

EIN 3D PARADIGMA FÜR BILDRECHERCHE UND -NAVIGATION

A 3D INTERACTION PARADIGM FOR IMAGE RETRIEVAL AND NAVIGATION

Christian Wilk
DaimlerChrysler AG
Forschung & Technologie 3
Postfach 2360, 89013 Ulm, Germany
fon 0731/505-2163, fax 0731/505-4110
christian.wilk@daimlerchrysler.com

Zusammenfassung:

Dieser Vortrag stellt ein Verfahren zum vereinfachten, intuitiven navigieren in Film- und Bilddatenbanken vor. Das Verfahren ist für eine bestimmte Klasse von Bildmaterial geeignet und setzt die Annotation der Bildmenge mit Metadaten voraus. Der Blickpunkt und die Blickrichtung (Point-of-View, POV) von dem aus das Bildmotiv aufgenommen wurde, wird als Index verwendet und den Bildmetadaten hinzugefügt. Der Anwender interagiert mit einem abstrahierten 3D Modell, das das Bildmotiv repräsentiert. Der POV aus dem das Modell betrachtet wird, generiert eine Abfrage an die Datenbank und liefert als Ergebnis alle Bilder, die mit einer gewissen Toleranz aus diesem POV aufgenommen wurden.

Abstract:

A method is presented for intuitive searching and navigation of an image database by means of 3D interaction. The point-of-view of the image object is used as an index to the images. The user retrieves and navigates through the images by interacting with an abstracted 3D model of the image's motif. The way the user looks at the 3D model generates a database query and retrieves all the corresponding images, which were shot from the same point-of-view.

1. Umfeld

Momentan sind etliche Standardisierungsbestrebungen im Gange, die zum Ziel haben, die für die Annotierung audiovisueller Medien verwendeten Metadaten zu standardisieren. Als Beispiel hierfür sei MPEG-7 genannt. MPEG-7 definiert eine Sammlung von XML Schemata, die multimediale Objekte beschreiben. Im Rahmen dieser Entwicklungen ist damit zu rechnen, daß innerhalb der nächsten Jahre ein Teil der neuen Bildmengen mit den dazugehörigen Metadaten erfasst wird. Die Art und Qualität der Metadaten umfasst eine große Spanne, die von den einfachsten – wie z.B. Größe und Farbverteilung des Bildes, die automatisch aus dem Bild extrahiert werden können – bis zu sehr fortschrittlichen reichen, die neuartige Kameras voraussetzen und bspw. automatisch den Ort anhand der GPS Koordinaten und den Point-of-View (POV), d.h. die Blickrichtung der Kamera aufnehmen.

2. Motivation

Aktuelle Content-Management Systeme und Multimediadatenbanken vereinfachen die Archivierung immer größerer Mengen an Bilddaten. Die Zugriffsverfahren bleiben jedoch weitgehend

die gleichen, nämlich ein Texteingabefeld, in dem ein Suchbegriff eingetragen wird. Anhand der den Bilddaten zugeordneten Metadaten werden aber auch neuartige Zugriffsverfahren möglich, die bisher nicht realisierbar waren.

Dem Anwender einer Bilddatenbank ist es nicht zumutbar, umständliche Abfragesprachen zu erlernen und diese über die Tastatur einzugeben. Deshalb wurden in den letzten Jahren etliche Verfahren entwickelt, die den Zugriff auf Bildbestände vereinfachen, z.B. die Spracheingabe von Schlüsselbegriffen oder die Suche anhand einer vom Anwender vorgegebenen Skizze, die die ungefähren Konturen des zu suchenden Motivs enthält. Aber auch diese Verfahren besitzen Nachteile:

- ⇒ Die Suche über Schlüsselwörter, sei es nun über Sprach- oder die übliche Tastatureingabe, ist immer begrenzt durch die Übereinstimmung der vom Suchenden vermuteten und dem vom Anbieter tatsächlich vorgenommenen Klassifizierung
- ⇒ Die ungefähre Eingabe der Bildkonturen anhand einer Skizze ist nur für einen eingeschränkten Anwenderkreis durchführbar, der die dazu benötigten zeichnerischen Fähigkeiten besitzt. Viele Anwender verhalten sich bei dem Versuch, das von ihnen gesuchte Bild zeichnerisch zu beschreiben, gehemmt.

Viele der herkömmlichen Verfahren zur Bildrecherche lehnen sich in ihrer Methodik an Verfahren an, die für die Dokumentenrecherche verwendet werden und auf textueller Basis, d.h. auf einer Menge von Stichworten oder Schlüsselbegriffen, arbeiten. Anders gesagt: den Verfahren fehlt die Unmittelbarkeit zum Zielmedium, denn die Interaktion/ Begriffsbildung findet über ein vermittelndes Medium, nämlich die Sprache statt. Ich möchte nun im weiteren Verlauf ein Verfahren vorstellen, dass auf diesen Umweg verzichtet und einen höheren Grad an visueller Direktheit besitzt als bisherige Verfahren. Das Verfahren stellt keine Generallösung dar, ist aber für eine grosse Klasse von Bildtypen anwendbar und lässt sich mit traditionellen Verfahren kombinieren.

3. Anforderungen

Das im folgenden vorgestellte Verfahren ist für eine spezielle Klasse von Bildern geeignet, die folgende Bedingung erfüllen:

- ⇒ Bilder – sowohl natürliche als auch synthetische - deren Motiv ein Objekt bzw. eine Menge von Objekten ist. Unter Objekt wird in diesem Zusammenhang ein konkreter Gegenstand der menschlichen Vorstellungswelt verstanden, also bspw. ein Auto, Gebäude etc. Eine abstrakte, rein grafische Gestaltung erfüllt diese Bedingung nicht.

4. Implementierung

Das Verfahren wurde prototypisch in der Programmiersprache Java umgesetzt. In den folgenden Abbildungen ist das Rechercheinterface sowie Ausschnittvergrößerungen daraus zu sehen.

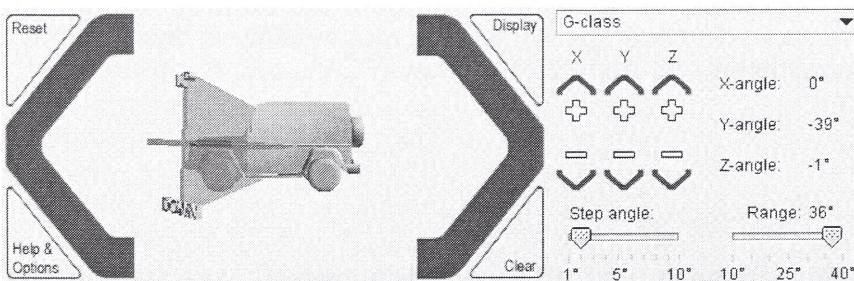


Abbildung 1: Interaktionsvisualisierung und -parametereinstellungen

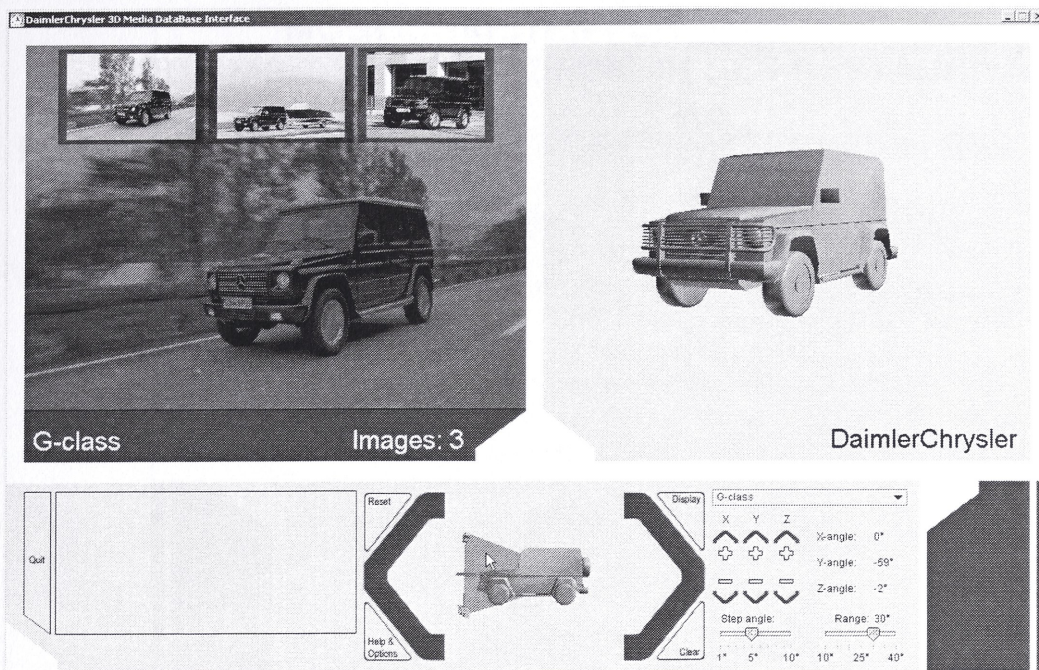


Abbildung 2: Benutzerinterface, Recherche/ Navigation

Die Interaktion läuft folgendermassen ab:

1. Der Anwender wählt aus dem Objektkatalog ein bestimmtes Objekt aus (in vorliegendem Fall wurde die Modellpalette eines Automobilherstellers als Beispiel verwendet)
2. In dem Fenster rechts oben erscheint das gerenderte 3D Modell des gewählten Objektes. In diesem Fenster kann der Anwender mit dem Modell frei interagieren. Da wegen der Abbildung der 2D Bewegungen der Maus viele Anwender Schwierigkeiten mit dieser Art der Interaktion haben, wurde die Möglichkeit der direkten Manipulation einer Drehachse um vordefinierte Gradschritte (siehe Abb. 1, rechts) hinzugefügt. Eine weitere Interaktionsmöglichkeit besteht für den Anwender darin, ein Objekt (transparentes Achsenkreuz), das die Blickrichtung/ POV veranschaulicht direkt zu manipulieren (siehe Abb. 1, links). In diesem Bildausschnitt bleibt im Gegensatz zu dem Bildfenster rechts oben das Motivobjekt fest und das POV-Objekt wird manipuliert.
3. Der Anwender wählt nun einen Blickpunkt anhand der drei realisierten Interaktionsverfahren aus und bestimmt den Grad der Ungenauigkeit, d.h. den Winkelbereich um die POV Achse herum, der für die Abfrage akzeptabel scheint (siehe Abb. 1, rechts).
4. Nach einem voreingestellten Zeitintervall wird automatisch oder auf Anwenderwunsch bei Betätigung des Knopfes „Display“ der Suchvorgang in der Datenbank gestartet. Die Ergebnisse der Suche erscheinen im Bildfenster auf der linken Seite. Der Anwender hat die Möglichkeit Bilder auszuwählen. Daraufhin erscheint das Bild in vergrößerter Darstellung und links unten werden weitere Informationen zu dem Bild dargestellt.

Eine Abfrage an das Datenbanksystem geschieht in diesem Szenario also nicht durch Eingabe eines Suchbegriffs oder einer SQL Anweisung, sondern durch direkte Interaktion mit einem abstrahierten 3D Modell, das das Bildmotiv veranschaulicht. Die Position aus der der Anwender ein 3D Modell betrachtet generiert eine entsprechende Abfrage an das Datenbanksystem.

Da zur Zeit keine automatisierten Bildanalyseverfahren zur Extraktion bzw. Kameras zur Aufzeichnung des POV existieren, ist eine manuelle Eingabe dieser Metadaten notwendig. Dies geschieht jedoch nahezu identisch zum vorher beschriebenen Rechercheablauf (siehe Abb. 3).

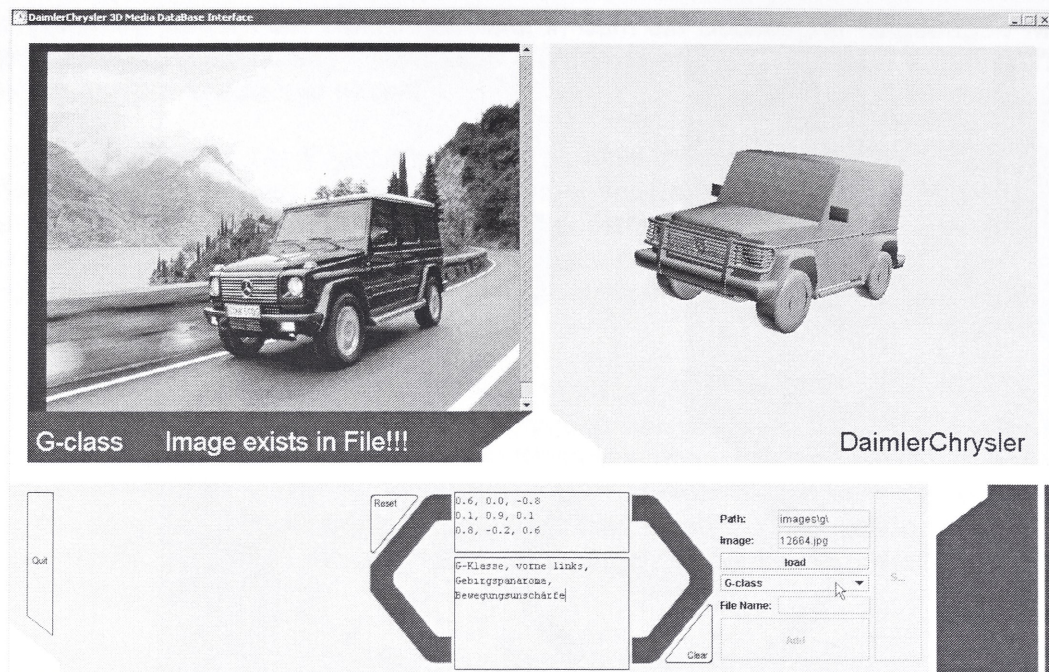


Abbildung 3: Benutzerinterface, Metadatenerfassung

Als erstes wählt der Anwender das zu erfassende Bild aus und interagiert im folgenden mit dem entsprechenden 3D Modell so lange, bis der POV mit dem im Bild dargestellten Objekt übereinstimmt. Ist das Bild noch nicht in der Datenbank vorhanden so wird es nun mit den eingegebenen Textinformationen sowie dem errechneten POV aufgenommen und steht für die Recherche zur Verfügung.

5. Ausblick

Die vorhandene prototypische Realisierung stellt einen ersten Schritt dar, der um weitere Komponenten erweitert wird. Momentan kann der Anwender das Objekt lediglich um einen festen Punkt rotieren. Weitere Möglichkeiten der Interaktion wie Verschieben, Zoomen und Bewegung im Raum werden hinzugefügt werden und erweitern somit die Recherchemöglichkeiten. Eine weitere Möglichkeit stellen rudimentäre 3D Modellierungsfunktionalitäten dar um bspw. Szenen mit mehreren Objekten definieren zu können.

Das hier vorgestellte Prinzip der Bewegung bzw. Positionierung innerhalb eines virtuellen Raumes und der damit definierte POV sowie die daraus automatisch generierte Abfrage an ein Informationssystem stellt ein universelles Prinzip dar, und lässt sich nahtlos auf weitere Informationsdomänen übertragen, z.B. Architektur, Geographie, Design, Porträt etc. und kann Verwendung finden in Museen, Presse- und Bilddatenbanken etc.

6. Literatur

[1] Wilk, Christian: **A4SM - An integrated digital movie production environment**, CBMI'99 European Workshop on Content-Based Multimedia Indexing, Toulouse 1999