

Inhaltsbasiertes Multimedia Retrieval: Überblick und Herausforderungen

Inhaltsbasierte Multimedia-Retrieval-Systeme sollen dem Benutzer das Suchen nach Multimediaobjekten in großen Datensammlungen mit Hilfe inhaltsbeschreibender Kriterien ermöglichen. Mit inhaltsbeschreibenden Kriterien sind hierbei in Multimediaobjekten dargestellte Objekte, Aktivitäten und Situationen gemeint, die sich in der Regel erst durch eine menschliche Interpretation erschließen. Wir sind zum Beispiel an Bildern interessiert, die den Magdeburger Dom darstellen, oder suchen eine Videosequenz über einen Surfer auf dem Gardasee. In diesem Beitrag werden in der Einführung zunächst die Hauptunterschiede zwischen Datenbank- und Retrieval-Suche herausgearbeitet. Daran schließt sich eine kurze Beschreibung der prinzipiellen Funktionsweise von Multimedia-Retrieval-Systemen an. Anschließend diskutieren wir kurz heutige Grenzen des inhaltsbasierten Multimedia Retrieval sowie aktuelle Forschungsansätze.

1 Einführung und Grundlagen

Unser Alltag ist geprägt durch eine Flut an digital verfügbaren Multimediadaten, die unter anderem Digitalfotos, Vektorgrafiken, XML-Dokumente, HTML-Dokumente, MPEG-kodierte Filme und Audiodaten (Musik, Sprache, Geräusche) umfassen. Die gegenwärtige Herausforderung besteht darin, diese große Menge an Daten geeignet zu organisieren und für eine effiziente Suche aufzubereiten. Ein besonderes Anliegen ist dabei die Unterstützung einer effektiven inhaltsbasierten Suche nach einzelnen Multimediaobjekten.

Leider sind herkömmliche relationale Datenbanksysteme für die Suche nach Multimediaobjekten wenig geeignet. Dies liegt vor allem daran, dass für die Selektion gewünschter Datenbankobjekte ein einfacher Datenvergleich auf Attributwerten, wie beispielsweise »dargestelltes_Objekt='Surfer mit Surfbrett'«, oft nicht möglich ist. Das grundsätzliche Problem einer *inhaltsbasierten*

(engl. *content-based*) Suche liegt darin begründet, dass die für die Selektion erforderliche Information nicht explizit vorliegt, sondern implizit in den Mediendaten kodiert ist. Die Information »Surfer auf Surfbrett« etwa versteckt sich hinter den Farbwerten der einzelnen Pixel. Für eine inhaltsbasierte Suche ist also eine *Interpretation* erforderlich, um implizit kodierte Informationen zu extrahieren und explizit als so genannte *Feature-Daten* für die Suche zu verwenden. Geeignete Feature-Daten für das Surfbrettbeispiel sind spezielle numerische Formmerkmale und Farbverteilungswerte, um einen Surfer anhand von Form und Farbe erkennen zu können. Da eine automatische Extraktion im Allgemeinen den Inhalt nicht exakt ermitteln kann, ist eine inhaltsbasierte Suche auf der Grundlage der jeweiligen Feature-Daten immer mit einer gewissen Ungenauigkeit bei der Relevanzberechnung verbunden. So ist es in unserem Beispiel leicht möglich, dass statt eines Surfers ein Segelboot gefunden wird.

Bei der Formulierung einer inhaltsbasierten Suchanforderung durch einen Nutzer müssen Feature-Daten spezifiziert werden, auf deren Grundlage dann die ähnlichsten Multimediaobjekte gesucht werden. Es muss also angegeben werden, was einen Surfer mit Surfbrett ausmacht. Diese Spezifikation erfolgt bei einer so genannten *query by example* durch Vorgabe eines oder mehrerer Multimediaobjekte, aus denen die benötigten Feature-Daten für den Vergleich extrahiert werden.

Im Gegensatz zu Datenbanksystemen können Retrieval-Systeme Multimediaobjekte also nicht ausschließlich anhand

von Bedingungen auf Attributen suchen, sondern müssen die durch die Objekte ausgedrückten, aber nur implizit vorhandenen Informationen berücksichtigen. Tabelle 1 fasst die Hauptunterschiede zwischen Datenbank- und Retrieval-Suche zusammen.

Um ein Retrieval-System zur Suche zu erstellen, benötigen wir somit Verfahren, die geeignete Feature-Daten aus den Multimediaobjekten extrahieren, ein Datenmodell, das einen Ähnlichkeitsvergleich zwischen Objekten sowie Objekten und Anfragen ermöglicht und dabei auch die Aggregation von Feature-Daten erlaubt, sowie eine Speicherungsmöglichkeit für Feature-Daten zur effizienten Suche. Im Folgenden beschreiben wir kurz die grundlegenden Konzepte und verweisen jeweils auf aktuelle Arbeiten.

1.1 Feature-Extraktion

Wie bereits oben erwähnt, werden für die inhaltsbasierte Suche Feature-Daten benötigt, die den Inhalt eines Objekts beschreiben. Für Texte sind dies die darin auftretenden Wörter. Bei anderen Medientypen wie etwa Bild-, Audio- und Videodaten ist die Bestimmung geeigneter Features nicht direkt möglich. Man versucht deshalb, Feature-Daten aus den Objekten zu extrahieren, bei denen man vermutet, dass diese den Inhalt des Bildes in abstrahierter Form beschreiben können und somit für einen inhaltsbasierten Vergleich geeignet sind.

Aus Rasterbildern werden üblicherweise Features wie Farbe (meist als Histogramm), Textur, Umrisse von Objekten und die räumliche Lage zwischen Objekten bestimmt. Bei Videos lassen sich Features aus den Tonkanälen (Geräusche, ggf. auch Spracherkennung), den Untertiteln (Text) und den Bilddaten extrahieren. Man könnte dabei theoretisch für jedes einzelne Bild Feature-Daten extrahieren. Da in einzelnen Videosequenzen

Unterscheidungsmerkmal	Datenbanksuche	Retrieval
Suche anhand	Daten	Informationen (Interpretation)
für Auswertung erforderl. Daten	explizit verfügbar	nur implizit verfügbar
Relevanzberechnung der Ergebnisse	exakt	ungenau

Tab. 1: Unterschiede zwischen Datenbank- und Retrieval-Suche

aber häufig eine Reihe sehr ähnlicher Bilder auftritt, würde dies zu sehr großen und stark redundanten Datenmengen führen. Deshalb wird ein Video in der Regel zunächst in einzelne Szenen (*shots*) segmentiert, und für jede Szene wird anschließend ein typisches Bild (*keyframe*) extrahiert und indiziert.

Bei den zeitabhängigen Medientypen Audio und Video werden häufig zeitabhängige Feature-Daten abgeleitet. Teilweise werden dafür Signalverarbeitungsalgorithmen eingesetzt. Prinzipiell müssen für jeden einzelnen Medientyp und für jede konkret geforderte Ähnlichkeitsunterstützung spezifische Features gefunden werden.

Ein grundlegendes Konzept bei der Feature-Extraktion ist die Bestimmung von Merkmalen, die bezüglich bestimmter Operationen *invariant* sind. So sollte es bei der Suche nach Surfern in Bildern in der Regel unwichtig sein, wie groß der Surfer abgebildet ist. Man fordert für die extrahierten Merkmale somit eine Größeninvarianz. Dies kann bei Objektumrissen zum Beispiel durch Verwendung relativer statt absoluter Abstände erreicht werden. Ein weiteres Beispiel können Störgeräusche in Audioobjekten sein: Die extrahierten Audio-Features sollten invariant zu diesen Störgeräuschen sein. Ähnliches gilt auch für Beleuchtungseffekte in Rasterbildern.

Neben numerischen Feature-Daten kommen auch nichtnumerische Daten zum Einsatz. Zum Beispiel können etwa aus Moleküldaten oder Grafiken extrahierte Graphstrukturen zum Ähnlichkeitsvergleich eingesetzt werden. Spezielle Ähnlichkeitsmaße auf Graphstrukturen sind Gegenstand intensiver Forschung in der Mathematik [Chartrand et al. 1998].

Ein Überblick über vielfältige aus Multimediaobjekten extrahierbare Features wird in [Schäuble 1997, Bimbo 1999, Veltkamp et al. 2001, Marques & Furht 2002, Castelli & Bergman 2002] gegeben.

1.2 Vektorraummodell

Klassische Information-Retrieval-Systeme ermöglichen die Suche nach Textdokumenten zu einem bestimmten Thema. *Multimedia Retrieval* hingegen versucht, Information-Retrieval-Techniken für die inhaltsbasierte Suche nach Multi-

mediaobjekten einzusetzen [Faloutsos 1996, Lu 1999, Schmitt 2002, Schmitt 2005]. Ein dabei häufig eingesetztes Retrieval-Modell ist das *Vektorraummodell*. In diesem Modell werden die Feature-Daten der Multimediaobjekte als Vektoren in einem in der Regel hochdimensionalen Vektorraum aufgefasst. Die Ähnlichkeit zwischen jeweils zwei Multimediaobjekten wird anhand ihrer Feature-Vektoren $q, o \in \mathbb{R}^n$ und einer Distanzfunktion oder eines Ähnlichkeitsmaßes berechnet. Genau genommen berechnet eine Distanzfunktion die Unähnlichkeit zweier Multimediaobjekte. Durch Einsatz einer Umkehrfunktion, zum Beispiel 2^{-d} für eine Distanz d , lassen sich Distanzwerte in Ähnlichkeitswerte aus dem Intervall $[0,1]$ überführen. Als Distanzfunktion kommt häufig die Minkowski-Distanzfunktion

$$\left(\sum_{i=1}^n |q[i] - o[i]|^m\right)^{\frac{1}{m}}$$

($m=1$ für Manhattan-Distanz und $m=2$ für euklidische Distanz) zum Einsatz, während als Ähnlichkeitsmaß häufig das Kosinusmaß

$$\cos(\vec{q}, \vec{o}) = \frac{\sum_{i=1}^n q[i] \cdot o[i]}{\sqrt{\sum_{i=1}^n q[i]^2} \sqrt{\sum_{i=1}^n o[i]^2}}$$

verwendet wird. Das Kosinusmaß berechnet den Kosinus des eingeschlossenen Winkels anhand des Skalarprodukts und der Normen der beteiligten Vektoren. Neben diesen Funktionen steht eine Vielzahl weiterer Funktionen, wie etwa die Earth-Mover-Distanz [Rubner & Tomasi 2001], zur Ähnlichkeitsberechnung zur Auswahl. Einen Überblick über Distanzfunktionen und Ähnlichkeitsmaße liefert [Schmitt 05]. Abbildung 1 illustriert die distanzbasierte Ähnlichkeitsberechnung

im Vektorraummodell auf Rasterbildern stellvertretend für verschiedene Medientypen.

1.3 Aggregation von Ähnlichkeitswerten

Daten eines Features sind algorithmisch ermittelte Werte, die eine bestimmte inhaltstragende Eigenschaft von Multimediaobjekten ausdrücken. Dazu gehört zum Beispiel die Farbverteilung eines Rasterbilds. Wenn eine Retrieval-Suchanfrage jedoch Suchbedingungen auf mehreren Eigenschaften formuliert, müssen Ähnlichkeitswerte von verschiedenen Feature-Daten in das Gesamtergebnis einfließen (siehe Abb. 2). Dabei stellt sich die Frage, wie mehrere Ähnlichkeitswerte beziehungsweise Distanzwerte eines Multimediaobjekts zu einem Wert aggregiert werden können. Häufig verwendete Aggregationsfunktionen sind die Minimums- und die Maximumsfunktion, die Summe und der arithmetische Durchschnitt. Kommt es bei den Ähnlichkeitswerten nur darauf an, Ergebnisobjekte in eine Reihenfolge zu bringen (*Rangbildung, ranking*), sind der Durchschnitt und die Summe natürlich äquivalent. Ein anderer Ansatz (siehe etwa [Fagin & Wimmers 2000]) verwendet T-Normen und T-Konormen der Fuzzy-Logik [Zadeh 1988, Galindo et al. 2005] zur Kombination von Ähnlichkeitswerten, die dort als Membership-Werte interpretiert werden. Mittels einer *Gewichtung*, etwa nach der faginschen Gewichtungformel [Fagin & Wimmers 2000, Schulz & Schmitt 2003, Schulz 2004] oder mittels einer gewichteten Summe, lassen sich verschiedenen Suchtermen unterschiedliche Gewichte zuordnen, die bei

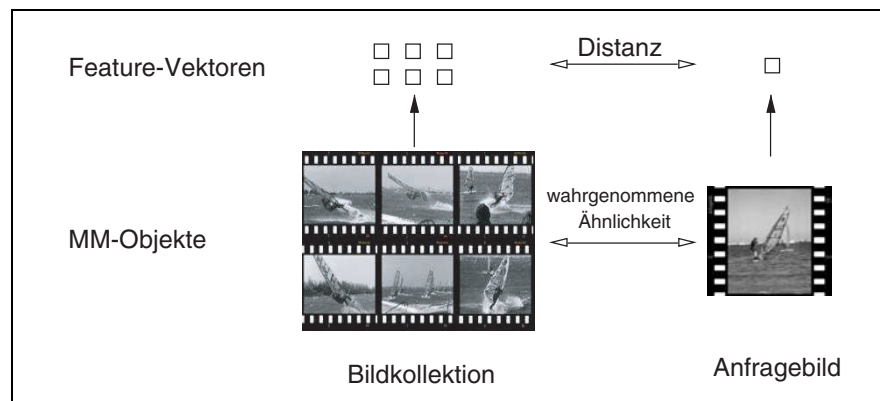


Abb. 1: Ähnlichkeitsberechnung mittels Distanzen auf Feature-Vektoren

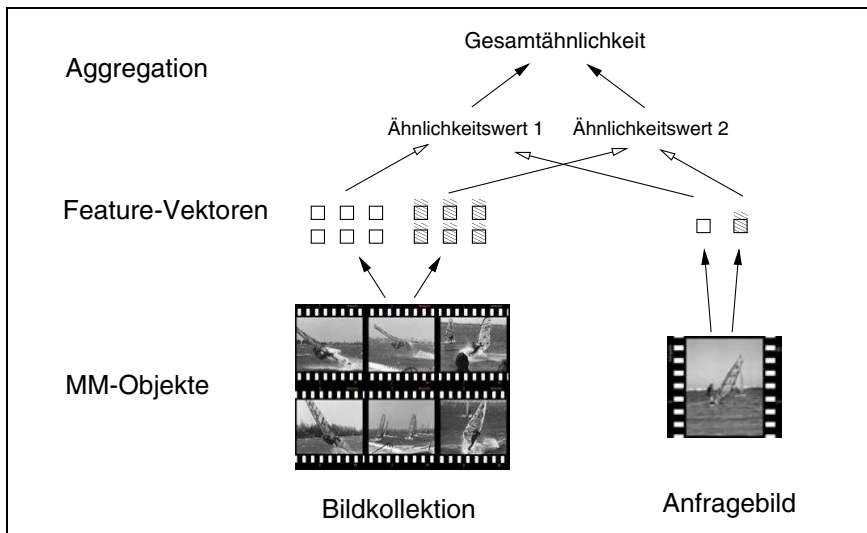


Abb. 2: Aggregation von Ähnlichkeitswerten

der Aggregation entsprechend ausgewertet werden. Wenn zum Beispiel Bilder mit einem Surfer mit Surfbrett gesucht werden sollen, wobei Form und Farbe von Bildobjekten als verschiedene Feature-Daten zur Verfügung stehen, dann würde man wahrscheinlich der Formbedingung (dreieckig) ein stärkeres Gewicht als der Farbbedingung zuordnen. Speziell für die effiziente Aggregation von Ähnlichkeitswerten aus zwei Sequenzen von Multimediaobjekten existieren spezielle Algorithmen, die z.B. in [Fagin et al. 2003, Henrich & Robbert 2003, Schmitt 2005] beschrieben werden.

1.4 Indizes zur schnellen Nächste-Nachbarsuche

Bei einer inhaltsbasierten Multimedia-Retrieval-Suche werden die bezüglich der Anfrage relevantesten Medienobjekte gesucht. Im Vektorraum bedeutet dies die Suche nach den Medienobjekten, deren Feature-Vektoren die geringsten Distanzen zum Feature-Vektor der Anfrage aufweisen. Die aufwendige Distanzberechnung zu allen Feature-Vektoren, die in einer Datenbank abgelegt sind, führt üblicherweise zu inakzeptabel langen Antwortzeiten. Daher wurden zur schnellen Nächste-Nachbarsuche im hochdimensionalen Vektorraum verschiedene Indexstrukturen entwickelt. So wurden beispielsweise baumartige Indizes vorgeschlagen [Böhm et al. 2001], um statt eines linearen einen logarithmischen Suchaufwand zu erzielen. Leider hat sich speziell bei der Verwendung der eukli-

dischen Distanz gezeigt, dass dies nur im niedrigdimensionalen Raum (bis zu etwa 10 Dimensionen) gelingt. In höherdimensionalen Vektorräumen ist der Suchaufwand durch diese Indexstruktur oft sogar höher, als wenn gar kein Index eingesetzt wird. Dieses Phänomen wird häufig als *Fluch der hohen Dimensionen* bezeichnet [Aggarwal et al. 2001].

Eine prominente Alternative zu den Baumindizes ist das VA-File [Weber et al. 1998], das die Idee der Distanzsubstitution aus [Faloutsos 1996] weiterentwickelt. Bei diesem Ansatz wird davon ausgegangen, dass eine Nächste-Nachbarsuche immer einen sequenziellen Aufwand erfordert. Allerdings wird dieser Aufwand minimiert, indem für jeden Feature-Vektor eine Signatur berechnet wird (siehe Abb. 3), die einen Feature-Vektor

approximierend, aber kompakt beschreibt. In einem ersten Schritt lässt sich mittels eines sequenziellen Durchlaufs über alle Signaturen ein Großteil der Multimediaobjekte von der weiteren Suche ausschließen. Dies wird durch die Berechnung von Distanzabschätzungen zwischen Feature-Vektor der Anfrage und Signatur ermöglicht. Im zweiten Schritt werden zu den restlichen Feature-Vektoren die exakten Distanzen berechnet und die nächsten Nachbarn entsprechend gefunden. Weitere so genannte Filterindizes sind in [Ferhatosmanoglu et al. 2000, Cha et al. 2002, Cha & Chung 2002, Balko et al. 2004] zu finden.

Viele Indizes wurden für die Nächste-Nachbarsuche mittels Feature-Vektoren und der euklidischen Distanz entwickelt. Allerdings gibt es Anwendungen, bei denen Feature-Daten nicht als Vektoren dargestellt werden können beziehungsweise eine nichteuklidische, möglicherweise sehr aufwendig zu berechnende Distanzfunktion, etwa die Earth-Mover-Distanz [Rubner & Tomasi 2001], zum Einsatz kommt. Für solche Fälle können Metrikindizes eingesetzt werden, die die Dreiecksungleichung einer Distanzfunktion zur Minimierung der Anzahl der Distanzberechnungen ausnutzen. Einen guten Überblick über Metrikindizes gibt die Arbeit [Chavez et al. 2001].

2 Multimedia-Retrieval-Systeme und ihre Grenzen

Seit Anfang der neunziger Jahre beschäftigt sich die Datenbankforschung mit inhaltsbasiertem Multimedia Retrieval, wobei der Schwerpunkt auf der Suche in

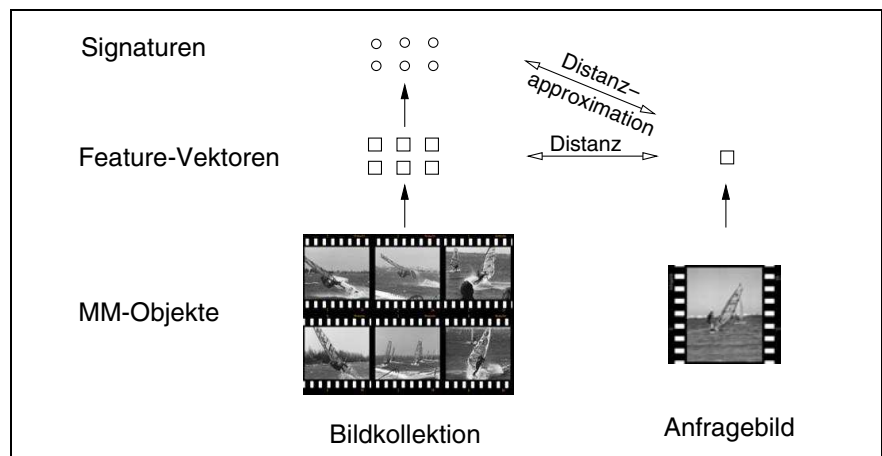


Abb. 3: Effiziente Ähnlichkeitsberechnung mittels Distanzen auf Signaturen und Feature-Vektoren

Bilddatenbanken [Smeulders et al. 2000] liegt. Eines der ersten kommerziellen Systeme entstand aus dem IBM-Projekt QBIC (Query By Image Content) [Flickner et al. 1995]. Dieses System hat als so genannter Image Extender im Datenbanksystem DB2 Eingang gefunden. In Folge dieses Projektes wurde eine große Anzahl ähnlicher inhaltsbasierter Bild-Retrieval-Systeme entwickelt, die meisten davon sind Prototypen aus dem universitären Umfeld. Einen guten Überblick, bei dem die Systeme bezüglich Anfrageformulierung, verwendeten Features, eingesetzten Indexstrukturen und Ergebnisdarstellung unterschieden werden, gibt [Veltkamp & Tanase 2000].

Neben dem IBM-Datenbanksystem DB2 haben weitere kommerzielle Datenbanksysteme ihre Funktionalität bezüglich einer inhaltsbasierten Suche nach Multimediaobjekten erweitert. Dies geschah im Kontext so genannter Universal-Server-Architekturen, die eine modulartige Funktionserweiterung von objektrelationalen Datenbanksystemen ermöglichen. In der ORACLE-Erweiterung, zum Beispiel, kann die Suche nach allen Bildern, die einen Surfer mit Surfbrett darstellen, anhand von 70% Form und 30% Farbe folgendermaßen formuliert werden:

```
SELECT photo_id
FROM photo_collection
WHERE ORDSYS.IMGSimilar(
  photo_sig, comparison_sig,
  'color="0.3", texture="0",
  shape="0.7", location="0",
  20)=1;
```

Die Feature-Daten werden hier als Signaturen (photo_sig, comparison_sig) bezeichnet. Eine Gewichtung über vier verschiedene Features wird unterstützt, und ein Schwellenwert – im obigen Beispiel 20 – selektiert alle Fotos, bei denen die gewichtete Distanzsumme den angegebenen Wert nicht übersteigt. In [Stolze 2002] werden die konkreten Bild-Retrieval-Erweiterungen der kommerziellen Datenbanksysteme ORACLE, DB2 und Informix kurz vorgestellt. Ein weiterer Ansatz, der als Teil des SQL-Standards versucht, die Nutzung von Multimedia-Retrieval-Erweiterungen zu vereinheitlichen, ist SQL/MM. Hierbei befasst sich Teil 2 des SQL/MM-Standards mit Text-Retrieval [ISO 1995] und Teil 5 mit Bild-Retrieval [ISO 1996, Stolze 2002].

2.1 Grenzen

Etwa zur Jahrtausendwende flaute die Euphorie bei der Entwicklung und dem Einsatz von inhaltsbasierten Multimedia-Retrieval-Systemen ab. Zu einem Teil lag diese Ernüchterung in nicht erfüllten Versprechen begründet. Es wurde die Erfahrung gemacht, dass in vielen Anwendungen die Suchergebnisse der Multimedia-Retrieval-Systeme nicht den hochgesteckten Erwartungen entsprachen. Häufig wurden als Ergebnis einer inhaltsbasierten Suchanfrage irrelevante Multimediaobjekte zurückgeliefert, während zusätzlich relevante Objekte oft unterdrückt wurden. Im Folgenden sollen zwei Aspekte aus diesem Problemkomplex kurz diskutiert werden, die die generellen Grenzen solcher Systeme aufzeigen: die semantische Lücke zwischen Nutzeranfrage und den verfügbaren Feature-Daten des gesuchten Objekts sowie die Breite des gewünschten Anwendungsbereichs.

Semantische Lücke: Sucht man in der deutschsprachigen Wikipedia-Ausgabe nach dem Begriff *semantische Lücke*, stößt man auf folgende Definition:

... *Unterschied zwischen Formulierung von Kontextwissen in einer mächtigen Sprache (z.B. natürliche Sprache) und dessen formaler und automatisiert reproduzierbaren Repräsentation in einer weniger mächtigen formalen Sprache (z.B. Programmiersprache).*

Die semantische Lücke in unserem Kontext (siehe Abb. 4) existiert zwischen der eingeschränkten Semantik der aus Rohdaten automatisch extrahierbaren Feature-Daten und der menschlichen Wahrnehmung von Inhalt und Ähnlichkeit. Da zurzeit nicht absehbar ist, ob diese Lücke je geschlossen werden kann, sind inhaltsbasierten Systemen in ihrer Anwendbarkeit enge Grenzen gesetzt. Im Allgemeinen können also automatisch extrahierte Feature-Daten die menschliche Interpretation nicht ausreichend nachbilden. Dem Retrieval-System fehlt

in der Regel das für die Interpretation benötigte menschliche (Welt- oder Allgemein-)Wissen über den Kontext. Daher spricht man in diesem Zusammenhang oft auch von *Low-Level-Feature-Daten*.

Enge versus breite Anwendungen: Bei der Diskussion der Anwendbarkeit einer inhaltsbasierten Multimediasuche ist es sinnvoll, *breite* von *engen* Anwendungen zu unterscheiden. Als ein Extrembeispiel für eine enge Anwendung sei angenommen, unsere Bildkollektion enthält ausschließlich Aufnahmen von grünen Wäldern und von Surfern. Offensichtlich lassen sich zur Ähnlichkeitsberechnung diese beiden Szenarien anhand einer Farbverteilung sehr einfach und zuverlässig voneinander unterscheiden. Die Feature-Daten reichen hier also zur Interpretation aus. Im Gegensatz dazu erhalten wir eine extrem breite Anwendung, wenn alle Bilder im Internet als Datenbasis betrachtet werden. Wie man leicht einsehen kann, wird es dann fast unmöglich, Surfer etwa von Segelbooten anhand von automatisch extrahierten Feature-Daten zuverlässig zu unterscheiden. Je breiter die Anwendung ist, desto stärker ist in der Regel auch die Variabilität der Inhalte. Mit der Variabilität wächst jedoch die Schwierigkeit, Inhalt zuverlässig durch Feature zu beschreiben, und somit wird es auch schwieriger, Inhalte zu erkennen bzw. wiederzufinden.

3 Forschungsthemen und Herausforderungen

Im vorherigen Abschnitt haben wir betont, dass das Hauptproblem der inhaltsbasierten Suche in Multimedia-Retrieval-Systemen in der Existenz der semantischen Lücke liegt. Zurzeit ist eine erfolgreiche Überwindung nicht abzusehen, es gibt allerdings einige Ansätze, diese Lücke zu verringern bzw. zu umgehen. Wir gehen im Folgenden auf vier grundlegende Ansätze kurz ein: Die Verwendung

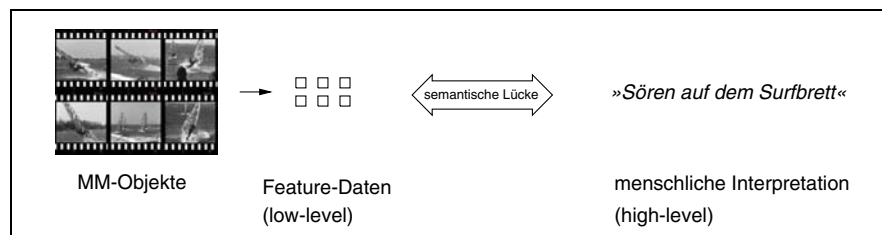


Abb. 4: Semantische Lücke

von zusätzlichen Informationen, die Verwendung von Lernverfahren, schwache Segmentierungsverfahren sowie spezielle Anfragesprachen. Auch wenn wir uns im Folgenden schwerpunktmäßig mit Bilddaten befassen, lassen sich die dargestellten Grundideen vielfach auf andere Multimediaobjekte übertragen.

3.1 Ausnutzen zusätzlicher Informationen

Die bisher diskutierten Feature-Daten werden gemäß der in [Schmitt 2002] vorgenommenen Klassifikation als inhaltsbasierte Metadaten bezeichnet. Neben ihnen lassen sich zum Wiederauffinden beziehungsweise Suchen von Multimediaobjekten zusätzliche Metadaten, nämlich *inhaltsbeschreibende* und *inhaltsunabhängige* Metadaten, verwenden. Inhaltsunabhängige Metadaten umfassen recht grundlegende Informationen wie Datum, Zeit und Ort der Aufnahme, Details zur Art der Aufnahme (Kamera, Bildgröße, Winkel, Beleuchtung etc.). Inhaltsbeschreibende Metadaten (Annotationen) tragen manuell erstellte Information wie Stichworte zum Bildinhalt, Untertitel, umfassen aber auch komplexere inhaltliche textuelle Beschreibungen. Viele Multimediaformate bieten die Möglichkeit, diese Informationen direkt in ein Multimediaobjekt einzubetten. Für inhaltsunabhängige und inhaltsbeschreibende Metadaten in Rasterbildern sind als Standard insbesondere EXIF- sowie FotoNotesTM- und JPEG-2000-Metadaten zu erwähnen. Für Videos existieren MPEG-7 [Kosch & Döllner 2005] und insbesondere das *StillRegion* und *StillRegionSpatialDecomposition Description Scheme (DS)* sowie das *Linguistic DS* [Hasida 2003], das die Einbindung externer Ressourcen, wie z.B. Ontologien, erlaubt. Bis auf EXIF basieren alle der genannten Formate auf XML.

Die Verwendung von Ontologien oder Thesauri erlaubt zum einen die Verwendung von Ober- oder Unterbegriffen bei der Suche. Zum anderen können somit auch Mehrdeutigkeiten aufgelöst werden. Wird zum Beispiel in einer Annotation das Wort *Bank* verwendet, ist nicht klar, ob im Bild eine Bank zum Sitzen oder ein Finanzinstitut abgebildet ist. Wird zusammen mit dem Stichwort jedoch ein Verweis auf eine eindeutige

Klasse innerhalb einer Ontologie, z.B. WordNet [Miller et al. 1990], abgelegt, kann diese Mehrdeutigkeit aufgelöst werden. Mittlerweile wurde eine Vielzahl solcher Ansätze zur semantischen Annotation vorgeschlagen [Hollink et al. 2003, Bloehdorn et al. 2005, Hentschel et al. 2006].

Ein grundlegendes Problem dieser Ansätze ist, dass die Annotation der Stichworte manuell erfolgen muss. Motiviert von dem Erfolg einer Vielzahl von Community-Plattformen sind mittlerweile jedoch einige Portale für die Annotation (bzw. das *tagging*) von Bildern und Videos entstanden, so zum Beispiel Flickr (www.flickr.com), YouTube (www.youtube.com) und Marvel (www.research.ibm.com/marvel), deren Daten frei verfügbar sind und das Hochladen und Annotieren von Bildern bzw. Videos ermöglichen. Der Grundgedanke solcher Ansätze ist es, die Motivation und den Selbstorganisationsprozess von Personengruppen mit gleichen Interessen zur Entwicklung eines effizient zu nutzenden Community-Portals zur Annotation von Bildsammlungen zu nutzen. Eine weitere Unterstützung bei der Annotation von Objektsammlungen kann auch durch Verfahren des maschinellen Lernens erfolgen. Eine kurze Diskussion hierzu enthält der nächste Abschnitt.

3.2 Lernverfahren

Bei den Lernverfahren, die in Multimedia-Retrieval-Systemen eingesetzt werden, kann man im Wesentlichen zwischen Verfahren unterscheiden, die während der Suche eingesetzt werden und die Rückmeldungen eines Nutzers über relevante beziehungsweise irrelevante Objekte einer Ergebnismenge (Relevance Feedback) zur Verbesserung des Suchergebnisses ausnutzen, und Verfahren zum Lernen von Metadaten, basierend auf (teilannotierten) Objekten.

Die meisten Verfahren zur Verwendung von Relevance Feedback basieren auf den Ideen von Rocchio für das Vektorraummodell [Rocchio 1971] oder Robertson für probabilistische Ansätze [Robertson 1977], die ursprünglich für die Textsuche entwickelt wurden. Somit handelt es sich streng genommen nicht um Verfahren des maschinellen Lernens. Die Grundidee ist es zu bestimmen, wie stark einzelne Features für den Nutzer bei einer gegebenen Suchanfrage relevant sind. Ei-

nen Überblick über Relevance-Feedback-Verfahren geben [Ruthven & Lalmas 2003, Ferecatu et al. 2004].

Mit maschinellen Lernverfahren versucht man, Zusammenhänge zwischen Low-Level-Feature-Daten von Medienobjekten und deren Annotation zu lernen. Aktuelle Verfahren werden zum Beispiel in [Natsev et al. 2005, Lu et al. 2005, Feng & Chua 2003] vorgestellt.

3.3 Schwache Segmentierung

In vielen Bild-Retrieval-Systemen wird ein Rasterbild bei der Feature-Extraktion als ganze Einheit betrachtet. Bei einer inhaltsbasierten Suche hingegen ist man in der Regel nur an Bildobjekten interessiert, die bestimmten Regionen innerhalb des Bildes zugeordnet werden können. Eine Objektsegmentierung der Bilder soll inhaltsbasiert diese Objektregionen ermitteln, etwa die Umrisse eines Surfsegels. Aus diesen lassen sich nun zur Verbesserung der Retrieval-Qualität regionenbasierte Features berechnen. Allerdings ist aus der Bildverarbeitung bekannt, dass eine inhaltsbasierte Objektsegmentierung im Allgemeinen nicht automatisch und gleichzeitig zuverlässig funktioniert (auch hier das Problem der semantischen Lücke). Eine manuelle oder semiautomatische Segmentierung hingegen ist oft zu aufwendig. Einen Ansatz zur Entschärfung dieses Problems bietet die so genannte *schwache Segmentierung*. Entsprechende Algorithmen werden etwa in [Mirmehdi & Petrou 2000, Pauwels & Frederix 2000] vorgestellt. Die Segmentierung soll hierbei nicht mehr exakt die Konturen von dargestellten Objekten finden. Stattdessen werden mittels eines geeigneten Homogenitätsprädikats gleichförmige Bildregionen ermittelt. Dabei sollte nach Möglichkeit eine homogene Region immer Bestandteil *einer* Objektregion sein, diese also nicht schneiden. Eine schwache Segmentierung kann automatisch berechnet werden. Obwohl die homogenen Regionen oft nicht exakt den Bildobjekten entsprechen, kann dadurch trotzdem die Suchqualität verbessert werden. Ein prominentes Bild-Retrieval-System, das auf der schwachen Segmentierung aufbaut, ist Blobworld [Carson et al. 2002]. Schwache Segmentierung ist nicht auf Rasterbilder begrenzt. Sie kann auch auf den Medientypen Video (Szenenerkennung)

und Audio zur Verbesserung der Retrieval-Qualität genutzt werden.

3.4 Anfragesprache

Ein großes Problem der inhaltsbasierten Suche ist die geeignete Formulierung des Informationsbedarfs, damit das Retrieval-System passende Ergebnisse ermitteln kann. Dabei spielen die Anfragesprache und -verarbeitung eine zentrale Rolle. Allgemein soll die Anfragesprache eine komplexe Suchformulierung auf unterschiedlich strukturierten Daten ermöglichen und dabei verschiedene Suchparadigmen in sich vereinen. Dies soll den Nutzer in die Lage versetzen, die semantische Lücke durch eine Anfrageformulierung auf der Grundlage der dem Retrieval-System verfügbaren Daten bzw. Features zu reduzieren. Genauer bedeutet dies die Forderung nach Sprachunterstützung folgender Aspekte:

1. *Retrieval-Techniken:* Diese ermöglichen ein Abschätzen der Zugehörigkeit bzw. der Wahrscheinlichkeit, mit der ein Objekt relevant zu einer Anfrage ist. Ergebnisobjekte werden also graduell bewertet und in eine bestimmte Relevanzreihenfolge gebracht. Weiterhin gehören Mechanismen des Relevance Feedback und zur Anfrage-Iteration hierzu.
2. *Datenbankanfragetechniken:* Da es zur Suche notwendig sein kann, auch relational strukturierte Daten heranzuziehen, ist eine diesbezügliche Unterstützung erforderlich.
3. *XML-Anfragetechniken:* Viele für Multimedia Retrieval relevante Daten lassen sich als XML-Daten darstellen. Dazu gehören etwa Strukturbeschreibungen komplexer Multimediaobjekte, aber auch Annotationen entsprechend dem MPEG-7-Standard. Zur Suche müssen daher XML-Anfragetechniken unterstützt werden.
4. *Ontologien:* Die Anfragesprache sollte Mechanismen für den Zugriff auf Ontologien und Thesauri unterstützen. Dies ermöglicht die Einbeziehung allgemein akzeptierter semantischer Zusammenhänge zwischen Begriffen der menschlichen Sprache in die Ergebnisberechnung.
5. *Nutzerschnittstelle:* Sowohl bei der Anfrageformulierung als auch bei der Ergebnispräsentation müssen speziel-

le Interaktionsschnittstellen zur Verfügung stehen. Diese umfassen Datenvisualisierungen, grafische Anfrageformulierung und Skizzieren von Multimediaobjekten zur Suche, Navigation und Browsen auf Ergebnisobjekten sowie Mechanismen zur Bewertung von Ergebnissen im Rahmen von Relevance Feedback.

6. *Räumliche und zeitliche Beziehungen:* In vielen Multimediaobjekten spielen räumliche und zeitliche Beziehungen eine wichtige Rolle. Dies wird besonders beim Medientyp Video ersichtlich.
7. *Kombination:* Vor allem muss die Anfragesprache die Kombination verschiedener Suchparadigmen (RDBMS, IR, XML) in *einem vereinheitlichten Formalismus* erlauben. Komplexe Anfragen müssen aus einfacheren Anfragen konstruierbar sein. Dazu ist auch ein geeignetes Modell zur unterschiedlichen Gewichtung von Suchbedingungen erforderlich. Dass die Kombination verschiedener Suchparadigmen eine aktuelle Herausforderung ist, zeigt etwa die derzeit aktuelle Bemühung, Retrieval-Techniken in eine XML-Anfragesprache zu integrieren [Trotman & Sigurbjörnsson 2005].

Leider existiert zurzeit keine derart universelle Anfragesprache zur inhaltsbasierten Suche nach Multimediaobjekten. Stattdessen sind einzelne Spezialsprachen im Einsatz, die jeweils nur einen kleinen Ausschnitt aus dem vorgestellten Anforderungsspektrum umsetzen. So existieren zum Beispiel reine Retrieval-Sprachen, die auf einem klassischen Retrieval-Modell aufbauen. Dazu gehören die meisten in [Veltkamp & Tanase 2000] untersuchten Systeme. Diese Systeme bieten in der Regel keine Anbindung an

eine klassische Datenbank- oder XML-Anfragesprache. Objektrelationale Anfragesprachen (SQL:2003) hingegen scheinen aufgrund der Einbindung von SQL/MM viele der Anforderungen zu erfüllen. Dies ist jedoch leider nicht der Fall, da in diesen Sprachen nur rudimentäre Retrieval-Mechanismen unterstützt werden. Weiterhin erkennt man in der ORACLE-Beispielanfrage (siehe Abschnitt 2), dass für die Kombination von Retrieval- und Datenbankbedingungen Ähnlichkeitswerte mittels eines Schwellenwertes auf boolesche Werte abgebildet werden müssen. Aufgrund dieser Abbildung gehen diese Ähnlichkeitswerte verloren und können nicht zur Kombination mit anderen Suchbedingungen genutzt werden. Insbesondere kann also keine komplexe Anfrage mittels Konjunktion und Disjunktion auf Retrieval-Bedingungen konstruiert werden.

Einen Ansatz zur Kombination von Retrieval- und klassischen Datenbanktechniken bietet der Einsatz von Fuzzy-Logik. Diese verallgemeinert die boolesche Logik um Aspekte der Unschärfe bzw. der graduellen Zugehörigkeit. Somit können komplexe Bedingungen aus einzelnen Ähnlichkeitsbedingungen konstruiert werden. Die *same-Algebra* [Ciaccia et al. 2000] und die Sprachen *WS-QBE*, *SDC* und *SA* [Schmitt & Schulz 2004, Schmitt et al. 2005] sind Vertreter dieses Ansatzes. In diesen Sprachen werden Ähnlichkeitswerte (*scores*) importiert und innerhalb der Fuzzy-Logik ausgewertet (siehe Abb. 5 links). Dies kann Probleme bereiten, wenn Ähnlichkeitswerte verschiedener Quellen nicht vergleichbar sind. In Abbildung 5 (rechts) sind zwei fiktive Scoring-Funktionen abgebildet, die bezüglich zweier verschiedener Features die empfundenen Ähnlichkeiten auf Ähnlichkeitswerte unterschiedlich abbilden. Werden diese

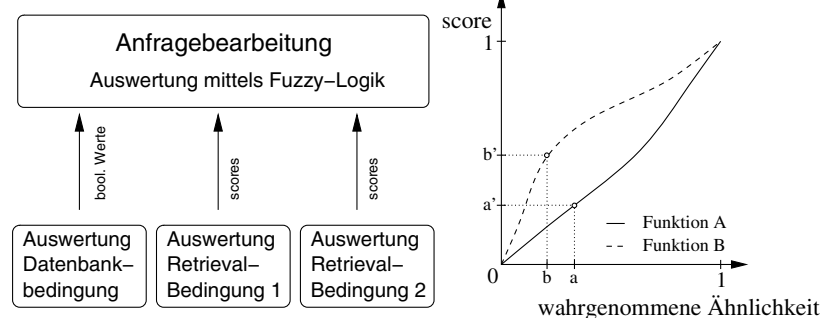


Abb. 5: Fuzzy-Auswertung mittels Import von booleschen und Ähnlichkeitswerten (links) sowie Ähnlichkeitswerten von zwei verschiedenen Quellen (rechts)

Werte zum Beispiel mit der T-Norm \min der Fuzzy-Logik konjunktiv kombiniert, werden Werte der Funktion A bevorzugt. Wenn in einem Beispiel etwa a die empfundene Ähnlichkeit bezüglich der Form und b die Ähnlichkeit bezüglich der Farbverteilung bezeichnet, dann wird sogar die Werteordnung aufgrund der nicht kompatiblen Funktionen vertauscht ($b < a$ aber $a' < b'$). Insgesamt reicht es also bei der Entwicklung einer Multimedia-Retrieval-Anfragesprache nicht aus, Ähnlichkeitswerte und boolesche Werte in einen einheitlichen Formalismus zu importieren. Stattdessen muss ein einheitlicher Formalismus zur Berechnung der Ähnlichkeitswerte verwendet werden.

4 Zusammenfassung

In diesem Beitrag haben wir einen kurzen Überblick über das Anliegen, die Prinzipien und den aktuellen Forschungsstand von inhaltsbasiertem Multimedia Retrieval gegeben. Als Hauptproblem wurde die semantische Lücke herausgestellt. Die kurz skizzierten Ansätze zur Verringerung dieser semantischen Lücke nähren die Hoffnung, dass inhaltsbasiertes Multimedia Retrieval in Zukunft besser als bisher in der Lage sein wird, Multimediaobjekte inhaltsbasiert zu finden. Eines ist auf jeden Fall gewiss: Das Problem der Suche nach Multimediaobjekten wird mit der steigenden Verfügbarkeit von Multimediale Daten weiter wachsen. Ernst zu nehmende alternative Ansätze zur Organisation und Suche sind zurzeit nicht in Sicht.

Danksagung

Die Autoren bedanken sich bei Anke Schneidewind, Ingolf Geist, Erik Buchmann und den Gutachtern für viele wertvolle Kritiken und Hinweise.

Literatur

[Aggarwal et al. 2001] Aggarwal, C. C.; Hinneburg, A.; Keim, D. A.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. In: Database Theory – ICDT'01, Proc. of the 8th Int. Conf., London, January 2001, LNCS, Band 1773, Springer-Verlag, Berlin, 2001, S. 420-434.

[Balko et al. 2004] Balko, S.; Schmitt, I.; Saake, G.: The Active Vertice Method: A Performant Filtering Approach to High-Dimensional Indexing. Data and Knowledge Engineering, 51(3):369-397, 2004.

[Bimbo 1999] Bimbo, A. D.: Visual Information Retrieval. Morgan Kaufmann, 1999.

[Bloehdorn et al. 2005] Bloehdorn, Stephan; Petridis, Kosmas; Saathoff, Carsten; Simou, Nikos; Tzouvaras, Vassilis; Avrithis, Yannis; Handschuh, Siegfried; Kompatsiaris, Yannis; Staab, Steffen; Strintzis, Michael G.: Semantic annotation of images and videos for multimedia analysis. In: Proc. of Second European Semantic Web Conf. (ESWC 2005), 2005.

[Böhm et al. 2001] Böhm, C.; Berchtold, S.; Keim, D.: Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. ACM Computing Surveys, 33(3):322-373, 2001.

[Carson et al. 2002] Carson, Chad; Belongie, Serge; Greenspan, Hayit; Malik, Jitendra: Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying. IEEE Trans. Pattern Anal. Mach. Intell., 24(8):1026-1038, 2002.

[Castelli & Bergman 2002] Castelli, V.; Bergman, L. D. (Hrsg.): Image Databases – Search and Retrieval of Digital Imagery. John Wiley & Sons, Inc., New York, USA, 2002.

[Cha & Chung 2002] Cha, G.-H.; Chung, C.-W.: The GC-Tree: A High-Dimensional Index Structure for Similarity Search in Image Databases. IEEE Transactions on Multimedia, 4(2):235-247, Juni 2002.

[Cha et al. 2002] Cha, G.-H.; Zhu, X.; Petkovic, D.; Chung, C.-W.: An Efficient Indexing Method for Nearest Neighbor Searches in High-Dimensional Image Databases. IEEE Transactions on Multimedia, 4(1):76-87, März 2002.

[Chartrand et al. 1998] Chartrand, G.; Kubicki, G.; Schultz, M.: Graph similarity and distance in graphs. Aequationes Mathematicae, 55(1-2):129-145, 1998.

[Chavez et al. 2001] Chavez, E.; Navarro, G.; Baeza-Yates, R. A.; Marroquin, J. L.: Searching in Metric Spaces. ACM Computing Surveys, 33(3):273-321, 2001.

[Ciaccia et al. 2000] Ciaccia, Paolo; Montesi, Danilo; Penzo, Wilma; Trombetta, Alberto: Imprecision and user preferences in multimedia queries: A generic algebraic approach. In: K.-D. Schewe, B. Thalheim (Hrsg.), FoIKS: Foundations of Information and Knowledge Systems, First International Symposium, FoIKS 2000, Burg, Germany, February 14-17, 2000, LNCS, Band 1762, Springer-Verlag, 2000, S. 50-71.

[Fagin & Wimmers 2000] Fagin, R.; Wimmers, E. L.: A Formula for Incorporating Weights into Scoring Rules. Special Issue of Theoretical Computer Science, 239(2):309-338, 2000.

[Fagin et al. 2003] Fagin, R.; Lotem, A.; Naor, M.: Optimal aggregation algorithms for middleware. Journal of Computer and System Sciences, 66(4):614-656, 2003.

[Faloutsos 1996] Faloutsos, Christos: Searching Multimedia Databases by Content. Kluwer Academic Publishers, Boston, Dordrecht, London, 1996.

[Feng & Chua 2003] Feng, HuaMin; Chua, Tat-Seng: A bootstrapping approach to annotating large image collection. In: MIR '03: Proc. of the 5th ACM SIGMM Int. Workshop on Multimedia Information Retrieval, New York, NY, USA, ACM Press, 2003, S. 55-62.

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

[Ferhatosmanoglu et al. 2000] Ferhatosmanoglu, H.; Tuncel, E.; Agrawal, D.; Abbadi, A. E.: Vector Approximation based Indexing for Non-uniform High Dimensional Data Sets. In: Proc. of the 9th ACM CIKM Int. Conf. on Information and Knowledge Management, November 6-11, 2000, McLean, VA, USA, New York, ACM Press, 2000, S. 202-209.

[Flickner et al. 1995] Flickner, M.; Sawhney, H. S.; Ashley, J.; Huang, Q.; Dom, B.; Gorkani, M.; Hafner, J.; Lee, D.; Petkovic, D.; Steele, D.; Yanker, P.: Query by Image and Video Content: The QBIC System. IEEE Computer, 28(9):23-32, 1995.

[Galindo et al. 2005] Galindo, J.; Urrutia, A.; Piattini, M.: Fuzzy Databases: Modeling, Design and Implementation. Idea Group Publishing, 2005.

[Hasida 2003] Hasida, Kioti: The linguistic DS: Linguistic description in MPEG-7. The Computing Research Repository (CoRR), cs.CL/0307044, 2003.

[Henrich & Robbert 2003] Henrich, A.; Robbert, G.: Ein Ansatz zur Übertragung von Rangordnungen bei der Suche auf strukturierten Daten. In: G. Weikum, H. Schöning, E. Rahm (Hrsg.), Datenbanksysteme in Business, Technologie und Web, BTW'03, 10. GI-Fachtagung, Leipzig, Februar 2003, Lecture Notes in Informatics (LNI) Volume P-26, Gesellschaft für Informatik, Bonn, 2003, S. 167-186.

[Hentschel et al. 2006] Hentschel, C.; Nürnberger, A.; Schmitt, I.; Stober, S.: SAFIRE: Towards standardized semantic rich image annotation. In: Proc. of 4th Int. Workshop on Adaptive Multimedia Retrieval (AMR 06), 2006 (erscheint).

[Hollink et al. 2003] Hollink, L.; Schreiber, G.; Wielemaker, J.; Wielinga, B.: Semantic annotation of image collections. In: Proc. of Workshop on Knowledge Markup and Semantic Annotation (KCAP'03), 2003.

[ISO 1995] ISO (International Organization for Standardization) & ANSI (American National Standards Institute): ISO/IEC JTC 1/SC 21/SQL/M LHR-004, N9790. ISO-ANSI Working Draft: SQL Multimedia and Application Packages (SQL/MM) Part2: Full-Text, September 1995.

[ISO 1996] ISO (International Organization for Standardization) & ANSI (American National Standards Institute): ISO/IEC JTC 1/SC 21/SQL/MM MAD-007. ISO-ANSI Working Draft: SQL Multimedia and Application Packages (SQL/MM) Part5: Still Image, Juni 1996.

[Kosch & Döller 2005] Kosch, Harald; Döller, Mario: MPEG: Überblick und Integration in

Multimedia-Datenbanken. Datenbank-Spektrum, 5 (15):26-35, 2005.

[Lu 1999] *Lu, Guojun*: Multimedia Database Management Systems. Artech House, Boston, London, 1999.

[Lu et al. 2005] *Lu, Jing; Ma, Shao ping; Zhang, Min*: Automatic image annotation based-on model space. In: Proc. of IEEE Int. Conf. on Natural Language Processing and Knowledge Engineering, 2005, S. 455-460.

[Marques & Furht 2002] *Marques, O.; Furht, B.*: Content-Based Image and Video Retrieval. Kluwer Academic Publishers, Boston, Dordrecht, London, 2002.

[Miller et al. 1990] *Miller, G.; Beckwith, R.; Fellbaum, C.; Gross, D.; Miller, K.*: Five papers on WordNet. Int. Journal of Lexicography, 3(4), 1990.

[Mirmehdi & Petrou 2000] *Mirmehdi, Majid; Petrou, Maria*: Segmentation of Color Textures. IEEE Trans. Pattern Anal. Mach. Intell., 22(2):142-159, 2000.

[Natsev et al. 2005] *Natsev, A. P.; Naphade, M. R.; Tesic, J.*: Learning the Semantics of Multimedia Queries and Concepts from a Small Number of Examples. In: Proc. of the 13th ACM Int. Conf. on Multimedia, ACM Press, 2005, S. 598-607.

[Pauwels & Frederix 2000] *Pauwels, Eric J.; Frederix, Greet*: Image Segmentation by Nonparametric Clustering Based on the Kolmogorov-Smirnov Distance. In: David Vernon (Hrsg.), Computer Vision – ECCV 2000, 6th European Conference on Computer Vision, Dublin, Ireland, June 26 - July 1, 2000, Proceedings, Part II, LNCS, Band 1843, Springer-Verlag, 2000, S. 85-99.

[Robertson 1977] *Robertson, S. E.*: The probability ranking principle. Journal of Documentation, 33:294-304, 1977.

[Rocchio 1971] *Rocchio, J. J. Jr.*: Relevance Feedback in Information Retrieval. In: G. Salton (Hrsg.), The SMART Retrieval System – Experiments in automatic Document Processing. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1971, Kapitel 14, S. 313-323.

[Rubner & Tomasi 2001] *Rubner, Yossi; Tomasi, Carlo*: Perceptual Metrics for Image Database Navigation. Kluwer Academic Publishers, Boston, Dordrecht, London, 2001.

[Ruthven & Lalmas 2003] *Ruthven, I.; Lalmas, M.*: A survey on the use of relevance feedback for information access systems. Knowledge Engineering Review, 18(2):95-145, 2003.

[Schäuble 1997] *Schäuble, Peter*: Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases. Kluwer Academic Publishers, Boston, Dordrecht, London, 1997.

[Schmitt 2002] *Schmitt, Ingo*: Einführung in Multimedia-Datenbanken: Tutorial. Datenbank-Spektrum, 1 (4): S. 28-35, 2002.

[Schmitt 2005] *Schmitt, I.*: Ähnlichkeitssuche in Multimedia-Datenbanken: Retrieval, Suchalgorithmen und Anfragebearbeitung. Oldenbourg Wissenschaftsverlag, München, 2005.

[Schmitt & Schulz 2004] *Schmitt, I.; Schulz, N.*: Similarity Relational Calculus and its Reduction to a Similarity Algebra. In: Dietmar Seipel; J. M. Turull-Torres (Hrsg.), Third Intern. Symposium on Foundations of Information and Knowledge Systems (FoIKS'04), Austria, February 17-20, LNCS, Band 2942, Springer-Verlag, Berlin, Heidelberg, 2004, S. 252-272.

[Schmitt et al. 2005] *Schmitt, I.; Schulz, N.; Herstel, T.*: WS-QBE: Eine QBE-Anfragesprache für komplexe Ähnlichkeitsanfragen. KI – Künstliche Intelligenz: Knowledge Representation, Reasoning and Management, (1):44-49, Februar 2005.

[Schulz 2004] *Schulz, N.*: Formulierung von Nutzerpräferenzen in Multimedia-Retrieval-Systemen. Dissertation, Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, 2004.

[Schulz & Schmitt 2003] *Schulz, N.; Schmitt, I.*: Relevanzziehung in komplexen Ähnlichkeitsanfragen. In: G. Weikum, H. Schöning, E. Rahm (Hrsg.), Datenbanksysteme in Business, Technologie und Web, BTW'03, 10. GI-Fachtagung, Leipzig, Februar 2003, Lecture Notes in Informatics (LNI) Volume P-26, Bonn, Gesellschaft für Informatik, 2003, S. 187-196.

[Smeulders et al. 2000] *Smeulders, A. W. M.; Worring, M.; Santini, S.; Gupta, A.; Jain, R.*: Content-Based Image Retrieval at the End of the Early Years. IEEE Trans. on Pattern Analysis and Machine Intelligence, (PAMI), 22(12), Dezember 2000.

[Stolze 2002] *Stolze, K.*: Still Image Extensions in Database Systems – A Product Overview. Datenbank-Spektrum, 2(2):40-47, 2002.

[Trotman & Sigurbjörnsson 2005] *Trotman, Andrew; Sigurbjörnsson, Börkur*: Narrowed Extended XPath I (NEXI). In: Norbert Fuhr, Mounia Lalmas, Saadia Malik, Zoltan Szlavik (Hrsg.), Advances in XML Information Retrieval, Third International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004, Dagstuhl Castle, Germany, December 6-8, 2004, Revised Selected Papers, LNCS, Band 3493, Springer-Verlag, 2005, S. 16-40.

[Veltkamp & Tanase 2000] *Veltkamp, Remco C.; Tanase, Mirela*: Content-Based Image Retrieval Systems: A Survey. Technischer Bericht UU-CS-2000-34, CS Dept., Utrecht University, 2000.

[Veltkamp et al. 2001] *Veltkamp, R. C.; Burkhardt, H.; Kriegel, H.-P. (Hrsg.)*: State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer Academic Publishers, Boston, Dordrecht, London, 2001.

[Weber et al. 1998] *Weber, R.; Schek, H.-J.; Blott, S.*: A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In: A. Gupta, O. Shmueli, J. Widom (Hrsg.), Proc. of the 24st Int. Conf. on Very Large Data Bases (VLDB'98), New York City, August 24-27, 1998, San Francisco, CA, Morgan Kaufmann Publishers, 1998, S. 194-205.

[Zadeh 1988] *Zadeh, Lofti A.*: Fuzzy Logic. IEEE Computer, 21(4):83-93, April 1988.



Ingo Schmitt arbeitet in der Datenbankgruppe der Otto-von-Guericke-Universität Magdeburg. Er promovierte 1998 zum Thema Schemaintegration für den Entwurf föderierter Datenbanken. Im Jahr 1999/2000 war er als Assistant Professor an der University of Windsor in Kanada tätig. Seit 1999 forscht er auf dem Gebiet Multimedia-Datenbanken. Aus seiner Habilitation Ende 2004 ging ein Lehrbuch über Ähnlichkeitssuche in Multimedia-Datenbanken hervor.



Andreas Nürnberger ist seit 2003 Juniorprofessor für Information Retrieval und Emmy-Noether-Nachwuchsgruppenleiter an der Otto-von-Guericke-Universität Magdeburg. Er hat an der Technischen Universität Braunschweig bis 1996 Informatik mit Nebenfach Betriebswirtschaftslehre studiert und promovierte 2001 zum Dr.-Ing. an der Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg. Direkt nach seiner Promotion ging er für zwei Jahre als Postdoctoral Researcher an die University of California, Berkeley. Der Schwerpunkt seiner aktuellen Forschungsarbeiten liegt auf der Entwicklung von Methoden zur besseren Nutzerunterstützung in Information-Retrieval-Systemen.

Priv.-Doz. Dr. Ingo Schmitt
 Jun.-Prof. Dr.-Ing. Andreas Nürnberger
 Otto-von-Guericke-Universität Magdeburg
 Fakultät für Informatik
 Universitätsplatz 2
 39106 Magdeburg
 schmitt@iti.cs.uni-magdeburg.de
 http://www.witi.cs.uni-magdeburg.de
 nuernb@iws.cs.uni-magdeburg.de
 http://irgroup.cs.uni-magdeburg.de