbasiert und lassen sich deswegen nicht sehr komfortabel nutzen. Das Photo-to-Search System, ein Projekt von Microsoft Research, versucht die vorhandenen Technologien besser zu nutzen, um die Interaktion mit der Suchmaschine zu vereinfachen und gleichzeitig die Qualität der Suchergebnisse zu verbessern. Heutige Geräten verfügen häufig über eine integrierte Kamera, damit lassen sich multimodale Suchen mit Bildern und Videos realisieren. In zukünftigen Versionen des Systems sollen auch Suchen mit Videos oder Sprache als Eingabe möglich sein. Ein typisches Anwendungsbeispiel: Man sieht ein Filmplakat auf der Straße und möchte mehr darüber erfahren. Also macht man ein Bild davon und startet die Suche. Als Ergebnis erhält man Links zu relevanten Seiten.

Photo-to-Search besteht aus vier Komponenten: Suche nach identischen Bildern (4.1.1), Hybride Bildersuche (4.1.2), Extraktion von Schlüsselbegriffen (4.1.3) und der Ergebnispräsentation (4.1.4). Abbildung 4.1 gibt einen Überblick über das System. Zunächst wird das Modul zur identischen Bildersuche gestartet. Ist die Suche erfolgreich folgen danach Extraktion von Schlüsselbegriffen und schlieslich die Präsentation der Ergebnisse. Ist die identische Bildersuche allerdings nicht erfolgreich, wird die hybride Bildersuche gestartet und danach erst Extraktion von Schlüsselbegriffen und Präsentation. Die einzelnen Module werden im Folgenden kurz beschrieben.

4.1.1 Identische Bilder Suche

Die Suche nach identischen Bildern muss möglichst schnell ablaufen, um das zu garantieren werden Merkmale aller Bilder aus der Datenbank extrahiert und ein Index zur schnellen Suche erstellt. Dazu wird von jedem Bild eine 32 Bit Signatur erstellt. Als Vergleichsmerkmal werden nur die Helligkeitswerte verwendet.

Die Erstellung der Hash Signatur ist in Abbildung 4.1 zu sehen. Zuerst wird das Bild in ein Graustufenbild umgewandelt, da die Sättigung und der Farbton der durch eingebaute Kameras geschossenen Bilder stark abweicht. Danach wird

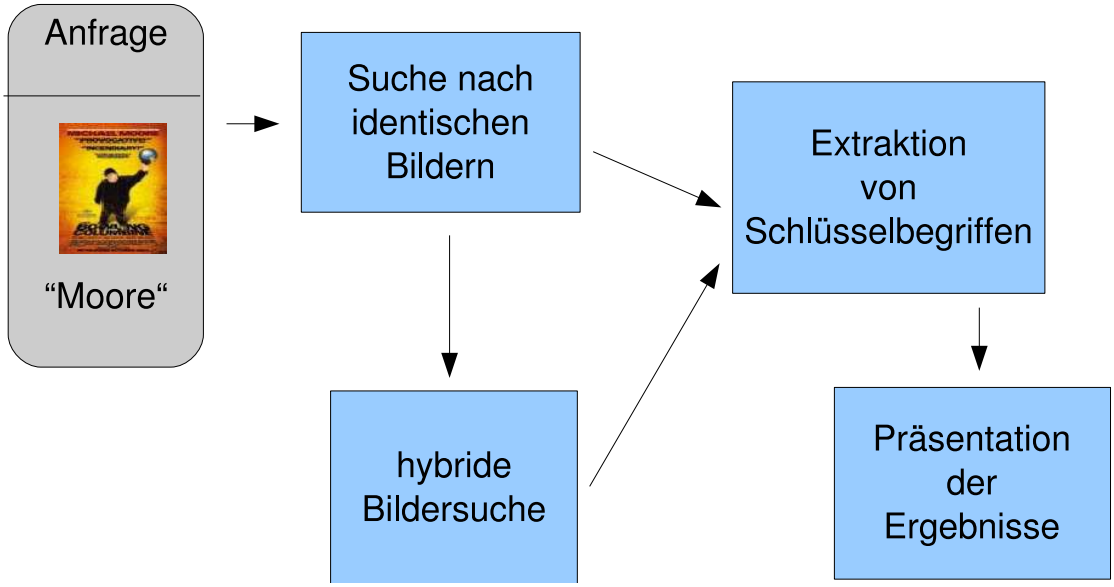

Abbildung 4.1: Aufbau des Photo-to-Search Systems

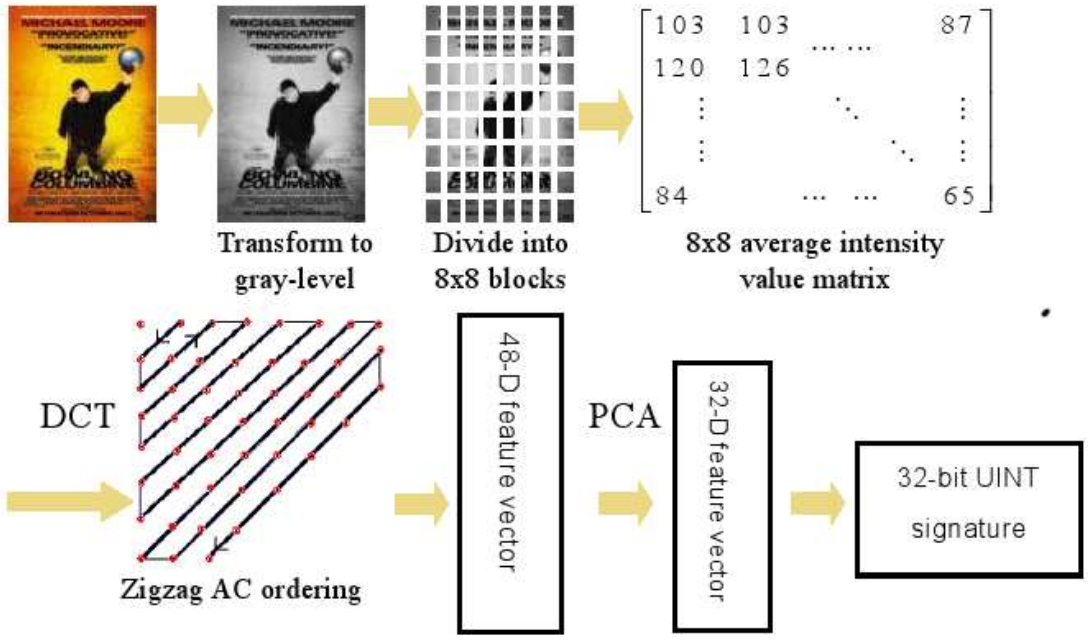

Abbildung 4.2: Erstellen der Hash-Signatur für die Identische Bilder Suche

das Bild in 8x8 Blöcke unterteilt und für jeden dieser Blöcke die durchschnittliche Helligkeit berechnet. Auf diese 8x8 Matrix wird nun eine zweidimensionale diskrete Kosinustransformation (2D DCT) angewendet. Die AC Koeffizienten werden im ZickZack angeordnet, wobei der DC Koeffizient vernachlässigt wird, da er die durchschnittliche Helligkeit im gesamten Bild angibt und diese Varianz zwischen den Bildern soll nicht berücksichtig werden. Die ersten 48 AC Koeffizienten werden in einem Vektor mit 48 Dimensionen gespeichert. Mit Hilfe einer Hauptkomponentenanalyse (PCA, Principal Component Analysis) wird der Vektor in einen Vektor mit 32 Dimensionen verwandelt. Um bei der Suche nicht jeden einzelnen Vektor vergleichen zu müssen, wird der Vektor mittels einer Hashfunktion in eine 32 Bit unsigned Integer Signatur abgebildet. Das jeweilige Bit wird auf 1 gesetzt falls der Wert größer 0 ist, ansonsten auf 0.

Bei einer Suchanfrage wird diese Berechnung ganz genauso durchgeführt. Bilder mit der gleichen Signatur werden als identisch angesehen. Obwohl das Verfahren sehr einfach ist, nur die Helligkeitswerte beachtet werden und als Vergleichsmerkmal nur eine 32 Bit Signatur verwendet wird, liefert das System bemerkenswerte Ergebnisse.

4.1.2 Hybride Bildersuche

Die hybride Bildersuche wird gestartet falls die Suche nach identischen Bildern keine Ergebnisse geliefert hat. Jetzt wird der Text der Suchanfrage genutzt um textuell relevante Bilder zu finden. Der Text der Suchanfrage besteht in der Regel nur aus einem oder zwei Wörtern, da die Eingabe auf mobilen Endgeräten eher lästig ist. Also werden zunächst mittels Wordnet weitere Suchbegriffe gesucht. Diese Suchbegriffe werden nun an eine textbasierte Bildersuche weitergegeben um eine Reihe von textuell relevanten Bildern zu erhalten. Dieses Set von Bildern wird basierend auf dem CBIR Ansatz mit dem Bild aus der Suchanfrage verglichen. Dabei werden nur low-Level Merkmale der Bilder betrachtet. Das Verfahren wählt die Bilder mit der geringsten Distanz zu dem Suchbild als visuell relevanten Bilder aus.

4.1.3 Extraktion von Schlüsselbegriffen

Die Schlüsselbegriffe werden offline extrahiert, dass heißt für jedes Bild in der Datenbank wurde die zugehörige Webseite durchsucht und Schlüsselbegriffe bzw Schlüsselphrasen anhand von strukturellen und statistischen Aspekten identifiziert. Die so gewonnenen Schlüsselbegriffe werden zusammen mit den Bildern gespeichert. Falls die Hybride Bildersuche benutzt wurde, also ein ganzen Set von relevanten Bildern vorhanden ist, müssen noch die globalen Schlüsselbegriffe gefunden werden. Dafür werden die Begriffe verwendet die am häufigsten in dem Bilderset vorkommen.

4.1.4 Ergebnispräsentation

Diese Komponente ist dafür zuständig dem Nutzer das Ergebnis der Suche zu präsentieren. Da entweder ein identisches Bild mit Hilfe der Identische Bilder Suche (4.1.1) oder eine ganze Menge von Bilder mit der Hybriden Bildersuche (4.1.2), gibt es zwei verschiedene Versionen der Präsentation.

- Die Ergebnisseite der Identischen Bilder Suche besteht aus den gefundenen Bildern und den zugehörigen Schlüsselwörtern bzw Schlüsselphrasen. Die Bilder sind mit den orginal Webseiten verlinkt und die Schlüsselwörter mit Einträgen aus Enzyklopädien oder dem Ergebnis traditioneller Suchmaschinen.

- Bei der Präsentation der Ergebnisse aus der Hybriden Suche werden entweder die gefundenen relevanten Bilder mit Links zu den orginal Seiten gezeigt, oder die gefundenen Schlüsselwörter wieder mit Links zu Enzyklopädien oder dem Ergebnis traditioneller Suchmaschinen.

4.2 Mutual Relevance Feedback

"Mutual Relevance Feedback for Multimodal Query Formulation in Video Retrievalïst ein System von IBM Research, es bietet eine interaktive Suche für digitale Video Datenbanken. Mit Hilfe von multimodalen Anfragen und einer Erweiterung von Relevance Feedback zu beidseitigem Feedback (Mutual Relevance Feedback) wird eine hohe Qualität der Suchergebnisse erreicht.

Zunächst werden die Videos aus der Datenbank in einzelne Shots unterteilt, welchen jeweils 1 - 2 key-frames zugeordnet werden. Nun werden die folgenden sechs Indexe erstellt um das Durchsuchen der Daten zu ermöglichen.

- **CBIR Bilder Index** Der Index wird für Anfragen mit Beispielbildern genutzt. Dafür wird für jedes key-frame wird ein Vektor mit diversen low-level Eigenschaften erstellt.

- **Phonetischer Sprach Index** Hier wird die Stimme in dem Video untersucht.

- **Semantische Konzepte** Die einzelnen Shots werden annotiert und diese Annotationen in MPEG-7 gespeichert.

- **Video OCR** Es werden Texteinblendungen in den Videos gesucht und diese mittels Texterkennung (OCR) in Text umgewandelt.

- **Sprach Erkennung** Es wird eine Spracherkennung (ASR, Automatic Speech Recognition) durchgeführt.

- **Untertitel** Ein Index für die Untertitel wird erstellt.

Abbildung 4.3: Screenshot Mutual Relevance Feedback

Suchanfragen

Die Anfragen werden zwar nur in ein normales Textfeld eingetragen, aber man kann durch einen Suffix angeben welche Modalität durchsucht werden soll. Die Anfrage 'sport AND basketball AND (NBA# OR NCAA#) AND 04.38333' sucht nach dem semantischen Konzept 'Sport', dem Wort 'Basketball' in der Spracherkennung und den Untertiteln, das Akronym 'NBA' im Phonetischem Sprach Index, dem Text 'NCAA' in den Einblendungen und nach Frames die dem Keyframe 04.38333 ähneln. Für jeden Index wird nun eine separate unimodale Suche gestartet. Die Ergebnislisten werden schliesslich zur einer Gesamtergebnissliste zusammengefasst, wobei die einzelnen Shots in Segmenten gruppiert werden. D.h. Shots die aus dem gleichen Video stammen und nahe beeinander liegen, werden zu einem Dokument zusammengefasst.

In Abbildung 4.2 ist ein Screenshot des Systems zu sehen. Oben links befindet sich das Eingabefeld für die Suche, im unteren Teil sind die Ergebnisse der Suche zu sehen und rechts oben befindet sich ein Videoplayer. Nachdem der Nutzer nun seine Suchanfrage eingegeben hat, erhält einer Liste relevanter Dokumente. Diese kann er sich jetzt anschauen und als positiv (relevant) oder negativ (irrelevant) markieren. Zusätzlich kann auch das jeweilige Storyboard des Videos durchsucht werden. Die durch solches Browsen gefundenen Videos können auch als relevant bzw. irrelevant markiert werden.

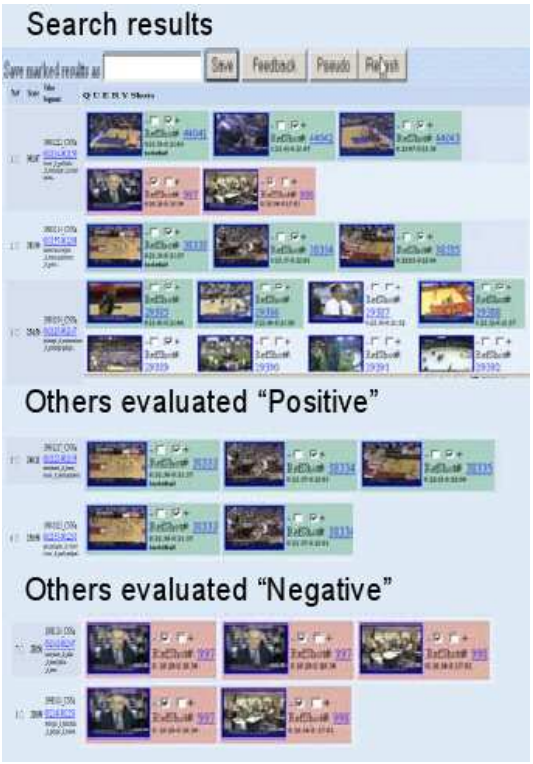

Abbildung 4.4: Screenshot Suchergebnisse Mutual Relevance Feedback

Beidseitiges Feedback

Das Feedback des Nutzers funkioniert über normales Relevance Feedback, wie schon beschrieben. Das System gibt dem Nutzer daraufhin ein visuelles Feedback, siehe Abbildung 4.2. Zusätzlich zu den aktuellen Suchergebnissen werden noch die Videos die in vorherigen Iterationen als relevant bzw irrelevant markiert wurde angezeigt. Dabei werden alle relevanten Videos grün hinterlegt und alle irrelevanten rot. Der Nutzer sieht nun direkt wieviele der grün bzw. rot hinterlegten Videos sich in den aktuellen Suchergebnissen wiederfinden. Sein Ziel ist es nun die Suchanfrage so zu verändern, dass möglichst viele der grün markierten Videos weit oben in seinem Suchergebniss sind und gleichzeitig möglichst wenige der rot hinterlegten Videos sich dort befinden. Das System gibt Vorschläge zur Erweiterung der Suchanfrage um genau das zu erreichen (siehe Abb. 4.2).

4.3 iView

Das iView System ist ein Client für eine multimodale und mehrsprachige digitale Video Bibliothek, welches speziell für Handheld Computer entwickelt wurde. Es werden zwei verschiedene Suchmodi angeboten, zum einen textbasierte Suche, aber auch Suche basierend auf geographischer Position. In diesem Modus kann der Benutzer ein Rechteck auf einer Weltkarte einzeichnen und bekommt dann

Nachrichtensendungen aus der gewählten Region geliefert ((Michael R. Lyu, 2003))

4.4 QBIC

Query By Image Content ist eine kommerzielle Bildersuche und Bestandteil des IBM DB2 Datenbank Systems. Es war eins der ersten Systeme die nicht nur die Anmerkungen zu den Bildern, sondern vor allem den visuellen Inhalt durchsucht haben. Man kann Suchanfragen mit Beispielbildern, Zeichnungen, Keywords oder auch nach bestimmten Farben suchen. Das System wird für die Online Gallerie des Hermitage Museums genutzt. (Janko Calic, 2005) Das System arbeitet mit Vektoren für die verschiedenen Merkmale.

5 Fazit

Chancen und Möglichkeiten

Mit multimodaler Suche ist es möglich die Qualität der Suchergebnisse im Vergleich zur rein textbasierten Suche enorm zu verbessern. Der Intention des Nutzers lässt sich besser erahnen als bei der Suche mit nur wenigen Stichworten. Multimodale Suche ist also ein Weg die 'semantische Lücke' zu schliessen. Außerdem haben wir gesehen, dass man mit multimodaler Suche die Interaktion zwischen Nutzer und System erheblich vereinfachen kann. So muss man bei dem Photo-to-Search System z.B. nicht mehr lästig Suchworte in ein mobiles Endgerät eingeben, sondern kann die eingebaute Kamera nutzen um Suchanfragen mit Bildern zu starten. Mit multimodaler Suche ergibt sich also auch die Möglichkeit andere Technologien zur Suche einzusetzen und somit besser zu nutzen

Herausforderungen

Die größte Herausforderung liegt zur Zeit in der Umsetzung eines solchen Systems. Die Technologie und der Markt sind sicherlich vorhanden.

Bei Systemen die auf Relevance Feedback basieren (zB. Mutual Relevance Feedback) wäre es wichtig bestimmte Prozesse zu beschleunigen bzw. zu automatisieren, um damit den Suchprozess zu beschleunigen. Um bei Mutual Relevance Feedback zu einem guten Ergebnis zu kommen, wird zur Zeit etwa eine Viertel Stunde benötigt. Das ist sicherlich nicht zufriedenstellend. Allerdings ist die Qualität des Suchergebnisses hervorragend, was in manchen Fällen sicherlich wichtiger ist.

Literaturverzeichnis

[Amir u. a. 2005]
AMIR, Arnon ; BERG, Marco ; PERMUTER, Haim: Mutual relevance feedback for multimodal query formulation in video retrieval. In: *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. New York, NY, USA : ACM Press, 2005. – ISBN 1–59593–244–5, S. 17–24

[Apostol Natsev 2005]
APOSTOL NATSEV, Jelena T.: Learning the semantics of multimedia queries and concepts from a small number of examples. In: *ACM Multimedia 2005: 598-607* (2005)

[Ashley u. a. 1995]
ASHLEY, Jonathan ; FLICKNER, Myron ; HAFNER, James ; LEE, Denis ; NIBLACK, Wayne ; PETKOVIC, Dragutin: The query by image content (QBIC) system. In: *SIGMOD '95: Proceedings of the 1995 ACM SIGMOD international conference on Management of data*. New York, NY, USA : ACM Press, 1995. – ISBN 0–89791–731–6, S. 475

[Janko Calic 2005]
JANKO CALIC, Stamatia Dasiopoulou and Yiannis Kompatsiaris: An Overview of Multimodal Video Representation for Semantic Analysis. In: *European Workshop on the Integration of Knowledge, Semantics and Digital Media Technologies (EWIMT 2005), IEE,* (2005), Dezember

[Kennedy u. a. 2005]
KENNEDY, Lyndon S. ; NATSEV, Apostol (. ; CHANG, Shih-Fu: Automatic discovery of query-class-dependent models for multimodal search. In: *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*. New York, NY, USA : ACM Press, 2005. – ISBN 1–59593–044–2, S. 882–891

[M. Ferecatu u. Boujemaa 2004]
M. FERECATU, M. C. ; BOUJEMAA, N.: Relevance feedback for image retrieval: a short survey. In: *in State of the Art in Audiovisual Content-Based Retrieval, Information Universal Access and Interaction, Including Datamodels and Languages, Report of the DELOS2 European Network of Excellence (FP6)* (2004)

[Menglei Jia 2006]
MENGLEI JIA, Xing X.: Photo-to-Search: Using Camera Phones to Inquire of the Surrounding World. In: *The 7th International Conference on Mobile Data Management (MDM'06), Nara, Japan* (2006), Mai

[Michael R. Lyu 2003]
MICHAEL R. LYU, Edward Y.: A wireless handheld multi-modal digital video library client system. In: *Multimedia Information Retrieval 2003: 231-238* (2003)

[Tong u. Chang 2001]
TONG, Simon ; CHANG, Edward: Support vector machine active learning for image retrieval. In: *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*. New York, NY, USA : ACM Press, 2001. – ISBN 1–58113–394–4, S. 107–118

[Wang u. a. 2005]
WANG, Xin-Jing ; MA, Wei-Ying ; ZHANG, Lei ; LI, Xing: Multi-graph enabled active learning for multimodal web image retrieval. In: *MIR '05: Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval*. New York, NY, USA : ACM Press, 2005. – ISBN 1–59593–244–5, S. 65–72

[Wikipedia 2006]
WIKIPEDIA: *Support-Vector-Maschine — Wikipedia, Die freie Enzyklopädie*. `http://de.wikipedia.org/w/index.php?title=Support-Vector-Maschine&oldid=16039121`. Version: 2006. – [Online; Stand 10. Juni 2006]