

Stand der Forschung zu Content Based Image Retrieval (CBIR) Systemen aus Sicht der Bildverarbeitung

Content Based Image Retrieval (CBIR) systems from an image processing's point of view

Dipl.-Inf. Thomas Lambeck
Applied Knowledge Representation and Reasoning Group
Artificial Intelligence Institute
Room 2037, Nöthnitzer Straße 46, Dresden-Räcknitz
Tel.: +49 (0) 351 463 38 430, Fax: +49 (0) 351 463 38 432
E-mail: thomas.lambeck@inf.tu-dresden.de, Internet: <http://awv.inf.tu-dresden.de/>

Zusammenfassung:

Ziel dieses Vortrages soll es sein, innerhalb des vorliegenden Kontextes einen einführenden Überblick der Grundlagen von CBIR Systemen zu geben sowie deren aktuelle Entwicklungen zu skizzieren. Dabei soll neben der Grundfrage, welche Zielstellungen mit diesen Systemen verfolgt werden und welche Beschränkungen herkömmlichen Suchverfahren innewohnen, vor allem ein schemenhafter Abriss der an einem solchen Prozess beteiligten Komponenten präsentiert werden. Dies beinhaltet etwa übliche Anfragemodelle, die Art der verwendeten Merkmale, verschiedene Abstraktionsstufen sowie weiterhin Elemente der Ergebnispräsentation und der in diesem Zusammenhang nötigen Navigation innerhalb der Ergebnismenge. Abschließend folgen einige Ausführungen zu existierenden Systemen hinsichtlich ihrer Charakterisierung und Praxistauglichkeit.

Abstract:

In this talk it's the author's goal to give an overview of the basics of CBIR systems. Next to the fundamental question of use cases such systems may aim at and regarding the limits of traditional search methods, a schematic introduction on the components participating in such processes will be presented. This includes common query types, selection and the use of features, different levels of abstraction as well as elements involved in result presentation and the required result set navigation. Finally, some thoughts on existing systems with regard to practical usage are given.

Dem Menschen fällt es vergleichsweise leicht, den auf Bildern dargestellten Inhalt zu erkennen und zu verarbeiten. Dabei spielen sich in unserem Gehirn komplexe Vorgänge ab, die neben grundlegenden Elementen wie Farbwahrnehmung, Erfassung der Gestalt, Größe und Position der abgebildeten Objekte vor allem vorhandenes Erfahrungswissen und Kontextinformationen verwenden. Diese Kontexte spielen unter anderem in der richtigen Zuordnung von Objekten bzw. korrekten Interpretation von Symbolen eine wesentliche Rolle. Als Beispiel hierfür kann ein Fall der deutschen Sprache herangezogen werden. In ihr treten mitunter Bedeutungsüberschneidungen auf, d.h. ein Symbol besitzt mehrere Semantiken. Dies wird etwa bei dem Wort „Bank“ sichtbar, dass je nach Kontext entweder als Sitzgelegenheit oder aber als Kreditinstitut interpretiert werden kann. Dieses Beispiel ist entsprechend auf visuelle Fragestellungen übertragbar.

Ein deutlicher Unterschied zwischen Mensch und Maschine wird gerade beim Einsatz von Kontext- und Erfahrungswissen sichtbar. Wird einem Probanden beispielsweise ein Foto eines Fußballspiels präsentiert, so fällt es ihm relativ leicht, die dargestellte Szenerie zu erkennen und die abgebildeten Elemente zu benennen. Dabei spielt das über Jahre gesammelte Erfahrungswissen dieser Person, seien es vergleichsweise einfache Sachverhalte wie etwa die grüne Farbe

des abgebildeten Grassbodens oder weitaus komplexere abstrakte Konzepte wie ein Fußballspiel mitsamt dem zugehörigen Regelwerk, eine wichtige Rolle bei der Szenenerkennung.

Gleichwohl schwieriger ist die gleiche Aufgabe für einen Computer zu bewältigen. Dieser verfügt zunächst über keinerlei Wissen, muss also ebenfalls angeleitet und trainiert werden. Je nach Qualität und Quantität dieses Lernprozesses kann man mit heutigen CBIR Systemen gute Ergebnisse erzielen, wenn es etwa um die Herausforderung geht, zu einem vorgelegten Bild aus einer vorhandenen Menge ähnliche Exemplare zu finden. Hierbei zeigt sich, dass die Aufgabenstellung der Bildsuche im Kontext des dargestellten Inhaltes nicht allgemein definierbar ist. Je nach Art der Anfrage an das System sowie der angestrebten Zielstellung differieren die Systeme bzw. die zugrunde liegenden Konzepte mitunter erheblich. Hierbei können zwei Hauptzweige unterschieden werden: zum einen die Ähnlichkeitssuche, zum anderen die Annotation.

In beiden Fällen ergibt sich zunächst die Wahl der Anfrageform. Üblich sind hierbei die Vorgabe eines konkreten Bildes, die Auswahl eines dem Nutzer gemäß seiner Zielsetzung entsprechenden Exemplars aus einer Bibliothek, das Skizzieren von Objekten durch den Anwender, die Vorgabe von Farbverteilungen innerhalb einer Grafik oder etwa die logische Verknüpfung von Elementen aus einem visuellen Thesaurus. Natürlich sind darüber hinaus weitere Methoden möglich.

Liegt ein Ausgangsbild vor, muss eine Auswahl der zu verwendenden Merkmale getroffen werden. Eine Merkmalrepräsentation wird angestrebt, da die Bilder im Merkmalraum untereinander häufig besser separierbar sind und darüber hinaus eine komprimierte Repräsentation des Ausgangsmaterials bilden. Dabei stellen etwa Farben, Formen sowie Texturen mögliche Merkmale dar; jedoch sind durchaus komplexere Varianten denkbar und in Verwendung. Es gibt verschiedene Bestrebungen, in diesem Zusammenhang einen Beschreibungsstandard zu entwickeln und zu etablieren. Zu nennen sind hierbei etwa MPEG 7, ICONCLASS oder MRML. Wichtig ist, dass die Auswahl der Merkmale hinsichtlich der Zielstellung des Suchprozesses und damit unter Beachtung der Ähnlichkeitsdefinition erfolgt.

Legt man wiederum einem Probanden zwei farbähnliche Bilder vor, so wird dieser die Ähnlichkeit bestätigen können. In diesem einfachen Fall kann auch der Computer gute Arbeit leisten. Setzt die Zielstellung jedoch auf weitaus abstraktere Konzepte, ergeben sich für den Computer im Allgemeinen zahlreiche Probleme, die wiederum zu einem schlechteren Ergebnis führen. Während ein Mensch etwa zwei farblich völlig verschiedene Bilder einer Kategorie wie „Essen“ zuordnen kann, müsste ein Computer dazu ebenfalls über entsprechendes Kontext- und Erfahrungswissen verfügen, da eine einfache Analyse unter Verwendung von Farben als Merkmal keine ausreichende Klassifikationsgrundlage darstellt. Unter dem „Farbaspekt“ sind die beiden exemplarischen Bilder daher nicht ähnlich, betrachtet man jedoch etwa ein Konzept bzw. eine Klassifikation, so können sie durchaus ähnlich sein. Dies zeigt, dass der Abstraktionsgrad einer Aufgabe neben anderen Gegebenheiten wie der Art und der Erzeugung der verwendeten Merkmale einen wesentlichen Faktor der Leistungsfähigkeit eines CBIR Systems darstellt.

Und auch die Darstellung der in einem Suchprozess gelieferten Resultate kann auf verschiedene Weise erfolgen. Üblich sind etwa eine einfache Listendarstellung, die topologische Anordnung basierend auf der im Bild dominierenden Farbe oder auch eine räumliche Anordnung. Wiederum ist eine Vielzahl weiterer Vorgehensweisen denkbar.

Aufgrund der Komplexität, welche mit der Aufgabe des CBIR verbunden ist, liegt eine allgemeine und umfassende Lösung dieser Problematik bisher nicht vor und wird nach Ansicht des Autors mittelfristig nicht vorliegen. Gleichwohl gibt es erfolgreiche Systeme und Konzeptstudien, die sich in spezialisierten Domänen bewährt haben und dort hervorragende Arbeit leisten. Es bleibt abzuwarten, welche neuen Konzepte und Modelle in den nächsten Jahren entwickelt werden und inwieweit es möglich sein wird, gerade auch abstrakte Vorgänge und Konzepte ausreichend zu erfassen.