MPEG 7 als Basis für eine Suche in Multimedialen Datenbanken

MPEG-7 based Retrieval on multimedia databases

Thomas Meiers

Fraunhofer Institut für Nachrichtentechnik, Heinrich-Hertz-Institut

Einsteinufer 37, 10587 Berlin

Tel: +49 – 30 – 31002 218, Fax: +49 – 30 -31002 212

E-mail: meiers@hhi.fhg.de, Internet: www.hhi.fraunhofer.de/im

Zusammenfassung:

Gegenstand dieses Beitrages ist die Navigation und Suche in großen Bildbeständen. Die Ähnlichkeit zweier Bilder wird durch MPEG-7 basierte Deskriptoren bestimmt, mit denen die Farb- und Kantenstatistik in Bildern ausgedrückt werden kann. Mit Hilfe einer multidimensionalen Skalierung (MDS) werden Bilder entsprechend ihrer gegenseitigen Ähnlichkeit in einem dreidimensionalen Raum angeordnet. Auf diese Weise erkennt der Nutzer die Struktur des Bildbestandes. Um auch große Bildbestände in einem dreidimensionalen Raum anzeigen zu können, werden mit Hilfe von Clusteringtechniken repräsentative Bilder ausgewählt. Dabei geht man hierarchisch vor – ähnlich wie bei Landkarten mit verschiedenem Maßstab. Beginnend mit einer Auswahl von repräsentativen Bildern aus dem gesamten Bildbestand wählt der Nutzer Beispielbilder aus, die seiner Suchintention am meisten entsprechen. Mit Hilfe von Lernverfahren stellt die Suchmaschine aus den Beispielbildern eine neue verfeinerte Auswahl von Bildern zusammen. Der Vorgang wird wiederholt, bis der Nutzer die von ihm gewünschten Bilder gefunden hat.

Abstract:

In this paper we address the user-navigation through large volumes of image data. Similarity-measures based on different MPEG-7 descriptors are introduced and multidimensional scaling is employed to display images in three dimensions according to their mutual similarities. With such a view the user can easily see similarity relations between images and understand the structure of the database. In order to cope with large volumes of images a clustering technique is introduced which identifies representative image samples for each cluster. Representative images are then displayed in three dimensions using multidimensional scaling structuring. The clustering technique proposed produces a hierarchical structure of clusters - similar to street maps with various resolutions of details. The user can zoom into various cluster levels to obtain more or less details if required. Further a new query refinement method is introduced. The retrieval process is controlled by learning from positive examples from the user, often called the relevance feedback of the user.

Wachsende Menge an multimedialen Daten, insbesondere Bilder

Seit etlichen Jahren wächst die Größe von digitalen Bild- und Video-Archiven ins Gigantische. Insbesondere Zeitschriftenverlage und Nachrichtenagenturen archivieren eine Fülle von Bildern

und Videoszenen. Mit der Größe der Archive wächst auch die Schwierigkeit archivierte Bilder zu finden.

Texte reichen zur Beschreibung oft nicht aus

Das Problem der Informationssuche wird verstärkt, wenn die gesuchten Informationen, wie beispielsweise Bilder, nicht sinnvoll oder ausreichend durch Stichwörter beschrieben und gefunden werden können.

So wird z.B. der Erfolg vieler Archivierungssysteme davon abhängen, dass sich ein Nutzer die Bilder auf einem Terminal nicht nur ansehen kann, sondern auch effizient und ökonomisch so sichten kann, dass er nicht von der Flut der Bildinhalte "erschlagen" wird. Innovative Verfahren der Nutzerführung und Informationsverwaltung müssen dem Nutzer den Wunsch „von den Augen ablesen". So müssen intelligente Sortier-, Filter- und Suchalgorithmen den Nutzer beim „Gang" durch die virtuellen Bildräume unterstützen und Bilder geeignet präsentieren. Angelehnt an das natürliche Auswahlverhalten von Kunden in einem realen Warenhaus muss die Suche und Vorauswahl durch intelligente Suchmaschinen hauptsächlich mittels visueller Beschreibungen – also nicht durch Textbeschreibungen - getroffen werden.

MPEG-7 als neuer Standard zur Beschreibung multimedialer Daten

Zu diesem Zweck wurde ein neues Standardisierungsprojekt MPEG-7 initiiert, in der eine visuelle Beschreibung von multimedialen Daten definiert wird. Diese Beschreibung kann grundsätzlich jeder Art von Multimedia-Daten beigelegt werden, so dass gespeicherte Daten, die mit diesen Informationen versehen sind, indiziert und gesucht werden können. So werden Farb-, Textur- und Formmerkmale von Bildern durch so genannte Deskriptoren beschrieben, deren Syntax im MPEG-7 Standard definiert wird. MPEG-7 wurde im Dezember 2001 zum Internationalen Standard erklärt.

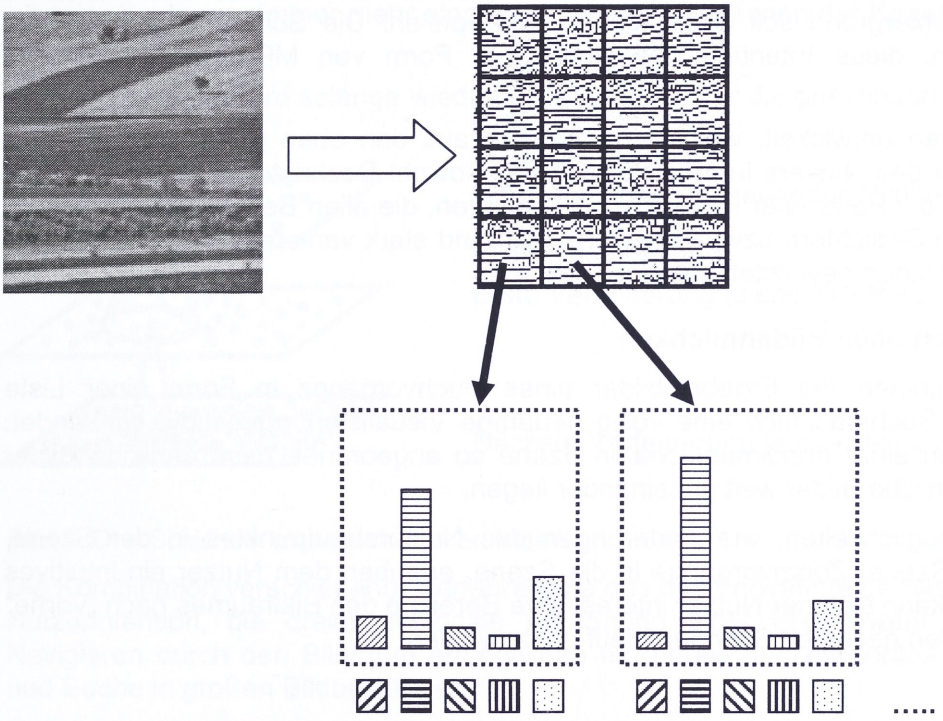


Bild 1: EdgeHistogram - Descriptor

Bild 1 zeigt einen solchen Deskriptor, der EgdeHistogram-Descriptor, der die Verteilung von Kanten in einem Bild angibt. Durch einen Filterung werden die Kanten eines Bildes berechnet (siehe oben rechts im Bild). Das Bild wird in 4x4 gleiche große Blöcke zerlegt. Für jeden Block

werden die Häufigkeiten der waagerechten, der senkrechten, der beiden diagonalen und von ungeordneten Kanten ermittelt. Man erhält also für jeden Block 5 Werte, insgesamt also 80 Werte.

Mit ähnlichen Verfahren werden Farbverteilungen ermittelt. Zum Beispiel der ScalableColor – Descriptor gibt die Häufigkeit von bestimmten Farben in einem Bild an, wobei er bzgl. der Anzahl der Farben skalierbar ist. Beim ColorLayout – Descriptor wird das Bild in 8x8 Blöcke zerlegt. Von jedem Block wird die Hauptfarbe bestimmt. Auf diese Weise erhält man eine grobe räumliche Farbverteilung des Bildes.

Neben der Beschreibung von Deskriptoren wird in dem MPEG-7 Standard auch ein Abstandsmaß für Deskriptoren angegeben. Auf diese Wiese ist es möglich, zwei Bilder bzgl. eines Deskriptors zu vergleichen. Haben zwei Bilder eine ähnliche Kantenverteilung, z.B. Aufnahmen von Gesichtern mit verschiedenem Hintergrund, so wird der Abstand in Bezug auf den EdgeHistogram – Descriptor gering sein, während er bzgl. der Farbdeskriptoren aufgrund des verschiedenen Hintergrunds groß sein kann.

Der MPEG-7 Standard definiert noch weitere Deskriptoren unter anderem für Formen von Objekten in Bildern.

Suche mit Hilfe von Beispielbildern

Ein übliches Suchverfahren ist Query by Example, kurz QBE. Hierbei sucht der Nutzer aus einer Vorgabe von Bildern diejenigen aus, die seiner Suchintention am meisten entsprechen. Anhand dieser Beispielbilder ermittelt die Suchmaschine aus dem Bildbestand diejenigen, die den Beispielbildern bzgl. von vorgegebenen Deskriptoren am ähnlichsten sind und zeigt sie. Dieser Vorgang wird so oft wiederholt bis der Nutzer die gesuchten Bilder gefunden hat.

Ein Hauptproblem bei dieser Ähnlichkeitssuche ist die sogenannte semantische Lücke zwischen der Nutzerintention und der Beschreibung der Bilder anhand von Farb- und Kantenverteilungen. Der Nutzer denkt semantisch; er sucht z.B. nach Personen als Portrait aber auch in Ganzkörperformat. Der Hintergrund soll dabei keine Rolle spielen. Die Schwierigkeit bzgl. der Suche besteht nun darin, diese Intention des Nutzers in Form von MPEG-7 Deskriptoren auszudrücken.

Hierzu wurde ein Verfahren entwickelt, welches aufbauend aus den oben genannten MPEG-7 Deskriptoren, die Intention des Nutzers lernt. Dabei werden sowohl Deskriptorenwerte statistisch derart gewichtet, dass sie die Parameter im Bild hoch gewichten, die allen Beispielbildern gemein sind (z.B. die Umrisse von Gesichtern bzw. Personen), während stark variierende Parameter, wie z.B. die Hintergrundfarbe niedrig gewichtet werden.

Virtueller Bildraum sortiert nach Bildähnlichkeit

Während viele Suchmaschinen die Ergebnisbilder eines Suchvorgangs in Form einer Liste ausgeben, wird in dieser Suchmaschine eine völlig neuartige Visualisierungstechnik verwendet. Die Bilder werden dabei in einer dreidimensionalen Szene so angeordnet, dass ähnliche Bilder benachbart sind und unähnliche Bilder weit auseinander liegen.

Völlig neue Navigationsmöglichkeiten, wie Änderungen des Nutzerstandpunktes in der Szene, beliebige Rotationen der Szene, Zoomvorgänge in die Szene, erlauben dem Nutzer ein intuitives Browsen im Bildraum. So kann sich der Nutzer interessante Bereiche des Bildraumes nach „vorne“ holen um neue Bilder für den nächsten Suchdurchlauf auszuwählen.

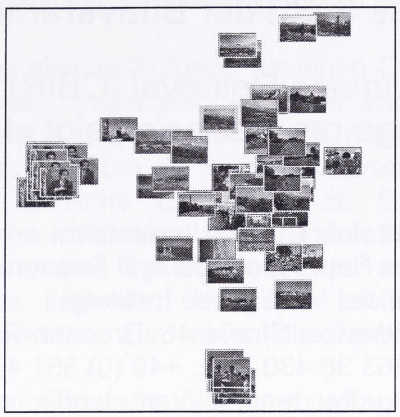


Bild 2: Bilder aufgrund ihrer Ähnlichkeit in einem dreidimensionalen Raum angeordnet

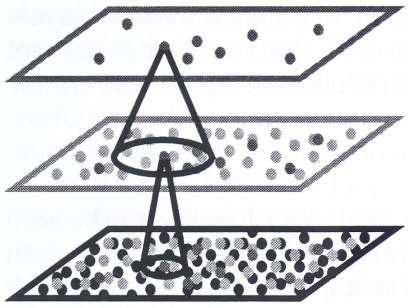
Metapher der Landkarten

Der Suchvorgang orientiert sich an der Metapher der Landkarten. So wie es Karten mit verschiedenem Maßstab gibt, werden die Bilder in verschieden fein aufgelösten Schichten mit repräsentativen Bildern aufgeteilt (siehe **Bild 3**).

Die Suche startet mit dem Zeigen aller etwa 100 Bilder der gröbsten Schicht. Sie geben einen Überblick über die Bildauswahl in dem Archiv.

Von diesen wählt der Nutzer die aus, die seinem Suchziel am ehesten entsprechen. Aus den Merkmalen der ausgewählten Bilder berechnet die Suchmaschine mit Hilfe von statistischen Verfahren die ähnlichsten Bilder aus der nächsten etwas dichteren Schicht und zeigt sie an. Dabei wird der Suchraum immer mehr eingeengt (siehe Kreisflächen der Kegel in den Schichten 2 und 3 in **Bild 3**)

Der Suchvorgang wird solange wiederholt, bis der Nutzer die passenden Bilder gefunden hat.



Überblick über den gesamten Bildbestand

Erste Verfeinerung (siehe Kreisfläche)

Nächste Verfeinerung (siehe kleine Kreisfläche)

Bild 3: Drei Schichten mit verschiedener Anzahl an Bildern.

Die Kombination verschiedener Verfahren wie das oben erwähnte Lernverfahren zur Erfassung der Nutzerintention, die dreidimensionale Anordnung einer Bildauswahl sowie das hierarchische Navigieren durch den Bildraum ermöglichen ein neues und effizientes Verfahren zur Navigation und Suche in großen Bildbeständen.

Dieses Projekt wurde durch das **Bundesministerium für Wirtschaft und Arbeit (BMWA)** gefördert.