Image Retrieval für klinische Bilddatenbanken

Jörg Dahmen[1], Thomas Lehmann[2], Klaus Spitzer[2], Hermann Ney[1]

[1] Lehrstuhl für Informatik VI der RWTH Aachen,
Ahornstraße 55, D-52056 Aachen
e-mail: dahmen@informatik.rwth-aachen.de

[2] Institut für Medizinische Informatik der RWTH Aachen,
Pauwelsstraße 30, D-52057 Aachen
e-mail: lehmann@imib.rwth-aachen.de

Zusammenfassung. Die ständig zunehmende Menge an multimedialen (Bild-)Daten in den verschiedensten Anwendungsbereichen stellt immer höhere Anforderungen an den Zugriff auf diese. Da eine manuelle Indizierung immer aufwendiger wird, sind in der Vergangenheit zahlreiche Systeme und Algorithmen entwickelt worden, mit denen Indizierung und Abfrage solcher Datenbanken erleichtert werden sollen. Ein möglicher Ansatz ist der Bildzugriff durch Bildanalyse, der eine automatische Indizierung von Bilddaten erlaubt. In dieser Arbeit wird ein Überblick über aktuelle Entwicklungen auf diesem Gebiet gegeben, insbesondere im Hinblick auf eine klinische Anwendbarkeit, und es wird ein neues Verfahren zur Leistungsbewertung solcher Systeme vorgeschlagen, das auf dem statistischen Ansatz von *Sensitivität* und *Spezifität* beruht.

Schlüsselwörter: Image Retrieval, Sensitivität, Spezifität

1 Einführung

Heute existierende Datenbanksysteme (DBS) bieten in der Regel nur die Möglichkeit, Anfragen textuell zu formulieren, üblicherweise über die Datenbankabfragesprache SQL. Um ein solches System zur Verwaltung von Bilddaten heranzuziehen, ist es daher nötig, jedes Bild manuell mit Schlüsselwörtern zu versehen. Eine solche Beschreibung ist jedoch aufwendig und zudem weder vollständig noch eindeutig, da abhängig von Erfahrungshorizont und Intention der Benutzer unterschiedliche Schlüsselwörter herangezogen werden. Auch eine nachträgliche Indizierung mit erweiterten Attributen ist oft nicht möglich, so daß in der Vergangenheit zahlreiche Versuche unternommen worden sind, neue Bildzugriffsverfahren zu entwickeln. Diese lassen sich unter dem Oberbegriff *Content-Based Image Retrieval* oder kurz *Image Retrieval* (ImR) zusammenfassen.

2 Aufbau eines Image Retrieval-Systems

ImR-Systeme beruhen in der Regel auf dem folgenden Ansatz. Bei der Eingabe in das DBS werden aus dem Bild Merkmale extrahiert, die den Merkmalsvektor bilden. Dieser wird nun zusätzlich zum eigentlichen Bild in der DB gespeichert. Anfragen geschehen fast immer über das 'Query by Example'-Paradigma (QBE). Aufgabe des Systems ist es dabei, alle Bilder zurückzuliefern, die einem vorgegebenen Anfragebild visuell ähnlich sind. Zu diesem Zweck wird der zu dem Anfragebild gehörige Merkmalsvektor bestimmt und mittels geeigneter Distanzmaße mit den in der DB gespeicherten verglichen. Unter *Query-Refinement* (QR) versteht man dabei die sukzessive Verfeinerung einer QBE-Anfrage. Eine spezielle Variante des QR ist das *Relevance Feedback*. In diesem Fall wird die Anfrage über die manuelle Einteilung des Suchergebnisses in erwünschte und unerwünschte Bilder verfeinert.

Die Definition der Ähnlichkeit zweier Bilder geschieht in der Regel über die Merkmale Farbe, Textur, Form sowie deren räumliche Anordnung [1]. Um das Merkmal Farbe sinnvoll einsetzen zu können, ist die Wahl eines geeigneten Farbraumes sehr wichtig. Nur perzeptive Farbräume wie das HVC-Farbsystem [2] haben die Eigenschaft, daß die visuell wahrnehmbaren Farbunterschiede den geometrischen Farbdistanzen entsprechen. Dies ist aber eine elementare Voraussetzung für ein farbbasiertes ImR [3][4]. Ausgangspunkt des texturbasierten Ansatzes ist HARALICK's Grundlagenwerk aus dem Jahr 1973 [5], in dem zahlreiche Texturdeskriptoren vorgestellt werden. Formbasierte Verfahren erlauben es dem Benutzer, das gesuchte Objekt zu skizzieren. Mittels eines Gradientenbildes kann dann untersucht werden, ob ein Bild dieses Objekt enthält oder nicht. Solche Verfahren sind, beispielsweise durch Verwendung der FOURIER-MELLIN-Transformation, invariant gegenüber Rotation, Translation und Skalierung oder anderen affinen Transformationen. Einige Systeme erlauben auch eine manuelle Segmentierung der Bilder in Regions of Interest (ROI). Diese Verfahren sind aber für große Datenmengen zu aufwendig, zumal die Suche dann auf Bilder beschränkt ist, deren ROIs zuvor ebenfalls manuell extrahiert wurden.

Zur Leistungsfähigkeit dieser Ansätze stellen SCASSELLATI et al. die These auf, die bislang verwendeten Formmerkmale korrelierten in keinster Weise mit der Wahrnehmung des Menschen [6]. In der Tat erzielen diese die schlechtesten Resultate [7], wohingegen das Merkmal Farbe nach FALOUTSOS et al. die besten Ergebnisse liefert [8]. Zur Komplexitätsreduktion werden Datenstrukturen wie der G-Baum [9] oder die Voronoi-Tesselierung [10] eingesetzt, die die Anzahl der zum ImR benötigten Vergleichsoperationen von N auf $log(N)$ reduziert.

3 Vorstellung von Image Retrieval-Systemen

Im folgenden werden einige ImR-Systeme vorgestellt, die nach Kriterien wie Beispielhaftigkeit und Innovativität ausgewählt wurden.

IBM's *QBIC*-System (Query by Image Content) war die erste kommerziell verfügbare ImR-Applikation und unterstützt im wesentlichen die in Kapitel 2 vorgestellten Merkmale [11]. Diesem System sehr ähnlich ist das *MARS*-System (Multimedia Analysis and Retrieval System) der University of Illinois [12]. Schwerpunkt von MARS sind Anfragen, die auf gewichteten Kombinationen mehrerer

Merkmale basieren. Eine automatische ROI-Extraktion findet in keinem der beiden Systeme statt.

VisualSEEK (Columbia University) beschränkt sich bislang auf die Analyse von Farben und ihrer räumlichen Anordnung zueinander [13], allerdings findet eine automatische ROI-Extraktion statt. *WebSEEK* ist eine Abwandlung des Systems für den WWW-Bereich. Das *Blobworld*-Projekt [14] der University of California at Berkeley, spezialisiert auf natürliche Bilder wie Landschafts- und Tieraufnahmen, arbeitet lediglich mit den Merkmalen Farbe und Textur, wobei auch und gerade räumliche Zusammenhänge berücksichtigt werden. Die Extraktion der durch Ellipsen (genannt Blobs) beschriebenen ROIs erfolgt automatisch unter Verwendung des Expectation-Maximization Algorithmus. *Photobook* (Massachusetts Institute of Technology) erlaubt unter anderem Anfragen über die sogenannte 'Appearance' eines Bildes [15]. Appearance-Merkmale sind solche, die eine möglichst exakte Rekonstruktion des Bildes erlauben (Semantics Preserving Compression); sie werden durch Anwendung einer Karhunen-Loève-Transformation (KLT) gewonnen und beispielsweise zur Gesichtserkennung genutzt. Photobooks Formerkennungsmodul basiert auf einem Finite Elemente-Ansatz und erlaubt die Detektion von verformten 2D-Objekten. Neu ist auch der Anfrageunterstützungsagent 'Four Eyes', der den Benutzer bei der Auswahl von für ein Problem besonders geeigneten Merkmalen unterstützen soll.

Erwähnenswert ist auch das *IRIS*-Projekt (Image Retrieval for Information Systems), das 1994 als Kooperation von IBM und der Universität Bremen gestartet wurde [16] und heute unter dem Name *ImageMiner* weiterentwickelt wird. Durch den Einsatz einer regelbasierten Wissensbank können Bilder gewisser Domänen automatisch textuell beschrieben werden. Damit ist IRIS das bislang einzige ImR-System, welches auf semantischer Ebene arbeitet [7]. Systeme wie *Virage* (Virage, Inc, San Mateo, USA) oder *CORE* (National University of Singapore) sind keine eigenständigen ImR-Applikationen, stellen aber Datenstrukturen und Funktionalitäten zur Verfügung, mit denen solche implementiert werden können.

4 Systemevaluierung über Sensitivität und Spezifität

Die Effizienz von ImR-Systemen wird üblicherweise über die Größen *Precision* (P) und *Recall* (R) definiert [14]. P gibt an, wieviel Prozent der auf eine Anfrage zurückgelieferten Bilder tatsächlich korrekt sind, R hingegen, wieviel Prozent der Bilder in der DB, die die Anfrage erfüllen, zurückgeliefert wurden. Man beachte, daß bei diesen Effizienzmaßen die Größe der DB unberücksichtigt bleibt, obwohl es offensichtlich ein erheblicher Unterschied ist, ob ein ImR-System aus einer Gesamtmenge von 32 oder 1000 Bildern eine Menge von M Bildern extrahiert.

Ein Leistungsvergleich der vorgestellten Systeme ist nicht möglich, da das Konzept von P und R einerseits Schwächen aufzeigt und andererseits nicht von allen Autoren genutzt wird. Darüberhinaus existieren noch keine standardisierten Trainings-/ und Testcorpora. Aussagen über die Leistungsfähigkeit heutiger Systeme auf medizinischen Bilddaten gibt es somit praktisch nicht.

Als Bewertungskriterium für die Effizienz eines ImR-Systems bietet sich das statistische Bewertungskonzept von *Sensitivität* (SE) und *Spezifität* (SP) an, welches in der Medizin zum Vergleich von Diagnoseverfahren herangezogen wird [2]. Die Größen P und R sind in diesem Zusammenhang völlig analog definiert, P wird nun aber als *positiver Vorhersagewert* (PV), R als *Sensitivität* bezeichnet. Zusätzlich wird jedoch noch der Begriff der *Spezifität* eingeführt. Diese ist, übertragen auf eine Anwendung im Bereich ImR, definiert als $SP = rn/(rn+fp)$. Dabei ist rn die Anzahl der korrekterweise nicht zurückgelieferten Bilder (also Bilder, die dem Anfragebild unähnlich sind und auch nicht zurückgeliefert wurden) und fp die Anzahl der Bilder, die zwar zurückgeliefert wurden, aber falsch sind.

In Verbindung mit standardisierten, gelabelten Testcorpora ermöglichen die Kriterien SE, SP und PV nun einen aussagekräftigen Vergleich verschiedener ImR-Applikationen, was am folgenden Beispiel gezeigt werden soll. Liefert ein ImR-System 16 Bilder zurück, von denen 4 korrekt sind und waren insgesamt 5 korrekte in der DB, so bedeutet dies stets $R = 80\%$, $P = 25\%$. Enthielte die DB 32 Bilder, so entspräche dies $SE = 80\%$ und $SP = 55.6\%$. Wird das gleiche Ergebnis auf 1000 Bildern erzielt, so gilt $SP = 98.8\%$, bei gleichbleibendem $SE = 80\%$. Der Tatsache, daß die Komplexität eines ImR-Problems proportional zur Anzahl der Bilder in der DB wächst, wird also durch SP Rechnung getragen.

5 Zusammenfassung und Ausblick

Wie bereits die große Anzahl an Publikationen in diesem Bereich zeigt, ist ImR nach wie vor ein aktuelles Forschungsgebiet. Obwohl eine experimentelle Untersuchung noch aussteht kann prognostiziert werden, daß heutige Systeme in einer Domäne wie der Medizin versagen werden; sie scheinen eher für heterogene Bildsammlungen, wie etwa ein privates Photoalbum, geeignet, da dem bislang effektivstem Merkmal 'Farbe' [8] nur hier diskriminative Eigenschaften zukommen. Es besteht also ein Bedarf an neuen ImR-Algorithmen bzw. an der Weiterentwicklung der bestehenden. Darüberhinaus ist die Entwicklung von standardisierten Testcopora zum Leistungsvergleich verschiedener Systeme unumgänglich. Interessant erscheint auch eine Anwendung der KLT auf medizinische Daten, allerdings sollte, da eine Klassentrennung angestrebt wird, das Verfahren um eine sich anschließende Lineare Diskriminanzanalyse (LDA) erweitert werden [10]. Abschließend sei bemerkt, daß ein ImR-System nur dann in der klinischen Routine eingesetzt werden kann, wenn es Standards wie DICOM oder HL-7 unterstützt. Bislang erfüllt kein ImR-System diese Anforderung.

Literatur

1. M.D. Marsicoi, L. Cinque, S. Levialdi, "Indexing Pictorial Documents by their Content: A Survey of Current Techniques", Image and Vision Computing 15, pp. 119-141, 1997

2. T. Lehmann, W. Oberschelp, E. Pelikan, R. Repges, “Bildverarbeitung für die Medizin”, Springer Verlag Berlin, 1997
3. J.R. Smith, S.F. Chang, “Tools and Techniques for Color Image Retrieval”, Procs. SPIE, Vol. 2670, pp. 426-437, 1996
4. H.J. Zhang, Y. Gong, C.Y. Low, S.W. Smoliar, “Image Retrieval based on Color Features: An Evaluation Study”, Procs. SPIE, Vol.2606, pp. 212-220, 1995
5. R.M. Haralick, K. Shanmugam, I. Dinstein, “Textural Features for Image Classification”, IEEE Trans. on Systems, Man and Cybernetics, Vol.3, No.6, pp. 610-621, 1973
6. B. Scassellati, S. Alexopoulos, M. Flickner, “Retrieving Images by 2-D Shape: A Comparison of Computation Methods with Human Perceptual Judgements”, Procs. SPIE, Vol.2185, pp. 2-14, 1994
7. J.P. Eakins, “Automatic Image Content Retrieval - Are we Getting Anywhere?”, 3rd Int. Conf. Electronic Library and Visual Information Research, De Montford University, Milton Keynes, UK, pp.123-135, 1996
8. C. Faloutsos, R. Barber, M. Flickner, J. Hafner et al., “Efficient and Effective Querying by Image Content”, Journal of Intelligent Information Systems, Vol.3, pp. 231-262, 1994
9. T.C. Wu, J. Cheng, “Retrieving similar Pictures from Iconic Databases using G-Tree”, Pattern Recognition Letters 18, pp. 595-603, 1997
10. D.L. Swets, J.J. Weng, “Efficient Content-Based Image Retrieval using Automatic Feature Selection”, Procs. International Symposium Computer Vision, Coral Gables, Florida, pp. 85-90, Nov. 1995.
11. W. Niblack, R. Barber, W. Equitz, M. Flickner et al., “The QBIC Project: Querying Images by Content using Color, Texture and Shape”, Procs. SPIE, Vol.1908, pp. 173-187, 1993
12. S. Mehrotra, Y. Rui, M. Ortega-Binderberger, T.S. Huang, “Supporting Content-Based Queries over Images in MARS”, IEEE Int. Conf. Multimedia Computing and Systems, Chateau Laurier, Canada, pp. 623-633, June 1997
13. J.R. Smith, S.F. Chang, “VisualSEEK: A Fully Automated Content-Based Image Query System”, ACM Multimedia 96, Boston, USA, pp. 87-98, Nov. 1996
14. C. Carson, S. Belongie, H. Greenspan, J. Malik, “Color- and Texure-Based Image Segmentation using EM and its Application to Image Querying and Classification”, to appear in: IEEE Pattern Analysis and Machine Intelligence
15. A. Pentland, R.W. Picard, S. Scarloff, “Photobook: Content-Based Manipulation of Image Databases”, Procs. SPIE, Vol.2185, pp. 34-47, 1994
16. P. Alshuth, T. Hermes, C. Klauck, J. Kreyß, M. Röper, ”IRIS - A System for Image and Video Retrieval”, Center for Advanced Studies Conference ’96, Toronto, Canada, Abstract: pp. 208, Paper: CDROM, Nov. 1996