

Content-Based Image Retrieval, bildinhaltsbasiertes Suchen in großen Bilddatenbanken

Henning Müller, David McG Squire, Wolfgang Müller
Henning.Mueller@cui.unige.ch

Computer Vision Group,
Centre Universitaire D'Informatique
Universität Genf
24 Rue du Général Dufour
CH-1211 Genf 4, Schweiz

Betreuer der Arbeit: Prof. Dr. Thierry Pun
Art der Arbeit: Doktor-Arbeit
Fachbereich GI: 2

Zusammenfassung

Dieser Artikel beschreibt einen neuen Ansatz im Bereich des content-based image retrieval (CBIR), dem bildinhaltsbasierten Suchen in Bilddatenbanken in der Regel ohne Annotationen. Gegenüber den herkömmlichen meist vektorbasierten Verfahren werden hier Methoden des Text oder Information Retrieval (IR) an die speziellen Bedürfnisse des Empfangs von Bildern angepasst. Benutzerexperimente belegen die Leistungsfähigkeit und Flexibilität des Verfahrens.

1. Einführung

Der Forschungsbereich des Content-Based Image Retrieval erfreut sich seit Anfang der 90er Jahre großer Beliebtheit. Die meisten Systeme wie IBMs QBIC (Query by Image Content) [3] suchen zu einem vorgegebenen Bild ähnliche Bilder aus einer Bilddatenbank. Andere Eingabemöglichkeiten für die Bildsuche sind Formen von Objekten oder Regionen in Bildern oder sogenannte BLOBs aus vorsegmentierten Bildern [1]. Verschiedene Systeme erlauben es, ein Bild oder mehrere Bilder des Anfrageergebnisses als positives und negatives Feedback auszuwählen und erleichtern so dem Benutzer das Browsen in der Datenbank. Durch die große Anzahl von Bildern z. B. im World Wide Web ist eine manuelle Indexierung quasi unmöglich geworden und Verfahren für die Verwaltung und das Auffinden von Bildern unumgänglich.

Das Text oder Information Retrieval (IR) ist dagegen schon eine sehr "alte" Forschungsdisziplin, die es seit etwa 40 Jahren gibt. Die Anfänge stammen vor allem aus Amerika aus dem Büchereiwesen und dem juristischen Bereich, in dem wegen des Case-Law große Textmengen verarbeitet werden müssen. Dieser Forschungsbereich hat eine lange Erfahrung in der Indexierung und dem effektiven Zugriff auf große Datenmengen. Diese Erfahrung wurde bisher leider von den CBIR Forschern weitgehend ignoriert. Viele der im CBIR benutzten Verfahren wie *Relevance Feedback* oder auch Evaluierungsverfahren wie *Precision/Recall (PR) Graphen* existieren im IR bereits seit längerer Zeit. Es erscheint daher logisch, systematisch die Methoden des IR auf eine Anwendbarkeit im CBIR zu untersuchen.

2. Systembeschreibung

Wir benutzen für den Zugriff auf die Bilder die bevorzugte Dateiform des IR, den Inverted File (IF) [5]. Dieser IF erlaubt es, den Suchraum auf die in der Anfrage vorhandenen Merkmale zu begrenzen. Das gibt uns die Möglichkeit, eine weitaus größere Zahl von möglichen Merkmalen im Vergleich zu anderen, auf Vektoren fester Länge basierenden Verfahren, zu benutzen. In dem Programm, das in diesem Artikel beschrieben wird, benutzen wir über 80.000 mögliche Merkmale, die alle sehr einfach gehalten sind. Jedes Bild in der Datenbank enthält zwischen 600 und 3000 dieser Merkmale.

Für die Bildsuche in der Datenbank benutzen wir Merkmale, die der menschlichen Perzeption möglichst ähnlich sind, und die die beiden Bereiche Farbe und Textur hinreichend gut beschreiben [4]. Zum globalen Farbvergleich benutzen wir eine Histogramm-Intersektion. Um der menschlichen Perzeption ähnlicher als RGB (Red, Green, Blue) zu sein, benutzen wir den HSV Farbraum (Hue, Saturation, Value) und quantisieren den Bereich in 16 Farbtöne, drei Sättigungen und drei Helligkeiten, sowie in zusätzlich vier Graustufen. Außerdem benutzen wir lokale Farbmerkmale in verschiedenen Größen. Dazu teilen wir das Bild in vier gleich große Bereiche und berechnen für jeden Bereich die häufigste Farbe als Farbmerkmal für diese Region. Anschließend wird jeder der vier Blöcke geteilt und das ganze für die Teilblöcke berechnet. Dieser Vorgang des Teilens wird viermal wiederholt. Für die Charakterisierung der Textur benutzen wir Gabor Filter, die nach [2] der menschlichen Wahrnehmung von Kanten am ähnlichsten sind. Hiermit berechnen wir die Energien von vier verschiedenen Orientierungen mit je drei verschiedenen Größen im Bild. Basierend auf der Energie des Filters werden die Werte in zehn verschiedene Gruppen eingeteilt. Ähnlich wie bei den

Farbmerkmalen werden die Gabor Filter nicht nur global sondern auch lokal berechnet. Ist die Energie unter einem Schwellenwert, so wird sie nicht betrachtet.

Für jedes Merkmal in einer Datenbankanfrage wird außerdem ein Gewichtungsfaktor berechnet, der sich aus Häufigkeit des Merkmals im Bild und in der Query, sowie dem Anteil der Dokumente in der Datenbank, die dieses Merkmal besitzen, zusammensetzt. Grundlage der Bewertung ist, daß

- ein Merkmal, das häufig in einem Bild ist, dieses Bild gut beschreibt,
- ein Merkmal, was in sehr vielen Bildern auftritt, Bilder nicht gut voneinander unterscheidet.

In Abbildung 2 werden die Ergebnisse verschiedener im IR benutzter Gewichtungsschemata gegenübergestellt. Die Formel für eines der in unseren Experimenten sehr guten Verfahrens, der *Classical Inverse Document Frequency*, lautet wie folgt :

$$score_{kq} = \sum_j \left(\left(\frac{1}{N} \left(\sum_{i=1}^N tf_{ij} R_i \right) \right) tf_{kj} \left(\log \left(\frac{1}{cf_j} \right) \right)^2 \right)$$

tf	- Häufigkeit eines Features in einem Bild	q	- Query mit $i=1..N$ Eingabebildern
cf	- Häufigkeit in der Kollektion	k	- Nummer des Ergebnisbildes
j	- Nummer eines Features	R	- Relevanz eines Eingabebildes

Wir sehen, daß sich die Bewertung eines Bilder aus einer Summe über alle in der Query vorhandenen Merkmale zusammensetzt. Die Häufigkeit des Merkmals in der Query und im Bild sind lineare Faktoren, während die Häufigkeit eines Merkmals in der gesamten Kollektion wesentlich stärker bewertet wird (\log^2). Methoden zum schnellen und effektiven Zugriff auf Bilddaten mit der Inverted File Technik können in [6] nachgelesen werden. Eine Demoversion des Systems findet sich unter [7].

3. Benutzerexperiment und Ergebnisse

Als Benutzerexperiment haben wir 14 Bilder aus der Bilddatenbank als Anfragen ausgewählt. Drei Benutzer haben zu jedem dieser Bilder alle für sie relevanten Bilder aus der Datenbank herausgesucht. Das Ergebnis ist, daß die Anzahl der ausgewählten Bilder abhängig von der Person sehr unterschiedlich ist (zwischen 4 und 32), und daß zum Teil sehr unterschiedliche Bilder ausgesucht werden. Im Mittel wurden 19 Bilder als relevant markiert. Als Bewertung der Performance des Systems benutzen wir *Precision/Recall (PR) Graphen*, die im Text Retrieval der Standard sind und auch im CBIR immer häufiger benutzt werden. PR Graphen beschreiben das Antwortverhalten eines Systems sehr gut. Als Datenbank benutzen wir 2500 Bilder des Schweizer Fernsehens "Télévision Suisse Romande". Die Datenbank enthält sehr verschiedene Bilder, kleine Gruppen von wenigen untereinander ähnlichen Bildern kommen vor (siehe Abbildung 1).



Abbildung 1 : Beispielbilder aus der Datenbank des Schweizer Fernsehens

Die in Abbildung 2 gezeigten Graphen sind über alle 14 Queries und alle drei Benutzer gemittelte *PR Graphen*. Die Ergebnisse für die einzelnen Queries weisen ein sehr unterschiedliches Verhalten für einfache und schwere Anfragen auf.

Als einfaches (automatisiertes) *Relevance Feedback* haben wir alle vom Benutzer als relevant angesehenen Bilder der 20 höchstbewerteten Bilder als Eingabe für eine zweite Anfrage genommen. 20 Antwortbilder wird als eine Standardmenge von Bildern beim Suchen in Datenbanken angenommen, weil es sich dabei um eine auf dem Bildschirm gut überblickbare Menge handelt. Mit diesem Verfahren des automatisierten Feedbacks kann man die Flexibilität des Programmes bezüglich der Benutzerwünsche messen.

Wie sich in Abbildung 2 zeigt, verbessert sich die Performance durch *Relevance Feedback* signifikant, das Verfahren paßt sich also während der Query gut an die Wünsche und Bedürfnisse verschiedener Benutzer an. Erste Experimente zeigen außerdem, daß ein erfahrener Benutzer vor allem mit negativem aber auch mit weitergehendem positivem Feedback deutlich bessere Ergebnisse erzielen kann.

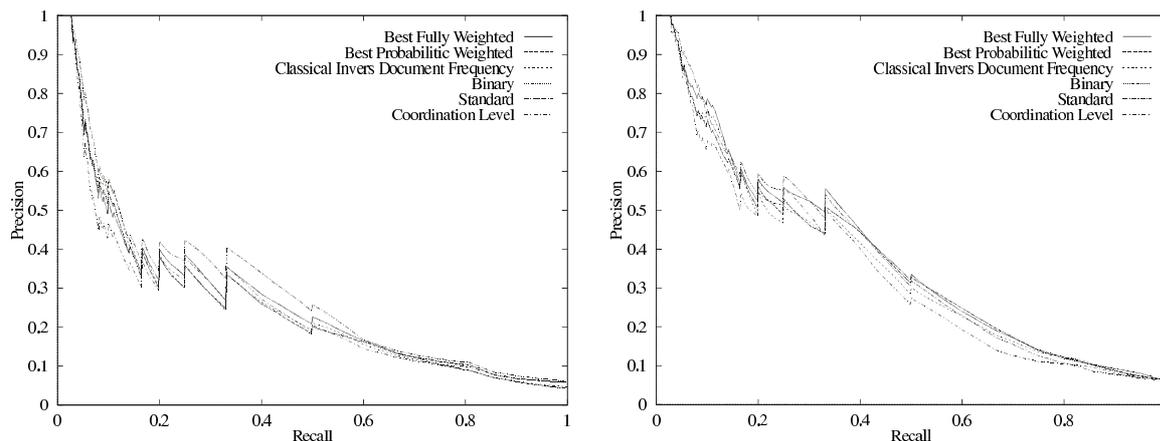


Abbildung 2 : Precision/Recall Graph ohne (a) und mit (b) Relevance Feedback

4. Ausblick

Es wird mit diesem Artikel gezeigt, daß viele Techniken aus dem IR auch im CBIR verwendet und somit hochdimensionale Merkmalsräume mit schnellem Zugriff benutzt werden können. Die Ergebnisse in Abbildung 2 zeigen die Stärken dieser Technik vor allem in der Benutzung von Relevance Feedback. Aber auch schon ohne Feedback sind im Schnitt alle als relevant markierten Bilder in den ersten 20 Prozent der Datenbank vertreten, und in den ersten 20 Bildern sind im Schnitt schon etwa sieben relevante Bilder, mit Relevance Feedback sogar etwa neun.

Potential für Verbesserungen erwarten wir neben dem Relevance Feedback auch in der Verbesserung der eingesetzten Features. Vor allem durch Segmentierung kann man mit einer Gruppe sehr selektiver, regionaler Merkmale rechnen. Mit einer Höherbewertung von "erfolgreichen" Features kann die Leistung des Systems sicherlich erhöht werden.

Einsatzbereich ist neben dem Zeitungs- und Fernseh-Journalismus auch die medizinische Bildverarbeitung wie z.B. die Dermatologie und Pathologie. Sowohl in der Pathologie als auch in der Dermatologie werden große Mengen von Bildern verglichen. Ein Tool zum automatischen Finden von ähnlichen Mustern oder Bildern mit ähnlichen Diagnosen kann sowohl in Lehre als auch in der Diagnose sinnvoll eingesetzt werden.

Literaturverzeichnis

- [1] Chad Carson, Serge Belongie, Hayit Greenspan and Jitendra Malik. Region-based image querying. In *Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, June 1997.
- [2] John G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1148-1161, 1993.
- [3] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele and Peter Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23-32, Septmeber 1995.
- [4] John R. Smith and Shi-Fu Shang. Tools and techniques for color image retrieval. In Ishwar K. Sehti and Ramesh C. Jain, editors, *Storage & Retrieval for Image and Video Databases IV*, volume 2670 of *IS&T/SPIE Proceedings*, pages 426-437, San Jose, CA, USA, March 1996.
- [5] David McG. Squire, Wolfgang Müller, Henning Müller and Jilali Raki. Content-based query of image databases, inspirations from text retrieval: Inverted files, frequency based weights and relevance feedback. In *The 10th Scandinavian Conference on Image Analysis (SCIA'99)*, Kangerlussuaq, Greenland, June 7-11 1999.
- [6] Henning Müller, David McG. Squire, Wolfgang Müller and Thierry Pun. Efficient access methods for content-based image retrieval with inverted files. In S. Panchanathan, S.-F. Chang and C.-C. J. Kuo, editors, *Multimedia Storage and Archiving Systems IV (VV02)*, volume 3846 of *SPIE Proceedings*, Boston, Massachusetts, USA, September 20-22 1999.
- [7] Web page: http://cuiwww.unige.ch/~viper/viper_test.pl, 1999